Challenges and opportunities for farmer-recorded data in health and welfare selection

C. Maltecca [1] K. Parker [1] J. Clay [2] J. Cole [3]

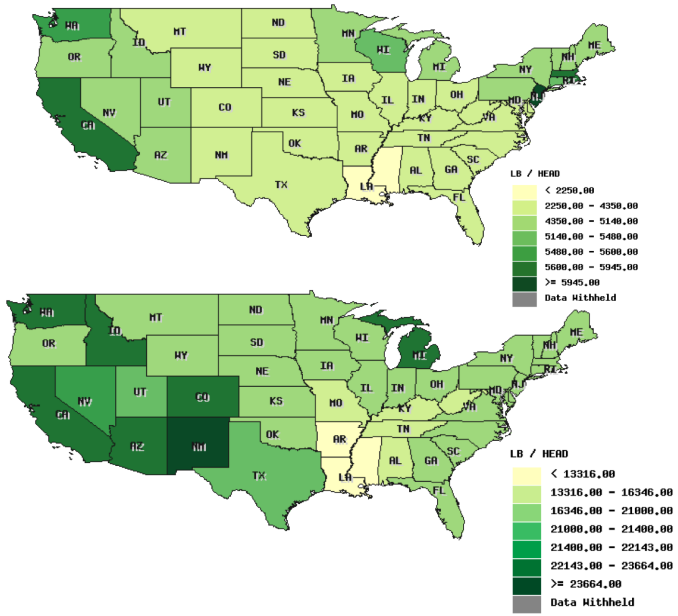
[1] North Carolina State University

[2] Dairy Records Management Systems

[3] USDA-AIPL

Aarhus 5-30-2013

Health traits

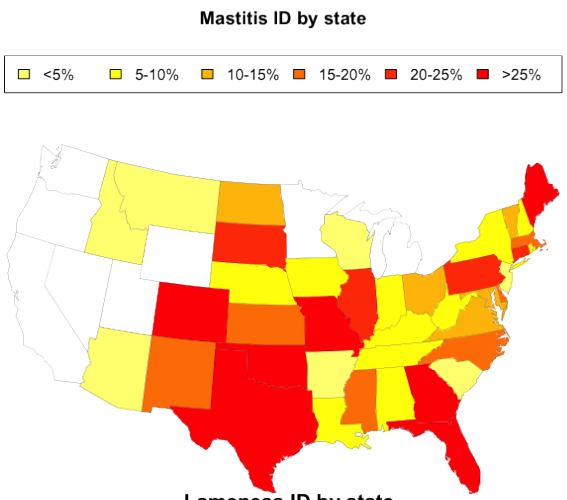


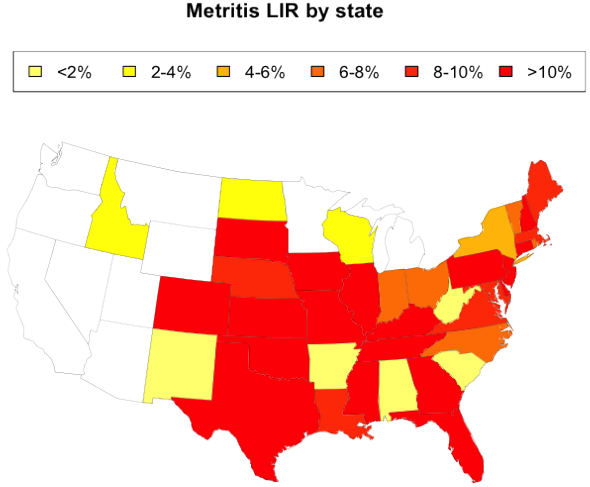
- Production Traits
  - Easy inexpensive to measure
  - Production more than tripled since 1950
- Health Traits
  - Lack of health-related phenotypes in the US an obstacle in achieving genetic improvement of health traits.
  - Antagonistic effects with production
  - Several confirmed the possibility of using on-farm recorded health information for genetic improvement.
- US producer-recorded data reflects true incidence of health events from epidemiological studies and relationships among occurrences are consistent

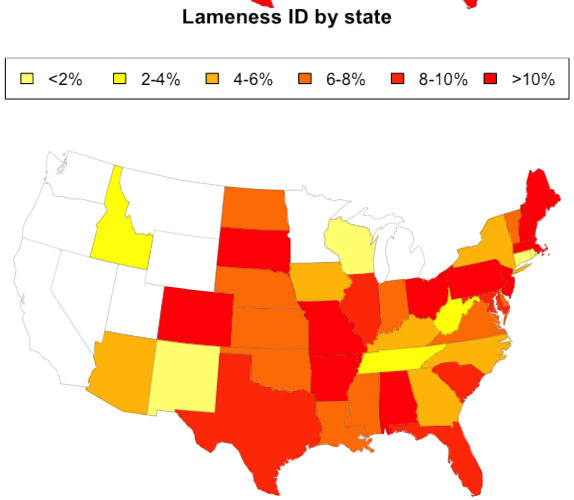
[3] Parker-Gaddis,et al. 2012

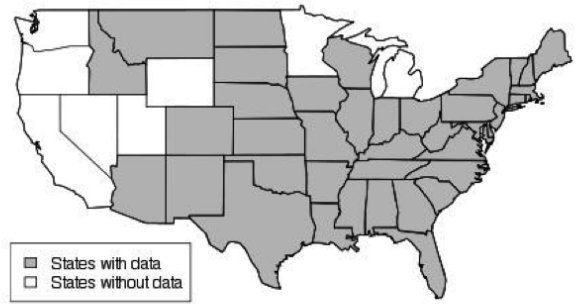
Health traits

The 3 most common diseases

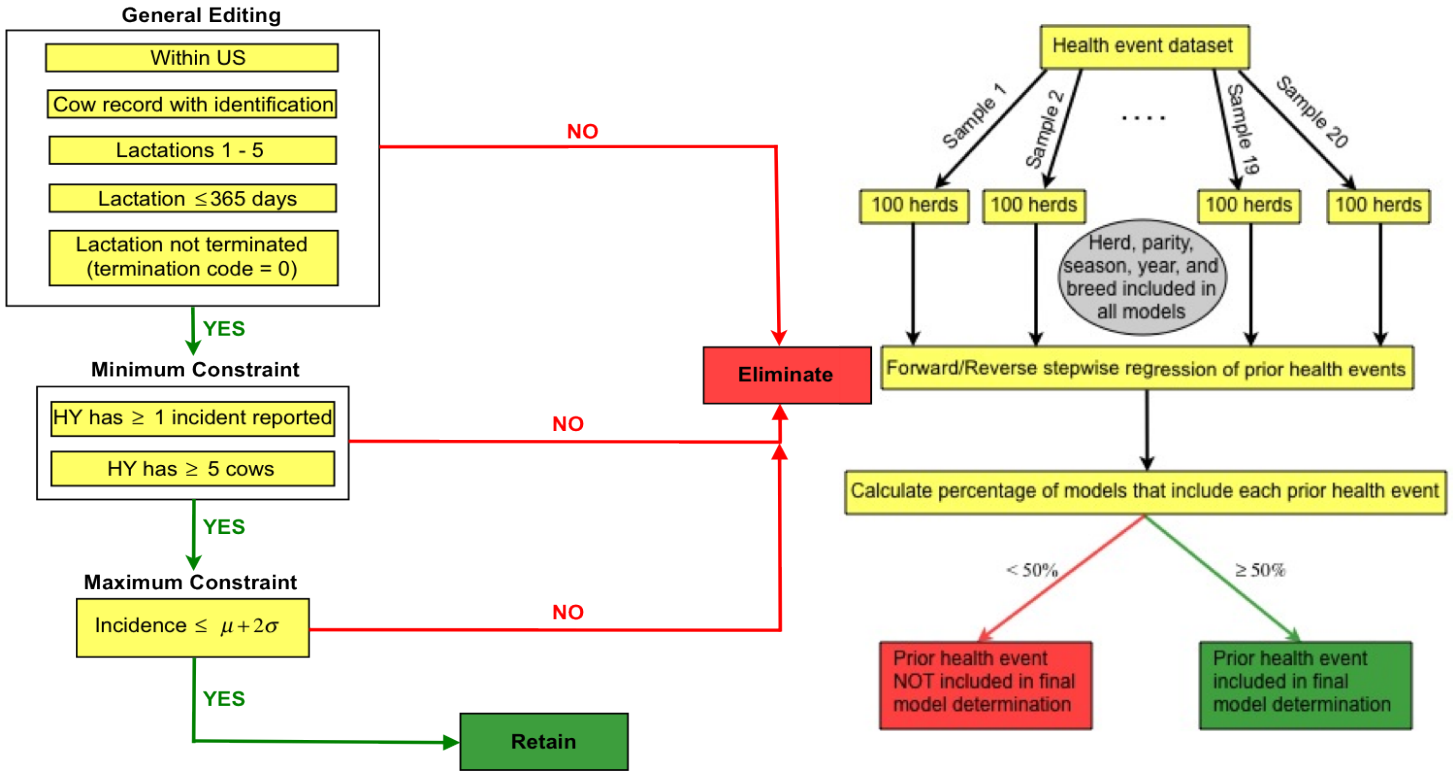




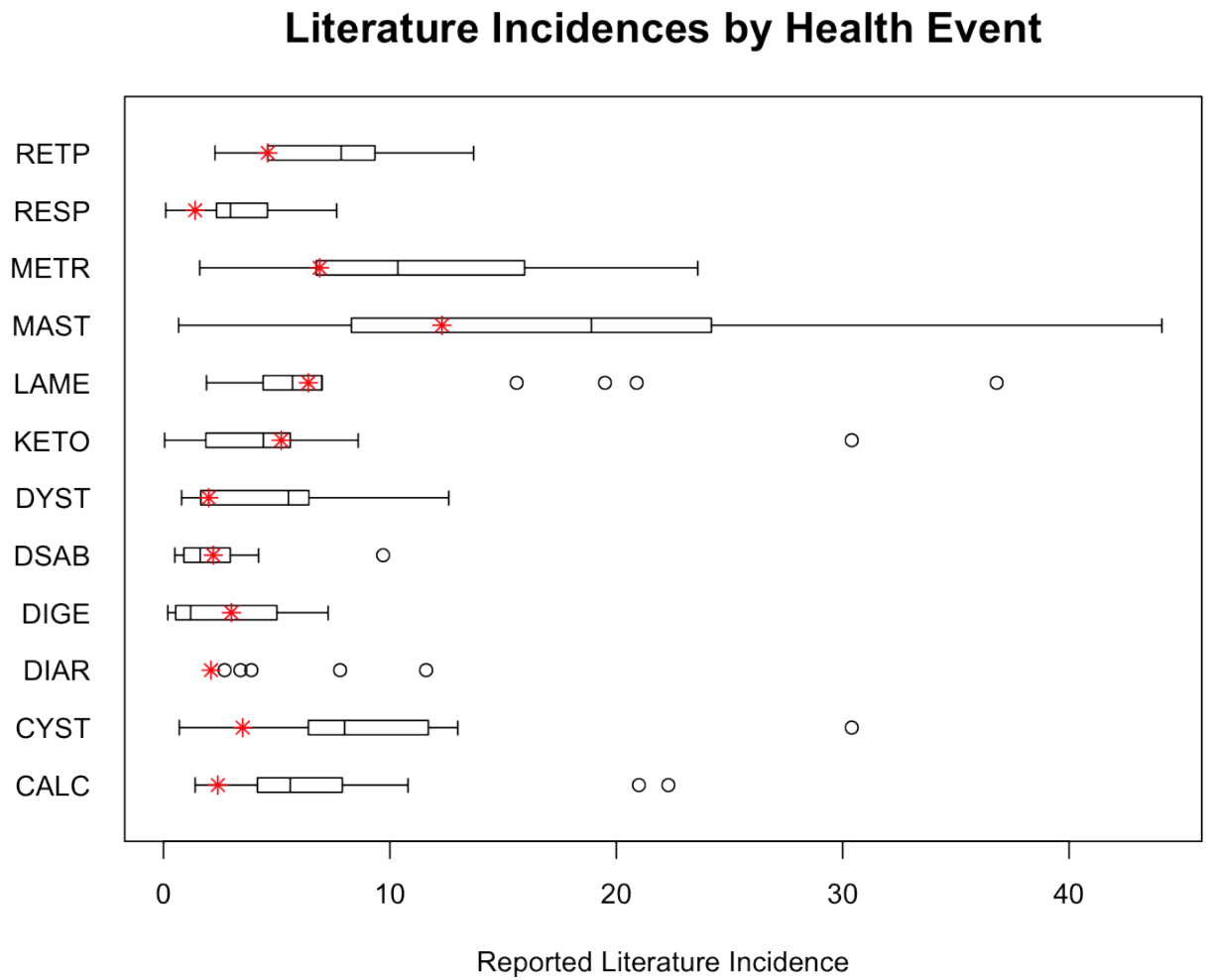




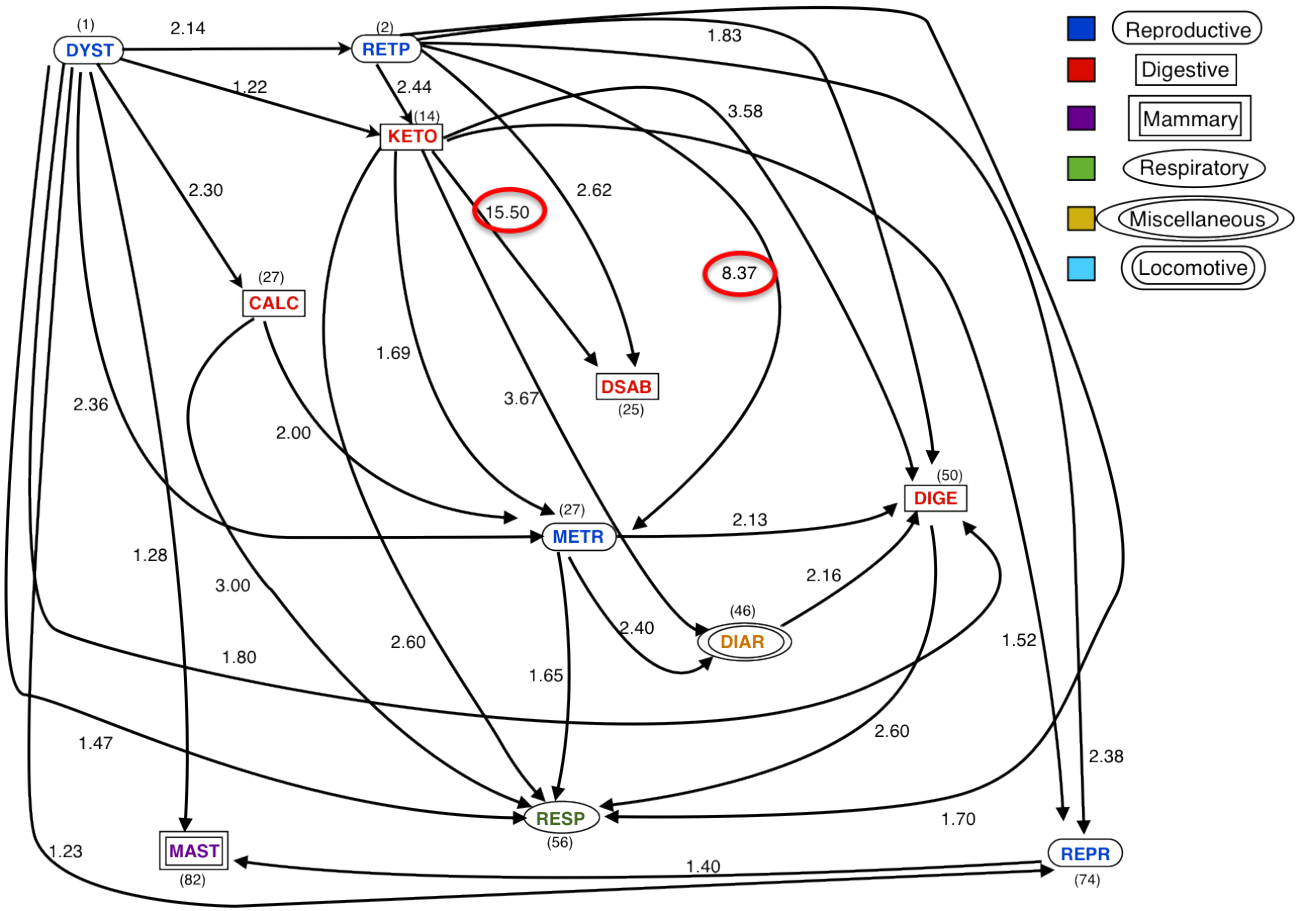
Health traits



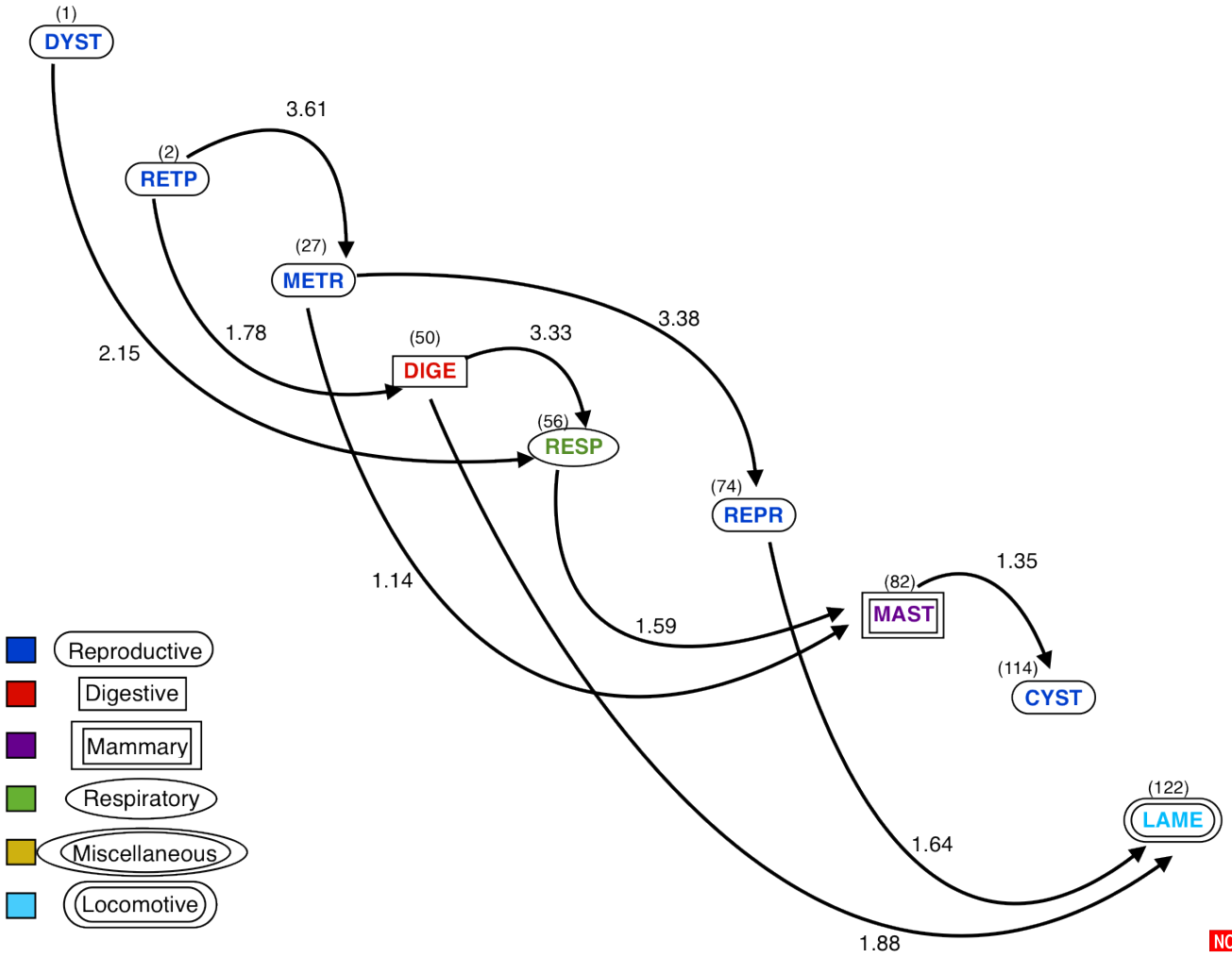
Health traits



Health traits



Health traits



Challenges with on-farm data

- Survey data still pose challenges in terms of data quality and appropriate use.
- A deeper understanding of causes and distribution for these data is needed.
- Endless pool of potential predictors
- Identifying a few key parameters for which a consistent and demonstrable improvement can be achieved

Beyond the first look

- An alternative perspective aiming at extracting the underlying health function of a cow.
  - PCA may be able to distinguish between groups of health events in order to further elucidate the complex nature of these traits.
  - Alternatively, a multiple correspondence analysis (MCA) can be performed directly with binary data
- A large role in the managing elements of dairy operation
  - Benchmarking management practices and herd characteristics related to disease incidence
  - Data quality control
  - Risk assessment

As part of a larger effort, we provide a preliminary characterization of both individual disease and herd characteristics related to disease incidence.

Disease Data

Health Event	Number of Records	Number of cows	Number of herd-year
Cystic Ovaries	22,937	131,194	3,369
Digestive disorders	156,520	97,430	1,780
Displaced abomasum	213,897	125,594	2,370
Ketosis	132,066	82,406	1,358
Lameness	233,392	144,382	3,191
Mastitis	274,890	164,630	3,859
Metritis	236,786	139,818	3,029
Reproductive disorders	253,272	151,315	3,360
Retained placenta	231,317	138,457	2,930

Herd Data

- Herd summary 4 time points throughout each year from 2000 through 2011
- Production, income, and feed cost summary
- Reproductive summary of the current breeding herd
- Reproductive summary of the total herd i
- Stage of lactation profile
- Genetic summary
- Production by lactation profile summary
- Dry cow profile summary
- Yearly summary of cows movements

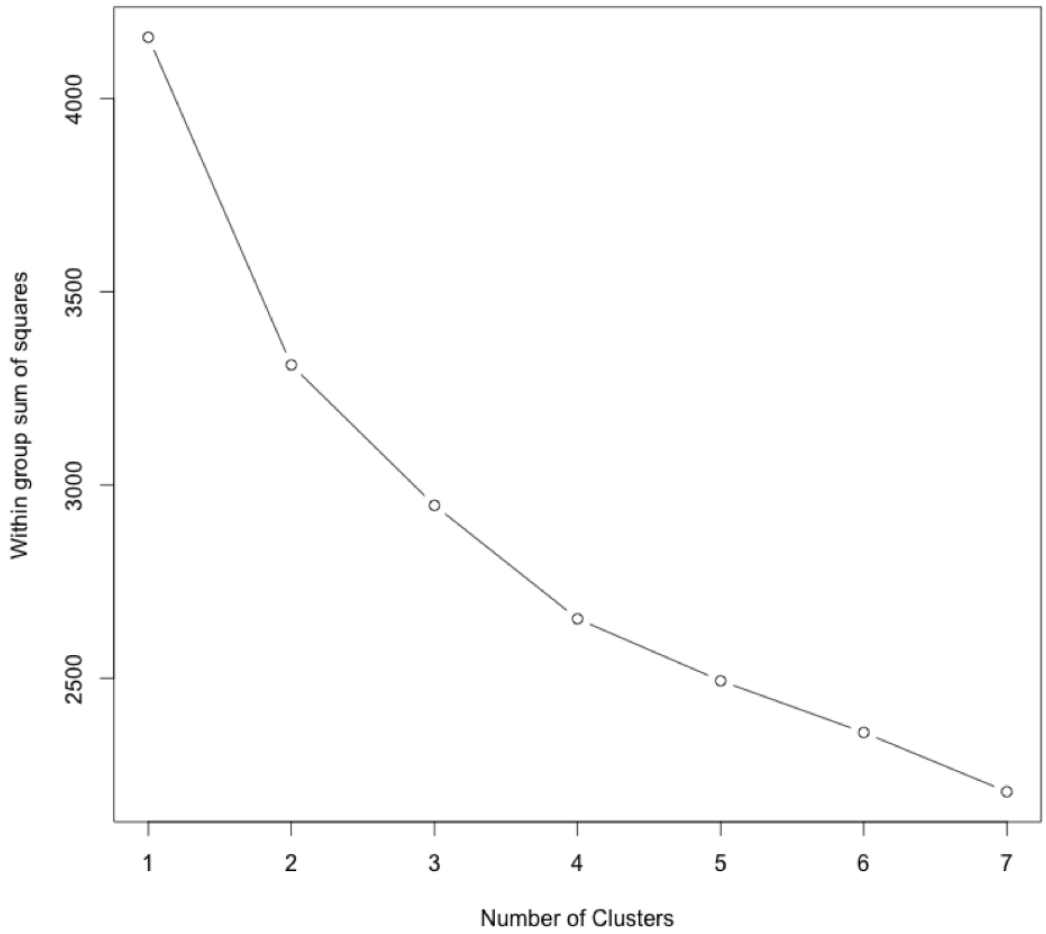
Indivduals grouped analysis

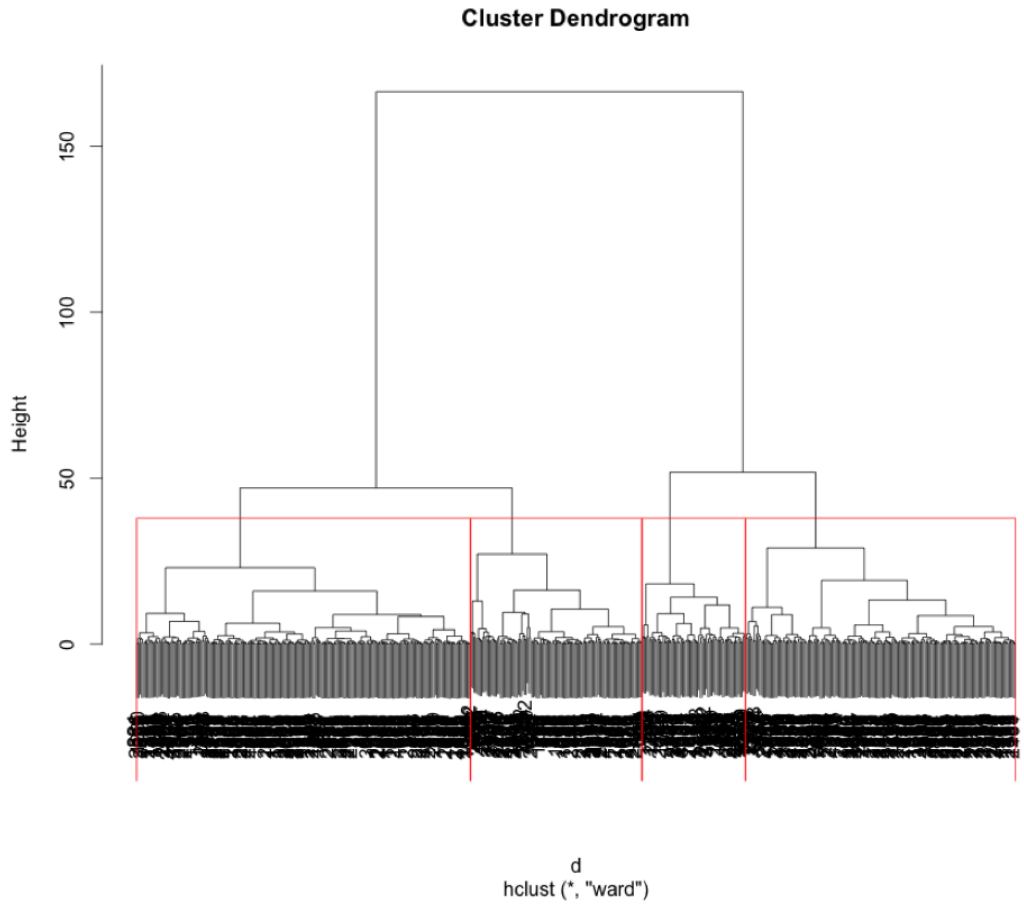
Investigate disease data clustering at individual level

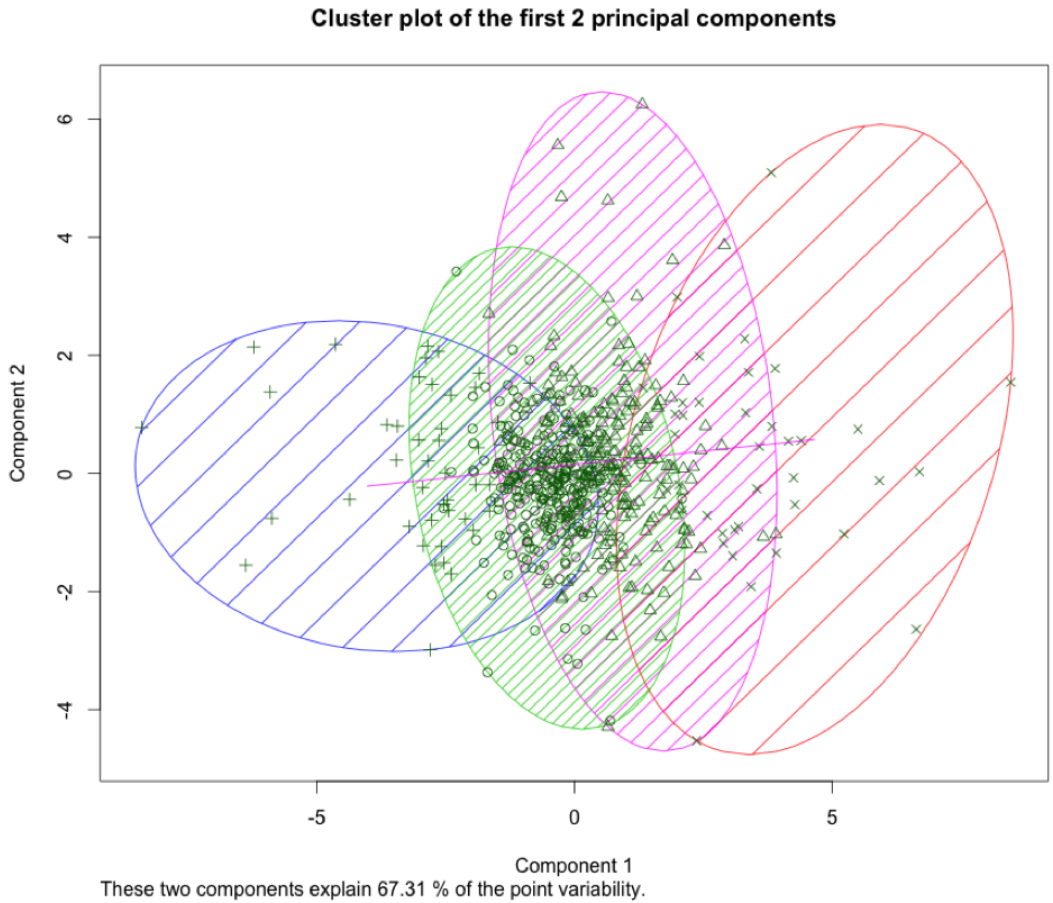
- MCA was performed with missing data imputation
- PCA was also performed
  - PCA requires quantitative variables, phenotypes used for this analysis were sire de-regressed estimated breeding values.
  - Estimated breeding values were obtained from a multiple-trait threshold sire analysis using the pedigree-based relationship matrix A.
- A scree plot to obtain an optimal number of clusters at the inflection point.
- Hierarchical cluster analysis based on k-means and Ward's minimum variance.

Herds grouped analysis

- PCA on 89 herd variables to determine if certain characteristics tended to occur together
- Herd variables were clustered in regard to the crude incidence of common health events
- Each health event was analyzed individually with optimum number of clusters estimated from data and observations split into the optimum number of clusters around medoids

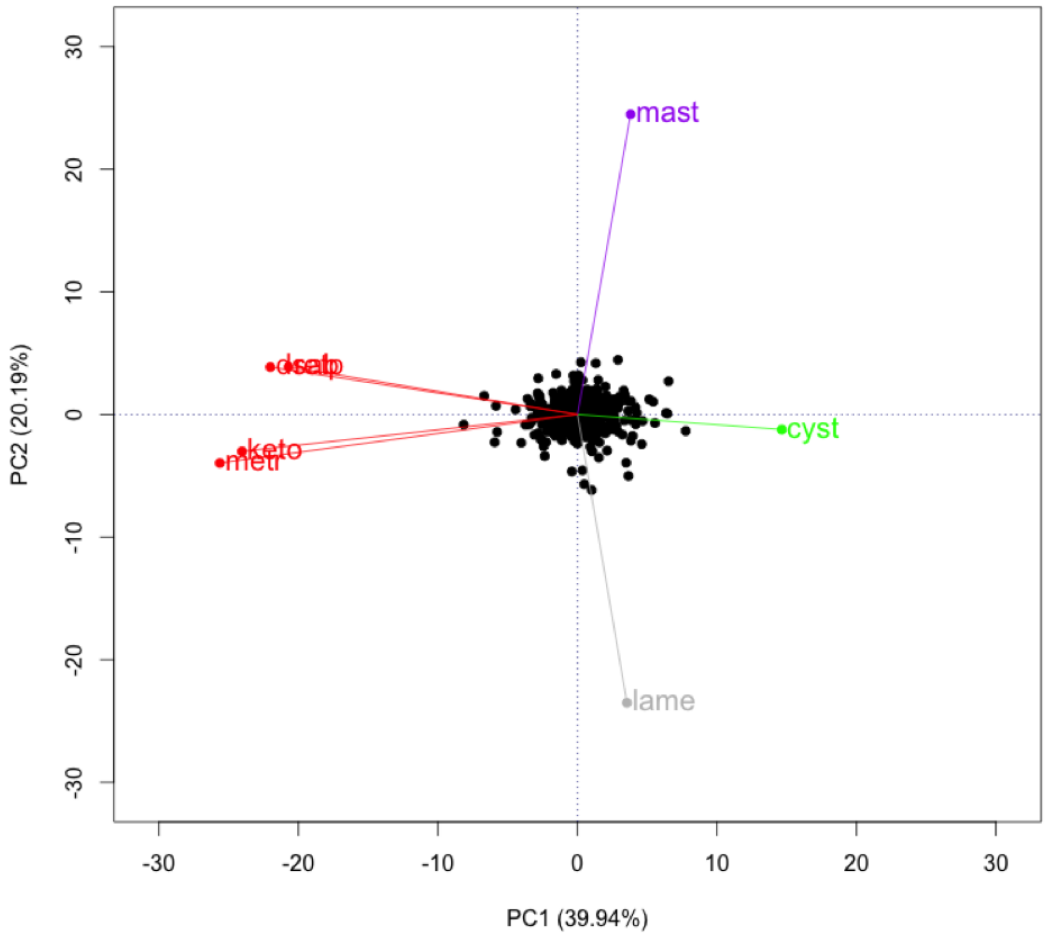

Within group sum of squares
2500
3000
3500
4000
1
2
3
4
5
6
7
Number of Clusters

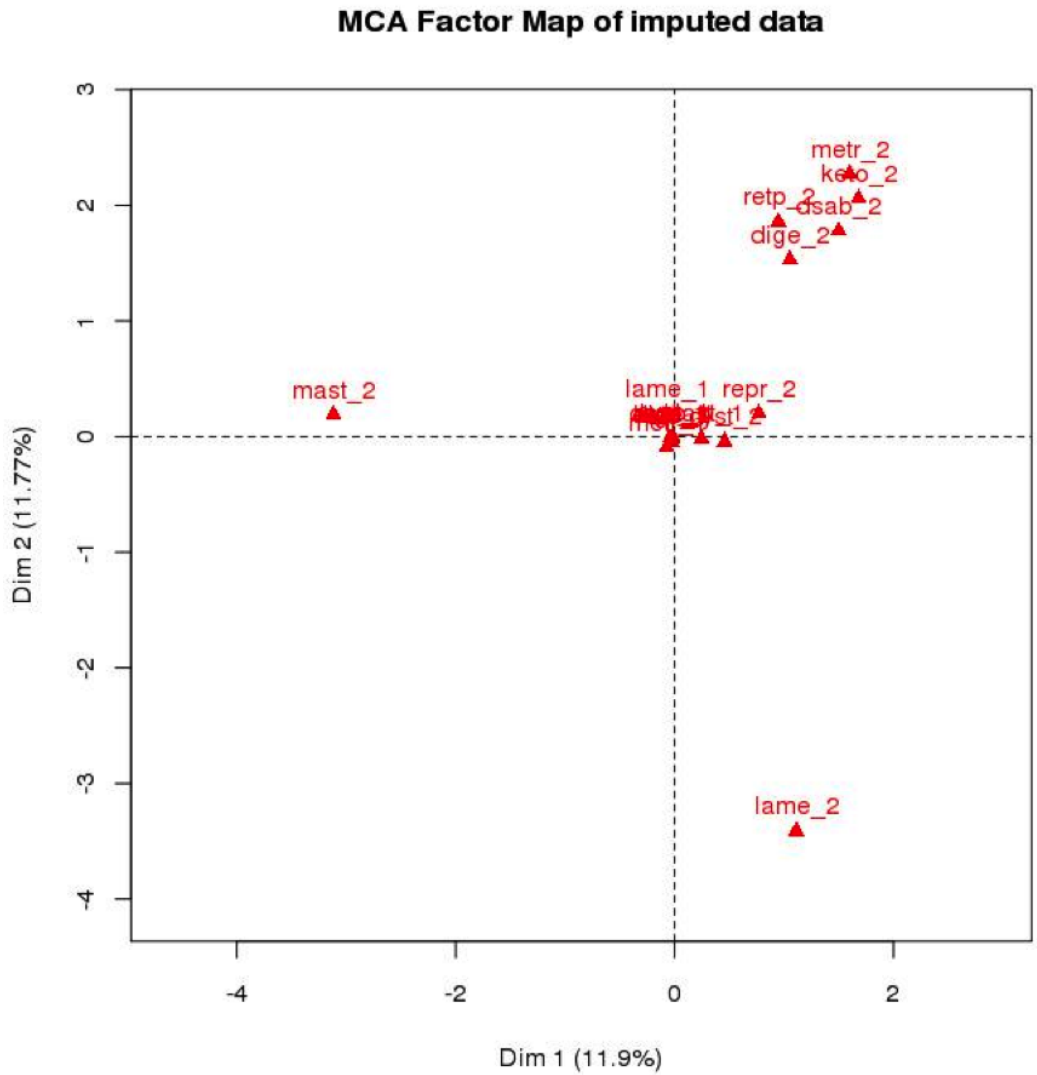

Cluster Dendrogram
Height
0
50
100
150
d
hclust (*, "ward")


Cluster plot of the first 2 principal components
Component 2
6
4
2
0
-2
-4
-5
0
5
Component 1
These two components explain 67.31 % of the point variability.

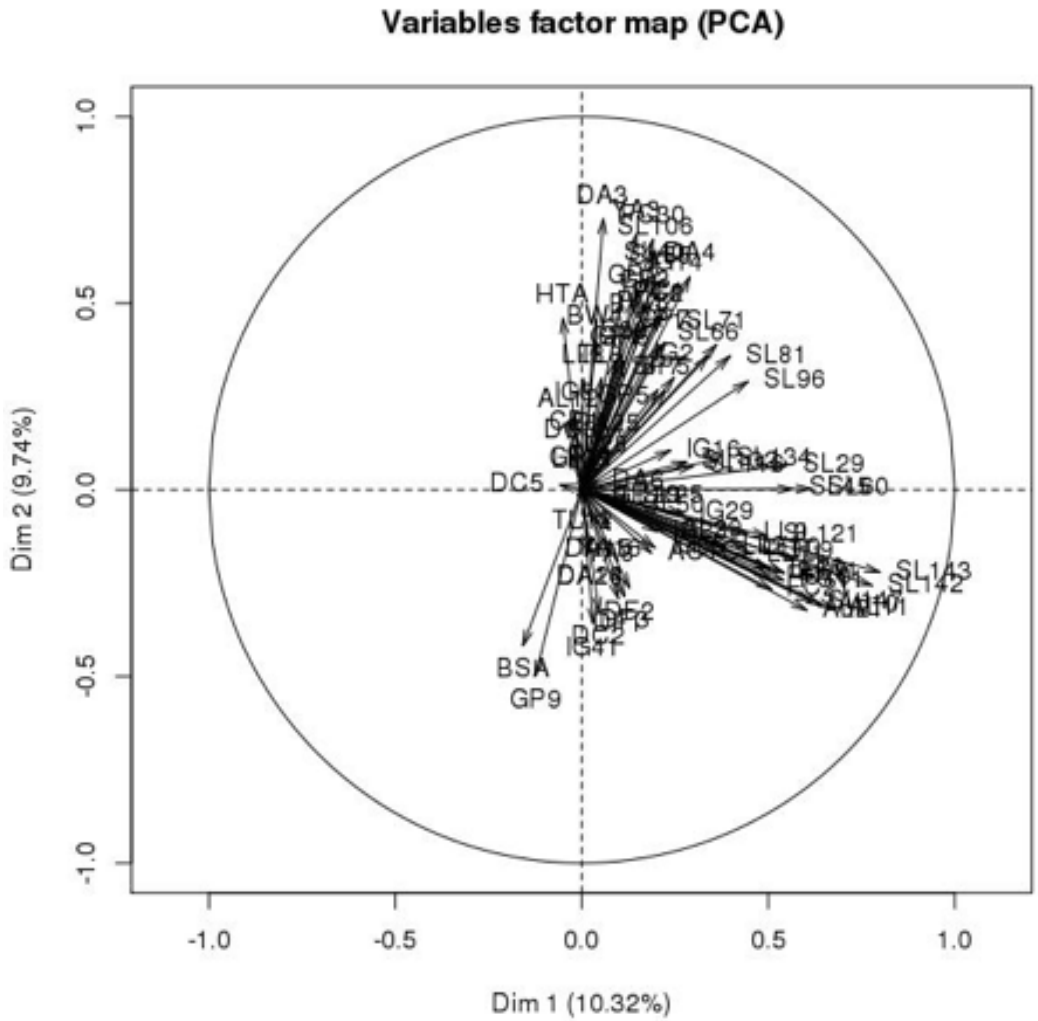
- Clustering for negative for MAST, negative for all events, negative values for metabolic and reproductive events, and positive values for all events.

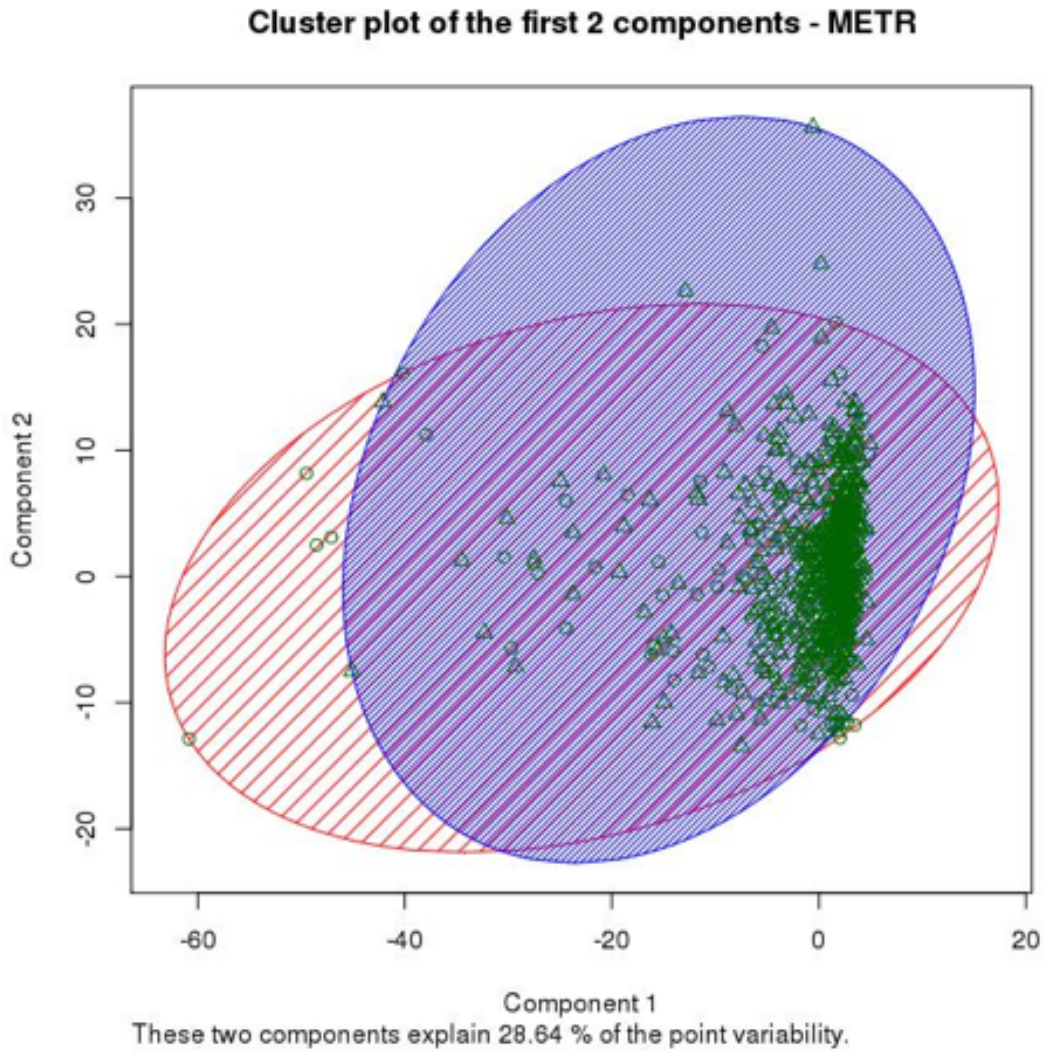
Cluster	CYST	DSAB	KETO	LAME	MAST	METR	RETP
1	-0.31	1.27	1.06	-0.01	-0.19	0.58	0.42
2	0.22	-0.62	-0.40	0.08	-0.09	-0.14	-0.11
3	0.60	-2.16	-1.33	0.16	0.09	-0.67	-0.53
4	0.02	0.30	0.18	-0.05	0.05	0.06	0.01

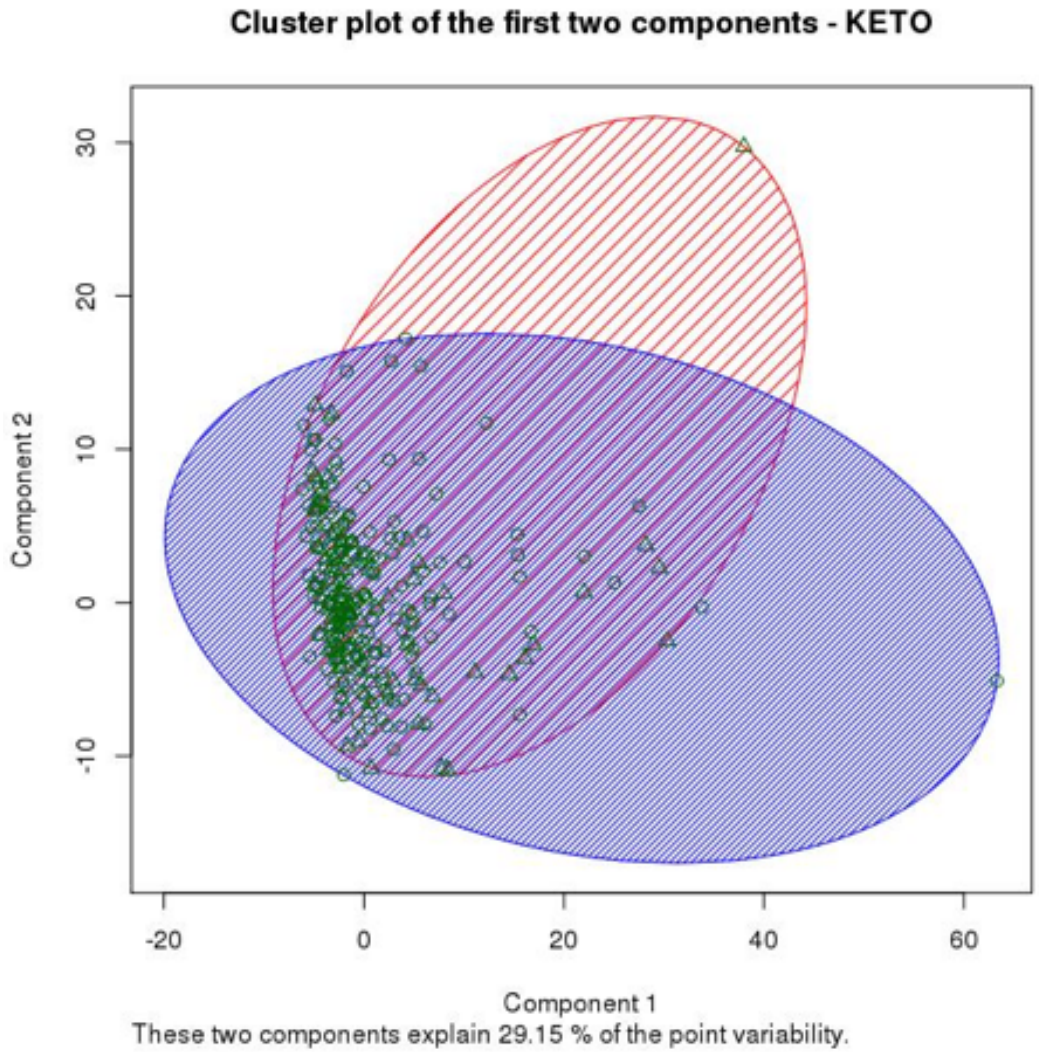

mast
dsetp
keto
metr
cyst
lame
PC2 (20.19%)
PC1 (39.94%)
30
20
10
0
-10
-20
-30
-30
-20
-10
0
10
20
30


MCA Factor Map of imputed data
metr_2
keto_2
retp_2
dsab_2
dige_2
mast_2
lame_1
repr_2
lame_2
Dim 2 (11.77%)
Dim 1 (11.9%)
3
2
1
0
-1
-2
-3
-4
-4
-2
0
2

Herd grouped analysis




Cluster plot of the first 2 components - METR
Component 2
30
20
10
0
-10
-20
-60
-40
-20
0
20
Component 1
These two components explain 28.64 % of the point variability.


Cluster plot of the first two components - KETO
Component 2
30
20
10
0
-10
-20
0
20
40
60
Component 1
These two components explain 29.15 % of the point variability.

Event	Inc	Group	TC	MY	FY	PY	DFS	CI	%SS	CALV	DMY	BW
RETP	0.005	(Low)	333	20884	792	653	84	13.9	36	561	66.6	1270
	0.10	(High)	444	22113	834	687	82	13.8	34	892	70.1	1306
MAST	0.008	(Low)	263	20872	794	652	85	14.0	36	349	66.5	1273
	0.16	(High)	554	21269	805	663	83	13.9	36	1110	68.3	1286
METR	0.01	(Low)	322	21026	803	660	86	14.0	35.3	445	66.8	1274
	0.14	(High)	578	21582	813	667	83	14.0	34.5	707	69.0	1310
KETO	0.006	(Low)	370	21744	833	681	79	13.8	32.7	427	69.3	1270
	0.10	(High)	441	22569	853	701	78	13.8	31.3	682	72.5	1311

Total cows Milk lbs. Fat lbs. Protein lbs. Avg. days to 1st service Actual calving interval Avg. % successful services Total number calving Avg. daily milk production Body weight

Health Event	Variable-1	Variable-2	Variable-3
KETO	MY	DTF	SPP
MAST	ATPC	VWP	TC
METR	VWP	TC	FC
RETP	ATPC	AHO	PCC

MY =Milk yield, DTF= Avg. days to 1st service (2nd lactation cows) SPP= Services per pregnancy (pregnant 1st lactation cows) ATPC=Average total pregnant cows VWP=Voluntary waiting period TC=Total cows FC=Feed cost per cwt milk AHO=Average percentage heats observed PCC=Pounds concentrate consumed

- Pointing to a combination of (few several) biological processes underlying disease resistance through Functional Annotation

Entry	Pathway	PCA	Class
bta04144	Endocytosis	2	Cellular_Processes;Cell Growth and Death
bta04145	Phagosome	2	Cellular_Processes;Transport and Catabolism
bta04210	Apoptosis	2	Cellular_Processes;Transport and Catabolism
bta00260	Glycine, serine and threonine metabolism	4	Metabolism; Amino Acid Metabolism
bta00500	Starch and sucrose metabolism	4	Metabolism; Carbohydrate Metabolism
bta00190	Oxidative phosphorylation	1	Metabolism; Energy Metabolism
bta00512	O-Glycan biosynthesis -	1	Metabolism; Glycan Biosynthesis and Metabolism
bta00531	Glycosaminoglycan degradation -	4	Metabolism; Glycan Biosynthesis and Metabolism
bta01100	Metabolic pathways -	4	NA
bta01100	Metabolic pathways -	2	NA
bta04910	Insulin signaling pathway -	4	Organismal Systems; Endocrine System
bta04062	Chemokine signaling pathway -	3	Organismal Systems; Immune System
bta04062	Chemokine signaling pathway -	2	Organismal Systems; Immune System
bta04062	Chemokine signaling pathway -	1	Organismal Systems; Immune System
bta04622	RIG-I-like receptor signaling pathway -	4	Organismal Systems; Immune System
bta04623	Cytosolic DNA-sensing pathway -	4	Organismal Systems; Immune System
bta04660	T cell receptor signaling pathway -	2	Organismal Systems; Immune System

- Opportunities exist to improve disease prediction and overall herd disease management by making use of patterns observed at both individual and herd level.
- Grouped information can be used in data editing and herd benchmarking, as well as a way to increase selection efficacy.
- Further evaluations of more comprehensive predictive models are nonetheless required.