**Optimisation of metabarcoding primers and protocols for early detection of invasive insect pests**

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

1 Agriculture Victoria Research, AgriBio, Bundoora, Victoria, Australia

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

**Corresponding author:**

**Alexander M. Piper**

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

**Running title:** Metabarcoding for insect pest detection

**Target journal:**

* Methods in Ecology and Evolution
* Molecular Ecology Resources
* Scientific Reports

# **Introduction**

Detection or non-detection can result in costly trade decision

*Drosophila suzukii* (Matsumara), or the Spotted Wing Drosophila is a newly significant worldwide pest of berries and stone fruit. Considered endemic to Japan, *D. suzukii* has recently spread throughout eastern Asia to Europe, North America, and South America (Asplen et al., 2015). While most drosophilid flies prefer rotting fruit, and therefore are considered pests only to households and the fermentation industry, *D. suzukii* prefers ripening fruit while it is still on the tree (Keesey, Knaden, & Hansson, 2015). Female flies have a large serrated ovipositor that allows penetration and egg laying inside these harder fruits, and damage to the fruit is caused by larval feeding inside the fruit surface, as well as oviposition holes exposing the fruit to secondary pathogens (Asplen et al., 2015; Rombaut et al., 2017).

*D. suzukii* was first detected in North America on Californian berry crops in 2008 (Goodhue, Bolda, Farnsworth, Williams, & Zalom, 2011) and has since spread throughout the north central and Eastern United States (Asplen et al., 2015). A similar case of highly efficient dispersal has been seen in Western Europe (Cini, Ioriatti, & Anfora, 2012). This rapid establishment and spread is attributed to the species high reproductive potential and multiple generations per year (Silva-Soares, Nogueira-Alves, Beldade, & Mirth, 2017), as well as being facilitated by prolific global fruit trade. Recent habitat modelling of the potential global distribution *of D. suzukii* suggests that the coastal band of south-eastern and south-western Australia as well as the North island of New Zealand may be susceptible to future invasion by this fly (Dos Santos et al., 2017). Previous risk assessment for introduction of *D. suzukii* into Australia have highlighted the significant potential Impacts on the strawberry, cherry, stone fruit and grape industries, worth a combined $6 billion, and further highlighted that early detection will be vital if eradication is to be successful (Department of Agriculture Fisheries and Forestry, 2013). It is therefore pertinent that rapid and effective detection methods be developed and validated for possible future incursions of *D. suzukii* in Australia.

In regions where it has established, surveillance for *D. suzukii* is generally conducted using attractants of apple cider vinegar (Landolt, Adams, Davis, & Rogg, 2012), live yeasts (Bellutti et al., 2017; Hamby & Becher, 2016), or synthetic attractants developed from a combination of volatile compounds from the former (Cha, Adams, Rogg, & Landolt, 2012). 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 (Van Timmeren, Diepenbrock, Bertone, Burrack, & Isaacs, 2017). The chances of detecting a new incursion of an exotic insect increases with the intensity of surveillance, and this is especially important for pests which have high dispersal and establishment rates (Epanchin-Niell, Haight, Berec, Kean, & Liebhold, 2012; Liebhold et al., 2016; Triska & Renton, 2018). However, the lures used to attract and trap *D. suzukii* are not particularly selective, and catch a diverse range of drosophilids and other insect taxa (ref). 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.

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 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 (Hodgetts et al., 2016), real-time PCR (Dhami & Kumarasinghe, 2014), PCR-RFLP (S. S. Kim, Tripodi, Johnson, & Szalanski, 2014), and Loop mediated Isothermal amplification (LAMP) for infield diagnostics (Y. H. Kim, Hur, Lee, Choi, & Koh, 2016). While both molecular and morphological identification methods have played an important role in confirming *D. suzukii* incursions in the past (Asplen et al., 2015), 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 (Batovska et al., 2018; Comtet, Sandionigi, Viard, & Casiraghi, 2015; Kocher et al., 2017). 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 (Comtet et al., 2015; Lawson Handley, 2015; Simmons et al., 2016).

Despite these advantages, a number of technical and regulatory challenges must first be overcome (Piper et al., 2019). Firstly, while The 658bp region of COI (Folmer, Black, Hoeh, Lutz, & Vrijenhoek, 1994) used for conventional DNA barcoding has a strong track record of delivering species level identification of insect pests (Ashfaq, Hebert, & Naaum, 2016), however the most cost-effective HTS platforms impose strict limitations in molecule length that can be sequenced and therefore smaller stretches of the conventional barcode loci or ‘mini-barcodes’ must be used (Brandon-Mong et al., 2015). This has resulted in the proliferation of new primer sets that have not yet been validated for effectiveness, or for their ability to diagnose pest insect species. Furthermore, 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 (Zinger et al., 2019).

In this article we develop a standardised laboratory and bioinformatic methods for the use of metabarcoding for detection of *D. suzukii* in a large background of native Australian insects. We conduct in-silico validation to define the taxonomic groups other than drosophila that this can be relied upon. 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 (Hardwick, Deveson, & Mercer, 2017). These positive controls can then be used to detect index switching and parameterise bioinformatics pipelines (Palmer, Jusino, Banik, & Lindner, 2017). 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.*

# **Methods**

*Assembly of databases*

To assemble a global list of insect pests, all records from the EPPO global database (<https://gd.eppo.int/>), EPPO Qbank (<https://qbank.eppo.int>), the Global invasive species database (<http://www.iucngisd.org/gisd/>), the Global register of introduced or invasive species (<http://www.griis.org/>), CABI invasive species compendium (https://www.cabi.org/cpc), VectorBase (<https://www.vectorbase.org/>), Australian department of agriculture top 40 priority pests (<http://www.agriculture.gov.au/pests-diseases-weeds/plant>), Plant health Australia national biosecurity status report (<http://www.planthealthaustralia.com.au/national-programs/national-plant-biosecurity-status-report/>), Delivering invasive alien species inventories Europe (<http://www.europe-aliens.org>), and the list from (Ashfaq et al., 2016). This list was filtered to retain only those taxa with genus species binomials and remove duplicate entries. Higher taxonomic levels for all species on the list were retrieved from the NCBI and GBIF reference taxonomies using the R package taxize (S. A. Chamberlain & Szöcs, 2013), and any taxa from outside the class Insecta were removed. All COI records were retrieved from BOLD and GenBank for the unique pest families present on this list using the bold (S. Chamberlain, 2017) and Rentrez (Winter, David, 2019) packages respectively. Following download all sequences with duplicated accession numbers were removed and sequences went through a series of filtering stages.

Due to the potential for non-homologous sequences to be misannotated as COI genes in public databases, all sequences were aligned against a profile hidden markov model (PHMM) (Eddy, 1998) of the COI gene and only those aligning with a score of over 100 using the viterbi algorithm (HOLMES & DURBIN, 2009) were retained. To assemble this model, the Midori-unique (Machida, Leray, Ho, & Knowlton, 2017) dataset was filtered to retain only sequences of the class Insecta and then trimmed to the standard DNA barcoding region (Folmer et al., 1994) using the insect R package (S. P. Wilkinson, Davy, Bunce, & Stat, 2018), and a PHMM was derived using the aphid R package (S. Wilkinson, 2019). During alignment, all bases outside the 658bp standard barcode region were removed. The second filter was to remove homologous sequences, but with misannotated taxonomy. To do this, all sequences were clustered at 99% and species level taxonomies compared within clusters, any sequences that differs from at least 80% of other sequences in the cluster was removed from the dataset Next, to reduce reference bias due to incomplete sampling of the taxonomic tree, all large groups of taxa were pruned to 3 representatives of each species, with sequences discarded from smallest to largest until only 3 sequences remained. Next, all sequences that did not have species genus binomials, were below 200bp of length, contained ambiguous bases or contained non-alphabet characters were removed.

*Identification of optimal diagnostic windows*

To identify optimal diagnostic windows for placement of mini-barcodes within the 658bp folmer region, a sliding window analysis was conducted on the curated reference sequences for each family using the SPIDER r package (Brown et al., 2012) and the number of diagnostic nucleotides and proportion of species that were monophyletic were summarised for each 220bp window. A density plot of the 10 highest ranked window positions within each family was then used to identify globally optimal regions for diagnostics of all pest insect families on the list. Published primer sets overlapping these regions were retrieved, and a number of additional sets were designed using Primer3 implemented in Geneious Prime v2019.2 (Geneious, New Zealand).

*Evaluation of primer sets*

To evaluate the resolution of published primer and novel primer sets across pest taxa, sequence alignments for each pest family were trimmed to the amplified region of each primer set using the insect R package (S. P. Wilkinson et al., 2018), and pairwise genetic distances were generated using the raw model in SPIDER. In order to summarise the taxonomic groups likely to be misidentified or ambiguously identified when using each mini-barcode primer set, summary statistics of identification success were generated using the nearNeighbour, BestCloseMatch, and reciprocal monophyly identification methods in SPIDER, and those failing identification were further explored using neighbour joining trees. In addition to the resolution of primer sets, the level of mismatch between the primers and template must be considered when selecting a primer set as this is thought to be the primary driver of bias in quantification and detection success (Piñol, Mir, Gomez-Polo, & Agustí, 2015; Piñol, Senar, & Symondson, 2019). To evaluate the primers for mismatch, a mismatch score for each primer and insect family was derived using the in-silico PCR function in PrimerMiner (Elbrecht & Leese, 2017a). While highly degenerate primers reduce primer-template mismatch and therefore perform well for the above measure, high degeneracy can also cause undesired off target amplification (Mioduchowska, Jan, Gołdyn, Kur, & Sell, 2018). To evaluate possible off-target effects of the primers, the primer tree R package (Cannon et al., 2016) was used to pick 1000 sequences at random from the trimmed datasets for each primer set and conduct a primer blast against the ncbi nt database, and neighbour joining trees were conducted to summarise the taxonomic scope of each primer.

*Evaluation of candidate primers against mock communities*

While 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 (Elbrecht & Leese, 2017b). Therefore, shortlisted primers from in-silico validation were further tested in metabarcoding assays on mock communities of known composition. To create these mock communities, isofemale lines (David et al., 2005) of *D. melanogaster*, *D. simulans*, *D. hydei* and S*captodrosophila lattivitata* were started from individual female drosophila caught in banana baited traps (Batista et al., 2017) around Victoria, Australia. F1 offspring from all isofemale lines were identified via DNA barcoding using the primers described by (Folmer et al., 1994). Following identification, all isofemale lines that were identified to be from the same species were combined to form larger ongoing colonies. *D. melanogaster*, *D. simulans* and *D. hydei* were maintained at 25c on a diet of instant drosophila medium (Carolina Biological Supply) and live brewer’s yeast, while *S. lattivitata* was maintained at 25c on a Scaptodrosophila specific diet described by Bock & Parsons (1980). Specimens of adult flies were collected weekly into absolute ethanol and a subset of which were further barcoded to confirm colony purity. 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), 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. Each 25µl reaction consisted of 5 μL 5X MyFi reaction buffer, 1uL of 10 nM forward and reverse primer, 0.8 μL MyFi DNA polymerase, 11.2 μL BSA and 2μL of variable concentration template DNA. Cycling conditions were 94°C for 2 min, 30 cycles of 94°C for 30s, 50°C for 45s, and 72°C for 45s, followed by a final extension step of 2 min at 72°C. 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).

*Design of SynMock positive control*

To design synthetic positive controls that would be commutable, and representative of the real biological diversity expected to be encountered In a field situation. 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).

*Field sampling*

For validation of the metabarcoding assay on a real background of Australian drosophila species, field trapping was conducted in 2 orchards located in Victoria, Australia. For field trapping of drosophila, red cup traps with holes similar to (Lee et al., 2012) were used, containing one of 4 attractants (Table 4). 57 Traps were deployed in a cherry orchard located in Mornington, Victoria in transects with alternating treatments at each step in the transect (Fig 7A). Trap catches were collected, and attractants replaced every 2 weeks for a total period of 10 weeks. A stone fruit orchard located in Tatura, Victoria was selected to provide comparison of native communities associated with an alternative *D. suzukii* host. Traps were deployed consecutively as different varieties of stone fruit in the orchard ripened over the course of a 10-week trapping period. Traps were deployed as sets with 3 treatments around a single tree and trap catch collected and attractants replaced every 2 weeks (Fig 7B). Trap set 1-3 were deployed in a row of Ice Princess Peaches and Trap set 4-6 were deployed in a row of August Flame Peaches and were maintained for 10 weeks total. Trap set 7-10 were deployed in a row of Autumn Bright Nectarines and Trap set 11-14 were deployed in a row of September Bright Nectarines and maintained for 6 weeks total. Finally, trap set 15-18 were deployed in a further row of September Bright Nectarines that still retained ripe fruit on the tree and maintained for 2 weeks total. In addition to trapping of adult specimens, recently fallen and damaged fruits were collected every 2 weeks from the trees surrounding the deployed traps to retrieve larvae. These fruits were crushed in salt solution, and drosophila larvae agitated out and collected using methods described by (Van Timmeren et al., 2017). Metabarcoding of larvae retrieved from fruit provides an alternative method of monitoring to field trapping, as larvae are the most common life stages to be intercepted in fruit produce.

Field samples were combined by week for each treatment and orchard, yielding 22 trap samples and 7 larval samples across both orchards. 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 probability. As insufficient 8bp unique dual indexes were available to facilitate the simultaneous analysis of this many samples, a twin tagging approach was used (Axtner et al., 2019). 3 modified versions of each PCR primer were produced, with a 2-4bp inline tag added to the 5’- end of the primer to allow differentiation of each PCR replicate. These inline tags were designed to be length variable to improve phasing during the critical first cycles of the sequencing process (Lundberg, Yourstone, Mieczkowski, Jones, & Dangl, 2013) and therefore allow a reduced PhiX spike in.

*Bioinformatics analysis*

Raw sequence reads were demultiplexed using bcl2fastq with 0 mismatches. Due to the use of a twin tagging strategy, a second round of demultiplexing was conducted using Seal in BBTools v38 to demultiplex each sample into individual replicates using the inline tag incorporated each primer. Demultiplexed sequencing reads were then trimmed of all non-biological sequences such as PCR primers and sequencing adapters using BBDuK in BBTools v38. Sequence quality profiles were used to filter reads with more than two expected errors or ambiguous ‘N’ bases and truncated at the first quality score below two. All remaining sequences >100 bp were then analysed using DADA2 v1.9.3, which produces amplicon sequence variants (ASVs) using a quality aware error model. As error rates can vary between flow cells and libraries, the DADA2 error model was determined separately for each sequencing run and visualised to ensure correct fit before reads were denoised. Following denoising, the inferred ASVs from each run were combined into a single table, which was further filtered to remove chimeras and collapse identical sequences with only length variation. The Kulcynski distance between replicates was calculated using the vegan R package (Oksanen et al., 2007) and any replicates with distance greater than 0.5 from the other 2 replicates were removed as they represented likely cross-contamination. All remaining replicates were then merged together retaining only those ASV’s that were present in at least 2/3 replicates. Taxonomy was assigned to the n ASVs to the lowest rank possible with a minimum confidence of 60 using the IDTAXA algorithm (Murali, Bhargava, & Wright, 2018), followed by further species level assignment of biological and synthetic sequences using exact matching between the query and validated in-house reference sequence. As the taxonomic training sets only covered Arthropoda, all sequences that could not be reliably assigned to this phylum were excluded from analysis, and COI sequences were aligned using MACSE v2.01 (Ranwez, Harispe, Delsuc, & Douzery, 2011) to further identify and remove any sequences containing frame shifts and stop codons that commonly indicate pseudogenes. Further analysis and plotting of the final ASV table was conducted using Phyloseq (McMurdie & Holmes, 2013) and tidyverse packages (Wickham, 2017).

# **Results**

Database assembly

In total, n unique taxa were retrieved from databases (supplementary fig 1) and following filtering n unique species across n families and n orders were retained. All sequences for these families were retrieved from bold and genbank, for a total of n and n sequences respectively. Following all stages of filtering, a total of n sequences were retained.

*Identifying optimal diagnostic windows*

220bp sliding window analysis of diagnostic nucleotides proportion of monophyletic trees identified 2 regions located 255-475bp and 340-560bp into the folmer DNA barcode region for which the highest density of top 10 ranked windows for all families on the pest list occurred. A range of published primers overlapping these regions were retrieved from the literature, and supplemented with novel primers (Table 1).

*Evaluation of primer sets*

All primer pairs showed comparable resolution across the, with the exception of the BF1-AgpestR2 primer (Figure 2). For all those that failed, generally the result of synonyms and indicate the need for curation of synonyms in reference sequence databases. When taxonomic synonyms were corrected for, the large majority of remaining failed to differentiate taxa came from the family Tephritidae, for which the full DNA barcode region also struggles with (Jiang, Jin, Liang, Zhang, & Li, 2014; Krosch, Schutze, Strutt, Clarke, & Cameron, 2017).

*In-vitro evaluation of primer sets against mock communities*

DNA was extracted from all mock and trap communities using a non-destructive modified version of the Qiagen DNeasy kits extraction protocol, all communities were soaked in the kit’s extraction buffer and proteinase overnight, and DNA was then purified from the soaking solution using the QiaCube DNA extraction robot according to the protocol of the extraction kit.

To test the importance of replication, multiple aliquots of

This run identified remarkably high levels of index switching (supplementary n), perhaps arising from too much DNA causing overcycling or bubble products. Nevertheless, the results seem sound?

*Developing a synthetic mock community positive control*

Resulting synthetic sequences varied for 88-92% similarity to any other sequence of the NCBI nucleotide database

# References

Ashfaq, M., Hebert, P. D. N., & Naaum, A. (2016). DNA barcodes for bio-surveillance: Regulated and economically important arthropod plant pests. *Genome*, *59*(11), 933–945. doi:10.1139/gen-2016-0024

Asplen, M. K., Anfora, G., Biondi, A., Choi, D. S., Chu, D., Daane, K. M., … Desneux, N. (2015). Invasion biology of spotted wing Drosophila (Drosophila suzukii): a global perspective and future priorities. *Journal of Pest Science*, *88*(3), 469–494. doi:10.1007/s10340-015-0681-z

Axtner, J., Crampton-platt, A., Lisa, A. H., Mohamed, A., Xu, C. C. Y., Yu, D. W., & Wilting, A. (2019). An efficient and robust laboratory workflow and tetrapod database for larger scale environmental DNA studies. *GigaScience*, *8*(4), giz029. doi:10.1093/gigascience/giz029

Batista, M. R. D., Uno, F., Chaves, R. D., Tidon, R., Rosa, C. A., & Klaczko, L. B. (2017). Differential attraction of drosophilids to banana baits inoculated with *Saccharomyces cerevisiae* and *Hanseniaspora uvarum* within a Neotropical forest remnant. *PeerJ*, *5*(March), e3063. doi:10.7717/peerj.3063

Batovska, J., Lynch, S. E., Cogan, N. O. I., Brown, K., Darbro, J. M., Kho, E. A., & Blacket, M. J. (2018). Effective mosquito and arbovirus surveillance using metabarcoding. *Mol Ecol Resour*, *18*(1), 32–40. doi:10.1111/1755-0998.12682

Bellutti, N., Gallmetzer, A., Innerebner, G., Schmidt, S., Zelger, R., & Koschier, E. H. (2017). Dietary yeast affects preference and performance in Drosophila suzukii. *Journal of Pest Science*. doi:10.1007/s10340-017-0932-2

Bock, I. R., & Parsons, P. A. (1980). Culture methods for species of the Drosophila (Scaptodrosophila) coracina group. *D. I. S.*, *55*(January), 147–148.

Brandon-Mong, G.-J., Gan, H.-M., Sing, K.-W., Lee, P.-S., Lim, P.-E., & Wilson, J.-J. (2015). DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin of Entomological Research*, *105*(6), 717–727. doi:10.1017/s0007485315000681

Brown, S. D. J., Collins, R. A., Boyer, S., Lefort, M.-C., Malumbres-Olarte, J., Vink, C. J., & Cruickshank, R. H. (2012). SPIDER: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, *12*, 562–565.

Cannon, M. V., Hester, J., Shalkhauser, A., Chan, E. R., Logue, K., Small, S. T., & Serre, D. (2016). In silico assessment of primers for eDNA studies using PrimerTree and application to characterize the biodiversity surrounding the Cuyahoga River. *Scientific Reports*, *6*(December 2015), 1–11. doi:10.1038/srep22908

Cha, D. H., Adams, T., Rogg, H., & Landolt, P. J. (2012). Identification and Field Evaluation of Fermentation Volatiles from Wine and Vinegar that Mediate Attraction of Spotted Wing Drosophila, Drosophila suzukii. *Journal of Chemical Ecology*, *38*(11), 1419–1431. doi:10.1007/s10886-012-0196-5

Chamberlain, S. (2017). bold: Interface to Bold Systems API. Retrieved from https://cran.r-project.org/package=bold

Chamberlain, S. A., & Szöcs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Research*, *2*, 191. doi:10.12688/f1000research.2-191.v2

Cini, A., Ioriatti, C., & Anfora, G. (2012). A review of the invasion of Drosophila suzukii in Europe and a draft research agenda for integrated pest management. *Bulletin of Insectology*, *65*(1), 149–160.

Comtet, T., Sandionigi, A., Viard, F., & Casiraghi, M. (2015). DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions*, *17*(3), 905–922. doi:10.1007/s10530-015-0854-y

David, J. R., Gibert, P., Legout, H., Pétavy, G., Capy, P., & Moreteau, B. (2005). Isofemale lines in Drosophila: An empirical approach to quantitative trait analysis in natural populations. *Heredity*, *94*(1), 3–12. doi:10.1038/sj.hdy.6800562

Department of Agriculture Fisheries and Forestry. (2013). Final pest risk analysis report for Drosophila suzukii, (April).

Dhami, M. K., & Kumarasinghe, L. (2014). A HRM real-time PCR assay for rapid and specific identification of the emerging pest spotted-wing Drosophila (Drosophila suzukii). *PLoS ONE*, *9*(6). doi:10.1371/journal.pone.0098934

Dos Santos, L. A., Mendes, M. F., Krüger, A. P., Blauth, M. L., Gottschalk, M. S., & Garcia, F. R. M. (2017). Global potential distribution of Drosophila suzukii (Diptera, Drosophilidae). *PLoS ONE*, *12*(3), 1–13. doi:10.1371/journal.pone.0174318

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*. doi:10.1093/bioinformatics/14.9.755

Elbrecht, V., & Leese, F. (2017a). PrimerMiner: an r package for development and in silico validation of DNA metabarcoding primers. *Methods in Ecology and Evolution*, *8*(5), 622–626. doi:10.1111/2041-210X.12687

Elbrecht, V., & Leese, F. (2017b). Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science*, *5*, 11. doi:10.3389/fenvs.2017.00011

Epanchin-Niell, R. S., Haight, R. G., Berec, L., Kean, J. M., & Liebhold, A. M. (2012). Optimal surveillance and eradication of invasive species in heterogeneous landscapes. *Ecology Letters*, *15*(8), 803–812. doi:10.1111/j.1461-0248.2012.01800.x

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299. doi:10.1371/journal.pone.0013102

Goodhue, R. E., Bolda, M., Farnsworth, D., Williams, J. C., & Zalom, F. G. (2011). Spotted wing drosophila infestation of California strawberries and raspberries: Economic analysis of potential revenue losses and control costs. *Pest Management Science*, *67*(11), 1396–1402. doi:10.1002/ps.2259

Hamby, K. A., & Becher, P. G. (2016). Current knowledge of interactions between Drosophila suzukii and microbes, and their potential utility for pest management. *Journal of Pest Science*, *89*(3), 621–630. doi:10.1007/s10340-016-0768-1

Hardwick, S. A., Deveson, I. W., & Mercer, T. R. (2017). Reference standards for next-generation sequencing. *Nature Reviews Genetics*, *18*(8), 473–484. doi:10.1038/nrg.2017.44

Hodgetts, J., Ostojá-Starzewski, J. C., Prior, T., Lawson, R., Hall, J., & Boonham, N. (2016). DNA barcoding for biosecurity: case studies from the UK plant protection program. *Genome*, *59*(11), 1033–1048. doi:10.1139/gen-2016-0010

HOLMES, I., & DURBIN, R. (2009). Dynamic Programming Alignment Accuracy. *Journal of Computational Biology*. doi:10.1089/cmb.1998.5.493

Jiang, F., Jin, Q., Liang, L., Zhang, A. B., & Li, Z. H. (2014). Existence of species complex largely reduced barcoding success for invasive species of Tephritidae: A case study in Bactrocera spp. *Molecular Ecology Resources*, *14*(6), 1114–1128. doi:10.1111/1755-0998.12259

Jordaens, K., Sonet, G., & Richet, R. (2013). Identification of forensically important Sarcophaga species ( Diptera : Sarcophagidae ) using the mitochondrial COI gene, 491–504. doi:10.1007/s00414-012-0767-6

Keesey, I. W., Knaden, M., & Hansson, B. S. (2015). Olfactory Specialization in Drosophila suzukii Supports an Ecological Shift in Host Preference from Rotten to Fresh Fruit. *Journal of Chemical Ecology*, *41*(2), 121–128. doi:10.1007/s10886-015-0544-3

Kim, S. S., Tripodi, A. D., Johnson, D. T., & Szalanski, A. L. (2014). Molecular Diagnostics of Drosophila suzukii ( Diptera : Drosophilidae ) using PCR-RFLP Molecular Diagnostics of Drosophila suzukii ( Diptera : Drosophilidae ) Using PCR-RFLP, *107*(3), 1292–1294.

Kim, Y. H., Hur, J. H., Lee, G. S., Choi, M. Y., & Koh, Y. H. (2016). Rapid and highly accurate detection of Drosophila suzukii, spotted wing Drosophila (Diptera: Drosophilidae) by loop-mediated isothermal amplification assays. *Journal of Asia-Pacific Entomology*, *19*(4), 1211–1216. doi:10.1016/j.aspen.2016.10.015

Kocher, A., Gantier, J. C., Gaborit, P., Zinger, L., Holota, H., Valiere, S., … Murienne, J. (2017). Vector soup: high-throughput identification of Neotropical phlebotomine sand flies using metabarcoding. *Molecular Ecology Resources*, *17*(2), 172–182. doi:10.1111/1755-0998.12556

Krosch, M. N., Schutze, M. K., Strutt, F., Clarke, A. R., & Cameron, S. L. (2017). A transcriptome-based analytical workflow for identifying loci for species diagnosis: A case study with Bactrocera fruit flies (Diptera: Tephritidae). *Austral Entomology*, *58*, 395– 408. doi:10.1111/aen.12321

Landolt, P. J., Adams, T., Davis, T. S., & Rogg, H. (2012). Spotted Wing Drosophila, Drosophila suzukii (Diptera: Drosophilidae), trapped with combinations of wines and vinegars. *Florida Entomologist*, *95*(2), 326–332. doi:10.1653/024.095.0213

Lawson Handley, L. (2015). How will the “molecular revolution” contribute to biological recording? *Biological Journal of the Linnean Society*, *115*(3), 750–766. doi:10.1111/bij.12516

Lee, J. C., Burrack, H. J., Barrantes, L. D., Beers, E. H., Dreves, A. J., Hamby, K. A., … Bruck, D. J. (2012). Evaluation of monitoring traps for Drosophila suzukii (Diptera: Drosophilidae) in North America. *Journal of Economic Entomology*, *105*(4), 1350–1357. doi:10.1603/EC12132

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., … Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity : application for characterizing coral reef fish gut contents, 1–14.

Liebhold, A. M., Berec, L., Brockerhoff, E. G., Epanchin-Niell, R. S., Hastings, A., Herms, D. A., … Yamanaka, T. (2016). Eradication of Invading Insect Populations: From Concepts to Applications. *Annual Review of Entomology*, *61*(1), 335–352. doi:10.1146/annurev-ento-010715-023809

Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., & Dangl, J. L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, *10*(10), 999–1002. doi:10.1038/nmeth.2634

Machida, R. J., Leray, M., Ho, S. L., & Knowlton, N. (2017). Data Descriptor: Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Scientific Data*, *4*, 170027. doi:10.1038/sdata.2017.27

Marquina, D. (2019). New mitochondrial primers for metabarcoding of insects , designed and evaluated using in silico methods, (August 2018), 90–104. doi:10.1111/1755-0998.12942

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4). doi:10.1371/journal.pone.0061217

Meusnier, I., Singer, G. A. C., Landry, J. F., Hickey, D. A., Hebert, P. D. N., & Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, *9*, 4–7. doi:10.1186/1471-2164-9-214

Mioduchowska, M., Jan, M., Gołdyn, B., Kur, J., & Sell, J. (2018). Instances of erroneous DNA barcoding of metazoan invertebrates: Are universal cox1 gene primers too “universal”? *Plos One*, *13*(6), e0199609. doi:10.1371/journal.pone.0199609

Murali, A., Bhargava, A., & Wright, E. S. (2018). IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome*, *6*(1), 140. doi:10.1186/s40168-018-0521-5

Oksanen, J., Kindt, R., Legendre, P., O’Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. (2007). The vegan package. *Community Ecology Package*, *10*, 631–637.

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

Piñol, J., Mir, G., Gomez-Polo, P., & Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, *15*(4), 819–830. doi:10.1111/1755-0998.12355

Piñol, J., Senar, M. A., & Symondson, W. O. C. (2019). The choice of universal primers and the characteristics of the species mixture determines when DNA metabarcoding can be quantitative. *Molecular Ecology*, *28*(May), 407– 419. doi:10.1111/mec.14776

Piper, A. M., Batovska, J., Cogan, N. O. I., Weiss, J., Cunningham, J. P., Rodoni, B. C., & Blacket, M. J. (2019). Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *GigaScience*, *8*(8), 1–22. doi:10.1093/gigascience/giz092

Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. P. (2011). MACSE: Multiple alignment of coding SEquences accounting for frameshifts and stop codons. *PLoS ONE*, *6*(9). doi:10.1371/journal.pone.0022594

Rennstam Rubbmark, O., Sint, D., Horngacher, N., & Traugott, M. (2018). A broadly-applicable COI primer pair and an efficient single tube amplicon library preparation protocol for metabarcoding. *Ecology and Evolution*, *8*, 12335– 12350. doi:10.1016/S1872-2067(12)60761-X

Rombaut, A., Guilhot, R., Xuéreb, A., Benoit, L., Chapuis, M. P., Gibert, P., & Fellous, S. (2017). Invasive Drosophila suzukii facilitates Drosophila melanogaster infestation and sour rot outbreaks in the vineyards. *R. Soc. Open Sci.* doi:10.1098/rsos.170117

Silva-Soares, N. F., Nogueira-Alves, A., Beldade, P., & Mirth, C. K. (2017). Adaptation to new nutritional environments: larval performance, foraging decisions, and adult oviposition choices in Drosophila suzukii. *BMC Ecology*, *17*(1), 21. doi:10.1186/s12898-017-0131-2

Simmons, M., Tucker, A., Chadderton, W. L., Jerde, C. L., Mahon, A. R., & Taylor, E. (2016). Active and passive environmental DNA surveillance of aquatic invasive species. *Canadian Journal of Fisheries and Aquatic Sciences*, *73*(1), 76–83. doi:10.1139/cjfas-2015-0262

Triska, M. D., & Renton, M. (2018). Do an invasive organism’s dispersal characteristics affect how we should search for it? *Royal Society Open Science*, *5*, 171784. doi:10.1098/rsos.171784

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

Van Timmeren, S., Diepenbrock, L. M., Bertone, M. A., Burrack, H. J., & Isaacs, R. (2017). A Filter Method for Improved Monitoring of Drosophila suzukii (Diptera: Drosophilidae) Larvae in Fruit. *Journal of Integrated Pest Management*, *8*(1), 1–7. doi:10.1093/jipm/pmx019

Wickham, H. (2017). The tidyverse. *R Package Ver. 1.1. 1*.

Wilkinson, S. (2019). aphid: an R package for analysis with profile hidden Markov models. *Bioinformatics*. doi:10.1093/bioinformatics/btz159

Wilkinson, S. P., Davy, S. K., Bunce, M., & Stat, M. (2018). Taxonomic identification of environmental DNA with informatic sequence classification trees. *PeerJ Preprints*, *6*, e26812v1. doi:10.7287/peerj.preprints.26812v1

Winter, David, J. (2019). rentrez: An R package for the NCBI eUtils API. *The R Journal*, *9*(2), 520. doi:10.32614/rj-2017-058

Zeale, M. R. K., Butlin, R. K., Barker, G. L. A., Lees, D. C., & Jones, G. (2011). Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*. doi:10.1111/j.1755-0998.2010.02920.x

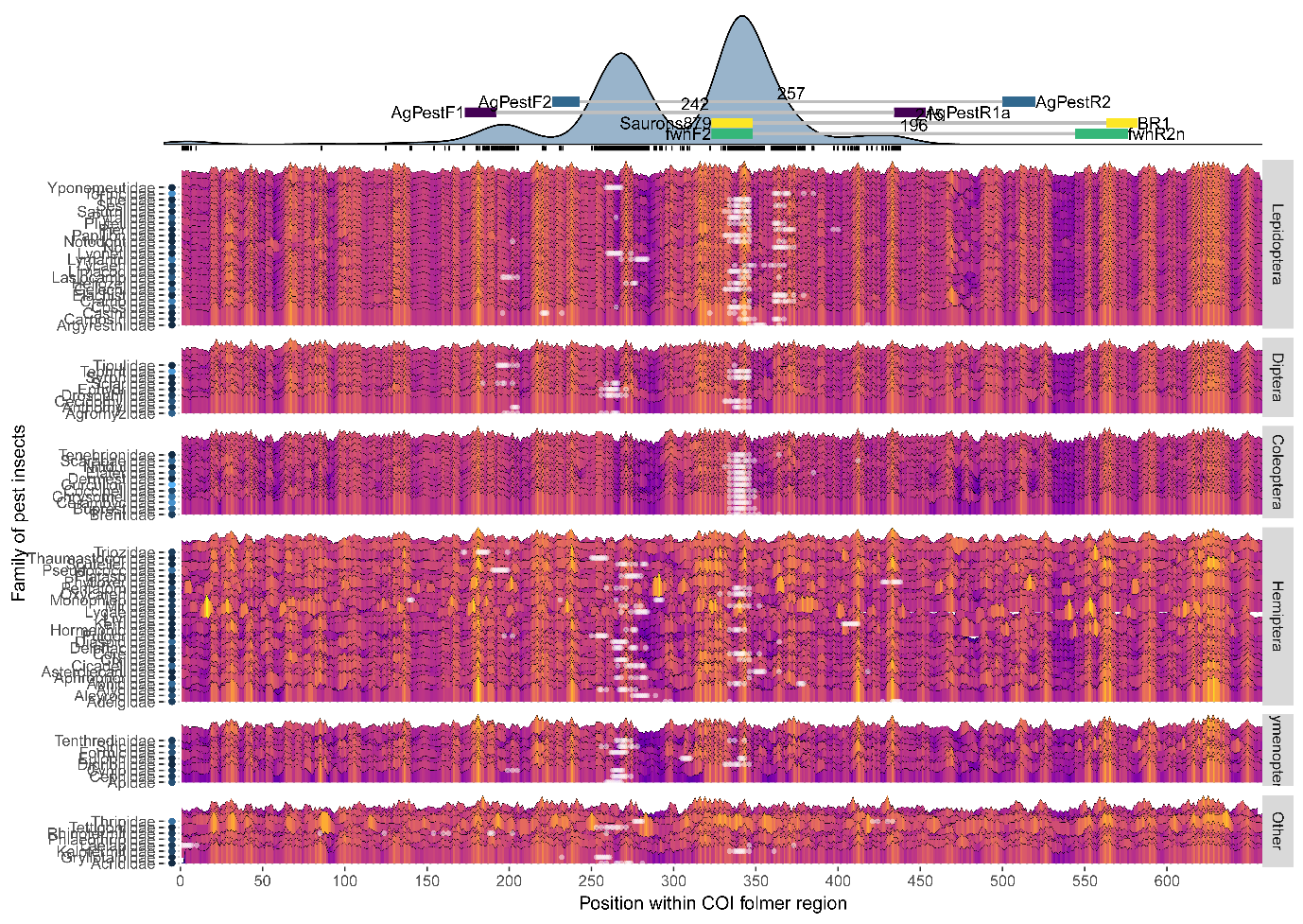
Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., … Taberlet, P. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, *28*(8), 1857–1862. doi:10.1111/mec.15060

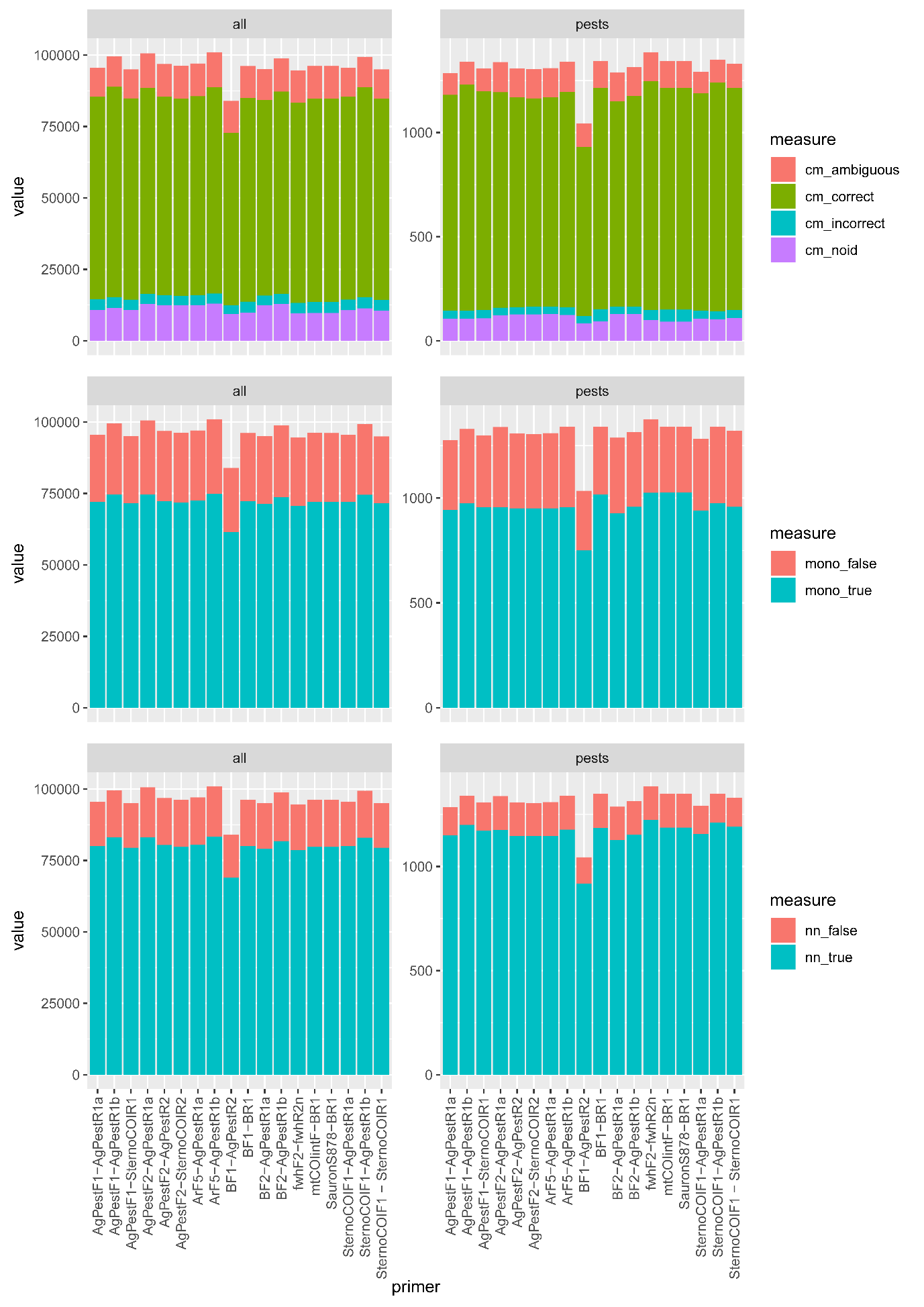
# Tables

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | F.seq | F.Start | R.seq | R.Start | amplicon | study |
| BF1-BR1 | ACWGGWTGRACWGTNTAYCC | 326 | ARYATDGTRATDGCHCCDGC | 563 | 218 | (Elbrecht & Leese, 2017b) |
| fwhF2-fwhR2n | GGDACWGGWTGAACWGTWTAYCCHCC | 323 | GTRATWGCHCCDGCTARWACWGG | 544 | 196 | (Vamos, Elbrecht, & Leese, 2017) |
| fwhF1-fwhR1 | YTCHACWAAYCAYAARGAYATYGG | -23 | ARTCARTTWCCRAAHCCHCC | 182 | 182 | (Vamos et al., 2017) |
| SternoCOIF1 - SternoCOIR1 | ATTGGWGGWTTYGGAAAYTG | 179 | ATRAARTTRATWGCTCCTA | 432 | 234 | Batovska et al (in prep) |
| mtCOIintF-BR1 | GGWACWGGWTGAACWGTWTAYCCYCC | 323 | ARYATDGTRATDGCHCCDGC | 563 | 215 | (Leray et al., 2013) |
| SauronS878-BR1 | GGDRCWGGWTGAACWGTWTAYCCNCC | 323 | ARYATDGTRATDGCHCCDGC | 563 | 215 | (Rennstam Rubbmark, Sint, Horngacher, & Traugott, 2018) |
| Uni-MinibarF1-Uni-MinibarR1 | TCCACTAATCACAARGATATTGGTAC | -22 | GAAAATCATAATGAAGGCATGAGC | 134 | 131 | (Meusnier et al., 2008) |
| UniMinibarF1−d-UniMinibarF1−d | TCYACTAATCATAAAGATATTGGYAC | -22 | AAAATTATAATAAARGCRTGRGC | 134 | 131 | (Jordaens, Sonet, & Richet, 2013) |
| ZBJ−ArtF1c-ZBJ−ArtR2c | AGATATTGGAACWTTATATTTTATTTTTGG | -8 | WACTAATCAATTWCCAAATCCTCC | 182 | 161 | (Zeale, Butlin, Barker, Lees, & Jones, 2011) |
| AgPestF1-AgPestR1a | ATYATWATTGGDGGDTTYGG | 173 | GTRATRAARTTDAYWGMHCC | 434 | 242 | This Study |
| AgPestF1-AgPestR1b | ATYATWATTGGDGGDTTYGG | 173 | ARAATWGADGADAYWCCWGC | 413 | 221 | This Study |
| AgPestF2-AgPestR1a | HGAYATRGCHTTYCCHCG | 226 | GTRATRAARTTDAYWGMHCC | 434 | 191 | This Study |
| AgPestF2-AgPestR1b | HGAYATRGCHTTYCCHCG | 226 | ARAATWGADGADAYWCCWGC | 413 | 170 | This Study |
| AgPestF2-AgPestR2 | HGAYATRGCHTTYCCHCG | 226 | RACWGMTCAVAYAAATARDGG | 500 | 257 | This Study |
| ArF5-AgPestR1a | GCICCIGAYATRKCITTYCCICG | 221 | GTRATRAARTTDAYWGMHCC | 434 | 191 | This Study |
| ArF5-AgPestR1b | GCICCIGAYATRKCITTYCCICG | 221 | ARAATWGADGADAYWCCWGC | 413 | 170 | This Study |
| ArF5-AgPestR2 | GCICCIGAYATRKCITTYCCICG | 221 | RACWGMTCAVAYAAATARDGG | 500 | 257 | This Study |
| BF2-AgPestR1a | GCHCCHGAYATRGCHTTYCC | 221 | GTRATRAARTTDAYWGMHCC | 434 | 194 | This Study |
| BF2-AgPestR1b | GCHCCHGAYATRGCHTTYCC | 221 | ARAATWGADGADAYWCCWGC | 413 | 173 | This Study |
| BF2-AgPestR2 | GCHCCHGAYATRGCHTTYCC | 221 | RACWGMTCAVAYAAATARDGG | 500 | 260 | This Study |
| SternoCOIF1-AgPestR1a | ATTGGWGGWTTYGGAAAYTG | 179 | GTRATRAARTTDAYWGMHCC | 434 | 236 | This Study |
| SternoCOIF1-AgPestR1b | ATTGGWGGWTTYGGAAAYTG | 179 | ARAATWGADGADAYWCCWGC | 413 | 215 | This Study |
| AgPestF1-SternoCOIR1 | ATYATWATTGGDGGDTTYGG | 173 | ATRAARTTRATWGCTCCTA | 432 | 240 | This Study |
| AgPestF2-SternoCOIR2 | HGAYATRGCHTTYCCHCG | 226 | ATRAARTTRATWGCTCCTA | 432 | 189 | This Study |
| BF1-AgPestR2 | ACWGGWTGRACWGTNTAYCC | 326 | RACWGMTCAVAYAAATARDGG | 500 | 155 | This Study |
| HexCOIF4- AgPestR2 | HCCHGAYATRGCHTTYCC | 223 | RACWGMTCAVAYAAATARDGG | 500 | 260 | (Marquina, 2019) |
| HexCOIF4- AgPestR1a | HCCHGAYATRGCHTTYCC | 223 | GTRATRAARTTDAYWGMHCC | 434 | 194 | (Marquina, 2019) |
| BF1-HexCOIR4 | ACWGGWTGRACWGTNTAYCC | 326 | TATDGTRATDGCHCCNGC | 563 | 218 | (Marquina, 2019) |
| fwhF2-HexCOIR4 | GGDACWGGWTGAACWGTWTAYCCHCC | 323 | TATDGTRATDGCHCCNGC | 563 | 215 | (Marquina, 2019) |
| SauronS878-HexCOIR4 | GGDRCWGGWTGAACWGTWTAYCCNCC | 323 | TATDGTRATDGCHCCNGC | 563 | 215 | Marquina, 2019) |

Table 1: Published and novel primers evaluated in this study

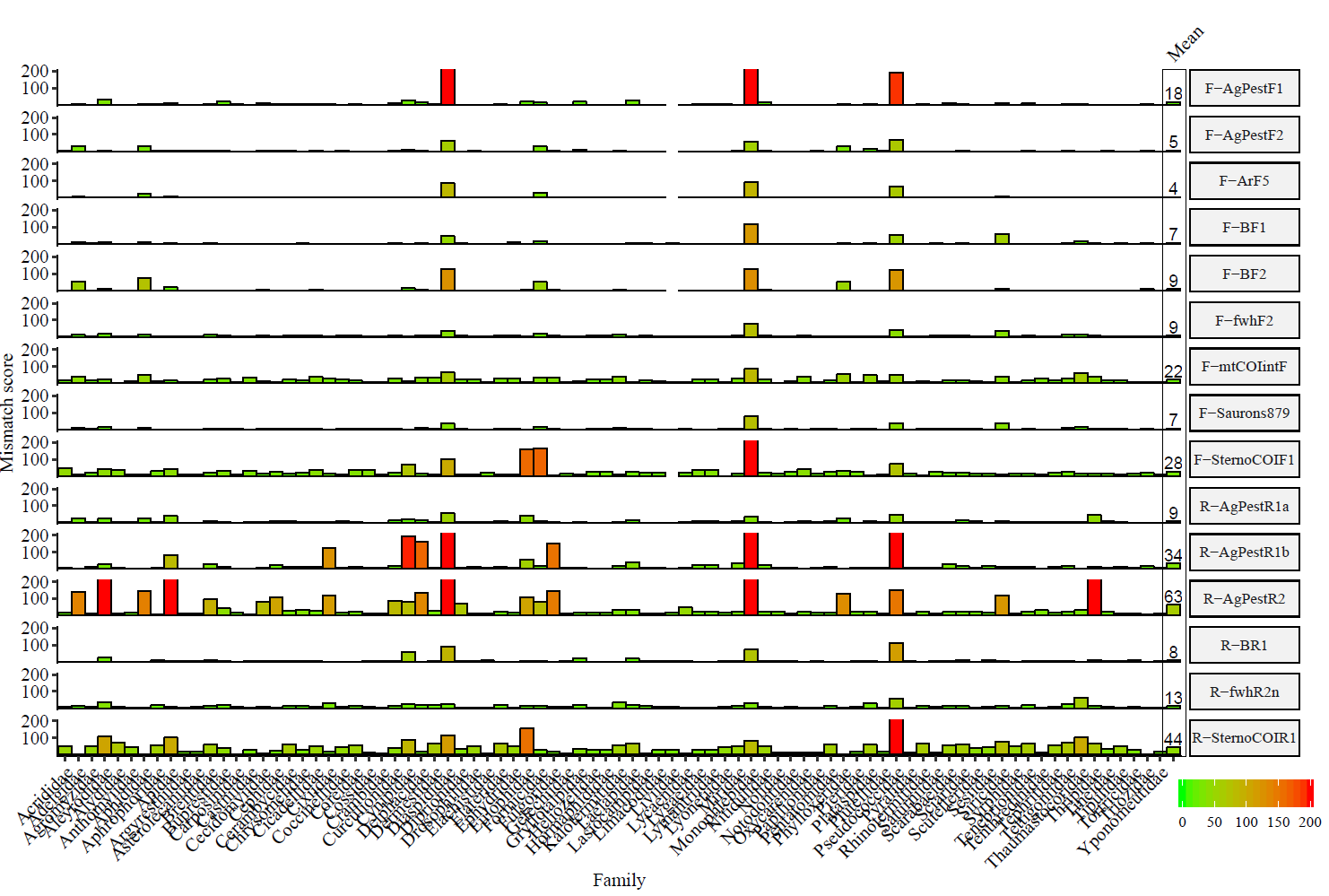
Figures:





*Figure 2 - Evaluation of novel and existing primers for resolution across pest insect families*

Figure3 - Evaluation of individual Primers for mismatch



A screenshot of a cell phone

Description generated with high confidence

A close up of a map

Description generated with very high confidence

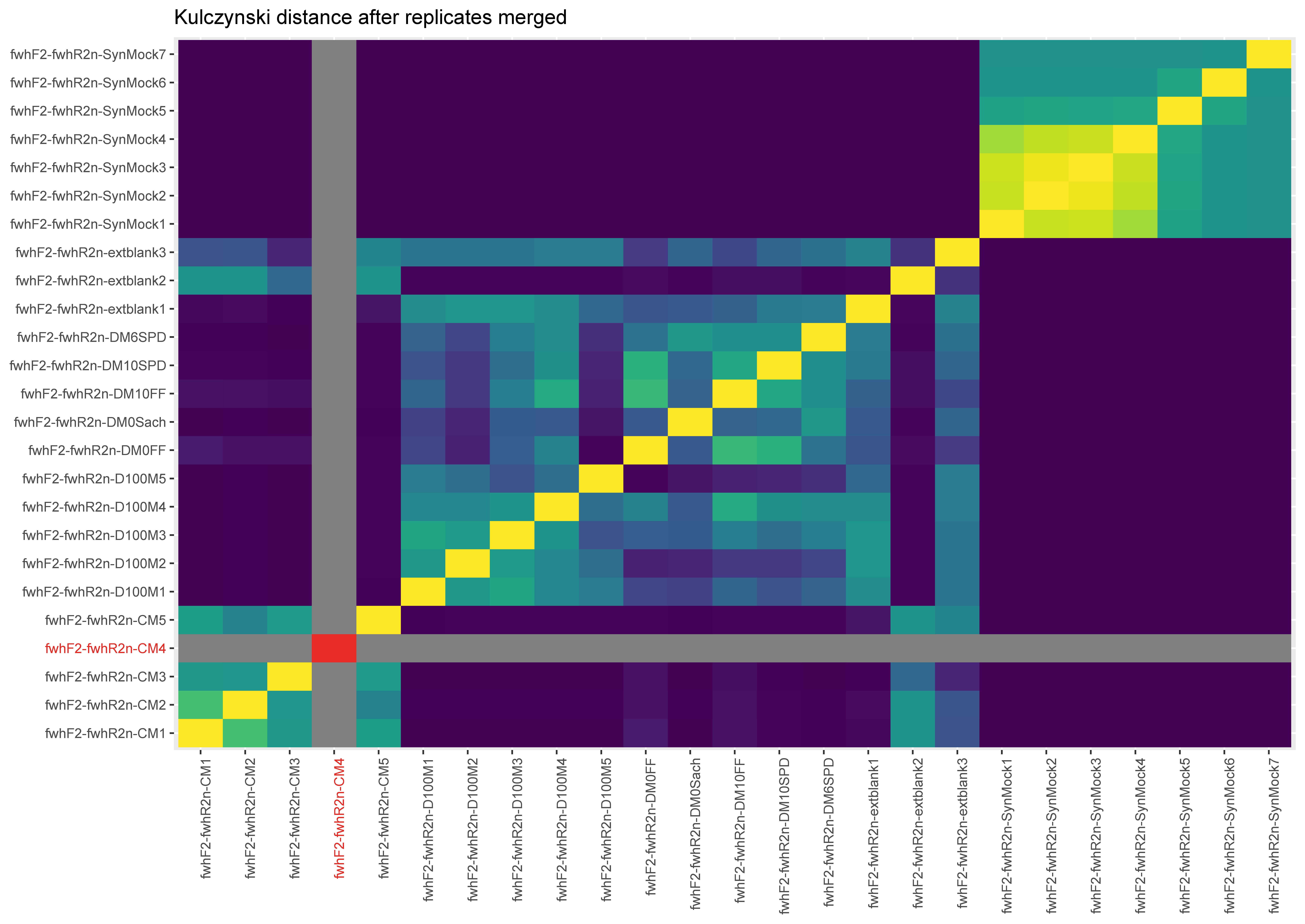
A close up of a map

Description generated with high confidence Figure 5- Metabarcoding analysis A close up of a map

Description generated with very high confidence

A screenshot of a cell phone

Description generated with high confidence



A picture containing writing implement, stationary

Description generated with very high confidence



Supplementary Figures

A close up of a device

Description generated with high confidence

Supplementary figure 1- Overlap

A screenshot of a cell phone

Description generated with high confidence