Molecular Detection and Quantification of the *Striga* Seedbank in Ethiopian Sorghum Field Soils

Jos Raaijmakers (j.raaijmakers@nioo.knaw.nl)
Netherlands Institute of Ecology: Nederlands Instituut voor Ecologie https://orcid.org/0000-0003-1608-6614

Getahun Mitiku
Netherlands Institute of Ecology: Nederlands Instituut voor Ecologie

Desalegn Etalo
Netherlands Institute of Ecology: Nederlands Instituut voor Ecologie

Paulien KleinGunnewiek
Netherlands Institute of Ecology: Nederlands Instituut voor Ecologie

Dominika Rybka
Ethiopian Institute of Agricultural Research

Taye Tessema
Netherlands Institute of Ecology: Nederlands Instituut voor Ecologie

Research Article

Keywords: Striga seedbank, qPCR, sorghum field

Posted Date: June 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-572695/v1

Abstract

Aims Striga hermonthica is a devastating parasitic weed in Sub-Saharan Africa (SSA) and its persistent soil seedbank is the major contributing factor for its prevalence and persistence. So far, there is little to no information on the *Striga* seedbank density in agricultural fields in SSA due to the lack of reliable detection and quantification methods.

Methods We developed a high-throughput method that combines density- and size-based separation techniques with quantitative polymerase chain reaction (qPCR)-based detection of *Striga* seeds in soil. The method was optimized and validated on two physicochemically different *Striga*-free Dutch agricultural soils by introducing increasing numbers of *Striga* seeds (0, 1, 3, 9, 27, 81 and 243 seeds).

Results The results showed that as little as one seed of *S. hermonthica* per 150 g of soil can be detected. This technique was subsequently tested on soil samples of 48 sorghum fields from different agro-ecological zones in Ethiopia to map the geospatial distribution of the *Striga* seedbank along a trajectory of more than 1500 km. Considerable variation in *Striga* seed densities was observed for these soils: in 75% of the field soils, *Striga* seeds were detectable up to 86 seeds per 150 g of soil. Correlation analyses further revealed a significant non-linear relationship between the seed density and *Striga* incidence assessed in the same sorghum field soils at the time of soil sampling.

Conclusions The method developed allows for high-through-put and accurate mapping of the *Striga* seedbank in physicochemically diverse field soils and can be used to predict *Striga* incidence and to assess the impact of management strategies on *Striga* seedbank dynamics.

Introduction

Striga is one of the major genera of parasitic plants in Africa, Asia and Australia. More than 50 species of *Striga* have been reported across the globe and *S. hermonthica, S. asiatica*, *S. gesneroides, S. aspera* and *S. forbesii* are the most common and destructive in cultivated cereal and legume crops (Scholes and Press 2008; Parker 2009). *Striga* is responsible for more crop losses in Africa than any other weed species. It is estimated that two-thirds of the total area of cereals and legumes in sub-Saharan Africa is infested with *Striga* and its spread has accelerated at an alarming rate (Parker 2012). The annual yield losses due to *Striga* alone were estimated at US $7 billion in Sub-Saharan Africa (SSA), posing a major threat to the livelihood of over 300 million people (Badu-Apraku and Akinwale 2011). Ethiopia is one of the epicenters of *Striga* infestation in sub-Saharan Africa and crop losses of 65–100% are commonly reported for different sorghum growing regions of the country (Ejeta et al., 2002; Bayu et al. 2005; Tesso et al. 2007; Abate et al. 2014). Management of *Striga* in many parts of the world is constrained to a great extent by *Striga* seeds residing in the soil, also referred to as the “seedbank”.

Striga produces 10,000-200,000 tiny seeds (0.2–0.3 mm, 4–7 µg) per plant, which survive in soil for at least two years (Hearne 2009). Managing *Striga* requires a better understanding of seedbank replenishment and depletion, also referred to as the seedbank dynamics. Replenishment encompasses

seed production by mature *Striga* plants and 'immigration' of seeds from neighbouring field soils via wind and farmer activities. Seedbank depletion is caused by suicidal germination (i.e. germination in absence of a host plant), pathogen infection, seed predation, seed aging, and 'emigration' of seeds to neighboring fields (van Mourik 2007). Hence, a methodology that allows accurate detection and quantification of the *Striga* seed density in agricultural fields is of paramount importance in the management of *Striga* in general and for understanding the seedbank dynamics in particular.

So far, some attempts have been made to quantify the *Striga* seedbank in agricultural fields in different African countries (van Delft et al. 1997; Oswald and Ransom 2001; Abunyewa and Padi 2003; Schulz et al. 2003; Sauerborn et al. 2003; Franke et al. 2006; van Mourik 2007). These studies employed laborious manual seed-soil separation methods accompanied by counting of the *Striga* seeds under a microscope. Such an approach is time consuming and is often prone to biases and error considering the large diversity of physicochemical properties of the field soils. Due to the lack of fast and robust methods for detection and quantification of *Striga* seedbank, the assessment of emerged *Striga* plants in an area is often used as a proxy for seedbank density (van Delft et al. 1997). Hence, reliable detection and quantification of the *Striga* seedbank is crucial to understand the dynamics of this parasitic weed and to evaluate the effectiveness of management strategies (Westerman et al. 2007; van Mourik et al. 2008).

The advancement of techniques to extract environmental DNA and RNA (eDNA and eRNA) from soils with different physicochemical characteristics followed by qPCR or sequencing has opened new means for sensitive and accurate detection and quantification of specific (micro)organisms and marker genes. Such techniques have been used for several years to detect and quantify pathogenic microorganisms in order to deploy or optimize early measures to prevent disease outbreaks in farms (Ophel-Keller et al. 2008; Taparia et al. 2020). Furthermore, integration of high-throughput eDNA extraction and qPCR can maximize the number of samples that can be processed in a single day, reducing labor costs and turnaround times (Prider et al. 2013). Recently, the use of qPCR has received attention for determining the seedbank of weeds present in a soil. Examples include, the use of molecular markers and DNA-based assays for the quantification and identification of seeds of different species of the parasitic weeds *Orobanche* and *Phelipanche* (Dongo et al. 2012; Aly et al. 2012; Prider et al. 2013). With the recent public release of the genome sequence of *Striga* (http://ppgp.huck.psu.edu/), we here developed a high-throughput molecular detection method for quantification of the *Striga* seedbank in field soils. The method encompasses several steps, starting with a density- and size-based separation of *Striga* seeds from the soil matrix followed by eDNA extraction and qPCR-based detection and quantification. Furthermore, the optimized protocol was then used to quantify and map the geospatial distribution of the *Striga* seedbank in sorghum field soils collected from different agro-ecological zones in Ethiopia covering a trajectory of more than 1500 km and to relate the seed densities to *Striga* incidence in these fields.

Results And Discussion

Selection of marker genes, primer design and specificity

From the parasitic plant genome project (http://ppgp.huck.psu.edu/) the *Striga* StHe0GB1 genome assembly was downloaded. The genome sequence was generated from *Striga* seeds that have been imbibed for various lengths of time, covering seeds that were conditioned as well as seeds that were treated for up to 6 h with the germination stimulant GR-24 (Westwood et al. 2012). From the assembly, several genes were selected and blasted against the nr database. We selected five genes (StHe0GB1_1, StHe0GB1_9, StHe0GB1_20, StHe0GB1_76 and StHe0GB1_93) with *Striga*-specific sequences as a putative marker gene for *Striga* seed detection and quantification (Table 1). For these 5 genes, a total of 14 primer pairs were designed targeting the *Striga*-specific sequences (Table 1). Using the NCBI primer blast web tool and the nr-database, the specificity of the forward and reverse primers was validated *in silico* (data not shown). To validate the efficacy of these primers experimentally, we extracted DNA from ~ 6000 *Striga* seeds, from 100 mg of *Striga*-free Dutch agricultural soil spiked with ~ 6000 *Striga* seeds and 100 mg of the soil sample without *Striga* seeds (control). All primer sets, except set 3 (targeting StHe0GB1_1), resulted in the PCR product of the predicted size (Table S1) for the samples containing the *Striga* seeds, whereas no amplification product was observed for any of the primer sets with DNA extracted from the soil samples without *Striga* seeds (Fig. 1a). Next, we tested the primer pairs in qPCR at two annealing temperatures (56 and 60 ^{o}C) to determine sensitivity, specificity and stability of the primers. All primers amplified the genomic DNA of *Striga* seeds at both temperatures but with different sensitivity, specificity and stability (Fig. 1b**)**. Primer set 14 (P14) targeting the StHe0GB1_93 gene, showed high sensitivity as manifested by a low quantification cycle (*Cq)* value and single melting curve for both annealing temperatures (Fig. 1b and S1). Furthermore, P14 resulted in a PCR product of the expected size for five independent *S. hermonthica* seed batches and one *S. asiatica* seed batch collected from different agroecological zones in Ethiopia (Fig. S2). Hence, primer set P14 was selected for testing the specificity and sensitivity of PCR-based detection and quantification of *Striga* seeds in soil samples.

Optimizing DNA extraction and qPCR efficiency in different agricultural soils

Accurate determination of the seed density of *Striga* in a field soil requires a substantial amount of soil sample. Previous assays to microscopically determine *Striga* seed densities used 100 g soil samples (van Mourik 2007). However, such a large sample cannot be directly accommodated in the currently available high-throughput DNA extraction kits. Moreover, the application of molecular techniques for the detection and quantification of eDNA from soils may be hampered by humic acids, polysaccharides, urea, phenolic compounds and heavy metals (Frostegård et al. 1999). Here, we introduced 65 *Striga* seeds into seven physicochemically different *Striga*-free Dutch agricultural soils (Table S2), each weighing 100 mg. qPCR analysis on eDNA extracted from these ‘spiked’ soil samples showed variation in the mean *Cq* value from 27.3 to 29.3 cycles (Fig. 2a). This variation could be due to differences in soil physicochemical properties between these soils affecting the efficiency of eDNA extraction and/or qPCR analysis. Correlation analysis between the *Cq* values and a number of physicochemical properties of the seven soil samples revealed that *Cq* values were positively correlated to Fe, Mg, S, C, N,C/N and organic matter (OM) content, whereas they were negatively correlated to pH, K and P contents (albeit not statistically significant $p \geq 0.05$).

To minimize interference of soil physicochemical properties, we then tested if separation of *Striga* seeds from the bulk soil prior to eDNA extraction and qPCR could improve the sensitivity of *Striga* detection and quantification. To this end, we adopted a density-dependent K_2CO_3 separation of the *Striga* seeds from the soil matrix followed by successive sieving through two filters with meshes of 425 and 75 µm, respectively. The *Striga* seeds and smaller soil particles and organic debris retained on the 75 µm filters were collected and dried at 35 ^{0}C for 48 hours followed by grinding and DNA extraction. This procedure reduced the soil volume by on average 99.7% for two physicochemically different Dutch soils tested. By reducing so effectively the soil volume, the soil mixture can be used directly for DNA extraction using the widely available commercial extraction kits.

Next, we introduced increasing densities of *Striga* seeds in soils D08 and D17 at final densities of 0, 1, 3, 9, 27, 81 and 243 seeds per 150 g of soil and processed these soil samples as described above. Results of the qPCR analysis revealed that even a single *Striga* seed introduced into 150 g of soil sample can be detected by qPCR in both soil types (Fig. 2b). Furthermore, the variations in *Cq* values for the same seed density in both soil types were minimal, suggesting efficient recovery of *Striga* seeds and qPCR efficiency in both soil types (clay, sand) (Fig. 2b). In the study by van Delft et al. (1997), where *Striga* seeds were manually counted, the flotation method had a recovery of up to 85%. Hence, our approach substantially improved *Striga* seed detection and provided a molecular confirmation of *Striga* seed presence.

Optimizing quantification of Striga seeds in agricultural soil

For accurate quantification of the *Striga* seedbank in naturally infested soils, standard curves are typically generated by using genomic DNA extracted from the weed seeds (Dongo et al. 2012; Aly et al. 2012, 2019). To be precise on gene copy number, we amplified and cloned the marker gene (StHe0GB1_93) into the pGEM®-T Easy vector to establish an absolute standard curve. Additionally, we established a second standard curve with genomic DNA extracted from *Striga* seeds introduced in 150 g D08 soil at six different densities. An initial number of gene copies (NGC) of 258129 single stranded (ss)-rpDNA μl^{-1} was calculated from the initial DNA concentration of the rpDNA (0.5 pg/µl) extracted from the transformed *E. coli* using Eq. 1 (Brankatschk et al. 2012). From the five point 10-fold serial dilution (0.5 pg/µl to 0.00005 pg/µl) of the purified rpDNA, mean *Cq* values of 16.51 and 30.82 were calculated for the highest (0.5 pg/µl) and lowest (0.00005 pg/µl) concentration, respectively corresponding to 258129 and 26 gene copies per μl^{-1} of ss-rpDNA. The absolute standard curve for StHe0GB1_93 is linear in the range tested (R^2 = 0.9959) with a slope of -3.5965 (Fig. 3a). From the slope, an amplification efficiency of 89.69% was determined for StHe0GB1_93. DNA concentration as low as 0.00005 pg/µl could be detected in the assay. This result revealed that our method is more sensitive than the recent study of Aly and coworkers (Aly et al. 2019), in which 0.001 ng/µl was the minimum DNA concentration that could be detected for genomic DNA of the parasitic weed *Orobanche cumana*. The high detection sensitivity obtained in our study with qPCR may be due to the use of plasmid DNA, which is devoid of other DNA and inhibitors from seed samples that can interfere in qPCR.

Following in the footsteps of the elegant study by Aly et al. (2019) to quantify *O. cumana* seeds in naturally infested soil samples, we plotted the seed number against the gDNA extracted from six *Striga* seed densities (1, 3, 9, 27, 81 and 243 seeds) introduced into 150 g of field soil D08. The obtained standard curve was then used to establish a relationship between the number of *Striga* seeds in a soil and the estimated copy number of the marker gene (StHe0GB1_93), calculated based on the *Cq* values of the different seedbank densities (Fig. 3b). The standard curve is linear in the range of *Striga* seed numbers tested (R^2 = 0.9942). Hence, this standard curve was then used for quantification of *Striga* seeds in naturally infested soil samples collected from sorghum growing fields in Ethiopia.

Striga seedbank density in naturally infested sorghum fields in Ethiopia

The method that was validated on artificially infested soils was used to detect and quantify *Striga* seeds in 48 naturally infested soil samples (referred to as E01 – E50, except soils E15 and E48) collected from sorghum field soils from different agroecological zones in Ethiopia and covering a trajectory along the sorghum belt of more than 1500 km (Fig. 4a). Following our new method described above, the results showed substantial variation in *Striga* seed density among the 48 Ethiopian soil samples (Fig. 4b and Table S3). The *Striga* seed densities ranged from 0 to 86 per 150 g of soil sample, with soil samples E22, E12 and E27 harboring the highest *Striga* seedbank densities of 86, 67 and 46 seeds per 150 g, respectively (Fig. 4b). *Striga* seeds were not detected in soil samples of 12 Ethiopian sorghum fields (E13, E16, E17, E19, E20, E21, E30, E33, E38, E40, E43, and E45) (Fig. 4b).

When looking into the geospatial distribution of the *Striga* seedbank in Ethiopian sorghum fields, most of the soils with relatively high *Striga* seed densities were collected from the Tigray region of Ethiopia. The majority of the samples that showed relatively low *Striga* seed densities were collected from sorghum growing areas of North Shewa. Here we would like to emphasize that the terms ‘relatively low’ and ‘relative high’ *Striga* seed densities are merely used to categorize the seedbank of our soil samples and this may not reflect the extent to which it poses an adverse effect on sorghum growth and yield. For example, if the seed density is presented per square meter of field soil with 300 kg top soil per square meter (considering only the top 20 cm and assuming a bulk density of agricultural soil of average 1.5 g/cm^3), then the lowest seed density detected (1 seed per 150 g of soil sample) still corresponds to approximately 2,000 *Striga* seeds per m^2. Translating the numbers shown in Fig. 4b to numbers that are relevant at field scale suggests the persistence of a high *Striga* seedbank in multiple fields in the sorghum belt of Ethiopia.

Relationship between Striga seedbank and Striga incidence

Determining the relationship between *Striga* seed densities and field infestation would be highly instrumental to predict the risk for crop losses in different agroecological conditions and to test the efficacy of specific management practices. At the same time, establishing causal relationships is difficult as these are highly dependent on the sorghum genotypes and management practices used by the farmers at the time of sampling and in future cultivations. Both linear and non-linear relationships

between the number of emerged *Striga* plants and the initial seedbank density were previously reported (Smith and Webb 1996; van Delft et al. 1997). Our analysis revealed a significant ($p \leq 0.0001$) positive correlation ($r = 0.561$) between the *Striga* seed density and the percentage of *Striga* emergence per square meter assessed in the same field where the soil samples were collected. Moreover, different regression analyses (linear, non-linear) revealed that the non-linear regression analysis provided the best relationship ($R^2 = 0.362$) between the number of *Striga* seeds per 150 g of soil and the number of emerged *Striga* seedlings counted per square meter of sorghum field (Fig. 5b). The asymptotic nature of this non-linear relationship appears to make biologically more sense than a linear relationship considering intraspecific competition for infection sites and/or outgrowth and emergence. Despite the overall positive correlation between these two parameters, however, some soil samples deviated to some extent from this relationship. For example, soil E04 showed high *Striga* incidence but low *Striga* seed density whereas soil E27 showed high seed density but low *Striga* incidence (Fig. 5b). The underlying mechanisms of this deviation are under investigation and can be due to soil physicochemical and/or microbiological attributes that act on the *Striga* seedbank or on *Striga* infection. A previous study also showed that even in fallow fields, one year after the last harvest, a decrease of 62% in the number of seeds was recorded for the top soil fraction; this was not the case for samples originating from below a depth of 10 cm, possibly reflecting the decrease in microbial activity with soil depth (van Delft et al. 1997). Soils E12 and E22 that showed high *Striga* seedbank density and high *Striga* incidence could be considered soils conducive for *Striga*, whereas soil E27 can be considered as a potential *Striga*-suppressive soil. Although, this regression analysis might not provide a conclusive means to categorize field soils as *Striga* conducive or suppressive, it can serve as a lead to further interrogate these soils for *Striga*-suppressive physicochemical or microbiological traits. Furthermore, *Striga* seeds were detected in soil samples collected from push-pull fields (E01, E07, E49 and E50, respectively having 1, 3, 12 and 1 *Striga* seeds per 150 g of soil) though no or low *Striga* incidence was observed during soil sampling (Fig. 4b). This result is in line with earlier observations of low *Striga* incidence in push pull fields but also suggests that *Striga* seeds may persist in push-pull soils that are assumed to diminish the seedbank of this parasitic weed. Whether the *Striga* seeds detected in these and other soils tested in this study are still viable remains to be determined. Hence, selecting marker genes that distinguish viable non-dormant from viable dormant seeds and designing primers that differentially amplify the genomic DNA/RNA extracted from these seeds is the next research priority to get an even more detailed insight into *Striga* seedbank dynamics.

Conclusions

In this study, we developed a high-throughput and robust molecular technique for the detection and quantification of the *Striga* seedbank in agricultural soils. This technique is a first important step to screen large numbers of samples to assess the impact of different intervention strategies on *Striga* seedbank dynamics and to unravel the impact of soil microbiological and physicochemical properties. The proof-of-principle experiment we performed by mixing known numbers of *Striga* seed in 150 g of two *Striga*-free Dutch agricultural soil samples showed that our procedure is efficient to detect and quantify a single *Striga* seed per 150 g of soil. The qPCR detection and quantification of *Striga* seeds in soils were

also further tested on soil samples collected from naturally infested sorghum fields and showed considerable variation in *Striga* seedbanks across the sorghum belt in Ethiopia. Correlation analysis also revealed a significant ($p \leq 0.0001$) positive correlation ($r = 0.561$) between the density of *Striga* seeds and the number of emerged *Striga* per square meter of sorghum field. The next challenge will be differentiating viable non-dormant and viable dormant seeds to further fine-tune the relationship between *Striga* seedbank dynamics and *Striga* incidence.

Materials And Methods

Soil sampling and study areas

The soil samples were collected from naturally *Striga* infested sorghum fields in Amhara (Kemise, North Shewa, South and North Wollo Zones) and Tigray (West, Central and South zones) regions of Ethiopia in October 2017 (Fig. 4a). For representative soil sampling, sorghum fields with four categories (zero, low, medium and high) of *Striga* field infestation were randomly selected. These categories were determined based on the number of emerged *Striga* plants counted for four quadrants of 1m *1m. Soil samples from the top layer (0–20 cm) around the root zone of sorghum plant in these quadrants were sampled separately and later combined together to form one composite sample per field. Utensils used for the sampling were washed with water and rinsed with 70% ethanol between successive samplings to avoid cross contamination of samples. In total, 48 composite soil samples covering a trajectory of more than 1500 km were collected from naturally *Striga* infested sorghum growing agro-ecological zones in Ethiopia. Among the soil samples, four soil samples from push-pull demonstration sorghum fields in North Shewa, Kemise and West Hararghae Zones of Ethiopia were included to investigate the effect of push-pull technologies on the *Striga* seedbank density in agricultural fields. Soil samples were brought to the lab in Holeta research centre, air dried and sieved through a 4-mm mesh sieve to remove stones and plant debris. Furthermore, seven *Striga*-free soil samples were collected from different parts of the Netherlands and used to investigate *Striga* DNA recovery and qPCR efficiency.

Selection of marker genes, primer design and specificity

From the parasitic plant genome project (http://ppgp.huck.psu.edu/) the *Striga* StHe0GB1 genome assembly was downloaded. We selected five genes (StHe0GB1_1, StHe0GB1_9, StHe0GB1_20, StHe0GB1_76 and StHe0GB1_93) with *Striga*-specific sequences as a putative marker gene for *Striga* seed detection and quantification. For these 5 genes, a total of 14 primer pairs were designed targeting the *Striga*-specific sequences. Using the NCBI primer blast web tool and nr-database, the specificity of the forward and reverse primers was validated *in silico*. To validate the efficacy of these primers experimentally, we extracted DNA from ~ 6000 *Striga* seeds, from 100 mg of *Striga*-free Dutch agricultural soil spiked with ~ 6000 *Striga* seeds and 100 mg of the soil sample without *Striga* seeds (control). Samples were ground manually with mortar and pestle in liquid nitrogen and kept at -80 ^{0}C until further use. The genomic DNA was extracted using DNeasy PowerSoil Kit (QIAGEN) according to the

manufacturer's instructions. The DNA quality and quantity were determined using a NanoDrop spectrophotometer.

PCR was carried out in 25 µl reaction volume using GoTaq hot start polymerase master mix (12.5 µl), primer mix (1 µl), templet DNA (0.5 µl) and water (11 µl) on a thermocycler equipped with heated lid. An initial denaturation for 2 minute at 95°C; 35 cycles with 30 sec at 95°C, 30 sec at 50°C, 30 sec at 72°C and a final elongation for 5 min at 72°C. The primer sets were also evaluated in qPCR. The qPCR mixes amounted to a total volume of 20 µl, consisting of 4 µl of the template DNA, 10 µl of SYBR Green, 1 µl of each forward and reverse primer (10 ppm), 2 µl of BSA (4 mg/ml) and 2 µl Sigma water. Two annealing temperatures (56 ^{0}C, 60 ^{0}C) were tested to assess the specificity of the primers. Bio-Rad qPCR machine was used with the following conditions: 3 minute at 95 ^{0}C followed by 35 amplification cycles of 5 sec at 95 ^{0}C, 15 sec at 56 ^{0}C or 60 ^{0}C, and 25 sec for the final elongation at 72 ^{0}C.

Establishment of standard curves

We developed a recombinant plasmid containing the marker gene (StHe0GB1_93) for establishing absolute standard curves aiming to establish the relationship between *Cq* values and gene copy numbers. Furthermore, we established a second standard curve using the genomic DNA extracted from six densities of *Striga* seeds introduced in 150 g of *Striga*-free agricultural soil (D08) to establish relationship between seed number and *Cq* value. The combined use of the two standard curves enabled us to establish relationship between *Cq* value and seed number when analyzing the naturally infested Ethiopian soil samples.

Recombinant plasmid DNA-based standard curve

The marker gene was first cloned in pGEM®-T Easy vector. Then, the vector containing the marker gene was transformed into the *E. coli* and positive colonies were identified using colony PCR and cultured in LB medium. The rpDNA was isolated and purified and the concentration of the rpDNA was determined. A five point 10 fold serial dilutions (0.5 pg/µl to 0.00005 pg/µl) of the purified rpDNA was subjected for qPCR assay to establish the relationship between *Cq* values and the calculated gene copies of the marker gene.

An initial number of gene copies µl^{-1} of a single strand (ss)-rpDNA (NGC ss-rpDNA) was calculated from the initial DNA concentration of the rpDNA (0.5 pg/µl), the length of the plasmid containing the target gene (3535 bp), the number of targets per DNA fragment (n_{target} [2 copies]), the Avogadro constant (6.022 * 10^{23} bp mol^{-1}), and the average weight of a double-stranded base pair (660 g mol^{-1} = 6.6 * 10^{11} ng mol^{-1}) (Eq. 1) (Brankatschk et al. 2012).

$$\mathrm{NGCss-rpDNA} = 2^{*}\frac{6.02\mathrm{x}1023\left(\frac{\mathrm{copy}}{\mathrm{mol}}\right)^{*}\mathrm{DNAamount\,(g)}}{\mathrm{DNAlength\,(bp)}^{*}660\left(\frac{\frac{g}{mol}}{dp}\right)}$$

1

The linear regression of the *Cq* value of each dilution versus their corresponding $\log_{10}$ gene copy ($N_{0\ Sample}$) was used to calculate the slop (b) and intercept (a) of the standard curve (Eq. 2) (Brankatschk et al. 2012). The amplification efficiency (*E*) was calculated from the slope of the standard curve using Eq. 3.

$$Cq\text{sample} = \text{a} + \text{b}^{*}\log\left(\text{Nosample}\right)$$

2

$$E = 10^{\left(-\frac{1}{b}\right)}$$

3

Striga seedbank density-based standard curve

Another standard curve was also established from the genomic DNA extracted above from soil sample (D08) mixed with six densities of *Striga* seeds (1, 3, 9, 27, 81 and 243 seeds). The gene copies of each density of the seeds were calculated from the average *Cq* value by using the regression formula generated above from rpDNA gene copies and the corresponding average *Cq* value. Then, the relationship between number of *Striga* seeds and the estimated gene copies was generated. Hence, this regression equation is used to convert the detected DNA of *Striga* seeds by qPCR to quantified number of seeds in naturally infested soils.

The above Eq. 2 was also rearranged and taken the reverse of *Log* of both sides to calculate the number of gene copies *Striga* seed DNA (NGC ssDNA) extracted from different densities of seed introduced in *Striga* free Dutch soil D08 as indicated in (Gallup 2011).

$$\text{NGCssDNA} = (10)^{\left(\frac{\text{CqssDNA}-\text{arpDNA}}{\text{brpDNA}}\right)} = (\text{EAMPrpDNA})^{(\text{arpDNA}-\text{CqssDNA})}$$

4

The number of gene copies of *Striga* seed DNA extracted from naturally infested field soils (NGC ss DNA soil) were calculated per the Eq. 5 as described in (Gallup 2011).

$$\text{NGC ss DNA soil} = (\text{EAMPssDNA})^{(\text{arpDNA}\log_{(\text{EAMP rpDNA}}\text{EAMPssDNA})-\text{CqssDNAsoil})}$$

5

The standard curve that established a relationship between *Striga* seed number and gene copy created above from artificially contaminated soil sample with different densities of *Striga* seeds was used to extrapolate the number of *Striga* seeds in naturally infested soil samples from the average *Cq* value-gene copy relationship.

Striga **seed separation from the soil matrix**

To reduce the influence of soil physicochemical properties on DNA recovery and qPCR efficiency, the methods that separate *Striga* seeds from the bulk soil to enhance the detection and accurate quantification of *Striga* seeds in soil was investigated. As a first step, using two *Striga* free Dutch agricultural soil samples (D08 and D17) that have contrasting physicochemical properties, we evaluated two seed-soil separation methods: 1) combined washing and sieving of the *Striga*-soil mixture and 2) density-dependent K_2CO_3 separation (flotation) followed by sieving. In the first method, 150 g of the soil samples were washed by gently mixing on five sieves arranged sequentially in successive order of pore size (300, 200, 180, 150 and 100 µm) and the soil samples retained on the last three smaller sieves were collected together. In the second method, density-based extraction by K_2CO_3 solution followed by size-dependent separation by sieving was performed to separate *Striga* seeds and some small and lighter soil particles from other heavy and larger organic debris and soil particles. The samples were divided into three 250 ml centrifuge bottles with 50 g of soil sample suspended in 150 ml of 5.5 M K_2C0_3 solution. Then, the soil samples were dispersed by shaking at 250 rpm for 15 min followed by sonication for 15 min by using Bransonic® Ultrasonic Cleaner sonicator containing a RF frequency of 47 KHZ ± 6%. The dispersed soil samples were centrifuged at 5,000 X *g* for 5 min at room temperature by using high speed centrifuge. The *Striga* seeds and other lighter organic matter floated on the top of the supernatant whereas the majority of the soil particles settled at the bottom. The supernatants from the three bottles of the same sample were collected into 1000 ml bottles and the process was repeated a second time to ensure full recovery of all the *Striga* seeds. Then, size-dependent separation of the *Striga* seeds, smaller soil particles and organic debris from larger particles were performed by using two meshes (pore sizes 425 and 75 µm) arranged in successive order. The *Striga* seeds and smaller particles retained on 75 µm were dried at 35 ^{0}C for 48 hours and collected for further grinding and DNA extraction.

The efficiency of the density and size-dependent method described above was assessed in proof-of-principle experiments involving introduction of known numbers of *S. hermonthica* seeds (0, 1, 3, 9, 27, 81 and 243 seeds) into 150 g of two soil samples (D08 and D17) with contrasting soil physicochemical properties. Then, the *Striga* seeds were re-separated from 150 g of soil samples and were ground manually by mortar and pestle under liquid nitrogen. The genomic DNA was also extracted using DNAeasy PowerSoil Kit (QIAGEN) according to the manufacturer's instructions as indicated above. Then, qPCR was carried out to assess the effectiveness of the recovery of the *Striga* seeds from the soil matrix.

Declarations

Acknowledgments

This research work was supported by the Bill and Melinda Gates Foundation through grant number OPP1082853: PROMISE 'Promoting Root Microbes for Integrated *Striga* Eradication' project via Ethiopian Institute of Agricultural Research and Netherlands Institute of Ecology.

References

1. Abate M, Mekbib F, Hussien T et al (2014) Assessment of genetic diversity in sorghum (*Sorghum bicolor* (L.) Moench) for reactions to *Striga hermonthica* (Del.) Benth. Aust J Crop Sci 8:1248–1256
2. Abunyewa AA, Padi FK (2003) Changes in soil fertility and *Striga hermonthica* prevalence associated with legume and cereal cultivation in the Sudan savannah zone of Ghana. Land Degrad Dev 14:335–343. https://doi.org/10.1002/ldr.555
3. Aly R, Bari VK, Londner A et al (2019) Development of specific molecular markers to distinguish and quantify broomrape species in a soil sample from infected field. Eur J Plant Pathol 155:1367–1371. https://doi.org/10.1007/s10658-019-01841-9
4. Aly R, Eizenberg H, Kocherman M et al (2012) Use of ITS nuclear sequences from *Phelipanche aegyptiaca* as a direct tool to detect single seeds of broomrape species in the soil. Eur J Plant Pathol 133:523–526. https://doi.org/10.1007/s10658-012-9965-9
5. Badu-Apraku B, Akinwale RO (2011) Cultivar evaluation and trait analysis of tropical early maturing maize under *Striga*-infested and Striga-free environments. Field Crops Research 121:186–194. doi:10.1016/j.fcr.2010.12.011
6. Bayu W, Binor S, Admassu L (2005) Tolerance of sorghum landraces and varieties to *Striga* (*Striga hermonthica*) infestation in Ethiopia. Acta Agron Hung 49:343–349. https://doi.org/10.1556/AAgr.49.2001.4.5
7. Brankatschk R, Bodenhausen N, Zeyer J, Bürgmann H (2012) Simple absolute quantification method correcting for quantitative PCR Efficiency variations for microbial community samples. Appl Environ Microbiol 78:4481–4489. https://doi.org/10.1128/AEM.07878-11
8. Dongo A, Leflon M, Simier P, Delavault P (2012) Development of a high-throughput real-time quantitative PCR method to detect and quantify contaminating seeds of Phelipanche ramosa and Orobanche cumana in crop seed lots. Weed Res 52:34–41. https://doi.org/10.1111/j.1365-3180.2011.00891.x
9. Franke AC, Ellis-Jones J, Tarawali G et al (2006) Evaluating and scaling-up integrated *Striga hermonthica* control technologies among farmers in northern Nigeria. Crop Prot 25:868–878. https://doi.org/10.1016/j.cropro.2005.11.014
10. Frostegård Å, Courtois S, Ramisse V et al (1999) Quantification of bias related to the extraction of DNA Directly from soils. Appl Environ Microbiol 65:5409–5420
11. Gallup JM (2011) qPCR inhibition and amplification of difficult templates. In: Kennedy S, Oswald N (eds) PCR trouble shooting and optimization: the essential guide. Caister Academic Press, Norfolk
12. Hearne SJ (2009) Control–the Striga conundrum. Pest Manag Sci 65:603–614. https://doi.org/10.1002/ps.1735
13. Ophel-Keller K, McKay A, Hartley D et al (2008) Development of a routine DNA-based testing service for soilborne diseases in Australia. Australas Plant Pathol 37:243–253. https://doi.org/10.1071/AP08029
14. Oswald A, Ransom JK (2001) *Striga* control and improved farm productivity using crop rotation. Crop Prot 20:113–120. https://doi.org/10.1016/S0261-2194(00)00063-6

15. Parker C (2009) Observations on the current status of *Orobanche* and *Striga* problems worldwide. Pest Manag Sci 65:453–459. https://doi.org/10.1002/ps.1713
16. Parker C (2012) Parasitic Weeds: A World Challenge. Weed Sci 60:269–276. https://doi.org/10.1614/WS-D-11-00068.1
17. Prider J, Keller KO, Mckay A (2013) Molecular diagnosis of parasite seed banks. In: Joel DM et al (eds) Parasitic Orobanchaceae. Springer, Berlin Heidelberg, pp 357–368. DOI 10.1007/978-3-642-38146-1$420
18. Sauerborn J, Kranz B, Mercer-Quarshie H (2003) Organic amendments mitigate heterotrophic weed infestation in savannah agriculture. Appl Soil Ecol 23:181–186. https://doi.org/10.1016/S0929-1393(03)00023-4
19. Scholes JD, Press MC (2008) *Striga* infestation of cereal crops – an unsolved problem in resource limited agriculture. Curr Opin Plant Biol 11:180–186. https://doi.org/10.1016/j.pbi.2008.02.004
20. Schulz S, Hussaini MA, Kling JG et al (2003) Evaluation of integrated *Striga hermonthica* control technologies under farmer management. Exp Agric 39:99–108. https://doi.org/10.1017/S0014479702001084
21. Smith MC, Webb M (1996) Estimation of the seedbank of *Striga* spp. (Scrophulariaceae) in Malian fields and the implications for a model of biocontrol of *Striga hermonthica*. Weed Res 36:85–92. https://doi.org/10.1111/j.1365-3180.1996.tb01804.x
22. Taparia T, Krijger M, Hodgetts J et al (2020) Six Multiplex TaqManTM-qPCR Assays for quantitative diagnostics of *Pseudomonas* Species, causative of bacterial blotch diseases of mushrooms. Front Microbiol 11:https://doi.org/10.3389/fmicb.2020.00989
23. Tesso T, Gutema Z, Deressa A, Ejeta G (2007) An integrated S*triga* management option offers effective control of *Striga* in Ethiopia. In: Ejeta G, Gressel J (eds) Integrating New Technologies for *Striga* Control: Towards Ending the Witch-hunt. World scientific publishing co., Singapore, pp 199–212
24. van Delft G-J, Graves JD, Fitter AH, Pruiksma M (1997) Spatial distribution and population dynamics of *Striga hermonthica* seeds in naturally infested farm soils. Plant Soil 195:1–15. https://doi.org/10.1023/A:1004214015281
25. van Mourik T, Bianchi FJJA, Werf WVD, Stomph TJ (2008) Long-term management of *Striga hermonthica*: strategy evaluation with a spatio-temporal population model. Weed Res 48:329–339. https://doi.org/10.1111/j.1365-3180.2008.00638.x
26. van Mourik TA (2007) *Striga hermonthica* seed bank dynamics: process quantification and modelling. Phd, S.n
27. Westerman PR, van Ast A, Stomph TJ, van der Werf W (2007) Long-term management of the parasitic weed *Striga hermonthica*: Strategy evaluation with a population model. Crop Prot 26:219–227. https://doi.org/10.1016/j.cropro.2006.01.017
28. Westwood JH, dePamphilis CW, Das M et al (2012) The parasitic plant genome project: new tools for understanding the biology of *Orobanche* and *Striga*. Weed Sci 60:295–306.

https://doi.org/10.1614/WS-D-11-00113.1

Figures

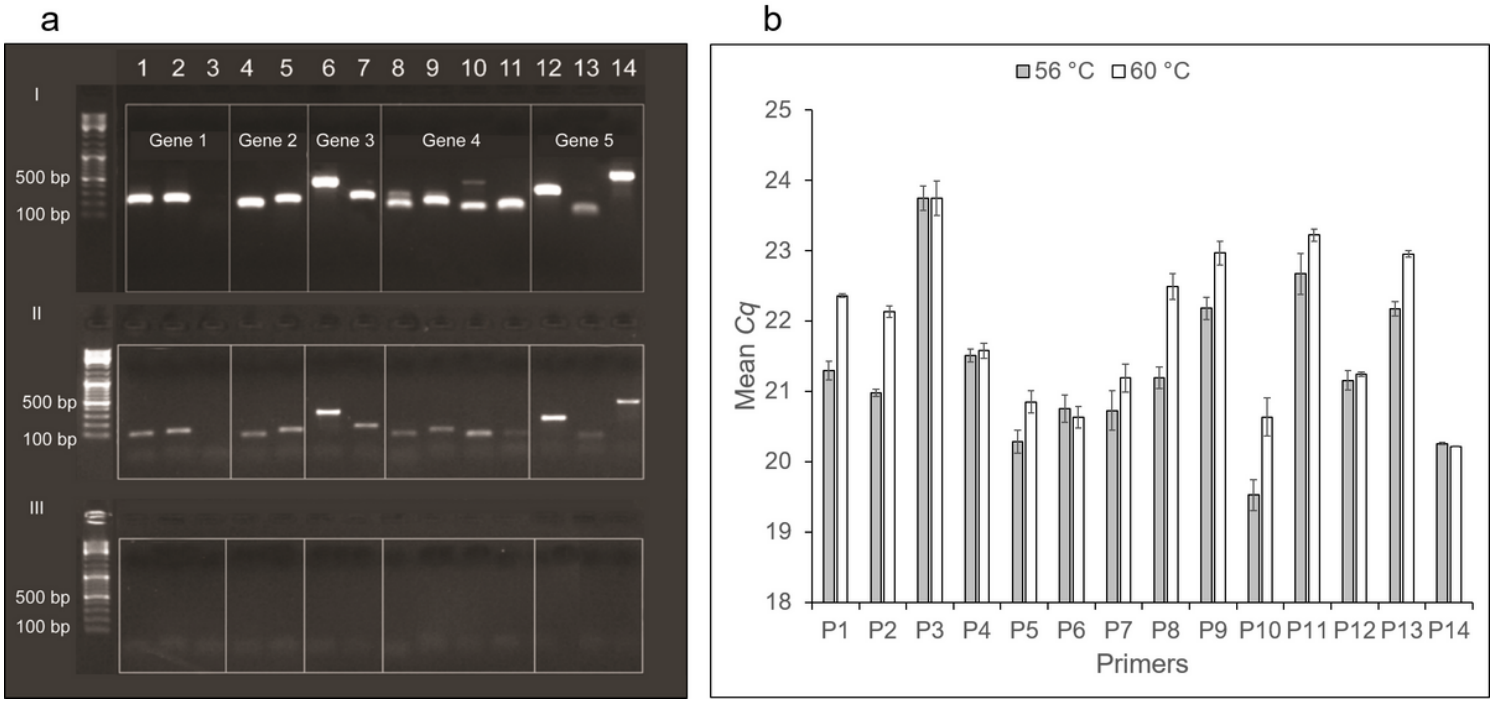

Figure 1

Striga marker gene selection. a) Gel electrophoresis of the PCR amplification products of five Striga hermonthica marker genes. A total of fourteen primer sets were tested with 3 sets for gene 1 (StHe0GB1_1), 2 sets for gene 2 (StHe0GB1_9), 2 sets for gene 3 (StHe0GB1_2), 4 sets for gene 4 (StHe0GB1_76) and 3 sets for gene 5 (StHe0GB1_93)). The template DNA was DNA extracted from (I) 50 mg Striga seeds, (II) 50 mg Striga seeds mixed in 100 mg of Dutch agricultural soil, and (III) 100 mg of Dutch agricultural soil (no Striga seed added; control). For the fourteen primer sets, the sizes of the predicted PCR products are 145, 161, 170, 115, 157, 111, 200, 112, 154, 101, 100, 276, 70 and 520 base pairs (see supplementary table 1). b) Mean Cq values of the qPCR analysis with the 14 primer sets using DNA extracted from Dutch agricultural soil mixed with S. hermonthica seeds. The qPCR analysis was tested at two different annealing temperatures (56 oC, 60 oC). Mean Cq values (± SE) of 3 biological replicates (with 2 technical replicates per biological replicate) are shown.

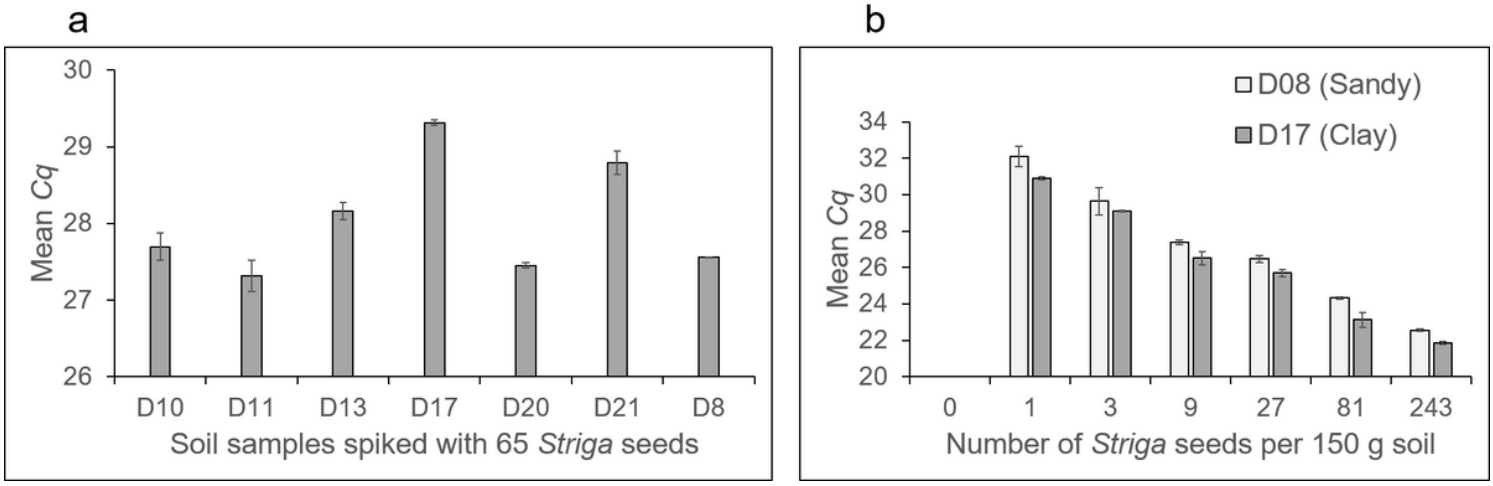

Figure 2

Influence of soil type on recovery of Striga seeds. a) qPCR detection of 65 Striga hermonthica seeds mixed into seven physicochemically different Dutch agricultural soils (D08, D10, D11, D13, D20, D21, D17). After mixing the seeds into these soils, total DNA was extracted and subjected to qPCR with primer set 14 (see figure 1B). b) qPCR detection of different Striga seed densities introduced into two physicochemically distinct Dutch agricultural soils (D08, D17). In contrast to the procedure used in panel A, soils containing the Striga seeds were first treated with K2CO3 for size-dependent separation of the Striga seeds from the soil matrix prior to DNA extraction. For both experiments, the mean Cq values (± SE) are shown for three biological replications and two technical replications per biological replication.

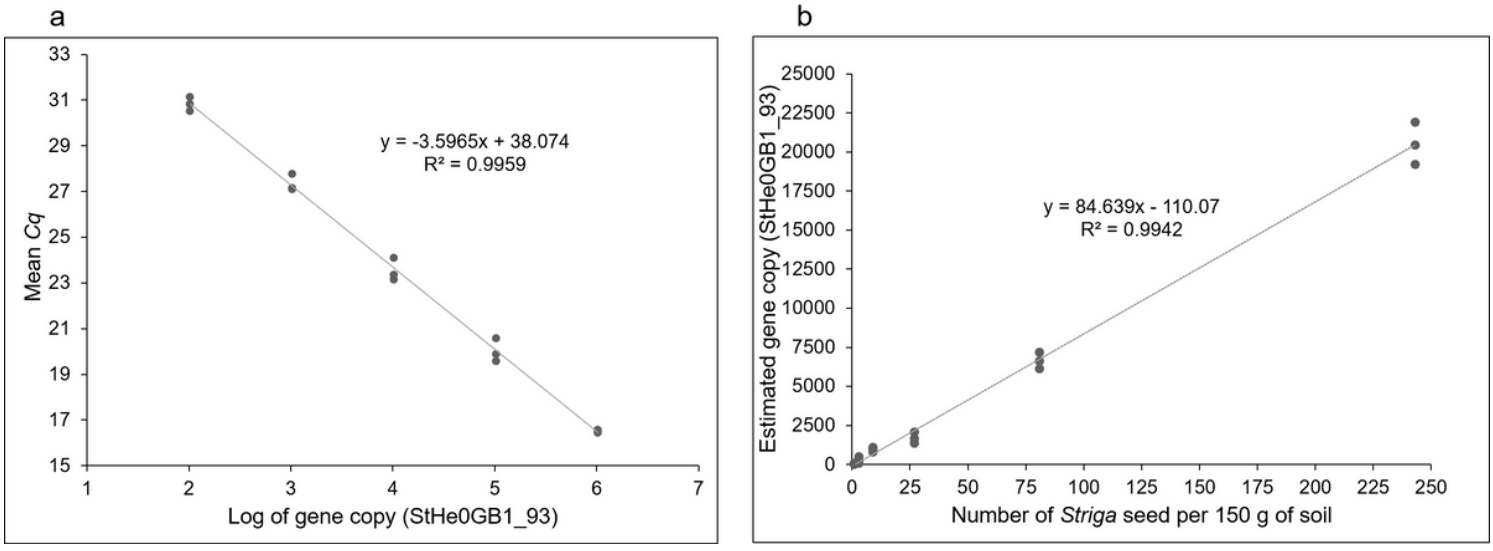

Figure 3

Standard curves to quantify Striga seeds in naturally infested soils. a) Relationship between the Cq values obtained in qPCR analysis of plasmid DNA containing the Striga marker gene StHe0GB1_93 (gene 5, Figure 1A) and the logarithm of the gene copy number. For each log gene copy number, 3 replicates were used in qPCR; b) relationship between different Striga hermonthica seed densities mixed into

agricultural soil and the estimated gene copy number. For each Striga seed density, three biological replications and two technical replications per biological replication were used.

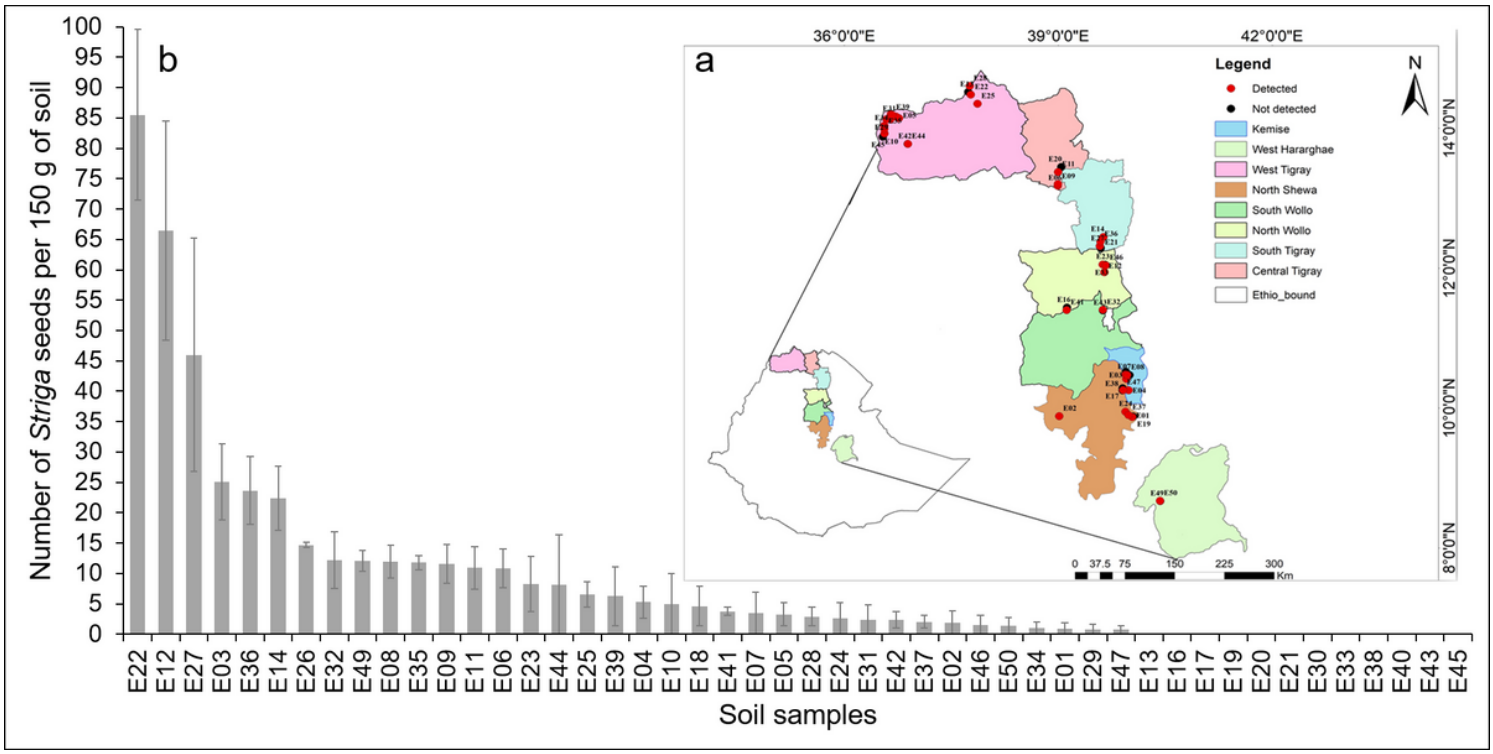

Figure 4

Striga field infestation in sorghum growing fields of Ethiopia. a) Map of Ethiopia showing the different sorghum growing agroecological zones and the agricultural field sites where a total of 48 soil samples (E numbers) were collected. The geospatial distribution of S. hermonthica seedbank in different sorghum growing agroecological zones of Ethiopia was also depicted by using the red dots for the sites where S. hermonthica seeds were detected and quantified in 150 g of soil sample whereas the black dots represent sites where S. hermonthica seeds were not detected in 150 g of soil sample. b) Number of Striga seeds detected by qPCR in 150 g of soil collected from each of these naturally infested field sites. Mean values (± SE) of 3 biological replicates (with two technical replicates per biological replicate) are shown.

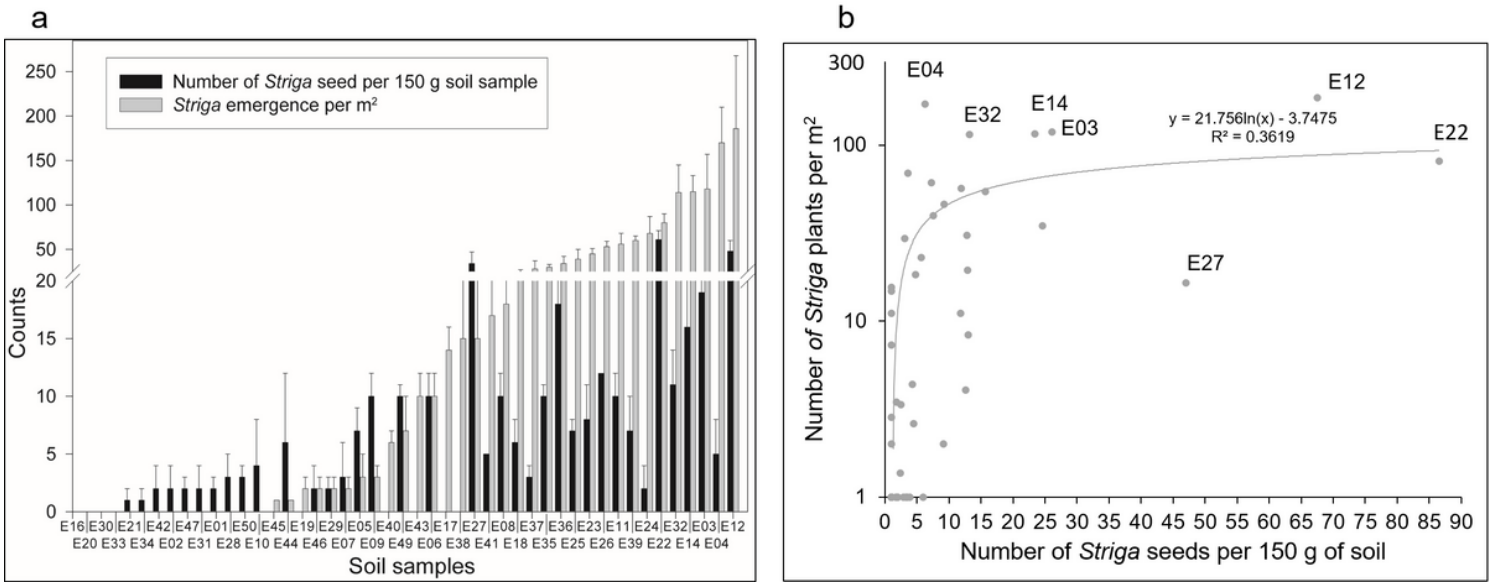

Figure 5

Relationship between Striga seed density and emerged Striga plants for 48 sorghum fields in Ethiopia (map shown in Figure 4a). For each sorghum field, the Striga seed density was quantified per 150 g of soil for 3 biological replicates as depicted in Figure 4b. Striga emergence was counted from four randomly chosen spots per field site and the number of emerged Striga plants was normalized by the number of sorghum plants counted per m2. a) Striga emergence and Striga seed densities of 48 naturally infested sorghum fields in Ethiopia. b) Non-linear relationship between the number of emerged Striga plants per m2 and the number of Striga seeds detected per 150 g of soil sample. Striga emergence on the Y-axis is shown on a log2 scale.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFiguresBentietal.pdf
- SupplementaryTablesBentietal.pdf