

*Highlights (for review)

- First time high-throughput sequencing is used as a detection method for EPN species.
- All EPN species detected by qPCR were also detected by HTS approach.
- High degree of correlation between HTS and qPCR species relative abundances.
- HTS has potential for use in studies of EPN population ecology.
- HTS suggests a cost-effective and accurate method assessing soil food webs.

**Comparing High Throughput Sequencing and Real Time qPCR for Characterizing
Entomopathogenic Nematode Biogeography**

**Alexandros Dritsoulas¹, Raquel Campos-Herrera², Rubén Blanco-Pérez² & Larry W.
Duncan¹**

¹Citrus Research and Education Center (CREC), Institute of Food and Agriculture Sciences
(IFAS), University of Florida (UF), 700 Experiment Station Road, FL 33850, USA.

²Instituto de Ciencias de la Vid y del Vino (CSIC-Universidad de La Rioja-Gobierno de La
Rioja), Finca La Grajera, Ctra. de Burgos Km. 6, 26007, Logroño, Spain.

Corresponding author

E-mail address: alexdrirts@ufl.edu (Alexandros Dritsoulas)

Phone: (352) 226 6355

ABSTRACT

Entomopathogenic nematodes (EPNs) are widely distributed in soils across all continents except Antarctica. Assessing the EPN community structure in an ecoregion can help reveal their biological control potential against important crop pests. Common methods for detecting EPNs in soil samples include baiting with sentinel insects, direct observation of extracted nematodes, or use of species-specific primer-probe combinations using qPCR. Less well studied is the use of high throughput sequencing (HTS), which has tremendous potential to characterize soil communities of EPNs and natural enemies of EPNs. Here, for the first time, we compared qPCR and HTS to characterize EPN food webs. The frequency and abundance of 10 EPN species and 13 organisms associated with EPNs from 50 orchard and natural area sites in two ecoregions of Portugal were evaluated using qPCR tools, and results were published in 2019. We applied an HTS approach to analyze frozen DNA samples from 36 sites in that study. Universal primers targeting ITS1 were used for nematode detection. All EPN species detected by qPCR were also detected by HTS. The EPN species and nearly all free-living nematodes detected by both processes were highly correlated ($P < 0.01$). *Steinernema feltiae*, the dominant EPN species, was detected by HTS in 55% more sites than by qPCR. HTS also detected more EPN species than did qPCR. Sample accuracy, measured by the fit of Taylor's Power Law to data from each method, was significantly better using HTS ($r^2=0.95$, $P < 0.01$) than qPCR ($r^2=0.76$, $P < 0.01$). The effect of biotic and abiotic variables on individual EPN species did not differ according to ANOVA and multiple regression analyses of both data sets while the drivers of EPN community structure did not differ when analyzing either data set with CCA. Our results combined with decreasing costs of metabarcoding, suggest that HTS may provide the most cost-effective and accurate means of assessing soil food webs of methods currently available.

KEYWORDS

Entomopathogenic nematodes, soil food webs, qPCR, metabarcoding, species detection, ITS1

1. INTRODUCTION

Entomopathogenic nematodes (EPN) in the genera *Steinernema* and *Heterorhabditis*, have been the subject of extensive research for more than a half century, due to their potential as biocontrol agents of many pest insects. Much of this work is oriented toward utilizing EPNs in either an inoculative (Parkman et al., 1993; Shields et al., 2009) or an inundate release strategy as a biopesticide (Duncan et al., 1996; Shapiro-Ilan et al., 2015). Development of mass production technology and easy-to-use formulations led to the expanded use of EPN and modest commercial successes in some markets (Dolinski et al., 2012; Georgis et al., 2006).

Each EPN species is symbiotically associated with a specific entomopathogenic bacterial species. These nematode-bacteria complexes have insecticidal effect against a broad range of insect hosts (Kaya and Gaugler, 1993). The nematodes infect the insects through body openings or by penetrating the cuticle, then release the symbionts from the nematode intestine. Insects are killed in a few days by septicemia, after which nematodes and bacteria reproduce within the cadaver. Similar to many nematode species, in response to harsh conditions such as overpopulation and resource depletion, EPN development arrests at a modified third-stage juvenile, a stress-resistant stage called “dauer” or “infective juvenile” (IJ). The IJ is the only free-living stage capable of exiting the cadaver in search of new hosts.

Use of EPNs in IPM programs requires proper fit of the nematode species to the cropping system and target pests (Shapiro-Ilan et al., 2006), preservation of genetic variability and properties such as persistence and virulence of the populations in cultures (Neumann and

Shields, 2011), and use of cultural practices beneficial to EPN functioning in the soil food web (Stuart et al., 2008). However, due to the cryptic nature of soil communities, relatively little is known about the magnitude of biological control provided by naturally occurring EPNs, or methods to exploit these services. Recent surveys of native EPNs and their natural enemies, combined with field experiments, identified some physical (Campos-Herrera et al., 2014, 2013b; El-Borai et al., 2016) and biotic (Campos-Herrera et al., 2019a) soil properties that potentially modulate the behavior of EPN populations and their contribution to pest suppression (Campos-Herrera et al., 2019b, 2014, 2013b; El-Borai et al., 2016).

Correct identification of EPN species is critical to understand observations made in ecological studies; however, morphological diagnostic methods are laborious and require taxonomic expertise. Indeed, IJ EPNs generally do not have adequate morphological characters for absolute species identification. Sentinel insects are typically employed to recover EPNs from soil samples, but infection is dependent on the susceptibility of the insect to particular EPN species, as well as environmental conditions such as soil moisture, temperature and porosity (Stuart et al., 2006). Real time, quantitative PCR has proven useful for studying EPN community structure directly from mass samples of nematodes extracted from soil (Campos-Herrera et al., 2015, 2013a; Spiridonov et al., 2007). The main principles of the method are the design of species-specific primers/probe for each species, and development of standard curves from pure cultures for quantification (Braun-Kiewnick and Kiewnick, 2018; Torr et al., 2007).

Technological developments continue to shift the predominant approaches to species identification for soil communities. The rapidly decreasing cost of gene amplicon sequencing in high-throughput (HTS) or next-generation sequencing (NGS) has numerous applications in soil and nematode community analysis. HTS of nematode communities has the potential to provide

increased taxonomic resolution and capture rare taxa that are missed using qPCR, or misidentified through morphological analysis (Treonis et al., 2018). Nevertheless, diversity assessment with species-level resolution remains an unresolved aspect of HTS. Potential errors and artefacts can arise at any step of the process: DNA amplification is limited by primers design, amplification efficiency can be species-specific, and formation of chimeric molecules can occur, especially when data include large numbers of unknown sequences (Porazinska et al., 2009). The resolution of different marker genes (Blaxter, 2003), clustering differences between bioinformatic pipelines (molecular operational taxonomic units “MOTUs” or amplicon sequence variants “ASVs”) and the availability of a high-quality reference database for species level taxonomic identification are additional challenges (Blaxter, 2003; Callahan et al., 2017; Qing et al., 2019).

The objective of this study was to compare two molecular approaches to detect EPN species: the qPCR and HTS using ITS1 sequences. We applied a HTS, metabarcoding approach using Illumina MiSeq sequencing platform to analyze frozen DNA samples from a previous survey (Campos-Herrera et al., 2019a). In that survey, the frequency and abundance of 10 EPN species and 13 organisms associated with EPNs from 50 orchard and natural area sites distributed between two ecoregions in Algarve (Portugal) were evaluated using qPCR tools. We tested the hypotheses that: (1) HTS can quantify soil organisms with comparable accuracy to that of qPCR and (2) the species detection threshold is lower for HTS than that for qPCR.

2. MATERIALS AND METHOD

2.1. Samples collection

Soil survey methods and qPCR survey results were given by Campos-Herrera et al. (2019). Briefly, 100 soil samples were recovered from 50 sites in the Portuguese Algarve region, comprising citrus, pine, palmetto and oak as the dominant plant species. Nematodes were extracted with sucrose centrifugation and then DNA was extracted using the PowerSoil[®] DNA Isolation Kit (MoBio) for that survey purposes. The same DNA samples used by Campos-Herrera et al. (2019) were shipped on dry ice to the University of Florida facilities where they arrived in chilled condition. Not all the frozen samples we received had enough DNA for the initial step of Library preparation and subsequently, only 56 samples from 36 sites with adequate total amount of extracted DNA from both or at least one of the two samples were processed

2.2. Library preparation

For HTS purposes and the proper identification of target organism, ribosomal DNA and ITS region were amplified (average length >730bp) from bulk DNA extractions using universal primers TW81 (5'- GTTTCGCTAGGTGAACCTGC-3') as forward primer and AB28 (5'- ATATGCTTAAGTTCAGCGGGT-3') as reverse primer (Iqbal et al., 2016). Primers were modified to include an overhang adapter sequence to enable sequencing, following the Illumina protocol for the 16S rRNA gene sequencing in microbial samples (16S Library Preparation Protocol at <http://support.illumina.com>).

Library preparation consist of four parts: (i) amplicon PCR, (ii) amplicon PCR cleanup, (iii) index PCR, and (iv) index PCR cleanup. Primarily, all samples normalized in 5 ng/ml DNA concentration and amplicon PCR consisted of initial denaturation 95°C for 3 min, twenty-five cycles of denaturation at 98°C for 30s, annealing at 55°C for 30s, elongation at 72°C for 60s, and

terminal elongation at 72°C for 10min. A single 25 µL PCR reaction contained 2.5 µL of template of 5 ng/µL (12.5 ng total), 12.5 µL of 2x KAPA HiFi HotStart ReadyMix (KAPA biosystems), 1µL of each 10 µM overhang primer, 8 µL of 10 mM Tris pH 8.5. Positive controls consisting of DNA extracted from a laboratory culture of the nematodes *Steinernema feltiae* and *Heterorhabditis bacteriophora* was also amplified, and negative controls consisting of purified, nuclease-free water were included for each PCR reaction. PCR products were verified on 0.8% agarose gels after staining with SYBR™ Safe DNA Gel Stain. PCR products were purified with 1.0× Agencourt AMPure XP beads (Beckman Coulter, Brea, CA) incubated for 5 minutes at room temperature, washed twice with 80% ethanol, and eluted in 50µL of 10 mM Tris pH 8.5. A second PCR added a specific index sequence to the amplicons for sample discrimination. Amplicons were used as template for a limited cycle PCR amplification to add dual-index barcodes and the P5 and P7 Illumina sequencing adapters (Nextera XT Index Kit [FC- 131- 1001]; Illumina, San Diego, CA, USA). The conditions were initial denaturation at 95°C for 3 min, 8 cycles of denaturation at 98°C for 30s, annealing at 55°C for 30s, and elongation at 72°C for 30 sec and a terminal elongation at 72°C for 10 min. Each 50 µL PCR reaction tube contained 5 µL of template, 25 µL of 2x KAPA HiFi HotStart ReadyMix (KAPA biosystems), 5 µL of Index Primers (N7XX), 5 µL of Index 2 Primers (S5XX). Finally, for the second clean up, PCR products were purified with 1.1× magnetic beads, eluted in 25 µL of 10 mM Tris pH 8.5. After library preparation, each individual library was quantified using Qubit 3.0 fluorometer the dsDNA BR kit (Life Technologies) and according to the average library size, the libraries were normalized in equal molar concentrations of 4nM and pooled together in a single library in aliquots of 5µL. The library was sequenced using 2 × 300 bp paired-end Illumina sequencing on the MiSeq platform at the Interdisciplinary Center for Biotechnology Research (ICBR) of University of Florida.

2.3. Bioinformatics

The data we received from the sequencing facility was already demultiplexed with the Illumina adapters trimmed and data separated by barcodes into respective sample identification codes. Initial quality assessment of each read was checked FASTQC v0.11. (Andrews et al., 2015), and all the quality information for individual read assessed were combined into a single viewable document using MULTIQC (Ewels et al., 2016). Due the average size of the target locus that was more 730bp, the subsequent reads did not meet the set merging criteria. We therefore used only the forward reads R1 for the bioinformatic pipeline. R1 reads derive from ITS 1 of ribosomal DNA which was used for the nematode identification. The resulting data set was de-replicated with the ASV-based approach, in which we used the DADA2 method for denoising, through QIIME2 v2019.4 pipeline, including removal of primer sequences, truncating sequences by length and removing chimeric sequences with a *de novo* approach too (Callahan et al., 2016), which resulted in a length of 250bp . We then generated count tables by mapping raw ASVs, assigning taxonomy generating input files for taxonomy assignment in QIIME2 from the NCBI database. The standalone database was generated including all the non-redundant nucleotide sequences from all traditional divisions of GenBank, EMBL, and DDBJ excluding GSS,STS, PAT, EST, HTG, and WGS (<ftp://ftp.ncbi.nlm.nih.gov/blast/db> ; nt.*tar.gz) employing an NCBI command-line tool to run BLAST, called BLAST+, integrated directly into our workflow.

2.4. Statistical Analysis

Campos-Herrera *et al.* (2019a) analyzed the means of two composite samples at each site with an area of ~0.5 ha. In the current study, samples from the 36 sites with adequate total amount of extracted DNA were analyzed differently according to whether one or two samples were available. The means were calculated for each site if both samples were available (20 sites), whereas information from a single sample was used for sites with just one sample (16 sites). The number of EPN copies measured in 12.5 ng DNA was adjusted based on the total amount of extracted DNA and then transformed to $\log(x + 1)$ for statistical analyses. Nematophagous fungi (NF) data and bacterial data reported by Campos-Herrera *et al.* (2019a) were transformed to square root (x) and $\log(x+1)$, respectively. One-way ANOVA and Tukey's HSD test were used to assess differences in the population densities of the most abundant EPN species found in the different types of vegetation. Taylor's Power Law was fitted to data for the most abundant nematodes by regressing the log-transformed sample variances against log-transformed means to assess sample measurement reliability for the two data sets (qPCR and HTS) using the 20 sites with adequate DNA from both samples for HTS (Duncan *et al.*, 2001). Nonparametric Spearman's Correlations were calculated between abundant nematode species measured by each method (JMP[®] Pro, v14.1.0.; SAS Institute Inc., Cary, NC). Stepwise multiple regression (backward elimination; $\alpha = 0.15$ for removal) of nematode species against selected soil organisms and abiotic properties from (Campos-Herrera *et al.*, 2019a) was performed using data from each method (Minitab, v. 17.3.1; State College, PA). Canonical correlation analysis was used to identify and measure the associations among explanatory and response variables setting orthogonal linear combinations of the variables within each set that best explain the variability. First, we used Pearson correlations (R) to selected soil properties to avoid variables with strong collinearity. Selected abiotic factors were included as explanatory variables or predictors. Tests of dimensionality for the canonical

correlation analysis, was employed to indicate the canonical dimensions that were statistically significant at the 0.05 level. The graphical results of the CCA were presented with bi-plot scaling (R Development Core Team, ‘Vegan’ package).

2.5. Phylogenetic tree.

For phylogenetic analysis, the newly obtained ITS1 sequences were aligned using the CLUSTAL W multiple alignment program (Thompson et al., 1994). Maximum likelihood (ML) analysis was performed with the program PhyML (Guindon and Gascuel, 2003) provided on the “phylogeny.fr” website (<http://www.phylogeny.fr/>). The probability of inferred branch was assessed by the approximate likelihood ratio test (aLRT), an alternative to the nonparametric bootstrap estimation of branch support (Anisimova and Gascuel, 2006). *Steinernema citrae* (Steinernematidae; DDBJ/EMBL/GenBank accession no. MF536116.1) was used as an outgroup for the construction of a phylogenetic tree.

3. RESULTS

Metabarcoding analysis after removing chimeras recovered 23578 unique amplicon sequence variants (ASVs) from twenty-one eukaryotic phyla, with 10.6% (2212 ASVs) assigned to the Nematoda. Unique nematode ASVs were found, encompassing 26 nematode families while more than half (1177) were identified to genus level or below, setting a threshold of 80% coverage. From all nematode ASVs, eighteen were identified as entomopathogenic nematodes.

Seven nematode species of interest were identified by HTS; four EPN species and four free living, bacteriophagous nematodes (FLBN) in the genus *Osccheius*, two of which were associated with EPNs in previous studies (Campos-Herrera et al., 2015b; Ye et al., 2018). *Osccheius onirici* and *O. tipulae* were detected by both HTS and qPCR, but *O. dolichura* and *O. myriophyla* were only detected using HTS, because primers-probes were not designed and not used for these two species. Eighty ASV's matching species in the *Acrobeloides* genus were recovered by HTS; however, blast results did not provide unambiguous identification, even though the query sequence was 100% identical to the reference sequence of the database. HTS did not detect any of the *Pristionchus* species detected by qPCR (Table 1).

Four EPN species were detected by HTS, compared to just two using qPCR. It is unremarkable that *Heterorhabditis megidis* was found only using HTS, because primers-probe for this species were not employed by Campos-Herrera et al., (2019a). Of the remaining EPN species detected by HTS, *S. affine*, failed to amplify using qPCR. The relative population densities of the four nematode species detected by both methods (*S. feltiae*, *H. bacteriophora*, *O. tipulate* and *O. onirici*) were significantly ($P < 0.0001$) correlated, with between 69%-95% of the variability explained by the models (Table 2).

The detection frequency provided by HTS was generally higher than that from qPCR. The most commonly encountered EPN *S. feltiae* was detected by HTS 62% more frequently (34 samples, 60.7%) than by qPCR (21 samples, 37.5%). NGS detected *S. feltiae* in all 13 samples from palmetto, 92.3% of 13 samples of oak, 41% of 7 citrus samples, and in 77% of 13 samples from pine. By contrast, qPCR detected *S. feltiae* in just 69% of oak habitats, in fewer than half and a quarter of the samples from citrus and pine, respectively, and did not detect *S. feltiae* in any palmetto habitat (Figure 2). *Osccheius onirici* was detected about twice as often (9 samples)

by NGS compared to qPCR (4 samples) and *O. tipulae*, the most frequently encountered species was found in all but seven NGS samples (96%) compared to all but 6 qPCR samples (89%). Despite the effect of detection frequency on the estimated relative abundance of the nematode in different vegetative habitats, results of ANOVA did not differ using either NGS or qPCR measurements (Figure 3). Plant species were shown to have a highly significant effect ($P < 0.01$) on *S. feltiae* abundance when data from both methods were subjected to one-way analysis of variance, with oak habitat supporting more *S. feltiae* than pine, palmetto or citrus orchards (Figure 3).

Taylor's law explained 95% of the variability in the *S. feltiae* variance-mean relationship derived from HTS sample statistics (HTS $P_{19} < 0.001$), compared to 76% for the qPCR-derived statistics (qPCR $P_{19} < 0.001$) (Figure 4). However, the *O. tipulae* variability explained by qPCR (68%) was more than 80% higher than that explained using NGS data (37%) (not shown). The slope of the regression line using HTS indicated that *S. feltiae* population is highly aggregated whereas that using qPCR suggested a more random spatial pattern (Figure 5).

Data from the two detection methods resulted in similar regression models of EPN against the biotic and abiotic habitat variables. For *S. feltiae*, elevation, electrical conductivity, soil moisture, pH and P level were significant abiotic variables using either data set. The NF *Arthrobotrys oligospora* was inversely related to *S. feltiae* measured by NGS and *Hirsutella rhossiliensis* was inversely related to the nematode using both data sets. *Oscheius onirici* was positively associated with *S. feltiae* measured by qPCR. More variation in the data was explained by NGS measurements (60%) than by those of qPCR (35%) (Table 3). Regression models for *Oscheius tipulae* were also very similar for the two data sets, with pH silt and clay, *A. oligospora* and *Purpureocillium lilacinum* being significant biotic and abiotic variables using either data set.

The models explained 59% of the variability for NGS data and 17% for those from qPCR. *Paenibacillus* sp. was not found to be related to any EPN species, but the total amount of the bacterium and the bacterial encumbrance rate per *S. feltiae* ($\log \text{Paenibacillus sp.} - \log S. feltiae$) were positively associated with soil pH, which ranged between 4.2 and 8.18 (Table 3).

Canonical correspondence analysis (CCA) of data from qPCR explained a greater amount of the total variation than did data from HTS (Figure 3). The CCA identified three significant abiotic variables: pH, clay and elevation ($P < 0.05$). The same variables were most important for the HTS model, and model significance was nearly the same using either data set (Table 4).

Blasting results of HTS data revealed that species such as *S. feltiae* and *O. tipulae* comprise multiple ASVs (Table 5), exhibiting patterns referred to as “head-tail” by Porazinska et al. (2010).

4. DISCUSSION

Characterizing EPN biogeography with a goal of conservation biological control requires fine-scale taxonomic resolution, because closely related EPN species can exhibit highly divergent phenotypes for key properties such as habitat adaptation (El-Borai et al., 2016) and insect host specificity (Lewis et al., 2006; Peters, 1996; Simões and Rosa, 1996). Here we showed that HTS technologies, targeting the rDNA ITS region, can achieve the required resolution within this nematode guild. This is the first report in which HTS was used to identify natural communities of EPN species.

Both original hypotheses were supported by the data. Using DNA from nematodes and other organisms extracted from soil samples, the HTS measurements of *S. feltiae* population

density were highly correlated with those obtained previously from qPCR (Campos-Herrera et al., 2019) and the detection frequency was significantly higher using HTS. Blasting the sequencing data did not reveal any of *Pristionchus* spp. reads, probably because of the primers' limitations. In the case of *Acrobeloides*, multiple ASVs were identified as *Acrobeloides* sp.; however, blast results did not provide unequivocal identification of the genus, suggesting either that ITS1 is not an informative region or there are erroneous reference data. These results explain why Campos-Herrera et al., (2012) designed qPCR probes targeting 18S SSU region for *Acrobeloides*-group detection. In the case of *S. feltiae*, a frequent lack of detection by qPCR resulted in apparent differences in the capacity of EPNs to colonize some vegetation habitats with few samples (e.g., palmetto); however, the differences between the two methods were not significant when discriminating the relative habitat preference among these plants. The intraspecific variability of the *S. feltiae* ITS region may account for the apparently lower detection limit of HTS compared to qPCR. Primers and probes designed for a species in one region may be relatively strain-specific (Spiridonov et al., 2004). Additionally, qPCR may be less sensitive than procedures using the Illumina MiSeq sequencer. Approximately 10 copies per reaction are required for detection by qPCR (Forootan et al., 2017), whereas the threshold for HTS is undetermined. Other comparative studies have also found MiSeq superior to qPCR, as well as HTS platforms such as Ion Torrent PGM and Roche-454, for detection of pathogens in mock samples (Frey et al., 2014).

We found conflicting evidence that the HTS measurements were more reliable than those of qPCR. Taylor's power law is an empirical law in ecology relating the variance and mean of the number of individuals of a species per unit area of habitat. The fit of the regression line to the data provides evidence of the measurement reliability and the regression slope is a quantifiable

measure of population dispersion. The slope using HTS measurements indicates that *S. feltiae* tended to be highly aggregated, whereas that from qPCR measurements revealed a tendency toward randomness. The fit of the model to the different data sets ($R^2=0.96$ vs 0.76) supported the interpretation provided by the HTS measurements as the more likely property of this species. Nevertheless, despite the good fit to the *S. feltiae* HTS data set, it is not apparent why neither qPCR nor HTS measurements of *O. tipulae* were especially well described by the power law.

The relationships measured between habitat properties and the two most frequently encountered nematodes, *S. feltiae* and *O. tipulae*, were very similar for the HTS and qPCR data sets; however, HTS data provided stronger support as measured by the coefficients of determination for stepwise multiple regression. An inverse relationship between the EPN and two common nematophagous fungi suggests that some habitat properties may favor the predaceous fungi at the expense of the EPN. For example, soil moisture was weakly inversely related to both fungal species ($P < 0.1$; not shown) and may partly account for greater abundance of *S. feltiae* in the wetter soils.

The multiple regression relationships between *S. feltiae*, *Paenibacillus* sp. and pH especially warrant additional study. *Paenibacillus* sp. was described as an ectoparasitic bacterium specific to *Steinernema diaprepesi* that exhibited properties of density dependent regulation of the nematode in laboratory experiments (El-Borai et al., 2005) and temporal surveys in the field (Campos-Herrera et al., 2019b). Basically, the bacterium adhered to the cuticle and impeded movement and host-finding of *S. diaprepesi* at high soil pH but detached from the cuticle at low pH. Campos-Herrera et al., (2019a) speculated that the detection of the bacterium in Algarve in the absence of *S. diaprepesi* indicated a lack of specificity by the primers-probe set or that the bacterium is less host specific than reported. If this bacterium in the

Algarve is associated with EPNs, *S. feltiae* would be a possibility. The observation that qPCR detected *S. feltiae* in only half of the sites inhabited by *Paenibacillus* sp. does not support the likelihood of a close association between the two, whereas HTS detected *S. feltiae* at all but one of the 16 sites where the bacterium was found. Moreover, both *S. feltiae* and the bacterium were associated with soil pH in the same manner as reported in previous laboratory and field experiments (Campos-Herrera et al., 2019b, 2014, 2013b; El-Borai et al., 2005). EPN abundance was inversely related to soil pH while the bacterium was highly positively associated with pH both in total abundance and when expressed as spore abundance per *S. feltiae* abundance (i.e. spore encumbrance rate).

Based on our comparative results, HTS seems preferable to qPCR for community analyses for multiple reasons. The cost of HTS here was about twenty percent higher than that of qPCR. However, HTS potentially reveals everything in a soil sample that can be amplified by universal primers, whereas qPCR found only that which was sought. *Heterorhabditis megidis* was found only by HTS, because qPCR was not attempted for this species. *H. megidis* is a cosmopolitan species, occurring worldwide in temperate regions from North America to Asia, with an apparent preference for turf and weedy habitats (Stock and Kaya, 1996; Stuart and Gaugler, 1994). It has been reported also in some Mediterranean countries like Greece (Menti et al., 1997), Turkey (Yilmaz et al., 2009) and Israel (Glazer et al., 1993), but this is the first report of *H. megidis* in Portugal.

The specificity of qPCR primers-probes is probably sometimes excessive for communities with significant intraspecific variability. Blasting the results of HTS suggested that most species consist of multiple ASVs. Porazinska et al., (2010), described species such as *S. feltiae* as exhibiting “head-tail” patterns where a single head ASV perfectly matches a NCBI

reference sequence and comprises the majority of the sequencing reads (“head” formed 99% of the *S.feltiae* reads) with tail ASVs represented by just a few sequences, revealing the real ITS 1 variation of a species. A phylogenetic tree verified that the eight *S. feltiae* ASVs probably belong to the same species because they are more closely related to each other than to any other closely related species. Literature and NCBI database survey, suggested that the closest related species to *S. feltiae* is the *Steinernema citrae* which was used as a root to our phylogenetic tree.

EPNs are a well-studied guild and the information included in the NCBI database, especially for the ribosomal gene, is relatively good compared to other groups. Nevertheless, an ongoing challenge of HTS and all molecular survey methods is the quality of reference databases, which contain mistaken identities and taxonomic gaps for known and undescribed species. HTS reveals these questionable sequences for further study, whereas they remain undetected by qPCR. In this study, *S. affine* consists of two ASVs which exhibited 98.1% and 98.0% identity to a reference sequence. Further study of the populations can resolve whether the 2% difference between the query and the reference sequences reveal an undescribed species or intraspecific variation of *S. affine*.

The congruence of results from those species relevant to both the qPCR and HTS tools used here, support the use of HTS for soil community analyses at the species level. The capacity of HTS to measure infinitely more species at a lower cost than can be achieved by qPCR will ensure wide adoption of metagenomic methods and hasten our understanding of EPN biogeography and those factors that modulate EPN presence and abundance.

5. ACKNOWLEDGEMENT

This study was supported in part by the US-Egypt Project cycle 17 (no. 172) entitled ‘Preparing and evaluating IPM tactics for increasing strawberry and citrus production.’ This article is derived from the Subject Data funded in part by NAS and USAID, and that any opinions, findings, conclusions, or recommendations expressed in it are those of the author alone, and do not necessarily reflect the views of USAID or NAS

6. REFERENCES

- Andrews, S., Krueger, F., Seckerson-Pichon, A., Biggins, F., Wingett, S., 2015. FastQC. A quality control tool for high throughput sequence data. Babraham Bioinformatics [WWW Document]. Babraham Institute. doi:v.0.11.5
- Anisimova, M., Gascuel, O., 2006. Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. *Systematic Biology* 55, 539–552. doi:10.1080/10635150600755453
- Blaxter, M., 2003. Counting angels with DNA. *Nature* 421, 122–123. doi:10.1038/421122a
- Braun-Kiewnick, A., Kiewnick, S., 2018. Real-time PCR, a great tool for fast identification, sensitive detection and quantification of important plant-parasitic nematodes. *European Journal of Plant Pathology*. doi:10.1007/s10658-018-1487-7
- Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal* 11, 2639–2643. doi:10.1038/ismej.2017.119
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13, 581–3. doi:10.1038/nmeth.3869
- Campos-Herrera, R., Ali, J.G., Diaz, B.M., Duncan, L.W., 2013a. Analyzing spatial patterns linked to the

- ecology of herbivores and their natural enemies in the soil. *Frontiers in Plant Science* 4, 378.
doi:10.3389/fpls.2013.00378
- Campos-Herrera, R., Blanco-Pérez, R., Bueno-Pallero, F.Á., Duarte, A., Nolasco, G., Sommer, R.J.,
Rodríguez Martín, J.A., 2019a. Vegetation drives assemblages of entomopathogenic nematodes and
other soil organisms: Evidence from the Algarve, Portugal. *Soil Biology and Biochemistry* 128,
150–163. doi:10.1016/j.soilbio.2018.10.019
- Campos-Herrera, R., El-Borai, F.E., Duncan, L.W., 2012. Wide interguild relationships among
entomopathogenic and free-living nematodes in soil as measured by real time qPCR. *Journal of*
Invertebrate Pathology 111, 126–135. doi:10.1016/j.jip.2012.07.006
- Campos-Herrera, R., El-Borai, F.E., Ebert, T.E., Schumann, A., Duncan, L.W., 2014. Management to
control citrus greening alters the soil food web and severity of a pest-disease complex. *Biological*
Control 76, 41–51. doi:10.1016/j.biocontrol.2014.04.012
- Campos-Herrera, R., Jaffuel, G., Chiriboga, X., Blanco-Pérez, R., Fesselet, M., Půža, V., Mascher, F.,
Turlings, T.C.J., 2015a. Traditional and molecular detection methods reveal intense interguild
competition and other multitrophic interactions associated with native entomopathogenic nematodes
in Swiss tillage soils. *Plant and Soil* 389, 237–255. doi:10.1007/s11104-014-2358-4
- Campos-Herrera, R., Pathak, E., El-Borai, F.E., Schumann, A., Abd-Elgawad, M.M.M., Duncan, L.W.,
2013b. New citriculture system suppresses native and augmented entomopathogenic nematodes.
Biological Control 66, 183–194. doi:10.1016/J.BIOCONTROL.2013.05.009
- Campos-Herrera, R., Půža, V., Jaffuel, G., Blanco-Pérez, R., Čepulyte-Rakauskiene, R., Turlings, T.C.J.,
2015b. Unraveling the intraguild competition between *Osccheius* spp. nematodes and
entomopathogenic nematodes: Implications for their natural distribution in Swiss agricultural soils.
Journal of Invertebrate Pathology 132, 216–227. doi:10.1016/j.jip.2015.10.007
- Campos-Herrera, R., Stuart, R.J., Pathak, E., El-Borai, F.E., Duncan, L.W., 2019b. Temporal patterns of
entomopathogenic nematodes in Florida citrus orchards: Evidence of natural regulation by
microorganisms and nematode competitors. *Soil Biology and Biochemistry* 128, 193–204.

- doi:10.1016/j.soilbio.2018.10.012
- Dolinski, C., Choo, H.Y., Duncan, L.W., 2012. Grower acceptance of entomopathogenic nematodes: Case studies on three continents. *Journal of Nematology* 44, 226–235.
- Duncan, L.W., McCoy, C.W., Stansly, P.A., Graham, J.H., Mizell, R.F., 2001. Estimating the Relative Abundance of Adult Citrus Root Weevils (Coleoptera: Curculionidae) with Modified Tedders Traps. *Environmental Entomology* 30, 939–946. doi:10.1603/0046-225X-30.5.939
- Duncan, L.W., McCoy, C.W., Terranova, A.C., 1996. Estimating sample size and persistence of entomogenous nematodes in sandy soils and their efficacy against the larvae of *diaprepes abbreviatus* in Florida. *Journal of Nematology* 28, 56–67.
- El-Borai, F., Killiny, N., Duncan, L.W., 2016. Concilience in entomopathogenic nematode responses to water potential and their geospatial patterns in Florida. *Frontiers in Microbiology* 7, 1–12. doi:10.3389/fmicb.2016.00356
- El-Borai, F.E., Duncan, L.W., Preston, J.F., 2005. Bionomics of a Phoretic Association Between *Paenibacillus* sp. and the Entomopathogenic Nematode *Steinernema diaprepesi*. *Journal of Nematology* 37, 18–25.
- Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)* 32, 3047–8. doi:10.1093/bioinformatics/btw354
- Forootan, A., Sjöback, R., Björkman, J., Sjögreen, B., Linz, L., Kubista, M., 2017. Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). *Biomolecular Detection and Quantification* 12, 1–6. doi:10.1016/j.bdq.2017.04.001
- Frey, K.G., Herrera-Galeano, J.E., Redden, C.L., Luu, T. V, Servetas, S.L., Mateczun, A.J., Mokashi, V.P., 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. doi:10.1186/1471-2164-15-96
- Georgis, R., Koppenhöfer, A.M., Lacey, L.A., Bélair, G., Duncan, L.W., Grewal, P.S., Samish, M., Tan,

- 453 L., Torr, P., van Tol, R.W.H.M., 2006. Successes and failures in the use of parasitic nematodes for
 454 pest control. *Biological Control* 38, 103–123. doi:10.1016/J.BIOCONTROL.2005.11.005
- 455 GLAZER, I., LIRAN, N., POINAR, G., SMITS, P., 1993. Identification and biological activity of newly
 456 isolated heterorhabditid populations from Israel. *Fundamental and Applied Nematology* 16, 467–
 457 472.
- 458 Guindon, S., Gascuel, O., 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies
 459 by Maximum Likelihood. *Systematic Biology* 52, 696–704. doi:10.1080/10635150390235520
- 460 Iqbal, M., Ehlers, R.-U., Waeyenberge, L., 2016. Molecular characterisation of novel isolates
 461 of entomopathogenic nematodes. *Nematology* 18, 277–291. doi:10.1163/15685411-00002959
- 462 Kaya, H.K., Gaugler, R., 1993. Entomopathogenic Nematodes. *Annual Review of Entomology* 38, 181–
 463 206. doi:10.1146/annurev.en.38.010193.001145
- 464 Lewis, E.E., Campbell, J., Griffin, C., Kaya, H., Peters, A., 2006. Behavioral ecology of
 465 entomopathogenic nematodes. *Biological Control* 38, 66–79. doi:10.1016/j.biocontrol.2005.11.007
- 466 Menti, H., Wright, D.J., Perry, R.N., 1997. Desiccation survival of populations of the entomopathogenic
 467 nematodes *Steinernema feltiae* and *Heterorhabditis megidis* from Greece and the UK. *Journal of*
 468 *Helminthology* 71, 41–46. doi:10.1017/s0022149x00000778
- 469 Neumann, G., Shields, E.J., 2011. Field persistence of *steinernema carpocapsae* Weiser (NY001),
 470 *Steinernema feltiae* Filipjev (Valko) and *Heterorhabditis bacteriophora* Poinar (Oswego) in Alfalfa
 471 Fields. *Great Lakes Entomologist* 44, 42–52.
- 472 Parkman, J.P., Hudson, W.G., Frank, J.H., Nguyen, K.B., Smart, G.C., 1993. Establishment and
 473 Persistence of *Steinernema scapterisci* (Rhabditida: Steinernematidae) in Field Populations of
 474 *Scapteriscus* spp. Mole Crickets (Orthoptera: Gryllotalpidae). *Journal of Entomological Science* 28,
 475 182–190. doi:10.18474/0749-8004-28.2.182
- 476 Peters, A., 1996. The natural host range of *Steinernema* and *Heterorhabditis* spp. and their impact on
 477 insect populations. *Biocontrol Science and Technology* 6, 389–402. doi:10.1080/09583159631361
- 478 Porazinska, D.L., Giblin-Davis, R.M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K., Powers, T.O.,

- Tucker, A.E., Sung, W., Thomas, W.K., 2009. Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources* 9, 1439–1450. doi:10.1111/j.1755-0998.2009.02611.x
- Porazinska, D.L., Giblin-Davis, R.M., Sung, W., Thomas, W.K., 2010. Linking operational clustered taxonomic units (OCTUs) from parallel ultra sequencing (PUS) to nematode species. *Zootaxa* 55–63. doi:10.11646/zootaxa.2427.1.6
- Qing, X., Wang, M., Karssen, G., Bucki, P., Bert, W., Braun-Miyara, S., 2019. PPNID: a reference database and molecular identification pipeline for plant-parasitic nematodes. *Bioinformatics*. doi:10.1093/bioinformatics/btz707
- Shapiro-Ilan, D.I., Cottrell, T.E., Mizell, R.F., Horton, D.L., Zaid, A., 2015. Field suppression of the peachtree borer, *Synanthedon exitiosa*, using *Steinernema carpocapsae*: Effects of irrigation, a sprayable gel and application method. *Biological Control* 82, 7–12. doi:10.1016/J.BIOCONTROL.2014.12.003
- Shapiro-Ilan, D.I., Gouge, D.H., Piggott, S.J., Fife, J.P., 2006. Application technology and environmental considerations for use of entomopathogenic nematodes in biological control. *Biological Control* 38, 124–133. doi:10.1016/J.BIOCONTROL.2005.09.005
- Shields, E.J., Testa, A., Neumann, G., Flanders, K.L., Schroeder, P.C., 2009. Biological Control of Alfalfa Snout Beetle with a multi-species application of locally adapted persistent entomopathogenic nematodes: The first success. *American Entomologist* 55, 250–257. doi:10.1093/ae/55.4.250
- Simões, N., Rosa, J.S., 1996. Pathogenicity and host specificity of entomopathogenic nematodes. *Biocontrol Science and Technology* 6, 403–411. doi:10.1080/09583159631370
- Spiridonov, S.E., Moens, M., Wilson, M.J., 2007. Fine scale spatial distributions of two entomopathogenic nematodes in a grassland soil. *Applied Soil Ecology* 37, 192–201. doi:10.1016/J.APSOIL.2007.06.009
- Spiridonov, S.E., Reid, A.P., Podrucka, K., Subbotin, S.A., Moens, M., 2004. Phylogenetic relationships within the genus *Steinernema* (Nematoda: Rhabditida) as inferred from analyses of sequences of the

- ITS1-5.8S-ITS2 region of rDNA and morphological features. *Nematology* 6, 547–566.
doi:10.1163/1568541042665304
- Stock, S.P., Kaya, H.K., 1996. A Multivariate Analysis of Morphometric Characters of *Heterorhabditis* Species (Nemata: Heterorhabditidae) and the Role of Morphometrics in the Taxonomy of Species of the Genus. *The Journal of Parasitology* 82, 806. doi:10.2307/3283895
- Stuart, R.J., Barbercheck, M.E., Grewal, P.S., Taylor, R.A.J., Hoy, C.W., 2006. Population biology of entomopathogenic nematodes: Concepts, issues, and models. *Biological Control* 38, 80–102. doi:10.1016/J.BIOCONTROL.2005.09.019
- Stuart, R.J., Gaugler, R., 1994. Patchiness in Populations of Entomopathogenic Nematodes. *Journal of Invertebrate Pathology* 64, 39–45. doi:10.1006/jipa.1994.1066
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22, 4673–4680. doi:10.1093/nar/22.22.4673
- TORR, P., SPIRIDONOV, S.E., HERITAGE, S., WILSON, M.J., 2007. Habitat associations of two entomopathogenic nematodes: a quantitative study using real-time quantitative polymerase chain reactions. *Journal of Animal Ecology* 76, 238–245. doi:10.1111/j.1365-2656.2006.01196.x
- Treonis, A.M., Unangst, S.K., Kepler, R.M., Buyer, J.S., Cavigelli, M.A., Mirsky, S.B., Maul, J.E., 2018. Characterization of soil nematode communities in three cropping systems through morphological and DNA metabarcoding approaches. *Scientific Reports* 8, 2004. doi:10.1038/s41598-018-20366-5
- Ye, W., Foye, S., MacGuidwin, A.E., Steffan, S., 2018. Incidence of *Oscheius onirici* (Nematoda: Rhabditidae), a potentially entomopathogenic nematode from the marshlands of Wisconsin, USA. *Journal of Nematology* 50, 9–26. doi:10.21307/jofnem-2018-004
- Yilmaz, H., Waeyenberge, L., Demir, I., Moens, M., Demirbağ, Z., 2009. A new entomopathogenic nematode species for Turkey, *Heterorhabditis megidis* Poinar, Jackson & Klein 1987 (Rhabditida: Heterorhabditidae). *Turkish Journal of Agriculture and Forestry* 33, 385–391. doi:10.3906/tar-0808-

531 26

532

Table 1. The frequency (percent of 56 samples) with which entomopathogenic nematodes (EPNs) and species previously associated with EPNs were detected by metabarcoding and qPCR.

		Metabarcoding	qPCR
EPN	<i>Steinernema feltiae</i>	60.7	37.5
	<i>Steinernema affine</i>	3.5	
	<i>Heterorhabditis bacteriophora</i>	7.1	3.5
	<i>Heterorhabditis megidis</i>	1.7	
Competitors	<i>Acrobeloides-group</i>		62.5
	<i>Oscheius onirici</i>	16	7.1
	<i>Oscheius tipulae</i>	87.5	89.2
	<i>Pristionchus maupaci</i>		3.5
	<i>Pristionchus pacificus</i>		30.3

Table 2. Non-parametric Spearman's correlations between the species' measurements from metabarcoding and qPCR

Species	Spearman ρ	Prob> ρ
<i>Steinernema feltiae</i>	0.8737	<.0001
<i>Heterorhabditis bacteriophora</i>	0.7201	<.0001
<i>Oschieus tipulae</i>	0.9026	<.0001
<i>Oschieus onirici</i>	0.6393	<.0001

Table 3. Significant variables from stepwise multiple regression of two nematode species measured by high throughput sequencing or qPCR and a bacterial species measured by qPCR, regressed against soil properties and potential biotic antagonists. Independent variables included *Oschieus tipulae*, *Oschieus onirici*, *Arthrobotrys oligospora*, *Arthrobotrys dactyloides*, *Purpureocillium lilacinum*, *Hirsutella rhossiliensis*, *Paenibacillus* sp., elevation, soil moisture, pH, electrical conductivity, percent clay, silt, sand and organic matter, K, P, Mg, Zn, Fe.

Species	<i>S. feltiae</i>		<i>O. tipulae</i>		<i>Paenibacillus</i> sp. abundance	<i>Paenibacillus</i> sp. encumbrance
Approach	HTS	qPCR	HTS	qPCR		
R ² / R ² adjusted	0.70/0.60	0.50/0.35	0.69/0.59	0.33/0.17	0.61/0.50	0.69/0.55
<i>A. oligospora</i>	(-)0.02			0.037		
<i>H. rhossiliensis</i>	(-)0.01	(-)0.04			(-)0.005	
<i>P. lilacinus</i>			0.001	0.06		
<i>O. onirici</i>		0.05				
Elevation	(-)0.001	(-)0.008	(-)0.001			0.001
<i>Paenibacillus</i> sp			0.021	0.014	-	-
H ₂ O	0.05	0.015	0.003		(-)0.031	(-)0.001
OM%					0.007	
EC	0.001	0.01				0.05
pH	(-)0.02	(-)0.005		0.026	0.01	<0.001
Sand			0.002	0.058	0.028	
Silt			0.007	0.028		
Clay						
P	0.001	0.06			0.045	(-)0.007
K					0.005	0.008

Table 4. Significance of the Canonical Correspondence Analysis model, axes, and variables.

	Metabarcoding		qPCR	
	Pr(>F)		Pr(>F)	
Model	0.043	*	0.03	*
CCA1	0.086	.	0.143	
CCA2	0.53		0.198	
pH	0.011	*	0.01	*
Clay	0.064	.	0.014	*
Elevation	0.163		0.023	*
P	0.074	.	0.462	

Table 5. *S. feltiae* ASVs table illustrating “head-tail” structure associated with the presence of within the species variation.

#	identity %	coverage	e-value	Description title	No. reads	Length	
1	100	100	1.94E-126	<i>Steinernema feltiae</i> isolate H9	50100	250	Head
2	100	100	1.94E-126	<i>Steinernema feltiae</i> isolate I2	2415	250	Tail
3	99.6	100	9.02E-12	<i>Steinernema feltiae</i> isolate H9	1082	250	
4	99.6	100	9.02E-12	<i>Steinernema feltiae</i> isolate 11A	686	250	
5	99.6	100	9.02E-12	<i>Steinernema feltiae</i> isolate H9	517	250	
6	97.22	100	2.54E-115	<i>Steinernema feltiae</i> isolate DONR	355	250	
7	99.6	100	9.02E-12	<i>Steinernema feltiae</i> isolate H9	291	250	
8	100	100	1.94E-126	<i>Steinernema feltiae</i> isolate 11A	76	250	

Figure 1. Phylogenetic relationships of ASVs identified as *S. feltiae* based on sequencing reads of ITS_1 region as inferred by maximum likelihood. *Steinernema citrae* (MF536110) was used as an outgroup.

Figure 2. The frequency of *S. feltiae* (percent of 56 samples) through different types of vegetation was detected by metabarcoding and qPCR.

Figure 3. Box plots of data from high throughput sequencing (A) and qPCR (B) measuring *Steinernema feltiae* populations in Portugal. Differences in means are designated by boxes without the same letters and were determined by Tukey's HSD test ($P \leq 0.05$) using $\log(X+1)$ transformed data.

Figure 4. Fit of Taylor's Power Law to sample statistics for *Steinernema feltiae* populations in Portugal measured using high throughput sequencing (A) and qPCR (B).

Figure 5. Canonical correspondence analysis depicting biplots of the regional distribution and relationships between significant abiotic factors and soil organisms.

Figure 1

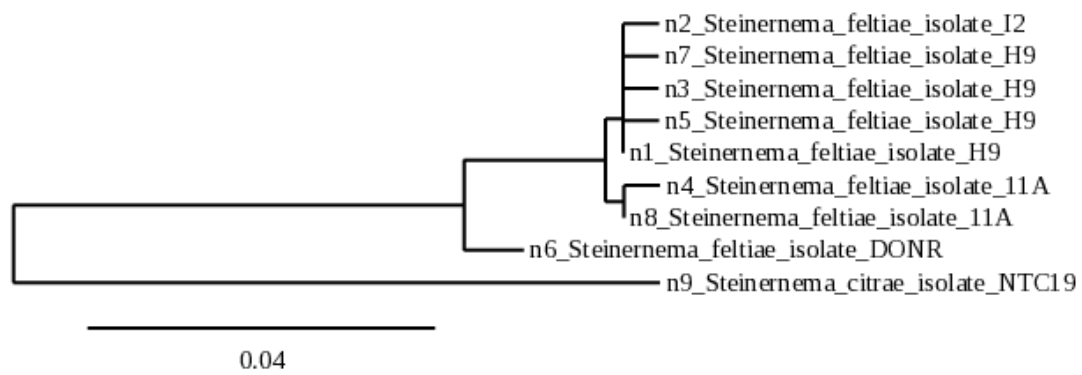


Figure 2

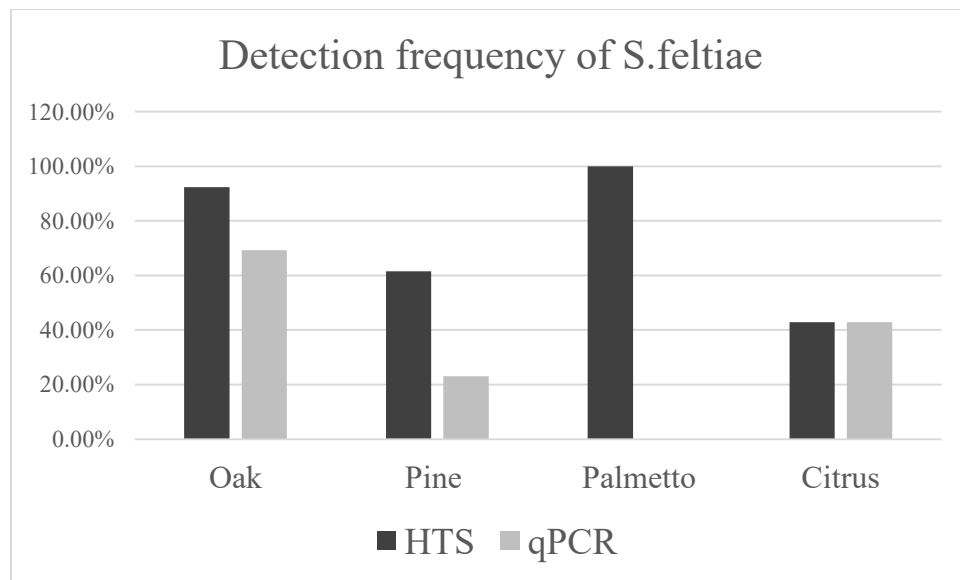


Figure 3

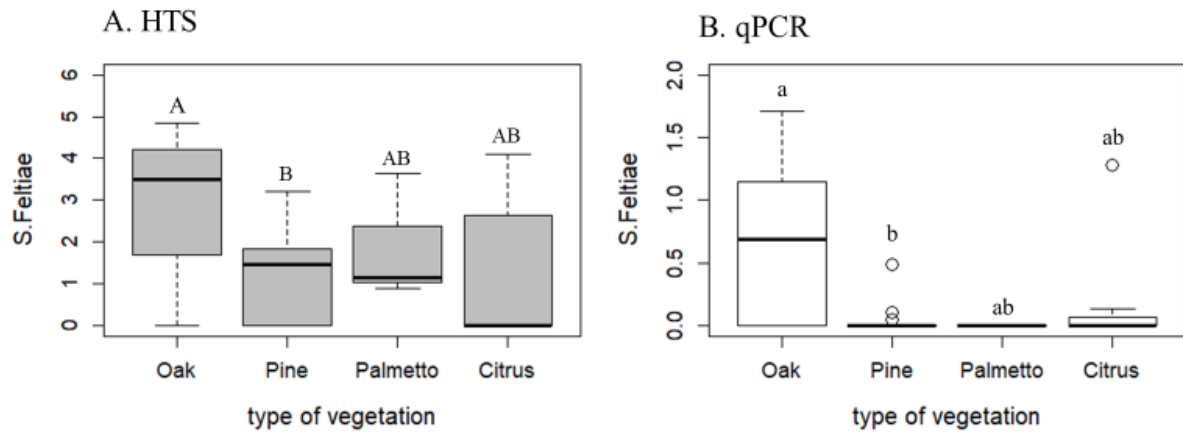
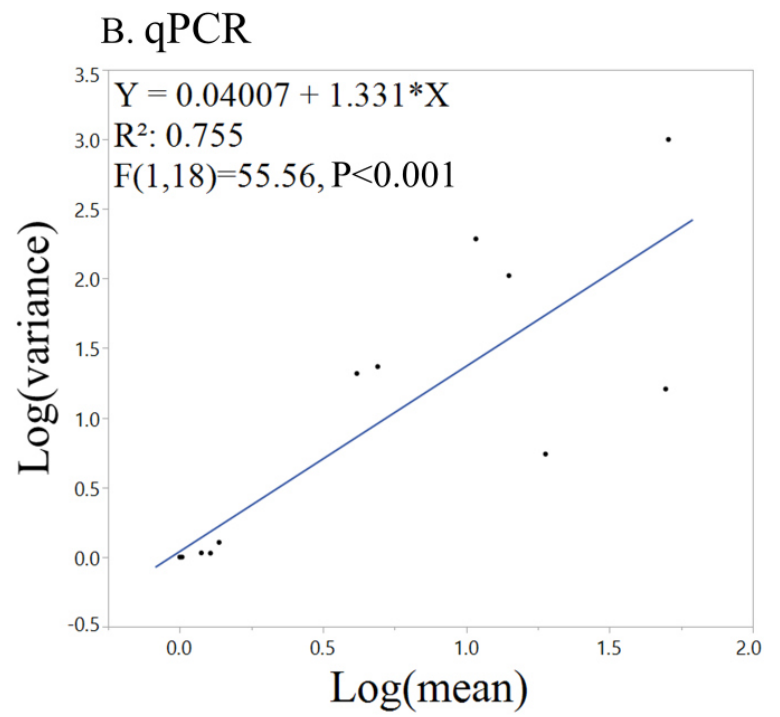
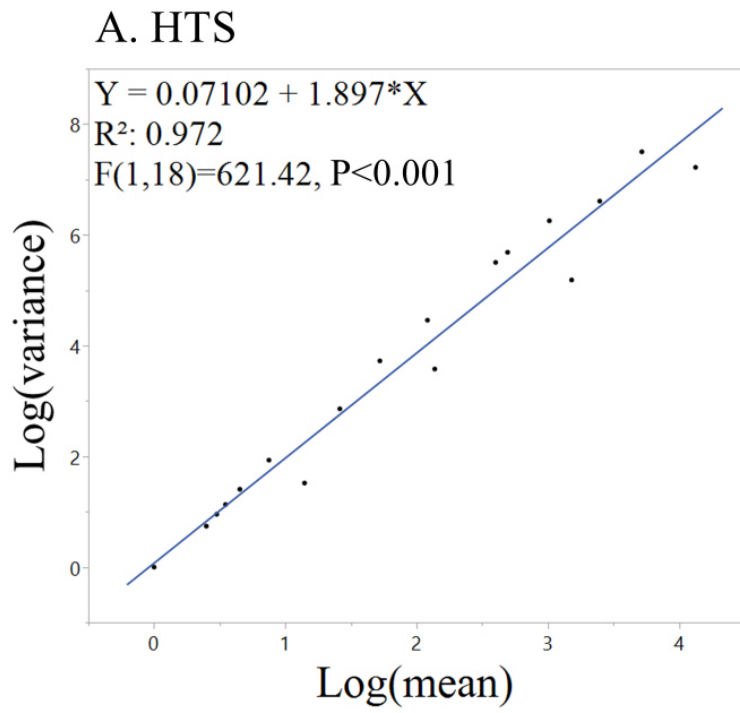


Figure 4



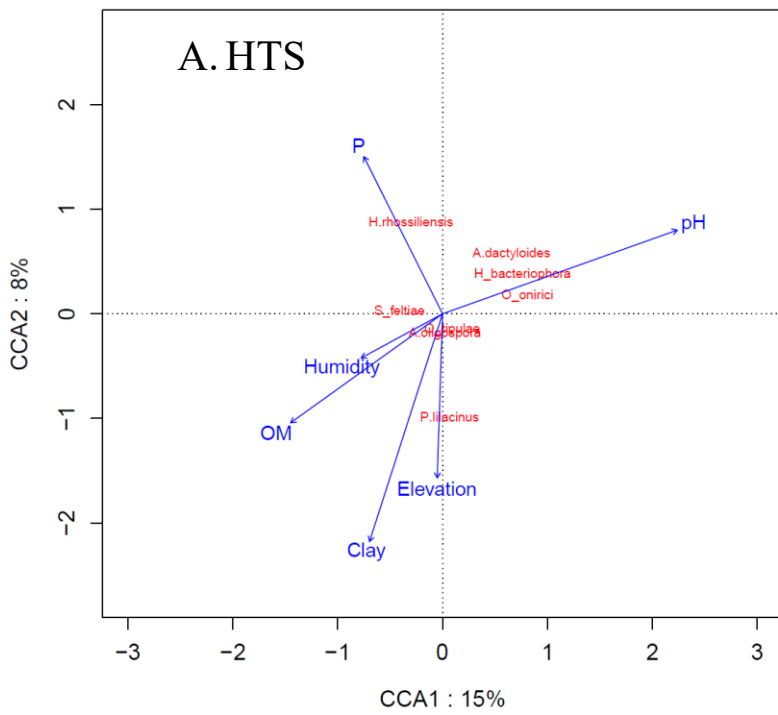


Figure 5

