SPAAM2 - Session Notes

Session 1 - Trash in, trash out: Optimizing and standardizing laboratory practices in ancient metagenomics

Chairs Irina Velsko
Icebreaker Presenters Jaelle Brealey & Zandra Fagernäs

Session Abstract

Optimizing laboratory practices is necessary to reduce the amount of contaminating DNA in a sample extract, and subsequently in sequencing data. All sequenced contaminating DNA takes up information that should be for sample DNA, reducing the efficiency of workflows, both financially and bioinformatically. Further, sample biomass influences the amount of "spill-over" that is observed in sequence data, where the contents of higher biomass samples may show up in lower biomass samples. In this session we will discuss how best to prevent or reduce introduction of contamination to sample extraction and library building protocols, as well as limiting sample cross-contamination. These techniques aim to improve sequencing efficiency by reducing the amount of sequenced contamination and increasing the amount of sequenced sample DNA.

General Discussion

- Poll: who works in the wet lab?
    - I am: 52% (16)
    - Nope: 48% (15)
    - Total Votes: 31

What in the lab causes biases in metagenomic analyses?

- **Q**: Different extraction methods?
    - **A**: Identified a lack of comparative analyses to investigate biases associated with lab methods in terms of tested methods and sample types (e.g., calculus, environmental samples). Some are already testing whether different extraction methods for DNA extraction from soil samples have an impact on microbial community composition. [Note, some results can be found from previous studies, list to be completed if needed: on bones, no impact on microbial profiles when using EDTA predigestion (Der Sarkissian et al. 2014); when using bleach treatment (Boessenkool et al. 2016); on coprolites, no impact on microbial profiles when using different extraction methods (Hagan et al. 2020)]. Could actually be extended to lab methods other than DNA extraction, i.e., USER treatment, library construction.
- **Q**: Lab-derived contamination?
    - **A**: In other words, is it possible to identify lab-specific signatures in the metagenomic data that they produce? Idea for a collaborative project, send out a bone to several labs to analyse and compare results, very interesting but major problem of financial resources and sharing samples across different labs (e.g., quarantine, authorisations, travel, other?). Possible solutions: using a same synthetic sample of known composition? bury a sheep bone for a year and dig it up? Multi-lab grant application?

others?

What about blanks?

- **Poll**: how many blanks do you include per batch?
  - 1: 56% (10)
  - 2: 22% (4)
  - >2: 22% (4)
  - Total Votes: 18
- **Q**: What are the practices regarding blanks?
  - **A**: No blanks included in experiments. Blanks as a standard for ancient metagenomics.
  - **A**: Blanks not sequenced because library quantification below threshold of detection of quantification methods. Blank sequencing as a standard for ancient metagenomics.
  - **A**: Blanks not sequenced because library quantification by qPCR shows Ct cycles for blanks significantly lower than for samples. Blank sequencing as a standard for ancient metagenomics.
  - **A**: Blanks sequenced
- **Q**: Why aren't blanks sequenced?
  - **A**: Due to budget constraints (main problem). A suggested solution would be to include blanks in the requested budget of grant application.
  - **A**: Due to low biomass, requires over-amplification of blanks. Even if the signal is low, there's always DNA in blanks.
  - **A**: When samples are pooled together, preferred practice of samples being pooled with equimolarity. Sometimes not possible with a blank given the low DNA content, and risk/worry of diluting the other samples?

- **Q**: What should be considered when sequencing blanks?
  - **A**: Sequencing depths. Some consider that blanks should be treated the same way as other samples: the same analytical pipelines should be applied, which includes considering the same criteria to decide on how deep to sequence them. A more economical solution would be to sequence blanks at lower depths than samples However, how much lower? what is acceptable? The complexity of the blank libraries can be first evaluated by shallow sequencing, helping with deciding how many reads to generate from blanks in a second sequencing round. Given examples: ~5 million reads are sequenced for blanks in ancient human analyses at the Leipzig lab, ~2 million reads in Jena.
  - **A**: I often add them to the sequencing run in a 1in10 dilution compared the samples, this avoids having to amplify them/clonality. If you're aiming for 10Mio reads per sample that should then give you 1Mio extraction blank reads.
  - **A**: Sequencing blanks as part of the same lane as “non-blank” samples or in a separate lane dedicated to blanks. The latter ensures that there is no barcode jumping and wrong identification of reads coming from samples/blanks (see, problems with HiSeq4000), but can cause delays in getting sequencing data. A solution to this is to use ligation adapters carrying one index each (double-indexed

libraries) + indexed PCR primer for PCR amplification, or similar set-up allowing to computationally exclude reads with non-matching internal indices post-sequencing. It was suggested ligation biases exist depending on adapters used (SI of Brealey et al. 2020 was cited).

- **Poll**: does your lab sequence on HiSeq 4000?
  - Yes: 40% (2)
  - No: 60% (3)
  - Total Votes: 5

- **Poll**: who uses internal barcodes?
  - We do: 7% (7)
  - We don't: 63% (12)
  - Total Votes: 19
- **A**: Uploading EBC sequence data on public repositories ENA/SRA. It remains unclear what metadata label should be applied to make them easily identifiable by others (e.g., ENA "Scientific Name") ?

How to limit the impact of pre-laboratory contamination?

- **Q**: Should we decontaminate samples pre-extraction?
  - **A**: There are several methods with pros and cons for hard substrates (e.g., bone, teeth, calculus), but for other sample types, it can be a challenge, or even impossible, e.g., soil and sediments. Suggested solution: avoid/reduce contamination when sampling (and storing?).
    - **A**: especially when working with sediment cores that are pulled through the water, the only option to prevent pre-laboratory contamination seems to be using clean sampling procedures.
- **Q**: What are people using as sampling kits?
  - **A**: It depends on sampling conditions: where the sampling takes place, what kinds of samples are collected, how fast the remains need to be sampled. On archaeological sites and museums, wear equipment (e.g., gloves, masks, sleeves, face shields, hair cap, full suit, other?) and all tools/surfaces are pre-UVed and bleached (e.g., core liners, cutting tools, other?). Scrape-off dirt from remains on site (some in the field disagree). (Sediment) cores sub-sampled in the clean lab. When many individuals have to be sampled in a short amount of time, it can be hard to always follow strict rules in terms of tool cleanliness. (Should we definitely exclude potentially interesting samples that have been collected in sub-optimal conditions? maybe not). In all cases, shared opinion that detailed sampling protocols should be included in Material and Method sections.
    - **A**: Note that bleaching can cause degradation of equipment, so other DNA removal reagents should also be used.
    - **A**: Something else to consider here are the storage conditions: if a sample was collected suboptimally, and then also stored suboptimally, such a sample might not be very useful anymore.

Should we use UDG treatment in ancient metagenomics?

- **Poll**: do you use UDG treatment of DNA extracts prior to library construction?
    - Yes: 23% (5)
    - No: 50% (11)
    - Sometimes: 27% (6)
    - Total Votes: 22
- **Q**: What are the pros and cons of UDG treatment?
    - **A**: Repair damage-derived nucleotide misincorporation sequence errors
    - **A**: Con, many working with environmental DNA never use UDG treatment because cytosine deamination signatures are used to authenticate data.
- **Q**: What are the alternatives?
    - **A**: Follow a 2 step-procedure: 1) screen libraries built on raw DNA extracts (not UDG treated); 2) deep-sequence on UDG-treated DNA extracts. Con, it costs (in time and DNA extract) to build 2 libraries.
    - **A**: Do not treat with UDG, then sequence more deeply, then eventually clip read ends.
    - **A**: Treat with UDG and identify contaminants differently from detecting C>T and G>A misincorporation? Proposed to be discussed in Session 4 (but not discussed at the end).
    - **A**: Use other types of enzymes than UDG, such as 2 enzyme repair-systems? Pfu polymerase for library amplification? The question was raised whether their impact has been tested systematically?
- **Poll**: do your lab use Phusion-like enzymes?
    - Yes: 20% (1)
    - No: 80% (4)
    - Total Votes: 5

To finish-up: Ideas for common SPAAM projects? (see above for details)

- Investigating biases due to extraction methods (~"wet Velsko2018"?): try out main extraction protocols and compare (damage, microbial community compositions) starting from calculus? teeth? Environmental samples? Others?
- Investigating biases due to lab-specific contamination.
- Comparing blank data from different labs (need to be reviewed: what blank sequence data is available? how have they been generated?).