PREDICTING COMPLEX TRAITS USING MICROBIOME INFORMATION: A COMPARISON OF METAGENOME DISTANCE MATRICES

Saborío-Montero, A.[1,2], Bach, A.[3], González-Recio, O.[4,5]
[1]Estudiante de Master en Mejora Genética Animal y Biotecnología de la Reproducción y [2]Doctorando en Ciencia y Tecnología de la Producción Animal, UPV, Camino de Vera, s/n 46022 Valencia, España. [3]Department of Ruminant Production, ICREA-IRTA, 08140 Caldes de Montbui, Spain.[4]Departamento de Mejora Genética Animal. INIA, Madrid 28040, España. [5]UPM, Ciudad Universitaria s/n, 28040 Madrid, España.
E-mail: alejandro.saboriomontero@ucr.ac.cr

INTRODUCTION

Recent studies have shown evidences revealing that the microbiome influences relevant complex traits for dairy livestock systems, such as feed efficiency or methane emissions. Besides, the genetic background of the animal partially controls the microbiota composition. The joint analysis of the genetic background of the host and its microbiota demands accounting for the distance (or dissimilarity) between communities of microorganisms between hosts. Several methods have been developed to ordinate these matrices; some of these ordination technics yield similar matrices while others yield considerable different ones, causing inconsistent conclusions. Similarities between matrices ordinated with different methods are related to particularities of the ordination methods and also to distance (or dissimilarities) metrics used for its ordination (i.e. Euclidean, Bray-Curtis, χ^2), which differ in the usage of alpha and beta diversity in its calculation. Alpha diversity refers to the number of taxa within a single microbial ecosystem, or operational taxonomic units (**OTUs**) within a sample, while beta diversity denotes differences in taxonomic abundance profiles from different ecosystems, or relative abundance of OTUs between samples. Consensus on what method is the most appropriate hasn't been reached yet, and might depend on data singularities and the purpose of the study. The aim of this study was to compare ordination methods for rumen microbiota distance (or dissimilarity) matrices, in order to estimate variance components for prediction of complex traits including the microbiota in its estimation.

MATERIALS AND METHODS

Data: Two data subsets were analysed either of simulated or real data. Simulated data were generated from observations in the real data. A data frame of 1000 genotyped Holstein animals with allelic variants for 9244 SNPs was used; the data for the relative abundance of 83 microbial OTUs was constructed using the (co)variances matrix from microbiota observation in the real data set. Phenotypes were simulated assigning random effects to SNPs and the OTUs, assuming a heritability and a microbiability of 0.30 and 0.50, respectively. A total of 100 replicates were simulated. Real data included phenotypic performance data, genotypic information (54609 SNPs) and relative abundances of 92 OTUs from rumen content of 70 Holstein cows from a single herd.

Analysis: Seven ordination methods were independently used to build the microbiota distance (or dissimilarity) matrices between cows. The seven methods were: the one reported in Ross et al. (2013) from now on identified as "Ross", multidimensional scaling (**MDS**), principal coordinates analysis (**PCoA**), detrended correspondence analysis (**DCA**), non-metric multidimensional scaling (**NMDS**), redundancy analysis (**RDA**) and constrained correspondence analysis (**CCA**). Mixed models were used following a Bayesian framework, using the following models in linear notation:

$$\mathbf{y} = \mathbf{1}'\boldsymbol{\mu} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{e} \quad [1]$$

Where: **y** = feed efficiency, μ = population mean, **1** = vector of ones of *n*x*1* dimensions, **u** = genetic background, **m** being the microbiota effect, **Z** and **W** the corresponding incidence matrices for the genetic and the microbiota effects, respectively, and **e** = residual error, with **u** ~**N** (0, $\mathbf{G}\sigma_u^2$), **m** ~**N** (0, $\mathbf{K}\sigma_m^2$) and **e** ~**N** (0,σ_e^2), where **G** is the genomic relationship matrix and **K** the microbiota distance (or dissimilarity) matrix between cows.

Additionally, another model accounting for the interaction between the genetic and the microbiota effects was tested:

$$\mathbf{y} = \mathbf{1}'\boldsymbol{\mu} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{Tu}x\mathbf{m} + \mathbf{e} \quad [2]$$

Where: **y**, $\boldsymbol{\mu}$, **Zu**, **Wm** and **e** are the same as in model 1 and **u***x***m** stands for the interaction between genetic background of the host and her microbiome, **T** represent the corresponding incidence matrix.

Models were solved in a Bayesian framework using the BGLR package in R (De Los Campos and Perez Rodriguez, 2016). The means and standard error of the replicates for the parameters of interest were obtained. Real data was analysed using the same models.

RESULTS AND DISCUSSION

Simulation. Heritability (h^2) estimates (and standard error of the means between brackets) from model 1 were: 0.312(±0.006), 0.311(±0.006), 0.312(±0.006), 0.256(±0.006), 0.225(±0.006), 0.315(±0.006) and 0.313(±0.005) according to ordination method of Ross, MDS, PCoA, DCA, NMDS, RDA and CCA, respectively (Figure 1A). Likewise the predictions for microbiability (m^2) were: 0.478(±0.006), 0.496(±0.003), 0.494(±0.003), 0.361(±0.007), 0.281(±0.009), 0.503(±0.003) and 0.500(±0.003) in the same order (Figure 1B). Correlations between estimated breeding values (**EBV**) and true breeding values (**TBV**) were similar for all ordination methods and were: 0.633(±0.003), 0.592(±0.004), 0.591(±0.004), 0.598(±0.004), 0.557(±0.004), 0.624(±0.003) and 0.631(±0.003) for ordination procedures of Ross, MDS, PCoA, DCA, NMDS, RDA and CCA, respectively. Correlations between estimated microbiome values (**EMV**) and true microbiome values (**TMV**) were: 0.975(±0.001), 0.844(±0.001), 0.845(±0.001), 0.807(±0.011), 0.517(±0.019), 0.949(±0.001) and 0.966(±0.001) in the same order. MDS and PCoA methods yielded exactly the same matrix; differences in results from these methods are due to iterative processes. Some of these methods were more accurate and precise for prediction of variance components and genetic parameters than previously proposed methods for ordination of microbiome distance matrices. DCA and NMDS distanced the most from simulated parameters. Only these two methods use Bray-Curtis dissimilarity in the ordination process. Estimates for h^2 and m^2 from model 2 were similar than those obtained from model 1, likewise correlations between EBV and TBV and between EMV and TMV from model 2 were also close to those obtained from model 1.The variance for simulated interaction effect (genotype *x* microbiome = 341.5) were underestimated by all methods and were: 196.5(±3.2), 251.5(±4.0), 256.7(±3.9), 185.0(±5.4), 208.5(±10.0), 243.0(±4.8) and 232.7(±4.5) for procedure of Ross, MDS, PCoA, DCA, NMDS, RDA and CCA, respectively. Correlations between the estimated interaction effect and the simulated value were generally low: 0.100(±0.007), 0.071(±0.021), 0.070(±0.021), 0.138(±0.017), 0.149(±0.023), 0.075(±0.009) and 0.081(±0.008) in the same order. Results show that h^2 and m^2 estimation were appropriate

only for some matrices (Ross, MDS/PCoA, RDA and CCA) and underestimated by others (NMDS and DCA). Model 2 was inefficient estimating the interaction effect.

Real data. Results from both models using real data are shown in Table 1. *h2* estimates ranged from 0.077 (Ross and MDS) to 0.083 (NMDS) for model 1 and from 0.059 (DCA) to 0.078 (RDA) for model 2, m^2 estimation ranged from 0.073 (MDS) to 0.103 (NMDS) for model 1 and from 0.056 (RDA) to 0.096 (NMDS) for model 2. Correlations between posterior means for genotype and phenotype ranged from 0.857 (DCA) to 0.912 (NMDS) for model 1 and from 0.799 (CCA) to 0.889 (Ross and MDS) for model 2, while between posterior means for microbiome and phenotype ranged from 0.210 (NMDS) to 0.910 (RDA) for model 1 and from 0.211 (NMDS) to 0.906 (CCA) for model 2.

Overall considerations. From simulation analysis it can be inferred that ordination methods of MDS/PCoA, RDA and CCA are as suitable or even better than previously reported by Ross et al. (2013) for prediction of microbiability, contrasting with DCA and NMDS which performed poorly as predictive methods for microbiability. From real data analysis, low heritability for feed efficiency and microbiability were obtained and mid to high correlations between genetic background of hosts and phenotype or microbiome and phenotype were obtained with methods that performed better, indicating that there might be a relationship linking genotype-microbiome-phenotype which could be used in prediction of complex traits; however, the number of samples (n=70) is still small and also the fact that all cows come from a single herd is a limitation to obtain accurate estimations of genetic parameters; for more robust results, the sample size needs to be incremented and more herds are to be included in further analysis.

REFERENCES

Ross, E.M., P.J. Moate, L.C. Marett, B.G. Cocks, and B.J. Hayes. 2013. Metagenomic Predictions: From Microbiome to Complex Health and Environmental Phenotypes in Humans and Cattle. PLoS One 8:e73056. doi:10.1371/journal.pone.0073056.

De Los Campos, G., and P. Perez Rodriguez. Bayesian Generalized Linear Regresion. R Documentation, cran.r-project.org/web/packages/BGLR/BGLR.pdf

Awknowlegments

We are thankful to Universidad de Costa Rica for financing the máster for the first author and also to Blanca de los Pirineos for providing samples and data.

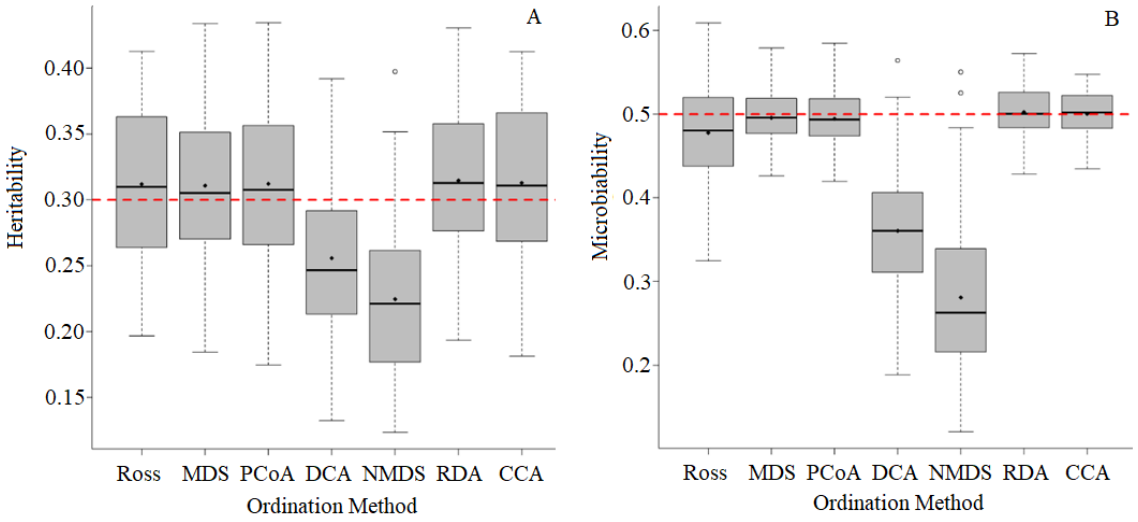

Figure 1. *Heritability (A) and microbiability (B) according to ordination method of Ross et al. 2013, multidimensional scaling (MDS), principal coordinates analysis*

(PCoA), detrended correspondence analysis (DCA), non-metric multidimensional scaling (NMDS), redundancy analysis (RDA) and constrained correspondence analysis (CCA).

Table 1. *Heritability, microbiability and correlations between genetic background and phenotype; and between microbiota and phenotype, estimated using models 1 and 2 according to method of ordination for microbiota from real data.*

	Ross	MDS	PCoA	DCA	NMDS	RDA	CCA
Model 1							
Heritability	0.077	0.077	0.082	0.082	0.083	0.083	0.081
Microbiability	0.075	0.073	0.074	0.102	0.103	0.076	0.077
Correlation G vs Phenotype	0.865	0.862	0.879	0.857	0.912	0.892	0.879
Correlation K vs Phenotype	0.483	0.666	0.669	0.360	0.210	0.910	0.906
Model 2							
Heritability	0.065	0.069	0.074	0.059	0.065	0.078	0.063
Microbiability	0.072	0.057	0.061	0.086	0.096	0.056	0.059
Correlation G vs Phenotype	0.889	0.889	0.866	0.839	0.802	0.861	0.799
Correlation K vs Phenotype	0.506	0.639	0.548	0.360	0.211	0.899	0.906

PREDICTING COMPLEX TRAITS USING MICROBIOME INFORMATION: A COMPARISON OF METAGENOME DISTANCE MATRICES

ABSTRACT: The aim of this study was to compare ordination methods for microbiota distance matrices, in order to estimate variance components for complex traits prediction including the microbiome in its estimation. Seven ordination methods for building distance (or dissimilarity) matrices were tested; real (n=70) and simulated (n=1000) data were analysed to estimate variance components including phenotypes, genotypes and microbiome information. The seven methods were: the one reported in Ross et al. (2013), multidimensional scaling (MDS), principal coordinates analysis (PCoA), detrended correspondence analysis (DCA), non-metric multidimensional scaling (NMDS), redundancy analysis (RDA) and constrained correspondence analysis (CCA). MDS and PCoA methods yielded exactly the same matrix. From simulation analysis it can be inferred that ordination methods of MDS/PCoA, RDA and CCA are as suitable as or even better than previously reported by Ross et al. (2013) for prediction of microbiability, contrasting with DCA and NMDS which performed poorly as predictive methods for microbiability. From real data analysis, low heritability for feed efficiency and microbiability were obtained and mid to high correlations between the genetic background of the hosts and the phenotypes or microbiota and phenotypes were obtained with methods that performed better in the simulation, indicating that it might be a relationship linking genotype-microbiome-phenotype which could be used in prediction of complex traits.

Keyword: feed efficiency, microbiability, heritability, ordination methods