Amplification of DNA mixtures

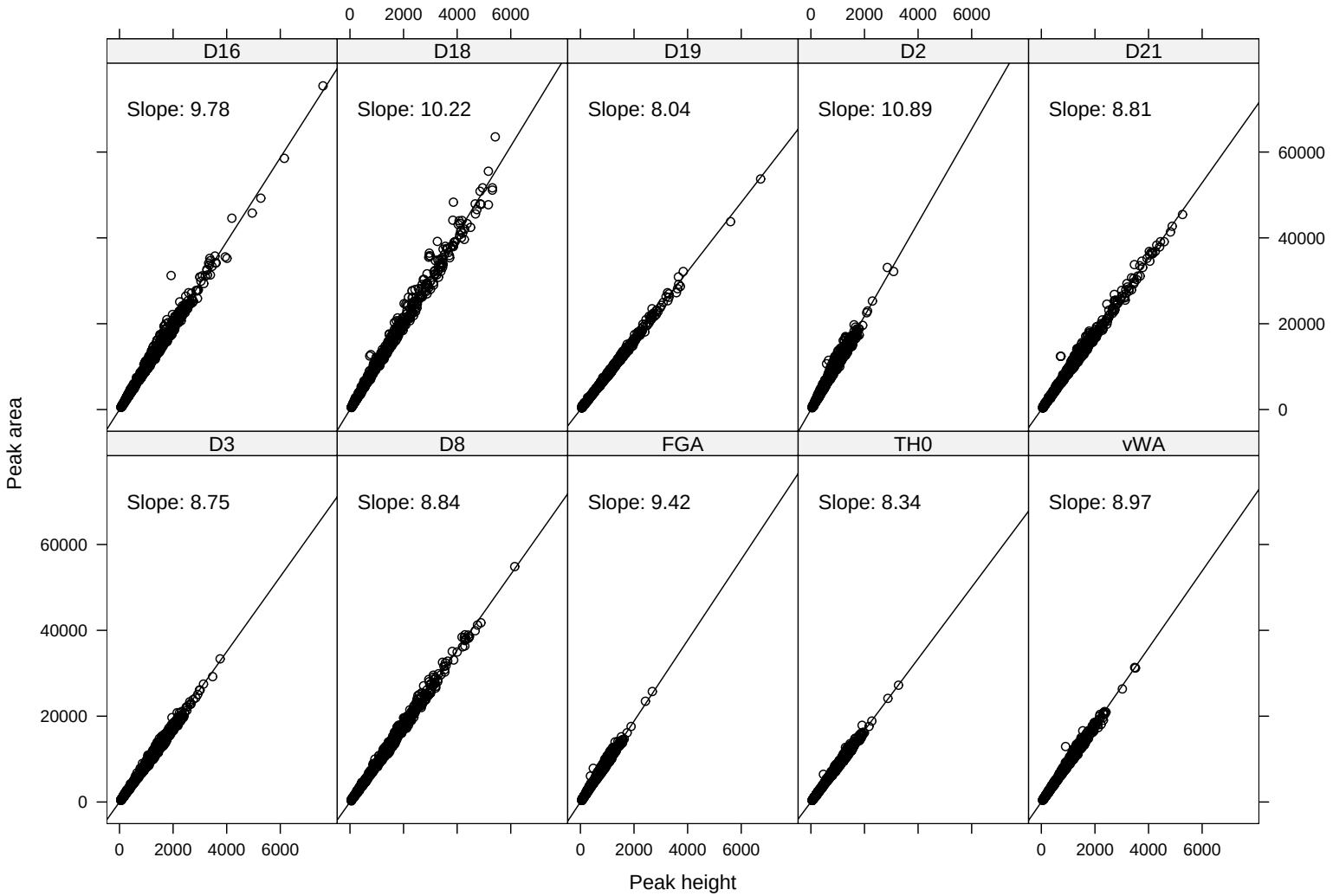
Missing data approach

Torben Tvedebrink

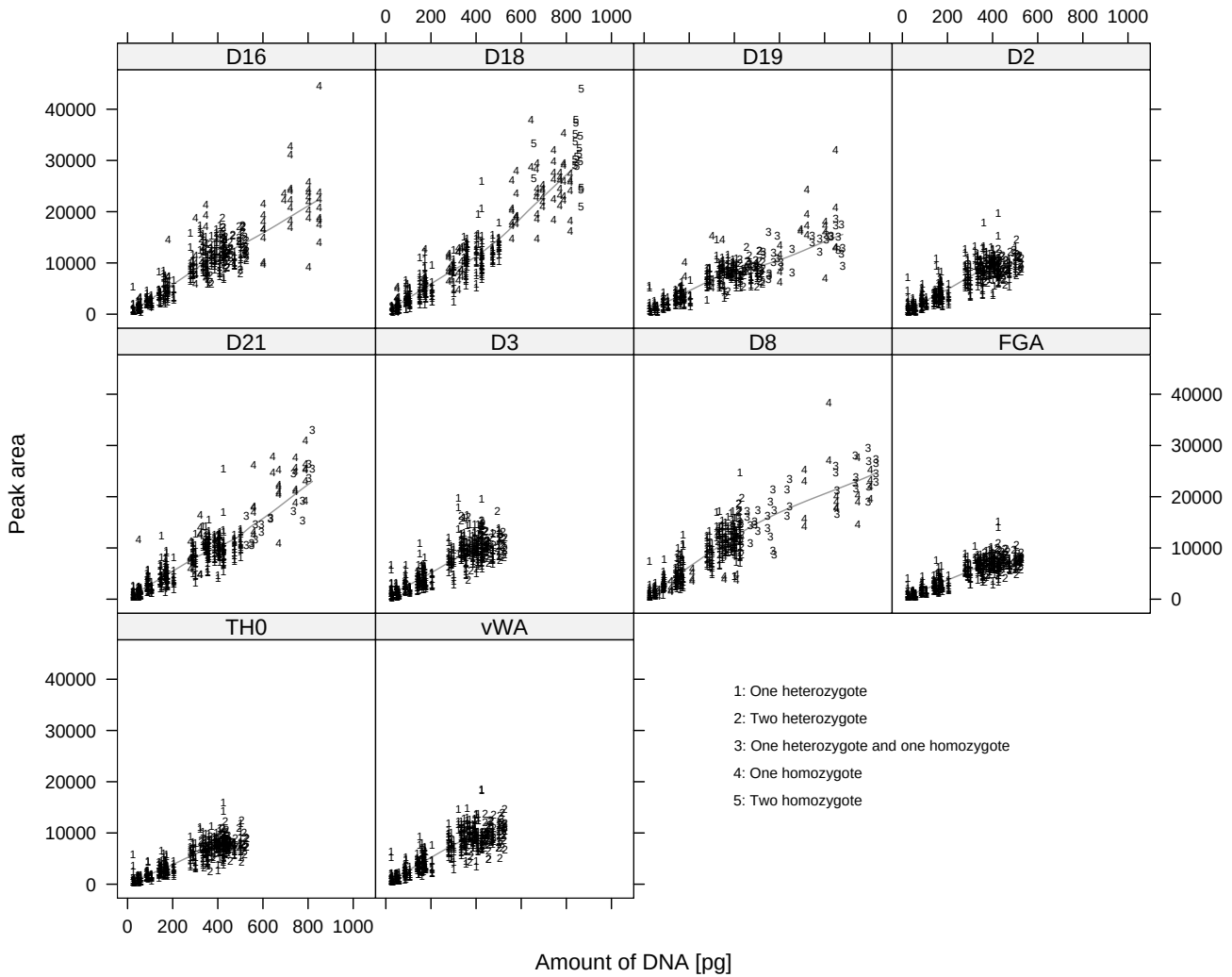
Department of Mathematical Sciences
Aalborg University

22nd ISFG Congress 2007 in Copenhagen : O-37

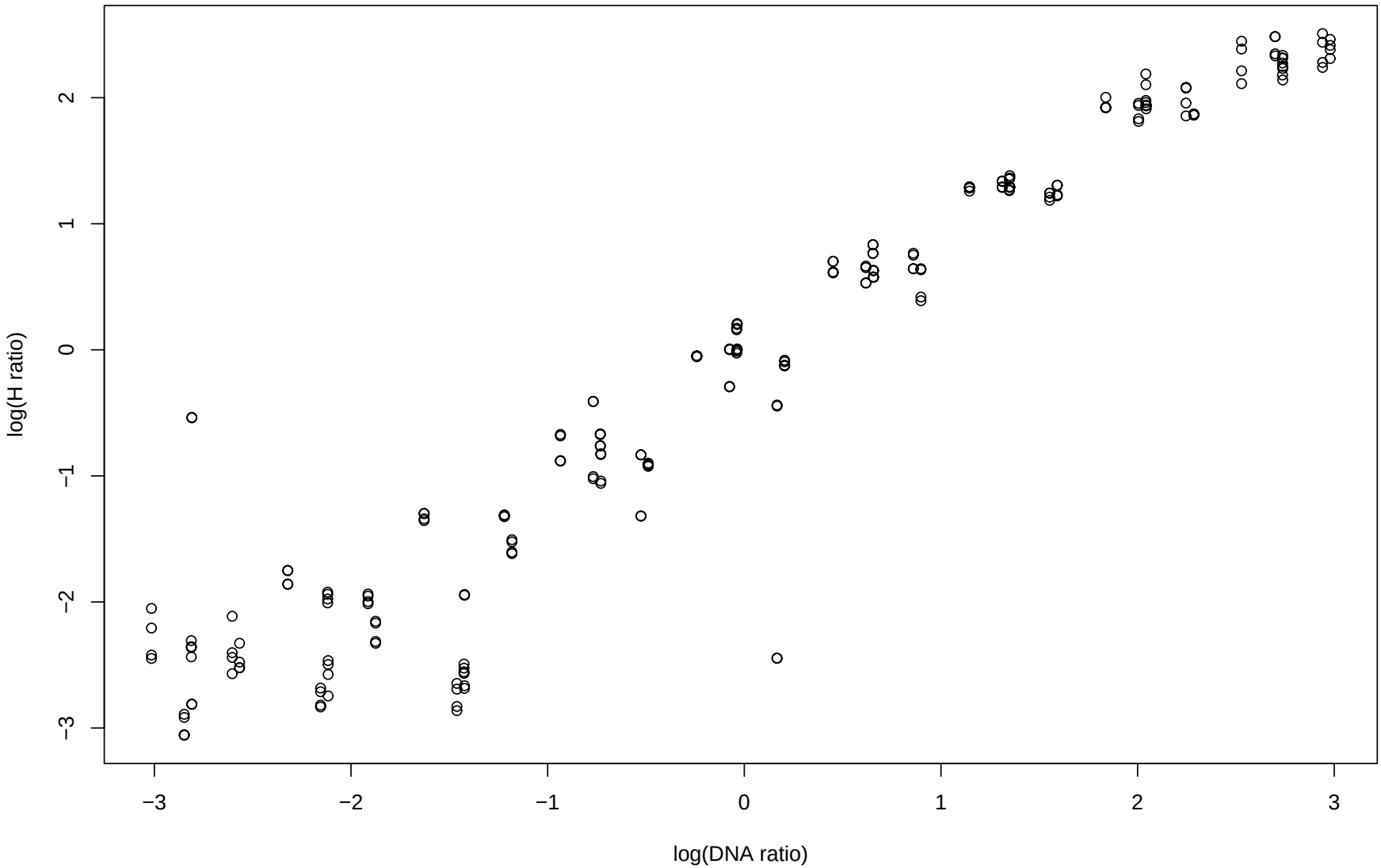
Data exploration



Data exploration



Data exploration



Data exploration

Summary of assumptions:

- Locus dependent proportionality of peak height and peak area.
- Mean of peak area increases with the amount of DNA.
- Variance is proportional to the mean.
- DNA concentration can be modelled by the weighted mean $H^{(k)}$.

(The plots are based on 71 controlled experiments conducted at Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen.)

Variation structure

Let the unobservable peak areas $\mathbf{A}$ be modelled as

$$\mathbf{A} = \boldsymbol{\mu} + \boldsymbol{\delta},$$

where $\boldsymbol{\mu} \propto H^{(k)}$ and $\boldsymbol{\delta} \propto H^{(k)}$. Both parameters depend on loci and the errors are assumed being independent.

Variation structure

Let the unobservable peak areas $\mathbf{A}$ be modelled as

$$\mathbf{A} = \boldsymbol{\mu} + \boldsymbol{\delta},$$

where $\boldsymbol{\mu} \propto H^{(k)}$ and $\boldsymbol{\delta} \propto H^{(k)}$. Both parameters depend on loci and the errors are assumed being independent.

Let the observable peak areas $\mathbf{M}$ be modelled as

$$\mathbf{M} = T\mathbf{A} + \boldsymbol{\varepsilon}.$$

Define $\tilde{\boldsymbol{\varepsilon}} = \mathrm{diag}(\mathbf{h})^{-1/2}\boldsymbol{\varepsilon}$ where $\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}) = \tilde{\boldsymbol{\Sigma}}$ has a compound symmetry structure.

Example

Assume we have two loci in our data. Then, we may observe

Person 1		Person 2		Locus	Allele	Height	Area	T_s-matrix
10	12	10	12	D16	10	1261	12381	$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$
10	12	10	12	D16	12	1249	12475	
13	13	12	16	D18	13	3141	32097	$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
13	13	12	16	D18	12	274	2833	
13	13	12	16	D18	16	146	1545	

Example

The compound symmetry structure on $\tilde{\Sigma}$ then yields

$$\tilde{\Sigma} = \left[\begin{array}{cc:cc:c} \nu_{D16} & \nu_{D16} & \nu_{D16.D18} & \nu_{D16.D18} & \nu_{D16.D18} \\ \nu_{D16} & \nu_{D16} & \nu_{D16.D18} & \nu_{D16.D18} & \nu_{D16.D18} \\ \hdashline \nu_{D16.D18} & \nu_{D16.D18} & \nu_{D18} & \nu_{D18} & \nu_{D18} \\ \nu_{D16.D18} & \nu_{D16.D18} & \nu_{D18} & \nu_{D18} & \nu_{D18} \\ \nu_{D16.D18} & \nu_{D16.D18} & \nu_{D18} & \nu_{D18} & \nu_{D18} \end{array}\right]$$

Evaluation of mixtures

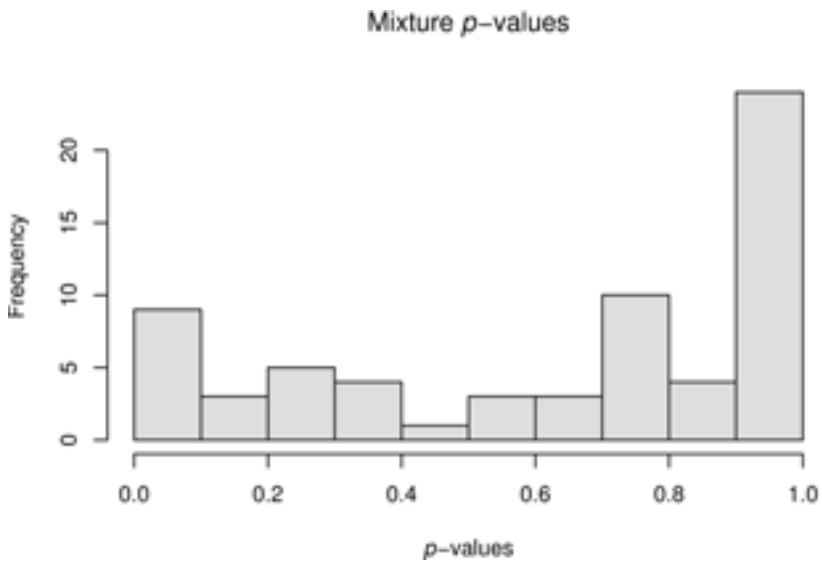
Given the estimated mean and variance of $\mathbf{M}$, we can compute the following two Mahalanobis distances with their approximate distributions

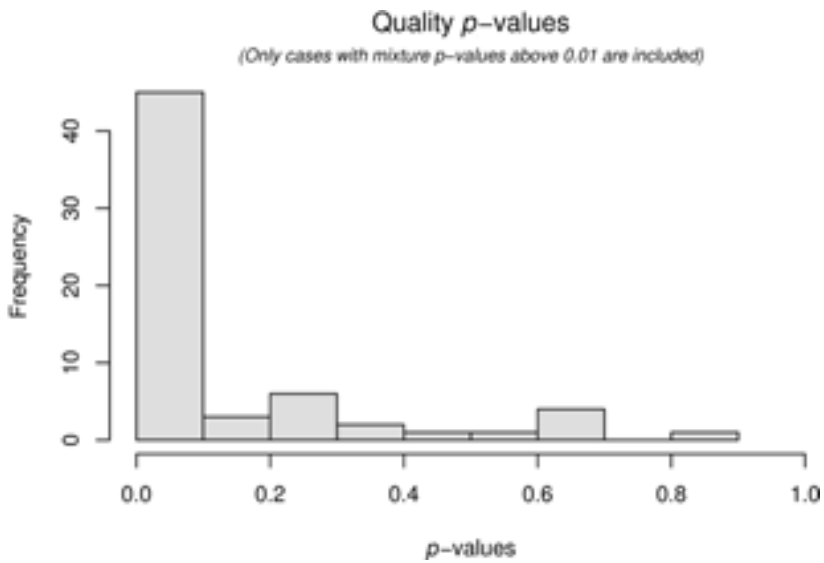
$$\begin{aligned}\left(\mathbf{M} - \mathrm{E}(\mathbf{M}|\,\bar{\mathbf{M}})\right)^{\top} \mathrm{cov}(\mathbf{M}|\,\bar{\mathbf{M}})^{-} \left(\mathbf{M} - \mathrm{E}(\mathbf{M}|\,\bar{\mathbf{M}})\right) &\sim \chi^2_{n-S} \\ (\,\bar{\mathbf{M}} - \mathrm{E}\,\bar{\mathbf{M}})^{\top} \mathrm{cov}(\,\bar{\mathbf{M}})^{-1} (\,\bar{\mathbf{M}} - \mathrm{E}\,\bar{\mathbf{M}}) &\sim \chi^2_S,\end{aligned}$$

where $\bar{\mathbf{M}}$ is the loci-sums.

Evaluation of mixtures

Evaluation of the model on 66 real crime case mixtures





Discussion

- Inter-loci correlations must be taken into account.
- Variation relative to observed peak heights.
- Further investigations will include degraded DNA.

Thank you for listening...

Acknowledgements
Poul Svante Eriksen, Department of Mathematical Sciences, Aalborg University
Helle Smidt Mogensen, Section of Forensic Genetics, University of Copenhagen
Prof Niels Morling, Section of Forensic Genetics, University of Copenhagen

Further information

More information can be found in the congress proceedings and a more detailed paper will be published in the near future.

The entire MSc-report is available at http://www.math.aau.dk/∼tvede