DNA extraction for streamlined metagenomics of diverse environmental samples

Clarisse Marotz*, Amnon Amir*, Greg Humphrey*, James Gaffney*, Grant Gogul*, and Rob Knight
Department of Pediatrics, University of California, San Diego, La Jolla, CA

*C.M., A.A., G.H., J.G., and G.G. contributed equally to this work.

BioTechniques 62:290-293 (June 2017) doi 10.2144/000114559
Keywords: DNA extraction; metagenomics; high throughput; microbiome

A major bottleneck for metagenomic sequencing is rapid and efficient DNA extraction. Here, we compare the extraction efficiencies of three magnetic bead–based platforms (KingFisher, epMotion, and Tecan) to a standardized column-based extraction platform across a variety of sample types, including feces, oral, skin, soil, and water. Replicate sample plates were extracted and prepared for *16S rRNA* gene amplicon sequencing in parallel to assess extraction bias and DNA quality. The data demonstrate that any effect of extraction method on sequencing results was small compared with the variability across samples; however, the KingFisher platform produced the largest number of high-quality reads in the shortest amount of time. Based on these results, we have identified an extraction pipeline that dramatically reduces sample processing time without sacrificing bacterial taxonomic or abundance information.

As microbiome analyses become applicable to an increasing number of scientific areas, a streamlined process for efficiently extracting DNA to generate *16S rRNA* gene amplicons or shotgun metagenomic sequencing data from a range of environmental sample types is increasingly important. DNA extraction is among the largest sources of experimental variability in *16S rRNA* sequence analysis (1–3). Complete bacterial lysis and DNA purification can be particularly time-consuming, representing a significant bottleneck for high-throughput analyses. However, certain time-sensitive analyses of large sample sizes (e.g., clinical diagnostics or water quality assessments) require expedited processing. We aimed to identify a DNA extraction method that yields faster results with no loss of taxonomic representation by comparing magnetic bead–based extraction robots with our benchmarked column-based extraction protocol.

To compare the extraction efficiency of several automated methods, we collected biological materials from a wide range of environmental samples. Replicate plates were prepared from aliquots of identical samples for comparison of extraction efficiency among different platforms. In total, 48 fecal samples, 12 soil samples, 12 marine sediment samples, 6 seawater samples, 5 skin samples, 5 oral samples, and 6 mattress dust samples were included. Material for extraction was collected by swabbing each sample with Puritan wooden handle cotton swabs according to the Earth Microbiome Project standard protocol (www.earthmicrobiome.org/emp500/emp500-sample-submission-guide/). Fecal samples comprised 2 mL homogenized human stool samples obtained from the Microbiome Quality Control Project (MBQC). Soil samples were selected from a published mouse decomposition study (4). Marine sediment samples were donated by Paul Jensen and collected at a depth of either 300 m or 700 m off the coast of San Diego. Seawater samples (4 L each) were collected off the Scripps Pier in San Diego, CA and filtered through a 0.22 µm Sterivex cartridge; the filters were then cut into segments and used for extraction (5). Oral and saliva samples were collected using five swabs simultaneously (one for each extraction method) as previously described (http://americangut.org/how-it-works/) (6) from a total of six volunteers over three time points. Finally, three mattresses were vacuumed in duplicate, and the dust collected was swabbed for extraction. *Vibrio fischerii* ES114 cells (~10^6) were used as a positive control, and 1 blank well (no input material) was used as a negative control. Together, these samples represent a wide range of biological materials, including

METHOD SUMMARY

Comparing three magnetic bead–based genomic DNA extraction platforms to the standard column-based extraction protocol, we identified a rapid, high-throughput pipeline for extraction of genomic DNA from a variety of environmental samples for *16S rRNA* gene amplicon sequencing analyses.

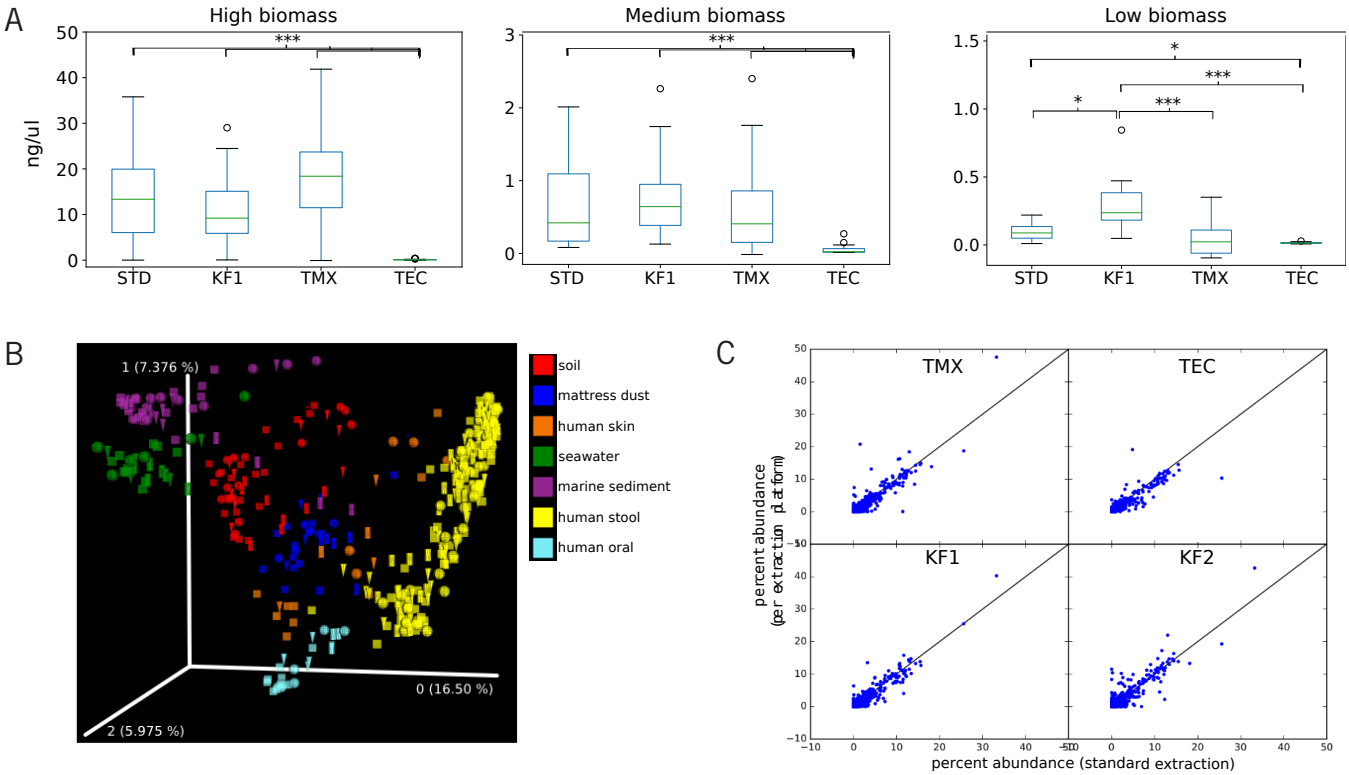

Figure 1. Inter-sample variability outweighs extraction method bias. (A) Average concentration of DNA (ng/µl) across extraction platforms. Statistical analysis was performed using the non-parametric binomial two-sided sign test; *<0.05; **<0.01; ***<0.001. Data are shown in a standard Tukey box plot, where the whiskers represent data within a 1.5 interquartile range (IQR). (B) Unweighted principal coordinates analysis (PCoA) of all samples colored by sample type. Extraction method is denoted by shape: sphere = column-based, cube = KingFisher, cylinder = Tecan, cone = TMX. (C) Scatter plot showing taxonomic abundance differences between magnetic bead–based extraction platforms (*y*-axis) and the standard column-based extraction method (*x*-axis). TMX: epMotion TMX; TEC: Tecan, KF1: KingFisher; KF2: KingFisher with overnight pause.

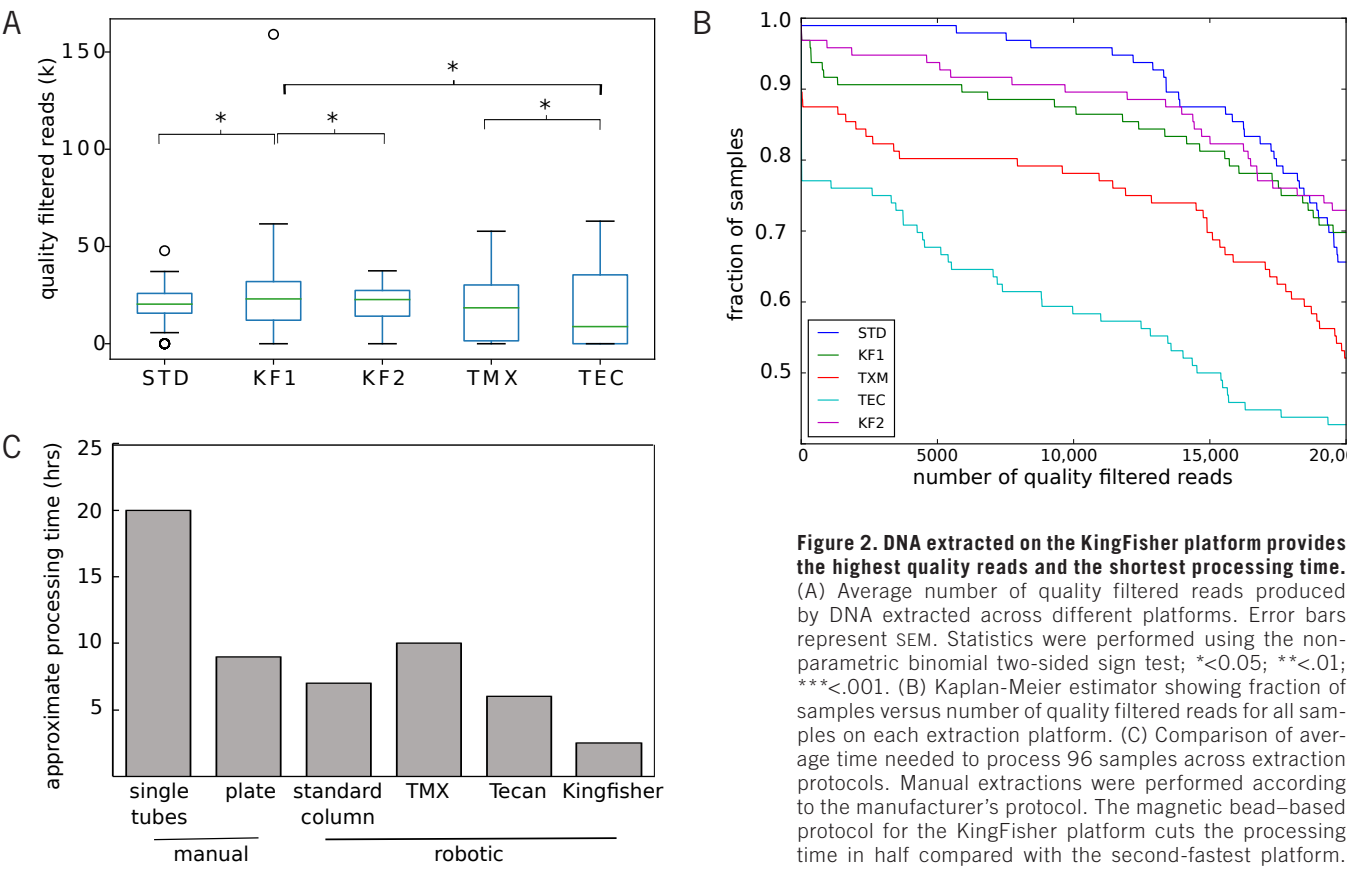

Figure 2. DNA extracted on the KingFisher platform provides the highest quality reads and the shortest processing time. (A) Average number of quality filtered reads produced by DNA extracted across different platforms. Error bars represent SEM. Statistics were performed using the non-parametric binomial two-sided sign test; *<0.05; **<.01; ***<.001. (B) Kaplan-Meier estimator showing fraction of samples versus number of quality filtered reads for all samples on each extraction platform. (C) Comparison of average time needed to process 96 samples across extraction protocols. Manual extractions were performed according to the manufacturer's protocol. The magnetic bead–based protocol for the KingFisher platform cuts the processing time in half compared with the second-fastest platform.

some (soil and fecal samples) that had already been well-characterized using widely accepted protocols. This diverse sample set permitted a broad assessment of the efficiencies of the different extraction pipelines.

We adapted the MoBio PowerMag Soil DNA isolation kit (Qiagen, Carlsbad, CA) for use with a magnetic bead plate in place of the silica membrane column. This replacement significantly reduces sample preparation time by eliminating the need to sequentially load the silicon membrane three times for each sample. Importantly, it requires appreciably less hands-on time, and total processing time is also reduced. Replicate plates were extracted in parallel to compare the standard column-based method to three magnetic bead-based nucleic acid purification platforms: epMotion 5075 TMX (Eppendorf, Hamburg, Germany), KingFisher Flex Purification System (ThermoFisher Scientific, Waltham, MA), and Tecan Freedom EVO Nucleic Acid Purification (Tecan, Morrisville, NC). An additional KingFisher run (KF2) was included, with an overnight incubation at 4°C immediately prior to magnetic bead clean-up to test the processing flexibility of this system. All samples were eluted in 100 µl PCR-grade H_2O. DNA yield was broadly similar across extraction platforms except for the Tecan system, where the yield was lower (Figure 1A). Notably, the Kingfisher extraction platform yielded the most gDNA from low biomass samples. DNA was prepared for *16S rRNA* gene amplicon sequencing as previously described (7). The demultiplexed fasta files were obtained from qiita.ucsd.edu (study ID 10178), and operational taxonomic units (OTUs) were detected using Deblur (8). Raw reads were submitted to the European Bioinformatics Institute under accession no. ERP021045.

We found that the bias introduced by extraction method is small compared with inter-sample variation (average distance across biological replicates, extraction method, and sample type, respectively: weighted UniFrac 0.11 ± 0.001, 0.19 ± 0.004, 0.29 ± 0.001; unweighted UniFrac 0.44 ± 0.01, 0.51 ± 0.003, 0.63 ± 0.001; Bray-Curtis 0.22 ± 0.02, 0.38 ± 0.007, 0.71 ± 0.003). Confirming this, principal coordinates analysis (PCoA) of unweighted UniFrac distance matrices revealed that sequences clustered by sample type rather than extraction platform (Figure 1B) (9). OTU abundance across the magnetic extraction platforms revealed broadly similar results to the standard column-based protocol, even when the KingFisher extraction protocol was paused overnight before bead clean-up (Figure 1C). Alpha diversity levels per sample were remarkably similar (within 2% of the average across all extraction platforms), with the exception of samples extracted with the Tecan platform, which produced on average 92% ± 2% of the average observed diversity. These results demonstrate that the taxonomic and community-level variation contributed by the different magnetic based extraction protocols was minor.

However, the number of quality sequencing reads across all sample types was highest in the KingFisher-

extracted samples (Figure 2, A and B). Importantly, the KingFisher protocol requires only a fraction of the time required by the other extraction platforms. This streamlined extraction pipeline saves an average of 72 h per 192 samples compared with single-tube manual extraction and cuts the processing time in half compared with the second fastest method (Figure 2C).

In conclusion, we prepared replicate 96-well plates loaded with swabs from a variety of sources including feces, skin, oral, soil, and water. The MoBio PowerMag Soil DNA isolation kit was adapted for comparison of three magnetic bead isolation platforms for representative *16S rRNA* sequencing results. Ultimately, we found the KingFisher Flex Purification system was the fastest, the most efficient for low biomass samples, and retained the largest number of high-quality amplicon sequencing reads. This streamlined protocol reduces the total time of DNA extraction to one-fourth of the original protocol while providing comparable sequencing results.

Author contributions

A.A., G.H., and R.K. conceived the idea for the study. G.H., J.G., and G.G. performed the experiments. C.M. and A.A. analyzed the data and drafted the manuscript. All authors approved the final manuscript prior to submission.

Acknowledgments

This work was funded by the John Templeton Foundation, the W.M. Keck foundation, and the Alfred P. Sloan foundation. The authors would like to thank Luke Thompson for contributing seawater samples and editing this manuscript.

Competing interests

The authors declare no competing interests.

References

1. **Wu, G.D., J.D. Lewis, C. Hoffmann, Y.Y. Chen, R. Knight, K. Bittinger, J. Hwang, J. Chen, et al.** 2010. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. BMC Microbiol. *10*:206.
2. **Yuan, S., D.B. Cohen, J. Ravel, Z. Abdo, and L.J. Forney.** 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. PLoS One *7*:e33865.
3. **Kennedy, N.A., A.W. Walker, S.H. Berry, S.H. Duncan, F.M. Farquarson, P. Louis, and J.M. Thomson.** 2014. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. PLoS One *9*:e88982.
4. **Metcalf, J.L., L.W. Parfrey, A. Gonzalez, C.L. Lauber, D. Knights, G. Ackermann, G.C. Humphrey, M.J. Gebert, et al.** 2013. A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. eLife *2*:e01104.
5. **Gilbert, J.A., D. Field, P. Swift, S. Thomas, D. Cummings, B. Temperton, K. Weynberg, S. Huse, et al.** 2010. The taxonomic and functional diversity of microbes at a temperate coastal site: a 'multi-omic'study of seasonal and diel temporal variation. PLoS One *5*:e15545.
6. **Bouslimani, A., C. Porto, C.M. Rath, M. Wang, Y. Guo, A. Gonzalez, D. Berg-Lyon, G. Ackermann, et al.** 2015. Molecular cartography of the human skin surface in 3D. Proc. Natl. Acad. Sci. USA *112*:E2120-E2129.
7. **Caporaso, J.G., C.L. Lauber, W.A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S.M. Owens, J. Betley, et al.** 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. *6*:1621-1624.
8. **Amir, A., D. McDonald, J.A Navas-Molina, E. Kopylova, J.T. Morton, Z.Z. Xu, E.P. Kightley, L.R. Thompson, et al.** 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems. *2*:e00191-16.
9. **Vázquez-Baeza, Y., M. Pirrung, A. Gonzalez, and R. Knight.** 2013. EMPeror: a tool for visualizing high-throughput microbial community data. Gigascience *2*:16.

Received 23 January 2017; accepted 02 May 2017.

Address correspondence to Rob Knight, Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093. E-mail: robknight@ucsd.edu