F:\sarah\fri_MV

Multivariate Linkage and Association

Sarah Medland and Manuel Ferreira

-with heavy borrowing from

Kate Morley and Frühling Rijsdijk

MV analyses can address

- Questions of common aetiology
  - Same gene (snp)
  - Co-incidental covariation due to LD between two different genes
  - Co-variation due to shared social/environmental risk factors

Aust. Albatross

MV analyses can address

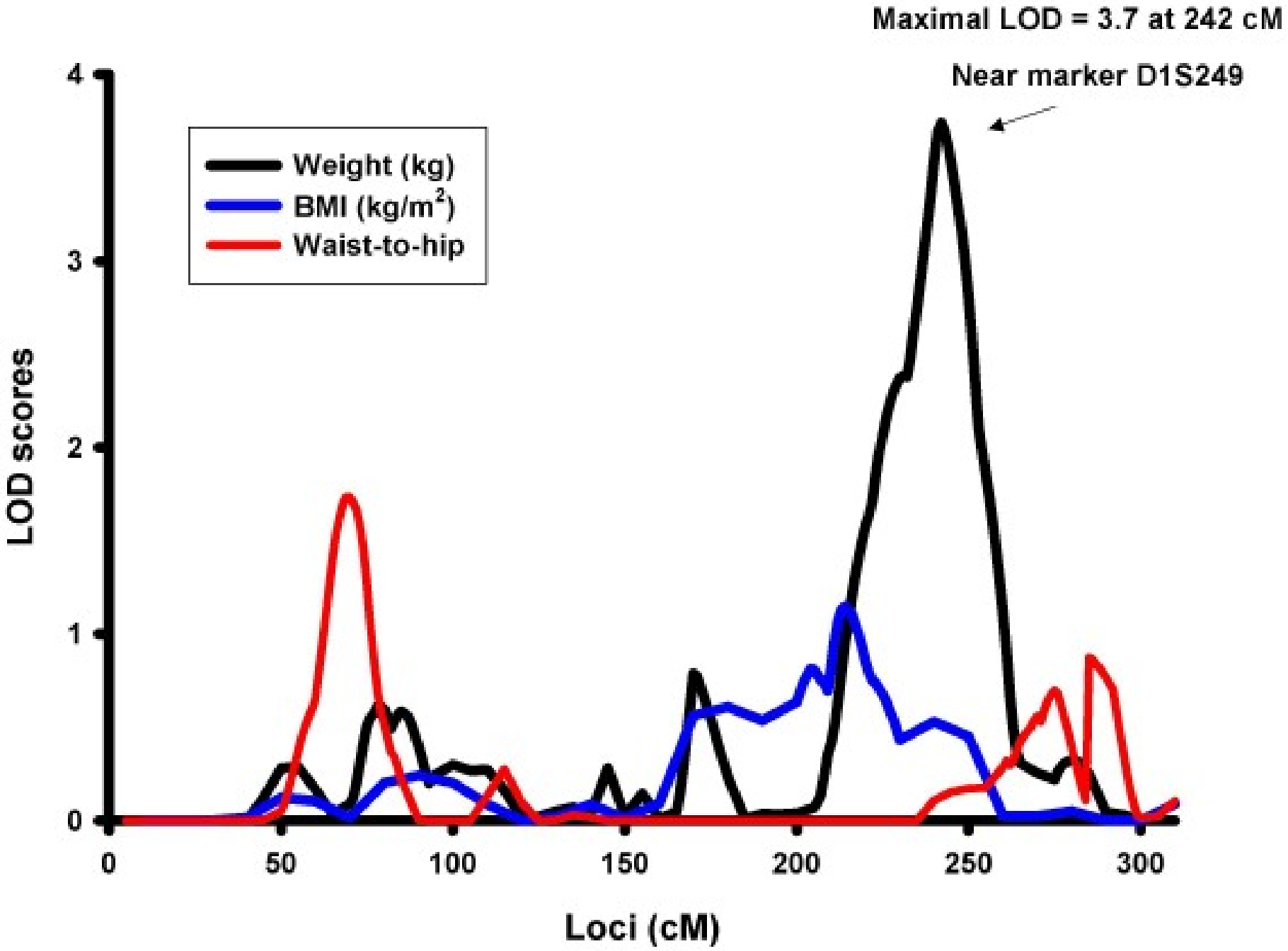
- Questions of common aetiology
  - Same gene (snp)
  - Co-incidental covariation due to LD between two different genes
  - Co-variation due to shared social/environmental risk factors
- **Pleiotropy** occurs when a single gene influences multiple phenotypic traits.

Bilby

Studying multiple phenotypes…

- Run multiple univariate analyses on correlated traits

Bandicoot


Maximal LOD = 3.7 at 242 cM
Near marker D1S249
Weight (kg)
BMI (kg/m2)
Waist-to-hip
LOD scores
0
1
2
3
4
0
50
100
150
200
250
300
Loci (cM)

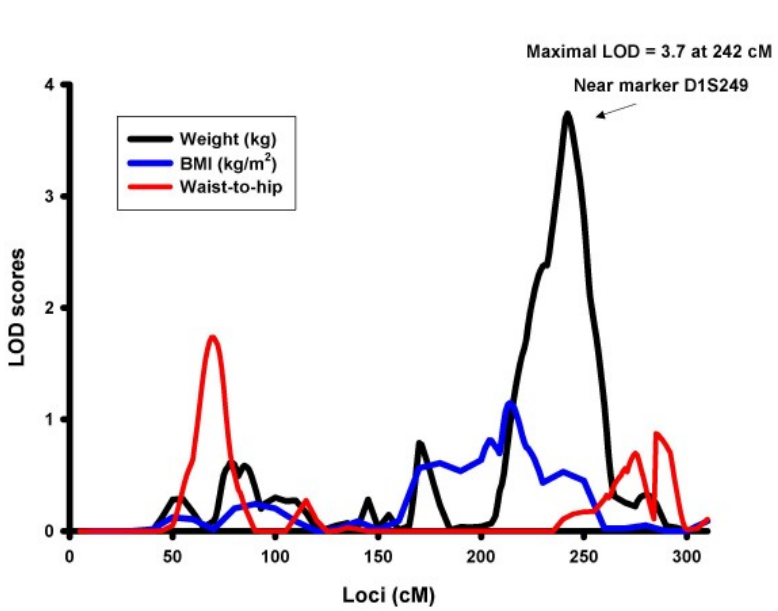
Studying multiple phenotypes…

- Run multiple univariate analyses
  - Correct for multiple testing…
    - Bonferroni ☹
    - Correction for equivalent number of independent variables
  - Doesn't really address the idea of common aetiology

Blue ringed octopus

Studying multiple phenotypes…

- Run multiple univariate analyses
  - Try and determine if the coincident linkage/association is statistically unlikely



Box jellyfish

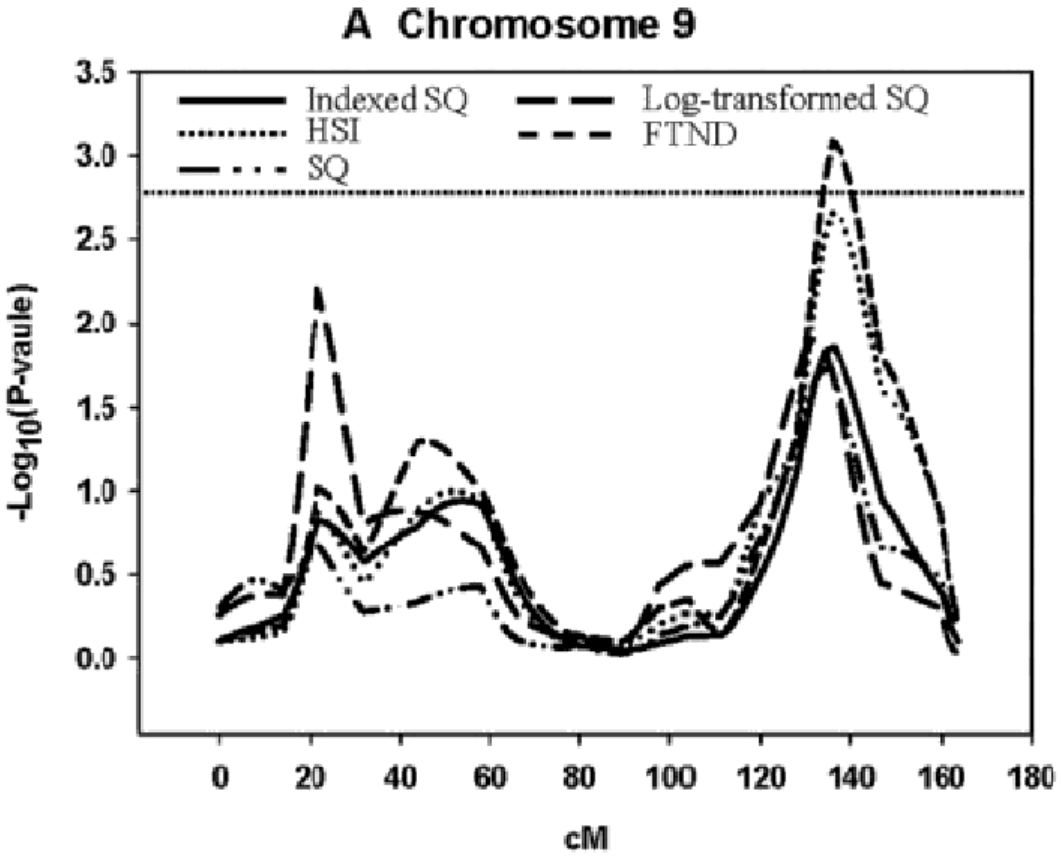
Studying multiple phenotypes…

- Run multiple univariate analyses
  - Try and determine if the coincident linkage/association is statistically unlikely
    - Simulate/Permute data and assess how often this group of traits reaches this pattern of sig. by chance

Brown Snake

Studying multiple phenotypes…

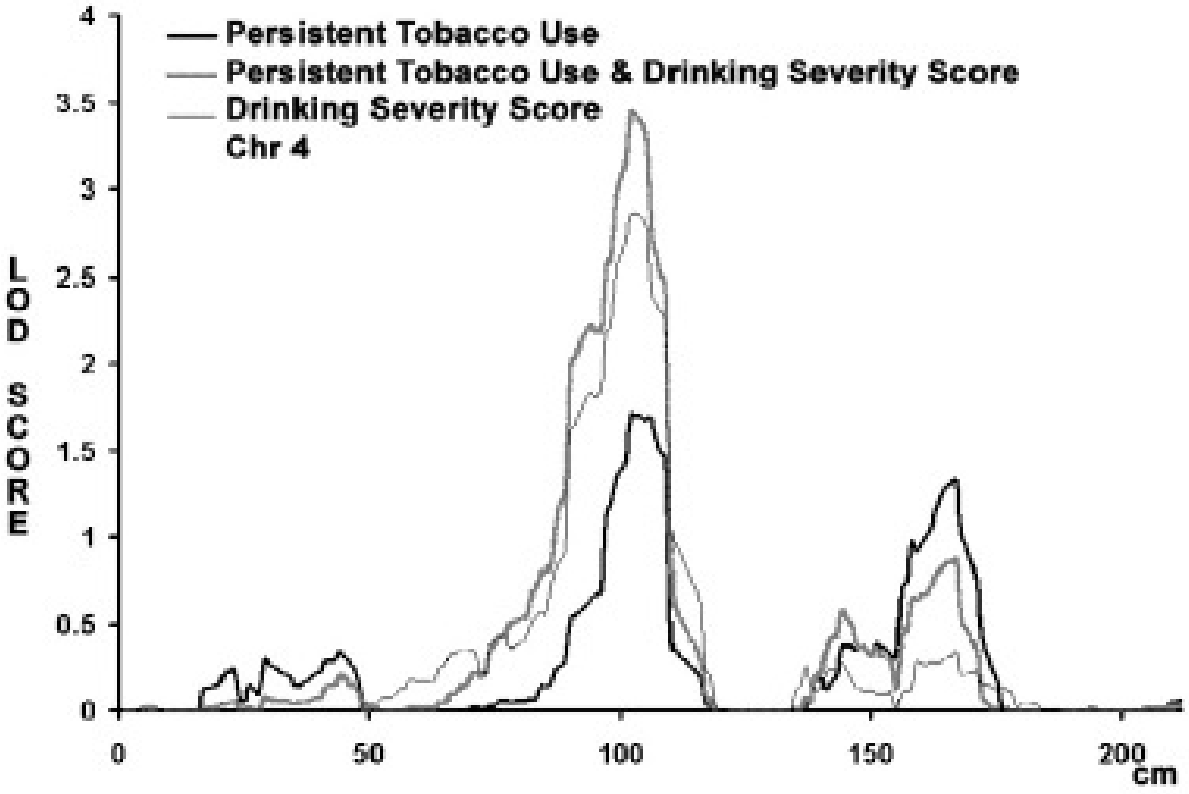
- Study a number of proxies ☹



Cockatoo

Studying multiple phenotypes…

- Make a composite phenotype ☹



Cuscus

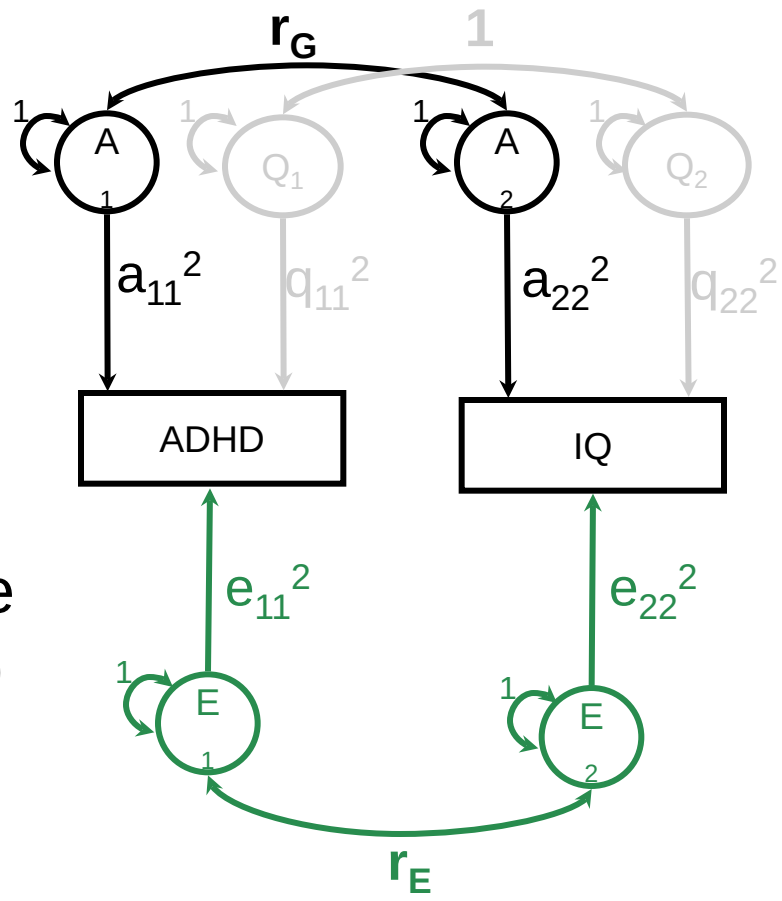
Studying multiple phenotypes…

- Make a factor score
  - combine both factor level and trait-specific effects
  - latent factor effects are inherently pleiotropic
  - residual effects are not
  - assumes factor loadings are constant across genome

Cuscus (again)

Alternative…

- Explicitly model the covariation between traits
  - Traits may be correlated due to shared genetic factors (A) or shared environmental factors (C or E)



Dingo

Explicitly model the covariation between traits…

- Directly assess pleiotropy
- Increased power
  - Esp. when the pattern of QTL effects is different from the background covariation
    - ie positive *r* but gene effects are in opposite directions
- Reduced multiple testing

Drop Bear

Multivariate analysis…

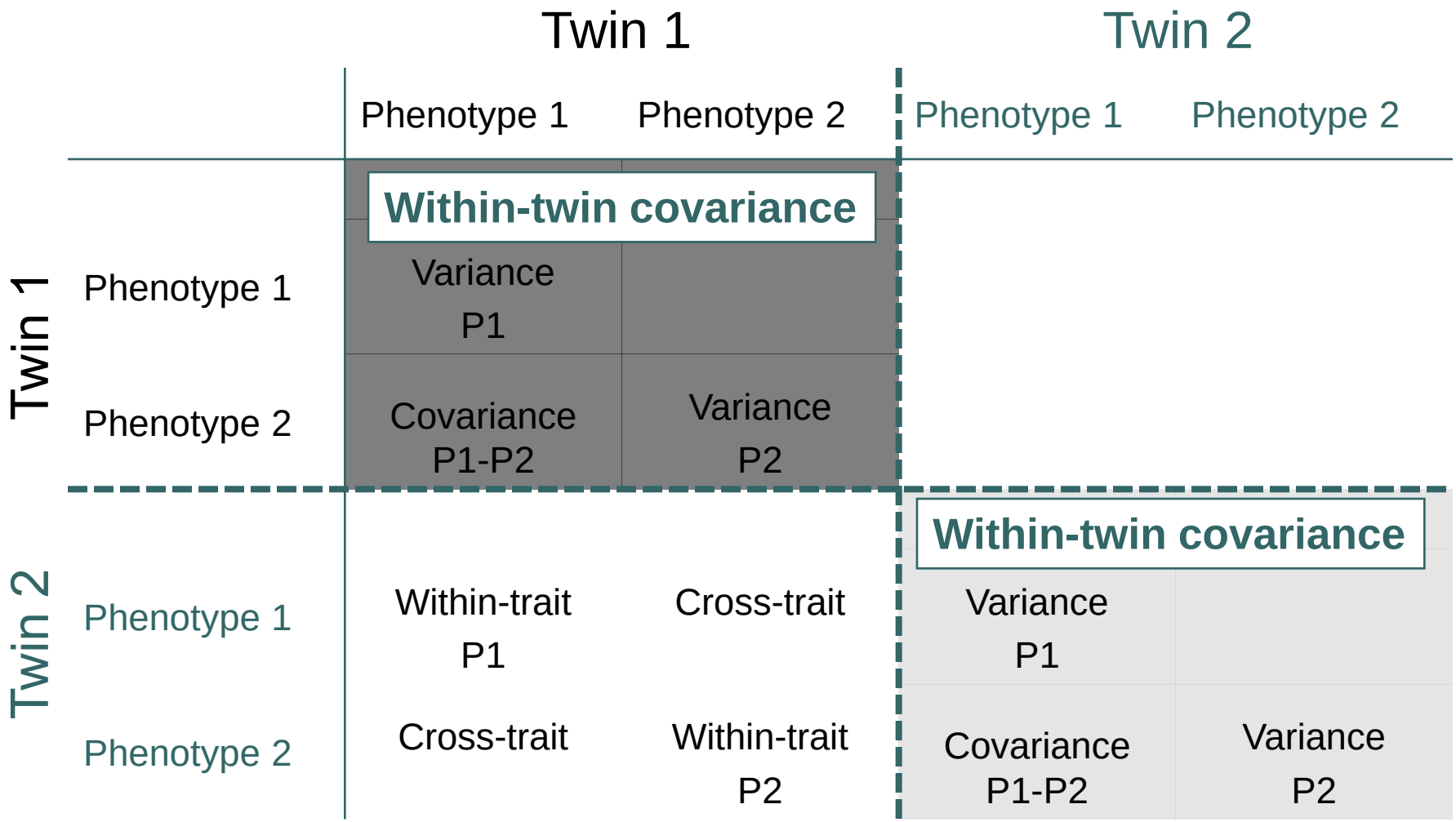
- In the context of linkage analysis when we have family data
  - two traits measured in twin pairs
  - Interested in:
    - Cross-trait covariance *within* individuals
    - Cross-trait covariance *between* twins
    - MZ:DZ ratio of cross-trait covariance between twins

Echidna

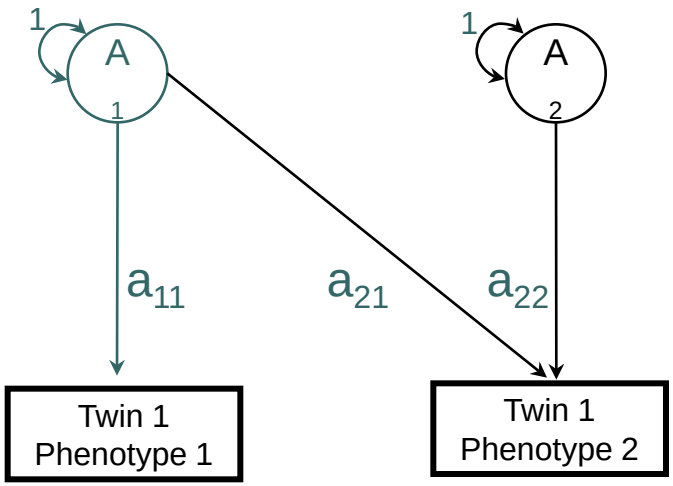
Observed Covariance Matrix

		Twin 1		Twin 2	
		Phenotype 1	Phenotype 2	Phenotype 1	Phenotype 2
Twin 1	Phenotype 1	Variance P1			
	Phenotype 2	Covariance P1-P2	Variance P2		
Twin 2	Phenotype 1	Within-trait P1	Cross-trait	Variance P1	
	Phenotype 2	Cross-trait	Within-trait P2	Covariance P1-P2	Variance P2

Observed Covariance Matrix



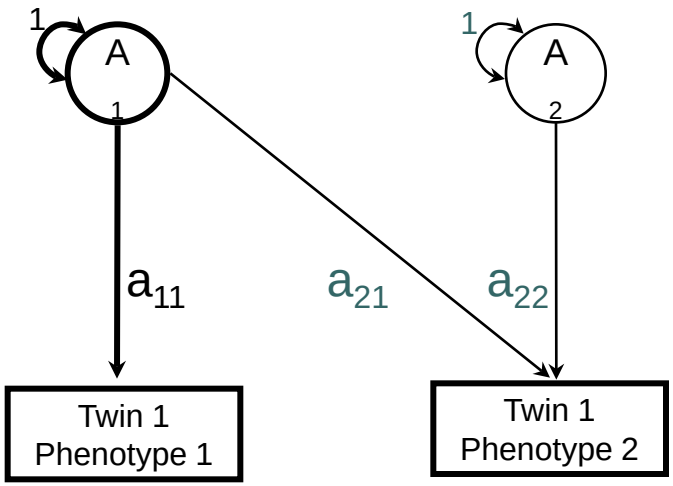
Within-Twin Covariances (A)



$$\begin{bmatrix} a_{11} & \\ a_{21} & a_{22} \end{bmatrix}$$

Twin 1	Phenotype 1	Phenotype 2
Phenotype 1		
Phenotype 2		

Within-Twin Covariances (A)

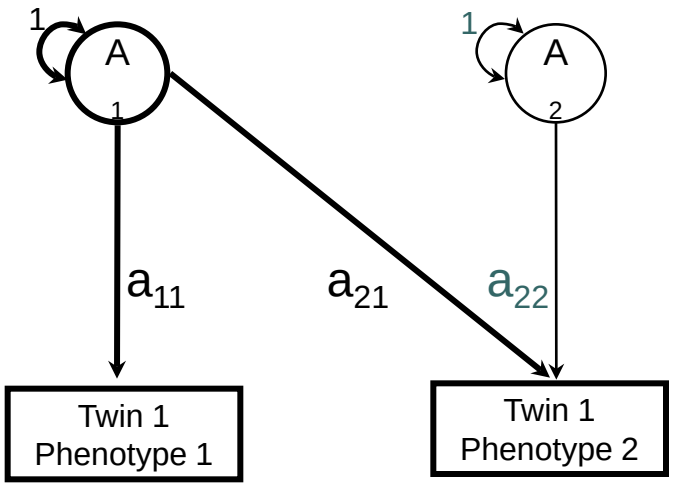


$$\begin{bmatrix} a_{11} & \\ a_{21} & a_{22} \end{bmatrix}$$

Twin 1

Twin 1	Phenotype 1	Phenotype 2
Phenotype 1	a_{11}^2	
Phenotype 2		

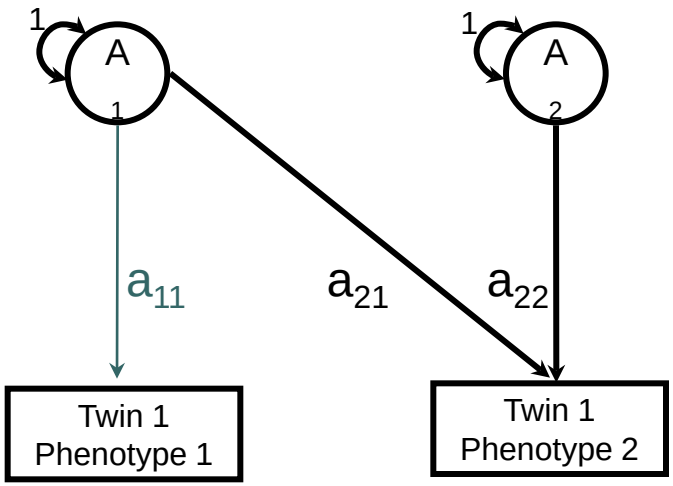
Within-Twin Covariances (A)



$$\begin{bmatrix} a_{11} & \\ a_{21} & a_{22} \end{bmatrix}$$

	Twin 1	
Twin 1	Phenotype 1	Phenotype 2
Phenotype 1	a_{11}^2	
Phenotype 2	$a_{11}a_{21}$	

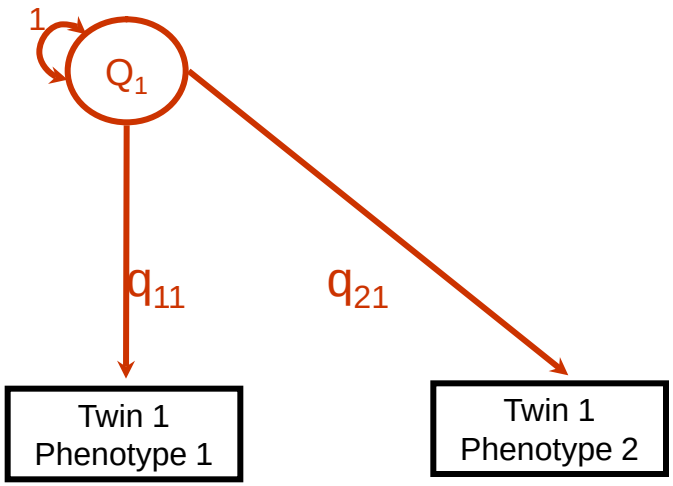
Within-Twin Covariances (A)



$$\begin{bmatrix} a_{11} & \\ a_{21} & a_{22} \end{bmatrix}$$

		Twin 1	
		Phenotype 1	Phenotype 2
Twin 1	Phenotype 1	a_{11}^2	
	Phenotype 2	$a_{11}a_{21}$	$a_{22}^2+a_{21}^2$

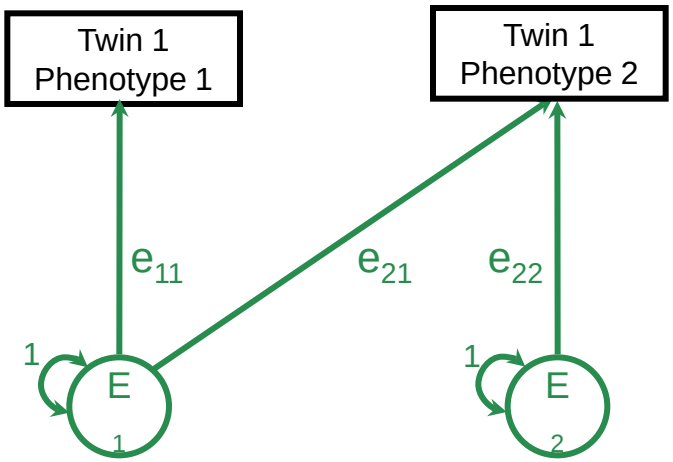
Within-Twin Covariances (Q)



$$\begin{bmatrix} q_{11} \\ q_{21} \end{bmatrix}$$

Twin 1	Phenotype 1	Phenotype 2
Phenotype 1	$a_{11}^2+q_{11}^2$	
Phenotype 2	$a_{11}a_{21}+q_{11}q_{21}$	$a_{22}^2+a_{21}^2+q_{21}^2$

Within-Twin Covariances (E)

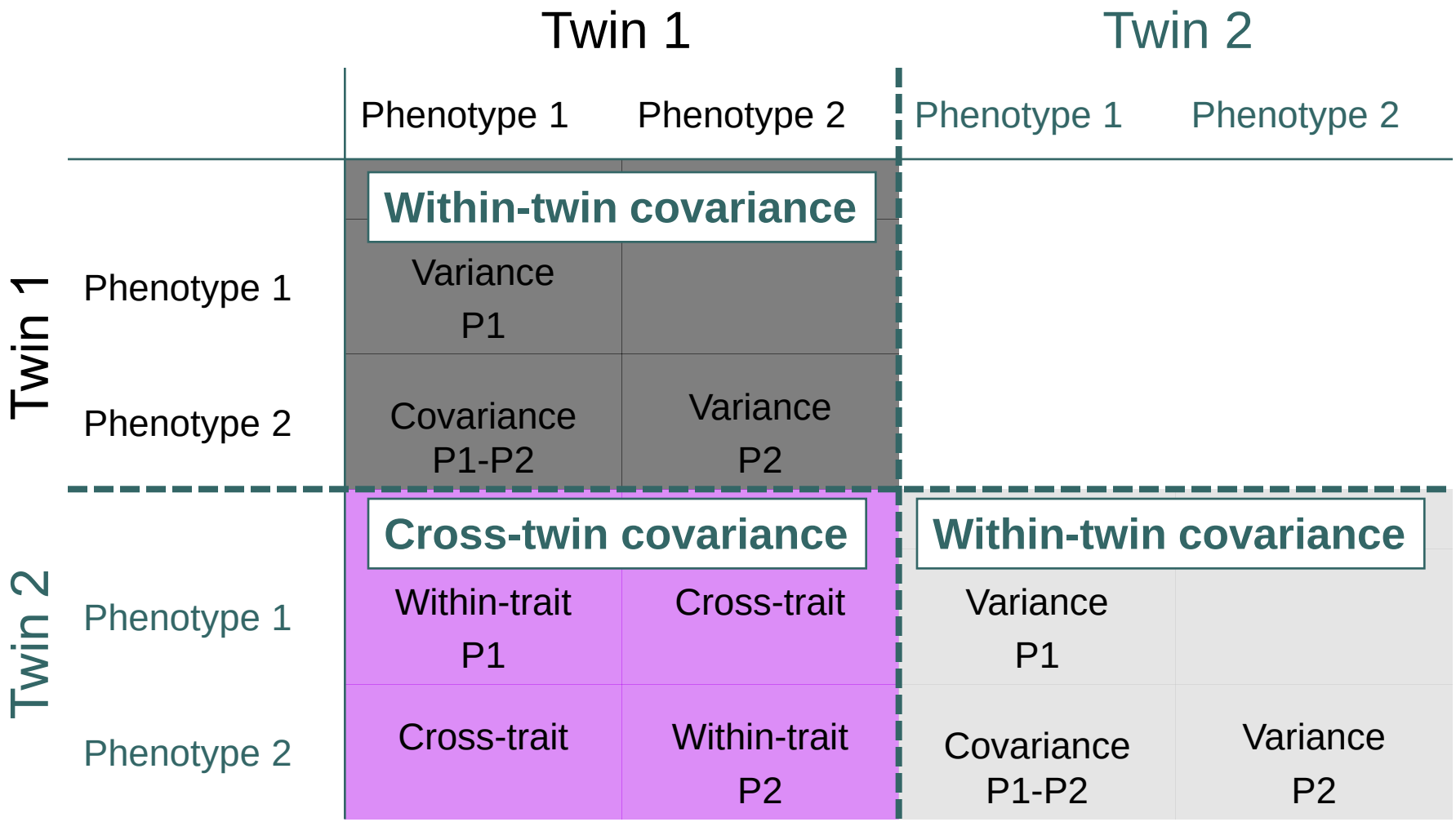


$$\begin{bmatrix} e_{11} & \\ e_{21} & e_{22} \end{bmatrix}$$

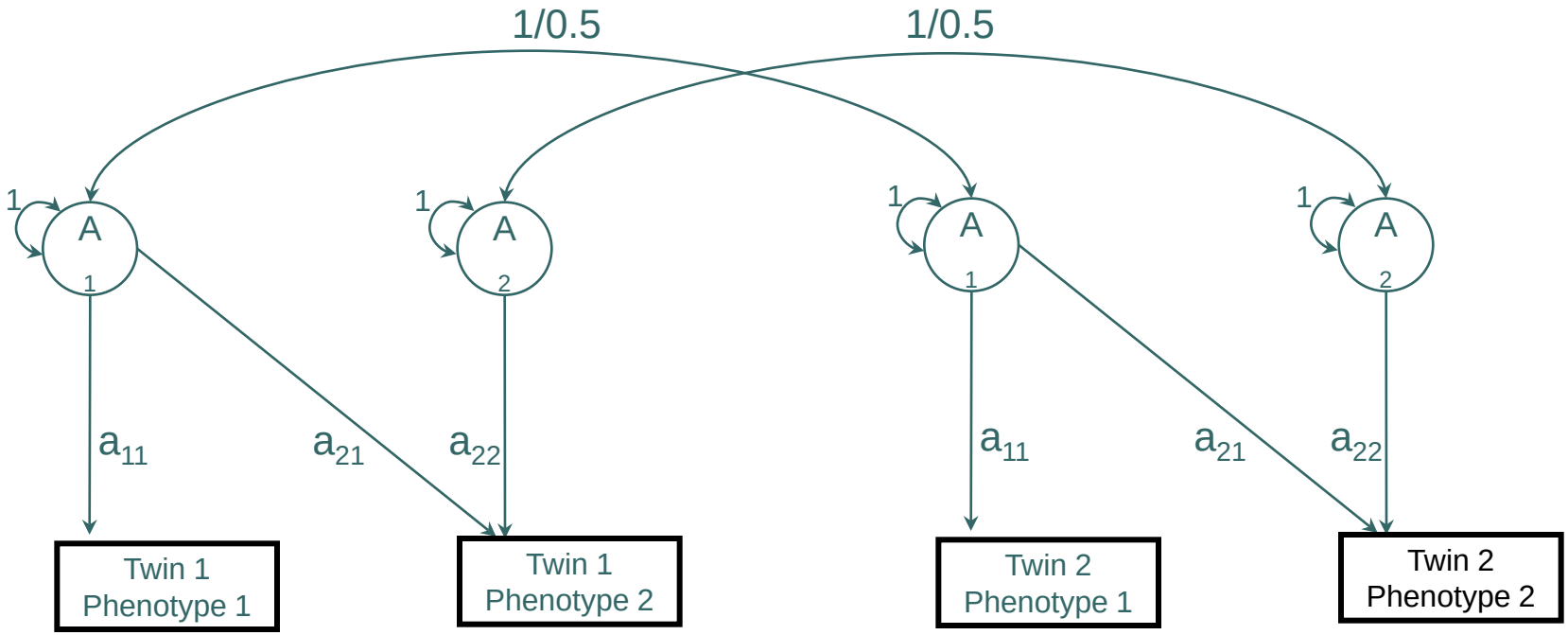
Twin 1

Twin 1	Phenotype 1	Phenotype 2
Phenotype 1	$a_{11}^2+q_{11}^2+e_{11}^2$	
Phenotype 2	$a_{11}a_{21}+q_{11}q_{21}+e_{11}e_{21}$	$a_{22}^2+a_{21}^2+q_{21}^2+e_{22}^2+e_{21}^2$

Observed Covariance Matrix

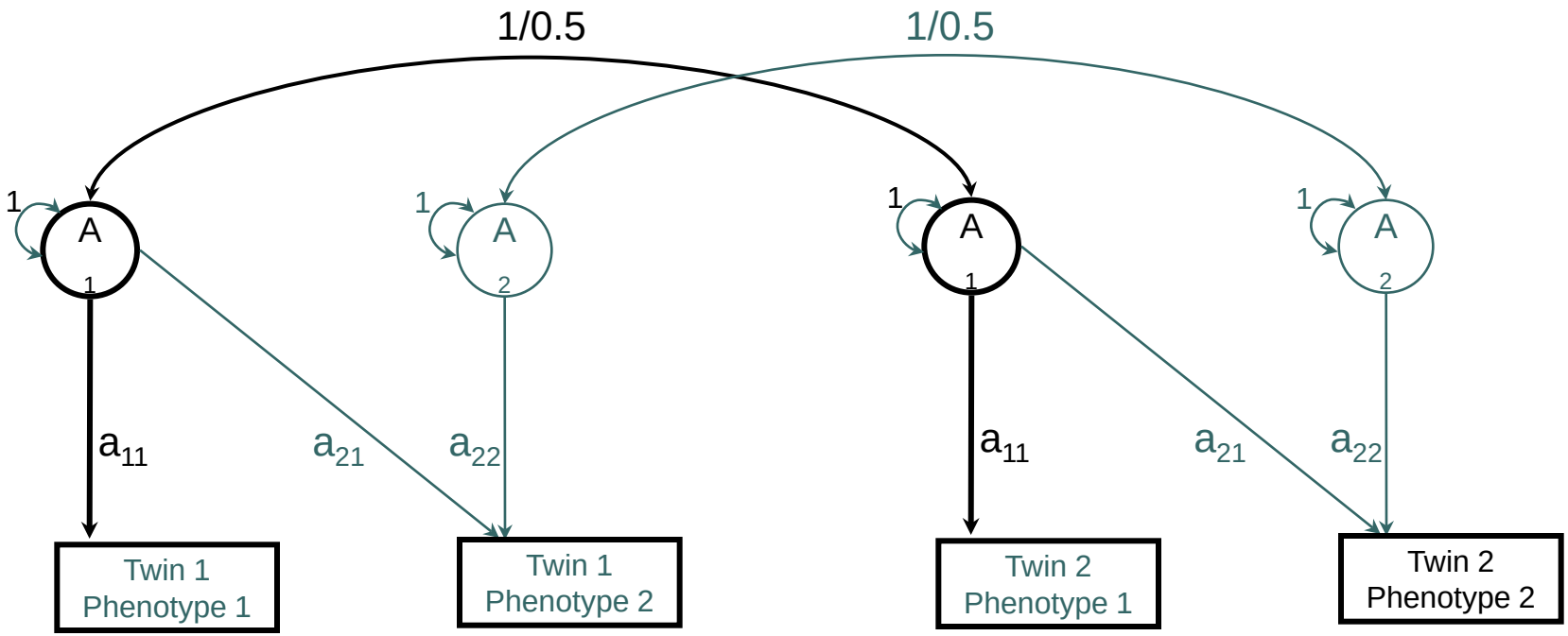


Cross-Twin Covariances (A)



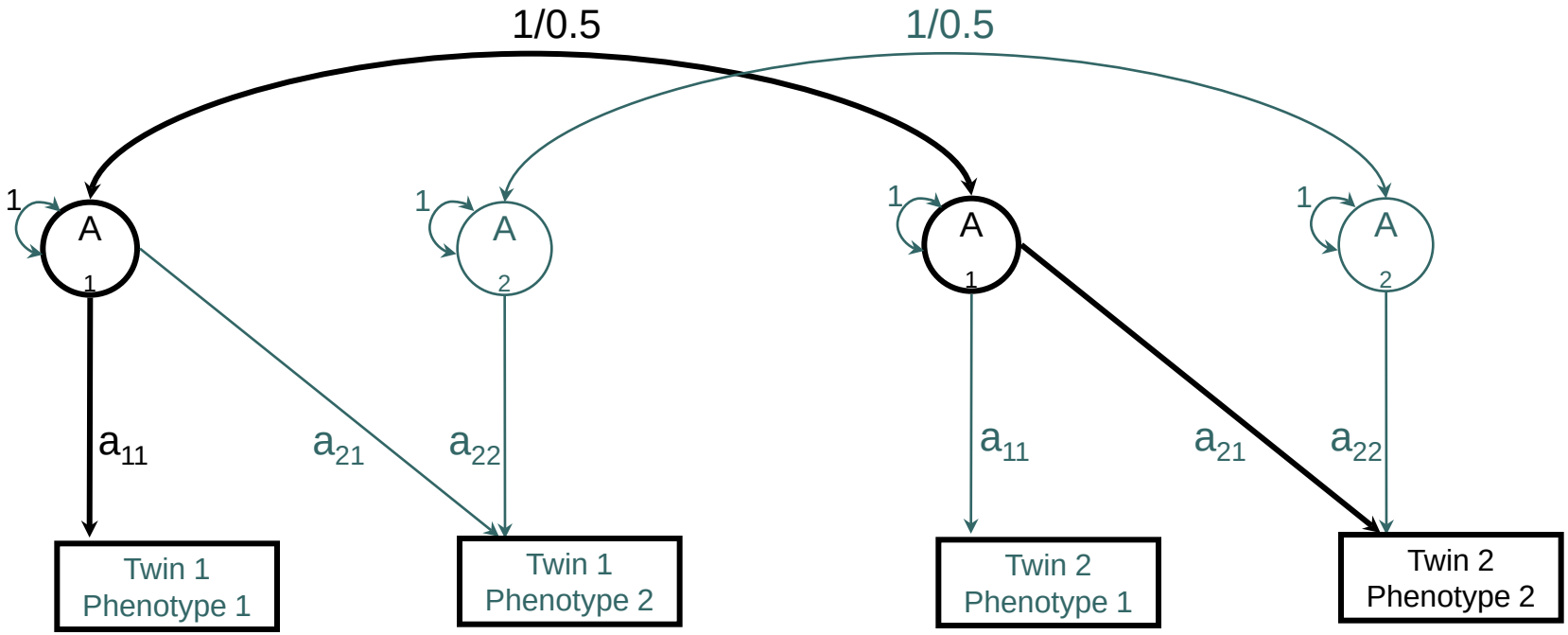
Twin 2 \ Twin 1	Phenotype 1	Phenotype 2
Phenotype 1		
Phenotype 2		

Cross-Twin Covariances (A)



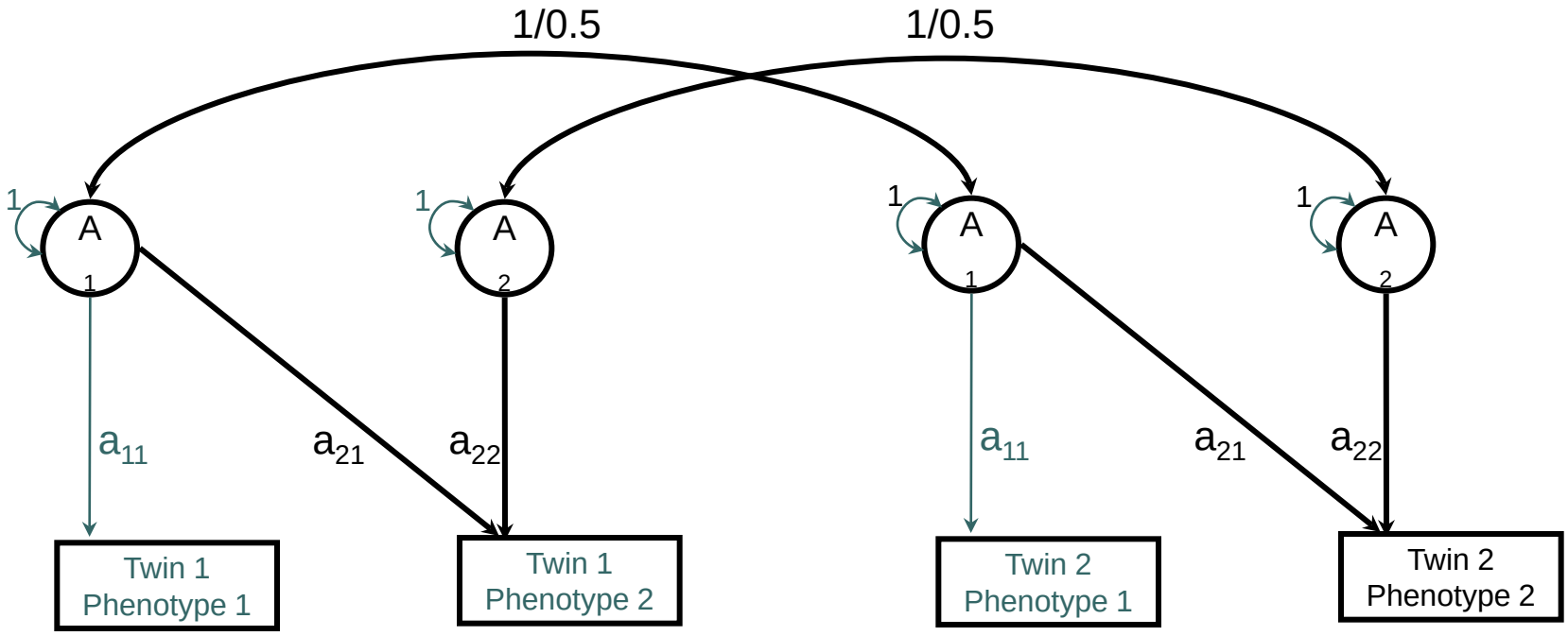
Twin 2 \ Twin 1	Phenotype 1	Phenotype 2
Phenotype 1	$1/0.5a_{11}^2$	
Phenotype 2		

Cross-Twin Covariances (A)



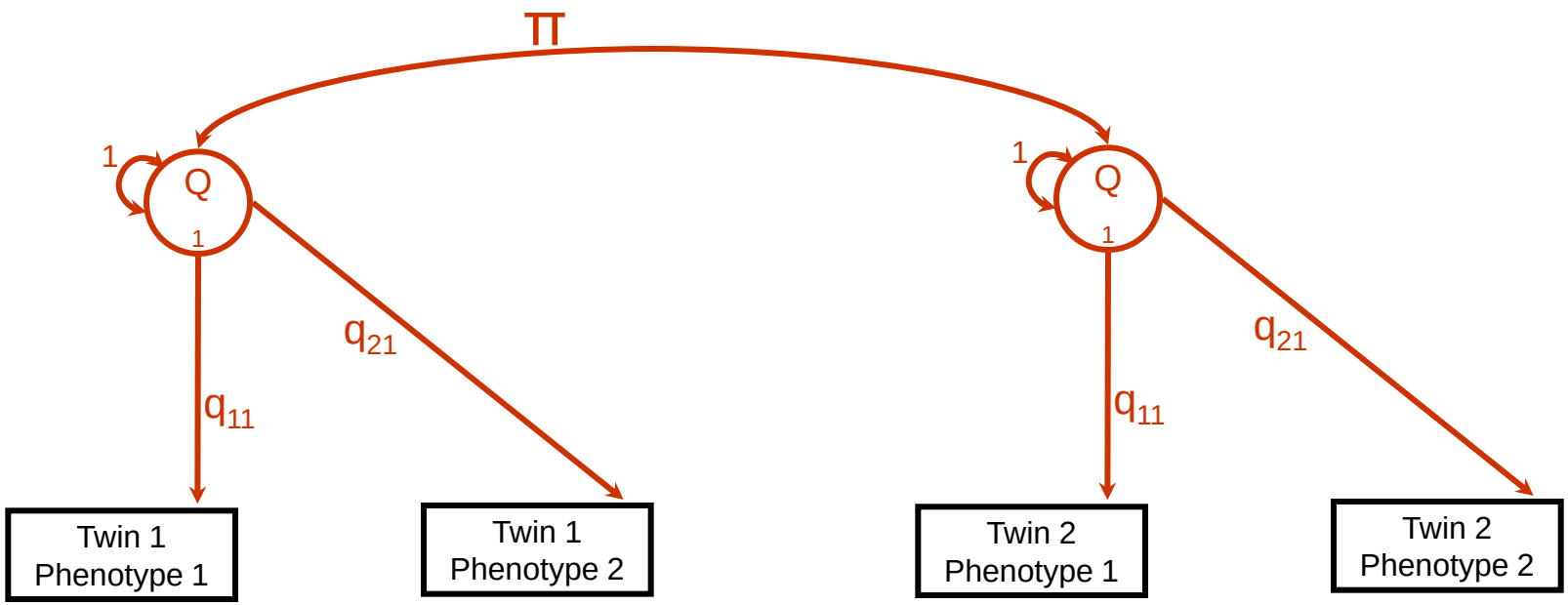
Twin 2 \ Twin 1	Phenotype 1	Phenotype 2
Phenotype 1	$1/0.5a_{11}^2$	
Phenotype 2	$1/0.5a_{11}a_{21}$	

Cross-Twin Covariances (A)



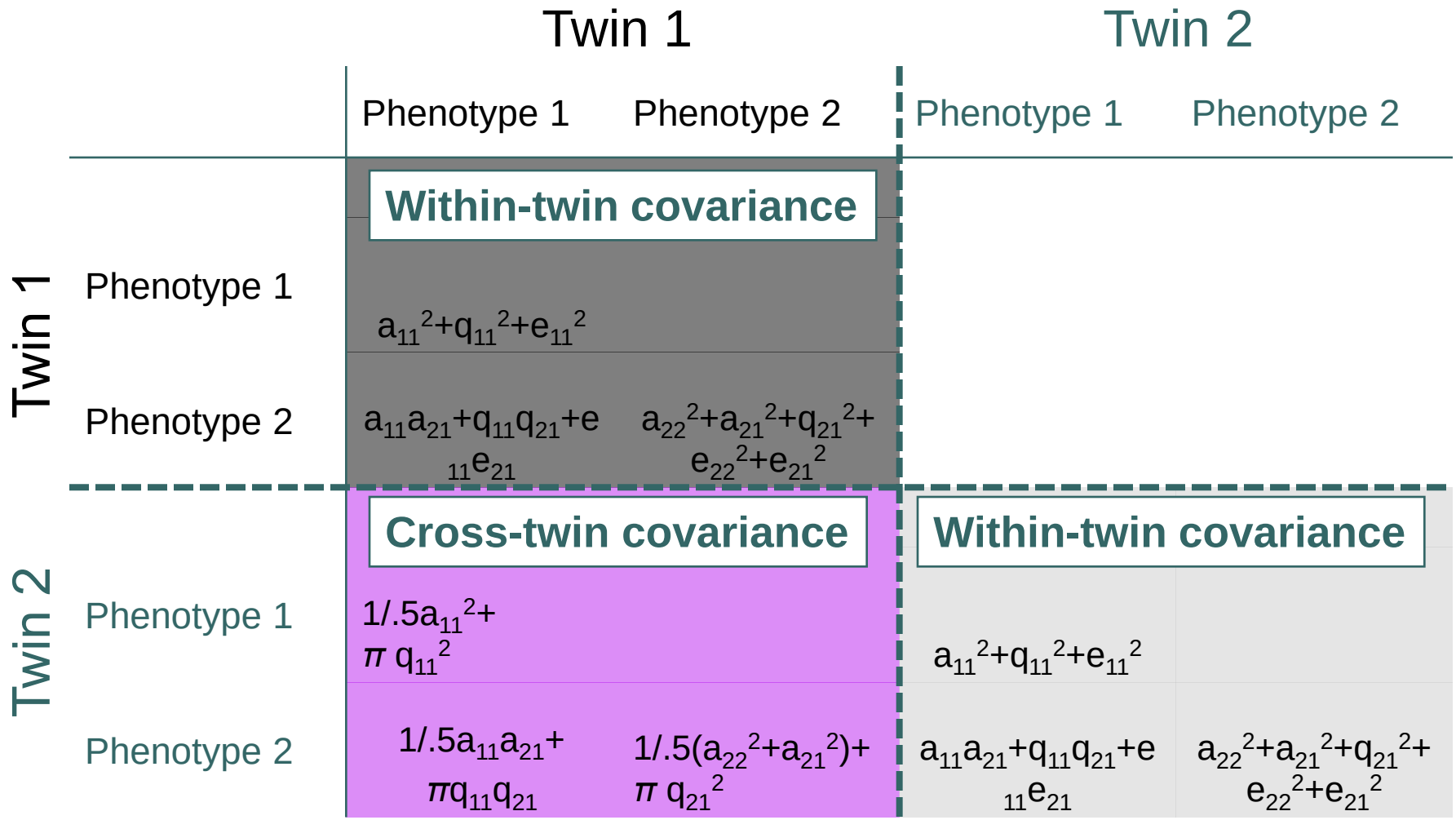
Twin 2 \ Twin 1	Phenotype 1	Phenotype 2
Phenotype 1	$1/0.5a_{11}^2$	
Phenotype 2	$1/0.5a_{11}a_{21}$	$1/0.5(a_{22}^2+a_{21}^2)$

Cross-Twin Covariances (Q)



Twin 2	Twin 1 Phenotype 1	Twin 1 Phenotype 2
Phenotype 1	$1/0.5a_{11}^2+\pi q_{11}^2$	
Phenotype 2	$1/0.5a_{11}a_{21}+\pi\, q_{11}q_{21}$	$1/0.5(a_{22}^2+a_{21}^2)+\pi\, q_{21}^2$

Predicted Model



Running MV linkage analysis…

- Numerous programs
  - Mx
  - Solar
  - Loki
  - Merlin
    - Repeated measures
- Computationally intensive
- Multiple boundary issues
  - p-values difficult to obtain

Emu (not e-moo)

MV association…

- Maximum likelihood – factor based approach
  - Mx
- Canonical Correlation approach
  - Plink
- Principal components approach
  - F-bat

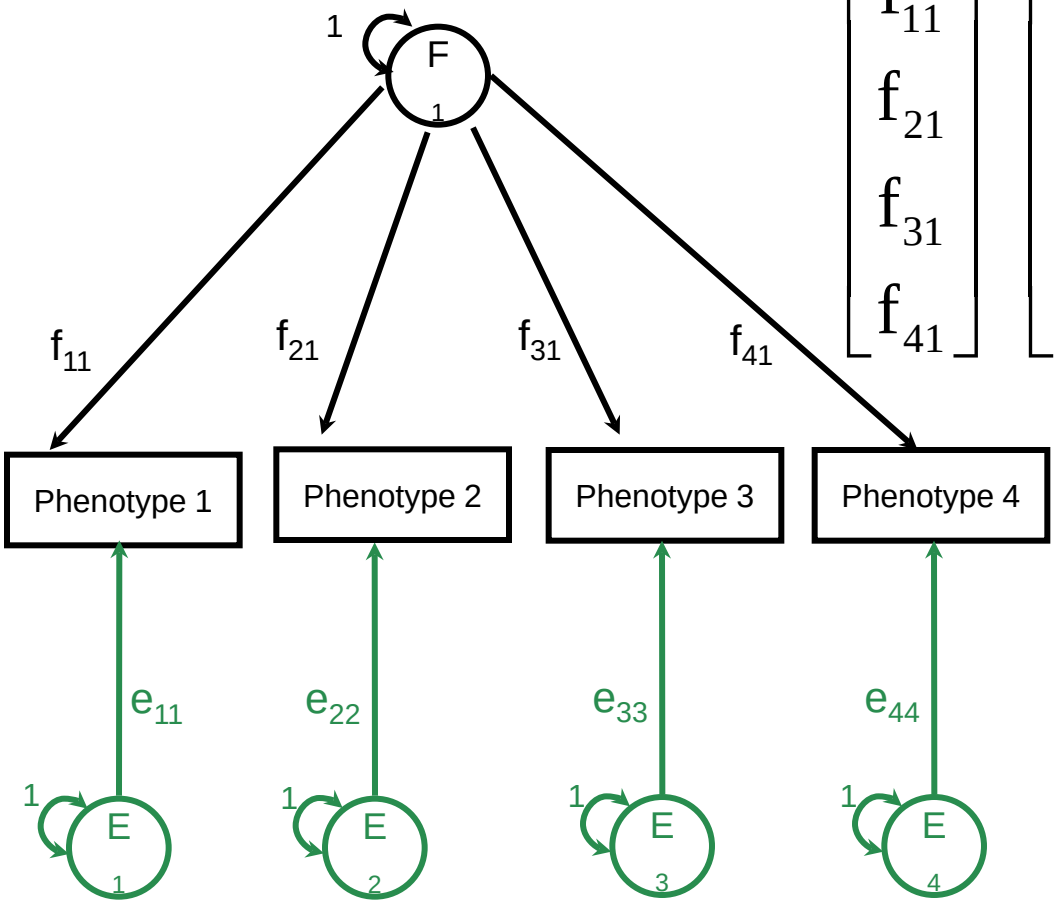
Funnel Web Spider

Maximum likelihood approach

- Unrelated individuals
  - Shared variance due to a common factor
  - Residual non-shared variance
- Family based data
  - ACE type models

Galah

Common factor model



$$\begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{41} \end{bmatrix} \begin{bmatrix} e_{11} & 0 & 0 & 0 \\ 0 & e_{22} & 0 & 0 \\ 0 & 0 & e_{33} & 0 \\ 0 & 0 & 0 & e_{44} \end{bmatrix}$$

Great white

Factor level association

$$\begin{bmatrix} \hat{\mu}_{11} \\ \hat{\mu}_{21} \\ \hat{\mu}_{31} \\ \hat{\mu}_{41} \end{bmatrix} = \left(\left[\beta_{factor} \right] \bullet \left[Genotype \right] \right) \otimes \begin{bmatrix} \mathrm{f}_{11} \\ \mathrm{f}_{21} \\ \mathrm{f}_{31} \\ \mathrm{f}_{41} \end{bmatrix} + \begin{bmatrix} \mathrm{m}_{11} \\ \mathrm{m}_{21} \\ \mathrm{m}_{31} \\ \mathrm{m}_{41} \end{bmatrix}$$

- Estimate a factor level beta
- Use the factor loadings as weights
- Add the uncorrected or grand mean
- 1 df

Green Tree Frog

Factor level association

$$\begin{bmatrix} \hat{\mu}_{11} \\ \hat{\mu}_{21} \\ \hat{\mu}_{31} \\ \hat{\mu}_{41} \end{bmatrix} = \left(\left[\beta_{factor} \right] \bullet \left[Genotype \right] \right) \otimes \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{41} \end{bmatrix} + \begin{bmatrix} m_{11} \\ m_{21} \\ m_{31} \\ m_{41} \end{bmatrix}$$

Dot product
see page 61 of the Mx manual

Green Tree Frog

Factor level association

$$\begin{bmatrix} \hat{\mu}_{11} \\ \hat{\mu}_{21} \\ \hat{\mu}_{31} \\ \hat{\mu}_{41} \end{bmatrix} = \left(\left[\beta_{factor} \right] \bullet \left[Genotype \right] \right) \otimes \begin{bmatrix} \mathrm{f}_{11} \\ \mathrm{f}_{21} \\ \mathrm{f}_{31} \\ \mathrm{f}_{41} \end{bmatrix} + \begin{bmatrix} \mathrm{m}_{11} \\ \mathrm{m}_{21} \\ \mathrm{m}_{31} \\ \mathrm{m}_{41} \end{bmatrix}$$

Kronecker product
see page 61 of the Mx manual

Green Tree Frog

Variable specific association

$$\begin{bmatrix} \hat{\mu}_{11} \\ \hat{\mu}_{21} \\ \hat{\mu}_{31} \\ \hat{\mu}_{41} \end{bmatrix} = \left([Genotype] \otimes \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} \right) + \begin{bmatrix} \mathrm{m}_{11} \\ \mathrm{m}_{21} \\ \mathrm{m}_{31} \\ \mathrm{m}_{41} \end{bmatrix}$$

- Estimate a separate beta for each trait
- Add the uncorrected or grand mean
- *n* df

Kookaburra

Simulated data set

- 10 traits, Moderately correlated ~.4
- 1 snp, MAF .2
- 500 individuals

Correlations

	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
V3	1	.390**	.403**	.366**	.396**	.418**	.410**	.340**	.429**	.412**
V4	.390**	1	.425**	.394**	.428**	.445**	.455**	.428**	.501**	.425**
V5	.403**	.425**	1	.387**	.479**	.453**	.444**	.405**	.410**	.403**
V6	.366**	.394**	.387**	1	.394**	.426**	.461**	.379**	.367**	.426**
V7	.396**	.428**	.479**	.394**	1	.403**	.425**	.416**	.370**	.438**
V8	.418**	.445**	.453**	.426**	.403**	1	.447**	.360**	.426**	.461**
V9	.410**	.455**	.444**	.461**	.425**	.447**	1	.387**	.444**	.412**
V10	.340**	.428**	.405**	.379**	.416**	.360**	.387**	1	.406**	.439**
V11	.429**	.501**	.410**	.367**	.370**	.426**	.444**	.406**	1	.445**
V12	.412**	.425**	.403**	.426**	.438**	.461**	.412**	.439**	.445**	1

**. Correlation is significant at the 0.01 level (2-tailed).

Kingfisher

factorlevel.mx… dataset.ped

```
Example Factor level association script - Boulder 2009 Sarah Medland
Data NI=15 NGroups=1
Rec file=dataset.ped
Labels fid iid a1 a2 genotype  V1 V2 V3 V4 V5 V6 V7 V8 V9 V10
Select genotype  V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 ;
Definition genotype ;

Begin Matrices;
F full 10 1 free          ! factor
R diag 10 10 free         ! residuals
M full 10 1 free          ! grand means
B full 1 1 free           ! association beta
G full 1 1                ! genotype
End Matrices;
```

Numbat

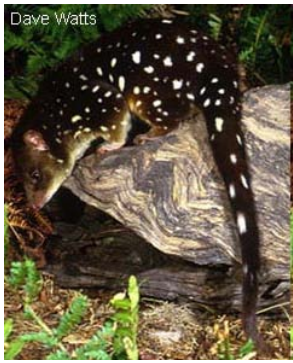
factorlevel.mx… dataset.ped

```
st .6 F 1 1 to F 10 1
st .4 R 1 1 to R 10 10
st  0 M 1 1 to M 10 1
sp  G genotype

Covariance (F*F')+(R*R') ;
Means M + (B.G)@F ;

Options multiple issat jiggle
end

drop B 1 1 1
end
```

Perentie Monitor

variablespecific.mx… dataset.ped

```
Example Factor level association script - Boulder 2009 Sarah Medland
Data NI=15 NGroups=1
Rec file=dataset.ped
Labels fid iid a1 a2 genotype  V1 V2 V3 V4 V5 V6 V7 V8 V9 V10
Select genotype  V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 ;
Definition genotype ;

Begin Matrices;
F full 10 1 free              ! factor
R diag 10 10 free             ! residuals
M full 10 1 free              ! grand means
B full 10 1 free              ! association beta
G full 1 1                    ! genotype
End Matrices;
```



Quoll

variablespecific.mx… dataset.ped

```
st .6 F 1 1 to F 10 1
st .4 R 1 1 to R 10 10
st  0 M 1 1 to M 10 1
sp  G genotype

Covariance (F*F')+(R*R') ;
Means M + (G@B).F ;

Options multiple issat jiggle
end

drop B 1 1 1 to B 1 10 1
end
```

Red back spider

Your task

- Run both FL and VS tests in Mx for the first data set
  - Edit the data file name
  - Calculate the p-value for the VS test using excel…
- Which variables are associated?
- What is the mean for variable 1 by genotype?

Results – factor level

F	B	M
0.608	0.169	0.061
0.670		0.068
0.654		0.066
0.614		0.062
0.640		0.065
0.659		0.066
0.665		0.067
0.604		0.061
0.652		0.066
0.659		0.066

Difference Chi-squared >> 4.116

Difference d.f. >>>>>>>>>> 1

Probability >>>>>>>>>>>> 0.042

Pademelon

Association with a factor score… Data set 1

Factor Matrix[a]

	Factor
	1
V1	.610
V2	.673
V3	.655
V4	.617
V5	.644
V6	.663
V7	.670
V8	.609
V9	.655
V10	.664

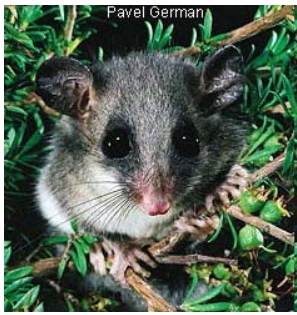
Extraction Method: Maximum Likelihood.

a. 1 factors extracted. 3 iterations required.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.100	.068		1.455	.146
	genotype	.168	.083	.090	2.023	.044

a. Dependent Variable: BART factor score 1 for analysis 1



Pigmy Possum

Results – variable specific

F	B	M
0.600	0.388	0.139
0.662	0.393	0.155
0.644	0.460	0.176
0.609	0.317	0.115
0.644	0.075	0.029
0.660	0.131	0.051
0.675	-0.042	-0.017
0.615	-0.087	-0.032
0.654	0.107	0.042
0.668	-0.019	-0.008

Difference Chi-squared	46.361
Difference d.f. >>>>	10
Probability >>>>>>>>	$1.2*10^{-6}$

Rainbow Lorikeet

Mean under different genotypes

-1	0	1
AA	AB	BB
-0.249	0.139	0.527

Salt water croc

Advantages to the ML approach

- Completely flexible
  - Can be applied to any model
    - Longitudinal models
    - Simplex/Autoregressive processed
  - Easy to add dominance etc
  - Covariates
  - Extends to family data

Stone fish

Disadvantages

- Correction for multiple testing?
- FL & VS tests provide complementary information
- Inflation of type 1 error
  - use a Bonferroni correction if you use both

Taipan

Tree Kangaroo

Wombat

Tiger Snake

Tassie Tiger

Thorny Devil

Zebra finch

Tassie Devil