

# Genomic relatedness strengthens genetic connectedness across management units

Haipeng Yu

Annual Meeting of the American Society of Animal Science

July 11, 2017

# Introduction

## "Genetic Connectedness"

- In animal breeding, it refers to a measure of genetic relatedness across management units (e.g., herds and flocks).
- Best linear unbiased prediction (BLUP): genetic links or relatedness

### Objective:

- Assess measures of genetic connectedness across management units with use of genomic information.

# Prediction Error Variance (PEV)

Inverse of MME coefficient matrix:

$$\begin{aligned}\mathbf{C}^{-1} &= \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{K}^{-1}\lambda \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{bmatrix}\end{aligned}$$

$$\begin{aligned}\text{PEV}(\hat{u}) &= (\mathbf{Z}'\mathbf{M}\mathbf{Z} + \mathbf{K}^{-1}\lambda)^{-1}\sigma_e^2 \\ &= \mathbf{C}^{22}\sigma_e^2\end{aligned}$$

where  $\mathbf{M}$  is the absorption matrix given by

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

# Statistics of Pairwise Genetic Connectedness

Prediction error variance of the difference (**PEVD<sub>ij</sub>**)

$$\begin{aligned}\text{PEVD}(\hat{u}_i - \hat{u}_j) &= [\text{PEV}(\hat{u}_i) + \text{PEV}(\hat{u}_j) - 2\text{PEC}(\hat{u}_i, \hat{u}_j)] \\ &= (\mathbf{C}_{ii}^{22} - \mathbf{C}_{ji}^{22} - \mathbf{C}_{ij}^{22} + \mathbf{C}_{jj}^{22})\sigma_\epsilon^2 \\ &= (\mathbf{C}_{ii}^{22} + \mathbf{C}_{jj}^{22} - 2\mathbf{C}_{ij}^{22})\sigma_\epsilon^2,\end{aligned}$$

Coefficient of determination (**CD<sub>ij</sub>**)

$$\text{CD}_{ij} = 1 - \lambda \frac{\mathbf{C}_{ii}^{22} + \mathbf{C}_{jj}^{22} - 2\mathbf{C}_{ij}^{22}}{\mathbf{K}_{ii} + \mathbf{K}_{jj} - 2\mathbf{K}_{ij}}.$$

Prediction error correlation (**r<sub>ij</sub>**)

$$r_{ij} = \frac{\text{PEC}(\hat{u}_i, \hat{u}_j)}{\sqrt{\text{PEV}(\hat{u}_i)\text{PEV}(\hat{u}_j)}}.$$

# Choice of K

- **Numerator Relationship Matrix (A)** :  $a_{ij} = 1 + F_i$ ;  $a_{ij}$  is equal to twice the kinship coefficient between animal  $i$  and  $j$
- **Genomic Relationship Matrix (G)** :  $\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{m}$ , where 
$$\mathbf{Z}_{\cdot j} = \frac{z_{\cdot j} - 2p_j}{\sqrt{2p_j(1-p_j)}}$$
- $\mathbf{G}_{0.5}$  : MAF( $p_j$ ) in based generation is treated as 0.5
- **Scaled G matrix (Gs)** :  $\mathbf{G}_{sij} = \frac{(\mathbf{G}_{smax} - \mathbf{G}_{smin})(\mathbf{G}_{ij} - \mathbf{G}_{min})}{\mathbf{G}_{max} - \mathbf{G}_{min}}$
- **H matrix:**

$$\begin{bmatrix} \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{G}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{bmatrix}$$

# Materials

## Mice Data:

- public real data from Wellcome Trust Centre for Human Genetics
- **1,814** individuals, **10,346** markers and **523** cages

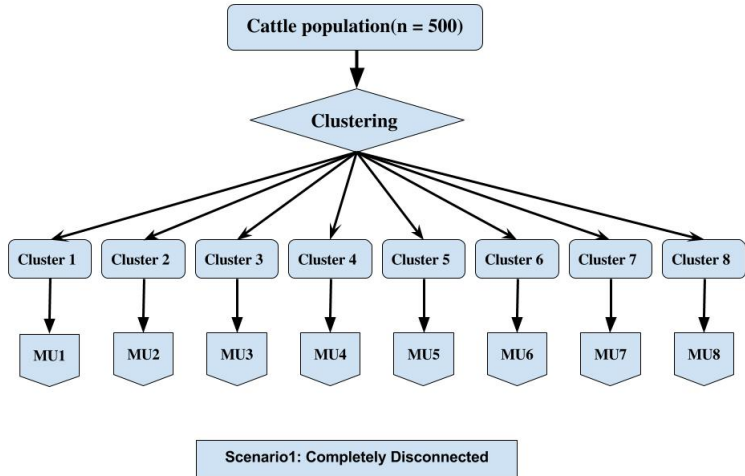
## Cattle Data:

- real data coupled with simulated management units , publicly available in the synbreed R package
- **500** individuals and **7,250** markers
- simulated management units from **8** clusters (K-medoid and average silhouette width algorithm)

# Average genetic connectedness measures across management units in the mice data

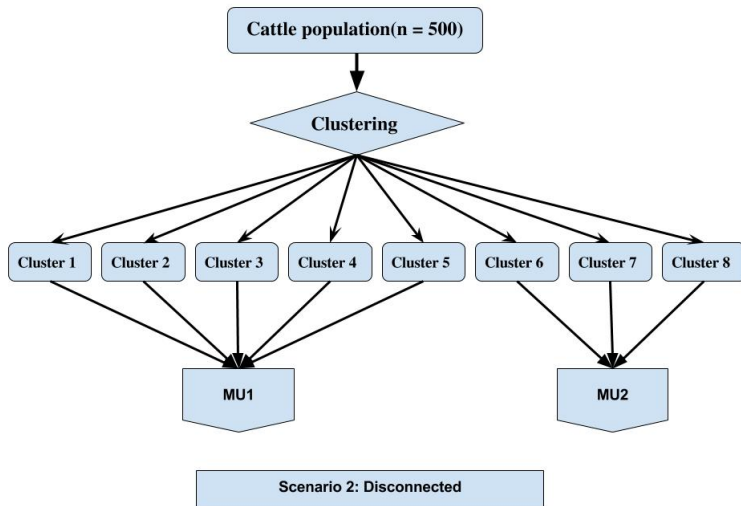
Methods	Kernels	Heritability ( $h^2$ )	
		0.8	0.2
PEVD	<b>A</b>	1.299	1.331
	<b>G</b>	0.456	1.037
	<b>G<sub>0.5</sub></b>	0.374	0.824
	<b>G<sub>s</sub></b>	0.532	1.254
CD	<b>A</b>	0.034	0.009
	<b>G</b>	0.662	0.234
	<b>G<sub>0.5</sub></b>	0.640	0.207
	<b>G<sub>s</sub></b>	0.690	0.270
$r_{ij}$	<b>A</b>	0.004	0.003
	<b>G</b>	-0.001	-0.001
	<b>G<sub>0.5</sub></b>	0.559	0.433
	<b>G<sub>s</sub></b>	0.496	0.270

# Cattle Management Units Simulation -Scenario 1

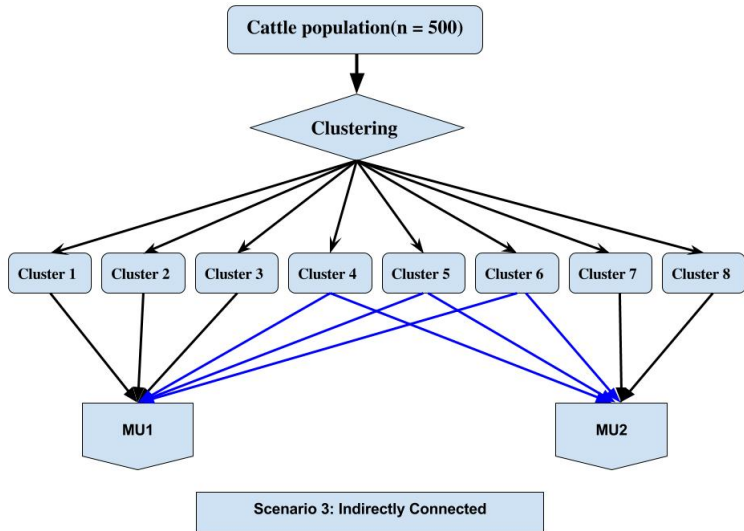




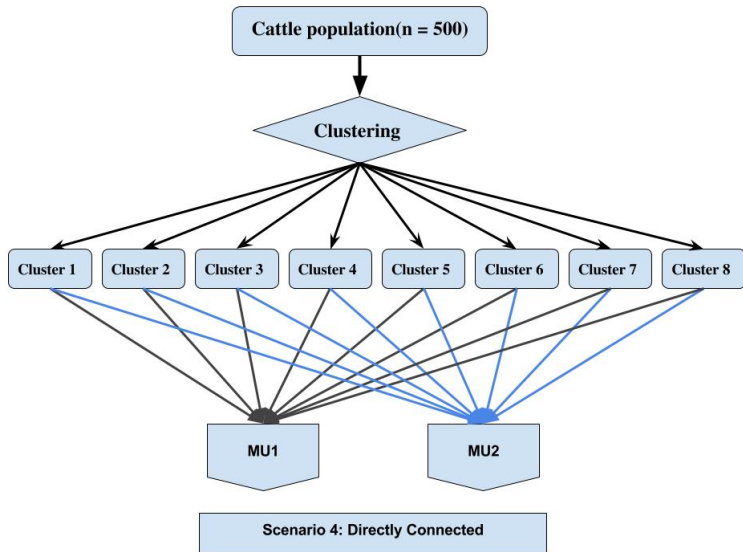
# Cattle Management Units Simulation -Scenario 2



# Cattle Management Units Simulation -Scenario 3



# Cattle Management Units Simulation -Scenario 4

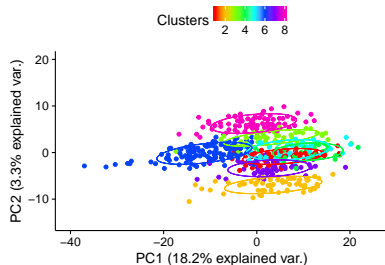
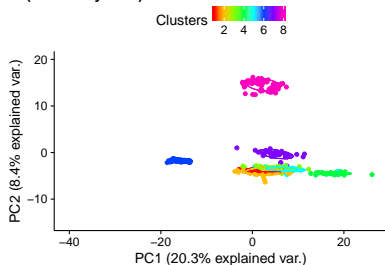


# Average genetic connectedness across management units for cattle data

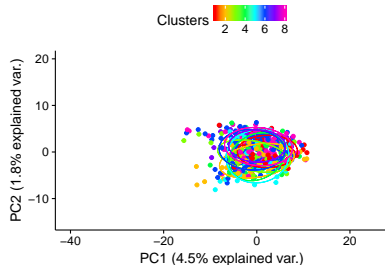
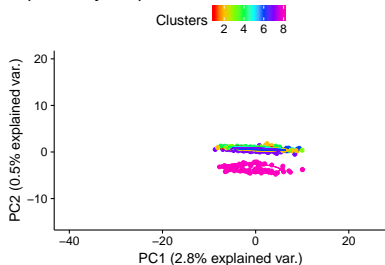
Scenarios	Methods	Kernels	Heritability ( $h^2$ )	
			0.8	0.2
S1	PEVD	<b>A</b>	0.077	0.102
		<b>G</b>	0.051	0.085
		<b>G<sub>0.5</sub> (G<sub>s</sub>)</b>	0.039 (0.066)	0.066 (0.110)
	CD	<b>A</b>	0.324	0.112
		<b>G</b>	0.539	0.224
		<b>G<sub>0.5</sub> (G<sub>s</sub>)</b>	0.528 (0.558)	0.195 (0.265)
	$r_{ij}$	<b>A</b>	0.017	0.005
		<b>G</b>	-0.014	-0.007
		<b>G<sub>0.5</sub> (G<sub>s</sub>)</b>	0.725 (0.468)	0.465 (0.174)
S4	PEVD	<b>A</b>	0.007	0.007
		<b>G</b>	0.005	0.007
		<b>G<sub>0.5</sub> (G<sub>s</sub>)</b>	0.004 (0.007)	0.005 (0.009)
	CD	<b>A</b>	0.125	0.048
		<b>G</b>	0.367	0.132
		<b>G<sub>0.5</sub> (G<sub>s</sub>)</b>	0.362 (0.384)	0.114 (0.158)
	$r_{ij}$	<b>A</b>	0.024	0.008
		<b>G</b>	-0.007	-0.002
		<b>G<sub>0.5</sub> (G<sub>s</sub>)</b>	0.741 (0.502)	0.470 (0.181)

# Principal Component Analysis for CD

SC1 (heritability = 0.8)

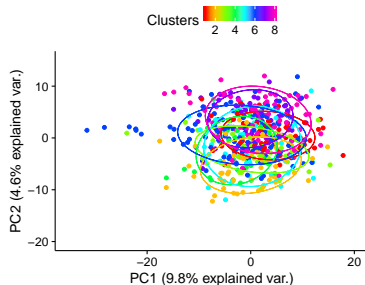
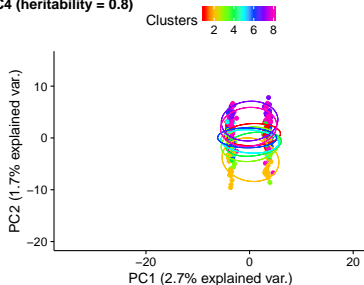


SC1 (heritability = 0.2)

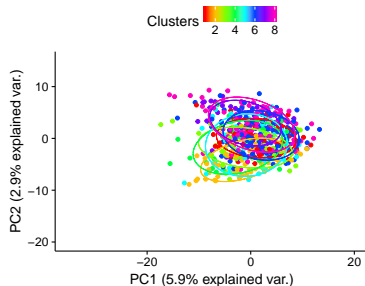
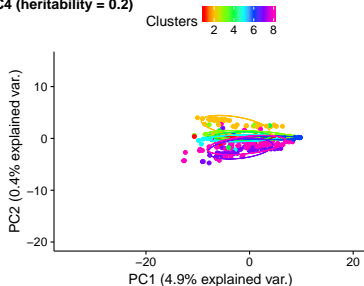


# Principal Component Analysis for CD

SC4 (heritability = 0.8)



SC4 (heritability = 0.2)



# Conclusion

- Genomic relatedness strengthens measures of genetic connectedness across management units compared to use of pedigree information.
- Coefficient of determination ( $R^2$ ) yields consistent measures of connectedness.
- Genomics provides relationships between units that appear disconnected according to the pedigree (independence assumption at baseline population)

# Acknowledgement

- Dr. Ron Lewis
- Dr. Matthew Spangler
- Dr. Dale Van Vleck
- Dr. Larry Kuehn
- Dr. Gota Morota