Meta-barcoding for assessment of risks posed by genetically modified crops to farmland arthropods

By

Trace Akankunda

A thesis submitted for the partial fulfilment of the requirements of the Master of Biotechnology (Plant Biotechnology)

The University of Adelaide

Faculty of Sciences

School of Agriculture, Food & Wine

Waite Campus

2014

Declaration

I declare that this thesis is a record of original work and contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text.

Akankunda Trace

Table of Contents

Preface

This research has been performed during the past 10 months as part of the requirements of the Master of Biotechnology (Plant Biotechnology) degree. In accordance with the requirements of the program, the research is presented in the format of a manuscript for submission to a peer-reviewed scientific journal. I have chosen to follow the format of the PLos One journal. My co-authors for the manuscript are Prof. Mike Keller, Dr. Otim Michael and Dr. Adam Croxford. Prof. Keller advised on the scope of the research and the experimental design. He advised about the analysis of results and reviewed the draft and final manuscript. Dr. Michael provided field supervision and consultation on the practical aspects of conducting an insect collection survey. Dr. Croxford provided technical assistance and direction in the molecular laboratory in order to generate and analyse both the reference and meta-barcoding sequence datasets used in this experiment. The manuscript in this thesis is intended as the first draft of a manuscript for future publication. The word count for the manuscript is 7500. Appendix A contains an alignment output of *Ceratitis anonae* and *Ceratitis rosa* extracted from an NCBI output. Appendix B of this thesis contains a summary table of the output from the Miseq sequencing run showing the quality of the run.

Abstract

The rate of adoption of genetically modified (GM) crops continues to grow at unprecedented rates 19 years after their first commercialisation. As global coverage of GM crops increases, concerns about their potential effects on the environment and specifically agro-ecosystem health intensify. To address these concerns, researchers have called for increased monitoring of agro-ecosystems to detect unforeseen adverse effects of GM crops. To date, only Europe has a statutory requirement for developers of GM products to conduct post market environmental monitoring (PMEM) in order to assess potential risks associated with their products. This might be due to lack of robust and cost effective methods for conducting PMEM. Here we propose the use of a modified meta-barcoding pipeline on an Illumina MiSeq platform as a comprehensive and cost effective approach for conducting PMEM on farmland arthropod communities. We test the method's capacity to generate baseline data on a selection of indicator arthropod groups following guidelines issued by the European Food Safety Authority (EFSA). We use arthropod communities of coffee plantations in the south and south-western regions of Uganda as an exemplar for the approach. We modify the sample preparation steps of the meta-barcoding pipeline to reduce sequencing cost and successfully adapt the MiSeq Reporter program to classify arthropods using COI sequence reads produced by the MiSeq. We compile baseline data on the diversity and distribution s of six generalist predators, two parasitoids, two pollinators, four common pests and three herbivores of the coffee crop system using incidence counts. We demonstrate the method's capacity to monitor arthropod communities at the genus and species level and discuss the application of the baseline data collected for GM risk assessment.

1. Introduction

Genetically modified (GM) crops have recorded the highest adoption rate among modern agricultural technologies, registering a more than 100% increase in global hectarage from 1996 when they were first commercialised (James et al., 2013). The report also announced that developing countries are now the major growers of GM crops. With the advent of GM crops, governments around the globe have come under enormous pressure from the public and civil society groups to establish proper regulatory frameworks capable of assessing the risks associated with GM crops to human health and the environment.

During risk assessment, country specific protection goals are identified by the regulatory bodies and the risks posed to them by GM crops are determined (EFSA GMO Panel 2010). Among the key environmental protection goals widely recognised by GM regulators are the farmland arthropod communities, chiefly due to their economic importance in agro-ecosystems (EFSA GMO panel 2010). Apart from being the main crop pollinators, some arthropods provide bio-control services as pest predators while others such as herbivores drive the nutrient recycling processes on farms (Isaacs et al., 2008). However, they are also the main threat to many crop systems and pose significant risk to agriculture as pests (Ishaaya et al., 2009).

In determining the potential risks associated with GM crops on farmland arthropod communities, both direct and indirect effects must be assessed (EFSA GMO Panel 2011). The direct effects of GM crops result from exposure of susceptible beneficial farmland arthropods to lethal quantities of the transgenic products. The routes of exposure of the transgenic products to beneficial arthropods depend on their ecological functions. Pollinators could be exposed via nectar (Nzeduru et al., 2012), detritivores via GM plant waste while predators and parasitoids via prey (Kapuscinski et al., 2007).

Additionally, the effects of GM crops on farmland arthropods can be indirect (EFSA GMO Panel 2006). Indirect harmful effects are secondary effects of the transgenic crops on arthropod communities that might in turn reverse the benefits accrued from adopting the GM crops (Tabashnik et al., 2013). These include the emergence of secondary pests and resistant species of the target pests for which the GM crops are developed (Baptiste et al., 2010). Surveys conducted on established Bt (Bacillus thuringiensis toxin) crop systems in China, USA and India have reported several cases of the rising pest status among previously minor pests, positively correlated with increase in regional adoption of Bt cotton (Zhao et al., 2011). GM crops have also been reported to accelerate field evolved pest resistance through natural selection for rare pest species that could aggravate crop damage (Tabashnik et al., 2013). A review of 77 studies on field evolved pest resistance to Bt crops grown across five continents confirmed five Bt resistant pests, namely; *Busseola fusca* (African stem borer), *Diabrotica virgifera* (Western corn rootworm), *Helocoverpa zea* (corn earworm), *Pectinophora gossypiella* (pink bollworm) *and Spodoptera frugiperda* (fall armyworm) that emerged after large scale adoption of Bt crops (Tabashnik et al., 2013).

Given the potential hazards resulting from both direct and indirect effects of GM crops, researchers have emphasised the need to routinely monitor agro-ecosystems before and after adoption of GM crops to detect changes in the natural structure of farmland arthropod communities that might cause unforeseen harms (Baptiste et al., 2010; Lu et al., 2012; Tabashnik et al., 2013). Monitoring farmland arthropod communities is necessary to allow for timely response to adverse effects once detected. This is not only important to guarantee profitability for farmers that adopt GM crops but also to increase the lifespan of the GM products developed.

Despite the importance of agro-ecosystem monitoring, to date only Europe has a statutory requirement that GM product developers submit a comprehensive Post Market Environment

Monitoring (PMEM) plan alongside their applications for deliberate release of GM crops (EFSA GMO Panel 2010). The European Food Safety Association (EFSA) directives, 2001/18/EC and EU Regulation 1829/2003 on PMEM of GM crops require that product developers conduct case specific and general surveillance on recipient agro-ecosystems (EFSA GMO Panel 2006). Case specific monitoring is conducted to confirm hypotheses made on potential risks associated with the GM crop during the risk assessment processes whereas, general surveillance is conducted to detect unforeseen adverse effects and is continuous until such a time when the GM crop gains a history of safe use (EFSA GMO Panel 2011). In the PMEM guidelines issued by the EFSA, ecological functional groups of arthropods that could be used as indicators of agro-ecosystem health are identified together with their measurement endpoints and proposed methods of assessment. The ecological functional groups include herbivores, detritivores, saprophytes, pollinators, predators and parasitoids. The measurement endpoints identified include, abundance, population change, growth, development, change in host range and decrease of natural pest regulation mechanisms. While, the methods suggested for collecting data on the measurement endpoint, include the use of questionnaires to record observations made by farmers of GM crops, use of issue alerts from farmers, use of existing environmental surveillance networks (EES) and searching through published literature for reports on GM related hazard incidences (EFSA Panel on GMO 2011).

However, the methods suggested are not effective for monitoring farmland arthropod communities especially in developing countries which are currently the major consumers of GM products (James et al., 2013). This is because very few farmers can accurately identify arthropods, later on observe patterns in their communities especially in developing countries where their ecosystem services are hardly appreciated (Munyuli et al., 2011). Given that the existing environmental monitoring networks were not specifically designed to monitor GM crops, the data they collect might not be informative for the risk assessment process (EFSA

GMO Panel 2010). In addition, monitoring systems are expensive to run and are therefore not well established in developing countries (Aheto et al., 2013). Finally, published scientific literature on the effect of GM crops on farmland arthropods is hard to access especially in developing countries where research is poorly funded (Obonyo et al., 2011), and so renders the method ineffective.

Here we propose, a more effective method of monitoring farmland arthropods by means of meta-barcoding using the universal 658bp Cytochrome Oxidase (COI) DNA marker (Yu et al., 2012). Meta-barcoding is a bio-monitoring approach that depends on DNA based taxonomic identification using high-throughput DNA sequencing on Next generation sequencing (NGS) platforms (Baird et al., 2012; Yu et al., 2012). The method has become widely recommended for monitoring arthropod communities precisely because it generates many times more taxonomically comprehensive datasets, in a short time and with less involvement of taxonomic experts (Ji et al., 2013). Up until 2014, the meta-barcoding pipeline was run on the 454 pyrosequencing platform (Yu et al., 2012) and later on the Hiseq Illumina platform (Zhou et al 2013). These platforms presented several challenges including a high overall cost of conducting a sequencing run, poor taxonomic resolution due to short sequence read length and delays in analysis of the sequence output due their dependence on sophisticated bioinformatics packages (Yu et al., 2012; Zhou et al., 2013). More recently, the meta-barcoding pipeline has been conducted on a MiSeq platform (Gibson et al., 2014), which offers significantly lower sequencing costs, and now with the release of a new Illumina MiSeq 600-cycle version 3 kit, is capable of generating a 600bp read length (Glenn et al., 2014). The platform also comes installed with a windows based MiSeq Reporter program (MSR) that has been customised to perform taxonomic assignment for bacteria using the Greengenes16srRNA database. To our best of knowledge, no publication has made sofar about the capacity of a meta-barcoding pipeline

conducted on a MiSeq platform to generate baseline data on farmland arthropod communities that is relevant for GM risk assessment.

In this study, we modify the meta-barcoding pipeline further, to eliminate the fragmentation step of the PCR amplicons during library preparation and attempt to analyse the MiSeq reads generated using the MSR program. We hope that by modifying the meta-barcoding pipeline, we can further lower the overall cost of processing bulk arthropod samples and significantly reduce the expertise and time required to analyse the MiSeq output. We attempt to generate baseline data on functional arthropods groups that inform the risk assessment process as suggested by the EFSA and use arthropod communities of the coffee crop system in Uganda as an exemplar for the approach. The coffee crop system in Uganda was chosen because tropical coffee systems are known to support a rich diversity of arthropods.

2. Methodology

2.1.Sampling sites and sampling design

Insects were collected from 48 plantations selected from four major coffee growing districts of Uganda: Masaka and Rakai in southern Uganda and Mitoma and Mbarara in south western Uganda. In each district, three subcounties were selected and collections made from coffee plantations in four parishes within each subcounty. Only medium to large scale plantations of greater than 1.5ha were selected for sampling because they were likely to support higher populations of arthropods compared to small scale plantations (Sandra et al., 2006). All necessary permits for conducting the survey were obtained from District Agricultural Officers of the respective districts. Grid references for the districts sampled and their distances from Kampala are presented in Table 1.

Table 1. The districts that were sampled during the months of May and June 2014, their respective distances from the capital city of Uganda (Kampala), latitude, longitude and altitude.

District	Distance from Kampala (km)	Latitude °S	Longitude ° E	Altitude (m)
Masaka	137	0.3400	31.7400	1,258
Rakai	225	0.7167	31.4000	1,176
Shema	300	0.7000	30.3000	1,500
Mitoma	346	0.5500	30.2000	1,545

Altitude values were obtained from www.elevationmap.net .

The collections were made in two replicates, at the start of May and at the end of June 2014. Arthropods were collected using a garden vacuum sampler (Grootaert et al., 2010), and occasionally important pests identified by farmers were hand captured since the vacuum sampler could not trap them. In order to standardise the sampling effort, collections were made by walking along a 30 metre transect in four minutes. Three transects were marked across each plantation and arthropods only collected from coffee trees avoiding any crops that were intercropped especially banana trees. The arthropods captured from each plantation were bagged,

labelled and sorted into common arthroods, large and small arhropods and a combination of Dipterans and Hymenopterans. A total of 288 bulk samples from 48 collection sites (replicated) were labelled and stored in 50 ml falcon tubes containing 70% ethanol. Photographs were taken of the common insects identified from the bulks using a camera mounted on a microscope.

2.2. DNA extraction for the reference samples

A total of 91 reference samples (37 common arthropods and a selection of 54 others) were isolated from their bulk samples, and legs, heads, antennae or wings were used as starting tissue for extracting DNA from each insect separately. The remainder of the insects' bodies were kept for inclusion in the bulk DNA extracts of their respective collection sites. DNA was extracted using the Isolate II Genomic DNA extraction kit (Bioline, Australia). Tissue was ground using acid-washed sand and micro-pestles in 1.5 µL eppendorf tubes, with an extra mechanical disruption step using sterile beads on a tissue lyser for 30 seconds. The micro-pestles were kept in ethanol, washed and rinsed with double distilled water after each extraction and soaked in bleaching reagent overnight to reduce contamination between samples. Following the manufacturer's protocol, the samples were incubated in lysis buffer and proteinase K at 56° C overnight. Samples were then treated with 95% ethanol and transferred to binding columns. They were washed twice by centrifugation and finally collected by a one-step elution process in 60µl of elution buffer. All reagents were used as supplied by the manufacturer except for the absolute ethanol and the wash buffer (GW2) which was provided as a concentrate and had to be reconstituted in 95% ethanol before use. DNA quantity and quality was assessed using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) and stored at -80 °C.

2.3. Bulk DNA extraction

Extractions were made from individual bulk insect samples collected from the coffee plantations. The starting tissue used was based on the size of the insect relative to the insects forming the bulk. In order to standardise the starting tissue for extraction, insects smaller than a honey bee

(*Apis sp.*) were used as whole bodies while two legs were used for those larger than a honey bee, as suggested by Ji (2013). The extraction process followed steps similar to that used for extracting the reference samples (described above) using the Isolate II genomic DNA extraction kit. The mechanical disruption process for the bulk extractions was longer than that for reference samples and elution was carried out in a two-step process in 100µl of elution buffer. Samples were quantified and stored at -80 °C. All extractions were made in the molecular laboratory of the Biosciences unit in the National Crop Resources Research Institute, Namulonge-Uganda, and shipped to Australia for sequencing.

2.4.PCR amplification and product clean up

A two-step PCR amplification strategy (Bell et al., 2011) using COI barcoding primers was used to process the DNA from the reference samples and bulk extractions made. The first PCR was conducted in 20 µl reaction volumes containing 2x BioMix (Bioline Australia), 0.4µl each of the forward and reverse primers (10µM), 7.2 µl of nuclease free water and 2 µl of DNA (10 ng/µl). The COI primers, LEP(F1),59 ATTCAACCAATCATAAAGATATTGG-39:LEP(R1), 59-TAAACTTCTGGATGTCCAAAAAATCA-39 (Hebert et al., 2004) were used, modified with an additional Illumina adapter sequence (Bell et al., 2011). The COI primers amplify a 658bp region across a wide range of taxa (Nzeduru et al., 2012). The amplification protocol comprised of 5 cycles 94°C/120 s, 94°C/40 s, 45°C/40 s, 72°C/60 s, 35 cycles of 94°C/40 s, 51°C/ 40 s, 72°C/60 s, and a final 5 min step at 72°C with the first five cycles used to increase the primer binding strength (Nzeduru et al., 2012). Representative samples were run on a 1% agarose gel to check for amplification before purification using the Agencourt AMPure XP PCR Purification system (Beckman Coulter, USA).

The second PCR reaction was set up using the Illumina Nextera XT Index Kit v2 (A,B,C,D) to add sample specific indices and Illumina adapters to the PCR products of the first PCR. A 12.5µL reaction was used containing 2.5µL (5×) MyFi Buffer, 0.5µL MyFi polymerase, 1.25µL

of each of the forward and reverse indexed primers (µ10M), 5.75µL nuclease free water and 1.25µL of the purified PCR product. Representative samples of the indexed PCR products were run alongside non-indexed PCR products on a 1% agarose gel to confirm inclusion of indices to PCR products.

2.5. Preparing library of reference samples for Sanger sequencing

Two 12 µl reaction mixtures were prepared from the indexed, purified PCR products of each reference sample. The reaction mixture comprised 50ng of DNA with 1 µL of either the forward or reverse primer (10µM) and water. The samples were sequenced by capillary electrophoresis on an ABI 3730xl sequencer at the Australian Genome Research Facility (Adelaide).

2.6. Preparing library of bulk samples for MiSeq sequencing

Indexed PCR products from each bulk sample were pooled to form three large bulks and purified using the Agencourt AMPure XP PCR1 Purification system (Beckman Coulter, USA). The clean PCR product bulks were then used to prepare 1/50 and 1/100 dilutions and quantified by qPCR using the KAPA SYBR FAST qPCR Kit (Kapa Biosystems, South Africa) in 10 µL reaction mixtures comprising 1uL of DNA, 0.1 µL of forward and reverse Illumina Sequencing primers at 10µM (Bell et al., 2011) and Kapa (2x). On the basis of the qPCR results, the bulked samples were pooled to make a 500 pg/µL aliquot that was analysed for fragment size and quantity using a High Sensitivity DNA Assay on the 2100 Bioanalyzer (Agilent Technologies). Using quantification results from the Bioanalyzer, a 4nM aliquot of the indexed full length COI amplicons was prepared and sequenced by paired-end sequencing on an Illumina MiSeq instrument using the Illumina 600-cycle Version 3 kit.

2.7. Bioinformatics processing for Sanger sequencing output

FASTA sequence files and raw base call sequence files provided by AGRF for each Sanger sequenced sample were imported into the Geneious program (Geneious 6.1.6), sorted and trimmed to remove poor reads and adapter/primer sequences. For each reference sample, a

complete 658bp COI sequence was obtained and BLAST searched against the NCBI database using the query centric BLAST function of the Geneious program. The BLAST search was limited to a single hit per sample and the results interpreted using rank specific cut off points suggested by Tuzhikov (2014) i.e. the true family of the query sequence at 80% sequence identity with 90% query coverage, genus at 95% identity and 90% query coverage and species at 97.5% identity and 95% query coverage.

2.8.Assembly of a local COI reference sequence database

In order to test the meta-barcoding pipeline's capacity to generate baseline data informative for GM risk assessment, reference sequences of ecological functional groups of arthropods suggested by EFSA (2010) were used to assemble a local COI database. The groups included generalist predators and parasitoids, pollinators, and herbivores of the coffee crop system. In assembling the local database, common arthropod families that had been identified to family level using Sanger sequencing results, were searched through the Barcoding Of Life Database (BOLD) for reference sequences submitted from East Africa. Those found were grouped into ecological functional groups and their reference sequences retrieved for inclusion into the local database. Also, insects that were confidently identified to species level using Sanger sequencing, were included in the database together with NCBI COI reference sequences of their closely related species in order to test the pipeline's accuracy in assigning taxonomic names to query sequence reads generated by the MiSeq. Vouchered COI reference sequences of different insect orders from NCBI were also included to increase taxonomic coverage of the database assembled. All reference sequences used to build the local database were first converted to a TSV format that is recognised by the MSR program.

2.9.Analysis of the bulk arthropod samples using MSR

Using MSR, the MiSeq reads generated were trimmed for short and poor quality sequences. Clean sequence reads were further de-multiplexed into FASTQ files for each bulk sample using the sample identification indices. The clean FASTQ files representing a total of 288 bulk samples were then BLAST searched against the assembled local database of COI reference sequences. MSR uses a Naive Bayesian Classifier algorithm to rapidly assign query DNA sequences to taxonomic ranks with 100 bootstrap trials used as an estimate of confidence in assigning a query sequence to the rank of a particular reference sequence (Wang et al., 2007). All the sequencing work and analysis was performed in the Wilkinson laboratory of the Plant Research Centre at the University of Adelaide (Waite Campus).

3. Results

3.1.Sanger sequenced reference samples of common arthropods

91 reference samples were identified into 40 arthropod families representing 12 orders (table 2). Of these, 37 were common arthropods and clustered into 25 families representing nine orders (fig 1). Five arthropods (table 3) were classified to species level using threshold percentage identity values suggested by Tuzhikov (2014) as a measure of confidence in assigning a reference sequence to a given rank.

3.2.Identification of Sanger sequenced references for assembling the local database

The common arthropod families were searched through the BOLD database for reference sequences of their genera, species or molecular taxonomic units (MOTUs) submitted from East Africa. Thirteen common arthropod families were represented in the BOLD database (Table 4) and these were grouped according to their ecological functional groups and their reference sequences included in the local database for analysing the bulk arthropod samples. Also, reference sequences in the BOLD database of genera belonging to arthropod families that are known to inhabit coffee plantations but had not been represented in the Sanger sequenced dataset were included in the local database. For arthropod families not represented in the BOLD collections, their Sanger sequences were BLAST searched in NCBI to retrieve different genera of the families in order to increase their taxonomic coverage in the local database assembled. Samples that were confidently identified to species level were also included in the local database (Table 3). In total, the local database assembled comprised 248 reference sequences.

3.3.Determining the taxonomic resolution achieved for identifying the MiSeq reads

The MiSeq sequencing run generated 18.84 M reads with 17.53 M passing filter and 88.4% of these reads above the Q30 quality score. For each bulk arthropod sample, non-overlapping forward and reverse reads of 275bp and 274bp respectively were obtained. Owing to the modification of the meta-barcoding pipeline applied, in which full length COI amplicons were sequenced by paired end sequencing to generate non-overlapping reads, the taxonomic resolution achieved for the MiSeq reads in the absence of the COI middle 109 base pairs was deduced by comparing Neighbour joining (NJ) trees constructed using sequences of reference samples obtained from both Sanger and MiSeq sequencing platforms. With the exception of three reference samples from families, Pompilidae, Anthocoridae and Hemerobidae, both NJ trees conserved the phylogenetic structure of the taxonomic clusters assigned to the reference samples (Fig 2). To further investigate the taxonomic resolution achievable when using the Bayesian classifier of the MSR to assign taxonomic names to the non-overlapping Miseq reads, COI sequences from NCBI of species closely related to the five reference samples identified to species level were added to the local database. The capacity of the MSR to distinguish the reference samples from their most closely related species was determined. All the five reference samples were accurately distinguished from their most closely related species. Four samples were resolved with greater than 97.5% accuracy and one sample (*Ceratitis anonae*) with 81% accuracy (table 5).

3.4.Diversity and Distribution profiles of selected ecological functional groups of arthropods

The bulk arthropod samples were BLAST searched against the assembled local database using the MSR (Illumina). A default read abundance threshold of 3.5% was adopted as a measure of confidence in the presence of a reference sequence in each bulk sample. An example of the analysis output made by the MSR for a single bulk sample is shown in Figure 2. The genera that were identified in the bulk arthropod samples collected from each district are presented in figure 3. Some genera were shared across the four districts with some like the *Trirhithrum, Drosophila* and *Ceratitis* consistently recording high incidence counts while *Orthocentrus* and *Ebelingia* consistently registered low incidence counts. *Rhinia, Orius, Xysticus, Aloconota* and *Empoasca* consistently recorded moderate incidence counts across the four districts surveyed. Some genera like *Chilocorus, Amata, Pitane* and *Dacus* only occurred in the Southern districts of Masaka and Rakai while *Mymarachne* only occurred in the South western districts of Mitoma and Shema.

With the exception of the pollinator group, only genera that registered significant presence in coffee plantations of at least three of the four surveyed districts ($\geq$ 3 incidence counts), were further classified into their ecological functional groups and used to summarise baseline data on farmland arthropod diversity and distribution (Table 4). A total of seventeen genera from thirteen arthropod families were categorised into five ecologically functional groups and baseline data on their distribution across the four districts was compiled. Baseline data on species diversity and distribution of the most common pest *Trirhithrum*, was also retrieved from the bulk arthropod samples by including reference sequences of different species of the genus into the local database. Four species, *T. coffeae, T. senex, T. nigerrimum* and *T. teres*, in order of their percentage incidence across the four districts surveyed, were identified from the bulk samples analysed (Fig. 5).

4. Discussion

The purpose of this study was to validate a modified meta-barcoding pipeline as a comprehensive and cost effective approach of assembling and monitoring farmland arthropod communities. We modified the sample preparation steps of the meta-barcoding pipeline to reduce the overall costs of processing bulk arthropod samples on the MiSeq platform. We also adapted the bioinformatics pipeline for analysing the large volume of data produced from the MiSeq by modifying the MSR program, originally customised for classifying microbial populations, to classify arthropods in the study. Bulk arthropod samples that were used as an exemplar for the approach were collected from coffee plantations in four major coffee growing districts of Uganda. Based on the EFSA guidelines for conducting PMEM on GM crops (EFSA GMO Panel 2010), we assembled a local database of common arthropod parasitoids, generalist predators, pollinators, common pests and herbivores (plant sucking pests) whose diversity and distribution across the sampled sites was analysed to generate baseline data that is informative for the GM risk assessment process

The baseline data required to assess risks posed by GM crops on farmland arthropods includes taxonomic classification, diversity and distribution of the arthropods inhabiting the crop system under surveillance. This study aimed to test the capacity of the MSR program to generate baseline data using MiSeq reads hence reducing cost, time and expertise required for this exercise. The MSR program uses a Naive Bayesian Classifier algorithm to assign taxonomic identification to query sequences based on a local database of reference sequences (Wang et al., 2007). The query sequences generated via our modified meta-barcoding pipeline were non-overlapping and lacked the middle 109 bp. We tested the loss in taxonomic resolution of the query sequences due to this sequence gap by comparing NJ phylogenetic trees constructed using reference sequences obtained from both Sanger and Miseq platforms. Both NJ trees conserved a uniform phylogenetic structure for the taxonomic clusters obtained (Fig. 2a and 2b) This result

demonstrates that for most of the reference samples sequenced, there was no loss in taxonomic resolution due to the sequence gap of the Miseq reads. However for three reference samples, the clustering changed on the NJ tree constructed using Miseq reads. This might indicate that the importance of the middle 109bp COI segment in assigning taxonomic identities to arthropods varies between individual arthropods but might be insignificant for most since the COI is highly polymorphic across the entire length of the barcode (Yu et al., 2012).

In order to further test the resolving power achievable in assigning taxonomic identification to the MiSeq reads using the Bayesian classifier algorithm of MSR, we included COI references sequences of species from NCBI closely related to five common arthropods collected from the survey that had been confidently identified to species level (Table 3). Table 5 shows the performance of the Bayesian classifier in distinguishing the known reference sequences from their closest related species. Of the five reference species included in the database, the Bayesian classifier discriminated four reference species from their most closely related species with more than 97.5% accuracy. Only one species, *Ceratitis anonae,* was discriminated from its most closely related species *Ceratitis rosa* with a lower accuracy of 81%. Further analysis using the NCBI alignment output of the COI reference sequences from *C. rosa* and *C. anonae* revealed that the two species are 99% genetically identical. The ability of the Bayesian classifier to discriminate species that are nearly 100% identical with 81% accuracy further demonstrates the high resolving power of the taxonomic assignment algorithm used by the MSR program. These results agree with a previous study conducted by (Wang et al., 2007), who demonstrated that the Bayesian classifier algorithm was capable of discriminating Bacterial 16SrNA sequences at the genus level with an overall accuracy above 88.7% using query sequences greater 400bp in length. In his study, Wang (2007) observed that the Bayesian classifier algorithm significantly improved in accuracy with increase in length of the query sequences. Therefore, the ability of the

modified meta-barcoding pipeline proposed in this study to generate near full length COI reference sequences of 549 bp contributed to the resolving power of the Bayesian classifier.

For each bulk sample collected, the MSR program generated a summary of its arthropod composition ranked from kingdom to genus by blasting against the locally assembled database. Figure 3 demonstrates the output from the MSR showing an arthropod community comprising three orders; Hymenoptera, Lepidoptera and Diptera, four families; Drosophilidae, Ichneumonidae, Arctiidae and Tephritidae, three genera; *Orthocentrus, Drosophila, Trirhithrum* and *Pitanes* and two species, *Pitane fervens* and *Trirhithrum nigerrimum*. These results show that the metagenomics analysis function in the MSR, originally customised for classification of bacteria using the 16S rRNA gene (Wang et al, 2007), was successfully adapted to identify and group the arthropods collected using COI sequence reads. MSR also provided statistics on the percentage composition of the identified taxa in each bulk sample using the number of sequence reads that matched a reference sequence. Due to PCR bias, the volume of sequence reads cannot yet be used to accurately measure relative abundance of the identified taxa (Zhou et al., 2013). Nevertheless, the percentage of reads assigned to each taxa, provides a measure of confidence in its presence within the bulk sample (*MiSeq Reporter Metagenomics Workflow Reference Guide* 2013). This is important for filtering out contaminants that might be introduced in a sample at any stage of sample preparation and sequencing. For this study, the default read abundance threshold of 3.5% was adopted as the measure of confidence in the presence of a taxonomic group in the bulk arthropod samples analysed. Using the bulk sample chosen for illustration (Fig 3), the *Drosophila* species, *D. melanogaster* would not be confidently placed in the bulk sample despite the high number of reads assigned to it by the classifier algorithm because its number of reads are proportionally below 3.5% of the total Miseq reads generated from the bulk sample. This level of stringency increases confidence in the taxonomic composition of the bulk samples assembled as a true representation of the arthropod communities under surveillance. Prior to this

study, no report has been made on the use of the MSR program to classify arthropod samples into taxonomic groups with much of the previous analysis limited to classifying microbial populations (Klindworth et al., 2013).

From the MSR analysis we were able to successfully construct partial arthropod communities from 48 coffee plantations across four major coffee growing districts of Uganda using a locally assembled database of 248 reference sequences (Fig. 4). The taxonomic coverage of the arthropod communities assembled can easily be increased by expanding the size of the database used to assign taxonomic identification to the sequenced reads. Therefore with a more comprehensive database, the arthropods identified in all the bulk samples collected would provide generic data from which GM regulators can easily identify non-target organisms for several GM crop events being assessed. Using figure 4, it was possible to identify the common arthropods from the rare arthropods, to determine the difference in the abundance of the common arthropods and to detect the difference and similarity in their distribution between the southern districts of Masaka and Rakai and southwestern districts of Mitoma and Shema. The ability to compare the abundance of different arthropod taxa across a landscape, allows the regulators to determine the level of exposure for potential non-target organisms based on their level of interaction with the GM crop. Resident arthropods of a given crop system would occur with a greater frequency as compared to tourist arthropods and are therefore more exposed to the GM crop (Schowalter et al., 2006; Tschenn et al., 2001).

Meta-barcoding provides the capacity to analyse several farms as individual bulk samples in a single sequencing operation which allows for easy comparison between the arthropod communities assembled from any given crop system. This allowed us to generate and compare diversity and distribution profiles of individual arthropod genera from five ecological functional groups across the four coffee growing districts of Uganda.

Table 6 is a summary on the diversity and distribution of the arthropod families belonging to the selected functional groups and that were confidently identified from the bulk arthropod samples. The relative abundance and distribution of individual genera belonging to the families identified was determined using incidence counts per site in a district. From table 6, baseline data was compiled on diversity and distribution of eight genera from six arthropod families of parasitoids and generalist predators. The parasitoid and the predator families identified have been previously reported to show significant presence in coffee plantations from surveys conducted in East Africa (Neuenschwander et al., 2003; Chay-Hernandez et al., 2007; Sobhy et al., 2014).

Generalist predators and parasitoids are the most commonly used indicators of agro-ecosytem health because their diversity provides insight into the stability of the ecosystem (Moonen et al., 2008). A higher diversity (species richness and evenness) of the generalist predators and parasitoids maintains pest populations under control and so keeps the agro-ecosystem less dependent on external agricultural inputs for optimal crop yield (Cabell et al., 2012; Moonen et al., 2008). The capacity to generate baseline data on several indicators from a single survey highlights the effectiveness of the proposed monitoring system. Baseline data obtained from several indicators can be used to generate Multiple Species Indices (MSI) for the measurement end points used when monitoring agro-ecosystems (Buckland et al., 2011). MSI are less prone to false alerts on adverse effects of GM crops (Glandorf et al., 2012). This is particularly useful in interpreting data collected on arthropod parasitoids and predator species because of the compensatory nature of their eco-system services in controlling pest populations. The loss of a single species does not necessarily imply an adverse effect on the ecosystem since another species in the same ecosystem functional group easily fills the missing gap (Moonen et al., 2008). Furthermore, the use of MSI from data generated using several indicators provides a stronger signal on the changes occurring to an agro-ecosystem and therefore increases sensitivity of the monitoring exercise to subtle effects of the GM crop if any (Glandorf et al., 2012). In

addition, the relative abundance profiles generated before introduction of GM crops into the crop system provide baseline data on the natural variation of individual arthropod genera caused by changes in weather patterns, the stage of plant growth and other natural factors (EFSA GMO Panel 2011). For example, the Apidae genus *Apis* revealed in the bulk samples by the MSR is a major coffee pollinator in Uganda (Munyuli et al., 2014). However, from the analysis made on the MSR output, the genus recorded a very low incidence during the survey period between May and June (Table 4). This is because pollinators only occur in the coffee plantations at a high frequency during flowering seasons, which in Uganda, take place between January and February and later July and August (Munyuli et al., 2013; Munyuli et al., 2014). Such natural variation must be accounted for and teased apart from variation caused by GM crop adoption in interpreting survey data collected for GM risk assessment (EFSA GMO Panel 2011).

Also, depending on the GM crop under surveillance, the baseline data generated can be used to monitor for direct effects of the transgenic product on non-target organisms based on relative abundance profiles assembled after the adoption of a GM product (EFSA GMO Panel 2011). For example the Staphylinidae and Coccinellidea predators (Table 6) being coleopterans are prone to effects of Cry 3 toxins which they might be exposed to from their prey (Romeis et al 2009). In case of adoption of a Cry 3 transgenic crop, data collected from subsequent surveys on relative abundance of the two potential NTOs can be compared with base line data collected before the GM introduction or sites cultivating conventional crops to determine if there are any direct effects of the Cry 3 toxin on the coleopteran predators in the agro-ecosystem.

Baseline data was also generated on species diversity and distribution of the most common coffee pest of the Tephritidae family, *Trirhithrum* (Table 6). As expected of common pests, *Trirhithrum* recorded the highest incidence level compared to genera in the rest of the functional

groups analysed. Four *Trirhithrum* species were recorded in the bulk arthropod samples analysed with *T. coffeae* recording the highest incidence followed by *T. senex*, *T. nigerrimum* and *T. teres.* The genus *Trirhithrum* and especially the species *T. coffeae* is a well-known pest of coffee plantations in East Africa where it damages the pulp of ripe coffee berries during oviposition (Wharton et al., 2000).

Baseline data collected on common pests is necessary for GM risk assessment to determine the effectiveness of a GM crop in controlling the target pest. Comparing the population size of a target pest before and after a GM event is introduced in an agro-ecosystem, reveals incidences if any, of emerging pest resistance to the GM crop (Tabashnik et al., 2013, EFSA GMO Panel 2010). However, this requires capacity to monitor the target pest at species level since related species of the target pest are the most likely sources of resistance to the GM crop (Zhanga et al., 2012; Underwood et al., 2006). Therefore, the capacity of the method to collect baseline data of a common pest at the species level, further demonstrates the effectiveness and power of the proposed monitoring system.

Finally, attempts to collect baseline data on herbivores revealed one genus of the Ciccadellidae family, *Empoasca*. The genus did not register significant presence in Masaka district and occurred at fairly low incidence levels throughout the rest of the districts as compared to the common pests. Several plant sucking pests are known to be secondary pests in crop systems and do not attract attention of farmers since they are easily controlled using pesticides unlike common pests (Dutcher et al., 2007; Lu et al., 2010). However, in the absence of common pests, such as would be the case when Bt crops targeting major pests are deployed, plant sucking pests have been reported to quickly rise in population size due to lack of competition for breeding ground or food and have in some cases become major pests (Baptiste et al., 2010). As recommended by the EFSA GMO Panel (2011), baseline data generated on secondary pests of a

crop system is a good indicator of reduction in the pest regulation activity of predators and parasitoids that warrants further investigation into a possible causal linkage to the GM crops under cultivation.

In this study, we used incidence counts as a proxy for measuring relative abundance. The use of incidence counts by subsampling as a proxy for relative abundance is a well-established method commonly used to determine species richness in populations (Chao et al., 2004). In meta-barcoding however, the method was only suggested by Yu (2012), as an alternative to the use of sequence read volumes to measure relative abundance of the arthropods identified in samples. Yu (2012) suggested that subsampling of the surveyed sites (nested sampling) would be required to provide a more accurate measure of abundance if incidence counts were to be used as proxies for measuring abundance of the arthropods captured. While the use of nested sampling in this study provided an estimate of the relative abundance of individual genera, further studies would be required to determine the loss in comparative statistical power when using presence/absence data as opposed to absolute counts of individual arthropods. Such a study would inform decisions on key questions like what frequency of occurrence is sufficient to draw conclusions of an impact on the arthropod community of the crop system under surveillance.

Therefore, using the proposed meta-barcoding pipeline, it is possible to meet the objectives of both case specific and general surveillance as recommended by the EFSA. As shown in this study, generating a local database of informative reference sequences for different measurement endpoints allows the GM regulators to monitor GM crops for potential adverse effects on the farmland arthropod communities. Inclusion of an NTO reference sequence provides its relative abundance profiles that allow the GM regulators to determine its level of exposure to the transgenic crop and to detect changes in its population size in the presence and absence of a GM crop. Inclusion of a secondary pest reference sequence allows the regulator to determine the

impact of the GM crop or its cultivation on regulation of minor pests in the crop system. Inclusion of a target pest reference sequence in the database allows the GM regulator to assess the efficiency of a transgenic crop in controlling the target pest or the efficiency of a risk management strategy such as a high dose/refuge strategy (Tabashnik et al., 2013) in supressing target pest resistance. Towards achieving the objectives of a general survey, comparisons made between the observed diversity and distribution profiles of the indicator species to their baseline data collected in the absence of GM crops, allow the GM regulators to detect changes in the farmland arthropod communities, determine if the changes are potentially harmful and provide insight into the causal linkage of these changes to the GM crop. The adversity of the changes in farmland arthropod communities and their causal linkage to the GM crop can then be proved by conducting case specific studies (EFSA GMO Panel 2011).

The approach is both time saving and cost effective because all the sequencing, identification, community assembly and comparison between arthropod communities can be performed on the MiSeq which is the cheapest Illumina sequencing instrument on the market (Quail et al., 2012). For this study, we used the MiSeq v3 kit which currently provides the longest paired-end sequencing reads at the lowest cost per run (Glenn et al., 2014). We further reduced the costs of the sequencing run by avoiding the fragmentation step of the COI amplicons during library preparation and instead performing paired-end sequencing on full length COI amplicons. With this modification we saved more than US$14,000 in the library preparation stage compared with using the Nextera DNA Sample Preparation Kit (96 Samples) at half the recommended reagent volumes to prepare a sequencing library for 288 bulk samples (*Quotation for supply of genetic analysis products* 2014). Still, we were able to generate 549bp sequence reads lacking 109bps in the middle without compromising the COI barcode's taxonomic resolution (Fig.2 and Table 3). The MSR sample processing time including trimming the MiSeq reads for short and poor quality sequences, de-multiplexing and blast searching the bulk samples against the assembled local

database was under two hours. The MSR output was ready to use and the speed of analysis allowed us to conduct several tests on the bulk arthropod samples. It is also worthwhile mentioning that the meta-barcoding pipeline using a MiSeq platform is being used to identify plants and vertebrates based on universal barcodes (Camp 2013) and could therefore be explored for monitoring other protection goals as determined by the GM regulatory authorities. Such a monitoring system could be customised and located in a centralised PMEM data reporting centre as suggested by the EFSA (Kleppin et al., 2011).

The main limitation of the proposed monitoring system is the lack of vouchered reference sequences in public databases especially from developing countries. This restricts the size of local database assembled in terms of taxonomic coverage and reduces accuracy of the Bayesian classifier algorithm of the MSR program. Given the low rate of COI reference sequence submissions to public databases from developing countries and the general inaccuracy in taxonomic assignment of the COI reference sequences in NCBI, future studies should look to assembling *de novo* local databases based on molecular taxonomic units of sequences derived from the bulk arthropod samples being analysed.

Also, in order to obtain more accurate baseline data, surveys should be conducted several times before the introduction of GM crops so as to establish the patterns in natural variation of the incidence and distribution of indicator arthropods.

5. Conclusion

In conclusion, we have demonstrated the capacity of a modified meta-barcoding pipeline on the MiSeq platform to assemble farmland arthropod communities and generate data on functional arthropod groups that can be used to monitor GM crops for potential adverse effects on agro-ecosystems. The approach we have proposed is both comprehensive and cost effective and the data generated is informative to the risk assessment process of GM crops.

6. Acknowledgements

I thank the International Centre for Genetics and Biotechnology (ICGEB) for sponsoring my Masters course in Plant Biotechnology at the University of Adelaide and funding the entire project. I wish to thank members of the Entomology Department Uni. Adelaide (Waite Campus), for discussing with me the ideas of this project at its conception stage. I also sincerely thank staff members of the National Crop Resource Research Institute, Uganda for hosting me during the sample collection phase of the project. My sincere gratitude extends to members of the Wilkinson Laboratory for helping me learn to conduct the laboratory experiments and the bioinformatics of the project. Finally I thank my colleagues of the Plant Biotechnology class for bearing with me through the years.

7. References

Aheto, DBT, Breckling, B, Johnnie van den Berg, Ching, LL, Wikmark, OG 2013, 'Implications of GM crops in subsistence-based agricultural systems in Africa', In: Breckling, B and Verhoeven (eds), Implications of GM crops in subsistence-based agricultural systems in Africa, Theorie in der Ökologie, Frankfurt, pp 93.

Baird, DJ, Hajibabaei, MD 2012, 'Biomonitoring 2.0: a new paradigmin ecosystem assessment made possible by next-generation DNA sequencing,' Molecular Ecology, vol. 21, pp.2039–2044.

Baptiste, JR, Richroch, AE 2010, 'Emergence of minor pests becoming major pests in GE cotton in China', GM Crops, vol.1, no.4, pp.214-219.

Bell, J 2011, 'Overview of tailed amplicon sequencing approach with MiSeq', Illumina, USA.

Buckland, S, Studeny, AC, Magurran ,AE, Illian, JB, Newson, SE 2011, 'The geometric mean of relative abundance indices: a biodiversity measure with a difference', Ecosphere, vol.2, no.9, pp.100.

Cabell, JF, Oelofse, M 2012, 'An indicator framework for assessing agroecosystem resilience', Ecology and Society, vol.17, no.1, pp.18.

Camp, A 2013, 'Southern hairy-nosed wombat diet reconstruction from scats using Next Generation Sequencing reveals toxic weed species', MA, University of Adelaide, Adelaide.

Chao, A, Chazdon, RL, Colwell, RK, Shen, TJen 2004, 'A new statistical approach for assessing similarity of species composition with incidence and abundance data', Ecology Letters, vol.8, no.2, pp.148-159.

Chay-Hernández, DA, Delfín-González, H and Parra-Tabla, V 2006, 'Ichneumonoidea (Hymenoptera) Community Diversity in an Agricultural Environment in the State of Yucatan, Mexico', Environmental Entomology, vol.35, no.5, pp.1286-1297.

Dutcher, J 2007, 'A Review of Resurgence and Replacement Causing Pest Outbreaks in IPM', In: Ciancio, A, Mukerji, KG (eds), A Review of Resurgence and Replacement Causing Pest Outbreaks in IPM, Springer, University of Georgia, USA.

EFSA Panel on GMO 2006, ‘Opinion on Genetically Modified Organisms on the Post Market Environmental Monitoring (PMEM) of genetically modified plants’, EFSA Journal, vol.319, pp.1-27.

EFSA Panel on GMO 2010, ‘Opinion on the assessment of potential impacts of genetically modified plants on non-target organisms’, EFSA Journal, vol.8, no.11, pp.1877.

EFSA Panel on GMO 2011, ‘Scientific Opinion on guidance on the Post-Market Environmental Monitoring (PMEM) of genetically modified plants’, EFSA Journal, vol.9, no.8, pp.2316.

Gibsona, J, Shokrallaa,S, Porterc, TM, Kinga, I, Steven van Konynenburga, Janzend,DH, Hallwachsd, W and Hajibabaeia,M 2014, ‘Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics’, Proceedings of the National Academy of Sciences, vol.111, no.22, pp.8007-8012.

Glandorf, D 2012, 'General Surveillance of genetically modified plants ', National Institute for Public Health and the Environment, Netherlands.

Glenn, T 2014, '2014 NGS Field Guide – Table 2 – Run time, reads, yields, and costs', John Wiley and Sons, USA.

Grootaert, P, Pollet, M, Dekoninck, W, Cees van Achterberg 2010, 'Sampling insects: general techniques, strategies and remarks', In: Eymann, J., Degreef, J, Häuser, C, Monje, J, C, Samyn Y And Vandenspiegel, D (eds), Sampling insects: general techniques, strategies and remarks, Abc Taxa, Belgium.

Hebert, PDN, Penton, EH, Burns, JM, Janzen, DH, Hallwachs, W 2004, ‘Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator’, Proceedings of the National Academy of Science, vol.101, pp.14812–14817.

Illumina Sequencing Pricelist 2013, Illumina, viewed 27 October 2014, <http://systems.illumina.com/systems/miseq/kits.ilmn>.

Isaacs, R, Tuell, J, Fiedler, A, Gardiner, M and Landis, D 2008, 'Maximizing arthropod-mediated ecosystem services in agricultural landscapes: the role of native plants', Frontiers in Ecology and the Environment, vol.7, no.4, pp.196-203.

Ishaaya, I, Horowitz, A 2009, Biorational Control of Arthropod Pests, Entomology, Israel, viewed 1 November 2014, <http://www.springer.com/life+sciences/entomology/book/978-90-481-2315-5>.

James, C, 2013, ‘Global Status of Commercialized Biotech/GM Crops’, ISAAA Brief, Ithaca, New York.

Ji, Y, Ashton, L, Pedley, M, Edwards, P, Tang, Y, Nakamura, A, Kitching, R, Dolman, PM, Woodcock, P, Edwards, FA, Larsen, TH, Hsu, WH, Benedick, S, Hamer, KC, Wilcove, DS, Bruce, C, Wang, X, Levi, T, Lott, M, Emerson, BC and Yu,DW 2013, 'Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding', Ecology Letters, vol.16, no.10, pp.1245-1257.

Kapuscinski, A 2007, Environmental risk assessment of genetically modified organisms, CABI, Vietnam, viewed 1 November 2014, <http://books.google.com.au/books/about/Environmental_Risk_Assessment_of_Genetic.html?id=HivfMgEACAAJ&redir_esc=y>.

Kleppin, L, Schmidt, G, Schröder, W 2011, 'Cultivation of GMO in Germany: support of monitoring and coexistence', Environmental Sciences Europe, vol. 23, no.4, pp.231-11.

Klindworth, A, Pruesse, E and Glöckner, FO 2013, 'Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies', Nucleic Acids, vol.7, no.41, pp1.

Lu , L, Wu, K, Jiang, Y, Guo, Y and Desneux, N 2012, 'Widespread adoption of Bt cotton and insecticide decrease promotes biocontrol services', Nature, vol.487, no.7407, pp362–365.

Lu, Y, Wu, K, Jiang, Y, Xia, B, Li, P, Feng, H, Wyckhuys, KAG, Guo, Y 2010, 'Mirid Bug Outbreaks in Multiple Crops Correlated with Wide-Scale Adoption of Bt Cotton in China', Science, vol.328, no.5982, pp. 1151-1154.

MiSeq Reporter Metagenomics Workflow Reference Guide 2013, Illumina, viewed 29 October 2014, <http://supportres.illumina.com/documents/documentation/software_documentation/miseqreporter/miseq-reporter-metagenomics-workflow-guide-15042317-b.pdf>.

Moonen, A, Ba`rberi, P 2008, 'Functional biodiversity: An agroecosystem approach', Agriculture, Ecosystems and Environment, vol.127, no.1-2, pp.7–21.

Munyuli, T, 2014,‘Influence of functional traits on foraging behaviour and pollination efficiency of wild social and solitary bees visiting coffee (Coffea canephora) flowers in Uganda’, Grana, vol.53, no.1, pp.69-89.

Munyuli, T, 2013, ‘Factors governing flower visitation patterns and quality of pollination services delivered by social and solitary bee species to coffee in central Uganda’, African Journal of Ecology, vol.49, no.4, pp.501 – 509.

Munyuli, T 2011, 'Farmers’ perceptions of pollinators’ importance in coffee production in Uganda', Agricultural Sciences, vol.2, pp.318-333.

Munyuli , T 2012, 'Butterfly Diversity from Farmlands of Central Uganda', Psyche: A Journal of Entomology, vol. 2012, Article ID 481509, pp. 23.

Nzeduru, C, Ronca, S, Wilkinson, MJ 2012, 'DNA Barcoding Simplifies Environmental Risk Assessment of Genetically Modified Crops in Biodiverse Regions', PLoS ONE, vol.7, no.5.

Obonyo, DN, Nfor, Lm and Uzochukwu, S 2011, ‘Identified Gaps in Biosafety Knowledge and Expertise in Sub-Saharan Africa, AgBioForum, vol.14, no.2, pp.71-82.

Quail , M, Smith, M, Coupland, P, Otto,T, Harris, S, Connor, TR, Bertoni, A, Swerdlow, HP and Gu,Y 2012, 'A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers', BMC Genomics, vol.13.1, no.341.

Quotation for supply of genetic analysis products 2014, Illumina, viewed 29 October 2014, <http://www.med.unc.edu/pharm/calabreselab/files/illumin_standing_quote>.

Romeis, J, Meissle, M, Raybould, A, Hellmich, RL 2009, 'Environmental impact of genetically modified crops on above-ground non target arthropods', In: Ferry, N., Gatehouse, A. M. R(eds), Environmental impact of genetically modified crops on above-ground non target arthropods, CAB International, USA, pp. 165-198.

Sandra, M 2006, 'Country Pasture/Forage Resource Profiles', Ministry of Agriculture, Animal Industry and Fisheries, FAO,Uganda.

Schowalter, T 2006, Insect Ecology: An Ecosystem Approach,Academic Press, USA, viewed 20 October 2014, < http://books.google.com.au/books?id=LQqHWCtj0F0C>.

Tabashnik, B, Brevault, T, Carriere, Y 2013, 'Insect resistance to Bt crops: lessons from the first billion acres', Nature Biotech, vol.31, no.6, pp.510-521.

Tschenn, J, Losey, JE, Jesse, LH, Obrycki, JJ, Hufbauer,R, 2001, 'Effects of Corn Plants and Corn Pollen on Monarch Butterfly (Lepidoptera: Danaidae) Oviposition Behavior', Environtal Entomology, vol.30, no.3, pp.495-500.

Tuzhikov, A, Alexander P, Valery, IS 2014, 'TUIT, a BLAST-based tool for taxonomic classification of nucleotide sequences', Bio Techniques, vol.56, no.2, pp.78-84.

Underwood, E, Fisher, BL 2006, 'The role of ants in conservation monitoring: If, when, and how', Biological Conservation, vol.1 3 2, pp.1 6 6 –1 8 2.

Waller, J, M,Bigger, B, Hillocks,R, J 2007 Coffee Pests, Diseases and Their Management ,CABI, viewed 1 November 2014, <http://books.google.com.au/books?id=RpcoIgUBOLoC>

Wang, Q, Garrity, MG, Tiedje, JM, Cole, J 2007, 'Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy', Applied and Environtal Microbiology, vol.73, no.16, pp.5261.

Wharton, R, Trostle, MK, Messing, RH, Copeland, RS, Kimani-Njogu, SW, Lux, S, Overholt, WA, Mohamed, S, Sivinski, J 2000, 'Parasitoids of medfly, Ceratitis capitata, and related tephritids in Kenyan coffee: a predominantly koinobiont assemblage', Bulletin of Entomological Research, vol.90, no.6, pp.517-26.

Yu, D, Ji, Y, Emerson, B, Wang, X, Ye, C, Yang, C, Ding, Z 2012, 'Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring', Methods in Ecology and Evolution, vol. 3, no, 4, pp. 613–6.

Zhanga, H, Tiana, W, Zhaoa, J, Jina, L, Yanga, J, Liua, C, Yanga, Y, Wua, S, Wub, K, Cuic, J, Tabashnikd, EB and Wua,Y 2012, 'Diverse genetic basis of field-evolved resistance to Bt cotton in cotton bollworm from China', Proceedings of the National Academy of Science, vol.109, no.26, pp.10275–10280.

Zhao, J, Ho, P, Azadi, H 2011, 'Benefits of Bt cotton counterbalanced by secondary pests? Perceptions of ecological change in China', Environtal Monitoring Assessment, vol.184. no. 11, pp7079.

Zhou, X, Li, Liu, Y, Yang, Q, Su, X, Zhou, L, Tang, M, Fu, R, Li, J and Huang, Q 2013, 'Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification', GigaScience, vol.2, no.4, pp.2.

Figure legends

Fig.1 Neighbour joining tree showing taxonomic classification of the common arthropods captured into 25 families representing nine orders. An asterisk (*) marks samples identified to species level.

Fig.2a Neighbour joining tree of reference samples constructed using COI sequences generated by Sanger sequencing. Asterisk (*) shows samples that clustered inconsistently when compared with a similar NJ tree constructed using Miseq reads of the same samples.

Fig.2b Neighbour joining tree of reference samples constructed using COI sequences generated on the Miseq platform. Asterisk (*) shows samples that clustered inconsistently when compared with a similar NJ tree constructed using Sanger sequenced reads of the same samples.

Figure 3. Composition of a bulk sample at four taxonomic ranks; order (a), family (b) genus (c) and species (d) as generated by the MSR software. The unclassified regions represent sequences in the bulk DNA sample for which no taxonomically informative reference sequence was found in the local database. The top species classification results (d) are demonstrated in a table format that shows the sample composition of the identified species as a percentage of the total number of reads generated from the bulk sample. Only taxa that met the abundance threshold of 3.5% are shown by the pie charts at each rank.

Fig.4 Incidence of genera that registered significant presence within the 12 coffee plantations sampled in each of the four districts, Mitoma, Shema, Masaka and Rakai. The colours code for individual genera.

Figure 5. Percentage occurrence of *Trirhithrum* species, *T. coffeae*, *T. senex*, *T. nigerrimum* and *T. teres* in four major coffee growing districts of Uganda based on presence absence data collected from 12 coffee plantations per district.

Table 2. Reference dataset of common arthropods (D) and other arthropods collected during the survey and identified using the NCBI database.

Sample ID	Order	Family	Putative genus and species name	Accession numbers
67D	Coleoptera	Cerambycidae	Arhopalus foveicollis	gi 606238908
98 D	Coleoptera	Chrysomelidae	Galerucella lineola	gi 485650176
31D	Coleopttera	Coccinellidae	Calvia quatuordecimguttata	gi 606237468
13D	Coleoptera	Curculionidae	Xylosandrus compactus	gi 293630437
34D	Coleoptera	Curculionidae	Sitona discoideus	gi 133722289
56D	Hemiptera	Dictyopharidae	Scolops pallidus	gi 589060443
40D	Araninae	Thomisidae	Xysticus luctuosus	gi 545749220
2D	Diptera	Culicidae	Anopheles sinensis	gi 648296150
39D	Diptera	Culicidae	Anopheles albitarsis	gi 309399364
37D	Diptera	Drosophilidae	Leucophenga neointerrupta	gi 513133103
59D	Diptera	Drosophilidae	Drosophila repleta	gi 110189392
43D	Diptera	Drosophilidae	Drosophila simulans	gi 40795386
9D	Diptera	Muscidae	Potamia littoralis	gi 312858287
48D	Diptera	Muscidae	Stomoxys sitiens	gi 188475059
44D	Diptera	Rhiniidae	Cosmina fuscipennis	gi 408716410
57D	Diptera	Sepsidae	Saltella nigripes	gi 183012472
3D	Diptera	Tephritidae	Trirhithrum coffeae	gi 339263573
70D	Ephemeroptera	Baetidae	Procloeon fragile	gi 269311455
17D	Hemiptera	Anthocoridae	Anthocoris confusus	gi 685163853
32D	Hemiptera	Cicadellidae	Empoasca vitis	gi 399157563
33D	Hemiptera	Issidae	Flavina sp	gi 304367649
54D	Hemiptera	Lygaeidae	Pachygrontha sp	gi 48256673
22D	Hemiptera	Membracidae	Leptobelus gazella	gi 332688488
30D	Hemiptera	Membracidae	Multareis cornutus	gi 589059803
50D	Hemiptera	Tropiduchidae	Tambinia sp	gi 379070375
80D	Hymenoptera	Apidae	Apis mellifera	gi 641804332
88 D	Hymenoptera	Braconidae	Aphidius colemani	gi 359292123
61D	Hymenoptera	Formicidae	Camponotus sp. MG005	gi 553835494
46D	Hymenoptera	Fomicidae	Myrmicaria brunnea	gi 485896044
26D	Hymenoptera	Formicidae	Camponotus sp. MG059	gi 553834761
64D	Hymenoptera	Formicidae	Camponotus maculatus	gi 148375291
1D	Hymenoptera	Ichneumonidae	Ichneutinae gen	gi 374873067
53D	Hymenoptera	Ichneumonidae	Trachyarus fuscipes	gi 307950253
10D	Lepidoptera	Arctiidae	Notarctia sp. Proxima	gi374910001
29D	Neuroptera	Hemerobiidae	Wesmaelius concinnus	gi 403311445
52D	Orthoptera	Tettigoniidae	Kuzicus suzukii	gi 325557155
62D	Orthoptera	Tettigoniidae	Tettigonia chinensis	gi 325557509
45F	Araneae	Salticidae	Myrmarachne japonica	gi 452091980

64E	Araneae	Salticidae	Myrmarachne japonica	gi 452091980
48F	Dictyoptera	Mantidae	Sphodromantis viridis	gi 259124209
73E	Diptera	Calliphoridae	Rhinia sp.	gi 408716412
47E	Diptera	Calliphoridae	Cosmina fuscipennis	gi 408716410
53G	Diptera	Calliphoridae	Chrysomya putoria	gi 408716354
95E	Diptera	Conopidae	Conops chinensis	gi 400234484
71E	Diptera	Culicidae	Anopheles campestris	gi 157734476
65F	Diptera	Diptera family	Diptera sp	gi 536461934
50F	Diptera	Dolichopodidae	Argyrochlamys impudicus	gi 225904778
55E	Diptera	Dolichopodidae	Argyrochlamys impudicus	gi 225904778
83E	Diptera	Drosophilidae	Drosophila repleta	gi 110189392
37E	Diptera	Drosophilidae	Colocasiomyia xenalocasiae	gi 649577458
53E	Diptera	Hybotidae	Elaphropeza lowi	gi 573463469
42F	Diptera	Hydrotaea	Hydrotaea armipes	gi 403310961
42E	Diptera	Muscidae	Limnospila albifrons	gi 451958544
64F	Diptera	Muscidae	Spilogona gibsoni	gi 479278779
48E	Diptera	Muscidae	Stomoxys calcitrans	gi 186694612
73F	Diptera	Muscidae	Haematobosca sanguinolenta	gi 511106847
45E	Diptera	Sarcophagidae	Sarcophaga africa	gi 380703356
40E	Diptera	Sciomyzoidea	Dictya gaigei	gi 353167294
46F	Diptera	Sphaeroceridae	Sphaeroceridae sp.	gi 321135514
41F	Diptera	Sphaeroceroidea	Epistomyia sp.	gi 399531301
63M	Diptera	Syrphidae	Melanostoma mellinum	gi 519668403
44E	Diptera	Tachinidae	Smidtia fumiferanae	gi 313663662
50E	Diptera	Tephritidae	Ceratitis aanonae	gi 339263275
54E	Diptera	Tephritidae	Capparimyia melanaspis	gi 339263257
21H	Diptera	Tephritidae	Trirhithrum coffeae	gi 339263575
50E	Diptera	Tephritidae	Ceratitis anonae	gi 339263275
63G	Diptera	Tephritidae	Bactrocera invaden	gi 380041083
66E	Hymenoptera	Apoidea	Apoidea sp.	gi 545334642
53H	Hymenoptera	Braconidae	Opius	gi 565337744
69E	Hymenoptera	Braconidae	Snellenius sp.	gi 349589678
39E	Hymenoptera	Cheloninae	Cheloninae sp	gi 411159121
39F	Hymenoptera	Cheloninae	Cheloninae sp	gi 411159479
21F	Hymenoptera	Formicidae	Camponotus maculatus	gi 148375291
44F	Hymenoptera	Formicidae	Tetraponera sp	gi 79915113
61E	Hymenoptera	Formicidae	Camponotus nr.	gi 382931578
62F	Hymenoptera	Formicidae	Camponotus sp.	gi 295308423
63F	Hymenoptera	Formicidae	Camponotus yiningensis	gi 383282047
63J	Hymenoptera	Formicidae	Camponotus yiningensis	gi 383282047
65E	Hymenoptera	Ichneumonidae	Ichneumonidae sp.	gi 411160977
96F	Hymenoptera	Ichneumonidae	Ichneumonidae sp.	gi 299015719
21I	Hymenoptera	Ichneumonidae	Hyposoter sp.	gi 374874457
56E	Hymenoptera	Ichneumonidae	Ichneumonidae sp.	gi 411164303

63K	Hymenoptera	Ichneumonidae	Campopleginae gen.	gi 374877907
68E	Hymenoptera	Ichneumonidae	Banchinae sp.	gi 342784831
57E	Hymenoptera	Pompilidae	Auplopus caerulescens	gi 190341720
67E	Hymenoptera	Pompilidae	Auplopus caerulescens	gi 190341720
63H	Lepidoptera	Arctiidae	Pitane fervens	gi 294466673
63I	Lepidoptera	Arctiidae	Amata phegea	gi 331256842
52E	Lepidoptera	Noctuidae	Noctuidae gen.	gi 374894871
41E	Rickettsiales	Anaplasmataceae	Wolbachia sp	gi 481068109

Table 3. Reference arthropod samples that were confidently identified to species level at 97.5% sequence identity and 95% query coverage using NCBI COI reference sequence database.

Sample ID	**Family**	**species name**	**Genbank Accession number**
53G	Calliphoridae	*Chrysomya putoria*	gi 408716354
43D	Drosophilidae	*Drosophila simulans*	gi 40795386
80D	Apidae	*Apis mellifera*	gi 641804332
3D	Tephritidae	*Trirhithrum coffeae*	gi 339263573
50E	Tephritidae	*Ceratitis anonae*	gi 339263275

Table 4. Common arthropod families selected from different ecological functional groups to be used for generating baseline data from the bulk arthropod samples.

Functional ecological groups	**Sample ID**	**Family**	**Putative genus and species name**	**Genbank Accession numbers**
Parasitoids	1D	Ichneumonidae	Ichneutinae genus	gi 374873067
	88 D	Braconidae	*Aphidius colemani*	gi 374873067
Generalist predators	17D	Anthocoridae	*Anthocoris confusus*	gi 685163853
	31D	Coccinellidae	*Calvia quatuordecimguttata*	gi 606237468
	40D	Thomisidae	*Xysticus luctuosus*	gi 545749220
	53G	Calliphoridae	*Chrysomya putoria*	gi 408716354
	61D	Formicidae	*Camponotus sp.*	gi 553835494
Pollinator	80D	Apidae	*Apis mellifera*	gi 641804332
Common pests	3D	Tephritidae	*Trirhithrum coffeae*	gi 339263573
	13D	Curculionidae	*Xylosandrus compactus*	gi 293630437
	83E	Drosophilidae	*Drosophila repleta*	gi 110189392
Herbivore (Plant sucking pest)	32D	Cicadellidae	*Empoasca vitis*	gi 399157563
	63H	Arctiidae	*Pitane fervens*	gi 294466673

Table 5. Performance of the Bayesian Classifier algorithm in assigning taxonomic identity to MiSeq reads of known reference samples using a database containing reference sequences of closely related species from the NCBI database.

Reference Sample ID	**Reference Genus**	**Total number of reads assigned to reference genus**	**Top Species Classification for genus**	**% of Total reads per genus assigned to species**
80D	*Apis*	44,497	*Apis mellifera*	99.7
			Apis dorsata	0.04
43D	*Drosophila*	40966	*Drosophila simulans*	99.3
			Drosophila melanogaster	0.3
53G	*Chrysomya*	39,897	*Chrysomya putoria*	98.0
			Chrysomya chloropyga	0.8
3D	*Trirhithrum*	26,482	*Trirhithrum coffeae*	97.9
			Trirhithrum nigerrimum	0.3
50E	*Ceratitis*	35,213	*Ceratitis anonae*	81.0
			Ceratitis rosa	17.5

Table 6. Diversity and distribution of selected arthropod families across 12 plantations per district in four major coffee growing districts of Uganda.

Ecological Functional Group	**Family**	**Genera**	**incidence /12 sites per District**			
			Mitoma	Shema	Masaka	Rakai
Parasitoid	Ichneumonidae	*Skiapus*	10	5	6	6
		Orthocentrus	4	3	2	3
Generalist predator	Thomisidae	*Xysticus**	7	5	1	5
		*Ebelingia**	3	4	2	3
	Anthocoridae	*Orius**	8	5	4	4
	Staphylinidae	*Aloconota**	3	5	6	2
	Coccinellidae	*Chilocorus*	-	3	3	4
	Calliphoridae	*Chrysomya*	4	6	12	12
Pollinator	Apidae	*Apis*	1	1	-	-
	Calliphoridae	*Rhinia*	7	6	5	2
Common pest	Tephritidae	*Trirhithrum*	12	9	12	11
		Dacus	-	-	6	5
		Ceratitis	9	8	12	11
	Drosophilidae	*Drosophila*	11	8	10	12
Herbivore	Cicadellidae	*Empoasca**	3	5	-	7
	Arctiidae	*Pitane*	-	-	4	6
	Nymphalidae	*Arygnnis*	-	-	7	9

- shows that the arthropod was not captured in any of the 12 coffee plantations sampled in the district

* shows arthropods that were only confidently typed to family level

Figure 1

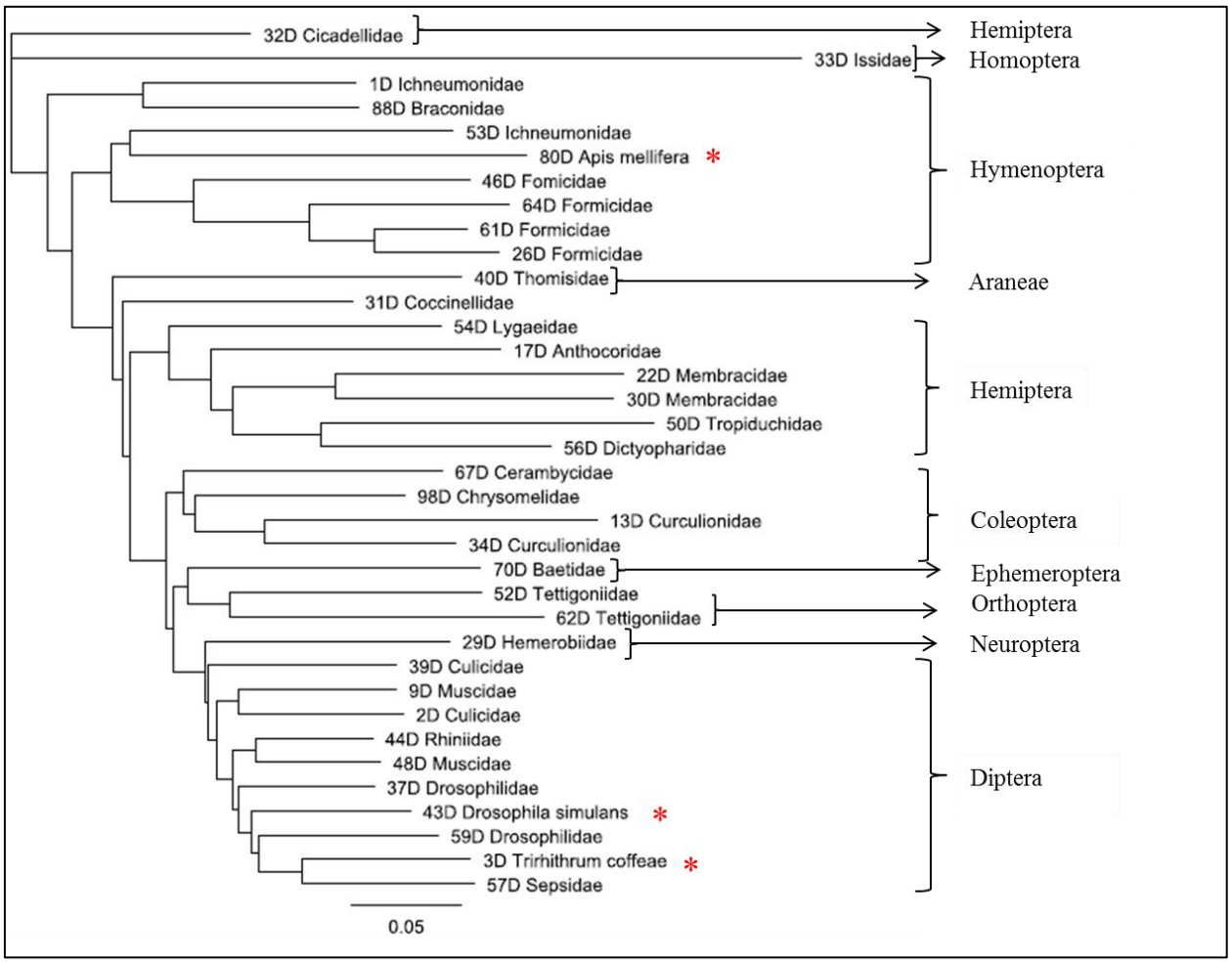

32D Cicadellidae
33D Issidae
1D Ichneumonidae
88D Braconidae
53D Ichneumonidae
80D Apis mellifera *
46D Fomicidae
64D Formicidae
61D Formicidae
26D Formicidae
40D Thomisidae
31D Coccinellidae
54D Lygaeidae
17D Anthocoridae
22D Membracidae
30D Membracidae
50D Tropiduchidae
56D Dictyopharidae
67D Cerambycidae
98D Chrysomelidae
13D Curculionidae
34D Curculionidae
70D Baetidae
52D Tettigoniidae
62D Tettigoniidae
29D Hemerobiidae
39D Culicidae
9D Muscidae
2D Culicidae
44D Rhiniidae
48D Muscidae
37D Drosophilidae
43D Drosophila simulans *
59D Drosophilidae
3D Trirhithrum coffeae *
57D Sepsidae
0.05
Hemiptera
Homoptera
Hymenoptera
Araneae
Hemiptera
Coleoptera
Ephemeroptera
Orthoptera
Neuroptera
Diptera

Figure 2a.

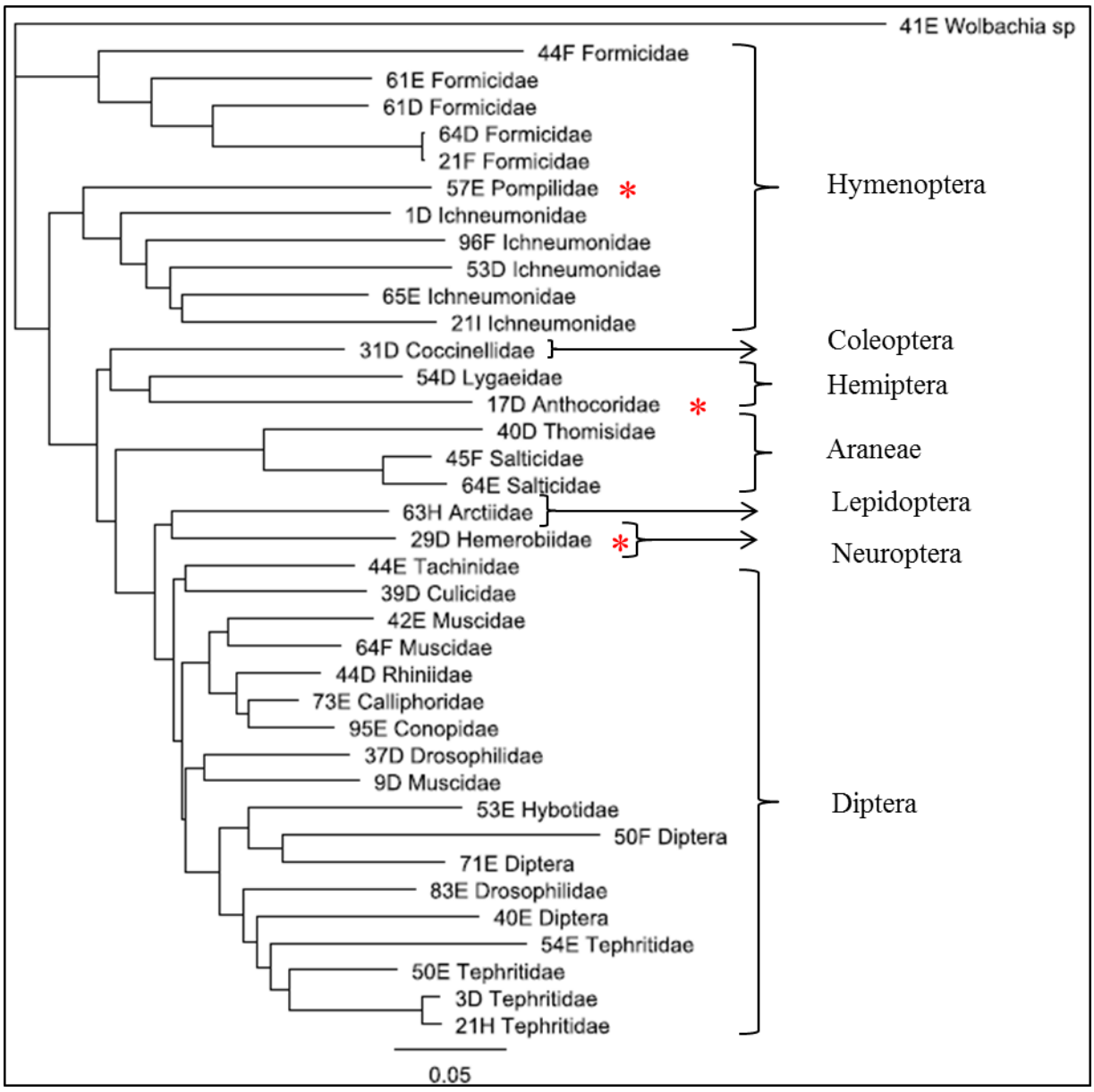

41E Wolbachia sp
44F Formicidae
61E Formicidae
61D Formicidae
64D Formicidae
21F Formicidae
57E Pompilidae *
1D Ichneumonidae
96F Ichneumonidae
53D Ichneumonidae
65E Ichneumonidae
21I Ichneumonidae
31D Coccinellidae
54D Lygaeidae
17D Anthocoridae *
40D Thomisidae
45F Salticidae
64E Salticidae
63H Arctiidae
29D Hemerobiidae *
44E Tachinidae
39D Culicidae
42E Muscidae
64F Muscidae
44D Rhiniidae
73E Calliphoridae
95E Conopidae
37D Drosophilidae
9D Muscidae
53E Hybotidae
50F Diptera
71E Diptera
83E Drosophilidae
40E Diptera
54E Tephritidae
50E Tephritidae
3D Tephritidae
21H Tephritidae
0.05
Hymenoptera
Coleoptera
Hemiptera
Araneae
Lepidoptera
Neuroptera
Diptera

Figure 2b.

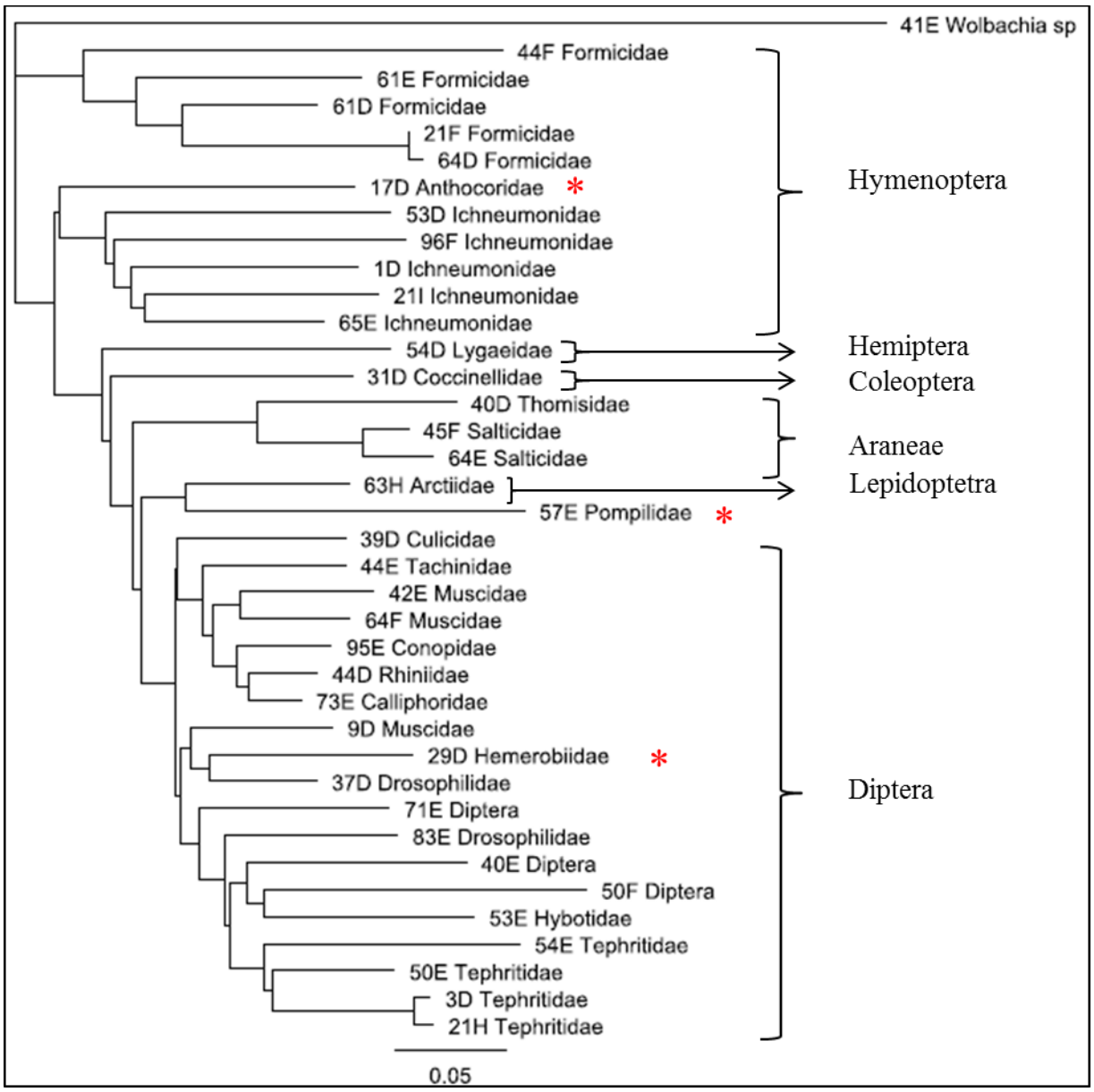
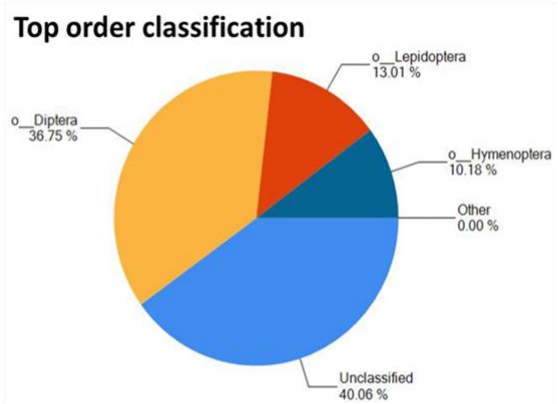
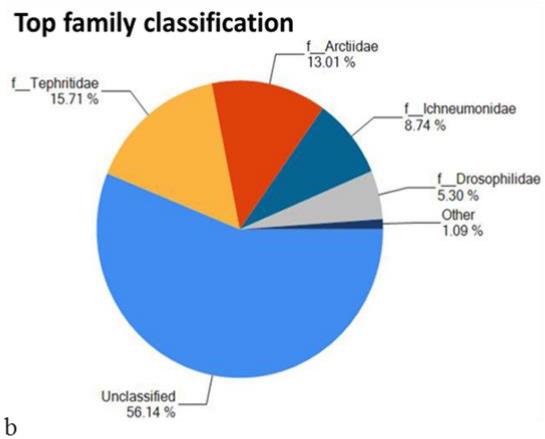

41E Wolbachia sp
44F Formicidae
61E Formicidae
61D Formicidae
21F Formicidae
64D Formicidae
17D Anthocoridae *
53D Ichneumonidae
96F Ichneumonidae
1D Ichneumonidae
21I Ichneumonidae
65E Ichneumonidae
54D Lygaeidae
31D Coccinellidae
40D Thomisidae
45F Salticidae
64E Salticidae
63H Arctiidae
57E Pompilidae *
39D Culicidae
44E Tachinidae
42E Muscidae
64F Muscidae
95E Conopidae
44D Rhiniidae
73E Calliphoridae
9D Muscidae
29D Hemerobiidae *
37D Drosophilidae
71E Diptera
83E Drosophilidae
40E Diptera
50F Diptera
53E Hybotidae
54E Tephritidae
50E Tephritidae
3D Tephritidae
21H Tephritidae
0.05
Hymenoptera
Hemiptera
Coleoptera
Araneae
Lepidoptetra
Diptera

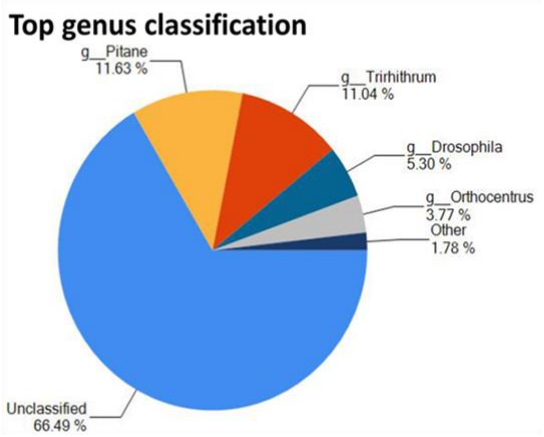
Figure 3



a



b



c

Top Species Classification Results

Classification	Number of Reads	% Total Reads
Unclassified at Species level	43,669	67.06 %
g__Pitane s__fervens	7,574	11.63 %
g__Trirhithrum s__nigerrimum	6,853	10.52 %
g__Orthocentrus Unclassified	2,452	3.77 %
g__Drosophila s__melanogaster	969	1.49 %
g__Amata s__phegea	448	0.69 %
g__Rhinia Unclassified	362	0.56 %
g__Zanzopsis Unclassified	351	0.54 %

d

Figure 4

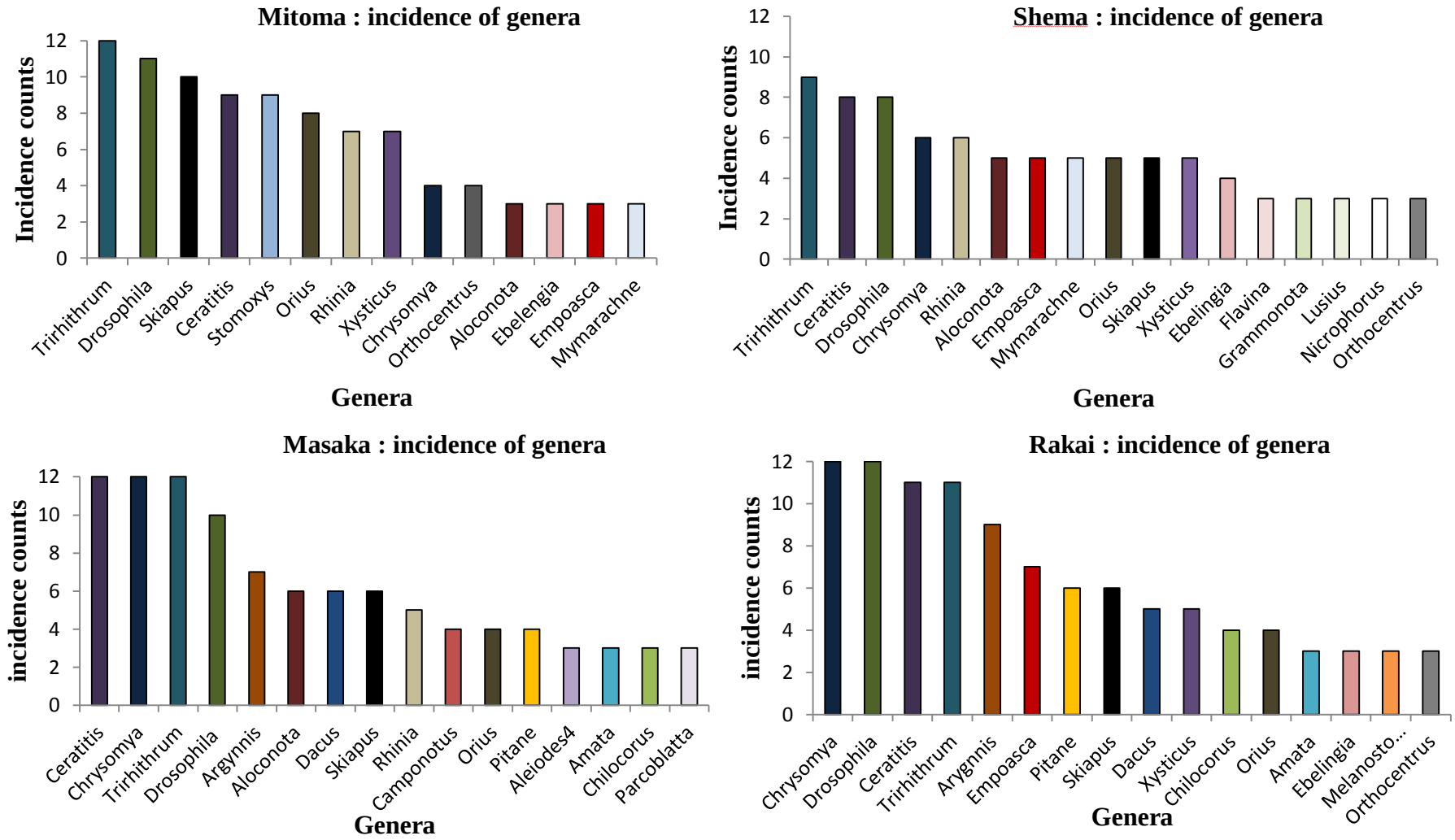
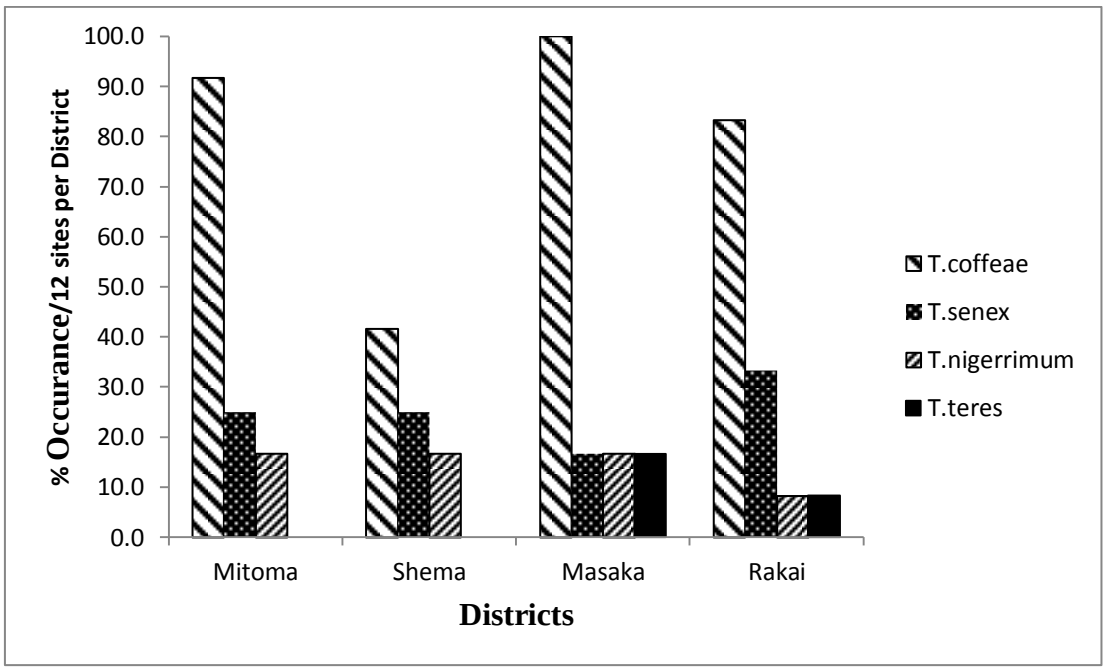

Mitoma : incidence of genera
Incidence counts
Genera
Trirhithrum
Drosophila
Skiapus
Ceratitis
Stomoxys
Orius
Rhinia
Xysticus
Chrysomya
Orthocentrus
Aloconota
Ebelengia
Empoasca
Mymarachne
Shema : incidence of genera
Incidence counts
Genera
Trirhithrum
Ceratitis
Drosophila
Chrysomya
Rhinia
Aloconota
Empoasca
Mymarachne
Orius
Skiapus
Xysticus
Ebelingia
Flavina
Grammonota
Lusius
Nicrophorus
Orthocentrus
Masaka : incidence of genera
incidence counts
Genera
Ceratitis
Chrysomya
Trirhithrum
Drosophila
Argynnis
Aloconota
Dacus
Skiapus
Rhinia
Camponotus
Orius
Pitane
Aleiodes4
Amata
Chilocorus
Parcoblatta
Rakai : incidence of genera
incidence counts
Genera
Chrysomya
Drosophila
Ceratitis
Trirhithrum
Argynnis
Empoasca
Pitane
Skiapus
Dacus
Xysticus
Chilocorus
Orius
Amata
Ebelingia
Melanosto...
Orthocentrus

Figure 5


100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
% Occurance/12 sites per District
Mitoma
Shema
Masaka
Rakai
Districts
T.coffeae
T.senex
T.nigerrimum
T.teres

8. Appendices

8.1.Appendix A

Alignment output of *Ceratitis anonae* and *Ceratitis rosa* extracted from an NCBI blast search of the *Ceratitis rosa* vouchered COI reference sequence (gi|407912445).

Fig3 An alignment extract from the NCBI blast search output showing the percentage identity of *Ceratitis anonea and Ceratitis rosa*. Red markings highlight the sites of polymorphism between the aligned sequences.

8.2. Appendix B

A summary table of the output from the Miseq sequencing run showing the cluster density, number of reads generated, reads passing the filter, percentage of the reads above the Q30 quality score and total yield.

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Phas/Prephas (%)	Reads (M)	Reads PF (M)	% > Q30	Yield	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
Read 1																
1	38	705 +/- 72	93.23 +/- 1.64	0.183 / 0.148	18.84	17.53	88.4	5.3 G	300	8.1 +/- 0.7	2.38 +/- 0.08	0.23 +/- 0.06	0.24 +/- 0.04	0.30 +/- 0.03	154 +/- 22	97.5 +/- 4.2
Read 2 (I)																
1	38	705 +/- 72	93.23 +/- 1.64	0.000 / 0.000	18.84	17.53	83.9	122.7 M	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	345 +/- 67	0.0 +/- 0.0
Read 3 (I)																
1	38	705 +/- 72	93.23 +/- 1.64	0.000 / 0.000	18.84	17.53	87.6	122.7 M	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	173 +/- 29	0.0 +/- 0.0
Read 4																
1	38	705 +/- 72	93.23 +/- 1.64	0.046 / 0.099	18.84	17.53	77.6	5.3 G	300	8.1 +/- 0.7	2.76 +/- 0.19	0.14 +/- 0.02	0.25 +/- 0.02	0.32 +/- 0.02	146 +/- 23	92.1 +/- 2.0