A mixed model approach for estimating drivers of microbiota community composition and differential taxonomic abundance

Amy R Sweeny[1]*, Hannah Lemon[1]*, Anan Ibrahim[1], Daniel H Nussey[1], Andrew Free[2$], Luke McNally[1$]

* These authors contributed equally

[$] Joint last authors

Affiliations:

1 Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL UK

2 Institute of Quantitative Biology, Biochemistry and Biotechnology, University of Edinburgh EH9 3FL UK

Author for correspondence: Amy R. Sweeny, a.r.sweeny@ed.ac.uk

Key words: microbiota, metabarcoding, generalised linear mixed-effects model, community composition, Bayesian estimation

Abstract

1. Next-generation sequencing (NGS) is a recent and powerful tool for quantifying complex within-host community dynamics in real time. Metabarcoding approaches quantifying the abundance particularly of members of microbial communities, accompanied by non-invasive faecal sampling, have facilitated a rapid expansion of wild studies of gut microbiota composition.
2. Generalised linear mixed-effects models (GLMMs) are powerful tools commonly applied to ecological and evolutionary data to dissect variance at multiple scales spanning the environment (e.g. season and geography) and host (e.g. age and genetics) levels. Despite being widely used in ecology, these approaches are as yet not well-integrated with tools and methods used to understand variance in metabarcoding data and microbiota data.
3. Here we outline a GLMM approach to quantifying how factors across multiple scales contribute to variation in gut microbiota composition.
4. We apply this framework to two years of data collected from the Soay Sheep population of St. Kilda to validate the approach. Specifically, we investigate the contribution of season and host age to gut microbiota community composition in these datasets. Our approach agreed with traditional approaches in showing a strong effect of age but a weak effect of season, while also a) providing the flexibility to identify fine-scale drivers of these effects and b) estimating specific taxa shifts associated with factors of interest.
5. These results highlight the conceptual benefits of this approach and provide the foundation for extension to more complex model formulations. We suggest that the approach presented here provides a powerful means to answer emerging questions surrounding the ecological and evolutionary roles of within-host communities using metataxonomic data.

1. Introduction

Understanding ecological and evolutionary dynamics in the wild relies upon decomposing the variance structures underlying phenotypes into relative contributions of genetic, environmental, and other factors. Mixed model approaches have been a powerful tool in this regard, in particular

48 for studies of wild populations experiencing complex natural environments (Zuur *et al.* 2009;
49 Charmantier *et al.* 2014). Generalised linear mixed-effects models (GLMM) provide a flexible
50 means for appropriately dealing with the complex data structures and relationships between
51 predictors of interest that arise in ecological data, such as over-dispersed count data, complicated
52 experimental design structures, or interactions between key variables (Bolker *et al.* 2009).
53 Broadly, GLMMs are specified with both fixed effects of experimental design or biological
54 interest as well as random effects which estimate technical or biological variance from the
55 population (Bolker *et al.* 2009). For example, in a GLMM analysis of birth weight in wild Soay
56 sheep, a phenotypic trait which is linked to early-life survival and lifetime fitness, studies have
57 estimated the differences in weight between sexes (males are heavier) as a fixed effect alongside
58 estimates of the variance in the birth weight of offspring among mothers as a random effect
59 (Wilson *et al.* 2005). Such studies have revealed appreciable maternal effects influencing
60 important offspring traits, despite the underlying causal maternal phenotypes being unknown
61 (Wilson *et al.* 2005). A powerful extension of GLMM framework for ecologists and evolutionary
62 biologists is the 'animal model', which estimates the genetic component of phenotypic variation
63 by specifying a pedigree or genetic relatedness matrix in the random effect structure (Wilson *et*
64 *al.* 2010). These approaches are crucial steps toward understanding wild phenotypes and the
65 evolutionary dynamics of natural populations (Charmantier *et al.* 2014).
66
67 The microorganisms living within a larger host play a key role in determining variation in host
68 phenotype and fitness, comparable in influence to a range of extrinsic environmental and intrinsic
69 host factors. The ecological dynamics of within-host communities of parasites and commensal
70 microbes can have dramatic effects on host health and fitness (Alberdi *et al.* 2016; Koskella *et al.*
71 2017). Interactions among the trillions of bacteria (microbiota) that reside in the host gut have
72 therefore been brought into focus recently as playing a key role in host physiology (Fontaine &
73 Kohl 2020). A growing number of studies from humans and model laboratory animals highlight
74 impacts of the gut microbiota on host behaviour, metabolism, as well as endocrine and immune
75 homeostasis (Sudo *et al.* 2004; Cox-Foster *et al.* 2007; Desbonnet *et al.* 2014; Wahlström *et al.*
76 2016). In natural systems, environmental factors, complex genetic architecture, and host
77 heterogeneity may all contribute to dynamics of within-host communities and their effect on
78 phenotypic variation. Studies of model organisms and laboratory experiments, whilst central to

our understanding of the mechanisms linking the gut microbiota and host phenotype, cannot replicate or capture the complexity and dynamism of natural systems. Therefore, to fully understand the host-microbiota interactions and their impacts on host health and fitness, we need to study these interactions in real ecological contexts. To date, common methods of quantifying microbial community structure in wild systems have lacked the power and flexibility offered by GLMMs that have been used to dissect variation in standard phenotypic traits.

Laboratory- and human-based workflows for microbiota analysis are increasingly being applied to wild systems (e.g. Morgan & Huttenhower 2012; Goodrich *et al.* 2014). Standard pipelines for investigating hypotheses concerning gut microbiota dynamics typically include collection of faecal samples from selected study subjects, and application of next-generation sequencing (NGS) techniques for meta-barcoding of informative bacterial genes for taxonomic assignment of sequenced reads. A number of bioinformatic tools and packages (e.g. MOTHUR (Schloss *et al.* 2009), deSeq (Anders & Huber 2010), DADA2 (Callahan *et al.* 2016)) are designed to process these sequencing data and assign taxonomy to reads. Grouping (binning) of reads by a set similarity threshold (typically 97%) into operational taxonomic units (OTUs) and assignment of taxonomic information to these OTUs is common for producing read counts per OTU for downstream analysis. More recently, use of amplicon sequence variants (ASVs, inferred exact biological sequences) as the taxonomic grouping has been implemented in many workflows for higher resolution metabarcoding information (Callahan *et al.* 2017). Subsequent microbiota community analysis commonly relies on transformation of these counts into relative proportions per sample or rarefaction such that a set library size is randomly subsampled from all samples (McMurdie & Holmes 2014; Weiss *et al.* 2017; McMurdie 2018)

Hypothesis testing using transformed counts from 16S taxonomic assignments is typically focused on community-level differences in taxonomic diversity and composition between experimental groups or time points of interest. Statistical approaches to this end include estimation of alpha diversity (the number of distinguishable taxa within a sample), distance measures (e.g. Bray-Curtis dissimilarity (Bray and Curtis, 1957)), and ordination with dimensionality reduction (e.g. principal coordinates analysis). These approaches can have several limitations with regard to data assumptions and statistical inference. Standardisations of data

110 based on proportions ignore heteroscedasticity from different library sizes across samples, while
111 those reliance on rarefaction restricts data such that the reads considered per sample are limited to
112 the minimum number of reads across all samples (McMurdie & Holmes 2014). This in turn can
113 significantly elevate rates of false positives or reduce performance in microbiome clustering
114 approaches. In addition to statistical pitfalls, these traditional approaches for assessing
115 community-level differences differ philosophically from GLMM-based approaches that involve
116 complex variance partitioning. Although these approaches have provided substantial insights into
117 microbiota community composition, particularly in laboratory systems where conditions are
118 strictly controlled, they typically lack flexibility to include, for example, host pedigree or spatial
119 information alongside other predictors of interest.
120
121 Here we apply a GLMM approach to an indicative wild microbiota dataset to illustrate the
122 advantages of this approach for analysis of microbiota data from natural populations. We validate
123 this approach investigating host and environmental factor contributions to microbiota community
124 composition, discuss benefits of the approach, and offer further suggestions for circumventing
125 possible limitations. Finally, we highlight conceptual extensions of the presented models for
126 application to more complex ecological and evolutionary questions and to additional types of
127 metataxonomic data.
128

129 2. A GLMM approach

130 Because gut microbes have important effects on host physiology, behaviour, and health, much
131 research has sought to identify individual microbial taxa that are responsible for alterations of
132 host phenotype and state. This has in many ways mirrored the goals of many genome-wide
133 association study (GWAS) analyses, which have sought to identify particular genetic variants
134 associated with phenotypes of interest, often with a goal of developing diagnostics or drug targets
135 (Visscher *et al.* 2012). However, just as GWAS analyses have shown that most phenotypes are
136 highly polygenic, being determined by a complex combination of genetic variants of small effects
137 (Goldstein 2009; Consortium 2009; Loh *et al.* 2015), the study of host-associated microbiomes
138 has often failed to find single taxa associated with host states (Clemente *et al.* 2012; Vayssier-
139 Taussat 2014). Instead many changes in host state are associated with general shifts in
140 microbiome composition, often termed dysbiosis (Clemente *et al.* 2012; Carding *et al.* 2015).

141 Phenomena such as dysbiosis shift the level at which we look for associations with host
142 phenotype move from a small number of microbial taxa to the whole microbiota. In addition, the
143 most pressing questions about host-associated microbiota in ecology and evolution are general
144 and focused on the entire microbiota community (Koskella *et al.* 2017). For example, what are
145 the relative roles of host physiology and environment in shaping the microbiota? How heritable is
146 the microbiota? How much does microbiota composition impact fitness? The shift in focus of
147 these questions from individual taxa to complete community poses an important conceptual and
148 statistical challenge.
149
150 As previously discussed, host-associated microbiotas often constitute hundreds or thousands of
151 different taxa. Whenever we need to estimate a large ensemble of related parameters a common
152 statistical approach is to treat them as random variables from some distribution (Gelman & Hill
153 2006). To understand how this approach applies to the microbiota let us consider the concrete
154 question of estimating how the composition of the gut microbiota might change with season in a
155 wild mammal. In traditional approaches to analysing microbiota datasets it is common to
156 visualise an ordination of the data distinguishing points by season, perform a permutational
157 ANOVA on a dissimilarity matrix to test if microbiotas from different seasons are more
158 dissimilar than those from the same season, and then use some test for differential abundance of
159 individual taxa across seasons to identify taxa with a major role in these changes (Segata *et al.*
160 2011; Mandal *et al.* 2015). In this approach, estimates for how individual taxa differ by season
161 are all independent of each other. Using a random effects model would approach this in a
162 fundamentally different manner, where a model would be fitted with a random effect for the
163 seasonal change for each microbial taxon. This approach has the advantage that all taxa inform
164 the estimate of the mean and variance of the distribution that the effects across taxa come from.
165 The estimates of parameters for individual taxa are then "shrunk" to this distribution. This
166 "shrinkage" is known to improve the accuracy of parameter estimation as long as there are large
167 numbers of groups for the random effects, which is generally true for most host-associated
168 microbiota owing to their large number of taxa.
169
170 While fitting such random effects models is known to improve parameter estimation owing to
171 shrinkage, their biggest advantage is in allowing us to shift the questions we ask to the whole

microbiota level, and partition complex & inter-related sources of variance. GLMM approaches have been used widely in ecological and evolutionary contexts to estimate repeatability, relative levels of spatiotemporal variance, social and common environment effects, as well as heritability and the role of host genetics. Answering such questions has proved hugely challenging in the microbiota field as most analyses rely on using dissimilarity and permutational tests, from which it is difficult to ascertain the relative importance of such complex variables. Here we make a first step towards applying the full power of GLMMs to studying host-associated microbiota. Specifically, we illustrate how to appropriately structure random effects across microbial taxa in a GLMM, and thus partition the sources of variance driving microbiota composition.

To see how such a model can be structured, we again return to the example of estimating the effect of season on a gut microbiota (Figure 1). Consider a scenario with two samples taken per host from a sample of hosts in a population, one in winter and one in summer, and with samples appropriately sequenced and reads bioinformatically assigned to ASVs. We will consider two different approaches to fitting a GLMM to such data. Firstly, we can directly analyse such count data by fitting a Poisson family GLMM with log-link. The predicted values on the link scale are given according to the following model

$$log(y_{h,asv,s}) = \beta_0 + \beta_1 s + u_h + u_{asv} + u_{h:s} + u_{asv:h} + u_{asv:s} + u_{asv:h:s}$$

Here $y_{h,asv,s}$ is the read count, β_0 is a global intercept, β_1 is the effect of season (s), u_h is a random effect on the total read count for a given host, u_{asv} is a random effect on the overall abundance (read count) of an ASV, $u_{h:s}$ is a random effect for the mean read count in each sample (i.e. host by season), $u_{asv:h}$ is a random effect on the abundance (read count) of an ASV in host h, $u_{asv:s}$ is a random effect on how an ASV changes with season and $u_{asv:h:s}$ is an additional row level random effect. This specification controls for variation in library size (total read count across samples) via the random effect $u_{h:s}$. By apportioning the variance attributed to these different random effects we can assess the relative contributions of these different factors to microbiota composition. For example, $u_{asv:h}$ can be interpreted as equivalent to individual 'repeatability' of ASV community composition and $u_{asv:s}$ can be interpreted as reflecting variance associated with compositional shifts across seasons.

While the Poisson model accounts for variation in library size across samples, there has also been a shift in microbiota research towards explicitly compositional data analysis, which removes any effects of library size (other than in quantifying uncertainty) prior to analysis. The centered log-ratio (CLR) described by Aitchsion (1982) represents one such transformation. The CLR takes the form of the log-ratio of the count for a given taxon in a given sample and the geometric mean of the counts of all taxa in that sample. The CLR can be calculated for any taxon *i* in a sample of *n* taxa as:

$$clr_i = log\left(\frac{x_i}{G(x)}\right)$$

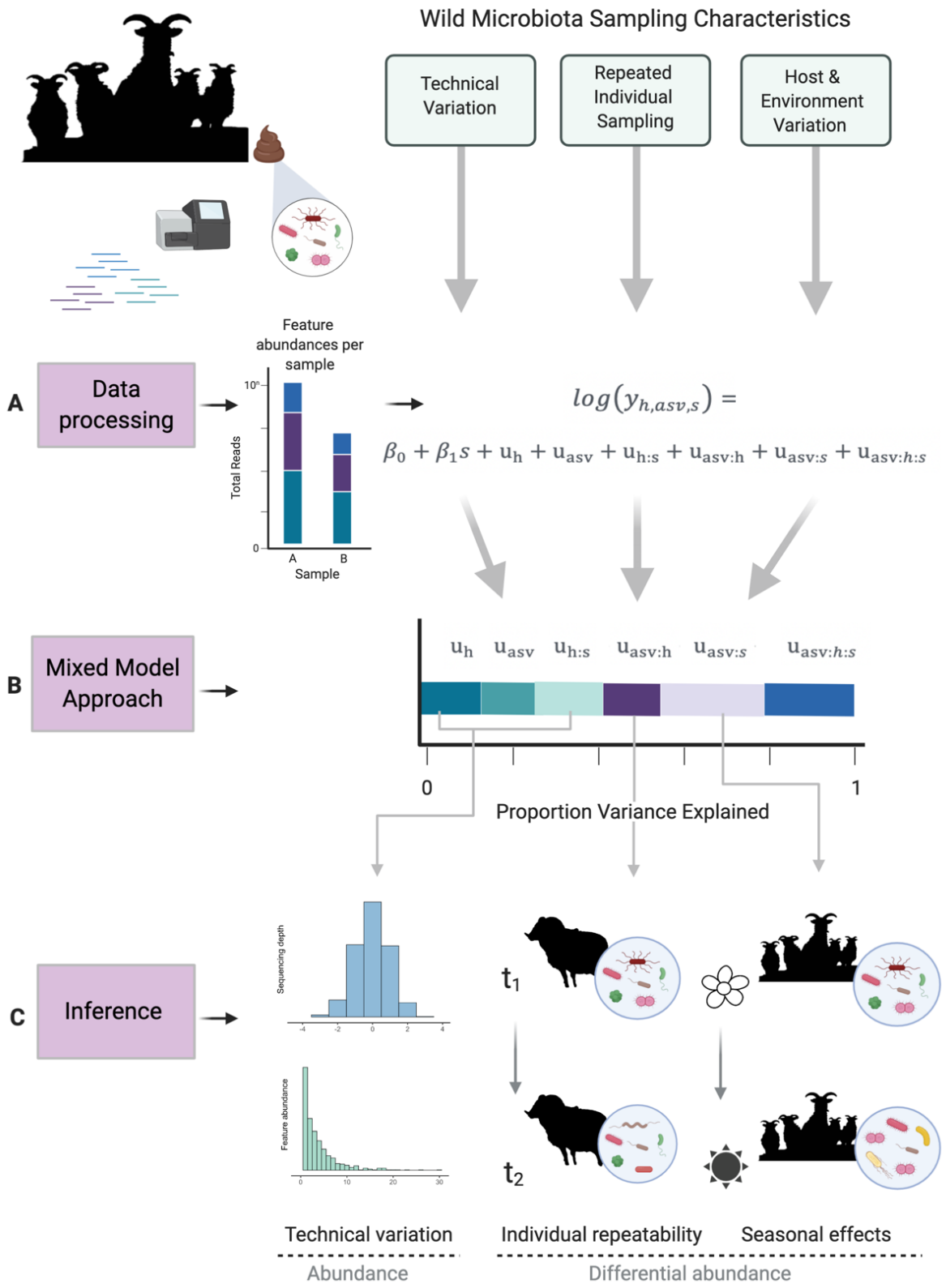
$$G(x) = \sqrt[n]{x_1 \times x_2 \times \dots x_n}$$

The CLR has several desirable properties. Firstly, as a log-ratio it is approximately Gaussian in distribution, allowing the use of a Gaussian linear mixed model. Secondly, the mean of the CLR of all taxa in any sample will be zero, allowing for simplification of our random effects structure. Using the CLR we can explore the effect of season on microbiota composition using the following Gaussian LMM:

$$clr_{h,asv,s} = \beta_0 + u_{asv} + u_{asv:h} + u_{asv:s} + u_{asv:h:s}$$

Note that here the random effects not including ASV as an element are removed as the mean of the CLR within samples, and therefore across hosts and seasons, is zero. This reveals the role of the different random effects here – the effects u_{asv}, $u_{asv:h}$, $u_{asv:s}$ and $u_{asv:h:s}$ describe effects on microbiota composition, while the effects u_h and $u_{s,h}$ are effects on the mean read count across all taxa. Finally, the three terms $u_{asv:h}$, $u_{asv:s}$ and $u_{asv:h:s}$ and their sizes relative to each other are of most biological interest. These three terms capture how much the abundance of a taxon are affected by consistent host effects, seasonal effects, and idiosyncratic effects (the row level effect of seasonal effects varying by host). These effects capture changes in a taxon's abundance across hosts and seasons, after controlling for the mean abundance of each taxon, sample, and season. We discuss the advantages and disadvantages of both the Poisson and CLR models below, but first illustrate their usage with an example analysis.

233


Wild Microbiota Sampling Characteristics
Technical Variation
Repeated Individual Sampling
Host & Environment Variation
A
Data processing
Feature abundances per sample
Total Reads
Sample
A
B
$log(y_{h,asv,s}) =$
$\beta_0 + \beta_1 s + u_h + u_{asv} + u_{h:s} + u_{asv:h} + u_{asv:s} + u_{asv:h:s}$
B
Mixed Model Approach
u_h u_{asv} $u_{h:s}$ $u_{asv:h}$ $u_{asv:s}$ $u_{asv:h:s}$
0
1
Proportion Variance Explained
C
Inference
Sequencing depth
Feature abundance
t_1
t_2
Technical variation
Abundance
Individual repeatability
Seasonal effects
Differential abundance

234

Figure 1. Overview of mixed model approach to wild microbiota analysis. Data processing (A) generates amplicon sequence variant (ASV)-level abundances for each sample. These raw abundances are used as the response for generalised linear mixed-effects models with Poisson error families. In the example illustrated, data includes sampling timepoints for a group of individuals taken during two seasons. Model syntax therefore specifies a fixed effect of age, and random effects for taxonomy (asv), sample id (host:season, h:s), individual differential abundance of asvs (asv:h), differential abundance of asvs across seasons (asv:s), and a residual variance at the row-level (asv:h:s). GLMM output can be used to partition the variance explained by each random effect term (B). These variance components can be interpreted as the relative contributions of both technical variation and host or environmental contributions to differential abundance as illustrated in (C). Created with BioRender.com

3. A worked example: age and season effects on gut microbiota in wild sheep

To test and illustrate our approach, we obtained faecal samples from Soay Sheep (*Ovis aries*) from the island of Hirta in the St. Kilda archipelago of the Outer Hebrides of Scotland. These animals are free-living and are part of a long-term study in which individuals have been marked and monitored longitudinally since 1985 (Clutton-Brock Pemberton 2004). All animals sampled had been caught and uniquely tagged within a few days of birth, so that their age and sex were known with certainty. Each year, fieldwork teams visit St Kilda in spring to monitor lambing and capture, mark and sample newborn lambs within a few days of birth. Subsequently, each August a larger field team visits to capture, mark and sample animals living in the study area using a series of corral traps (Clutton-Brock & Pemberton 2004).

Two sets of faecal samples were collected, in 2013 and 2016, to allow comparison of the gut microbiota of individuals of different ages (2013) and from the same individuals sampled in different season (2016). The 2013 samples were collected during the August catch and included 30 samples from lambs (around 4 months old) and 28 samples from older adults (ages 2-13). The 2016 samples were collected from a set of 36 females age 1-13 years who were sampled in both spring (around the time of parturition) and then 3-4 months later in August. Microbial DNA was extracted from samples, amplified using bacterial 16S rRNA V4 region primers, and sequenced using the Ilumina MiSeq platform to generate 250 base pair (bp) paired-end reads. Sequences were processed using the DADA2 pipeline (v1.12.1) in R to call amplicon sequence variants (ASVs) (Callahan *et al.* 2016).Full details of sampling, sequencing, and data processing methods are provided in the Electronic Supplementary Material (ESM 1.1-1.2).

Prior to further analysis an abundance filter was applied to both datasets, eliminating ASVs with a total abundance of <100 reads, resulting in 2,023 ASVs for 2013 and 2,364 for 2016 (filtering threshold sensitivity was validated examining models from one year – 2016 –against multiple cut-offs , see ESM 1.4.2 & Figure S2). All analyses were conducted in R version 3.6.1 (R Core Team 2019). To provide a comparison with more traditional statistical approaches to meta-barcoding data, we normalised read abundances to compositional proportion data and calculated pairwise dissimilarities among samples using Bray-Curtis in the 'phyloseq' package of R (McMurdie & Holmes 2013). To assess the extent to which age and season predicted microbiota composition these dissimilarities were used in principle coordinates analysis (PCoA). Permutational analysis of variance (PERMANOVA) was used to study the difference in group means and was carried out using the Vegan function 'adonis' (Anderson 2001).

3.1. Specification of GLMMs:

We applied separate GLMMs to the 2013 and 2016 datasets. First, an aggregate dataset for each year was created from the sample metadata, taxonomic classifications for each ASV, and an ASV-by-sample abundance matrix. We fit GLMMS with Poisson errors and log links to each dataset using the Bayesian package 'MCMCglmm' (Hadfield 2010) using the approach introduced above (Section 2). Models fit to 2013 data inclusive of two age classes from a single season (one sample per host) were specified as follows:

$$log(\bar{y}_{h,asv}) = \beta_0 + a\beta_a + u_h + u_{asv} + u_{asv:a} + u_{asv:h}$$

Where $y_{h,asv,s}$ is the read count, β_0 is a global intercept, β_a is the effect of age (*a,* binary factor: lamb & adult), u_h is a random effect on the mean read count in each host (sample), u_{asv} is a random effect on the overall abundance (read count) of an ASV, $u_{asv:a}$ is a random effect on how an ASV changes with host age and $u_{asv:h}$ is an additional row level random effect.

Models fit to 2016 data inclusive of one group of individuals sampled in both spring and summer (two samples per host) were specified as follows:

$$log(\bar{y}_{h,asv,s}) = \beta_0 + s\beta_s + u_h + u_{asv} + u_{h:s} + u_{asv:h} + u_{asv:s} + u_{asv:h:s}$$

Here $y_{h,asv,s}$ is the read count, β_0 is a global intercept, β_s is the effect of season (*s,* binary factor: spring & summer), u_h is a random effect on the total read count for a given host, u_{asv} is a random effect on the overall abundance (read count) of an ASV, $u_{h:s}$ is a random effect for the mean read count in each sample (i.e. host by season), $u_{asv:h}$ is a random effect on the abundance (read count) of an ASV in host *h,* $u_{asv:s}$ is a random effect on how an ASV changes with season and $u_{asv:h:s}$ is an additional row level random effect.

Using this GLMM approach, we calculated the relative contributions to sources of variance in the data from variation in total library size across hosts (u_h), samples (2013: equivalent to host; 2016: $u_{h:s}$), seasons (2016: u_s) or age classes (2013: u_a), taxonomic variance among microbes in mean abundance (u_{asv}), and community composition of taxa (ASVs) across individual sample history (2016: $u_{asv:h}$), seasons (2016: $u_{asv:s}$), and age classes (2013: $u_{asv:a}$). We followed Nakagawa & Schielzeth (2013) for calculation of r^2 from GLMMs with Poisson error distributions. Where multiple samples are present per individual (2016), repeatability can be estimated as the proportion of variance attributable to differential taxonomic composition across individuals divided by the sum of the variance explained by all other component terms estimating differential abundance $u_{asv:h}/(u_{asv:s} + u_{asv:h} + u_{asvh:s})$.

To extract information of specific bacterial taxa contributing to differential abundance across age groups or season, we used Poisson model outputs and subtracted the posterior distributions across group levels (2013: Age; 2016: Season) for each ASV and calculated a mean difference and 95% credible intervals from each resultant distribution to estimate the differential abundance for that taxonomic level. For example, the mean of the posterior distribution for ASV1:summer – that for ASV1:spring can be interpreted as the relative shift of ASV from spring to summer, where shifts can be considered significant when credible intervals do not span zero. These values were used to infer ASVs which exhibit the largest deviations from the means.

3.2. Results:

The microbiota communities of Soay Sheep were dominated by two phyla, Firmicutes and Bacteroidetes (Figure S1), as has been previous observed in most vertebrates (Ley *et al.* 2008) Principle coordinates analysis (PCoA) based on Bray-Curtis dissimilarity indicates clustering of samples by age and by season (Figure 2). The result of a PERMANOVA test on the 2013 data set shows a significant difference in group centroids for lambs and adults (pseudo-F = 7.161, $p < 0.001$). 11.34% of the variance in gut microbiota composition is explained by differences between lambs and adults. PERMANOVA results for the 2016 data show that group centroids for April and August are significantly distinct (pseudo-F = 2.026, p = 0.002), but season only explains 2.81% of the observed variance.

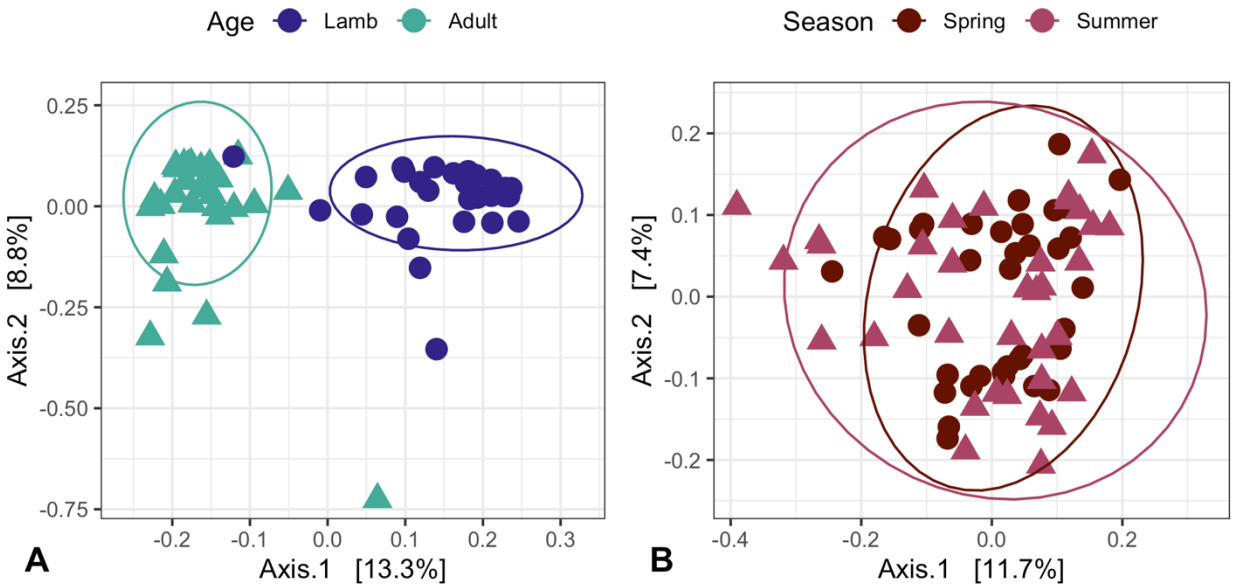
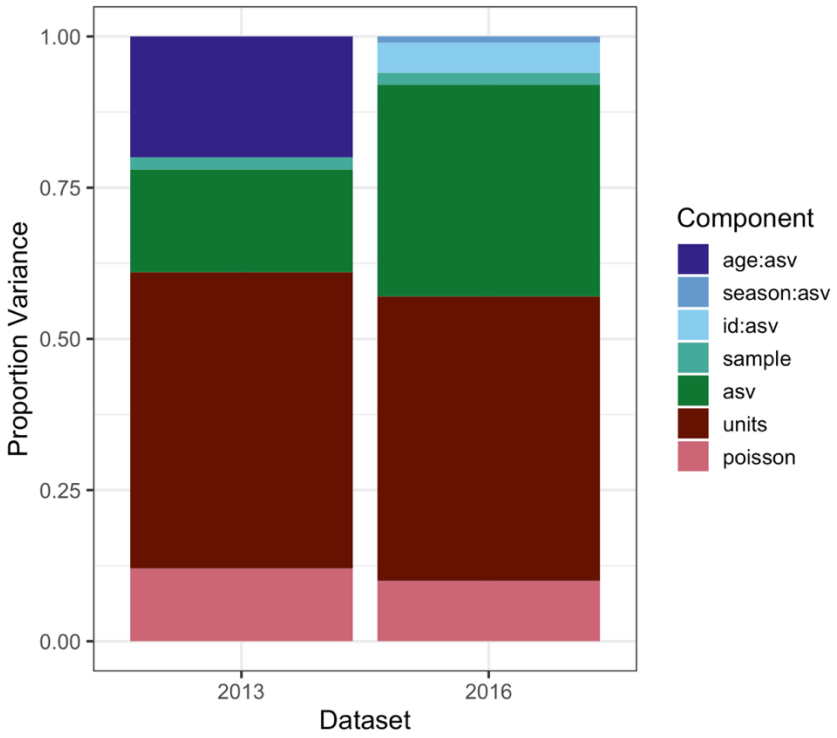


Figure 2. Soay Sheep gut microbiota beta diversity in adults and lambs from 2013 (A) and from April and August of 2016 (B). Principle coordinates (PCoA) plots represent Bray-Curtis dissimilarity indicating clustering of samples by group. Ellipsoids represent a 95% confidence interval surrounding each group.

Poisson GLMMs from 2013 data showed comparable results to ordination approaches, where a high proportion of variance is explained by community composition differences between age classes ($u_{asv:a}$: 19.88% CI 18.47-21.39%; Figure 2). Additional effects estimated by the model showed a substantial proportion of variance explained by taxonomic variation in ASV abundance (u_{asv} 2013: 17.35%), a small portion of variation explained by variation in mean library size

351 across samples (u_h 2013: 1.59 %), and considerable residual variance (estimated by the 'units'
352 term; $u_{asv:h}$ 2013: 49.42%; Figure 3).
353
354 Poisson GLMMs from 2016 data likewise showed comparable results to ordination approaches,
355 where community composition changed only modestly between seasons ($u_{asv:s}$: 1.24% CI 0.99 –
356 1.44%). In 2016 as well as 2013, additional effects estimated by the model showed a substantial
357 proportion of variance explained by taxonomic variation in ASV abundance (u_{asv} 2016:
358 34.74%), a small portion of variation explained by variation in mean library size across samples
359 ($u_{h:s}$ 2013: 1.99 %), and considerable residual variance ($u_{asv:h:s}$ 2013: 46.67%; Figure 3).
360 Repeated sampling of individuals in 2016 additionally showed moderate variance explained by
361 interindividual variation in community composition ($u_{asv:h}$ 2016: 5.49 %). Individuals in this
362 year exhibited repeatability of 11.1% for their microbiota community composition across
363 sampling time points.
364



365
366
367 **Figure 3**. Proportion variance explained by GLMM component terms for 2013 dataset including samples
368 from two age classes and 2016 dataset including samples from the same individuals over two seasons
369
370

371 We repeated Poisson models for both years using GLMMs with Gaussian family errors fitted to
372 CLR ratio-transformed ASVs to investigate robustness of this approach to transformations for
373 computational efficiency (ESM 1.4.3). For both datasets, GLMMs using CLR-normalised values
374 as the response variable showed agreement with both traditional approaches and raw read count
375 models (Figure S3) and computationally more efficient run times (ESM Section 2.1; Figure S3).
376
377 Using best linear unbiased predictions (BLUPs) from posterior distributions of GLMMs, we
378 estimated differential abundances of specific taxa across age classes and seasons (Figure 4). For
379 ASV-by-age effects (2013 $u_{asv:a}$, Figure 3), these taxa-specific differential abundances suggest
380 that ASVs demonstrating significant shifts between lambs and adults belong primary to two
381 phyla, Firmicutes and Bacteroidetes (Figure 4A-B). 683 out of 2,023 (33.76%) ASVs present in
382 2013 data showed significant shifts between lambs and adults (50.81% positive shifts, 49.19%
383 negative shifts). Bacteroidetes represented 52.16% of significant positive shifts into adulthood,
384 and Firmicutes represented 72.02% of the significant negative shifts into adulthood (Figure 4B;
385 Table S4). In contrast, aligned with modest ASV-by-season effects (2016 $u_{asv:s}$, Figure 3 &
386 Table S4), very few ASVs (24 of 2,364; 1.02%) showed significant shifts in abundance between
387 spring and summer sampling (Figure 4C-D; Table S4).
388
389 Our results show that there are substantive developmental shifts in the Soay sheep microbiota
390 between lambs and adults, and that the majority of taxa shifting in abundance belong to the
391 Bacteroidetes and Firmicutes. However, this raises the question of whether this shift is because
392 Bacteroidetes are *generally* more abundant in adults and firmicutes more abundant in lambs, or if
393 the ASVs that show these patterns *just happen* to be in these phyla. To illustrate how GLMMs
394 can be used to address questions of this sort, we modified our models for the 2013 dataset to
395 include additional taxonomic effects of family and phylum (ESM 1.4.4), allowing us to identify
396 taxonomic levels most responsible for differential abundances. This modified model for 2013
397 data showed that that among the 20% overall variance associated with compositional differences
398 between age classes (Figure 3), 13% is attributable to variance at the sequence variant (ASV)
399 level, 5% at the family level, and 2% at the phylum level (Table S3). Low proportion of variance
400 at the phylum level when multiple taxonomic levels are taken into account is due to variation in
401 the direction of ontogenic shifts at the family and ASV levels even within the two phyla most

402 associated with differential ASV abundance (Figure 5). For example, although two families
403 within the Bacteroidetes do show positive shifts into adulthood, some show no difference across
404 age classes, and one (Bacteroidaceae) shows a negative shift (Figure 5). Similarly, although
405 Firmicutes primarily are implicated in negative significant shifts in GLMMs accounting only for
406 ASV (Figure 4A-B), models decomposing the family and phylum levels show that only one
407 family is associated with a significant shift (Figure 5). These results illustrate how GLMMs can
408 be used to decompose the taxonomic levels at which ecological effects occur in microbiota
409 communities.
410

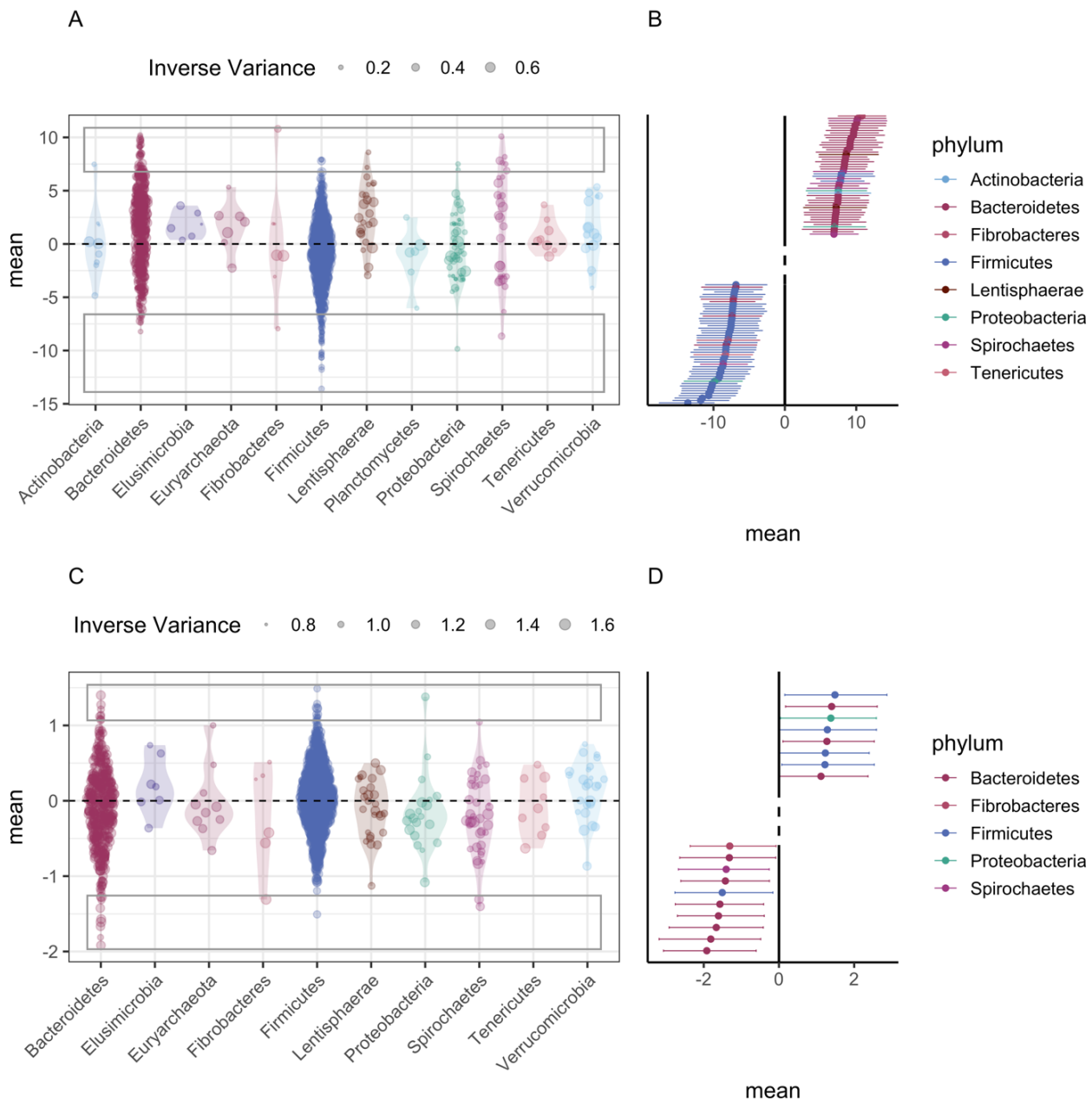


Figure 4. Differential abundances across age classes (A-B) or season (C-D) for individual ASVs calculated from GLMMs with Poisson error families and taxonomic level specified as ASV-only. A & C represent all ASV-level effects. Violin plots represent distribution of effect estimates, size of point represents the inverse variance of the estimate. Rectangles indicate the top 50 and bottom 50 differential abundances in order of magnitude presented in Forest plots (B & D). Forest plots represent point estimate and 95% confidence intervals for the ASVs involved in the 50 (age class) or 20 (season) strongest increases and decreases of abundance.

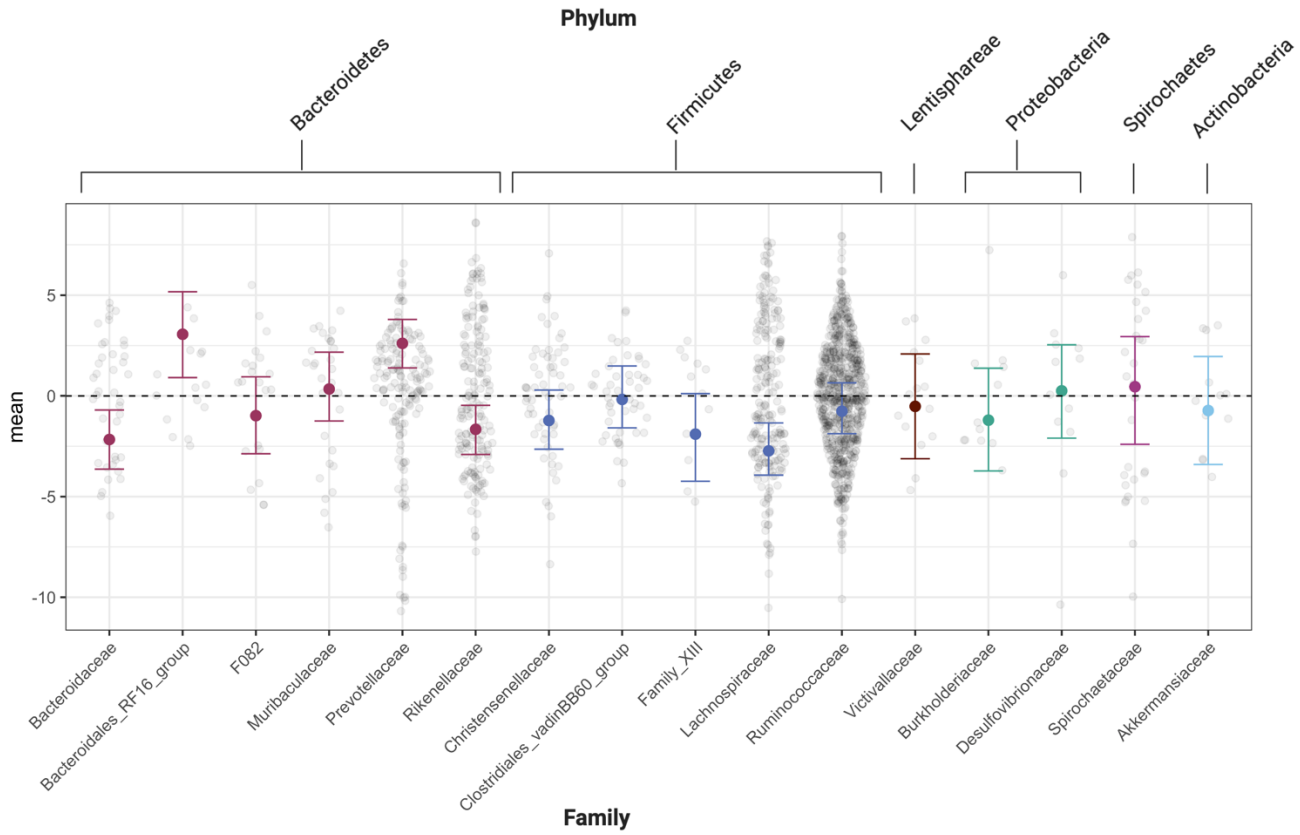


Figure 5. Differential abundances across age classes accounting for multiple levels of taxonomy. Data represents 2013 model estimates represent from GLMMs with Poisson error families and taxonomic level specified as ASV, family, and phylum. Points in sina distributions represent ASV-level estimates, point and error bars represent mean and 95% CI for family-level estimates, and colours indicate phylum categories.

4. Discussion & wider applications

We have described a novel approach to analysing meta-taxonomic data derived from NGS using a GLMM framework, and have illustrated this method using data describing variation in the gut microbiota community in wild sheep. Our approach represents an important step forward for researchers interested in using metataxonomic approaches to understand variation in community structure and in complex, non-experimental settings. It allows the well-established power and flexibility of GLMM-based approaches to be harnessed to decompose drivers of variation in NGS-derived data on the taxonomic composition of samples. Our example analyses provide very simple illustrations of how the approach can be used to estimate the contribution of host-related or environmental factors (specifically, age and season) to variation in the community structure of the gut microbiota. We show that results are comparable to those derived from widely-applied ordination-based approaches, and discuss the implications of observed variation in gut microbiota structure with age and season briefly below. However, these analyses are intended mainly as templates to help illustrate the approach, and barely scratch the surface of types of the important outstanding questions the method could be used to tackle with larger-scale data sets. Applying GLMMs to taxonomic-level sequence count data provides a rich toolkit from fields like quantitative genetics to dissect the contributions of different environmental and host factors to variation in community structure, with the potential to advance our understanding of community ecology, host-microbe or –pathogen interactions, and evolutionary dynamics.

In our analyses of wild Soay sheep, both GLMM and ordination-based approaches identified a stronger effect of age than season on the gut microbiota in our example data sets from wild Soay sheep. Changes in the structure of the gut microbiota across development in early life and during senescence in later adulthood are well-established in human studies (Popkes & Valenzano 2020), but remain poorly understood in natural populations. Our data clearly illustrate the gut microbiota community structure changes between recently weaned lambs and adults, and argues for further longitudinal studies in natural systems to test whether shifts in gut community structure could play a role in patterns of demographic ageing in the wild. In contrast, a growing number of studies in the wild have documented seasonal gut microbiota changes (Amato *et al.* 2015; Maurice *et al.* 2015; Ren *et al.* 2017; Orkin *et al.* 2019). The absence of a strong seasonal

458 different in our data may be related to the relative homogeneity of the herbivorous diet of Soay
459 sheep, as most previous wild studies are of omnivores with strong seasonal shifts in diet
460 preference. Alternatively, it may be because the spring and summer seasons we sampled in are
461 both periods of relatively high food abundance and quality, and more pronounced change would
462 be observed in autumn and winter when habitat quality and food availability changes more
463 dramatically. In future studies, repeated sampling of the same individuals over time will be
464 crucial to understand the effects of age, environmental and other variables on gut community
465 structure. Unlike classical ordination-based approaches, our GLMM-based method allows
466 within-individual changes over time (e.g. with season or age) to be separated from among-
467 individual variation in average community structure across samples, via inclusion of individual-
468 level random effects. The application of our approach to such longitudinal metataxonomic data
469 sets will help researchers robustly estimate within-individual patterns of change in community
470 structures over time or space, whilst also estimating how repeatable community structure across
471 hosts.
472
473 A compelling benefit of GLMM approaches can help to contextualise wild metabarcoding data
474 within host-microbe evolutionary dynamics. Individual repeatability is widely estimated in
475 ecological and quantitative genetic literature as a key component of phenotype, and estimating
476 consistency of microbiota community structure likewise offers insight into the extent to which
477 host control and environmental selection determine species composition (Foster *et al.* 2017). The
478 GLMM structure presented here highlights the inclusion of individual repeatability, but is also
479 readily extendible to mixed-effect model extensions of more a more complex nature to address
480 emerging questions in the field. For example, kinship & inbreeding effects on microbiota
481 composition in have been explored previously in microbiome studies but via ordination methods
482 using genetic clusters as grouping units (Yuan *et al.* 2015). Incorporating well-developed
483 methods of including host pedigree or relatedness matrices as a random effect in a mixed-effects
484 model (animal model) with GLMM approaches to microbiota community analysis can offer
485 insights into heritability and inbreeding effects on community composition as it compares to
486 other forces in the population (Wilson *et al.* 2010). Previous work in Soay sheep highlight the
487 ability of these approaches to determine to a high-resolution heritability in antibody response
488 (Sparks *et al.* 2019), resistance to helminths (Smith *et al.* 1999) and body size traits (Bérénos *et*

489 *al.* 2015). Conceptually we can easily extend these methods to models in which microbiota reads
490 are the response, as well as incorporating factors which can confound heritability estimates such
491 as maternal effects or shared nest or litter effects (Kruuk & Hadfield 2007). Finally, maternal and
492 social effects as well spatial autocorrelation in ecological datasets can be incorporated into this
493 model framework (Hayward *et al.* 2010; Pemberton 2010; Stopher *et al.* 2012; Albery *et al.*
494 2020)
495
496 Although we used GLMMs with Poisson error distributions to illustrate the benefits of a GLMM
497 approach, we also acknowledge several limitations with this method. First, even applied to
498 illustrative datasets with a modest number of random effects, run-time for MCMCglmm-based
499 Poisson models is very slow due to the high number of rows for reads-by-ASV-by-sample data
500 frames. We therefore anticipate that dependent on sample size, diversity of ASVs present, and
501 complexity of covariates required these models may become prohibitively slow. For this reason,
502 we explored log ratio (CLR) transformations for read counts as responses (ESM Section 1.4.3).
503 CLR transformations are increasingly common, acknowledge the compositional nature of count
504 data, allow estimation of error associated with CLR estimates for each ASV per sample, and
505 allow gaussian error distributions to be fit in models structurally similar to Poisson models
506 (Gloor *et al.* 2017). CLR models applied to illustrative datasets here showed comparable results
507 to Poisson models with regard to proportion variance attributable to model components. There
508 was slight divergence in the relative proportions of variance explained by CLR vs Poisson
509 models in that the age effect was associated with a slightly more modest proportion variance in
510 CLR models, which may be attributable to the transformations accounting for the limitations of
511 the count data as bound by upper limits of the sequencing platform. We suggest that CLR
512 transformations therefore can offer a more computationally-efficient route for GLMM
513 approaches without sacrificing model complexity. This may be particularly useful for instances
514 where it is desirable to extract posterior distributions to estimate differential abundance of
515 specific taxa given storage of posterior distributions for all random effect levels introduces
516 significant demands on RAM. These possible issues may also be circumvented by exploring
517 packages beyond MCMCglmm including 'ASREML' (Gilmour *et al.* 2009) for gaussian models
518 which estimates mean and 95% confidence intervals (rather than saving posterior distributions)
519 for random effects or 'brms' (Bürkner 2017) which uses Hamiltonian sampling rather than

520 Markov chain Monte Carlo methods and is more amenable to parallelisation. We believe there
521 are several relevant options given the significant interest for understanding functional
522 implications of host microbiota community composition.
523
524 Full realisation of the role of microbiota communities in the wild depends on the ability to
525 decompose global and specific aspects of community structure effects. Ecological and
526 evolutionary tool kits with regard to structure of GLMMs and estimation of inter-related
527 explanatory variable contribution to phenotype variance are applied here to investigate suitability
528 for detecting general shifts in community composition associated with factors such as host age
529 and sampling season (Bolker *et al.* 2009). We further developed GLMMs applied to these
530 datasets for (a) inference of specific ASVs implicated in these community-level shifts and (b) the
531 taxonomic level from which most variance is explained. First, estimation of group comparisons
532 for each ASV simultaneously highlight which (if any) ASVs exhibit significant changes between
533 group levels for factors of interest (Figure 4). Although individual ASVs are unlikely to be
534 responsible in isolation for host states observed, leveraging model output in this way can provide
535 information for hypothesis generation for taking forward relevant information from community
536 composition shifts to understand their impacts on host phenotype (Clemente *et al.* 2012;
537 Vayssier-Taussat 2014). Next, it is important to consider how taxonomy and phylogeny of the
538 microbial community present may either confound results or provide a foundation for inference
539 (Washburne *et al.* 2018). We offer an initial suggestion of accounting for evolutionary
540 relationships among bacterial species here by expanding model structures to estimate variance of
541 model components not just at the level of ASV composition but also the family and phylum level
542 (Figure 5). For example, these model extensions highlight that the phylum level would be a
543 reductive view of composition shifts associated with age of Soay sheep and that there is
544 substantial variation at the family level within each phylum. Differential abundances obtained in
545 this fashion should represent similar insights to linear discriminant analysis (LDA) used in
546 approaches such as LefSE (Segata *et al.* 2011). We suggest that these estimations can be further
547 developed in more powerful ways using quantitative genetic methods designed to account for
548 phylogeny and taxonomic-relatedness via means such as approaches incorporating both random
549 effects specifying nested taxonomic rank and a bacterial phylogeny matrix to account for both
550 continuous additive genetic variation as well as specific taxonomic level effects (Hadfield &

551 Nakagawa 2010). Building this infrastructure into microbiota community GLMM analyses
552 alongside inter-related effects of host genetics, spatiotemporal variation, host variation, and
553 environmental factors can provide the flexibility to robustly address emerging questions
554 surrounding the role of the microbiome in ecology and evolution.
555
556 Beyond applications to microbiota community analyses, approaches outlined in this manuscript
557 are applicable more broadly to multiple types of metataxonomic data being collected across
558 myriad systems and topics of interest. For example, there has been great interest in describing the
559 dynamics of the parasite community as an ecosystem and their influence on host health
560 (Pedersen & Fenton 2007; Graham 2008; Rynkiewicz *et al.* 2015). Recent work has leveraged
561 non-invasive faecal sampling to apply metabarcoding to gastrointestinal parasite communities of
562 nematodes (Avramenko *et al.* 2015; Aivelo & Medlar 2017; Avramenko *et al.* 2017) and
563 coccidia (Vermeulen *et al.* 2016). Another area of interest in disease ecology and evolution is
564 high-resolution classification of diet composition and its relationship to host phenotypes.
565 Bayesian mixed model approaches have also recently been applied to analysis of presence and
566 absence of *Cyanistes caeruleus* (blue tit) diet components and align conceptually with
567 approaches presented in this manuscript (Shutt *et al.* 2020). GLMM approaches to
568 metabarcoding data maintain key similarities to other multivariate community ecology
569 approaches to abundance data (Wang *et al.* 2012) while integrating benefits of ecological and
570 evolutionary approaches to quantifying phenotypic variation. We therefore suggest that
571 approaches presented in this manuscript can be applied across a range of systems and data types
572 for powerful and flexible understanding of complex drivers of community dynamics.
573
574
575
576
577
578
579

References

Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**, 139–160.

Aivelo, T. & Medlar, A. (2017). Opportunities and challenges in metabarcoding approaches for helminth community identification in wild mammals. *Parasitology*, **145**, 608–621.

Alberdi, A., Aizpurua, O., Bohmann, K., Zepeda-Mendoza, M.L. & Gilbert, M.T.P. (2016). Do Vertebrate Gut Metagenomes Confer Rapid Ecological Adaptation? *Trends in Ecology & Evolution*, **31**, 689–699.

Albery, G.F., Kirkpatrick, L., Firth, J.A. & Bansal, S. (2020). Unifying spatial and social network analysis in disease ecology (D. Farine, Ed.). *Journal of Animal Ecology*, **30**, 2009–18.

Amato, K.R., Leigh, S.R., Kent, A., Mackie, R.I., Yeoman, C.J., Stumpf, R.M., Wilson, B.A., Nelson, K.E., White, B.A. & Garber, P.A. (2015). The gut microbiota appears to compensate for seasonal diet variation in the wild black howler monkey (Alouatta pigra). *Microbial ecology*, **69**, 434–443.

Anders, S. & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome biology*, **11**, R106.

Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 32–46.

Avramenko, R.W., Redman, E.M., Lewis, R., Bichuette, M.A., Palmeira, B.M., Yazwinski, T.A. & Gilleard, J.S. (2017). The use of nemabiome metabarcoding to explore gastro-intestinal nematode species diversity and anthelmintic treatment effectiveness in beef calves. *International journal for parasitology*, 1–10.

Avramenko, R.W., Redman, E.M., Lewis, R., Yazwinski, T.A., Wasmuth, J.D. & Gilleard, J.S. (2015). Exploring the Gastrointestinal 'Nemabiome': Deep Amplicon Sequencing to Quantify the Species Composition of Parasitic Nematode Communities (E. Serrano Ferron, Ed.). *PLoS One*, **10**, e0143559–18.

Bérénos, C., Ellis, P.A., Pilkington, J.G., Lee, S.H., Gratten, J. & Pemberton, J.M. (2015). Heterogeneity of genetic architecture of body size traits in a free-living population. *Molecular Ecology*, **24**, 1810–1830.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.-S.S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, **24**, 127–135.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. 1–1.

Callahan, B.J., McMurdie, P.J. & Holmes, S.P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *Isme Journal*, **11**, 2639–2643.

614 Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. & Holmes, S.P. (2016).
615 DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods*, **13**, 581–
616 583.

617 Carding, S., Verbeke, K., Vipond, D.T., Corfe, B.M. & Owen, L.J. (2015). Dysbiosis of the gut
618 microbiota in disease. *Microbial Ecology in Health and Disease*, **26**, 1–10.

619 Charmantier, A., Garant, D. & Kruuk, L.E.B. (Eds.). (2014). *Quantitative Genetics in the Wild*. Oxford
620 University Press, Oxford, UK.

621 Clemente, J.C., Ursell, L.K., Parfrey, L.W. & Knight, R. (2012). The Impact of the Gut Microbiota on
622 Human Health: An Integrative View. *Cell*, **148**, 1258–1270.

623 Consortium, T.I.S. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar
624 disorder. *Nature*, **460**, 748–752.

625 Cox-Foster, D.L., Conlan, S., Holmes, E.C., Palacios, G., Evans, J.D., Moran, N.A., Quan, P.-L., Briese,
626 T., Hornig, M., Geiser, D.M., Martinson, V., vanEngelsdorp, D., Kalkstein, A.L., Drysdale, A., Hui,
627 J., Zhai, J., Cui, L., Hutchison, S.K., Simons, J.F., Egholm, M., Pettis, J.S. & Lipkin, W.I. (2007). A
628 metagenomic survey of microbes in honey bee colony collapse disorder. *Science*, **318**, 283–287.

629 Desbonnet, L., Clarke, G., Shanahan, F., Dinan, T.G. & Cryan, J.F. (2014). Microbiota is essential for
630 social development in the mouse. *Molecular psychiatry*, **19**, 146–148.

631 Fontaine, S.S. & Kohl, K.D. (2020). Optimal integration between host physiology and functions of the
632 gut microbiome. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*,
633 **375**, 20190594–9.

634 Foster, K.R., Schluter, J., Coyte, K.Z. & Rakoff-Nahoum, S. (2017). The evolution of the host
635 microbiome as an ecosystem on a leash. *Nature*, **548**, 43–51.

636 Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*.
637 Cambridge University Press.

638 Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2009). ASReml User Guide. 1–398.

639 Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V. & Egozcue, J.J. (2017). Microbiome Datasets Are
640 Compositional: And This Is Not Optional. *Frontiers in microbiology*, **8**, 2224.

641 Goldstein, D.B. (2009). Common Genetic Variation and Human Traits. *N. Engl. J. Med.*, 1696–1703.

642 Goodrich, J.K., Di Rienzi, S.C., Poole, A.C., Koren, O., Walters, W.A., Caporaso, J.G., Knight, R. &
643 Ley, R.E. (2014). Conducting a microbiome study. *Cell*, **158**, 250–262.

644 Graham, A.L. (2008). Ecological rules governing helminth-microparasite coinfection. *Proceedings of the*
645 *National Academy of Sciences of the United States of America*, **105**, 566–570.

646 Hadfield, J.D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The
647 MCMCglmm R Package. *Journal of Statistical Software*, **33**, 1–22.

648 Hadfield, J.D. & Nakagawa, S. (2010). General quantitative genetic methods for comparative biology:
649 phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of*
650 *Evolutionary Biology*, **23**, 494–508.

651 Hayward, A.D., Pilkington, J.G., Pemberton, J.M. & Kruuk, L.E.B. (2010). Maternal effects and early-life
652 performance are associated with parasite resistance across life in free-living Soay sheep.
653 *Parasitology*, **137**, 1261–1273.

654 Koskella, B., Hall, L.J. & Metcalf, C.J.E. (2017). The microbiome beyond the horizon of ecological and
655 evolutionary theory. *Nature Ecology & Evolution*, **1**, 1606–1615.

656 Kruuk, L.E.B. & Hadfield, J.D. (2007). How to separate genetic and environmental causes of similarity
657 between relatives. *Journal of Evolutionary Biology*, **20**, 1890–1903.

658 Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L.,
659 Tucker, T.A., Schrenzel, M.D., Knight, R. & Gordon, J.I. (2008). Evolution of mammals and their
660 gut microbes. *Science*, **320**, 1647–1651.

661 Loh, P.-R., Bhatia, G., Gusev, A., Finucane, H.K., Bulik-Sullivan, B.K., Pollack, S.J., de Candia, T.R.,
662 Lee, S.H., Wray, N.R., Kendler, K.S., O'Donovan, M.C., Neale, B.M., Patterson, N. & Price, A.L.
663 (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast
664 variance-components analysis. *Nature Publishing Group*, 1–11.

665 Mandal, S., Van Treuren, W., White, R.A., Eggesbø, M., Knight, R. & Peddada, S.D. (2015). Analysis of
666 composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology*
667 *in Health and Disease*, **26**, 1–8.

668 Maurice, C.F., Knowles, S.C.L., Ladau, J., Pollard, K.S., Fenton, A., Pedersen, A.B. & Turnbaugh, P.J.
669 (2015). Marked seasonal variation in the wild mouse gut microbiota. *Isme Journal*, **9**, 2423–2434.

670 McMurdie, P.J. (2018). Normalization of Microbiome Profiling Data. *Microbiome Analysis*, pp. 143–168.
671 Methods and Protocols. Humana Press, New York, NY, New York, NY.

672 McMurdie, P.J. & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and
673 Graphics of Microbiome Census Data (M. Watson, Ed.). *PLoS One*, **8**, e61217.

674 McMurdie, P.J. & Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is
675 Inadmissible (A.C. McHardy, Ed.). *PLoS computational biology*, **10**, e1003531.

676 Morgan, X.C. & Huttenhower, C. (2012). Chapter 12: Human Microbiome Analysis (F. Lewitter & M.
677 Kann, Eds.). *PLoS computational biology*, **8**, e1002808–14.

678 Nakagawa, S. & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized
679 linear mixed-effects models. *Methods in Ecology and Evolution*, **4**, 133–142.

680 Orkin, J.D., Campos, F.A., Myers, M.S., Cheves Hernandez, S.E., Guadamuz, A. & Melin, A.D. (2019).
681 Seasonality of the gut microbiota of free-ranging white-faced capuchins in a tropical dry forest. *Isme*
682 *Journal*, **13**, 183–196.

683 Pedersen, A.B. & Fenton, A. (2007). Emphasizing the ecology in parasite community ecology. *Trends in*
684 *Ecology & Evolution*, **22**, 133–139.

685 Pemberton, J.M. (2010). Evolution of quantitative traits in the wild: mind the ecology. *Philosophical*
686 *Transactions of the Royal Society of London B: Biological Sciences*, **365**, 2431–2438.

687 Popkes, M. & Valenzano, D.R. (2020). Microbiota–host interactions shape ageing dynamics.
688 *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **375**, 20190596–9.

689 Ren, T., Boutin, S., Humphries, M.M., Dantzer, B., Gorrell, J.C., Coltman, D.W., McAdam, A.G. & Wu,
690 M. (2017). Seasonal, spatial, and maternal effects on gut microbiome in wild red squirrels.
691 *Microbiome*, **5**, 163–14.

692 Rynkiewicz, E.C., Pedersen, A.B. & Fenton, A. (2015). An ecosystem approach to understanding and
693 managing within-host parasite community dynamics. *Trends in Parasitology*, **31**, 212–221.

694 Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A.,
695 Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J. &
696 Weber, C.F. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-
697 Supported Software for Describing and Comparing Microbial Communities. *Applied and*
698 *environmental microbiology*, **75**, 7537–7541.

699 Segata, N., Izaard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S. & Huttenhower, C. (2011).
700 Metagenomic biomarker discovery and explanation. *Genome biology*, **12**, R60.

701 Shutt, J.D., Nicholls, J.A., Trivedi, U.H., Burgess, M.D., Stone, G.N., HADFIELD, J.D. & Phillimore,
702 A.B. (2020). Gradients in richness and turnover of a forest passerine's diet prior to breeding: A mixed
703 model approach applied to faecal metabarcoding data. *Molecular Ecology*, **29**, 1199–1213.

704 Smith, J.A., Wilson, K., Pilkington, J.G. & Pemberton, J.M. (1999). Heritable variation in resistance to
705 gastro-intestinal nematodes in an unmanaged mammal population. *Proceedings of the Royal Society*
706 *of London. Series B, Biological sciences*, 1283–1290.

707 Sparks, A.M., Watt, K., Sinclair, R., Pilkington, J.G., Pemberton, J.M., McNeilly, T.N., Nussey, D.H. &
708 Johnston, S.E. (2019). The genetic architecture of helminth-specific immune responses in a wild
709 population of Soay sheep (Ovis aries) (K. Bomblies, Ed.). *PLoS genetics*, **15**, e1008461–29.

710 Stopher, K.V., Walling, C.A., Morris, A., Guinness, F.E., Clutton-Brock, T.H., Pemberton, J.M. &
711 Nussey, D.H. (2012). Shared spatial effects on quantitative genetic parameters: accounting for spatial
712 autocorrelation and home range overlap reduces estimates of heritability in wild red deer. *Evolution*,
713 **66**, 2411–2426.

714 Sudo, N., Chida, Y., Aiba, Y., Sonoda, J., Oyama, N., Yu, X.-N., Kubo, C. & Koga, Y. (2004). Postnatal
715 microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response in
716 mice. *The Journal of physiology*, **558**, 263–275.

717 Vayssier-Taussat, M. (2014). Shifting the paradigm from pathogens to pathobiome: new concepts in the
718 light of meta-omics. 1–7.

719 Vermeulen, E.T., Lott, M.J., Eldridge, M.D.B. & Power, M.L. (2016). Evaluation of next generation
720 sequencing for the analysis of Eimeria communities in wildlife. *Journal of Microbiological Methods*,
721 **124**, 1–9.

722 Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. (2012). REVIEW Five Years of GWAS
723 Discovery. *The American Journal of Human Genetics*, **90**, 7–24.

724 Wahlström, A., Sayin, S.I., Marschall, H.-U. & Bäckhed, F. (2016). Intestinal Crosstalk between Bile
725 Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metabolism*, **24**, 41–50.

726 Wang, Y., Naumann, U., Wright, S.T. & Warton, D.I. (2012). mvabund- an Rpackage for model-based
727 analysis of multivariate abundance data. *Methods in Ecology and Evolution*, **3**, 471–474.

728 Washburne, A.D., Morton, J.T., Sanders, J., McDonald, D., Zhu, Q., Oliverio, A.M. & Knight, R. (2018).
729 Methods for phylogenetic analysis of microbiome data. *Nature Microbiology*, 1–10.

730 Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., González, A., Lozupone, C., Zaneveld, J.R.,
731 Vázquez-Baeza, Y., Birmingham, A., Hyde, E.R. & Knight, R. (2017). Normalization and microbial
732 differential abundance strategies depend upon data characteristics. *Microbiome*, **5**, 1–18.

733 Wilson, A.J., Coltman, D.W., Pemberton, J.M., Overall, A.D.J., Byrne, K.A. & Kruuk, L.E.B. (2005).
734 Maternal genetic effects set the potential for evolution in a free-living vertebrate population. *Journal*
735 *of Evolutionary Biology*, **18**, 405–414.

736 Wilson, A.J., Réale, D., Clements, M.N., Morrissey, M.M., Postma, E., WALLING, C.A., Kruuk, L.E.B.
737 & Nussey, D.H. (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology*, **79**,
738 13–26.

739 Yuan, M.L., Dean, S.H., Longo, A.V., Rothermel, B.B., Tuberville, T.D. & Zamudio, K.R. (2015).
740 Kinship, inbreeding and fine-scale spatial structure influence gut microbiota in a hindgut-fermenting
741 tortoise. *Molecular Ecology*, **24**, 2521–2536.

742 Zuur, A., leno, E.N., Walker, N., Saveliev, A.A. & Smith, G.M. (2009). *Mixed Effects Models and*
743 *Extensions in Ecology with R*. Springer Science & Business Media.

Electronic Supplementary Information

A mixed model approach for estimating drivers of microbiota community composition and differential taxonomic abundance

Amy R Sweeny*, Hannah E Lemon*, Anan Ibrahim, Daniel H Nussey, Andrew Free[$], Luke McNally[$]

* These authors contributed equally

[$] Joint last authors

1. Methods

1.1 Data collection

Faecal samples were collected at at two different time points and each provided a small data set with which to test for differences in the gut microbiota community structure among individuals of different ages and within individuals in different seasons. In August 2013, during the summer catch field season, 58 faecal samples were collected rectally from animals at capture. 30 of these samples were taken from lambs (N = 15 Male, 15 Female) and 28 were taken from adults (age range 2 – 13 years; N = 6 Male, 22 Female). In 2016 faecal samples were collected from pasture from 36 marked adult female sheep (age range 1- 13 years) in spring (between 4th and 27th April). These samples were collected by observing identifiable, marked individuals at close quarters until they defecated and then collecting the faecal samples from the pasture within a minute. The same individuals were resampled during the summer catch around three months later (12th – 18th August) using rectal faecal collection. Samples were stored at -20C within approximately 3 hours of collection.

1.2 Laboratory processing

DNA extractions for the 2013 samples were carried out using the Qiagen 'QIAamp DNA Stool' kit. DNA extractions for the 2016 samples were carried out using the MoBio PowerFecal DNA Isolation kit (Mo Bio, Carlsbad, CA, USA) following the manufacturers guidelines. DNA extractions were stored at -20C until they were amplified using polymerase

chain reaction. The V4 region of the bacterial 16S rRNA gene was amplified using the 515F forward primer and Golay barcoded 806R reverse primer series (Caporaso et al, 2012). Next-generation DNA sequencing was carried out at Source BioScience LifeSciences for the 2013 samples and at Edinburgh Genomics for the 2016 samples using Illumina MiSeq v2 to generate 250 base pair (bp) paired-end reads.

1.2.1 2013 Samples

Adaptation to the standard Qiagen Stool DNA extraction protocol: All centrifugation steps were carried out at 15,000 rcf/g, Step 4 – Centrifuge for 1.5 minutes, Step 5 – Supernatant volumes below 1.2ml were supplemented with Buffer ASL, Step 7 & 8 – Centrifuge for 4.5 minutes, Step 12 – Incubate at 95C for 10 minutes, Step 14 & 15 – Centrifuge for 1.5 minutes, Step 16 – Centrifuge for 4.5 minutes. All PCRs were performed in 25ul reactions: 12.25ul H_2O, 5ul Q Solution, 2.5ul PCR Buffer (10x), 1ul dNTP Mix, 0.25ul Hot star taq polymerase, 1ul Forward Primer (10uM), 1ul Reverse Primer (10uM) and 2ul DNA. The PCR conditions were as follows: Initial denaturation at 96C for 15 minutes then 35 cycles of denaturing at 94C for 45 seconds, annealing at 50 degrees for 60 seconds, extension at 72C for 90 seconds and a final extension at 72C for 10 minutes. PCR products were cleaned using Zymo DNA 96 Clean and Concentrator kit (all carried out by Source BioScience LifeSciences).

1.2.2 2016 Samples

All PCRs were performed in 25ul reactions using Roche reagents: 18.5ul H_2O, 1ul $MgCl_2$(25mM), 2.5ul PCR Buffer (10x), 0.5ul dNTP Mix (10mM), 0.25ul Taq DNA polymerase (5U/ul), 0.625ul Forward Primer (10uM), 0.625ul Reverse Primer (10uM) and 1ul DNA. The PCR conditions were as follows: Initial denaturation at 94C for 3 minutes then 25 cycles of denaturing at 94C for 45 seconds, annealing at 50 degrees for 60 seconds, extension at 72C for 90 seconds and a final extension at 72C for 10 minutes. PCR products were cleaned up using Wizard SV Gel and PCR Clean-up System (reference).

1.3 Data Processing: DADA2 Pipeline Parameters

Sequences were processed using the DADA2 pipeline in R (v1.12.1) to call amplicon sequence variants (ASVs) (Callahan *et al.* 2016). Briefly, forward and reverse reads were examined for quality, trimmed and filtered (See Supplementary Material). Sequences for both

datasets were independently denoised using the DADA2 algorithm (Callahan et al 2016), forward and reverse reads merged, ASVs inferred and tabulated, and putative chimeras removed. Taxonomy was assigned to each ASV using the naive Bayesian classifier of DADA2 and the Silva Project v132 database formatted for DADA2.

Bioinformatics processing followed the workflow of the DADA2 Pipeline Tutorial (1.12) (available at https://benjjneb.github.io/dada2/tutorial.html). Sequences for both datasets were examined for quality to determine trimming parameters. Sequences were trimmed at 240bp (Forward) and 200bp (Reverse) in 2016, and at 200bp (Forward) and 150bp (Reverse) in 2013, where sequences were of poorer quality. Standard filtering parameters were used (maxN=0, truncQ=2, rm.phix=TRUE and maxEE=2), with the exception of an increase of maxEE=10 for 2013 sequences to improve read retention during filtering. The average proportion of reads retained per sample at the end of the bioinformatics processing was ~0.7 for 2016 data and ~0.6 for 2013 data.

Sample metadata, taxonomy tables, and an ASV abundance matrix were integrated into a phyloseq object for downstream analysis (McMurdie & Holmes 2013). 2013 data contained 5,156 ASVs and 2,454,172 total reads (average 42,313.31; range 23,462-56,289). 2016 phyloseq objects contained 8,593 ASVs and 3,911,335 total reads (average 42,313.31; range 32,940-103,743). Prior to analysis an abundance filter was applied to datasets of both years, eliminating ASVs with a total abundance of <100 reads, resulting in 2,023 ASVs for 2013 and 2,364 for 2016. We ran a series of mixed models using multiple filtering criteria to ensure model outputs were not dependent on data processing steps (see Supplementary Material).

1.4 Statistical analysis

1.4.1 Traditional analysis additional tests

In addition to beta diversity, alpha diversity (diversity within an individual) was calculated using the Simpson index . This was calculated across categorical variables: age and season and homogeneity of dispersion was tested using a PERMDISP with the function 'betadisper' in the Vegan package (Anderson, 2001; Oksanen et al, 2013).

1.4.2 Poisson GLMM Abundance Threshold Validation

As a validation of data filtering in our mixed model approach, we ran 2016 models on data using the following criteria: (i) no abundance filter, (ii) a threshold of 50 total reads, (iii) a threshold of 100 total reads, and (iv) a threshold of 500 total roads to ensure model outputs were not dependent on data processing steps. Neither the rank order nor the absolute values of proportion variance explained by component terms of interest were notably affected by abundance filtering levels (Figure S2). The primary effect of increased abundance filtering was a reduction in variance explained by distributional Poisson variation in the data, where and the highest variance explained by the row-level residual variance of differential abundance across samples (component "units", Figure S2).

1.4.3 Centered log ratio approach to sheep data

We used the package 'ALDEx2' (Fernandes *et al.* 2013) to calculate a centered log-ratio for each asv in each sample. In addition to clr values, ALDEx2 estimates technical variation for each asv from a probability distribution using Monte-Carlo instances drawn from the Drichlet distribution, which maintains the compositional nature of the data. MCMCglmm allows specification of this technical variance as measurement error within model syntax. The variance for each asv:sample observation was therefore calculated from probability distributions generated from 128 Monte-Carlo instances and was specified as the measurement error for the response variable in CLR models. CLR model syntax to estimate the same terms as Poisson GLMMs (Main text Section 3.1), with the exclusion of the effects estimating variation in mean library size across age $a\beta_a$ (2013) or season $s\beta_s$ (2016) and across sample u_h (2013) or $u_{h:s}$ (2016), given that this technical variation is contained within the measurement error structure described above. Model syntax for ratio-normalised values were therefore as follows:

2013:

$$clr_{h,asv} = \beta_0 + u_{asv} + u_{asv:a} + u_{asv:h}$$

2016:

$$clr_{h,asv,s} = \beta_0 + u_{asv} + u_{asv:h} + u_{asv:s} + u_{asv:h:s}$$

Models were specified in MCMCglmm with gaussian error families and proportion of variance explained by each model component was calculated as described in main text (Section 3.1)

1.4.4 Taxonomic resolution models

We fit a set of models to further investigate the drivers of differential abundance across host (age) and environment (season) groups by including multiple levels of taxonomy (ASV, family, phylum). These models were fit as GLMMs with Poisson error families for both 2013 and 2016 as described above, with additional random effects describing family and phylum effects as follows for 2013. Resultant model structure each dataset were as follows:

2013:

$$log(\bar{y}_{h,asv}) = \beta_0 + a\beta_a + u_h + u_{asv} + \boldsymbol{u_f} + \boldsymbol{u_p} + u_{asv:a} + \boldsymbol{u_{f:a}} + \boldsymbol{u_{p:a}} + u_{asv:h}$$

2016:

$$log(\bar{y}_{h,asv,s}) = \beta_0 + s\beta_s + u_{h:s} + u_{asv} + \boldsymbol{u_f} + \boldsymbol{u_p} + u_{asv:s} + u_{asv:s} + \boldsymbol{u_{f:s}} + \boldsymbol{u_{p:s}} + u_{asv:h:s}$$

Where bolded terms represent additional terms incorporated to models presented in the main text (Section 3.1). In each model, $\boldsymbol{u_f}$ and $\boldsymbol{u_p}$ represent taxonomic variation in overall abundance at the family and phyla levels, and family- or phyla- level variation in community composition structure across age classes ($\boldsymbol{u_{f:a}} + \boldsymbol{u_{p:a}}$) or seasons ($\boldsymbol{u_{f:s}} + \boldsymbol{u_{p:s}}$) are described by additional interaction terms. Differential abundance calculations using posterior distributions were carried out for the family and phylum level as described in the main text for differential abundance of ASVs.

2. Supplementary Results & Figures

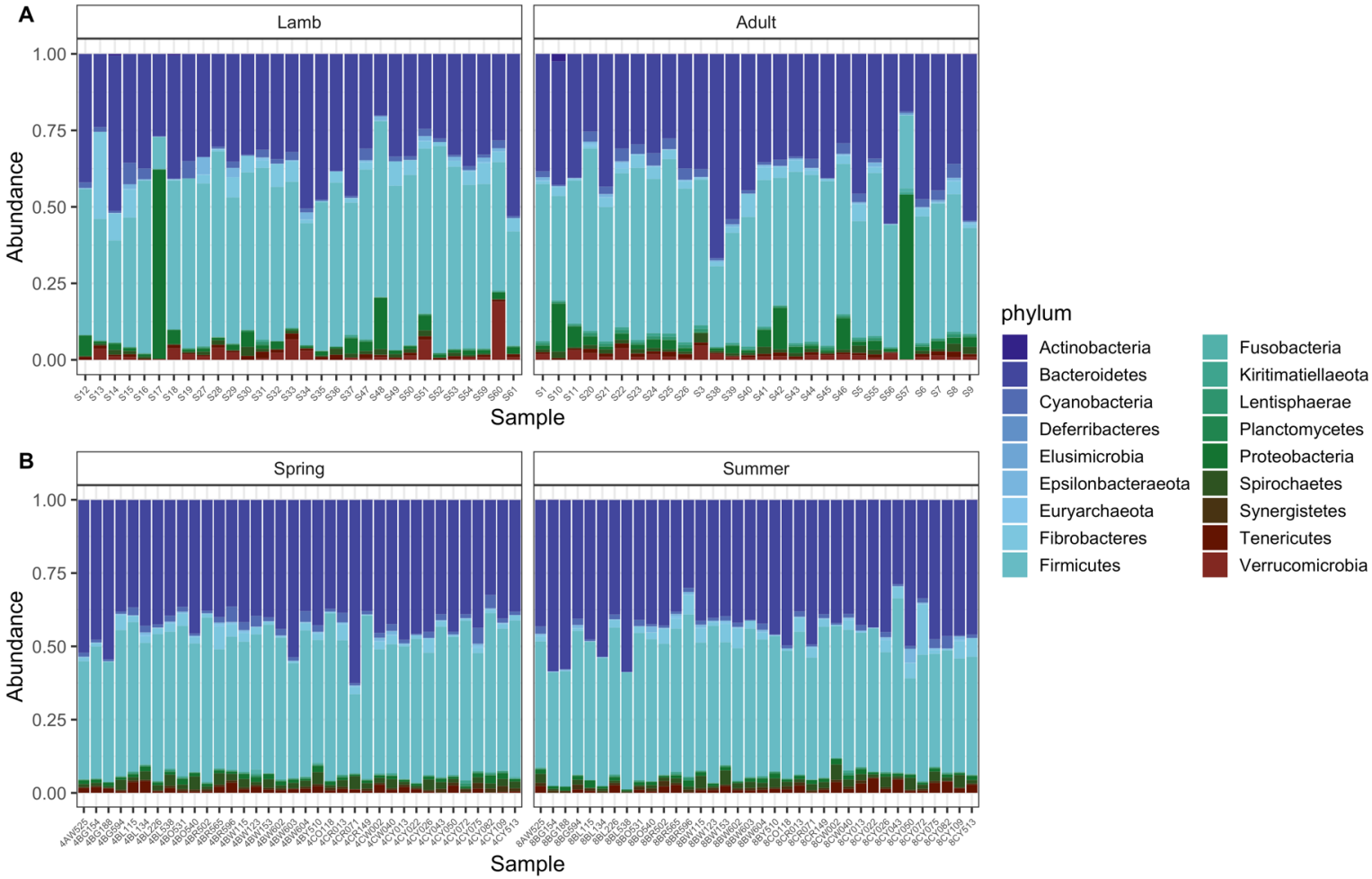


Figure S1. Relative abundance of bacterial phyla present in Soay Sheep data from (A): 2013 and (B): 2016.

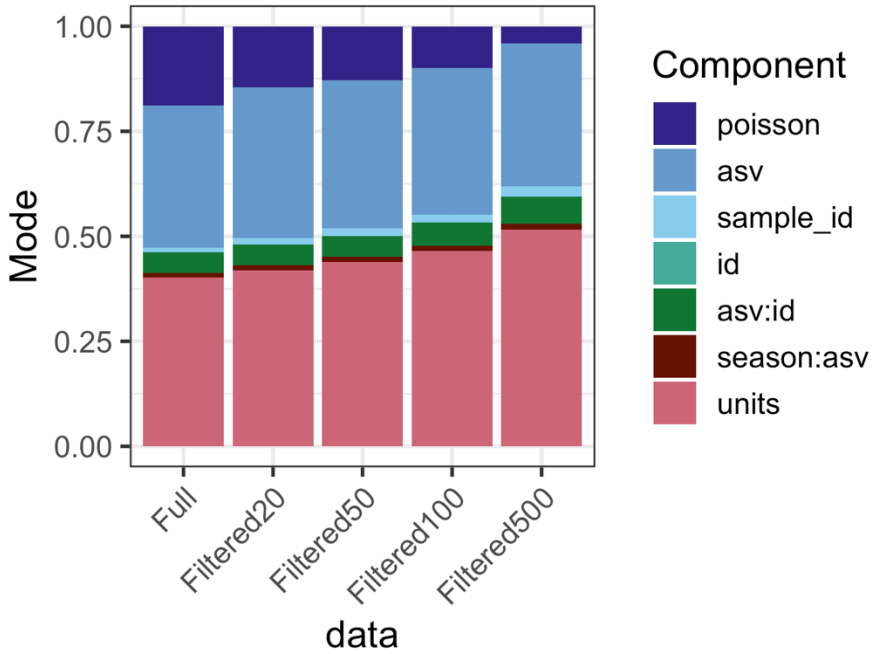


Figure S2. Proportion variance of GLMM component terms across multiple levels of initial abundance filtering.

2.1 Centered log-ratio models

GLMMs using CLR normalised values as the response variable showed agreement with both traditional approaches and raw read count models (Figure S3), but were computationally more efficient allowing for the direct estimation from model output of the specific taxa driving observed effects. Proportion of variance attributable to differential community composition across age classes ($u_{asv:a}$, 11.0%) was substantial but more conservative compared to Poisson models, while proportional variance for season effects ($u_{asv:t}$, 0.8%) was comparable to Poisson models. As in Poisson models, taxonomic variation in ASV abundance explained a high proportion of variance (u_{asv}; 2013: 28.0%, 2016: 44.0%); differential taxonomic abundances across individuals ($u_{asv:h}$) explained 4.0% amount of variance in 2016 (Figure S3). Finally, technical variation estimated as measurement error in CLR models explained a moderate proportion of variance in both years (mev; 2013: 9.9%, 2016: 8.3%).

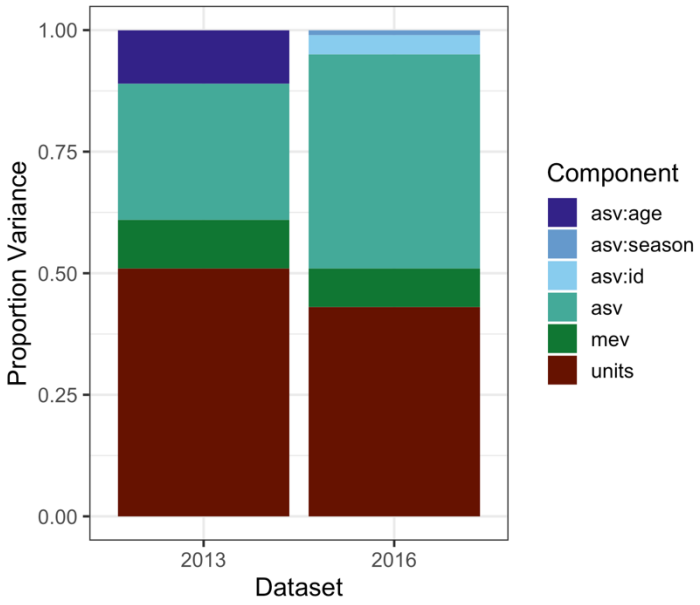


Figure S3. Proportion variance of GLMM component terms for gaussian models using ratio-transformations.

Table S1. MCMCglmm output for Poisson GLMMs. Blank cells indicate NAs.

dataset	variable	post.mean	conf.low	conf.high	eff.samp	pMCMC	effect	propVariance	lHPD	uHPD
2013	Intercept	-5.458	-5.819	-5.120	1000.000	0.001	fixed			
	ageAdult	1.126	0.652	1.637	839.332	0.001	fixed			
	sample	0.725	0.443	1.014	1000.000		random	0.0159	0.0110	0.0244
	asv	7.336	6.479	8.088	1000.000		random	0.1735	0.1552	0.1885
	age:asv	8.551	7.807	9.266	860.142		random	0.1988	0.1847	0.2139
	units (id:asv)	21.132	20.729	21.619	981.480		residual	0.4941	0.4845	0.5086
2016	Intercept	-2.651	-2.918	-2.359	1000.000	0.001	fixed			
	seasonSummer	0.066	-0.258	0.392	1048.847	0.684	fixed			
	sample	0.588	0.408	0.807	1000.000		random	0.0200	0.0146	0.0287
	asv	9.547	8.969	10.142	1000.000		random	0.3474	0.3317	0.3611
	id:asv	1.491	1.285	1.670	1027.140		random	0.0549	0.0468	0.0611
	season:asv	0.332	0.274	0.396	1000.000		random	0.0124	0.0099	0.0144
	units (id:asv:season)	12.792	12.530	13.054	1000.000		residual	0.4668	0.4525	0.4779

Table S2. MCMCglmm output for CLR Gaussian GLMMs. Blank cells indicate NAs.

dataset	variable	post.mean	conf.low	conf.high	eff.samp	pMCMC	effect	propVariance	lHPD	uHPD
2013	Intercept	0.889	0.802	0.984	1000.000	0.001	fixed			
	age:asv	2.136	1.977	2.281	942.252		random	0.147	0.139	0.160
	asv	3.286	2.984	3.575	782.964		random	0.224	0.213	0.246
	mev	1.000	1.000	1.000	0.000		random	0.069	0.068	0.071
	units (id:asv)	7.935	7.848	8.031	1000.000		residual	0.554	0.540	0.564
2016	Intercept	1.025	0.916	1.113	1000.000	0.001	fixed			
	asv:id	0.731	0.630	0.809	1000.000		random	0.0517	0.0444	0.0569
	season:asv	0.150	0.122	0.177	809.734		random	0.0108	0.0084	0.0124
	asv	5.786	5.457	6.112	1000.000		random	0.4105	0.3939	0.4220
	mev	1.000	1.000	1.000	0.000		random	0.0701	0.0688	0.0721
	units (id:asv:season)	6.531	6.429	6.641	1000.000		residual	0.4609	0.4475	0.4733

Table S3. MCMCglmm output for Poisson GLMM with nested taxonomic structure for 2013 age effects. Blank cells indicate NAs.

dataset	variable	post.mean	conf.low	conf.high	eff.samp	pMCMC	effect	propVariance	lHPD	uHPD
2013	Intercept	-6.659	-7.645	-5.744	1000.000	0.001	fixed			
	ageAdult	2.590	1.307	3.637	1000.000	0.001	fixed			
	sample	0.960	0.549	1.445	1000.000		random	0.017	0.012	0.031
	asv	7.748	6.839	8.486	1000.000		random	0.158	0.144	0.178
	age:asv	6.427	5.786	7.038	1000.000		random	0.133	0.121	0.148
	phylum	0.289	0.000	1.099	844.192		random	0.000	0.000	0.023
	sample:phylum	0.967	0.692	1.217	1000.000		random	0.018	0.015	0.025
	age:phylum	1.190	0.168	2.504	970.971		random	0.019	0.004	0.051
	family	0.587	0.000	1.920	608.939		random	0.000	0.000	0.040
	sample:family	2.221	1.892	2.570	1000.000		random	0.048	0.040	0.055
	age:family	2.597	1.185	4.243	733.308		random	0.054	0.024	0.084
	units (id:asv)	19.419	19.014	19.797	863.139		residual	0.408	0.384	0.429

Table S4. Representation across phyla of significant ASV-level compositional shifts. Blank cells indicate that the phyla had no ASVs with significant abundance shifts.

	Age: Adults v Lambs				Season: Summer v Spring			
	Positive Shifts		Negative shifts		Positive Shifts		Negative shifts	
phylum	n ASV	% total ASVs	n ASV	% total ASVs	n ASV	% total ASVs	n ASV	% total ASVs
Bacteroidetes	181	52.16	62	18.45	3	37.50	9	56.25
Firmicutes	89	25.65	242	72.02	4	50.00	1	6.25
Kiritimatiellaeota	14	4.03	1	0.30			1	6.25
Cyanobacteria	13	3.75	6	1.79			2	12.50
Spirochaetes	13	3.75	7	2.08			2	12.50
Lentisphaerae	12	3.46						
Proteobacteria	9	2.59	3	0.89	1	12.50		
Tenericutes	6	1.73	12	3.57				
Verrucomicrobia	5	1.44						
Actinobacteria	1	0.29	1	0.30				
Elusimicrobia	1	0.29						
Euryarchaeota	1	0.29						
Fibrobacteres	1	0.29	1	0.30			1	6.25
Synergistetes	1	0.29						
Planctomycetes			1	0.30				

Supplementary Material References

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**, 581–583.

Fernandes, A.D., Macklaim, J.M., Linn, T.G., Reid, G. & Gloor, G.B. (2013). ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq (J. Parkinson, Ed.). *PLoS One*, **8**, e67019.