H3ABioNet

Pan African Bioinformatics Network for H3Africa

16SrRNA Intermediate Bioinformatics Online Course: Int_BT

Module 3:

Sample collection, extraction and library prep for 16S NGS analyses

Part 3.3

16S rRNA high throughput sequencing: Sample processing

16S rRNA high throughput sequencing: sample processing

Journal of Oral Microbiology

COACTION PUBLISHING

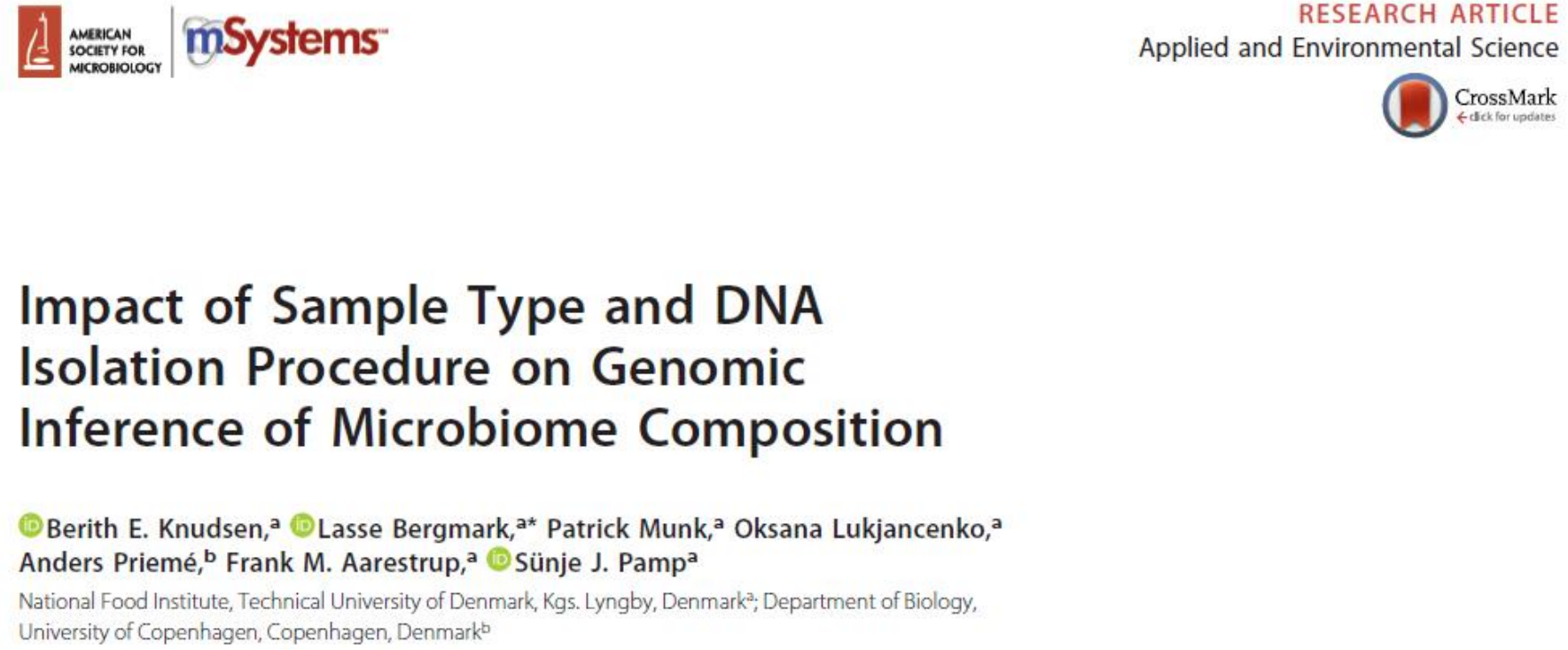
ORIGINAL ARTICLE

Influence of DNA extraction on oral microbial profiles obtained via 16S rRNA gene sequencing

Loreto Abusleme[1,2†], Bo-Young Hong[1†], Amanda K. Dupuy[3], Linda D. Strausbaugh[3] and Patricia I. Diaz[1*]

[1]Division of Periodontology, Department of Oral Health and Diagnostic Sciences, The University of Connecticut Health Center, Farmington, CT, USA; [2]Laboratory of Oral Microbiology, Faculty of Dentistry, University of Chile, Santiago, Chile; [3]Department of Molecular and Cell Biology, The Center for Applied Genetics and Technologies, The University of Connecticut, Storrs, CT, USA

16S rRNA high throughput sequencing: sample processing



Journal of Oral M

AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems

RESEARCH ARTICLE
Applied and Environmental Science

CrossMark click for updates

ORIG
Infl
obt

Lore
Linc

[1]Divis
Conn
Denti
for A

Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition

Berith E. Knudsen,[a] Lasse Bergmark,[a*] Patrick Munk,[a] Oksana Lukjancenko,[a] Anders Priemé,[b] Frank M. Aarestrup,[a] Sünje J. Pamp[a]

National Food Institute, Technical University of Denmark, Kgs. Lyngby, Denmark[a]; Department of Biology, University of Copenhagen, Copenhagen, Denmark[b]

16S rRNA high throughput sequencing: sample processing

OPEN

Methodology challenges in studying human gut microbiota – effects of collection, storage, DNA extraction and next generation sequencing technologies

Marina Panek[1], Hana Čipčić Paljetak[1], Anja Barešić[2], Mihaela Perić[1], Mario Matijašić[1], Ivana Lojkić[3], Darija Vranešić Bender[4], Željko Krznarić[5] & Donatella Verbanac[1]

Received: 6 November 2017
Accepted: 9 March 2018
Published online: 23 March 2018

16S rRNA high throughput sequencing: sample processing

OPEN ACCESS Freely available online

PLOS ONE

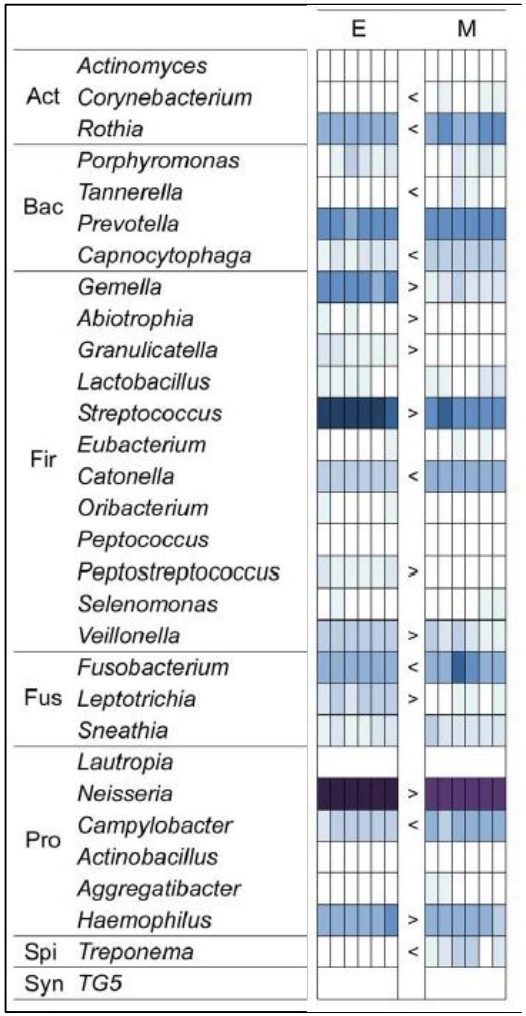
Comparison of DNA Extraction Methods in Analysis of Salivary Bacterial Communities

Vladimir Lazarevic*, Nadia Gaïa, Myriam Girard, Patrice François, Jacques Schrenzel

Genomic Research Laboratory, Division of Infectious Diseases, Geneva University Hospitals, Geneva, Switzerland

Abstract

Culture-independent high-throughput sequencing-based methods are widely used to study bacterial communities. Although these approaches are superior to traditional culture-based methods, they introduce bias at the experimental and bioinformatics levels. We assessed the diversity of the human salivary microbiome by pyrosequencing of the 16S rDNA V1–3 amplicons using metagenomic DNA extracted by two different protocols: a simple proteinase K digestion without a subsequent DNA clean-up step, and a bead-beating mechanical lysis protocol followed by column DNA purification. A high degree of congruence was found between the two extraction methods, most notably in regard to the microbial community composition. The results showed that for a given bioinformatics pipeline, all the taxa with an average proportion >0.12% in samples processed using one extraction method were also detected in samples extracted using the other method. The same taxa tended to be abundant and frequent for both extraction methods. The relative abundance of sequence reads assigned to the phyla Actinobacteria, Spirochaetes, TM7, Synergistetes, and Tenericutes was significantly higher in the mechanically-treated samples than in the enzymatically-treated samples, whereas the phylum Firmicutes showed the opposite pattern. No significant differences in diversity indices were found between the extraction methods, although the mechanical lysis method revealed higher operational taxonomic unit richness. Differences between the extraction procedures outweighed the variations due to the bioinformatics analysis pipelines used.



16S rRNA high throughput sequencing: sample processing

RESEARCH ARTICLE

Back to Basics – The Influence of DNA Extraction and Primer Choice on Phylogenetic Analysis of Activated Sludge Communities

Mads Albertsen, Søren M. Karst, Anja S. Ziegler, Rasmus H. Kirkegaard, Per H. Nielsen*

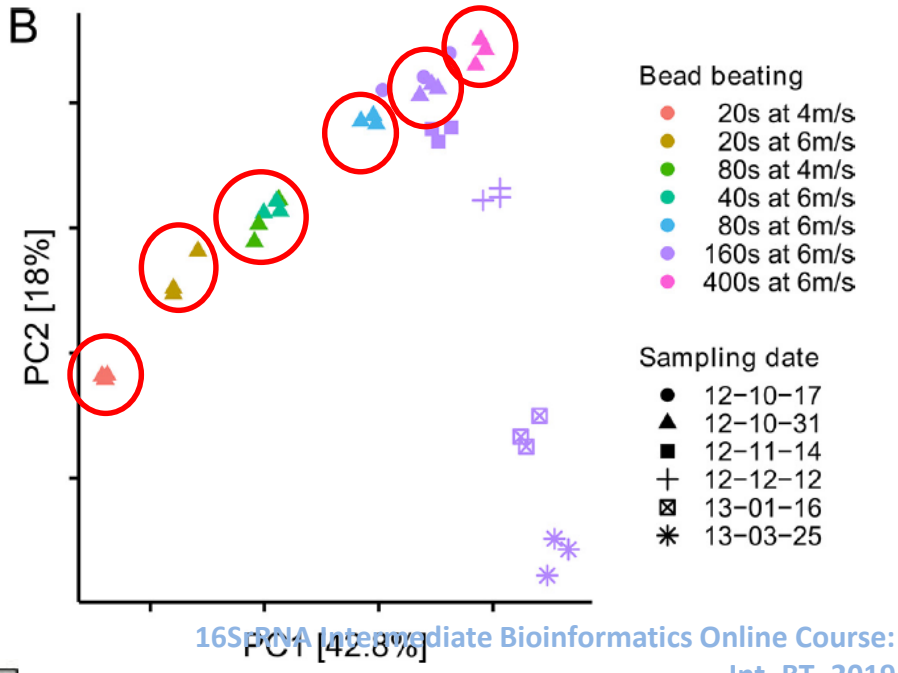
Albertsen et al., (2015), *PLoS ONE* 10(7): e0132783

Increased bead beating had a dramatic influence on the observed community composition

A

	20s 4m/s	20s 6m/s	80s 4m/s	40s 6m/s	80s 6m/s	160s 6m/s	400s 6m/s
Betaproteobacteria	35.5	34.4	33.4	31.4	26	24.5	22.6
Actinobacteria	5.6	7	10.4	10	13.8	17.6	21.7
Chloroflexi	8.5	10	12.3	12.3	14.3	14	13.2
Alphaproteobacteria	5.9	7.4	7.8	8.1	9.7	10.4	11.7
Bacteroidetes	18.9	16.4	12.8	13.7	11.7	9.9	8.6
Firmicutes	2.9	3.4	3.9	4.4	5.4	5.8	5.3
Deltaproteobacteria	3.5	2.9	2.6	2.8	2.6	2.6	2.4
Acidobacteria	1.9	1.9	1.7	1.8	2	1.9	2
Gammaproteobacteria	2.7	2.7	2.5	2.4	2.3	2	1.9
Nitrospirae	1	1.8	2.1	2.2	2	1.8	1.6
Chlorobi	2.8	2.7	2.2	2.3	2.1	1.8	1.6

Compared to the time series samples, the effect of bead beating was larger than the effect of sampling 5 months apart.



16S rRNA high throughput sequencing: sample processing

METHODOLOGY Open Access

Quantification of variation and the impact of biomass in targeted 16S rRNA gene sequencing studies

Jeffrey M. Bender[1†], Fan Li[2†], Helty Adisetiyo[2], David Lee[3], Sara Zabih[3], Long Hung[2], Thomas A. Wilkinson[2], Pia S. Pannaraj[1], Rosemary C. She[4], Jennifer Dien Bard[1], Nicole H. Tobin[3] and Grace M. Aldrovandi[3*]

Abstract

Background: Recent advances in sequencing technologies and bioinformatics tools have allowed for large-scale microbiome studies that are rapidly advancing medical research. However, small changes in technique or analysis can significantly alter the results and lead to conflicting findings. Quantifying the technical versus biological variation expected in targeted 16S rRNA gene sequencing studies and how this variation changes with input biomass is critical to guide meaningful interpretation of the current literature and plan future research.

Results: Data were compiled from 469 sequencing libraries across 19 separate targeted 16S rRNA gene sequencing runs over a 2.5-year time period. Following removal of contaminant sequences identified from negative controls, 244 samples retained sufficient reads for further analysis. Coefficients of variation for intra- and inter-assay variation from repeated measurements of a bacterial mock community ranged from 8.7 to 37.6% (intra) and 15.6 to 80.5% (inter) for all but one genus of bacteria whose relative abundance was greater than 1%. Intra- versus inter-assay Bray-Curtis pairwise distances for a single stool sample were 0.11 versus 0.31, whereas intra-assay variation from repeat stool samples from the same donor was greater at 0.38 (Wilcoxon $p = 0.001$). A dilution series of the bacterial mock community was used to assess the effect of input biomass on variability. Pairwise distances increased with more dilute samples, and estimates of relative abundance became unreliable below approximately 100 copies of the 16S rRNA gene per microliter. Using this data, we created a prediction model to estimate the expected variation in microbiome measurements for given input biomass and relative abundance values.

Conclusions: Well-controlled microbiome studies are sufficiently robust to capture small biological effects and can achieve levels of variability consistent with clinical assays. Relative abundance is negatively associated with measures of variability and has a stronger effect on variability than does absolute biomass, suggesting that it is feasible to detect differences in bacterial populations in very low-biomass samples. Further, by quantifying the effect of biomass and relative abundance on compositional variability, we developed a tool for defining the expected variance in a given microbiome study.

Keywords: Biomass, Technical variation, Biological variation, Precision, Accuracy

16S rRNA high throughput sequencing: sample processing

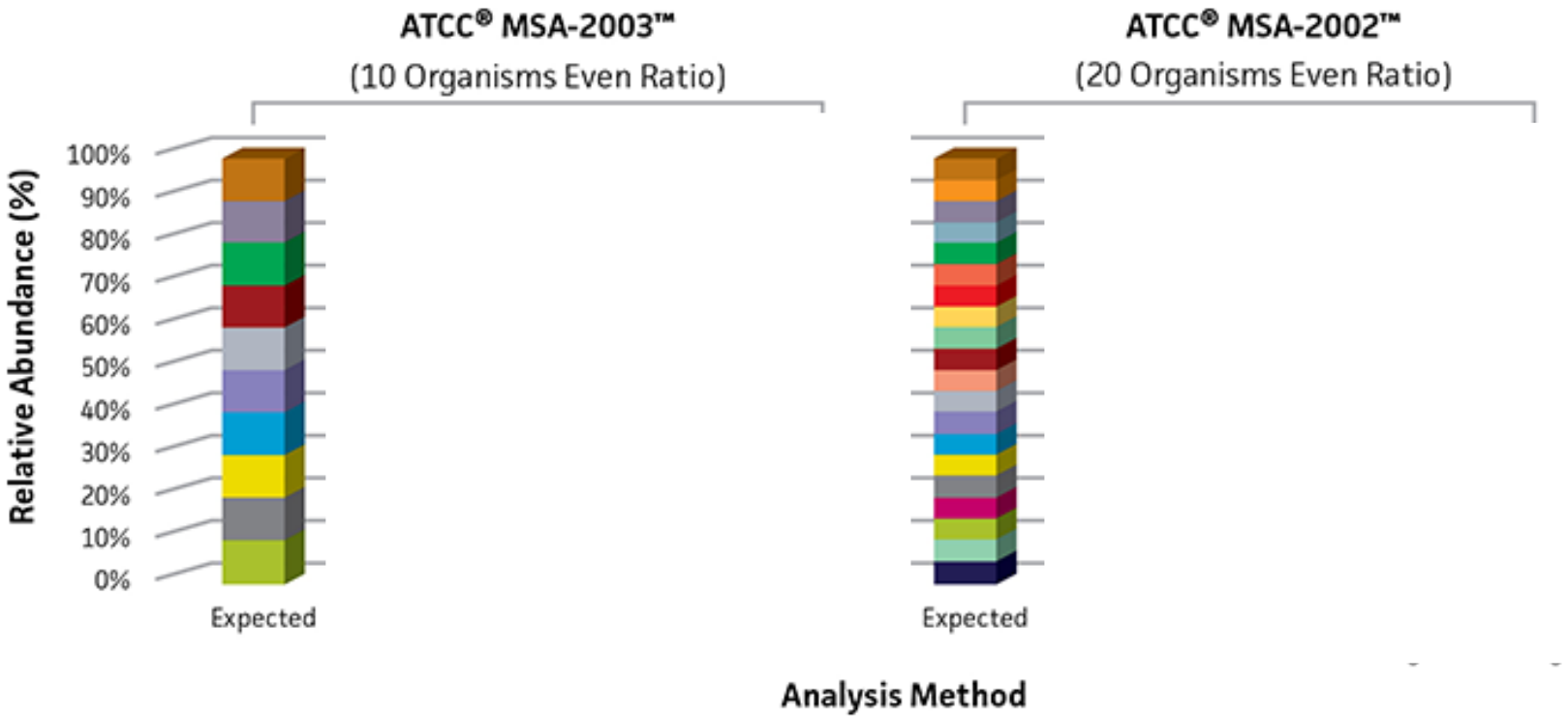


Figure 1. The whole cell mock microbial communities can be used to compare different DNA extraction kits. Here, we show a proof-of-concept study. Total DNA from two aliquots of the whole cell microbiome standards was extracted using two commercial extraction kits. Following extraction, the genomic DNA was analyzed using 16S rRNA and shotgun genomic sequencing analysis on the Illumina® MiSeq® platform, and the resulting data were analyzed using the ATCC bioinformatics tool in One Codex.

http://www.lgcstandards-atcc.org/Microbiome?geo_country=za

16S rRNA high throughput sequencing: sample processing

H3ABioNet

NOT EXACTLY ROCKET SCIENCE: November 11, 2014

Contaminomics: Why Some Microbiome Studies May Be Wrong

by Ed Yong

Source: http://phenomena.nationalgeographic.com/2014/11/11/contaminomics-why-some-microbiome-studies-may-be-wrong/

16S rRNA high throughput sequencing: sample processing

Weiss *et al. Genome Biology* 2014, **15**:564
http://genomebiology.com/2014/15/12/564

Genome Biology

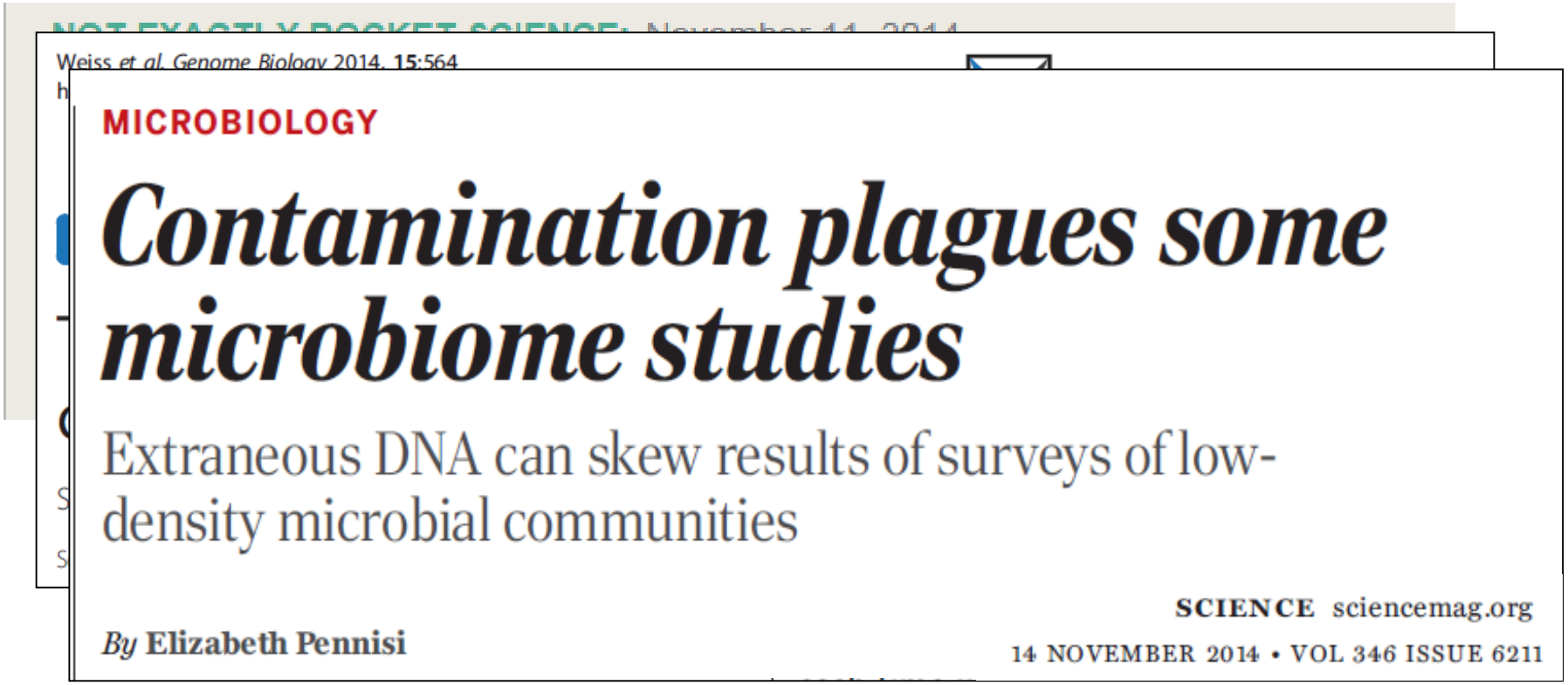
RESEARCH HIGHLIGHT

Tracking down the sources of experimental contamination in microbiome studies

Sophie Weiss[1], Amnon Amir[2], Embriette R Hyde[2], Jessica L Metcalf[2], Se Jin Song[2] and Rob Knight[2,3,4*]

See related research, http://www.biomedcentral.com/1741-7007/12/87

16S rRNA high throughput sequencing: sample processing



Weiss et al. Genome Biology 2014, 15:564

MICROBIOLOGY

Contamination plagues some microbiome studies

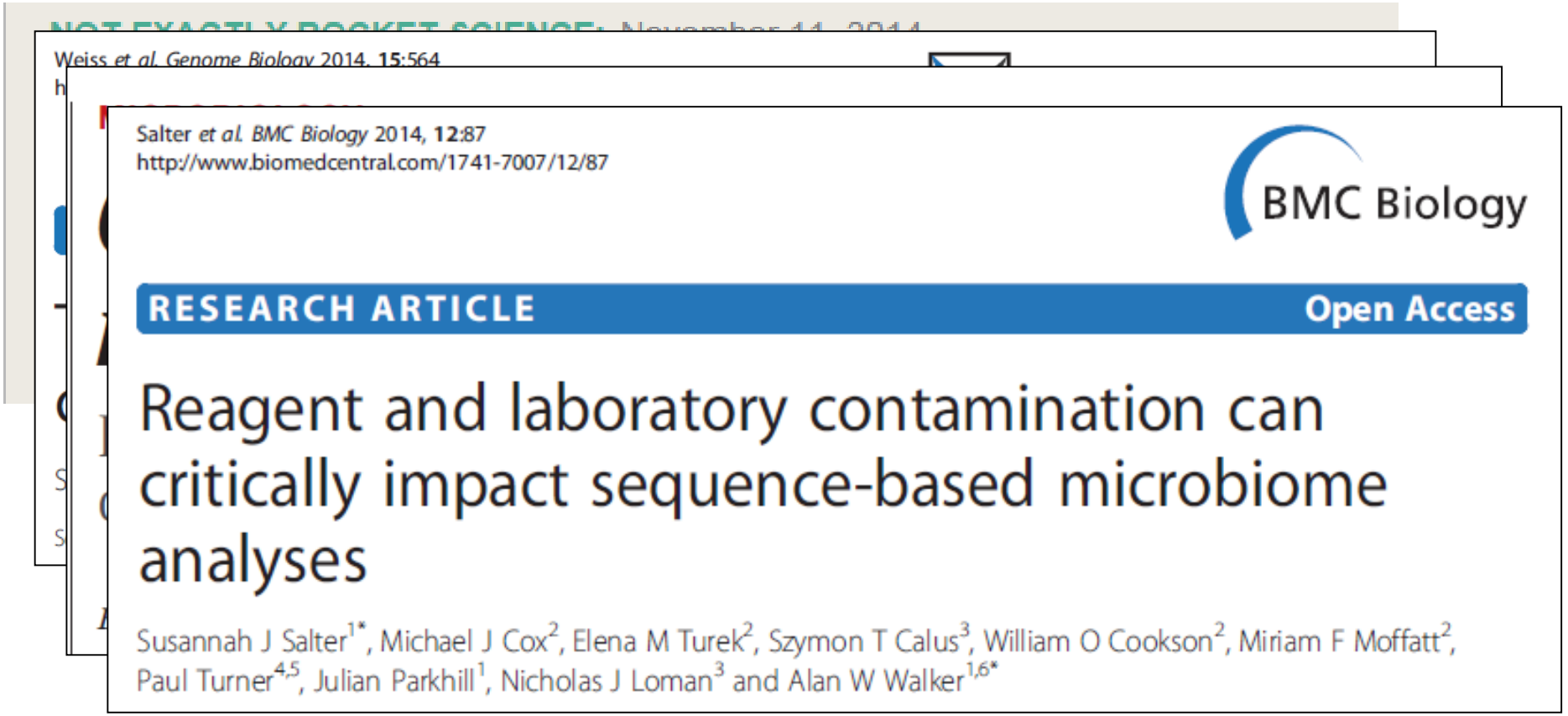
Extraneous DNA can skew results of surveys of low-density microbial communities

By Elizabeth Pennisi

SCIENCE sciencemag.org

14 NOVEMBER 2014 • VOL 346 ISSUE 6211

16S rRNA high throughput sequencing: sample processing



Weiss *et al. Genome Biology* 2014, **15**:564

Salter *et al. BMC Biology* 2014, **12**:87
http://www.biomedcentral.com/1741-7007/12/87

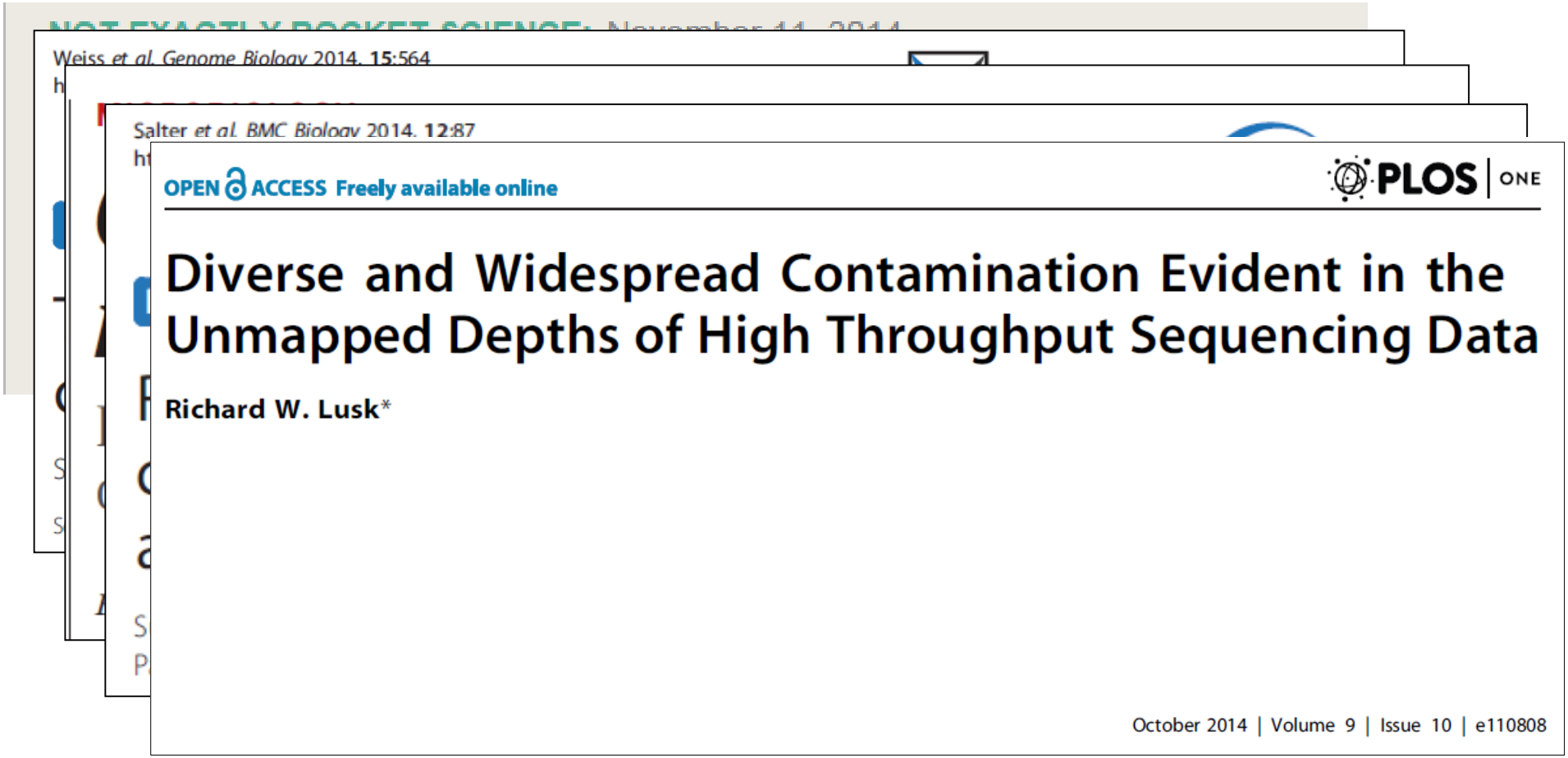
BMC Biology

RESEARCH ARTICLE Open Access

Reagent and laboratory contamination can critically impact sequence-based microbiome analyses

Susannah J Salter[1*], Michael J Cox[2], Elena M Turek[2], Szymon T Calus[3], William O Cookson[2], Miriam F Moffatt[2], Paul Turner[4,5], Julian Parkhill[1], Nicholas J Loman[3] and Alan W Walker[1,6*]

16S rRNA high throughput sequencing: sample processing



16S rRNA high throughput sequencing: sample processing

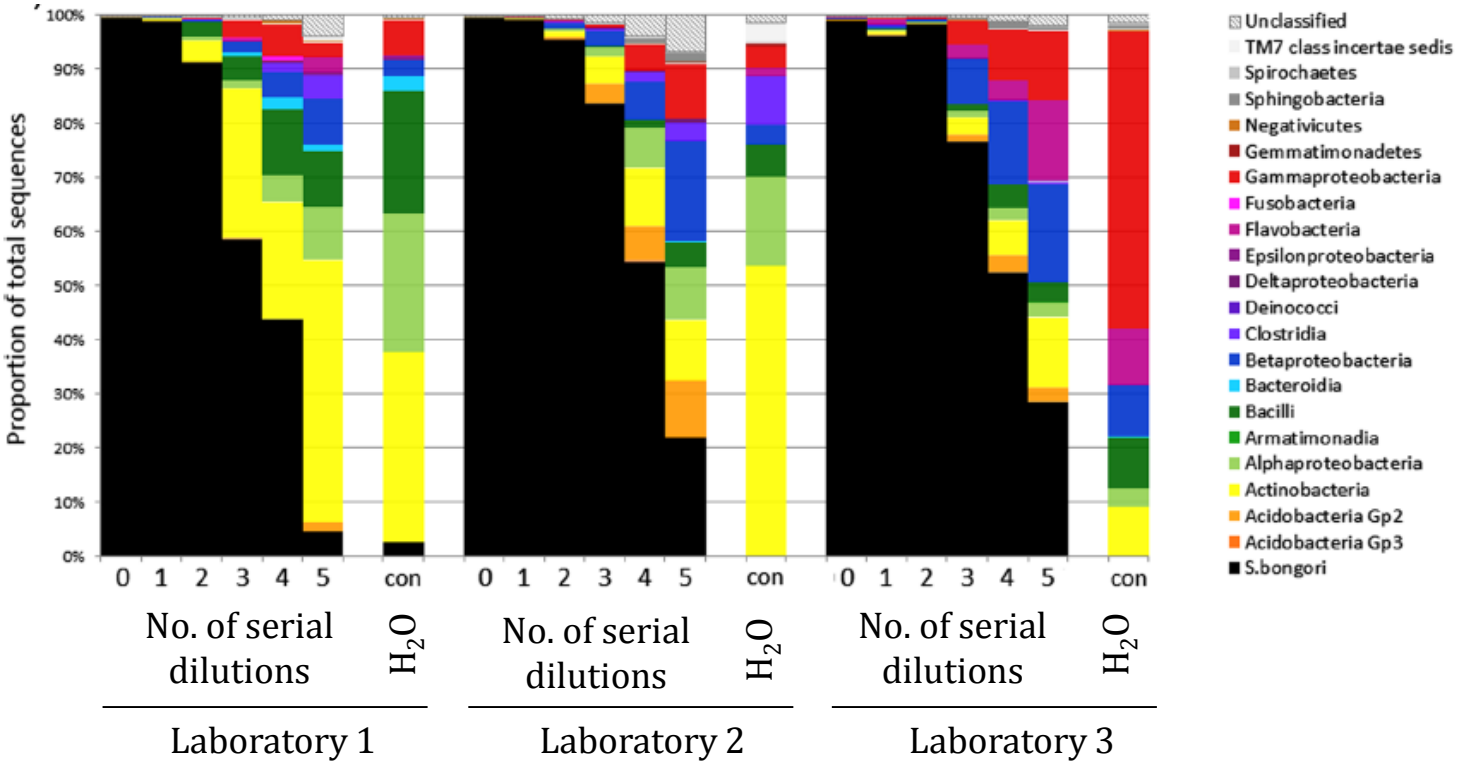
Table 1 List of contaminant genera detected in sequenced negative 'blank' controls

Phylum	List of constituent contaminant genera
Proteobacteria	Alpha-proteobacteria:
	Afipia, Aquabacterium[e], *Asticcacaulis, Aurantimonas, Beijerinckia, Bosea, Bradyrhizobium*[d], *Brevundimonas*[c], *Caulobacter, Craurococcus, Devosia, Hoeflea*[e], *Mesorhizobium, Methylobacterium*[c], *Novosphingobium, Ochrobactrum, Paracoccus, Pedomicrobium, Phyllobacterium*[e], *Rhizobium*[c,d], *Roseomonas, Sphingobium, Sphingomonas*[c,d,e], *Sphingopyxis*
	Beta-proteobacteria:
	Acidovorax[c,e], *Azoarcus*[e], *Azospira, Burkholderia*[d], *Comamonas*[c], *Cupriavidus*[c], *Curvibacter, Delftia*[e], *Duganella*[a], *Herbaspirillum*[a,c], *Janthinobacterium*[e], *Kingella, Leptothrix*[a], *Limnobacter*[e], *Massilia*[c], *Methylophilus, Methyloversatilis*[e], *Oxalobacter, Pelomonas, Polaromonas*[e], *Ralstonia*[b,c,d,e], *Schlegelella, Sulfuritalea, Undibacterium*[e], *Variovorax*
	Gamma-proteobacteria:
	Acinetobacter[a,d,c], *Enhydrobacter, Enterobacter, Escherichia*[a,c,d,e], *Nevskia*[e], *Pseudomonas*[b,d,e], *Pseudoxanthomonas, Psychrobacter, Stenotrophomonas*[a,b,c,d,e], *Xanthomonas*[b]
Actinobacteria	*Aeromicrobium, Arthrobacter, Beutenbergia, Brevibacterium, Corynebacterium, Curtobacterium, Dietzia, Geodermatophilus, Janibacter, Kocuria, Microbacterium, Micrococcus, Microlunatus, Patulibacter, Propionibacterium*[e], *Rhodococcus, Tsukamurella*
Firmicutes	*Abiotrophia, Bacillus*[b], *Brevibacillus, Brochothrix, Facklamia, Paenibacillus, Streptococcus*
Bacteroidetes	*Chryseobacterium, Dyadobacter, Flavobacterium*[d], *Hydrotalea, Niastella, Olivibacter, Pedobacter, Wautersiella*
Deinococcus-Thermus	*Deinococcus*
Acidobacteria	Predominantly unclassified Acidobacteria Gp2 organisms

The listed genera were all detected in sequenced negative controls that were processed alongside human-derived samples in our laboratories (WTSI, ICL and UB) over a period of four years. A variety of DNA extraction and PCR kits were used over this period, although DNA was primarily extracted using the FastDNA SPIN Kit for Soil. Genus names followed by a superscript letter indicate those that have also been independently reported as contaminants previously. [a]also reported by Tanner *et al.* [12]; [b]also reported by Grahn *et al.* [14]; [c]also reported by Barton *et al.* [17]; [d]also reported by Laurence *et al.* [18]; [e]also detected as contaminants of multiple displacement amplification kits (information provided by Paul Scott, Wellcome Trust Sanger Institute). ICL, Imperial College London; UB, University of Birmingham; WTSI, Wellcome Trust Sanger Institute.

Salter et al., (2014), *BMC Biology*; **12**; 1-12

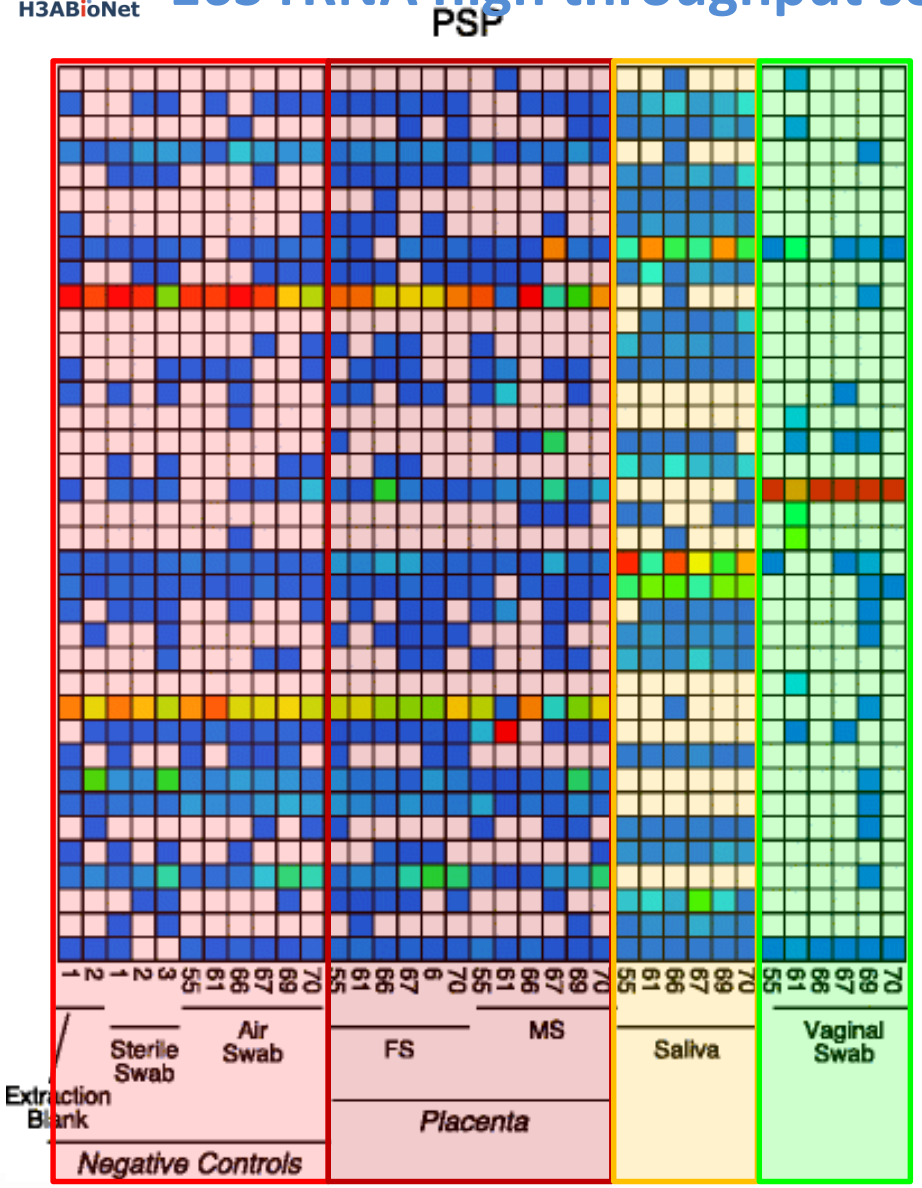
16S rRNA high throughput sequencing: sample processing

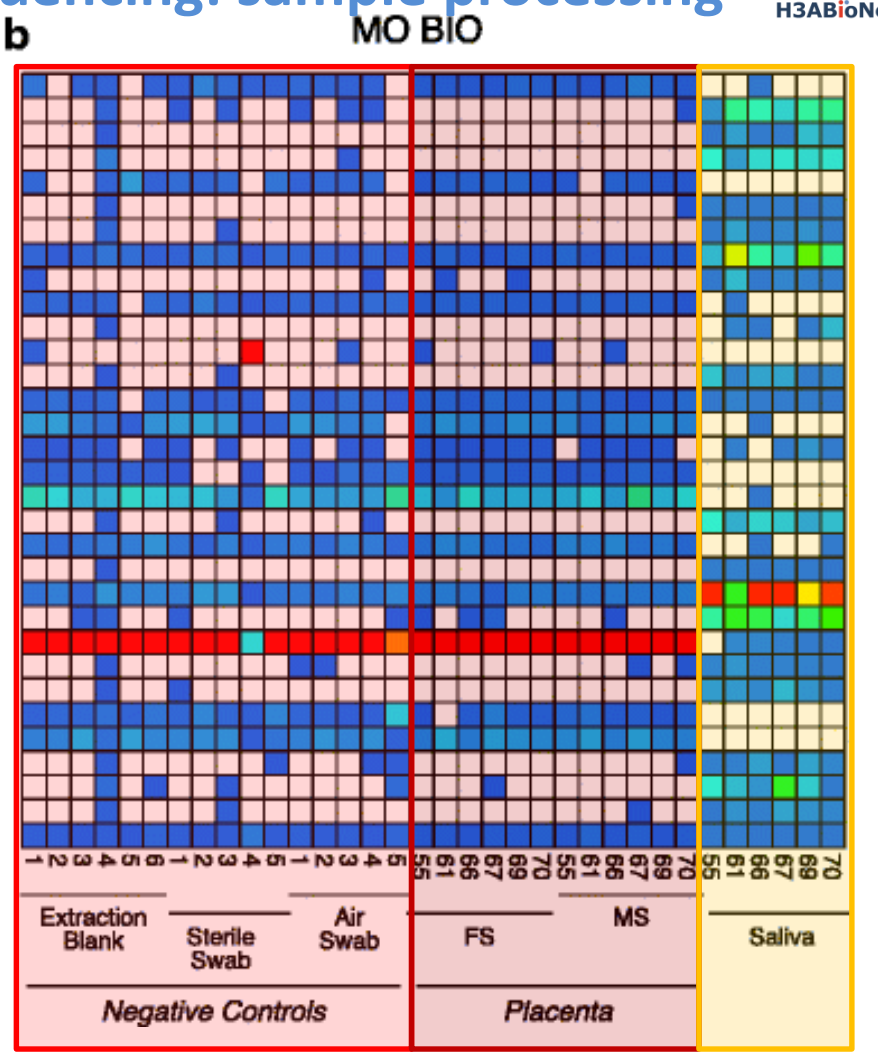


Targeted the 16S rRNA gene of *Salmonella bongori* (in black) which had undergone five rounds of serial ten-fold dilutions

Salter et al., (2014), *BMC Biology*; **12**; 1-12

16S rRNA high throughput sequencing: sample processing





Lauder et al. (2016) Microbiome. 4(29) https://doi.org/10.1186/s40168-016-0172-3

16S rRNA high throughput sequencing: sample processing

H3ABioNet

Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data

Nicole M. Davis[1], Diana Proctor[2,6], Susan P. Holmes[3], David A. Relman[1,2,4]*, Benjamin J. Callahan[5]*^

* Equal contributions
^ Corresponding: benjamin.j.callahan@gmail.com

1. Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305

2. Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305

3. Department of Statistics, Stanford University, Stanford, CA 94305

4. Infectious Diseases Section, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA 94304

5. Department of Population Health and Pathobiology, College of Veterinary Medicine, North Carolina State University, Raleigh, NC 27607

6. Department of Orofacial Sciences, University of California, San Francisco School of Dentistry, San Francisco, CA 94143

16S rRNA high throughput sequencing: sample processing

The 16S rRNA real time PCR protocol is published at:
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0017035

301460	F05	Unkn		24.26819	0.116
Primestore_FTDint_P1 (SC 9 B8)	F06	Unkn	Primestore_int	36.03133	0.000
502527	F07	Unkn		29.54712	0.003
101398	F08	Unkn		29.55681	0.003
102598	F09	Unkn		23.255	0.233
104454	F10	Unkn		24.64554	0.089
518306	F11	Unkn		22.84429	0.310
302194	F12	Unkn		21.13398	1.015
102539	G03	Unkn		33.36721	0.000
104435	G04	Unkn		21.43081	0.910
118685	G03	Unkn		24.20348	0.121
104212	G04	Unkn		19.27805	3.674
108148	G05	Unkn		24.50339	0.098
w ater	G06	Unkn	MilliQ	35.66736	0.000
505016	G07	Unkn		35.65833	0.000
103051	G08	Unkn		27.24958	0.015
104614	G09	Unkn		19.42097	3.327
108290	G10	Unkn		29.70401	0.003
512561	G11	Unkn		24.04204	0.135
303017	G12	Unkn		20.00377	2.222
507636	H03	Unkn		21.52387	0.854
508423	H04	Unkn		25.33993	0.066
506335	water	Unkn	MilliQ	36.39297	0.000
301009	H04	Unkn		21.97896	0.565

16S rRNA high throughput sequencing: sample processing

Microbiology: making the best of PCR bias

Michael Eisenstein

Many factors can skew the results of a widely used amplification technique for microbiome analysis, but researchers are finding strategies for getting at the truth.

https://www.nature.com/articles/nmeth.4683.pdf?origin=ppub

16S rRNA high throughput sequencing: sample processing

SCIENTIFIC REPORTS

OPEN

Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract

Received: 21 November 2017
Accepted: 8 June 2018
Published online: 26 June 2018

Simon Graspeuntner[1], Nathalie Loeper[1], Sven Künzel[2], John F. Baines[2,3] & Jan Rupp[1,4]

Next-generation sequencing-based methods are extensively applied in studies of the human microbiota using partial 16 S rRNA gene amplicons. However, they carry drawbacks that are critical to consider when interpreting results, including differences in outcome based on the hypervariable region(s) used. Here, we show that primers spanning the V3/V4 region identify a greater number of taxa in the vaginal microbiota than those spanning the V1/V2 region. In particular, taxa such as *Gardnerella vaginalis*, *Bifidobacterium bifidum* and *Chlamydia trachomatis*, all species that influence vaginal health and disease, are not represented in V1/V2-based community profiles. Accordingly, missing or underestimating the frequency of these species overestimates the abundance of other taxa and fails to correctly assess the bacterial diversity in the urogenital tract. We elaborate that covering these taxa using the V3/V4 region leads to profound changes in the assignment of community state types. Altogether, we show that the choice of primers used for studying the vaginal microbiota has deep implications on the biological evaluation of the results.

16S rRNA high throughput sequencing: sample processing

METHODOLOGY ARTICLE Open Access

The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies

J Paul Brooks[1,2*], David J Edwards[1], Michael D Harwich Jr[3], Maria C Rivera[4], Jennifer M Fettweis[3], Myrna G Serrano[2,3], Robert A Reris[1], Nihar U Sheth[2], Bernice Huang[3], Philippe Girerd[5], Vaginal Microbiome Consortium (additional members), Jerome F Strauss III[5], Kimberly K Jefferson[2,3] and Gregory A Buck[2,3]

Abstract

Background: Characterizing microbial communities via next-generation sequencing is subject to a number of pitfalls involving sample processing. The observed community composition can be a severe distortion of the quantities of bacteria actually present in the microbiome, hampering analysis and threatening the validity of conclusions from metagenomic studies. We introduce an experimental protocol using mock communities for quantifying and characterizing bias introduced in the sample processing pipeline. We used 80 bacterial mock communities comprised of prescribed proportions of cells from seven vaginally-relevant bacterial strains to assess the bias introduced in the sample processing pipeline. We created two additional sets of 80 mock communities by mixing prescribed quantities of DNA and PCR product to quantify the relative contribution to bias of (1) DNA extraction, (2) PCR amplification, and (3) sequencing and taxonomic classification for particular choices of protocols for each step. We developed models to predict the "true" composition of environmental samples based on the observed proportions, and applied them to a set of clinical vaginal samples from a single subject during four visits.

Results: We observed that using different DNA extraction kits can produce dramatically different results but bias is introduced regardless of the choice of kit. We observed error rates from bias of over 85% in some samples, while technical variation was very low at less than 5% for most bacteria. The effects of DNA extraction and PCR amplification for our protocols were much larger than those due to sequencing and classification. The processing steps affected different bacteria in different ways, resulting in amplified and suppressed observed proportions of a community. When predictive models were applied to clinical samples from a subject, the predicted microbiome profiles were better reflections of the physiology and diagnosis of the subject at the visits than the observed community compositions.

Conclusions: Bias in 16S studies due to DNA extraction and PCR amplification will continue to require attention despite further advances in sequencing technology. Analysis of mock communities can help assess bias and facilitate the interpretation of results from environmental samples.

Keywords: Assessments of microbial community structure via metagenomics, DNA extraction bias, PCR bias, Quality control, Next generation sequencing

16S rRNA high throughput sequencing: sample processing

H3ABioNet

AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

COMMENTARY
Applied and Environmental Science

A Lot on Your Plate? Well-to-Well Contamination as an Additional Confounder in Microbiome Sequence Analyses

AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE
Novel Systems Biology Techniques

Quantifying and Understanding Well-to-Well Contamination in Microbiome Research

Jeremiah J. Minich,[a] Jon G. Sanders,[b] Amnon Amir,[b] Greg Humphrey,[b] Jack A. Gilbert,[c,d] Rob Knight[b,e,f,g]

16S rRNA high throughput sequencing: sample processing

In summary:

- Different DNA extraction methods may result in differences in bacterial profiles obtained following sequencing.
- Identify (and/or optimize) an extraction method best suited for your sample types and research questions – and stick to it.
- Low DNA yields may result from the extraction method used or the nature of the biological specimen. These are at higher risk for increased levels of “background” or “contaminant” profiles.
- Contaminant profiles may be different for different laboratories and reagents. Hence, ensure that you have optimal negative controls processed alongside your biological specimens.
- Be aware of PCR bias and cross-contamination between wells.
- Finally, explore different methods for identifying and removing contaminants.