Peter Gill, James Curran, Cedric Neumann,
Amanda Kirkham, Tim Clayton, Jonathan Whitaker, Jim Lambert
Forensic Science International: Genetics 2 (2008) 91–103

Interpretacja złożonych profili DNA za pomocą modeli empirycznych i metoda mierzenia rzetelności tych modeli

Eliminacja nierozstrzygającego profilu DNA

Posługując się tradycyjnymi metodami, można sformułować opinię tylko o profilu DNA, który rzeczywiście jest całkowicie lub częściowo zgodny z profilem podejrzanego. Podawana wartość dowodowa zawsze ma więc iloraz wiarygodności (LR) większy niż jeden. Często jednak profile DNA nie są jednoznaczne: mogą być profilami niepełnymi (do spełnienia hipotezy oskarżenia (H_p) może brakować kilku alleli) lub mieszaninami. Mogą także występować stuttery, co wpływa na utrudnienie interpretacji.

Każdy ekspert w dziedzinie badania DNA rutynowo podejmuje decyzje co do tego, czy formułować opinię w przypadku brakujących alleli lub wtedy, gdy dodatkowe allele w profilu nie są zgodne z profilem podejrzanego. Może to jednak prowadzić do sytuacji, w której ten sam dowód może zostać uwzględniony przez jednego eksperta, podczas gdy inny poda wynik nierozstrzygający, który nie potwierdzi ani nie wykluczy winy podejrzanego. W przypadku profili złożonych oznacza to, że tradycyjne obliczenie można przeprowadzić, tylko upraszczając sytuację. Na przykład, gdy profil jest niepełny, trzeba założyć, że jeżeli prawdziwa jest hipoteza oskarżenia, doszło do wypadania (*dropout*) alleli. Przykładowo genotyp podejrzanego to *ab*, a profil śladu z miejsca zdarzenia to *a*, w takiej sytuacji prawdopodobieństwo występujące w liczniku (H_p) jest mniejsze niż jeden. Nie można więc zastosować tradycyjnego sposobu obliczenia LR [1, 2] (szczególnie jeżeli pik oznaczony ze śladu ma wystarczającą wielkość), do stwierdzenia, że prawdopodobieństwo wypadania wynosi w rzeczywistości zero ($\Pr(D) \approx 0$) [2].

W sytuacji gdy mamy do czynienia ze złożonym profilem albo jeżeli występuje w nim wiele prążków niezgodnych z porównawczym profilem DNA podejrzanego, należy wziąć pod uwagę, że albo istnieje dodatkowy dawca (dawcy), albo doszło do kontaminacji (mierzonej przez $\Pr(C)$ [1]). W konsekwencji profil taki może być przedstawiony w opinii jako nierozstrzygający, co zostanie poparte sformułowaniem: „nie można dostosować żadnych znaczących obliczeń statystycznych". Jest to po prostu inny sposób stwierdzenia, że profil jest zbyt złożony do interpretacji.

Nasuwa się zatem pytanie, czy wynik LR jest wiarygodny. W celu zbadania jego rzetelności korzysta się z testów Tippetta, opracowanych przez Tippetta i wsp. [3]. Testy te początkowo wykorzystywano przy badaniach farb i lakierów. Evett i Weir [14] zastosowali tę koncepcję do wczesnych przykładów analizy DNA, szczegółowy opis zasady działania testu przedstawili zaś Backleton i wsp. [5]. Wcześniej testy Tippetta były wykorzystywane jako ogólny sprawdzian rzetelności danej techniki albo dla przetestowania losowo generowanych zdarzeń (typowo 1000 oddzielnych spraw).

Wykorzystując wcześniej opisane testy, autorzy przytaczanego artykułu opracowali alternatywną strategię sprawdzania rzetelności LR dla konkretnego przypadku. Jest to jej pierwsze zastosowanie w analizie mieszaniny STR lub gdy profil jest niepełny. Przy jej zastosowaniu wymagane jest sformułowanie opinii przez eksperta, szczególnie jeżeli chodzi o ocenę prawdopodobieństwa wypadania alleli, jednak proces ten został przez autorów w znacznym stopniu sformalizowany. W obliczeniach uwzględnia się wszystkie allele bez potrzeby przypisywania ich konkretnym dawcom albo kwalifikowania ich jako artefaktów – stutterów. Jest to szczególnie ważne w przypadku mniejszościowego profilu, ponieważ trudno jest wówczas odróżnić prążki od stutterów.

Autorzy tekstu zaproponowali też nową metodę pozwalającą stwierdzić, czy proponowane ulepszenia modeli są rzeczywiście efektywne i sensowne. W rezultacie powstała nowa koncepcja, dająca pojęcie o rzetelności samego pomiaru ilorazu wiarygodności.

Przestaje zatem być konieczne stosowanie przy opiniowaniu kategorii „nierozstrzygający", wynikającej ze złożoności wyniku. Można bowiem przeprowadzić ocenę prawdopodobieństwa dowolnego zestawu hipotez każdego profilu (od dowolnej liczby potencjalnych dawców).

Hipotezy formułuje się na podstawie analizy okoliczności sprawy oraz uzyskanych profili DNA. Często proces ten nie jest prosty i może wymagać rozważenia wielu par twierdzeń. Gill i wsp. [6] zademonstrowali, jak można uprościć go, poprzez przedstawienie minimalnego ilorazu wiarygodności dla zdefiniowanego zestawu par twierdzeń. Dotychczas złożoność takich obliczeń wykluczyła tego rodzaju podejście do rozważania materiału dowodowego.

Na ile rzetelna jest odpowiedź?

Po obliczeniu ilorazu wiarygodności określone dla każdego przypadku testy Tippetta dają pogląd na wiarygodność samego testu. Autorzy chcieliby zatem odpowiedzieć na pytanie: „jeżeli podejrzany nie jest dawcą śladu dowodowego, na ile prawdopodobne jest to, że otrzyma się podobną wartość LR, jeśli dawcą będzie przypadkowa osoba?".

W celu uzyskania odpowiedzi skorzystano z symulacji komputerowej. Zamiast obliczyć LR odnoszący się do podejrzanego i warunkowe profile w ramach hipotezy oskarżenia (H_p) autorzy eksperymentu zastąpili je przypadkowymi osobami – które stanowią twierdzenie w ramach hipotezy obrony (H_d). Wykazano, że w omawianych przykładach uzyskane wartości LR są znacznie mniejsze niż jeden, a szansa, że ślad dowodowy pochodzi od przypadkowej osoby, była nieistotna.

Interesujące jest także rozważenie alternatywnego scenariusza: gdyby plama dowodowa naprawdę pochodziła od podejrzanego, jak często LR byłby mniejszy niż jeden? Jak często wartość dowodowa przemawiałaby na korzyść hipotezy obrony, jeżeli podejrzany naprawdę jest sprawcą? To także sprawdza się drogą symulacji komputerowej – pod uwagę bierze się wszystkich dawców w ramach H_p i tworzy nowy profil dowodowy uwzględniający Pr(D) i Pr(C). Następnie oblicza się LR i powtarza się symulację odpowiednio dużą liczbę razy (w omawianej pracy 1000). Wykazano, że otrzymywane LR były większe niż jeden, a LR obliczane dla konkretnej sprawy pokrywa się z symulowanymi przedziałami.

W przytaczanym artykule wykazano, że zasady te można zastosować do złożonych przypadków, dla których dotychczas nie można było sformułować opinii.

Rola oceny eksperta w odniesieniu do materiału dowodowego w postaci profilowania DNA

Klasyczna interpretacja mieszanin DNA ma na celu oznaczenie alleli oraz określenie liczby dawców. Decyzje ekspertów bywają dwubiegunowe, to znaczy, że prawdopodobieństwo zdarzenia wynosi albo zero, albo jeden. Na przykład:

Ustalanie, który pik to stutter: w rzeczywistości niewielki pik, który znajduje się w pozycji stutter (–4 bp od większościowego allela) jest albo podprążkiem, albo allelem, albo też mieszaniną stuttera i allela. Ekspert, na podstawie wielkości przylegającego allela (macierzystego), podejmuje zazwyczaj ostateczną decyzję. Zgodnie z powszechną wytyczną stosowany jest próg 15% wyznaczony na podstawie eksperymentalnych obserwacji – stuttery mają zazwyczaj wielkość 15% rozmiarów macierzystego allela (+4 bp).

Dla przykładu można podać próbkę niebędącą mieszaniną: jeżeli pik w pozycji stuttera ma rozmiary mniejsze niż 15% wielkości przylegającego (+4 bp) allela, wówczas jest oznaczony jako stutter, tj. Pr(St) ≈ 1 [7] i nie jest dalej uwzględniany w obliczeniach. Analogicznie, jeżeli pik przekracza 15% wielkości allela macierzystego, wówczas będzie oznaczony jako allel, tj. Pr(St) ≈ 0.

Tworzenie par heterozygotycznych: Równowaga heterozygotyczna to stosunek wysokości (Ht) albo pola powierzchni dwóch pików. Można zdefiniować ją jako:

$$Hb = \frac{Ht_{najmniejsza}}{Ht_{największa}} \times 100\%$$

Jeżeli mniejszy z dwóch alleli mieści się w 60% (Hb > 60%) większego allela [7], oznacza to, że są komplementarne i jest możliwe, że wywodzą się z tego samego locus i od jednej osoby (zasada ta nie ma zastosowania, kiedy ma się do czynienia z małą ilością DNA – Whitaker i wsp. [8] proponują, aby przy wartościach rzędu 10 000 rfu równowaga heterozygotyczna wynosiła Hb > 20% dla danych zbieranych z zastosowaniem 34 cykli PCR). Wytyczna ta jest często używana do rozdzielania składnika mniejszościowego mieszaniny od składnika większościowego.

Określenie współczynnika wiarygodności

Iloraz wiarygodności (LR) jest zwykle uznawany za najlepszy sposób przedstawienia dowodów w postaci DNA [4, 9–11].

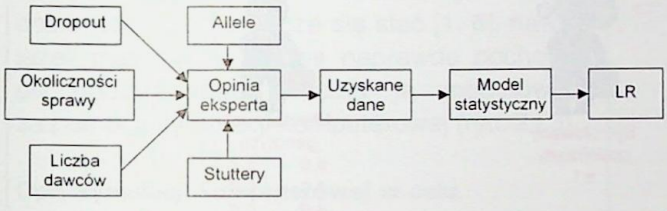
Prawdopodobieństwo dowodu (E) jest oceniane względem dwóch przeciwstawnych hipotez, zwykle określanych jako hipoteza oskarżenia (H_p) i hipoteza obrony (H_d). Ekspert ocenia prawdopodobieństwo zaistnienia danego dowodu przy każdej z tych hipotez za pomocą modelu statystycznego.

Zastosowane podejścia mogą być niesformalizowane, dlatego może się zdarzyć, że różni eksperci mogą przypisać różne prawdopodobieństwa danemu przypadkowi, co będzie szczególnie częste w problemach złożonych. Różnorodność w przedstawianiu LR może

zostać ograniczona przez zastosowanie wytycznych (np. użycie progów dla stuttera i równowagi heterozygotycznej) w celu ukierunkowania opinii eksperta. Te wytyczne tworzą podstawy do dokumentów systemu zarządzania jakością (QMS). Generalizując, można więc stwierdzić, że celem zarządzania jakością jest zapewnienie ekspertowi wytycznych w zakresie formułowania opinii oraz zasad wykorzystania modelu statystycznego, który służy do obliczania ilorazu wiarygodności przedstawianego następnie w sądzie.

Model (ryc. 1) składa się z dwóch części:

a) *dane wyjściowe*: ekspert określa allele, ocenia występowanie stutterów, bierze pod uwagę okoliczności sprawy, formułuje opinię na temat liczby dawców. Jest to ocena wstępna, podczas której wykorzystuje się QMS w celu zminimalizowania różnic między ekspertami.



Ryc. 1. Doświadczenie i opinia eksperta prowadzą bezpośrednio do stosowanego obecnie modelu obliczania LR

b) *model statystyczny*: na podstawie ww. założeń ekspert podejmuje decyzje o wyborze modelu statystycznego. Model ten wykorzystuje stały, niezmienny wzór. Od tego uzależniona jest siła dowodu.

Jak rzetelny jest dany model; jak można porównać różne modele?

Zasady statystyczne, będące podstawą interpretacji dowodów w postaci DNA, są rozszerzane i ulepszane. Już w trakcie powstawania algorytmu może okazać się, że należy go udoskonalić. W przytaczanej pracy stosuje się na przykład parametr prawdopodobieństwa wypadania Pr(*D*) jako uogólnioną średnią z całego danego profilu. Bierze się jednak pod uwagę, że w przyszłości powinien on zostać zmodyfikowany w sposób, który umożliwi zróżnicowanie Pr(*D*) w zależności od dawcy i locus.

Według autorów przytaczanego artykułu trzeba odpowiedzieć na następujące pytania:

- jak rzetelna jest dana metoda interpretacji?
- jak można zmierzyć jej rzetelność?
- jak można zmierzyć wpływ modyfikacji w modelach statystycznych na jej rzetelność?
- jak takie zmiany wpłynęłyby na modele i ich działanie i jak można by zmierzyć wpływ poszczególnych ulepszeń?

Autorzy omawianej pracy posługują się modelem stosowanym do złożonego przypadku [6] i sugerują sposoby testowania rzetelności tego modelu. Próbują także odpowiedzieć na dwa pierwsze pytania. Dla zilustrowania teorii wykorzystują model interpretacji DNA zaproponowany przez Gilla i wsp. [6] i rozszerzony przez Currana i wsp. [12].

Testowanie efektywności z wykorzystaniem wykresów Tippetta

Wykresy Tippetta [3–5, 13–15] stworzono początkowo na potrzeby kryminalistycznej analizy lakierów. Były używane w celu oceny jakości badań statystycznych w wielu obszarach analizy kryminalistycznej, takich jak badania szkła, daktyloskopia, badania środków kryjących, automatyczna identyfikacja twarzy, rozpoznawanie mowy i badanie pisma ręcznego.

Warto rozważyć, w jaki sposób wykresy te mogą być wykorzystywane w interpretacji wyników analizy DNA. Autorzy omawianego artykułu proponują zastosowanie wykresu Tippetta w określonych sytuacjach, aby zapewnić miarę działania dla każdego konkretnego przypadku. Zgodnie z zawartą w niniejszej pracy tezą, że efektywność jest mierzona dwoma parametrami (Neumann i wsp. [17]), można założyć że:

1. Prawdopodobieństwo uznania błędnego dowodu na korzyść oskarżenia jest określane jako prawdopodobieństwo wykazania przez iloraz wiarygodności wartości powyżej 1 dla hipotezy obrony (H_d). Można to wyrazić jako $\Pr(LR > 1 \mid H_d \equiv prawda)$. W tej sytuacji interesująca jest także największa wartość przyjmowana przez LR. Nazywa się ją górną granicą (LR_{Hdmax}). Odzwierciedla ona, jak przesadzony może być błędny dowód na korzyść oskarżenia.

2. Prawdopodobieństwo błędnego dowodu na korzyść obrony jest ideą komplementarną do wyżej przedstawionej tezy (1). Określa się to jako prawdopodobieństwo, że LR przyjmuje wartość mniejszą niż jeden. Można to wyrazić jako $\Pr(LR < 1 \mid H_p \equiv prawda)$. Interesuje nas także najmniejsza wartość przyjmowana przez LR. Nazywa się ją dolną granicą (LR_{Hpmin}) i odzwierciedla ona, na ile przesadzony może być błędny dowód na rzecz obrony.

Opis modelu badawczego

Każda obserwacja allela w profilu DNA opisana jest przez trzy parametry [1]:

– prawdopodobieństwo wypadania, Pr(D) albo przeciwnego zdarzenia, Pr($\bar{D}$),
– prawdopodobieństwo kontaminacji, Pr(C) albo przeciwnego zdarzenia, Pr($\bar{C}$),
– prawdopodobieństwo stutteru, Pr(St) albo przeciwnego zdarzenia, Pr(St).

Dodatkowo w profilu powinna być brana pod uwagę liczba dawców (n_c) [18].

Obecnie nie ma statystycznego modelu, który łączy jednocześnie wszystkie te parametry, dlatego też wszystkie istniejące modele należy uważać za niepełne (w rzeczy samej należy uznać, że pełny model jest nieosiągalny). Podstawy teoretyczne są jednak dobrze ustalone, a parametry prawdopodobieństwa wypadania i prawdopodobieństwa kontaminacji można uniwersalnie stosować do wszystkich profili DNA w ramach schematu początkowo opisanego przez Gilla i wsp. [1]. Można wprowadzić je także do systemu eksperckiego [6, 12]. Gill i wsp. [1, 19, 20] dostarczyli prostych metod oszacowania tych prawdopodobieństw dla danej liczby dawców.

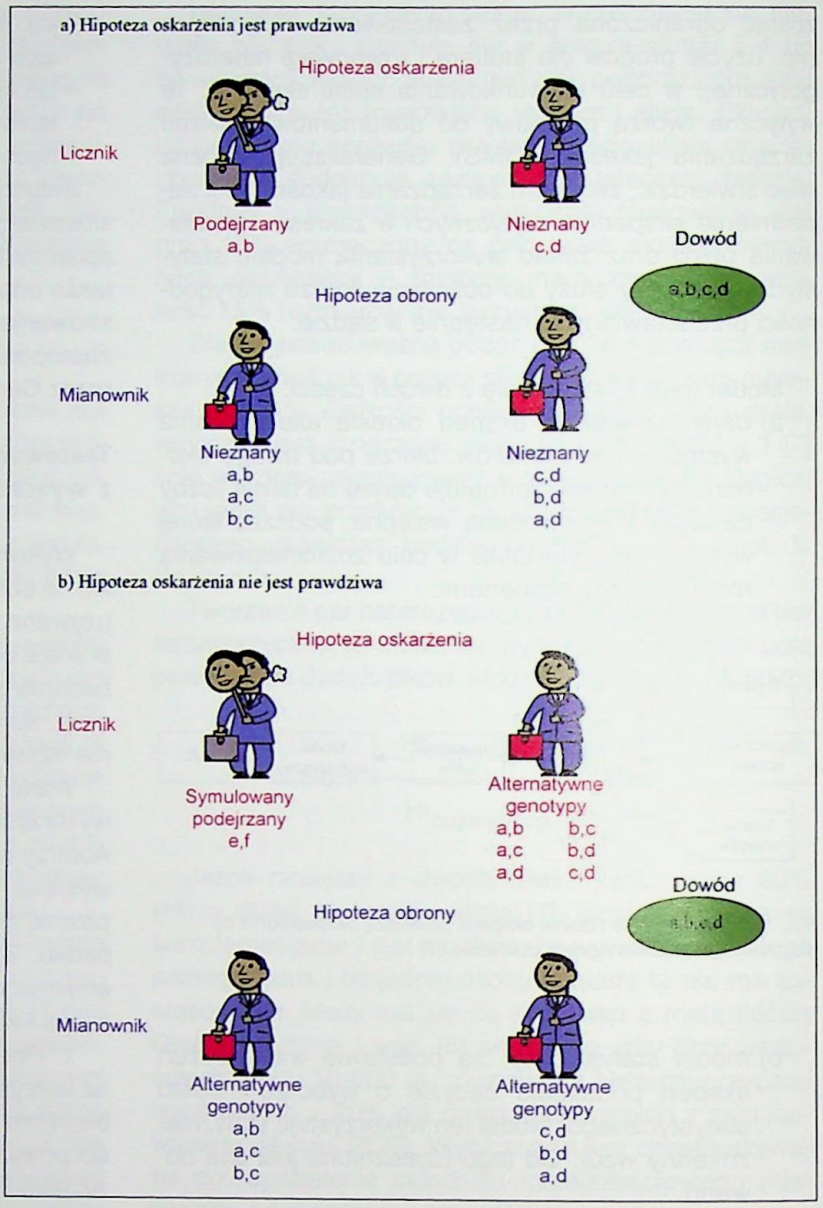
Iloraz wiarygodności porównuje prawdopodobieństwo dowodu (E) w odniesieniu do dwóch alternatywnych wersji, gdzie H_p to hipoteza oskarżenia, a H_d hipoteza obrony. Mierzone są:

– prawdopodobieństwo dowodu (E), jeżeli ślad pochodzi od podejrzanego (H_p),
– prawdopodobieństwo dowodu (E), jeżeli ślad nie pochodzi od podejrzanego (H_d).

$$LR = \frac{Pr(E \mid H_p)}{Pr(E \mid H_d)}$$

Przykład: oskarżenie utrzymuje, że ślad z miejsca zdarzenia jest prostą mieszaniną materiału od oskarżonego/podejrzanego (S) i nieznanej osoby (U). Można to przedstawić jako H_p: $S + U$.

Obrona może twierdzić, że prawdziwym przestępcą nie jest podejrzany, lecz nieznana przypadkowa osoba. W zapisie autorów wzór wygląda następująco: H_d: $U_1 + U_2$. Oznaczenia U_s wskazują, że nie muszą to być te same wartości, co U w ramach H_p.



Ryc. 2. Ilustracja symulacji Tippetta: 1. Jeżeli hipoteza oskarżenia jest prawdziwa: podejrzany to *ab*, a nieznane allele to *c* i *d*. Dowód to *abcd*, a obliczenie ilorazu wiarygodności oparte jest na założeniu, że to standardowa mieszanina od dwóch dawców (bez wypadania alleli i kontaminacji), a LR jest większy niż jeden. 2. Jeżeli hipoteza oskarżenia nie jest prawdziwa: podejrzany jest niewinny, a sprawcą, którego nazywa się „symulowanym podejrzanym", jest przypadkowa osoba. Stworzono przypadkową osobę o genotypie *e* i *f*. Zakłada się, że mieszanina jest dwuskładnikowa. Żadnego z tych alleli nie stwierdzono w śladzie z miejsca zdarzenia – hipoteza oskarżenia może być wyjaśniona tylko przy założeniu, że allele te wypadły, pozostawiając sześć alternatywnych kombinacji (trzech par) nieznanego genotypu. Zakładając, że ma się do czynienia z mieszaniną dwóch osób, wyjaśnieniem jednej pary dla każdej kombinacji może być fakt, że są one kontaminantami; np. jeżeli $U = a$ i b, to allele c i d można wyjaśnić tylko wówczas, jeżeli są kontaminantami. Alternatywne genotypy w hipotezie obrony są takie same, jak dla hipotezy 1, ponieważ allele dowodowe są takie same. Hipoteza H_p *jest prawdą* jest mniej prawdopodobna, stąd LR jest mniejszy niż 1. Za pomocą symulacji komputerowej tworzone są cały czas przypadkowe genotypy „symulowanego podejrzanego" i odpowiadające im 1000 LR, które są wykreślane jako dystrybuanta.

Testowanie rzetelności LR

Należy zmierzyć rzetelność testu ilorazu wiarygodności. Oczywiste założenie testu brzmi: „jeżeli hipoteza oskarżenia jest prawdziwa, wówczas iloraz wiarygodności powinien być znacznie wyższy niż dla hipotezy nieprawdziwej". Wykres Tippetta może być wykorzystany do zmierzenia zmiany wartości LR w zależności od prawdopodobieństwa hipotezy. W tym celu autorzy zakładają, że oskarżenie fałszywie wskazało podejrzanego, i symulują zastąpienie profilu DNA podejrzanego przypadkowo wybranym profilem. W takiej sytuacji możliwe jest przetestowanie wpływu niezgodności (zastąpionego) profilu podejrzanego ze śladem z miejsca zdarzenia na wartość LR. W standardowej interpretacji DNA takiej sytuacji przypisany byłby LR równy zero, ponieważ prawdopodobieństwo potwierdzenia H_p wynosi zero. Jednakże, przy uwzględnieniu dodatkowych czynników jak wypadanie alleli i kontaminacja, istnieje niezerowe prawdopodobieństwo, że tak może się stać [1, 6], nawet jeżeli materiał w śladzie naprawdę pochodzi od podejrzanego. Można ją modelować za pomocą symulacji komputerowej (ryc. 2 i 3).

Opis symulacji komputerowej w celu uzyskania wykresów Tippetta

W celu uzyskania wykresu Tippetta konieczne jest skonstruowanie empirycznej dystybuanty (*cdfs*) dla LR przy założeniu, że H_p jest prawdziwa i dla LR przy założeniu, że H_d jest prawdziwa. W celu uzyskania zmienności wartości LR wykorzystuje się mechanizm losowy. Dla przykładu można założyć, że profil dowodowy składa się z 4 alleli *a*, *b*, *c* i *d*, a podejrzany to typ *a* i *b*. W takim przykładzie hipotezy zapisane są następująco:

H_p: $S + U$
H_d: $U_1 + U_2$

- **Symulacja wykresu Tippetta dla założenia, że H_p *jest prawdziwa***

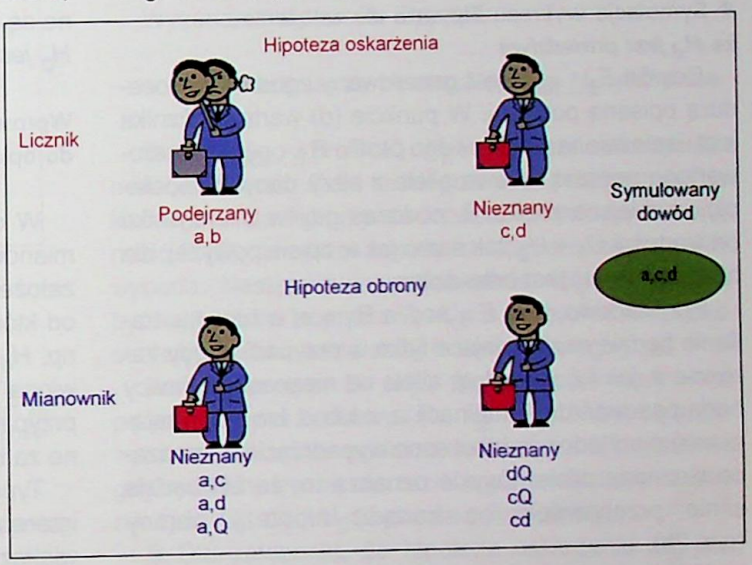
Zmiany wartości dowodu (E) zgodnie z następującymi zasadami (ryc. 2a):

a) symulacja 1000 mieszanin DNA ($E_{1...1000}$),
b) pobieranie profilu podejrzanego (S) – dla każdego locus *i* oraz dla każdego locus *j* allela A (*i* oraz *j*) wchodzi w skład profilu E_n z prawdopodobieństwem Pr($\bar{D}$),
c) allele niewyjaśnione przez S dodawane są do E_n,
d) kalkulacja $LR_{1...1000}$. Argumentem w liczniku jest $S+U$, a w mianowniku U_1+U_2.

Pod uwagę bierze się nieznane allele (lub potencjalny stutter), niezmiennie w każdej symulacji, ponieważ celem jest zbadanie wpływu zróżnicowania alleli podejrzanego poprzez ich wypadanie.

Przykład: Dla danych locus $i = 1$, $E = acd$ oraz $S = ab$ oraz przy założeniu, że H_p *jest prawdziwe*, musiało nastąpić wypadnięcie *b*. W obu hipotezach autorzy zakładają, że mieszanina pochodzi od dwóch osób. Należy jednak zaproponować alternatywne wyjaśnienie dla alleli *c* i *d*.

Jednym wyjaśnieniem jest ich pochodzenie od nieznanego dawcy (*U*). Jeśli zastosuje się Pr($\bar{D}$) dla każdego allelu (S), otrzyma się 4 możliwe wyniki: *a*, *b*, *ab*



Ryc. 3. Wykres Tippetta (z wypadaniem alleli w profilu dowodowym). Zakłada się, że hipoteza oskarżenia jest rzeczywiście prawdziwa i prawdziwymi dawcami są podejrzani *ab* i nieznany dawca *cd*. Dowody są typu *a*, *c* i *d*. Przy założeniu, że H_p *jest prawdziwe* oznacza to, że allel *b* od podejrzanego musiał wypaść. Symuluje się dowód ($E_{1...1000}$) z czterech alleli pochodzących od podejrzanego (*a* i *b*) i nieznanego dawcy (*c* i *d*) i uwzględnia przy tym fakt, że podejrzane allele mogły wypaść na podstawie danej wartości Pr(D). Można także włączyć kombinacje *ac* i *ad* pochodzące od nieznanego dawcy, jeżeli pozostałe allele (odpowiednio *d* i *c*) są allelami *drop-in*. Jeśli to prawdopodobieństwo wynosi mniej niż jeden, otrzyma się alternatywne profile dowodowe, takie jak -, *b*, *c* i *d* (jeżeli allel *a* wypadł) albo *a*, -, *c* i *d* (jeżeli allel *b* wypadł). Przy założeniu, że H_p jest prawdziwe, dowód *a*, *c* i *d* można wyjaśnić w liczniku, tylko jeżeli allel *b* wypadł. W mianowniku kombinacje genotypu mogą albo oznaczać wypadanie alleli, albo jego brak (np. rozważane są wszystkie kombinacje par alleli *a*, *c* i *d* – dla zachowania zwięzłości pokazanych jest tylko kilka kombinacji; Q oznacza nieznany allel, który wypadł). Dla każdej symulacji ($E_{1...1000}$) tworzony jest nowy profil dowodowy poprzez zastosowanie danego prawdopodobieństwa wypadania do genotypu podejrzanego (*a* i *b*) przy zachowaniu nieznanych alleli (*cd*) jako stałych (co powtórzono we wszystkich 10 loci). Umożliwia to obliczenie LR, pod warunkiem że H_p *jest prawdziwe*. Metoda ustalania H_d *jest prawdziwe* jest identyczna, z tym że licznik pozostanie niezmienny, argumentem będzie natomiast mianownik z losowo wybieranym profilem nieznanego dawcy ($R_1...R_{1000}$) z allelami (*ef*) (ryc. 2b).

lub *zero*. Połączenie pozostałych alleli z pozostałymi możliwościami daje wynik: *acd*, *bcd*, *abcd* lub *cd*. Procedura jest następnie powtarzana dla każdego locus $i = 1...n$, a LR obliczone jak podano wyżej w punkcie (d). LR dla całego profilu w dowodzie E jest obliczane jako iloczyn $LR = \prod_{i=1...n} LR_j$.

W kolejnym etapie ocenia się LR, uwzględniając powyższą hipotezę z wykorzystaniem znanego genotypu S oraz symulowanego dowodu E_n – model zaproponowany przez Gilla i wsp.

Procedura jest powtarzana określoną liczbą razy w celu ustalenia dystrybuanty LR. W przytaczanej pracy zastosowano 1000 symulowanych dowodów. Zdaniem autorów 1000 powtórzeń wystarczy, aby zbadać zróżnicowanie LR.

- **Symulacja wykresu Tippetta dla założenia, że *H_d jest prawdziwa***

Dowód $E_{1...1000}$ jest generowany zgodnie z procedurą opisaną powyżej. W punkcie (d) wartość licznika jest uzależniona od losowego profilu $R_{1...1000}$ (generowanego poprzez losowe allele z bazy danych populacji), w miejsce profilu S, podczas gdy w mianowniku od wartości $U_1 + U_2$, tak samo jak w opisie powyżej, dla hipotezy że H_p *jest prawdziwe*.

Przykładowo, jeśli $E = acd$, a $R_1 = ef$ w liczniku, trafienie będzie miało miejsce tylko w przypadku, gdy zarówno *e*, jak i *f*, wypadną; allele od nieznanego dawcy będą pasować do kombinacji *a*, *c* lub *d*; lub gdy nastąpi minimum jedno jednoczesne wypadnięcie tłumaczące nieznane allele. Zwykle oznacza to, że LR będzie silnie przemawiało na korzyść hipotezy obrony (ryc. 2b).

Należy zwrócić uwagę, że w tej procedurze genotyp podejrzanego (*ab*) może być wybrany losowo. Prawdopodobieństwo podobnego zdarzenia na wszystkich loci jest jednakże bardzo małe dla nowoczesnego multipleksu. W tradycyjnej interpretacji prawdopodobieństwo dowodu dla hipotezy oskarżenia byłoby w większości przypadków bliskie zeru. Jest to spowodowane faktem, że allele prawdziwego sprawcy nie będą takie same jak losowo generowane. W modelu LCN, nawet gdyby doszło do takiego zdarzenia, z powodu wypadnięcia lub kontaminacji wynik nie musiałby się jednak zmienić.

Procedurę powtarza się wiele razy w celu określenia przebiegu dystrybuanty LR przy założeniu, że H_d *jest prawdziwe*. Większość wartości LR powinna być mniejsza od 1, ponieważ dowód jest bardziej prawdopodobny przy założeniu, że hipoteza obrony jest prawdziwa, niż przy założeniu odwrotnym, ale mogą być także zaobserwowane LR większe od 1. Należy zwrócić uwagę, że autorzy unikają określeń prawdziwy/fałszywy do opisu tych sytuacji. Ponieważ dopuszczają możliwość wystąpienia alleli od prawdziwego sprawcy, szansa, że w jednym locus nastąpi trafienie, jest niewielka.

Procedura w przypadku mieszanin od wielu dawców

Omawianą metodę można zastosować w przypadku mieszaniny od wielu dawców. Zasada jest w skrócie następująca: zakłada się, że H_d *jest prawdziwe*, a E jest symulowane poprzez obliczenie $Pr(D)$ dla każdego allela w każdym locus znanych osób (ryc. 2). Argumentem w liczniku są profile przypadkowo wybranych osób, które zastępują profile osób znanych. Przy założeniu, że H_p *jest prawdziwe*, symuluje się E od znanych osób po uwzględnieniu wypadania w taki sam sposób (ryc. 3). Argumentem w liczniku są profile znanych osób. Allele z nieznanych źródeł traktowane są jako stałe w liczniku – zarówno dla założenia, że H_p *jest prawdziwe* jak i H_d *jest prawdziwe*.

Wprowadzenie parametrów kwalifikujących do opisania siły dowodu

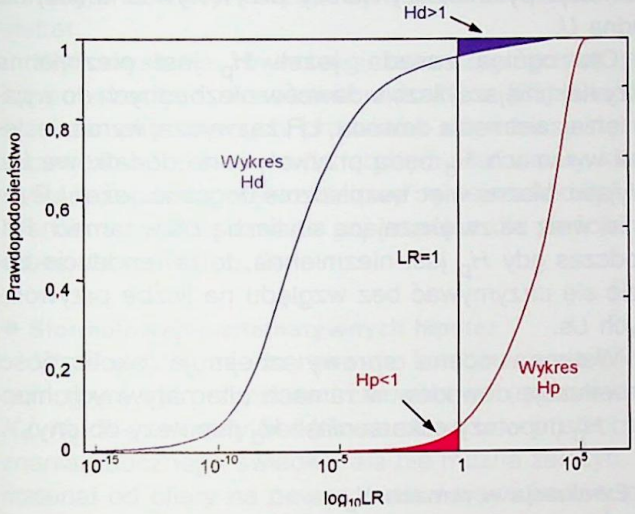
W ogólnych zilustrowanych przykładach (ryc. 2 i 3) mianownik LR, $Pr(E \mid H_d)$ jest obliczany z przyjęciem założenia, że dwie nieznane osoby to prawdziwi dawcy, od których materiał znalazł się w śladzie dowodowym, np. H_d: $U_1 + U_2$. Oczekuje się, że dystrybuanta LR powinna wykazywać wartości powyżej jeden. Rzadkie przypadki, kiedy LR wynosi mniej niż jeden, są uważane za błędne dowody, świadczące na korzyść obrony.

Typowy wykres ukazuje rycina 4. Przede wszystkim interesująca jest proporcja obszaru, dla którego LR jest większy niż jeden w hipotezie H_d. Zgodnie z Neumannem i wsp. [17] definiuje się to jako prawdopodobieństwo błędnego dowodu na korzyść oskarżenia $Pr(LR_{H_d} > 1)$. Ciekawa jest również największa wartość zaobserwowanego LR, $LR_{H_{dmax}}$. Dowodowy LR jest kwalifikowany przez te dwa parametry. Rozważany oddzielnie LR jest nieco uproszczonym wskaźnikiem siły dowodu. Aby umieścić go w prawidłowym kontekście, autorzy proponują równoległe podawanie oceny działania modelu interpretacji DNA. Wskaźnikiem oceny H_d jest więc kombinacja dwóch parametrów: prawdopodobieństwa $Pr(LR_{H_d} > 1)$ sprzężonego z $LR_{H_{dmax}}$.

W odwrotnym przypadku obliczone może być także prawdopodobieństwo mylnego dowodu na korzyść obrony, wyrażonego dwoma parametrami $Pr(LR_{H_p} < 1)$ i $LR_{H_p min}$.

Definiując efektywność dowolnego testu statystycznego w kategoriach parametrów $LR_{H_p min}$, $LR_{H_{dmax}}$, $Pr(LR_{H_d} > 1)$ i $Pr(LR_{H_p} < 1)$, można obiektywnie ocenić wpływ „ulepszeń" w modelach statystycznych poprzez porównanie otrzymanych wyników. W „ulepszonym" modelu wykresy dystrybuanty widoczne na ryci-

nie 3 odsunęłyby się od siebie. W odwrotnym przypadku, w modelu, który źle działa, zbliżyłyby się. Ponieważ w niniejszej publikacji rozważane są modele specyficzne dla sprawy, ważne jest, aby każda ocena ogólnej rzetelności danego modelu została dokonana z użyciem różnych kategorii spraw – na przykład mieszaniny od dwóch dawców z nierówną ilością składników wobec mieszaniny od dwóch dawców w równych proporcjach wobec mieszaniny materiału od trzech dawców.



Ryc. 4. Cechy wykresu Tippetta.
Wykres przedstawia dystrybuantę rozkładu Pr($E \mid H_p$) oraz Pr($E \mid H_d$) i ilustruje możliwe do zmierzenia parametry oceny. Należy zauważyć, że to wysokość krzywych w punkcie LR = 1 daje informację o prawdopodobieństwie.

Taka ewaluacja pomogłaby ustalić nie tylko względną sprawność (efektywność) różnych modeli, lecz także dałaby ocenę ich sprawności i ograniczeń w odniesieniu do pewnych kategorii przypadków. Ocenia się, że wielkość otrzymanego LR i odległość dwóch krzywych powinna być skorelowana.

Podsumowanie procesu oceny złożonego profilu DNA

Definicja modelu. W pierwszym etapie należy:

1. Ustalić hipotezy H_p i H_d.
2. Ustalić minimalną liczbę dawców potwierdzającą dowód.
3. Ocenić wstępnie dowód w odniesieniu do H_p, korzystając z modelu ilościowego, aby upewnić się, że jest racjonalny (np. poprzez oględziny). Należy upewnić się, że wysokości/pola powierzchni pików w ramach H_p są zgodne, oraz uwzględnić inne profile DNA i dowody niezawierające DNA.

Kolejny krok to dokonanie ewaluacji H_d za pomocą modelu jakościowego, w którym rozważane są wszystkie genotypy bez względu na wysokość/pole powierzchni pików. W razie konieczności należy zmienić hipotezę, aby obliczyć górną granicę LR, gdzie Pr(St) = 1 lub w celu obliczenia dolnej granicy modelu LR, gdzie Pr(St) = 0. W razie wystąpienia większej liczby alleli w profilu dowodowym konieczne będzie zwiększenie minimalnej liczby dawców.

Następnie należy dokonać ewaluacji LR w przedziale $0 \leq \Pr(D) \leq 1$ i ustalić racjonalny przedział dla Pr(D), wykorzystując:

– opinię eksperta wydaną na podstawie oględzin,
– symulację liczby dawców relatywnej do Pr(D),
– obliczenia zaobserwowanego Pr(D) w ramach H_p.

Na końcu należy wybrać racjonalną wartość ilorazu wiarygodności LR_{min} porównywalną do oceny przez eksperta Pr(D).

- **Testowanie rzetelności modelu**

Wykonanie wykresów Tippetta:

a. Dla założenia, że hipoteza obrony H_d, *jest prawdziwa*:

Dla każdego z profili K_1, K_2 i K_3 należy stworzyć symulację dowodu E dla każdego z alleli j na locus i. Dla każdego allela A (i i j) prawdopodobieństwo będzie wynosić Pr(D), tzn. jeżeli tak się nie stanie, wówczas allel wypada. Następnie wszystkie allele, które nie mogą być wyjaśnione poprzez K_1–K_3, a które określono jako stuttery, są dodawane do E (niezmienionego).

Znane genotypy K_1, K_2 i K_3 zastępuje się trzema przypadkowymi osobami (R_1, R_2 i R_3), które są argumentem w liczniku podczas testowania LR, podczas gdy w alternatywnej hipotezie argumentem w mianowniku jest U_1, U_2 i U_3 (i w razie potrzeby U_4).

b. Dla założenia, że hipoteza oskarżenia H_p *jest prawdziwa*:

Należy stworzyć symulację dowodu E od znanych i nieznanych osób zgodne z opisem dla założenia H_d *jest prawdziwa* (biorąc pod uwagę, że symulowane dowody w tym przypadku dadzą inny zestaw wyników). Argumentem w liczniku są profile znanych osób (K_1, K_2, K_3) (i w razie potrzeby allele nieznane lub stuttery), podczas gdy w alternatywnej hipotezie w mianowniku są U_1, U_2, U_3 (i w razie potrzeby U_4).

Aby otrzymać zakres LR, należy powtórzyć powyższe dwie czynności więcej niż 1000 razy.

c. Wykreślenie i porównanie dystrybuant.

Analiza prawdziwego przypadku

Okoliczności sprawy, wstępna ocena, oznaczanie alleli, liczba dawców itp.

Zdarzenie miało miejsce w pubie, gdzie zmarły spędził wieczór z kilkoma przyjaciółmi. Na parkingu doszło do sprzeczki między ofiarą (K_1) i kilkoma innymi osoba-

mi, w wyniku czego ofiara poniosła śmierć. Następnie domniemani sprawcy opuścili miejsce zdarzenia i poszli do innego pubu, gdzie widziano ich, jak idą do toalety, aby się umyć.

Chociaż na odzieży podejrzanych wykryto wiele plam krwi, kluczowe znaczenie dla dochodzenia miało odkrycie dwóch plam w łazience pubu. One to bowiem zostały zanalizowane z użyciem systemu SGM Plus. Profile z każdej plamy miały cechy mieszaniny pochodzącej od trzech osób. Podejrzanego, zmarłego i jeszcze jednej ofiary nie można było wyeliminować jako potencjalnych dawców (tab. 1).

Analiza plamy MC18

Wstępna ocena profilu DNA wykazała, że jest to złożona mieszanina materiału pochodzącego od co najmniej trzech osób. Pokrywały się one z profilami trzech znanych osób oznaczonych K_1, K_2, K_3. Ponadto wystąpiło wiele dodatkowych alleli, które mogły być stutterami albo kombinacjami stutter/allel. Są one zaznaczone jako dwuznaczne allele w tabeli 1.

- **Obliczenie górnej i dolnej granicy ilorazów wiarygodności – stutterów i liczby dawców**

Obecna (standardowa) praktyka umożliwia ekspertowi przypisanie prawdopodobieństwa stuttera w sposób binarny głównie na podstawie opinii eksperta. To znaczy, że Pr(St) przyjmuje wartość 0 albo 1 na podstawie tej opinii. Opinia będzie przede wszystkim, ale nie tylko, ukierunkowana przez względne rozmiary stutterów w porównaniu z pikiem allela +4 bp [7]. W niniejszej pracy zostanie wprowadzona nowa metoda uwzględniania wpływu stuttera w obliczeniach.

„Górną granicę LR" można obliczyć, przyjmując Pr(St) = 1 poprzez uznanie, że wszystkie allele na niskim poziomie w pozycjach stutter są prawdziwymi stutterami bez zawartości alleli i można je pominąć w dalszych obliczeniach, traktując jako artefakty. W pewnym sensie ta górna granica będzie „skoncentrowana na oskarżeniu".

Można także obliczyć „dolną granicę LR", przyjmując Pr(St) = 0. Znaczy to, że wszystkie allele w pozycjach stutterów są uważane za właściwe allele (ze stutterami lub bez) i dlatego muszą być traktowane jako dowody w dalszych obliczeniach. W pewnym sensie ta dolna granica będzie „skoncentrowana na obronie". Podejście ukierunkowane na dolną granicę wymaga, aby zarówno do licznika, jak i mianownika, dodać nieznaną osobę (*U*) i to zazwyczaj zmniejszy wartość LR [18]. Powstaje pytanie, czy należy przywoływać więcej niż jedną *U*.

Oto ogólna zasada: jeżeli H_p jest niezmienna przy najmniejszej liczbie dawców niezbędnych do wyjaśnienia zaistnienia dowodu, LR zazwyczaj wzrośnie, jeżeli w ramach H_d będą przywoływane dodatkowe *U*s. Wyjątki: Można więc bezpiecznie uogólnić, jeżeli LR rośnie, wraz ze zwiększającą się liczbą *U*s w ramach Hd, podczas gdy H_p jest niezmienna, to ta tendencja będzie się utrzymywać bez względu na liczbę przywołanych *U*s.

Wstępna ocena sprawy obejmuje okoliczności i ewaluację dowodów w ramach alternatywnych hipotez: H_p (hipotezy oskarżenia) i H_d (hipotezy obrony).

- **Ewaluacja w ramach H_p**

W ramach H_p prowadzi się ewaluację jakościową, co oznacza, że ocenia się wysokości/pola powierzchni pików alleli w profilu DNA: proces ten jest prowadzony zgodnie z opinią eksperta w celu zapewnienia, że zaobserwowany dowód w postaci profilu DNA jest zgodny z założoną hipotezą. Na tym etapie uwzględnia się równowagę heterozygotyczną.

Oskarżenie utrzymywało, że profil DNA jest mieszaniną materiału od trzech osób – odpowiednio K_1, K_2, K_3, gdzie K_1 to zmarła ofiara, K_2 był obecny na miejscu zdarzenia, ale nie został oskarżony o morderstwo, a K_3 jest podejrzany o pchnięcie ofiary nożem. Jeżeli to wszystko stworzy warunkową hipotezę, wówczas $Pr(E \mid H_p) \approx 1$ (pod warunkiem, że proporcje alleli są rozsądne) [18]. Analizę względnych proporcji mieszaniny (M_x) albo stosunek mieszaniny (M_r) można łatwo

Tabela 1

Profile SGM+ dla każdego śladu z miejsca zdarzenia oraz domniemanych dawców

	AMEL	D3	VWA	D16	D2	D8	D21	D18	D19	TH01	FGA
K_1	X,Y	15,18	14,18	12,12	23,24	13,16	30,31	14,16	13,14	7,8	24,26
K_2	X,Y	17,17	16,16	11,12	24,24	13,14	29,30	14,14	15,16.2	9,9	21,22
K_3	X,Y	17,19	15,19	11,13	16,17	10,11	28,30	12,16	14,15	9,3;9,3	20,23
Ślad MC18											
Jednoznaczne allele		15,17,18,19	14,15,16,18,19	11,12,13	16,17,23,24	10,11,13,14,16	28,30,31	12,14,16	13;14;15;16,2	7,8,9,9	20,22,23,24,26
Niejednoznaczne allele (stuttery)		14,16	17		22	12,15	29	13,15	12		22,25
Ślad MC15											
Jednoznaczne allele	X,Y	15,17,18,19	14,15,16,18,19	12,13	16,17,23,24	10,11,13,14,16	28,30,31	12,14,16		7;8;9,3	
Niejednoznaczne allele (stuttery)		14	17	11	22	12,15	29	13,15	13,14,15,12		20,21,23,24,26
Brakujące allele w Hp									16,2	9	25,22

ustalić na podstawie proporcji alleli [22]. To daje współczynniki dawców (równoważne z M_r) genotypów większościowy: mniejszościowy: mniejszościowy: – stutter/nieznana domieszka jako odpowiednio 0,67:0,11:0, 19:0,04 (tab. 2). Oględziny potwierdziły, że wszystkie większościowe allele odpowiadają w każdym locus ofierze, lecz w każdym locus zaobserwowano allele odpowiadające dwóm mniejszościowym dawcom. Nie było niezgodności. Ponadto zaobserwowano niewielki komponent, który można było zakwalifikować jako stutter.

Wykorzystanie równowagi heterozygotycznej jest już powszechnie stosowane w praktyce [23]. Dla dobrze zamplifikowanych produktów używana jest wytyczna Hb > 0,6 służąca do łączenia alleli w pary, kiedy występuje dawca większościowy/ mniejszościowy i allele nie są zamaskowane ani wspólne dla dwóch lub więcej dawców (tab. 2).

- **Sformułowanie alternatywnych hipotez**

W ramach H_p zaobserwowano kombinację alleli, które mogłyby być przypisane trzem osobom (odpowiednio K_1, K_2, K_3). Istniały dodatkowe dowody w postaci zeznania naocznego świadka, ale nie można założyć, że materiał od ofiary na pewno znalazł się w śladzie, ponieważ plamę znaleziono w pewnej odległości od miejsca zdarzenia. Alternatywne wyjaśnienie, H_d, jest takie, że kombinacja alleli pochodziła od trzech albo więcej nieznanych osób (U_1, U_2, U_3). Te dwie hipotezy tworzą górną granicę LR, gdzie zakłada się, że allele w pozycjach stutter są naprawdę stutterami. Tak wyglądałoby klasyczne podejście do interpretacji mieszanin.

- **Stuttery**

W rzeczywistości prawdopodobieństwo, że pik naprawdę jest stutterem, Pr(St) musi wypadać gdzieś między wartościami 0 i 1. Dolna granica LR jest formułowana poprzez założenie, że wszystkie stuttery to allele, a więc Pr(St) = 0 oraz że te dodatkowe allele musiały pochodzić od dodatkowej osoby/osób. Rozważane hipotezy to H_p: $K_1 + K_2 + K_3 + U$ i w alternatywnej hipotezie H_d: $U_1 + U_2 + U_3 + U_4$.

Podsumowując, *H*p ocenia się za pomocą modelu ilościowego (wysokość/pole powierzchni piku), aby zapewnić, że hipoteza dotycząca dowodu jest zgodna i rozsądna (jak opisano wcześniej). Natomiast aby ocenić *H*d, używa się modelu jakościowego (ignorując wysokość/pole powierzchni piku), który obejmuje wszystkie kombinacje alleli [4]. To podejście jest zawsze bardzo konserwatywne, ponieważ obejmuje wiele kombinacji, które mogą tylko zwiększyć prawdopodobieństwo dowodu w ramach H_d. Nie ma odrzucania genotypów na podstawie cech ilościowych [23].

- **Model statystyczny**

Model statystyczny jest uwarunkowany założeniami opisanymi powyżej: oblicza się ilorazy wiarygodności górnej i dolnej granicy odnoszące się do racjonalnego przedziału prawdopodobieństwa wypadania alleli Pr(*D*) (ryc. 5). Prezentowana tu metoda została szczegółowo opisana przez Gilla i wsp. [6], w omawianym przypadku użyto lekko zmodyfikowanego wzoru w celu oszacowania prawdopodobieństwa wypadania.

- **Przedstawianie LR w odniesieniu do Pr(*D*)**

W niniejszym przykładzie LR zmniejszał się w miarę, jak wartość Pr(*D*) się zwiększała (ponieważ wszystkie allele zaobserwowano w ramach H_p). Rozsądny przedział dla Pr(*D*) został oszacowany drogą symulacji [6]. Przedstawiony w opinii LR = 107 (odpowiadał Pr(*D*) = 0,25).

- **Ślad MC15**

Podobną analizę przeprowadzono na śladzie MC15 (tab. 3). Zastosowano to samo uwarunkowanie. Profil występujący w większości zawierał 80% całkowitego profilu – wszystkie połączone w pary allele składnika większościowego były zgodne z allelami ofiary. Allele od dawcy mniejszościowego były zbyt niskie, zgodnie z wytycznymi dla wartości *H*b, aby można je było połączyć z jakimikolwiek allelami od dawcy większościowego.

Od dwóch mniejszościowych dawców pochodziło odpowiednio 6 i 12% profilu, łącznie ze stutterami podobnej wielkości. Ten M_x był granicą dla przedstawiania opinii o mieszaninie na zasadzie mniejszościowy/większościowy, ponieważ allele były zbliżone do zakłóceń tła i granic identyfikacji. Aby ustalić rozsądny przedział prawdopodobieństwa wypadania, zastosowano przesłankę opisaną wcześniej dla MC18. Tym razem LR wzrastał wraz ze wzrostem Pr(*D*) – niektórych alleli brakowało, więc Hp można wytłumaczyć tylko zaistnieniem dropoutu.

W ramach H_p wypadły w sumie trzy allele: 16,2 w D19, allel 9 w TH01 i allel 22 w FGA, stąd zaobserwowana wartość w ramach H_p dla prawdopodobieństwa wynosi Pr(*D*) = 3/41 albo około 0,07. Dolna granica LR = 10^4 w 95 centylu (ryc. 6). Była to wielkość zamieszczona w opinii.

Zdefiniowawszy model, autorzy tekstu wprowadzają oddzielną metodę testowania jego działania: w tym przykładzie model obejmuje wybór Pr(*D*) użytego do obliczenia LR, przy odnotowaniu, że LR jest całkowicie uwarunkowany założeniami dla modelu. Konkretnie jesteśmy zainteresowani liczbą wyników, które dają LR > 1 w ramach H_d i odwrotnie – liczbą wyników, które dają LR < 1 w ramach H_p wraz z odpowiadającą wy-

Tabela 2

Analiza próbki MC18. Dla każdego locus podane są przypisane allele znanych osób *K*1, *K*2, *K*3 oraz nieznanej osoby (*U*)

D3		wysokość	% całości		K_1	K_2	K_3	*U*
14S		50	2		15,18	17,17	17,19	14,16,18
15		715	34	proporcja mieszaniny	0,65	0,16	0,13	0,06
16S		67	3	*Hb*	0,89			
17		479	23					
18		638	31					
19		136	7					
	suma	2085						
VWA					K_1	K_2	K_3	*U*
14		876	32		14,18	16,16	15,19	17
15		249	9	proporcja mieszaniny	0,68	0,1	0,18	0,04
16		274	10	*Hb*	0,91		0,99	
17S		97	4					
18		967	36					
19		251	9					
	suma	2714						
D16					K_1	K_2	K_3	*U*
11		534	21		12,12	11,12	11,13	brak
12		1786	69	proporcja mieszaniny	0,59	0,21	0,21	0
13		265	10					
	suma	2585						
D2					K_1	K_2	K_3	*U*
16		189	11		23,24	24,24	16,17	22
17		171	10	proporcja mieszaniny	0,74	0,02	0,21	0,03
22S		55	3	*Hb*	0,95		0,9	
23		638	37					
24		673	39					
	suma	1726						
D8					K_1	K_2	K_3	*U*
10		241	9		13,16	13,14	10,11	12,15
11		192	7	proporcja mieszaniny	0,61	0,16	0,16	0,07
12S		127	5	*Hb*	0,74		0,8	
13		1092	41					
14		127	5					
15S		58	2					
16		808	31					
	suma	2645						
D21					K_1	K_2	K_3	*U*
28		304	13		30,31	29,3	28,30	brak
29S		134	6	proporcja mieszaniny	0,63	0,12	0,26	
30		1146	49	*Hb*	0,64			
31		734	32					
	suma	2318						
D18					K_1	K_2	K_3	*U*
12		187	8,9		14,16	14,14	12,16	13,15
13S		87	4,2	proporcja mieszaniny	0,71	0,121	0,17852	0,08
14		997	47,6	*Hb*	0,75			
15S		80	3,8					
16		744	35,5					
	suma	2095						

cd. tabeli 2

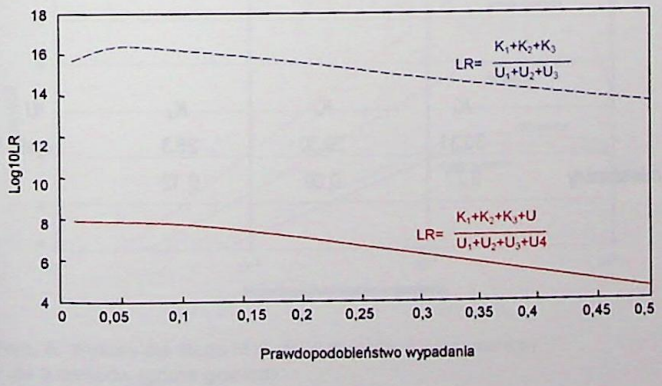
D19				K_1	K_2	K_3	U
12S		57	3	13,14	15;16,2	14,15	12
13		775	41	proporcja mieszaniny	0,82	0,08	0,09	0,03
14		818	43	*Hb*	0,95			
15		159	8					
16.2		76	4					
	suma	1885						
TH01				K_1	K_2	K_3	U
7		670	38		7,8	9,9	9,3;9,3	brak
8		636	36	proporcja mieszaniny	0,75	0,06	0,2	
9		99	6	*Hb*	0,95			
9,3		348	20					
	suma	1753						
FGA				K_1	K_2	K_3	U
20		99	8		24,26	21,22	20,23	25
21		49	4	proporcja mieszaniny	0,65	0,11	0,21	0,03
22		76	7	*Hb*	0,85	0,64	0,68	
23		145	12					
24		412	35					
25S		39	3					
26		349	30					
	suma	1169		SUMA	0,67	0,11	0,19	0,04

sokością LR. Autorzy zbadali wykorzystanie testów Tippetta [4, 14] do oceny działania modelu dla konkretnego przypadku.

Wyniki testu Tippetta

Założenie, że dawcami są cztery osoby i że stuttery to allele

Z danych MC18 (ryc. 7) symulacja w ramach H_d (po lewej stronie) nigdy nie przekroczyła 1, stąd proporcja wydarzeń, gdzie $LR_{Hd} > 1$, wynosi zero. Ponieważ wielkość próbki jest ograniczona, obliczono gęstość prawdopodobieństwa za pomocą nieparametrycznej estymacji gęstości (ang. *kernel density estimates*, w skr. KDE). Z wykorzystaniem tej metody $Pr(LR_{Hd} > 1) = 0{,}008\%$. Największy zaobserwowany LRs wśród 1000 symulacji to $LR_{Hdmax} \approx 0{,}02$. Symulacja w ramach H_p (po prawej stronie) daje 213 z 1000 LR, które są mniejsze niż 1 w ramach H_p (21,3%). Używając KDE, szacuje się to jako 22,7%. Na podstawie obserwacji 1000 symulacji, $LR_{Hpmin} = 10^{-27}$; $LR_{Hpmax} = 10^{7,5}$. Podany w opinii LR to 10^7, który mieści się w obserwowanym przedziale. Najważniejsza sprawa to możliwość wykazania, że model jest rzetelny, ponieważ wszystkie oszacowania LR_{Hd} są mniejsze niż jeden (ryc. 8).



Ryc. 5. Analiza górnej i dolnej granicy ilorazów wiarygodności dla śladu MC18 z wykorzystaniem modeli „czterech dawców" i „trzech dawców"

Co więcej – wykazano, że wykresy H_p i H_d są wyraźnie oddzielone z niewielkim nakładaniem się (około 4%), co oznacza, że model dobrze działa.

Dla wyników MC15 1 z 1000 LR jest większy od 1 w ramach H_d. Za pomocą KDE jest to szacowane jako 0,09%. Największy LR zaobserwowany w ramach H_d wynosił $LR_{Hdmax} = 10{,}3$. Dodatkowo istnieje 0 na 1000 LR mniejszych niż 1 w ramach H_p. LR zaobserwowany w ramach H_p wynosił $LR_{Hpmin} = 1137$; $LR_{Hpamx} = 10^{12}$, a LR w opinii to 10^4, na dolnej granicy przedziału. Model wydaje się rzetelny, ponieważ dystrybuanty H_p i H_d nie nakładają się oraz istnieje jeden przypadek $LR_{Hd} > 1$.

Tabela 3

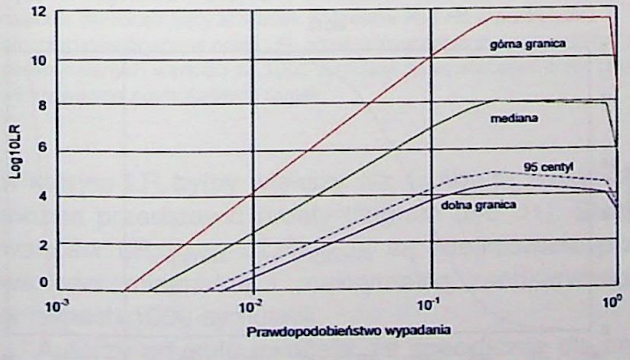
Analiza MC15. Dla każdego locus przedstawione są allele znanych osób K_1, K_2, K_3 i nieznanej osoby (U).

D3		wysokość	% całości		K_1	K_2	K_3	U
14S		79	4		15,18	17,17	17,19	14,16,18
15		993	46	proporcja mieszaniny	0,77	0,07	0,12	0,04
16S		0	0	*Hb*	0,69			
17		286	13					
18		689	32					
19		135	6					
	suma	2182						
VWA					K_1	K_2	K_3	U
14		1036	47		14,18	16,16	15,19	17
15		98	4	proporcja mieszaniny	0,81	0,07	0,08	0,04
16		163	7	*Hb*	0,72		0,87	
17S		79	4					
18		746	34					
19		85	4					
	suma	2207						
D16					K_1	K_2	K_3	U
11S		256	12		12,12	11,12	11,13	brak
12		1724	83	proporcja mieszaniny	0,76	0,14	0,10	0
13		109	5					
	suma	2089						
D2					K_1	K_2	K_3	U
16		64	5		23,24	24,24	16,17	22
17		96	8	proporcja mieszaniny	0,84	0,02	0,13	0,03
22S		0	0	*Hb*			0,67	
23		507	42					
24		534	44					
	suma	1201						
D8					K_1	K_2	K_3	U
10		152	6		13,16	13,14	10,11	12,15
11		140	6	proporcja mieszaniny	0,77	0,04	0,12	0,07
12S		76	3	*Hb*			0,92	
13		929	40					
14		58	2					
15S		84	4					
16		901	39					
	suma	2340						
D21					K_1	K_2	K_3	U
28		120	6		30,31	29,30	28,3	brak
29		89	4	proporcja mieszaniny	0,77	0,09	0,12	
30		1010	51					
31		763	38					
	suma	1982						

cd. tabeli 3

D18				K_1	K_2	K_3	U
12S	99	5,2		14,16	14,14	12,16	13,15
13	61	3,2	proporcja mieszaniny	0,98	0	0,1	0,09
14	707	37,1					
15	107	5,6					
16	930	48,8					
suma	1904						
D19				K_1	K_2	K_3	U
12	53	4		13,14	15;16,2	14,15	12
13	546	40	proporcja mieszaniny	0,81	dropout	0,14	0,04
14	655	48					
15	98	7					
suma	1352						
TH01				K_1	K_2	K_3	U
6S	0			7,8	9,9	9,3;9,3	brak
7	727	48	proporcja mieszaniny	0,89	dropout	0,11	
8	625	41	*Hb*	0,86			
9	0	0					
9.3	165	11					
suma	1517						
FGA				K_1	K_2	K_3	U
20	90	8		24,26	21,22	20,23	25
21	52	4	proporcja mieszaniny	0,79	0,04	0,16	0
22	0	0	*Hb*	0,71	dropout	0,87	
23	103	9					
24	556	47					
25S	0	0					
26	392	33					
suma	1193		SUMA	0,8	0,06	0,12	0,04

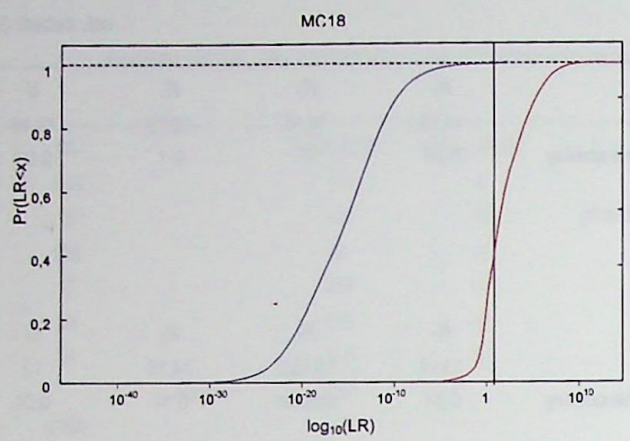
Jeżeli występuje suffix (S), wówczas allel może być stutterem. Jeżeli heterozygota jest niejednoznaczna dla danego dawcy (tzn. nie ma maskowania), wówczas oblicza się Hb jako proporcję mniejszą od jednego (kolor czerwony).



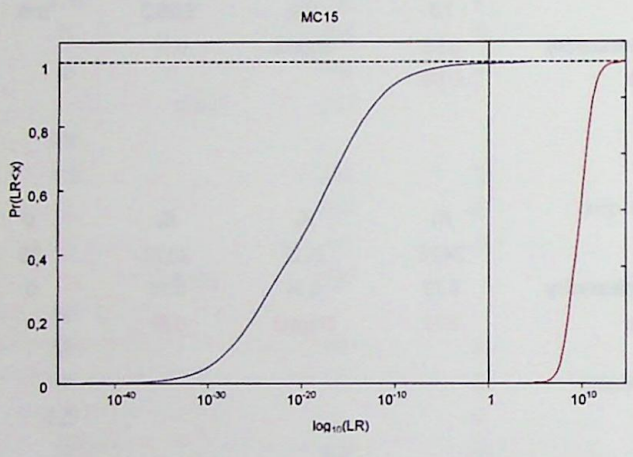
Ryc. 6. Wykres dla śladu M15 dla 4 dawców (dolna granica) i dla 3 dawców (górna granica)

Wyniki trzyosobowych testów Tippetta przy założeniu, że stuttery to nie allele

Alternatywny scenariusz do rozważenia to model trzyosobowy (ryc. 9 i 10) z założeniem, że wszystkie allele w pozycjach stutter to wyłącznie stuttery, bez komponentu alleli. Jak oczekiwano, zwiększa się odległość między wykresami H_p i H_d. Największy efekt to zmniejszenie wartości LR_{Hdmax}. Wzrost LR podanych w opinii jest odzwierciedlony w rozkładzie LR_{Hp} na wykresach Tippetta.



Ryc. 7. Wykres Tippetta dla MC18: model czteroosobowy



Ryc. 8. Wykres Tippetta dla MC15: model czteroosobowy

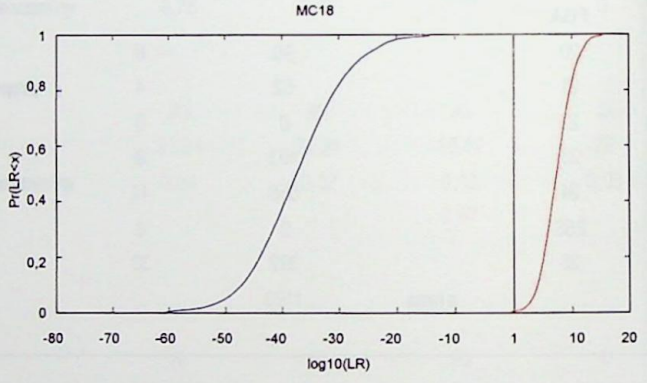
Omówienie wyników

Eliminacja „nierozstrzygającego" profilu DNA

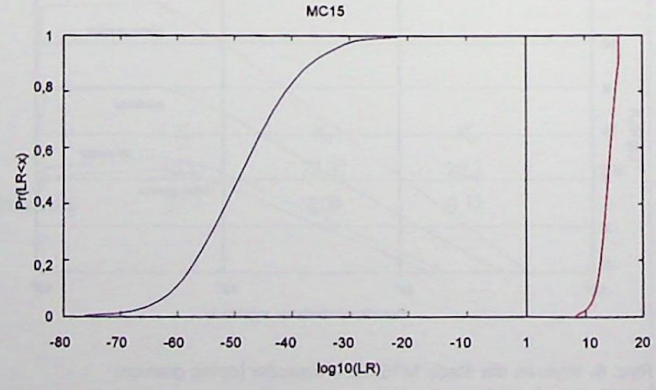
Opisana metoda nie tylko oblicza siłę dowodu, kiedy podejrzany jest dawcą profilu DNA, lecz także oblicza siłę dowodu, kiedy materiał od podejrzanego nie znalazł się w plamie. Najważniejsze, że można zapewnić LR < 1, ponieważ można zanalizować każdy zestaw alleli i zbadać go względem każdego zestawu hipotez, nawet jeżeli allele nie są zgodne z allelami podejrzanego. Jest to spowodowane faktem, że możemy przypisać prawdopodobieństwo „brakującym" allelom w ramach H_p za pomocą Pr(*D*), a allele spoza profilu, które nie są zgodne z allelami podejrzanego, mogą być oznaczone jako zdarzenia „kontaminacji" [1, 6]. Pr(*C*) jest szacowane na podstawie badania kontroli negatywnych [19], podczas gdy Pr(*D*) jest szacowane z sumowania liczby alleli uwarunkowanych liczbą dawców [6].

Biorąc pod uwagę fakt, że obliczenia nie są ograniczone liczbą dawców, w konsekwencji nie ma potrzeby wprowadzania kategorii „nierozstrzygający" dla celów formułowania opinii, która byłaby oparta na stwierdzonej kompleksowości wyniku. Można bowiem dokonać oceny prawdopodobieństwa każdego profilu w ramach dowolnego zestawu hipotez.

Po obliczeniu ilorazu wiarygodności, specyficzne dla sprawy testy Tippetta, dają pogląd na rzetelność testu. Stosuje się symulację komputerową w celu obliczenia przedziału wartości LR, które mogą być oczekiwane dla konkretnej sprawy, kiedy H_p *jest prawdziwa* wobec twierdzenia H_d *jest prawdziwa*. To z kolei prowadzi do oszacowania występowania błędu typu I: jaka jest możliwość zaistnienia mylnego materiału dowodowego na korzyść obrony, oraz błędu typu II: jaka jest szansa mylnego dowodu na korzyść oskarżenia. Przy tradycyjnych metodach obliczenia prawdopodobieństwa jest to istotnie najważniejszy miernik rzetelności danej metody badawczej. Wykazano także, jak można obliczyć przedział LR dla obu założeń.



Ryc. 9. Wykresy Tippetta MC18. Model trzyosobowy, po odliczeniu stutterów, wykazuje zwiększoną separację wykresów H_p i H_d.



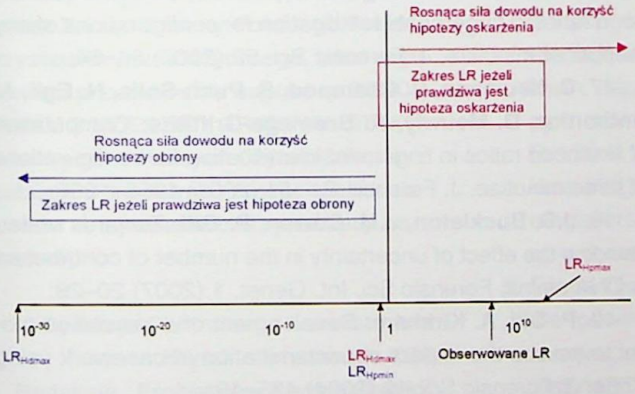
Ryc. 10. Wykresy Tippetta MC15. Analiza trzyosobowej mieszaniny po odliczeniu stutterów.

Autorzy prezentowanego artykułu sugerują, że tę informację można włączyć do przyjaznej dla sądu opinii, która wskazuje rzetelność oszacowania ilorazu wiarygodności. Gdyby wykorzystać wykres nieparametrycznej estymacji gęstości H_d MC18 dla czterech dawców jako przykład (ponieważ nie było przypadku LR > 1 na 1000 symulacji), opinia brzmiałaby następująco: „Jeżeli zamiast z materiału od osób x, y, z ślad składałby się z materiału od trzech przypadkowych osób, wówczas istniałaby tylko szansa jak jeden do 12 500 zaistnienia mylnego dowodu na korzyść wersji obrony".

Ta nowa opinia może być następnie powiązana z tradycyjną opinią LR:

„W tym wypadku siła dowodu jest milion razy bardziej prawdopodobna, jeżeli dawcami materiału w śladzie są osoby x, y, z, niż gdyby dawcami były trzy przypadkowe osoby".

Bez wątpienia potrzebne jest szersze omówienie, aby poprawić jasność i znaczenie tych dwóch opinii. Konieczne byłoby zakwalifikowanie opinii przez dalsze wyjaśnienie: jeżeli dowód zaistniałby w ramach H_d *jest prawdą*, wówczas jego siła byłaby bardzo mała, ponieważ 1000 symulacji nie dało żadnego przypadku,



Ryc. 11. Sugerowany „przyjazny dla sądu" model MC18 dla czterech dawców demonstrujący stosunek wykresów *Hp* i *Hd* w porównaniu do przedstawionego w opinii LR, gdzie odpowiednie maksima/minima zaobserwowanych wartości na 1000 symulacji są wyznaczane przez pionowe krawędzie prostokątnych ramek.

w którym LR byłby większy niż 1. Aby to zilustrować, można przedstawić prosty diagram (ryc. 11). Skrajne wartości LR_{Hpmin} i LR_{Hdmax} są zdefiniowane przez wartości minimalną i maksymalną zaobserwowane w ramach 1000 symulacji.

Autorzy artykułu uważają, że specyficzne dla omawianego przypadku wykresy Tippetta są także użyteczne w celu testowania różnych modeli statystycznych, które mogą być uwarunkowane zupełnie innymi zasadami. W opisanych przykładach porównywano model uwzględniający wystąpienie stuttera (od czterech dawców) z modelem dla trzech dawców (który uwzględniał stuttera). Ten drugi wykazał oczekiwaną „poprawę", ponieważ zmniejszono niepewność w modelu. Ciekawe było to, że największy wpływ, jaki zaobserwowano, to zmniejszenie przedziału wykresu LR dla H_d, w przykładzie MC18 (ryc. 7–9) z $1–10^{-30}$ do $0{,}1–10^{-60}$. W przykładach tego typu autorzy wolą myśleć w kategoriach modeli górnej i dolnej granicy, ponieważ nie stosują w każdym z nich tych samych danych. Najlepszy byłby taki model, który zwiększyłby odległość między wykresami H_p i H_d, przy identycznych warunkach, jeżeli chodzi o liczbę dawców, alleli i stutterów (czyli danych).

Przyszłe prace

Stało się jasne, że można stworzyć wykres H_p *jest prawdą* na wiele różnych sposobów: jeden z ekspertów (C. Neumann) zasugerował, że można też stworzyć wykres *Hp jest prawdą*, jeżeli *E* jest stałe poprzez zmiany Pr(*D*) i Pr(*C*) w „rozsądnym" przedziale. Ten pogląd na pewno nie jest bezpodstawny. Z perspektywy sądu prawdopodobnie najistotniejszą symulacją jest wykres H_d *jest prawdą* – gdzie oceniamy szansę, że materiał w próbce mógłby pochodzić od przypadkowej osoby, a mimo to LR byłby większy niż jeden. LR przedstawiony w opinii pochodzi jednak obecnie z rozkładu H_p *jest prawdą* – autorzy sugerują, że metody przedstawiania opinii w sądzie można ulepszyć, jeżeli zostaną oparte na wykresach Tippetta H_d *jest prawdą*. Obszar, nad którym należy pracować w przyszłości, to studium porównawcze różnych metod mające na celu stworzenie wykresów Tippetta i wykazanie ich względnych możliwości odnoszących się do założeń modelowania.

Implikacje przeszukań złożonych profili w krajowych bazach danych DNA

Autorzy sugerują także, że metodę opracowaną przez Gilla i wsp. [6] można rozszerzyć na przeszukiwanie krajowych baz danych w celu wytypowania potencjalnych trafień.

Można założyć, że popełniono przestępstwo, nie ma podejrzanego, ale profil DNA *E* jest złożony (mieszany), niepełny, i nie da się odpowiednio określić dawcy większościowego i mniejszościowego. Pierwszym krokiem jest przypisanie liczby dawców profilowi DNA, oszacowanie Pr(*D*) i zastosowanie Pr(*C*). H_p byłaby sformułowana następująco $S + U_1 + U_2$, a H_d: $U_0 + U_1 + U_2$ (może być konieczne przeszukanie z kilkoma zestawami różnych propozycji-par). Jeżeli baza danych przestępców zawiera 3m ($R_1 ... R_{3m}$) osób, dokonuje się ewaluacji każdej osoby po kolei, poprzez podstawienie R_n zamiast *S* w liczniku obliczenia LR. Jest to taki sam

protokół, jak sugerowany dla wykresu Tippetta w ramach H_d *jest prawdą*, tyle że w tym scenariuszu *E* (zaobserwowany profil z miejsca zdarzenia) utrzymywano by taki sam we wszystkich testach. Bez względu na złożoność profilu DNA iloraz wiarygodności można by sformułować dla każdego kandydata tak, aby stworzyć uszeregowaną listę na podstawie bazy danych przestępców. Większość testów stworzy rozkład nieciągły (ryc. 7).

Potencjalne trafienia to po prostu osoby, których LR jest większy niż jeden. W szczególności, oczekiwać można, że sprawca zostałby zidentyfikowany jako punkt leżący wyraźnie poza głównym rozkładem, gdzie LR jest dużo większy niż jeden. Poprzez zmianę warunkowej hipotezy, zgodnie z okolicznościami sprawy, metodę łatwo można by rozszerzyć na przeszukania rodzinne baz danych, w których krewni sprawców mogli by być zidentyfikowani na podstawie mieszanin albo niepełnych profilów (a nie tylko jednorodnych śladów) opisane przez Biebera i wsp. [24].

Załącznik A. Dodatkowe dane

Dodatkowe dane związane z tym artykułem można znaleźć w wersji on-line doi:10.1016/j.fsigen.2007.10.160.

oprac. Renata Zbieć i Ewa Nogacka
ryciny i tabele na podstawie oryginału – R. Zbieć

BIBLIOGRAFIA

1. **P. Gill, J.P. Whitaker, C. Flaxman, N. Brown, J. Buckleton:** An inwestigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, Forensic Sci. Int. 112 (2000) 17–40;

2. **J. Buckleton, C. Triggs:** Is the 2p rule always conservative? Forensic Sci. Int. 159 (2006) 206–209;

3. **C.F. Tippett, V. J. Emerson, M.J. Fereday, F. Lawton, A. Richardson, L.T. Jones, S.M. Lampert:** The evidential value of the comparison of paint flakes from sources other than vehicles, J. Forensic Sci. Soc. 8 (1968) 61–65.

4. **I.W. Evett, B.S. Weir:** Interpreting DNA Evidence, Sinauer, Sunderland, MA, 1998;

5. **J. Buckleton, C.M. Triggs, S.J. Walsh:** Forensic DNA Evidence Interpretation, CRC Press, Boca Raton, FL, 2005;

6. **P. Gill, A. Kirkham, J. Curran:** LoComatioN: a software tool for the analysis of low copy number DNA profiles, Forensic Sci. Int. 166 (2006), 128–138;

7. **P. Gill, R. Sparkes, C. Kimpton:** Development of guidelines to designate alleles using an STR multiplex system, Forensic Sci. Int. 89 (1997) 185–197;

8. **J.P. Whitaker, E.A. Cotton, P. Gill:** A comparison of the characteristics of profiles produced with the AMPFISTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis, Forensic Sci. Int. 123 (2001) 215–223;

9. NRC, NRC II: The evaluation of forensic DNA evidence. National Academy Press, Washington, DC, 1996;

10. **B. Robertson, G.A. Vignaux:** Interpreting Evidence: Evaluating Forensic Science in the Courtroom, John Wiley & Sons Ltd., 1995;

11. **B.W. Robertson. G.A. Vignaux:** DNA evidence: wrong answers or wrong questions, Genetica 96 (1995) 145–152;

12. **J.M. Curran, P. Gill, M.R. Bill:** Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substi-ucture, Forensic Sci. Int. 148 (2005) 47–53;

13. **I.W. Evett:** Interpretation of evidence, Int. Crim. Police Rev. 444 (1993), 12–17;

14. **I.W. Evett, P.D. Gill, J.K. Scrariage, B.S. Weir:** Establishing the robustness of short-tandem-repeat statistics for forensic applications, Am. J. Hum. Genet. 58 (1996) 398–407;

15. **I.W. Evett, J.S. Buckleton:** Statistical analysis of STR data, in: A. Carracedo, B. Brinkmann, W. Bar (Eds.), Advances in Forensic Haemogenetics, Springer-Verlag, Berlin, 1996, pp. 79-86;

16. **C. Neumann, C. Champod, R. Puch-Solis, N. Egli, A. Anthonioz, A. Bromage-Griffiths:** Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae, J. Forensic Sci. 52 (2007) 54–64;

17. **C. Neumann, C. Champod, R. Puch-Solis, N. Egli, A. Anthonioz, D. Meuwly, A. Bromage-Griffiths:** Computation of likelihood ratios in fingerprint identification for configurations of three minutiae, J. Forensic Sci. 51 (2006) 1255–1266;

18. **J.S. Buckleton, J.M. Curran, P. Gill:** Towards understanding the effect of uncertainty in the number of contributors to DNA stains, Forensic Sci. Int. Genet. 1 (2007) 20–28;

19. **P. Gill, A. Kirkham:** Development of a simulation model to assess the impact of contamination in casework using STRs, J. Forensic Sci. 49 (2004) 485–491;

20. **P. Gill, R. Sparkes, J.S. Buckleton:** Interpretation of simple mixtures when artefacts such as stutters are present–with special reference to multiplex STRs used by the forensic science service, Forensic Sci. Int. 95 (1998) 213–224;

21. **C.M. Triggs, J.M. Curran:** The sensitivity of the Bayesian HPD method to the choice of prior, Sci. Justice 46 (2006) 169–178.

22. **P. Gill, R. Sparkes, R. Pinchin, C. TM, J.P. Whittaker, J. Buckleton:** Interpreting simple STR mixtures using allele peak area, Forensic Sci. Int. 91 (1998) 41–53;

23. **M. Bill, P. Gill, J.M. Curran, T. Clayton, R. Pinchin, M. Healy, J. Buckleton:** PENDULUM-a guideline-based approach to the interpretation of STR mixtures, Forensic Sci. Int. 148 (2005) 181–189;

24. **F.R. Bieber, C.H. Brenner, D. Lazer:** Human genetics. Finding criminals through DNA of their relatives, Science 312 (2006) 1315–1316.