Identifying plant DNA in the faeces of a generalist insect pest to inform trap cropping strategy

Jerry Asalma Nboyine, Stéphane Boyer, David J Saville, Stephen David Wratten

Identifying plant DNA in the faeces of a generalist insect pest to inform trap cropping strategy

Nboyine Jerry Asalma [1, 2], Boyer Stephane [3*], Saville David J [4], Wratten Stephen David [1]

[1] Bio-Protection Research Centre, P. O. Box 85084, Lincoln University, Christchurch 7647, New Zealand

[2] CSIR- Savanna Agricultural Research Institute, P. O. Box 52, Tamale. Ghana

[3] Institut de Recherche sur la Biologie de l'Insecte (IRBI) – UMR 7261 CNRS / Université de Tours, Parc Grandmont, 37200 Tours, France

[4] Saville Statistical Consulting Limited, P. O. Box 69192, Lincoln 7640, New Zealand

*Corresponding author's email: stephane.boyer@univ-tours.fr

Phone no.: +33 (0)2 47 36 74 55

Abstract

Monocropping elevates many insects to the status of economic pests. In these agroecosystems, non-crop habitats are sometimes deployed as trap crops to reduce pest damage. This environmentally-friendly alternative to pesticides can be particularly fitting when dealing with native invaders that may be afforded legal protection or enjoy public sympathy as is the case for the ground wētā *Hemiandrus* sp. 'promontorius' (Orthoptera) in New Zealand. However, this approach requires knowledge of the insects' diet to select the most appropriate plant species for trap cropping.

Here, ingested plant DNA in the faeces of wētā, was analysed to help develop strategies for mitigating its damage in New Zealand vineyards. DNA was extracted from faeces of wētā collected from six different vineyards over four seasons. Using a DNA metabarcoding approach, we amplified the *rbc*L gene region and sequenced the amplicons on an Illumina MiSeq platform. The identity of plants in the diet of this insect was determined by comparing the sequences generated with those available in the GenBank database and crosschecking the results with a database of plants known to be present in New Zealand.

A total of 47 plant families and 79 genera were detected. Of the genera identified, *Vitis, Poa, Festuca*, *Anthoxanthum*, *Anagallis, Camelina, Epilobium, Menyanthes, Pedicularis, Urtica, Garrya, Pinus* and *Tilia* were the major ones (i.e., they were present in more than 50% of the faecal samples). The composition of the above plant taxa in faecal materials was significantly different between collection sites or dates, except for *Menyanthes*. The occurrence of the latter was significantly different between collection sites. These results

indicate that effectively mitigating wētā damage to vines requires the use of a diverse mix of plant species for trap cropping as wētā seem to be highly generalist in their feeding behaviour even when plant diversity is relatively low.

Key words: Wētā, Diet analysis, DNA metabarcoding, faeces, pest management, *rbc*L, vineyards, New Zealand

1. Introduction

Agricultural intensification has led to monocultures of high yielding plant species/cultivars over vast areas of land. This provides abundant resources for insects feeding on those monocultural species, elevating them to the status of economic pests (Rusch et al. 2016). To reduce pest damage while maintaining a monocultural state, high amounts of inputs are often applied, especially prophylactic use of insecticides and herbicides. These practices have led to major biodiversity losses and unwanted adverse effects on arable land and the surrounding environment (Rockström et al. 2009). Although the risks to human health and the environment from these chemicals have resulted in some cases in shifts to more sustainable non-pesticide pest management practices (Ekström and Ekbom 2011), most food production worldwide still relies heavily on high-input practices. Alternative strategies, although still under-deployed, have as a key component the enhancement of functional farmland plant diversity (Rusch et al. 2016). This is because areas of non-crop habitats in farmland can influence pest populations by harbouring pests' natural enemies (Gurr et al. 2016). Non-crop vegetation in or around farmland may also attract, divert or intercept the targeted insect pest(s) and reduce their damage to the main crop. These latter processes include trap cropping as well as supplemental management strategies such as trap vacuuming, trap harvesting, sticky traps and pesticide application to trap crops (Moreau and Isman 2012). These pest management principles have been used worldwide in a variety of cropping systems including viticulture (e.g. Villanueva-Rey et al. 2014). For instance, although vineyards are almost monocultures, it is common for at least one grass species to cover the inter-row areas. In addition, strips of flowering plants (e.g., buckwheat, *Fagopyrum esculentum* Moench.) are sometimes sown under vines or in the inter-rows to enhance populations and fitness of natural enemies for managing important vine insect pests such as larvae of the leafroller complex (*Epiphyas postvittana, Ctenopseustis* spp., *Planotortrix* spp., etc.), leafhoppers (*Erythroneura* spp.) and other phytophagous insects (Shields et al. 2016). Inter-row vegetation and any surviving weeds could also act as alternative food sources for generalist insect pests, thereby potentially reducing economic damage.

However, successfully using this approach to manage pests, hinges on identifying and deploying appropriate non-crop species (Gurr et al. 2016). Hence, deployment of a less suitable non-crop vegetation will not result in reduced pest damage to the main crop (Villa et al. 2016).

Generally, identification of candidate trap-crop species may involve the time-consuming observation of the insect's feeding behaviour, or alternatively, analysing its gut content or faeces for the most abundant plant species (Pompanon et al. 2012). Several classical methods of gut content or faecal analysis are available (e.g., microhistological analysis, near infra-red reflectance spectroscopy, stable isotopes etc.), but they often lack taxonomic resolution (Soininen et al. 2009). More recently, advances in DNA barcoding, combined with high-throughput DNA sequencing, allow for the identification and characterisation of the composition of an animal's diet with much higher precision (Soininen et al. 2009; Alberdi et al. 2018; Pompanon et al. 2012; Boyer et al. 2013).

Insects emerge as pest when they are introduced to a new habitat, just as the introduction of new crop plants can also lead to novel associations where native species become pests (Lefort et al. 2015). This is the case for one species of wētā which is native to New Zealand but has become a pest in vineyards (Nboyine et al. 2016). Wētā *Hemiandrus* sp. 'promontorius' are present in New Zealand's vineyards throughout the year and causes significant damage to vines at the specific period of budburst when they feed on the very young leaves (Nboyine et al. 2017).

Wētā are a well-known and iconic group of New Zealand insects comprising about 70 species in the families Anostostomatidae and Rhaphidophoridae. Their name is derived from that of Wētāpunga, the god of ugly things in the Māori mythology. As such, these insects are considered *Taonga* (i.e. treasure) and must be protected. All wētā are endemic to New Zealand and many of them are at risk of extinction because of the degradation of their natural habitat and the introduction of mammalian predators (in particular rats, mice and stoats). Because many species are threatened, wētā have become useful indicators of environmental health and the focus of numerous conservation initiatives. They are also ideal candidates for citizen science and science outreach projects because they are easier to work with than most other insects, they can be relatively large and are very appealing to the public. As a consequence, wētā are one of the only insect group that is well recognised and highly valued by the general public.

Due to their endemic status, their significance in the Maori culture and the sympathy they generate from the public, it is not conceivable to control wētā populations with insecticides and alternative methods .

Figure 1: Photograph of a female ground wētā *Hemiandrus* sp. 'promontorius' with egg clutch after excavation of her burrow.

The current work therefore aimed at analysing ingested plant DNA in the faeces of a generalist orthopteran pest, a ground wētā (*Hemiandrus* sp. '*promontorius'*: Orthoptera, Anostostomatidae), in New Zealand vineyards to help identify candidate plant families/ genera for inclusion in its management strategy, for example, as potential trap plants. Using generic PCR primers, we anticipate to detect a range of plant taxa eaten by wētā and to obtain a good coverage of the insect's plant-based diet. According to the existing literature (e.g. Johns et al. 2001), we hypothesise that wētā feed on grass and vines, but also on a number of plant species that may be less common in New Zealand vineyards. Seasonal variation in diet is expected as some of the targeted plants may only be available at certain periods of the year.

2. Materials and methods

2.1 Wētā collection sites

Six vineyard blocks located in three different vineyard locations were sampled in the Awatere Valley, Marlborough, New Zealand at elevations ranging from 8 to 46 m a.s.l.: O- Block (Castle Cliffs, -41.6103 ^{O}S, 174.1276 ^{O}E) was 4.61 ha ; D- Block (Castle Cliffs, -41.6075 ^{O}S, 174.1328 ^{O}E) was 37.88, ha **;** H- Block (Castle Cliffs, -41.6131 ^{O}S, 174.1359 ^{O}E) was 2.98 ha ; L- Block (The Favourite**,** -41.6198 ^{O}S, 174.1071 ^{O}E) was 16.88, ha ; N- Block (The Favourite**,** -41.6260 ^{O}S, 174.1105 ^{O}E) was 44.41 ha and CR- Block (Caseys Road, -41.6880 ^{O}S, 174.120 ^{O}E) was 11.98, ha. These vineyards were subjected to conventional management practices, with weeds, insect pests and diseases being controlled with pesticides. The inter-rows were densely sown with grass mixtures dominated by *Lolium perenne* L., *Festuca arundinacea* Schreb. and *Poa pratensis* L., while under-vine areas sometimes harboured a few sparsely growing dicotyledonous weeds and grasses. In spring, under-vine areas were sprayed with conventional herbicides to remove weeds and maintain the soil bare. Pine tree (*Pinus* L. spp.) hedges bounded at least one side of each sampled block.

2.2 Sampling wētā from vineyards for faecal analysis

Sampling was performed during the day while wētā are generally buried in individual galleries (Fig 1). On random locations in the vineyard, the upper layer of soil was swiftly removed with a movement of the shovel to expose galleries inhabited by wētā and draw the insects out. Wētā mid-instar larvae were hand-collected as they came out of their galleries. Each of the six vineyard blocks were sampled over four seasons, namely in July 2014, October 2014, January 2015 and April 2015. Sampling on one vineyard block took about 2 hours and all blocks were sampled within 4 days at each season to avoid any difference in food availability due to plant phenology. In each season, 60 individual insects (i.e., 10 from each of the six vineyard blocks) were collected and placed singly in a labelled plastic arena (9 cm height × 15 cm width × 15 cm length) lined with a double layer of tissue paper. The arenas were stored at room temperature (20 ^{o}C) for 24 h, after which the insects were released. Individual wētā mostly produced one faecal pellet which was stuck to the tissue paper. Each pellet was carefully transferred into a labelled 60 mm diameter Petri dish (excluding the tissue) and stored at -80 ^{o}C pending DNA extraction.

2.3 DNA extraction

DNA was extracted from 72 faecal samples (i.e., three randomly selected pellets per site per season) using a Zymo Research Fecal DNA MicroPrep^{TM} kit. The manufacturer's protocol was followed with slight modifications. To extract DNA from wētā faeces, 500 µl lysis solutions were pipetted into 72 individual BashingBead^{TM} lysis tubes each containing faecal material. The DNA from the faecal material produced by an individual wētā was extracted individually as its weight was less than the 150 mg recommended by the

manufacturer. The tubes were secured in a bead beater and processed at 50 oscillations per second for 5 minutes, followed by centrifuging at 10,000 g for 1 minutes. The supernatants (400 µl) were transferred to Zymo-Spin™ IV spin filters in collection tubes and centrifuged at 7,000 g for 1 minute. Faecal DNA binding buffer (1,200 µl) was then added to the filtrates after which the resulting mixtures were transferred to Zymo-Spin ™ IC columns in collection tubes and centrifuged at 10,000 g for 1 minute. This was followed by the addition of 200 µl DNA pre-wash buffer and 500 µl faecal DNA wash buffer to the columns and centrifuging for 1 minute at 10,000 g after adding each reagent. The columns were transferred into clean 1.5 ml microcentrifuge tubes and 30 µl of DNA elution buffer were added directly to each column matrix. The tubes were centrifuged for 30 seconds at 10,000 g to elute the DNA. The latter was transferred into Zymo-Spin™ IV-µHRC spin filters in clean 1.5 ml microcentrifuge tubes and left for 30 minutes before centrifuging at 8,000 g for 1 minute for purification. The purified DNA was then amplified through polymerase chain reaction (PCR).

2.4 PCR and electrophoresis

The universal primer pair [rbcL19 and rbcLZ1 (Poinar et al. 1998)], which amplifies a ≥ 150 base pairs (bp) fragment of the ribulose bisphosphate carboxylase large subunit (rbcL) chloroplast DNA gene region was used to detect ingested plant DNA in wētā faeces. Primers were designed to include the recommended overhang adapters for Illumina sequencing. The PCR amplification was performed in 40 µl reaction mixtures containing 6 µl DNA extract, 6.8 µl water, 20 µl GoTaq® Green 2×, 2 µl bovine serum albumin (BSA, 10 mg/ml), 2 µl MgCl2 (25mM,) and 1.6 µl each of the forward and reverse primers (10 µM). The protocol for the thermocycling was: 94 °C for 5 min, 45 cycles of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 30 s, and a final elongation at 72 °C for 10 min. Positive (*Lolium perenne* DNA) and negative (PCR grade water and wētā DNA extract) controls were included in each PCR to check for the success of amplification and DNA contamination, respectively. All PCR products underwent gel electrophoresis to check for successful amplification. Products of expected fragment size were cleaned with an Agencourt® AMPure® XP PCR purification kit following the manufacturer's instructions and standardized at 2ng/µL. This procedure was also applied to the wētā DNA negative control. Unique molecular identifiers (MID) were added to each of the 72 samples as well as the negative control before high-throughput DNA sequencing on one run of Illumina MiSeq using the 300 × 300 paired end protocol as recommended by the manufacturer (https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html). A 600 cycle kit was used to sequence the amplicons on the MiSeq instrument. Read 1 was sequenced to 320 base pairs, and

Read 2 sequenced to 280 base pairs. Identifier ligation and Illumina sequencing were performed by New Zealand Genomics Ltd, Auckland, New Zealand.

2.5 Data analysis

Paired end reads were merged using the software VSEARCH version 1.9.5. For quality control reads were truncated at the first low-quality base (i.e. quality score < 3) if present to ensure high-quality tails and accurate merging of the paired end reads. Merged sequences from the Miseq run that were shorter than 150 bp were discarded because the expected length was 210 bp (~150 bp for the internal amplicon plus 30 bp for each primer). At this stage, we discarded any sequence with more than one expected error in the sequence as well as singletons (i.e., Operational Taxonomic Units (OTUs) represented by a single read). To make the downstream analysis faster, non-unique sequences were then collapsed with a one base mismatch allowance. These unique sequences were clustered into Molecular Operational Taxonomic Units (MOTUs) using a conservative 97% identity threshold. Chimeric sequences were then removed using the UCHIME *de novo* method. To determine the identity of plant taxa in the diet of wētā, each MOTU had its representative sequence searched against the GenBank nucleotide database using BLASTN version 3.2.31. Identifications accepted as correct matches and used for subsequent analyses in this study were those where BLAST search returned values of query coverage > 80% (i.e. identification based on at least 120 base pairs out of 150), and identity > 90% (i.e. identification at genus level). Because rbcL is not perfectly resolutive at the species level in plants and the DNA fragment used was very short, the risk of obtaining assignations that matched several different taxa with the same score was higher than that commonly encountered in DNA barcoding studies. To minimise the risk of multiple assignations, we conducted a barcoding gap analysis using the *localMinima* function in the R package SPIDER (Brown et al. 2012) to determine the appropriate species identity threshold based on our own data. This analysis found the species identification threshold for our dataset to be 1.8% (Fig. 2). The accepted identifications were further cross checked with a database of plants present in New Zealand (Allan Herbarium 2000). Sequences with no match in BLAST or with a match not recorded in the database of plants present in New Zealand were removed from the dataset and not used in subsequent analyses. See decision map in Fig. 2 for details.

Data were analysed as frequency of occurrence (FOO) and relative read abundance (RRA) (Deagle et al. 2019). To calculate FOO, the data were converted into presence (1)/ absence (0) before performing statistical analyses. To limit the potential inclusion of contaminants, a filtering step was also performed, in which 'presence' was assigned to MOTUs that occurred at least four times (i.e., 4 reads) in one faecal sample, while 'absence' was assigned to those that were detected less than four times and only present in one faecal sample.

193 Because our main interest was in the detection of food items that could potentially be used in a trap cropping
194 strategy. Statistical analyses of FOO focused on major food items, which were defined as those genera which
195 were detected in more than 50% of the faecal samples. These major taxa were *Vitis* sp. (vines), *Poa* spp. (grass).,
196 *Festuca* spp. (grass), *Anthoxanthum* spp. (grass), *Anagallis* spp. (weed), *Camelina* spp. (weed), *Epilobium* spp.
197 (weed), *Menyanthes* spp. (weed), *Pedicularis* spp. (weed), *Urtica* spp. (weed), *Garrya* spp. (tree), *Pinus* spp.
198 (tree), and *Tilia* spp. (tree).
199 Food items were categorized in two groups: 'Cultivated' plants, when grown for economic reasons (vines) or to
200 provide other beneficial services such as erosion control (grasses), and 'Uncultivated' plants, which were weeds
201 and trees growing inside or outside the vineyards, respectively.
202 Accumulation curves were built based on the cumulative number of plant families and genera detected in
203 relation to the number of samples analysed using a bootstrap method to estimate diet coverage. Generalised
204 linear models were used to determine the effect of sites and dates of sampling on the detection of each of the
205 eight major taxa. The binomial distribution (with a binomial total of three faecal samples for each sampling unit)
206 and logit link function were chosen for these analyses. The response variables were the diet (i.e. the eight major
207 plant items), while the fitted model comprised date and site. Main effect means for either date or site that were
208 significantly different were separated using least significant differences (LSD). Significant differences between
209 the proportions of groups, subgroups and genera of plants were determined by computing the 95% confidence
210 intervals (C. I.) of their mean. Relative read abundance was compared between sites and season using an
211 analysis of deviance on a multivariate generalised linear model. A negative binomial distribution was chosen for
212 this analysis based on the dispersion of the residuals. A 5% probability level was used for all tests.
213

214 **3. Results and discussion**

215 A total of 7,413,745 quality reads were obtained, of which 7,026,022 reads provided quality matches with 540
216 MOTUs from GenBank. Applying a species identification threshold of 1.8% resulted in 40 MOTUs identified at
217 species level and 500 that could only be identified at genus level. Of the latter, 366 MOTUs (i.e., 5,024,502
218 reads) corresponded to genera known to be present in New Zealand (Fig. 2). Overall, 80% of plant taxa present
219 in the faecal materials could be confidently identified at the genus level, while only 3% could be identified at
220 species level. The *rbc*L gene was therefore resolutive enough for genus-level identification of plants, making it
221 particularly suitable for studies focusing on ecosystems with moderate levels of plant biodiversity such as
222 intensive agricultural landscapes. Based on these results, data was analysed only at the genus and family level

and the following analyses were based on a total of 401 MOTUs represented by 6,129,201 reads (green boxes on Figure 2).

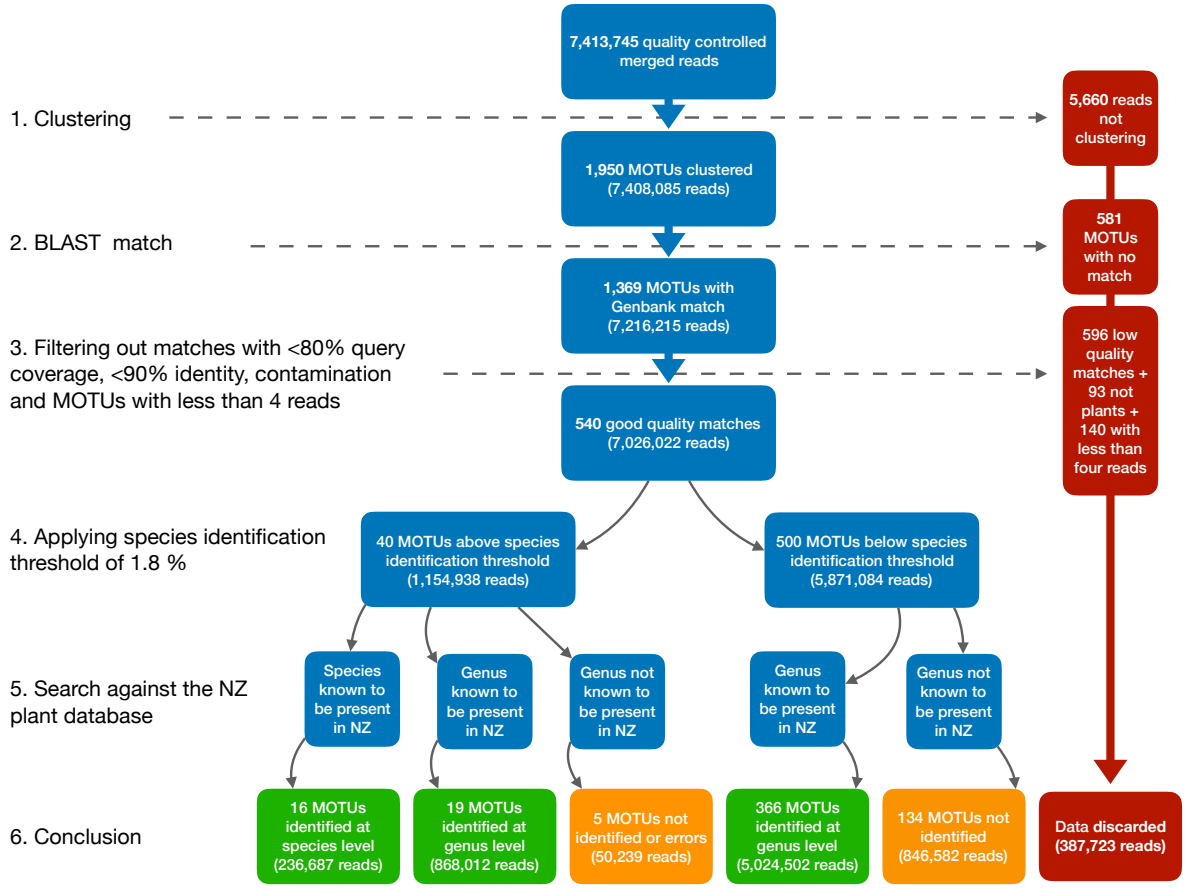

Figure 2: Decision tree and number of reads and MOTUs retained or discarded at each step of the bioinformatics analysis.

In the current study, the *rbc*L gene region was targeted using general primers. A major advantage of using a general primer set, is that a priori knowledge of the range of potentially consumed species (i.e., taxonomic coverage) by herbivores is not required (Pompanon et al. 2012). In such single locus studies, the P6 Loop of the *trn*L intron or the *rbc*L region are usually recommended because these regions are easily amplified and are well-conserved in land plants, thus allowing to achieve a high taxonomic resolution when using a metabarcoding approach (Alberdi et al. 2018; Pompanon et al. 2012). In addition, these regions are relatively short (ranging between 12 – 134 bp and 150 bp respectively), which makes them more likely to be amplified from degraded DNA samples such as faeces and gut contents (Pompanon et al. 2012). In the New Zealand context, rbcL sequences are available for the great majority of native and naturalised plant genera (Lear et al. 2018), which makes it the better candidate. A recent study by McClenaghan et al. (2015) successfully described the diet of

239 different species of grasshoppers (Orthoptera: Acrididae) in Ontario by using the same primers and identifying
240 plants at the family level and to a lesser extent, at genus and species levels.
241 The identified taxa belonged to 47 plant families and 79 genera. According to a bootstrapped estimate using
242 good quality sequences from 72 faecal samples, an estimated 93.7% of all plant genera and 96.4% of all plant
243 families likely present in the diet of the wētā were successfully detected. Our analysis was therefore sufficient to
244 determine the overall diet of that species at these taxonomic levels. Of the families detected, Poaceae comprised
245 12 genera, while the families Amaranthaceae, Asteraceae, Podocarpaceae and Rosaceae recorded four genera
246 each. Except Lamiaceae, the remaining 41 families displayed at most two genera (Table 1). Overall (i.e.,
247 irrespective of sampling site or season) the genera *Vitis, Poa*, *Festuca*, *Anthoxanthum*, and *Tilia* were more
248 frequently detected than any other taxon (Fig. 3A).
249 The mean detection rate of cultivated plants (grasses and vines) ($P < 0.05$; C. I. = 0.98 – 1) were significantly
250 higher than that of uncultivated plants (weeds and trees) ($P < 0.05$; C. I. = 0.67 – 0.0.75). Pairwise comparisons
251 of the mean proportional detections of the different categories of plants showed that, vines and grass (*Poa* sp.,
252 *Festuca* spp., *Anthoxanthum* spp.) occurred more often than trees (*Pinus* spp., *Tilia* spp, *Garrya* spp., etc.)
253 which were detected more often than dicotyledonous weeds (*Anagallis* spp., *Camelina* spp., *Epilobium* spp.,
254 *Menyanthes* spp., *Pedicularis* spp., *Urtica* spp. etc.) (Fig. 3C). It is important to note that weeds were rare in the
255 vineyards studied (J. Nboyine, pers. Observ.). However, every wētā collected had eaten at least one of these
256 weed species in spite of the unlimited availability of grasses and vines. A similar pattern was observes for trees
257 (*Pinus* spp., *Tilia* spp, *Garrya* spp., etc.), which were also represented in every faecal sample. The high diversity
258 of plant families and genera identified from the faecal samples, confirmed the status of this wētā as a generalist
259 feeder. Species in the genus *Hemiandrus* are usually omnivores, feeding on a diverse range of plants including
260 green leaves of trees and shrubs, but also on other invertebrates (Wyngaarden 1995; Johns 2001). Diets
261 comprising a mixture of plant and/or animal species is a common feeding behaviour among generalist
262 orthopterans and other omnivore arthropods (Coll and Guershon 2002). This gives such insects a better nutrient
263 balance than is possible by feeding on a single plant taxon, resulting in increased growth and survival (Coll and
264 Guershon 2002; Berner et al. 2005). In addition, toxic secondary metabolites produced as defence mechanisms
265 against herbivory by some plant species are diluted in mixed diets, reducing their effect on the insect (Ali and
266 Agrawal 2012).
267

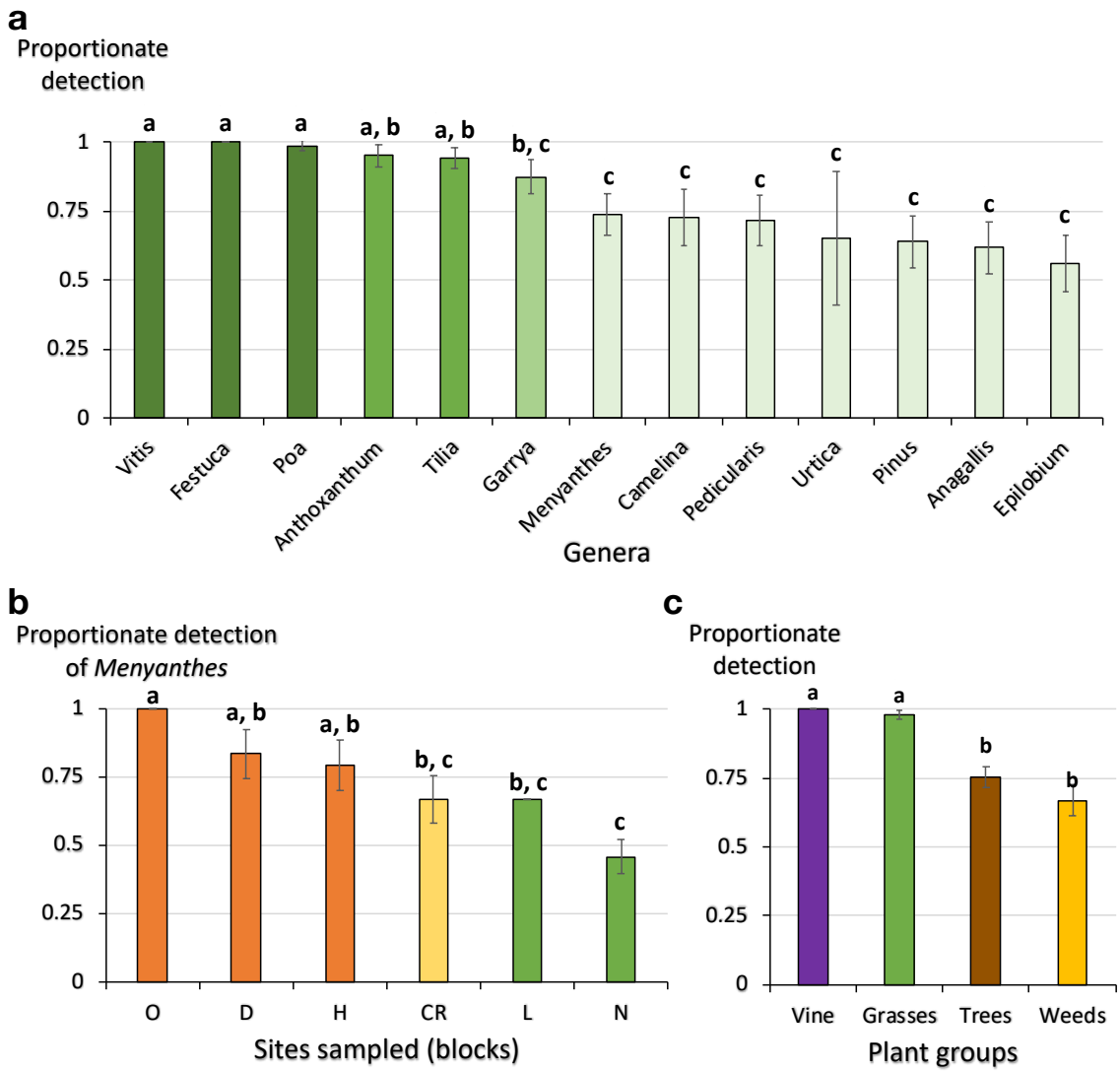

Figure 3: Detection of plant DNA through molecular analysis of frass from wētā (*Hemiandrus* sp. 'promontorius'). A) Proportion of the major plant genera detected. Data are means ± 95% confidence intervals. Bars with no letters in common are significantly different at the 5% probability threshold. B) Proportion of wētā frass tested positive for *Menyanthes* spp. at different sites sampled. Data are means ± standard error of means (S. E.). Bars with no letters in common are significantly different at the 5% level of significance. C) Proportion of the different plant categories detected in wētā frass. Data are means ± 95% confidence intervals. Bars with no letters in common are significantly different at the 5% probability threshold. Trees: *Pinus* spp., *Tilia* spp, *and Garrya* spp.; weeds: *Anagallis* spp., *Camelina* spp., *Epilobium* spp., *Menyanthes* spp., *Pedicularis* spp., and *Urtica* spp.; grasses: *Poa* spp., *Festuca* sp. and *Anthoxanthum* sp.; vines: *Vitis* sp.

With regards to the relative abundance of reads (RRA), 45% of all reads belonged to Poaceae and 41% to vines leaving only 14% of reads to the 31 remaining families (Fig 4). These percentages reflect the composition of

281 plants communities generally observed in the studied vineyards. The inter-rows of the vineyards studied were
282 dominated by grasses, which are low in protein content (below 50% of DM) and high in carbohydrates. As the
283 grasses mature, protein content declines to less than 10% while carbohydrate increases (Lledó et al. 2015).
284 Proteins are a major requirement of the diet of *Hemiandrus* spp. (Johns 2001). Being an omnivore, this insect
285 could balance its protein intake by preying on other insects. The latter were however, killed by the regular
286 applications of insecticides in the vineyard. Therefore, sustainable intake of protein for this wētā appeared to
287 rely on balanced feeding on weeds and tree species that have been detected, but because these were mostly rare
288 in vineyards, it alternatively fed on vine buds. Hence, management practices that encourage patches of weed
289 growth in vineyards could probably minimise wētā feeding on vines.

290 No dietary variation was detected in relation to date of sampling based on FOO and RRA (LRT = 126.8, p =
291 0.262). This was true when analysing the full dataset and when focusing on the major food items (plant genera
292 detected in more than 50% of the samples analysed). Indeed, the proportions of faecal samples that tested
293 positive for DNA of *Vitis, Poa, Festuca*, *Anthoxanthum*, *Anagallis, Camelina, Epilobium, Pedicularis, Urtica,*
294 *Garrya, Pinus,* and *Tilia* did not change significantly with date. This result is possibly due to a limited number
295 of samples analysed for each season as cumulative curves show that the 18 samples analysed per season allowed
296 detection of an estimated 60.6% of all plant genera and 80.5% of all plant families in the diet of wētā.
297 A significant difference was observed in relation to sampling location for RRA (LRT = 111.9, p = 0.026). In
298 terms of detection, only *Menyanthes*, displayed a significant change in occurrence in relation to sampling
299 location (P = 0.028). Detection of this flowering annual weed in wētā faeces was highest in the O- Block and
300 lowest at the N- Block. The detection rate of this genus in the D-, H-, and L- Blocks was significantly lower
301 than that recorded in the O- Block but higher than in the N- Block (Fig. 3B). These small geographical
302 variations, reflect slight differences in the blocks' plant communities and confirm that wētā are highly
303 generalists and capable of feeding widely on the plants present in their environment.
304

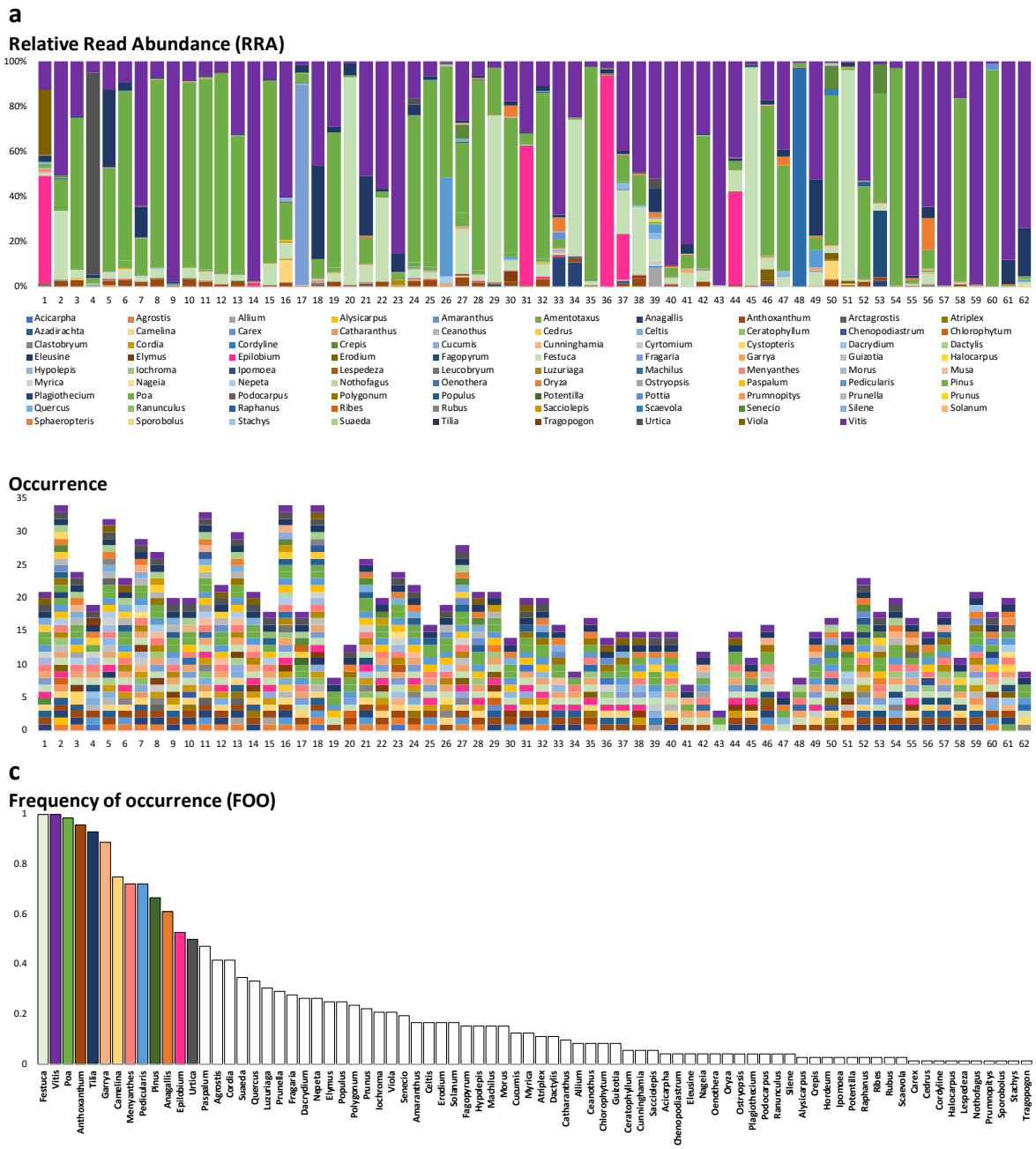

Figure 4: Wētā individual diet analysis. Relative read abundance (A) and Occurrence (B) of plant genera as measured from each individual faecal sample. Frequency of occurrence (C) of the different plant genera, with major food items represented in colour.

Our study allowed for the identification of the main food items in the diet of wētā. This approach could be followed by food choice experiments to ascertain wētā food preferences between the small number of trap crop candidates identified here, and to select one or two optimal trap crops. Alternatively, and because wētā appear highly generalists, seeds of cultivars from various taxa could be sown in mixtures to provide divers alternative

314 food for this insect. In this case, the main selection criteria may be the plant individual needs (climatic, soil etc.)
315 to ensure they provide enough resources at the critical time of vine budburst, thereby reducing damage to vines.

316 The deliberate use of weed species to attract natural enemies for insect pests' population regulation has been
317 studied extensively (Sarkar et al. 2018). The findings here suggest that this approach to pest management (when
318 adopted by vine growers) would have the added advantage of reducing crop damage by generalist insect
319 herbivores and omnivores such as wētā, which may use weeds as alternative foods. However, to limit the
320 multiplication of wētā, trap cropping should be restricted to the vine sensitive period (around budburst) and the
321 trap crop removed as soon as the vine leaves are tough enough that they are no longer targeted by wētā. Some
322 unexpected genera were detected in the faeces analysed. For example, *Tilia* spp. (an ornamental mostly found in
323 urban parks and gardens), *Populus* spp *Solanum* spp., *Ipomoea* spp., *Cucumis* spp., *Quercus* spp., etc. are not
324 usually common around vineyards in New Zealand. Considering that the identified taxa were assigned up to the
325 genus level, it is quite hard to anticipate the origin of those taxa in the samples analysed. Perhaps, they came
326 from other sources. For example, because wētā are omnivorous and known to also feed on other insects, the
327 presence of tree DNA may be explained by secondary predation. Further studies would be needed to clarify this
328 observation. This notwithstanding, the molecular diet analysis used here highlighted the high level of diet
329 mixing in this species and hinted of potential plant families or genera that can be used for trap cropping. Based
330 on these results, effective management of wētā will likely require sowing more than one plant species as trap
331 crops to adequately satisfies the nutritional needs of this insect. The proposed method also presents some
332 limitations. Firstly, the use of a short single-locus molecular marker only allowed identification of most plant
333 MOTUs at the genus level and gives no information about predation on other invertebrates. This last point is
334 important because prey nutrient content can modify omnivorous insect's propensity to engage in herbivory
335 (Ugine et al. 2019). Secondly, the RRA may not be an accurate quantitative measure of actual amount of each
336 species consumed (Deagle 2019). Thirdly, there may be poor representations of particular species for which
337 primer affinity was low (Alberdi et al. 2018). Nonetheless, this type of analysis provides extremely valuable
338 information, and as NGS technology improves, some of the above limitations are becoming less problematic
339 (see Alberdi et al. 2018; Pompanon et al. 2012).

340

341 **4. Conclusion**

342 In summary, the current work examined how the results of faecal DNA analyses could potentially contribute to
343 developing trap cropping strategy for managing a generalist insect pest, thus reducing the high pesticide input in

most modern agriculture. Primers targeting a short fragment of the *rbc*L gene region were used to successfully identify the range of plants eaten by wētā, at least to the genus level. A wide variety of plant families were found in the diet of the target insect, in spite of grasses being abundant in vineyards. Such feeding behaviour is common among generalist insects, both herbivores and omnivores, and it is thought to ensure a balanced intake of major nutrients (proteins and carbohydrates). Hence, non-pesticidal management strategies for generalist insect pests could rely on trap crops that offset existing nutrient imbalances. For wētā, non-crop species with high protein content are recommended in agricultural systems dominated by plants with high carbohydrate content, and they should be planted to coincide with periods of damage to the economic crop. If these plants are potential weeds, they can be removed, for example with herbicides, once the main period of pest-induced damage has passed. The method used here could be applied to other agricultural pests, in particular, those feeding on seasonal resources. By collecting insects and analysing their diet outside their damaging period, it is possible to acquire the necessary knowledge to apply an efficient trap cropping strategy. Many non-crop plants in vines or other crops deliver a wide range of ecosystem services throughout the growing season (Shields et al. 2016), some of which are not well understood. Managing non-crop plants in agriculture is key to achieve 'sustainable intensification' and could be better informed through molecular diet analyses of pest species.

Acknowledgements

The authors gratefully acknowledge Ollie Davidson, Joanne Brady, Craig Payne, Cliff Pilcher, Gerard Shand, Robert Blathwayt and the entire staff and management of Constellation Brands NZ for their support during the study. Staff in the Bio-Protection Research Centre, Lincoln University are also gratefully acknowledged. We would like to thank two anonymous reviewers for their thoughtful suggestions that have considerably helped improving the manuscript.

Funding

This study was funded by the New Zealand Government's Callaghan Innovations PhD Research and Development Grant, in partnership with Constellation Brands NZ (Project no. CONB1201). Other funding came from a Bio-Protection Research Centre grant and a Lincoln University fee scholarship.

Declaration on conflict of interest

The authors declare that they have no conflict of interest.

374

375 **Ethical approval**

376 This article does not contain any study with animals performed by any of the authors.

377

378 **Author contribution statement**

379 JAN, SB and SDW conceived and designed the research. JAN and SB conducted the research. JAN, DS and SB
380 analysed the results. JAN and SB wrote the manuscript and prepared the figures. All authors read, contributed to
381 and approved the final manuscript.

382

383 **References**

384 Alberdi A, Aizpurua O, Bohmann K, Gopalakrishnan S, Lynggaard C, et al. (2019) Promises and pitfalls of
385 using high-throughput sequencing for diet analysis. Mol Ecol Resour 19: 327–348. doi:10.1111/1755-
386 0998.12960.

387 Ali JG, Agrawal AA (2012) Specialist versus generalist insect herbivores and plant defense. Trends Plant Sci
388 17: 293-302. doi: 10.1016/j.tplants.2012.02.006

389 Allan Herbarium LR (2000) Ngā Tipu o Aotearoa - New Zealand Plant Names Database. Landcare Research,
390 New Zealand.

391 Berner D, Blanckenhorn WU, Körner C (2005) Grasshoppers cope with low host plant quality by compensatory
392 feeding and food selection: N limitation challenged. Oikos 111:525-533. doi: 10.1111/j.1600-
393 0706.2005.14144.x

394 Boyer S, Wratten SD, HolyoakeA, Abdelkrim J, Cruickshank RH (2013) Using next-generation sequencing to
395 analyse the diet of a highly endangered land snail (*Powelliphanta augusta*) feeding on endemic
396 earthworms. PLoS One 8, e75962. doi: 10.1371/journal.pone.0075962

397 Brown SDJ, Collins RA, Boyer S, Lefort M-C, Malumbres-Olarte J, Vink CJ, Cruickshank RH (2012) Spider:
398 an R package for the analysis of species identity and evolution, with particular reference to DNA
399 barcoding. Mol Ecol Resour 12:562–5. doi: 10.1111/j.1755-0998.2011.03108.x

400 Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Percy DM, Hajibabaei
401 M, Barrett SC (2011) Discriminating plant species in a local temperate flora using the rbcL+ matK
402 DNA barcode. Methods Ecol Evol 2:333-340. doi: 10.1111/j.2041-210X.2011.00092.x

403 Coll M, Guershon M (2002) Omnivory in terrestrial arthropods: mixing plant and prey diets. Annu Rev Entomol
404 47:267-297. doi: 10.1146/annurev.ento.47.091201.145209

405 Deagle BE, Thomas AC, McInnes JC, Clarke LJ, Vesterinen EJ, et al. (2019) Counting with DNA in
406 metabarcoding studies: How should we convert sequence reads to dietary data? Mol Ecol 28: 391–406.
407 Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.14734.
408 Ekström G, Ekbom B (2011) Pest control in agro-ecosystems: an ecological approach. Critical Rev Plant Sci
409 30:74-94. doi: 10.1080/07352689.2011.554354
410 Johns P (2001) Distribution and conservation status of ground wētā, *Hemiandrus* species (Orthoptera:
411 Anostostomatidae). Department of Conservation Wellington,, New Zealand.
412 Lear G, Dickie I, Banks J, Boyer S, Hannah L, et al. (2018) Methods for the Extraction, Storage , Amplification
413 and Sequencing of DNA from Environmental Samples. N Z J Ecol 42: 10-50S.
414 doi:10.20417/nzjecol.42.9.
415 Lefort M.-C, Worner SP, Rostas M, Vereijssen J, Boyer S (2015) Responding positively to plant defences, a
416 candidate key trait for invasion success in the New Zealand grass grub *Costelytra zealandica*. N Z J
417 Ecol 39: 128-132.
418 Lledó S, Rodrigo S, Poblaciones MJ, Santamaria O (2015) Biomass yield, mineral content, and nutritive value
419 of Poa pratensis as affected by non-clavicipitaceous fungal endophytes. Mycol Progress 14: 67.
420 10.1007/s11104-015-2596-0
421 McClenaghan B, Gibson JF, Shokralla S, Hajibabaei M (2015) Discrimination of grasshopper (Orthoptera:
422 Acrididae) diet and niche overlap using next-generation sequencing of gut contents. Ecol Evol 5(15):
423 3046-3055. doi: 10.1002/ece3.1585
424 Moreau TL, Isman MB (2012) Combining reduced-risk products, trap crops and yellow sticky traps for
425 greenhouse whitefly (Trialeurodes vaporariorum) management on sweet peppers (Capsicum annum).
426 Crop Protection 34: 42-46. doi: 10.1016/j.cropro.2011.11.011
427 Nboyine J, Boyer S, Saville D, Smith M, Wratten S (2016) Ground wētā in vines of the Awatere Valley,
428 Marlborough: biology, density and distribution. N Z J Zool 43: 336-350.
429 Nboyine JA, Boyer S, Saville DJ, Wratten SD (2017) Agroecological management of a soil-dwelling
430 orthopteran pest in vineyards. Insect Sci. doi: 10.1111/1744-7917.12425
431 Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz BA, Bland H, Evershed RP, Possnert G, Pääbo
432 S (1998) Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*.
433 Science 281: 402-406. doi: 10.1126/science.281.5375.402

434 Pompanon F, Deagle BE, Symondson WO, Brown DS, Jarman SN, Taberlet P (2012) Who is eating what: diet
435 assessment using next generation sequencing. Molecul Ecol 21: 1931-1950. doi: 10.1111/j.1365-
436 294X.2011.05403.x
437 Raubenheimer D, Simpson S (2003) Nutrient balancing in grasshoppers: behavioural and physiological
438 correlates of dietary breadth. J. Exper Biol 206: 1669-1681. doi: 10.1242/jeb.00336
439 Rockström J, Steffen W, Noone K, Persson Å, Chapin III FS, Lambin E, Lenton T, Scheffer M, Folke C,
440 Schellnhuber HJ (2009). Planetary boundaries: exploring the safe operating space for humanity. Ecol
441 Society 14.
442 Rusch A, Chaplin-Kramer R, Gardiner MM, Hawro V, Holland J, Landis D, Thies C, Tscharntke T, Weisser
443 WW, Winqvist C (2016) Agricultural landscape simplification reduces natural pest control: A
444 quantitative synthesis. Agric Ecosyst Environ 221: 198-204. doi: 10.1016/j.agee.2016.01.039
445 Sarkar S, Wang E, Wu S, Lei Z (2018) Application of trap cropping as companion plants for the management of
446 agricultural pests: a review. Insects 9(4):128.
447 Shields MW, Tompkins J-M, Saville DJ, Meurk CD, Wratten S (2016) Potential ecosystem service delivery by
448 endemic plants in New Zealand vineyards: successes and prospects. PeerJ 4, e2042. doi:
449 10.7717/peerj.2042
450 Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebø JH, Ims RA,
451 Yoccoz NG (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with
452 high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Front
453 Zool 6: 16. doi: 10.1186/1742-9994-6-16
454 Ugine TA, Krasnoff SB, Grebenok RJ, Behmer ST, Losey JE (2019) Prey nutrient content creates omnivores
455 out of predators. Ecology letters 22(2):275-83.
456 Van Wyngaarden, F (1995) The ecology of the Tekapo ground wētā (Hemiandrus new sp., Orthoptera:
457 Anostostomatidae) and recommendations for the conservation of a threatened close relative. Master of
458 Science in Environmental Science, University of Canterbury, New Zealand.
459 http://hdl.handle.net/10092/7752
460 Villa M, Santos SA, Mexia A, Bento A, Pereira JA (2016) Ground cover management affects parasitism of
461 Prays oleae (Bernard). Biol Control 96: 72-77. doi: 10.1016/j.biocontrol.2016.01.012

462 Villanueva-Rey P, Vázquez-Rowe I, Moreira MT, Feijoo G (2014) Comparative life cycle assessment in the
463 wine sector: biodynamic vs. conventional viticulture activities in NW Spain. J Cleaner Prod 65: 330-
464 341. doi: 10.1016/j.jclepro.2013.08.026
465
466

467 **Statement of data availability**

468 Data and analyses (R codes) are available on a Figshare repository (doi: 10.6084/m9.figshare.5777853)
469
470

471 **Table 1:** Plant taxa identified from wētā faeces and their detection rate (i.e., proportion of wētā tested positive

472 for each taxa).

473

Family	Genera detected	Common name	Detection
Alstroemeriaceae	*Luzuriaga* Ruiz & Pav. 1802	-	27.42
Amaranthaceae	*Amaranthus* L.	Amaranth	24.19
	Atriplex L.	Saltbush	9.68
	Chenopodiastrum (L.) S. Fuentes,	Nettle-leaved Goosefoot	3.23
	Suaeda Forssk. ex J. F. Gmel.	-	35.48
Amaryllidaceae	*Allium* L.	Onion	4.84
Apocynaceae	*Catharanthus* G. Don	Periwinkles	8.06
Asparagaceae	*Chlorophytum* Ker Gawl.	Spider/ airplane plant	8.06
	Cordyline Comm. ex R.Br.	Cabbage-palm	1.61
Asteraceae	*Crepis* L.	Narrowleaf hawksbeard	3.23
	Guizotia Cass. 1829	Annual herb	8.06
	Senecio L.	Ragworts	17.74
	Tragopogon L.	-	1.61
Betulaceae	*Ostryopsis* Decne.	-	4.84
Boraginaceae	*Cordia* L.	Shrubby tree	41.94
Brassicaceae	*Camelina* Crantz	False flax	74.19
	Raphanus L.	Radish	3.23
Calyceraceae	*Acicarpha*	-	4.84
Cannabaceae	*Celtis* L.	Nettle trees/ Hackberries	17.74
Caryophyllaceae	*Silene* L.	Catchfly	3.23
Ceratophyllaceae	*Ceratophyllum* L.	Hornwort	6.45
Convolvulaceae	*Ipomoea* L.	-	1.61
Cucurbitaceae	*Cucumis* L.	Muskmelon	11.29
Cupressaceae	*Cunninghamia* R.Br.	-	8.06
Cyperaceae	*Carex* L.	Sedges	1.61
Dennstaedtiaceae	*Hypolepis* Bernh. 1806	-	14.52
Fabaceae	*Alysicarpus* Desv.	Moneyworts	1.61
	Lespedeza Michx.	Bush clovers	1.61
Fagaceae	*Quercus* L.	Oak tree	30.65
Garryaceae	*Garrya* Douglas ex Lindl.	Ashy silktassel	87.10
Geraniaceae	*Erodium* Aiton	Redstem filaree	12.90
Goodeniaceae	*Scaevola* L.	Fairy fun-flower	1.61
Grossulariaceae	*Ribes* L.	Currants	1.61
Lamiaceae	*Nepeta* L.	-	22.58
	Prunella L.	Self-heal	30.65
	Stachys L.	Woolly hedgenettle	1.61
Lauraceae	*Machilus* Nees	-	14.52
Malvaceae	*Tilia* L.	Lime tree	93.55
Menyanthaceae	*Menyanthes* L.	-	70.97
Moraceae	*Morus* L.	Mulberry	12.90
Myricaceae	*Myrica* L.	-	12.90
Nothofagaceae	*Nothofagus* Blume		1.61
Onagraceae	*Epilobium* L.	Hoary willowherb	53.23

	Oenothera L.	Tall evening primrose	3.23
Orobanchaceae	*Pedicularis* L.	Warrior's plume	72.58
Pinaceae	*Cedrus* Trew	-	1.61
	Pinus L.	Pine	67.74
Poaceae	*Agrostis* L.	Bentgrass	43.55
	Anthoxanthum L.	Grass	95.16
	Dactylis L.	Grass	4.84
	Eleusine (L.) Gaertn.	Goosegrass	3.23
	Elymus L.	Grass	27.42
	Festuca L.	Fescue	100
	Hordeum L.	Grass	1.61
	Oryza L.	Rice	3.23
	Paspalum L.	Crowngrasses	48.39
	Poa L.	Grass	98.39
	Sacciolepis Nash	Cupscale grass	6.45
	Sporobolus R.Br.	Grass	1.61
Podocarpaceae	*Dacrydium* Lamb.	-	30.65
	Halocarpus C. J. Quinn	Tree	1.61
	Nageia Gaertn, 1788	Shrub	4.84
	Podocarpus L'Hér ex Pers. 1807	-	4.84
Polygonaceae	*Fagopyrum* Mill.	Buckwheat	14.52
	Polygonum L.	Japanes knotweed	24.19
Primulaceae	*Anagallis* L.	Pimpemel	59.68
Prumnopityaceae	*Prumnopitys* Phil.	-	1.61
Ranunculaceae	*Ranunculus* L.	Buttercup/ crowfoot	3.23
Rhamnaceae	*Ceanothus* L.	-	9.68
Rosaceae	*Fragaria* L.	Strawberries	30.65
	Potentilla L.	Cinquefoils	1.61
	Prunus L.	Bitter berry	22.58
	Rubus L.	Raspberries, Blackberries	3.23
Salicaceae	*Populus* L.	-	24.19
Solanaceae	*Iochroma* Benth.	Shrub	20.97
	Solanum L.	Tomato	19.35
Urticaceae	*Urtica* L.	Annual nettle	69.35
Violaceae	*Viola* L.	Crow-foot violet	19.35
Vitaceae	*Vitis* L.	Grape vine	100

474

475

Supplementary information:

Accumulation curves representing the cumulative number of plant families (left) and genera (right) detected against the number of faecal samples analysed (n = 72). Horizontal solid lines represent the estimated total number of plant families (left) and genera (right) expected with limitless sampling, based on bootstrapped estimates. Vertical dashed lines indicate sampling efforts for the current experiment, intercepting the curves at 40 families and 61 genera for seasonal sampling and 47 families and 79 species for overall sampling.