Microbiome Analysis Enables Non-invasive Monitoring of Rocky Mountain Elk Populations

Samuel Pannoni[1], Kelly Proffitt[2], and William Holben[3]

[1]University of Montana System
[2]Montana Fish Wildlife and Parks
[3]University of Montana Missoula Division of Biological Sciences

April 28, 2020

Abstract

Rocky Mountain elk (Cervus elaphus nelsoni) seasonal migration, body-condition and sex ratios are important parameters for characterizing elk populations but have thus far been outside the scope of non-invasive methods. Fecal microbiomes can be surveyed non-invasively from scat samples and are associated with changes in diet, stress, age, disease and physical condition of the host, as well as differences between sexes. With this in mind, we surveyed the fecal microbiome of Montana elk that varied geographically (i.e. populations), by body condition, age and by sex. Our goal was to explore an approach for evaluating linkages between the host animal and its microbiome composition, and to develop bioinformatic techniques useful for characterizing host categories and population parameters based on microbiome analysis. We built a supervised-machine learning classifier based on bacterial taxa with cross validation to predict each fecal microbiome's affiliation to known host categories. The microbiome classifier predicted host population, sex, age and body-condition with promising cross validation results. Monitoring wildlife microbiomes represents a breakthrough for non-invasive conservation biology, and we provide proof of concept for obtaining low cost, fine scale, management-relevant information from scat samples.

a breakthrough for non-invasive conservation biology, and we provide proof of concept for obtaining low cost, fine scale, management-relevant information from scat samples.

INTRODUCTION

Wildlife species are threatened by numerous anthropogenic forces that erode the sustainability of populations, creating an increased reliance on management for species survival(Scott, Goble, Haines, Wiens, & Neel, 2010). However, methods for informing management practices are often invasive to the animal, sometimes making sample collection limited or not feasible for financial, legal or ethical reasons, which can negatively impact the value and quality of the results(Bissonette, 2017). Fortunately, breakthroughs in molecular and genomic technologies have allowed culture-independent study of microbial communities (microbiomes; Foster, Schluter, Coyte, & Rakoff-Nahoum, 2017; Miller, Svanbäck, & Bohannan, 2018), leading to numerous opportunities to improve our ability to understand animal populations through non-invasive sampling and monitoring of wildlife fecal microbiomes(Lynch & Hsiao, 2019; U. G. Mueller & Sachs, 2015). These opportunities include the discovery that microbiomes reflect a myriad of interactions between the host organism and its environment(Carroll, Threadgill, & Threadgill, 2009; Cho & Blaser, 2012; Hooper, Littman, & Macpherson, 2012). Findings from other animal and human systems have expanded on this exciting potential, describing associations between microbiomes and host sex(Bolnick et al., 2014; Vemuri et al., 2019), diet(De Filippo et al., 2010; Muegge et al., 2011; Wu et al., 2011), stress(Rea, Dinan, & Cryan, 2016; Stothart et al., 2016), age(S. Mueller et al., 2006; Yatsunenko et al., 2012), disease(Rausch et al., 2011; Round & Mazmanian, 2009) and physical-condition(Kohl, Amaya, Passement, Dearing, & Mccue, 2014; Sonnenburg & Sonnenburg, 2014; Turnbaugh et al., 2009). These studies provide an encouraging outlook for unobtrusively determining health parameters of animals of interest in conservation biology and management by sampling the fecal microbiome, rather than the animal itself.

The host and its microbial complement are co-evolving entities, now considered to be a "holobiont", that collectively comprise a single evolutionary unit that selection acts upon(Bordenstein et al., 2015; Brooks et al., 2016). It is now worthwhile to investigate wildlife species through this holobiont concept, capturing the complex roles and interactions of the host, microbiome and ecosystem inside and out—and non-invasively—to better conserve their place in our wildlands(Trevelline, Fontaine, Hartup, & Kohl, 2019; West et al., 2018).

To realize the conservation potential of the holobiont generally, we sought to further develop and apply bioinformatic tools utilizing host-related microbiome information in wild Rocky Mountain elk (*Cervus e. nelsoni*) populations. Multiple competing interests make elk management a good first target for improved monitoring, since populations are distributed broadly and animals can be both desirable (e.g. for hunting and eco-tourism) and a nuisance (e.g. from overpopulation, crop damage and disease spread perspectives) depending on the season and location. Elk and other large North American ungulates are an integral part of ecosystems and can provide the majority of biomass in predator diets(Metz, Smith, Vucetich, Stahler, & Peterson, 2012; Stewart, Bowyer, Ruess, Dick, & Kie, 2006). However, their distribution and abundance is threatened by emergent diseases like Brucellosis, Chronic Wasting Disease (CWD) and Tuberculosis (TB)(Proffitt et al., 2011). The importance of elk management strategies is thus both ecological and economic and, in elk as in other species of concern, future success depends on developing and deploying more efficient means of informing management and policy(Scott et al., 2010; Shafer et al., 2015).

In the current study, we performed a biogeographical survey of elk fecal microbiomes using 16S rRNA gene sequencing across four discrete Montana populations. Bacterial presence and abundance in the fecal microbiome was then compared to individual estimates of host ingesta-free body-fat (hereafter body-fat), sex, age and population location using machine learning techniques. Our intent was to show the plausibility of a non-invasive approach for studying these metrics based on the individual's fecal microbiome.

Microbiome communities generally have large numbers of diverse bacterial taxa, and theory suggests that specific community members are associated with vital host functions, with benefits provided to the host and the endosymbiotic community through a "leaky" surplus or common goods(Boon et al., 2014; Morris,

Lenski, & Zinser, 2012). This creates a highly linked endosymbiotic community. Since these microbial communities are subject to genetic drift within and between elk in a herd, when the total microbiome is explored with unsupervised pattern recognition techniques (e.g. with principal coordinates analysis, PCoA), only the strongest community signals are captured from the data. When multiple elk populations are compared, this diversity can be seen as biogeographic clustering likely due to unique drift, selection and horizontal transfer of microbes among animals in a population, leaving other function-specific taxa shared among populations masked in the microbiome (Supplemental Fig. 1). In light of this problem, we used a feature selection (FS) algorithm to reduce the dimensionality of the microbiome to better explore bacteria associated with specific host parameters including sex, age, body-fat and host population (Supplemental Table 1).. Reduced communities were then used to train linear discriminate classifiers (FS-LDA) with leave-one-out cross validation (CV) to predict each microbiome's affiliation to known host body-fat categories, sex, age or population, respectively. Using this combination of supervised and unsupervised machine learning approaches allowed us to recognize and separate the complex pattern responses and host states in the microbiome and also provided validation of the models (through CV). This method can be further refined for non-invasive classification into the future, allowing management-relevant information to be gained for individuals with more ease and efficiency than with animal capture-based techniques.

METHODS

Sample Collection

For our study, we received fecal pellet samples from wild Montana elk from four populations, namely the Bitterroot Mountains, Sapphire Mountains, Black's Ford area of the Madison River, and the Tobacco Root Mountains (Supplemental Fig. 3). Collection of scat samples, body-condition data and radio-collaring of elk were conducted in February 2014 by MTFWP. body-fat data was not collected for males (N=19) since these measurements are not informative for bull elk. The sampling used currently available and accepted invasive methods for wildlife immobilization, measurements of body-fat, sex classification and age(R. C. Cook et al., 2001, 2010).

Sample Preparation, DNA extraction and sequencing.

Frozen elk fecal pellets (-20° C) were prepared for DNA extractions by separating a standard weight (250 mg) from one randomly selected pellet per individual using a sterile petri dish (10 cm) and sterile safety razor blade for each sample. This fraction was placed into the designated sample tube from the Qiagen PowerSoil DNA extraction kit (Qiagen Inc., Germantown, MD) and processed further using the manufacturer's recommended protocol. Resulting purified metagenomic DNA was eluted with PCR-grade water and stored at -20° C.

To assess the bacterial community present in the fecal DNA extraction, we used a generally conserved (i.e. "universal") 16S/18S barcoded primer set and PCR focusing on V4 & V5 variable regions of the rRNA gene. The barcoded rRNA gene primers were 536F and 907R(Holben, Feris, & Kettunen, 2004). Once amplified, the samples were gel purified using the QIAGEN Gel Purification kit (QIAGEN, Germantown, MD) following the manufacturer's recommended protocol for downstream direct sequencing. This gel purification step removed any 18S eukaryotic DNA amplicons and artifactual PCR artifacts, thereby isolating and purifying the desired ~400bp 16S bacterial amplicons for sequencing. Illumina MiSeq 300 base-pair (bp) paired-end sequencing was conducted for all sampled individuals using the v3 reagent kit.

Sequence analysis.

Primers from the MiSeq paired end reads were trimmed using cutadapt(Martin, 2011). The DADA2 package was used in R to quality filter and trim, merge reads, call amplicon sequence variants (ASVs) and assign taxonomy to the ASVs(Callahan et al., 2016). ASVs were taxonomically assigned with DADA2 instance of the NaiveBayes classifier using the Ribosomal Database Project II release(Cole, Wang, Cardenas, & Fish, 2009). An ASV matrix was produced containing counts corresponding to the abundance of each ASV present

3

in each elk sample and its taxonomic classification. The Phyloseq package was used to summarize ASV tables into taxonomic bar plots(McMurdie & Holmes, 2014).

Feature Selection and Cross Validation

Metagenomic and 16S studies produce large amounts of data because of the need to sample microbial communities as deeply and completely as possible, but not all taxa have predictive power during statistical analysis for determining host states. We used a form of the Sequential Forward Floating Search algorithm (i.e. Feature Selection—FS) to select for informative genera from the elk microbiome(Pudil, Novovičová, & Kittler, 1994). This algorithm selects a subset of genera from the total pool using a heuristic method that maintains (or minimally reduces) the performance of the complete data set. The complete data matrix contains "noisy" genera that obscured the biological patterns present. FS avoids nesting issues where features (in this case bacterial taxa) are falsely fixed early in the selection process (an issue with other feature selection methods which results in reduced performance)(Saeys, Inza, & Larrañaga, 2007). By allowing all features to be added or subtracted as the algorithm progresses (essentially "floating" the selections), features are allowed to interact to produce dynamic and unbiased performance results not dependent on starting conditions. The FS algorithm employed herein uses J3 scores, a form of scatter matrices that rewards close clustering within groups of data points and rewards increased distance between groups of data points using Euclidean distances in multidimensional space. We produced tables of FS taxa sequentially with 2 through 30 features. For each feature table, a linear discriminate classifier was created and tested with CV, which uses a leave-one-out method of training and testing to reduce over-fitting the model to the training data set(Liu, Chen, Sheng, & Liu, 2014; Saeys et al., 2007). This method removes a sample from the training data, builds the model with remaining samples then tries to predict the classification of the removed sample. This leave-one-out method is iterated over all samples (N-1) to calculate the CV accuracy by summing the number of CV events in the denominator and summing the successful classification events in the numerator (e.g. $25/26 = 96.15\%$ percent correct). The intent of training the model in this way is to allow it to function on future data sets of similar character with very little optimization necessary, potentially producing an optimized model for determining these associations blindly from non-invasive scat samples.

We balanced classifier performance with over-fitting by comparing the CV performance differences between multiple numbers of features for signs of overfitting(Braga-Neto & Dougherty, 2004). We visualized this relationship with box plots and linear discriminant analysis (LDA(Liu et al., 2014); Supplemental Fig. 4). To help choose the optimal number of features for the visualizations, we computed a Pareto Front for multi-parameter optimization including accuracy, number of features and variance (not shown). We have thus named our workflow FS-LDA.

RESULTS

Study populations, field measurements and 16S survey.

We sampled 110 elk across 4 populations in Montana. Collection of scat samples, sex classification, age estimation and body-condition data were conducted in February 2014 by Montana Fish, Wildlife and Parks (MTFWP). Not all measurements were taken for all individuals. During the winter, elk body-condition is highly relevant to cow elk survival since elk depend on stores of body-fat to supplement their energy demands, especially in pregnant females(R. C. Cook et al., 2013). Body condition data was collected during capture by MTFWP personnel using a portable ultrasound machine to estimate levels of ingesta-free body fat (body-fat; see Methods). Female elk body-fat ranged from ~5% to ~13% (Supplemental Figure 5). Sequencing of partial 16S rRNA amplicons from each fecal sample provided a survey of bacterial presence and abundance in the microbiome (Supplemental Fig. 2).

Classifier for elk population.

Elk populations included the Bitterroot Mountains, Sapphire Mountains, Black's Ford area of the Madison River, and the Tobacco Root Mountains (Supplemental Fig. 3). Analysis of feature-selected elk microbiome

data (Fig. 1) and total elk microbiome data (Supplemental Fig. 1) showed strong patterns of elk population structure (biogeography). These signals of biogeography in the microbiome will naturally be mixed with (or potentially mask) other signals such as those of starvation, disease, and other unmeasured stresses affecting the host and its microbes. Although the unsupervised clustering showed natural groupings of elk populations (Supplemental Fig. 1), FS was found to further "denoise" the signal of biogeography by filtering non-associated taxa from the analysis (Fig. 1), providing tighter clustering and an improved visualization overall. The elk FS-LDA population classifier performed with 81 percent CV accuracy using 23 bacterial genera.

The strength of the relationship between biogeography and the microbiome is also supported by consistent accuracy values across all FS dimensions from 2 to 30 (supplemental Fig. 4). This pattern suggests that the predictive genera selected during FS vary in presence or abundance collectively by geographic location and that this biogeographic relationship has low sensitivity to particular genera since many dimensions display this relationship.

Classifier for body-fat.

Body-fat is a good predictor of survival and reproductive outcome in female elk(J. G. Cook, Cook, Davis, & Irwin, 2016). FS-LDA was used to classify bacteria from female elk across three populations (body-fat data was not collected for the Bitterroot population) and was trained on two classes, high ([?]8%) or low (<8%) body-fat (Fig. 2, left panel). The performance of the classifier for 2 elk body-fat categories was supported by low levels of overlap between clusters and by high CV accuracy (91%, where 50% would be random). Classifier accuracy was high despite combining data from 3 elk populations, which included the strong confounding signal of biogeography in the total microbiome. The results support a strong biological connection between microbiomes and host body condition. When more-specific classes were used to train the model (splitting the continuous distribution of body-fat into 3 and 4 bins), there was a strong trend in the ordination plot, but lower overall CV accuracies were obtained (77.8% and 58.3% respectively; Fig. 2, middle & right panel). This suggests that a finer-resolution biological pattern between the microbiome and elk body-fat may exist, but FS-LDA may be suboptimal for the continuous structure of the body-fat data. It is possible that more specific classification of microbiome-body-fat interactions could be achieved with further exploration of continuously distributed data classifier methods.

Classifier for sex determination

Male and female elk across the 4 Montana populations were included in this analysis (N=106). Due to female elk being more abundant and prioritized during sampling (due to the polygynous mating system of elk), the sample data was skewed toward females (87:19). This made classification difficult to interpret due to unequally weighted classifier training groups. Nonetheless, microbiome clustering delineating sex was obtained with high CV accuracy after normalizing bin sizes between males and females by either sampling males with replacement (bootstrapping) up to 87 samples (Fig. 3, top panel), or by randomly rarefying female samples to 19 individuals (Fig. 3, bottom panel). Bootstrapping produced ordination clustering by sex with 90 percent CV accuracy. Multiple random bootstrap iterations were performed with little change to the outcome of the classifier indicating the method was not sensitive to the random effects of sampling with replacement (some data not shown). Random rarefaction produced similar results in the visualization, with a comparable CV accuracy of 89 percent, which was qualitatively repeated across multiple random rarefaction iterations (Fig. 3, bottom panel, iterative data not shown).

Classifier for age

Age information was collected by observing tooth eruption and wear patterns for female elk in the sapphire population (N=34) and estimates ranged from 3-10 years(Hamlin, Pac, Sime, DeSimone, & Dusek, 2000). The FS-LDA model was trained on 2 age bins. The model used 8 taxa and performed with 87% CV accuracy (Fig. 4).

DISCUSSION

This study illustrates the strong linkages between the host animal (elk), its microbiome (across individuals), and the environment. Previous research in other model systems has shown the potential for both biogeographic structuring of the microbiome(Linnenbrink et al., 2013) and host-associated responses independent of biogeography(Cho & Blaser, 2012) . Analyses of interactions of the host microbiome with more-subtle variables like host health parameters are rarely conducted alongside a large geographical survey or on large wild animals. Here, we have demonstrated that the elk microbiome contains bacterial taxa that respond to both a strong correlate (population) and simultaneously occurring, but more subtle correlates (individual body-fat, sex and age class). Thus, we have developed and described herein an approach that allows us to disentangle microbiome responses to multiple host parameters of varying strength from the same bacterial sequence data set. In this pilot study, elk fecal microbiome composition has been shown to associate with body-fat, sex, age class and biogeography, an outcome that can guide the future use and development of microbiome-based monitoring of wildlife populations. These findings illustrate the potential of using this non-invasive, microbiome-based approach to monitor wildlife on the landscape and to potentially detect other biologically important attributes such as zoonotic diseases, predation stress and resource competition by observing effects on the intestinal microbiome.

Are differences in microbiome composition driven by host genetic differences or environmental factors?

Population genetic studies of nearby elk populations have shown limited female-specific gene flow and low standing genetic diversity for both sexes (mitochondiral $F_{ST} = 0.161$; nuclear $F_{ST} = 0.002$)(Hand et al., 2014). This observation supports the hypothesis that microbiome compositional differences seen between populations of female elk across Montana are not due to host genetic differences between the populations but are likely due to environmental factors and may be maintained by a lack of transmission between elk groups due to limited female movement.

Can microbiome-based measurements meet or exceed information derived from traditional physical measurements of the host?

Because measurements obtained by invasive sampling (e.g. immunological, endocrine, nutritional and metabolic data) are an interdependent part of a larger physiological regulatory network in the host, these traditional biomarkers are often correlated(Warne, Proudfoot, & Crespi, 2015). Host parameters are generally measured invasively or observed non-invasively (e.g. body-fat and age are invasively measured, and predation is observed). Conversely, microbiome compositional analysis of a single, non-invasively obtained fecal sample containing hundreds or thousands of interacting bacterial taxa can be used to associate a host's condition to its microbiome compositional response by applying the FS-LDA approach employed here. Once these key taxa and patterns are identified, one can subsequently obtain temporal fluctuations and other trends through multiple sampling events. Additionally, the microbiome can be used in direct conjunction with measurement of these other biomarkers to produce a deeper understanding of host condition and response when sampling is not limited (e.g. by pairing with blood or tissue sampling). To understand the biological context of these markers and determine where covariance diverges, it is advisable to compare a suite of markers (many hundreds of bacterial taxa) at once, across space and time for a given measurement, since sensitivity and accuracy may vary under such conditions.

In addition to direct host influence on microbiome composition, microbiome community structure can also be interpreted as an emergent property of external environmental variables experienced by the host, such as climatic variation, dietary composition and availability, and landscape heterogeneity. Some of these variables have been shown to produce temporal and spatial variation in animal microbiomes(Kartzinel et al., 2015; Linnenbrink et al., 2013). In the future, the combination of supervised and unsupervised machine learning approaches may allow us to recognize associations between microbiome taxa and habitat type or diet availability to produce insights to guide further investigation of these ecological interactions.

How can microbiome associations be used for animal conservation and population management?

6

With this approach, information can be collected relatively inexpensively (and therefore population-wide) on individual animals using the fecal microbiome. This increased scope of sampling may allow a more direct management of animal health or may be used to identify and ameliorate disease by its ability to provide a bigger "net" with simple field techniques and lower per-sample cost. The approach may be more relevant in small genetically isolated populations or during planned translocations, where individual animals are more impactful to population persistence, but individual-based wildlife management should become more important as the threat of disease increases with climate change and other anthropogenic disturbances.

Although not specifically tested here, microbiome-derived, individual-based measurements could provide data for individual-based population models incorporating survival, age and reproduction. The ability to collect individual-based data has not been explored in many systems due to cost and logistical limitations, which is why these measurements are often not collected and models are often simplified with a single survival parameter estimate for an entire age class (when available). This simplification prevents the application of individual-based modeling and may decrease our understanding of drivers of demographic change and reduce confidence in predicting future population responses. Microbiome analysis may help overcome these limitations by providing additional model information (e.g. individual population affiliation, age, sex, and body condition) from the same non-invasive fecal sample conducted within a DNA mark-recapture framework(Lukacs & Burnham, 2005; Marucco, Boitani, Pletscher, & Schwartz, 2011).

More recent population models have been evolving in complexity to include environmental variation affecting population parameters, and these models have therefore become more accurate(Schaub & Abadi, 2011), but also more data hungry. However, our limited understanding and ability to measure how local environmental effects influence individual animals in the short term has caused this source of variation to remain largely undefined in wildlife population modeling. As the relationships between environmental-animal-microbiome feedbacks are delineated, the microbiome could function as a proxy for environment-animal interactions and help in assessing species-specific environmental needs and monitoring goal-oriented habitat improvements. Indeed, further understanding the connection between environment and animal health, and predicting its effects using microbiome data can help resolve some of the general challenges inherent in measuring and understanding demographic changes in wildlife populations.

Age and the microbiome.

We demonstrated that the elk age differences can be detected in the microbiome, which suggests its future use in non-invasive population modeling or in augmenting invasive methods for age estimation. Age of individuals has been difficult to measure in many wildlife scenarios where adults do not possess any physical characteristics that can be passively observed after they near reproductive age. The relative ease of fecal microbiome sampling may provide the means to investigate the age structure of populations.

Transition matrices are a vital component of many population models and require estimates of individuals entering or leaving age-specific reproductive classes of the model system. As noted above, this is often the most difficult information to acquire and the ability to readily obtain this data would greatly improve the accuracy and applicability of modeling wild systems. Although we did not sample the 'yearling' age category and therefore could not include it in the age model, we expect this early age class to be the most different from the adult age class due to the more recent transition from milk to forage and the physiological changes that occur in the rumen as an elk matures. This suggests that once this age class is included in the model, a complete non-invasive method for grouping unknown samples into biologically relevant age classes may be possible.

Population effects on microbiome composition.

The microbiome shows strong patterns of biogeography delineating different elk herds. These patterns are consistent across most tested dimensions of the microbiome (2 taxa through 30 taxa), indicating that biogeography is insensitive to which bacteria are selected in the LDA (Supplemental Fig. 4). This is promising for understanding migratory animal herd affiliations and monitoring individual movement between herds using the fecal microbiome. This pattern also suggests that environmental and site-specific effects are

7

important in structuring fecal microbiome communities in elk.

Sex effects on the microbiome.

Delineating males from females using the feature selected fecal microbiome data can also be valuable to conservation. When sampling fecal pellets non-invasively, it is often impossible to know the sex of the animal without additional testing (e.g. PCR amplification of host markers such as the SRY gene). The ability to readily assign sex to each fecal sample allows us to partition other microbiome-based findings by sex at no additional cost, which can be desirable. Models for microbiome-based estimates of sex can also be built or validated with PCR methods using a subset of the fecal samples when needed. This has implications for future non-invasive fecal sampling, where these data can be used by managers to estimate sex ratios in a population during a fecal sampling event, independently verifying ratios observed by observational counts, or filling in gaps in information for populations where 'sightability' is low, such as within forest cover. In addition, this approach when combined with microbiome-based population delineation, could potentially be used to specifically monitor male and female elk movement independently between herds (e.g. monitor so-called "satellite bulls" or female specific migration) as mentioned in the previous section.

Microbiome-based estimation of ingesta-free body-fat.

Unlike biogeographic and other data, which can accurately be represented as discrete variables (e.g. common point locations, sex), body-fat data are typically distributed along a continuous gradient from low to high (Supplemental Figure 5), and as such we expected overlap near the imposed categorical bins in the LDA. A model for validating continuous data predictions is currently needed for body-fat data. The leave-one-out CV approach applied here attempts to fit and test the model against body-fat categories, which leads to the appearance of reduced model accuracy because body-fat values are continuous. We expect improved confidence in body-fat model results once an appropriate validation scheme for continuous variables is developed, such as a "fuzzy" boundary approach(Nguyen, 1997).

Although predictions of body-fat performed less well when using strict binning procedures compared to sex and population, exploring associations of a set of taxa to a binned range of body-fat remains informative, especially with repeated bin permutations. If the same taxa remain highly associated with a given biological measurement regardless of the imposed binning, this provides support for those taxa having a role in the functional phenomena behind the measurement (even if the direct function is unknown). We concede that FS-LDA is probably not the best method for developing predictive biomarkers for continuous data as currently presented, but our method does suggest the ability to discover associations that can be further developed and tested for functional relevance. For example, the continuity in bacterial features selected by the model across different numbers body-fat groups suggests biological relevance of the association of these bacteria with host body-fat (Supplemental Fig. 5). This method provides proof of concept and encourages the use and development of microbiome analyses for conservation science once statistical methods for continuous data are discovered or constructed.

Conclusion.

The overarching goal of this research was to explore an approach for identifying host animal measurements associated with fecal microbiome composition and to develop bioinformatic techniques useful for their analysis. Microbial biomarkers represent a less invasive alternative for acquiring information on wildlife populations than traditional sampling methodologies and could augment or even outperform traditional methods once strongly associating microbial taxa are identified and vetted for stability across population-space and time. The research reported herein provides the foundation for continued development and expansion of microbial biomarkers into a diverse range of wildlife species (including non-mammals) for monitoring and conservation. The products of such efforts could eventually provide insights and novel solutions to current wildlife management issues in general and allow threatened and endangered species to be studied with less perturbation.

ACKNOWLEDGEMENTS

REFERENCES

Bissonette, J. A. (2017). Avoiding the scale sampling problem: A consilient solution. *Journal of Wildlife Management* ,*81* (2), 192–205. doi: 10.1002/jwmg.21187

Bolnick, D. I., Snowberg, L. K., Hirsch, P. E., Lauber, C. L., Org, E., Parks, B., . . . Svanback, R. (2014). Individual diet has sex-dependent effects on vertebrate gut microbiota. *Nature Communications* , *5* . doi: 10.1038/ncomms5500

Boon, E., Meehan, C. J., Whidden, C., Wong, D. H. J., Langille, M. G. I., & Beiko, R. G. (2014). Interactions in the microbiome: Communities of organisms and communities of genes. *FEMS Microbiology Reviews* ,*38* (1), 90–118. doi: 10.1111/1574-6976.12035

Bordenstein, S. R., Theis, K. R., Furlan, M., Whiteson, K., Erb, M., & Pogliano, J. (2015). Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLOS Biology* ,*13* (8), e1002226. doi: 10.1371/journal.pbio.1002226

Braga-Neto, U., & Dougherty, E. (2004). Is cross-validation valid for small-sample microarray classification? *Bioinformatics* . Retrieved from http://bioinformatics.oxfordjournals.org/content/20/3/374.short

Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J., Bordenstein, S. R., & Kim, J. (2016). Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLOS Biology* , *14* (11), e2000225. doi: 10.1371/journal.pbio.2000225

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* (7), 581–583. doi: 10.1038/nmeth.3869

Carroll, I. M., Threadgill, D. W., & Threadgill, D. S. (2009). The gastrointestinal microbiome: a malleable, third genome of mammals.*Mamm Genome* , *20* (7), 395–403. doi: 10.1007/s00335-009-9204-7

Cho, I., & Blaser, M. (2012). The human microbiome: at the interface of health and disease. *Nature Reviews Genetics* . Retrieved from http://www.nature.com/nrg/journal/v13/n4/abs/nrg3182.html

Cole, J., Wang, Q., Cardenas, E., & Fish, J. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis.*Nucleic Acids* . Retrieved from http://nar.oxfordjournals.org/content/37/suppl_-1/D141.short

Cook, J. G., Cook, R. C., Davis, R. W., & Irwin, L. L. (2016). Nutritional ecology of elk during summer and autumn in the Pacific Northwest. *Wildlife Monographs* , *195* (1), 1–81. doi: 10.1002/wmon.1020

Cook, R. C., Cook, J. G., Murray, D., Zager, P., Johnson, B. K., & Gratson, M. W. (2001). Development of Predictive Models of Nutritional Condition for Rocky Mountain Elk. *Journal of Wildlife Management* ,*65* (4), 973–987. Retrieved from http://www.jstor.org/stable/3803046

Cook, R. C., Cook, J. G., Stephenson, T. R., Myers, W. L., Mccorquodale, S. M., Vales, D. J., ... Miller, P. J. (2010). Revisions of Rump Fat and Body Scoring Indices for Deer, Elk, and Moose. *Journal of Wildlife Management* , *74* (4), 880–896. doi: 10.2193/2009-031

Cook, R. C., Cook, J. G., Vales, D. J., Johnson, B. K., McCorquodale, S. M., Shipley, L. A., ... Schmitz, L. (2013). Regional and seasonal patterns of nutritional condition and reproduction in elk.*Wildlife Monographs* , (184), 1–45. doi: 10.1002/wmon.1008

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., ... Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America* , *107* (33), 14691–14696. doi: 10.1073/pnas.1005963107

Foster, K. R., Schluter, J., Coyte, K. Z., & Rakoff-Nahoum, S. (2017). The evolution of the host microbiome as an ecosystem on a leash.*Nature* , Vol. 548, pp. 43–51. doi: 10.1038/nature23292

Hamlin, K. L., Pac, D. F., Sime, C. A., DeSimone, R. M., & Dusek, G. L. (2000). Evaluating the Accuracy of Ages Obtained by Two Methods for Montana Ungulates. *The Journal of Wildlife Management* ,*64* (2), 441. doi: 10.2307/3803242

Hand, B. K., Chen, S. Y., Anderson, N., Beja-Pereira, A., Cross, P. C., Ebinger, M., ... Luikart, G. (2014). Sex-Biased Gene Flow Among Elk in the Greater Yellowstone Ecosystem. *Journal of Fish and Wildlife Management* , *5* (1), 124–132. doi: Doi 10.3996/022012-Jfwm-017

Holben, W., Feris, K., & Kettunen, A. (2004). GC fractionation enhances microbial community diversity assessment and detection of minority populations of bacteria by denaturing gradient gel electrophoresis.*Applied And* . Retrieved from http://aem.asm.org/content/70/4/2263.short

Hooper, L. V. L., Littman, D. R. D., & Macpherson, A. J. (2012). Interactions Between the Microbiota and the Immune System.*Science* , *336* (6086), 1268–1273. doi: 10.1126/science.1223490

Kartzinel, T. R., Chen, P. a., Coverdale, T. C., Erickson, D. L., Kress, W. J., Kuzmina, M. L., ... Pringle, R. M. (2015). DNA metabarcoding illuminates dietary niche partitioning by African large herbivores.*Proceedings of the National Academy of Sciences* , *112* (26), 8019–8024. doi: 10.1073/pnas.1503283112

Kohl, K. D., Amaya, J., Passement, C. A., Dearing, M. D., & Mccue, M. D. (2014). Unique and shared responses of the gut microbiota to prolonged fasting: A comparative study across five classes of vertebrate hosts. *FEMS Microbiology Ecology* , *90* (3), 883–894. doi: 10.1111/1574-6941.12442

Linnenbrink, M., Wang, J., Hardouin, E. E. A., K??nzel, S., Metzler, D., & Baines, J. F. (2013). The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Molecular Ecology* ,*22* (7), 1904–1916. doi: 10.1111/mec.12206

Liu, Z., Chen, D., Sheng, L., & Liu, A. (2014). Correction: Class Prediction and Feature Selection with Linear Optimization for Metagenomic Count Data. *PloS One* . Retrieved from http://journals.plos.org/plosone/article?id=10.1371/j

Lukacs, P. M., & Burnham, K. P. (2005). Review of capture-recapture methods applicable to noninvasive genetic sampling. *Molecular Ecology* , Vol. 14, pp. 3909–3919. doi: 10.1111/j.1365-294X.2005.02717.x

Lynch, J. B., & Hsiao, E. Y. (2019, September 27). Microbiomes as sources of emergent host phenotypes. *Science* , Vol. 365, pp. 1405–1409. doi: 10.1126/science.aay0240

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* , *17* (1), 10. doi: 10.14806/ej.17.1.200

Marucco, F., Boitani, L., Pletscher, D. H., & Schwartz, M. K. (2011). Bridging the gaps between non-invasive genetic sampling and population parameter estimation. *European Journal of Wildlife Research* , Vol. 57, pp. 1–13. doi: 10.1007/s10344-010-0477-7

McMurdie, P. J., & Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Computational Biology* , *10* (4), e1003531. doi: 10.1371/journal.pcbi.1003531

Metz, M. C., Smith, D. W., Vucetich, J. A., Stahler, D. R., & Peterson, R. O. (2012). Seasonal patterns of predation for gray wolves in the multi-prey system of Yellowstone National Park. *Journal of Animal Ecology* , *81* (3), 553–563. doi: 10.1111/j.1365-2656.2011.01945.x

Miller, E. T., Svanback, R., & Bohannan, B. J. M. (2018, December 1). Microbiomes as Metacommunities: Understanding Host-Associated Microbes through Metacommunity Ecology. *Trends in Ecology and Evolution* , Vol. 33, pp. 926–935. doi: 10.1016/j.tree.2018.09.002

Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss.*MBio* , *3* (2), e00036-12. doi: 10.1128/mBio.00036-12

Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., Gonzalez, A., Fontana, L., . . . Gordon, J. I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* , *332* (6032), 970–974. doi: 10.1126/science.1198719

Mueller, S., Saunier, K., Hanisch, C., Norin, E., Alm, L., Midtvedt, T., . . . Blaut, M. (2006). Differences in fecal microbiota in different European study populations in relation to age, gender, and country: a cross-sectional study. *Applied and Environmental Microbiology* ,*72* (2), 1027–1033. doi: 10.1128/AEM.72.2.1027-1033.2006

Mueller, U. G., & Sachs, J. L. (2015, October 1). Engineering Microbiomes to Improve Plant and Animal Health. *Trends in Microbiology* , Vol. 23, pp. 606–617. doi: 10.1016/j.tim.2015.07.009

Nguyen, H. T. (1997). Fuzzy sets and probability. *Fuzzy Sets and Systems* , *90* (2), 129–132. doi: 10.1016/S0165-0114(97)00078-X

[dataset] Pannoni, S. B., Holben, W. E., Proffitt, K. M. (2020). Microbiome Analysis Enables Non-invasive Monitoring of Rocky Mountain Elk Populations. *Dryad, Dataset* , https://doi.org/10.5061/dryad.4j0zpc880

Proffitt, K. M., Gude, J. A., Hamlin, K. L., Garrott, R. A., Cunningham, J. A., & Grigg, J. L. (2011). Elk distribution and spatial overlap with livestock during the brucellosis transmission risk period. *Journal of Applied Ecology* , *48* (2), 471–478. doi: 10.1111/j.1365-2664.2010.01928.x

Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters* . Retrieved from http://www.sciencedirect.com/science/article/pii/0167865594901279

Rausch, P., Rehman, A., Kunzel, S., Hasler, R., Ott, S. J., Schreiber, S., . . . Baines, J. F. (2011). Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proceedings of the National Academy of Sciences* ,*108* (47), 19030–19035. doi: 10.1073/pnas.1106408108

Rea, K., Dinan, T. G., & Cryan, J. F. (2016). The microbiome: A key regulator of stress and neuroinflammation. *Neurobiology of Stress* , Vol. 4, pp. 23–33. doi: 10.1016/j.ynstr.2016.03.001

Round, J. L., & Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nature Reviews. Immunology* , *9* (5), 313–323. doi: 10.1038/nri2515

Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics* . Retrieved from http://bioinformatics.oxfordjournals.org/content/23/19/2507.short

Schaub, M., & Abadi, F. (2011). Integrated population models: a novel analysis framework for deeper insights into population dynamics.*Journal of Ornithology* . Retrieved from http://link.springer.com/article/10.1007/s10336-010-0632-7

Scott, J. M., Goble, D. D., Haines, A. M., Wiens, J. A., & Neel, M. C. (2010). Conservation-reliant species and the future of conservation.*Conservation Letters* , *3* (2), 91–97. doi: 10.1111/j.1755-263X.2010.00096.x

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergstrom, L., Bruford, M. W., Brannstrom, I., ... Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology and Evolution* , *30* (2), 78–87. doi: 10.1016/j.tree.2014.11.009

Sonnenburg, E. D., & Sonnenburg, J. L. (2014). Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metabolism* ,*20* (5), 779–786. doi: 10.1016/j.cmet.2014.07.003

Stewart, K. M., Bowyer, R. T., Ruess, R. W., Dick, B. L., & Kie, J. G. (2006). *Herbivore Optimization by North American Elk: Consequences for Theory and Management* . 1–24. doi: 10.2193/0084-0173(2006)167[1:HOBNAE]2.0.CO;2

Stothart, M. R., Bobbie, C. B., Schulte-Hostedde, A. I., Boonstra, R., Palme, R., Mykytczuk, N. C. S., & Newman, A. E. M. (2016). Stress and the microbiome: linking glucocorticoids to bacterial community dynamics in wild red squirrels. *Biology Letters* , *12* (1). Retrieved from http://rsbl.royalsocietypublishing.org/content/12/1/2015

Trevelline, B. K., Fontaine, S. S., Hartup, B. K., & Kohl, K. D. (2019, January 30). Conservation biology needs a microbial renaissance: A call for the consideration of host-associated microbiota in wildlife management practices. *Proceedings of the Royal Society B: Biological Sciences* , Vol. 286. doi: 10.1098/rspb.2018.2448

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., ... Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature* , *457* (7228), 480–484. doi: 10.1038/nature07540

Vemuri, R., Sylvia, K. E., Klein, S. L., Forster, S. C., Plebanski, M., Eri, R., & Flanagan, K. L. (2019, March 15). The microgenderome revealed: sex differences in bidirectional interactions between the microbiota, hormones, immunity and disease susceptibility.*Seminars in Immunopathology* , Vol. 41, pp. 265–275. doi: 10.1007/s00281-018-0716-7

Warne, R. W., Proudfoot, G. A., & Crespi, E. J. (2015). Biomarkers of animal health: integrating nutritional ecology, endocrine ecophysiology, ecoimmunology, and geospatial ecology. *Ecology and Evolution* ,*5* (3), 557–566. doi: 10.1002/ece3.1360

West, A. G., Waite, D. W., Deines, P., Bourne, D. G., Digby, A., Mckenzie, V. J., & Taylor, M. W. (2018). *The microbiome in threatened species conservation* . doi: 10.1016/j.biocon.2018.11.016

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., ... Lewis, J. D. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science (New York, N.Y.)* , *334* (6052), 105–108. doi: 10.1126/science.1208344

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., ... Clemente, J. C. (2012). *Human gut microbiome viewed across age and geography* . doi: 10.1038/nature11053
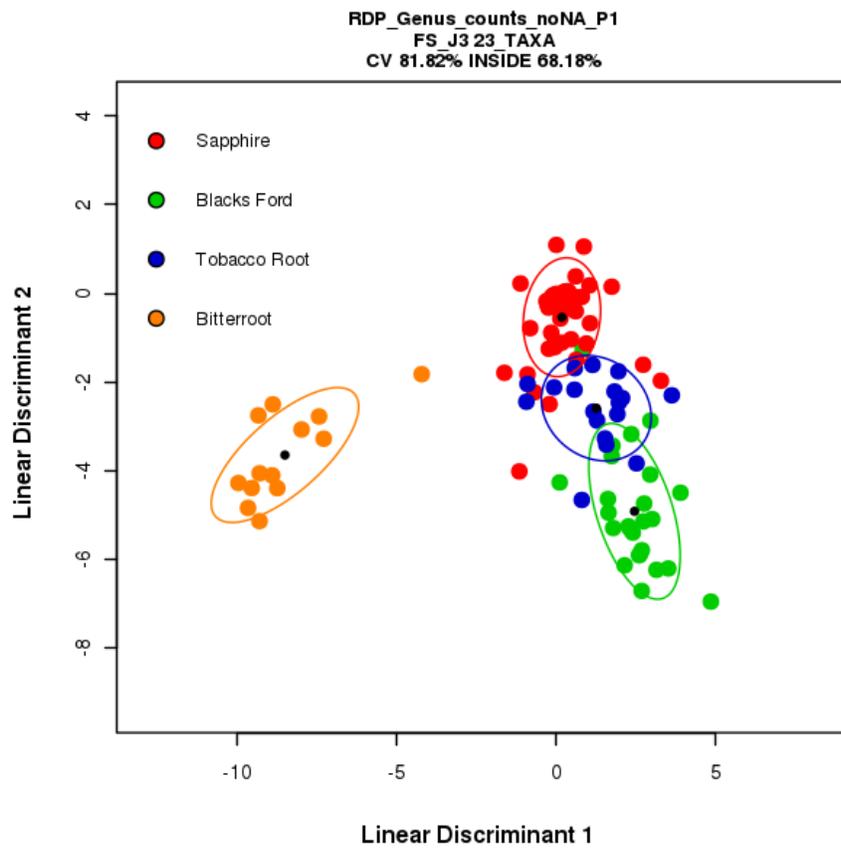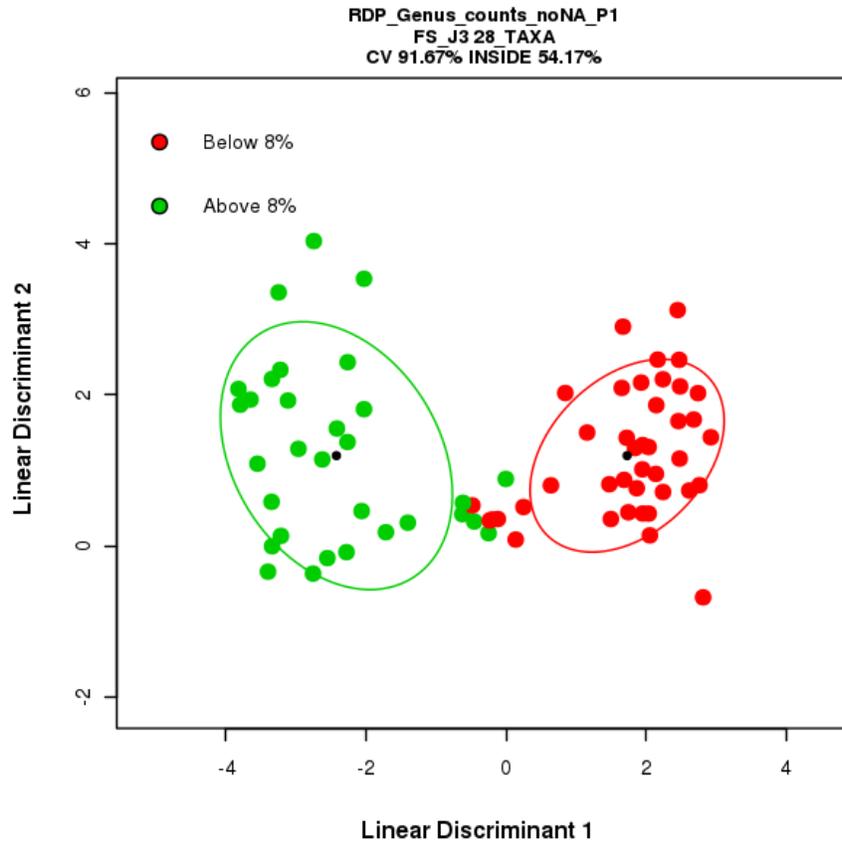
DATA ACCESSIBILITY

The raw sequence dataset will be deposited to the Dryad repository under accession number doi:10.5061/dryad.4j0zpc880 (Pannoni, 2020). R scripts used to process sequences and build classifiers are available upon request.
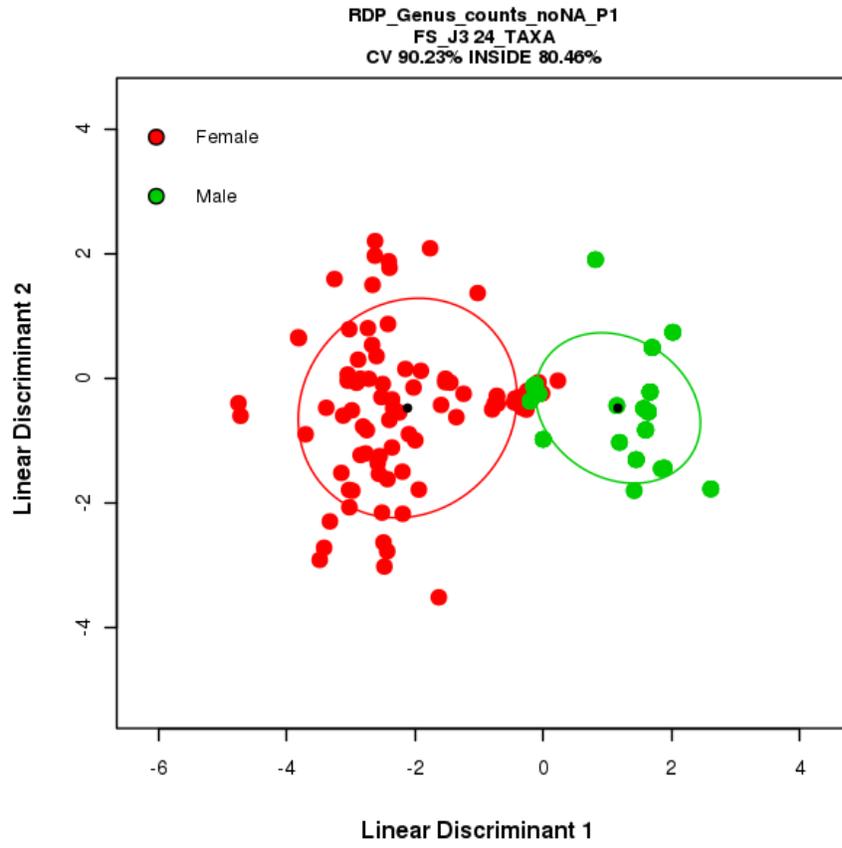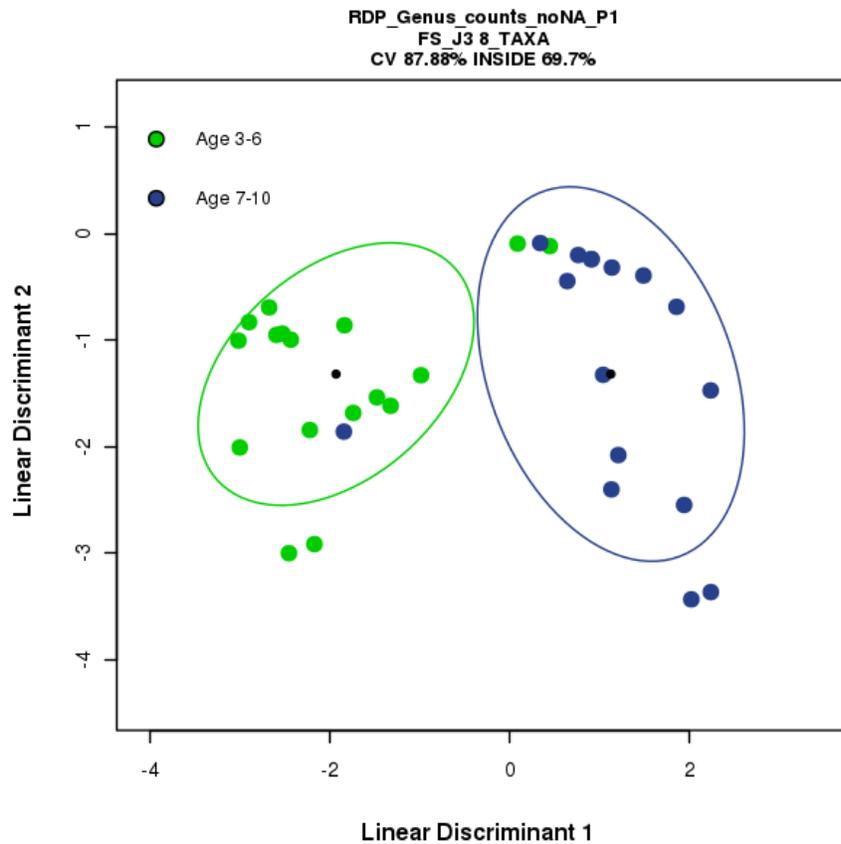
AUTHOR CONTRIBUTIONS

S.P. and W.H. conceived the project. S.P. and K.P. developed and implemented the research plan. K.P. coordinated the field work, elk handling and sampling aspects of the project. S.P. performed the sample processing and data analysis. The first draft of the paper was written by S.P. with input from W.H. and K.P. All authors contributed to discussing the results and editing the paper.

FIGURES

13

RDP_Genus_counts_noNA_P1
FS_J3 28_TAXA
CV 91.67% INSIDE 54.17%

Below 8%

Above 8%

Linear Discriminant 2

Linear Discriminant 1

RDP_Genus_counts_noNA_P1
FS_J3 24_TAXA
CV 90.23% INSIDE 80.46%

RDP_Genus_counts_noNA_P1
FS_J3 8_TAXA
CV 87.88% INSIDE 69.7%

SUPPLEMENTARY INFORMATION (separately attached)

Supplementary index:

S. Fig. 1 - PCoA of elk total microbiomes

S. Fig. 2 - Microbiome bar-plots of Genera

S. Fig. 3 - Sampling location map

S. Fig. 4 - CV box and whisker plots for population analysis

S. Fig. 5 - Body-fat of female elk

S. Table 1 - Bacteria used in each FS-LDA Classifier