



**UNIKLINIK
KÖLN**

Institut für
Rechtsmedizin

Zum Stand der Entwicklung und Anwendung probabilistischer Software für die Interpretation von komplexen DNA-Mischspuren

Peter M. Schneider

35. Spurenworkshop



Complex Mixtures



Komplexe Mischungen

Mixture Deconvolution



Auftrennung einer Mischung

Biostatistische Ansätze

bei komplexen Mischungen

- **Trefferwahrscheinlichkeit**
= Genotyp-Wahrscheinlichkeit bei Einzelprofilen
= match probability
- **Kombinierte Einschlusschance $P(I)$**
= combined probability of inclusion (CPI)
= RMNE (Random Man Not Excluded)
- **Likelihood Quotient (LQ)**
= Chance für die Beobachtung der Spur unter der Annahme von zwei alternativen Hypothesen

Statistische Modelle zur Interpretation von Mischspuren

1. Das **binäre** Modell
(binary model)
2. Das **semikontinuierliche** Modell
(semi-continuous model)
3. Das **(voll) kontinuierliche** Modell
(fully) (continuous model)

Statistische Modelle zur Interpretation von Mischspuren (1)

Das binäre Modell

- behandelt Allele als vorhanden oder nicht vorhanden
- und berücksichtigt keine Peakhöhen.
- Bei möglichem Drop-out muss das betroffene System weggelassen werden, oder es wird die "2p" Regel zur Berechnung eingesetzt.
- Dieses Modell bildet die Grundlage der "klassischen" Mischspurenberechnung unter Verwendung des LQ.
 - z.B. Statistefix

Statistische Modelle zur Interpretation von Mischspuren (2)

Das semikontinuierliche Modell

- setzt eine Wahrscheinlichkeit ein für den Fall, dass ein (eigentlich vorhandenes) Allel nicht nachgewiesen wird – die Dropout-Wahrscheinlichkeit $\text{Pr}(D)$
- Berücksichtigt nur die Allel- bzw. Genotypkombinationen und ignoriert Peakhöhen

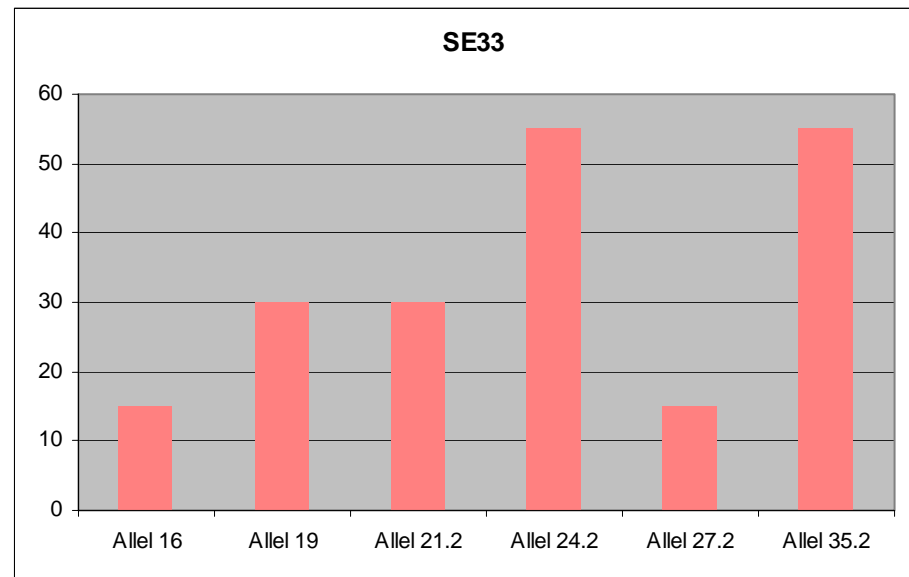
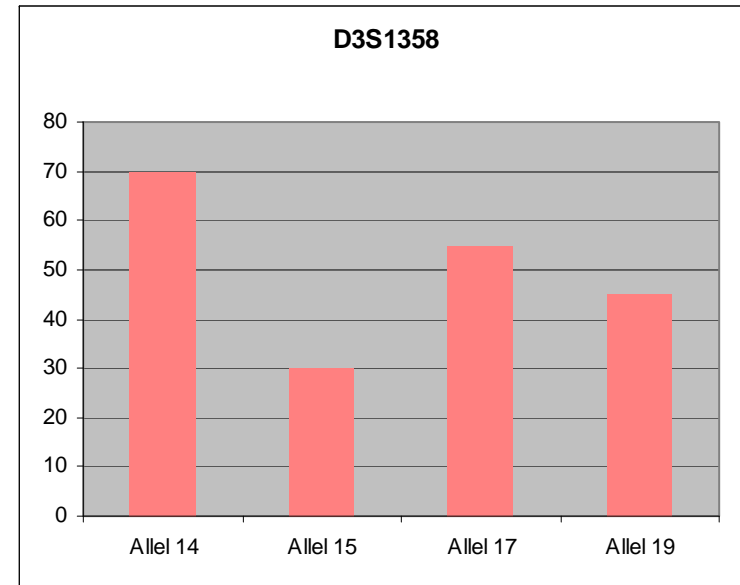
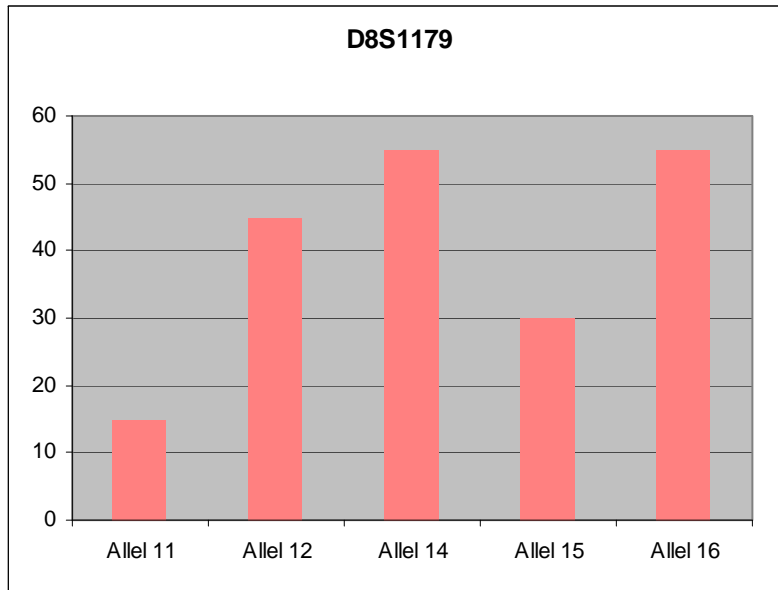
	D3S1358	D8S1179	SE33
Spur	14, 15, 17, 19	11, 12, 14	16, 19, 23.2
Person A	15, 17	11, 14	16, 23.2
Person B	14, 19	12, 14	19, 28.2
LQ PA + ?	> 1	> 1	> 1
LQ PB + ?	> 1	> 1	< 1

Statistische Modelle zur Interpretation von Mischspuren (3)

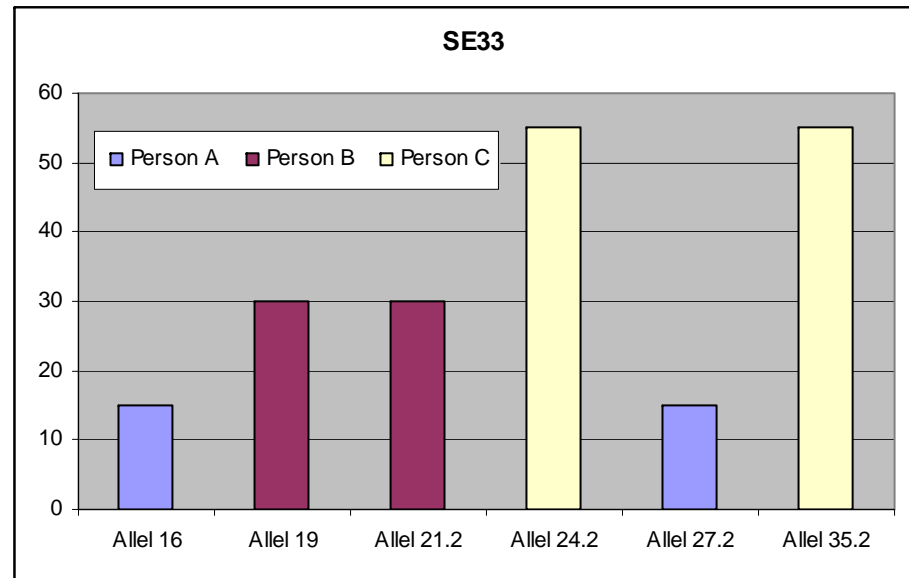
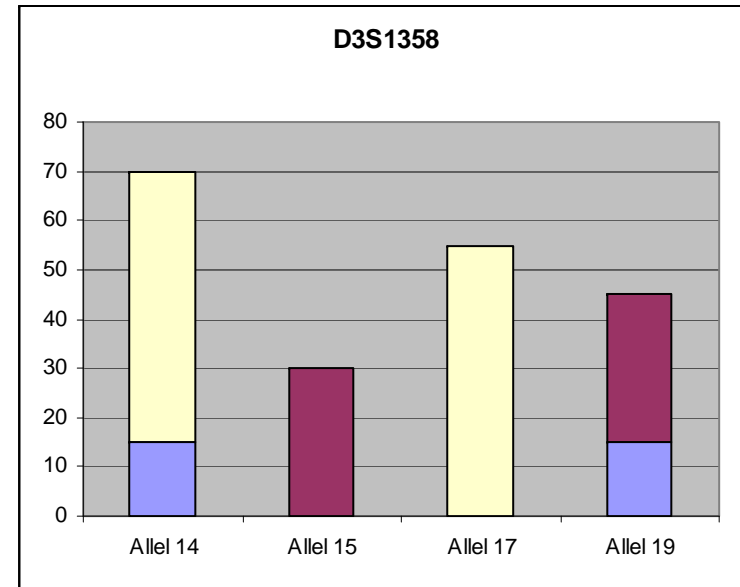
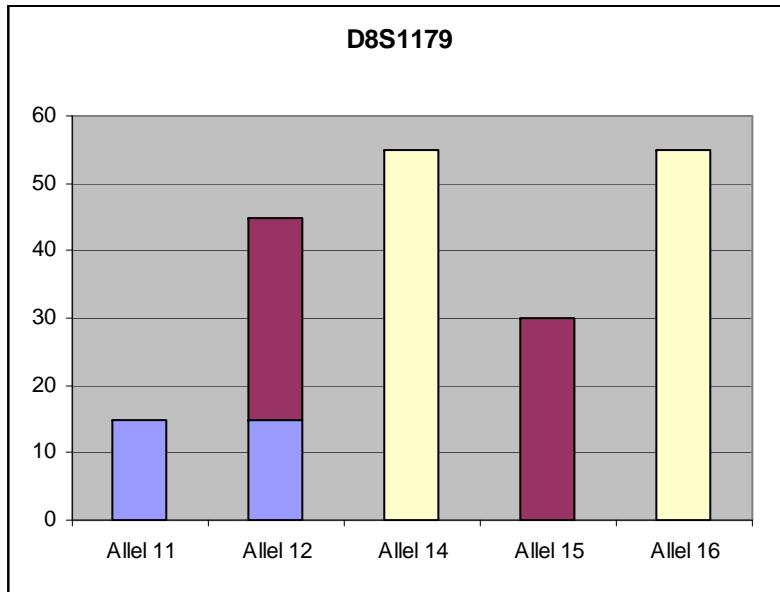
Das **kontinuierliche** Modell

- behandelt das Auftreten von Dropout und weiteren stochastischen Ereignissen auf der Grundlage individueller Peakhöhen;
- verwendet die Information von Peakhöhen, Peakbalancen und Stutterbanden zur Auftrennung eines Mischprofils in die Einzelprofile;
- erfordert eine individuelle Anpassung der Software-Parameter an die Kit- und Labor-spezifischen Eigenschaften durch eine Serie von Validierungsexperimenten;
- verwendet mathematische Simulationen (Markov chain Monte Carlo – MCMC) zur Absicherung der möglichen alternativen Szenarien.

Rechnerische "Entmischung" eines komplexen Mischprofils



Rechnerische "Entmischung" eines komplexen Mischprofils



15% Person A
35% Person B
50% Person C

Software zur Mischspurenanalyse

- Es gibt bereits eine Reihe von Programmen zur Interpretation komplexer Mischspuren
- mit semikontinuierlichen Modellen, z.B.
 - Forensim / LRmix / LRmix Studio (H. Haned, NFI)
 - LikeLTD (D. Balding)
 - Lab Retriever (N. Rudin, K. Inman, K. Lohmueller)
 - Forensic Statistics Tool – FST (A. Mitchell, OCME)
- mit kontinuierlichen Modellen, z.B.
 - True Allele (Cybergenetics, M. Perlin, USA)
 - STRmix (J. Buckleton et al., ESR, New Zealand)
 - LiRa (R. Puch-Solis, LGC)
 - DNA-VIEW Mixture Solution (Ch. Brenner, USA)
 - EuroForMix (Ø. Bleka, Oslo – EUROFORGEN-NoE)

[Open source]

LRmix STUDIO

- Anwenderfreundliche Windows-Version von Forensim/LRmix mit graphischem Interface, benötigt Java (NFI / H. Haned)
 - Website: <http://lrmixstudio.org>
 - Verwendet nur Genotyp-Information
 - Standard: $\Pr(D) = 0.1$ und $\Pr(C) = 0.05$
 - kann durch "sensitivity analysis" empirisch bestimmt werden
 - Dateninput als "csv" Texttabelle
 - Analyse von Replikaten möglich

STRmix

- Kommerzielle Software von J. Buckleton, D. Taylor und J. Bright (ESR, NZL)
 - Website: <http://strmix.esr.cri.nz>
 - Genemapper-Tabelle inkl. Peakhöhen, Fragmentlängen etc. als Input
 - Analyse von Replikaten möglich
 - Verwendet MCMC-Simulation zur Mischspurenanalyse, daher bis zu 1 Std. Rechenzeit
 - Soll als "Expertensystem" für alle Arten von DNA-Spuren eingesetzt werden

Vergleichsstudie (A. Kloosterman – ICFIS 2014)

- Berechnung einer Serie von 2-Personen-Mischungen
- Haupt- zu Nebenanteil 9:1, Identifiler mit 1 - 0,5 - 0,25 ng

		log LQ	
		STRmix	LRmix
Person A (minor)	1 ng (2x Dropout)	11,47 / 11,59 / 11,42	6,75
	0,25 ng (14x Dropout)	0,253 / 0,501 / -0,548	0,568
Person B (major)	1 ng	19,22 / 19,21 / 19,21	11,39
	0,25 ng	17,33 / 17,13 / 13,26	13,36

Vor- und Nachteile

- Software auf der Basis kontinuierlicher Modelle
 - Beschreibung einer Mischspur durch Darstellung und Bewertung aller Einzelkomponenten
 - Dadurch deutliche höhere LQs
 - Probleme bei Minimal Spuren mit stark erhöhter stochastischer Variation
 - Längere Rechenzeiten durch MCMC-Simulationen zur Verifizierung der Hypothesen
- Software auf Basis semikontinuierlicher Modelle
 - Verwenden das "klassische" binäre Modell wenn kein Dropout vorliegt ($\Pr(D) = 0$)
 - Ergebnisse auch bei Minimal Spuren mit stark erhöhter stochastischer Variation möglich
 - Keine langen Rechenzeiten

Ausblick

- Kriterien zur Qualitätsbewertung von DNA-Profilen?
 - Grenzen der Bewertung von DNA-Profilen mit massiven Dropout-Ereignissen
- Empirische Bestimmung von $Pr(D)$ und $Pr(C)$
 - Anwenderspezifische Labordaten für Multiplexkit und analytische Parameter
- Neue ISFG DNA-Kommission zur Erarbeitung von Empfehlungen zur Software-Validierung!