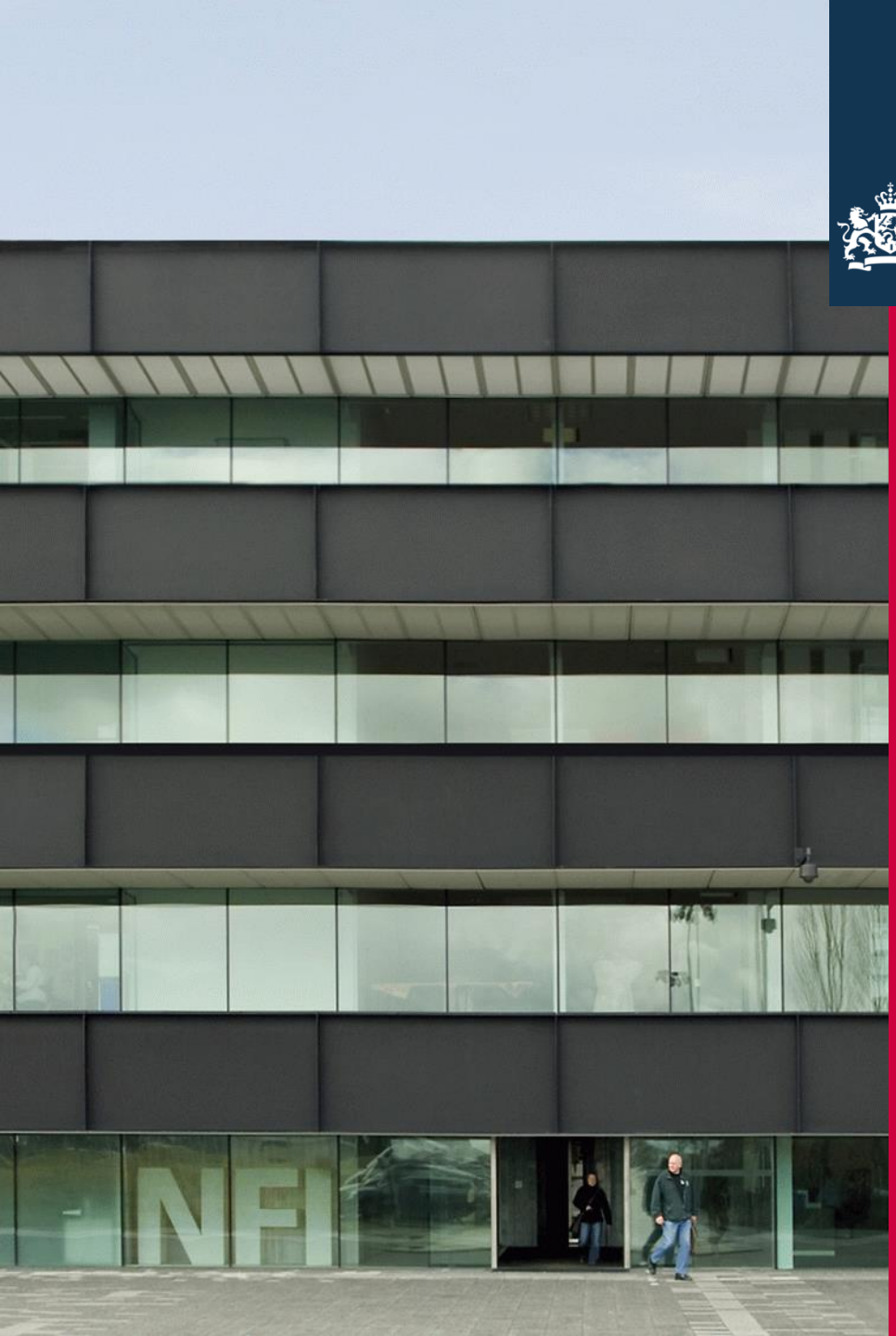


Netherlands Forensic Institute
Ministry of Security and Justice

The Impact of MPS on National Databases: Considerations, Challenges and Opportunities

Dr.Ir. C.P. (Kees) van der Beek MBA
Custodian Dutch DNA-database

Newcastle 15 July 2016

Contents of the presentation

- Development of forensic DNA-databases
- Present storage and comparison of DNA-profiles
- Allele differentiation by MPS and consequences for DNA-databasing
- Differentiation of identical twins
- Whole Mt DNA sequencing
- Conclusions

Development of forensic DNA-databases

- Replacement of the original DNA-fingerprinting technique by the PCR technique which could be automated
- Development of multiplex PCR kits
- Standardization of DNA-markers to enable international comparison of DNA-profiles
  - Europe: European Standard Set (ESS)
  - USA: CODIS Core Loci

Development of forensic DNA-databases

- All these developments have lead to national DNA-databases containing millions of STR-DNA-profiles of stains and persons:
  - China: > 20 million
  - USA: > 15 million
  - UK: > 5 million
  - France: > 3 million
  - Germany: > 1 million

European forensic DNA-databases (dec 15)

Country	Population size	Persons	Stains
Albania	3.600.000		
Armenia	3.000.000		
Austria	8.100.000	197.941	81.798
Belgium	10.400.000	35.991	43.224
Bosnia & Herzegovina	4.400.000		
Bulgaria	7.900.000		
Croatia	4.300.000	31.199	5.320
Cyprus	772.000	414	13.590
Czech Republic	10.553.800	171.519	16.561
Denmark	5.500.000	112.829	39.773
Estonia	1.311.800	47.618	10.534
Finland	5.475.866	157.303	19.624
France	66.030.000	3.068.243	312.815
Georgia	4.700.000	5, 819	998
Germany	81.000.000	849.907	284.066
Greece	10.600.000	8.362	12.708
Hungary	9.982.000	137.661	6.111
Iceland	315.000		
Ireland	4.200.000	7	1.746
Italy	58.000.000		
Kosovo	1.800.000		
Latvia	2.000.000	52.541	5.476
Liechtenstein	37.000		

Country	Population size	Persons	Stains
Lithuania	2.960.000	76.317	4.354
Luxembourg	570.000	2.361	4.352
Macedonia	2.100.000	17.094	5.132
Malta	400.000	30	430
Montenegro	650.000		
Netherlands	17.000.000	224.669	66.592
Northern Ireland	1.685.000		
Norway	5.000.000	66.076	11.206
Poland	38.200.000	42.753	5.833
Portugal	10.300.000	4.664	1.937
Romania	22.000.000	32.149	1.228
Russia	143.800.000		
Scotland	5.500.000	311.107	18.725
Serbia	7.335.000		
Slovakia	5.500.000	51.826	10.848
Slovenia	2.000.000	31.003	6.941
Spain	46.700.000	319.837	90.349
Sweden	9.845.155	151.931	30.064
Switzerland	7.779.000	176.758	63.941
Turkey	66.800.000		
UK (England & Wales)	53.700.000	4.691.350	488.126
Ukraine	47.600.000		
Total	801.401.621	11.071.460	1.664.402

Present storage and comparison of DNA-profiles

- DNA-profiles are stored and compared based on the number of STR’s in the alleles of a locus
- Incomplete repeats are indicated by a number of basepairs after the decimal sign
- In CODIS only alleles which have been included in the configuration are accepted and only four positions are available to designate an allelic value

Storage and comparison of DNA-profiles

	Target NEFLS2288 [redacted] Keyboard	Candidate NEFLS2288 [redacted] Forensic, Unknown	
STR Locus	Target	Candidate	Locus Match Stringency
D10S1248	14,16		Not Searched
vWA	15,17	15,17	High
D16S539	11	11	High
D2S1338	16,23	16,23	High
Amelogenin	X,Y	X,Y	High
D8S1179	14,15	14,15	High
D21S11	29,30	29,30	High
D18S51	12,16	12,16	High
D22S1045	15,16		Not Searched
D19S433	14,15	14,15	High
TH01	8,9.3	8,9.3	High
FGA	22	22	High
D2S441	14		Not Searched
D3S1358	15,16	15,16	High
D1S1656	13		Not Searched
D12S391	19,20		Not Searched
11 STR Loci Match		Match Stringency: High	

Allele differentiation by MPS

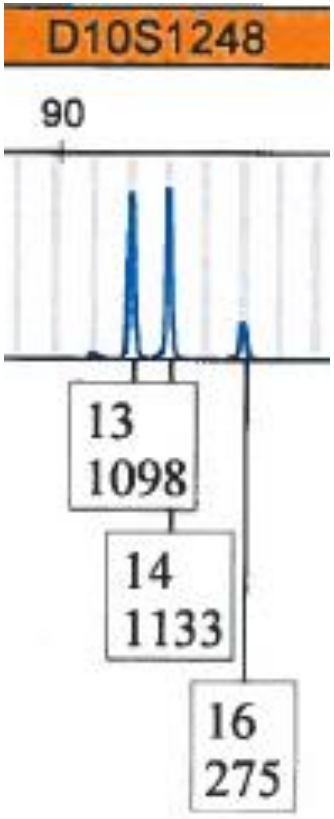
Allele sequence **D21S11**	Allele call	Total frequency	European frequency	In-house database frequency[a]
[TCTA]5 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]9	27	0.010	0.000	0.032
[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]10	28	0.135	0.152	0.190
[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]11	29	0.163	0.185	0.204
[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]10	29	0.067	0.054	
[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]7 TCA [TCTA]3	29.3	0.010	0.011	–
[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]11	30	0.163	0.185	0.237
[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12	30	0.087	0.076	
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]11	30	0.038	0.022	
[TCTA]7 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]10	30	0.010	0.011	
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]10 TA TCTA	30.2	0.029	0.033	0.034
[TCTA]5 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]11TA TCTA	30.2	0.019	0.022	
[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12	31	0.019	0.022	0.098
[TCTA]4 [TCTG]7 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12	31	0.010	0.011	
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12	31	0.010	0.011	
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]11 TA TCTA	31.2	0.058	0.065	0.079
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]13	32	0.010	0.011	0.014
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12 TA TCTA	32.2	0.058	0.033	0.024
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]13 TA TCTA	33.2	0.087	0.087	0.024
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]14 TA TCTA	34.2	0.010	0.011	0.001
[TCTA]11 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12	36	0.010	0.000	–

Rockenbauer et al (2014) FSI Genetics 8, 68-72

Consequences of allele differentiation

- The evidential value of DNA-profiles will increase because Random Match Probabilities will go down due to lower allele frequencies and as a consequence international searches, familial searches and searches with partial and mixed profiles will yield less false positives
- Mixtures will be more easy to analyze

Analysis of mixtures



Possible contributors without allele differentiation

13,14 + 13,16

13,14 + 14,16

13,14 + 16,16

Possible contributors if allele 14 consists of 80% 14A + 20% 14B

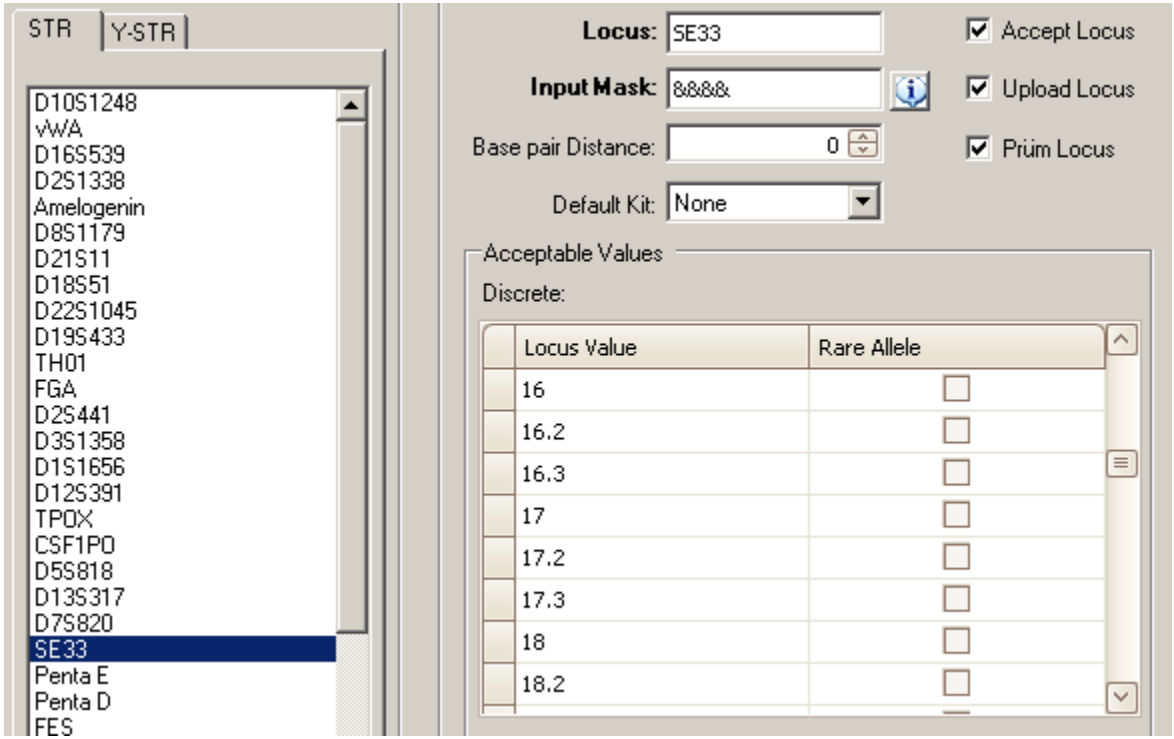
13,14A + 14B,16

Allele differentiation by MPS

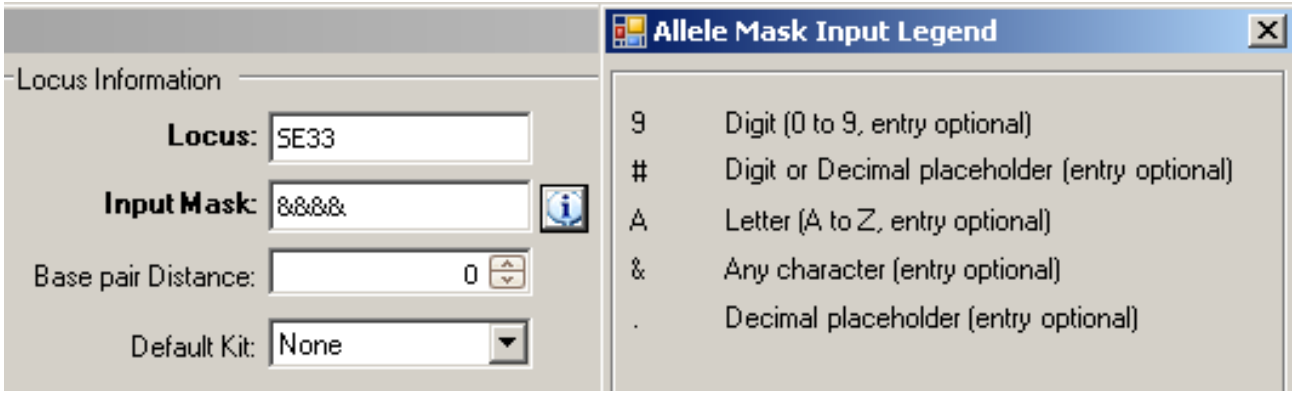
Allele sequence **D21S11**	Allele call	Total frequency	European frequency	In-house database frequency[a]
[TCTA]5 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]9	27	0.010	0.000	0.032
[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]10	28	0.135	0.152	0.190
[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]11	29	0.163	0.185	0.204
[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]10	29	0.067	0.054	
[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]7 TCA [TCTA]3	29.3	0.010	0.011	–
[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]11	30A	0.163	0.185	0.237
[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12	30B	0.087	0.076	
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]11	30C	0.038	0.022	
[TCTA]7 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]10	30D	0.010	0.011	
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]10 TA TCTA	30.2	0.029	0.033	0.034
[TCTA]5 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]11TA TCTA	30.2	0.019	0.022	
[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12	31	0.019	0.022	0.098
[TCTA]4 [TCTG]7 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12	31	0.010	0.011	
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12	31	0.010	0.011	
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]11 TA TCTA	31.2	0.058	0.065	0.079
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]13	32	0.010	0.011	0.014
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12 TA TCTA	32.2	0.058	0.033	0.024
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]13 TA TCTA	33.2	0.087	0.087	0.024
[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]14 TA TCTA	34.2	0.010	0.011	0.001
[TCTA]11 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12	36	0.010	0.000	–

Rockenbauer et al (2014) FSI Genetics 8, 68-72

Locus configuration in CODIS



Locus configuration in CODIS



Allele differentiation by MPS

In addition:

- MPS-alleles can look longer or shorter than CE-alleles due to insertions or deletions in the flanking regions or because the CE-alleles were not sequenced in the right direction in the past
- In addition there is also variation in the flanking regions due to SNP’s

Consequences of allele differentiation

- A new STR-nomenclature has to be developed to incorporate the additional variation
- The new names have to be translated back into the old names to enable comparison with the DNA profiles in existing DNA-databases
- DNA-databases must be modified to also include the new MPS derived profiles and to compare them to each other

MPS nomenclature for STR's

Forensic Science International: Genetics 22 (2016) 54–63

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements

Walther Parson[a,b,*], David Ballard[c], Bruce Budowle[d,e], John M. Butler[f], Katherine B. Gettings[f], Peter Gill[g,h], Leonor Gusmão[i,j,k], Douglas R. Hares[l], Jodi A. Irwin[l], Jonathan L. King[d], Peter de Knijff[m], Niels Morling[n], Mechthild Prinz[o], Peter M. Schneider[p], Christophe Van Neste[q], Sascha Willuweit[r], Christopher Phillips[s]

Nomenclature proposal of the group of Peter de Knijff in Leiden (NL)

D13S317[CE12]-Chr13-GRCh37-g.82.722.160:82.722.223-TATC[13]AATC[1]ATCT[3]-g.x.136G>A

Five elements:

1. Locus name and CE STR allele name
2. Chromosome and human genome assembly number
3. STR repeat region coordinates of reference allele
4. Full description of the STR motif
5. Location of flanking region SNPs

Differentiation of identical twins

- Presently identical twins cannot be distinguished genetically
- Some criminals get away with that

KaDeWe heist suspects released because of twin DNA

Published: 19 Mar 2009 11:51 GMT+01:00
Updated: 19 Mar 2009 11:51 GMT+01:00

Share Tweet Share reddit

Five weeks after their arrest, Berlin police have released two suspects in the spectacular robbery of the city's KaDeWe department store because DNA tests can't tell the identical twins apart, daily *Berliner Morgenpost* reported on Thursday.

Differentiation of identical twins

- Rare individual somatic mutations can be found by whole genome sequencing (WGS) to distinguish identical twins
- However whole genome sequencing can result in unexpected findings (e.g. the BRCA-1 allele which results in an in-creased susceptibility to breast cancer)
- This can be prevented by instructing the MPS-software to show only the differences between the twins

Whole mtDNA sequencing

- At this moment only part of the mtDNA is sequenced
- Sequencing all 16.569 basepairs of the mtDNA will increase its distinctive character and hence its evidential value
- However this may also reveal disease causing mutations
- The NFI will instruct its MPS software to ignore the positions of such mutations

Conclusions

- DNA-databases will have to be adjusted to accept and compare MPS derived STR profiles to each other and to CE derived STR profiles
- Targetted MPS in stead of WGS will reveal only forensically important sequences
- Ethically unacceptable sequences can be separated from the forensically important sequences by instructing the MPS software to ignore them