



OPEN

Laboratory protocol is important to improve the correlation between target copies and metabarcoding read numbers of seed DNA in ground beetle regurgitates

Veronika Neidel[✉] & Michael Traugott

DNA metabarcoding is increasingly important for studying feeding interactions, yet it remains unresolved whether reporting read counts or occurrences is to be preferred. To address this issue for gut content samples, basic experimental data on the relationship between read numbers and initial prey DNA amounts and how both change over digestion time is needed. Using regurgitates of the carabid *Pseudoophonus rufipes* the digestion of *Taraxacum officinale* seeds was documented for 128 h post-feeding to determine how the number of (1) seed DNA copies and (2) metabarcoding reads change over digestion time, and (3) how they correlate to each other. Additionally, we tested (4) whether PCR cycle-numbers during library preparation affect this correlation. The number of copies and reads both decreased with digestion time, but variation between samples was high. Read and copy numbers correlated when using a library preparation protocol with 35 cycles ($R^2 = 42.0\%$), yet a reduction to 30 cycles might significantly improve this correlation, as indicated by additional PCR testing. Our findings show that protocol optimization is important to reduce technical distortions of read numbers in metabarcoding analysis. However, high inter-sample variation, likely due to variable digestive efficiency of individual consumers, can blur the relationship between the amount of food consumed and metabarcoding read numbers.

Over the past decade, the use of DNA metabarcoding for the study of feeding interactions has increased considerably¹. DNA metabarcoding (hereafter simply, ‘metabarcoding’) is a method that uses DNA barcodes for the simultaneous identification of multiple species contained in mixed samples, typically via high-throughput sequencing (HTS) techniques^{2–4}. For each taxon detected within a dietary sample, a specific number of sequence reads is generated during HTS, which has prompted the idea that the number of reads and the proportion of reads between taxa could potentially hold information on the biomass or number of prey items as well as the relative prey proportions ingested by a consumer⁵. Currently, there is an ongoing discussion whether metrics based on read numbers are indeed superior to mere occurrences (presence/absence data) when reporting the results of metabarcoding studies of trophic interactions⁶. For faecal samples specifically the aspect of quantification has been addressed with ambiguous results⁷, however, there is a striking lack of information regarding the significance of read numbers for samples based on gut content.

Yet, gut content, including dissected guts, whole-body extracts (homogenized individuals) and regurgitates, plays an important role for molecular diet analysis, especially when it comes to the study of arthropods’ feeding interactions^{8–10}. Metabarcoding analysis of gut content samples is considered more challenging than of faecal samples, due to the effect of unknown digestion times on recoverable prey DNA.

In studies using diagnostic PCR assays, the effect of digestion time on prey DNA is taken into account by reporting the presence/absence data translated to detectability at specific times after feeding^{11–13} or through detectability half-lives^{14,15}. Generally, the detectability of prey DNA in the gut content of a consumer decreases

Applied Animal Ecology Research Unit, Department of Zoology, University of Innsbruck, Technikerstraße 25, 6020 Innsbruck, Austria. ✉email: Veronika.Neidel@uibk.ac.at

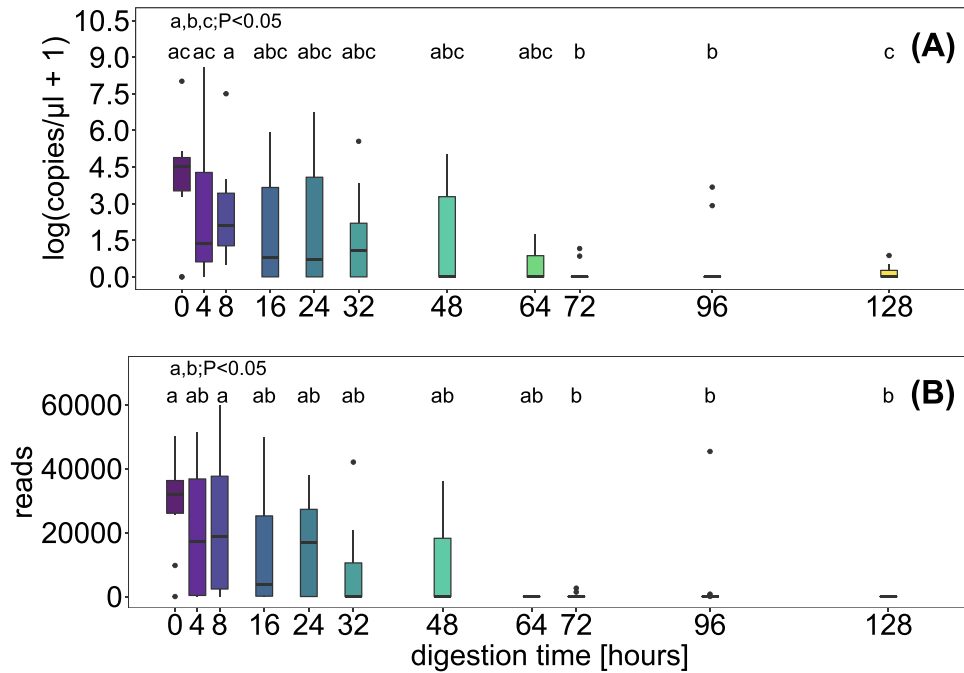


Figure 1. Target copies/ μl DNA extract measured with droplet digital PCR (A, $n = 111$) and read numbers obtained in HTS (B, $n = 102$) of the plant-specific *trnL* c-h target fragment in regurgitates of *Pseudoophonus rufipes*, collected at 0, 4, 8, 16, 24, 32, 48, 64, 72, 96 and 128 h after feeding on three seeds of *Taraxacum officinale*. Boxplots of measured values, letters indicate time groups that were significantly different in pairwise Wilcoxon tests with BH correction for multiple comparisons.

over digestion time, suggesting a continuous breakdown of DNA. To study how the copy numbers in predators' guts are affected by prey numbers and digestion time, diagnostic gut content analysis has been done with quantitative or real-time PCRs (qPCR). For example, prey-specific DNA target molecules of the silverleaf whitefly *Bemisia tabaci* were shown to increase linearly with the number of consumed individuals and to decrease over time of digestion in the guts of the ladybird beetle *Propylea japonica*¹⁶. Prey DNA in the guts of the ladybird beetle *Coleomegilla maculata* was found to decrease exponentially, but with different decay rates for different prey items¹⁷. An exponential decay over digestion time was also reported for the number of reads in a shotgun sequencing study using dissected guts of ladybird beetles fed with aphids¹⁸. For metabarcoding assays, however, no studies are available that report on the amount of prey DNA in gut content samples, its change over time of digestion and how these values are related to read numbers of metabarcoding analysis.

Here, we use a set of regurgitate samples of the omnivorous ground beetle *Pseudoophonus rufipes* De Greer, 1774 (Coleoptera: Carabidae), that map the first 128 h of digesting seeds of *Taraxacum officinale* (Asteraceae), across 11 points in time after feeding. For this sample set, we determine how (1) the number of DNA target copies contained in DNA extracts, and (2) the number of reads, produced by a standard metabarcoding protocol, change over digestion time, and (3) how they correlate to each other. Initial copy numbers were measured with droplet digital PCR (ddPCR), which provides a direct measure of copy numbers for each sample¹⁹. In environmental monitoring surveys, ddPCR was found to be more sensitive²⁰, and cheaper than qPCR²¹, and at the same time, equally accurate^{21,22}.

Further, we (4) compare the copy- and read-numbers to the relative target-fragment amount in the library-PCR products, which can be viewed as a check-point between the DNA extracts and the HTS-output. Therefore, we combined end-point PCRs with capillary electrophoresis (celPCRs). During capillary electrophoresis, the signal strength of each PCR product is estimated in relative fluorescent units (RFU), a measure that strongly correlates with the initial copy numbers in DNA extracts²³.

Based on the observed distortions in copy numbers, we conducted additional PCR assays to study, if (5) increasing or lowering the PCR cycle number can mitigate amplification biases during library preparation and provide a means for getting better estimations of initial target DNA copy numbers.

Results

Prey DNA copy numbers vary strongly between samples, means decrease over time of digestion. Absolute copy concentrations of the *trnL* c-h target fragment in DNA extracts of regurgitates from *P. rufipes* after the consumption of three *T. officinale* seeds ranged from 0.65 to 5490 copies/ μl extract. The mean copy numbers per time point were found to decrease with digestion time ($n = 111$; Pearson $r = -0.437$, 95% CI $[-0.577, -0.273]$, $p < 0.001$, Fig. 1A). The standard deviation, however, was quite high and took values that were 1.4–2.9 times the mean value (Supplementary Table S3). Overall, there was a significant difference in copies measured at the different times after feeding (Kruskal–Wallis chi-squared = 31.187, $df = 10$, $p < 0.001$). In pair-

wise comparisons of the individual time groups, no significant difference between the average copy numbers measured between 0 and 64 h after feeding was found. Average copy numbers at 72 and 96 were significantly lower than those measured at 0, 4 and 8 h after feeding (Fig. 1 and Supplementary Table S5a).

Decrease of prey DNA reads over time of digestion. We recovered between 7 and 59,968 sequence reads per sample that matched a reference database entry at a level of 98% and above (mean \pm sd = 26,256.5 \pm 15,964.7 reads). The resulting species identities indicated that the targeted fragment of the *trnL c-h* marker region did not allow for species-level determination: only a fraction of the recovered sequences was identified as *T. officinale*, the seed provided to the beetles as an experimental diet (mean 0.03% of sequences per sample with match \geq 98% for *T. officinale*). Of the total reads assigned to the target fragment length of 135–160 bp, 23 plant-families were detected, of which 87.6% were assigned to be Asteraceae, all other families made up 3.6% or less of the total read count over all samples (Table S6, Figure S1). Per sample, a mean of 78.6 \pm 37.3% of the sequences were assigned to the family Asteraceae, including the afore mentioned ones assigned to *T. officinale*. Thus, we assumed these reads to be the result of the consumed seeds and used the respective read counts as a reference value for analysis. The Asteraceae reads were negatively correlated with the digestion time (Pearson $r = -0.465$, 95% CI [-0.605; 0.298], $p < 0.001$). As with copy numbers, we observed high variability in measured reads between samples at the same time after feeding on three seeds of *T. officinale* (Supplementary Table S4). There was a significant difference in Asteraceae reads measured at different times after feeding (Kruskal–Wallis chi-squared = 34.689, df = 10, $p < 0.001$), and pairwise comparison revealed a significant difference between read numbers at 0 and 12 h compared to 72 h and longer (Fig. 1B, Supplementary Table S5b).

Initial target copy numbers were correlated with HTS-read count. The linear model including $\log(\text{copies}/\mu\text{l} + 1)$ as a predictor of the recovered read numbers could explain 42.0% of the observed variation in the sequence reads assigned to the family Asteraceae in our HTS analysis ($R^2 = 0.43$, $F(1, 62) = 46.6$, $p < 0.001$; Fig. 2). A 1% increase in copy numbers was predicted to lead to an increase of 53.5 reads (95% CI [33.1, 68.2], $p < 0.001$, Supplementary Table S1A).

Signal strengths of library PCR correlated to read counts and initial target copy number. When modelling the copy numbers from the library preparation PCR product signal strengths, the linear regression explained 58.8% of the observed variation ($R^2 = 0.588$, $F(1, 86) = 122.9$, $p < 0.001$, Supplementary Table S2). Copy numbers increase by 11.5% with every 0.1 increase in RFU-values (95% CI [8.3, 15.0], $p < 0.001$, Fig. 3A). The model prediction of read numbers from the signal strengths of the PCR product [RFU] was also statistically significant ($R^2 = 0.536$, $F(1, 62) = 71.61$, $p < 0.001$). The read numbers increase by 553.4 reads with every 0.1 increase in RFU-values (95% CI [395.9, 697.7], $p < 0.001$, Fig. 3B, Supplementary Table S1B).

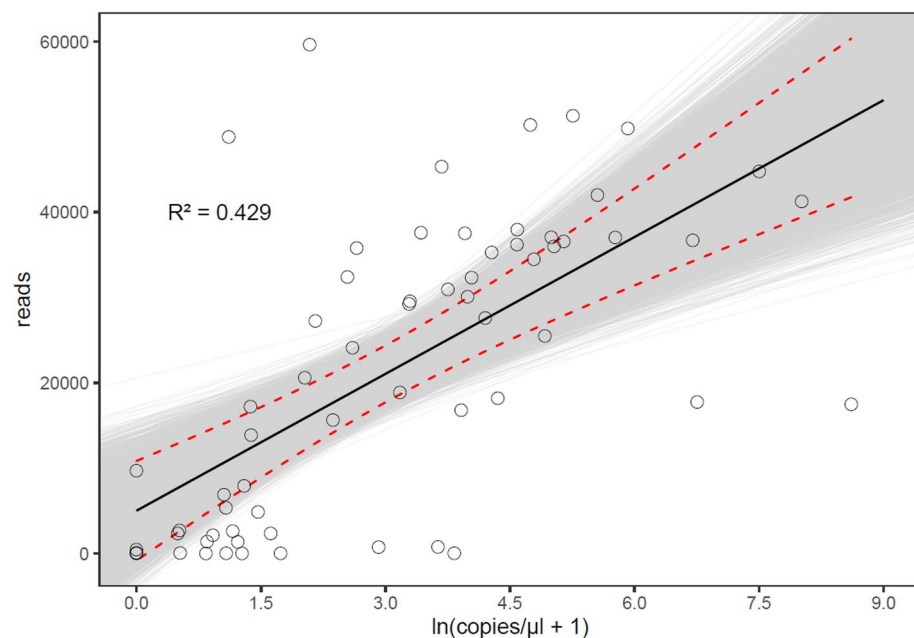


Figure 2. Correlation of copies/ μl and read numbers of the plant-specific *trnL c-h* target fragment in regurgitates of *Pseudoophonus rufipes* collected at 0, 4, 8, 16, 24, 32, 48, 64, 72 and 96 h after feeding on 1–3 seeds of *T. officinale*. Dots in the scatterplot (○) are measured values, black bold lines are predicted values of a linear regression model of copies and reads, dashed lines are 95% confidence intervals, shaded area marks bootstrapped 95% confidence interval, and R^2 indicates goodness of model-fit.

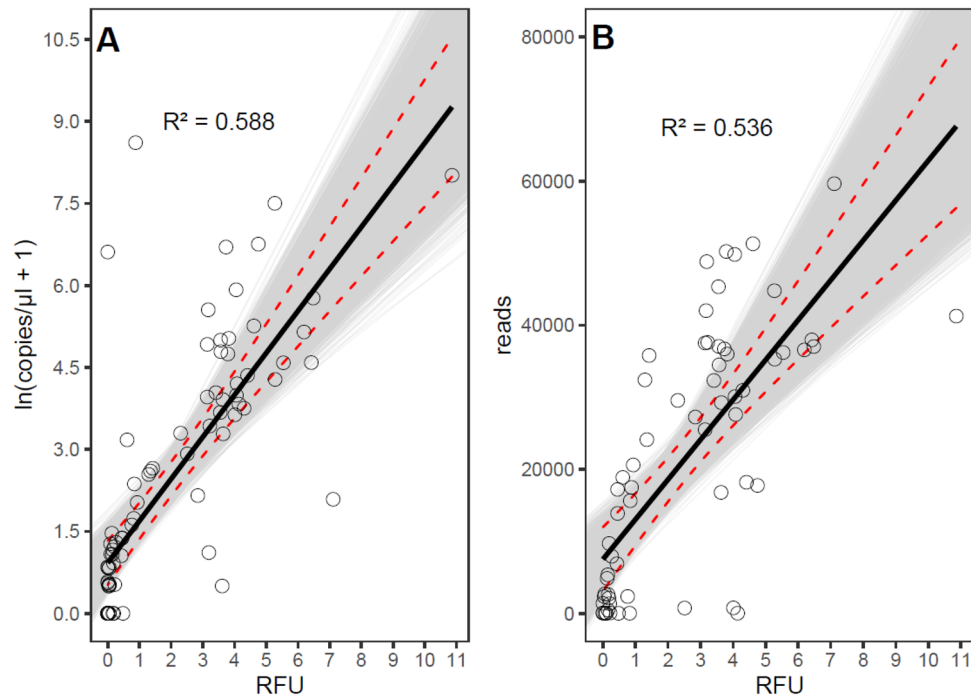


Figure 3. Signal strengths [RFU] of celPCR products of the library preparation PCR for high-throughput sequencing, plotted against (A) copies/ μl and (B) *Asteraceae* read numbers of the plant-specific *trnL c-h* target fragment, measured in regurgitates of *Pseudoophonus rufipes*, collected at 0, 4, 8, 16, 24, 32, 48, 64, 72, and 96 h after feeding on 1–3 seeds of *Taraxacum officinale*. In the scatterplots, dots (O) represent measured values, black bold lines are predicted values of linear regression models, dashed lines are 95% confidence intervals, shaded area marks boot-strapped 95% confidence interval, and R^2 indicates goodness of model-fit.

Reduced reaction volume in PCR yields a better estimation of initial target copy numbers. RFU-values of singleplex PCRs with a volume of 10 μl and 35 cycles had a greater explanatory power for the initial copy numbers (Fig. 4, Supplementary Table S2D), than the library preparation PCR with a reaction volume of 15 μl and 35 cycles (Fig. 3B, Supplementary Table S2A). Variation in copy numbers was explained to 84.9% by RFU-values as predictor ($R^2=0.849$, $F(1, 90)=504.1$, $p<0.001$). Copy numbers increased by 48.45% with every 0.1 RFU increase (95% CI [38.6, 59.7], $p<0.001$).

PCR cycle number strongly affects the correlation between RFUs and initial copy numbers. The increase and decrease of cycles also changed the relationship between the copy numbers and the PCR product signal strengths (RFU). While all tested PCR cycle variants reflected the initial copies well, there was a varying degree of explanatory power (Fig. 4). The model with RFUs derived from celPCRs with 30 cycles showed the strongest correlation between RFU-values and copy numbers ($R^2=0.95$, $F(1, 90)=1760$, $p<0.001$). RFU-values obtained from celPCRs with 25 cycles explained the variation to 73.9% ($F(1, 90)=258$, $p<0.001$). The model for RFUs after 40 cycles could account for only 53.3%, ($F(1, 90)=104.9$, $p<0.001$) and had, thus, the lowest explanatory power of the assessed thermocycling protocols.

Discussion

We measured the absolute target copy numbers and the HTS read numbers of a plant-specific DNA marker in regurgitate samples of the carabid *P. rufipes* at different times after feeding on dandelion seeds to study its change during digestion. The intuitive assumption, that target DNA copy numbers would be high early, and low late during the digestive progress, was only partly confirmed by our results. The mean copy numbers in samples collected at the same time after feeding indeed decreased with increasing digestion time due to a shrinking number of samples with positive detections. Moreover, we found that read numbers from metabarcoding were only moderately correlated to the target copy numbers contained in the original DNA extracts. Thus, on the level of the individual sample, neither copy numbers nor HTS read numbers of gut content samples held reliable information on whether a feeding event had taken place very recently before sampling. This, once more, stresses the need to maximise biological replication in molecular diet studies²⁴. While some samples failed to amplify DNA even soon after feeding, a high number of target copies was more frequently associated with a shorter digestion time. Thus, if high read counts occur in a major part of the tested individuals, a prey item is likely to be consumed frequently and thus be of importance to the consumer species. Consequently, we suggest, that when working with gut content samples of arthropods for the study of feeding interactions with plant seeds, sufficient samples numbers need to be collected to capture differences between field sites or consumers.

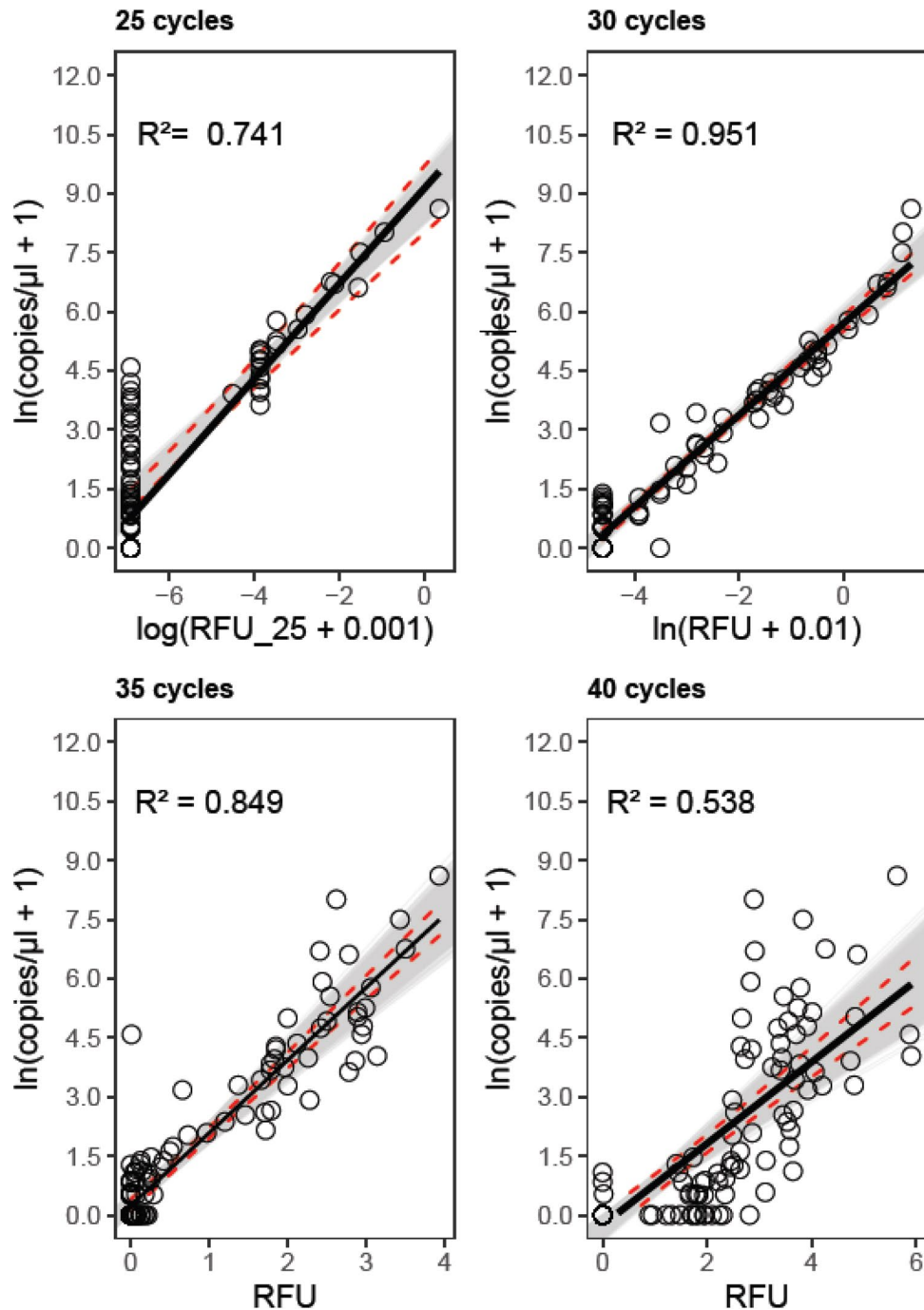


Figure 4. Correlations of plant-specific *trnL*-target copies/μl DNA extract and signal strength (RFU) of celPCRs products after 25, 30, 35 or 40 cycles with a reaction volume of 10 μl. Dots (○) are measured values, black bold lines are predicted values of linear regression models, dashed lines are 95% confidence intervals, shaded area are boot-strapped 95% confidence intervals and R^2 indicates goodness of model-fit.

While we did not compare different meal sizes directly, our results indicate that meals of the same size ingested by different consumer individuals of the same taxon can yield a wide range of different target copy numbers, and further read numbers in HTS. For whole-body samples of ladybird beetles, Weber and Lundgren¹⁷ found that prey DNA means estimated for prey items increased with an increasing number of items consumed. Yet, the data had wide deviation within the meal size variants, leading to a strong overlap among them, which was also reflected in the prey DNA copy numbers over time of digestion. Zhang et al.¹⁶ reported a linear relationship between estimated copy numbers and ingested number of larvae, however, SDs were not reported. Both studies used animals as experimental prey, while we fed consumers with plant seeds. For gut content samples of carabids, the food DNA detection probability curves over time of digestion follow a constant decrease for animal prey but less so for plant seeds²⁵, which might also be the case regarding copy numbers.

Several technical, methodological and biological factors can cause variation between samples, for example extraction efficiency^{5,26} or PCR inhibition^{27,28}. For the present study, DNA was extracted from regurgitate samples using an automated extraction platform with a magnetic-bead-based extraction technology, which allows for a standardized and efficient extraction process^{29,30}. Also, we used droplet-digital PCR for absolute target copy quantification, a technology that is less prone to PCR inhibition than endpoint or qPCR^{31,32}. Thus, we consider our methods robust to such biases.

Therefore, we assume the variation is due to biological factors, such as inter-individual differences in the digestive efficiency of the beetles³³. Seed feeding in carabid beetles is facilitated by bacterial endosymbionts in guts, which can vary in their occurrence between individual beetles of the same species³⁴, potentially affecting digestive capacity of individuals. It is possible that slight variation in size, age or maturation stage of the seeds we provided as food, could have affected copy number detections. However, to our knowledge, there is no literature available on the amount of chloroplast DNA in plant seeds.

It is possible that the regurgitated amount of gut content differed between individual beetles. Regurgitates are a non-lethal type of gut content sample³⁵, that resembles whole-body extracts regarding the change in detection probability over time of digestion³⁶. Regurgitation also occurs naturally in carabids: some species regurgitate the digestive enzymes contained in the crop, which is a part of the foregut, onto their prey for extra-oral digestion³⁷. The crop of carabids is the part of the digestive system, where the most recent meal is stored and digested^{37–39}. Therefore, it can be assumed that regurgitates are representative of the last meal consumed by the beetles. Sunderland³⁹ found remains of previous meals in carabids' hind guts. Thus, dissected guts as well as whole-body extracts are likely to contain DNA of the most recent meal and, additionally, of previous ones. However, for whole-body samples, the entire consumer is homogenized, leading to an overabundance of intact consumer DNA in the extract, potentially outcompeting the digested food remains during extraction. Dissected guts and regurgitates, even more so, contain less consumer tissue, thus an improved yield of food DNA is to be expected^{29,36}. Whether extracts of entire digestive systems, such as in whole-bodies or dissected guts, differ from regurgitates regarding prey read numbers needs to be assessed.

It has been suggested that faeces are better suited for quantitative estimations than gut content samples because, as the final product of digestion, they are supposed to be more stable and, thus, directly comparable regarding the condition of target DNA⁴⁰. However, as previously mentioned, the aptitude of prey DNA from faeces for quantitative metabarcoding is still debated^{6,7}. Comparative studies actually suggest that gut content yields higher read counts in metabarcoding, potentially due to the more advanced degree of prey decomposition and the presence of inhibitory substances in faeces⁴¹. The disadvantage of gut content samples, however, is that the unknown time lapse since feeding in field-collected samples, can cause uncertainties for the estimation of ingested biomass. For a shotgun sequencing analysis, Paula et al.⁴² demonstrated how to estimate likely combinations of prey individual numbers consumed and time lapse since feeding based on experimentally determined DNA decay rates. Given there is little variation in expected DNA amounts per prey individual, and DNA digestion follows an exponential decay, the approach appears promising. However, neither of these two conditions applies to our present study system.

We found that the number of target copies/ μ l DNA extract derived from regurgitates was moderately correlated with the target plant read numbers recovered during high-throughput sequencing. PCR stochasticity is a frequently mentioned source of bias for the quantitative estimation from read numbers^{43–45}, thus, the distortion between initial and final target fragment estimates did not come as a surprise. We implemented two check-points between ingested biomass and read numbers: droplet digital PCR for absolute copy number quantification and capillary electrophoresis for measuring the relative amount of the target PCR products. Our relative measure of the amount of PCR product, RFU-values, was moderately related to both, the initial copy numbers and the read numbers. Moreover, its correlation with either was higher than that observed among the two measures. As the RFU-values were a stronger predictor of the final read numbers than the initial copy numbers, we suggest optimizing the library preparation protocol in a way that RFU-values strongly reflect initial copy numbers. Our additional celPCR screenings of the samples, with 25, 30 and 40 cycles, revealed the best correlation between initial copy numbers and RFU-values after 30 cycles. This is in agreement with reports from pollen metabarcoding, where library preparation with 30 cycles improved the estimation of pollen grains from read numbers with *trnL* primers⁴⁶. Moreover, Valentini⁴⁷ recommend using 35 cycles at most when working with *trnL* primers to avoid the amplification of environmental DNA contaminations. Generally, high numbers of PCR cycles can lead to saturation, which results in a mismatch between RFU-values from celPCRs and absolute target copy numbers²³. At the same time, too low cycle numbers have been found to impair estimations as well⁴⁴. Correlations between copy number estimates and read numbers, or their respective proportions, have also been reported in studies using qPCR^{48–50}. When implementing capillary electrophoresis in the HTS-library preparation workflow, we strongly recommend to do it after sample indexing or on subsamples only, to reduce the risk of cross-contamination. Based on our findings, we suggest that an adjustment of the overall cycle numbers used during library preparation can play a role in the degree of bias introduced into HTS data. While we think that protocol optimization through cycle number adjustment can increase the fit between copy and read counts, it will not reduce amplification biases due to, for example, primer mismatches⁵¹.

To study the detection of plant seed DNA in regurgitates of carabids, we targeted a conserved region in the chloroplast genome, the *trnL* P6 loop, which allows for successful amplification of a wide range of different plant taxa⁵⁰, with the drawback of having a low taxonomic resolution⁵². Moreover, multi-locus pollen analysis suggests that plastid marker regions, such as *trnL*, are superior to nuclear markers regarding quantitative estimations from read counts^{46,53}. The particular primer combination used in our study, which has highly conserved priming sites among land plants⁵⁴, amplifies a rather short (~200 bp) PCR fragment. It has previously been used in a metabarcoding assay, resulting in a low taxonomic resolution, allowing primarily for family-level identification⁵⁵,

which was also true for our samples. In another study, the same primer combination has been used to target plants in a diagnostic multiplex PCR for the study of the diverse diet of omnivorous carabid beetles in arable land⁵⁶.

Although the beetles in the feeding experiment were provided with seeds of only one plant species, the metabarcoding analysis yielded several reads of other plant families. Among the more abundant plant species, we found reads assigned to the family of *Poaceae*. As grass seeds are among the preferred seeds of *P. rufipes*⁵⁷, the reads could stem from trace amounts of field-consumed seeds still contained in the beetles' guts, despite the extensive digestion interval of at least 7 days prior to the feeding experiments. Also, the samples of the present experiment, were sequenced alongside other samples containing plant DNA. Therefore, it cannot be ruled out that the unexpected plant-family detections were derived through index switching. Index switching is a source of bias known to occur when combinatorial dual indexes are used for high-throughput sequencing with Illumina HiSeq and refers to the mis-assignment of indexes between libraries^{58,59}. The use of unique dual indexes is a good solution to avoid this problem.

Relative estimation of copy numbers through RFU-values derived from celPCRs as demonstrated here, could also be implemented for diagnostic assays in gut content analysis. The high resolution of PCR product patterns separated by capillary electrophoresis facilitates the distinction between different fragment sizes also in multiplex diagnostic PCR approaches employing multiple group-specific primers⁶⁰ or general primers with species-specific differences in amplicon size^{52,61}. For both single- and multiplex diagnostic approaches, RFU-values have been found to correlate to target copy numbers²³.

Here, we found that copy numbers and, further downstream, also read numbers, recovered from gut content samples showed a trend of decrease over time of digestion. However, recovered reads from individual samples may be difficult to use for biomass estimations, as there is a high inter-individual variation between samples. Internal standards or normalization could be a way to go about technical distortions, while biological factors that are part of digestion still need further investigation. For now, we recommend to use read number-derived estimates only with caution, and to plan metabarcoding experiments, based on gut content samples, to include as many replicate samples as possible.

Material and methods

Regurgitates. For the present study, we used gut content samples of adult carabid beetles of the species *P. rufipes*, collected in August 2014 and August 2015²⁵. The carabids were collected with pitfall traps in arable land near Innsbruck/Austria. Adult beetles were kept individually in plastic cups (71 × 58 mm, height × diameter) with moist tissue paper, in a climate chamber at an artificial light regime of L:D 14:10 h and temperatures of 22 °C/12 °C (day/night). They were maintained on a diet of 1/3 mealworm (*Tenebrio molitor* L.) larva every fifth day until being starved for 5 days prior to the feeding experiments.

During the feeding experiment, the beetles were offered five seeds of *Taraxacum officinale* and a drop of water in fresh cups for 2 h in darkness. After feeding on at least one seed, beetles were left to digest without food for 0, 4, 8, 16, 24, 32, 48, 64, 72, 96 or 128 h. Per digestion time interval, 10–15 individuals were used for sample collection. Beetles were put individually into 1.5 ml microtubes and stimulated to regurgitate by dipping the tip of the tube briefly into hot water⁶². Each beetle was used for regurgitation only once during the entire experiment and set free afterwards.

For DNA extraction, regurgitates were mixed with 200 µl lysis buffer (200 µl 1 × TES, 5 µl Proteinase K (20 mg/ml), 1 mg Polyvinylpyrrolidone) and incubated at 58 °C for 3 h. DNA was extracted on a BioSprint[®] 96 extraction platform (Qiagen, Hilden, Germany) with the DNA Blood and Tissue Kit (Qiagen) according to the manufacturer's protocol with elution in 200 µl 1 × TE buffer. The DNA extracts were stored at -29 ± 1 °C before PCR testing.

Primers. For all PCR analysis (schematic overview in Fig. 5), we used the general plant primers *c* B49317 and *h* B49466^{54,63}, which targets an approximately 200 bp long fragment of the *trnL* (UAA) region in the chloroplast genome⁵⁶.

Absolute quantification of target copy numbers with droplet digital (dd)PCR. The copies of the *trnL c-h* DNA fragment in DNA extracts of regurgitates were quantified by droplet digital PCR. Reaction mixes had a total volume of 22 µl, each containing 2.2 µl DNA template, 11.0 µl EvaGreen Supermix (Bio-Rad Laboratories, Inc., Hercules, CA, USA), 0.114 µM of each primer and 8.3 µl molecular grade water. Droplets were generated on the QX200 Automatic Droplet Generator System (Bio-Rad). Thermocycler conditions with optimized annealing temperature for the *trnL* primer pair were set according to the QX200 ddPCR EvaGreen Supermix user manual for 3-step PCRs. The final program started with 5 min of activation at 95 °C, followed by 40 cycles of 30 s denaturation at 95 °C, 1 min annealing at 58 °C and 1 min extension at 72 °C, followed by a first step of stabilization at 4 °C for 5 min and a second step at 90 °C for 5 min and finished with 12 °C on hold. Samples were analysed on the QX200 Droplet Digital[™] System (Bio-Rad).

celPCR: endpoint PCRs and capillary electrophoretic analysis. For endpoint PCRs, the 10 µl reaction volume contained 2.5 µl DNA template, 5.0 µl 2xType-it Multiplex PCR Master Mix (Qiagen), 0.5 µM of each primer, 0.5 µg, and 1.0 µl molecular grade water. Samples were screened in four PCR replicates, each with a different number of cycles without replications within each protocol. Thermocycling programs started with an initial denaturation for 5 min at 95 °C, followed by 25, 30, 35 or 40 cycles of 20 s denaturation at 92 °C, 90 s annealing at 55 °C and 60 s extension at 70 °C, and finished with a final elongation of 5 min at 70 °C. Samples that tested negative in the celPCRs with 40 and 35 cycles were not screened in celPCR-variants with 25 and 30 cycles and neither were they subject to target copy quantification in ddPCR or high-throughput sequencing (HTS).

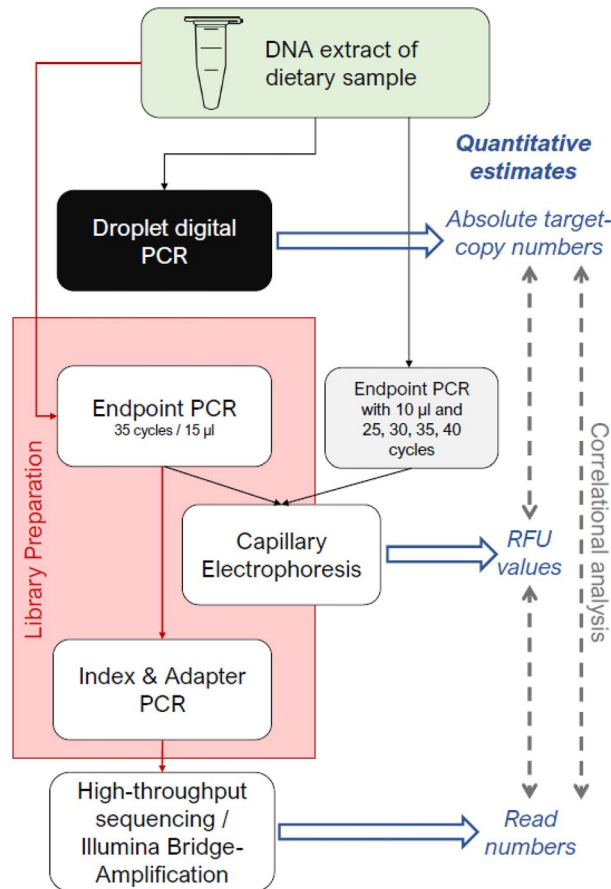


Figure 5. Flow diagram of analysis methods used on regurgitate samples of *Pseudoophonus rufipes* to get absolute and relative estimates of copy/read numbers contained in DNA extracts, PCR products and metabarcoding libraries. Measures of target DNA estimates at different analysis steps are indicated in blue.

Batches of up to 94 samples were screened alongside one negative control and one positive control, containing molecular grade water or DNA of *T. officinale*, respectively.

PCR products were separated and visualized with a QIAxcel Advanced capillary electrophoresis system (Qiagen), using the process profile AM 320 with an injection time of 30 s. We set the parameters of the ScreenGel (Qiagen) software analysis profile to smoothing filter 15 pts, baseline 40 s, minimum 0.25 s, suspend integration 0.50 min, and a threshold 0.1 relative fluorescent units (RFU). Peaks below this threshold were manually inserted and measured.

HTS-library preparation. 64 samples testing positive for the *trnL* target amplicon were further analysed by high-throughput sequencing. Amplicon libraries were prepared with the Nextera[®] XT Indices (Illumina, San Diego, CA, USA) in a combinatorial dual indexing strategy with a 2-step PCR protocol. In the first PCR, the general plant primers *trnL c-h* with appended adapter sequences, as described in the Nextera user manual, were used. The 15 µl reaction mix was prepared with 4 µl DNA template, 7.5 µl 2xType-it Multiplex PCR Master Mix (Qiagen), 0.5 µM of each primer, 0.33 µg BSA and 1.5 µl molecular grade water. Thermocycler conditions were set as described above for celPCRs with 35 cycles.

To remove primer dimers, 7 µl of undiluted PCR product were added to a mix of 0.1 µl Exonuclease I (10 u/µl) (NEB New England Biolabs, Ipswich, MA, USA), 0.1 µl rSAP (NEB) and 1.8 µl water and incubated for 15 min at 37 °C and 15 min at 80 °C in a thermocycler.

PCRs for attaching indices and sequencing adapters were done in total volumes of 25 µl for 5 µl cleaned PCR-1 product, 12.5 µl 2xType-it Multiplex PCR Master Mix (Qiagen), 1 µM of each adapter, 0.4 µg BSA and 1.5 µl water. Index-PCR conditions started with an initial denaturation at 95 °C for 5 min, followed by 8 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s and extension at 72 °C for 60 s, and final elongation at 72 °C for 5 min. Finally, libraries were cleaned with SPRIselect (Beckmann Coulter Inc., Brea, CA, US), following the manufacturer's protocol for left side selection with a ratio of 1:0.8. Samples were pooled with equal volumes (10 µl per sample) and 7 µl of the final pool were sent for sequencing to the VBCF Vienna BioCenter Core Facilities (Vienna, Austria). Samples were sequenced on the Illumina MiSeq System (Illumina), PE300, with a reported total read number of 25.9 million reads/Q > 30 19.3; Bloomca PhiX (7.1).

Bioinformatic processing of high-throughput sequencing data. High-throughput sequencing data were quality checked with FastQC (v0.11.8)⁶⁴. Forward and reverse reads were extracted with samtools (v1.9, using htlib 1.9)⁶⁵. Primers were removed with cutadapt (v3.5)⁶⁶. Usearch (v11.0.667)⁶⁷ was used for merging (settings: fastq_pc 90, fastq_maxdiffs 30) and quality filtering (settings: fastq_maxee 0.5) of paired-end reads, further, for identifying unique sequences, discarding singletons and sorting merged sequences by length, keeping only sequences with at least 130 bp. Finally, sequences were blasted with NCBI-blast+ (v2.8.1)⁶⁸ against the full NCBI nucleotide database (updated 29.10.2021). Sequences of organisms other than plants were discarded. Moreover, only sequences that matched the reference database at a level of 98% identity and above were used for further analysis.

Data analysis and statistics. A total of 145 samples was initially screened for the target fragment in end-point celPCRs with 40 and 35 cycles. Of these samples, 41 did not amplify DNA in either of the cycle-number variants, partly because of long digestion times, during which detection probability decreases. A total of 106 samples were additionally screened in PCRs with 25 and 30 cycles, and 0 RFU were assumed for samples that tested negative in the previous rounds of screening. We used 92 samples for measurement of copies/ μ l extract with ddPCR, and combined them with 40 samples, that were assumed to contain 0 copies because they repeatedly tested negative in celPCRs. Of the samples screened multiple times, 64 were further screened via high-throughput sequencing of the *trnL* c-h target region.

Statistical analysis was done with R statistical software (v3.5.0, 2018-04-23)⁶⁹ and the plots were created using ggplot2⁷⁰ with ggthemes⁷¹.

To study the change of copies/ μ l and read numbers of the target plant over time of digestion, we used only samples of beetles that had consumed three seeds during the feeding trials. The relationship of the copies and read numbers with the time of seed digestion by carabid beetles was assessed via the non-parametric Pearson correlation, because the data did not meet the assumptions for linear regression, even after data transformation. We used the Kruskal–Wallis test to test whether there were significant differences in copy numbers (and read) between different time points after feeding. To reveal which times differed significantly, we used the pairwise.wilcox function with the BH (Benjamini–Hochberg) correction, at the significance level of $\alpha = 0.05$.

To study the correlation among the different quantitative measurements of the DNA fragment, we used all available samples, regardless of the numbers of seeds consumed but did not use the set of samples with “assumed 0 s”, as in the analysis over time. We assessed the relationship between initial copy numbers per μ l DNA extract and reads or the signal strengths measured via capillary electrophoresis for end-point PCR-products (RFU-values) with linear regression models. Therefore, we used the log-transformed copy numbers (natural log (copies + 1)) as the response variable and reads or RFU-values as the predictors. Model assumptions were tested through visual inspection of residual plots and by applying the ‘ols_test_normality’-function of the package olsrr⁷² to test for normality and the function ‘ncvTest’ (Non-Constant Error Variance or Breusch–Pagan test) included in the car package⁷³ to test for heteroscedasticity. If the assumption of linearity was not met, RFU-values were log-transformed as well. For all models, we calculated robust standard errors with ‘coefest’ in the lmtest package⁷⁴ combined with the function ‘vcovHC’ of the sandwich package^{75,76} and bootstrapped confidence intervals, using the functions ‘boot’ and ‘boot.ci’ of the boot package^{77,78}.

Data availability

Original data of droplet digital PCRs, celPCRs and high-throughput sequencing are available at <http://doi.org/10.17605/OSF.IO/C49N3>.

Received: 14 July 2022; Accepted: 30 January 2023

Published online: 03 February 2023

References

- de Sousa, L. L., Silva, S. M. & Xavier, R. DNA metabarcoding in diet studies: Unveiling ecological aspects in aquatic and terrestrial ecosystems. *Environ. DNA* **1**, 199–214. <https://doi.org/10.1002/edn3.27> (2019).
- Pompanon, F. *et al.* Who is eating what: Diet assessment using next generation sequencing. *Mol. Ecol.* **21**, 1931–1950. <https://doi.org/10.1111/j.1365-294X.2011.05403.x> (2012).
- Liu, M. X., Clarke, L. J., Baker, S. C., Jordan, G. J. & Burridge, C. P. A practical guide to DNA metabarcoding for entomological ecologists. *Ecol. Entomol.* **45**, 373–385. <https://doi.org/10.1111/een.12831> (2020).
- Traugott, M., Thalinger, B., Wallinger, C. & Sint, D. Fish as predators and prey: DNA-based assessment of their role in food webs. *J. Fish Biol.* **98**, 367–382. <https://doi.org/10.1111/jfb.14400> (2021).
- Clare, E. L. Molecular detection of trophic interactions: Emerging trends, distinct advantages, significant considerations and conservation applications. *Evol. Appl.* **7**, 1144–1157. <https://doi.org/10.1111/eva.12225> (2014).
- Deagle, B. E. *et al.* Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data?. *Mol. Ecol.* **28**, 391–406. <https://doi.org/10.1111/mec.14734> (2019).
- Ando, H. *et al.* Methodological trends and perspectives of animal dietary studies by noninvasive fecal DNA metabarcoding. *Environ. DNA* **2**, 391–406. <https://doi.org/10.1002/edn3.117> (2020).
- Masonick, P., Hernandez, M. & Weirauch, C. No guts, no glory: Gut content metabarcoding unveils the diet of a flower-associated coastal sage scrub predator. *Ecosphere*. <https://doi.org/10.1002/ecs2.2712> (2019).
- Eitzinger, B. *et al.* Assessing changes in arthropod predator–prey interactions through DNA-based gut content analysis-variable environment, stable diet. *Mol. Ecol.* **28**, 266–280. <https://doi.org/10.1111/mec.14872> (2019).
- Kim, T. N. *et al.* Using high-throughput amplicon sequencing to determine diet of generalist lady beetles in agricultural landscapes. *Biol. Control*. <https://doi.org/10.1016/j.biocontrol.2022.104920> (2022).
- Wallinger, C. *et al.* The effect of plant identity and the level of plant decay on molecular gut content analysis in a herbivorous soil insect. *Mol. Ecol. Resour.* **13**, 75–83. <https://doi.org/10.1111/1755-0998.12032> (2013).

12. Seabra, S. G. *et al.* PCR-based detection of prey DNA in the gut contents of the tiger-fly, *Coenosia attenuata* (Diptera: Muscidae), a biological control agent in Mediterranean greenhouses. *Eur. J. Entomol.* **118**, 335–343. <https://doi.org/10.14411/eje.2021.035> (2021).
13. Panni, S. & Pizzolotto, R. Fast molecular assay to detect the rate of decay of *Bactrocera oleae* (Diptera: Tephritidae) DNA in *Pterostichus melas* (Coleoptera: Carabidae) gut contents. *Appl. Entomol. Zool.* **53**, 425–431. <https://doi.org/10.1007/s13355-018-0564-x> (2018).
14. Greenstone, M. H., Payton, M. E., Weber, D. C. & Simmons, A. M. The detectability half-life in arthropod predator–prey research: What it is, why we need it, how to measure it, and how to use it. *Mol. Ecol.* **23**, 3799–3813. <https://doi.org/10.1111/mec.12552> (2014).
15. Fülöp, D., Szita, E., Gerstenbrand, R., Tholt, G. & Samu, F. Consuming alternative prey does not influence the DNA detectability half-life of pest prey in spider gut contents. *PeerJ* <https://doi.org/10.7717/peerj.7680> (2019).
16. Zhang, G. F., Lu, Z. C., Wan, F. H. & Lovei, G. L. Real-time PCR quantification of *Bemisia tabaci* (Homoptera: Aleyrodidae) B-biotype remains in predator guts. *Mol. Ecol. Notes* **7**, 947–954. <https://doi.org/10.1111/j.1471-8286.2007.01819.x> (2007).
17. Weber, D. C. & Lundgren, J. G. Detection of predation using qPCR: Effect of prey quantity, elapsed time, chaser diet, and sample preservation on detectable quantity of prey DNA. *J. Insect Sci.* <https://doi.org/10.1673/031.009.4101> (2009).
18. Paula, D. P. *et al.* Detection and decay rates of prey and prey symbionts in the gut of a predator through metagenomics. *Mol. Ecol. Resour.* **15**, 880–892. <https://doi.org/10.1111/1755-0998.12364> (2015).
19. Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83**, 8604–8610. <https://doi.org/10.1021/ac202028g> (2011).
20. Wood, S. A. *et al.* A comparison of droplet digital polymerase chain reaction (PCR), quantitative PCR and metabarcoding for species-specific detection in environmental DNA. *Mol. Ecol. Resour.* **19**, 1407–1419. <https://doi.org/10.1111/1755-0998.13055> (2019).
21. Nathan, L. M., Simmons, M., Wegleitner, B. J., Jerde, C. L. & Mahon, A. R. Quantifying environmental DNA signals for aquatic invasive species across multiple detection platforms. *Environ. Sci. Technol.* **48**, 12800–12806. <https://doi.org/10.1021/es5034052> (2014).
22. Kim, T. G., Jeong, S. Y. & Cho, K. S. Comparison of droplet digital PCR and quantitative real-time PCR for examining population dynamics of bacteria in soil. *Appl. Microbiol. Biotechnol.* **98**, 6105–6113. <https://doi.org/10.1007/s00253-014-5794-4> (2014).
23. Thalinger, B., Pütz, Y. & Traugott, M. Endpoint PCR coupled with capillary electrophoresis (celPCR) provides sensitive and quantitative measures of environmental DNA in singleplex and multiplex reactions. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0254356> (2021).
24. Mata, V. A. *et al.* How much is enough? Effects of technical and biological replication on metabarcoding dietary analysis. *Mol. Ecol.* **28**, 165–175. <https://doi.org/10.1111/mec.14779> (2019).
25. Sint, D., Guenay, Y., Mayer, R., Traugott, M. & Wallinger, C. The effect of plant identity and mixed feeding on the detection of seed DNA in regurgitates of carabid beetles. *Ecol. Evol.* **8**, 10834–10846. <https://doi.org/10.1002/ece3.4536> (2018).
26. Nielsen, J. M., Clare, E. L., Hayden, B., Brett, M. T. & Kratina, P. Diet tracing in ecology: Method comparison and selection. *Methods Ecol. Evol.* **9**, 278–291. <https://doi.org/10.1111/2041-210x.12869> (2018).
27. Schrader, C., Schielke, A., Ellerbroek, L. & John, R. PCR inhibitors—Occurrence, properties and removal. *J. Appl. Microbiol.* **113**, 1014–1026. <https://doi.org/10.1111/j.1365-2672.2012.05384.x> (2012).
28. Juen, A. & Traugott, M. Amplification facilitators and multiplex PCR: Tools to overcome PCR-inhibition in DNA-gut-content analysis of soil-living invertebrates. *Soil Biol. Biochem.* **38**, 1872–1879. <https://doi.org/10.1016/j.soilbio.2005.11.034> (2006).
29. Wallinger, C. *et al.* Evaluation of an automated protocol for efficient and reliable DNA extraction of dietary samples. *Ecol. Evol.* **7**, 6382–6389. <https://doi.org/10.1002/ece3.3197> (2017).
30. Marotz, C. *et al.* DNA extraction for streamlined metagenomics of diverse environmental samples. *Biotechniques* **62**, 290–293. <https://doi.org/10.2144/000114559> (2017).
31. Dingle, T. C., Sedlak, R. H., Cook, L. & Jerome, K. R. Tolerance of droplet-digital PCR vs real-time quantitative PCR to inhibitory substances. *Clin. Chem.* **59**, 1670–1672. <https://doi.org/10.1373/clinchem.2013.211045> (2013).
32. Racki, N., Dreo, T., Gutierrez-Aguirre, I., Blejec, A. & Ravnar, M. Reverse transcriptase droplet digital PCR shows high resilience to PCR inhibitors from plant, soil and water samples. *Plant Methods* <https://doi.org/10.1186/s13007-014-0042-6> (2014).
33. Juen, A. & Traugott, M. Detecting predation and scavenging by DNA gut-content analysis: A case study using a soil insect predator-prey system. *Oecologia* **142**, 344–352. <https://doi.org/10.1007/s00442-004-1736-7> (2005).
34. Lundgren, J. G. & Lehman, M. Bacterial gut symbionts contribute to seed digestion in an omnivorous beetle. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0010831> (2010).
35. Waldner, T. & Traugott, M. DNA-based analysis of regurgitates: A noninvasive approach to examine the diet of invertebrate consumers. *Mol. Ecol. Resour.* **12**, 669–675. <https://doi.org/10.1111/j.1755-0998.2012.03135.x> (2012).
36. Kamenova, S. *et al.* Comparing three types of dietary samples for prey DNA decay in an insect generalist predator. *Mol. Ecol. Resour.* **18**, 966–973. <https://doi.org/10.1111/1755-0998.12775> (2018).
37. Cheeseman, M. T. & Pritchard, G. Spatial organization of digestive processes in an adult carabid beetle, *Scaphinotus marginatus* (Coleoptera: Carabidae). *Can. J. Zool.* **62**, 1200–1203. <https://doi.org/10.1139/z84-173> (1984).
38. Sunderland, K. D. Diet of some predatory arthropods in cereal crops. *J. Appl. Ecol.* **12**, 507–515. <https://doi.org/10.2307/2402171> (1975).
39. Sunderland, K. D., Lovei, G. L. & Fenlon, J. Diets and reproductive phenologies of the introduced ground beetles *Harpalus affinis* and *Clivina australasiae* (Coleoptera: Carabidae) in New Zealand. *Aust. J. Zool.* **43**, 39–50. <https://doi.org/10.1071/zo9950039> (1995).
40. Deagle, B. E. & Tollit, D. J. Quantitative analysis of prey DNA in pinniped faeces: Potential to estimate diet composition?. *Conserv. Genet.* **8**, 743–747. <https://doi.org/10.1007/s10592-006-9197-7> (2007).
41. Snider, A. M., Bonisoli-Alquati, A., Perez-Umphrey, A. A., Stouffer, P. C. & Taylor, S. S. Metabarcoding of stomach contents and fecal samples provide similar insights about Seaside Sparrow diet. *Ornithol. Appl.* <https://doi.org/10.1093/ornithapp/duab060> (2022).
42. Paula, D. P., Timbo, R. V., Togawa, R. C., Vogler, A. P. & Andow, D. A. Quantitative prey species detection in predator guts across multiple trophic levels by mapping unassembled shotgun reads. *Mol. Ecol. Resour.* **23**, 64–80. <https://doi.org/10.1111/1755-0998.13690> (2023).
43. Elbrecht, V. & Leese, F. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0130324> (2015).
44. Krehenwinkel, H. *et al.* Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Sci. Rep.* **7**, 17668. <https://doi.org/10.1038/s41598-017-17333-x> (2017).
45. Piñol, J., San Andrés, V., Clare, E. L., Mir, G. & Symondson, W. O. C. A pragmatic approach to the analysis of diets of generalist predators: The use of next-generation sequencing with no blocking probes. *Mol. Ecol. Resour.* **14**, 18–26. <https://doi.org/10.1111/1755-0998.12156> (2014).
46. Baksay, S. *et al.* Experimental quantification of pollen with DNA metabarcoding using ITS1 and trnL. *Sci. Rep.* <https://doi.org/10.1038/s41598-020-61198-6> (2020).

47. Valentini, A. *et al.* New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: The trnL approach. *Mol. Ecol. Resour.* **9**, 51–60. <https://doi.org/10.1111/j.1755-0998.2008.02352.x> (2009).
48. Murray, D. C. *et al.* DNA-based faecal dietary analysis: A comparison of qPCR and high throughput sequencing approaches. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0025776> (2011).
49. Hansen, B. K. *et al.* From DNA to biomass: Opportunities and challenges in species quantification of bulk fisheries products. *ICES J. Mar. Sci.* **77**, 2557–2566. <https://doi.org/10.1093/icesjms/fsaa115> (2020).
50. Pawluczuk, M. *et al.* Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. *Anal. Bioanal. Chem.* **407**, 1841–1848. <https://doi.org/10.1007/s00216-014-8435-y> (2015).
51. Piñol, J., Mir, G., Gomez-Polo, P. & Agusti, N. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Mol. Ecol. Resour.* **15**, 819–830. <https://doi.org/10.1111/1755-0998.12355> (2015).
52. Czernik, M. *et al.* Fast and efficient DNA-based method for winter diet analysis from stools of three cervids: Moose, red deer, and roe deer. *Acta Theriol.* **58**, 379–386. <https://doi.org/10.1007/s13364-013-0146-9> (2013).
53. Richardson, R. T. *et al.* Quantitative multi-locus metabarcoding and waggle dance interpretation reveal honey bee spring foraging patterns in Midwest agroecosystems. *Mol. Ecol.* **28**, 686–697. <https://doi.org/10.1111/mec.14975> (2019).
54. Taberlet, P. *et al.* Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkl938> (2007).
55. Briem, F. *et al.* Identifying plant DNA in the sponging-feeding insect pest *Drosophila suzukii*. *J. Pest. Sci.* **91**, 985–994. <https://doi.org/10.1007/s10340-018-0963-3> (2018).
56. Frei, B., Guenay, Y., Bohan, D. A., Traugott, M. & Wallinger, C. Molecular analysis indicates high levels of carabid weed seed consumption in cereal fields across Central Europe. *J. Pest. Sci.* <https://doi.org/10.1007/s10340-019-01109-5> (2019).
57. Luff, M. L. The biology of the ground beetle *Harpalus rufipes* in a strawberry field in Northumberland. *Ann. Appl. Biol.* **94**, 153–164. <https://doi.org/10.1111/j.1744-7348.1980.tb03907.x> (1980).
58. Illumina. Effects of index Misassignment on multiplexing and downstream analysis. <https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf?linkId=36607862> accessed 2022-11-10 (2018).
59. Guenay-Greunke, Y., Bohan, D. A., Traugott, M. & Wallinger, C. Handling of targeted amplicon sequencing data focusing on index hopping and demultiplexing using a nested metabarcoding approach in ecology. *Sci. Rep.* <https://doi.org/10.1038/s41598-021-98018-4> (2021).
60. Staudacher, K., Wallinger, C., Schallhart, N. & Traugott, M. Detecting ingested plant DNA in soil-living insect larvae. *Soil Biol. Biochem.* **43**, 346–350. <https://doi.org/10.1016/j.soilbio.2010.10.022> (2011).
61. Espunyès, J. *et al.* Comparing the accuracy of PCR-capillary electrophoresis and cuticle microhistological analysis for assessing diet composition in ungulates: A case study with Pyrenean chamois. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0216345> (2019).
62. Wallinger, C. *et al.* Detection of seed DNA in regurgitates of granivorous carabid beetles. *Bull. Entomol. Res.* **105**, 728–735. <https://doi.org/10.1017/s000748531500067x> (2015).
63. Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. Universal primers for amplification of 3 noncoding regions of chloroplast DNA. *Plant Mol. Biol.* **17**, 1105–1109. <https://doi.org/10.1007/bf00037152> (1991).
64. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. (2010).
65. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* <https://doi.org/10.1093/gigascience/giab008> (2021).
66. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J. Next Gener. Seq. Data Anal.* <https://doi.org/10.14806/ej.17.1.200> (2011).
67. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461> (2010).
68. Camacho, C. *et al.* BLAST plus: Architecture and applications. *BMC Bioinform.* <https://doi.org/10.1186/1471-2105-10-421> (2009).
69. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2018).
70. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2016).
71. Arnold, J. B. ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. R package version 4.2.0 <https://CRAN.R-project.org/package=ggthemes> (2019).
72. Hebbali, A. olsrr: Tools for Building OLS Regression Models. R package version 0.5.3. <https://CRAN.R-project.org/package=olsrr> (2020).
73. Fox, J. & Weisberg, S. *An {R} Companion to Applied Regression*. 2nd ed. (Sage, 2011).
74. Zeileis, A. & Hothorn, T. Diagnostic checking in regression relationships. *R News* **2**, 7–10 (2002).
75. Zeileis, A. Econometric computing with HC and HAC covariance matrix estimators. *J. Stat. Softw.* **11**, 1–17. <https://doi.org/10.18637/jss.v011.i10> (2004).
76. Zeileis, A., Köll, S. & Graham, N. Various versatile variances: An object-oriented implementation of clustered covariances in {R}. *J. Stat. Softw.* **95**, 1–36. <https://doi.org/10.18637/jss.v095.i01> (2020).
77. boot: Bootstrap R (S-Plus) Functions v. R package version 1.3-28 (2021).
78. Davison, A. C. & Hinkley, D. V. *Bootstrap Methods and Their Applications*. (Cambridge University Press, 1997).

Acknowledgements

We thank Corinna Wallinger for providing the samples analysed in the present study, the FACCE ERA-NET C-IPM project BioAWARE "Could Biodiversity Assure Weed regulAtion for Resilient Ecosystem service provision?" for financially supporting VN, and the University of Innsbruck for supporting this work through their publishing fund. The computational results presented here have been achieved in part using the LEO HPC infrastructure of the University of Innsbruck.

Author contributions

M.T. and V.N. conceived the ideas and designed methodology; V.N. collected and analysed the data and led the writing of the manuscript. Both authors contributed critically to the drafts and gave final approval for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-29019-8>.

Correspondence and requests for materials should be addressed to V.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023