Sociobiology
An international journal on social insects

RESEARCH ARTICLE - ANTS

Sequencing the Ant fauna of a Small Island: Can Metagenomic Analysis Enable Faster Identification for Routine Ant Surveys?

JD Majer[1,2,3], MA Castalanelli[4], JL Ledger[4], NR Gunawardene [2], BE Heterick[2,3]

1 - School of Biological Sciences, University of Western Australia, Perth, WA 6009, Australia
2 - Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia
3 - Western Australian Museum, Perth, WA 6000, Australia
4 - EcoDiagnostics Pty. Ltd., 46 - 48 Banksia Rd, Welshpool WA 6106, Australia

Article History

Edited by
Gilberto M. M. Santos, UEFS, Brazil
Received 09 February 2018
Initial acceptance 28 March 2018
Final acceptance 10 Abril 2018
Publication date 02 October 2018

Keywords
Next Generation Sequencing, Barrow Island, Evaluation, Biosecurity, Ant incursion.

Corresponding author
Jonathan D. Majer
School of Biological Sciences
University of Western Australia
Perth, WA 6009, Australia
E-Mail: jonathan.majer@uwa.edu.au

Abstract

All known ant species from a small Western Australian island were subjected to DNA barcoding of the CO1 gene, with a view to using the database to identify ants by Next Generation Sequencing in subsequent, routine surveys. A further aim was to evaluate whether the data could be used to see if any new species had arrived on the island since the total fauna had been inventoried. Of the 125 unique ant species then known from the island, 72 were successfully barcoded. Those that were refractory to amplification were largely the result of sample age and/or contamination. Following this base-line barcoding, ants were sampled from 14 regular sampling sites and ant sequences were obtained from the bulked 'metagenomic soup'. Prior to doing this, a parataxonomist had identified all ant species in the samples and returned them to the 'soup'. Successful identification for each site varied from 38% (Sites 12 and 27) to 100% of species (Site 10). Comparison of the number of species recovered with the number of sequences obtained from each sample showed a positive correlation between the two variables. When a site had >1,000 sequences, the average recovery rate was 79%, which is in contrast to the lowest four recovery rates (Site samples 12, 22, 26 and 27), which had fewer than 440 amplicon sequences. The ability to detect individuals that occur at low frequencies is also important. We analysed each site individually to determine if a species was detected and how that related to the proportion of individuals in the pooled sample. Where a species was present at <4% of the total sample, it was only detected 10% of the time, indicating that adequate sequencing depth is critical to species recovery. We conclude that this technique was only partially successful in replacing conventional taxonomy and that it could have limited ability to detect incursions unless the new arrival is abundant. Current barcoding is no longer limited to the CO1 gene and other genes are characterised for identification of intractable groups where CO1 does not provide appropriate levels of resolution.

Introduction

Recent advances in sequencing technology (Next Generation Sequencing or NGS) have expanded the possibilities of using DNA barcoding to identify species in complex environmental surveys (Telfer et al., 2015). Using improved NGS techniques, often referred to as "Metagenomics", the effects of environmental change (e.g., deforestation, resource extraction, site development) can be examined at the level of whole ecosystems (e.g., Beng et al., 2016; Gibson et al., 2014; Ji et al., 2013; Smith et al., 2005; Yang et al., 2014; Yu et al., 2012). This can be achieved by establishing a baseline record of biodiversity within an ecosystem and then routinely comparing the

current biodiversity to that of the baseline, a procedure that is exemplified by the approaches of Yu et al. (2012) and Kocher et al. (2017). Rather than sorting and separating specimens, DNA is extracted from a batch or 'soup' of specimens from a field collection (often >100 specimens in a vial) and the DNA sequenced using NGS techniques (Pochon et al., 2013; Yoon et al., 2016). This technique allows hundreds of different taxa to be sequenced simultaneously and the sequence data that is generated can be compared, as a batch, to the DNA reference library. Outputs from studies of this type provide huge data-sets, and lead to a good understanding of sample diversity at a reasonable cost (Ji et al., 2013). More importantly, previously unrecorded species that are rare, cryptic, or are invasive may be identified because their genetic signature differs from the reference database.

The Barrow Island Invertebrate Surveillance project provided an opportunity to evaluate the application of this technique in an invertebrate species context, concentrating on the ant fauna. Barrow Island is Western Australia's second largest offshore island; it is a Class A Reserve and also happens to be Australia's only land-based oil field. The presence of industrial interests on the island, and also its important conservation value, has meant that the island is not publicly accessible. This has resulted in the exclusion, or control and eradication of non-indigenous or invasive species. In 2009, Chevron Australia Pty Ltd and its Joint Venture Participants undertook the construction of a liquefied natural gas plant on the island. One of the conditions under which approval for the plant was granted was the implementation of a rigorous biosecurity effort to ensure that no non-indigenous species (NIS) were allowed to establish on the island and, if any new species were to be introduced, have a 0.8 probability of detection if they are present. To fulfill this condition a non-indigenous species surveillance program was implemented. If NGS procedures were to be used for diagnostics, it would be critical for the technique to be sensitive and reliable enough to detect previously unrecorded species. The aim of this pilot study is to evaluate whether this is achievable using a single gene.

A series of systematic surveys on flora and fauna has been performed using purpose-designed sampling protocols in order to provide baseline data on the existing terrestrial invertebrate species on Barrow Island. As part of the fauna surveys, terrestrial invertebrates were sampled between 2005 and 2008 (Majer et al., 2013). Callan et al. (2011) initially recorded a total of 1,873 species and morphospecies, with subsequent surveys and taxonomic developments increasing the count to 2,670 species with 25 invertebrate species considered non-indigenous to BWI (Thomas et al., 2017). The Barrow Island collection represents one of the few areas in Australia where sampling of invertebrates has occurred before and after development.

The ant species on Barrow Island are well-documented, totalling 125 (since upgraded to 129) species, none of which are endemic to the island (Heterick, 2013). The presence of a voucher specimen library of dry and wet preserved ant species from the island enabled us to establish a reference DNA barcoding library for the Barrow Island ants. DNA was extracted and barcoded for specimens of each species to establish a DNA reference database. As regular, repeated surveys of invertebrate species are still ongoing on Barrow Island, a pilot study was then conducted to test whether the ant specimens collected in subsequent samples could be verified using the NGS technique and the DNA reference database. Specifically, we evaluated the efficacy of universal forward primer CI-J-1718 and the reverse primers HCO and CI-N-2191 in recovering and identifying multiple ant species within a trap using a Roche's 454 GS-Junior metabarcoding approach**.** GS-Junior was selected due to its ability (at the time) to sequence >400 bps per direction, and this was seen as a cost effective solution.

We intended that the reference database of ant species barcodes could be used by future researchers to rapidly determine the species composition within samples taken from the field, a process that normally takes two weeks of a taxonomist's time. In view of the fact that our surveys were designed to detect whether any NIS had been introduced during the construction of the gas liquification plant, it was thought that the technique also might have the potential to identify non-indigenous ant species if present within the collection. The NGS procedure that we utilised has since been superseded by more refined techniques such as Illumina Mi and Hi-Seq. Nevertheless, we consider it timely to report on our experiences with this procedure and to consider the feasibility of routinely using newer barcoding procedures for identifying ants and, ultimately, other invertebrates in bulk samples.

Methods

Invertebrate sampling: All ant species from the Barrow Island voucher specimen collection were identified morphologically by BEH. Specimens were vouchered between 2005 and 2008 at Curtin University and are now curated at the Western Australian Museum, with a duplicate set at the Western Australian Department of Primary Industries and Regional Development. The majority of the voucher collection is dry-preserved, with some species requiring vouchers from more recent surveys due to the deterioration of the original voucher specimen.

Single specimens of all 125 ant species (Appendix 1) were submitted for barcoding in order to provide baseline data for the ant fauna of the Island. Then, in order to test whether it is feasible to identify ants from unsorted trap samples, 14 of the Barrow Island sites were sampled in September 2013 using multiple trapping methods. These were: 1) Night Hand Collection (NHC), a method whereby trained field workers collect ants by hand in the evening (this method typically yields low abundances of ant species but more cryptic diversity); 2) Window Trap (WIN), which is a water trap with

a Perspex window that captures flying insects (one drawback with this method is that many ant species are attracted by the water); 3) Suction Samples (SUC), a method whereby a garden blower/vacuum suctions small insects from low shrubs and branches and which usually yields high abundances of shrub-foraging ant species; and 4) Barrier Pitfall Trap, Bait Trap and Litter Trap (BBL), a combination of three sampling methods that focus on ground-dwelling ants and often yield high abundances of a few species. For the purpose of this exercise, all samples from a site were combined. All other invertebrate groups were removed from the samples, and the ants were identified to morphospecies by a parataxonomist before being returned to 14 vials representing the bulked samples for each of the sampling sites. The specimens were preserved in 70% ethanol.

Barcoding procedure: A non-destructive DNA extraction method, ANDE (Castalanelli et al., 2010), was used to extract DNA from the morphologically identified ant specimens and from the bulked, unsorted trap samples. Amplification of the target barcoding region from individual ant specimens was performed using the Cytochrome Oxidase 1 (CO1) primers outlined in Table 1. Following PCR amplification of the target region and subsequent DNA sequencing, sequences were edited using Geneious Pro 8.0.3 (Biomatters Ltd) and aligned with the reference data set using Geneious' built-in alignment algorithm. Geneious Pro 8.0.3 was used to detect the presence of NuMTs by translating each CO1 sequence with the standard invertebrate and *Drosophila* codes. Forward and reverse sequences were manually edited, primer sequences removed, and the final quality checked. Consensus sequences were used to interrogate all available public sequence databases to determine if the morphological and molecular results used to determine the identifications were congruent.

Table 1. Cytochrome Oxidase 1 primers used to generate DNA barcodes for the reference database.

Primer	Sequence 5' – 3'
CI-J-1718	GGA GGA TTT GGA AAT TGA TTA GTT CC
CI-N-2191	CCC GGT AAA ATT AAA ATA TAA ACT TC
LCO	GGT CAA CAA ATC ATA AAG ATA TTG G
HCO	TAA ACT TCA GGG TGA CCA AAA AAT CA

DNA originating from each site sample was amplified using the forward primer CI-J-1718 and the reverse primers HCO and CI-N-2191, with the additional M13 sequences added to the 5` end of the forward and reverse primers. A second round of PCR amplification was performed to attach the Roche Lib-A adapters to the previous PCR product. A unique MIDTag barcode specific to each individual Barrow Island site sample was also incorporated, allowing multiple samples to be pooled together for sequencing on the GS Junior. The GS Junior sequencing run was set up as per the manufacturer's instruction, and run for 200 flows.

The NGS run was conducted using Roche GS Junior (454). This NGS platform was selected due to its lower cost and its ability to generate sequence lengths >400bp per direction. The sequencing run was performed at the Western Australian State Agriculture Biotechnology Centre (SABC).

Analysis of the NGS data was conducted using an EcoDiagnostics Pty. Ltd. in-house bioinformatics pipeline which de-convoluted the DNA sequences into individual site samples and then compared sets of sequences from each sample to the CO1 reference database.

Results

Baseline ant data: a summary of the sample sequencing outcomes for the 126 ant species is shown in Table 2 and sequences for each of these species are deposited in Genbank. In total, 72 species were successfully DNA barcoded and the remaining 53 were unsuccessful. Five samples returned a DNA barcode that was incongruent with the morphological result and were considered contaminated due to their molecular similarity to the barcode for a species of gastropod that is commonly found in some of the invertebrate samples. Of the remaining 48 species that failed to generate a sequence age of the specimens was a possible reason for failure. For 13 species that failed to amplify, sequences from public databases (i.e., NCBI; Genbank) were available and hence substituted for the failed amplification.

Table 2. Summary of sample sequencing outcomes.

Sequencing Result	Count
Successfully sequenced	72
No sequence generated	53
Sample contaminated	5
External public database sequence (e.g. NCBI)	13
Total	125

Pooled sample analysis: The ant species found in each of the 14 pooled samples and the species that were recovered by NGS from each site are shown in Table 3. The sequencing run generated 42,098 high quality sequences; any sequences <400bp in length were removed, reducing the data set to 23,072 sequences from the 14 site samples. The number of sequences dramatically varied between site samples, with between 13 (Site sample 12) and 10,046 (Site sample 17; Table 4) amplicon sequences being recovered.

Valid assignments were made when similarity of the NGS sequence to one of the reference species was greater than 95%. Since the samples were taken in September 2013, which represents only a small time capsule of the total invertebrate surveillance effort, only 39 out of the 126 ant species were present within the 14 site samples (Table 3). Six species in the 14 site samples did not have a corresponding CO1 reference sequence. Despite several sites missing one to three reference species (Table 4), the majority of site samples had >92% of their sequences assigned to a reference (Table 4).

Table 3. Comparison between morphological identification and molecular detection using Next Generation Sequencing.

Species/Site ID	9	10	11	12	14	16	17	20	21	22	23	24	26	27
Anochetus rectangularis	B													
Camponotus capito			M									N		M
Camponotus gibbinotus		B								M	M			M
Camponotus fieldeae	B		B			M	B			M	M	B	B	M
Camponotus scratius					B									
Camponotus simpsoni										M				
Cardiocondyla atalanta											M			
Iridomyrmex anceps		B	B			B		M	B	B	B	B		B
Iridomyrmex cephaloinclinus				B										
Iridomyrmex chasei					M		B		M	M				
Iridomyrmex dromus	N					N	N				N			
Iridomyrmex exsanguis	B		B	M			M		B	M	B			B
Iridomyrmex minor				B	B	B		B	B	B	B	M	M	B
Iridomyrmex mjobergi											M			
Iridomyrmex sanguineus	M		M					M	N	M	M	M	M	M
Melophorus biroi									M					M
Melophorus paramorphomenus												B		
Melophorus hirsutipes									B					M
Meranoplus curvispina				M						M				
Monomorium laeve										M	M	M		
Monomorium leae								M						
Monomorium sydneyense				B				B		B			B	
Ochetellus flavipes								B						
Ochetellus sp. JDM 527			B					B		B			M	
Paratrechina longicornis		B										B		
Pheidole turneri							B							
Pheidole variabilis	B	B		B			B	B	B	M		B		B
Polyrhachis ammonoeides					M			B	M					
Polyrhachis inconspicua	M			M			M			M				M
Polyrhachis seducta											M			
Polyrhachis sp. JDM 808			M	M				M	N	M	M	M	M	
Rhytidoponera crassinoda											M			M
Rhytidoponera taurus											M		M	M
Tetramorium spininode		M												
Tetramorium striolatum complex		M												

* M = Morphological identification only; N = Next generation sequencing detection only;
B = Both morphological identification and Next Generation Sequencing detection

The only exception was site sample 9, where the percentage of unassigned species was 31% (Table 4). Analysis of these unassigned sequences showed that in the majority of cases they clustered with *Camponotus*, *Iridomyrmex,* and *Polyrhachis* clades but weren't closely related to any particular species (>15% pairwise divergence from its closest neighbor). One issue known to occur during PCR when multiple templates are present is cross amplification of two species (Hass et al., 2017), i.e. the front half is of one species the back half is of another. Combined, they create a unique chimeric sequence that cannot be assigned to a reference.

The term "recovery" is here defined as the number of species identified using the NGS barcoding approach compared to the number of morphologically identified species. Recovery for each site varied from 38% (Sites 12 and 27) to 100% (Site 10; Table 4). Six sites had additional species that

were not recognised by morphological methods. Of particular note were the highly similar species *Iridomyrmex exsanguis* and *Iridomyrmex dromus* (Table 3).

Comparison of the number of species recovered with the number of sequences obtained from each trap sample shows a positive correlation between the two variables (Fig 1). When a site had >1,000 sequences, the average recovery rate was 79%, which is in contrast to the lowest four recovery rates (Site samples 12, 22, 26 and 27), which had fewer than 440 sequences. One exception was Site 24, which produced a low number of sequences but still had 67% recovery rate (Table 4).

Table 4. Overview of the NGS sequencing data, including the number of species present at each site and whether they were recovered by NGS. Also shown are the outcomes of the NGS sequencing run output for each site, including the number of sequences that were assigned with > 95% similarity to a reference sequence.

Site ID	9	10	11	12	14	16	17	20	21	22	23	24	26	27
No. of species identified morphologically	6	6	7	8	4	3	6	10	8	15	13	9	7	13
No. of species identified with NGS	6	6	6	3	2	3	6	7	7	7	7	7	3	5
No. of morphologically identified species in NGS result	5	6	6	3	2	2	5	7	5	7	6	6	3	5
No. of species missed by NGS	1					1	1		2		1	1		
Percentage of species identified with NGS	83	100	86	38	50	67	83	70	63	47	46	67	43	38
No. of sequences	2,431	1,498	651	13	876	3,796	10,046	481	1,871	256	434	308	26	385
No. of species missing a CO1 reference			1	1		2			3	2	1	2	1	3
Percentage of assigned sequences	69	95	96	100	97	97	97	95	97	95	98	93	96	96

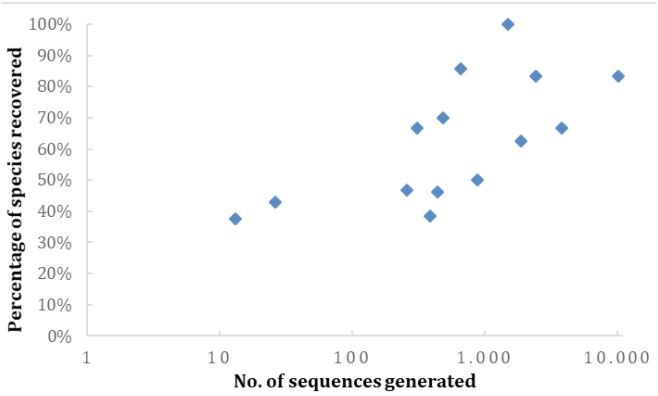

Fig 1. Number of 454 sequences generated per sample in relation to the percentage of species recovered.

Validating a highly sensitive technique with the ability to detect individuals that occur at low frequencies is one of the most important functions of any biosecurity venture. To examine the sensitivity of the NGS technique, we analysed each site individually to determine if a species was detected (hit) and how that related to the number of individuals per species (determined by dividing the number of species identified per site by the total number of individuals per site sample [termed species occurrence).

Figure 2 indicates that where a species was present at <4% of the total sample size, it was only detected 10% of the time. As the frequency at which a species occurred increased, so too did the rate at which that species was detected. The only exceptions were *Iridomyrmex minor* at Sites 24 and 26, *Iridomyrmex chasei* at Site 24, *Camponotus fieldeae* at Site 16, *Monomorium laeve* at Site 24, and *Polyrhachis ammonoeides* at Site 14.

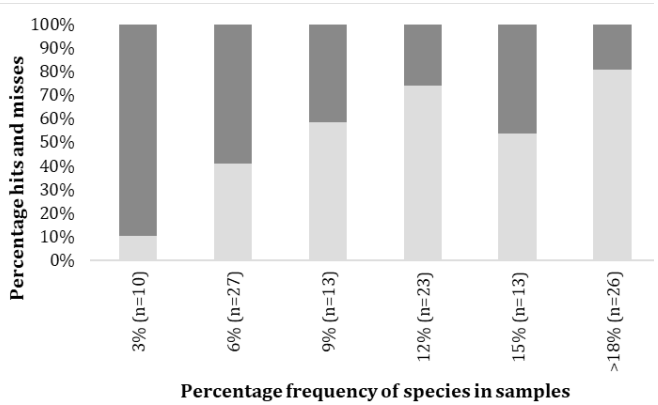

Fig 2. Sensitivity of the Next Generation Sequencing technique, illustrated by showing the rate at which a species was detected in preserved Barrow Island ant material. The "Hit" (light part of bar) shows the degree of success relative to the frequency at which that particular ant species occurred within the material.

Table 5. Comparison of species recovery over all sites and number of mutations within the CI-J-1715 priming site. The species that were not recovered are underlined.

Species	Species Recovery Rate (%)	Number of Mutations (26bp)	Number of Mutations @ 5` (5bp)
Anochetus rectangularis	100	1	0
Camponotus capito	50	3	1
Camponotus gibbinotus	25	3	1
Camponotus fieldeae	56	4	1
Camponotus scratius	100	0	0
Camponotus simpsoni	0	2	1
Cardiocondyla atalanta	0	5	1
Iridomyrmex anceps	89	3	1
Iridomyrmex cephaloinclinus	100	2	1
Iridomyrmex chasei	50	2	1
Iridomyrmex exsanguis	63	NA*	Missing datum
Iridomyrmex minor	80	NA	NA
Iridomyrmex mjobergi	0	0	0
Iridomyrmex sanguineus	50	2	1
Melophorus biroi	0	NA	NA
Melophorus paramorphomenus	100	4	1
Melophorus hirsutipes	50	NA	Missing datum
Meranoplus curvispina	0	4	1
Monomorium laeve	0	3	2
Monomorium leae	0	2	0
Monomorium sydneyense	100	2	1
Ochetellus flavipes	100	3	1
Ochetellus sp. JDM 527	75	5	1
Paratrechina longicornis	100	3	1
Pheidole turneri	100	1	0
Pheidole variabilis	78	1	0
Polyrhachis ammonoeides	33	2	1
Polyrhachis inconspicua	0	NA	Missing datum
Polyrhachis seducta	0	NA	NA
Polyrhachis sp. JDM 808	43	2	0
Rhytidoponera crassinoda	0	3	1
Rhytidoponera taurus	66	4	1
Tetramorium spininode	100	1	0
Tetramorium striolatum complex	100	0	0

* NA indicates that no data was available to examine bind efficiency.

Discussion

This study demonstrates the encouraging potential of NGS metabarcoding to characterise ant species from bulk trap samples, despite the study being resource-limited (only a single sequencing run was costed). A number of technical difficulties have been highlighted in this pilot study. These include variation in the number of sequences recovered between trap sample sites (13 to 10,046) and lack of sequencing depth, which clearly affects the ability to recover species. This can be largely overcome by the rapid development of improved technologies, and we acknowledge that the sequencing platform and chemistry used in this study has largely been superseded and discontinued. Newer procedures will lead to significant improvements in the recovery of sequences from mixed trap samples that here varied between 38 and 100% for individual trap samples. Another major factor that needs to be overcome is preferential amplification. Future projects need to select priming sites that are either void of mutations or have minimal mutations and are shown to detect all intended target species in a comprehensive fashion. Increased depth and more suitable primers may lead to the development of robust and practical monitoring methods with a very high diagnostic sensitivity and specificity.

The occasional incongruence between species identified by NGS and morphological species was probably caused by contamination from unrelated taxa due to their molecular similarity to the barcode concerned. This probably arose in some of our samples because traces of gastropod DNA, which were often found in our samples prior to removal of the ants, were preferentially amplified. These invertebrates were much larger than the ants and exuded a DNA-rich slime. Because of the nature of the project and its restricted resourcing, only three primer pairs could be used to try and generate sequences. These three primer pairs (Table 1) are generalist primers that have been shown to successfully amplify genetic material from invertebrates (Simon et al., 1994). However, experience has shown that 20% or more of the samples will be refractory to amplification due to primer mismatch, poor quality DNA, and PCR inhibition. Future work should involve the design of more specific primers and, if possible, fresh samples that haven't been collected in pitfall traps. (Pitfall trapped material reveals rapid degradation of DNA and may also have high levels of contaminating DNA present (Castalanelli et al., 2011).)

The variation in number of sequences between sites may be in part be attributed to the preservation of specimens in a lower grade of ethanol (i.e., 70%), which was used throughout the NIS project for specimen preservation. This may have contributed to the lack of amplification success and also to technical difficulties in making the PCR products from each trap sample of similar molarity prior to NGS library preparation.

Recovery and sensitivity: The ability to detect an individual accurately within a particular sample is one of the most important functions of biosecurity surveillance, regardless of whether it involves morphological or molecular techniques. Therefore, we believe that the most important aspect of this pilot study is understanding recovery and sensitivity.

Underpinning this is the need for good baseline data from conventional morphological taxonomic approaches. As mentioned earlier, *Iridomyrmex exsanguis* and *Iridomyrmex dromus* were not distinguished by morphological methods (Table 3). These two species are morphologically very similar, but seem to have a different nest structure and behaviour around the nest. Physically, they can only be distinguished with difficulty. The *Iridomyrmex exsanguis* worker is always pale yellow and has a noticeable propodeal angle, i.e., is truncate when viewed in profile. *Iridomyrmex dromus* is commonly pale also, but the colour can range from depigmented yellow to almost black, and viewed in profile the propodeum lacks a noticeable angle, i.e., is not truncate. Anyone considering using NGS procedures should be aware of this sort of subtlety in taxonomic differentiation.

Apart from the anomaly with site 24, these results suggest that an important contributor to recovery for a given sample is the number of sequences; namely, the more sequences, or greater sequencing depth, the greater the recovery. These results are congruent with other studies that used newer technologies. For instance, Brandon-Mong et al. (2015) evaluated the MiSeq (Illumina) and showed that for lepidopteran specific primers, 106,070 sequences recovered 60%, which was increased to 80% when the number of sequences was increased to 685,208.

The results presented here can generally be firstly explained by preferential amplification. While species occurrence is likely to cause some bias towards preferential amplification, it seems that primer design is the most probable cause. Since full-length DNA barcoding only allowed us to examine the primer binding site for CI-J-1715, only this section was scrutinised (Table 5). The data suggest that as the mutations within the priming site increase, the chance of recovery decreases. The noteworthy examples where primers clearly contributed to good sequence recovery were *Anochetus rectangularis*, *Camponotus scratius*, *Pheidole turneri*, *Tetramorium spininode*, and *Tetramorium striolatum*, all of which had either zero mutations or a single mutation and 100% recovery. In comparison, *Monomorium laeve* and *Cardiocondyla atalanta*, which had three and four mutated sites, respectively, failed to be recovered; more importantly *Monomorium laeve* had two mutations that occurred within the 5` binding site which is the most important part of the priming site (Table 5). Interestingly, the species that were not recovered (Table 5: underlined) tended to have mutations that ranged from two to five and a species occurrence of <7%; suggesting that priming mutations and rarity compounded the failure to recover a species.

Lessons learned: This pilot study successfully generated a reference ant DNA barcode database that is fundamental to the development of improved NGS metabarcoding and DNA based individual specimen identification approaches. Our investigation revealed a number of issues of which users should be aware and take care to address, namely:

- Age of material – try to use as fresh a sample as possible;
- Appropriate, high-grade preservative – use preservatives that maximise preservation of genetic material;
- Morphologically based taxonomy – underpin investigation with a sound taxonomic database;
- Contamination of samples by non-target taxa – be aware of other taxa, including plant material, that occur in the samples and how this could contaminate the DNA;
- Cross amplification between related taxa – be aware of this possibility; and
- Cost of barcoding has to be considered at the outset of investigations such as this.

Returning to the original thrust behind this investigation, can barcoding using the CO1 gene be used to detect ant incursions with an 80% confidence of detection? With recovery rates averaging only 79% and sometimes falling as low as 38%, the answer is no. However, though there are acknowledged gaps in the database that we have generated, these can be rectified with further study to increase the robustness of data interpretation and species identification. Furthermore, barcoding is no longer limited to CO1; more recently, other genes have been preferred for the intractable groups (e.g., see tables in Purty & Chatterjee, 2016).

Acknowledgements

The Gorgon Gas Development is operated by an Australian subsidiary of Chevron and is a joint venture of the Australian subsidiaries of Chevron (47.3%), ExxonMobil (25%), Shell (25%), Osaka Gas (1.25%), Tokyo Gas (1%) and JERA (0.417%). The authors are grateful to Chevron Australia Pty Ltd for on-island support with logistics, and the Gorgon Joint Venturers for financial support to undertake the surveillance component of this project. Funding from Curtin University was provided in order to undertake the barcoding of specimens. Christopher Taylor, Morgan Lythe, Brad Scanlon, and Callum Smithyman all assisted the project, either in the field and/or in the laboratory.

References

Beng, K.C., Tomlinson, K.W., Shen, X.H., Surget-Groba, Y., Hughes, A.C., Corlett, R.T. & Silk, J.F. (2016). The Utility of DNA Metabarcoding for Studying the Response of Arthropod Diversity and Composition to Land-use Change in the Tropics. Scientific Reports, 6: 24965. doi: 10.1038/srep24965

Biomatters Ltd. Geneious version 8.0.3. Biomatters. Available from http://www.geneious.com/

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: 717-27. doi: 10.1017/S0007485315000681

Callan, S. K., Majer, J. D., Edwards, K. & Moro, D. (2011). Documenting the Terrestrial Invertebrate Fauna of Barrow Island, Western Australia. Australian Journal of Entomology, 50: 323-343. doi: 10.1111/j.1440-6055.2011.00818.x

Castalanelli, M. A., Severtson, D. L., Brumley, C. J., Szito, A., Foottit, R. G., Grimm, M., Munyard, K. & Groth, D .M. (2010). A Rapid Non-destructive DNA Extraction Method for Insects and Other arthropods. Journal of Asia-Pacific Entomology, 13: 243-248. doi: 10.1016/j.aspen.2010.04.003

Castalanelli, M. A., Mikac, K. M., Baker, A. M., Munyard, K., Grimm, M. & Groth, D. M. (2011). Multiple Incursions and Putative Species Revealed Using a Mitochondrial and Nuclear Phylogenetic Approach to the *Trogoderma variabile* (Coleoptera: Dermestidae) trapping program in Australia. Bulletin of Entomological Research, 101: 333-343. doi: 10.1017/S0007485310000544

Gibson, J., Shokralla, S., Porter, T. M., King, I., Van Konynenburg, S., Janzen, D. H., Hallwachs, W. & Hajibabael, M. (2014). Simultaneous Assessment of the Macrobiome and Microbiome in a Bulk Sample of Tropical Arthropods through DNA Metasystematics. PNAS, 111: 8007-8012. doi: 10.1073/pnas.1406468111

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E. & Methé, B. (2011). Chimeric 16S rRNA Sequence Formation and Detection in Sanger and 454-pyrosequenced PCR Amplicons. Genome Research, 21: 494-504. doi: 10.1101/gr.112730.110

Heterick, B. E. (2013). A Taxonomic Overview and Key to the Ants of Barrow Island, Western Australia. Records of the Western Australian Museum, Supplement, 83: 375-404. doi: 10.18195/issn.0313-122x.83.2013.375-404

Ji, Y, Ashton, L., Pedley S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., Edwards, F. A. & Larsen, T. H. (2013). Reliable, Verifiable and Efficient Monitoring of Biodiversity via Metabarcoding. Ecology Letters, 16: 1245-1257. doi: 10.1111/ele.12162

Kocher, A., Gantier, J. C., Gaborit, P., Zinger, L., Holota, H., Valiere, S., Dusfour, I., Girod, R., Bañuls, A. L. & Murienne, J. (2017). Vector Soup: High-Throughput Identification of Neotropical Phlebotomine Sand flies Using Metabarcoding. Molecular Ecology Resources, 17: 172-182. doi: 10.1111/1755-0998.12556

Majer, J. D., Callan, S. K., Edwards, K., Gunawardene, N. R. & Taylor, C. K. (2013). Baseline Survey of the Terrestrial Invertebrate Fauna of Barrow Island. Records of the Western Australian Museum, Supplement, 83: 13-112. doi: 10.18195/issn.0313-122x.83.2013.013-112

Pochon, X., Bott, N. J., Smith, K. F. & Wood, S. A. (2013). Evaluating Detection Limits of Next-generation Sequencing for the Surveillance and Monitoring of International Marine Pests. PLoS ONE, 8(9): e73935. doi: 10.1371/journal.pone.0073935

Purty, R. S. & Chatterjee, S. (2016). DNA Barcoding: An Effective Technique in Molecular Taxonomy. Austin Journal of Biotechnology and Bioengineering, 3: 1059.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Flook, P. (1994). Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene Sequences and a Compilation of Conserved Polymerase Chain Reaction Primers. Annals of the Entomological Society of America, 87: 651-701. doi: 10.1093/aesa/87.6.651

Smith, M. A., Fisher, B. L. & Hebert P. D. N. (2005). DNA Barcoding for Effective Biodiversity Assessment of a Hyperdiverse Arthropod Group: the Ants of Madagascar. Philosophical Transactions of the Royal Society of London, B Biological Sciences, 360: 1825-34. doi: 10.1098/rstb.2005.1714

Telfer, A. C., Young, M. R., Quinn, J., Perez, K., Sobel, C. N., Sones, J. E., Levesque-Beaudin, V., Derbyshire, R., Fernandez-Triana, J., Rougerie, R. & Thevanayagam, A., (2015). Biodiversity Inventories in High Gear: DNA Barcoding Facilitates a Rapid Biotic Survey of a Temperate Nature Reserve. Biodiversity Data Journal, 3: e6313. doi: 10.3897/BDJ.3.e6313

Thomas, M. L., Gunawardene, N., Horton, K., Williams, A., O'Connor, S., McKirdy, S. & Van der Merwe, J. (2017). Many Eyes on the Ground: Citizen Science is an Effective Early Detection Tool for Biosecurity. Biological Invasions, 19: 1-15. doi: 10.1007/s10530-017-1481-6

Yang, C., Wang, X., Miller, J. A., De Blécourt, M., Ji, Y, Yang, C, Harrison, R.D. & Douglas, W. Y. (2014). Using Metabarcoding to Ask if Easily Collected Soil and Leaf-litter Samples Can be used as a General Biodiversity Indicator. Ecological Indicators, 46: 379-89. doi: 10.1016/j.ecolind.2014.06.028

Yoon, T. H., Kang, H. E., Kang, C. K., Lee, S. H., Ahn, D. H., Park, H. & Kim, H. W. (2016). Development of a Cost-effective Metabarcoding Strategy for Analysis of the Marine Phytoplankton Community. PeerJ, 4: e2115. doi: 10.7717/peerj.2115

Yu, D. W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C. & Ding, Z. (2012). Biodiversity Soup: Metabarcoding of Arthropods for Rapid Biodiversity Assessment and Biomonitoring. Methods in Ecology and Evolution, 3: 613-623. doi: 10.1111/j.2041-210X.2012.00198.x

SUPPLEMENTARY MATERIAL

Appendix 1. List of species supplied for DNA barcoding, indicating which were successfully barcoded, which failed and species for which sequences were obtained from the NCBH public database.

Genus	Species/sub-species	Author	Sequence	GenBank Accession Number
Aenictus	*turneri*	Forel	Pass	TBC
Amblyopone	sp.	[queen]	Pass	TBC
Anochetus	*rectangularis*	Mayr	Pass	TBC
Anochetus	*renatae*	Shattuck & Slipinska, 2012	Failed	
Arnoldius	sp. JDM 433		Failed	
Brachyponera	*lutea*	(Mayr)	Pass	TBC
Camponotus	*capito*	Mayr	Pass	TBC
Camponotus	*donnellani*	Shattuck & McArthur	Failed; Old	
Camponotus	*evae zeuxis*	Forel	Failed	
Camponotus	*fieldeae*	Forel	Pass	TBC
Camponotus	*gibbinotus*	Forel	Pass	TBC
Camponotus	*lownei*	Forel	Pass	TBC
Camponotus	*rubiginosus* complex sp. JDM 1158		Failed; Old	
Camponotus	*scratius*	Forel	Pass	TBC
Camponotus	*simpsoni*	McArthur	Pass	TBC
Cardiocondyla	*atalanta*	Forel	Pass	TBC
Cardiocondyla	*nuda*	(Mayr)	Contaminated	
Carebara	sp. JDM 1131		Failed	
Crematogaster	*laeviceps chasei*	Forel	Pass	TBC
Crematogaster	*queenslandica*	Forel	Failed	
Discothyrea	*clavicornis*	Emery	Contaminated	
Doleromyrma	*rottnestensis*	(Wheeler)	Failed	
Hypoponera	*queenslandensis*	(Forel)	Pass	TBC
Iridomyrmex	*agilis*	Forel	Failed; Old	
Iridomyrmex	*anceps*	Roger	Pass	TBC
Iridomyrmex	*bicknelli*	Emery	Pass	TBC
Iridomyrmex	*cephaloinclinus*	Shattuck	Pass	TBC
Iridomyrmex	*chasei*	Forel	Pass	TBC
Iridomyrmex	*coeruleus*	Heterick & Shattuck	Pass	TBC
Iridomyrmex	*difficilis*	Heterick & Shattuck	Pass	TBC
Iridomyrmex	*discors*	Forel	Pass	TBC
Iridomyrmex	*dromus*	Clark	Pass	TBC
Iridomyrmex	*exsanguis*	Forel	Pass	TBC
Iridomyrmex	*gibbus*	Heterick & Shattuck	Failed; Old	
Iridomyrmex	*hartmeyeri*	Forel	Failed	
Iridomyrmex	*minor*	Forel	Pass	TBC
Iridomyrmex	*mjobergi*	Forel	Pass	TBC
Iridomyrmex	*sanguineus*	Forel	NCBI	
Iridomyrmex	*suchieri*	Heterick & Shattuck	Pass	TBC
Iridomyrmex	*tenuiceps*	Heterick & Shattuck	Pass	TBC
Iridomyrmex	*xanthocoxa*	Heterick & Shattuck	Pass	TBC
Leptanilla	*swani*	Wheeler [males only]	Failed	
Leptogenys	*tricosa*	Taylor	Pass	TBC
Leptogenys	sp. JDM 1128		Failed	

Appendix 1. List of species supplied for DNA barcoding, indicating which were successfully barcoded, which failed and species for which sequences were obtained from the NCBH public database. (Continuation)

Genus	Species/sub-species	Author	Sequence	GenBank Accession Number
Lioponera	*brevis*	Clark	NCBI	
Lioponera	*clarki*	Crawley	Failed	
Lioponera	*iovis*	(Forel)	Pass	TBC
Lioponera	*longitarsus*	Mayr	Failed	
Lioponera	*ruficornis*	Clark	Pass	TBC
Lioponera	sp. JDM 942		Pass	TBC
Melophorus	*aeneovirens*	Wheeler	Pass	TBC
Melophorus	*biroi*	Forel	NCBI	
Melophorus	*hirsutipes*	Heterick, Castalanelli & Shattuck	Pass	TBC
Melophorus	*ludius*	Forel	NCBI	
Melophorus	*microtriches*	Heterick, Castalanelli & Shattuck	Pass	TBC
Melophorus	*paramorphomenus*	Heterick, Castalanelli & Shattuck	Pass	TBC
Melophorus	*parvimolaris*	Heterick, Castalanelli & Shattuck	Pass	TBC
Melophorus	*rufoniger*	Heterick, Castalanelli & Shattuck	Pass	TBC
Melophorus	*sulla*	Forel	NCBI	
Melophorus	*teretinotus*	Heterick, Castalanelli & Shattuck	Pass	TBC
Melophorus	*turneri*	Forel	Pass	TBC
Meranoplus	*dimidiatus*	Smith	Failed	
Meranoplus	*fenestratus*	Smith	Failed	
Meranoplus	*mjobergi*	Forel	Contaminated	
Meranoplus	*curvispina*	Forel	Pass	TBC
Meranoplus	sp. JDM 865		Failed	
Meranoplus	sp. JDM 1133		Failed	
Monomorium	*antipodum* (NW form)	Forel	NCBI	
Monomorium	*antipodum* complex sp. JDM 717		Contaminated	
Monomorium	*arenarium*	Heterick	Pass	TBC
Monomorium	*disetigerum*	Heterick	Failed; Old	
Monomorium	*euryodon*	Heterick	Pass	TBC
Monomorium	*fieldi*	Forel	NCBI	
Monomorium	*insolescens*	Wheeler	NCBI	
Monomorium	*laeve*	Mayr	NCBI	
Monomorium	*leae*	Forel	Pass	TBC
Monomorium	*punctulatum*	Heterick	Pass	TBC
Monomorium	*rubriceps* gp sp. JDM 1175		NCBI	
Monomorium	*sydneyense*	Forel	Pass	TBC
Monomorium	*sydneyense* complex sp. JDM 101		Pass	TBC
Nylanderia	*glabrior*	(Forel)	Pass	TBC
Ochetellus	*flavipes*	Kirby	Pass	TBC
Ochetellus	sp. JDM 527		Pass	TBC
Odontomachus	*ruficeps*	Smith	Pass	TBC
Opisthopsis	*haddoni rufoniger*	Forel	NCBI	

Appendix 1. List of species supplied for DNA barcoding, indicating which were successfully barcoded, which failed and species for which sequences were obtained from the NCBH public database. (Continuation)

Genus	Species/sub-species	Author	Sequence	GenBank Accession Number
Paraparatrechina	*minutula*	Forel	Pass	TBC
Paraparatrechina	*minutula* gp sp. JDM 916		Failed; Old	
Paratrechina	*longicornis*	(Latreille)	Pass	TBC
Pheidole	*mjobergi*	Forel	Pass	TBC
Pheidole	*rugosula*	Gregg	Pass	TBC
Pheidole	*turneri*	Forel	Pass	TBC
Pheidole	*variabilis*	Mayr	Pass	TBC
Pheidole	sp. JDM 684		Pass	TBC
Platythyrea	sp.	[male]	Pass	TBC
Polyrhachis	*ammonoeides*	Roger	Pass	TBC
Polyrhachis	*bohemia*	Kohout	Failed	
Polyrhachis	*gravis*	Clark	Failed; Old	
Polyrhachis	*inconspicua*	Emery	NCBI	
Polyrhachis	*lata*	Emery	NCBI	
Polyrhachis	*melaneura*	Kohout	Failed	
Polyrhachis	*seducta*	Kohout	Pass	TBC
Polyrhachis	(*Campomyrma*) sp. JDM 703		Failed; Old	
Polyrhachis	(*Campomyrma*) sp. JDM 1009		Pass	TBC
Polyrhachis	(*Campomyrma*) sp. JDM 1010		Pass	TBC
Polyrhachis	(*Chariomyrma*) sp. JDM 807		Failed	
Polyrhachis	(*Chariomyrma*) sp. JDM 808		Failed	
Probolomyrmex	*latalongus*	Shattuck, Gunawardene & Heterick, 2012	Contaminated	
Pseudoneoponera	*denticulata*	(Kirby)	Pass	TBC
Rhytidoponera	*convexa* complex sp. JDM 1129		Failed; Old	
Rhytidoponera	*crassinoda*	Forel	Pass	TBC
Rhytidoponera	*taurus*	Forel	Pass	TBC
Rhytidoponera	*tyloxys*	Brown & Douglas	Failed	
Solenopsis	*belisarius*	Forel	Failed	
Solenopsis	*clarki*	Crawley	Pass	TBC
Stigmacros	*punctatissima*	McAreavey	Failed	
Stigmacros	*termitoxena*	Wheeler	Failed; Old	
Strumigenys	*emmae*	Emery	Failed	
Tapinoma	*minutum broomense*	Forel	Pass	TBC
Tapinoma	sp. JDM 981		Failed	
Tetramorium	cf. *megalops*	Bolton	Pass	TBC
Tetramorium	*sjostedti*	Forel	Pass	TBC
Tetramorium	*spininode*	Bolton	Pass	TBC
Tetramorium	*striolatum* complex		Pass	TBC
Tetramorium	*striolatum* complex sp. JDM 36		Pass	TBC
Tetraponera	*punctulata*	Smith	Pass	TBC
Zasphinctus	*duchaussoyi*	André	Pass	TBC