

Proper experimental design requires randomization/balancing of molecular ecology experiments

Miklós Bálint¹  | Orsolya Márton^{1,2} | Marlene Schatz³ | Rolf-Alexander Düring³ | Hans-Peter Grossart^{4,5} 

¹Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany

²Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary

³Institut für Bodenkunde und Bodenerhaltung, Gießen, Germany

⁴Leibniz Institute for Freshwater Ecology and Inland Fisheries, Stechlin, Germany

⁵Institute of Biochemistry and Biology, Potsdam University, Potsdam, Germany

Correspondence

Miklós Bálint, Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany.

Email: mbalint@senckenberg.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: BA 4843/2-1

Abstract

Properly designed (randomized and/or balanced) experiments are standard in ecological research. Molecular methods are increasingly used in ecology, but studies generally do not report the detailed design of sample processing in the laboratory. This may strongly influence the interpretability of results if the laboratory procedures do not account for the confounding effects of unexpected laboratory events. We demonstrate this with a simple experiment where unexpected differences in laboratory processing of samples would have biased results if randomization in DNA extraction and PCR steps do not provide safeguards. We emphasize the need for proper experimental design and reporting of the laboratory phase of molecular ecology research to ensure the reliability and interpretability of results.

KEYWORDS

batch effect, bias, DNA extraction, environmental DNA, laboratory practice, lake community, metabarcoding, nondemonic intrusions, PCR, sediment

1 | INTRODUCTION

Ecological studies regularly ensure that the experimental setup is randomized and/or balanced. This allows to interpret results with respect to the original questions and to minimize the influence of confounding factors. The importance of randomized experimental setups (Fisher, 1936) along with balanced designs (Student & Student 1938) is well known. Consequently, such designs are enforced today in manipulative or observational ecological research (Hurlbert, 1984; Smith, Anderson, & Pawley, 2017).

This is often handled differently with laboratory experiments in molecular biology. By laboratory experiments, we mean the laboratory processing (versus obtaining) of samples to generate quantitative molecular genetic data: DNA extractions, polymerase chain reactions, DNA sequencing, etc., in order to obtain haplotype frequencies,

taxonomically informative marker gene counts, gene expression measures, SNP tables, etc. Although methodological errors associated with molecular data are frequently discussed (e.g., Aird et al., 2011; Leray & Knowlton, 2017; Robasky, Lewis, & Church, 2014; Schirmer et al., 2015; Schnell, Bohmann, & Gilbert, 2015), batch effects have received considerably less attention. Early genome-wide association studies (GWAS) are examples of how they may be ignored in basic experimental design and what the consequences are: The analyses are expensive, but the obtained data cannot be interpreted (or are misinterpreted) due to confounding effects of laboratory procedures (Sebastiani et al., 2010). The early problems lead to the current recognition of randomized and/or balanced laboratory experimental designs in medical genomics (Lambert & Black, 2012; Leek et al., 2010; Yang et al., 2008).

Complex and expensive molecular genetic datasets are increasingly generated in ecology. It is important that these data are generated

appropriately as important conclusions and recommendations are drawn from them, often addressing issues of global importance for nature, society, and economy. Randomization or balancing in laboratory experiments is essential to avoid batch effects and other nondemonic intrusions (see Hurlbert, 1984). This issue has been already raised by Meirmans (2015) in a recent opinion paper on population genetics. Meirmans (2015) notes that “It is perfectly possible that such randomization is already practised in genotyping laboratories everywhere and I am simply unaware of it. [...], if this is the case, this is nowhere evident in the literature”. We have similar impressions and the screening of one randomly selected 2016 issue from each of five relevant journals supports this assumption (Molecular Ecology, The ISME Journal, Ecology and Evolution, Journal of Biogeography, Soil Biology and Biochemistry, Appendix S1). Only two of the 59 relevant studies report some form of randomization during the laboratory processing of samples. This small literature survey is surely not representative of overall molecular ecology research, but the pattern is worrying as a simple Web of Science search for the keyword combination “molecul* AND ecol*” resulted in over 1,740 hits only from 2016.

The omission of randomization in the laboratory may allow chance events to systematically influence results. Such chance events are common everytime and everywhere: electric fallouts happen, sudden flaws incapacitate laboratory personnel, DNA extraction kits are not delivered in time or have been stored inappropriately, just to mention some. If samples are processed in batches, the coincidence of these events confounds the results and renders interpretations unreliable. The potential diversity of such events is so high that nothing can protect against them except randomization of laboratory procedures, potentially in combination with balanced designs.

Hurlbert (1984) notes that most of the time chance events have immeasurably small effects on the results. However, by nature, they are also completely unpredictable, both in frequency and effect size.

As molecular ecology studies mostly work with high observation numbers (thousands of SNPs over genomes, thousands of operational taxonomic units—OTUs—in hundreds of samples, etc.), even small chance events may result in statistically significant results (Carver, 1993). Here, we demonstrate this with taxonomically informative marker gene fragments amplified from environmental DNA (eDNA metabarcoding). The eDNA was preserved in lake sediments and provides a perspective on lake ecosystem history over several decades. We looked at three aspects of methodological or biological interest: extracted DNA concentration, PCR efficiency, and community properties (Figure 1). We evaluated several sources of variation: (1) expected laboratory biases (DNA extraction kit, Deiner, Walser, Mächler, & Altermatt, 2015; Barlow et al., 2016), (2) unexpected laboratory biases (e.g., a sudden change in laboratory personnel), and (3) an ecologically interesting predictor (either the age of the sediment or the effects of a power plant).

2 | MATERIALS AND METHODS

2.1 | Sampling

Two sediment cores of the same location from Lake Stechlin were taken on 14 May 2015 with a gravity corer (UWITEC[®], Mondsee, Austria) and Perspex tubes (inner diameter 9 cm, lengths 60 cm). Lake Stechlin (latitude 53°10'N, longitude 13°02'E) is a dimictic meso-oligotrophic lake (maximum depth 69.5 m; area 4.5 km²) in the lowlands of northern Germany. GDR's first nuclear power plant was built here between 1960 and 1966 and operated until 1990, connecting the lake with the nearby mesotrophic Lake Nehmitz and discharging its cooling water into Lake Stechlin. After coring, the cores were sliced immediately in the field in approximately 0.5-cm intervals. The first core was designated to eDNA. All sampling tools were H₂O₂-sterilized

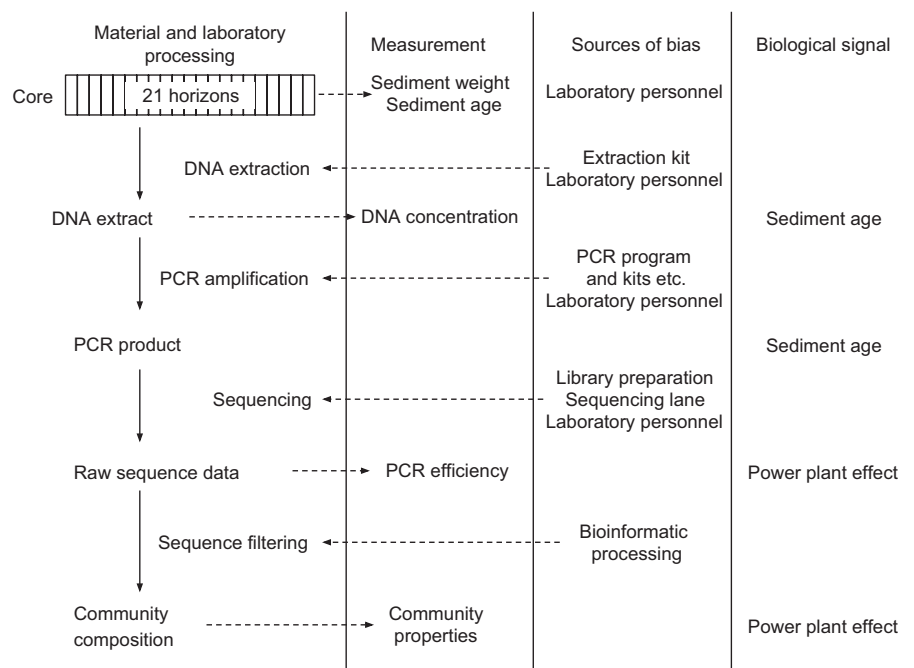


FIGURE 1 Analysis scheme with predictors of variation in high-throughput-sequenced eDNA amplicon data. The first column lists typical analysis steps and the endpoints of these, the second column contains options to evaluate these endpoints through measurements, the third column lists several laboratory biases that may exert batch effects, and the fourth column contains biological factors of interest

after cutting each horizon. Sediment for DNA extraction was taken only from the central part of the core to avoid contamination by contact with the corer's wall. Samples were immediately stored in 15-ml Falcon tubes (Neolab Migge GmbH, Heidelberg, Germany) at -20°C until DNA extraction. Horizons from the second core were used for organohalogenic pesticide measurements.

2.2 | Date approximation

Approximate dates were obtained by comparing DDT decomposition compound concentrations with sedimentation rates inferred with ^{137}Cs (Casper, 1994): 1.2 mm/year between 1986 and 1996 and 1.7 mm/year between 1963 and 1986. We assumed that DDT deposition started with World War II when a military training camp was operated near the lake and it effectively stopped in 1990 when agrochemical subventions of the GDR ceased with the reunification of Germany. The pesticide concentrations, sedimentation rates, and inferred dates can be consulted in the file *Stechlin_organohalogene.csv*, deposited in Figshare (<https://figshare.com/s/32dbca0a906c7f06449b>, <https://doi.org/10.6084/m9.figshare.4579681>).

Halocarbon compound extraction was performed by shaking 300 mg freeze-dried sediment sample once in acetone and petroleum ether ($40-60^{\circ}\text{C}$) and then only in petroleum ether ($40-60^{\circ}\text{C}$), based on ISO 10382:2002. The clear supernatants were unified and vortexed after centrifugation, then a 10 ml aliquot was transferred to SPME amber screw top vials and evaporated under a gentle stream of nitrogen until dry and dissolved again in 100 μl methanol and mixed with 10 ml of a 0.01 mol $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ /3.4 mol NaCl salt solution. As internal standards, ^{13}C 2.4 DDT, ^{13}C 4.4 DDT, α -HCH D6, Trifluralin D14, 4.4 DDD D8, 4.4 DDE D8, and ^{13}C HCB were used. Finally, samples were extracted by Head Space Solid Phase Microextraction (SPME- HS) with a PDMS 100 fiber and analyzed by GC/MS ion trap in selected ion monitoring mode (SIM). Separation and detection were accomplished using a Trace Ultra Gas Chromatograph (Thermo Fisher Scientific Inc., Schwerte, Germany) provided with a RTX-Dioxin 2 fused-silica capillary column with 0.25 μm film thickness, 0.25 mm ID, and 60 m length coupled with an ion trap mass spectrometer in SIM (Thermo Fisher Scientific Inc.).

2.3 | DNA extraction

We selected the youngest 21 horizons (the upper 13.5 cm of the core) for DNA extractions. Sample order was randomized before DNA extraction to minimize sampling biases. Four DNA extractions were carried out from each horizon with two commercial kits (two replicated extractions with both Macherey-Nagel NucleoSpin Soil—Macherey-Nagel, Düren, Germany, and MoBio PowerSoil—Carlsbad, CA, USA). The protocols of both kits were modified to specifically target extracellular DNA: Instead of lysis, a saturated phosphate buffer was used to extract sediment-bound DNA (Taberlet et al., 2012). All four extraction replicates of a horizon were performed in the same DNA extraction batch (24 extractions, see column *extract_order* in *sample_infos.csv* deposited in Figshare

(<https://figshare.com/s/32dbca0a906c7f06449b>, <https://doi.org/10.6084/m9.figshare.4579681>). We altogether included four DNA extraction negative controls into the experiment (dH_2O instead of sediment), and these were randomly distributed within the extraction batches. No controls for field contamination were taken. Extracted DNA concentrations were estimated on a Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). We could not measure DNA in any of the extraction controls.

2.4 | PCR amplifications

DNA templates were rerandomized before PCR setup. Four PCR negative controls (with dH_2O instead of DNA template) and two positive controls (containing DNA from tissues or cell cultures of *Hypsiboas punctatus* (66.5 ng/ μl), *Ponticola kessleri* (306 ng/ μl), *Aspius aspius* (280 ng/ μl), *Coregonus* sp. (398 ng/ μl), *Pacifastacus leniusculus* (425 ng/ μl), *Aphanomyces astaci* (104 ng/ μl), a parasitic Chytridiomycota (4.5 ng/ μl), a saprotrophic Chytridiomycota (25.3 ng/ μl), *Yamagishiella* sp. (15.7 ng/ μl), *Fragilaria crotonensis* (3.3 ng/ μl), *Staurastrum planktonicum* (2.7 ng/ μl), *Chaetomium* sp. (7.6 ng/ μl), and *Lutra lutra* (17.4 ng/ μl)) were included. The first positive control was an equimolar mixture, with a nominal concentration of 5 ng/ μl from each source. The second positive control was a complex nonequimolar mixture with the following dilutions: *Hypsiboas punctatus* (32x), *Ponticola kessleri* (90x), *Aspius aspius* (4x), *Coregonus* sp. (1x), *Pacifastacus leniusculus* (2x), *Aphanomyces astaci* (50x), a parasitic Chytridiomycota (256x), a saprotrophic Chytridiomycota (8x), *Yamagishiella* sp. (64x), *Fragilaria crotonensis* (500x), *Staurastrum planktonicum* (520x), *Chaetomium* sp. (128x), and *Lutra lutra* (16x). We used AmpliTaq MasterMix for the PCRs (Thermo Fisher Scientific). We used general eukaryote primers that amplify a short fraction of the V7 region of the 18S gene region (Guardiola et al., 2015): forward—TYTGTCTGTTTRATTSCG, reverse—CACAGACCTGTTATTGC. The primers contained the Illumina sequencing primers (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG). The PCRs were run in 15 μl reaction volume (AmpliTaq MasterMix: 7.5 μl , water: 4 μl , each 5 $\mu\text{mol/L}$ primer 1 μl , DNA template 1.5 μl). DNA concentrations of individual templates are provided in the “conc” column *sample_infos.csv* file (accessible through <https://figshare.com/s/32dbca0a906c7f06449b>, <https://doi.org/10.6084/m9.figshare.4579681>). The cycling conditions were 95°C (10 min), 44 cycles of 95°C (30 s), 45°C (30 s), 72°C (30 s), and final extension at 72°C (10 min). The PCR products were visualized on a 2% agarose gel and purified with Agencourt AMPure XP beads (Beckman Coulter GmbH, Krefeld, Germany).

2.5 | Multiplexing strategy and sequencing

We indexed all samples for multiplexed sequencing in a subsequent short PCR with primers that contained a fraction of the Illumina sequencing primer (TCGTCGGCAGCGTC and GTCTCGTGGGCTCGG), an eight-bp nucleotide index, and the Illumina plate adapters (P5: AATGATACGGCACCGAGATCTACAC, P7: CAAGCAGAA

GACGGCATACGAGAT). The final products are indexed, ready to sequence Illumina libraries. Index combinations and sequences are provided in the file multiplexing_indices.xlsx at Figshare (<https://figshare.com/s/32dbca0a906c7f06449b>, <https://doi.org/10.6084/m9.figshare.4579681>). The procedure follows the Illumina 16S metabarcoding protocol (Illumina 2016). This protocol eliminates index jumps during library preparation (although a few index jumps are still known to happen on the sequencing plate (Schnell et al., 2015). The indexing PCRs were run in 15 µl reaction volume (AmpliTaQ MasterMix: 7.5 µl, each 5 µmol/L primer 1 µl, PCR product 6.5 µl). The cycling conditions were 95°C (10 min), 8 cycles of 95°C (30 s), 52°C (30 s), 72°C (30 s), and final extension at 72°C (10 min). We checked the efficiency of each PCR run on a 2% agarose gel. The indexed libraries were purified with Agencourt AMPure XP beads (Beckman Coulter GmbH, Krefeld, Germany). The indexed libraries were mixed and purified on four QIAamp MinElute columns (Qiagen, Hilden, Germany). We did not normalize the PCR template concentrations to obtain a rough estimate of PCR and sequencing efficiency through the read numbers. Our sequencing kit potentially produces about 1 million paired-end reads with 2 × 150 bp length. Illumina sequencing was performed at the Berlin Center for Genomics in Biodiversity Research (www.begendiv.de) with the MiSeq sequencing kit v2 nano (300 cycles). Unprocessed sequence data were deposited in the European Nucleotide Archive as PRJEB19403.

2.6 | Sequence processing and data analysis

Raw sequence data were processed with OBITools (Boyer et al., 2015). Potential contamination and false detection biases were controlled for by following the recommendations of (Boyer et al., 2015; Giguet-Covex et al., 2014; Pansu et al., 2015) in R 3.3.1. ("R: The R Project for Statistical Computing"). All OBITools and R commands are documented in the file *stechlin_analyses.pdf* at Figshare (<https://figshare.com/s/32dbca0a906c7f06449b>, <https://doi.org/10.6084/m9.figshare.4579681>), with the full code accessible through the GitHub repository <https://github.com/MikiBalint/LaboratoryDesign.git>. Commands

were run with GNU "parallel" when possible (Tange, 2011). The resulting OTU abundance table is provided in the *stechlin_assigned_190915.tab* file through Figshare (<https://figshare.com/s/32dbca0a906c7f06449b>, <https://doi.org/10.6084/m9.figshare.4579681>).

We fitted linear mixed-effect models with lme4 (Bates, Mächler, Bolker, & Walker, 2015) on extracted DNA concentration, PCR efficiency, and measures of diversity (the first three integers from Hill's diversity series (Hill, 1973) to estimate the effects of potential laboratory biases and biological factors of interests. The first three Hill numbers correspond to species richness (H1), the exponent of Shannon diversity (H2), and the inverse of the Simpson diversity (H3). The identity of the sediment horizon was used as the random effect in these models. We used multispecies generalized linear models (GLMs) with the "mvabund" R package (Wang, Naumann, Wright, & Warton, 2012) to investigate the effects of the predictors on community composition. The multispecies GLM cannot handle random effects. The community composition effects were visualized with a latent variable model-based ordination performed with the boral R package (Hui, 2016). Both compositional analyses assume a negative binomial distribution of the data, accounting for the sparse and overdispersed nature of read counts (Bálint et al., 2016). The input data matrices are accessible through Figshare (<https://figshare.com/s/32dbca0a906c7f06449b>, <https://doi.org/10.6084/m9.figshare.4579681>).

The models can be written up as

1. $conc \sim weight + kit + person + age + I(age^2) + (1|depth.nominal)$
2. $PCR\ efficiency \sim conc + kit + person + age + (1|depth.nominal)$
3. $diversities \sim PCR\ efficiency + person + kit + nuclear + (1|depth.nominal)$
4. $Community\ composition \sim reads + kit + person + nuclear,$

where *conc* is the extracted DNA concentration, *weight* is the sediment weight used for DNA extraction, *kit* is the DNA extraction kit, *person* is the laboratory personnel, *depth.nominal* is the identity of the sediment horizon, *PCR efficiency* is estimated from HTS read numbers, and *nuclear* is the operational period of the nuclear power plant (Figure 1).

TABLE 1 Summary of predictor contributions to variation

	DNA concentration	PCR efficiency	H1	H2	H3	Community composition
Sediment weight	0.1***	—	—	—	—	—
Extraction kit	12.9***	2,719,604***	672.6***	197.9***	67.8***	906
Laboratory personnel	1.7**	81,413,118**	37.9	0.9	1.9	2,018*
DNA concentration	—	58,222***	—	—	—	—
PCR efficiency	—	—	14,834.7***	1,156.3***	172.3***	1,163**
Age/power plant effect	1.8***	198,964***	1,758.3***	867***	435.5***	5,488*

*Statistically marginally significant result ($p < .1$).

**Statistically significant result ($p < .05$).

***Statistical significance not tested.

3 | RESULTS

The results are summarized in Table 1 and Figure 2. Regarding DNA concentrations, the DNA extraction kit (equivalent to the expected laboratory biases) accounted for most variation, followed by the age of the sediment horizon (biological signal) and the laboratory personnel (equivalent to the unexpected laboratory bias). The starting weight (amount) of the sediment had limited effects on the extraction efficiency, and the effect of the laboratory personnel was marginally significant. PCR efficiency (evaluated as non-normalized HTS read numbers from PCR amplification) was mostly explained by the personnel identity (unexpected laboratory bias), followed by the DNA extraction kit (expected laboratory biases), the age of the sediment horizon, and the DNA template concentration used for the PCR. Here, the effect of the laboratory personnel was statistically significant. The most important contributors to variation in the first three Hill numbers consisted of PCR efficiency and the effects of the nuclear power plant. The DNA extraction kit contributed relatively little to the observed variation in the diversity indices. The nuclear power plant effects, however, represented the largest contributors to the explained variation in community composition, followed by the identity of the laboratory personnel and PCR efficiency. The DNA extraction kits contributed the least to the explained compositional variation (Figure 3). The effects of the laboratory personnel were statistically marginally significant. Additional results and effect plots are available in file *stechlin_analyses.pdf* at Figshare.

4 | DISCUSSION

Our results demonstrate that nondemonic intrusions (Hurlbert, 1984) in the laboratory may produce in strong, statistically significant effects

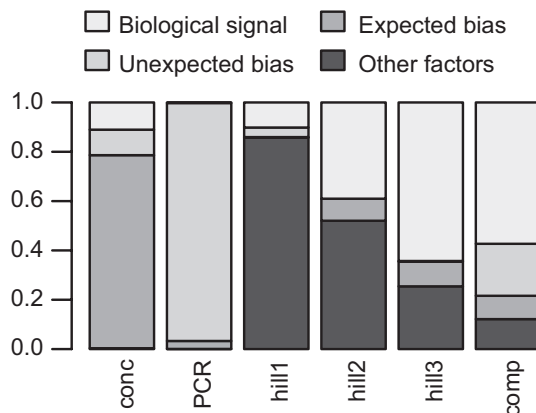


FIGURE 2 Partitioning of variance explained by expected and unexpected laboratory biases, and biological signal. The bars represent explained variance in DNA concentration (conc), PCR efficiency (PCR), diversity indices (hill1–3), and community composition (comp). Predictors: biological signal; effects of sediment age (conc, PCR) or the power plant operation periods (hill1–3, comp); unexpected bias: effects of laboratory personnel; expected bias: effects of DNA extraction kit; other factors: sediment weight (conc), DNA concentration (PCR); PCR efficiency (hill1–3, comp)

that may severely confound results. Such effects render equivocal interpretations impossible if they coincide with effects targeted by the study. For example, interpretation of power plant effects on community composition would be difficult if samples are processed in batches and the sudden change in laboratory personnel coincides with a shift between operation periods. Similarly, strong personnel effects are well known when scoring microsatellite genotypes (or allozyme electrophoresis patterns): they are well known to be influenced by personal judgments and authors should also report how they dealt with it. Different pipelines and/or processing parameters (e.g., thresholds to discard rare OTUs) may produce different results in metabarcoding, for example, in diversity measures (Golob, Margolis, Hoffman, & Fredricks, 2017). We expect that all samples within a study are processed with the same pipeline or parameters; thus, the pipeline effect is uniform among all samples, allowing comparisons. However, differences in results caused by pipelines and parameters may influence metastudies. We hypothesize that these effects may reflect the study date due to developments in pipeline use. The uniform reprocessing of the data before such interstudy comparisons is thus important (Meiser, Bálint, & Schmitt, 2014).

The results allow to evaluate whether observed community properties are differentially influenced by methodological biases and biological signal. The first three Hill numbers correspond with commonly used indices of biodiversity (H1: richness, H2: the exponent of Shannon diversity, and H3: inverse of Simpson diversity), and they differ only in the way they penalize rare species. The variation explained by the biological signal increases as rare species are increasingly penalized (Figure 2), and this points to the importance of handling rare sequence variants during the analyses (Bálint et al., 2016).

We do not state that biases with comparable extent always appear in unrandomized, not balanced laboratory experiments, but they certainly have the potential to do so. This is clear in our example: The effects of unexpected laboratory biases exceed the effects of known laboratory biases (DNA extraction kit effects) and biological signal in several models (Figure 2). Such effects potentially influence all molecular ecology studies and threaten the interpretability of results. Their importance and extent are known in biomedicine (Fungtammasan et al., 2015; Lambert & Black, 2012; Leek et al., 2010; Yang et al., 2008), and it needs to be urgently considered in molecular ecology.

Generally, randomization of samples before major laboratory steps (particularly DNA extraction) is simple and low cost. The only case where this might be disputable is the processing of highly contamination-prone materials where it is almost a laboratory rule that DNA extraction is performed consecutively from the most contamination-prone toward the least contamination-prone samples (although to our knowledge, the validity of this still needs to be tested). Obviously, nondemonic intrusions (including contamination) in the laboratory easily become collinear with the processing order and this makes biological signals difficult to interpret (Salter et al., 2014). Rearrangement (e.g., a new randomization) of DNA extracts before PCR reactions or sequencing might be more difficult and error-prone, especially with large numbers of samples. Trade-offs between

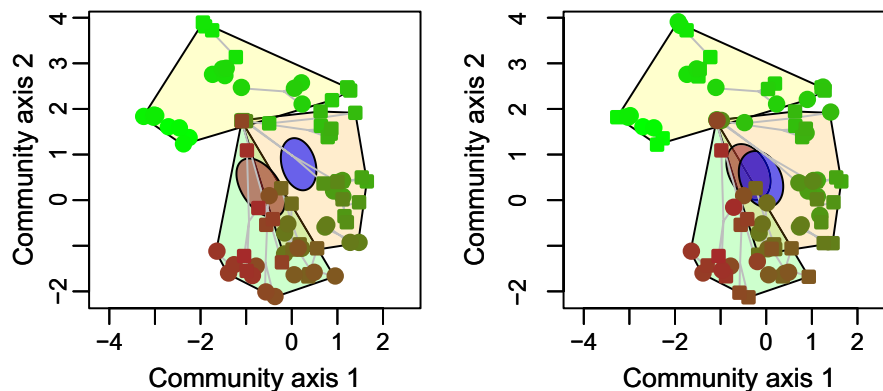


FIGURE 3 Compositional changes in historic communities explained by expected and unexpected laboratory biases, and biological signal. Points represent communities reconstructed from replicated DNA extractions from 21 sediment horizons, representing the last ~70 years of the lake's history. Symbol color indicates age: Dark brown is the oldest, and light green is the youngest communities. Replicated DNA extracts of a horizon are connected by gray lines. The operational phases of the nuclear power plant are marked with hulls: green—before building the plant, orange—during power plant operation, and yellow—after operation. (a) Symbols mark the effects of laboratory personnel on community composition, and the two ellipses show the 95% confidence interval of the corresponding group centroids. (b) Symbols and ellipses mark the effects of the DNA extraction kits

rerandomization and possible contaminations should be considered in this case.

We recommend the followings: First, researchers involved in molecular ecology laboratory work need to properly design and report laboratory procedures. Guidelines in biomedicine exist and may be readily adapted, for example, how and why samples were assigned to certain processing batches and processing timeframes, evaluation of technical differences among equipment, and blinding the experiment by concealing information about sample identity from the laboratory personnel (Masca et al., 2015). Second, ecologists who rely on molecular data generated by laboratory personnel or companies must ensure (and should not take for granted) that principles of experimental design are followed in the laboratory. This is the easiest when giving samples to a laboratory as the ecologist can already rearrange and relabel his/her samples (but controls of PCR, sequencing, orders, etc. may require further communication). Third, editors and reviewers of manuscripts and grants should enforce the reporting of laboratory experimental design. This is as much necessary for reproducible research as the proper presentation of sampling schemes, details of manipulative experiments, and data analysis. We do not intend to provide a list of important laboratory biases as there are potentially infinite variations, but some that easily come to mind are (1) the order in which DNA extractions are performed (one may expect a "learning effect" that may change the results even of the same person as he/she processes more samples), (2) the position of the sample on PCR plate (reactions placed on the outer part of plates sometimes tend to evaporate more water during PCR), (3) PCR machine (especially if a different brand, or purchased at a different time), and (4) the lane of a high-throughput sequencer. Therefore, molecular ecologists must ensure randomization or properly balanced designs in every step of laboratory work and present the details. There is no excuse for avoiding this as more and more globally important decisions require reliable molecular ecology data in nature and biodiversity conservation.

ACKNOWLEDGMENTS

The authors express gratitude to the working community of the Senckenberg Conservation Genetics Group (Gelnhausen, Germany) for hosting and supporting the laboratory work. Claudia Wittwer, Silke Van den Wyngaert, Martin Jansen, and Bernardino Cocchiararo provided tissues for generating the positive controls. We thank Susan Mbedi from BeGenDiv for suggestions related to multiplexed sequencing library preparation and sequencing. MB is supported by DFG grant BA 4843/2-1.

CONFLICT OF INTEREST

The authors state no conflict of interest.

AUTHOR CONTRIBUTIONS

MB conceived the ideas. MB and HPG designed the methodology and obtained the cores. MB and OM performed the molecular laboratory work. MS and RAD performed the organohalogen measurements. MB processed the sequences, analyzed the data, and lead the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA SOURCES

(accessible through Figshare, <https://figshare.com/s/32dbca0a906c7f06449b>, <https://doi.org/10.6084/m9.figshare.4579681>):

sample_infos.csv: the description of samples, negative, and positive controls.

multiplexing_indices.xlsx: PCR plate setup and nucleotide indices used for sample multiplexing.

stechlin_assigned_190915.tab: OTU abundance table.

Stechlin_organohalogene.csv: organohalogen pesticide concentrations in the sediments.

lab-methods_OTU_anova.RData: ANOVA table of the multispecies generalized linear model (100 bootstraps).

Lab_LV_model_40000-iter.RData: ordination results with a latent variable model (40 000 iterations).

ORCID

Miklós Bálint  <http://orcid.org/0000-0003-0499-8536>

Hans-Peter Grossart  <http://orcid.org/0000-0002-9141-0325>

REFERENCES

- Aird, D., Ross, M. G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., ... Gnrke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, 12, R18. <https://doi.org/10.1186/gb-2011-12-2-r18>
- Bálint, M., Bahram, M., Eren, A. M., Faust, K., Fuhrman, J. A., Lindahl, B., ... Tedersoo, L. (2016). Millions of reads, thousands of taxa: Microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews*, 40, 686–700. <https://doi.org/10.1093/femsre/fuw017>
- Barlow, A., Gonzalez Fortes, G. M., Dalen, L., Pinhasi, R., Gasparyan, B., Rabeder, G., ... Hofreiter, M. (2016). Massive influence of DNA isolation and library preparation approaches on palaeogenomic sequencing data. *BioRxiv*, <https://doi.org/10.1101/075911>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, <https://doi.org/10.18637/jss.v067.i01>
- Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2015). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, 16, 176–182.
- Carver, R. P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education*, 61, 287–292. <https://doi.org/10.1080/00220973.1993.10806591>
- Casper, P. (1994). Die Cäsium-Datierung von Sedimenten unterschiedlicher mikrobieller Aktivität, Erweiterte Zusammenfassungen der Jahrestagung, DGL - Deutsche Gesellschaft für Limnologie, 386–389.
- Deiner, K., Walser, J.-C., Mächler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, 183, 53–63. <https://doi.org/10.1016/j.biocon.2014.11.018>
- Fisher, R. A. (1936). Design of experiments. *BMJ*, 1, 554. <https://doi.org/10.1136/bmj.1.3923.554-a>
- Fungtammasan, A., Ananda, G., Hile, S. E., Su, M. S.-W., Sun, C., Harris, R., ... Makova, K. D. (2015). Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Research*, 25, 736–749. <https://doi.org/10.1101/gr.185892.114>
- Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., ... Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature Communications*, 5, 3211.
- Golob, J. L., Margolis, E., Hoffman, N. G., & Fredricks, D. N. (2017). Evaluating the accuracy of amplicon-based microbiome computational pipelines on simulated human gut microbial communities. *BMC Bioinformatics*, 18, 283. <https://doi.org/10.1186/s12859-017-1690-0>
- Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangenstein, O. S., & Turon, X. (2015). Deep-Sea, deep-sequencing: Metabarcoding extracellular DNA from sediments of marine canyons. *PLoS ONE*, 10, e0139633. <https://doi.org/10.1371/journal.pone.0139633>
- Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54, 427–432. <https://doi.org/10.2307/1934352>
- Hui, F. K. C. (2016). boral- Bayesian ordination and regression analysis of multivariate abundance data. *Methods in Ecology and Evolution*, 7, 744–750. <https://doi.org/10.1111/2041-210X.12514>
- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54, 187–211. <https://doi.org/10.2307/1942661>
- Illumina (2016). 16S Metagenomic sequencing library preparation. Part # 15044223 Rev. A pp 28. Retrieved from http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf
- Lambert, C. G., & Black, L. J. (2012). Learning from our GWAS mistakes: From experimental design to scientific method. *Biostatistics*, 13, 195–203. <https://doi.org/10.1093/biostatistics/kxr055>
- Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., ... Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews. Genetics*, 11, 733–739. <https://doi.org/10.1038/nrg2825>
- Leray, M., & Knowlton, N. (2017). Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ*, 5, e3006. <https://doi.org/10.7717/peerj.3006>
- Masca, N. G., Hensor, E. M., Cornelius, V. R., Buffa, F. M., Marriott, H. M., Eales, J. M., ... Teare, M. D. (2015). RIPOSTE: A framework for improving the design and analysis of laboratory-based research. *ELife*, 4, <https://doi.org/10.7554/eLife.05519>
- Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, 24, 3223–3231. <https://doi.org/10.1111/mec.13243>
- Meiser, A., Bálint, M., & Schmitt, I. (2014). Meta-analysis of deep-sequenced fungal communities indicates limited taxon sharing between studies and the presence of biogeographic patterns. *The New Phytologist*, 201, 623–635. <https://doi.org/10.1111/nph.12532>
- Pansu, J., Giguet-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., ... Choler, P. (2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. *Molecular Ecology*, 24, 1485–1498. <https://doi.org/10.1111/mec.13136>
- R: The R Project for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Robasky, K., Lewis, N. E., & Church, G. M. (2014). The role of replicates for error mitigation in next-generation sequencing. *Nature Reviews. Genetics*, 15, 56–62.
- Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., ... Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, 12, 87. <https://doi.org/10.1186/s12915-014-0087-z>
- Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research*, 43, e37. <https://doi.org/10.1093/nar/gku1341>
- Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated—reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, 15, 1289–1303. <https://doi.org/10.1111/1755-0998.12402>
- Sebastiani, P., Solovieff, N., Puca, A., Hartley, S. W., Melista, E., Andersen, S., ... Perls, T. T. (2010). Genetic signatures of exceptional longevity in humans. *Science*, 2010, <https://doi.org/10.1126/science.1190532>
- Smith, A. N. H., Anderson, M. J., & Pawley, M. D. M. (2017). Could ecologists be more random? Straightforward alternatives to haphazard spatial sampling. *Ecography*, 1251–1255.
- Student & Student (1938). Comparison between balanced and random arrangements of field plots. *Biometrika*, 29, 363.

- Taberlet, P., Prud'homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., ... Coissac, E. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Molecular Ecology*, 21, 1816–1820. <https://doi.org/10.1111/j.1365-294X.2011.05317.x>
- Tange, O. (2011). GNU parallel - The command-line power tool. *The USENIX Magazine*, 31, 42–47.
- Wang, Y., Naumann, U., Wright, S. T., & Warton, D. I. (2012). mvabund - an R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution*, 3, 471–474. <https://doi.org/10.1111/j.2041-210X.2012.00190.x>
- Yang, H., Harrington, C. A., Vartanian, K., Coldren, C. D., Hall, R., & Churchill, G. A. (2008). Randomization in laboratory procedure is key to obtaining reproducible microarray results. *PLoS ONE*, 3, e3724. <https://doi.org/10.1371/journal.pone.0003724>

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Bálint M, Márton O, Schatz M, Düring R-A, Grossart H-P. Proper experimental design requires randomization/balancing of molecular ecology experiments. *Ecol Evol*. 2018;8:1786–1793. <https://doi.org/10.1002/ece3.3687>