Incorporation of DNA Tests into a Genetic Evaluation

Stephen D. Kachman

Department of Statistics
University of Nebraska–Lincoln

July 2, 2008

Introduction

- Objective is to evaluate the genetic potential of an animal
  - Phenotypic information
  - Pedigree information

Introduction

- Objective is to evaluate the genetic potential of an animal
  - Phenotypic information
  - Pedigree information
- Incorporate DNA information
  - Marker Information
  - Molecular Breeding Values

Introduction

- Objective is to evaluate the genetic potential of an animal
  - Phenotypic information
  - Pedigree information
- Incorporate DNA information
  - Marker Information
  - Molecular Breeding Values
- Implementation
  - Fit alongside current approaches
  - Practical
  - Flexible

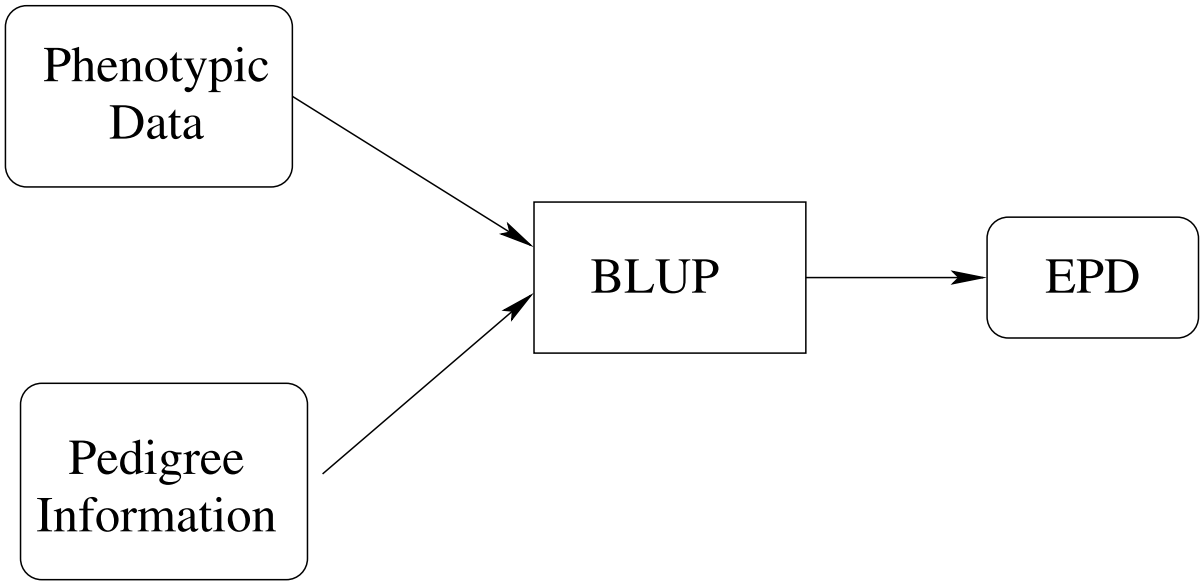
- Fostered by the NBCEC
- Team
  - Rohan Fernando (Iowa State)
  - Rob Tempelman (Michigan State)

Outline

- Quick overview of current approach to genetic evaluation
- DNA marker information
- Incorporation into genetic evaluations
  - MBVs as correlated traits
  - Similarities and differences with other correlated traits
  - Challenges and how to address them
- Initial analysis
- Some implications
- Where do we go from here?

Current Phenotypic Based Analysis

- Sources of information:
  - Phenotypic records on the trait of interest
  - Phenotypic records on correlated traits
  - Pedigree information
- Evaluation
  - Best Linear Unbiased Prediction



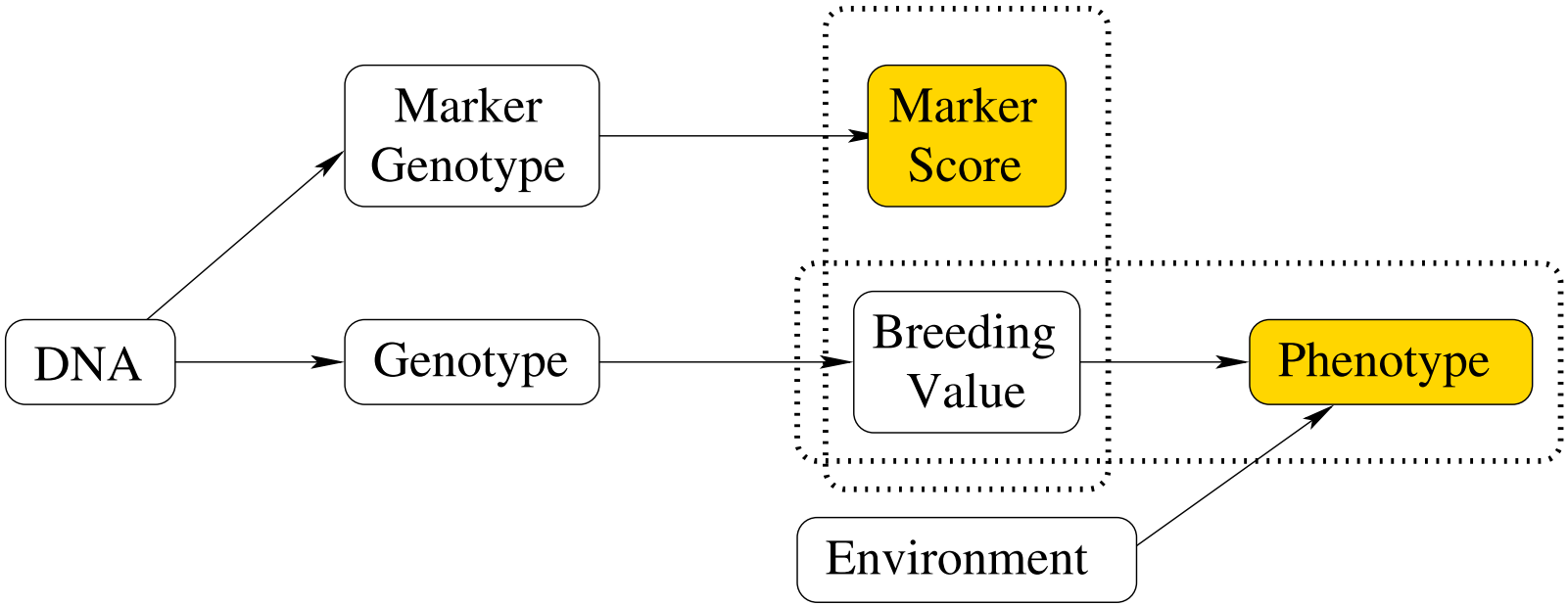
- Overall this approach works well
- Phenotypic data may be limited
  - Difficult to collect
  - For example: Carcass traits
- Need complementary sources of information
  - DNA Information
  - Summarized into MBVs or Marker Scores

- DNA Information
  - Single marker
  - Several markers

- DNA Information
  - Single marker
  - Several markers
  - Thousands of markers
  - Sequence data

- DNA Information
  - Single marker
  - Several markers
  - Thousands of markers
  - Sequence data
- Summarized into marker scores
  - Estimated from reference populations
  - Flexibility to handle evolving molecular technology
  - Flexibility to handle evolving statistical methodology

Marker Scores and Phenotypes



- Conceptually marker scores can be viewed as a correlated trait
  - Similar to using Birth Weight when evaluating Calving Ease
    - Observable trait
    - Underlying genetic component which is correlated to the trait of interest

Tenderness Example

- Warner-Bratzler Shear Force (kg)
- 410 steers
- 14 sires
- 36 contemporary groups
- Marker scores
  - Pfizer GeneSTAR Tenderness
  - Igenity Tenderness
  - MMI Tru-Tenderness

Models

- Phenotypic Model (Single Trait Model)
  - Only include phenotypic data
  - Fixed: Contemporary group
  - Random: Direct additive effect
  - Residual

Models

- Phenotypic Model (Single Trait Model)
  - Only include phenotypic data
  - Fixed: Contemporary group
  - Random: Direct additive effect
  - Residual
- Marker Model (Three Trait Model)
  - Fixed: Company baseline effect
  - Random: Direct additive effect

Models

- Phenotypic Model (Single Trait Model)
  - Only include phenotypic data
  - Fixed: Contemporary group
  - Random: Direct additive effect
  - Residual
- Marker Model (Three Trait Model)
  - Fixed: Company baseline effect
  - Random: Direct additive effect
- Full Model (Four Trait Model)

Models

- Phenotypic Model (Single Trait Model)
  - Only include phenotypic data
  - Fixed: Contemporary group
  - Random: Direct additive effect
  - Residual
- Marker Model (Three Trait Model)
  - Fixed: Company baseline effect
  - Random: Direct additive effect
- Full Model (Four Trait Model)
- Reduced Models
  - Use all available phenotypic and marker scores
  - Exploit commonalities in the genetic basis of the marker scores

Phenotypic Accuracies

Generation	Median	Range
Sire	0.30	0.00–0.39
Progeny	0.21	0.00–0.24

- Accuracies are greater in the sires compared to the progeny.

Full Model

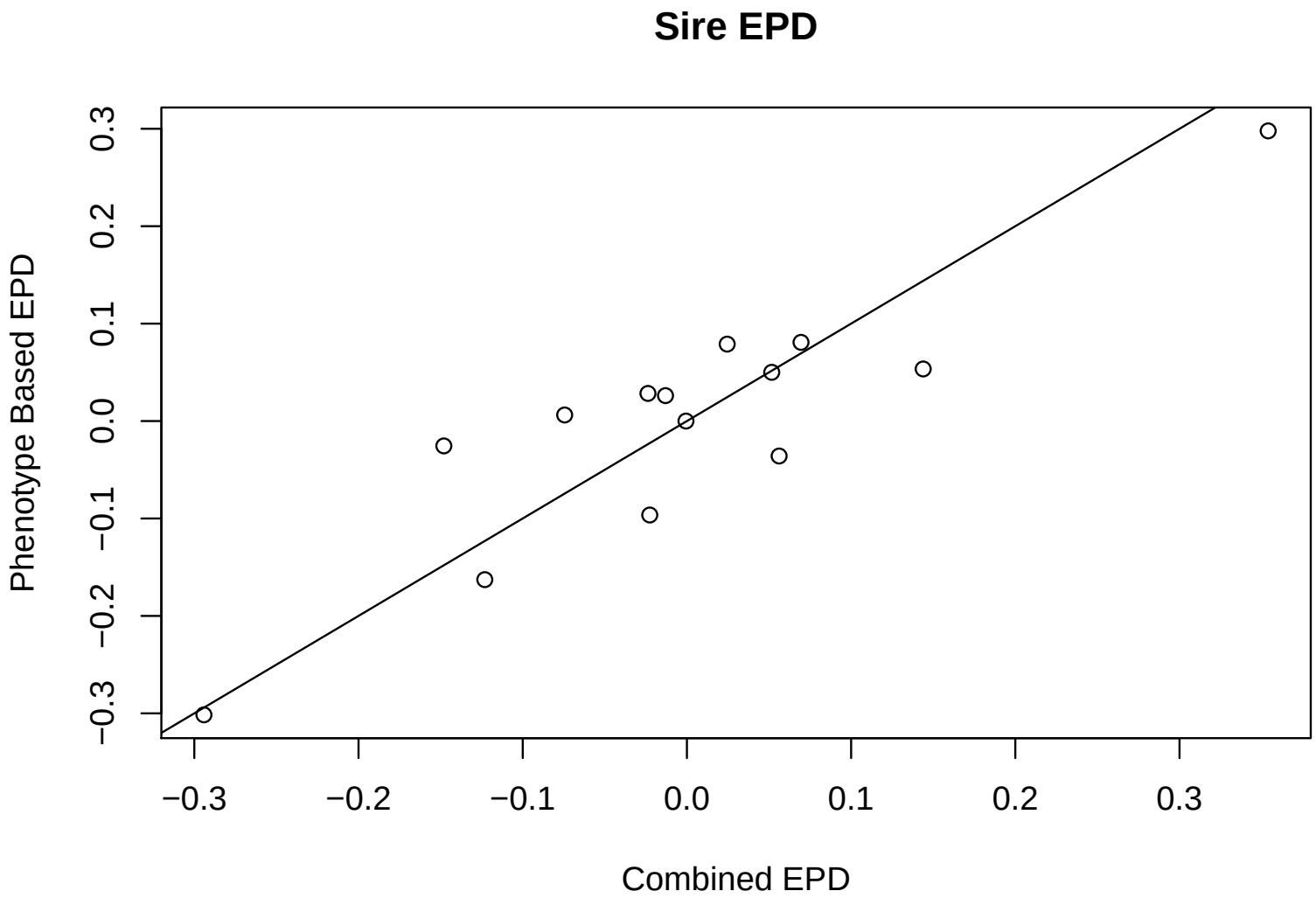
- Included four traits
  - Shear Force
  - Company A Marker Score
  - Company B Marker Score
  - Company C Marker Score
- Genetic variances for the shear force and marker scores
- Genetic Correlations of shear force with each of the marker scores
- Genetic correlations between the marker scores
- No residual variation in the marker score

Median Accuracies

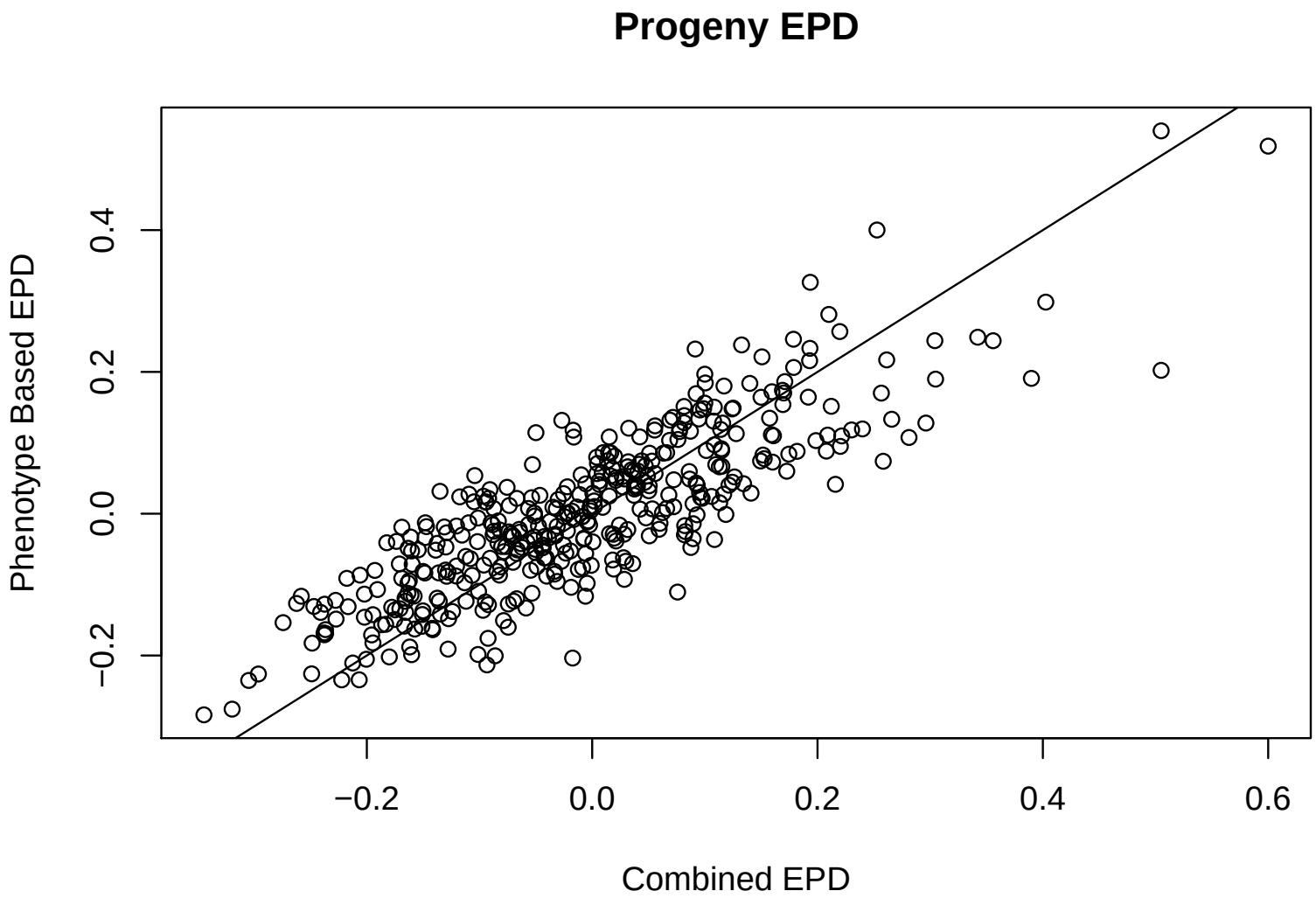
	Model	
Generation	Phenotypic	Full
Sire	0.30	0.34
Progeny	0.21	0.30

- Accuracy increases with the addition of marker information
- Benefit is greatest for the individual who is genotyped

Full versus Phenotype EPDs (r=0.90)



Full versus Phenotype EPDs (r=0.85)



Marker only model

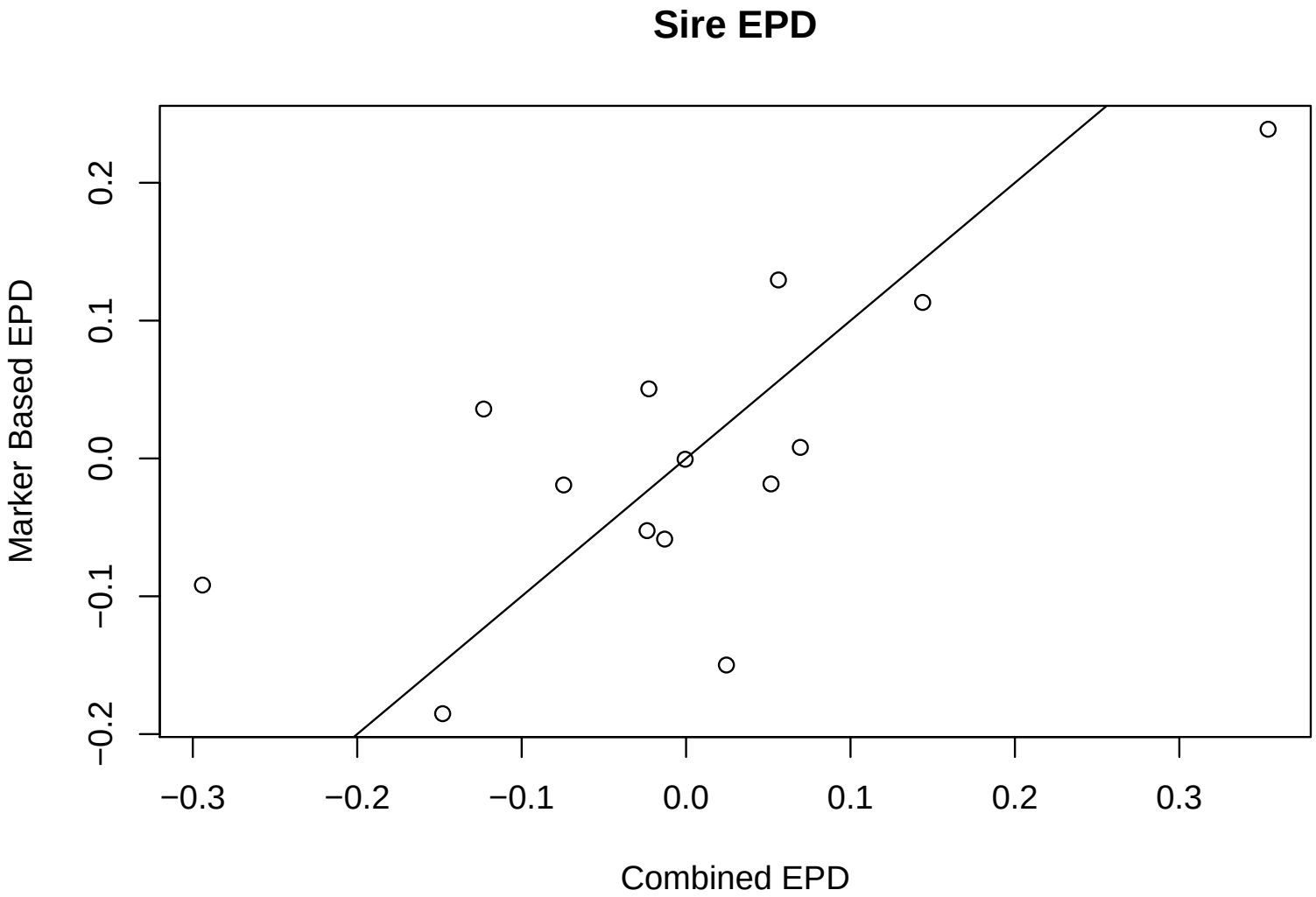
- Used genetic parameter estimates from the full model
  - Estimation of genetic correlations between the shear force and marker scores requires that both be observed in the same data set.
- Excluded all shear force data when predicting the shear force EPDs

Median Accuracies

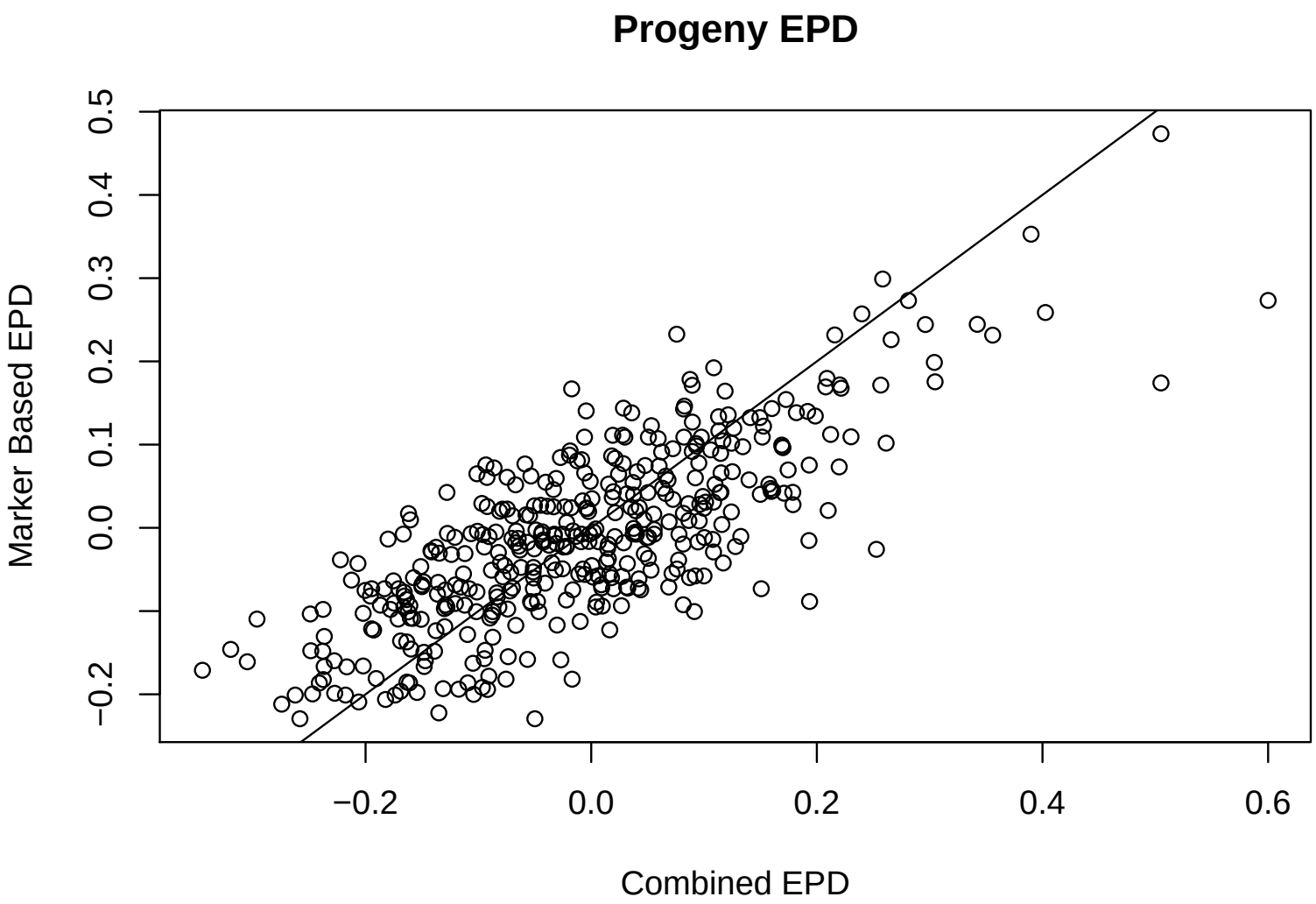
	Model		
Generation	Phenotypic	Full	Marker
Sire	0.30	0.34	0.13
Progeny	0.21	0.30	0.16

- Marker only accuracy is greatest in the individual who is genotyped.
- Once an individual has been genotyped, genotyping additional relatives doesn't have a direct impact on accuracy.

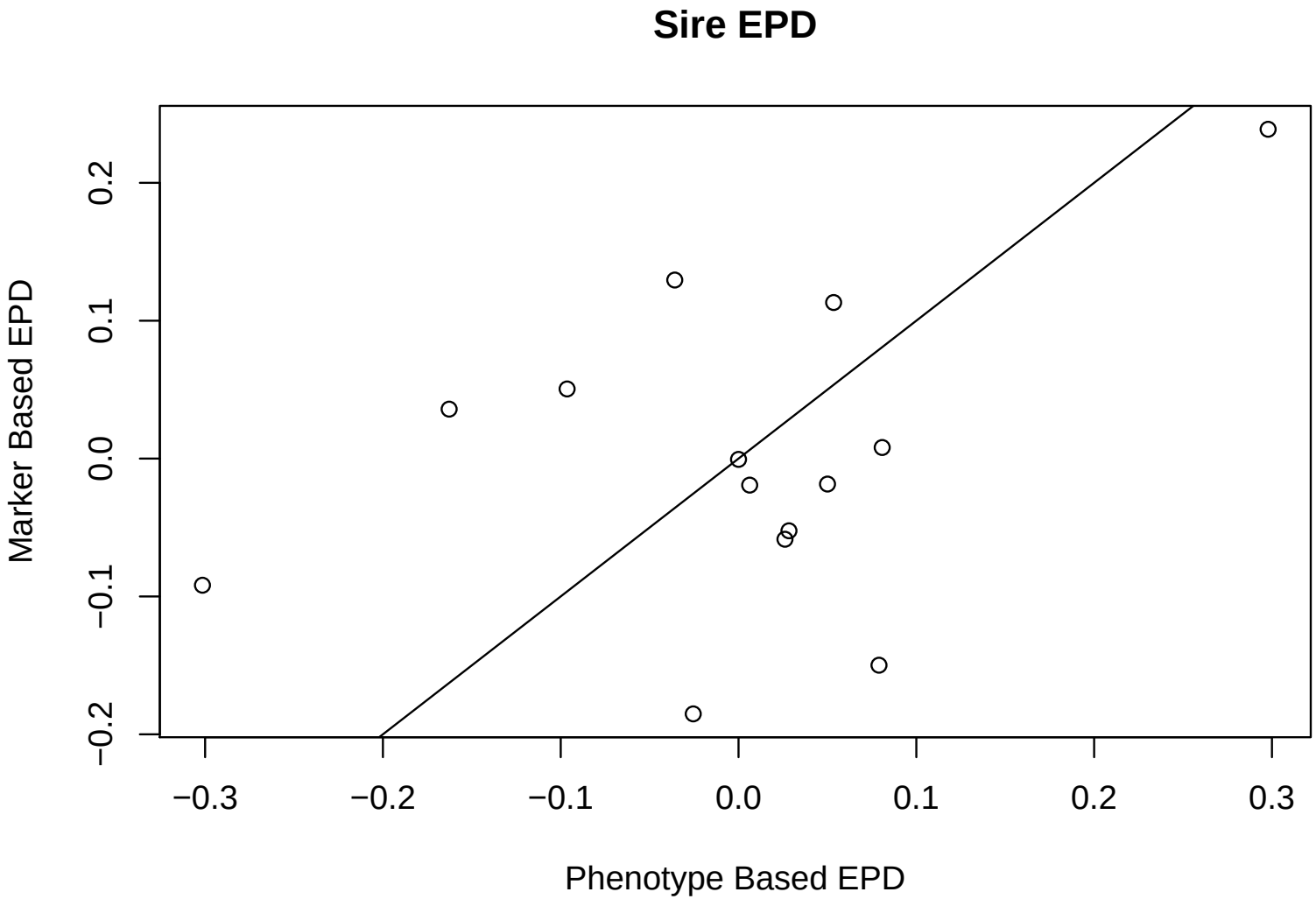
Full versus Marker EPDs (r=0.72)



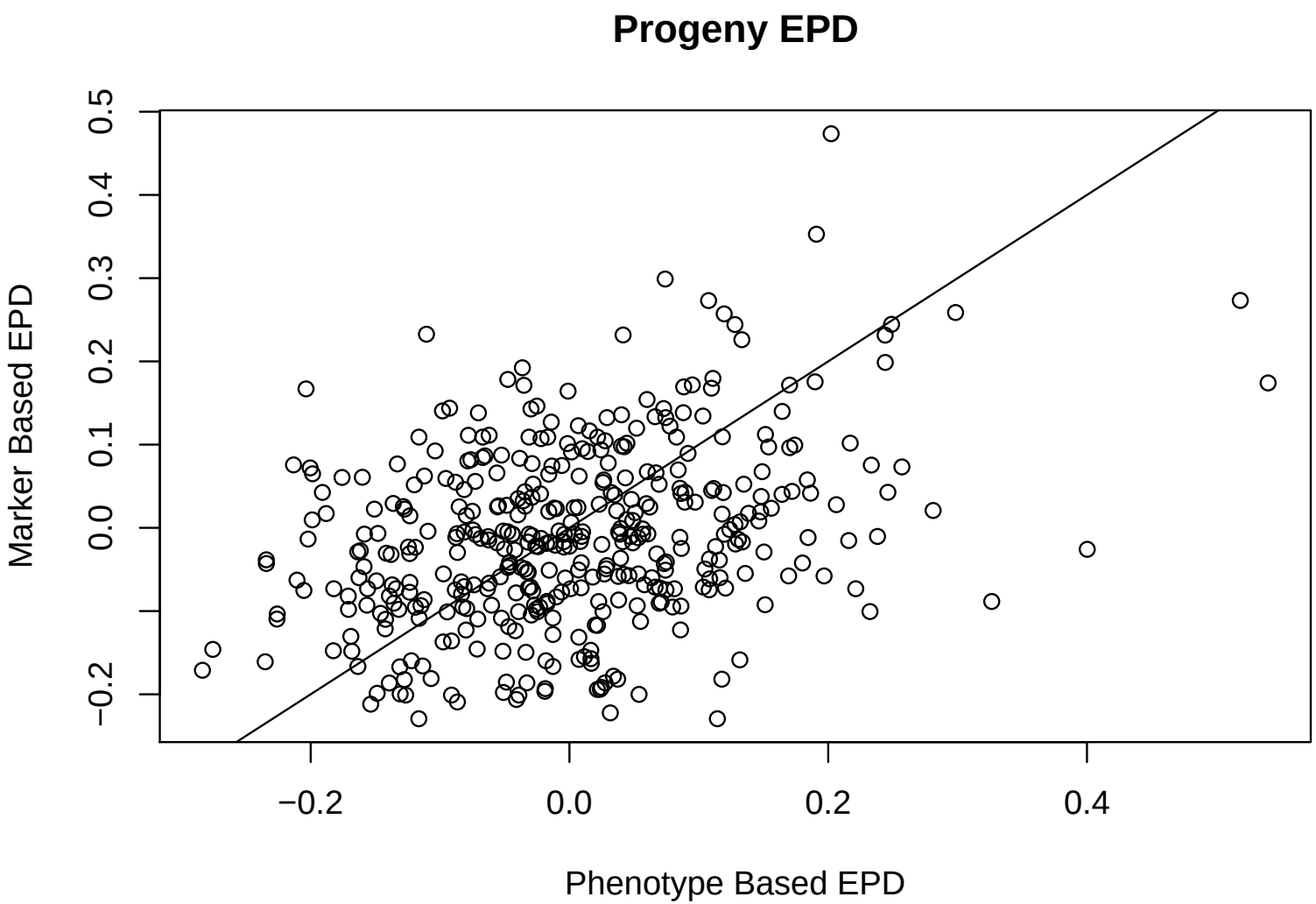
Full versus Marker EPDs (r=0.77)



Phenotypic versus Marker EPDs (r=0.40)



Phenotypic versus Marker EPDs (r=0.34)



- Given the actual genetic parameters
  - Full model produces the BLUP of the EPD
  - Same criteria used for the current phenotypic based evaluations
  - Accuracies and Interim EPDs are available using standard approaches

Need for a Reduced Model

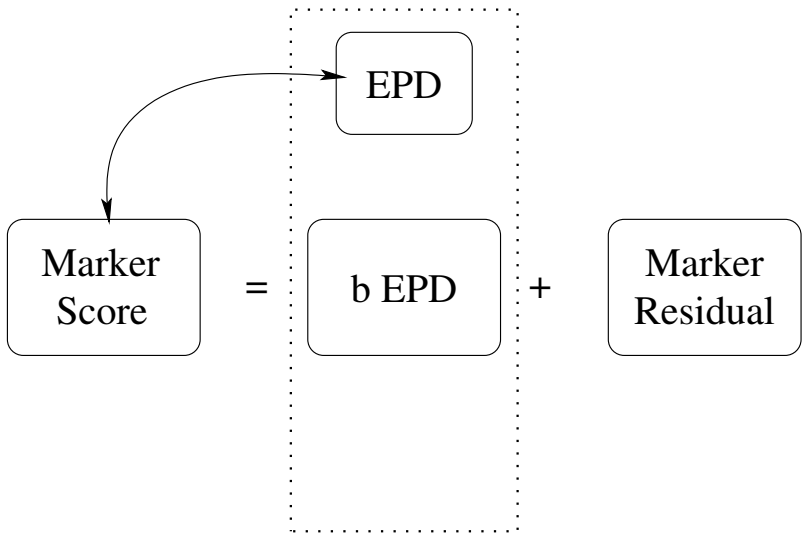
- A direct implementation of this approach would require a new trait whenever a new (or modified) marker score is introduced
  - Increases the computational and memory requirements
- Reduced models will allow for the evolution of marker scores while keeping the computational and memory requirements within reason

Conceptual framework

- Given the common genetic basis of the marker scores it is expected that they will share a number of common features
- Partition a marker score into two independent components

Conceptual framework

- Given the common genetic basis of the marker scores it is expected that they will share a number of common features
- Partition a marker score into two independent components
  1. Component associated with the true EPD of the trait
  2. Residual component



EPD Component

- Each marker score will have its own weight component (b)
  - Function of the genetic correlation between the marker score and the phenotypic trait.
- Allows for different units between the marker score and phenotypic trait.
- Allows for multiple marker scores on an individual animal.
- Maintains the sparsity and the size of the estimating equations.

Residual Component

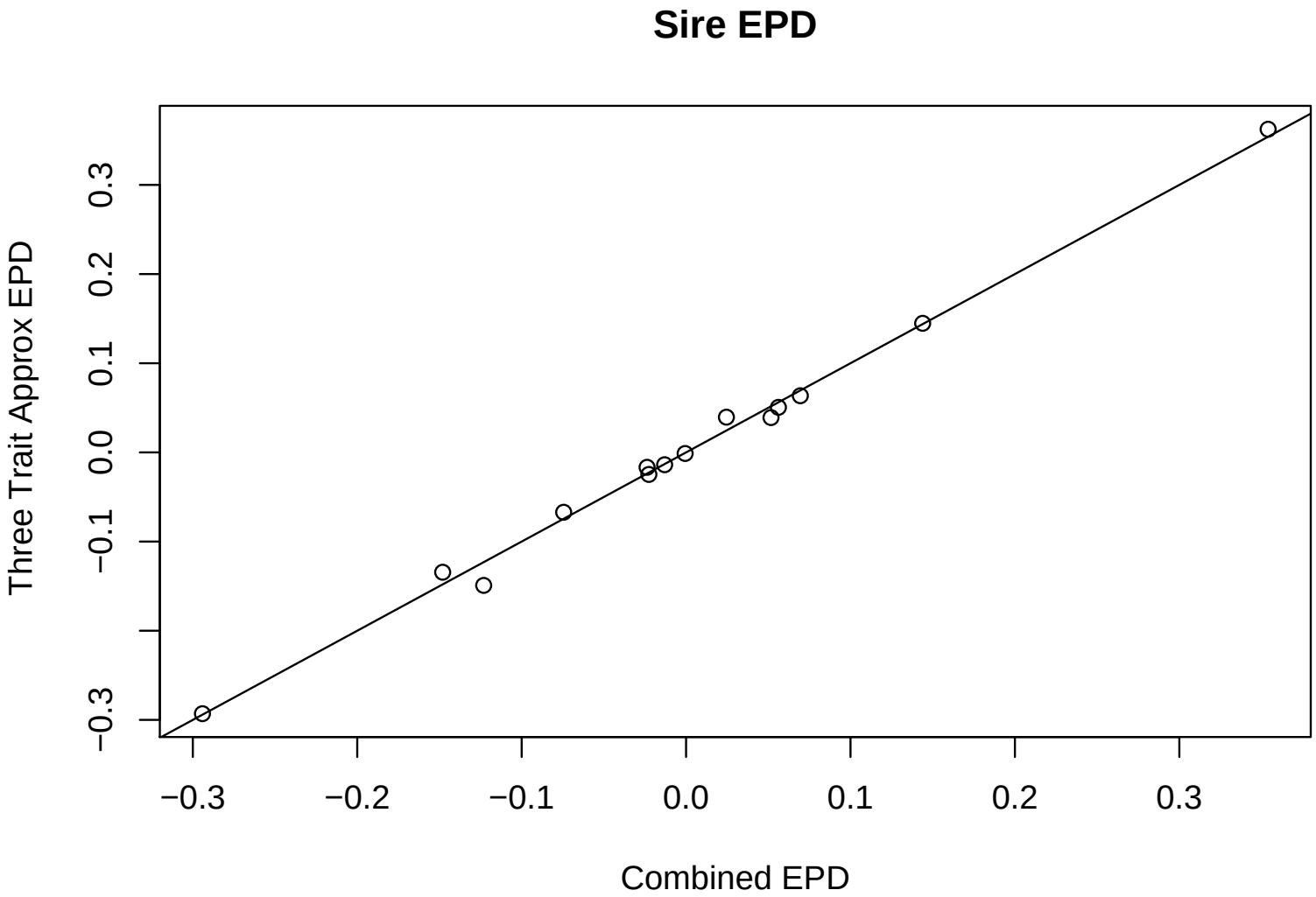
- Models the proportion of variability in the true EPD not accounted for by the marker scores.
  - The greater the correlation between a marker score and the EPD the smaller the residual variance.
- The correlation in the residual components accounts for redundant information in multiple marker scores.
  - Allow a quantification of the gain in accuracy from using multiple marker scores.

- If marker scores are only available on unrelated animals.
  - There will be only a trivial impact on the computational requirements.
- For related animals
  - Introduces an extra trait for each marker
- However, because of the independence of the residual component from the effects of interest
  - Should be possible to reduce the number traits, with a minimal impact on the estimated EPDs

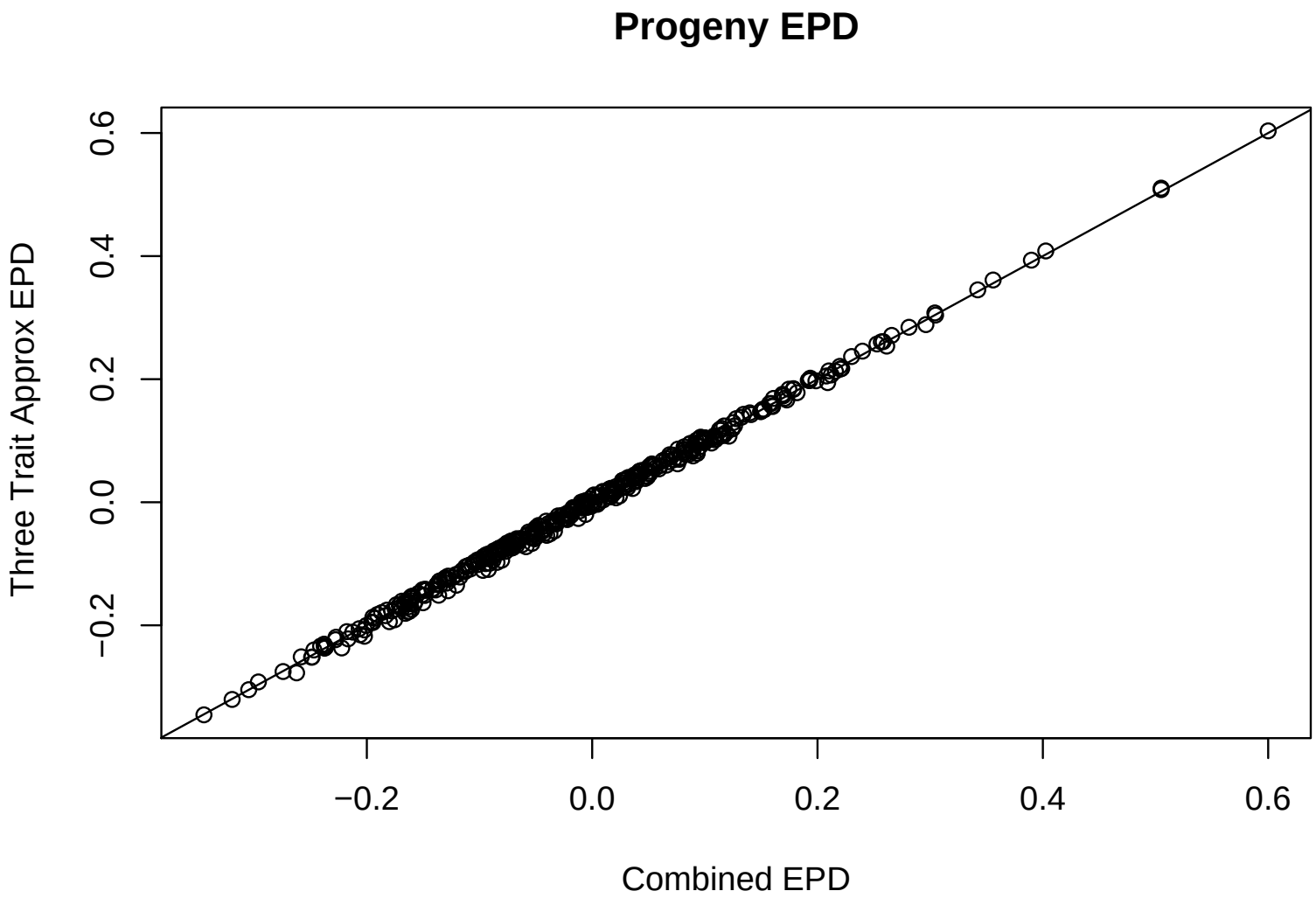
Reduced Models Examined

- Because observations were collected on half-sibs the residual components will be correlated
- Two reduced models were compared to four trait combined model.
  - Three trait model
  - Two trait model

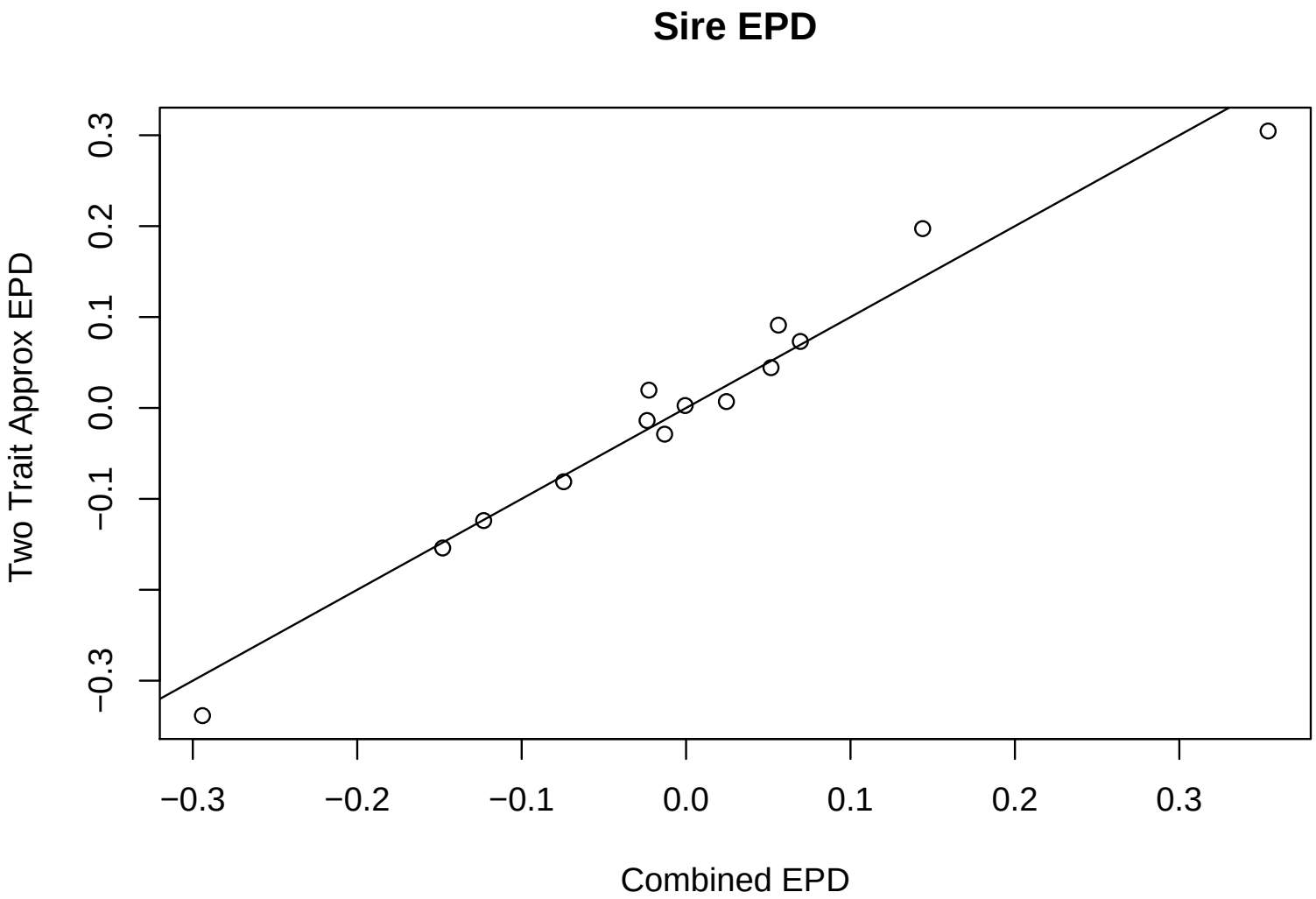
Full versus Three Trait EPDs (r=0.99)



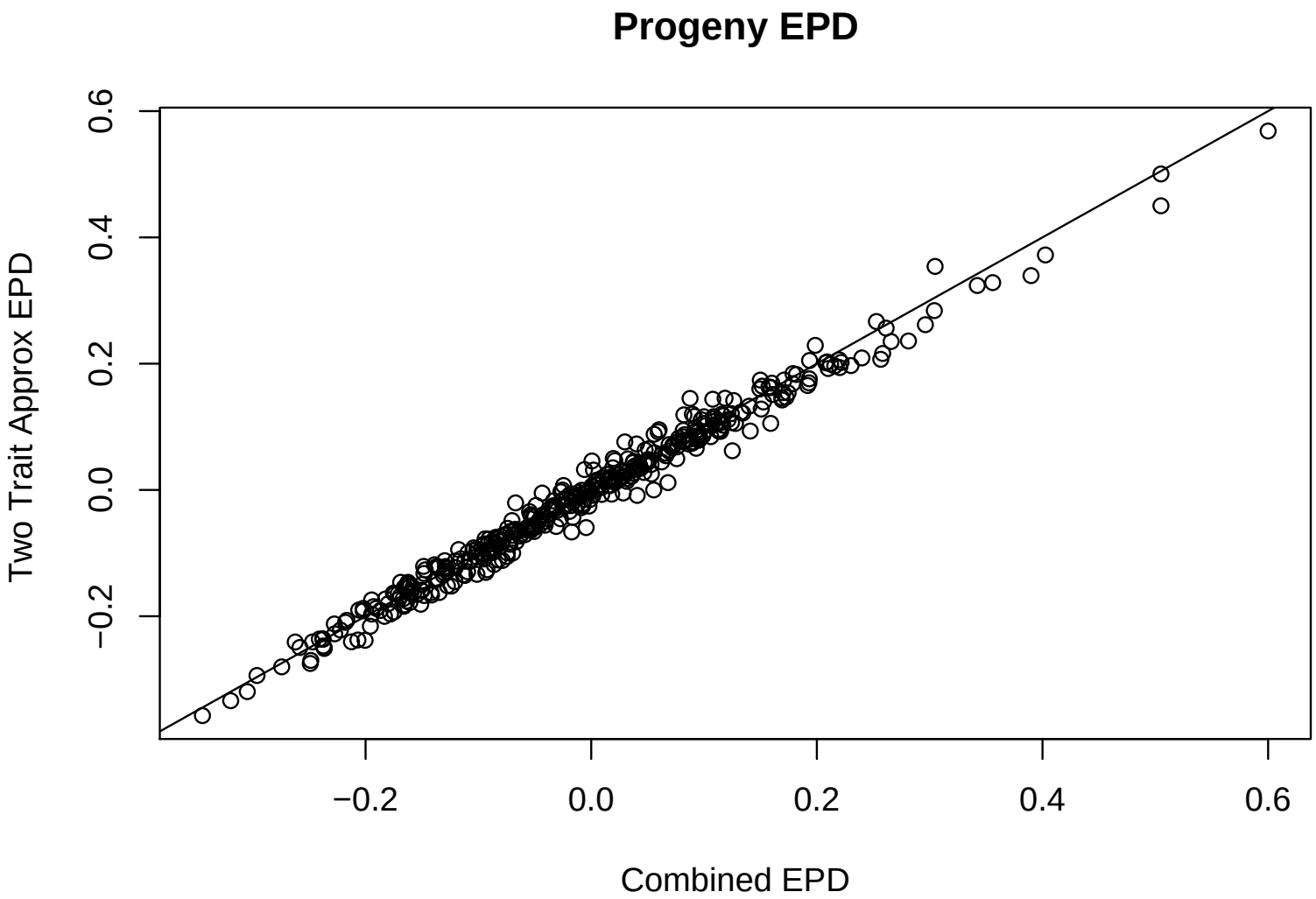
Full versus Three Trait EPDs (r=0.99)



Full versus Two Trait EPDs (r=0.98)

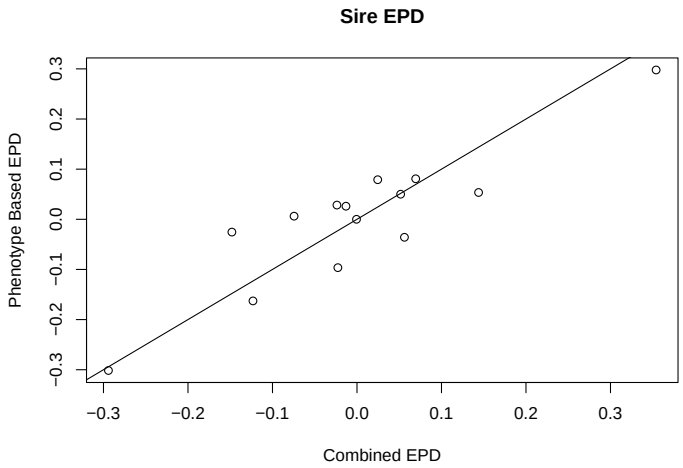


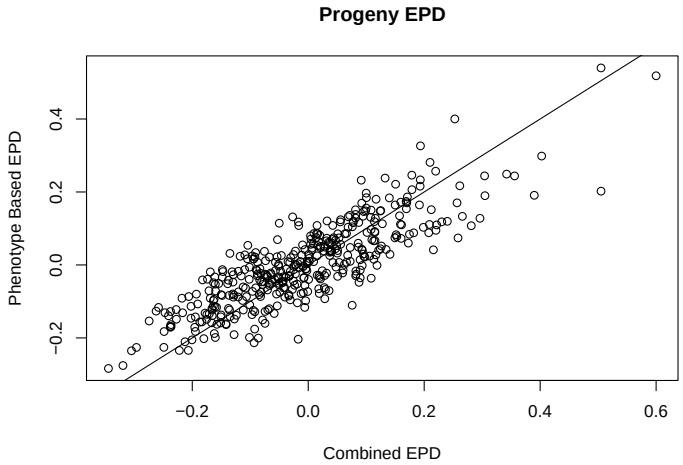
Full versus Two Trait EPDs (r=0.99)



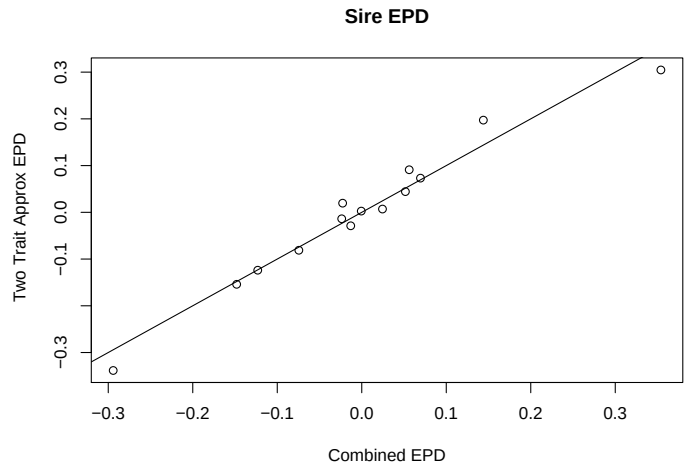
Comparison with the Phenotype Model

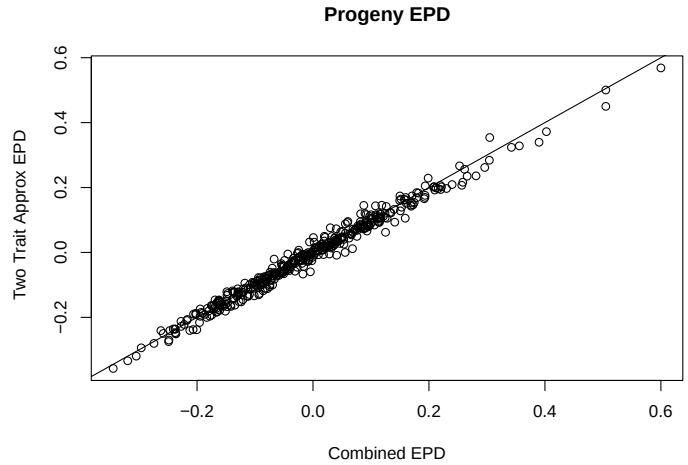
Phenotype Model





Reduced Model





Accuracies

- Calculated using a h^2 of 0.4 and a genetic correlation between the marker scores and shear force of 0.45.
- Single Marker Score with 10 sires and 10 progeny/sire

		Accuracy	
Phenotype	Genotype	Sire	Progeny
Progeny	None	0.27	0.24
Progeny	Progeny	0.28	0.29
Progeny	Sire	0.31	0.25
Progeny	Both	0.33	0.29

Without Phenotype Information

		Accuracy	
Phenotype	Genotype	Sire	Progeny
None	Progeny	0.07	0.11
None	Sire	0.10	0.02
None	Both	0.10	0.11

Single Progeny/Sire

		Accuracy	
Phenotype	Genotype	Sire	Progeny
Progeny	None	0.05	0.20
Progeny	Progeny	0.05	0.25
Progeny	Sire	0.13	0.21
Progeny	Both	0.13	0.25

Without Phenotype Information

		Accuracy	
Phenotype	Genotype	Sire	Progeny
None	Progeny	0.02	0.10
None	Sire	0.10	0.02
None	Both	0.10	0.10

- Once an individual has been genotyped there is minimal benefit in terms of accuracy to genotyping relatives.
  - Indirect benefit, if the genotyped ancestor has phenotypic information.
- Greatest benefit from genotyping is for animals with limited phenotypic information.

Interim EPDs

- Simplified, an interim EPD:
  - Takes the EPD from the current evaluation
    - Typically based on the parental EPDs
  - Along with new information
    - Typically based on an individual's own adjusted record deviated from the adjusted records of its contemporaries,
  - The new information is then given a weight and added to the individual's current EPD to produce an interim EPD.

Marker Score Interim EPDs

- A marker score on an individual is a new piece of information
  - Marker score would need to be adjusted in the same way a phenotypic record is adjusted
  - Using the appropriate weight the adjusted marker score is added to the current EPD to produce a marker score interim EPD.
    - The weight will be a function of the genetic variances and covariances.

- Methodology is in place.
- Methodology is based on a robust and familiar statistical foundation.
  - EPDs, Accuracies, and Interim EPDs will be available
- Extension of the current approach to genetic evaluation.
  - Can make use of lessons already learned

What needs to be done?

- Estimation of genetic parameters
  - Resource populations for estimation of genetic parameters
- Reporting criteria
  - Evaluate the effect of selective reporting
- Criteria for determining when a marker score is ready to be included
  - Step beyond validation
- Evaluate the trade off between computational requirements and model complexity
- Software development