

# Jumping the green wall: the use of PNA-DNA clamps to enhance microbiome sampling depth in wildlife microbiome research

Luis Viquez-R<sup>1</sup>, Ramona Fleischer<sup>1</sup>, Kerstin Wilhelm<sup>1</sup>, Marco Tschapka<sup>1</sup>, and Simone Sommer<sup>1</sup>

<sup>1</sup>University of Ulm

June 22, 2020

## Abstract

As microbiome research moves away from model organisms to wildlife, new challenges for microbiome high throughput sequencing arise caused by the variety of wildlife diets. High levels of contamination are commonly observed emanating from the host (mitochondria) or diet (chloroplast). Such high contamination levels affect the overall sequencing depth of wildlife samples thus decreasing statistical power and leading to poor performance in downstream analysis. We developed an amplification protocol utilizing PNA-DNA clamps to maximize the use of resources and to increase the sampling depth of true microbiome sequences in samples with high levels of plastid contamination. We chose two study organisms, a bat (*Leptonycteris yerbabuenae*) and a bird (*Mimus parvulus*), both relying on heavy plant-based diets that sometimes lead to traces of plant-based faecal material producing high contamination signals from chloroplasts and mitochondria. On average, our protocol yielded a 13-fold increase in bacterial sequence amplification compared with the standard protocol (Earth Microbiome Protocol) used in wildlife research. For both focal species, we were able significantly to increase the percentage of sequences available for downstream analyses after the filtering of plastids and mitochondria. Our study presents the first results obtained by using PNA-DNA clamps to block the PCR amplification of chloroplast and mitochondrial DNA from the diet in the gut microbiome of wildlife. The method involves a cost-effective molecular technique instead of the filtering out of unwanted sequencing reads. As 33% and 26% of birds and bats, respectively, have a plant-based diet, the tool that we present here will optimize the sequencing and analysis of wild microbiomes.

## Jumping the green wall: the use of PNA-DNA clamps to enhance microbiome sampling depth in wildlife microbiome research

Luis Viquez-R<sup>1\*</sup>, Ramona Fleischer<sup>1</sup>, Kerstin Wilhelm<sup>1</sup>, Marco Tschapka<sup>1,2</sup>, and Simone Sommer<sup>1</sup>

1. Institute for Evolutionary Ecology and Conservation Genomics, University of Ulm, Ulm, Germany;

2. Smithsonian Tropical Research Institute, Balboa, Ancón, Panamá

Running Headline: PNA-DNA clamps for wildlife microbiome research

Correspondence author: Luis Viquez-R

Correspondence author email: [luis.viquez@alumni.uni-ulm.de](mailto:luis.viquez@alumni.uni-ulm.de)

## Abstract

As microbiome research moves away from model organisms to wildlife, new challenges for microbiome high throughput sequencing arise caused by the variety of wildlife diets. High levels of contamination are commonly observed emanating from the host (mitochondria) or diet (chloroplast). Such high contamination levels affect the overall sequencing depth of wildlife samples thus decreasing statistical power and leading to poor performance in downstream analysis.

We developed an amplification protocol utilizing PNA-DNA clamps to maximize the use of resources and to increase the sampling depth of true microbiome sequences in samples with high levels of plastid contamination. We chose two study organisms, a bat (*Leptonycteris yerbabuenae*) and a bird (*Mimus parvulus*), both relying on heavy plant-based diets that sometimes lead to traces of plant-based faecal material producing high contamination signals from chloroplasts and mitochondria.

On average, our protocol yielded a 13-fold increase in bacterial sequence amplification compared with the standard protocol (Earth Microbiome Protocol) used in wildlife research. For both focal species, we were able significantly to increase the percentage of sequences available for downstream analyses after the filtering of plastids and mitochondria.

Our study presents the first results obtained by using PNA-DNA clamps to block the PCR amplification of chloroplast and mitochondrial DNA from the diet in the gut microbiome of wildlife. The method involves a cost-effective molecular technique instead of the filtering out of unwanted sequencing reads. As 33% and 26% of birds and bats, respectively, have a plant-based diet, the tool that we present here will optimize the sequencing and analysis of wild microbiomes.

## Keywords

wildlife microbiome, chloroplast, mitochondria, 16s, PNAs, PCR blockers

## Introduction

A new world of research opportunities has emerged with the advancement of sequencing techniques. One of the fields that have benefited most is the study of whole microbial communities, so-called microbiomes. This novel method allows the study of bacterial mats and gut microbiome linings without the need to cultivate each bacterium separately (Caporaso *et al.* 2012). Together with recently developed and improved bioinformatic pipelines (Mothur, QIIME 2, etc.), we now have the means to classify and assign taxonomy with a reasonable level of confidence (Bolyen *et al.* 2019).

As microbiome research moves away from model organisms and extends into natural settings, new challenges of wildlife research and those arising because of the variety of wildlife diets need to be tackled. One of the drawbacks is that the separation of bacterial sequences from non-bacterial signals, i.e. those from mitochondria (from the host) and chloroplasts (from the diet) can sometimes be tricky (Barott *et al.* 2011; Lundberg *et al.* 2013). According to the widely accepted endosymbiosis theory (Margulis [then known as Sagan (1967), Gray 2017], mitochondria and chloroplasts were originally derived during evolution from hijacked bacteria engulfed by other bacteria. Because of this bacterial origin, some DNA sequences of organelles are strikingly bacteria-like. This is also the case with reads obtained from high throughput sequencing of 16S rRNA genes, the usual target of microbiome studies. In the worst case, the resulting read coverage consists of many reads assigned to mitochondria or chloroplasts.

Several ways are available to circumvent this problem; the most common path is to increase the sequencing depth and then filter out the reads assigned to the organelles. However, this technique results in an expensive price tag for sequencing and may lead to highly skewed read numbers depending on the provenance of the samples. Another option has recently arisen: the use of DNA-PNA clamps as PCR blockers to prevent the amplification of the specific mitochondrial or chloroplast sequences (Fitzpatrick *et al.* 2018). PNAs (peptide nucleic acids) are DNA-mimicking molecules with outstanding hybridization properties (Nielsen & Egholm 1999). The backbone of the molecules is constructed of N-(2-amino-ethyl) glycol (AEG) instead of the sugar-phosphate backbone of DNA (Nielsen, Egholm & Buchardt 1994). The nucleobases attached to this backbone are the same as those in DNA, thereby allowing hybridization between the probe and the bacterial DNA. PNAs are thus a powerful molecular tool in microbiome research for dealing with samples with a high content of either host or plant remnants in faecal pellets (Lundberg *et al.* 2013).

In this study, we tested the PNA-DNA clamps as a method for improving microbiome discovery rates in bats (Tequila bat *Leptonycteris yerbabuenae*) and Galapagos mockingbirds (*Mimus parvulus*). We chose these two study organisms because they both rely on heavy plant-based diets that sometimes can lead to masses

of plant-based faecal material producing high contamination signals from chloroplast and mitochondria. Our study presents the first results obtained by using PNA-DNA clamps to block the PCR amplification of chloroplast and mitochondrial DNA from the diet in the gut microbiome of wildlife. The method involves a cost-effective molecular technique, instead of the filtering out of the unwanted sequencing reads.

## Material and methods

### *Sample collection*

In 2018, we netted Tequila bats (*L. yerbabuenae*) while they were returning from a night's foraging trip. A mist net was positioned at the entrance of the roosting cave located in the Pinacate and Gran Desierto de Altar Biosphere Reserve in Northern Sonora, Mexico. Bats were immediately removed from the net and kept in a soft cloth bag until processed (< 60 minutes). Animals were handled following guidelines from the ASM for animal care (Sikes *et al.* 2016) and local regulations (Permit Number: SGPA/DGVS/06361/17). A single faecal pellet was collected from the cloth bag and preserved in a safe-lock 1.5-ml Eppendorf tube containing 500 µl DNA/RNA shield (Zymo Research Europe GmbH, Germany). The tube was shaken to ensure the maximum impregnation of the sample with the buffer and then stored in a cool place until it could be frozen at -20°C. For the present study, we used samples from eight randomly chosen individuals.

Mockingbird (*M. parvulus*) individuals were captured between 2007 and 2008 at various sites across the Galapagos Islands. Birds were trapped by using mist nets or potter traps. Faecal pellets from the birds were collected in ethanol and stored at -20°C. Further details about the capturing procedure are given in Hoeckel *et al.* (2010) and Štefka *et al.* (2011). In the present study, we used samples from ten randomly chosen individuals inhabiting the islands of Santiago, Santa Cruz, and Marchena (Fleischer *et al.*., in review).

### *DNA extraction*

We extracted the faecal pellets by using the NucleoSpin® Soil extraction kit (Macherey-Nagel, Düren, Germany) following the manufacturer's guidelines. For the Tequila bat samples, we homogenized the sample (2 x 150 sec at 50 Hz) by using a SpeedMill PLUS (Analytik Jena, Jena, Germany). To maximize DNA yield, we conducted consecutive elutions (2 x 50 µl) with a preheated (ca. 45°C) SE buffer. For the mockingbird samples, the samples were washed in 50 µl SE Buffer and then homogenized using the same procedure as with the Tequila bat samples. We stored the extracted DNA at -20°C.

### *PNA-DNA clamp design*

The probes in our study were designed based on the work of Lundberg *et al.* (2013) who developed PNA-DNA clamps to block mitochondrial (mPNA) and chloroplast (pPNA) DNA, these clamps are known as universal clamps. Recently, Fitzpatrick *et al.* (2018) reported that the universal pPNA showed a mismatch in six plant lineages by means of an experimental and bioinformatic analysis. Preliminary results from our study showed that the plant contamination material in our bat samples belonged to one of these lineages, namely Saguaro Columnar cacti (Cactaceae: *Carnegiea gigantea*). Therefore, following the recommendations of Fitzpatrick *et al.* (2018), we developed a special clamp for the bat samples (Fig 2). This is not a trivial matter, since even a single base mismatch between the chloroplast sequence and the clamp can increase levels of plastid contamination in the sequencing output (Fitzpatrick *et al.* 2018). For the mockingbird, the universal clamps were used to block both chloroplast and mitochondrial DNA. All clamps were ordered from PNA Bio (Newbury Park, USA).

### *DNA amplification, library preparation and sequencing*

To investigate the gut microbiomes of bats (n=8) and birds (n=10), we followed the earth microbiome protocol (Caporaso *et al.* 2010). Moreover, we added four samples consisting of a ZymoBIOMICS microbial community standard D6300 (Zymo Research Europe, Freiburg, Germany). These were used as positive controls for microbiome amplification and allowed us to examine whether the clamps had any effect over the yield of a normal sample depleted of chloroplast and mitochondria. The extracted DNA was amplified with the universal bacterial primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-

GGACTACHVGGGTWTCTAAT-3'). We used a two-step amplification process following the amplicon tagging scheme of Fluidigm (Access Array System for Illumina Sequencing Systems, (c)Fluidigm, San Francisco, USA). In the first step, we amplified a 291 bp fragment of the hypervariable V4 region of the 16S rRNA gene by using tagged (CS) target specific (TS) primers: CS1-NNNN-TS-515F and CS2-TS-806R. We added four random bases to our forward primers to facilitate cluster identification during the first cycles on the Illumina Miseq System. In the second step, the tags (CS1 and CS2) were used to add a sample-specific 10 bp barcode and the Illumina system adapters.

The initial 15  $\mu$ l PCR volume contained 1.5  $\mu$ l (5-15 ng) extracted DNA, 7.5  $\mu$ l DNA polymerase AmpliTaq Gold 360 Master Mix (Applied Biosystems, Darmstadt, Germany), 1.5  $\mu$ l (0.2  $\mu$ M) primers and 4.5  $\mu$ l sterile water. The PCR protocol consisted of an initial activation step at 95°C for 10 minutes, followed by 30 cycles at 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 45 seconds, and a final elongation at 72°C for 10 minutes. When clamps were implemented, the water volume was reduced to 1.5  $\mu$ l; the 1.5  $\mu$ l from each clamp (mPNA and either cpPNA or pPNA) was added to this first step to give a final concentration of 1  $\mu$ M.

The modified PCR protocol included a step in order to allow the binding of the PNA to the target sequences (Fig 3). For the second barcoding PCR (20  $\mu$ l), we used 3  $\mu$ l initial PCR product, 10  $\mu$ l AmpliTaq Gold 360 Master Mix, 4  $\mu$ l (0.4  $\mu$ M) barcode primers (Fluidigm) and 3  $\mu$ l sterile water. PCR conditions were the same as before, but only 10 cycles were performed.

We used the NucleoMag® NGS Clean-Up and Size Select Kit (Macherey-Nagel, Düren, Germany) on a GeneTheatre® (Analytik Jena, Jena, Germany) to clean the PCR products according to the manufacturer's guidelines. We assessed the quality of the amplicons by using the QIAxcel Advanced System® (QIAGEN, Hilden, Germany) and then proceeded to quantify the DNA concentration by using the PicoGreen QuantiFluor® dsDNA System (Promega, Madison, USA) on a TECAN Infinite F200 PRO® plate reader (Tecan, Männedorf, Switzerland). We normalized the library to include 20 ng DNA from each sample. Finally, we diluted the library to 3 nM for sequencing. The library was spiked with 5% PhiX sequencing control V3 (Illumina, San Diego, USA). Paired-end sequencing of the amplicons was performed as recommended by Illumina (Miseq Reagent Kit v2 – Reagent Preparation Guide) and loaded at a final library concentration of 6 pM. Paired-end sequencing was performed over 2 x 250 cycles.

### Bioinformatics analysis

We conducted the de-multiplexing and de-noising of the samples in the QIIME2 (version 2019.10) pipeline (Bolyen *et al.* 2019) and used the DADA2 method (Callahan *et al.* 2016) to get rid of artefacts and chimeras. Only amplicon sequence variants (ASVs) that survived the filtering step were kept for subsequent analyses. We trained a new SILVA V4 Classifier (SSU release 132 515-806) by using QIIME2 tutorials as a reference (Quast *et al.* 2012). ASVs were then assigned a taxonomy with the highest resolution possible (level 7). Following the taxonomic assignment, we split the analysis into two parts: we kept the original output of the taxonomy assignment (unfiltered) and then we filtered the chloroplast and mitochondria assigned reads (filtered). This step was necessary to evaluate the effect of the clamps on the percentage of reads that were allocated to the chloroplasts and mitochondria before and after application of the clamps. The script for our analysis is deposited in Github ([https://github.com/luisvqz/V4\\_pna.clamps\\_4\\_wildlife](https://github.com/luisvqz/V4_pna.clamps_4_wildlife)). Further analyses were performed in R [version 3.4.4 (2018)] by using the *phyloseq* package (McMurdie & Holmes 2013) in a Linux environment. Plots were generated in R by using the package *ggplot2* (Wickham 2016).

## Results

In bats, we obtained on average 40278 ( $\pm$  3719, n= 8) reads per individual and in birds 71367 ( $\pm$  26981, n= 10) (Supplementary information: Table S1). From early on, we were able to detect that a large percentage of the reads in the unclamped samples were allocated to a few ASVs and, after performing the taxonomic assignment, we were certain that those reads matched known sequences of chloroplasts and mitochondria from publicly available databases. The chloroplast sequences obtained from the bat samples were a 100% match with chloroplast sequences obtained from chloroplast of Saguaro Columnar cacti (GenBank Accession

number: KT164771).

After filtering out the chloroplast and mitochondria assigned reads from the data set, we found that, by using the PNA-DNA clamps, we were able to retain a significantly larger portion of the reads after the filtering step (Fig 4). Although the effectivity varied between individual samples, we always detected an improvement of read coverage available for downstream analyses while using the clamps compared with the unclamped results in pairwise comparisons. On average, the percentage of reads kept improved by 13-fold for the bat (with the clamps cpPNA and mPNA) and by 34-fold for the bird (cPNA and mPNA) (Table S1). The two extreme cases were the bat sample Lepto-195 with a 65-fold improvement and the bird sample MM-143195 with a 216-fold improvement. The control samples, i.e. the bacterial mock community without chloroplasts and mitochondria, showed no fold change indicating that the use of the clamps did not affect the Zymo Mock community in any way (Fig 4). These samples also confirmed that the use of the clamps affected neither the read number nor the percentage of reads kept after the subsequent filtering step.

## Discussion

Challenges associated with plastid contamination represent a major concern in microbiome analyses (Beckers *et al.* 2016; Jackreiet *et al.* 2017; Gaona *et al.* 2020). Our results indicate that the use of DNA-PNA clamps significantly improves the microbiome sequencing output of faecal samples obtained from species with a diet harbouring a large amount of chloroplast and mitochondrial DNA. This effect has also been shown by Fitzpatrick *et al.* (2018) in plant surface microbiomes; however, our study is the first to test the usefulness of clamps in wildlife microbiome studies relying on faecal pellets. Microbiome studies have recently been growing at an accelerated pace. As we move away from model organisms, the diets of the animals under study become more and more diverse. As a rough estimate, 26% of bats and 33% of birds (Ko *et al.* 2014) follow a plant-based diet. Therefore, techniques that allow us to bypass the remnant plant material in faecal samples are becoming more and more important for microbiome studies.

One important factor to keep in mind when using PNA-DNA clamps is the need to have some information about the diet of the study species. PNA-DNA clamp specificity varies between groups. In our case, we had previous knowledge that, in our study area, the diet of Tequila bats consists of almost 100% columnar cacti, particularly from one species, namely *Carnigea gigantea* (LV and MT, personal observation and unpublished data). In the bat case, visual inspection of the faecal pellets also revealed that a large percentage of the pellets was undigested pollen grain clusters. This facilitated the development of the cpPNA clamp thanks to the information available from other studies (Fitzpatrick *et al.* 2018).

Our technique allows the more cost-effective use of sequencing capacity. By employing PNA-DNA clamps, we have been able to target the “true microbiome” more directly and waste fewer reads related to by-products from the diet of the animal. Having higher read numbers enables better statistical power in the analysis and decreases data losses in the subsequent steps in downstream processing. Even though the sequencing price tag is becoming cheaper every day (NHGRI 2020), without the PNA-DNA clamps, we would have had to double or triple or even increase by ten-fold our sequencing depth to make the latter reasonable enough to allow downstream analyses. The cost of the clamps varies between providers but, in general, the use of the clamps will always be more cost-effective than aiming at larger sequencing depth. With the expansion of microbiome studies to non-model organisms, we believe that additional tools like the one presented in this paper will streamline the future advancement of the field.

## Acknowledgments

LV was partly funded by Consejo Nacional de Ciencia y Tecnología; Ministerio de Ciencia, Tecnología y Telecomunicaciones; Rufford Small Grants; Elisabeth Kalko Stiftung, Columbus Zoo, Arizona Game and Fish and Whitley Fund for Nature (In collaboration with RA Medellín, UNAM). We would like to thank the many field assistants and collaborators that joined us on our numerous capturing trips in México: RA. Medellín, A. Menchaca, D. Zamora, B. Iñarritu, A. Vogeler, I. Barba, O. Calva, L. Torres, E. Ramírez and many others. We are also grateful for the technical advice provided by R. Jiménez, U. Stehle and S. Brändel during lab work. Our thanks are also due to A. Risely, R. Jiménez, A. Heni, M. Gillingham and P. Santos

for their comments and advice in the early stages of this manuscript. Finally, we would like to thank the Secretaría Nacional de Medio Ambiente y Recursos Naturales, Dirección General de Vida Silvestre, and Consejo Nacional de Áreas Naturales Protegidas for their help with permits and logistic support.

### Author contributions

LV, SS, MT and KW conceived the ideas and designed the methodology; LV, RF and MT collected the data; LV, RF and KW performed the experiments; LV and RF analysed the data; LV and SS wrote the manuscript. All authors contributed to the manuscript edits and gave their final approval for submission.

### Data Accessibility Statement:

-DNA sequences: The DNA sequences will be stored in Genbank upon acceptance of the manuscript

-Mapping files, data sheets and the scripts for all the microbiome analysis (Qiime2 and R) are store in GitHub ([https://github.com/luisvqz/V4\\_pna\\_clamps\\_4\\_wildlife](https://github.com/luisvqz/V4_pna_clamps_4_wildlife) ) and are already available along with the supplementary table.

### Competing Interests Statement

The authors declare no competing interests.

### References

- BAROTT, K. L., RODRIGUEZ-BRITO, B., JANOUŠKOVEC, J., MARHAVER, K. L., SMITH, J. E., KEELING, P. & ROHWER, F. L. 2011. Microbial diversity associated with four functional groups of benthic reef algae and the reef-building coral *Montastraea annularis*. *Environmental Microbiology*, 13, 1192-1204.
- BECKERS, B., OP DE BEECK, M., THIJS, S., TRUYENS, S., WEYENS, N., BOERJAN, W. & VAN GRONSVELD, J. 2016. Performance of 16s rDNA Primer Pairs in the Study of Rhizosphere and Endosphere Bacterial Microbiomes in Metabarcoding Studies. *Frontiers in Microbiology*, 7, 650.
- BOLYEN, E., RIDEOUT, J. R., DILLON, M. R., BOKULICH, N. A., ABNET, C. C., AL-GHALITH, G. A., ALEXANDER, H., ALM, E. J., ARUMUGAM, M., ASNICAR, F., BAI, Y., BISANZ, J. E., BITTINGER, K., BREJNROD, A., BRISLAWN, C. J., BROWN, C. T., CALLAHAN, B. J., CARABALLO-RODRÍGUEZ, A. M., CHASE, J., COPE, E. K., DA SILVA, R., DIENER, C., DORRESTEIN, P. C., DOUGLAS, G. M., DURALL, D. M., DUVALLET, C., EDWARDSON, C. F., ERNST, M., ESTAKI, M., FOUQUIER, J., GAUGLITZ, J. M., GIBBONS, S. M., GIBSON, D. L., GONZALEZ, A., GORLICK, K., GUO, J., HILLMANN, B., HOLMES, S., HOLSTE, H., HUTTENHOWER, C., HUTTLEY, G. A., JANSSEN, S., JARMUSCH, A. K., JIANG, L., KAEHLER, B. D., KANG, K. B., KEEFE, C. R., KEIM, P., KELLEY, S. T., KNIGHTS, D., KOESTER, I., KOSCIOLEK, T., KREPS, J., LANGILLE, M. G. I., LEE, J., LEY, R., LIU, Y.-X., LOFTFIELD, E., LOZUPONE, C., MAHER, M., MAROTZ, C., MARTIN, B. D., MCDONALD, D., MCIVER, L. J., MELNIK, A. V., METCALF, J. L., MORGAN, S. C., MORTON, J. T., NAIMEY, A. T., NAVAS-MOLINA, J. A., NOTHIAS, L. F., ORCHANIAN, S. B., PEARSON, T., PEOPLES, S. L., PETRAS, D., PREUSS, M. L., PRUESSE, E., RASMUSSEN, L. B., RIVERS, A., ROBESON, M. S., ROSENTHAL, P., SEGATA, N., SHAFFER, M., SHIFFER, A., SINHA, R., SONG, S. J., SPEAR, J. R., SWAFFORD, A. D., THOMPSON, L. R., TORRES, P. J., TRINH, P., TRIPATHI, A., TURNBAUGH, P. J., UL-HASAN, S., VAN DER HOOFT, J. J. J., VARGAS, F., VÁZQUEZ-BAEZA, Y., VOGTMANN, E., VON HIPPEL, M., WALTERS, W., et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* . 37, 852-857
- CALLAHAN, B. J., MCMURDIE, P. J., ROSEN, M. J., HAN, A. W., JOHNSON, A. J. A. & HOLMES, S. P. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581.
- CAPORASO, J. G., KUCZYNSKI, J., STOMBAUGH, J., BITTINGER, K., BUSHMAN, F. D., COSTELLO, E. K., FIERER, N., PEÑA, A. G., GOODRICH, J. K., GORDON, J. I., HUTTLEY, G. A., KELLEY, S. T., KNIGHTS, D., KOENIG, J. E., LEY, R. E., LOZUPONE, C. A., MCDONALD, D., MUEGGE, B. D., PIRRUNG, M., REEDER, J., SEVINSKY, J. R., TURNBAUGH, P. J., WALTERS, W. A., WIDMANN,

- J., YATSUNENKO, T., ZANEVELD, J. & KNIGHT, R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, 7, 335-336.
- CAPORASO, J. G., LAUBER, C. L., WALTERS, W. A., BERG-LYONS, D., HUNTLEY, J., FIERER, N., OWENS, S. M., BETLEY, J., FRASER, L., BAUER, M., GORMLEY, N., GILBERT, J. A., SMITH, G. & KNIGHT, R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*, 6, 1621-4.
- FITZPATRICK, C. R., LU-IRVING, P., COPELAND, J., GUTTMAN, D. S., WANG, P. W., BALTRUS, D. A., DLUGOSCH, K. M. & JOHNSON, M. T. J. 2018. Chloroplast sequence variation and the efficacy of peptide nucleic acids for blocking host amplification in plant microbiome studies. *Microbiome*, 6, 144.
- GAONA, O., CERQUEDA-GARCÍA, D., MOYA, A., NERI-BARRIOS, X. & FALCÓN, L. I. 2020. Geographical separation and physiology drive differentiation of microbial communities of two discrete populations of the bat *Leptonycteris yerbabuenae*. *MicrobiologyOpen*, June 2020,1113-1127.
- HOECK, P. E. A., BOLLMER, J. L., PARKER, P. G. & KELLER, L. F. 2010. Differentiation with drift: a spatio-temporal genetic analysis of Galápagos mockingbird populations (*Mimus* spp.). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 1127-1138.
- INSTITUTE, N. H. G. R. *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)* [Online]. Available: <http://www.genome.gov/sequencingcosts> [Accessed April 13th 2020].
- JACKREL, S. L., OWENS, S. M., GILBERT, J. A. & PFISTER, C. A. 2017. Identifying the plant-associated microbiome across aquatic and terrestrial environments: the effects of amplification method on taxa discovery. *Molecular Ecology Resources*, 17, 931-942.
- KO, C.-Y., SCHMITZ, O. J., BARBET-MASSIN, M. & JETZ, W. 2014. Dietary guild composition and disaggregation of avian assemblages under climate change. *Global Change Biology*, 20, 790-802.
- LUNDBERG, D. S., YOURSTONE, S., MIECZKOWSKI, P., JONES, C. D. & DANGL, J. L. 2013. Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, 10, 999-1002.
- MCMURDIE, P. J. & HOLMES, S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, 8, e61217.
- NIELSEN, P. E. & EGHOLM, M. 1999. An Introduction to Peptide Nucleic Acid. *Current Issues Molec. Biol.*, 1, 89-104.
- NIELSEN, P. E., EGHOLM, M. & BUCHARDT, O. 1994. Peptide nucleic acid (PNA). A DNA mimic with a peptide backbone. *Bioconjugate Chemistry*, 5, 3-7.
- SAGAN, L. 1967. On the Origin of Mitosing Cells. *J. Theoret. Biol.*, 14, 225-274.
- SIKES, R. S. & THE ANIMAL CARE AND USE COMMITTEE OF THE AMERICAN SOCIETY OF MAMMALOGISTS. 2016. Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *Journal of Mammalogy*, 97, 663-688.
- ŠTEFKA, J., HOECK, P. E. A., KELLER, L. F. & SMITH, V. S. 2011. A hitchhikers guide to the Galápagos: co-phylogeography of Galápagos mockingbirds and their parasites. *BMC Evolutionary Biology*, 11, 284.
- TEAM, R. C. 2018. *R: A language and environment for statistical computing* [Online]. Vienna, Austria: R Foundation for Statistical Computing. Available: <https://www.R-project.org/>.
- WICKHAM, H. 2016. ggplot2: Elegant Graphics for Data Analysis. New York, NY: Springer.

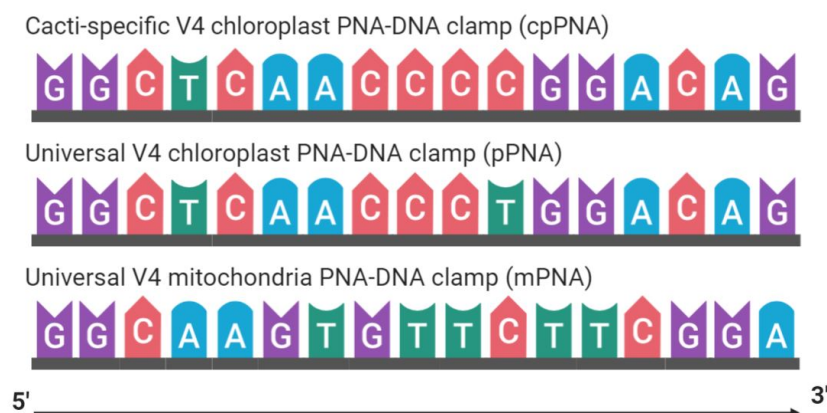
## Figures legends

Figure 1. A Tequila Bat (*Leptonycteris yerbabuenae*) visiting a Cactus flower (Image for journal cover)

Figure 2. Clamps used for this study. The cacti-specific chloroplast clamp (cpPNA) has a single base modification at the 11<sup>th</sup> position (T C, Fitzpatrick et al (2018)) compared with the universal chloroplast clamp (pPNA). Both universal clamps are based on the study of Lundberg et al (2013).

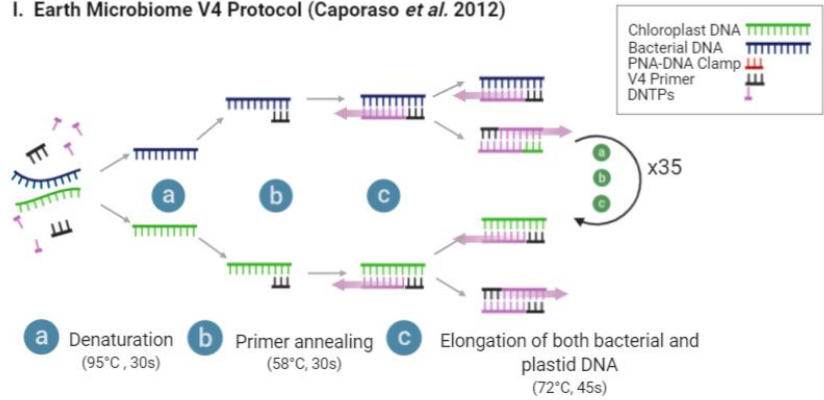
Figure 3. PCR protocol for the implementations of the cpPNA-DNA clamps. I- Normal workflow according to the Earth Microbiome Protocol (Caporaso et al. 2012); II. Our modified protocol with the addition of one extra step for clamp annealing (Step b).

Figure 4. Percentage of reads kept after the filtering out of the chloroplast and mitochondrial DNA using QIIME2. Bat samples were clamped by implementing cpPNA/mPNA and birds by using pPNA/mPNA. The Zymo Mock community was treated with cpPNA/mPNA.





I. Earth Microbiome V4 Protocol (Caporaso *et al.* 2012)



II. Modified PNA-DNA Clamps Protocol (This study)

