**Metabarcoding analysis of insect bulk traps for pest diagnostics: highlighting strengths and weaknesses.**

**Disentangling Metabarcoding bias.**

Francesco Martoni1,2, Alex Piper1, Brendan Rodoni1, Mark Blacket1

*1Agriculture Victoria Research, AgriBio Centre for AgriBio Science, Bundoora, Victoria, Australia*

2Corresponding author: Francesco Martoni – email: francesco.martoni@agriculture.vic.gov.au

**Abstract**

Agriculture, plant protection and invasive species management all rely heavily on diagnostics for a quick and precise identification of insect pests and vectors of plant pathogens in order to intercept them, eradicate them, or model their movements to predict their future incursions. A correct species identification, however, can often be challenging depending on the state of preservation of the insect in the trap, the life stage of the insect, with immatures often harder to recognize, or when putative pests are trapped together with a high number of often harmless non-pest species. In these instances, in fact, traditional morphological assessment of presence/absence of pest species requires a time- and cost-expensive employment of strong taxonomical expertise to manually examine the morphology of each specimen trapped.

Metabarcoding has recently emerged as a method to obtain a rapid and high-throughput taxonomical identification of many taxa simultaneously. However, this technique is known to provide a semi-quantitative assessment, which is sufficient to confirm presence/absence of species but not to estimate the number of individuals present.

Here, not only we compared how different DNA extractions and PCR primers might impact the record of possible pest species in a metabarcoding analysis, but we also investigated steps carrying higher risk to introduce bias in the number of reads produced. The results obtained suggest that DNA extraction, more than PCR primers selection, is responsible for the semi-quantitative nature of metabarcoding.

1. **Introduction**

In order to be sustainable, human population growth requires an equal growth in food production. Consequently, this adds pressure to agricultural systems, that are required to meet future food production with limited resources, such as water (Wallace 2000), and to face new threats. Biological invasions, driven by globalization and climate changes are becoming one of the main menaces to agriculture (Meyerson and Mooney 2007; Hulme 2009; Chown *et al.* 2015) with insect invasive species being a major threat to agroecosystems (Paini *et al.* 2016). This has led to pest interception and quarantine procedures to be implemented at the entry pathways to many countries (Martin *et al.* 2016; Early *et al.* 2016). However, when such procedures fail, the establishment of an alien insect species into a new environment, its containment, monitoring or eradication attempts can be a substantial cost to industry and governments as well as to private landowners (Bradshaw *et al.* 2016). One of the main surveillance programme costs is often the extensive trapping and monitoring activities carried out both in agricultural properties (Low-Choy 2015) and in protected environments, such as National Parks (e.g. Davidovitch *et al.* 2009).

Non-specific surveillance traps, such as air-sampling suction traps or liquid traps, tend to capture high numbers of non-target species, often endemics, that raise the number of specimens to be morphologically examined and identified by highly trained entomologists and taxonomists (Piper *et al.* 2019). In order to cope with the lack of taxonomical expertise, often limited to a few experts worldwide per taxonomic group, a range of molecular techniques has been developed in recent years to allow more standardized identification of high numbers of insect species (Piper *et al.* 2019). For example, the development of the DNA barcoding technique (Hebert *et al.* 2003) allowed comparison of a short standardized genetic region from an unidentified specimen to a vast number of known species deposited in reference databases. DNA barcoding has been successfully applied to identification of invasive insect pests (refs), however difficulties remain scaling this approach to the sheer numbers of specimens that can be caught in surveillance traps. Therefore with the advent of high throughput sequencing technologies, the focus is now shifting from the single specimen sequencing of DNA barcoding to identifying entire communities of species in parallel using DNA metabarcoding (Piper *et al.* 2019 and references therein).

Moving from analysis of single specimens to mixed communities allows a dramatic scale up in the number of both individual specimens and surveillance traps that can be processed (ref) and provide valuables information on non-pest and native species within the ecosystem. However, with DNA metabarcoding individual specimens are no longer the base unit of measurement, and instead the datasets returned are relative counts molecules. This brings new challenges for quantifying the number of individual specimens in a trap from metabarcoding sequences alone as the relationship between counts of molecules and the individuals they arise from is affected by a number of biases that systematically distort the measured sequence counts of each species from their true abundances.

Elucidating the various causes of biases in metabarcoding studies has seen significant attention in recent literature, and the primary contributor is thought to be mismatches between PCR primers and template molecules, particularly at the 3’ end of the primer where extension takes place (refs). Generally, metabarcoding studies conducted on invertebrate samples have mostly used the subunit 1 of the cytochrome oxidase gene (COI) as their target locus (e.g. Andújar et al., 2018; Elbrecht et al., 2017; Yu et al., 2012). This is mostly due to the availability of reference sequences, with COI being the most sequenced locus widely available for the highest number of insect species (Piper *et al.* 2019). Primer-template mismatch is particularly problematic for protein coding genes such as COI, due to natural degeneracy in the genetic code leaving no strictly conserved nucleotide sites for design of universal PCR primers (refs). This has led to the design and evaluation of highly degenerate primers in order to mitigate the effects of mismatch on detection efficiency and abundance estimates (ref). However, in addition to PCR bias, a range of physical characteristics of the community under study also play a role, with perhaps the most obvious of these is the large variation in specimen body sizes within insect community assemblages (refs). DNA extraction from complex communities for metabarcoding analysis has often involved destructive homogenisation of tissues, and this means large‐sized organisms will contribute a larger quantity of DNA molecules to the pool than small ones (refs). Nevertheless, when individual specimen size is taken into account the influence of PCR mismatch generally outweighs DNA extraction bias for macroinvertebrates (refs), at least for destructive homogenisation-based DNA extraction.

Recently, non-destructive DNA extraction has emerged as an alternative that allows voucher specimens to be retained following sequencing for morphological confirmation (refs). This is of particular importance in the context of biosecurity and diagnostic metabarcoding, allowing DNA sequences to be linked to an insect sample that can be used for morphological comparisons and can be preserved in entomological collections for future records. While It is well established for homogenised samples that large sized organisms will contribute more than small this may differ for non-destructive metabarcoding, where contribution of DNA may instead depend on surface/volume ratio (ref). Furthermore, additional aspects related to permeability of exoskeletons could alter detection efficiencies such as level of sclerotization of exoskeletons (ref). Moreover, with this move from destructive to non-destructive DNA extraction it is unclear if previous results and assumptions about the bias generating process still hold.

Here we applied two different non-destructive metabarcoding protocols from the recent literature, as well as two combinations of degenerate PCR primers on mock communities composed by a mix of possible insect pests and harmless by-catch species. We aimed to measure and partition the taxon specific bias introduced by the DNA extraction, PCR, and library preparation stages, and evaluate the downstream effects on the two main diagnostic-related aspects: sensitivity and quantitation

1. **Materials and Methods**
   1. **Samples**

For this study we used specimens belonging to insect colonies reared at the AgriBio laboratory of Agriculture Victoria together with a single field collected insect species, a psyllid identified only at the superfamily level (Table 1). Insect specimens from the colonies had been preserved in absolute ethanol deposited at the Victorian Agricultural Insect Collection (VAIC). A total of 16 taxa spanning four orders were grouped into eight pools, labelled I1-I8 (Table 1).

**Table 1:** Composition of the eight pools used for this study.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Pool | | | | | | | |
| Species | **Order** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| *Drosophila simulans* | Diptera | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *D. melanogaster* | Diptera | 5 | 10 | 25 | 50 | 5 | 10 | 25 | 50 |
| *D. hydeii* | Diptera | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Bactrocera tryoni* | Diptera | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Scaptodrosophila lativittata* | Diptera | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Bradysia ocellaris* | Diptera | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| *Aphidius colemani* | Hymenoptera | 36 | 5 | 6 | 12 | 36 | 4 | 9 | 20 |
| *Lysiphlebus testaceipes* | Hymenoptera | 14 | 0 | 4 | 13 | 14 | 1 | 1 | 5 |
| *Carpophilus nr dimidiatis* | Coleoptera | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *C. davidsoni* | Coleoptera | 25 | 50 | 5 | 10 | 25 | 50 | 5 | 10 |
| *Diuraphis noxia* | Hemiptera | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Metopolophium dirhodum* | Hemiptera | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Rhopalosiphum padi* | Hemiptera | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Psyllid sp. | Hemiptera | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Acizzia alternata* | Hemiptera | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *A. solanicola* | Hemiptera | 10 | 25 | 50 | 5 | 10 | 25 | 50 | 5 |
| Total | **4** | **100** | **100** | **100** | **100** | **101** | **100** | **101** | **100** |

* 1. **Molecular analysis**
     1. **DNA extraction, amplification and library preparation from insect pools.**

DNA was non-destructively extracted from the eight pools of insects using the QuickExtract kit (Biosearch Technologies, California, USA) for pools 1-4 and the Blood and Tissue kit (Qiagen, Germany) for pools 5-8 (pools I1-I8).

Non-destructive DNA extraction using QuickExtract was performed as follows: ethanol was removed from the pooled insects using a 200ul pipette, which were air-dried in tubes for 10 minutes. 500ul of QuickExtract was added to the pools ensuring all insects were submerged. Pools were then vortexed for 30 seconds, incubated at 65°C for 6 minutes, vortexed for 15 seconds and incubated at 98°C for 2 minutes. All the liquid present in the vial was then pipetted to a new Eppendorf tube and stored at -20°C. Following extraction, pooled insect specimens were preserved in absolute ethanol as vouchers.

Non-destructive DNA extraction using Qiagen Blood and Tissue kit was initially performed following the first steps of the protocol presented in Martoni *et al.* (2019) and Bahder *et al.* (2015). Briefly, ethanol was removed from the insect pools (as above), insects were then submerged in an ATL buffer / Proteinase K mix with a ratio of 9/1 and then left for approximately 17 hours (overnight) at 56°C. The overnight buffer was removed from the insects (as above), and pooled insect specimens were preserved in absolute ethanol.

PCRs were performed using the Bioline MyFi DNA Polymerase kit (Meridian Bioscience, Ohio, United States of America) using 2.5 uL of DNA template. PCRs were conducted on the 20 samples and 3 controls using two separate degenerate primer pairs overlapping a similar region of COI. Primer pairs used were fwhF2 (GGDACWGGWTGAACWGTWTAYCCHCC)- fwhR2n (GTRATWGCHCCDGCTARWACWGG) and fwhF2 – HexCOIR4(TATDGTRATDGCHCCNGC) (Vamos *et al.* 2017; Marquina *et al.* 2018). The PCR was run with the same cycling conditions for both primer pairs, with an initial 5 minute denaturation at 95°C, followed by 30 cycles, of denaturation at 95°C for 45 seconds, annealing at 50°C for 30 seconds and extension at 72°C for 30 seconds, followed by a final extension at 72°C for 7 minutes. Successful PCR amplification was verified on a 1% agarose gel.

PCR products were used directly as template for a real time qPCR to attach illumina sequencing adapters containing unique dual indexes to each sample (Figure SM1). Phusion High-Fidelity DNA Polymerase Taq (New England Biolabs, Massachusetts, USA) was used to attach adapters with a qPCR profile of 30s at 98°C followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 65°C for 30s and elongation at 72°C for 30s. The qPCR was interrupted after 8 cycles, immediately after the exponential increase phase ceased. Then, qPCR product purification was performed using AMPure XP kit (Beckman Coulter, California, USA) following the manufacturer instructions. Library fragment size (amplicon + indexes) and presence of primer dimers was verified on an Agilent TapeStation (Agilent Technologies, California, USA). Additional cleanings of the amplicons were performed in order to remove any primer dimer present. Finally, the samples were equimolarly pooled based on their concentrations, measured via Qubit Fluorometric Quantification, using a Qubit dsDNA HS Assay kit (ThermoFisher Scientific, Massachusetts, USA). The final pooled libraries were then diluted 7 pM with a 15% PhiX spike-in and sequenced on the Illumina MiSeq platform (2 x 250 bp reads) (Illumina, California, USA).

* + 1. **DNA extraction, amplification and library preparation from insect species.**

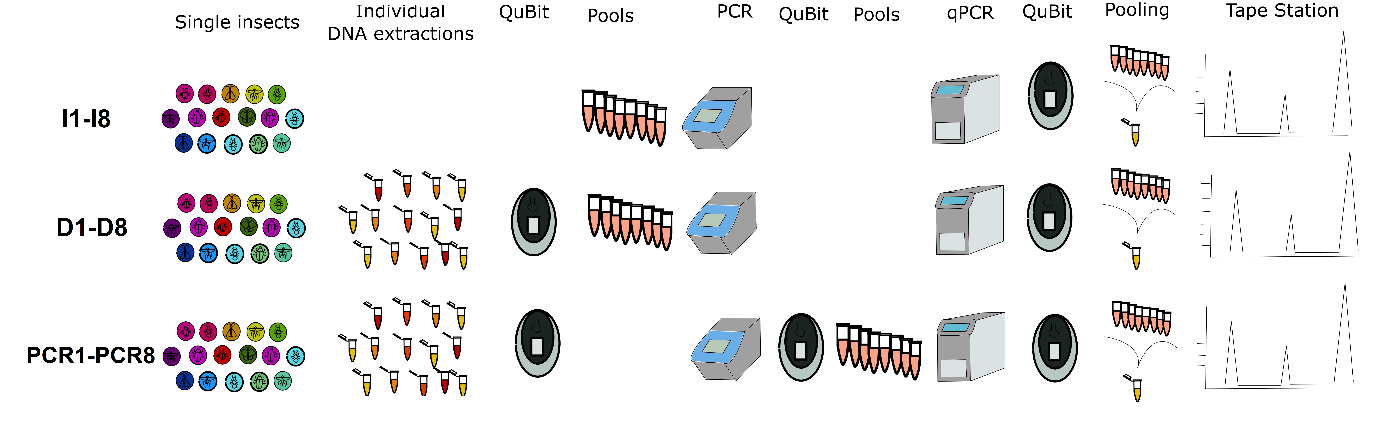
DNA was also extracted from each one of the 16 insect species used by pooling together 5-20 individuals (depending on their dimensions) and using the Blood and Tissue kit (Qiagen, Germany) destructively, by homogenising the insects using a micro-pestle. This was performed destructively and in bulks because the aim was to obtain good quality DNA from each species.

The DNA from each insect species was then quantified using a QuBit (XXXXXX) and pooled together imitating the composition of the original insect pools. An arbitrary amount of 4ng of DNA was used as the unit corresponding to 1 insect (i.e., 4ng of DNA for each insect composing the original pools). The DNA pools obtained were named D1-D8.

In the same PCR, using the same protocol described above, were amplified the original 8 insect pools (from now on, I1-I8), the DNA pools (D1-D8) and the DNA of each of the 16 species (S1-S16). Each set of samples was amplified using both primer sets.

The PCR products obtained from the samples S1-S16 was then quantified using QuBit (XXXXXX) and pooled imitating the composition of the original insect pools by using 2ng of DNA as a unit corresponding to 1 insect (i.e., 2ng of DNA for each insect composing the original pools). The DNA pools obtained were named PCR1-PCR8.

A second library was prepared following the same protocol of the first one. This library included a total of 48 samples: the original insect pools (I1-I8), the DNA pools (D1-D8) and the PCR pools (PCR1-PCR8); all samples for each primer set. Additionally, two blank extractions controls were included. The samples were pooled and run on the same MiSeq machine previously used, using the same protocol.



**Figure 1:** Workflow of the experiment for the three types of pools. Insect pools (I1-I8), DNA pools (D1-D8) and PCR pools (PCR1-PCR8). Each of these were amplified using two primer sets.

* + 1. **Bioinformatics analysis**

Bioinformatic analysis was conducted using R v3.5.1 (The R Core Team, 2019) and a modified pipeline of Callahan *et al.* (2016). In brief, raw sequence reads were demultiplexed using bcl2fastq with 0 mismatches. Demultiplexed sequencing reads were then trimmed of PCR primers and using BBDuK in BBTools v38, and 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 uses a quality aware error model to resolve distinct amplicon sequence variants (ASVs) from erroneous background noise.

Taxonomy was assigned to the n ASVs to the lowest rank possible with a minimum confidence of 60 using the IDTAXA algorithm (Murali *et al.* 2018), followed by further species level assignment using exact matching between the query and validated in-house reference sequences. Further analysis and plotting of the final ASV table were conducted using Phyloseq (McMurdie & Holmes 2013) and tidyverse packages (Wickham, 2017).

1. **Results**

Metabarcoding on insect pools was performed using a non-destructive DNA extraction method in order to retain a morphological voucher that is often required in biosecurity. Non-destructive DNA extractions were performed using two standard kits, QuickExtract (samples I1-I4) and DNeasy Blood and tissue kit (samples I5-I8), requiring approximatively 1 hour and 1 day of the operators’ time, respectively.

While presence/absence of the insects was generally very precise, when comparing the number of reads obtained for each insect species to the number of reads expected based on the percentage of such species in the pool, a clear bias in the number of reads was recorded. Discrepancies appeared across all species between the observed relative abundances and the expected relative abundances (based on the individual insects included in each pool).

When comparing samples I1-I8 with D1-D8, the bias introduced at the DNA extraction step could be examined. The variation in reads is..

When comparing samples D1-D8 with PCR1-PCR8, the bias introduced at the PCR stage could be examined. The variation in reads is..

Furthermore, when quantifying the DNA concentration of the PCR products obtained from each single species, different values could be recorded for each species and for different number of cycles used (Table 3a and 3b).

**Table 2: DNA concentration of PCR products.** The graphs show the different DNA concentrations obtained for each insect species from the same template concentration (4ng/uL) using two primer pairs and two different number of PCR cycles: 30 (**3a**) and 40 (**3b**).

**Discussion**

In this work, metabarcoding analysis of insect pools could successfully identify insect species from pools of 100-101 individuals, also for those species that were present with single individuals. Furthermore, the use of non-destructive DNA extraction methods in this workflow allows to retain voucher specimens of the insects. This is of paramount importance in the context of biosecurity and diagnostics, allowing to link DNA sequences to an insect sample that can be used for morphological comparisons and can be preserved in entomological collection for future records. Hence, the results presented here confirm that metabarcoding analysis can be successfully applied to insect pools to obtain a species identification, highlighting possible applications in the fields of diagnostics and biosecurity to determine presence/absence of pests. Here we also highlighted a number of factors that should provide further information when dealing with metabarcoding of insect traps, especially where a quantitative output could be a desired result.

The relative abundance of reads obtained for each insect species was not proportional to the number of individuals, but it is biased toward certain species. This bias in the number of output reads highlights the semi-quantitative nature of metabarcoding. In order to understand if such a bias could be eliminated, we separately analysed the number of reads obtained at each step of the workflow.

Firstly, to test bias introduced by DNA extraction, DNA non-destructively extracted from insect pools was compared with DNA destructively obtained from each species singularly. The most importance difference between these two methodologies in terms of DNA extracted, is that non-destructive techniques act on the external surface of the insect exoskeletons, as opposed to destructive methods that can access the full DNA contained in the insect bodies. Furthermore, the DNA pools composed here allowed to compare the effect of metabarcoding analysis on the ideal supposition that each insect species could generate the same amount of DNA per unit, irrespectively of the insect size, body consistency and biomass. While we are well aware of variations in body size and structure for insects, we aimed to test if starting from the same concentrations, all DNAs would react the same to PCR amplification and sequencing.

The results showed that bias in the final number of output reads appeared to be strongly influenced by the DNA extraction step when using non-destructive protocols. This issue appears to be unavoidable when voucher specimens are required, and destructive DNA extraction is not an option. Possible reasons for the different amounts of DNA extracted can be due to the size of the different insects, their biomass, the surface coming in contact with the extraction buffer and the body consistency, ranging from soft to hard. Part of these issues, such as the biomass, however, should be expected to impact metabarcoding analyses also when destructive DNA extractions are performed. Indeed, larger specimens would add higher volumes of DNA to the pool, resulting in a higher percentage of DNA from a single large beetle than from 10 psyllids or aphids. Furthermore, the unknown composition of insect traps makes it challenging to include mock communities spike-ins before DNA extractions, despite these having been proved useful for quantification elsewhere (ref). Indeed, for such mock communities to provide a useful comparative standard, the content of the insect pools should be known (ref), which cannot be the case in diagnostic analyses.

When comparing the results obtained using two different non-destructive DNA extraction protocols, QuickExtract and Blood and Tissue kit, the results did not show much variations. However, repeating the same analysis using the same DNA extracts after two months from extraction – during which the templates were kept in -20C freezer – the results changed dramatically. The QuickExtract template appeared to have degraded a lot, resulting in a number of species dropping out and not being recorded. On the other hand, the results obtained from the Blood and Tissue kit were identical to the first analysis in terms of presence/absence records. This suggest DNA extraction products obtained with QuickExtract kit should be used immediately for analysis, or possibly kept at lower temperatures, such as at -80C.

Beside DNA extraction, we wanted to assess if bias was introduced also at the PCR stage. The samples obtained from pooled DNA were amplified together with the DNA of each insect species separately and this was then pooled together in pools with identical composition. This highlighted that read bias was introduced at this stage, as well.

Primer bias was confirmed also when quantifying separately the PCR product from each different insect species. Since this DNA had been normalized before PCR, variation in subsequent concentration is the result of primers bias. Testing two primer pairs that are considered generic in the literature (Vamos et al. 2017; Marquina et al. 2018), we aimed to test if primer bias could be observed in pools of possible agricultural pests and if this would lead to false negative results, leading to some species not being recorded. However, both primer pairs tested here allowed us to record all the species present in all the pools, with a sensitivity of up to 1 in 101 for many of the species tested. This suggests that the primers used here are indeed generic enough to record the different species included in the current study, which represented a broad array of insect taxa. Nonetheless, the fact these primers are generic enough to amplify hundreds of species (ref) shouldn't be misinterpreted as a proof they will amplify each species in the same measure. Here, the PCRs performed on each species singularly confirmed both primer pairs showed to amplify the DNA of some species up to 10x more than that of others. This suggests further work is required to test additional primer pairs that might produce a lower primer bias. However, as mentioned above, in a diagnostic context, priority is given to a precise presence/absence assessment, especially when testing for the presence of unexpected pests. In this instance, the broad target semi-quantitative output of the MiSeq amplicon sequencing confirmed here is to be prioritised over a more quantitative approach, perhaps more time-consuming such as qPCR, targeting each single species.

Finally, it is safe to assume that the bias recorded by comparing the final number of reads with the expected pool composition, is mostly due to the qPCR step performed during library preparation, to attach dual unique indexes to the target sequences. Hence, these final 6-8 cycles on a qPCR appear to add a non-significant variation to the expected reads amounts. This suggest this last step is the one introducing the smallest amount of bias.

**References**

* Bahder BW, Bollinger L, Sudarshana MR& Zalom FG. 2015. Preparation of mealybugs (Hemiptera: Pseudococcidae) for genetic characterization and morphological examination. Journal of Insect Science 15, 104.
* Batovska et al. In Prep. Developing a non-destructive metabarcoding protocol for detection of target insects in bulk trap catches.
* Borrell, Y.J., Miralles, L., Huu, H.D., Mohammed-Geba, K., and Garcia-Vazquez, E. (2017). DNA in a bottle—Rapid metabarcoding survey for early alerts of invasive species in ports. PLOS ONE 12, e0183347.
* Chown SL, Hodgins KA, Griffin PC, et al. Biological invasions, climate change and genomics. Evol Appl 2015;8:23–46.
* Davidovitch L, Stoklosa R,Majer J, et al. Info-gap theory and robust design of surveillance for invasive species: The case study of Barrow Island. J Environ Manage 2009;90:2785–93.
* Early R, Bradley BA, Dukes JS, et al. Global threats from invasive alien species in the twenty-first century and national response capacities. Nat Commun 2016;7:12485.
* Gloor GB, Macklaim JM, Pawlowsky-Glahn V and Egozcue JJ (2017) Microbiome datasets are compositional: and this is not optional. Frontiers in microbiology, 8:2224.
* Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Procedings of the Roayal Society B -Biological Sciences 270: 313–321.
* Hulme PE. Trade, transport and trouble: Managing invasive species pathways in an era of globalization. J Appl Ecol 2009;46:10–8.
* Klymus, K.E., Marshall, N.T., and Stepien, C.A. (2017). Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. PLOS ONE 12, e0177643.
* Low-Choy S. Getting the story straight: Laying the foundations for statistical evaluation of the performance of surveillance. In: Jarrad F, Low-Choy S, Mengersen K , eds. Biosecurity Surveillance: Quantitative Approaches. 6th ed.CABI; 2015:43–73.
* Martin RR, Constable F, Tzanetakis IE. Quarantine regulations and the impact of modern detection methods. Annu Rev Phytopathol 2016;54:189–205.
* Martoni F, Valenzuela I and Blacket M. (2019) Non-destructive DNA extractions from fly larvae (Diptera: Muscidae) enable molecular identification of species and enhance morphological features. Austral Entomology XXX.
* Marquina D, Andersson AF, and Ronquist F. New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Molecular Ecology Resources 19(12). DOI: 10.1111/1755-0998.12942
* McLaren MR, Willis AD, Callahan BJ (2018) Consistent and correctable bias in metagenomic sequencing experiments. eLife 2019;8:e46923. DOI: https://doi.org/10.7554/eLife.46923
* McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8, e61217.
* Meyerson LA, Mooney HA. Invasive alien species in an era of globalization. Front Ecol Environ 2007;5:199–208.
* Muralil A, Bhargava A, and Wright ES. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. Microbiome volume 6, Article number: 140 (2018)
* Paini DR, Sheppard AW, Cook DC, et al. Global threat to agriculture from invasive species. Proc Natl Acad Sci U S A 2016;113:7575–9.
* Piper A, Batovska J, Cogan NOI, Weiss J, Cunningham JP, Rodoni BC, and Blacket MJ. Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. GigaScience, 8, 2019, 1–22. doi: 10.1093/gigascience/giz092
* Pochon, X., Zaiko, A., Fletcher, L.M., Laroche, O., and Wood, S.A. (2017). Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. PLOS ONE 12, e0187636.
* Vamos E, Elbrecht V, and Leese F. Short COI markers for freshwater macroinvertebrate metabarcoding. Metabarcoding and Metagenomics 1: e14625. DOI: 10.3897/mbmg.1.14625
* Wallace JS (2000). Increasing agricultural water use efficiency to meet future food production. Agriculture, Ecosystems & Environment, 82, (1–3), 105-119. <https://doi.org/10.1016/S0167-8809(00)00220-6>.
* Wickham H. (2017) R for data science : import, tidy, transform, visualize, and model data. Grolemund, Garrett (First ed.). Sebastopol, CA. ISBN 9781491910399. OCLC 968213225.