**High-throughput detection of Drosophila suzukii within mixed trap catches using DNA Metabarcoding**

Alexander M. Piper1,2, Paul Cunningham1, 2, Noel O.I. Cogan1,2, Mark J. Blacket1

1 Agriculture Victoria Research, AgriBio Centre, 5 Ring Road Bundoora 3083, Victoria, Australia

2 School of Applied Systems Biology, La Trobe University, Bundoora 3083, Victoria, Australia

**Corresponding author:**

**Alexander M. Piper**

Email: [alexander.piper@agriculture.vic.gov.au](mailto:alexander.piper@agriculture.vic.gov.au)

**Running title:** Metabarcoding of spotted wing drosophila

**Target journal:**

* Communications Biology
* Scientific Reports
* Pest management science
* Biological invasions

# Introduction

The dual influences of anthropogenic dispersal and climate change are increasing the rate of insect pest invasion, creating a global burden on food security 1. A particularly striking example of this phenomenon is the global invasion of Spotted Wing Drosophila (SWD), *Drosophila suzukii* (Matsumara), a newly significant worldwide pest of small and stone fruits 2. Considered endemic to Japan, SWD has rapidly spread throughout eastern Asia to Europe, North America, and South America 2. Whereas most drosophila flies are considered pests only to households and the fermentation industry, SWD has evolved preferences for fruit that is still on the tree 3, using a serrated ovipositor to penetrate and lay eggs inside these harder fruits. SWD was first detected in North America on Californian berry crops in 2008 4 and has since spread throughout the north central and Eastern United States 2. A similar case of highly efficient dispersal has been seen in Western Europe 5. This rapid establishment and spread is attributed to the species high reproductive potential and multiple generations per year 6, as well as being facilitated by prolific global fruit trade.

For geographic regions such as Australia where SWD has not yet established, early detection surveillance is becoming increasingly important, as rapid detection increases the likelihood of containment and eradication 7. Whereas in regions where it is considered established, population monitoring focussed trapping is conducted as part of integrated pest management programmes 8. Various lures based on attractants of apple cider vinegar 9, live yeasts 10,11, or synthetic attractants developed from a combination of volatile compounds from the former 12 are used to trap drosophila. However, these traps are not particularly selective for SWD, and catch a diverse range of drosophilids and other insect taxa 8,13. In addition to trapping of adult flies, infested fruit can be crushed in a salt solution to agitate and float any larvae and eggs from to the surface, which can then be collected via filtration for identification 14, however diverse insect groups can be found in fruit, especially if it is collected on the ground. When put in the context of a large-scale surveillance program, this insufficient selectivity has the potential to generate tens of thousands of insects weekly, which must be processed and identified in order to locate a potential incursion of *D. suzukii* of which becomes an extremely labour-intensive task.

While male *D. suzukii* can be easily identified from native Australian drosophila by a single black spot on their wings (a feature also shared by other closely related exotic drosophila *D. biarmipes* and *D. subpulchrella*), however female flies and larval stages require microscopic or molecular methods of identification. To supplement morphological examination, a number of molecular diagnostic assays have been developed that allow identification of all life stages of *D. suzukii* without the requirement for taxonomic expertise. These molecular methods include DNA barcoding 15, real-time PCR 16, PCR-RFLP 17, and Loop mediated Isothermal amplification (LAMP) for infield diagnostics 18. While both molecular and morphological identification methods have played an important role in confirming *D. suzukii* incursions in the past 2, a major shortfall for surveillance programs covering large geographic scales is the time-consuming and costly process of conducting single reactions on individual specimens to obtain an identification. In order to dramatically increase the throughput of insect identification for biosecurity surveillance, recent studies have looked to ‘DNA metabarcoding’, or the coupling of DNA barcoding to next generation sequencing to allow barcodes to be generated in a massively massively-parallel manner 19–21. This process, termed ‘metabarcoding’ generates a large number of individual DNA barcode sequences in a single reaction, enabling the simultaneous identification of all organisms in complexed mixed communities, such as the contents of drosophila trap samples. Moreover, the ability to survey entire communities in a single reaction may enable a truly generic diagnostic system for detection of not just target pests but also other unanticipated species that are not being actively searched for 21–23. The broad taxonomic scope further enables metabarcoding assays to detect not just target species of concern, but also reveal the presence of previously unrecorded introduced species that have been missed by former targeted diagnostic approaches (Batovska et al 2020).

Despite these advantages, for biosecurity and pest management ensuring the accuracy of detections must be a priority, especially in light of new methods for false positives to be introduced. While the use of reference standards, controls, and replication forms a cornerstone of diagnostics applications, their use in many published metabarcoding studies to date have been lacking 24. Flexible scope or methods-based validation has recently been proposed for metabarcoding. This places emphasis on defining critical components of the workflow that are monitored with ongoing QC. We develop a robust bioinformatic pipeline and appropriate QC checkpoints for detection.

In this article we develop a standardised laboratory and bioinformatic methods for metabarcoding based diagnostics, and evaluate its ability for detection and quantification of *D. suzukii* in a large background of native Australian insects. In addition to following best practices in use of negative controls and replication, we use a novel PHMM based approach to generate a synthetic COI sequences for use as a positive control for run-to-run quality control 25. These positive controls can then be used to detect index switching and parameterise bioinformatics pipelines 26. Finally, we demonstrate how DNA metabarcoding can overcome the processing bottleneck for large scale biosecurity surveillance, providing the ability to identify hundreds of thousands of insects weekly. In addition to providing a high-throughput detection method for exotic drosophila, by-catch data from our metabarcoding assay sheds light on the composition of native drosophila communities in Australian orchard environments. We recommend that our metabarcoding assay be immediately incorporated into national diagnostic protocols to allow efficient and effective large-scale surveillance for new introductions of *D. suzukii.*

**Results**

## **Testing of 4 primer sets**

4 primers sets of appropriate size for use with the 2x150bp restrictions of high throughput sequencing platforms were selected from those determined as high performing in (Piper et al 2020). The primer sets chosen were BF1-BR1 27, fwhF2-fwhR2n 28, fwhF2-HexCOIR4 29, fwhF2-Sauron . These has previously identified a range of primers that should perform well on insect pests. While this in-silico evaluation can greatly reduce the scope of potential primers for evaluation, in-silico methods do not consider factors such as dimer production, off target amplification, and template secondary structure which can render primers that seem suitable in-silico unsuitable for field application 27. A subset of n mock communities and n trap communities were chosen to evaluate the primers on. Following sequen

## **Evaluation on field samples**

The best performing primers were found to be both fwhF2-fwhR2n and fwhF2-HexCOIR4. fwhF2-fwhR2n was chosen to go ahead with. For validation of the metabarcoding assay on a real background of Australian drosophila species, field trapping was conducted in a cherry (Preferred host for *D. suzukii*) and stonefruit (Secondary host for *D. suzukii*) orchard located in Victoria, Australia for a total of 10 weeks through the fruit ripening season, with traps collected every 2 weeks

**Quality control checkpoints**

DNA extraction, preparation of mastermix, amplifiifaction were conducted in different rooms using dedicated equipment and pipettes as per EPPO standard PM 7/98 (2019) DNA was extracted as previous, with the exception that lysates of each community were replicated twice at before running through columns, and each finalised extract was replicated 3 times prior to PCR in order to test the contribution of technical replicates to detection success. As insufficient 8bp unique dual indexes were available to facilitate the simultaneous analysis of this many samples, a twin tagging approach was used where 3 modified versions of the forward and reverse tags were created with a 2-4bp inline tag added to the 5’- end of the primer to allow differentiation of each PCR replicate (Supplementary n). These inline tags were designed to be length variable to improve phasing during the critical first cycles of the sequencing process 30 and therefore allow a reduced PhiX spike in. All field samples were sequenced on a 5th of an Illumina NovaSeq 6000 S2 flow cell.

A series of quality control checkpoints were implemented

* OTUs had to have occurred in at least both of the extraction replicates, and 2/3 of the PCR replicates to count as a detection

*Comparison between platforms*

3 miseq runs, producing n, n and n reads passing filter. In addition, a novaseq S2 flowcell lane producing n reads. The number of reads passing all filtering stages can be seen in supplementary figure n. A total of n ASV’s were inferred by DADA2 across the whole run, using pseudo-pooling for increased sensitivity to rare variants. number removed. While this filtering removed a large number of unique sequence variants, the overall abundance of bad sequences was low (supplementary figure n). Higher proportions removed from NovaSeq, which could reflect the issue of robustly modelling error rates with binned quality scores, or overconfidence in G’s. Top hit identity distribution from BLAST search between cleaned ASV’s and reference database, compared to IDTAXA classification with confidence threshold 60

***Evaluation of index switching***

Resulting synthetic sequences varied for 88-92% similarity to any other sequence of the NCBI nucleotide database. This arrived on an index switch rate of .. compared to other methods of index switching which include looking at the ratio of switched to unswitched reads.

**Discussion**

* Comparison between primers. Both fwhF2-fwhR2n, and fwhF2-HexCOIR4 performed well in both bias and community characterisation. fwhF2-fwhR2n was chosen for the rest of the study due to the favourable bias profile towards the target species and the fact it has been previously used in a number of studies and shown promising results. The Sauron878-HexCOIR4 combination showed off target amplification of fungal sequences, and therefore was not chosen for further validation. This is most likely due to the high degeneracy. BF1-BR1 while not showing major problems, has less consistent bias that the fwhF2-fwhR2n
* In vitro validation allowed successful detection of target drosophila species in both mock communities and background. How did it go for quantification?
* When extra tags were incorporated into the primers it did not significantly increase bias.
* Replicate similarity between PCR replicates was high, indicating more biological replicates is more important. On the other hand there was some difference in extraction replicates.
* Replicates and synthetic positive control suggest that this cross contamination was introduced prior to PCR. It would be nice to use synthetic spike ins into each sample to allow better detection of cross contamination or sample switching events. Worth replicating at the DNA extraction stage rather than PCR replicates. Better would be more traps In the field as this would increase detection probability much higher. With move to higher throughput sequencers read depth is becoming less of a problem.
* For the field samples, the samples using the DC lure failed completely, and so did many of the apple cider vinegar. We believe this is due to the highly acidic nature of the attractant, which can cause depurination and thus degredation of DNA molecules 31. On the other hand, the traps that caught insects into propylene glycol adequately preserved the specimens. Studies suggest that this preservation action of propylene glycol can be effective for periods of up to 6 months 32. Future studies should either integrate propylene glycol, perhaps alongside controlled released dispensers such as 33
* Novaseq introduces extra problems compared to miseq – including binned error rates and poly Gs being inserted
* While we have specifically validated this assay for detection of flies In the drosophila suzukii complex, the primers are universal and can be used well beyond this (Thesis chapter 1). For example, this assay could be expanded to other emerging drosophilid pests such as *Zaprionus indianus*, *Zaprionus tuberculatus* and *Chymomyza amoena*.
* Similarly metabarcoding allows the targeting of diverse gene regions for purposes beyond taxonomic assessment. An approach like this could be used in concert with novel pest control strategies. For example, inclusion of primers targeting the medea element could allow monitoring of the spread of a medea based gene drive 34 through the population. Similarly, it could be used to track the prevalence of certain Wolbachia haplotypes, or male only strains in SIT.
* One of the main limitations currently is access to HTS technology and turn around times, particularly for pest management. While metabarcoding can be extremely cost effective when large volumes of samples are available, getting this many samples together on a regular basis can prove a logistical challenge. Promising new advances in improving the error rates of lower cost nanopore technology would get around this 35, allowing democratisation of access to metabarcoding based diagnostic strategies.

# Methods

***Assembly of mock communities and collection of field samples***

To create mock communities of known composition for primer validation, colonies of *D. melanogaster*, *D. simulans*, *D. hydei* and S*captodrosophila lattivitata* were started from individual female drosophila caught in banana baited traps 36 around Victoria, Australia (See supplementary for rearing details). Further Ethanol preserved specimens of *D. suzukii, D. Immigrans, D. serrata, D. busckii D. subpulchrella* and *D. biarmipes* were obtained from Cornell Drosophila Stock Centre, USA, Ehime University Drosophila Species Stock Centre, Japan and the National Institute of Agricultural Botany, UK. Various numbers of the individuals from the above species were combined to form the mock communities (Table 2).

. Red cup traps similar to 37 were employed containing one of 4 attractants (Table ), and in addition larvae were retrieved from recently fallen and damaged fruits. All field collected samples were combined by week for each treatment and orchard, yielding 22 trap samples and 7 larval samples across both orchards for DNA metabarcoding. *D. suzukii*, *D. subpulchrella* and *D. biarmipes* individuals were spiked into n traps (Supplementary n).

## **Evaluation of 4 primer sets**

, from which DNA was extracted using a modified non-destructive version of the Qiagen DNeasy kit. In short communities dried of all ethanol overnight in 15mL falcon tubes, then immersed in lysate buffer and proteinase K, with the volume of buffer used increasing proportionally to the size of the insect community ensuring at least 1mL of buffer was above the communities. The falcon tubes containing insect communities and lysate buffer were then placed in a shaking incubator and incubated at 56c and 220rpm for 24 hours. Lysate was then manually loaded onto 96 well Qiagen DNeasy extraction blocks using a multichannel pipette, and the remainder of the kit protocol was followed within the QiaCube automated DNA purification workstation.

Following DNA extraction, the COI locus was amplified from each community using the 4 candidate primer sets shortlisted from in-silico evaluation (see supplementary information for reaction information) . The amplicons were verified on a 2% w/v agarose gel, then diluted 1:10, with 1uL of the diluted amplicon used for indexing PCR. 7 cycles of Real-time PCR was used to attached 8bp unique dual indexes and sequencing adapters to each of the amplicons, followed by a melt curve to quantify the resulting indexed products. Libraries were pooled in equimolar ratios from the meltcurve data using a Biomek liquid handling robot and pooled libraries were purified using a 0.8:1 ratio of AMPure XP beads. The libraries were then sized and quantified using a 2200 TapeStation (Agilent Technologies) and a Qubit 3.0 Fluorometer (Life Technologies) respectively, and then diluted 7 pM for sequencing on the Illumina MiSeq platform (2 x 150 bp reads).

## **Index switching - validation and synthentic positive control**

Index switching was measured using 2 different methods. Firstly indices were summarised from fastq headers and calculating the ratio of valid (applied during library preparation) to invalid (pairs that could only arise due to switching) combinations 38. Secondly, in the positive control case where a DNA should be unambiguously assignable to a single sample due to it being a synthetic molecule, the sequence was mapped to its expected sample as well as any other samples and the ratio calculated.

To design synthetic positive controls, PHMM’s were derived from the curated alignments of 13 target families, covering 7 orders using the aphid R package and novel sequences were generated from the per base probabilities contained within the profiles. Novel sequences were then checked for absence of stop codons which could cause removal during bioinformatic processing and verified through BLAST searches to be greater than 8% diverged from any sequence on the GenBank nucleotide database for the entire 658bp sequence and within the ~220bp amplified region of candidate primers. To improve the GC balance for synthesis and allow further differentiation from biological sequences the letters PAC short for Positive Amplification Control was spelt in amino acids (CCT GCC TGC) and added to each side of the 658bp synthetic folmer region before synthesis as GeneBlock fragments by Integrated DNA Technologies (Iowa, USA).

## **Bioinformatics analysis**

Raw sequence reads were demultiplexed using bcl2fastq with 0 mismatches, followed by a second round of demultiplexing for the inline tags using Seal in BBTools v38. Demultiplexed sequencing reads were then trimmed of PCR primer sequences using BBDuK in BBTools v38. Sequence quality profiles were used to filter reads with more than one expected errors, a kmer complexity less than 8, or those containing ambiguous ‘N’ Bases and all remaining sequences of size 124bp were then analysed using DADA2 v1.9.3. As error rates can vary between sequencing platforms and libraries, the loess error model of DADA2 was determined separately for each sequencing run and visualised to ensure correct fit before reads were denoised. For the NovaSeq data, due to the use of binned quality scores, the error matrix of nucleotide transitions was manually modified to set each error rate to the max detected from that or a higher quality score for that transition, in order to enforce monotonicity. Following denoising, the inferred ASVs from each sequencing run were combined into a single table and filtered to remove chimeras using the *removeBimeraDenovo* function in DADA2. To further remove any non-specific amplification products and pseudogenes, the ASV’s were aligned against a PHMM of the folmer COI region, and checked for frame shifts and stop codons. Taxonomy was assigned to the n ASVs to the lowest rank possible with a minimum confidence of 60 using the IDTAXA algorithm 39 trained on the database of (Piper et al 2020), followed by further species level assignment of biological and synthetic mock community sequences using exact matching between the query and reference databases.

## **Statistical analysis**

To determine the probability of detecting the target species for a given number of replicates, the Bayesian hierarchical occupancy model of 40 was fit to the dataset, considering each sample as the ‘site’ variable. The probability of DNA presence in each extraction given species presence (stage 1 true positive), the probability of DNA presence in each extraction given species absence (stage 1 false positive), and probability of species detection given presence in extraction (stage 2 true positive), and probability of species detection given absence in extraction (stage 2 false positive) was modelled as a function of covariates such as community size, species diversity and trap type. Bayesian variable selection was used to identify important covariates impacting detection probability at both the DNA extraction, and PCR stages. To ascertain the effects of adding additional biological, extraction and PCR replicates for non-target species, species accumulation curves and jaccard dissimilarity between replicates were calculated using the vegan R package 41.

All replicates were then merged together retaining only those ASV’s that were present in at least 2/3 replicates.

Metabarcoding datasets are inherently compositional, meaning all elements in a sample sum to 1, and only the ratios between elements carry meaning. Compositional datasets violate the assumptions of many standard statistical methods that rely on Euclidean geometry in real space. Therefore In order to account for compositionality, the bias for each primer was calculated as the difference in log-ratios between each taxa and the samples geometric mean using the metacal R package 42.

Species richness, species evenness and faiths phylogenetic diversity was calculated for each trapped community using the picante R package. Association between tips of the phylogeny and covariates such as trap.

Further analysis and plotting of the final ASV table was conducted using Phyloseq 43 and tidyverse packages 44.

## Confirmation of metabarcoding detections

Metabarcoding detections of

**Acknowledgments**

The authors would like to acknowledge Jessica Henneken, Shahidul Islam, Kevin Farnier and Farrukh Baig for assistance in trapping, Lea Rako and Shane McEvey for taxonomic assistance and x for allowing traps placed on their property. We also like to acknowledge the Drosophila Species Stock Center (Cornell University, Ithaca, USA), Ehime University Drosophila stock centre, and EMR UK for providing drosophila specimens used in this work. Field work in this study was conducted on the lands of the Yorta Yorta and Bunurong people for whichh we would like to acknowledge elders past, present and emerging. Financial support for this study was provided by Agriculture Victoria's Improved Market Access for Horticulture programme (CMI105584). A.M.P. is further supported by Australian Government Research Training Program Scholarship.

**Availability of data and materials:**

All code required to reproduce all statistical analyses and generate all figures is contained in the manuscript’s GitHub repository <https://github.com/alexpiper/Drosophila_metabarcoding>

# Author contributions

A.M.P, J.P.C. and M.J.B. conceptualised the study, A.M.P performed all field collections, laboratory procedures, and bioinformatic and statistical analyses. A.M.P. wrote the manuscript with input and supervision from J.P.C, NC, and M.J.B. All authors read and reviewed the final version of the manuscript.

# References

1. Savary, S. *et al.* The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* **3**, 430–439 (2019).

2. Asplen, M. K. *et al.* Invasion biology of spotted wing Drosophila (Drosophila suzukii): a global perspective and future priorities. *J. Pest Sci. (2004).* **88**, 469–494 (2015).

3. Keesey, I. W., Knaden, M. & Hansson, B. S. Olfactory Specialization in Drosophila suzukii Supports an Ecological Shift in Host Preference from Rotten to Fresh Fruit. *J. Chem. Ecol.* **41**, 121–128 (2015).

4. Goodhue, R. E., Bolda, M., Farnsworth, D., Williams, J. C. & Zalom, F. G. Spotted wing drosophila infestation of California strawberries and raspberries: Economic analysis of potential revenue losses and control costs. *Pest Manag. Sci.* **67**, 1396–1402 (2011).

5. Cini, A., Ioriatti, C. & Anfora, G. A review of the invasion of Drosophila suzukii in Europe and a draft research agenda for integrated pest management. *Bull. Insectology* **65**, 149–160 (2012).

6. Silva-Soares, N. F., Nogueira-Alves, A., Beldade, P. & Mirth, C. K. Adaptation to new nutritional environments: larval performance, foraging decisions, and adult oviposition choices in Drosophila suzukii. *BMC Ecol.* **17**, 21 (2017).

7. Liebhold, A. M. *et al.* Eradication of Invading Insect Populations: From Concepts to Applications. *Annu. Rev. Entomol.* **61**, 335–352 (2016).

8. Burrack, H. J. *et al.* Multistate comparison of attractants for monitoring Drosophila suzukii (Diptera: Drosophilidae) in blueberries and caneberries. *Environ. Entomol.* **44**, 704–712 (2015).

9. Landolt, P. J., Adams, T., Davis, T. S. & Rogg, H. Spotted Wing Drosophila, Drosophila suzukii (Diptera: Drosophilidae), trapped with combinations of wines and vinegars. *Florida Entomol.* **95**, 326–332 (2012).

10. Bellutti, N. *et al.* Dietary yeast affects preference and performance in Drosophila suzukii. *J. Pest Sci. (2004).* (2017) doi:10.1007/s10340-017-0932-2.

11. Hamby, K. A. & Becher, P. G. Current knowledge of interactions between Drosophila suzukii and microbes, and their potential utility for pest management. *J. Pest Sci. (2004).* **89**, 621–630 (2016).

12. Cha, D. H., Adams, T., Rogg, H. & Landolt, P. J. Identification and Field Evaluation of Fermentation Volatiles from Wine and Vinegar that Mediate Attraction of Spotted Wing Drosophila, Drosophila suzukii. *J. Chem. Ecol.* **38**, 1419–1431 (2012).

13. Tonina, L. *et al.* Comparison of attractants for monitoring Drosophila suzukii in sweet cherry orchards in Italy. *J. Appl. Entomol.* **142**, 18–25 (2018).

14. Van Timmeren, S., Diepenbrock, L. M., Bertone, M. A., Burrack, H. J. & Isaacs, R. A Filter Method for Improved Monitoring of Drosophila suzukii (Diptera: Drosophilidae) Larvae in Fruit. *J. Integr. Pest Manag.* **8**, 1–7 (2017).

15. Hodgetts, J. *et al.* DNA barcoding for biosecurity: case studies from the UK plant protection program. *Genome* **59**, 1033–1048 (2016).

16. Dhami, M. K. & Kumarasinghe, L. A HRM real-time PCR assay for rapid and specific identification of the emerging pest spotted-wing Drosophila (Drosophila suzukii). *PLoS One* **9**, (2014).

17. Kim, S. S., Tripodi, A. D., Johnson, D. T. & Szalanski, A. L. Molecular Diagnostics of Drosophila suzukii ( Diptera : Drosophilidae ) using PCR-RFLP Molecular Diagnostics of Drosophila suzukii ( Diptera : Drosophilidae ) Using PCR-RFLP. **107**, 1292–1294 (2014).

18. Kim, Y. H., Hur, J. H., Lee, G. S., Choi, M. Y. & Koh, Y. H. Rapid and highly accurate detection of Drosophila suzukii, spotted wing Drosophila (Diptera: Drosophilidae) by loop-mediated isothermal amplification assays. *J. Asia. Pac. Entomol.* **19**, 1211–1216 (2016).

19. Batovska, J. *et al.* Effective mosquito and arbovirus surveillance using metabarcoding. *Mol Ecol Resour* **18**, 32–40 (2018).

20. Kocher, A. *et al.* Vector soup: high-throughput identification of Neotropical phlebotomine sand flies using metabarcoding. *Mol. Ecol. Resour.* **17**, 172–182 (2017).

21. Comtet, T., Sandionigi, A., Viard, F. & Casiraghi, M. DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biol. Invasions* **17**, 905–922 (2015).

22. Simmons, M. *et al.* Active and passive environmental DNA surveillance of aquatic invasive species. *Can. J. Fish. Aquat. Sci.* **73**, 76–83 (2016).

23. Lawson Handley, L. How will the ‘molecular revolution’ contribute to biological recording? *Biol. J. Linn. Soc.* **115**, 750–766 (2015).

24. Zinger, L. *et al.* DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Mol. Ecol.* **28**, 1857–1862 (2019).

25. Hardwick, S. A., Deveson, I. W. & Mercer, T. R. Reference standards for next-generation sequencing. *Nat. Rev. Genet.* **18**, 473–484 (2017).

26. Palmer, J. M., Jusino, M. A., Banik, M. T. & Lindner, D. L. Non-biological synthetic spike-in controls and the AMPtk software pipeline improve fungal high throughput amplicon sequencing data. *PeerJ* 213470 (2017) doi:10.1101/213470.

27. Elbrecht, V. & Leese, F. Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Front. Environ. Sci.* **5**, 11 (2017).

28. Vamos, E. E., Elbrecht, V. & Leese, F. Short COI markers for freshwater macroinvertebrate metabarcoding. 1–20 (2017) doi:10.3897/mbmg.1.14625.

29. Marquina, D., Andersson, A. F. & Ronquist, F. New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. *Mol. Ecol. Resour.* 325688 (2018) doi:10.1111/1755-0998.12942.

30. Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D. & Dangl, J. L. Practical innovations for high-throughput amplicon sequencing. *Nat. Methods* **10**, 999–1002 (2013).

31. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).

32. Nakamura, S., Tamura, S., Taki, H. & Shoda-Kagaya, E. Propylene glycol: a promising preservative for insects, comparable to ethanol, from trapping to DNA analysis. *Entomol. Exp. Appl.* **168**, 158–165 (2020).

33. Larson, N. R., Strickland, J., Shields, V. D. C. & Zhang, A. Controlled-Release Dispenser and Dry Trap Developments for Drosophila suzukii Detection. *Front. Ecol. Evol.* **8**, 1–14 (2020).

34. Buchman, A., Marshall, J. M., Ostrovski, D., Yang, T. & Akbari, O. S. Synthetically engineered *Medea* gene drive system in the worldwide crop pest *Drosophila suzukii*. *Proc. Natl. Acad. Sci.* **0**, 201713139 (2018).

35. Karst, S. M., Ziels, R. M., Kirkegaard, R. H. & Albertsen, M. Enabling high-accuracy long-read amplicon sequences using unique molecular identifiers and Nanopore sequencing. *bioRxiv* 645903 (2019) doi:10.1101/645903.

36. Batista, M. R. D. *et al.* Differential attraction of drosophilids to banana baits inoculated with *Saccharomyces cerevisiae* and *Hanseniaspora uvarum* within a Neotropical forest remnant. *PeerJ* **5**, e3063 (2017).

37. Lee, J. C. *et al.* Evaluation of monitoring traps for Drosophila suzukii (Diptera: Drosophilidae) in North America. *J. Econ. Entomol.* **105**, 1350–1357 (2012).

38. Wilcox, T. M., Schwartz, M. K. & Lowe, W. H. Evolutionary Community Ecology: Time to Think Outside the (Taxonomic) Box. *Trends Ecol. Evol.* 1–11 (2018) doi:10.1016/j.tree.2018.01.014.

39. Murali, A., Bhargava, A. & Wright, E. S. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome* **6**, 140 (2018).

40. Griffin, J. E., Matechou, E., Buxton, A. S., Bormpoudakis, D. & Griffiths, R. A. Modelling environmental DNA data; Bayesian variable selection accounting for false positive and false negative errors. *J. R. Stat. Soc. Ser. C Appl. Stat.* (2019) doi:10.1111/rssc.12390.

41. Oksanen, J. *et al.* The vegan package. *Community Ecol. Packag.* **10**, 631–637 (2007).

42. McLaren, M. R., Willis, A. D. & Callahan, B. J. Consistent and correctable bias in metagenomic sequencing measurements. *bioRxiv* 559831 (2019) doi:10.1101/559831.

43. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, (2013).

44. Wickham, H. The tidyverse. *R Packag. ver. 1.1. 1* (2017).

Figures:

1. In vitro test of 4 primers – PCA and Bias plot (compared to geometric mean) as multifigure (Supp?)
2. Bias of 3 modified primers (Supplementary)
3. Summary of QC statistics / detection probability for targets – Look at the relationship between detection success and total community size, species diversity, species evenness, phylogenetic diversity. Perhaps summaries of regression models here. Hopefully show decreasing detection probability with absolute community size, and not any of the others (see alberdi et al paper)
4. Summary of Detection probability / QC statistics for entire dataset (overlap between replicates with different processing strategies, seq depth, number of ASV’s kept, numbers of reads kept through process (see alberdi et al scrutinizing paper for good example of multifigs of these)
5. Heatmap of detection taxa detected across whole study (Supplementary Information)
6. Circular phylogeny of all species detected in the study, with metadata around the outside (Type of trap, orchard, week, number of samples) – This could also be done in a phylogeny free approach using metacoder: <https://onlinelibrary.wiley.com/doi/full/10.1111/ddi.12972>

Figures:

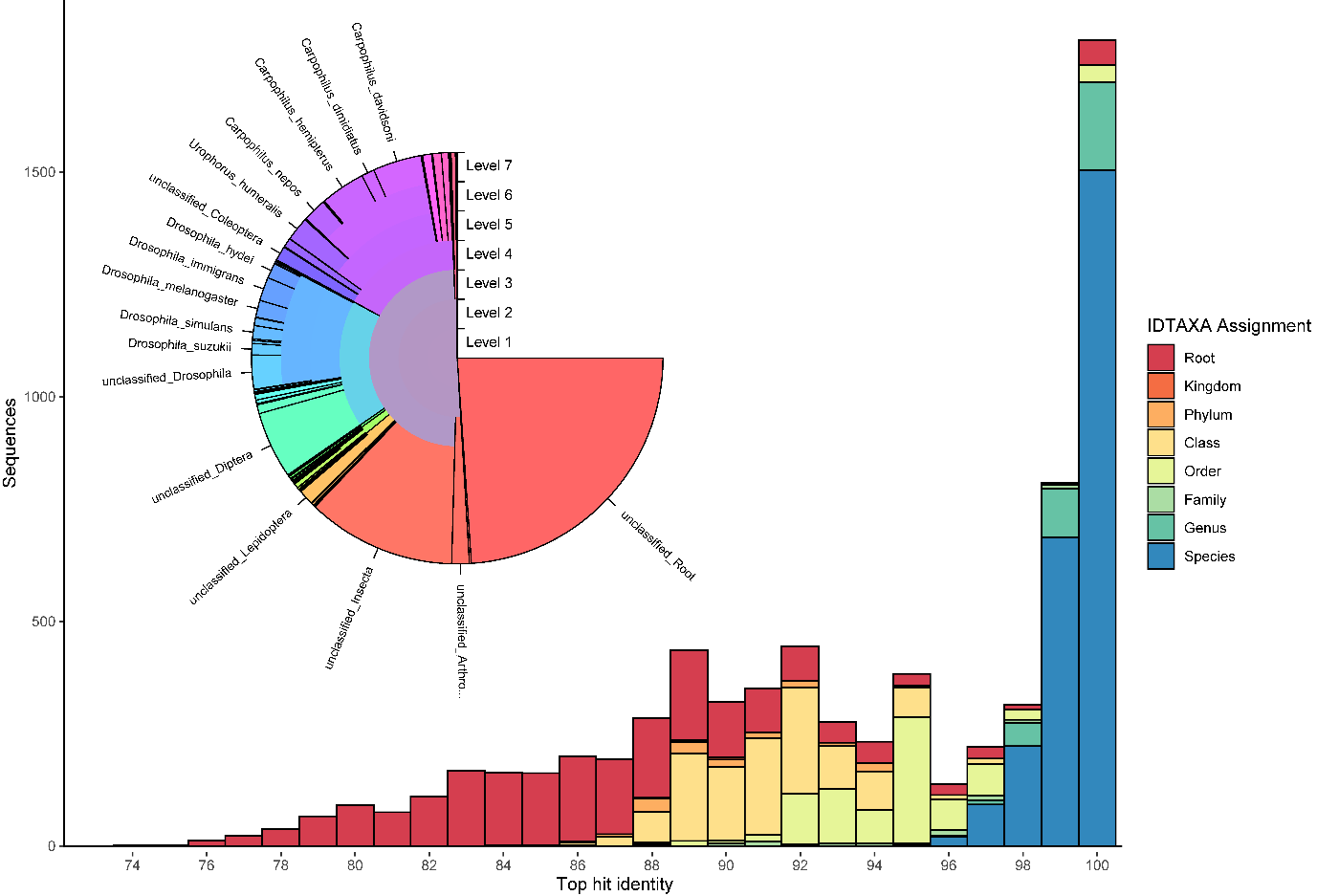


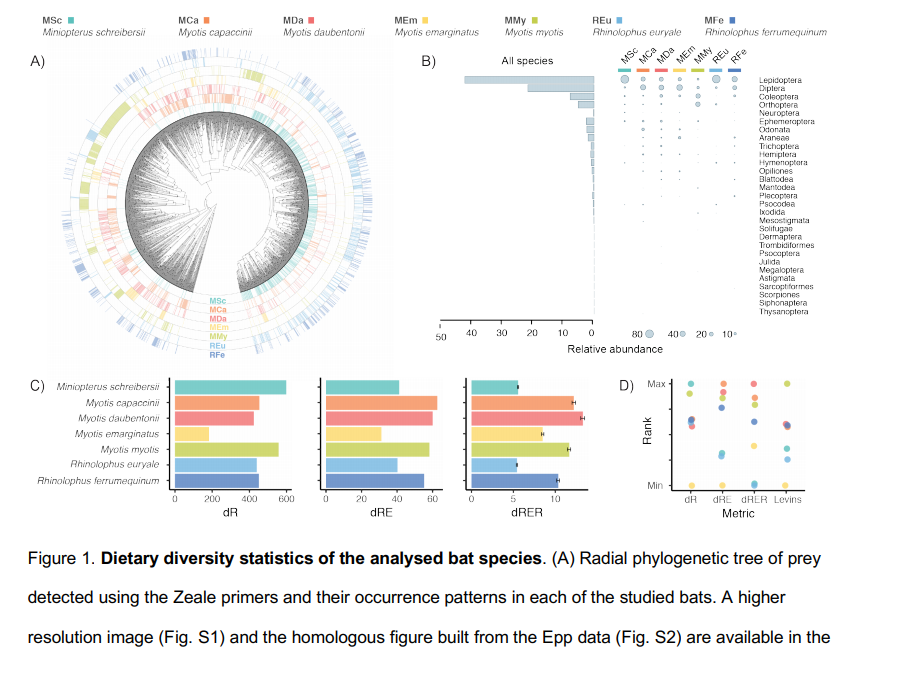
Figure n: Top hit identity distribution from BLAST search between cleaned ASV’s and reference database, compared to IDTAXA classification with confidence threshold 60. Inset - Taxonomy of sequences inferred by IDTAXA algorithm.

A close up of a map

Description generated with very high confidence

A close up of a map

Description generated with high confidence Figure 5- Metabarcoding analysis

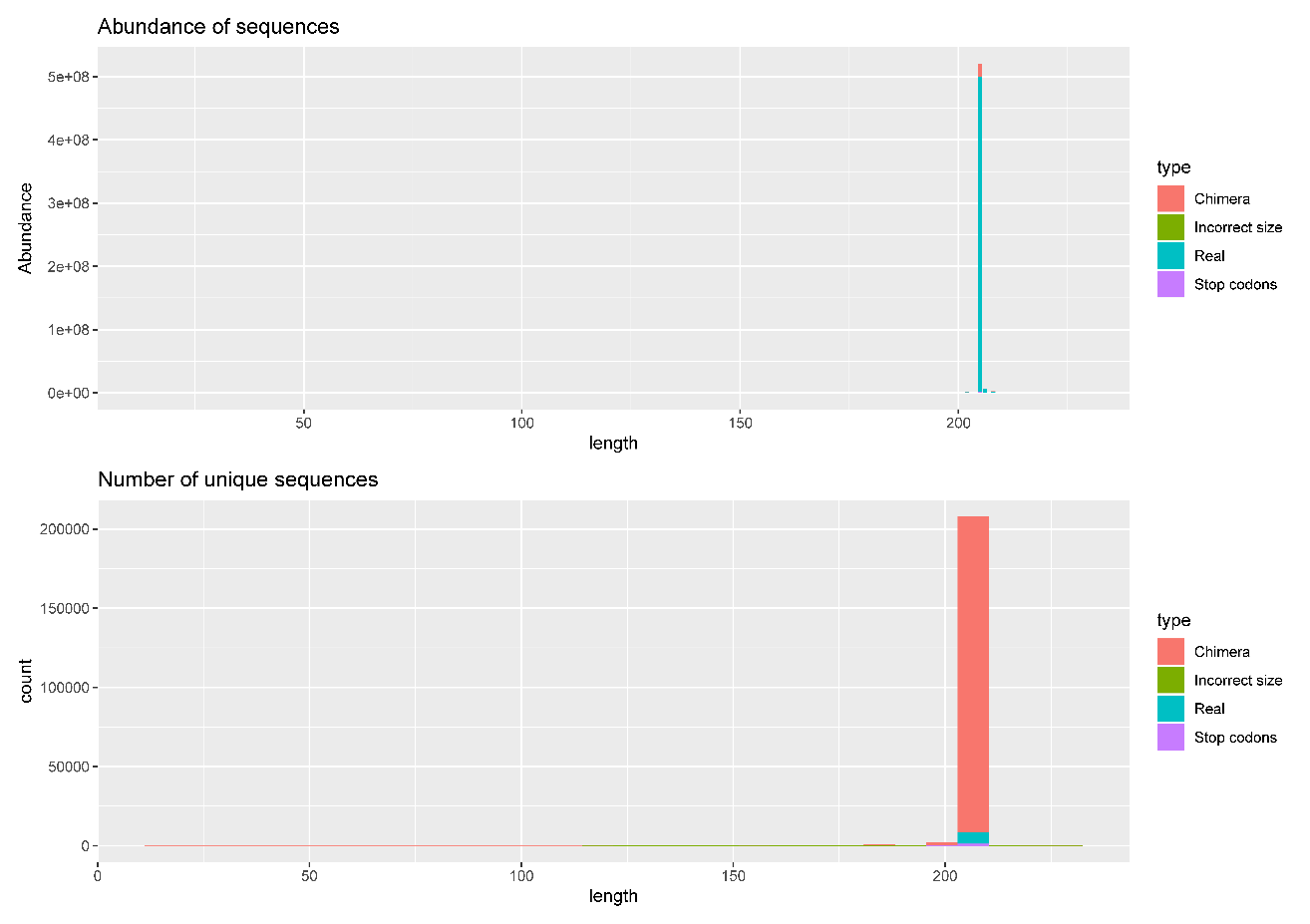


Good example of final summary figure from <https://www.biorxiv.org/content/10.1101/704759v1.full.pdf> . The stats for c ) could by richness, shannons diversity and faiths PD for the different trap

# Supplementary Figures



Supplementary figure n : Drosophila total catch with each attractant by week



Supplementary figure n : Length distribution of inferred sequence variants and their abundance. (b) chimeric sequences represented a high number of unique ASV’s, (a) they however represent a low total abundance.

A close up of a map

Description generated with very high confidence Supplementary figure n : Quantitative performance of 2 candidate primer sets modified with internal tag sequences for higher levels of multiplexing

A picture containing writing implement, stationary

Description generated with very high confidence

Supplementary figure n : Quantitative performance of SynMock positive control over a serial dilution