MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2017) **17**, 904–914 doi: 10.1111/1755-0998.12643

eDNA metabarcoding: a promising method for anuran surveys in highly diverse tropical forests

CARLA M. LOPES,* THAIS SASSO,† ALICE VALENTINI,‡ TONY DEJEAN,‡ MARCIO MARTINS,† KELLY R. ZAMUDIO§ and CÉLIO F. B. HADDAD*

*Departamento de Zoologia, Instituto de Biociências, Universidade Estadual Paulista "Júlio de Mesquita Filho", Av. 24A, n. 1515, Rio Claro, SP 13506-900, Brazil, †Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP 05508-090, Brazil, ‡SPYGEN, Savoie Technolac, BP 274, Le Bourget-du-Lac 73375, France, §Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

Abstract

Understanding the geographical distribution and community composition of species is crucial to monitor species persistence and define effective conservation strategies. Environmental DNA (eDNA) has emerged as a powerful noninvasive tool for species detection. However, most eDNA survey methods have been developed and applied in temperate zones. We tested the feasibility of using eDNA to survey anurans in tropical streams in the Brazilian Atlantic forest and compared the results with short-term visual and audio surveys. We detected all nine species known to inhabit our focal streams with one single visit for eDNA sampling. We found a higher proportion of sequence reads and larger number of positive PCR replicates for more common species and for those with life cycles closely associated with the streams, factors that may contribute to increased release of DNA in the water. However, less common species were also detected in eDNA samples, demonstrating the detection power of this method. Filtering larger volumes of water resulted in a higher probability of detection. Our data also show it is important to sample multiple sites along streams, particularly for detection of target species with lower population densities. For the three focal species in our study, the eDNA metabarcoding method had a greater capacity of detection per sampling event than our rapid field surveys, and thus, has the potential to circumvent some of the challenges associated with traditional approaches. Our results underscore the utility of eDNA metabarcoding as an efficient method to survey anuran species in tropical streams of the highly biodiverse Brazilian Atlantic forest.

Keywords: 12S mitochondrial gene, amphibians, biodiversity, monitoring methods, next-generation sequencing, site occupancy–detection modelling

Received 11 June 2016; revision received 19 November 2016; accepted 6 December 2016

Introduction

Detecting species in the environment provides critical data on species' distributions, community composition and the presence of endangered or invasive species, data that are required for designing effective conservation strategies. Surveys of amphibian species are typically based on the use of pitfall traps, visual encounters and audio surveys (Dodd 2010). However, these traditional methods differ in efficacy depending on observer experience as well as specific traits of target species, life stages, seasonality, meteorological conditions, population densities and the complexity of the environment (Goldberg *et al.* 2011; Dejean *et al.* 2012). Reducing overall variance in detection probability is an important goal for the success of any monitoring programme.

Environmental DNA (eDNA) metabarcoding is a noninvasive and powerful approach that helps overcome many of the challenges of traditional survey methods for detection of aquatic species. The identification of species presence through DNA traces left behind by organisms in the environment has been increasingly applied in both still and running water systems, to detect a single focal species (Ficetola *et al.* 2008), describe community composition (Valentini *et al.* 2016) and to monitor invasive (Jerde *et al.* 2011; Dejean *et al.* 2012; Tréguier *et al.* 2014) or endangered (Thomsen *et al.* 2012) species for conservation purposes, even when these are at low population densities (Thomsen & Willerslev 2015).

Despite the growing applications of eDNA metabarcoding, to date, most aquatic eDNA survey studies have

Correspondence: Carla M. Lopes,
E-mail: cmlopes82@hotmail.com

been performed in temperate zones (Hoffmann *et al.* 2016), where environmental conditions may favour the persistence of DNA traces in the field (Dejean *et al.* 2011; Barnes *et al.* 2014) compared to tropical environments. Tropical regions typically have high species richness. Brazil, for example, has the highest amphibian diversity in the world, with over 1000 species reported occurring in the country (Segalla *et al.* 2014), and currently an average of 15 new species are described each year (Frost 2016). The Brazilian Atlantic forest harbours over 500 of those amphibian species (7% of all known species), 88% of which are endemic to these forests (Myers *et al.* 2000; Haddad *et al.* 2013). In contrast to most temperate amphibian communities, the species diversity of any tropical site is typically not fully known, and often the taxonomy of species complexes is unresolved, posing a significant challenge for eDNA taxonomic assignments that depend on a reference database. In megadiverse tropical regions such as the Brazilian Atlantic forest, successful results with eDNA metabarcoding will require molecular markers that are capable of detecting highly degraded DNA fragments from environmental samples, but that are also sufficiently variable to distinguish a high diversity of species occupying the same environment. In addition, eDNA surveys are prone to imperfect detection of species, and thus, appropriate site occupancy–detection modelling is important for proper eDNA-based detection surveys (Schmidt *et al.* 2013). All these issues need to be carefully considered during laboratory procedures, bioinformatic pipeline development and interpretation of eDNA data (Goldberg *et al.* 2016).

In this study, we applied eDNA metabarcoding in a rapid inventory of tropical stream anurans in the Brazilian Atlantic forest. We chose streams in which the anuran fauna is well known, and chose three common focal species with which to compare detection by eDNA and simultaneous short-term visual and audio surveys of the stream reaches at the time of sampling. Our specific goals were to (i) test the feasibility of eDNA metabarcoding for anuran detection in a tropical stream environment, (ii) assess the performance of eDNA for overall detection of abundant vs. locally rare species, (iii) compare the performance of rapid, single-day surveys using eDNA sampling vs. short-term visual and audio inventories of species presence and (iv) estimate the effect of filtering water samples of different volumes on species detection probability.

Material and methods

Area of study

We sampled water for eDNA metabarcoding analyses from four Atlantic coastal forest streams in the Núcleo Picinguaba in the Parque Estadual da Serra do Mar, São Paulo state, Brazil (Fig. 1). Núcleo Picinguaba ranges from sea level to 1340 m. The general climate is tropical wet, with virtually no dry season and average annual temperature of 22 °C. Annual rainfall is over 2200 mm, with monthly means ranging from >200 mm from October to April, to >80 mm from March to September (Morellato *et al.* 2000; Joly *et al.* 2012).

From 22 April 2015 to 25 April 2015, we performed single-day visual encounter surveys (VESs) together with audio strip transects (ASTs), and water sampling for eDNA metabarcoding in four streams (hereafter ST1, ST2, ST3 and ST4). Sampling was performed at the end of wet season (which runs from November through March), when reproductive peak of anuran species occurring in Picinguaba begins to decline (Hartmann 2004). Sampled streams encompass the variation found in second-order streams in the study area, including width, declivity, water flow and structure. All streams belonged to different drainages and traversed well-preserved primary forest. The two most distant streams (ST1 and ST4) were 3.6 km apart.

Forty four anuran species have been recorded in the Núcleo Picinguaba, of those, nine species (*Aplastodiscus eugenioi*, *Bokermannohyla* sp. aff. *circumdata*, *Cycloramphus boraceiensis*, *Hylodes asper*, *Hylodes phyllodes*, *Phasmahyla cruzi*, *Scinax trapicheiroi*, *Thoropa taophora*, and *Vitreorana uranoscopa*) are known to inhabit our focal streams, at least at some point during their life cycles (egg deposition, larval development or adult microhabitat preferences; Table S1, Supporting information).

Physicochemical and geographical variables

We measured water temperature, pH and dissolved oxygen (mg/L), and took geographical coordinates and elevation measurements for each sampling site at the time of eDNA sampling (Table S2, Supporting information).

Short-term visual and audio surveys

Visual encounter surveys (VESs) and audio strip transects (ASTs) were performed together using the methodology described in Heyer *et al.* (1994) for comparison with the single time point eDNA sampling results. We chose three species as focal taxa: *Hylodes phyllodes*, *Hylodes asper* and *Cycloramphus boraceiensis*. These species were chosen because they are common in the sampled streams; that is the detection probability is high, but the presence/absence and relative abundances of these species vary among the streams (Lopes 2016). All three species complete their entire life cycle (egg deposition, larval development and postmetamorphic stages) in or near streams, and are easy to identify in the field. *Hylodes*

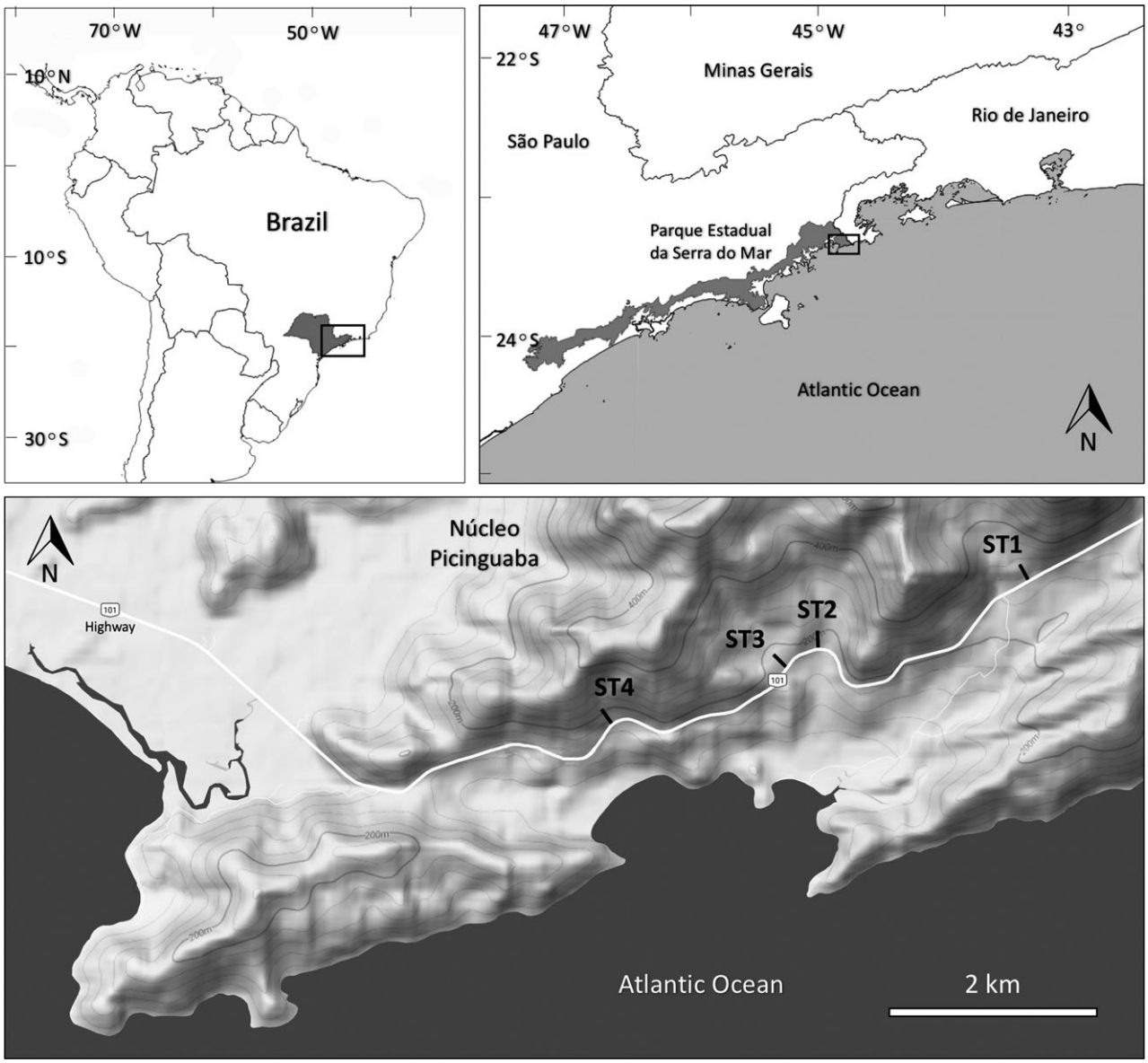

Fig. 1 Streams of Núcleo Picinguaba sampled using eDNA and short-term visual and audio surveys. ST1, Stream 1; ST2, Stream 2; ST3, Stream 3; ST4, Stream 4. Sampling sites S1 and S2 are in the extremes of the transects represented for each stream (S1 closer to the highway and S2 in the opposite side of the transect). The map was extracted from Google Maps and slightly modified.

phyllodes and *H. asper* are active during daytime, and *C. boraceiensis* is nocturnal (Haddad *et al*. 2013).

Visual and audio surveys (VES and AST) were simultaneously performed by three observers, once during daytime and once at night for each stream. We performed these short-term surveys after eDNA sampling to avoid disturbance and contamination of the water. A transect of 95–115 m was defined along each stream to search for the focal species. We searched for postmetamorphic individuals while walking slowly upstream, during 30–60 min, checking all visually accessible spots in the streambed.

eDNA sampling

Water sampling for eDNA analyses was performed at two sites along each stream transect, approximately 100 m apart (Fig. 1). We sampled each locality twice, filtering one sample of 20 L and one sample of 60 L of water, to compare the power of detection depending on the volume of water analysed. Filtering was carried out with a peristaltic pump model 410 (Solinst Canada Ltd., Georgetown, Ontario, Canada) and Envirochek HV® 1 μm sampling capsules (Pall Corporation, Port Washington, NY, USA). Water was filtered directly from

streams, applying a debit of 1.6 L/min. Filters were filled with 150 mL of lysis buffer (Tris-HCl 0.1 M, EDTA 0.1 M, NaCl 0.01 M and N-lauroyl sarcosine 1%, pH 7.5–8). To certify that there was no contamination of equipment or cross-contamination among streams, negative sampling controls were performed for each stream, filtering 5 L of distilled water in the field.

Laboratory procedures

The samples were processed in two steps. First, we extracted, amplified, purified, constructed the library and sequenced the DNA of two samples from each stream to verify our methods, and then we applied the same protocols to the remaining samples. For DNA extraction, filtration capsules were left at 56 °C for 2 h, agitated manually for 5 min and then emptied into 50-mL tubes. In total, approximately 120 mL was retrieved in three tubes that were centrifuged for 15 min at 15 000 *g*. Supernatant was removed with a sterile pipette, leaving 15 mL of liquid. Subsequently, 33 mL of ethanol and 1.5 mL of 3 M sodium acetate were added to each 50-mL tube. The tubes were stored at −20 °C for at least 2 h. The three tubes per site were centrifuged at 15 000 *g*, for 15 min at 6 °C, and the supernatant was discarded. After this step, 360 μL of ATL Buffer of the DNeasy Blood & Tissue Extraction Kit (Qiagen GmbH, Hilden, Germany) was added in the first tube, the tube was vortexed, and the supernatant was transferred to the second tube. This step was repeated for all tubes. The supernatant from the third tube was transferred to a 2-mL tube, and the DNA extraction was performed following the manufacturer's instructions. To monitor possible contaminations, one negative extraction control was performed, and the control extract was amplified and sequenced along with the field samples.

DNA amplifications of a short fragment of the 12S rRNA mitochondrial gene were performed in a final volume of 25 μL, using 3 μL of DNA extract as template. The amplification mixture contained 1 U of AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, CA, USA), 10 mM of Tris-HCl, 50 mM of KCl, 2.5 mM of MgCl2, 0.2 mM of each dNTP, 0.2 μM of batra_F (5′-ACACCGCCCGTCACCCT-3′) and batra_R (5′-GTAYACTTACCATGTTACGACTT-3′) primers (Valentini *et al.* 2016), 4 μM of human blocking primers batra_blk (5′-TCACCCTCCTCAAGTATACTTCAAAGGCA-SPC3I-3′; Valentini *et al.* 2016) and 0.2 μg/μL of bovine serum albumin (BSA, Roche Diagnostic, Basel, Switzerland). The primers were 5′ labelled with a unique eight-nucleotide tag (with at least three differences between tags) allowing the assignment of sequences to the respective samples during the sequence filtering process. Tags for forward and reverse primers were identical for each PCR replicate. The PCR mixture was denatured at 95 °C for 10 min, followed by 50 cycles of 30 s at 95 °C, 30 s at 55 °C and 1 min at 72 °C, followed by a final elongation at 72 °C for 7 min. The PCR was run in 12 replicates per sample. One negative PCR control (ultrapure water, with 12 replicates as well) was analysed in parallel to the samples to monitor possible contaminations. Amplifications were done in a dedicated room, with negative air pressure, and physically separated from the DNA extraction rooms.

Amplified samples were titrated using capillary electrophoresis (QIAxcel; Qiagen GmbH, Hilden, Germany) and purified using a MinElute PCR purification kit (Qiagen GmbH, Hilden, Germany). Before sequencing, purified DNA was titrated again using capillary electrophoresis. The purified PCR products were pooled in equal volumes, to achieve an expected sequencing depth of 300 000 reads per sample. Library preparation and sequencing were performed at Fasteris facilities (Geneva, Switzerland). Two libraries, a preliminary one to test our methods and determine coverage, and a final library with remaining samples, were prepared using the Metafast protocol (https://www.fasteris.com/dna/?q=content/metafast-protocol-amplicon-metagenomic-analysis). The paired-end sequencing (2 × 125 bp) was carried out in two Illumina HiSeq 2500 runs (Illumina, San Diego, CA, USA) using the HISEQ SBS Kit version 4 (Illumina, San Diego, CA, USA) following the manufacturer's instructions.

Local reference database

To improve taxonomic assignment of potential amphibian DNA traces found in our samples, we assembled a local amphibian reference database of the 12S rRNA mitochondrial gene. Tissue samples were obtained from the Célio F. B. Haddad amphibian tissue collection (CFBHT), at Universidade Estadual Paulista (UNESP), Rio Claro, São Paulo, Brazil. The reference database included one to three individuals of species known to occur within Núcleo Picinguaba (Table S1, Supporting information). Whenever possible, we chose tissues from Núcleo Picinguaba. When not available, we chose tissues collected at geographically close locations. Total DNA was extracted from 10 mg of muscle tissue, following a modified protocol from Sambrook & Russel (2001).

Reference sequences were amplified using the primers MVZ59 (5′-ATAGCACGTAAAAYGCTDAGATG-3′; Graybeal 1997); tVal (5′-TGTAAGCGARAGGCTTTKGTTAAGCT-3′; Wiens *et al.* 2005); 12SA-L (5′-AAACTGGGATTAGATACCCCACTAT-3′; Palumbi *et al.* 1991); 12SF-H (5′-CTTGGCTCGTAGTTCCCTGGCG-3′; Goebel *et al.* 1999); H978 and 148 (ML Lyra, unpublished data), following the protocols described in Faivovich *et al.*

(2004). PCR products were purified using Exonuclease I and Shrimp Alkaline Phosphatase (Thermo Fischer Scientific, Waltham, MA, USA), following the guidelines of the suppliers. Both DNA strands were sequenced at Macrogen Inc. (Seoul, South Korea).

Sequences were visually inspected and edited, and consensus sequences were constructed using CODONCODE ALIGNER 5.1.5 (CodonCode Corporation, Centerville, MA, USA).

Bioinformatics and statistical analysis

The eDNA sequence reads were filtered and annotated using the programs implemented in the OBITOOLS package (Boyer *et al.* 2016). Direct and reverse strands were assembled to construct consensus sequences, using the *illuminapairedend* program. Sequences unambiguously identified by their molecular tags (no error allowed) and primers (2 bp errors per primer allowed) were assigned to PCR products, applying the *ngsfilter* program. Strictly identical sequences were clustered together, keeping the information of their read counts per PCR product (*obiuniq* program), and only sequences longer than 20 bp or total read counts higher than 10 were kept for the subsequent steps of the analyses (*obigrep* program). The *obiclean* program was applied to label each sequence as 'head', 'internal' or 'singleton' (Shehzad *et al.* 2012) in each PCR product, to identify possible amplification/sequencing errors. We extracted the relevant part of 12S sequences from our local reference database and from anuran species from the release *vrt*124 of EMBL database, using the ECOPCR (Ficetola *et al.* 2010) and OBITOOLS programs. This 12S reference database was used to assign species identity to the sequences retrieved in eDNA samples (program *ecotag*).

Following the filtering process, we used the software R 3.2.4 Revised (R Development Core Team 2016) to remove PCR and sequencing errors. Low-frequency sequence reads (<0.1%) and sequences classified as internal were excluded from the data set per PCR product. We compared the profile of negative controls and stream samples to exclude low-quantity DNA PCR products (<150 sequence reads in total) and potential sequences from cross-contamination sources (frequency <0.3% per sample per Molecular Operational Taxonomic Unit – MOTU). Sequences with less than 96% of identity with a sequence from the reference database were removed from the analyses.

All statistical analyses were performed using R software. PCR replicates were separated in categories, based on: (i) stream of origin (ST1, ST2, ST3 and ST4); (ii) the site sampled within each stream (S1 and S2); and (iii) the volume of water filtered (20 or 60 L). The proportion of sequence reads for each taxon at each category was calculated based on the sum of read counts among PCR replicates.

We applied the site occupancy–detection model of Royle & Link (2006), which accounts for the possibility of false positives, to compare the estimated probability of detection and site occupancy for our three focal species (*H. phyllodes*, *H. asper* and *C. boraceiensis*) retrieved in samples of 20 and 60 L of filtered water. The other species were not included in site occupancy analyses because they were detected only a few times and thus did not provide sufficient power for statistical analysis. Two matrices of presence/absence, one for 20 L samples and another for 60 L samples, were constructed for each focal species separately, considering the eight sampling sites together (two sites per stream) and the 12 PCR replicates for each sample. We used the JAGS program (Plummer 2003) in the package R2jags (Su & Yajima 2015), to apply Bayesian inference, considering the prior of maximum probability of false presences as 0.05, running four chains of 100 000 iterations, 50 000 as burn-in, and saving 1000 iterations per chain. For more details, see Lahoz-Monfort *et al.* (2015).

To test for a correlation between the number of frogs observed by VES and AST methods (a proxy for abundance) and the number of sequence reads, we performed simple linear regressions based on the number of individuals of each focal species per stream (independent variable), and the log10-transformed sum of sequence read counts for each species per stream (dependent variable). Regressions were performed separately for day and night surveys, and for 20 and 60 L water samples.

Results

Physicochemical and geographical variables

We observed a slight variation in physicochemical measures among distinct sites and streams sampled. Water temperatures ranged from 20.4 to 20.9 °C ($\bar{x}$ = 20.7 °C; s = 0.16), dissolved oxygen in water from 4.43 to 7.23 mg/L ($\bar{x}$ = 5.48; s = 0.85) and pH from 6.5 to 7.0 ($\bar{x}$ = 6.72; s = 0.22). Geographical and physicochemical characteristics of sampling sites in the studied streams are described in Table S2 (Supporting information).

Short-term visual and audio surveys

For ST1, the species *C. boraceiensis* and *H. asper* were recorded during short-term VES and AST. In ST2, only *H. phyllodes* was observed, and for ST3 and ST4 the three focal species were observed (Table 1). The species most frequently found and in highest numbers across all streams was *H. asper*, followed by *C. boraceiensis* and *H. phyllodes*.

Table 1 Number of individuals identified in four Picinguaba streams by means of short-term visual encounter surveys and audio strip transects

	Cycloramphus boraceiensis	*Hylodes phyllodes*	*Hylodes asper*
ST1			
Day	—	—	20
Night	7	—	4
ST2			
Day	—	9	—
Night	—	4	—
ST3			
Day	—	2	9
Night	8	2	6
ST4			
Day	—	5	16
Night	28	—	1
Total	43	22	56

ST1, Stream 1; ST2, Stream 2; ST3, Stream 3; ST4, Stream 4; S1, Site 1; S2, Site 2.

eDNA survey

Of the 70 specimens sequenced for the 12S mitochondrial gene, 48 unique sequences were obtained for the local reference database, representing 36 species. The length of the fragment used for metabarcoding analyses ranged from 50 to 54 bp. From the 44 species recorded in Núcleo Picinguaba (Hartmann 2004), we were able to obtain 12S rRNA sequences for 40 species (sequences of 36 species generated by us, and sequences of four species retrieved from GenBank). The only species for which we could not obtain tissues or sequences were *Brachycephalus hermogenesi*, *Brachycephalus vertebralis*, *Myersiella microps* and *Chiasmocleis atlantica*. None of these are stream-dwelling species; thus, we proceeded with the reference database that did not include these sequences.

A total of 10 104 512 sequence reads were obtained before the filtering process among all samples and negative controls. After the filtering pipeline, the sampling, extraction and PCR negative controls were completely clean and no sequence reads remained in those samples. From the initial 192 PCR replicates of samples, no sequence read remained in 52 PCR replicates at the end of filtering process, including all PCR replicates from Stream 3, Site 2, 20 L (ST3_S2_20). We are not able to state whether this resulted from the DNA preservation procedure, poor amplification in PCR replicates, or whether it is a true nondetection of species. For the remaining PCR replicates, a total of 1 286 149 sequence reads were retained after the filtering process, corresponding to an average of 9186 reads per PCR replicate.

In the final eDNA data set, we detected 17 sequences, corresponding to 11 taxa: *Aplastodiscus eugenioi*, *Bokermannohyla*, *Bokermannohyla* sp. aff. *circumdata*, *Cycloramphus boraceiensis*, *Hylodes asper*, *Hylodes phyllodes*, *Phasmahyla cruzi*, *Scinax ruber*, *Scinax trapicheiroi*, *Thoropa taophora* and *Vitreorana uranoscopa*. Only one sequence was not identified to the species level, corresponding to the genus *Bokermannohyla*. It had 100% of identity with two sequences deposited in GenBank (AY843673 – *Bokermannohyla hylax* and AY549338 – *Bokermannohyla* sp.). The species *Thoropa taophora* was initially identified by the *ecotag* program only to the genus *Thoropa* for the eDNA data. However, it is known that some sequences in GenBank are identified as *Thoropa miliaris* and not up to date with recent taxonomic revisions (Feio *et al*. 2006) in which the *Thoropa* species from Picinguaba was assigned to *T. taophora*. Therefore, we assigned those sequences to *T. taophora*. The same happened with the species *Phasmahyla cruzi*, which was identified only to genus, because the GenBank sequence AY843716 was identified as *P. guttata*, but only one species of *Phasmahyla* occurs in the region, which is currently identified as *P. cruzi*.

One sequence was assigned to the taxon *Scinax ruber* (GenBank reference sequence Accession no. AY819447), comprising 47 reads found in only one PCR replicate. This species is distributed in the Amazon basin, and it is not known to occur in the region of Picinguaba. We considered that laboratory contamination was an unlikely source to explain the presence of this species in our sample, because of the strict protocols we followed, the fact that our samples were analysed at SPYGEN (France), where no samples containing *Scinax* were ever analysed, and because none of our negative controls showed any evidence of this species. However, it is also biologically highly unlikely that this Amazonian species is present at our site, because in addition to its distribution, *S. ruber* does not breed in densely canopied streams. The issue of minimum detection thresholds is currently an active field of research, and at this point, there are no universal standards in place for detecting false positives. However, researchers in this new field agree that inference of species presence should be based on the strength of evidence from various data sources, including the frequency and consistency of positive eDNA samples, what is known about species' distributions, habitats and behaviours (Goldberg *et al.* 2016). Therefore, given the preponderance of evidence in this case, we excluded this species from downstream analyses, although the origin of this sequence remains uncertain.

The species *H. phyllodes* was detected in all samples and represented the highest proportion of sequence reads, except for ST1_S1_20. The proportion of sequence reads for this species ranged from 100% in samples ST2_S1_20, ST2_S2_60 and ST2_S2_20 to 33.40% in ST1_S1_20 (Fig. 2). The species *H. asper* and

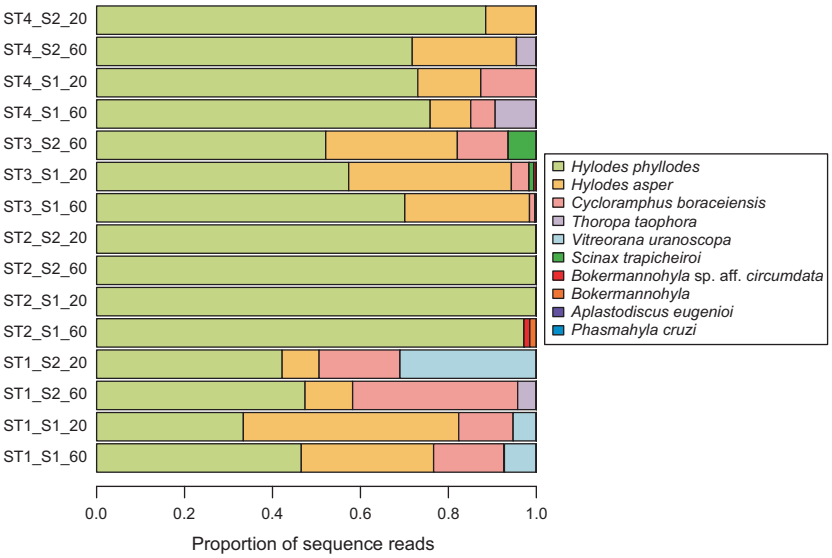

Fig. 2 Proportion of sequence reads for taxa observed in eDNA samples of the studied streams of Núcleo Picinguaba. ST1, Stream 1; ST2, Stream 2; ST3, Stream 3; ST4, Stream 4; S1, Site 1; S2, Site 2; 20, 20 L; 60, 60 L.

C. boraceiensis also showed high proportions of sequence reads among samples (0 – 48.95% and 0 – 37.54%, respectively). On the other hand, sequences of the genus *Bokermannohyla*, and the species *Aplastodiscus eugenioi* and *Phasmahyla cruzi* were observed in only one sample each, in low frequency (1.46% – ST2_S1_60; 0.29% – ST3_S1_60; and 0.09% – ST4_S1_60, respectively; Fig. 2).

In general, a higher number of sequence reads and more positive PCR replicates were recovered in 60 L filtered samples compared to 20 L filtered samples, both for each species retrieved and for total values among all species. Exceptions were observed for *V. uranoscopa*, which was detected in two PCR replicates of ST1_S2_20, but was not detected in ST1_S2_60, and *S. trapicheroi* and *B.* sp. aff. *circumdata* which were detected in ST3_S1_20 but were not detected in ST3_S1_60 (Table S3, Supporting information). The estimated detection probabilities (p11) for *H. phyllodes*, *H. asper* and *C. boraceiensis* were higher for 60 L filtered samples (p11 = 0.761, 0.649 and 0.596, respectively), than for 20 L filtered samples (p11 = 0.614, 0.570 and 0.154, respectively). The estimated proportion of sites occupied (psi) for *H. phyllodes*, *H. asper* and *C. boraceiensis* for 60 L samples were 0.922, 0.603 and 0.406, respectively, and for 20 L were 0.805, 0.402 and 0.409, respectively (Fig. 3).

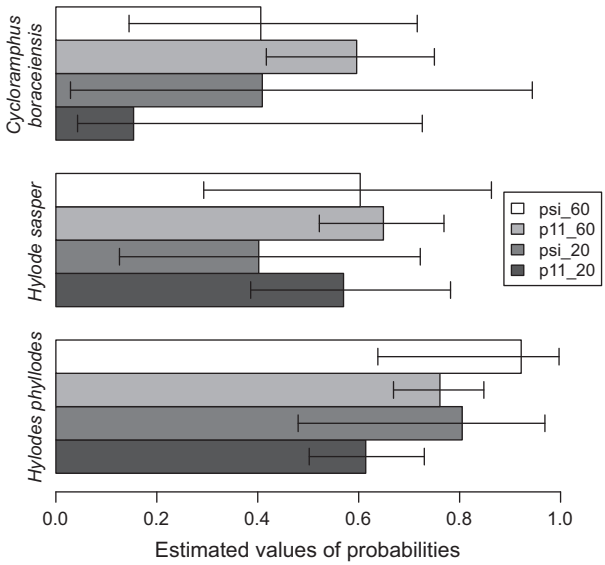

Fig. 3 Estimated values of probability of species detection (p11) and occupancy of sites (psi) for 20 and 60 L water samples, for the three most common species occurring in the studied streams of Núcleo Picinguaba. 95% confidence intervals are shown.

Comparison of methodologies

The eDNA metabarcoding approach detected the species *H. phyllodes* in all four streams and all sampling sites, while the short-term VES and AST failed to detect this species in ST1 (day and night surveys) and during the night in ST4. *Hylodes asper* was detected in all streams except ST2, using both eDNA (all sampling sites) and VES and AST (day and night surveys) methods. *Cycloramphus boraceiensis* was detected by eDNA in all samples of ST1 and ST3, and at S1 of ST4. The short-term VES and AST failed to detect this species during the day in ST1, ST3 and ST4. Both eDNA and VES and AST methodologies did not detect *C. boraceiensis* in ST2.

Despite a slight positive trend of association between the number of individuals observed by short-term VES and AST and the sequence read counts from the eDNA samples, none of the regression analyses were significant: eDNA 20 L × day VES and AST (r^2_{adj} = 0.05906; 95% CI = −0.15560 to 0.59162; $P > 0.05$); eDNA 20 L × night

VES and AST (r^2_{adj} = −0.08441; 95% CI = −0.2979532 to 0.4201404; $P > 0.05$); eDNA 60 L × day VES and AST (r^2_{adj} = 0.02882; CI = −0.1947797 to 0.6115977; $P > 0.05$); and eDNA 60 L × VES and AST (r^2_{adj} = −0.06204; 95% CI = −0.2761722 to 0.4787062; $P > 0.05$).

Discussion

Physicochemical and geographical conditions can influence the detectability of DNA because they affect rates of DNA degradation and dispersion in the environment (Dejean *et al.* 2011; Barnes *et al.* 2014; Pilliod *et al.* 2014; Klymus *et al.* 2015). DNA traces in water degrade quickly and therefore only allow the detection of species that were recently present in the environment; the maximum time that amphibian DNA remained detectable in a controlled experiment was 25 days (Dejean *et al.* 2011). We did not find marked physicochemical or elevation differences among the four streams analysed that could alter detection probabilities in this study.

The final portion of the 12S mitochondrial gene used in the metabarcoding analyses showed high power of taxonomic discrimination, despite the small size of the amplified fragment (~50 bp). Only one sequence, corresponding to the genus *Bokermannohyla*, was not identified to species level, likely because of taxonomic problems of identification of sequences available in GenBank.

We detected all nine species known to be directly associated with the streams we surveyed with one single visit for eDNA sampling. The capacity to detect a species using eDNA varies, among other factors, according to the release rates of DNA by organisms in the environment, which depends on species biomass, number of individuals in the population, activity levels and life stages (Pilliod *et al.* 2014; Goldberg *et al.* 2015; Klymus *et al.* 2015). It is expected that species with greater population sizes will release more DNA in the environment, and consequentially the probability of detection using eDNA increases (Thomsen *et al.* 2012; Pilliod *et al.* 2013). Sequences corresponding to *Hylodes phyllodes* were retrieved in all eDNA samples for the four streams in high proportions. The sequences corresponding to the species *Hylodes asper* and *Cycloramphus boraceiensis* were also commonly recovered in the samples of ST1, ST3 and ST4, and in considerable proportions (Fig. 2 and Table S3, Supporting information). The higher proportions of sequences and positive PCR replicates observed for these three species may reflect the fact that they are the most common species in this system of streams throughout the year (Lopes 2016). Furthermore, their reproductive modes (egg deposition, larval development) and adult microhabitat preferences (associated with high gradient streams) are closely tied to stream water (Haddad *et al.* 2013). Although a single-day eDNA sampling is not ideal for inventory of uncommon species, our eDNA analyses detected even the presence of species less commonly observed in these streams (e.g. *Aplastodiscus eugenioi*, *Phasmahyla cruzi*, *Vitreorana uranoscopa*; Lopes 2016), demonstrating the power of this approach to detect the presence of both abundant but also less frequent species in the environment.

Our results show that for tropical streams, filtering a larger volume of water will, in general, result in a higher capacity to detect species. The estimated probabilities of detection for *H. phyllodes* and *H. asper* were 14.7% and 7.9% higher, respectively, in 60 L samples of filtered water compared to 20 L samples. Difference in detection probability was even higher for *C. boraceiensis* (44.2%), which was also the species with lowest estimates of site occupancy (~41%) among the three species analysed (Fig. 3). However, the recovery of species varied between the two volumes sampled at the same site, and from different sites within the same stream (Table S3, Supporting information). Thus, our results underscore the importance of filtering a considerable volume of water at various sampling sites in streams, particularly for target species with low population densities, to increase the capacity of detection using eDNA samples. Mächler *et al.* (2016), analysing eDNA water samples to search for three species of macroinvertebrates in Glatt River, Switzerland, found significant relationship between sampling volume and detection rate only for one species (*Gammarus pulex*), which is also the species with the lowest levels of detection probability calculated among the three species analysed. The authors concluded that detection rates may vary among species and volume of water filtered, and it should be considered during experimental design of eDNA studies.

For our three focal species, despite a slightly positive trend, we found no significant relationship between the number of individuals recorded in our rapid field inventories and the number of sequence reads retrieved in eDNA samples. These results might be explained by the fact that on more than one occasion the short-term visual encounter survey together with audio strip transect methods failed to detect the presence of *H. phyllodes* and *C. boraceiensis*, whereas these species were positively detected by eDNA. A number of factors underlie the low detection in our short-term VES and AST and the positive detection in the eDNA results. At each stream, we performed only a single VES and AST at each time of the day (one diurnal and one nocturnal), and this sampling effort is clearly not enough to accurately detect all species present in all streams where they occur. Time of day can also influence the results of VES and AST, depending on the time that species are more or less active, but it seems to not affect DNA concentration in the environment (Pilliod *et al.* 2013). Furthermore, during our short-

term surveys, we did not quantify larvae or eggs. One potential advantage of eDNA surveys is the ability to detect alternate life stages (eggs and/or tadpoles) that may persist in streams during longer periods of time, and also contribute to the release of DNA in the environment. However, these life stages are less easily found and identified by traditional methods. The nine species in our streams have their peak of reproduction during the wet season (from November through March; Hartmann 2004), and our sampling was performed at end of this season (April). Thus, our data indicate that eDNA surveys might increase the efficacy of detection of species even outside their reproductive peak, especially in cases of rapid species inventories. Finally, the eDNA signal in a river can come from a few kilometres above the sampling point (Deiner & Altermatt 2014) so that the spatial scale is different between eDNA detection and VES and AST.

For our three focal species (*H. phyllodes*, *H. asper* and *C. boraceiensis*), the eDNA metabarcoding methodology had a greater capacity of species detection in the field than our short-term VES and AST methods. A higher sensitivity of eDNA compared to traditional methods has been observed for amphibian and fish species from both lentic and lotic aquatic systems in other studies. The detection of invasive Asian carp with eDNA methods in Chicago area waterways in the United States demonstrated that the front of invasion was ahead of that previously determined with standard electrofishing surveys, indicating the application of eDNA is highly suitable to large spatial scales and for rapid surveys compared to traditional survey methods (Jerde *et al.* 2011). Likewise, Dejean *et al.* (2012) demonstrated that the eDNA method is more sensitive and reduces the sampling effort for surveys of an invasive frog (*Lithobates catesbeianus*) in southwestern France, compared to traditional surveys based on calling and visual encounters. Similarly, Pilliod *et al.* (2013) compared eDNA detection data with kick-netting data for larvae of two amphibian species (*Ascaphus montanus* and *Dicamptodon aterrimus*) in 13 streams of central Idaho, USA, and showed higher detection rates with eDNA data, especially for *D. aterrimus*, the species that occurred at lower density. Finally, Valentini *et al.* (2016) estimated that at least four successive visits to the field using traditional survey methods are necessary to obtain the same success of detection for amphibians in Mediterranean ponds compared to one single sampling using eDNA metabarcoding. Overall quantitative costs are difficult to compare between traditional and eDNA metabarcoding methodologies, although commonly the latter are still more expensive. However, a number of factors need to be considered to choose which methodology to apply, such as the impact of the survey to the species and the environment, personnel training needed, hours of laboratory and field work, infrastructure and equipment costs, and rates of false-positive/false-negative species detection for each method. In general, the cost-benefit ratio of an eDNA metabarcoding analysis should be more advantageous when surveying highly diverse environments, large sampling areas, when short period of time available, when surveying several species at the same time and/or when surveying species that are difficult to find or identify in the field.

Clearly, eDNA has much to offer in the survey of species and can circumvent some of the challenges associated with traditional methods. However, due to the sensitivity of the eDNA metabarcoding, careful considerations and high standard quality control is required during field work, laboratory procedures and data analyses to avoid false-positive/false-negative data and misinterpretations. Also, for new applications of eDNA, researchers should consider previously established methods and protocols depending on the taxonomic target group, kind of environment studied and the objectives of the study (Goldberg *et al.* 2016). All these points were carefully considered by us during the development of this study.

Our results showed that no major modifications to laboratory protocols were necessary to conduct an eDNA survey in a tropical environment. This is an efficient method to survey amphibian species in tropical streams, despite differences in overall species diversity and abiotic factors between tropical and temperate regions. Environmental DNA metabarcoding is a feasible noninvasive approach to survey several sites, for a wide taxonomical range, on a large spatial scale, in a relatively short time. The potential of this method will only increase as we better understand the factors that alter detection probability in different environments and for species with different traits.

Acknowledgements

This study was funded by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, process #401729/2013-3) and by the São Paulo Research Foundation (FAPESP grant #2013/50741-7). We thank Fábio P. de Sá, Lucas N. Bandeira and Bianca Berneck for help with fieldwork, colleagues from Laboratório de Herpetologia and Gentile F. Ficetola for help with analyses and Aurélie Bonin for reviewing the manuscript. C.F.B.H., M.M. and K.Z. are grateful to CNPq, and C.M.L. is grateful to CNPq and FAPESP (grant #2016/14054-3) for research fellowships.

References

Barnes MA, Turner CR, Jerde CL *et al.* (2014) Environmental conditions influence eDNA persistence in aquatic systems. *Environmental Science & Technology*, **48**, 1819–1827.

Boyer F, Mercier C, Bonin A *et al.* (2016) OBITools: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, **16**, 176–182.

Deiner K, Altermatt F (2014) Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE*, **9**, e88786.

Dejean T, Valentini A, Duparc A *et al.* (2011) Persistence of environmental DNA in freshwater ecosystems. *PLoS ONE*, **6**, e23398.

Dejean T, Valentini A, Miquel C *et al.* (2012) Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. *Journal of Applied Ecology*, **49**, 953–959.

Dodd CKJ (2010) *Amphibian Ecology and Conservation: A Handbook of Techniques* (ed. Dodd CKJ). Oxford University Press, Oxford.

Faivovich J, García PCA, Ananias F *et al.* (2004) A molecular perspective on the phylogeny of the *Hyla pulchella* species group (Anura, Hylidae). *Molecular Phylogenetics and Evolution*, **32**, 938–950.

Feio RN, Napoli MF, Caramaschi U (2006) Considerações taxonômicas sobre *Thoropa miliaris* (Spix, 1824), com revalidação e redescrição de *Thoropa taophora* (Miranda-Ribeiro, 1923) (Amphibia, Anura, Leptodactylidae). *Arquivos do Museu Nacional, Rio de Janeiro*, **64**, 41–60.

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology Letters*, **4**, 423–425.

Ficetola GF, Coissac E, Zundel S *et al.* (2010) An in silico approach for the evaluation of DNA barcodes. *BMC Genomics*, **11**, 434.

Frost D (2016) *Amphibian Species of the World: An Online Reference*. American Museum of Natural History, New York.

Goebel AM, Donnelly JM, Atz ME (1999) PCR primers and amplification methods for 12S ribosomal DNA, the control region, cytochrome oxidase 1, and cytochrome *b* in bufonids and other frogs, and an overview of PCR primers which have amplified DNA in amphibians successfully. *Molecular Phylogenetics and Evolution*, **11**, 163–199.

Goldberg CS, Pilliod DS, Arkle RS, Waits LP (2011) Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. *PLoS ONE*, **6**, e22746.

Goldberg CS, Strickler KM, Pilliod DS (2015) Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biological Conservation*, **183**, 1–3.

Goldberg CS, Turner CR, Deiner K *et al.* (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, **7**, 1299–1307.

Graybeal A (1997) Phylogenetic relationships of bufonid frogs and tests of alternate macroevolutionary hypotheses characterizing their radiation. *Zoological Journal of the Linnean Society*, **119**, 297–338.

Haddad CFB, Toledo LF, Prado CPA *et al.* (2013) *Guia dos Anfíbios da Mata Atlântica – Diversidade e Biologia*. Anolisbooks, São Paulo.

Hartmann MT (2004) Biologia reprodutiva de uma comunidade de anuros (Amphibia) na Mata Atlântica (Picinguaba, Ubatuba, SP). PhD Thesis, Universidade Estadual Paulista "Júlio de Mesquita Filho", Rio Claro.

Heyer WR, Donnelly MA, McDiarmid RW, Hayek L-AC, Foster MS (1994) *Measuring and Monitoring Biological Diversity*. Smithsonian Institution Press, Washington.

Hoffmann C, Schubert G, Calvignac-Spencer S (2016) Aquatic biodiversity assessment for the lazy. *Molecular Ecology*, **25**, 846–848.

Jerde CL, Mahon AR, Chadderton WL, Lodge DM (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, **4**, 150–157.

Joly CA, Assis MA, Bernacci LC *et al.* (2012) Florística e fitossociologia em parcelas permanentes da Mata Atlântica do sudeste do Brasil ao longo de um gradiente altitudinal. *Biota Neotropica*, **12**, 123–145.

Klymus KE, Richter CA, Chapman DC, Paukert C (2015) Quantification of eDNA shedding rates from invasive bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypophthalmichthys molitrix*. *Biological Conservation*, **183**, 77–84.

Lahoz-Monfort JJ, Guillera-Arroita G, Tingley R (2015) Statistical approaches to account for false positive errors in environmental DNA samples. *Molecular Ecology Resources*, **16**, 673–685.

Lopes TS (2016) Spatial and temporal occurrence of stream frogs in the Atlantic forest and their detection through environmental DNA. Master Thesis, Universidade de São Paulo, São Paulo.

Mächler E, Deiner K, Spahn F, Altermatt F (2016) Fishing in the water: effect of sampled water volume on environmental DNA-based detection of macroinvertebrates. *Environmental Science and Technology*, **50**, 305–312.

Morellato LPC, Talora DC, Takahasi A *et al.* (2000) Phenology of Atlantic rain forest trees: a comparative study. *Biotropica*, **32**, 811–823.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Palumbi SR, Martin A, McMillan WO, Stice L, Grabowski G (1991) *The Simple Fool's Guide to PCR*. University of Hawaii, Honolulu.

Pilliod DS, Goldberg CS, Arkle RS, Waits LP, Richardson J (2013) Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, **70**, 1123–1130.

Pilliod DS, Goldberg CS, Arkle RS, Waits LP (2014) Factors influencing detection of eDNA from a stream-dwelling amphibian. *Molecular Ecology Resources*, **14**, 109–116.

Plummer M (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna.

R Development Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Royle JA, Link WA (2006) Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, **87**, 835–841.

Sambrook J, Russel DW (2001) Rapid isolation of yeast DNA. In: *Molecular Cloning, a Laboratory Manual* (eds Sambrook J, Russel DW), pp. 631–632. Cold Spring Harbor, New York, NY.

Schmidt BR, Kéry M, Ursenbacher S, Hyman OJ, Collins JP (2013) Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution*, **4**, 646–653.

Segalla MV, Caramaschi U, Cruz CAG *et al.* (2014) Brazilian amphibians: list of species. *Herpetologia Brasileira*, **3**, 37–48.

Shehzad W, McCarthy TM, Pompanon F *et al.* (2012) Prey preference of snow leopard (*Panthera uncia*) in South Gobi, Mongolia. *PLoS ONE*, **7**, e32104.

Su Y-S, Yajima M (2015) *R2jags: Using R to Run "JAGS."* Available from: https://cran.r-project.org/web/packages/R2jags/index.html (accessed 5 April 2016).

Thomsen PF, Willerslev E (2015) Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, **183**, 4–18.

Thomsen PF, Kielgast J, Iversen LL *et al.* (2012) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, **21**, 2565–2573.

Tréguier A, Paillisson J-M, Dejean T *et al.* (2014) Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. *Journal of Applied Ecology*, **51**, 871–879.

Valentini A, Taberlet P, Miaud C *et al.* (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, **25**, 929–942.

Wiens JJ, Fetzner JW, Parkinson CL, Reeder TW (2005) Hylid frog phylogeny and sampling strategies for speciose clades. *Systematic Biology*, **54**, 778–807.

All authors contributed to research design and discussions; C.M.L., T.S. and M.M. did field sampling; C.M.L. and T.S. did reference database; A.V. did eDNA laboratory procedures; C.M.L. and A.V. analysed the data; C.M.L. wrote the study with the contribution of all coauthors.

Data accessibility

DNA sequences of 12S rRNA mitochondrial gene for the reference database: GenBank Accession nos KY202765–KY202834.

NGS eDNA unfiltered data: Dryad Digital Repository doi:10.5061/dryad.q5s4v.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Anuran species recorded in the Núcleo Picinguaba in the Parque Estadual da Serra do Mar, São Paulo state (SP), Brazil.

Table S2 Geographical and physicochemical characteristics of sampling sites in streams of Núcleo Picinguaba, Parque Estadual da Serra do Mar, São Paulo, Brazil.

Table S3 Summary of number of reads and number of positive PCR replicates for each species recovered in each eDNA sample.