Jun 05, 2020

Amplicon library protocol for metabarcoding-based diet analysis in blue tits (Cyanistes caeruleus) using faecal DNA

DOI
dx.doi.org/10.17504/protocols.io.2jdgci6

James Nicholls[1], James Nicholls[2]

[1]CSIRO; [2]University of Edinburgh

James Nicholls
CSIRO

OPEN ACCESS

DOI: https://dx.doi.org/10.17504/protocols.io.2jdgci6

Protocol Citation: James Nicholls, James Nicholls 2020. Amplicon library protocol for metabarcoding-based diet analysis in blue tits (Cyanistes caeruleus) using faecal DNA. **protocols.io https://dx.doi.org/10.17504/protocols.io.2jdgci6**

Manuscript citation:
Shutt, J.D., Nicholls, J.A., Trivedi, U.H., Burgess, M.D., Stone, G.N., Hadfield, J.D. & Phillimore, A.B. (2020) Gradients in richness and turnover of a forest passerine's diet prior to breeding: a mixed model approach applied to faecal metabarcoding data.Molecular Ecology 29: 1199-1213.
Shutt, Nicholls et al. (2020) Mol. Ecol.

Protocol status: Working

Created: May 02, 2019

Last Modified: June 05, 2020

Protocol Integer ID: 22853

Keywords: metabarcoding, amplicon library, next-generation sequencing, faeces, faecal dna amplicon library protocol for metabarcoding, using faecal dna amplicon library protocol, invertebrate dna, targeting invertebrate dna, avian dna in order, using faecal dna, amplifying avian dna, faecal dna, diet analysis in blue tit, particular invertebrate, barcoding gene, second mitochondrial gene, amplicon, plastid gene, plant dna, cyanistes caeruleus, starting dna, 16s rrna, sequencing, dna

Abstract

Amplicon library protocol for metabarcoding-based diet analysis in blue tits (*Cyanistes caeruleus*) using faecal DNA.
This methodology uses a two-stage PCR method to produce amplicon libraries for sequencing on an Illumina MiSeq. Amplicons are required to be short (roughly 225-260 base pairs of target sequence, including priming sites) as the starting DNA, obtained from faeces, is highly degraded. Primers and PCR conditions are presented for three loci. These loci are:
a. cytochrome oxidase subunit I (COI) - the 5′ end of the Folmer region, the standard animal barcoding gene. Targets animal (in particular invertebrate) DNA.
b. 16S rRNA - a second mitochondrial gene, again targeting invertebrate DNA but also amplifying avian DNA in order to identify the species producing the faeces.
c. rbcL - a plastid gene, designed to target plant DNA.

Materials

MATERIALS

ssODN (Ultramer DNA Oligonucleotides) **Integrated DNA Technologies, Inc. (IDT)**

Herculase II Fusion polymerase **Agilent Technologies Catalog #**600679

SequalPrep Normalization Plate Kit 96-well **Thermo Fisher Scientific Catalog #**A1051001

Troubleshooting

Overview of methodology

1 Amplicons were produced using a two stage PCR. The initial PCR used locus-specific primers with 5′ tails containing part of either the Illumina Nextera P5 or P7 adaptor sequence. Reagent concentrations, annealing temperatures and number of PCR cycles varied by locus (see table below). This first PCR was done in duplicate for each sample and locus, to help control for random variation in what template DNA amplifies within the species mix extracted from the faeces. The two duplicates per sample/locus were pooled and mixed well before an aliquot was removed for use as template in the second stage PCR.
The second PCR used primers containing the remainder of the respective P5/P7 Nextera adaptor including an 8 base pair index (following the published i5 and i7 indices used in the Nextera XT kit). This indexing was done separately for each locus-by-sample-plate combination. The dual-indexing PCR allows amplicons from multiple samples to be multiplexed post-amplification and sequenced on the same MiSeq run. For the blue tit study up to 280 samples were multiplexed on a single MiSeq run, using 18 different i7 indices and 16 different i5 indices. This indexing PCR used the same conditions for all three loci, with the exception of the number of cycles (see table below). Amplicons derived from different loci but the same original faecal sample were labelled with the same index combination.
All PCRs were performed in 96-well plates. Each plate contained 94 samples, a positive control (containing a mix of DNA from a known insect species and known plant species) and a negative control (see below for more detail on controls). An aliquot of the positive control, negative control and a small subset of samples were run on a 2% agarose gel after the second round PCR was complete; if the positive failed or the negative contained product then PCRs were repeated for the whole plate.

Primer sets used for first round of PCR

2 **COI:**
This protocol uses the forward primer **LepF1** (found by Brandon-Mong *et al.* 2015 to be one of the best of multiple mini-barcode forward primers) with a modified version of the reverse primer **ZBJ-ArtR2c**. This amplifies a target region of 178bp in most invertebrate taxa; 227bp including the locus-specific priming sites.
Details:
LepF1, used as is from Hebert *et al.* (2004).
sequence: 5′ ATT CAA CCA ATC ATA AAG ATA TTG G 3′
ZBJ-ArtR2c-deg, a modified version of the primer ZBJ-ArtR2c presented in Zeale *et al.* (2011). Modifications, introduced for this study, involve degeneracy at third codon positions towards the 3′ end of the primer using data from Clarke *et al.* (2014) and Piñol *et al.* (2014).

sequence: 5′ WAC TAA TCA ATT WCC AAA HCC HCC 3′

16S:
This study uses a novel combination of published 16S primers, specifically targeting invertebrate 16S, using the forward primer **16S1F-deg** and reverse primer **Ins16S_1R**. In theory the primer 16S1F-deg shouldn't work in birds but in practice it does amplify blue tit DNA. This primer combination amplifies a fragment approx. 218bp in insects (254bp including priming sites); the fragment is roughly 90bp shorter in spiders. In blue tits the expected size is 268bp (304bp with priming sites).
Details:
16S1F-deg, used as is from Deagle *et al.* (2007).
sequence: 5′ GAC GAK AAG ACC CTA 3′
Ins16S_1R, used as is from Clarke *et al.* (2014).
sequence: 5′ TCT TAA TCC AAC ATC GAG GTC 3′

rbcL:
This study uses a combination of published primers targeting the plastid gene rbcL, using the forward primer **rbcL1** and reverse primer **rbcLB**. These primers are both originally from Palmieri *et al.* (2009), and assessed as useful minibarcode primers by Little (2014). They amplify a fragment 184bp long (226bp with locus-specific primers).
Details:
rbcL1, used as is from Palmieri *et al.* (2009).
sequence: 5′ TTG GCA GCA TTY CGA GTA ACT CC 3′
rbcLB, used as is from Palmieri *et al.* (2009).
sequence: 5′ AAC CYT CTT CAA AAA GGT C 3′

All locus-specific forward primers have a 5′ tail, which comprises part of the standard Illumina P5 adapter used for Nextera library preps. All locus-specific reverse primers have a 5′ tail which comprises part of the standard Illumina P7 adapter used for Nextera library preps. These 5′ tail sequences are as follows:

LepF1 has the 5′ tail TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
ZBJ-ArtR2c-deg has the 5′ tail GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

16S1F-deg has the 5′ tail TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Ins16S_1R has the 5′ tail GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

rbcL1 has the 5′ tail TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
rbcLB has the 5′ tail GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

Primer sets used for second round of PCR

3 Indexing primers for the second PCR are:
P5 adaptor primer:
AATGATACGGCGACCACCGAGATCTACAC<i5>TCGTCGGCAGCGTC
P7 adaptor primer:
CAAGCAGAAGACGGCATACGAGAT<i7>GTCTCGTGGGCTCGG

Different P7 and P5 primers were made up using an 8bp index substituted for the "<i5>" or "<i7>" in the above sequences. Indices used were taken from the published set used within the Nextera XT library prep kit.

All primers were synthesised by IDT, using their Ultramer synthesis option with HPLC purification.

PCR conditions

4 The Herculase II Fusion polymerase (Agilent catalogue number 600679) was used for all PCRs, with 0.1µL of polymerase and 1µL of DNA extraction in a 10µL reaction. For the second indexing PCR 1µL of template from the pool of the two duplicate first round PCRs was used in a 10µL reaction. Final concentrations of other reagents are:

		Mg^{2+}(mM)	dNTPs (mM of each)	each primer (µM)	BSA (mg/mL)
	rbcL	2.5	0.2	0.2	0.5
	16S	2.5	0.2	0.2	0.5
	COI	2	0.2	0.2	0.5
	indexing PCR	2	0.2	0.2	0.5

Thermocycler conditions were as follows:
a. initial denaturation of 3 minutes at 94°C
b. cycling of:
30 seconds at 94°C
30 seconds at annealing temperature
40 seconds at 72°C
c. final extension of 5 minutes at 72°C
d. hold at 10°C

Annealing temperatures and the number of cycles were as follows for each locus-specific PCR and the second-round indexing PCR.

		Ta (°C)	no. 1st round cycles	no. 2nd round (indexing) cycles
	rbcL	56	25	20
	16S	54	25	10
	COI	51	40	10
	indexing PCR	63	-	-

Reactions included a 40 second extension time, which is longer than would be normally required given the length of the target amplicons using the Herculase II Fusion polymerase, in order to reduce the chance of chimera formation (following the recommendation of Lenz & Becker 2008). A slow ramping rate (1°C per second) was also used to reduce chimera formation (following Stevens *et al.* 2013).

Post-PCR clean-up, pooling and sequencing

5 After the second round PCR, samples within each locus-by-sample-plate combination were cleaned up (removing salts, unincorporated primers, and any possible PCR dimer) and normalised to a constant concentration using SequalPrep Normalisation plates (ThermoFisher catalogue number A1051001). These plates have been found to provide an effective and high-throughput method for normalisation (see Harris *et al.* 2010). The manufacturer's protocol was followed, using the sequential elution option with same aliquot of buffer used across four samples to facilitate pooling (i.e. same 20µL of elution buffer used to elute 4 wells). All buffer aliquots used to elute samples from a single plate were then pooled into one tube, giving one tube of eluted amplicon library for each locus-by-sample-plate combination.
An estimate of the molarity of each of these locus-by-sample-plate combinations was generated via quantification using a Qubit and the estimate of average amplicon length (363bp for COI, 440bp for 16S, 362bp for rbcL). Each locus-by-sample-plate combination was diluted (or concentrated using a speedy-vac) to 10nM.
Sequencing pools were constructed by taking equal volumes from 9 locus-by-sample-plate combinations (i.e. three plates of samples each for three loci). For the blue tit experiment this equated to amplicons from three loci derived from between 275 and 278 faecal samples, inclusive of multiple controls and replicates. Three such pools were produced to accommodate all 793 samples, 30 replicates and 24 controls (9 x PCR

positives, 9 x PCR negatives and 6 x extraction negatives). Amplicons within each pool were sequenced on an Illumina MiSeq, using 150 bp paired-end reads.

Experimental controls

6 Control samples were introduced at various stages of the molecular work. Six different negative controls were introduced at regular intervals when performing the DNA extractions (using all the same reagents as samples, but with no faeces added). These six extraction negatives were carried through the remainder of the molecular and informatics methods, to provide indication of any contaminants that may have been introduced during the molecular lab processes. A separate negative control was also included in each sample plate (n = 9) containing pure water in place of DNA extract, as was a positive control containing a mix of template DNA from one known species of insect (*Dryocosmus israeli*) and one known species of plant (*Inga pezizifera*), neither of which occurred in the same country as the blue tits were sampled from. As mentioned above, these PCR negatives and PCR positives and a small subset of samples were run on agarose gels before the PCR plate was taken through to the next stage of the protocol; the PCR was repeated for the whole plate if either the negative contained any evidence of an amplicon band or the positive lacked a band. These PCR negatives and positives were also carried through the informatics pipeline defining molecular operation taxonomic units. As the positive control contained known species, it additionally acted as a control to confirm that sample indexing (at the lab stage) and de-multiplexing of samples (at the informatics stage) had been performed correctly.

References

7 Brandon-Mong, G.-J., Gan, H.-M., Sing, K.-W., Lee, P.-S., Lim, P.-E. & Wilson, J.-J. (2015). DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin of Entomological Research* **105**: 717-727.

Clarke, L.J., Soubrier, J., Weyrich, L.S. & Cooper, A. (2014). Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources* **14**: 1160-1170.

Deagle, B.E., Gales, N.J., Evans, K., Jarman, S.N., Robinson, S., Trebilco, R. & Hindell, M.A. (2007). Studying seabird diet through genetic analysis of faeces: a case study on Macaroni penguins (*Eudyptes chrysolophus*). *PLoS One* **2**: e831.

Harris, J.K., Sahl, J.W., Castoe, T.A., Wagner, B.D., Pollock, D.D. & Spear, J.R. (2010). Comparison of normalization methods for construction of large, multiplex amplicon pools for next-generation sequencing. *Applied and Environmental Microbiology* **76**: 3863-3868.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 14812-14817.

Lenz, T.L. & Becker, S. (2008). Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci - implications for evolutionary analysis. *Gene* **427**: 117-123.

Little, D.P. (2014). A DNA mini-barcode for land plants. *Molecular Ecology Resources* **14**: 437-446.

Palmieri, L., Bozza, E. & Giongo, L. (2009). Soft fruit traceability in food matrices using real-time PCR. *Nutrients* **1**: 316-328.

Piñol, J., San Andrés, V., Clare, E.L., Mir, G. & Symondson, W.O.C. (2014). A pragmatic approach to the analysis of diets of generalist predators: the use of next-generation sequencing with no blocking probes. *Molecular Ecology Resources* **14**: 18-26.

Stevens, J.L., Jackson, R.L. & Olson, J.B. (2013). Slowing PCR ramp speed reduces chimera formation from environmental samples. *Journal of Microbiological Methods* **93**: 203-205.

Zeale, M.R.K., Butlin, R.K., Barker, G.L.A., Lees, D.C. & Jones, G. (2011). Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources* **11**: 236-244.