Hidden Treasure: Bar-coding NZ native invertebrates to develop a reference database

Allan Wilson Centre for Molecular Ecology and Evolution

Summer Studentship 2014/2015 Report by Jacqueline Tizard

Introduction

Using Te Hauturu-o-Toi (Little Barrier Island) as a model ecosystem, the 'Hidden Treasure' project aims to use modern DNA sequencing in conjunction with traditional ecology to improve conservation management in New Zealand (New Zealand's next top model ecosystem, 2011).

The objective of this summer studentship project was to extract DNA from ~ 400 invertebrates, amplify the COI barcode region and sequence it and to then edit the data and conduct a preliminary analysis. These sequences could then be added to the DNA reference dataset of organisms found on Little Barrier Island. The hypothesis was that the COI barcode region can be used to distinguish between different morphotypes of Collembola.

We focussed on a class of organisms called Collembola, commonly known as springtails. Springtails are small (2-3mm) wingless hexapods. At least 380 species and subspecies of Collembola are found in New Zealand, and many more are still undescribed (Greenslade, Boyer, & Wratten, 2013). Despite being distributed worldwide, when compared to other hexapod groups Collembola are only moderately diverse in terms of species number (Cicconardi, Fanciulli, & Emerson, 2013). Recently however, molecular genetic studies have shown that springtail biodiversity may be greatly underestimated due to the high number of cryptic species (Emerson, Cicconardi, Fanciulli, & Shaw, 2011).

Following the quick development of molecular systematics, mitochondrial DNA and ribosomal DNA have become useful in the study of Collembolan relationships (Frati, Swofford, Sullivan, & Simon, 1997). The mitochondrial cytochrome c oxidase subunit I (COI) gene is most commonly used for phylogenetic studies of animals because it appears to be one of the most conservative protein-coding genes in the mitochondrial genome (Folmer, Black, Lutz, Hoeh, & Vrijenhoek, 1994).

Methods

DNA extraction

Collembola were collected from leaf litter samples from 10 plots from Hauturu (Drummond et al submitted). Morphotypes were classified through visual inspection by Mark Stevens, South Australian Museum. Each sample tube contained multiple specimens from the same morphotype but each DNA sample was taken from only one individual specimen, in general two per morphotype. DNA samples of Collembola already extracted were obtained from colleagues at Plant & Food Research.

Polymerase chain reaction and sequencing

Polymerase chain reaction (PCR) amplification was performed on DNA extracted from131 samples. To amplify the *COI* region of DNA, the primers LCO1490: 5'-ggtcaacaaatcataaagatattgg-3' and HC02198: 5'-taaacttcagggtgaccaaaaaatca-3' (Folmer, Black, Lutz, Hoeh, & Vrijenhoek, 1994) were

used for all samples. Typically 3 μL of the DNA extract was used as template for a 25 μL PCR reaction, using 5Units/uL *Taq* polymerase (Kapa2G) per reaction. Each 25 μL reaction consisted of 5 μL of buffer, 0.5 μL of each of the two primer stock solutions (10 μmol), 1 μL BSA (10 mg/mL), 2.5 μL of dNTP and 10.3 μL of sterile distilled water.

PCR cycling conditions were set at the following standard conditions: Initial denaturation at 95°C for 5 minutes, followed by 5 cycles of denaturation at 95°C for 30 seconds, annealing at 45°C for 30 seconds, extension at 72°C for 60 seconds. This was followed by another 30 cycles of denaturation at 95°C for 30 seconds, annealing at 48°C for 30 seconds, extension at 72°C for 60 seconds. Final extension was set at 72°C for 2 minutes and samples were then held at 4°C. Those samples which failed to produce a PCR product were amplified again at either a 1/10 dilution, a 1/100 dilution or at an increased DNA concentration of 7uL. Successful PCR products were visualised on a 1.5% agarose gel with SB buffer, using SYBR® Safe DNA gel stain under UV light. The Zymo research DNA Clean & Concentrator™ was used to purify resulting PCR products. PCR product was then Sanger sequenced by Marcogen in both directions using LCO1490 and HC02198 as sequencing primers.

Bioinformatics

Resulting sequences were then edited manually and aligned in Geneious R6 (Biomatters). From the raw sequence data contigs were assembled from the two strands. Contigs were individually analysed by eye to assemble consensus sequences. Consensus sequences were then aligned using Muscle. Indels were checked for and the sequences were manually trimmed by searching for the primer sequences. Collembola sequences were downloaded from the Bold database and aligned. Diplura was also used as an outgroup for comparison. A Phylogenetic tree was then built using the genetic distance model Tamura-Nei and the neighbor-joining tree building method.

Results and discussion

From a total of X individuals, 80 resulted in DNA extractions that successfully yielded amplified *COI* products (green samples in Appendix 1). To date however none of the 80 products have yielded any sequence data. This may have been due to ethanol contamination from the cleanup procedure. Ethanol is routinely used for DNA extraction even though it can inhibit PCR at concentrations of 1% or more (Demeke & Jenkins, 2009). Drying the DNA pellet prior to resuspension removes residual ethanol. While procedures were followed exactly to the instructions in the kit, perhaps there was not adequate drying time. When the same clean up kit was used on PCR product of mite DNA from a side project while awaiting sequencing results, those cleaned products also showed possible evidence of ethanol contamination when run on a gel. Another possible cause for the sequencing to fail is if the primers that were sent away with the product to be sequenced did not have adequate time to resuspend in solution. Due to time constraints we were unable to send away any more samples to be sequenced. However, the samples are likely to eventually be re-sequenced.

In absence of any of my own sequences to analyse, I was able to use sequences from *COI* Collembola DNA previously extracted and sent for sequenced by Leah Tooman and Laura Ward (see appendix 2). Forty four sequences were aligned and compared. Diplura was used as an outgroup out of interest despite the fact that we were not actually trying to infer an evolutionary order just a set of relationships (figure 1). Due to not knowing to which orders our specimens actually belonged to we cannot compare relationships amongst geographical distributions. However, by building a

phylogenetic tree using known COI sequences of species from all four Collembola orders; Entomobryomorpha, Poduromorpha, Symphypleona, Neelipleona I could make inferences about which order the samples belonged to and whether or not DNA analysis reflects the morphological groupings.

Traditional concepts of the phylogeny of Collembola are based on morphology and have been challenged by molecular phylogenetics. For example, Arthropleona was replaced by two orders Poduromorpha and Entomobryomorpha and Neelidae was separated from Symphypleona as Order Neelipleona (Xionga, Gaoa, Yina, & Luana, 2008).

Using the tree approximately 26 Operational Taxonomic Units (OTUs) were identified from the 44 sequences. In the phylogenetic tree some individual sequences from the same morphotype samples clustered together in the tree (samples 026-195, 236-1, 026-201 and 239-137) which may substantiate using morpotypes to group specimens In other instances however, sequences that had been identified as being from the same morphotype did not cluster together (samples 025-100, 014-233, 26-297 and 026-200). This could be due to incorrect sorting of individual Collembola into morphotype groups or it may indicate the presence of cryptic species. Many Collembolan species that have been morphologically defined have been found to have highly divergent DNA sequence diversity (Timmermans, Ellers, & Marien, 2006) and this may be the case for some of the morphotypes that were grouped together in this study. The Collembola barcoding campaign by Porco et al., (2012) illustrated a high degree of cryptic diversity within common species that had previously been considered single taxa when classification had been based on morpholgy alone.

It should be noted that it is tenuous to base a phylogeny on only one section of DNA (the COI barcoding region). It would of course be preferable to have whole genomes to compare but this is not feasible or practical. Barcoding regions of DNA allows us to quickly and reasonably accurately identify species. This dataset represent the initiation of a reference set of Collembola COI sequences for use to type New Zealand Collembola within subsequent metabarcoding projects.

Conclusion

The results indicate that using morphology to group Collembola has its merits but is limited. Many species that have the same or very similar morphology are actually genetically very different and reproductively isolated. Traditional taxonomic approaches are therefore unable to detect a significant level of Collembolan biodiversity. DNA barcoding discloses high level of cryptic diversity and therefore it offers an additional contribution to the study of Collembolan taxonomy and taxonomy in general.

Acknowledgements

I would like to thank Dr Richard Newcomb for the opportunity to complete this project and for his invaluable expertise and knowledge he shared with me. Laura Ward was hugely helpful, providing my training and much assistance both in and out of the lab as were all other members of the lab group. I would also like to thank the Allan Wilson Centre for funding this summer scholarship.

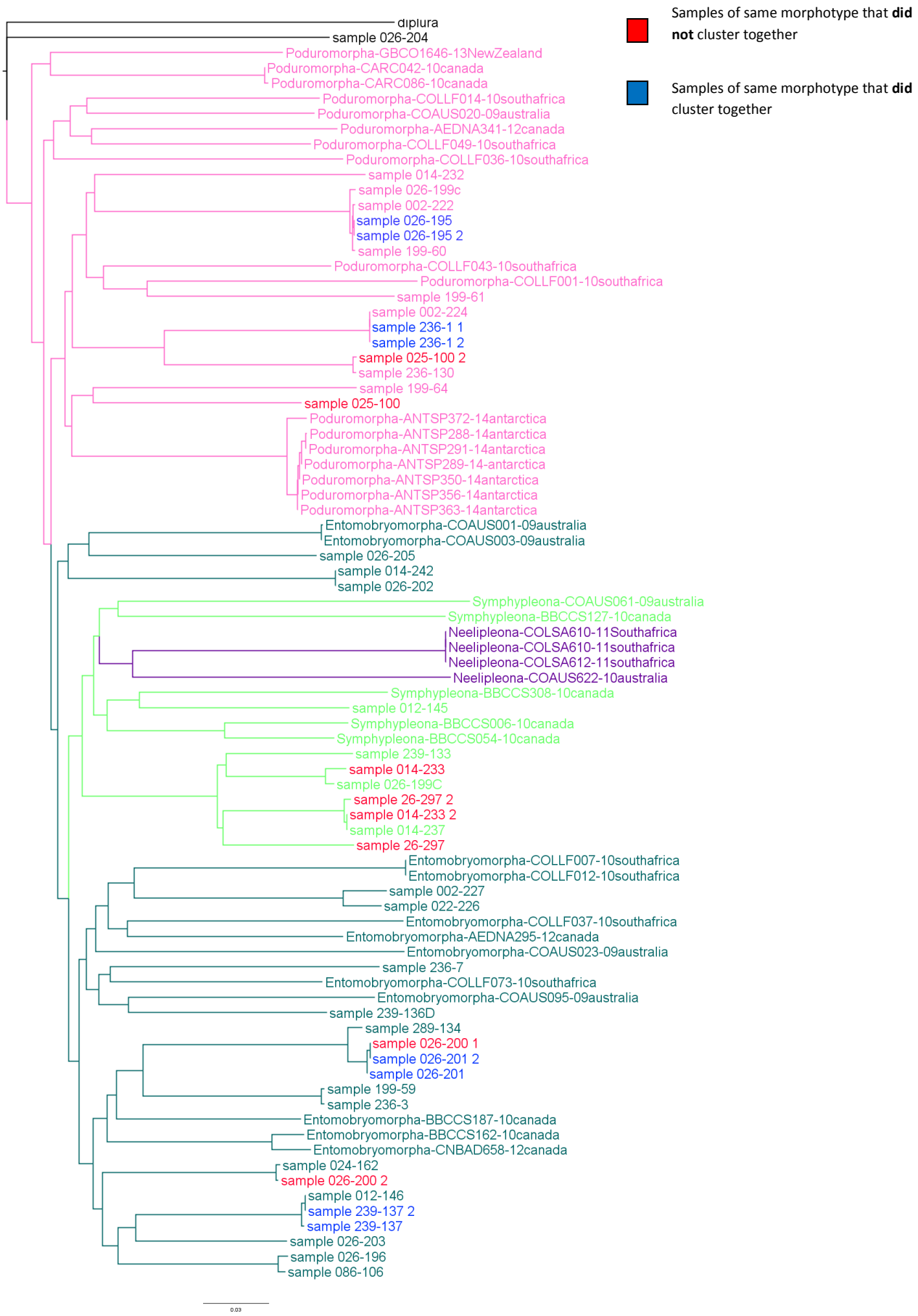

Figure 1: Tamura-Nei Neighbour-Joining Tree for region COI generated in Geneious R7 (Biomatters).

References

Cicconardi, F., Fanciulli, P. P., & Emerson, B. C. (2013). Collembola, the biological species concept and the. *Molecular ecology , 22* (1), 5382-5396.

Demeke, T., & Jenkins, G. R. (2009). Influence of DNA extraction methods, PCR inhibitors and quantification methods on real-time PCR assay of biotechnology-derived traits. *Analytical and Bioanalytical Chemistry* , doi: 10.1007/s00216-009-3150-9.

Emerson, B. C., Cicconardi, F., Fanciulli, P. P., & Shaw, P. J. (2011). Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. *Philosophical transactions B , 366*, 2391-2402.

Folmer, O., Black, M., Lutz, R., Hoeh, W., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* , 294-299.

Frati, F., Swofford, D. L., Sullivan, J., & Simon, C. (1997). Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. *J. Mol. Evol.* (44), 143-158.

Geneious version R6 created by Biomatters. Available from http://www.geneious.com/

Greenslade, P., Boyer, S., & Wratten, S. (2013). New records of springtails in New Zealand pasture: how well are our pastoral invertebrates known? *New Zealand Journal of Agricultural Research , 56* (2), 93-101.

New Zealand's next top model... ecosystem. (2011). *Pheno*, (1), 4-5.

Porco, D., Beods, A., Greenslade, P., Janion, C., Skarzynski, D., Stevens, M., et al. (2012). Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding - See more at: http://www.publish.csiro.au.ezproxy.auckland.ac.nz/view/journals/dsp_journal_fulltext.cfm?nid=120&f=IS12026#sthash.78h22Txb. *Invertebrate Systematics , 26* (6), 470-477.

Timmermans, M., Ellers, J., & Marien, J. (2006). Genetic structure in Orchesella cincta (Collembola): strong subdivision of european populations inferred from mtDNA and AFLP markers. *Molecular ecology , 14*, 2017-2024.

Xionga, Y., Gaoa, Y., Yina, W.-y., & Luana, Y.-x. (2008). Molecular phylogeny of Collembola inferred from ribosomal RNA genes. *Molecular Phylogenetics and Evolution , 49*, 728–735.

Appendix 1

25-Nov	26-Nov	27-Nov	1-Dec	2-Dec	3-Dec
021-118	11E	4H	7C	011-190	G9 0224-165
013-34	10D	7E	7G	020-213	C10 026-194
016-257	021-116	8G	9A	G2 024-170	H10 024-174
987-181	10A	007-55	10F	E4 014-235	A1 026-194B
202-32	11C	987-182	10H	B5 002-228	E1 236-8
018-171	11D	4G	11A	E2 024-170	H1 199-70
240-45	9C	8H	12D	H2 026-198	B3 021-144
023-155	021-117	11G	12G	B3 086-104	C3 026-199
003-140	021-113	5F	226-219	C5 014-241	D3 026-199
240-47	241-93	201-210	987-178	B4 199-63	H3 026-199D
013-38	020-212	8A	019-10	F5 026-195C	C5 086-103
008-148	003-141	011-191	016-254D	G4 199-69	B6 086-105
007-58	987-185	10C	226-220	D5 239-136B	F6 236-09
016-259	022-41	9E	021-119	G5 024-169	A7 025-99
007-49	9D	7B	226-122	B6 014-236	B7 025-99
011-189	12B	9G	027-107	D6 236-4	C7 025-99
087-31	12A	12C	007-52	A1 236-1A	F7 025-97
007-54	12E	12H	954-243	C1 025-96	C8 024-167
022-44	10B	9B	016-256	C6 026-194D	D8 024-167
987-180	11B	240-48	230-159	H1 236-6	A11 239-131
8B	022-43	987-187	988-14	A2 086-102	D11 02-229
240-46	10G	8D	087-32	B2 239-132	/

RE-RUNS

5-Dec	8-Dec	9-Dec	11-Dec	15-Dec	16-Dec
8D	022-44	021-116	021-116	022-44	G2 024-170
9G	11G	8D	8D	9G	B5 002-228
11G	4H	007-55	007-55	016-254D	E2 024-170
4H	987-182	013-38	013-38	007-55	C5 014-241
987-182	9E	021-113	021-113	013-38	B4 199-63
9E	008-148	021-119	021-119	12B	F5 026-195C
007-55	10D	023-155	023-155	12C	D5 239-136B
011-191	027-107	240-45	240-45	027-107	G5 024-169
016-257	011-191	011-189	011-189	9D	B6 014-236
241-93	016-257	12H	12H	9E	D6 236-4
021-117	9G	987-180	987-180	10B	A1 236-1A
013-34	021-117	8B	8B	11B	C1 025-96
12B	013-34	11B	11B	016-257	H1 236-6
12C	016-254D	10G	10G	8B	A2 086-102
11B	12B	10B	10B	9C	G9 0224-165
10G	12C	9C	9C	10G	H10 024-174
10B	11B	011-191	011-191	008-148	A1 026-194B
9C	10G	016-257	016-257	011-189	H1 199-70
9D	10B	11G	11G	987-180	B3 021-144
12H	9C	4H	4H	10D	C3 026-199
9B	9D	008-148	008-148	013-34	D3 026-199
/	/	10D	10D	/	H3 026-199D
1/10 dillution	1/100 dillution	more DNA	FRESH STUFF		retry plates

DID WORK
NO WORK
CLEANED
RE-TRIED
FAILED :(

Appendix 2

sent 10/10/13													
	1	2	3	4	5	6	7	8	9	10	11	12	
A	026-194B	199-60	021-144	289-134	025-100	086-105	012-146	024-167	086-106	002-224	239-136D	002-222	
B	199-61	199-60	021-144	289-134	025-100	086-105	012-146	026-194	236-7	236-130	014-233	26-197	
C	199-61	239-137	026-199	026-195	086-103	026-201	026-202	025-95	026-204	014-242	014-233		
D	236-5	239-137	026-199	026-195	086-103	026-201	236-1	012-145	026-204	002-230	0224-165		
E	236-8	014-232	026-200	024-168	026-205	024-172	236-1	026-203	014-238	026-199C	239-136C		
F	236-8	014-232	026-200	024-168	026-205	236-09	239-131	002-227	025-95	024-162	239-139		
G	199-70	02-231	026-199D	014-237	199-64	022-226	239-132	239-133	025-94	014-235	239-139		
H	199-70	199-68	026-199D	014-237	199-64	022-226	086-104	026-196	086-101	239-136	026-198		
	TUBE:	Extraction:	SAMPLE:										
	1-5H	Plate 1	199-64										
	1-7D	Plate 1	236-1										
	1-8B	Plate 1	026-194										
	1-8G	Plate 1	239-133										
	1-9B	Plate 1	236-7										
	1-9D	Plate 1	026-204										
	1-9F	Plate 1	025-95										
	1-10C	Plate 1	014-242										
	1-10E	Plate 1	026-199C										

	1-11H	Plate 1	026-198										
	96	TUBE	236-2										
	101	TUBE	199-59										
	102	TUBE	199-59										
	103	TUBE	236-3										
	104	TUBE	236-3										
	105	TUBE	024-164										
	106	TUBE	024-166										