**Metabarcoding mosquitoes: MinION sequencing of bulk samples gives accurate species profiles for vector surveillance (Culicidae)**

**Rebecca Ker LOH\*#1, Tyrone Ren Hao TAN#1, Huiqing YEO1, Tze Xuan YEOH1, Theodore Tze Ming LEE2, Sujatha Narayanan KUTTY3, Nalini PUNIAMOORTHY1\***

1Department of Biological Sciences, National University of Singapore, 16 Science Drive 4, Singapore 117558, Singapore

2Military Medicine Institute, Singapore Armed Forces, 701 Transit Rd #01-03, Singapore 778910 Singapore

3Tropical Marine Science Institute, National University of Singapore, 18 Kent Ridge Road, Singapore 119227, Singapore

\*Corresponding Authors

Email: rebeccalohker@gmail.com and nalini@nus.edu.sg

# These authors share first authorship

**AFFILIATIONS:**

1Department of Biological Sciences, National University of Singapore, 16 Science Drive 4, Singapore 117558, Singapore

2MMI, Singapore Armed Forces, 701 Transit Rd #01-03, Singapore 778910 Singapore

3Tropical Marine Science Institute, National University of Singapore, 18 Kent Ridge Road, Singapore 119227, Singapore

**CONTENTS:**

Pages S1-S13

Table S1-S5

Text S1-S5

Figure S1

**Tables**

**Table S1:** Sampling details, including sample codes and specimen counts. One of our 10 traps was destroyed due to natural causes and was excluded from further analysis (marked “NA”).

**Table S2:** List of 13-base pair primer indices used for PCR triplicates. Negative controls of DNA extractions are labelled as NC while negative controls of PCR are labelled as PCRN. Selected samples which were exceptionally specimen-rich were kept as three distinct DNA extracts/subsamples (Subsample number: R1-R3) and processed separately to prevent the potential issue of highly abundant species overwhelming that signals of rare species (these are later combined during data analysis because they are pseudo-replicates).

**Table S3:** Full list of statistical models compared using the Information Theory approach and AICc, to find variables the source of CO2 trap lure (“trap”), storage method (“storage”), and body segments (“part) influencing the number of species (“number”) detected on sequencing platforms: Illumina, MinION.

**Table S4:** List of models that fulfil ΔAICc<2 criterion.

**Table S5:** Table of final Poisson models after stepwise simplification \* indicating that p-value <0.05.

**Texts**

**Text S1:** Sampling location and permission.

**Text S2:** Detailed PCR protocol.

**Text S3:** Additional modifications made to MinION library prep protocol.

**Text S4:** Brief overview of proof-of-concept experiment, demonstrating the effectiveness of our DNA barcode reference library curated by expert taxonomists.

**Text S5:** Statistical modelling and data analysis conducted on R software.

**Figures**

**Figure S1:** Heatmap of mosquito species detected on both sequencing platforms.

**Table S1:** Sampling details, including sample codes and specimen counts. One of our 10 traps was destroyed due to natural causes and was excluded from further analysis (marked “NA”).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sampling date (sample number) | Source of for CO2 trap lure (B=biogenic, G=gas cylinder) | Number of mosquito specimens collected | Storage method(E=ethanol, D=dry) | Body segments(B=bodies, H=heads) | Sample code | Number of mosquitoes in each bulk sample |
| 16 August 2021 (S1) | B | 34 | E | B | S1BEB | 17 |
| H | S1BEH | 17 |
| D | B | S1BDB | 17 |
| H | S1BDH | 17 |
| G | 92 | E | B | S1GEB | 46 |
| H | S1GEH | 46 |
| D | B | S1GDB | 46 |
| H | S1GDH | 46 |
|  |  |  |  |  |  |  |
| 17 August 2021(S2) | B | 38 | E | B | S2BEB | 19 |
| H | S2BEH | 19 |
| D | B | S2BDB | 19 |
| H | S2BDH | 19 |
| G | 36 | E | B | S2GEB | 18 |
| H | S2GEH | 18 |
| D | B | S2GDB | 18 |
| H | S2GDH | 18 |
|  |  |  |  |  |  |  |
| 18 August 2021(S3) | B | 26 | E | B | S3BEB | 13 |
| H | S3BEH | 13 |
| D | B | S3BDB | 13 |
| H | S3BDH | 13 |
| G | 34 | E | B | S3GEB | 17 |
| H | S3GEH | 17 |
| D | B | S3GDB | 17 |
| H | S3GDH | 17 |
|  |  |  |  |  |  |  |
| 19 August 2021(S4) | B | NA | NA | NA | NA | NA |
| NA | NA | NA |
| G | 92 | E | B | S4GEB | 46 |
| H | S4GEH | 46 |
| D | B | S4GDB | 46 |
| H | S4GDH | 46 |
|  |  |  |  |  |  |  |
| 20 August 2021(S5) | B | 38 | E | B | S5BEB | 19 |
| H | S5BEH | 19 |
| D | B | S5BDB | 19 |
| H | S5BDH | 19 |
| G | 356 | E | B | S5GEB | 178 |
| H | S5GEH | 178 |
| D | B | S5GDB | 178 |
| H | S5GDH | 178 |

**Table S2:** List of 13-base pair primer indices used for PCR triplicates. Negative controls of DNA extractions are labelled as NC while negative controls of PCR are labelled as PCRN. Selected samples which were exceptionally specimen-rich were kept as three distinct DNA extracts/subsamples (Subsample number: R1-R3) and processed separately to prevent the potential issue of highly abundant species overwhelming that signals of rare species (these are later combined during data analysis because they are pseudo-replicates).

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample codes** | **Subsample number** | **PCR triplicates sample codes** | **Primer indices** |
| S5BDB | NA | S5BDB\_1 | ACTCCCAACCAAC |
| S5BDB | NA | S5BDB\_2 | GAGCCGTGGTGTT |
| S5BDB | NA | S5BDB\_3 | GGAGATAATCGAA |
| S5BDH | NA | S5BDH\_1 | GGCCTATTAAGTT |
| S5BDH | NA | S5BDH\_2 | GTTGCTGCCAGAA |
| S5BDH | NA | S5BDH\_3 | CTGGCCAAGACGC |
| S5BEB | NA | S5BEB\_1 | CTTAAGAAGGTGC |
| S5BEB | NA | S5BEB\_2 | ACACAACTGAGGT |
| S5BEB | NA | S5BEB\_3 | GTTCGGTCCACCG |
| S5BEH | NA | S5BEH\_1 | TGCTAATACCAAT |
| S5BEH | NA | S5BEH\_2 | CCGCACCATGACC |
| S5BEH | NA | S5BEH\_3 | GAGACACGGACTT |
| S1BDB | NA | S1BDB\_1 | TTAAGCACCTGAA |
| S1BDB | NA | S1BDB\_2 | CGTGTTACACCTT |
| S1BDB | NA | S1BDB\_3 | GAGATGGTAGTTA |
| S1BDH | NA | S1BDH\_1 | AGCAATCGCGCAC |
| S1BDH | NA | S1BDH\_2 | CGCACCGAATAAT |
| S1BDH | NA | S1BDH\_3 | CCTGAATCATCTA |
| S1BEB | NA | S1BEB\_1 | AGTAACTTGTTCC |
| S1BEB | NA | S1BEB\_2 | AACCATTCTCACC |
| S1BEB | NA | S1BEB\_3 | GGTGAAGGTGGTC |
| S1BEH | NA | S1BEH\_1 | TAATAGGCTTCTG |
| S1BEH | NA | S1BEH\_2 | AAGCGGAGCCAAT |
| S1BEH | NA | S1BEH\_3 | CTTCTATTGTTGT |
| S2BDB | NA | S2BDB\_1 | ACCGGATCTGCGA |
| S2BDB | NA | S2BDB\_2 | CTACAGAGTTGGT |
| S2BDB | NA | S2BDB\_3 | GCTGCGGACACAC |
| S2BDH | NA | S2BDH\_1 | ATAATATTCGGCA |
| S2BDH | NA | S2BDH\_2 | CTCCTCCAGTTGC |
| S2BDH | NA | S2BDH\_3 | CGCTACCGCACCA |
| S2BEB | NA | S2BEB\_1 | CAAGTCCACTATG |
| S2BEB | NA | S2BEB\_2 | AAGGCTGTATAAG |
| S2BEB | NA | S2BEB\_3 | GTATGTTGGATAA |
| S2BEH | NA | S2BEH\_1 | TCCAGAGAAGAGA |
| S2BEH | NA | S2BEH\_2 | CAGGGTATGAATA |
| S2BEH | NA | S2BEH\_3 | GACCACACCGGAT |
| S3BDB | NA | S3BDB\_1 | AATGGCCGGTTCA |
| S3BDB | NA | S3BDB\_2 | ACACGGCTGTCAT |
| S3BDB | NA | S3BDB\_3 | GCCTCTTCGAATA |
| S3BDH | NA | S3BDH\_1 | GCCCTATAACACA |
| S3BDH | NA | S3BDH\_2 | GGTGTATCCTCCA |
| S3BDH | NA | S3BDH\_3 | CTAAGACTAAGGA |
| S3BEB | NA | S3BEB\_1 | CCTACAAGAATAC |
| S3BEB | NA | S3BEB\_2 | AATGCAATATGTC |
| S3BEB | NA | S3BEB\_3 | GTCGTGAAGCTAG |
| S3BEH | NA | S3BEH\_1 | ATGTTAAATGGAG |
| S3BEH | NA | S3BEH\_2 | CTTTACTACAATG |
| S3BEH | NA | S3BEH\_3 | CGGATTGGTAGGC |
| S5GDB | R1 | S5GDBR1\_1 | ACTTAACGTTAGC |
| S5GDB | R1 | S5GDBR1\_2 | GCAGAATTGTCTT |
| S5GDB | R1 | S5GDBR1\_3 | GGATAGAGGTATG |
| S5GDH | R1 | S5GDHR1\_1 | CAGGAACAACGCC |
| S5GDH | R1 | S5GDHR1\_2 | GAGAAGCCGGCTG |
| S5GDH | R1 | S5GDHR1\_3 | CGTAAGAGGCGGA |
| S5GDB | R2 | S5GDBR2\_1 | AGCCACCGAAGCG |
| S5GDB | R2 | S5GDBR2\_2 | GTATGTTACGCTG |
| S5GDB | R2 | S5GDBR2\_3 | GGCACCACGCGTT |
| S5GDH | R2 | S5GDHR2\_1 | CCGAGATTGGCCG |
| S5GDH | R2 | S5GDHR2\_2 | GAGTGTGGTCAAG |
| S5GDH | R2 | S5GDHR2\_3 | CGTGCTTATCTTG |
| S5GDB | R3 | S5GDBR3\_1 | AGGCCGTCCTTGT |
| S5GDB | R3 | S5GDBR3\_2 | AACCAGCGCCTAG |
| S5GDB | R3 | S5GDBR3\_3 | GGCGTGCTCTTCA |
| S5GDH | R3 | S5GDHR3\_1 | GTATTCGCCTGGT |
| S5GDH | R3 | S5GDHR3\_2 | GTTGGTTGTGTTC |
| S5GDH | R3 | S5GDHR3\_3 | CTTACAAGCAGTT |
| S5GEB | R1 | S5GEBR1\_1 | GAAGGCGGCACCT |
| S5GEB | R1 | S5GEBR1\_2 | ACACACCACCTCC |
| S5GEB | R1 | S5GEBR1\_3 | TAATGTTGCCAAC |
| S5GEH | R1 | S5GEHR1\_1 | TCTTCCTCCTCAA |
| S5GEH | R1 | S5GEHR1\_2 | CCAACTATCAGCT |
| S5GEH | R1 | S5GEHR1\_3 | GAGAACTTCGGAG |
| S5GEB | R2 | S5GEBR2\_1 | GACTTCTGTGGAC |
| S5GEB | R2 | S5GEBR2\_2 | ACGTACATCCACA |
| S5GEB | R2 | S5GEBR2\_3 | TACCGGCCTAATC |
| S5GEH | R2 | S5GEHR2\_1 | TTCATCGGTCTTA |
| S5GEH | R2 | S5GEHR2\_2 | TTGGATGCGCGCT |
| S5GEH | R2 | S5GEHR2\_3 | GCCATTGCCGTAA |
| S5GEB | R3 | S5GEBR3\_1 | GCAAGCCTGATTC |
| S5GEB | R3 | S5GEBR3\_2 | ACTCACTTGTAAT |
| S5GEB | R3 | S5GEBR3\_3 | TAGTAAGGAAGAT |
| S5GEH | R3 | S5GEHR3\_1 | TGGTCTCATGCGT |
| S5GEH | R3 | S5GEHR3\_2 | CCGGCGGAATGAG |
| S5GEH | R3 | S5GEHR3\_3 | GAGAGTGGATACC |
| S1GDB | NA | S1GDB\_1 | ACATTGTTATAGC |
| S1GDB | NA | S1GDB\_2 | CGCCTTACTGCCA |
| S1GDB | NA | S1GDB\_3 | GCGTTGCAGCTCC |
| S1GDH | NA | S1GDH\_1 | AGTGGACCAACAA |
| S1GDH | NA | S1GDH\_2 | CTATGCGTCAAAT |
| S1GDH | NA | S1GDH\_3 | CCTTAGCATTGTT |
| S1GEB | NA | S1GEB\_1 | CAAGACTCTACAC |
| S1GEB | NA | S1GEB\_2 | AACCTGGAAGCCT |
| S1GEB | NA | S1GEB\_3 | GTAGGTGTGGCTT |
| S1GEH | NA | S1GEH\_1 | TATCGCAAGAACA |
| S1GEH | NA | S1GEH\_2 | AGGCCACATCCAG |
| S1GEH | NA | S1GEH\_3 | GAAGGATAGCCTG |
| S2GDB | NA | S2GDB\_1 | AATAGAACTGCCT |
| S2GDB | NA | S2GDB\_2 | ACAACAACGACAA |
| S2GDB | NA | S2GDB\_3 | GCCTCCTGGTGAG |
| S2GDH | NA | S2GDH\_1 | ATGCTAGACATCG |
| S2GDH | NA | S2GDH\_2 | CTGTATCTCGGTT |
| S2GDH | NA | S2GDH\_3 | CGGAATTAGATAT |
| S2GEB | NA | S2GEB\_1 | CCGTTGGAGAACT |
| S2GEB | NA | S2GEB\_2 | AAGGTACCGGCGA |
| S2GEB | NA | S2GEB\_3 | GTCAACAATTCTG |
| S2GEH | NA | S2GEH\_1 | CCTCCAACCGCTG |
| S2GEH | NA | S2GEH\_2 | GGATACCTCCTCT |
| S2GEH | NA | S2GEH\_3 | CGTTAACCATAAT |
| S4GDB | NA | S4GDB\_1 | TTCATAACAGAAG |
| S4GDB | NA | S4GDB\_2 | TCCGTAATGCGCG |
| S4GDB | NA | S4GDB\_3 | GCAGCAGGTTATT |
| S4GDH | NA | S4GDH\_1 | GCTTTGCTTGGCG |
| S4GDH | NA | S4GDH\_2 | GTTCTCCGGCCTG |
| S4GDH | NA | S4GDH\_3 | CTCCTTATTATCC |
| S4GEB | NA | S4GEB\_1 | CTCCAATTCCTTC |
| S4GEB | NA | S4GEB\_2 | ACAAGGTTGACGC |
| S4GEB | NA | S4GEB\_3 | GTGTCAAGGTCCG |
| S4GEH | NA | S4GEH\_1 | CTTGAGGTGAAGA |
| S4GEH | NA | S4GEH\_2 | GGCCACACGTTAA |
| S4GEH | NA | S4GEH\_3 | CGTTGTATTGCTA |
| S3GDB | NA | S3GDB\_1 | TTAGGTGAATTTG |
| S3GDB | NA | S3GDB\_2 | GCCATGCTATTAT |
| S3GDB | NA | S3GDB\_3 | GCAATAAGGAGCG |
| S3GDH | NA | S3GDH\_1 | GCGCTCTCCTACG |
| S3GDH | NA | S3GDH\_2 | GGTTCCATGTAAG |
| S3GDH | NA | S3GDH\_3 | CTATGTCCGTATT |
| S3GEB | NA | S3GEB\_1 | CGAGGATTCGATC |
| S3GEB | NA | S3GEB\_2 | AATTCACGGAGAA |
| S3GEB | NA | S3GEB\_3 | GTGGTGTCCGCAG |
| S3GEH | NA | S3GEH\_1 | CAATCGATGGGCC |
| S3GEH | NA | S3GEH\_2 | GACAAGATTCCTA |
| S3GEH | NA | S3GEH\_3 | CGGTCGCGCCATT |
| NC | NA | NC1\_1 | GTCACATAGATTG |
| NC | NA | NC1\_2 | CATTGTAGAGTGA |
| NC | NA | NC1\_3 | TGCGATGTCTGAC |
| NC | NA | NC2\_1 | TAACACTGGCTAC |
| NC | NA | NC2\_2 | CCACCTCTTCGCG |
| NC | NA | NC2\_3 | TGGTCAACCTTGC |
| NC | NA | NC3\_1 | TACTGACCTCGCC |
| NC | NA | NC3\_2 | CCAGCGCAATTCC |
| NC | NA | NC3\_3 | TGGTTAGAACCAG |
| NC | NA | NC4\_1 | TAGGTAAGAGTCT |
| NC | NA | NC4\_2 | CCGCGGTAAGTAA |
| NC | NA | NC4\_3 | TTATTCCTTCCGC |
| NC | NA | NC5\_1 | TCTAATTGGACTC |
| NC | NA | NC5\_2 | CCTAGGCGGAGGT |
| NC | NA | NC5\_3 | TTTCGGCATCGGC |
| NC | NA | NC6\_1 | TCTCCATCAACCT |
| NC | NA | NC6\_2 | CCTCTCGGCGGTT |
| NC | NA | NC6\_3 | TTGCACGCAAGGC |
| NC | NA | NC7\_1 | TGGTCTTGGAACC |
| NC | NA | NC7\_2 | AACCGTATTGAAC |
| NC | NA | NC7\_3 | TTGCCGCCGACCT |
| NC | NA | NC8\_1 | TGTATAGCGGAGC |
| NC | NA | NC8\_2 | AAGGCGCCATGGC |
| NC | NA | NC8\_3 | TTGGTGGCTTGAT |
| PCRN | NA | PCRN\_1 | TTACGGAGGCTCA |
| PCRN | NA | PCRN\_2 | AATTCCGTGGCGA |
| PCRN | NA | PCRN\_3 | TTGTGGCGGCGTC |

**Table S3:** Full list of statistical models compared using the Information Theory approach and AICc, to find variables the source of CO2 trap lure (“trap”), storage method (“storage”), and body segments (“part) influencing the number of species (“number”) detected on sequencing platforms: Illumina, MinION.

|  |  |  |
| --- | --- | --- |
| **Dataset utilised**  | **Model ID** | **Model parameters** |
| Illumina | ill\_0 | number ~ 1 + (1|day) |
| ill\_1 | number ~ trap\*storage\*part + (biomass|day) |
| ill\_2 | number ~ trap\*storage\*part + (1|day) |
| ill\_2a | number ~ trap\*storage + trap\*part + (1|day) |
| ill\_2b | number ~ trap\*storage + storage\*part + (1|day) |
| ill\_2c | number ~ trap\*part + storage\*part + (1|day) |
| ill\_3 | number ~ trap + storage + part + (1|day) |
| ill\_4 | number ~ trap\*storage + (1|day) |
| ill\_5 | number ~ trap + storage + (1|day) |
| ill\_6 | number ~ trap\*part + (1|day) |
| ill\_7 | number ~ trap + part + (1|day) |
| ill\_8 | number ~ storage\*part + (1|day) |
| ill\_9 | number ~ storage + part + (1|day) |
| ill\_10 | number ~ trap + (1|day) |
| ill\_11 | number ~ storage + (1|day) |
| ill\_12 | number ~ part + (1|day) |
|  |  |  |
| MinION | min\_0 | number ~ 1 + (1|day) |
| min\_1 | number ~ trap\*storage\*part + (biomass|day) |
| min\_2 | number ~ trap\*storage\*part + (1|day) |
| min\_2a | number ~ trap\*storage + trap\*part + (1|day) |
| min\_2b | number ~ trap\*storage + storage\*part + (1|day) |
| min\_2c | number ~ trap\*part + storage\*part + (1|day) |
| min\_3 | number ~ trap + storage + part + (1|day) |
| min\_4 | number ~ trap\*storage + (1|day) |
| min\_5 | number ~ trap + storage + (1|day) |
| min\_6 | number ~ trap\*part + (1|day) |
| min\_7 | number ~ trap + part + (1|day) |
| min\_8 | number ~ storage\*part + (1|day) |
| min\_9 | number ~ storage + part + (1|day) |
| min\_10 | number ~ trap + (1|day) |
| min\_11 | number ~ storage + (1|day) |
| min\_12 | number ~ part + (1|day) |

**Table S4:** List of models that fulfil ΔAICc<2 criterion.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model ID** | **Intercept** | **Df** | **Loglik** | **AICc** | **Delta** | **weight** |
| ill\_10 | 2.386 | 3 | -87.803 | 182.4 | 0.00 | 0.514 |
| ill\_5 | 2.343 | 4 | -87.529 | 184.3 | 1.99 | 0.190 |
| min\_10 | 2.33 | 3 | -86.514 | 179.8 | 0.00 | 0.526 |

**Table S5:** Table of final Poisson models after stepwise simplification \* indicating that p-value <0.05.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **Estimate** | **Standard error** | **z-value** | **p-value** | **Marginal R2** | **Conditional R2** |
| Species counts ~ Source of CO2 trap lure + (1|Day) **Illumina dataset**  |
| Trap  | -0.781  | 0.130  | -6.02  | 1.75e-9\*  | 0.472  | 0.646  |
| Species counts ~ Source of CO2 trap lure + (1|Day) **MinION dataset**  |
| Trap  | -0.794  | 0.134  | -5.95  | 2e.67-9\*  | 0.465  | 0.646  |
| Species counts ~ Sequencing platform + (Biomass|Day) **Combined dataset**  |
| Platform  | -0.056  | 0.075  | -0.75  | 0.453  | 0.003  | 0.462  |

**Text S1:** Sampling location and permission.

Adult mosquitos were sampled in the forest with the permission of the Singapore Armed Forces as a joint biosurveillance project. Exact sampling locations are confidential.

**Text S2:** Detailed PCR protocol.

Each PCR reaction volume is composed of 8 µL of commercial Mastermix from CWBio, 1 µL of 1 mg/mL BSA, 1 µL of 20 µM of each primer and 5 µL of DNA extract (diluted to the concentration of 2 ng/µL). The thermocycling condition was as follows: a starting denaturation step of 3 min at 94 ⁰C, followed by 25 cycles of 30 s at 94 ⁰C, 1 min at 45 ⁰C and 30 s at 72 ⁰C, and a final extension step of 3 min at 72 ⁰C.

**Text S3:** Additional modifications made to MinION library prep protocol.

The first modification was the exclusion of FFPE DNA repair mix in the end-repair reaction and instead, the reaction consisted of 50 µL of DNA, 7 µL of Ultra II End-prep reaction butter (New England Biolabs, USA), 3 µL of Ultra II End Prep enzyme mix (New England Biolabs, USA). The second modification was using a 1 × ratio of Ampure XP beads for all steps.

**Text S4**: Brief overview of proof-of-concept experiment, demonstrating the effectiveness of our DNA barcode reference library curated by expert taxonomists.

Before initiating the present study, we conducted a proof-of-concept experiment to test our DNA barcode reference library. This carefully curated library includes DNA barcodes of mosquitoes captured locally and identified by expert taxonomists (Yeo et al., 2021; Yeo et al., unpublished), or collated from regional DNA barcoding projects in the literature(Chan et al., 2014). In this experiment, mosquitoes were individually identified to species resolution by expert taxonomists and subsequently pooled to create six mock communities comprising 50 specimens of 4 –10 species per sample. These samples were then processed using the DNA metabarcoding protocol. By leveraging on our carefully curated DNA reference library, we successfully and accurately determined the species composition of all mock community. Notably, this approach even facilitated the detection of cryptic species.

**Text S5:** Statistical modelling and data analysis conducted on R software.

Statistical analysis was performed in R v. 4.2.1 (R Core Team, 2018). Pseudo-replicates were combined prior to further analysis. In order to find out the effect of different variables on the species richness detected, we fitted generalised linear mixed effect models to the data using the *glmer* function in the *lme4* package (Bates et al., 2015). The explanatory collection variables “Source of CO2 trap lure”, “Body part”, and “Storage method” were included as fixed effects to assess whether changes in the collection variables would affect the number of mosquito species detected. For the choice of sequencing platform, “Sequencing platform” was included as the sole fixed effect to assess whether the choice of sequencing platform affects the number of mosquito species detected. Maximal models were simplified stepwise to find statistically significant variables. This was corroborated with the results of comparing models by assessing delta-AICc values (i.e., the Information Theoretic approach), using the *model.sel* function in the *MuMIn* package (Bartoń, 2020). To find out the effect of using different sequencing platforms on species composition, we used permutational multivariate analysis of variance (PERMANOVA).

**Figure S1:** Heatmap of mosquito species detected on both sequencing platforms.



**Literature Cited**

Bartoń, K. (2020). *MuMIn: Multi-modal inference. Model selection and model averaging based on information criteria (AICc and alike)* (R package version 1.43.17) [Computer software]. http://cran.r-project.org/web/packages/MuMIn/index.html

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–51. https://doi.org/10.18637/jss.v067.i01

Chan, A., Chiang, L.-P., Hapuarachchi, H. C., Tan, C.-H., Pang, S.-C., Lee, R., Lee, K.-S., Ng, L.-C., & Lam-Phua, S.-G. (2014). DNA barcoding: Complementing morphological identification of mosquito species in Singapore. *Parasites & Vectors*, *7*(1), 569. https://doi.org/10.1186/s13071-014-0569-4

R Core Team. (2018). *R: A Language and Environment for Statistical Computing.*

Yeo, H., Yeoh, T. X., Ding, H., Lee, T. T. M., & Puniamoorthy, N. (2021). Morphology and mini-barcodes: The inclusion of larval sampling and NGS-based barcoding improves robustness of ecological analyses of mosquito communities. *Journal of Applied Ecology*, *58*(10), 2087–2100. https://doi.org/10.1111/1365-2664.13966