ORDINATION METHODS TO BUILD MICROBIOTA SIMILARITY MATRICES FOR COMPLEX TRAITS PREDICTION

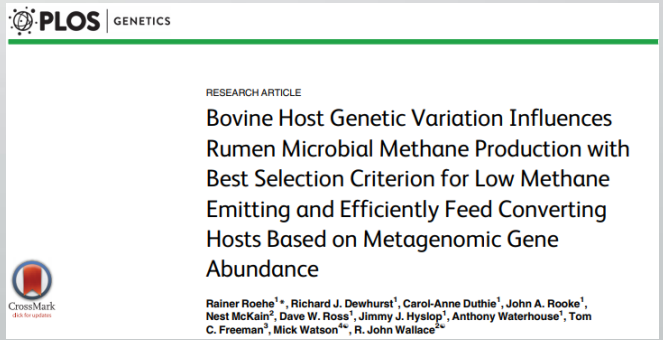
Alejandro Saborío-Montero
Oscar González-Recio

EAAP 2018

There are evidences of:

- Microbiome influencing complex traits in ruminants

- Partial control of host genotype over microbiome composition



- These evidences pose the hypothesis of microbiome as a source of information to predict complex traits

- Complex traits as feed efficiency and methane emissions could be included into genetic evaluations taking into account a microbiome effect to select highly profitable and sustainable animals

OPEN ACCESS Freely available online PLOS ONE

Metagenomic Predictions: From Microbiome to Complex Health and Environmental Phenotypes in Humans and Cattle

Elizabeth M. Ross[1,2,3*], Peter J. Moate[4], Leah C. Marett[4], Ben G. Cocks[1,2,3], Ben J. Hayes[1,2,3]

Animal (2015), **9:9**, pp 1431–1440 © The Animal Consortium 2015. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.
doi:10.1017/S1751731115000968

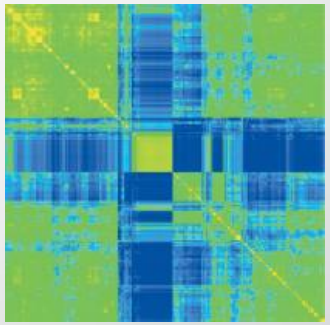
animal

Animal board invited review: genetic possibilities to reduce enteric methane emissions from ruminants

N. K. Pickering[1a], V. H. Oddy[2], J. Basarab[3], K. Cammack[4], B. Hayes[5,6,7], R. S. Hegarty[8], J. Lassen[9], J. C. McEwan[1], S. Miller[10,11b], C. S. Pinares-Patiño[12c] and Y. de Haas[13†]

- Single given OTU vs whole microbiota in statistical analysis.

- Microbiota relationship matrices (MRM) consider microbiota as a whole; However there are differences in the methods of ordination of matrices leading to differences in MRM outputs.

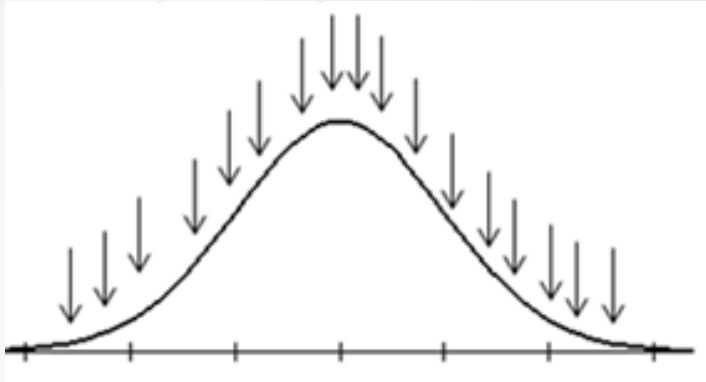


- The comparison of MRM has not been evaluated previously in an animal breeding context.

- We propose to compare the MRM obtained from different ordination methods to disentangle which are the most appropriate to be included in statistical models analyzing genotype and microbiome to predict feed efficiency.

- Develop and test some statistical approaches to evaluate feed efficiency in dairy cows including host genotype and microbiome simultaneously.
- Compare ordination methods to build MRM using simulation
- Estimate the proportion of phenotypic variance for feed efficiency explained by microbiome variance (microbiability), considering the interaction between microbiota and host genotype

- Simulated data
  - 1000 Holstein cows and 92 OTUs
  - Microbiota effect (co(variances) matrix from real data)
  - Genetic effect (assigning effects to simulated QTL)
  - Phenotype



- Real data
  - 70 Holstein cows
  - Microbiota
    - V3-V4 hypervariable regions 16S rRNA (92 OTU)
  - Genotype
    - 54609 SNPs
  - Phenotype
    - Parity, DMI, milk yield, fat, protein body weight, feed efficiency.

Genomic + Microbiota

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

- Residual
- Microbiota effect
- Incidence matrix (microbiota)
- Additive genetic effect
- Incidence matrix (genotype)
- Population mean
- An nx1 vector of ones
- Feed efficiency

- u ~N (0, GRMσ_u^2),
- m ~N (0, **MRM**σ_m^2) y
- e ~N (0,σ_e^2)

Bayesian resolution approach

BGLR package in R environment

Effects were included as RKHS

Genomic + Microbiota + Genomic*Microbiota

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

- Residual
- genetic-microbiota interaction
- Incidence matrix (interaction)
- Microbiota effect
- Incidence matrix (microbiota)
- Additive genetic effect
- Incidence matrix (genotype)
- Population mean
- An nx1 vector of ones
- Feed efficiency

- u ~N (0, GRMσ_u^2),
- m ~N (0, **MRM**σ_m^2)
- u x m ~N (0, GRM#**MRM**σ_{uxm}^2)
- e ~N (0,σ_e^2)

Reference MRM

OPEN ACCESS Freely available online

PLOS ONE

Metagenomic Predictions: From Microbiome to Complex Health and Environmental Phenotypes in Humans and Cattle

Elizabeth M. Ross[1,2,3]*, Peter J. Moate[4], Leah C. Marett[4], Ben G. Cocks[1,2,3], Ben J. Hayes[1,2,3]

1 Biosciences Research Division, Department of Environment and Primary Industries, Bundoora, Victoria, Australia, **2** Dairy Futures Cooperative Research Centre, Bundoora, Victoria, Australia, **3** La Trobe University, Bundoora, Victoria, Australia, **4** Future Farming Systems Division, Department of Environment and Primary Industries, Ellinbank, Victoria, Australia

Metagenomic profiles for a group of samples are defined as an $n \times m$ matrix $\mathbf{X}$ with elements x_{ij}, the log transformed and standardised count for sample i for contig j, with n samples and m contigs. The relationship between samples can then be described by a matrix $\mathbf{G} = \mathbf{XX'}/m$.

OTU

Sample

x_{11}	x_{12}	x_{13}	.	.	.	x_{ij}
x_{21}	x_{22}	x_{23}	.	.	.	x_{ij}
x_{31}	x_{32}	x_{33}	.	.	.	x_{ij}
.	.	.	.	.	.	x_{ij}
.	.	.	.	.	.	x_{ij}
.	.	.	.	.	.	x_{ij}
x_{ij}	x_{ij}	x_{ij}	x_{ij}	x_{ij}	x_{ij}	x_{ij}

Each element (x_{ij}) from the X matrix is the log-transformed and standardized count for relative abundance for sample *i* in OTU *j*

Ordination methods used

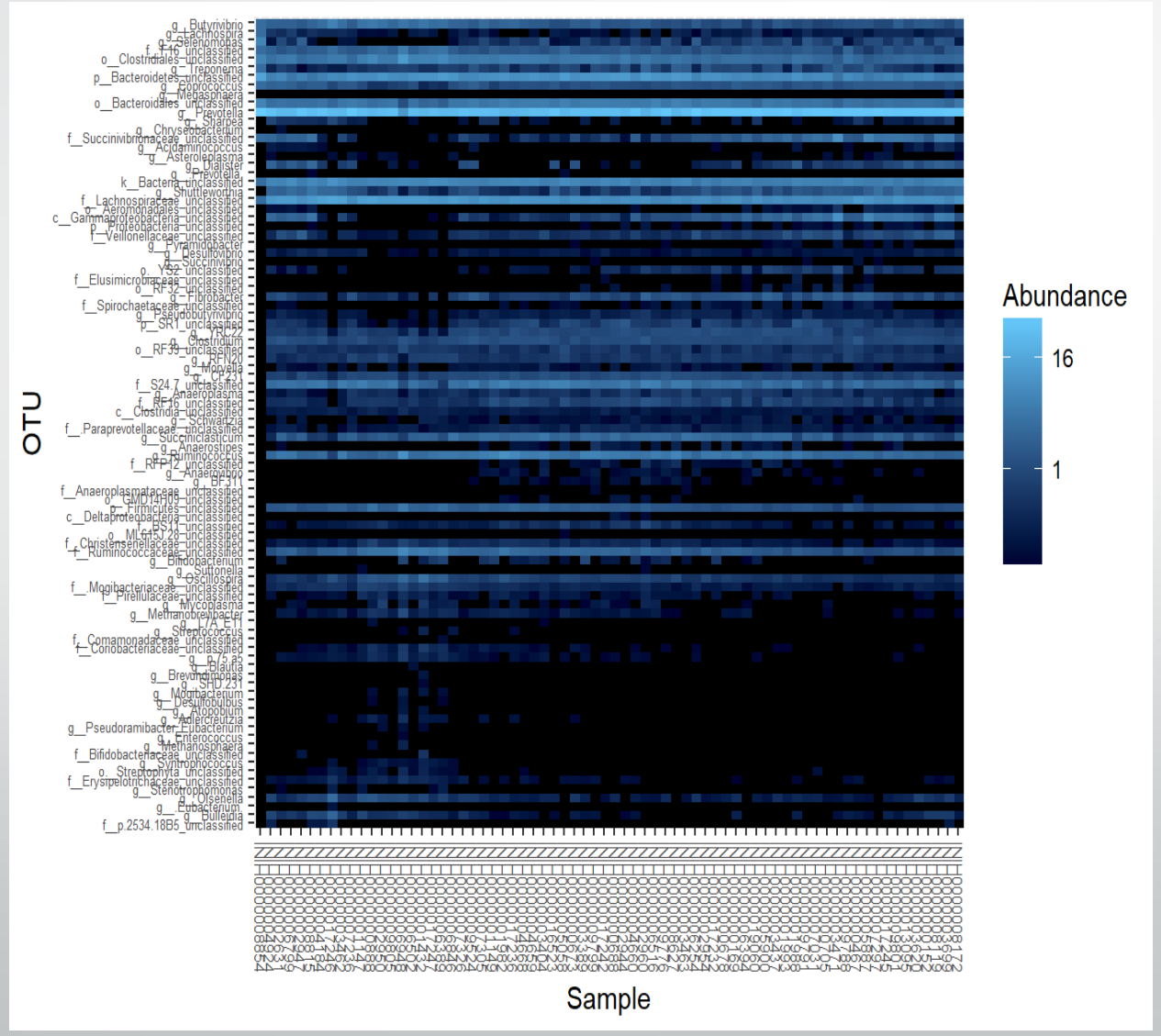
- Metric Multidimensional Scaling (MDS/PCoA)
- Detrended Correspondence Analysis (DCA)
- Non-Metric Multidimensional Scaling (NMDS)
- Redundancy Analysis (RDA)
- Constrained Correspondence Analysis (CCA)

Table 1. Relevant metrics, procedures and miscellaneous characteristics regarding ordination methods[1].

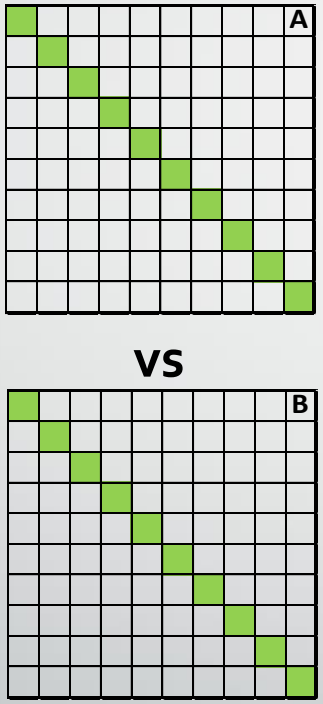
Trait	MDS	DCA	NMDS	RDA	CCA
Distance	Euclidean	Chi Square	Bray-Curtis	Euclidean	Chi Square
Arch effect correction	No	Yes	No	No	No
Canonical analysis	No	No	No	Yes	Yes
Computation time	Low	Low	High	Low	Low

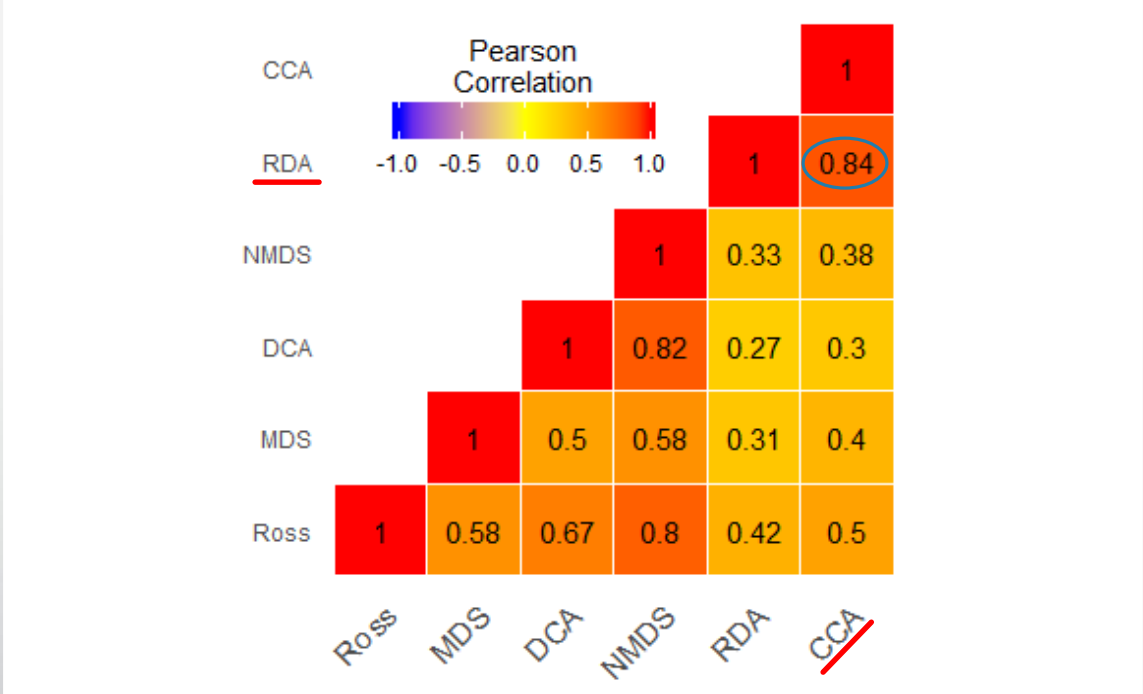
[1]MDS = Multidimensional Scaling, DCA = Detrended Correspondence Analysis, NMDS = Non-Metric Multidimensional Scaling, RDA = Redundancy Analysis, CCA = Constrained Correspondence Analysis.

Relative abundance of OTUs

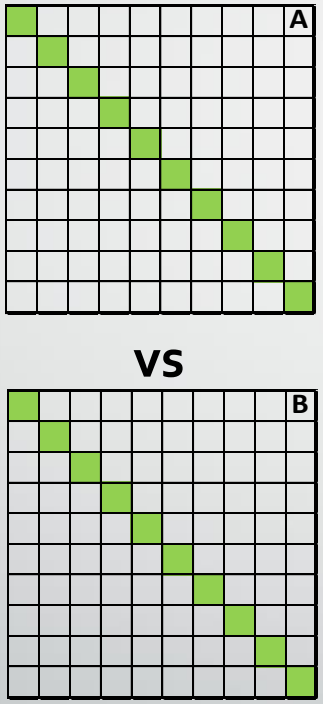


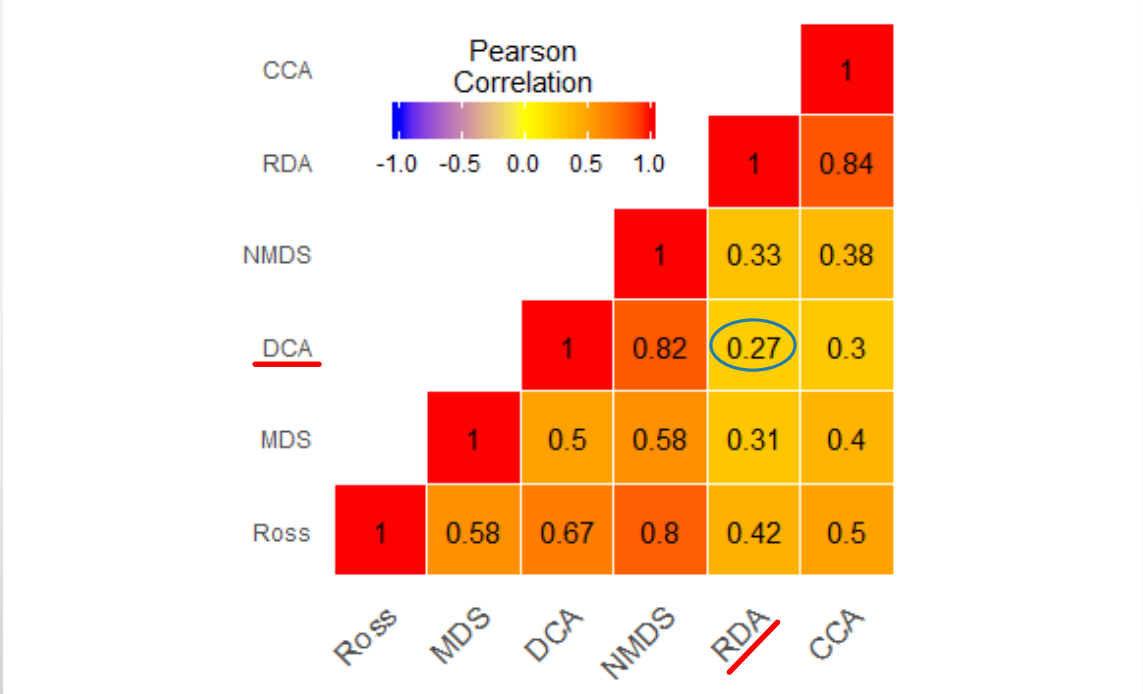
Association between diagonal elements



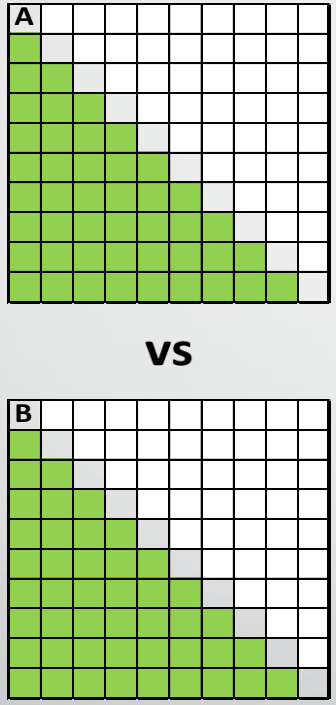


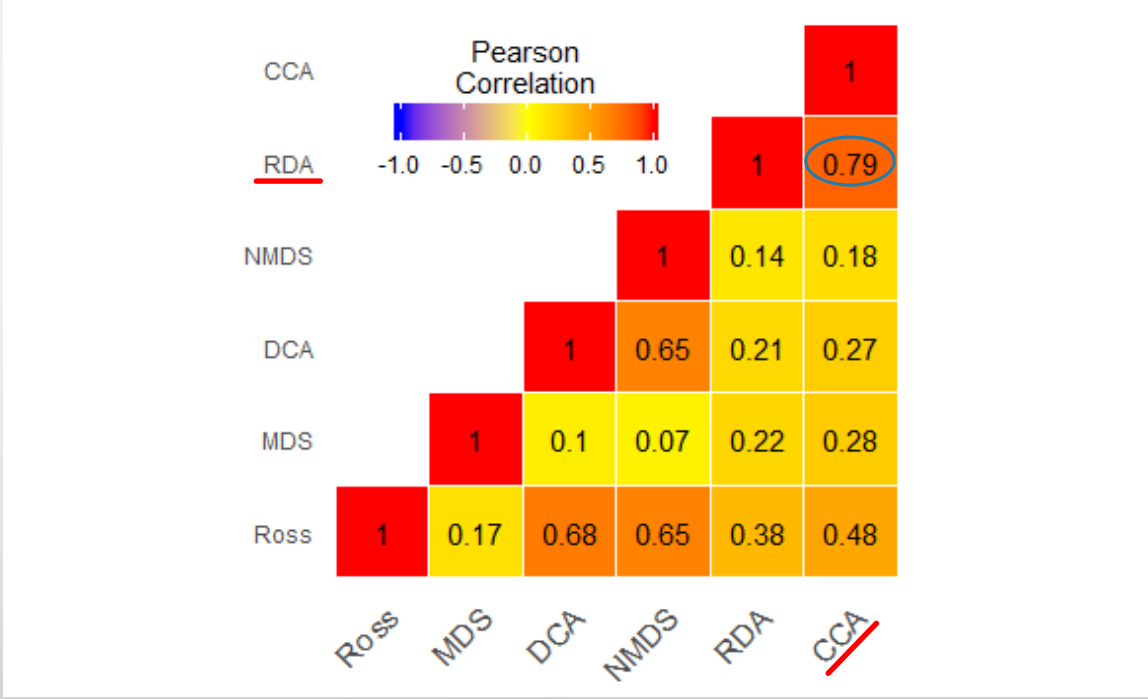
Association between diagonal elements



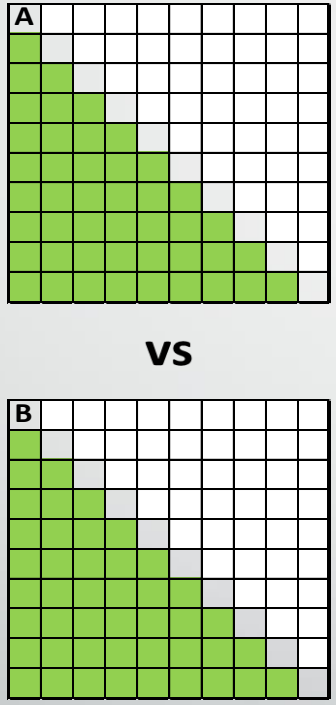


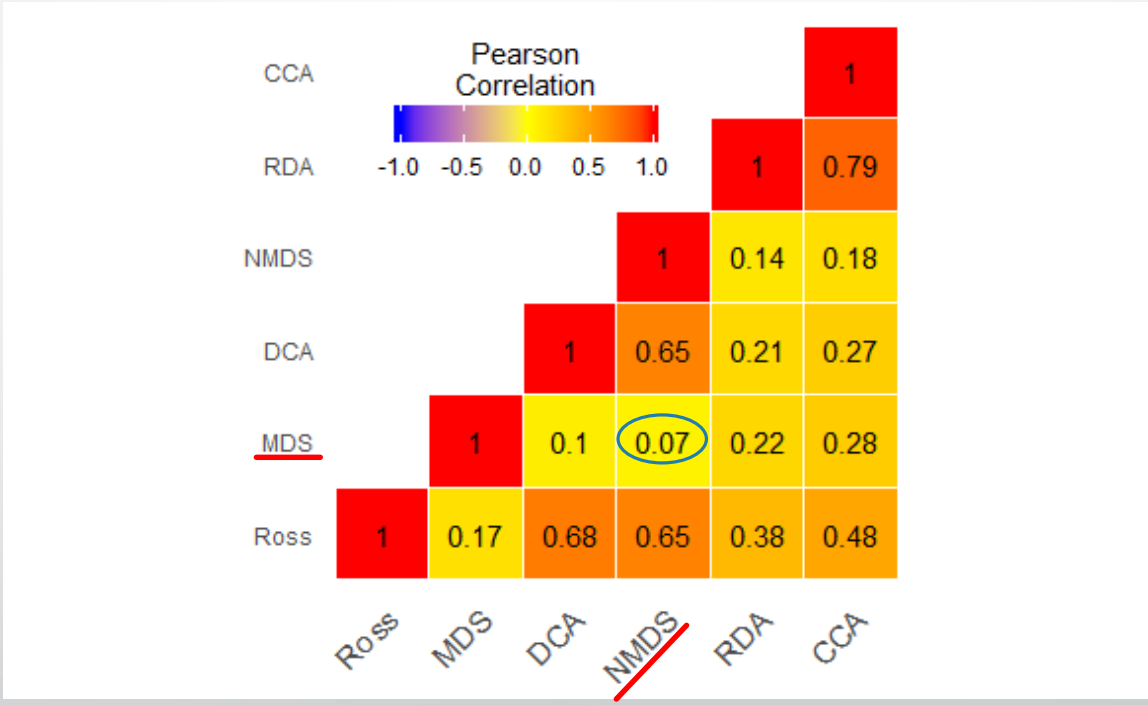
Association between non-diagonal elements





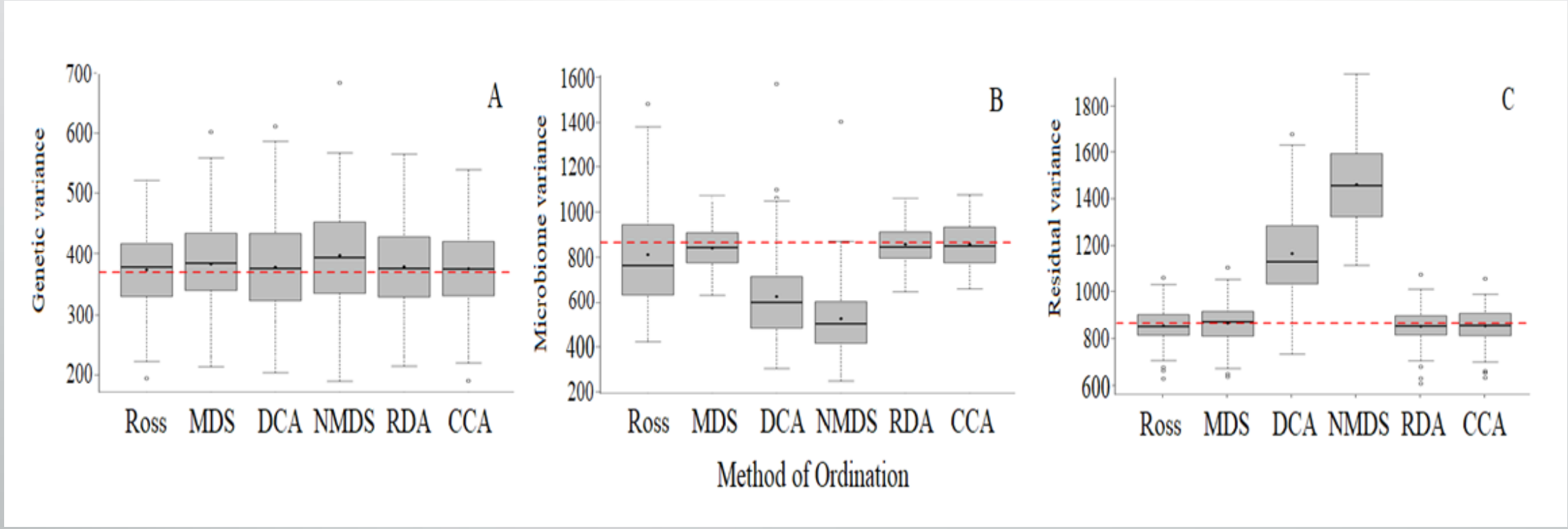
Association between non-diagonal elements





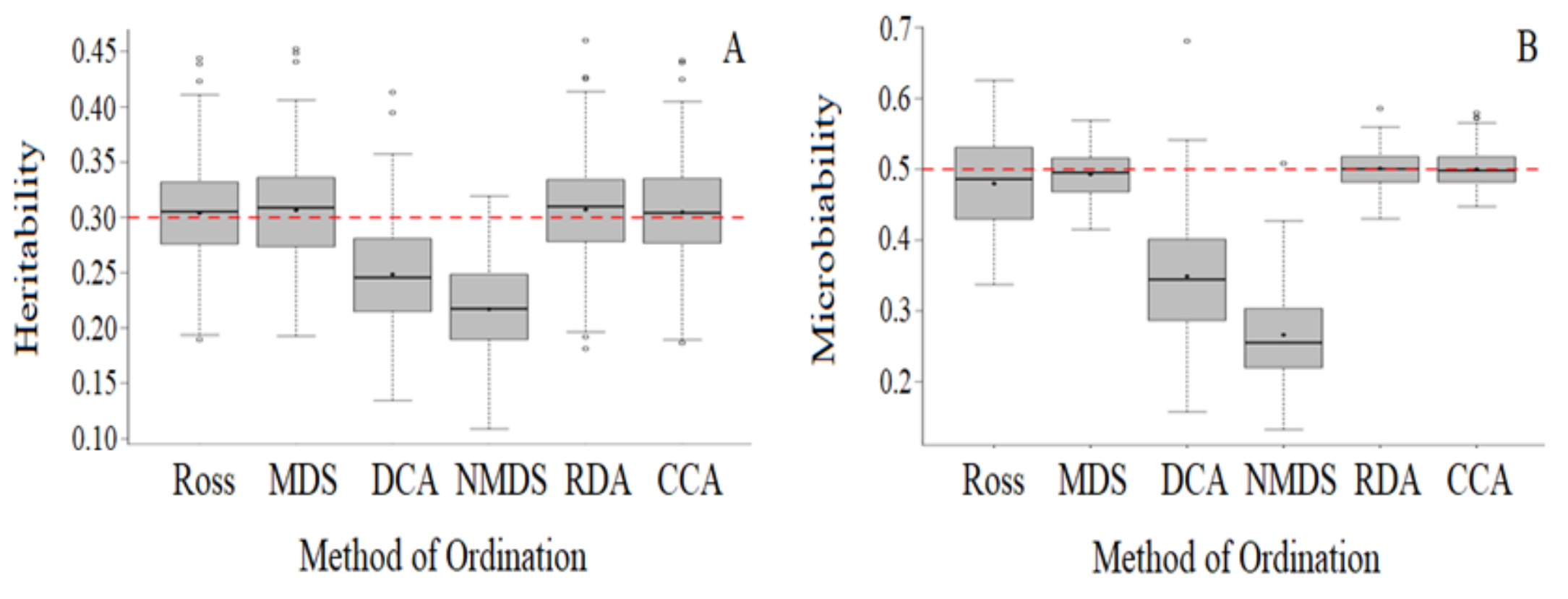
Variance components

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

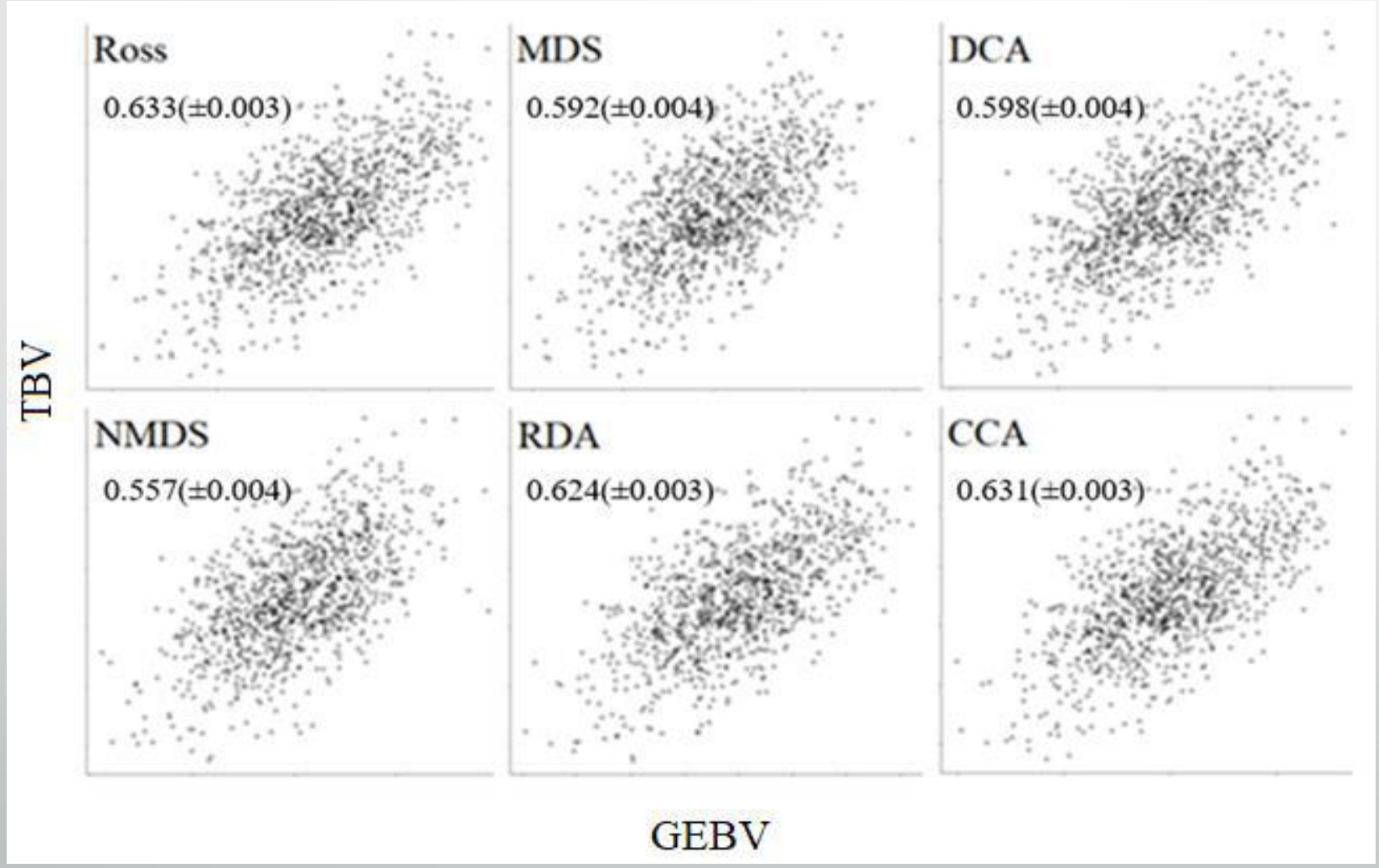


Heritability and Microbiability

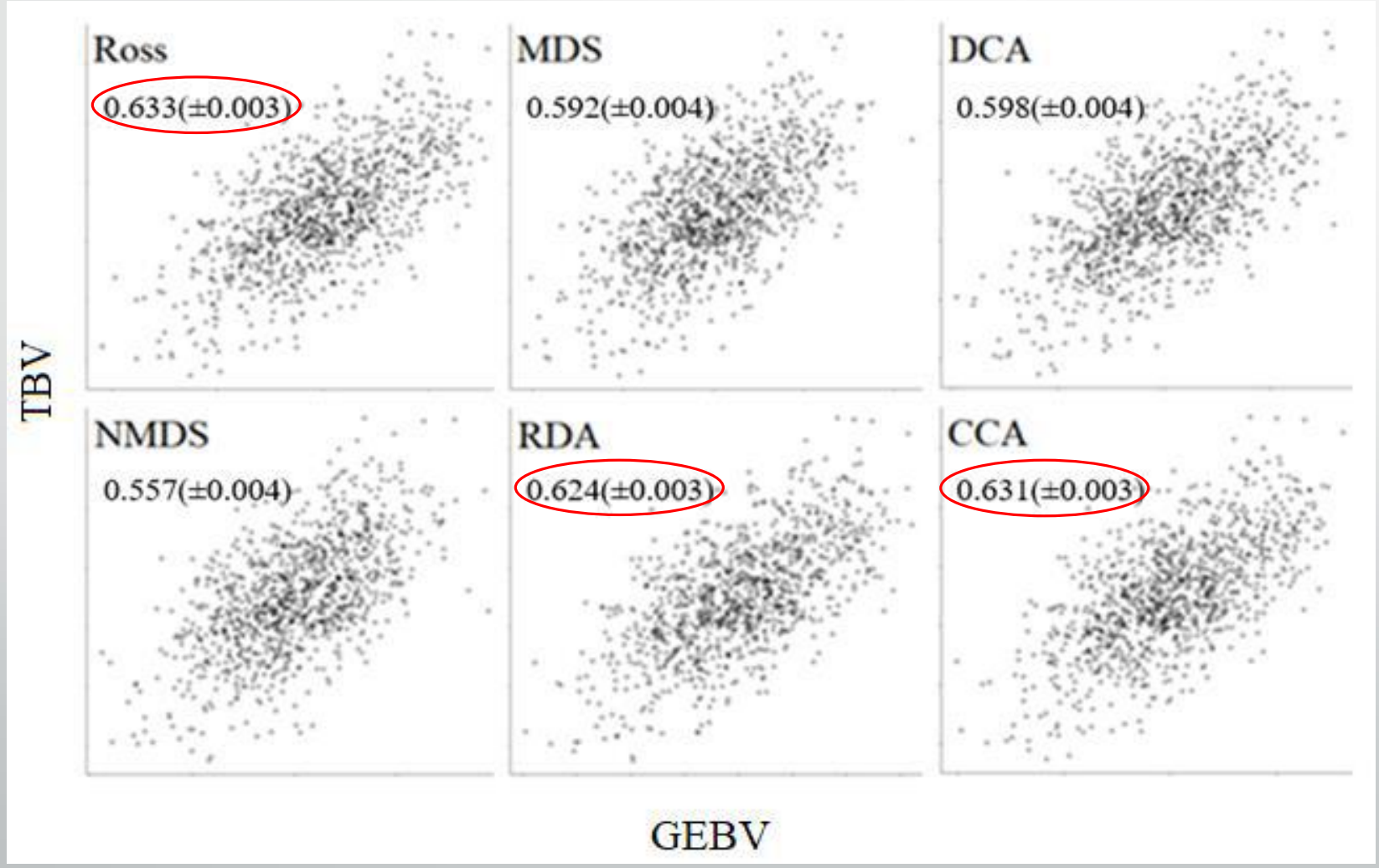
$$\mathbf{y} = 1'\boldsymbol{\mu} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{e}$$



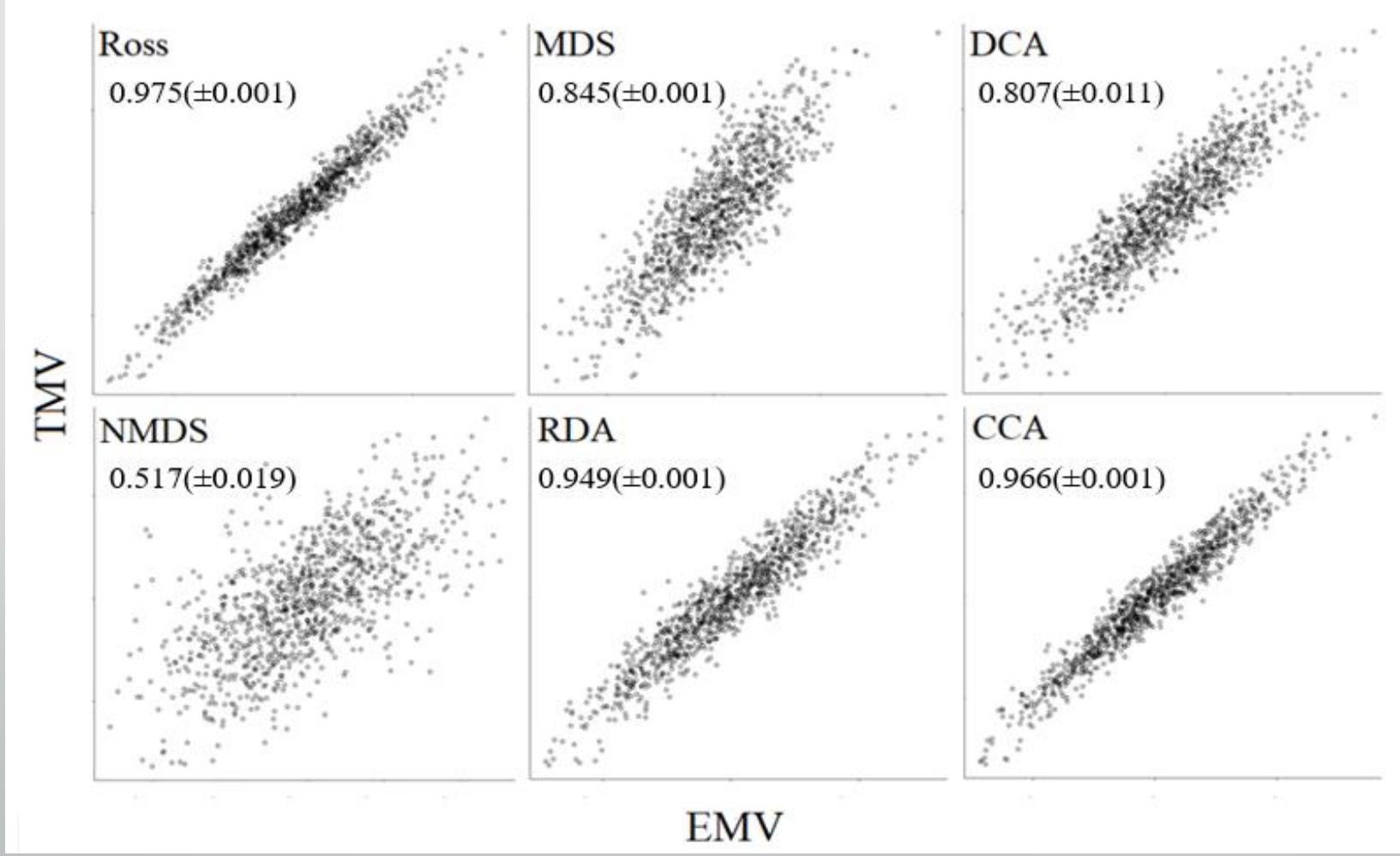
Accuracy (GEBV vs TBV)



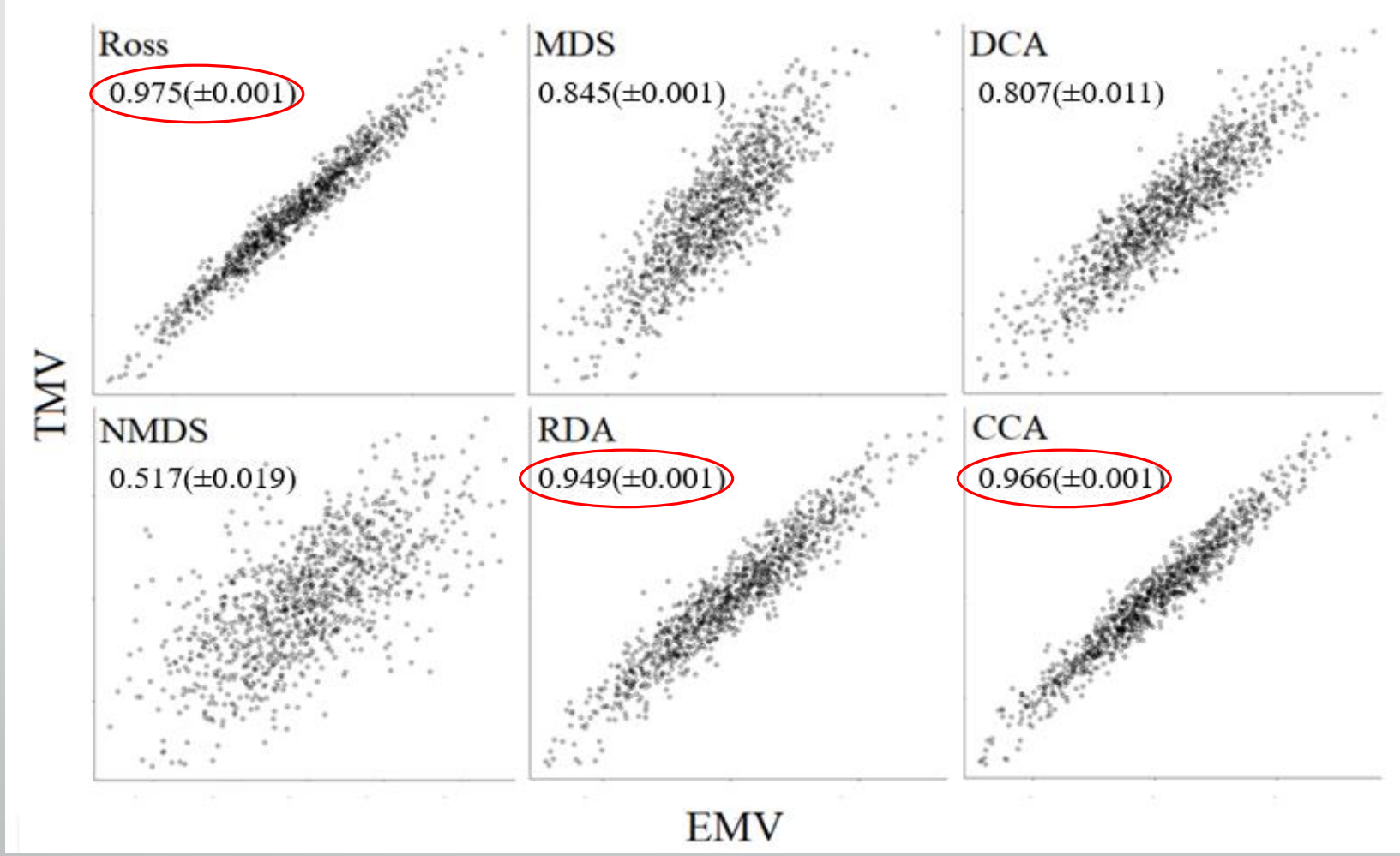
Accuracy (GEBV vs TBV)



Accuracy (EMV vs TMV)

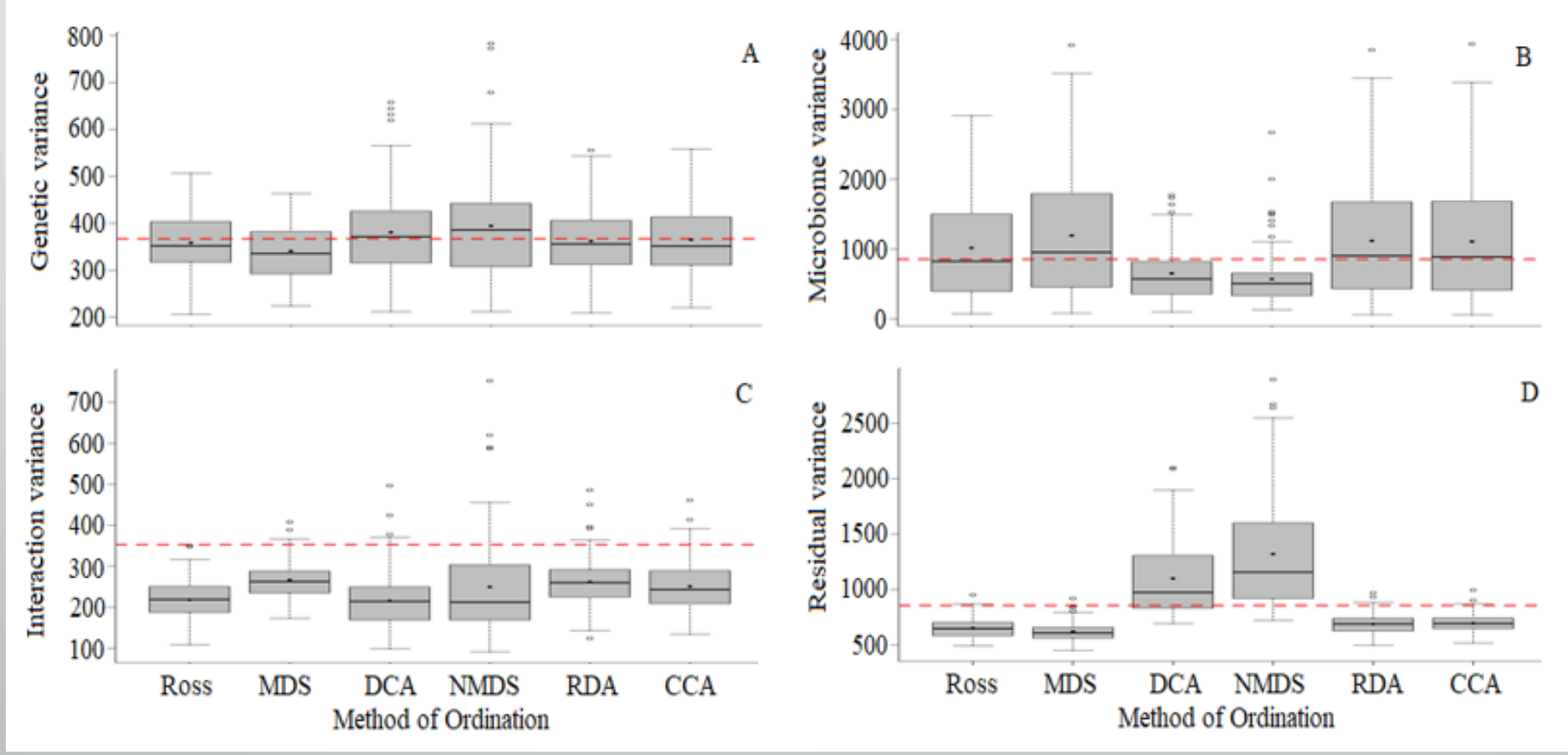


Accuracy (EMV vs TMV)



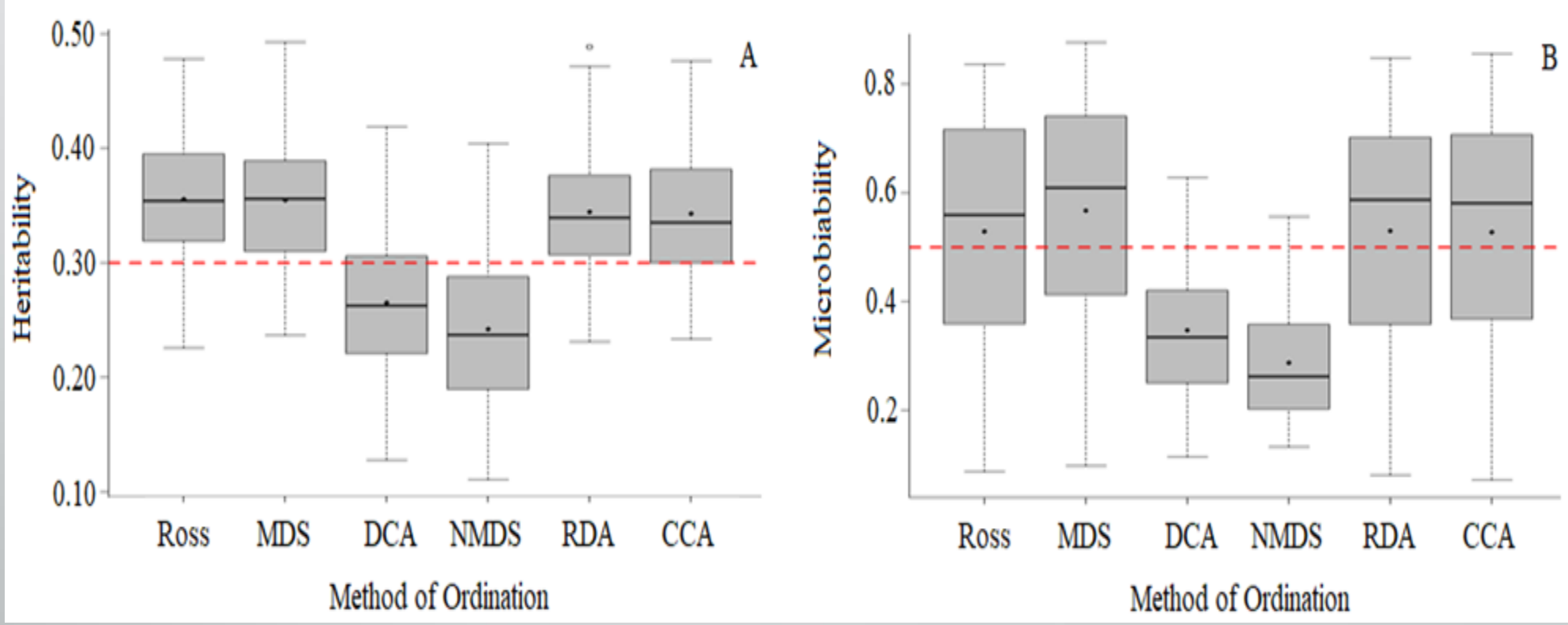
Variance components

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



Heritability and Microbiability

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



Real Data

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

Table 7. Heritability, microbiability and correlations between GEBV and phenotype; and between EMV and phenotype for feed efficiency, estimated using a model independently including genomic and microbiome effects according to method of ordination for microbiota using real data from 70 cows[1].

Parameters	Ross	MDS	DCA	NMDS	RDA	CCA
$\mathbf{y} = 1'\boldsymbol{\mu} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{e}$						
Heritability	0.313	0.304	0.305	0.293	0.333	0.337
Microbiability	0.230	0.198	0.263	0.240	0.346	0.318
Correlation GEBV vs Phenotype	0.942	0.960	0.926	0.955	0.956	0.956
Correlation EMV vs Phenotype	0.538	0.698	0.368	0.218	0.987	0.993

[1]GEBV= Genomic Estimated Breeding Values, EMV= Estimated Microbiome Values, Ross = ordination method of Ross et al. 2013, MDS = Multidimensional Scaling, DCA = Detrended Correspondence Analysis, NMDS = Non-Metric Multidimensional Scaling, RDA = Redundancy Analysis, CCA = Constrained Correspondence Analysis.

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

Table 9. Heritability, microbiability and correlations between GEBV and phenotype; and between EMV and phenotype for feed efficiency, estimated using a model with interaction between genomic and microbiome effects according to method of ordination for microbiota using real data from 70 cows[1].

Parameters	Ross	MDS	DCA	NMDS	RDA	CCA
$\mathbf{y} = 1'\boldsymbol{\mu} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{Tu}x\mathbf{m} + \mathbf{e}$						
Heritability	0.269	0.273	0.266	0.269	0.297	0.281
Microbiability	0.192	0.198	0.220	0.242	0.282	0.272
Correlation GEBV vs Phenotype	0.934	0.952	0.923	0.947	0.945	0.946
Correlation EMV vs Phenotype	0.525	0.700	0.368	0.214	0.989	0.996

[1]GEBV= Genomic estimated breeding values, EMV= Estimated microbiome values, Ross = ordination method of Ross et al. 2013, MDS = Multidimensional scaling, DCA = Detrended correspondence analysis, NMDS = Non-metric multidimensional scaling, RDA = Redundancy analysis, CCA = Constrained correspondence analysis.

Statistical model comparison

Table 10. Information criteria estimated for models with and without interaction between genetic and microbiome effect according to method of ordination for the microbiota relationship matrix for real data from 70 cows, a GBLUP model* is included as reference [1].

	Ross	MDS	DCA	NMDS	RDA	CCA	$y = 1'\mu + Zu + e^*$
$y = 1'\mu + Zu + Wm + e$							
LogLikAtPostMean	-84.2	-84.9	-84.0	-85.4	-84.1	-84.2	-85.1
PostMeanLogLik	-88.8	-89.8	-88.0	-88.9	-89.1	-89.2	-88.7
pD	9.3	9.8	7.9	7.1	10.1	10.1	7.1
DIC	187.0	189.3	183.9	185.0	188.3	188.5	184.4
$y = 1'\mu + Zu + Wm + Tuxm + e$							
LogLikAtPostMean	-84.1	-84.2	-83.9	-85.4	-83.6	-83.7	-85.1
PostMeanLogLik	-89.5	-90.0	-88.8	-89.6	-89.5	-89.7	-88.7
pD	10.9	11.6	9.8	8.3	11.9	11.9	7.1
DIC	190.0	191.7	187.5	187.5	191.0	191.3	184.4

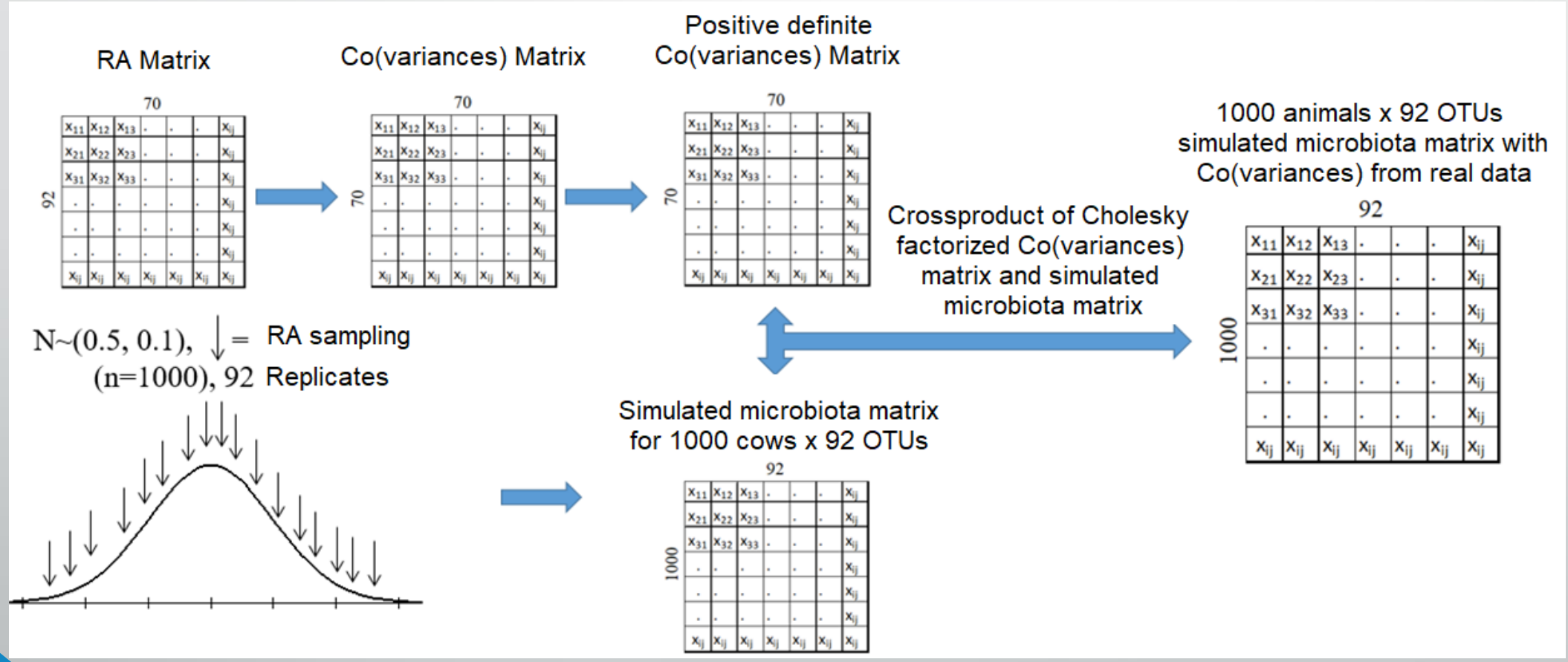
[1]Ross = ordination method of Ross et al. 2013. MDS = Multidimensional scaling. DCA = Detrended correspondence analysis. NMDS = Non-metric multidimensional scaling. RDA = Redundancy analysis. CCA = Constrained correspondence analysis. logLikAtPostMean = log-likelihood evaluated at posterior mean. pD = estimated effective number of parameters. PostMeanLogLik = posterior mean of the Log-Likelihood. DIC = deviance information criteria. * $y = 1'\mu + Zu + e$ = GBLUP model accounting for the genomic effect, not including microbiome effect.

Model	DIC
$y = 1'\mu + Zu + e$	184.4
$y = 1'\mu + Zu + Wm + e$	183.9 - 189.3
$y = 1'\mu + Zu + Wm + Tuxm + e$	187.5 - 191.7

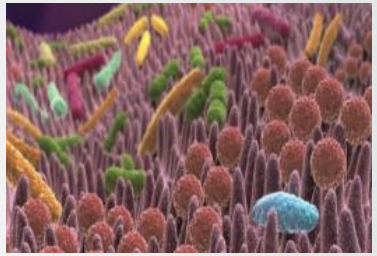
- The obtained MRM using MDS, RDA and CCA were as suitable as, or even better than the benchmark matrix in terms of the estimation of variance components, heritability and microbiability using simulation analysis.
- The genomic breeding values were accurately predicted when a microbiome effect was accounted for; the benchmark matrix and the canonical ordination methods of CCA and RDA showed higher accuracies than MDS, DCA and NMDS.
- It is possible to include a whole microbiota effect in the statistical analysis of feed efficiency.
- From the deviance information criteria, there is not enough evidence to reject any of the models that include microbiota information.

Thanks

Simulation of microbiota



Simulation of microbiota effect



- 50 out of the 92 OTUs were randomly selected. An effect ($\boldsymbol{\beta_j}$) was sampled from a normal distribution (N ~ (0, 1)) and assigned to each of the 50 selected OTUs.
- The microbiota effect ($\boldsymbol{m}$) for each animal was simulated as follow:

$$\boldsymbol{m_i} = \sum_j \boldsymbol{\beta_j} \times \boldsymbol{OTU_{ij}}$$

- Where $\boldsymbol{\beta_j}$ i s the effect of OTU_j and OTU_{ij} is the relative abundance of OTU j in animal i. The resulting $\{m_i\}$ was scaled to have a variance of σ_m^2

Simulation of genotype effect

- A dataframe with 1000 genotyped Holstein cows with allelic variants for 9244 SNPs was used
- The additive genetic effects were determined by 1000 QTL which were simulated from a normal distribution (~N (0, 1))
- The true breeding values (**u**) were calculated by adding all QTL effects which were subsequently scaled to a realized variance of σ_u^2

Simulation of phenotype

- Phenotypes were simulated assigning a residual variance to obtain a heritability of 0,3 and a microbiability of 0,5

- Simulated for an independent effect model and for an interaction effect model as follow:

$$y_i = \mu + u_i + m_i + e_i$$

$$y_i = \mu + u_i + m_i + u_i \times m_i + e_i$$

- Where μ is the population mean, u_i is the genomic effect, m_i is the microbiota effect, $u_i \times m_i$ is the genomic-microbiota interaction effect and e_i is the residual.