



HAL
open science

Metabarcoding targeting the EF1 alpha region to assess Fusarium diversity on cereals

Anne-Laure Boutigny, Angelique Gautier, Ryan Basler, Florian Dauthieux,
Stephen Leite, Romain Valade, Jaime Aguayo, Renaud Ioos, Valerie Laval

► **To cite this version:**

Anne-Laure Boutigny, Angelique Gautier, Ryan Basler, Florian Dauthieux, Stephen Leite, et al..
Metabarcoding targeting the EF1 alpha region to assess Fusarium diversity on cereals. PLoS ONE,
2019, 14 (1), 10.1371/journal.pone.0207988 . hal-02625914

HAL Id: hal-02625914

<https://hal.inrae.fr/hal-02625914>

Submitted on 26 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

Metabarcoding targeting the EF1 alpha region to assess *Fusarium* diversity on cereals

Anne-Laure Boutigny¹✉, Angélique Gautier²✉, Ryan Basler², Florian Dauthieux², Stephen Leite², Romain Valade³, Jaime Aguayo¹, Renaud Iosif¹, Valérie Laval²*✉

1 ANSES Laboratoire de la santé des végétaux, Unité de Mycologie, Malzéville, France, **2** INRA, UMR1290 BIOGER_CPP, Thiverval-Grignon, France, **3** ARVALIS Institut du végétal, Thiverval-Grignon, France

✉ These authors contributed equally to this work.

* valerie.laval@inra.fr



Abstract

Fusarium head blight (FHB) is a major cereal disease caused by a complex of *Fusarium* species. These species vary in importance depending on climatic conditions, agronomic factors or host genotype. In addition, *Fusarium* species can release toxic secondary metabolites. These mycotoxins constitute a significant food safety concern as they have health implications in both humans and animals. The *Fusarium* species involved in FHB differ in their pathogenicity, ability to produce mycotoxins, and fungicide sensitivity. Accurate and exhaustive identification of *Fusarium* species *in planta* is therefore of great importance. In this study, using a new set of primers targeting the *EF1α* gene, the diversity of *Fusarium* species on cereals was evaluated using Illumina high-throughput sequencing. The PCR amplification parameters and bioinformatic pipeline were optimized with mock and artificially infected grain communities and further tested on 65 field samples. *Fusarium* species were retrieved from mock communities and good reproducibility between different runs or PCR cycle numbers was observed. The method enabled the detection of as few as one single *Fusarium*-infected grain in 10,000. Up to 17 different *Fusarium* species were detected in field samples of barley, durum and soft wheat harvested in France. This new set of primers enables the assessment of *Fusarium* diversity by high-throughput sequencing on cereal samples. It provides a more exhaustive picture of the *Fusarium* community than the currently used techniques based on isolation or species-specific PCR detection. This new experimental approach may be used to show changes in the composition of the *Fusarium* complex or to detect the emergence of new *Fusarium* species as far as the *EF1α* sequence of these species show a sufficient amount of polymorphism in the portion of sequence analyzed. Information on the distribution and prevalence of the different *Fusarium* species in a given geographical area, and in response to various environmental factors, is of great interest for managing the disease and predicting mycotoxin contamination risks.

OPEN ACCESS

Citation: Boutigny A-L, Gautier A, Basler R, Dauthieux F, Leite S, Valade R, et al. (2019) Metabarcoding targeting the EF1 alpha region to assess *Fusarium* diversity on cereals. PLoS ONE 14(1): e0207988. <https://doi.org/10.1371/journal.pone.0207988>

Editor: Ulrich Melcher, Oklahoma State University, UNITED STATES

Received: October 2, 2017

Accepted: November 9, 2018

Published: January 11, 2019

Copyright: © 2019 Boutigny et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information file. Sanger DNA sequences are deposited in R-syst:fungi database (http://138.102.89.206/new_rsyst_chp/). Metabarcoding data are deposited in the European Nucleotide Archive (ENA) (accession number: PRJEB27158). QPCR protocol is published in <https://doi.org/10.1101/272534>.

Funding: This work was supported by AAP PSPE 2012 (Appel à proposition Pour et Sur le Plan Ecophyto de Juin 2012) www.onema.fr.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Fusarium head blight (FHB) is a major cereal disease causing severe damage to crops worldwide including Europe. In addition to reducing grain yield and quality, the harvested grains are often contaminated with mycotoxins that are a major health and food safety concern due to their toxicity to humans and other animals. Up to 17 fungal species have been associated with FHB [1], belonging either to the toxinogenic genus *Fusarium* or to the non-toxinogenic genus *Microdochium*. For example, in France, the most common species associated with FHB on soft wheat, durum wheat and barley are *F. graminearum sensu lato*, *F. avenaceum*, *F. poae* and *M. nivale*, but species such as *F. tricinctum* and *F. culmorum* have also been found frequently, along with less frequent species such as *F. equiseti*, *F. acuminatum*, *F. sambucinum*, *F. sporotrichioides*, *F. verticillioides*, *F. heterosporum*, *F. subglutinans* and *F. oxysporum* [2, 3]. These *Fusarium* species are frequently associated with one another in the infected grains and constitute a disease complex [4]. A similar species profile was also found in other European countries, with local and temporal specificities due to several factors such as climatic conditions, agronomic factors or host genotype [5–12].

Most *Fusarium* species are able to produce one or more mycotoxins with varying degrees of toxicity [5]. *Fusarium* species involved in FHB differ in their epidemiology, pathogenicity, ability to produce mycotoxins, and fungicide sensitivity. However, in a context of strong political and public awareness on the issues of food, health and environmental safety, the current trend is to reduce the quantities of fungicides used on crops, while maintaining efficient and sustainable agriculture. Climate change and evolving cultural farming practices may also contribute to the modification of *Fusarium* species distribution and in turn, mycotoxin content. Information on the distribution and predominance of the different *Fusarium* species in a given geographical area, and in response to various environmental and cultural factors, is therefore of great importance for managing the disease and evaluating mycotoxin contamination risks.

The analysis of *Fusarium* species composition in field samples is mainly performed using morphological identification or species-specific PCR-based methods [2, 13–15]. Morphological identification of *Fusarium* species is based on the microscopic observation of various criteria (mycelia, growth, macroconidia, microconidia, etc.), and requires time and high taxonomic skills. In addition, competition between species during the isolation step can lead to an underestimation of the occurrence of slower-growing species and of morphologically indistinguishable (i.e. cryptic) species, like the *Fusarium graminearum* species complex (FGSC) occurring on cereal spikes and grains. Species-specific PCR-based methods are reliable methods to detect/quantify various *Fusarium* species in field samples [13–15], but constitute *a priori* approaches restricted to known and specific targets. Likewise, target-specific PCR tests cannot be used to search for emerging, i.e. ill-described or yet undescribed, *Fusarium* species.

Typically, identification of fungal species can be achieved by analysis of one or a combination of DNA sequences [16], but still requires preliminary isolation of the fungus in pure culture before amplification by PCR and sequencing of the sequence. The development of more efficient and more affordable high-throughput sequencing techniques now enables the generation of thousands of DNA sequences from environmental samples, without any isolation and culturing step required. Metabarcoding, i.e. the simultaneous amplification and sequencing of barcode sequences directly from environmental samples, has become a very popular and powerful technique to explore the diversity of microorganisms, including slow-competing or uncultivable taxa. Numerous studies have recently described or re-assessed fungal diversity in samples as diverse as soil, rhizosphere, phyllosphere or field samples using metabarcoding [17–24]. High-throughput sequencing methods are evolving at a fast pace, and Illumina MiSeq sequencing is currently the most successful and extensively used technology globally due to its

low error rate and the lowest cost per million bases, with a high number of reads generated. It also has a major advantage over other technologies by generating the longer DNA sequences required for barcoding.

For fungi, the internal transcribed spacer (ITS) region of rDNA has been recognized as the official barcode [16]. ITS sequences are largely represented in international databases [16, 25] and have often been used in fungal diversity studies [17, 26]. However, this barcode does not always enable species discrimination for some genera, in particular within *Fusarium* [27]. The translation elongation factor 1- α (*EF1 α*) gene, which is involved in the protein translation machinery of eukaryotic cells, has proven to be efficient to distinguish related *Fusarium* species [28, 29], and was selected as the barcode of choice for the FUSARIUM-ID online identification tool [27]. Owing to its phylogenetic utility and the fact that it is present as a single copy in *Fusarium*, this gene appears to be a good candidate for *Fusarium* metabarcoding approaches.

In this study, two PCR primers were designed to amplify an informative region of the *EF1 α* gene in the genus *Fusarium* and a bioinformatic pipeline was developed to assess overall *Fusarium* diversity in field samples by metabarcoding. Various control samples were used to validate the precision, sensitivity and robustness of the technique. As a proof of principle, analyses were performed on field samples and the results were compared with real-time PCR species-specific detection tests.

Materials and methods

Biological samples

Mock communities. Five different fungal mock communities were prepared (Table 1) to test the efficacy of the analytical process (i.e. PCR primer specificity, PCR reaction, bioinformatic pipeline) in differentiating *Fusarium* species. Mock communities 1 to 4 included various *Fusarium* species and other fungal plant pathogens from diverse taxonomic groups from the Anses, ARVALIS-Institut du végétal and INRA laboratory collections. All of the strains used in this publication were single-spored. They were characterised by morphological identification when possible and by sequencing of at least two genes: *ITS* and *EF1 α* (S1 Fig). The sequences of *ITS*, *EF1 α* and some of *RNA polymerase II* second largest subunit (*RPB2*) are publicly available on R-Syst::fungi database (http://138.102.89.206/new_rsyst_chp/, <https://github.com/r-syst/databases/tree/master/r-syst::fungi>). To search for these sequences enter internal lab id (available S1 Table column 2) in 'advance search' 'strain' and then click to view the results. Using Phylogeny.fr [30] and the following set up: MUSCLE alignment, Gblock curation, Maximum Likelihood PhyML, a phylogenetic tree was built with *EF1 α* sequences of our strains and *EF1 α* sequences from collection strains of the Fusarium ID database [27] (S2 Fig). To confirm the name, the same way an *RPB2* phylogenetic tree was obtained of some of our strains (S3 Fig). Mock community 5 contained 10 species in the FGSC previously obtained from the ARS culture collection (USDA ARS, Peoria, Illinois, USA). To prepare each mock community, 100 ng of DNA of each species were mixed in sterile water, except for mock community 5 for which 10 ng of DNA of each species were mixed.

Artificially infected grains. Infected grain communities were prepared (Table 2) to (i) test the efficacy of the analytical process to differentiate *Fusarium* species in complex samples, (ii) verify consistency between replicates and DNA extractions, and (iii) evaluate the sensitivity of the technique. Healthy wheat grains were infected *in vitro* with *Fusarium* strains from the ANSES, ARVALIS and INRA laboratory collections. Each *Fusarium* strain was grown on potato dextrose agar (PDA) plates for 7 to 10 days. Wheat grains were sterilized by autoclaving, and plated on PDA. For each *Fusarium* strain, a series of plates containing 6 to 8 wheat grains was prepared. Three plugs taken from the edge of the fungal cultures on PDA were evenly

Table 1. DNA mock communities description.

Mock 1		Mock 2		Mock 3	
Species	ID	Species	ID	Species	ID
<i>Fusarium oxysporum</i>	FOR 4	<i>Fusarium equiseti</i> (FIESC 14)	C33	<i>Fusarium culmorum</i>	MTK0055
<i>Fusarium sambucinum</i>	C39	<i>Fusarium graminearum</i>	MTK0019	<i>Fusarium poae</i>	MTK0005
<i>Fusarium sp.</i> (FSSC 11)	FJMO	<i>Fusarium sambucinum</i>	C39	<i>Fusarium pseudonygamai</i>	FoxFMO14
<i>Fusarium sp.</i>	22374	<i>Cladosporium delicatulum</i>	C58	<i>Fusarium sambucinum</i>	C39
<i>Fusarium staphyleae</i>	22316	<i>Cladosporium sphaerospermum</i>	96	<i>Fusarium sporotrichioides</i>	Fs23
<i>Alternaria alternata</i>	C61	<i>Colletotrichum orbiculare</i>	C-orbiculare	<i>Calonectria pseudonaviculata</i>	93
<i>Botrytis cinerea</i>	C87	<i>Eutypa lata</i>	0002-11	<i>Epicoccum nigrum</i>	100
<i>Helminthosporium tritici-repentis</i>	MTK0038	<i>Gaeumannomyces graminis var. tritici</i>	MTK0077	<i>Leptosphaeria maculans</i>	102
<i>Phaeoacremonium mortoniae</i>	0002-12	<i>Microdochium nivale</i>	Mn08N58	<i>Microdochium majus</i>	MTK0025
<i>Puccinia triticina</i>	B77Saba	<i>Pseudocercospora herpotrichoides</i>	MTK0049	<i>Mortierella elongata</i>	95
<i>Parastagonospora nodorum</i>	MTK0081			<i>Pseudonectria buxi</i>	94
<i>Sclerotinia sclerotiorum</i>	C62				
Mock 4		Mock 5			
Species	ID	Species	ID		
<i>Fusarium culmorum</i>	MTK0009	<i>Fusarium cortaderiae</i>	NRRL31171		
<i>Fusarium langsethiae</i>	FL39	<i>Fusarium meridionale</i>	NRRL29010		
<i>Fusarium sambucinum</i>	C39	<i>Fusarium acaciae-mearnsii</i>	NRRL26755		
<i>Fusarium subglutinans</i>	Fsub69	<i>Fusarium austroamericanum</i>	NRRL28718		
<i>Fusarium temperatum</i>	Fpro32-2	<i>Fusarium aethiopicum</i>	NRRL46722		
<i>Fusarium sp.</i> (FTSC)	FT12	<i>Fusarium brasilicum</i>	NRRL31238		
<i>Fusarium avenaceum</i>	MTK0070	<i>Fusarium vorosii</i>	NRRL38208		
<i>Fusarium verticillioides</i>	Fmoni62	<i>Fusarium gerlachii</i>	NRRL38380		
<i>Leptosphaeria biglobosa</i>	C68	<i>Fusarium mesoamericanum</i>	NRRL25797		
<i>Microdochium bolleyi</i>	98	<i>Fusarium asiaticum</i>	NRRL13818		
<i>Zymoseptoria tritici</i>	MTK0030				

This table presents the species name and identification number (ID) of the strains used to prepare the 5 mock communities. DNA of these strains was extracted and mixed in equimolar quantity after Qubit quantification.

<https://doi.org/10.1371/journal.pone.0207988.t001>

placed on each plate containing the grains, and incubated for 10 days to allow fungal colonization of the grains. After infection, grains were dried for 24 h at 50 °C.

A first set of 100 grain samples containing 10 grains infected with *F. graminearum* (LSV M808) and 1 grain infected with *F. langsethiae* (LSV M846), *F. poae* (LSV M861) and *F. tricinctum* (LSV M860) in different proportions was prepared (Table 2). Each sample was prepared with durum wheat (cultivar CG09) and soft wheat (cultivar Galopain). The samples in their entirety were ground in a grinding jar using a TissueLyser (Qiagen, France). The resulting powder (particle size not determined) was collected in a 50 mL falcon tube and mixed manually before collecting 200mg for DNA extraction.

A second set of artificially infected samples was prepared by mixing 8 grains individually infected with *F. poae* (FP1), *F. langsethiae* (FL33), *F. sp. FTSC* (FT12), *F. culmorum* (C17), *F. avenaceum* (Fa), *F. graminearum* (FG20), *F. subglutinans* (FU12238) and *F. verticillioides* (FU09041), respectively, with a total of 100, 200, 500, 1000, 2000, 5000 and 10,000 healthy grains (Table 2). Each sample was prepared with durum wheat (cultivar Miradou). The samples in their entirety were ground using a PULVERISETTE 14 Variable-Speed Rotor Mill (Fritsch, Germany). The particle size of the flour was approximately 5µm.

Table 2. Description of artificially infected grain samples prepared in this study.

Sample name	Number of artificially infected grains								Number of uninfected grains
	<i>F. graminearum</i>	<i>F. langsethiae</i>	<i>F. poae</i>	<i>F. sp. FTSC</i>	<i>F. culmorum</i>	<i>F. avenaceum</i>	<i>F. subglutinans</i>	<i>F. verticillioides</i>	
G10 P1 L1 T1	10	1	1	1					87
G10 L1 P1	10	1	1						88
G10 L1 T1	10	1		1					88
G10 P1 T1	10		1	1					88
F8 100	1	1	1	1	1	1	1	1	92
F8 200	1	1	1	1	1	1	1	1	192
F8 500	1	1	1	1	1	1	1	1	492
F8 1000	1	1	1	1	1	1	1	1	992
F8 2000	1	1	1	1	1	1	1	1	1992
F8 5000	1	1	1	1	1	1	1	1	4992
F8 10000	1	1	1	1	1	1	1	1	9992

This table describe which *Fusarium* species were used to inoculate grain and the number of infected grains and uninfected grains per samples.

<https://doi.org/10.1371/journal.pone.0207988.t002>

Field samples. In 2014 and 2015, 65 field samples of durum wheat, soft wheat and barley were collected in different locations in France. Samples were selected to represent a variety of cereal-growing regions with different climates and cultural farming practices, and following different chemical treatment strategies (S4 Fig). The samples were used to test the applicability of the protocol and analysis steps. No permits were required for the field samplings and the authors have permission to conduct the study. Each sample, corresponding to 3 kg of grains, was homogenised and ground using a hammer mill (TITAN 2000, fao). Then, 100 g of flour was ground again using an MM400 Retsch mixer mill (Retsch, Germany). The particle size of the flour was approximately 5µm.

Methods

DNA extraction. For the mock communities, fungal DNA was obtained from aerial mycelium aseptically scraped from cultures grown on either PDA or malt agar. The mycelium was lyophilized for 24h in an Edwards Modulyo 4K Lyophilizer (Edwards, United Kingdom) and 100 mg of lyophilized mycelium was placed in a Fast-Prep tube (2 mL) containing 130 mg of glass beads (4.5 mm in diameter; Dutscher, France) and ground using FastPrep (MP Biomedicals, France) for 30s at maximum shaking frequency. DNA was extracted using a DNeasy Plant Mini kit (Qiagen, France) according to the manufacturer’s instructions, except that the incubation time was extended to 1 h at 65°C, and the volumes of AP1 and P3 buffers were doubled.

DNA was extracted in the same way from both artificially infected grains and field grain samples. Three independent extractions were performed per sample for artificially infected samples with four species (Table 2), whereas only a single extraction was carried out for artificially infected samples with eight species (F8 100 to F8 10,000; Table 2) and field samples. 100 mg of the homogenized flour was re-suspended in 400 µL of AP1buffer (DNeasy Plant Mini kit, Qiagen, France) and sonicated for 30s. Then, 10 µL of proteinase K (10 mg/mL) and 10 µL of RNase A (Qiagen) were added to the suspension. After 20 minutes of incubation at 65°C, the protocol recommended by the manufacturer was followed. All DNA extracts were quantified using a Qubit 2.0 Fluorometer (Invitrogen, USA).

Primer design. A local database of partial *EF1 α* gene sequences from different *Fusarium* species was constructed, containing sequences downloaded from the Fusarium ID database [27]. In total, the database contained 162 sequences, covering 74 species. Using Phylogeny.fr [30], a phylogenetic tree was built and 6 divergent sequences were selected and aligned using MultAlin [31] with *EF1 α* sequences from other genera such as *Zymoseptoria*, *Puccinia*, *Microdochium*, *Parastagonospora*, and *Leptosphaeria* in order to find regions i) sufficiently conserved in *Fusarium* to design a genus-specific primer pair, and ii) flanking a DNA region showing a high level of interspecific polymorphism within the genus *Fusarium*. Several sets of candidate primers were designed and their specificity was first assessed *in silico* by BLAST with the NCBI database, then *in vitro* by PCR using DNA from a panel of 54 strains including 31 *Fusarium* species and 19 other genera of fungi (S5 Fig). The best primer pair was then selected: EF1-F2 (5' TCATC GGCCACGTCGACTCT3') and EF1-R3 (5' TACCAGCCTCGAAC TCACCA3') (S6 Fig).

PCR and Illumina sequencing. In order to obtain a PCR fragment suitable for Illumina MiSeq sequencing, primers EF1-F2 and EF1-R3 were tagged with specific sequences (PYROFUS-EF1-F2-MISEQ: 5' **CTTTC CCTACACGACGCTCTTCCGATCTATTAACCCTCACTAAAGGGA**TCATCGGCCACGTCGAC TCT3' and PYROFUS-EF1-R3-MISEQ 5' **GGAGTTC AGACGTGTGCTCTTCCGATCT CCCTATAGTGAGTCGTATTATACCAGCCTCGAACTCACC A3'**). The PCR reaction contained 1 x mastermix for the Type-it Microsatellite PCR Kit (Qiagen, France), 5 μ L Q-Solution and 0.25 μ M of each primer in a total volume of 50 μ L. PCR reactions were performed with either 60 ng (mock communities) or 125 ng (artificially infected grains and field samples) of DNA. PCR reactions were performed using an Applied Biosystems 9700 thermocycler. The amplification parameters consisted of an initial denaturation step at 95°C for 5 min; then either 31 or 35 cycles of denaturation at 95°C for 60 s, annealing at 65°C for 90 s, and elongation at 72°C for 60 s; and a final extension step at 72°C for 10 min. The size of the final PCR amplicon before sequencing was around 750 bp. For field samples, in order to obtain a sufficient amount of PCR product for Illumina MiSeq sequencing, 3 or 6 independent PCRs of the same sample were pooled before purification. The PCR products were purified by PEG-8000 precipitation and concentration (26.2% polyethylene glycol 8000, 6.6 mM MgCl₂ and 0.6 M NaOAc) and sent to the GenoToul platform (INRA Toulouse) for Illumina MiSeq (V3 2 x 250 bp) sequencing. Three independent sequencing runs were performed.

Species detection by real-time PCR. Total DNA from each of the 65 field samples was analyzed by species-specific real-time PCR targeting *F. avenaceum*, *F. culmorum*, *F. graminearum*, *F. langsethiae*, *F. poae*, *F. sporotrichioides* and *F. tricinctum* [32]. The artificially infected samples were also analyzed with these tests in order to verify the presence of the *Fusarium* species inoculated. The primers and probes used were custom-synthesized by Eurogentec (Seraing, Belgium) and are listed in Table 3.

Real-time PCR assays were performed using 2 x qPCR MasterMix with ROX and Uracil N Glycosylase (UNG) from Eurogentec (Angers, France). PCR reactions were performed in a total of 20 μ L consisting of 12.5 μ L of mastermix, with primers and probes at the concentrations described in Table 3, and 5 μ L of DNA extract (100 ng for field samples) were added to the mix. Standard calibration curves were constructed with DNA from a pure culture of the target species with concentrations ranging from 1 ng/ μ L to 0.001 ng/ μ L and were included in each experiment. Real-time PCR reactions were performed on an ABI PRISM 7900 Sequence Detection System (Applied Biosystems, Foster City, CA, USA) in Applied Biosystem 96-well plates. Cycling conditions included an initial denaturation step at 95°C for 10 min, followed by 40 cycles each at 95°C for 15 s then one minute in specific hybridization/polymerization conditions (Table 3). Each DNA template was analyzed in triplicate.

Table 3. Primer and probe sequences and real-time PCR amplification conditions for each *Fusarium* species (table extracted from [32]).

Target species	Name	Sequence*	Final concentration	Annealing temperature
<i>F. graminearum</i>	EF1-FCFG_F	TCGATACGCGCCTGTTACC	300nM	62°C
	EF1-FG_R	ATGAGCGCCCAGGGAATG	300nM	
	grami2-EF1_rev	AGCCCCACCGGGAAAAAATTACGACA	100nM	
<i>F. culmorum</i>	EF1-FC_F2	CGAATCGCCCTCACACG	300nM	62°C
	EF1-FC-R2	GTGATGGTGCGCGCCTAG	300nM	
	culmo2-EF1-R2	ATGAGCCCCACCAGAAAAAATTACGACAA	100nM	
<i>F. poae</i>	EF1-FP2_F	CTCGAGCGATTGCATTTCTTT	300nM	60°C
	EF1_FP2_R	GGCTTCCTATTGACAGGTGGTT	300nM	
	EF1-FP	CGGAATCGTACGTGTCAATCAGTT	100nM	
<i>F. sporotrichioides</i>	EF1-FS_F3	GGCTCATACCCCGCCG	300nM	60°C
	EF1-FS_R2	GCGCCCATGTAAATGGATG	300nM	
	EF1 FS	TGGGAAGGGCAAAGCGCCTGT	100nM	
<i>F. langsethiae</i>	EF1 FL F3	GCCGTGTCGTAATTTTTTTTGTG	300nM	62°C
	EF1 FL R3	AAATGGCTATGTGGGAAGGAAG	300nM	
	EF1_FL	GGGCTCATACCCCGCCACTCGA	300nM	
<i>F. avenaceum</i>	EF1-FA_F2	CATCTTGCTAACTCTTGACAGACCG	300nM	64°C
	EF1-FA_R3	GGTAATGAATGCGTTTCGAA	300nM	
	Ef1 FA	AGCGAGTCGTGGGAATCGATGGG	150nM	
<i>F. tricinctum</i>	EF1-FT_F2	CCACGATTCGCTCCCTCAC	300nM	62°C
	EF1-FT_R2	GGTAAGATACCCACCAGAAAAA	300nM	
	EF1 FT	AGCGGGTAATGGATGCGTTTCGAGT	150nM	

*All probes were labelled with FAM fluorophore and BHQ1 Quencher

<https://doi.org/10.1371/journal.pone.0207988.t003>

Filtering, processing and assigning a taxonomic classification to sequences. A stringent pipeline taking as a basis for the MiSeq SOP [33] was used to filter and process raw sequences utilizing MOTHUR v.1.36 [34] (S7 Fig). After FastQC analysis [35], the reverse sequences were associated with poor quality files. Around 100 bp of each read received quality scores lower than 20. We eliminated these poor-quality reads from further analyses. By contrast, forward sequences met good quality standards and were processed. First, only the sequences showing the expected primer and barcode sequences were retained. Then, sequences with low quality (shorter than 100 bp or including low quality scores) were removed. After preprocessing the sequences, multiple sequence files were merged and unique sequences represented by singletons removed. Chimeric sequences were removed with the Uchime tool [36] available in MOTHUR, using self as the reference. Uncorrected pairwise distances were determined between the remaining sequences with the Needleman alignment method [37] with match 1.0, mismatch -1.0, gapopen -2.0 and gapextend 1.0 using 1000 iterations. The distance matrix was then clustered with the average neighbor algorithm to assign sequences to operational taxonomic units (OTUs).

A custom *EF1α* gene fungal reference database with 9,196 sequences was created by extracting Fungal *EF1α* sequences from GenBank using the query: (fungi[All Fields] OR fungi [Organism] OR fusarium[Title]) AND (translation elongation factor[Title] OR tef1[Title] OR translation elongation factor-1 alpha[Title] OR elongation[Title] OR EF1-a[Title] OR tef-1 [Title]) AND (CBS*[Title] OR NRRL*[Title] OR MUCL*[Title] OR voucher*[Title] OR DAOM*[Title] OR MAFF*[Title] OR MIAE*[Title] OR IPO*[Title] OR ATCC*[Title] OR LMSA*[Title]) AND (100[SLEN]: 3000[SLEN]) NOT Fusarium sp.[Title] NOT IonTorrent

[All Fields] NOT (TEF3[Title] OR translation elongation factor 3[Title] OR tef-3[Title] OR elongation factor 3[Title] OR EF-3[Title]). An additional set of partial *EF1 α* gene sequences was generated in-house by PCR, using the EF1-F2/EF1-R3 primers and DNA from the *Fusarium* strains available in the INRA and ANSES collections. The custom *EF1 α* gene fungal reference database was called 'merge.reference.fasta' in the pipeline described in the S7 Fig. There were nine taxonomic levels associated with each fungal sequence in the following order: kingdom, phylum, subphylum, class, subclass, order, family, genus, and species, with undefined levels recorded as unclassified. The classification of taxonomic levels for sequences was manually curated from the Mycobank database [38]. The curated taxonomy file was called 'merge.taxonomy.fasta' in the pipeline described in the S7 Fig. The consensus taxonomy for representative sequences for each OTU was performed using MOTHUR's classifier [39] to assign the taxonomy of sequences from the custom *EF1 α* fungal reference database using 2% divergence. An excel database was created containing OTU number, read count per OTU per sample, representative sequence name, representative sequence, and OTU representative assignment. Then, the OTUs were carefully checked and assigned to the species taxonomic level for *Fusarium* OTUs and genus taxonomic level for non-*Fusarium* OTUs. OTUs with the same species or genus assignment were merged and were represented as a percentage of the total sequences in the sample, with a minimum acceptance confidence level. Errors in the sequence of mock communities were calculated using Sanger sequences as references (MOTHUR v.1.36; [34]). The calculated errors were then used to determine the minimum acceptance confidence level to retain OTUs (S7 Fig).

Statistical analysis. A linear regression model in R version 3.3 with Rcmdr [40] was used to evaluate the influence of PCR cycle number, run and PCR pooling on the percentage of final *Fusarium* OTUs.

To compare the diversity present in each wheat sample, only *Fusarium* species determined from OTUs as described above were considered, as the experiment was developed specifically for *Fusarium* analysis. Analyses were performed with R (R version 3.3.1 (2016-06-21)) using the 'Phyloseq' R package specifically designed for analysis of microbiological communities [41]. Using 'Phyloseq', a Heatmap of abundance of each species per sample was generated.

Results

Design of a *Fusarium*-specific PCR primer pair

The different PCR primers were evaluated on a panel of DNA extracts representing 30 *Fusarium* species and 23 non-*Fusarium* species. Primers EF1-F2 and EF1-R3 were selected due to their specificity for the *Fusarium* genus (S5 Fig), and their ability to generate a single sharp band after electrophoresis on an agarose gel (data not shown). The primers amplified a 640 bp region of the *EF1 α* gene in the *Fusarium* species, displaying a high level of interspecific polymorphism. This region was successfully amplified with DNA from the 30 *Fusarium* species tested. No PCR amplification was observed with DNA from uninfected wheat. Weak amplification signals were observed for a few non-target taxa, such as *Alternaria alternata*, *Botrytis cinerea*, *Sclerotinia sclerotiorum*, *Zymoseptoria tritici*, *Leptosphaeria maculans*, *Pseudocercospora herpotrichoides*, *Phaeoacremonium mortoniae*, *Epicoccum nigrum*, *Eutypa lata*, *Cladosporium delicatulum*, and *Colletotrichum orbiculare*. (S5 Fig).

Analysis of the DNA mock communities

The five fungal mock communities were used as positive controls in three independent sequencing experiments. In these three runs, a total of 3,491,565 raw reads were generated for the mock communities. After filtering, 81,604 (2.3% of the raw sequences) were kept

for the analysis. Mock communities were used to determine error rates. As the *EF1α* Sanger sequence of each strain included in each mock community was available, a maximum error rate generated by Illumina MiSeq sequencing was calculated for each mock community using Mothur (S7 Fig). In our experiments, the maximum error rate obtained was 0.3%. In this respect, for each sample analyzed (artificially infected grain samples and field samples), OTUs representing less than 0.3% of the total reads were removed from the data.

For each OTU, results were expressed as a percentage of sequences out of the total number of sequences analyzed per sample. The expected *EF1α* sequences from all the *Fusarium* species added in mock 1, mock 2, mock 3 and mock 4 could be identified (Fig 1). The percentage of *Fusarium* OTUs did not differ significantly ($p > 0.05$) between the different PCR cycle conditions tested (31 or 35 cycles), and between runs (3 independent runs). Two independent PCRs were also performed with DNA from the same mock community and sequenced in one run, and no difference could be observed between the percentages of *Fusarium* OTUs retrieved. Sequences from a few non-*Fusarium* species that were included in the mock communities were identified: *Phaeoacremonium* (mean 1.0%, SEM 0.1%) for mock 1, *Cladosporium* (mean 7.3%, SEM 2.9%), *Colletotrichum* (mean 10.7%, SEM 2.8%), *Eutypa* (mean 0.4%, SEM 0.3%), and *Pseudocercospora* (mean 0.4%, SEM 0.1%) for mock 2, and *Leptosphaeria* (mean 2.2%, SEM 0.2%) for mock 3, *Zymospetoria* (mean 0.2%, SEM 0.002%) for mock 4. No other species/genera were identified.

In mock community 5, which contained 10 species from the FGSC, only three species could be identified based on the sequence analyses: *F. acaciae-mearnsii* (mean 5.3%, SEM 0.1%), *F. cortaderiae* (mean 10.9%, SEM 0.1%), and *F. meridionale* (mean 5.6%, SEM 0.1%). All the other reads were assigned to *F. graminearum* (77.8%, SEM 0.1%). This correlates with the differences observed in the 200 bp sequence of the *EF1α* gene originating from Illumina Miseq sequencing, as *F. cortaderiae*, *F. meridionale* and *F. acaciae-mearnsii* have 4, 3 and 3 base pair differences, respectively. The other species of the FGSC showed fewer than 3 base pair differences in the 200 bp sequence of the *EF1α* gene used for the analysis, and could not be differentiated using the 2% sequence divergence threshold.

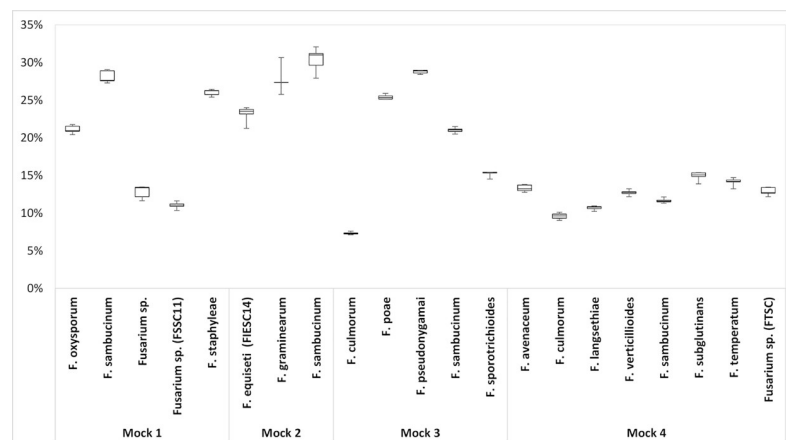


Fig 1. Boxplots of relative abundance of reads per species for mock communities 1 to 4. For each species, each Boxplot includes 8 independent samples for each mock (two PCR cycle conditions and 2 replicates for sequencing run 1, one PCR cycle condition and 3 replicates for sequencing run 2, one PCR cycle condition and 1 replicate for sequencing run 3).

<https://doi.org/10.1371/journal.pone.0207988.g001>

Analysis of the artificially infected grain samples

A total of 6,995,829 raw sequences were generated for the infected grain communities, and after filtering, 172,894 unique sequences were kept (2.4% of the raw sequences).

In samples contaminated with 4 species, the results showed that all these species could be retrieved in the mixed samples, regardless of the level and type of artificial contamination, the wheat varieties or the DNA extraction replicate (Fig 2). The results obtained comparing these 24 DNA samples (4 types of artificial contamination, 2 varieties, 3 extractions), sequenced in one run, were reproducible: *F. graminearum* (mean 89.4%, SEM 4.5%), *F. langsethiae* (mean 7.2%, SEM 2.5%), *F. poae* (mean 5%, SEM 1.6%), and *F. tricinctum* (mean 1.6%, SEM 0.5%). *F. graminearum* represented the highest percentage of reads as 10 grains infected with *F. graminearum* were mixed with only one single grain infected with the other species. *EF1α* sequences of *F. poae*, *F. tricinctum* or *F. langsethiae* could be detected in samples with only one single grain infected, mixed with 100 healthy grains. Real-time PCR analysis confirmed the presence of the 4 different *Fusarium* species in the samples (data not shown).

As the results were reproducible regardless of the wheat variety, it was decided to use only one variety to assess the sensitivity of the technique (F8-100 to F8-10,000, Table 2). In addition, as no significant effect of extraction and PCR cycle number were observed during the preliminary tests, these samples were extracted once, and tested with PCR cycling condition described

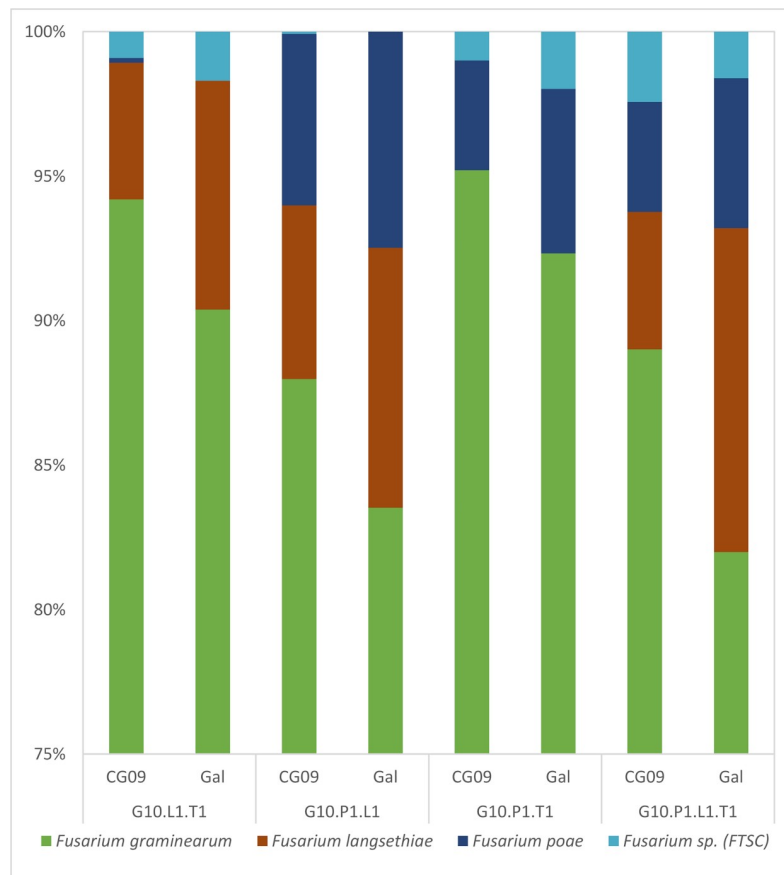


Fig 2. Percentage of sequences in *Fusarium* OTUs retrieved from artificially infected grain samples. G10 = 10 grains infected with *F. graminearum*; L1 = 1 grain infected with *F. langsethiae*; P1 = 1 grain infected with *F. poae*; T1 = 1 grain infected with *F. tricinctum*; CG09 and GAL = wheat varieties. The scale was adjusted as *F. graminearum* accounted for more than 75% of the reads for each sample.

<https://doi.org/10.1371/journal.pone.0207988.g002>

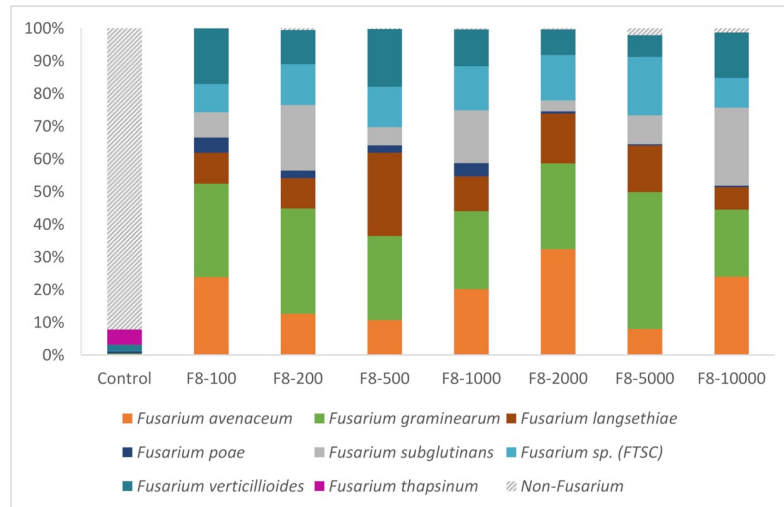


Fig 3. Percentage of sequences in *Fusarium* OTUs retrieved from artificially-infected grain samples. Eight grains infected with *F. poae*, *F. langsethiae*, *Fusarium sp. (FTSC)*, *F. culmorum*, *F. avenaceum*, *F. graminearum*, *F. subglutinans* and *F. verticillioides* were added to increased the number of uninfected grains. The same lot of uninfected grains was used as the control sample.

<https://doi.org/10.1371/journal.pone.0207988.g003>

in the materials and methods. The products of 6 independent PCRs were pooled per sample to improve sensitivity. *EF1α* sequences of *F. avenaceum*, *F. graminearum*, *F. langsethiae*, *F. poae*, *F. subglutinans*, *Fusarium sp. (FTSC)* and *F. verticillioides* were identified in all samples (99.3% ± 0.74% SEM all sequences) (Fig 3). No sequences of *F. culmorum* were observed. A two- to four-fold increase in non-*Fusarium* sequences was observed with higher levels of dilutions (F8-5000 and F8-10,000). In the control sample, 38,977 sequences were kept after filtering, of which 92.1% did not correspond to *Fusarium* OTUs. 4.6% of the reads were identified as *Fusarium thapsinum* and 1.9% as *F. verticillioides*. *F. thapsinum* was not identified in the other samples although the wheat seeds used all originated from the same sample.

Analysis of field samples

In all, 65 field samples were analyzed using the protocol developed in this study. Samples were collected from 12 fields of durum wheat, 30 fields of soft wheat, and 23 fields of barley in 2014 and 2015 (S1 Fig). A total of 5,115,242 raw sequences were generated. A total of 141,103 unique sequences remained after filtering (2.8% of the raw sequences). In order to obtain a sufficient amount of PCR amplicons before Illumina sequencing, independent PCR amplifications of the same DNA template were pooled. A first test showed that the percentage of *Fusarium* OTUs or the total percentage of non-*Fusarium* OTUs did not differ significantly ($p > 0.5$) between pooling 3 versus 6 PCR reactions before sequencing. Thus, 6 independent PCRs were performed for all samples and pooled before sequencing as this increased the amount for detection. The Heatmap (Fig 4) shows the abundance of each species detected. The abundance corresponds to the percentage of the number of reads per species out of the total number of reads in a sample. Up to 11 *Fusarium* species could be detected in wheat and up to 16 in barley. *F. graminearum* and *F. poae* were present in the majority of the samples. *Fusarium sp. (FTSC)*, *F. culmorum*, *F. langsethiae*, *F. verticillioides* and *F. avenaceum* were also detected in many samples, regardless of the type of cereal. Some *Fusarium* species were only detected in wheat, such as *F. proliferatum*, whereas by contrast, other *Fusarium* species, such as *F. thapsinum*, *F. torulosum*, *F. subglutinans*, *F. redolens*, *F. oxysporum*, and *F. mahasenii*, were only observed in barley.



Fig 4. Heatmap analysis of *Fusarium* species recovered from field samples. The colour reflects the read abundance: from fewer reads in yellow to increased number of reads in red.

<https://doi.org/10.1371/journal.pone.0207988.g004>

The DNA extracts from the 65 field samples were also analyzed by seven species-specific real-time PCR tests targeting *F. graminearum*, *F. culmorum*, *F. poae*, *F. tricinctum*, *F. langsethiae*, *F. sporotrichioides* and *F. avenaceum*, respectively (S1 Table). These real-time PCR tests were developed in 2009 using the *Fusarium* species taxonomy that was in force at that time. Since then, new cryptic species have been described, like *F. sibiricum* which is closely related to *F. sporotrichioides*, or taxa from the *F. graminearum* or *F. tricinctum* species complex. These cryptic species will therefore not be differentiated with the real-time PCR tests used here. Comparison of the real-time PCR results and the metabarcoding analyses is summarized in Table 4. Only a few discrepancies were observed. The level of discrepancy varied according to the *Fusarium* species. Similar results were obtained using the two techniques for *F. culmorum*, *F. poae* and *F. sporotrichioides* (< 8% of discrepant results), whereas a higher percentage of discrepant results was observed for *F. avenaceum*, *F. graminearum*, *F. langsethiae* and *F. tricinctum* (18.46% to 36.92%). Except for *F. tricinctum*, the metabarcoding always detected more positive samples than real-time PCR did.

This table summarised the number of samples in which *Fusarium* species were detected by qPCR or the metabarcoding techniques developed in this paper. For each species, the number of samples that were positively detected by both qPCR and the metabarcoding technique (+/+), the number samples negatively detected by both techniques (-/-) and the number of

Table 4. Discrepancy analysis results for 65 field cereal samples by real-time PCR and PCR/Illumina sequencing.

Detection qPCR / metabarcoding	<i>F. avenaceum</i>	<i>F. culmorum</i>	<i>F. graminearum</i>	<i>F. langsethiae</i>	<i>F. poae</i>	<i>F. sporotrichioides</i>	<i>F. tricinctum</i>
+/+	23	24	49	17	62	0	40
-/-	26	40	4	24	1	60	11
+/-	2	1	0	1	2	5	8
-/+	14	0	12	23	0	0	6
Discrepancies (in % of samples of non-corresponding detection out of total samples)	24.62	1.54	18.46	36.92	3.08	7.69	21.54

<https://doi.org/10.1371/journal.pone.0207988.t004>

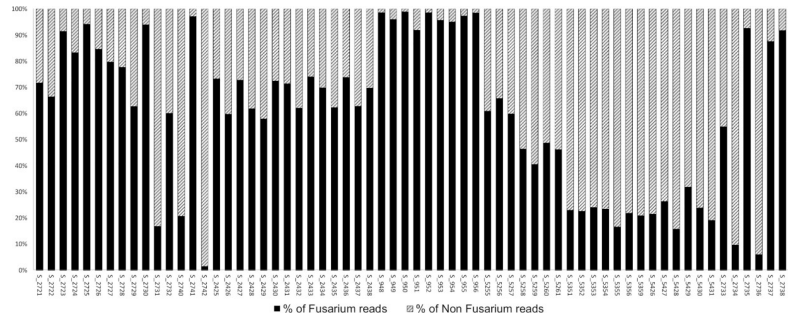


Fig 5. Percentage of *Fusarium* and non-*Fusarium* reads per sample after bioinformatic analyses. Percentage of *Fusarium* reads are presented in black and percentage of non *Fusarium* reads in hatched per field sample.

<https://doi.org/10.1371/journal.pone.0207988.g005>

samples positively detected only by qPCR (+/-) or only by metabarcoding (-/+) are presented. A percentage of discrepancies in detection is calculated at the bottom of the columns for each species.

EF1α sequences from other genera were also identified. The percentage of reads for non-*Fusarium* species varied from 5% to 95% of total reads after filtering (Fig 5). The percentage of non-*Fusarium* reads increased when contamination by *Fusarium* species was low. For example, small amounts of *Fusarium* species were detected by real-time PCR in samples S_2742, S_2734 and S_2736, in which up to 95% of non-*Fusarium* species were identified by metabarcoding. The non-*Fusarium* taxa identified in durum wheat were *Alternaria*, *Cladosporium*, *Colletotrichum*, *Ilyonectria*, and *Lewia*; whereas *Alternaria*, *Campyloctarpon*, *Chalastospora*, *Cladosporium*, *Colletotrichum*, *Diaporthe*, *Lewia*, *Trichoderma* and *Zymoseptoria* were detected in bread wheat and *Alternaria*, *Cladosporium*, *Colletotrichum*, *Gnomoniopsis*, *Lewia*, *Trichoderma* and *Zymoseptoria* were detected in barley.

Discussion

This study presents the development and validation of a metabarcoding-based protocol to assess *Fusarium* diversity on cereals. The protocol includes a method for extracting good quality DNA, followed by PCR amplification of a genus-specific, polymorphic region of the *EF1α* gene, followed by Illumina Miseq sequencing of the amplicons and analysis of the reads to identify *Fusarium* species. A new specific PCR primer pair was designed that targets the *EF1α* gene, allowing species-level discrimination, which confirms the potential of this gene previously reported [27]. In metabarcoding techniques, the initial PCR is an important step to correctly describe a community due to bias resulting from primer mismatches or PCR inhibitors [42]. As we selected conserved primers preferentially targeting the *EF1α* gene from species that belong to a single genus, *Fusarium*, primer mismatch should be avoided. An advantage of using the *EF1α* gene to study *Fusarium* communities is that it appears to be present as a single copy in this genus [27], thus circumventing problems caused by the occurrence of non-orthologous copies or a variable copy number among species, like in the case of the *ITS* gene, which is mainly used for metabarcoding [18]. Recently specific primers for the *EF1α* gene were used to assess *Fusarium* diversity in field samples from soil and wheat kernels in association with 454 pyrosequencing [24]. In a different study [22], the authors used another single-copy protein-coding locus (RPB2) to target the genus *Fusarium* associated with the rhizosphere of native grassland plants using 454 technology [22]. In our study, Illumina Miseq technology was used in order to deepen the sequencing compared to 454 technology. In addition, Illumina technology enabled the simultaneous processing of a larger number of samples. Indeed, in our

study, up to 154 field samples could be sequenced at the same time, with a mean number of 88,663 reads per sample before bioinformatic analysis.

DNA mock communities and artificially infected grain samples were used to validate our metabarcoding protocol. A sufficient number of reads could be retrieved and assigned to species in our DNA mock communities using the forward Illumina MiSeq sequences. Reverse Illumina MiSeq sequences could have given complementary information, but could not be used due to poor quality. One difficulty using high-throughput sequencing technologies is to validate OTUs with low numbers of reads, as they can be the result of PCR and/or sequencing errors [43]. DNA mock communities were used to determine an error rate, which was calculated by comparing the Sanger sequence of each strain included in each mock community with the sequence of the OTUs obtained. The maximum error rate for the 5 mock communities was estimated to be 0.3%. Thus, to ensure that correct OTUs were retained in all samples, all OTUs with a number of reads below 0.3% of the total reads per sample were removed. Rare OTUs were thus removed from all samples with a similar rule in order to avoid overestimation of the diversity present in the samples. *EF1 α* sequences from all 20 *Fusarium* species used were successfully retrieved after analysis of the DNA mock communities. In addition, 3 species out of 10 in the FGSC (*F. cortaderiae*, *F. meridionale* and *F. acaciae-mearnsii*) could also be identified using our protocol. Interestingly, even though absolute quantification is not possible due to PCR bias, all *Fusarium* species in mock communities 2 and 4 yielded approximately identical percentages of total reads. This might be due to the fact that *EF1 α* used for barcoding is a single copy gene. However, the amplification efficiency of the primers developed in this study might be different and may depend on the target species. In mock community 3, *F. culmorum* accounted for only 5% of the reads.

All *Fusarium* species, except *F. culmorum*, could be detected in the artificially infected grain samples. Several markers of the strain used for inoculation were sequenced and the sequences obtained were identical to the sequences of other *F. culmorum* strains. However, *F. culmorum* was detected in 24 field samples showing that our metabarcoding protocol is also able to specifically identify *F. culmorum* in complex samples. A close look at the results for artificially infected grain samples showed that the OTU assigned to *F. graminearum* had a score of 56. This score is a bootstrap value; the higher it is, the greater the probability that the correct name is given to an OTU. The other OTUs had a score of 75–100. After careful checking, some reads grouped in this OTU could be assigned to *F. culmorum*. The errors seemed to be linked to a PCR chimera between *F. graminearum* and *F. culmorum* not recognized by the bioinformatic pipeline. However, careful examination of the final identification scoring should prevent these errors.

The analysis of a sample using our metabarcoding approach will yield a large batch of barcode sequences. Assignment to a species name using a barcode sequence will first depend on the quality of the reference database used and second on the phylogenetic resolution of the chosen barcode. The phylogenetic resolution of the *EF1 α* barcode used in this publication allowed us to identify species commonly present in FHB disease. However some closely related (or cryptic-) species might be difficult to differentiate due to barcode resolution. For example, to the *EF1 α* region that was used here will not enable to discriminate some strains of the FGSC complex (*F. vorosii*, *F. brasilicum*, *F. gerlachii*) from *F. graminearum*. Another example are the closely related species *F. sibiricum* and *F. sporotrichioides* [44]. It is obvious that the quality of the overall assignment pipeline will depend on the quality of the reliability of the DNA sequence database used, and on the fact that the species names are regularly updated according to the state of the art regarding taxonomy in this intensively studied genus.

As the protocol proposed was aimed at assessing *Fusarium* diversity on field samples, it was important to test the reproducibility and the sensitivity of the technique. In this study,

metabarcoding proved to be reproducible using DNA mock community samples, independent of the amplification or the sequencing run. Numbers of PCR cycles have been reported as an important factor increasing contamination [45], but no such effects were observed in our experiments by testing different cycle numbers. Analysis of artificially infected wheat samples showed the high sensitivity of metabarcoding for *Fusarium*. One infected grain could be detected in a total of 10,000 grains.

As a proof of principle, further analyses were performed on 65 cereal samples harvested across the major cereal-growing regions of France. Overall, 17 different *Fusarium* species were detected in these samples of barley, durum and soft wheat. In wheat samples, *F. graminearum* and *F. poae* were the dominant species, followed by *Fusarium* sp. (FTSC), *F. culmorum* and *F. avenaceum*, which are common *Fusarium* species previously recovered from wheat in France [2, 3]. In barley samples, *F. poae* and *Fusarium* sp. (FTSC) were the dominant species, followed by *F. langsethiae*, *F. graminearum* and *F. avenaceum*. These species were previously reported from barley samples collected in France, except *F. langsethiae* [2, 3]. As *F. langsethiae* was described in 2004 [46] and it resembles *F. poae* morphologically, *F. langsethiae* was probably present on barley sampled in France in 2004 but was likely assigned to *F. poae* in a previous survey [2]. To our knowledge, this is the first time species such as *F. thapsinum*, *F. torulosum*, *F. redolens* and *F. mahasenii* have been reported on barley, and *F. proliferatum* on wheat collected in France. In a survey conducted on FGSC diversity in France in 2011, *F. graminearum sensu stricto* was predominant on wheat, barley and maize, but three isolates of *F. cortaderiae* and two isolates representing *F. graminearum* × *F. boothii* hybrids were also identified from maize [47]. The tool developed in this study should recover the *F. cortaderiae* isolates in France but not the hybrid. As the symptoms on cereal grains are not specific to the *Fusarium* species causing the disease, real-time PCR tests have been developed for *Fusarium* species detection and quantification in field samples [15]. Some of these tests are multiplexed in order to reduce the costs of analysis [48, 49]. Using our metabarcoding protocol, up to 12 *Fusarium* species could be detected in field samples and up to 150 samples could be processed at the same time. Species like *F. torulosum*, *F. sambucinum* and 3 species of the FGSC were also detected. As far as we know, no specific real-time PCR tests are available for these species. It is difficult to compare real-time PCR tests with the metabarcoding technique to detect *Fusarium* species because the efficiency of the species-specific primers used in real-time PCR tests and the length of the amplified fragment vary depending on the species. We used the real-time PCR only to verify that *Fusarium* species were present in our field samples. Discrepancies were observed between the results of the two techniques, with up to 37% non-corresponding detection observed (Table 4). In the majority of cases, discrepancies were due to detection of species by metabarcoding but not by real-time PCR. It is difficult to conclude whether these correspond to true or false positives. Due to PCR amplification biases, each technique may be favourable to the detection of one species rather than another. A possible explanation for these discrepant results may be that the limit of detection of each test (metagenomics and the different species-specific qPCR test) may be variably affected by the presence of inhibiting compounds in the DNA extracts. Since the level of contamination by some of the *Fusarium* species was sometimes very low, it is likely that some of the analyses were at the limit of detection for some of the targets.

Some of the reads generated were not assigned to species of *Fusarium*. As the identification of other genera was not included in the objectives of this study, the corresponding OTUs were only assigned to the genus level even though a species name could be determined. Importantly, when low *Fusarium* contamination was found in field samples, the amounts of non-*Fusarium* species increased. It is also interesting to note that in artificially infected samples, when *Fusarium* contamination decreased, an increase in non-*Fusarium* species was observed. This could be due to the lack of target species that decrease competition for the primer, leading to less

specific amplification. However, it was shown that reagent and laboratory contamination increased in the results, especially when there was a low amount of starting material [50]. As a result, special care should be taken when performing metabarcoding experiments.

Conclusion

Using mock communities and artificially infected samples, a protocol was developed to assess the composition of *Fusarium* species in field samples without *a priori* criteria. For the first time, the sensitivity and reproducibility of the technique proved to be suitable to detect most of the *Fusarium* species with low contamination levels (up to 1 seed in 10,000). Moreover, 3 species out of the 10 tested in the FGSC could be distinguished. The possibility of multiplexing samples makes this protocol suitable for large-scale *Fusarium* epidemiological surveys, such as those studying the impact of climate change, or of evolving cultural farming practices.

Supporting information

S1 Fig. Strain used in this study.

(PDF)

S2 Fig. Phylogenetic tree using EF1 α sequences.

(PDF)

S3 Fig. Phylogenetic tree using RPB2 sequences.

(PDF)

S4 Fig. Description of field samples.

Sampling year, variety treatment, cultural practices and location are presented for each sample.

(PDF)

S5 Fig. Test of primers' specificity.

This file summarizes the fungal species that were tested and detected by the PCR primer proposed in this publication. After PCR, the results were visualised on an agarose gel. D represent a faint band on the gel, + means detected and – means not detected.

(PDF)

S6 Fig. Position of the primers on EF1 α sequence.

(PDF)

S7 Fig. Description of the pipeline for Miseq Illumina reads analysis.

(PDF)

S1 Table. Real-time PCR results of fields samples analysis.

(XLSX)

Acknowledgments

This work was part of the PSPE 2018 program (*Pour et Sur le Plan Ecophyto 2018*). We would like to thank the GeT Genotoul platform in Toulouse, Midi-Pyrénées for the MiSeq sequencing experiments. We are grateful to the Genotoul Bioinformatics platform in Toulouse, Midi-Pyrénées for providing help computing and data storage.

Author Contributions

Conceptualization: Anne-Laure Boutigny, Angélique Gautier, Romain Valade, Renaud Ioo, Valérie Laval.

Data curation: Angélique Gautier, Ryan Basler.

Formal analysis: Angélique Gautier, Ryan Basler.

Funding acquisition: Anne-Laure Boutigny, Romain Valade, Renaud Ioo, Valérie Laval.

Investigation: Anne-Laure Boutigny, Angélique Gautier, Ryan Basler, Florian Dauthieux, Stephen Leite.

Methodology: Anne-Laure Boutigny, Angélique Gautier, Ryan Basler, Florian Dauthieux, Romain Valade, Jaime Aguayo, Renaud Ioo, Valérie Laval.

Project administration: Anne-Laure Boutigny.

Resources: Anne-Laure Boutigny, Angélique Gautier, Romain Valade, Renaud Ioo, Valérie Laval.

Software: Ryan Basler.

Supervision: Valérie Laval.

Validation: Anne-Laure Boutigny, Angélique Gautier, Florian Dauthieux, Stephen Leite, Romain Valade, Renaud Ioo, Valérie Laval.

Visualization: Anne-Laure Boutigny, Angélique Gautier, Ryan Basler, Valérie Laval.

Writing – original draft: Anne-Laure Boutigny, Romain Valade, Renaud Ioo, Valérie Laval.

Writing – review & editing: Anne-Laure Boutigny, Angélique Gautier, Ryan Basler, Romain Valade, Jaime Aguayo, Renaud Ioo, Valérie Laval.

References

1. Parry DW, Jenkinson P, McLeod L. *Fusarium* ear blight (scab) in small grain cereals—a review. *Plant Pathol.* 1995; 44: 207–238.
2. Ioo R, Belhadj A, Menez M. Occurrence and distribution of *Microdochium nivale* and *Fusarium* species isolated from barley, durum and soft wheat grains in France from 2000 to 2002. *Mycopathologia.* 2004; 158: 351–362. <https://doi.org/10.1007/s11046-004-2228-3> PMID: 15645174
3. Ioo R, Belhadj A, Menez M, Faure A. The effects of fungicides on *Fusarium* spp. and *Microdochium nivale* and their associated trichothecene mycotoxins in French naturally-infected cereal grains. *Crop Prot.* 2005; 24: 894–902.
4. Bottalico A. *Fusarium* diseases of cereals: species complex and related mycotoxin profiles in Europe. *J Plant Pathol.* 1998; 80: 85–103.
5. Bottalico A, Perrone G. Toxigenic *Fusarium* species and mycotoxins associated with head blight in small-grain cereals in Europe. *Eur J Plant Pathol.* 2002; 108: 611–624.
6. Waalwijk C, Kastelein P, de Vries I, Kerényi Z, Van der Lee T, Hesselink T, et al. Major changes in *Fusarium* spp. in wheat in The Netherlands. *Eur J Plant Pathol.* 2003; 109: 743–754.
7. Xu XM, Parry DW, Nicholson P, Thomsett MA, Simpson D, Edwards SG, et al. Predominance and association of pathogenic fungi causing *Fusarium* ear blight in wheat in four European countries. *Eur J Plant Pathol.* 2005; 112: 143–154.
8. Xu XM, Nicholson P, Thomsett MA, Simpson D, Cooke BM, Doohan FM, et al. Relationship between the fungal complex causing *Fusarium* head blight of wheat and environmental conditions. *Phytopathology.* 2008; 98: 69–78. <https://doi.org/10.1094/PHYTO-98-1-0069> PMID: 18943240
9. Audenaert K, Van Broeck R, Bekaert B, De Witte F, Heremans B, Messens K, et al. *Fusarium* head blight (FHB) in Flanders: population diversity, inter-species associations and DON contamination in commercial winter wheat varieties. *Eur J Plant Pathol.* 2009; 125: 445–458.
10. Chandelier A, Nimal C, André F, Planchon V, Oger R. *Fusarium* species and DON contamination associated with head blight in winter wheat over a 7-year period (2003–2009) in Belgium. *J Plant Pathol.* 2011; 130: 403–414.

11. Nielsen LK, Jensen JD, Nielsen GC, Jensen JE, Spliid NH, Thomsen IK, et al. *Fusarium* head blight of cereals in Denmark: species complex and related mycotoxins. *Phytopathology*. 2011; 101: 960–969. <https://doi.org/10.1094/PHYTO-07-10-0188> PMID: 21323468
12. Hietaniemi V, Rämö S, Yli-Mattila T, Jestoi M, Peltonen S, Kartio M, et al. Updated survey of the *Fusarium* species and toxins in Finnish cereal grains. *Food Addit Contam*. 2016; 33: 831–848.
13. Doohan FM, Parry DW, Jenkinson P, Nicholson P. The use of species-specific PCR-based assays to analyse *Fusarium* ear blight of wheat. *Plant Pathol*. 1998; 47: 197–205.
14. Waalwijk C, Van der Heide R, de Vries I, Van der Lee T, Schoen C, Costrel-de-Corainville G, et al. Quantitative detection of *Fusarium* species in wheat using TaqMan. *Eur J Plant Pathol*. 2004; 110: 481–494.
15. Nicolaisen M, Suproniene S, Nielsen LK, Lazzaro I, Spliid NH, Justesen AF. Real-time PCR for quantification of eleven individual *Fusarium* species in cereals. *J Microbiol Methods*. 2009; 76: 234–40. <https://doi.org/10.1016/j.mimet.2008.10.016> PMID: 19047000
16. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci USA*. 2012; 109: 6241–6246. <https://doi.org/10.1073/pnas.1117018109> PMID: 22454494
17. Buée M, Reich M, Murat C, Morin E, Nilsson RH, Uroz S, et al. 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytol*. 2009; 184: 449–456. <https://doi.org/10.1111/j.1469-8137.2009.03003.x> PMID: 19703112
18. Amend AS, Seifert KA, Bruns TD. Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Mol Ecol*. 2010; 19: 5555–5565. <https://doi.org/10.1111/j.1365-294X.2010.04898.x> PMID: 21050295
19. Jumpponen A, Jones KL. Seasonally dynamic fungal communities in the *Quercus macrocarpa* phyllosphere differ between urban and nonurban environments. *New Phytol*. 2010; 186: 496–513. <https://doi.org/10.1111/j.1469-8137.2010.03197.x> PMID: 20180911
20. Liu X, Zhang J, Gu T, Zhang W, Shen Q, Yin S, et al. Microbial community diversities and taxa abundances in soils along a seven-year gradient of potato monoculture using high throughput pyrosequencing approach. *PLoS ONE*. 2014; 9: e86610. <https://doi.org/10.1371/journal.pone.0086610> PMID: 24497959
21. Arfi Y, Buée M, Marchand C, Levasseur A, Record E. Multiple markers pyrosequencing reveals highly diverse and host-specific fungal communities on the mangrove trees *Avicennia marina* and *Rhizophora stylosa*. *FEMS Microbiol Ecol*. 2012; 79: 433–444. <https://doi.org/10.1111/j.1574-6941.2011.01236.x> PMID: 22092266
22. LeBlanc N, Kinkel LL, Kistler HC. Soil fungal communities respond to grassland plant community richness and soil edaphic. *Microb Ecol*. 2015; 70: 188–195. <https://doi.org/10.1007/s00248-014-0531-1> PMID: 25399511
23. Sapkota R, Knorr K, Nistrup Jørgensen L, O'Hanlon KA, Nicolaisen M. Host genotype is an important determinant of the cereal phyllosphere mycobiome. *New Phytol*. 2015; 207: 1134–1144. <https://doi.org/10.1111/nph.13418> PMID: 25898906
24. Karlsson I, Edel-Hermann V, Gautheron N, Durling MB, Kolseth A-K, Steinberg C, et al. Genus-specific primers for study of *Fusarium* communities in field samples. *Appl Environ Microbiol*. 2016; 82: 491–501. <https://doi.org/10.1128/AEM.02748-15> PMID: 26519387
25. Nilsson RH, Abarenkov K, Veldre V, Nylinder S, De Wit P, Brosché S, et al. An open source chimera checker for the fungal ITS region. *Mol Ecol Resour*. 2010; 10: 1076–1081. <https://doi.org/10.1111/j.1755-0998.2010.02850.x> PMID: 21565119
26. Nilsson RH, Ryberg M, Abarenkov K, Sjökvist E, Kristiansson E. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *FEMS Microbiol Lett* 2009; 296: 97–101. <https://doi.org/10.1111/j.1574-6968.2009.01618.x> PMID: 19459974
27. Geiser DM, Gasco J, Kang MM, Mkalowska S, Veeraraghavan I, Ward N, et al. FUSARIUM-IDv.1.0: a DNA sequence database for identifying *Fusarium*. *Eur J Plant Pathol*. 2004; 110: 473–479.
28. O'Donnell K, Kistler HC, Tacke BK, Casper HH. Gene genealogies reveal global phylogeographic structure and reproductive isolation among lineages of *Fusarium graminearum*, the fungus causing wheat scab. *Proc Natl Acad Sci USA*. 2000; 97: 7905–7910. <https://doi.org/10.1073/pnas.130193297> PMID: 10869425
29. O'Donnell K, Ward TJ, Aberra D, Kistler HC, Aoki T, Orwig N, et al. Multilocus genotyping and molecular phylogenetics resolve a novel head blight pathogen within the *Fusarium graminearum* species complex from Ethiopia. *Fungal Genet Biol*. 2008; 45: 1514–1522. <https://doi.org/10.1016/j.fgb.2008.09.002> PMID: 18824240

30. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 2008; 36: 465–469.
31. Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 1988; 16: 10881–10890. PMID: [2849754](https://pubmed.ncbi.nlm.nih.gov/2849754/)
32. Elbelt S, Siou D, Gelisse S, Cruaud C, Lannou C, Lebrun MH, et al. Optimized real-time qPCR assays for detecting and quantifying the *Fusarium* and *Microdochium* species responsible for wheat head blight, as defined by MIQE guidelines. *BioRx*: <https://doi.org/10.1101/272534>
33. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol.* 2013; 79: 5112–5120. <https://doi.org/10.1128/AEM.01043-13> PMID: [23793624](https://pubmed.ncbi.nlm.nih.gov/23793624/)
34. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009; 75: 7537–7541. <https://doi.org/10.1128/AEM.01541-09> PMID: [19801464](https://pubmed.ncbi.nlm.nih.gov/19801464/)
35. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
36. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics.* 2011; 27: 2194–2200. <https://doi.org/10.1093/bioinformatics/btr381> PMID: [21700674](https://pubmed.ncbi.nlm.nih.gov/21700674/)
37. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol.* 1970; 48: 443–453. PMID: [5420325](https://pubmed.ncbi.nlm.nih.gov/5420325/)
38. Crous PW, Gams W, Stalpers JA, Robert V, Stegehuis G. MycoBank: an online initiative to launch mycology into the 21st century. *Stud Mycol.* 2004; 50: 19–22.
39. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 2007; 73: 5261–5267. <https://doi.org/10.1128/AEM.00062-07> PMID: [17586664](https://pubmed.ncbi.nlm.nih.gov/17586664/)
40. Fox J. The R Commander: A Basic Statistics Graphical User Interface to R. *J Stat Softw.* 2005; 14: 1–42.
41. McMurdie PJ, Holmes S. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE.* 2013; 8: e61217. <https://doi.org/10.1371/journal.pone.0061217> PMID: [23630581](https://pubmed.ncbi.nlm.nih.gov/23630581/)
42. Tedersoo L, Anslan S, Bahram M, Põlme S, Riit T, Liiv I, et al. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys.* 2015; 10: 1–43.
43. Taylor DL, Walters WA, Lennon NJ, Bochicchio, Krohn AJ, Caporaso JG, et al. Accurate estimation of fungal diversity and abundance through improved lineage-specific primers optimized for Illumina amplicon sequencing. *Appl Environ Microbiol.* 2016; 82: 7217–7226. <https://doi.org/10.1128/AEM.02576-16> PMID: [27736792](https://pubmed.ncbi.nlm.nih.gov/27736792/)
44. Yli-Mattila T, Ward TJ, O'Donnell K, Proctor RH, Burkin AA, Kononenko GP, et al. *Fusarium sibiricum* sp. nov, a novel type A trichothecene-producing *Fusarium* from northern Asia closely related to *F. sporotrichioides* and *F. langsethiae* *Int J Food Microbiol.* 2011; 147: 58–68. <https://doi.org/10.1016/j.ijfoodmicro.2011.03.007> PMID: [21459470](https://pubmed.ncbi.nlm.nih.gov/21459470/)
45. Polz MF, Cavanaugh CM. Bias in Template-to-Product Ratios in Multitemplate PCR. *Appl Environ Microbiol.* 1998; 64: 3724–3730. PMID: [9758791](https://pubmed.ncbi.nlm.nih.gov/9758791/)
46. Torp M, Nirenberg HI. *Fusarium langsethiae* sp. nov. on cereals in Europe. *Int J Food Microb.* 2004; 95: 247–256.
47. Boutigny AL, Ward TJ, Ballois N, Iancu G, Ios R. Diversity of the *Fusarium graminearum* species complex on French cereals. *Eur J Plant Pathol.* 2014; 138: 133–148.
48. Yli-Mattila T, Paavananen-Huhtala S, Jestoi M, Parikka P, Hietaniemi V, Gagkaeva T, et al. Real-time PCR detection and quantification of *Fusarium poae*, *F. graminearum*, *F. sporotrichioides* and *F. langsethiae* in cereal grains in Finland and Russia. *Arch Phytopathology Plant Protect.* 2008; 41: 243–260.
49. Bluhm BH, Cousin MA, Woloshuk CP. Multiplex real-time PCR detection of fumonisin-producing and trichothecene-producing groups of *Fusarium* species. *J Food Protect.* 2004; 67: 536–543.
50. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 2014; 12: 87–99. <https://doi.org/10.1186/s12915-014-0087-z> PMID: [25387460](https://pubmed.ncbi.nlm.nih.gov/25387460/)