

Bachelorproef

Wat is de effectiviteit van het gebruik van likelihood ratios (LR) ten opzichte van het aantal loci bij het onderscheiden van DNA-matches in en rond de 'grijze zone' tussen *High 7*, *High 8-9* en *High 10* in forensische DNA-analyses?

Nationaal instituut voor Criminalistiek en Criminologie –
Vilvoordsesteenweg 100, 1120 Brussel

Séverine Steuve – Beheerder van de Nationale DNA-
databanken

Rani Bens – Gerechtelijk deskundige DNA-databanken

Tine Dorné - Forensisch Statistisch deskundige

Hannah Decloedt

Departement Gezondheid, Design & Technologie

Bachelor Biomedische Laboratoriumtechnologie

Farmaceutische en Biologische Laboratoriumtechnologie

Voorwoord

Voor u ligt de bachelorproef: “Wat is de effectiviteit van het gebruik van likelihood ratios (LR) ten opzichte van het aantal loci bij het onderscheiden van DNA-matches in en rond de 'grijze zone' tussen *High 7*, *High 8-9* en *High 10* in forensische DNA-analyses?”. Deze scriptie werd geschreven om mijn laatste jaar van de opleiding Biomedische Laboratoriumtechnologie aan Erasmushogeschool Brussel af te ronden. Ik volgde gedurende twee jaar het traject ‘forensisch onderzoek’, en het was mij dan ook een genoegen om stage te doen in een forensische instelling, namelijk het Nationaal Instituut voor Criminalistiek en Criminologie (NICC).

Mijn stage bij het NICC, van 8 januari 2024 tot 29 maart 2024, heeft me waardevolle inzichten gegeven, zowel op persoonlijk als professioneel vlak. Ik heb geleerd om in teamverband te werken en leerde meer over de werking van de forensische DNA-databanken. Aangezien bijna alles nieuw voor mij was, was ik zeer leergierig.

Graag wil ik mijn promotor, Séverine Steuve, bedanken voor deze unieke kans en om mij met open armen te ontvangen. Ook dank aan mijn copromotor, Rani Bens, om mij gedurende de stageperiode te begeleiden en het grondig doorlezen en bijsturen van mijn bachelorproef. Daarnaast wil ik Tine Dorné bedanken voor haar feedback op mijn bachelorproef, haar hulp met het programma RStudio en de interpretatie van de statistische analyses. Mijn dank gaat ook uit naar Laurent Köhler om mij te laten deelnemen aan het FOR2BIOB-project en de hulp bij de interpretatie van de data. Tot slot wil ik ook mijn familie en vrienden, in het bijzonder mijn mama om mij de kans te geven om te studeren en mij hierin te steunen.

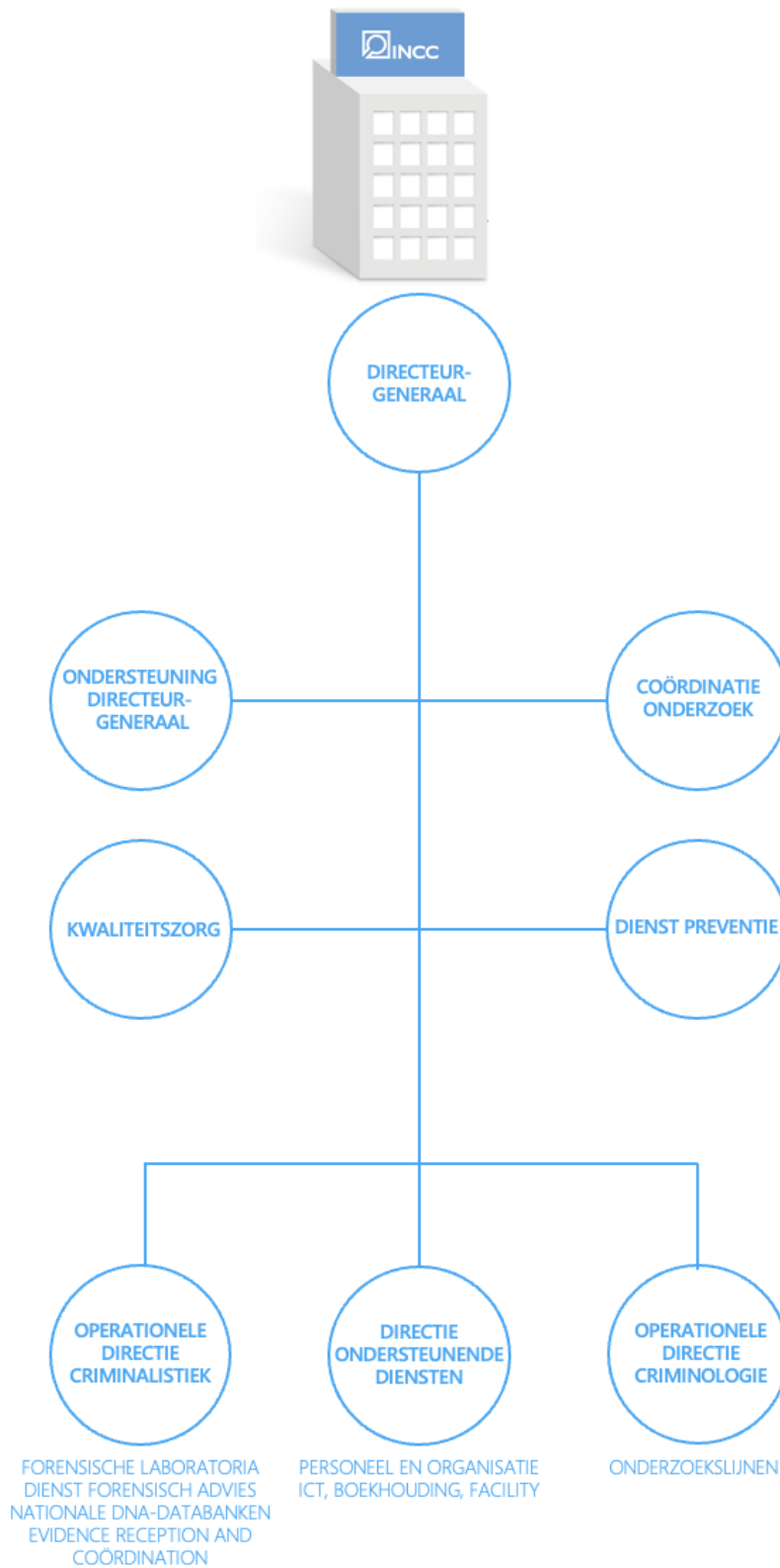
Ik wens u veel leesplezier toe!

Hannah Decloedt

Kaapstad, 12 mei 2024

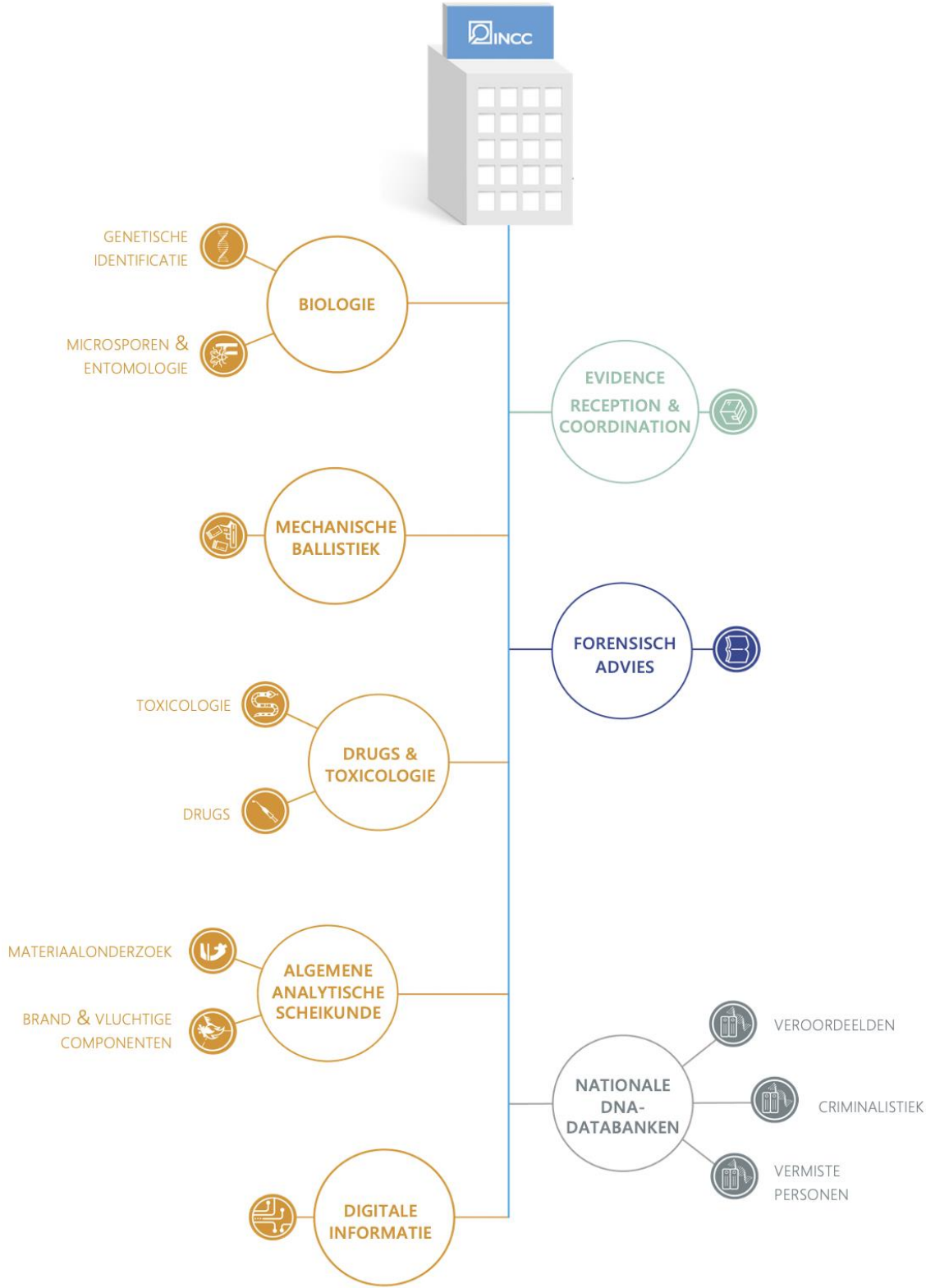
Organigram

Organisatiestructuur



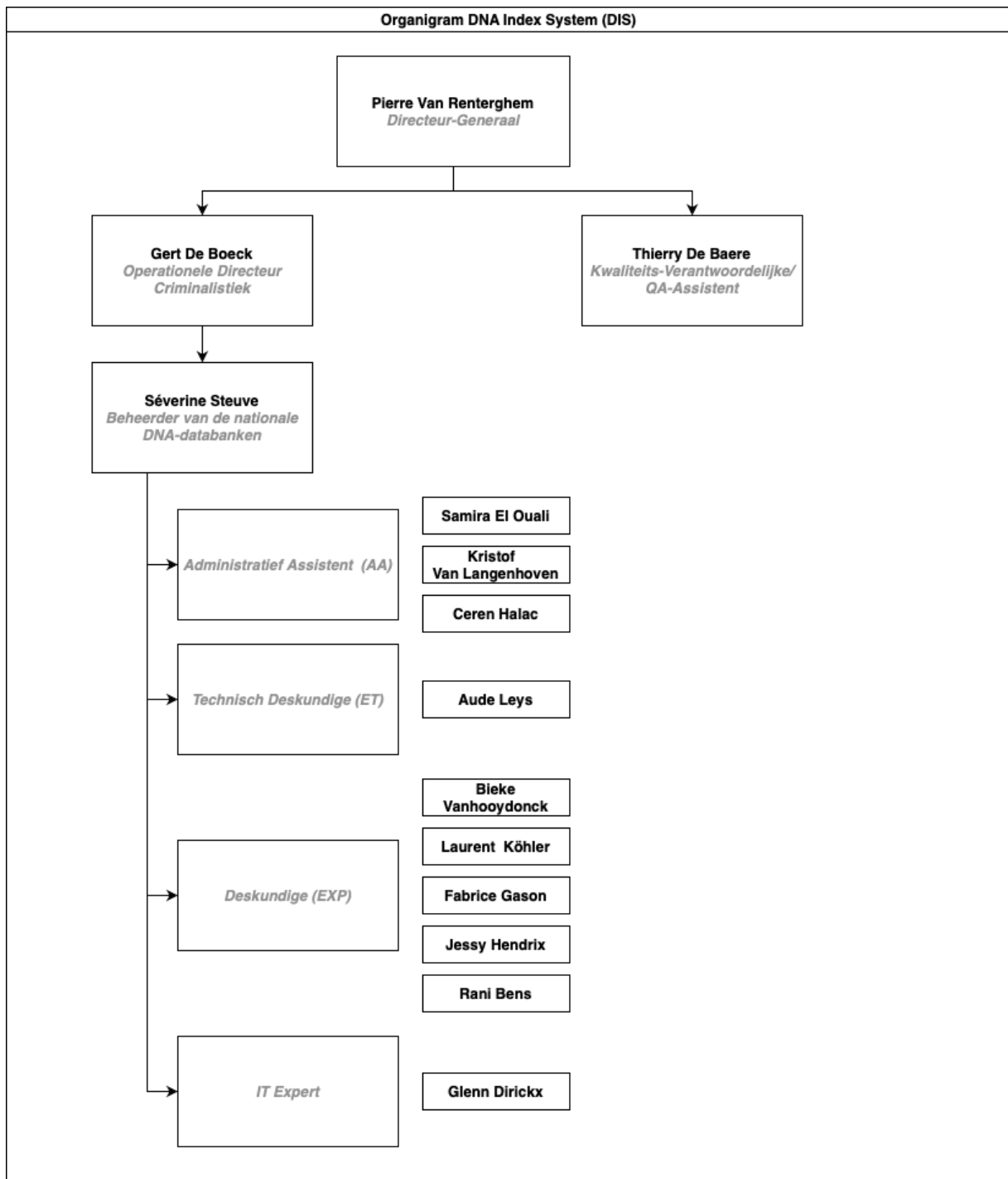
Figuur 1: Organisatiestructuur bij [Powerpoint presentatie] en "Organisatiestructuur" door NICC, n.d, Geraadpleegd op 2 februari 2024, van <https://nicc.fgov.be/organisatiestructuur>.

Activiteitsveld Operationele Directie Criminalistiek



Figuur 2: Activiteitsveld Operationele Directie Criminalistiek.

Organigram DNA Index System



Figuur 3: Organigram dienst DNA Index System (DIS).

Inhoudsopgave

Voorwoord	
Organigram	
<i>Organisatiestructuur</i>	
<i>Activiteitsveld Operationele Directie Criminalistiek</i>	
<i>Organigram DNA Index System</i>	
Lijst van afkortingen	
Abstract	
1 Inleiding	1
2 Literatuuronderzoek	2
2.1 <i>Kadering van de DNA-databanken</i>	2
2.1.1 <i>Wat is DNA?</i>	2
2.1.2 <i>Rol van DNA-profielen in forensisch onderzoek</i>	3
2.1.3 <i>Kadering in de Wetgeving</i>	3
2.2 <i>Werking van de DNA-databanken</i>	5
2.2.1 <i>De profielen</i>	5
2.2.2 <i>Behandeling van de matches in CODIS</i>	9
2.2.3 <i>Statistiek</i>	12
2.2.4 <i>Rapporteren van DNA-matches</i>	14
3 Probleemstelling	16
4 Materiaal en methoden	17
4.1 <i>Algemeen</i>	17
4.2 <i>Selectie van matches in CODIS</i>	17
4.3 <i>LR berekenen</i>	18
4.4 <i>Statistische analyse</i>	19
4.4.1 <i>Boxplot</i>	19
4.4.2 <i>Dichtheidsplot</i>	19
4.4.3 <i>Kruskal-Wallis test</i>	20
4.4.4 <i>Post-hoc Dunn-test</i>	20
5 Resultaten	21
5.1 <i>Selectie van matches in CODIS</i>	21
5.2 <i>LR berekenen</i>	21
5.3 <i>Statistische analyse</i>	22
5.3.1 <i>Boxplot</i>	22
5.3.2 <i>Dichtheidsplot</i>	23
5.3.3 <i>Boxplot met enkele unieke LR-waarde per cluster</i>	24
5.3.4 <i>Dichtheidsplot met enkele unieke LR-waarde per cluster</i>	25
5.3.5 <i>Kruskal-Wallis test met alle LR-waarden en enkele unieke LR-waarde per cluster</i>	26
5.3.6 <i>Post-hoc Dunn-test met alle LR-waarden</i>	26
5.3.7 <i>Post-hoc Dunn-test met enkele unieke LR-waarde per cluster</i>	27

6	Discussie	28
6.1	<i>Twee pieken bij de dichtheidsplot.....</i>	28
6.2	<i>Hoge LR-waarden</i>	28
6.3	<i>Verskil tussen groepen van High matches.....</i>	29
6.4	<i>Overzicht van de verschillende methoden om een DNA-match te evalueren</i>	29
6.5	<i>Beperkingen van het onderzoek.....</i>	31
7	Besluit.....	32
8	Referenties.....	33
9	Bijlagen.....	
	<i>Bijlage 1: Verbale schaal voor waarschijnlijkheid.....</i>	
	<i>Bijlage 2: Formule om de random match probability per locus te berekenen.....</i>	
	<i>Bijlage 3: Code van de dichtheidsplot in RStudio.....</i>	
	<i>Bijlage 4: Code van de Kruskal-Wallis test in RStudio.....</i>	
	<i>Bijlage 5: Code van de post-hoc Dunn-test in RStudio.....</i>	
	<i>Bijlage 6: Resultaten LR-waarden.....</i>	

Lijst van afkortingen

BM	Beperkte merkers
BR	Bijkomende resultaten
DIS	DNA Index System
DNA	Desoxyribonucleïnezuur
ESS	European Standard Set
LR	Likelihood Ratio
MtDNA	Mitochondriaal DNA
NICC/INCC	Nationaal Instituut voor Criminalistiek en Criminologie/ Institut National de Criminalistique et Criminologie
STR	Short Tandem Repeats

Abstract

Evaluating the match configuration of *High* 7-10 by using likelihood ratios compared to the number of loci

Author: Hannah Decloedt

(Co-)supervisors: Séverine Steuve, Rani Bens, Tine Dorné

Supervising lecturer: Annelies Crabbe

Introduction: As forensic DNA databases expand, ensuring the application of accurate inclusion, comparison and matching criteria has become more important. The aim of this study is to optimize the efficiency of the Belgian forensic DNA-databases by evaluating the effectiveness of likelihood ratios (LR) compared to the number of loci for distinguishing DNA-matches within and around the 'grey zone'.

Methods: To achieve this goal matches within and around the grey zone, matches with 7, 8, 9 or 10 matching loci, were selected. The likelihood ratio was calculated for these selected matches. Statistical analyses, including visualization through boxplots and density plots, were conducted to compare the likelihood ratio distributions. Additionally, the Kruskal-Wallis test was performed to determine if there are significant differences among the groups of *High* matches. If a significance is found, the post-hoc Dunn-test will be performed to differentiate between which groups a significant difference occurs.

Results: A total of 71 *High* 7 matches, 106 *High* 8 matches, 42 *High* 9 matches, and 100 *High* 10 matches were selected for analysis. The results of the boxplot indicate that the median $10\log(LR)$ value for the *High* 7 matches corresponds to an LR value of approximately 10^{10} , which is relatively high for this group. The boxplots and density plots visually illustrate variability among the different groups of *High* matches, with the most overlap observed between *High* 8 and 9. The Kruskal-Wallis test indicates a significant difference between the four groups of *High* matches. The post-hoc Dunn-test confirms a significant difference between all groups except between *High* 8 and 9.

Conclusion: There is a significant difference between LR values of *High* matches within and around the grey zone except between *High* 8 and *High* 9. *High* 7 matches show notably high LR values, often equal to or greater than the previously used match determining cut-off value of 10^9 . A proposal is to calculate the LR value for *High* 7 matches before evaluating the match. If the LR value is lower than 10^9 , an email could be sent to the laboratory requesting additional results or analyses. If the LR value is higher than 10^9 , the match could be reported with the note of caution.

Keywords: National Forensic DNA databases, likelihood ratios, *High* matches, loci, matching criteria.

Evaluëren van de matchconfiguratie van *High 7-10* door het gebruik van likelihood ratios in vergelijking met het aantal loci

Auteur: Hannah Decloedt

(Co)promotors: Séverine Steuve, Rani Bens, Tine Dorné

Begeleidende docent: Annelies Crabbe

Introductie: Naarmate de forensische DNA-databanken zich uitbreiden, wordt het steeds belangrijker om nauwkeurige inclusie-, vergelijkings- en matchingscriteria toe te passen. Het doel van deze studie is om de efficiëntie van de Belgische forensische DNA-databanken te optimaliseren door de effectiviteit van likelihood ratio's (LR) in vergelijking met het aantal loci te evalueren voor het onderscheiden van DNA-matches in en rond de 'grijze zone'.

Methode: Om dit doel te bereiken werden matches in en rond de grijze zone, matches met 7, 8, 9 of 10 overeenkomstige loci, geselecteerd. De likelihood ratio werd berekend voor deze geselecteerde matches. Statistische analyses, waaronder visualisatie door boxplots en dichtheidsplots, werden uitgevoerd om de verdelingen van de likelihood ratio te vergelijken. Daarnaast werd de Kruskal-Wallis test uitgevoerd om te bepalen of er significante verschillen zijn tussen de groepen *High matches*. Als er een significant verschil wordt gevonden, zal de post-hoc Dunn-test worden uitgevoerd om te onderscheiden tussen welke groepen een significant verschil optreedt.

Resultaten: In totaal werden 71 *High 7 matches*, 106 *High 8 matches*, 42 *High 9 matches* en 100 *High 10 matches* geselecteerd voor analyse. De resultaten van de boxplot geven aan dat de mediaan $10\log(LR)$ -waarde voor de *High 7 matches* overeenkomt met een LR-waarde van ongeveer 10^{10} , wat relatief hoog is voor deze groep. De boxplots en dichtheidsplots illustreren visueel de variabiliteit tussen de verschillende groepen *High matches*, met de meeste overlap tussen *High 8* en *9*. De Kruskal-Wallis test geeft een significant verschil aan tussen de vier groepen *High matches*. De post hoc Dunn-test bevestigt een significant verschil tussen alle groepen onderling behalve tussen *High 8* en *9*.

Conclusie: Er is een significant verschil tussen LR-waarden van *High matches* in en rond de grijze zone, behalve tussen *High 8* en *High 9*. *High 7 matches* hebben opvallend hoge LR-waarden, vaak gelijk aan of groter dan de eerder gebruikte drempelwaarde van 10^9 . Een voorstel is om de LR-waarde voor *High 7 matches* te berekenen voordat de match wordt geëvalueerd. Als de LR-waarde lager is dan 10^9 , kan er een e-mail naar het laboratorium worden gestuurd met het verzoek om aanvullende resultaten of analyses. Als de LR-waarde hoger is dan 10^9 , kan de match worden gerapporteerd met de opmerking 'beperkte merkers'.

Sleutelwoorden: Nationale DNA-databanken, likelihood ratios, *High matches*, loci, matching criteria.

1 Inleiding

Forensisch onderzoek heeft een periode van snelle veranderingen doorgemaakt als gevolg van de evolutie in DNA-profilering. Sinds de herziening van de DNA-wet, die oorspronkelijk werd opgesteld in 1999 en werd gewijzigd in 2011, zijn de voorwaarden voor het gebruik van DNA in forensisch onderzoek ruimer geworden. Dit leidde tot een toename van het aantal registraties van profielen, meer en routinematiger gebruik van DNA en een toename van internationale verbanden.

Door de toenemende groei van de forensische DNA-databanken, wordt het gebruik van correcte inclusie-, vergelijkings- en matchingcriteria steeds belangrijker. Dit heeft geleid tot reflectie en het waarborgen van het onderhoud van de databank in een evoluerende context. Als ISO 9001 gecertificeerd, ondergaan de Belgische DNA-databanken voortdurend een herziening van de procedures. Het is echter verplicht om te evalueren of de vergelijkingsregels nog steeds in overeenstemming zijn met de veranderende omvang van de databanken.

Het doel van het onderzoek is om de efficiëntie van de nationale forensische DNA-databanken te optimaliseren. Hierdoor kwam de volgende onderzoeksvraag tot stand: Wat is de effectiviteit van het gebruik van likelihood ratios (LR) ten opzichte van het aantal loci bij het onderscheiden van DNA-matches in en rond de 'grijze zone' tussen *High 7*, *High 8-9* en *High 10* in forensische DNA-analyses? Om tot een antwoord op deze onderzoeksvraag te komen, zal er eerst een inleidende uitleg gegeven worden over de oprichting en werking van de DNA-databanken, de vergelijking van DNA-profielen en de behandeling van mogelijke matches tussen DNA-profielen.

2 Literatuuronderzoek

2.1 Kadering van de DNA-databanken

2.1.1 Wat is DNA?

Het genetisch materiaal in het menselijk lichaam bestaat uit Desoxyribonucleïnezuur (DNA), een macromolecule die opgebouwd is uit een aaneenschakeling van de vier basen adenine, guanine, cytosine en thymine. De volgorde van deze basen bepaalt de genetische informatie in het DNA. Elke lichaamscel van een individu, ongeacht het celtype, bevat hetzelfde DNA. Deze informatie is cruciaal in forensisch onderzoek want hierdoor kan DNA uit wangslimvlies vergeleken worden met DNA dat verkregen is uit andere cellen zoals bloed, speeksel, sperma, haar, botten, ... Daarbovenop blijft DNA onveranderd vanaf de bevruchting tot aan de ontbinding van het lichaam (De Kinder, 2014; Meulenbroek & Aben, 2019).

Verschillende types DNA

Er worden twee types DNA onderscheiden: nucleair DNA en mitochondriaal DNA.

Het nucleair DNA is terug te vinden in de kern van de cel en vormt de meerderheid van het totale DNA. Het bestaat uit autosomen en geslachtschromosomen, namelijk het X- en Y-chromosoom. Een deel van het nucleair DNA wordt overgeërfd van de vader en een deel van de moeder. Hierdoor is het nucleair DNA van elk individu volledig uniek, met uitzondering van eenzijdige tweelingen (De Kinder, 2014).

Op basis van de STR-merkers van het Y-chromosoom kan ook een profiel opgesteld worden. Het Y-chromosoom is enkel aanwezig in de kern van de mannelijke cellen en wordt volledig via de vader overgeërfd. Alle nakomelingen in dezelfde vaderlijn vertonen een identiek Y-STR profiel (De Kinder, 2014; Nationaal Instituut voor Criminalistiek en Criminologie, n.d.).

Mitochondriaal DNA (mtDNA) bevindt zich in de mitochondriën, die fungeren als de energiecentrales van de cel. MtDNA is compacter dan nucleair DNA, en elke cel bevat meerdere kopieën van dit mtDNA. Dit type DNA wordt via de moeder overgeërfd. Alle nakomelingen in eenzelfde moederlijn delen hetzelfde mtDNA (De Kinder, 2014).

DNA-profielen worden opgesteld van zowel nucleair DNA, Y-STR en mtDNA. Deze DNA-profielen kunnen gebruikt worden in forensisch onderzoek.

2.1.2 Rol van DNA-profielen in forensisch onderzoek

Genetische profielen kunnen worden gebruikt om aangetroffen sporen op een plaats delict te linken aan elkaar en/of aan referentiestalen van verdachten, veroordeelden of slachtoffers. De donors van sporen aangetroffen op de plaats delict, ook sporenprofielen genoemd, zijn niet gekend. De sporenprofielen kunnen vergeleken worden met referentieprofielen, waarvan de oorsprong van het DNA-profiel wel gekend is. Als de twee profielen verschillend zijn van elkaar, kan er afgeleid worden dat de stalen niet van dezelfde persoon afkomstig zijn (Nationaal Instituut voor Criminalistiek en Criminologie, n.d.).

De forensische nationale DNA-databanken kunnen personen die betrokken zijn bij een misdrijf (of hun familieleden) identificeren en kunnen dossiers aan elkaar koppelen (De Kinder, 2014; ENSFI DNA Working Group, 2022; Nationaal Instituut voor Criminalistiek en Criminologie, 2022).

2.1.3 Kadering in de Wetgeving

Invloed van de EU-wetgeving op oprichting van de DNA-databanken

De raad van de Europese Unie nodigde zijn lidstaten in 1997 uit om DNA-databanken op te richten (ENSFI DNA Working Group, 2022). De forensische nationale DNA-databanken vallen onder de regelgeving van de wet van 22 maart 1999 betreffende de identificatieprocedure via DNA-analyse in strafzaken (Belgisch Staatsblad, 1999).

In 2002 werden de twee forensische nationale DNA-databanken, “criminalistiek” en “veroordeelden”, operationeel. De wijziging van het beheer van de databanken “criminalistiek” en “veroordeelden” door de dienst DNA Index System (DIS) van het Nationaal Instituut voor Criminalistiek en Criminologie werd ingevoerd door de wet van 7 november 2011, die het Wetboek van strafvordering en van de wet van 22 maart 1999 wijzigde. Deze wijzigingen zijn in werking getreden op 1 januari 2014 (Belgisch Staatsblad, 2011, 2013).

Op basis van de wet van 21 december 2013 werd de derde databank (“Vermiste personen”) opgericht en is op 28 juni 2015 in werking getreden (Belgisch Staatsblad, 2014, 2015).

De automatische overdracht van DNA-profielen, opgesteld door erkende DNA-analyselaboratoria, naar de forensische nationale DNA-databanken wordt door de DNA wet van 7 november 2011 en 21 december 2013 voorzien (Belgisch Staatsblad, 2011, 2014).

De wetten bepalen dat het de taak van de dienst DIS is om alle overgemaakte profielen te vergelijken en onder bepaalde voorwaarden te bewaren (Belgisch Staatsblad, 2011, 2014). De wet van 9 april 2017 heeft verdere wijzigingen aangebracht in het Wetboek van Strafvordering en de Wet van 22 maart 1999 met betrekking tot de identificatieprocedure via DNA-onderzoek in strafzaken (Belgisch Staatsblad, 2018).

Drie forensische nationale DNA-databanken

Er zijn drie forensische nationale DNA-databanken: criminalistiek, veroordeelden en vermiste personen.

De DNA-databank “criminalistiek” bevat de genetische profielen van aangetroffen sporen en verdachtenprofielen. Verdachtenprofielen worden eenmalig vergeleken met de nationale DNA-databanken en worden niet standaard opgeslagen. Enkel wanneer er een door het labo bevestigde match binnen hetzelfde dossier is, of als er een match binnen de DNA-databank is, zal het verdachtenprofiel opgeslagen worden. Sporenprofielen worden vergeleken met de DNA-databanken en worden opgenomen, zelfs als er geen positieve overeenkomst wordt gevonden (Belgisch Staatsblad, 1999).

De profielen van personen die voor bepaalde misdrijven zijn veroordeeld of veroordeeld werden voor een poging tot het plegen van bepaalde misdrijven, worden opgenomen in de DNA-databank “veroordeelden” (Belgisch Staatsblad, 1999).

In de DNA-databank “vermiste personen” worden de genetische profielen van stoffelijke resten van ongeïdentificeerde personen, sporen van vermiste personen en DNA-profielen van verwanten van de vermiste personen opgenomen (Belgisch Staatsblad, 2014; Nationaal Instituut voor Criminalistiek en Criminologie, 2022).

Internationale gegevensuitwisseling

De profielen die opgeslagen zijn in de forensische nationale DNA-databanken, worden vergeleken met de nieuw toegevoegde profielen op nationaal niveau en met de buitenlandse DNA-databanken. De DNA-profielen van verwanten van een vermiste persoon en mengprofielen, een sporenprofiel van meer dan één persoon, worden enkel op nationaal niveau vergeleken (Nationaal Instituut voor Criminalistiek en Criminologie, 2022).

In 2005 ondertekende België het verdrag van Prüm, waardoor de internationale uitwisseling en vergelijking van DNA mogelijk werd tussen zeven verschillende Europese landen (De Kinder Jan, 2014). Op 23 juni 2008 heeft de Raad van de Europese Unie het Verdrag van Prüm omgezet in EU-wetgeving, de EU Prüm-beslissing (Kaderbesluiten 2008/615/JBZ en 2008/616/JBZ). De nieuwe EU-wetgeving verplicht elke lidstaat van de EU om een forensische DNA-databank op te zetten en deze databank beschikbaar te stellen voor geautomatiseerde zoekopdrachten door andere EU-lidstaten. Op 1 januari 2024 was België operationeel met 24 andere EU-lidstaten (De Kinder, 2014; ENSFI DNA Working Group, 2022; Nationaal Instituut voor Criminalistiek en Criminologie, 2022; Steuve, n.d.-a).

Nationale DNA-cel

De forensische nationale DNA-databanken werken volledig anoniem, er wordt geen informatie verwerkt die betrekking heeft tot de identiteit van betrokkenen. Om het staalnummer van een referentieprofiel aan een persoon te kunnen linken, wordt er met een unieke DNA-code gewerkt. De nationale DNA-cel is verantwoordelijk voor het toekennen, beheren en centraliseren van de DNA-codes. Daarnaast houdt de nationale cel ook bij welke DNA-codes overeenkomen met welke staalnummers en identiteiten (De Kinder, 2014, 2015).

2.2 Werking van de DNA-databanken

2.2.1 De profielen

Opstellen van DNA-profielen

De DNA-profielen mochten voor de nieuwe DNA wet van 7 maart 2024 enkel verkregen worden door middel van polymerase chain reaction analyse van microsatellietsequenties in niet-coderende DNA-regio's (Belgisch staatsblad, 2024).

Kern-DNA bestaat ongeveer uit drie miljard bouwstenen waarvan er slechts 2% coderend is (De Kinder, 2014; Meulenbroek & Aben, 2019). Coderend DNA is betrokken bij de productie van eiwitten en bevat persoonsgebonden informatie zoals de fenotypische kenmerken en ziektepatronen. De wet bepaalde dat DNA-analyses in strafzaken alleen mochten plaatsvinden binnen het niet-coderende domein van het DNA. Het niet-coderende DNA codeert niet voor eiwitten en er werd verondersteld dat het geen directe genetische informatie bevat. Het geslacht is het enige uiterlijke kenmerk dat via DNA-analyse in België mocht worden vastgesteld (De Kinder, 2014). Recentere studies tonen aan dat delen binnen de niet-coderende regio's een functionele rol kunnen hebben zoals genregulatie en genexpressie. In de DNA wet van 7 maart 2024 wordt beschreven dat DNA-fenotypering met betrekking tot oog- en haarkleur, bio-geografische afkomst en leeftijd uitgevoerd mag worden. Het mag enkel aangewend worden als onderzoeksinstrument en laatste redmiddel in dossiers met betrekking tot zware feiten of ter identificatie van onbekende lichamen (Belgisch staatsblad, 2024; Vanhooydonck & Van Renterghem, 2024).

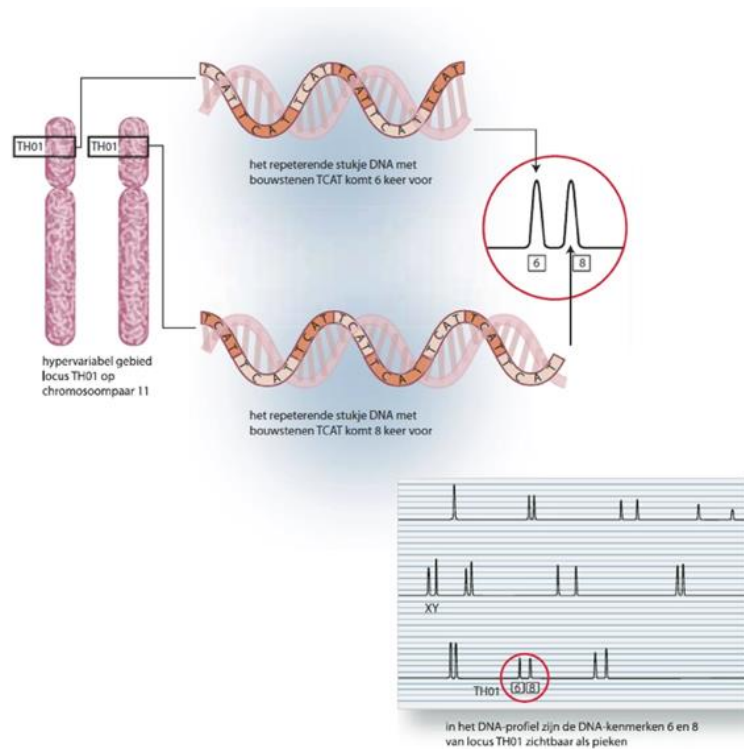
Short tandem repeats

Het niet-coderende deel van het DNA bevat 'hypervariabele' loci, waarbij de genetische variatie tussen individuen veel sterker is. Op figuur 4 is een voorbeeld te zien van het aantal repeats op de hypervariabele locus TH01. Op deze loci wordt eenzelfde stukje DNA van twee tot zes baseparen steeds herhaald, gekend als Short Tandem Repeats (STR) of microsatellieten (de Poot, 2021; Putman & Carbone, 2014).

Het aantal herhalingen van de STR's kan verschillen van persoon tot persoon. STR-merkers zijn verdeeld over alle chromosomen, met uitzondering van de geslachtschromosomen, en ze zijn genetisch onafhankelijk van elkaar. Door de totale lengte van de Short Tandem Repeats op verschillende locaties in het genoom te bepalen, ontstaat een DNA-profiel. Door op meerdere locaties op de chromosomen de STR's te bepalen, zal het profiel steeds meer discriminerend worden (de Poot, 2021; Nationaal Instituut voor Criminalistiek en Criminologie, n.d.).

Een DNA-profiel bestaat uit loci die aanwezig zijn op de autosomen, de chromosomen die geen geslachtschromosomen zijn. Voor elke locus worden de geërfde allelen van zowel de moeder als de vader weergegeven. Ook de geslachtsmerker amelogenine, een gen dat voorkomt op het X en Y chromosoom, wordt gebruikt om het DNA-profiel op te stellen (de Poot, 2021; Nationaal Instituut voor Criminalistiek en Criminologie, n.d.).

Het genetisch profiel wordt in een tabel weergegeven met de waargenomen herhalingen voor elke geanalyseerde locus. Om de overeenkomst tussen twee STR-profielen aan te tonen, wordt er een statistische interpretatie gegeven van de verkregen resultaten (Nationaal Instituut voor Criminalistiek en Criminologie, n.d.).



Figuur 4: Landelijke DNA-databank (2014). [Hypervariabel gebied locus TH01 op chromosoom 11]. Geraadpleegd op 3 mei 2024, van <https://dnadatabank.wordpress.com/wp-content/uploads/2014/08/dna-hypervariabele-gebieden1.jpg>.

Kern-DNA als krachtig identificatiemiddel

Het genetisch profiel dat opgesteld is uit kern-DNA heeft het grootste onderscheidend vermogen en daarom wordt de voorkeur aan dit type analyse gegeven. De resultaten die bekomen worden via analyse van het mitochondriaal DNA en het Y-chromosoom bieden minder mogelijkheid tot bevestiging (De Kinder, 2014).

Analyse van mitochondriaal DNA kan uitgevoerd worden bij haren omdat het kern-DNA daar moeilijker kan worden opgespoord. Mitochondriaal DNA is beter bestand tegen degradatie, waardoor het ook bij zeer oude stalen en sterk gedegreerde beenderen en tanden wordt aangewezen (De Kinder, 2014; Nationaal Instituut voor Criminalistiek en Criminologie, n.d.).

Analyse van het Y-chromosoom wordt voornamelijk uitgevoerd bij zedenfeiten. In dergelijke zaken worden in biologische sporen vaak DNA gevonden van zowel het slachtoffer (vaak vrouwelijk) als de dader (vaak mannelijk). Het mannelijke DNA kan door het vrouwelijke DNA 'overschaduwd' worden maar door het Y-chromosoom te onderzoeken, kan het mannelijke profiel in kaart gebracht worden (De Kinder, 2014; Nationaal Instituut voor Criminalistiek en Criminologie, n.d.).

Criteria voor opname in de Belgische forensische DNA-databanken

Om in de Belgische forensische DNA-databanken te worden opgenomen, moeten de DNA-profielen voldoen aan specifieke kwaliteitscriteria dat door de beheerder van de DNA-databanken zijn vastgelegd.

Alleen de profielen die vanuit kern-DNA zijn opgesteld, worden in de databanken opgeslagen. De analyse van de European Standard Set (ESS)¹ of loci en amelogenine vormt het referentiekader van de forensische nationale DNA-databanken. Tot december 2009 bevatte de ESS slechts zeven loci², maar werd daarna uitgebreid met vijf loci³. Over het algemeen worden enkel zuivere profielen, afkomstig van één enkele persoon, en mengprofielen, afkomstig van maximaal twee personen, opgeslagen in de DNA-databanken (De Kinder, 2014; ENSFI DNA Working Group, 2022).

Als referentieprofielen van verdachten en veroordeelden zuiver zijn en resultaten bevatten voor alle geanalyseerde merkers, kunnen ze worden opgeslagen in de forensische DNA-databank. Verdachtenprofielen worden niet opgeslagen als er geen positieve overeenkomst is gevonden. Enkel als zuivere sporenprofielen voor minstens vijf van de zeven originele ESS-loci of minstens tien merkers in totaal resultaten bevatten, zullen ze opgeslagen worden. Mengprofielen worden opgeslagen als ze van maximaal twee personen afkomstig zijn en voor de zeven originele ESS-loci of voor minstens tien merkers in totaal resultaten bevatten. Op verzoek van een DNA-analysesdeskundige kan uitzonderlijk ook een minder volledige profiel worden toegestaan bijvoorbeeld wanneer er een zeldzaam allel aanwezig is (De Kinder, 2015; Steuve, 2023).

¹ De ESS bestaat uit de genetische merkers D3S1358, vWA, FGA, TH01, D8S1179, D21S11, D18S51, D1S1656, D2S441, D10S1248, D12S391 en D22S1045.

² D3S1358, vWA, FGA, TH01, D8S1179, D21S11 en D18S51.

³ D1S1656, D2S441, D10S1248, D12S391 en D22S1045.

Automatische vergelijking van DNA-profielen

De DNA-profielen worden beheerd met behulp van de software 'CODIS', die ter beschikking is gesteld door de Amerikaanse FBI (De Kinder, 2015).

De recent opgeslagen DNA-profielen worden automatisch vergeleken met de reeds eerder opgeslagen DNA-profielen in de forensische databank. Er zijn verschillende matchconfiguraties mogelijk in CODIS: *High match*, *Moderate match* en *mismatch*. Afhankelijk van de configuratie, zal CODIS mogelijke overeenkomsten voorstellen die voldoen aan de ingestelde parameters (Bens, 2023a; De Kinder, 2015; Steuve, n.d.-a).

Een zuiver profiel dat volledig overeenstemt met een ander profiel op basis van ten minste vijf gemeenschappelijke genetische merkers zonder amelogenine wordt beschouwd als een mogelijke match volgens de *High 5* configuratie (De Kinder, 2015).

Wanneer voor elke gemeenschappelijke merker alle allelen van één profiel worden teruggevonden in het andere profiel wordt dit beschouwd als een mogelijke match volgens de *Moderate 8* configuratieparameter, waarbij minimaal acht gemeenschappelijke genetische merkers zonder amelogenine worden vereist. Over het algemeen is een match tussen een zuiver profiel en een mengprofiel een moderate match (Bens, 2023a; De Kinder, 2015).

Een mogelijke match volgens de configuratieparameter "*High 6 + 1 mismatch*" vereist minimaal zes gemeenschappelijke genetische merkers zonder amelogenine en staat één mismatch toe. Er wordt gesproken van een mismatch wanneer allelen van een gemeenschappelijke genetische merker van twee DNA-profielen niet overeenkomen. Het toestaan van een mismatch kan helpen bij het opsporen van mogelijke transcriptiefouten in de profielen (Bens, 2023a; De Kinder, 2015).

De opgeslagen profielen worden automatisch vergeleken via de CODIS-software, die mogelijke overeenkomsten voorstelt. Hierna zal een deskundige deze voorstellen valideren (De Kinder, 2015).

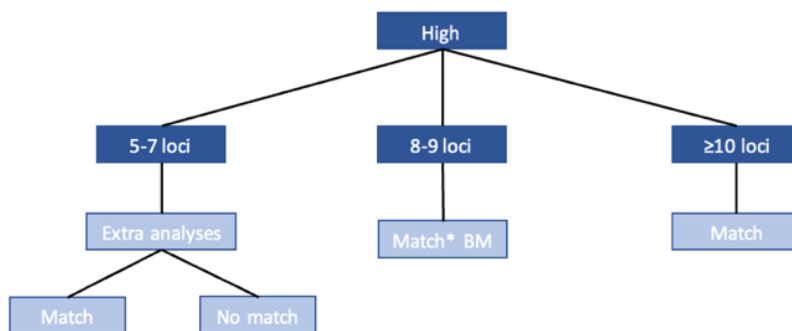
2.2.2 Behandeling van de matches in CODIS

De matches die CODIS voorstelt op basis van bovengenoemde configuraties, zullen ook nog manueel beoordeeld moeten worden. Afhankelijk van hun configuratie worden de matches op een bepaalde

manier behandeld. Aangezien het onderzoek zich richt op *High* matches, zal er dieper in gegaan worden op de behandeling van *High* matches. Er wordt een onderscheid gemaakt tussen de behandeling van matches op nationaal en internationaal niveau, maar tijdens het onderzoek zal de focus liggen op de nationale matches. Wat betreft moderate matches, zijn er ook specifieke regels, maar hier zal niet verder op ingegaan worden.

Behandeling van een “High” match op nationaal niveau

Op figuur 5 wordt de behandeling van de *High* matches schematisch weergegeven. Deze workflow helpt de experts van de DNA-databank bij het nemen van een correcte beslissing over de mogelijke match.



Figuur 5: Schema behandeling van High matches op nationaal niveau; BM: Beperkte merkers.

Aantal gemeenschappelijke genetische merkers is 5, 6 of 7 (zonder amelogenine)

Kandidaat matches met slechts vijf, zes of zeven gemeenschappelijke merkers (zonder amelogenine) mogen niet zomaar gerapporteerd worden. Er dient contact opgenomen te worden met het laboratorium voor verdere informatie. Het laboratorium kan mogelijk bijkomende resultaten (BR) verstrekken of bijkomende analyses uitvoeren die kunnen helpen om de match te bevestigen of weerleggen (Bens, 2023a).

Als het labo niet over bijkomende resultaten beschikt, zal de match als “info”⁴ gerapporteerd worden indien het gaat om een match met een niet-geïdentificeerd sporenprofiel. De match zal niet gerapporteerd worden als het een match is met enkel persoonsprofielen of een match met een geïdentificeerd sporenprofiel (Bens, 2023a).

⁴ Zie ‘2.2.4 Rapporteren van DNA-matches’ voor verdere uitleg over het infoformulier.

Als er bijkomende resultaten zijn en dus ook meer gemeenschappelijke genetische merkers, kan dit leiden tot een "match", "no match", "info" of "data problem" (Bens, 2023a).

Als er geen bijkomende resultaten beschikbaar zijn of de bijkomende resultaten niet genoeg zijn om een match te bevestigen of weerleggen, kunnen er eventueel bijkomende analyses gedaan worden. De bijkomende analyses kunnen opnieuw resulteren in een "match", "no match", "info" of "data problem" (Bens, 2023a).

Als er geen bijkomende resultaten kunnen verkregen worden of geen bijkomende analyses gedaan kunnen worden, is er sprake van een "data problem" (Bens, 2023a).

Aantal gemeenschappelijke genetische merkers is 8 of 9 (zonder amelogenine)

Kandidaat matches met acht of negen gemeenschappelijke merkers worden gerapporteerd als een "match", maar in het verslag wordt vermeld dat dit een match is met beperkte merkers (BM) en dat deze match met voorzichtigheid behandeld moet worden. Matches met acht of negen gemeenschappelijke bevinden zich dus in de grijze zone.

Aantal gemeenschappelijke genetische merkers is ≥ 10 (zonder amelogenine)

Als er tien of meer gemeenschappelijke merkers zijn, wordt dit beschouwd als een "match".

Transitiviteit

Wanneer er significante overeenkomsten worden gevonden tussen minstens twee profielen, zal een clusterformulier⁵ worden opgesteld (Steuve, n.d.-b). Stel dat er bijvoorbeeld een *High 11* match is tussen twee profielen, profiel 1 en profiel 2, in het cluster. Een bijkomend profiel, profiel 3, kan een *High 7* match hebben met profiel 1 en een *High 10* match met profiel 2. De match tussen profiel 2 en 3 voldoet aan de matchingscriteria. Omdat de match tussen profiel 1 en 2 ook aan de criteria voldoet, wordt de overeenkomst tussen profiel 1 en 3 ook als een match gerapporteerd. Dit fictief voorbeeld wordt weergegeven in figuur 6.

⁵ Zie '2.2.4 Rapporteren van DNA-matches' voor verdere uitleg over het clusterformulier.

Locus	Profiel 1	Profiel 2	Profiel 3
Amelogenin	X,Y	X,Y	X,Y
CSF1PO			
D1S1656			15, 17
D2S441			10,11
D2S1338		18,23	18,23
D3S1358	14,16	[H] 14,16	
D5S818	11,13	[H] 11,13	
D7S820	9,10	[H] 9,10	
D8S1179	13	[H] 13	[H] 13
D10S1248		13,16	13,16
D12S391			15,17
D13S317	9,12	[H] 9, 12	
D16S539		11, 13	11, 13
D18S51	12	[H] 12	[H] 12
D19S433	13	[H] 13	[H] 13
D21S11	28,29	[H] 28,29	[H] 28,29
D22S1045			
FGA	20,22	[H] 20,22	[H] 20,22
TH01	8,9	[H] 8,9	[H] 8,9
TPOX	8,9		
vWA	18	[H] 18	[H] 18

Figuur 6: Transitiviteit; [H]: High match; De overeenstemmende merkers met profiel 1 werden in het groen aangeduid.

2.2.3 Statistiek

Aantal matching loci

Het aantal matching loci hebben een invloed op de random match probability (ENSFI DNA Working Group, 2022). De random match probability is de kans dat een DNA-profiel van een willekeurig persoon uit de populatie overeenstemt met het DNA-profiel dat gevonden is op de plaats delict, terwijl die niet de donor van het spoor is. Deze kans wordt berekend door de frequentie van het profiel in de relevante populatie te bepalen. Een relevante populatie kan bijvoorbeeld de populatie zijn die overeenstemt met de locatie van het misdrijf of de demografische kenmerken van de verdachten (National Research Council, 1996; Ng et al., 2018). Hoe meer overeenstemmende STR-loci, hoe groter de discriminatie waarde en hoe lager de random match probability. Om deze reden moeten DNA-profielen die in de DNA-databank zijn opgenomen een minimum aantal loci hebben en wordt aan het DNA-laboratorium gevraagd om bijkomende resultaten of bijkomende analyses te verstrekken, zodat de match kan worden bevestigd of weerlegd. Matches tussen twee genetische profielen met acht of minder overeenstemmende loci zullen gemakkelijker leiden tot valse positieven (Ćurić et al., 2012; ENSFI DNA Working Group, 2022; Gill et al., 2006; National Research Council, 1996; Ng et al., 2018; Udogadi et al., 2020).

Likelihood Ratio

De statistische significantie van een mogelijke overeenkomst tussen twee genetische profielen kan geëvalueerd worden door middel van de berekening van de likelihood ratio (Mortera, 2020). De likelihood ratio wordt erkend als de krachtigste en relevantste statistiek die gebruikt wordt om het gewicht van DNA-bewijs te berekenen (Wivell et al., 2023). De LR is de verhouding tussen de waarschijnlijkheid van de resultaten onder de hypothese van de aanklager (H1) en de waarschijnlijkheid van de resultaten onder de hypothese van de verdediging (H2).

$$LR = \frac{P(E|H1)}{P(E|H2)}$$

De hypothese van de aanklager (H1) ondersteunt dat de twee profielen afkomstig zijn van dezelfde persoon en de hypothese van de verdediging (H2) ondersteunt dat de twee profielen afkomstig zijn van twee verschillende, niet-verwante personen, willekeurig genomen binnen een bepaalde populatie (Bens, 2023b; National Research Council, 1996). De frequentie van de allelen van de STR merkers die gebruikt zijn om de genetische profielen op te stellen, zijn in de LR-waarde verwerkt. Hierdoor worden zeldzame allelen in acht genomen. Zeldzame allelen zijn allelen die minder dan 1% voorkomen in de referentiepopulatie (Gibson, 2012; Mortera, 2020).

Een LR groter dan één betekent dat het DNA-bewijs de hypothese van de aanklager ondersteunt. Als de LR kleiner is dan één betekent dat het bewijs van de verdediging ondersteunt (Wivell et al., 2023). De LR-waarde is dus een maat voor de sterkte van het bewijs met betrekking tot de hypothese dat de twee profielen afkomstig zijn van dezelfde bron. De LR-waarde is het omgekeerde van de random probability match (National Research Council, 1996). Hoe hoger de LR-waarde, hoe lager de random match probability (Ng Jillian et al., 2018).

De Popstats module van de CODIS-software wordt gebruikt om statistische berekeningen, waaronder de random match probability en likelihood ratio, uit te voeren (Bens, 2023b; Butler, 2012).

In samenwerking met de Belgische DNA-labo's is een verbale schaal opgesteld voor de waarschijnlijkheid van matches. Deze verbale schaal wordt weergegeven in bijlage 1 (Steuve, n.d.-b). Volgens deze schaal worden overeenkomsten tussen zuivere profielen als significant beschouwd als ze minstens een miljard keer waarschijnlijker is dat de overeenkomstige DNA-profielen bekomen werden

van eenzelfde donor eerder dan dat ze afkomstig zijn van verschillende niet-verwante personen willekeurig genomen uit de Europese populatie. Een LR-waarde van 10^9 is dus de drempelwaarde die een indicatie geeft voor de matching criteria. Een LR-waarde groter dan de drempelwaarde zal beschouwd worden als een echte match, onafhankelijk van het aantal loci. Een voorgestelde overeenkomst die als onvoldoende significant wordt beschouwd, zal als worden behandeld als een negatief resultaat (De Kinder, 2014; Nationaal Instituut voor Criminalistiek en Criminologie, n.d.; Steuve, n.d.-c).

Aantal matching loci versus likelihood ratio

Om twee genetische profielen te vergelijken, is het aantal STR-merkers dus niet het enige relevante criterium. Het type STR-merkers dat wordt gebruikt om twee profielen te vergelijken, is ook een belangrijk criterium dat in acht genomen dient te worden, aangezien sommige STR's meer polymorf zijn dan andere. Dit betekent dat er veel verschillende varianten of allelen van STR-loci bestaan, waarbij sommige varianten frequenter voorkomen dan andere. De regels voor het vergelijken van twee genetische profielen zouden moeten evolueren naar een likelihood ratio berekening die het STR-polymorfe aspect omvat. Bij de berekening van de likelihood ratio wordt er rekening gehouden met het voorkomen van zeldzame allelen binnen een genetisch profiel. Bij de criteria die gebaseerd zijn op het aantal merkers wordt er geen rekening gehouden met zeldzame allelen (ENSFI DNA Working Group, 2022; Hicks et al., 2010; Mortera, 2020; Udogadi et al., 2020).

2.2.4 Rapporteren van DNA-matches

Na de validatie van de voorstellen van CODIS, kunnen de overeenkomsten gerapporteerd worden. Enkel de overeenkomsten tussen verschillende gerechtelijke dossiers zullen de forensische nationale DNA-databanken rapporteren. Overeenkomsten in eenzelfde gerechtelijk dossier worden door de DNA-laboratoria, die de betrokken profielen hebben opgesteld, behandeld (De Kinder, 2015).

Mogelijke overeenkomsten die een ongeïdentificeerd sporenprofiel bevatten en niet bevestigd of weerlegd kunnen worden door het DNA-laboratorium, worden gerapporteerd als "info" (Bens, 2023a; Steuve, n.d.-a). Het infoformulier wordt opgesteld om de mogelijke overeenkomst mee te delen aan de magistraat, maar de kans dat deze mogelijke overeenkomst daadwerkelijk een match is, is te laag. Deze mogelijke overeenkomst wordt enkel ter informatie meegedeeld, omdat deze nuttig kan zijn voor het verdere onderzoek. Het is belangrijk om deze informatie te plaatsen binnen de algemene context van het dossier en van de resultaten verkregen door het DNA-laboratorium (Steuve, n.d.-b).

Wanneer er significante overeenkomsten worden gevonden tussen minstens twee profielen, zal een clusterformulier worden opgesteld. Het clusterformulier geeft alle significante overeenkomsten weer in één formulier. Dit kunnen zowel nationale of internationale overeenkomsten zijn. Verschillende types van DNA-profielen kunnen voorkomen in een clusterformulier zoals referentieprofielen, sporenprofielen, DNA-profielen van niet-geïdentificeerde stoffelijke resten of sporen van een vermiste persoon (Steuve, n.d.-b).

De resultaten worden onder de vorm van gestandaardiseerde kennisgevingen gerapporteerd, namelijk een eenvoudige of omstandige kennisgeving. Een negatief resultaat, geen (significante) overeenkomst, wordt onder de vorm van een eenvoudige kennisgeving gerapporteerd aan de desbetreffende lokale DNA-cel. Indien er een positief resultaat is, overeenkomst met een of meerdere DNA-profielen uit andere dossiers, worden de resultaten meegedeeld onder de vorm van een omstandige kennisgeving. Deze wordt aan desbetreffende DNA-cel of via de nationale DNA-cel aan de lokale DNA-cel overgemaakt als het gaat over een overeenstemming met een DNA-profiel van een Belgische veroordeelde of verdachte. Bij een positief resultaat wordt het clusterformulier samen met een omstandige kennisgeving verstuurd (Steuve, n.d.-a).

3 Probleemstelling

Zoals al eerder aangehaald werd, bevinden *High 8* en *High 9* matches zich in een grijze zone. Hoewel deze matches worden gerapporteerd, wordt de opmerking 'beperkte merkers' toegevoegd dat de lezer waarschuwt over de kwaliteit van deze match. In dit onderzoek zullen we nagaan of de likelihood ratio een betrouwbaarder criterium is om na te gaan of deze matches wel of niet gerapporteerd mogen worden. Bovendien zullen we ook de grenzen rond deze grijze zone, *High 7* en *High 10*, nader bekijken om juiste conclusies te kunnen trekken.

4 Materiaal en methoden

4.1 Algemeen

In het kader van het onderzoek zullen een aantal matches in en rond de grijze zone geselecteerd worden. Voor deze geselecteerde matches zal de likelihood ratio berekend worden. Om de berekende LR-waarden van de *High 7, 8, 9* en *10* matches te vergelijken, zullen er verschillende statistische analyses worden uitgevoerd. Om de resultaten visueel weer te geven zullen een boxplot en een dichtheidsplot gemaakt worden. Ook zal de Kruskal-Wallis test worden uitgevoerd, een statistische test om te onderzoeken of de groepen significant verschillend zijn. Als de Kruskal-Wallis test een significant resultaat oplevert, zal de post-hoc Dunn-test worden uitgevoerd om te identificeren welke groepen er statistisch significant van elkaar verschillen.

4.2 Selectie van matches in CODIS

Allereerst werd een dataset van matches in en rond de grijze zone geselecteerd in de module Match Manager van CODIS. Hiervoor werden de volgende filters gebruikt:

- Match Stringency = "*High*"
- Matching STR loci = "*7, 8, 9* en *10*"

Bovenstaande filters werden toegepast om de *High 7, 8, 9* en *10* matches te isoleren.

- Match Date \geq 1 januari 2021
- Match Date \leq 31 december 2023

Aangezien het onderzoek gericht is op het evalueren van de huidige matching criteria zijn alleen matches gebruikt die volgens de meest recente matching criteria zijn behandeld. Let wel, een profiel in de match kan verouderd zijn, maar dit heeft geen invloed op de matching criteria zelf. Daarnaast mag de dataset niet te groot zijn om deze nog manueel te kunnen verwerken. Om deze redenen werden enkel matches van de afgelopen drie jaar bekeken.

- Search Program = "*Autosearcher*"

Enkel nationale matches werden geselecteerd aangezien we bij internationale matches niet weten aan welke kwaliteitsvoorwaarden deze voldoen. Elk land heeft zijn eigen criteria.

- STR differences = "0"

Mismatches werden niet toegestaan tijdens de selectie van de profielen, aangezien bij een mismatch altijd contact opgenomen dient te worden met het labo. Het labo moet de match bevestigen of weerleggen. Hierdoor vallen matches met een mismatch niet binnen de grijze zone die zal onderzocht worden.

4.3 LR berekenen

Om de statistische significantie van een mogelijke overeenkomst tussen DNA-profielen te evalueren aan de hand van de likelihood ratio, wordt er gebruik gemaakt van de Popstats module in CODIS. In bijlage 2 wordt de formule om de random match probability per locus te berekenen weergegeven. De RMP's van elke locus in de match worden vermenigvuldigd om te RMP te verkrijgen van het totale profiel. De LR-waarde is $1/RMP$.

De LR-waarden werden berekend met behulp van de gevalideerde methode die is opgesteld door de dienst DIS. Alleen de overeenstemmende merkers werden in de LR-berekening meegenomen. De frequentie (f) en het omgekeerde van de frequentie (1/f) werden automatisch door Popstats berekend. De frequentie geeft aan hoe vaak een gemeenschappelijk profiel binnen een bepaalde bevolkingsgroep voorkomt. Het omgekeerde van de frequentie is de LR-waarde. Bij een lage frequentie en dus een hoge LR-waarde is de kans groter dat een overeenkomst tussen profielen significant is en niet te wijten aan toeval.

Gezien de aanzienlijke variabiliteit in LR-waarden, werd de logaritme van de LR-waarden berekend met behulp van de log 10-functie in Excel.

$$\text{Log } 10(\text{LR}) = y \text{ dus } 10^y = \text{LR}$$

Het gebruik van de logaritmes kan helpen om de data beter te visualiseren en de impact van extreme waarden te verminderen, waardoor een meer betrouwbare vergelijking tussen verschillende categorieën mogelijk is.

4.4 Statistische analyse

Alle statistische analyses, behalve de boxplot, werden uitgevoerd in RStudio. RStudio is een geïntegreerde ontwikkelomgeving voor R, een programmeertaal voor statistisch computergebruik en grafische afbeeldingen.

De Kruskal-Wallis test en de post hoc Dunn-test werden uitgevoerd met een significantieniveau α van 0,05. Bij een significantieniveau van 0,05 is het risico om te concluderen dat er een verschil is tussen de groepen wanneer er geen werkelijk verschil is, gelijk aan 5%.

4.4.1 Boxplot

Om de berekende LR-waarden van de *High 7, 8, 9 en 10 matches* te visueel vergelijken, werden boxplots gemaakt. Boxplots bieden een samenvatting van de verdeling van een dataset, inclusief informatie over centrum, spreiding en uitschieters.

De boxplots zijn opgesteld met behulp van de online tool BoxplotR die gebruik maakte van de shiny package van RStudio. De dataset werd geüpload in een CSV-bestand, waarna de online tool automatisch de boxplots en de bijbehorende tabel met de waarden van de ondergrens, de bovengrens, kwartielen en mediaan genereerde.

4.4.2 Dichtheidsplot

De resultaten werden ook aan de hand van een dichtheidsplot visueel voorgesteld. Een dichtheidsplot toont de verdeling van een numerieke variabele, zoals LR-waarden, met behulp van een kernel dichtheidsschatting. Aan de hand van dichtheidsplots kunnen belangrijke kenmerken zoals meerdere pieken, asymmetrie, overlap en scheefheid in de gegevensverdeling geïdentificeerd worden.

De dichtheidsplot werd gemaakt met behulp van RStudio. In bijlage 3 is de volledige code weergegeven. Na het oproepen van de pakketten werden de gegevens in een CSV-bestand geïmporteerd en verzameld. Vervolgens werd aan de hand van het 'ggplot' pakket de dichtheidsplot gemaakt waarbij *High 7 tot 10* van elkaar onderscheiden worden aan de hand van kleur. Om de plot weer te geven, werd gebruik gemaakt van de "print" functie.

4.4.3 Kruskal-Wallis test

De Kruskal-Wallis test werd uitgevoerd om de mogelijke verschillen tussen de vier groepen verder te onderzoeken. De Kruskal-Wallis test is een test die geschikt is wanneer er geen normale verdeling van de gegevens is. Deze test vergelijkt of er statistisch significante verschillen zijn tussen twee of meer groepen. De Kruskal-Wallis test geeft niet aan welke specifieke groepen er significant van elkaar verschillen, het geeft enkel weer dat er minstens twee groepen verschillend zijn.

De Kruskal-Wallis test werd uitgevoerd met behulp van RStudio, waarbij de volledige code weergegeven is in bijlage 4. Na het oproepen van het pakket werden de gegevens in een csv-bestand geïmporteerd en verzameld. Vervolgens kon de Kruskal-Wallis test worden uitgevoerd.

Om te bepalen of er statistisch significante verschillen waren tussen de groepen van *High matches*, werd de nulhypothese en de alternatieve hypothese gedefinieerd. De nulhypothese stelt dat er geen significante verschillen zijn tussen de groepen, de alternatieve hypothese stelt dat er wel significante verschillen zijn tussen de groepen. Daarna werd de p-waarde vergeleken met het vooraf bepaalde significantieniveau. Als de p-waarde kleiner is dan of gelijk aan het significantieniveau, wordt nulhypothese verworpen en wordt de alternatieve hypothese aanvaard. Als de p-waarde groter is dan het significantieniveau, is er niet genoeg bewijs om de nulhypothese te verwerpen.

4.4.4 Post-hoc Dunn-test

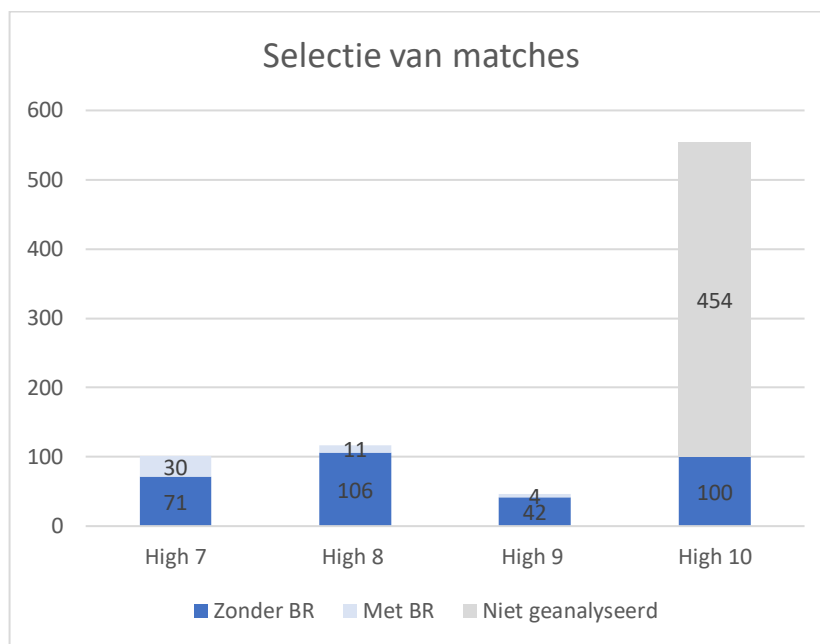
Als de Kruskal-Wallis test een significant resultaat oplevert, wordt de post-hoc Dunn-test uitgevoerd om te identificeren welke groepen er statistisch significant van elkaar verschillen. De post-hoc Dunn-test is een aanvullende analyse die na de Kruskal-Wallis test wordt uitgevoerd.

De post-hoc Dunn-test werd uitgevoerd in RStudio. In bijlage 5 is de volledige code weergegeven. Vervolgens werd de post-hoc Dunn-test uitgevoerd aan de hand van het FSA pakket. Tijdens de post-hoc Dunn-test worden meerdere vergelijkingen gemaakt tussen groepen, waardoor het risico op type I-fouten (de nulhypothese wordt onterecht verworpen) toeneemt. Om dit te corrigeren, werd een Bonferroni-correctie toegepast. Deze correctie verlaagt de kans op het onjuist verwerpen van de nulhypothese door de alfa-waarde te verdelen door het aantal vergelijkingen dat wordt uitgevoerd, waardoor de drempel voor statistische significantie strenger wordt. Na het toepassen van de Bonferroni-correctie werd het resultaat weergegeven via de print-functie.

5 Resultaten

5.1 Selectie van matches in CODIS

Op basis van de al genoemde filtercriteria werden nationale matches van zuivere profielen geselecteerd met een match datum in de jaren 2021, 2022 en 2023. Op figuur 7 worden de aantallen geselecteerde en geanalyseerde matches per groep van *High matches* grafisch weergegeven.



Figuur 7: Grafische weergave van de aantallen van de geselecteerde en geanalyseerde matches per groep van High matches; BR: Bijkomende resultaten.

Matches waarbij bijkomende resultaten werden toegevoegd, werden niet meegenomen in de selectie omdat ze meer gemeenschappelijke loci hebben dan aangegeven in CODIS.

Er waren 554 *High 10* matches die voldeden aan de filtercriteria. Aangezien 554 matches te veel is om manueel te verwerken, werden er honderd matches geselecteerd voor analyse. De overige matches uit de selectie werden niet geanalyseerd. Om eventuele vooringenomenheid te vermijden, werden via de RAND-functie van Excel honderd matches willekeurig geselecteerd uit deze groep. Hierdoor kon de mogelijke variabiliteit tussen de verschillende jaren uitgesloten worden.

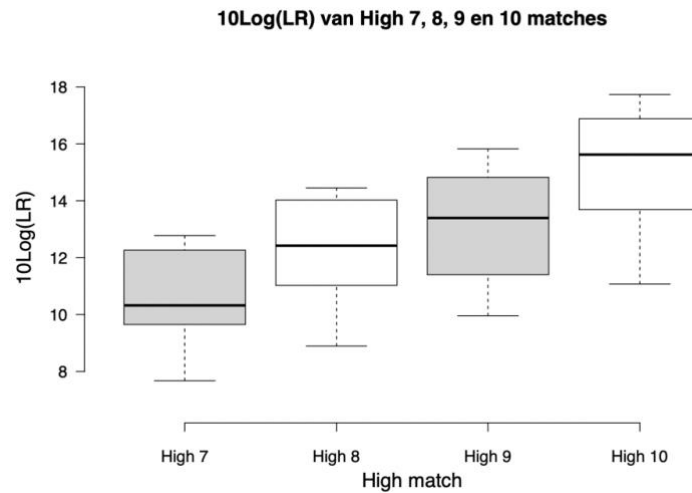
5.2 LR berekenen

In bijlage 6 worden de berekende LR-waarden en het bijhorende matchnummer weergegeven.

5.3 Statistische analyse

5.3.1 Boxplot

Op figuur 8 worden de boxplots voor de verschillende groepen van *High* matches weergegeven.



Figuur 8: Boxplot 10Log(LR) van High 7, 8 9 en 10 matches.

In tabel 1 worden de 10Log (LR)-waarden voor de ondergrens, de bovengrens, kwartielen en mediaan weergegeven.

Tabel 1: Boxplot 10Log(LR) van High 7, 8 9 en 10 matches statistieken.

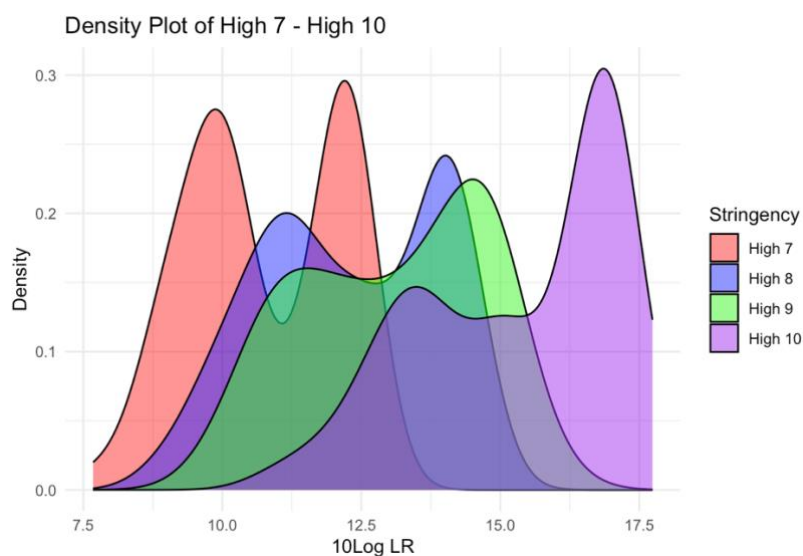
10Log (LR)	High 7	High 8	High 9	High 10
Bovengrens	12,78	14,45	15,83	17,74
3^e kwartiel	12,26	14,03	14,83	16,88
Mediaan	10,38	12,42	13,40	15,62
1^e kwartiel	9,72	11,03	11,40	13,70
Ondergrens	7,68	8,89	9,96	11,08
Aantal gegevens	71	106	42	100

De mediaan van 10Log(LR) voor de *High 7* matches (aangeduid met een dikke zwarte lijn) is 10,38, wat ongeveer overeenstemt met een LR-waarde van 10^{10} . Aangezien drempelwaarde momenteel op 10^9 ligt, is deze mediaan vrij hoog voor *High 7* matches, gegeven dat er slechts 7 overeenkomstige loci zijn die mee verwerkt kunnen worden in de LR berekening. De mediaan ligt niet in het midden van de boxplot, wat erop wijst de resultaten niet normaal verdeeld zijn. De gegevens zijn laag gecentraliseerd.

Dit ondersteunt de keuze voor de Kruskal-Wallis test, aangezien deze test gebruikt wordt bij niet-normaal verdeelde data. De medianen van de *High* matches in de grijze zone leunen dicht bij elkaar aan en liggen ongeveer centraal in de boxplot. Bij de *High* 10 matches ligt de mediaan weer hoger met een LR-waarde van ongeveer 10^{15} . Er zijn geen uitschieters in deze data.

5.3.2 Dichtheidsplot

Op figuur 9 wordt de dichtheidsplot weergegeven.



Figuur 9: Dichtheidsplot $10\text{Log}(\text{LR})$ van High 7, 8 en 10 matches.

Bij elke groep zijn twee afzonderlijke pieken te onderscheiden in de dichtheidsplot. Op figuur 9 is duidelijk te zien dat de pieken van *High* 8 en *High* 9 sterk overlappen. Bij *High* 7 is het verschil tussen de twee pieken meer uitgesproken dan bij de andere groepen. Aangezien het verschil tussen de waarden van *High* 7 opmerkelijk is, werd gezocht naar een mogelijke verklaring hiervoor.

Eerst werd gezocht naar de aanwezigheid de meest polymorfe merkers in de matches. Een gen is polymorf wanneer er binnen een populatie meer dan één allel bestaat op de locus van dat gen. SE33, D21S11 en D18S51 zijn de meest polymorfe merkers en vertonen veel variabiliteit. Door de grote variabiliteit bij deze merkers, is er meer kans op het voorkomen van zeldzame allelen. De aan- of de afwezigheid van zeldzame allelen hebben een sterke invloed op de LR-waarde.

De aanwezigheid van loci SE33, D21S11 en D18S51 in de geselecteerde matches werd nagegaan in Popstats. De meeste matches bevatten een overeenkomst op loci D18S51 en D21S11. Er zijn weinig

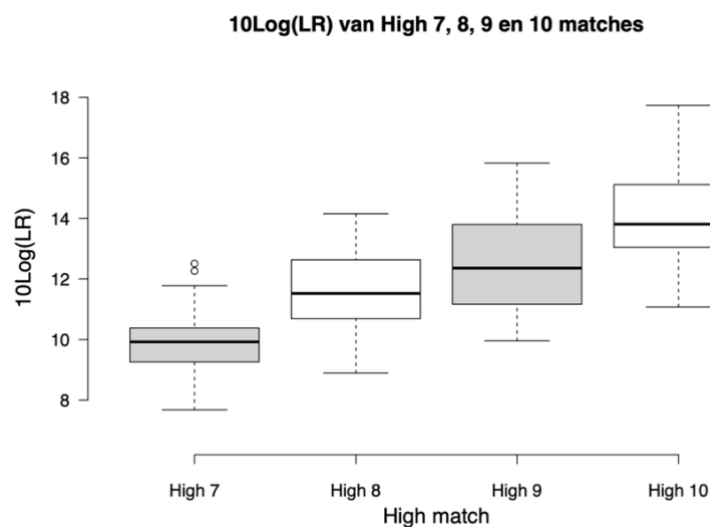
matches met locus SE33 omdat deze niet behoort tot de European Standard Set of loci. Aangezien bijna alle matches een overeenkomst hebben met dezelfde polymorfe loci, werd op basis hiervan nog geen verklaring gevonden.

Na verder onderzoek bleek dat de LR-waarden van matches die tot dezelfde cluster behoren, vaak overeenkomen. Aangezien in een cluster hetzelfde DNA-profiel voorkomt, wat vaak resulteert in dezelfde matchende loci en dus hetzelfde aantal herhalingen per locus, is dit een logisch resultaat. Om deze reden werd voor iedere cluster de LR-waarde slechts één keer gebruikt.

Uit verder onderzoek blijkt dat één bepaalde cluster (Cluster G, zie bijlage 6), vaak voorkomt. De LR-waarden van de matches in cluster G liggen vrij hoog. In Popstats werd een van de matches uit cluster G met een LR-waarde van 1 837 000 000 000 in detail geanalyseerd. In de match vertonen de loci D18S51, TH01, D21S11 en FGA weinig voorkomende allelen. Deze combinatie van zeldzame allelen resulteert in een relatief zeldzaam profiel. Het verschil tussen de LR-waarden voor High 7 matches ligt hoger door het voorkomen en de grote bijdrage van dit zeldzame profiel. Dit resulteerde in twee grote pieken op de dichtheidsplot voor High 7. Voor High 8, 9 en 10 matches was het verschil tussen de minimale en maximale LR-waarden kleiner.

5.3.3 Boxplot met enkele unieke LR-waarde per cluster

Op figuur 10 worden de boxplots voor de verschillende groepen van High matches weergegeven, waarbij de waarden van de clusters slechts één keer worden meegenomen in de dataset.



Figuur 10: Boxplot 10Log(LR) van High 7, 8 9 en 10 matches met enkele unieke LR-waarde per cluster.

In tabel 3 worden de waarden voor de ondergrens, de bovengrens, kwartielen en mediaan weergegeven.

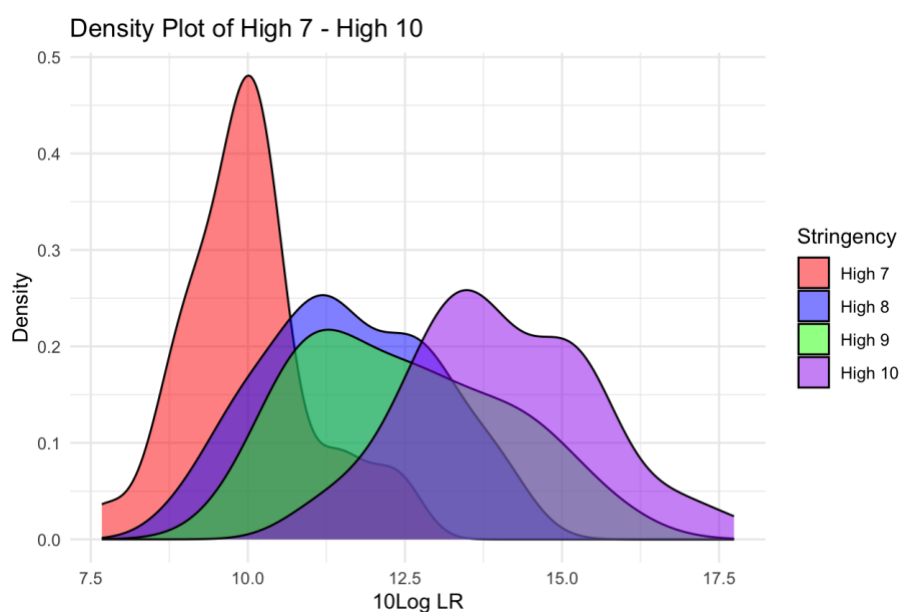
Tabel 2: Boxplot 10Log(LR) van High 7, 8 9 en 10 matches statistieken met enkele unieke LR-waarde per cluster.

	High 7	High 8	High 9	High 10
Bovengrens	11,78	14,15	15,83	17,74
3^e kwartiel	10,38	12,64	13,80	15,12
Mediaan	9,92	11,52	12,36	13,81
1^e kwartiel	9,26	10,69	11,17	13,05
Ondergrens	7,68	8,89	9,96	11,08
Nr. of data points	33	61,00	27,00	50,00

De gemiddelde 10Log(LR)-waarden zijn lager wanneer slechts één waarde per cluster wordt meegenomen in de dataset in vergelijking met het meenemen van alle clusterwaarden. Desondanks blijft de mediaan van de *High 7* matches vrij hoog, maar deze keer zijn er twee uitschieters, aangeduid met een kleine cirkel. De medianen van de *High* matches in de grijze zone leunen hier ook dicht bij elkaar aan. Het verschil tussen de groepen lijkt kleiner dan bij het voorgaande resultaat.

5.3.4 Dichtheidsplot met enkele unieke LR-waarde per cluster

Op figuur 11 wordt de dichtheidsplot weergegeven.



Figuur 11: Dichtheidsplot 10Log(LR) van High 7, 8 9 en 10 matches met enkele unieke LR-waarde per cluster.

Na het nemen van één waarde per cluster zijn er bij *High 8-9* en *High 10* geen duidelijk onderscheiden pieken te zien. Daarentegen vertoont *High 7* een grote piek, wat aangeeft dat de waarden bij *High 7* zeer dicht bij elkaar liggen.

5.3.5 Kruskal-Wallis test met alle LR-waarden en enkele unieke LR-waarde per cluster

Om de resultaten te kunnen interpreteren, werden de nulhypothese en de alternatieve hypothese gedefinieerd:

- H_0 : Er is geen significant verschil tussen de medianen van de vier groepen van *High matches*.
- H_1 : Er is wel een significant verschil tussen de medianen van de vier groepen van *High matches*.

De p-waarde, berekend door RStudio, is $2,2e-16$. Zowel bij het meenemen van alle waarden voor de clusters als bij het nemen van slechts één waarde per cluster, werd dezelfde p-waarde verkregen.

$$P(2,2e-16) < \alpha(0,05)$$

De p-waarde is kleiner dan het significantieniveau dus de nulhypothese wordt verworpen. Er is een significant verschil tussen de medianen van de vier groepen van *High matches*.

5.3.6 Post-hoc Dunn-test met alle LR-waarden

In tabel 4 worden de oorspronkelijke p-waarden en de gecorrigeerde p-waarden (na Bonferroni-correctie) weergegeven.

Tabel 3: Resultaten post hoc Dunn-test.

Vergelijking	Oorspronkelijke p-waarde	Gecorrigeerde p-waarde
<i>High 7 – High 8</i>	4,563977e-06	2,738386e-05
<i>High 8 - High 9</i>	5,717500e-02	3,430500e-01
<i>High 9 - High 10</i>	2,749917e-06	1,649950e-05
<i>High 7 - High 9</i>	6,942260e-08	4,165356e-07
<i>High 7 - High 10</i>	7,117750e-35	4,270650e-34
<i>High 8 - High 10</i>	4,251850e-18	2,551110e-17

De gecorrigeerde p-waarden voor de vergelijkingen tussen *High 7 – 8*, *High 9 – 10*, *High 7 – 9*, *High 7 – 10* en *High 8 – 10* zijn kleiner dan het significantieniveau 0,05. Tussen deze groepen wordt een significant verschil gevonden. De p-waarde voor de vergelijking tussen *High 8 – 9* is groter dan 0,05, wat duidt dat er geen significant verschil is tussen *High 8* en *High 9*.

5.3.7 Post-hoc Dunn-test met enkele unieke LR-waarde per cluster

In tabel 5 worden de oorspronkelijke p-waarden en de gecorrigeerde p-waarden (na Bonferroni-correctie) weergegeven.

Tabel 4: Resultaten Post-hoc Dunn-test met enkele unieke LR-waarde per cluster.

Vergelijking	Oorspronkelijke p-waarde	Gecorrigeerde p-waarde
<i>High 7 - High 8</i>	4,538457e-05	2,723074e-04
<i>High 8 - High 9</i>	6,067945e-02	3,640767e-01
<i>High 9 - High 10</i>	1,550567e-03	9,303400e-03
<i>High 7 - High 9</i>	4,042397e-07	2,425438e-06
<i>High 7 - High 10</i>	2,636596e-20	1,581957e-19
<i>High 8 - High 10</i>	4,512798e-10	2,707679e-09

De gecorrigeerde p-waarden voor de vergelijkingen tussen *High 7 – 8*, *High 9 – 10*, *High 7 – 9*, *High 7 – 10* en *High 8 – 10* zijn kleiner dan het significantieniveau 0,05. Tussen deze groepen wordt een significant verschil gevonden. De p-waarde voor de vergelijking tussen *High 8 – 9* is groter dan 0,05, wat duidt dat er geen significant verschil is tussen *High 8* en *High 9*.

6 Discussie

6.1 Twee pieken bij de dichtheidsplot

Bij de dichtheidsplot van alle LR-waarden zijn er twee opvallende pieken er bij *High 7*, maar ook twee kleinere pieken bij *High 8*, 9 en 10. Twee pieken per groep wijzen op de aanwezigheid van twee subgroepen in de data, met één subgroep die lage en de andere hoge LR-waarden heeft. Om dit te kunnen verklaren werd gekeken naar de polymorfe merkers SE33, D18 en D21 (zie '5.3.2 Dichtheidsplot'), maar dit leverde aanvankelijk geen verklaring op. Daarna werd gekeken of de twee pieken aan de hand van de clusters verklaard kon worden. De veelvoorkomende cluster G bevat een zeldzaam DNA-profiel, wat resulteerde in hogere LR-waarden en dus twee pieken. Na het gebruik van één unieke LR-waarde per cluster in de dataset is bij *High 7* een duidelijke piek waarneembaar, terwijl voor de andere *High matches* geen duidelijk te onderscheiden piek zichtbaar is. Bij *High 7* liggen de resultaten binnen de groep dichter bij elkaar dan de resultaten binnen de andere groepen. Het opnemen van alle LR-waarden van de clusters in de dataset heeft weldegelijk een invloed op de resultaten, vooral wanneer er een cluster een zeldzaam profiel bevat dat vaak voorkomt.

6.2 Hoge LR-waarden

De berekende LR-waarden bleken na het nemen van één LR-waarde per cluster niet volledig overeen te stemmen met de verwachtingen. De LR-waarden voor de geselecteerde *High 7 matches* liggen in de meeste gevallen hoger dan de drempelwaarde van 10^9 , zelf na het nemen van één waarde per cluster. De waarden die boven deze drempelwaarde liggen, duiden volgens de verbale schaal die wordt gehanteerd door de Belgische DNA-labo's, een betrouwbare match aan (zie bijlage 1). Alle LR-waarden van geanalyseerde *High 7 matches* waren groter dan 10^6 , wat wijst op een uitermate sterke waarschijnlijkheid dat de DNA-profielen afkomstig zijn van dezelfde persoon. Momenteel worden matches met zeven gemeenschappelijke loci echter niet als betrouwbaar beschouwd volgens de Belgische matching criteria, tenzij ze in een cluster voorkomen en door middel van transitiviteit een match vormen.

Aangezien de LR-waarden voor *High 7* hoger liggen dan de drempelwaarde kan de vraag gesteld worden of het nog nodig is om voor deze *High 7 matches* bijkomende resultaten of analyses te verstrekken (Bens, 2023a; Steuve, n.d.-b). Een mogelijke aanbeveling voor DIS kan zijn om bij *High 7 matches*, wanneer de LR-waarde onder een bepaalde drempelwaarde ligt, te vragen om bijkomende resultaten of bijkomende analyses.

6.3 Verschil tussen groepen van *High matches*

De Kruskal-Wallis test toont tweemaal aan dat er een significant verschil is tussen de vier groepen van *High matches*. De post-hoc Dunn-test bevestigt tweemaal dat er een significant verschil tussen alle groepen is behalve tussen *High 8* en *9*. Dit komt overeen met de huidige matching criteria, waarbij *High 7* bijkomende analyses of resultaten vereisen en wordt onderscheiden van *High 8* en *9 matches*, die worden gerapporteerd met de opmerking ‘beperkte merkers’. Volgens de matching criteria werden *High 8* en *9 matches* al beschouwd als een grijze zone met overlappende LR-waarden. *High 10* wordt als een goede match beschouwd, en dit werd ook met de resultaten aangetoond aangezien de laagste LR-waarde hoger is dan 10^9 (Bens, 2023a).

Uit onderzoek van Laurent et al. (2023)⁶ is gebleken dat er overlap is tussen de LR-distributies voor matches met drie, vier, vijf en zes gemeenschappelijke loci. Echte matches op basis van een gelimiteerd aantal merkers hebben vaak een even hoge LR-waarde hebben als échte matches op basis van meer merkers. Dit toont aan dat het beoordelen van de match op basis van het aantal overeenstemmende loci arbitrair is. Het bepalen van een match enkel op basis van het aantal overeenstemmende loci is dus geen goed criterium. Daarentegen werd in het onderzoek van Laurent et al. (2023) ook aangetoond dat profielen met zes of minder gemeenschappelijke merkers meer kans hebben op vals-positieve matches. Tijdens het onderzoek van deze bachelorproef werd enkel met échte matches gewerkt, die minstens zeven gemeenschappelijke loci bevatten. Op basis van de resultaten en conclusies van dit onderzoek en het onderzoek van Laurent et al. (2023) kan gesuggereerd worden om een combinatie te gebruiken van LR-waarden en het aantal overeenstemmende loci.

6.4 Overzicht van de verschillende methoden om een DNA-match te evalueren

Om een DNA-match te evalueren, zijn er verschillende methoden: aantal gemeenschappelijke loci, LR-waarden op basis van de aan- of afwezigheid van pieken, LR-waarden op basis van exacte piekhoogten, dynamische LR-waarden en een combinatie van LR-waarden en aantal overeenstemmende loci. Om de meest geschikte aanbevelingen te kunnen maken voor de matching criteria van de dienst DIS, worden zowel de conclusies van dit onderzoek als verschillende reeds bestaande methoden in overweging genomen.

⁶ Het artikel van dit onderzoek is nog niet gepubliceerd omdat het in afwachting is van goedkeuring door peer review.

Een DNA-match kan geëvalueerd worden op basis van het aantal gemeenschappelijke loci. Dit is de methode die wordt gebruikt door de Belgische nationale databanken. Volgens hun matching criteria moet een match minimaal acht gemeenschappelijke merkers hebben om als een match beschouwd te worden. Bovendien kan een match met minder gemeenschappelijke loci toch als betrouwbare match beschouwd worden door middel van transitiviteit. Zoals al werd aangehaald is deze methode om DNA-matches te evalueren arbitrair.

Daarnaast heeft eerder onderzoek aanbevolen om methoden met een continue verdeling en LR te gebruiken bij het beoordelen van matches (National Research Council, 1996). Uit recenter onderzoek van Lund & Iyer (2017) blijkt dat de interpretatie van bewijs in de vorm van een LR persoonlijk en subjectief kan zijn.

Een ander interessant onderzoek van Puch-Solis et al. (2013) stelt een model voor dat de waarschijnlijkheidsdistributies voor het niveau van piekhoogten van een DNA-profiel gebruikt. Door het gebruik van exacte piekhoogten in plaats van methoden die de LR-waarde berekenen alleen op basis van de aan- of afwezigheid van pieken, wordt de subjectiviteit verminderd. Dit onderzoek suggereert dat beslissingen niet alleen gebaseerd moeten zijn op LR-waarden wanneer geen rekening wordt gehouden met de exacte piekhoogte bij het berekenen ervan. Het nadeel van deze aanpak is dat het exacte verschil tussen de methodes voor het berekenen van LR-waarden nog niet bekend is.

In de studie van Laurent et al. (2023) werd voorgesteld om een dynamische LR-waarde te gebruiken om te bepalen of een mogelijke match al dan niet gerapporteerd dient te worden, afhankelijk van het aantal matchende loci en de grootte van de DNA-databank. Door de $10\log(LR)$ drempel aan te passen naarmate de database groeit, en dus het niveau matchende loci te verhogen, blijft het maximale aantal vals-positieve matches beperkt. Een soortgelijke aanpak wordt al sinds 2016 toegepast in NDIS, dat het uploaden en doorzoeken toestaat van een forensisch DNA-profiel met minimaal acht van de originele CODIS-kernloci in combinatie met een $10\log(LR)$ van 7.

Het combineren van zowel de LR-waarde als het aantal overeenstemmende loci biedt zijn voordelen omdat het een soort dubbel controlemechanisme vormt. Dit verhoogt de nauwkeurigheid van de matchbeslissingen. Het toepassen van een dynamische LR-waarde is waardevol omdat het rekening houdt met de grootte van de databank en het aantal matchende loci, maar het moet regelmatig worden herzien om relevant te blijven. Het continu berekenen van LR-waarden bij elk match-voorstel zou te veel tijd vergen.

Op basis van dit gelimiteerd onderzoek en al gepubliceerde methoden is een mogelijke aanpak voor DIS om de LR-waarde te berekenen voor *High 7* matches. Als de LR-waarde lager is dan 10^9 , zou een mail naar het labo gestuurd kunnen worden met de vraag voor bijkomende resultaten of bijkomende analyses. Als een LR-waarde hoger is dan 10^9 , zou de match gerapporteerd kunnen worden met de opmerking 'beperkte merkers'. Om een concreet antwoord te vormen op de onderzoeksvraag moet er meer onderzoek verricht worden.

6.5 Beperkingen van het onderzoek

De DNA-laboratoria maken gebruik van het programma 'LRMixStudio' om LR-waarden te berekenen, wat resulteert in een nauwkeurigere berekening op basis van de piekhoogtes. Deze LR-waarden kunnen verschillen van de LR-waarden dat berekend werden met behulp van Popstats. Het is momenteel onbekend hoe groot dit verschil precies is en of de drempelwaarde die door DNA-databanken wordt gebruikt, overeenkomt met die van de DNA-laboratoria. Om het effect van de verschillende methoden voor het berekenen van LR-waarden te bepalen, is verder onderzoek nodig.

Om een concreet antwoord te vormen op de onderzoeksvraag, is verder onderzoek nodig met een groter aantal matches. In dit onderzoek werd slechts één waarde per cluster gebruikt, waardoor de dataset verkleind werd en mogelijks niet representatief is. Daarnaast werden in dit onderzoek alleen matches geëvalueerd. Als er verder onderzoek gevoerd wordt, zou het interessant zijn om de LR-waarde voor niet-matches te bestuderen.

7 Besluit

Dit onderzoek trachtte de onderzoeksvraag te beantwoorden: Wat is de effectiviteit van het gebruik van likelihood ratios (LR) ten opzichte van het aantal loci bij het onderscheiden van DNA-matches in en rond de 'grijze zone' tussen *High 7*, *High 8-9* en *High 10* in forensische DNA-analyses? De bevindingen, verkregen via onderzoek, stemden niet volledig overeen met de verwachtingen. De LR-waarden lagen hoger dan verwacht en kunnen mogelijks verklaard worden door het voorkomen van een veelvoorkomend zeldzaam profiel. Zoals verwacht werd een significant verschil tussen de groepen van *High matches* gevonden behalve tussen *High 8* en *High 9*.

Op basis van dit gelimiteerd onderzoek en al gepubliceerde methoden is een mogelijke aanpak voor DIS om de LR-waarde te berekenen voor *High 7 matches*. Als de LR-waarde lager is dan 10^9 , zou een mail naar het labo gestuurd kunnen worden met de vraag voor bijkomende resultaten of bijkomende analyses. Als een LR-waarde hoger is dan 10^9 , zou de match gerapporteerd kunnen worden met de opmerking 'beperkte merkers'. Om een concreet antwoord te vormen op de onderzoeksvraag moet er meer onderzoek verricht worden.

8 Referenties

- Belgisch Staatsblad. (1999). Wetten, decreten, ordonnanties en verordeningen. In 1999-05-20. <https://www.ejustice.just.fgov.be/cgi/article.pl>
- Belgisch Staatsblad. (2011). Wetten, decreten, ordonnanties en verordeningen. In 2011-11-30. <https://www.ejustice.just.fgov.be/cgi/article.pl>
- Belgisch Staatsblad. (2013). Wetten, decreten, ordonnanties en verordeningen. In 2013-08-12. <https://www.ejustice.just.fgov.be/cgi/article.pl>
- Belgisch Staatsblad. (2014). Wetten, decreten, ordonnanties en verordeningen. In 2014-01-30. https://www.ejustice.just.fgov.be/cgi/article_body.pl?language=nl&caller=summary&pub_date=14-01-30&numac=2014009023
- Belgisch Staatsblad. (2015). Wetten, decreten, ordonnanties en verordeningen. In 2015-09-16. https://www.ejustice.just.fgov.be/cgi/article_body.pl?language=nl&caller=summary&pub_date=15-09-16&numac=2015000479
- Belgisch Staatsblad. (2018). Wetten, decreten, ordonnanties en verordeningen. In 2018-02-28. https://www.ejustice.just.fgov.be/cgi/article_body.pl?language=nl&caller=summary&pub_date=18-02-28&numac=2018010955
- Belgisch staatsblad. (2024). Wetten, decreten, ordonnanties en verordeningen. In 2024-03-24. https://www.ejustice.just.fgov.be/cgi/article_body.pl?language=nl&caller=summary&pub_date=24-04-02&numac=2024003125
- Bens, R. (2023a). *Behandeling van de nationale matchen in CODIS*.
- Bens, R. (2023b). *Gebruik van Popstats*.
- Butler, J. M. (2012). Chapter 8 - DNA Databases: Uses and Issues. In *Advanced Topics in Forensic DNA Typing: Methodology* (pp. 2013–2070). <https://www.sciencedirect.com/science/article/abs/pii/B9780123745132000087>
- Ćurić, G., Gašić, V., Plužarić, V., & Smiljčić, D. (2012, October). *Genetic parameters of five new European Standard Set STR loci (D10S1248, D22S1045, D2S441, D1S1656, D12S391) in the population of eastern Croatia*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3496899/>
- De Kinder, J. (2014). *Wat U moet weten over forensisch DNA-onderzoek*. Nationaal Instituut voor Criminalistiek en Criminologie. <https://nicc.fgov.be/upload/files/OverhetNICC/nicc-dna-magazine-nl-lr.pdf>
- De Kinder, J. (2015). *Activiteitenverslag 2014 Nationale DNA-databanken*. <https://nicc.fgov.be/upload/images/ODcriminalistiek/dna-databank-jv-pub-nl-screen.pdf>
- de Poot, C. (2021). *Het gebruik van DNA in het opsporingsproces*. <https://research.hva.nl/en/publications/het-gebruik-van-dna-in-het-opsporingsproces>
- ENFSI DNA Working Group. (2022). *DNA database management review and recommendations*.
- Gibson, G. (2012). *Rare and Common Variants: Twenty arguments* (Vol. 13, Issue 2, pp. 135–145).

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4408201/#:~:text=Rare%20allele%20model%3A%20many%20rare,alleles%20in%20the%20human%20population.>
- Gill, P., Fereday, L., Morling, N., & Schneider, P. M. (2006). *The evolution of DNA databases—Recommendations for new European STR loci* (Vol. 156, Issues 2–3, pp. 242–244). <https://www.sciencedirect.com/science/article/abs/pii/S0379073805003427?via%3Dihub>
- Hicks, T., Taroni, F., Curran, J., Buckleton, J., Ribaux, O., & Castella, V. (2010). Use of DNA profiles for investigation using a simulated national DNA database: Part I. Partial SGM Plus® profiles. *Forensic Science International: Genetics*, 4(4). <https://doi.org/10.1016/j.fsigen.2009.10.002>
- Landelijke DNA-databank. (2014). *Hypervariabel gebied locus TH01 op chromosoom 11*. <https://dnadatabank.wordpress.com/wp-content/uploads/2014/08/dna-hypervariabele-gebieden1.jpg>
- Laurent, F.-X., Nikolaeva, A., Gout, C., & Hitchin, S. (2023). *Enhancing the Potential of Very Partial DNA Database Matches for Investigation Using Dynamic Cutoff Lr Thresholds – a Proof of Concept*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4620448
- Lund, S. P., & Iyer, H. (2017). *Likelihood Ratio as Weight of Forensic Evidence: A Closer Look* (Vol. 122, pp. 1–32). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7339646/>
- Meulenbroek, L., & Aben, D. (2019). *Grootschalig DNA-verwantschapsonderzoek*.
- Mortera, J. (2020). DNA Mixtures in Forensic Investigations: The Statistical State of the Art. In *Annual Review of Statistics and Its Application* (Vol. 7, Issue 1, pp. 111–142). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3553171#
- Nationaal Instituut voor Criminalistiek en Criminologie. (n.d.). *Genetische identificatie*. Retrieved January 30, 2024, from <https://nicc.fgov.be/genetische-identificatie>
- Nationaal Instituut voor Criminalistiek en Criminologie. (2022). *Nationale DNA-databanken: jaarverslag 2022*. <https://nicc.fgov.be/nationale-dna-databanken>
- National Research Council. (1996). Statistical Issues. In *The Evaluation of Forensic DNA evidence*.
- Ng, J., Oldt, R. F., & Kanthaswamy, S. (2018). *Assessing the FBI's Native American STR database for random match probability calculations*. <https://pubmed.ncbi.nlm.nih.gov/29175583/>
- Puch-Solis, R., Rodgers, L., Mazumder, A., Pope, S., Evett, I., Curran, J., & Balding, D. (2013). *Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters*. 7(5), 555–563. <https://pubmed.ncbi.nlm.nih.gov/23948327/>
- Putman, A. I., & Carbone, I. (2014). *Challenges in analysis and interpretation of microsatellite data for population genetic studies* (Vol. 4, Issue 22, pp. 4399–4428). <https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.1305>
- Steuve, S. (n.d.-a). *Beheer van de Nationale DNA-databanken*.
- Steuve, S. (n.d.-b). *Verbale schaal voor waarschijnlijkheid*.
- Steuve, S. (n.d.-c). *Verklaring van de formulieren Cluster, Info en Pedigree*.
- Steuve, S. (2023). *Registratiecriteria van DNA-profielen*.
- Udogadi, N. S., Abdullahi, M. K., Adams, T. B., Imose, O. P., & Esewi, A. D. (2020, July). *Forensic DNA Profiling: Autosomal Short Tandem Repeat as a Prominent Marker in Crime Investigation*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7444828/>
- Vanhooydonck, B., & Van Renterghem, P. (2024). *De nieuwe DNA-wet: een grote stap voorwaarts voor het DNA-onderzoek in Belgische strafdossiers* (Vol. 45, Issue 2, pp. 182–195).

Wivell, R., Kelly, H., Kokoszka, J., Daniels, J., Dickson, L., Buckleton, J., & Bright, J.-A. (2023).
An Investigation into Compound Likelihood Ratios for Forensic DNA Mixtures.

9 Bijlagen

Bijlage 1: Verbale schaal voor waarschijnlijkheid

Tabel 5: Verbale schaal voor waarschijnlijkheid

LR	De DNA-resultaten ondersteunen...
>1 miljard (10^9)	Met een aan zekerheid grenzende waarschijnlijkheid
>1 miljoen (10^6) – miljard (10^9)	Uitermate sterk
>100 000 – 1 miljoen (10^6)	Heel sterk
>10 000 – 100 000	Sterk
>1000 – 10 000	Matig
>100 – 1000	Zwak
1 - 100	De resultaten zijn niet informatief genoeg

Bijlage 2: Formule om de random match probability per locus te berekenen

The used calculation for the RMP are:	
For homozygous alleles:	$p^2 + p \times (1 - p) \times \theta$
For heterozygous alleles:	$2 \times p \times q$

Figuur 12: Formule random match probability per locus. Overgenomen uit Popstats module van CODIS. P: frequentie allel X; Q: frequentie allel Y; θ : maat voor de effecten van populatieonderverdeling (in teelt) = 0,01. De geschatte waarde van θ werd bepaald aan de hand de dataset van National Research Council (1996).

Bijlage 3: Code van de dichtheidsplot in RStudio

```
##### DENSITY PLOT #####  
  
# Om deze code te kunnen uitvoeren heb je enkele pakketten nodig, deze  
# roep je op via de library() code  
library(dplyr)  
library(ggplot2)  
  
# Verzamel gegevens uit CSV-bestand en importeer via Import Dataset
```

```
# Omdat de lege velden in de csv file niet automatisch gezien worden
als 'geen waarden' moeten we deze eruit halen om problemen in verdere
stappen te vermijden.
```

```
Filtered_High_7 <- Gegevens$High.7[is.finite(Gegevens$High.7)]
Filtered_High_8 <- Gegevens$High.8[is.finite(Gegevens$High.8)]
Filtered_High_9 <- Gegevens$High.9[is.finite(Gegevens$High.9)]
Filtered_High_10 <- Gegevens$High.10[is.finite(Gegevens$High.10)]
```

```
# Omdat de Density plot de waarden afleest in een andere vorm maken
we eerst een dataframe met 2 kolommen
```

```
Gegevens_long <- bind_rows(
  data.frame(Stringency = "High 7", Value = Filtered_High_7),
  data.frame(Stringency = "High 8", Value = Filtered_High_8),
  data.frame(Stringency = "High 9", Value = Filtered_High_9),
  data.frame(Stringency = "High 10", Value = Filtered_High_10)
)
```

```
# Om er zeker van te zijn dat de waarden in chronologische volgorde
weergegeven worden (High 7 naar High 10) gaan we er een stringency aan
toekennen
```

```
Gegevens_long$Stringency <- factor(Gegevens_long$Stringency, levels
= c("High 7", "High 8", "High 9", "High 10"))
```

```
# Met deze functie maken we de density plot
```

```
p <- ggplot(Gegevens_long, aes(x = Value, fill = Stringency)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("High 7" = "red", "High 8" = "blue",
"High 9" = "green", "High 10" = "purple")) +
  labs(title = "Density Plot of High 7 - High 10", x = "10Log LR", y
= "Density") +
  theme_minimal()
```

```
# Toon de plot
```

```
print(p)
```

Bijlage 4: Code van de Kruskal-Wallis test in RStudio

```
##### KRUSKAL WALLIS TEST #####

# Om deze code te kunnen uitvoeren heb je het FSA pakket nodig, dit
roep je op via library(FSA)

library(FSA)

# Verzamel gegevens uit CSV-bestand

# Voer de Kruskal-Wallis test uit
# De Kruskal-Wallis test gaat aan de hand van statistische
berekningen nagaan of er een significant verschil is tussen de 4
groepen

kruskal_result <- kruskal.test(list(Gegevens$High.7,
Gegevens$High.8, Gegevens$High.9, Gegevens$High.10))
print(kruskal_result)

# Omdat de Kruskal-Wallis test niet zegt tussen welke groepen er een
significant verschil is moeten we nog een Dunn-test uitvoeren
# De Dunn-test gaat kijken naar alle groepen onderling of er een
significant verschil is
```

Bijlage 5: Code van de post-hoc Dunn-test in RStudio

```
##### DUNN-TEST #####

# Omdat de Dunn-test de waarden afleest in een andere vorm maken we
eerst een dataframe met 2 kolommen
posthoc_data <- data.frame(
  Group = rep(c("High.7", "High.8", "High.9", "High.10"), each =
nrow(Gegevens)),
  Value = c(Gegevens$High.7, Gegevens$High.8, Gegevens$High.9,
Gegevens$High.10))
```

```

# De Dunn-test maakt gebruik van categoriën, daarom moeten we van de
Group = (High.7, High.8,...) als een categoriën/factoren definiëren

posthoc_data$Group <- as.factor(posthoc_data$Group)

# Omdat de Dunn-test functie de lege velden in de csv file niet
automatisch ziet als geen waarden moeten we deze uit de tabel halen.
posthoc_data <- na.omit(posthoc_data)

# Voer post-hoc Dunn-test uit met FSA-pakket
# We maken een Bonferonni correctie om multiplicity-effect te
vermijden
dunn_result <- dunnTest(x = posthoc_data$Value, g =
posthoc_data$Group, method = "bonferroni")

# Toon het resultaat van de Dunn-test
print(dunn_result)

```

Bijlage 6: Resultaten LR-waarden

Tabel 6: Resultaten LR-waarden

Match #	# Matching STR Loci	LR	Log 10 van LR	Cluster
2278248	7	598 400 000 000	11,77699158	A
2278250	7	598 400 000 000	11,77699158	A
2278251	7	598 400 000 000	11,77699158	A
2326364	7	11 490 000 000	10,06032003	/
2399822	7	47 440 000	7,67614468	B
2435430	7	3 235 000 000 000	12,50987429	C
2435431	7	3 235 000 000 000	12,50987429	C
2435433	7	21 040 000 000	10,32304574	D
2449923	7	598 400 000 000	11,77699158	A
2449925	7	598 400 000 000	11,77699158	A

2449926	7	5 984 000 000 000	12,77699158	A
1700207	7	622 300 000	8,793999801	E
1700209	7	622 300 000	8,793999801	E
1700211	7	622 300 000	8,793999801	E
2504962	7	2 459 000 000	9,390758529	F
2512692	7	1 837 000 000 000	12,26410916	G
2512693	7	1 837 000 000 000	12,26410916	G
2512746	7	1 837 000 000 000	12,26410916	G
2512747	7	1 837 000 000 000	12,26410916	G
2521071	7	12 500 000 000	10,09691001	H
2521073	7	12 500 000 000	10,09691001	H
2558052	7	18 670 000 000	10,27114432	I
2568890	7	1 837 000 000 000	12,26410916	G
2568891	7	1 837 000 000 000	12,26410916	G
2568944	7	1 837 000 000 000	12,26410916	G
2568945	7	1 837 000 000 000	12,26410916	G
2570875	7	1 837 000 000 000	12,26410916	G
2570902	7	1 837 000 000 000	12,26410916	G
2614892	7	1 837 000 000 000	12,26410916	G
2614893	7	1 837 000 000 000	12,26410916	G
2614946	7	1 837 000 000 000	12,26410916	G
2614947	7	1 837 000 000 000	12,26410916	G
2634773	7	7 117 000 000	9,852296966	J
2642818	7	4 456 000 000	9,648945182	K
2678816	7	1 242 000 000	9,094121596	L
2685595	7	1 490 000 000	9,173186268	M
2691868	7	6 192 000 000	9,791830948	N
2691869	7	6 192 000 000	9,791830948	N
2692868	7	4 456 000 000	9,648945182	K
2698849	7	1 837 000 000 000	12,26410916	G
2698866	7	1 837 000 000 000	12,26410916	G
2698878	7	1 837 000 000 000	12,26410916	G
2783149	7	42 320 000 000	10,62654566	O
2783150	7	163 400 000 000	11,21325205	O

2791441	7	215 500 000 000	11,33344727	P
2791446	7	15 470 000 000	10,18949031	E
2791447	7	15 470 000 000	10,18949031	E
2801726	7	8 361 000 000	9,922258223	R
2838210	7	1 837 000 000 000	12,26410916	G
2838237	7	1 837 000 000 000	12,26410916	G
2846099	7	3 306 000 000	9,519302849	Q
2872773	7	8 361 000 000	9,922258223	R
2910550	7	1 837 000 000 000	12,26410916	G
2910579	7	1 837 000 000 000	12,26410916	G
565541	7	1 201 000 000	9,079543007	S
3151628	7	306 300 000	8,486146997	/
3152143	7	17 540 000 000	10,24402959	T
3254775	7	221 700 000 000	11,34576569	U
3268822	7	1 821 000 000	9,260309946	V
3268823	7	4 120 000 000	9,614897216	V
3276690	7	1 781 000 000	9,250663919	W
3382643	7	9 251 000 000	9,966188681	X
3390795	7	565 900 000	8,752739694	Y
3393618	7	29 900 000 000	10,47567119	Z
433288	7	24 180 000 000	10,3834563	/
756604	7	14 040 000 000	10,14736711	/
532872	7	4 550 000 000	9,658011397	/
1487718	7	24 080 000 000	10,38165648	/
786078	7	8 058 000 000	9,906227263	/
579706	7	986 200 000	8,993964998	/
3427333	7	7 174 000 000	9,855761372	AA
2287862	8	338 400 000 000	11,52943035	AB
2323824	8	8 251 000 000 000	12,91650659	AC
2323825	8	99 300 000 000	10,99694925	AD
2326362	8	142 500 000 000 000	14,15381486	AE
2396218	8	11 850 000 000 000	13,07371835	AF
2399821	8	3 350 000 000	9,525044807	B
2433840	8	954 200 000 000	11,97963941	AG

2435432	8	3 715 000 000 000	12,56995882	D
2445257	8	80 060 000 000 000	13,90341559	AH
2445807	8	222 300 000 000	11,34693946	AI
2451682	8	87 800 000 000	10,94349452	AJ
2512720	8	106 000 000 000 000	14,02530587	G
2512721	8	106 000 000 000 000	14,02530587	G
2512722	8	106 000 000 000 000	14,02530587	G
2512723	8	106 000 000 000 000	14,02530587	G
2512724	8	283 900 000 000 000	14,45316539	G
2512725	8	283 900 000 000 000	14,45316539	G
2515787	8	35 440 000 000	10,54949371	AK
2515796	8	35 440 000 000	10,54949371	AK
2565491	8	27 680 000 000 000	13,44216609	AL
2565493	8	4 558 000 000	9,658774321	AM
2565490	8	27 680 000 000 000	13,44216609	AL
2568055	8	8 795 000 000	9,944235844	AN
2568918	8	106 000 000 000 000	14,02530587	G
2568919	8	106 000 000 000 000	14,02530587	G
2568920	8	106 000 000 000 000	14,02530587	G
2568921	8	106 000 000 000 000	14,02530587	G
2568922	8	106 000 000 000 000	14,02530587	G
2568923	8	106 000 000 000 000	14,02530587	G
2570889	8	106 000 000 000 000	14,02530587	G
2570890	8	106 000 000 000 000	14,02530587	G
2570891	8	283 900 000 000 000	14,45316539	G
2578933	8	390 300 000 000	11,59139855	AO
2614211	8	8 795 000 000	9,944235844	AN
2614920	8	106 000 000 000 000	14,02530587	G
2614921	8	106 000 000 000 000	14,02530587	G
2614922	8	106 000 000 000 000	14,02530587	G
2614923	8	106 000 000 000 000	14,02530587	G
2614924	8	283 900 000 000 000	14,45316539	G
2614925	8	283 900 000 000 000	14,45316539	G
2617720	8	518 700 000 000	11,71491625	AP

2624817	8	99 300 000 000	10,99694925	AD
2640419	8	509 700 000 000	11,70731463	AQ
2640421	8	4 182 000 000	9,621384028	AQ
2640420	8	233 300 000 000	11,36791474	AR
2642822	8	1 408 000 000 000	12,14860265	K
2688788	8	85 760 000 000 000	13,93328477	AS
2688789	8	121 100 000 000	11,08314414	AT
2690648	8	305 100 000 000	11,48444221	/
2692883	8	53 730 000 000	10,73021684	/
2692881	8	1 408 000 000 000	12,14860265	K
2692869	8	3 583 000 000 000	12,55424681	AU
2692867	8	3 484 000 000 000	12,54207815	AV
2698864	8	106 000 000 000 000	14,02530587	G
2698865	8	106 000 000 000 000	14,02530587	G
2700712	8	34 030 000 000	10,53186195	AW
2701945	8	119 200 000 000	11,07627626	AX
2736195	8	63 820 000 000	10,8049568	AY
2780810	8	781 200 000	8,892762235	/
2790077	8	3 583 000 000 000	12,55424681	AU
2791444	8	12 430 000 000 000	13,09447113	P
2791448	8	3 038 000 000 000	12,48258777	AZ
2801725	8	206 000 000 000	11,31386722	BA
2838224	8	106 000 000 000 000	14,02530587	G
2838225	8	106 000 000 000 000	14,02530587	G
2838226	8	283 900 000 000 000	14,45316539	G
2847593	8	11 760 000 000	10,07040732	BB
2849710	8	338 400 000 000	11,52943035	AB
2862652	8	325 000 000 000	11,51188336	V
2906990	8	85 760 000 000 000	13,93328477	AS
2910564	8	106 000 000 000 000	14,02530587	G
2910565	8	106 000 000 000 000	14,02530587	G
2910566	8	283 900 000 000 000	14,45316539	G
2923204	8	29 780 000 000	10,47392469	T
2925110	8	291 400 000 000	11,46448955	BC

2932382	8	100 600 000 000 000	14,00259798	BD
2933842	8	109 500 000 000	11,03941412	BE
2933843	8	109 500 000 000	11,03941412	BE
2980183	8	18 740 000 000	10,27276959	BF
2980462	8	11 660 000 000	10,06669855	BG
3038896	8	109 500 000 000	11,03941412	BE
3038897	8	109 500 000 000	11,03941412	BE
3039735	8	37 170 000 000 000	13,57019256	BH
1285645	8	9 579 000 000 000	12,98132017	BI
3049659	8	48 970 000 000	10,6899301	BJ
3097785	8	338 500 000 000	11,52955867	BK
3097782	8	1 533 000 000 000	12,18554215	BL
3143731	8	583 800 000 000	11,76626409	BM ⁷
3143725	8	1 533 000 000 000	12,18554215	BL
3152135	8	4 337 000 000 000	12,63718942	T
2189329	8	334 000 000 000	11,52374647	BN
3237972	8	4 537 000 000 000	12,65676878	BO
2633758	8	2 567 000 000 000	12,40942587	/
3253228	8	2 731 000 000 000	12,4363217	BP
3268829	8	13 810 000 000 000	13,14019368	V
3276691	8	19 640 000 000	10,29314148	W
3365620	8	7 843 000 000	9,894482215	BQ
3390790	8	48 970 000 000	10,6899301	BJ
3393617	8	4 757 000 000 000	12,67733315	Z
3409195	8	85 760 000 000 000	13,93328477	AS
3411538	8	16 670 000 000	10,2219356	BK
3427334	8	1 066 000 000	9,027757205	/
3433592	8	106 600 000 000	11,0277572	BR
3440789	8	3 445 000 000	9,537189226	BS
3448446	8	11 800 000 000 000	13,07188201	BT
3464710	8	63 940 000 000	10,80577263	BU
2279883	9	2 841 000 000 000	12,45347123	BV
2287865	9	87 640 000 000 000	13,94270237	BW

⁷ Hier wordt verwezen naar de naam van de cluster en niet naar de afkorting van beperkte merkers.

2287869	9	87 640 000 000 000	13,94270237	BW
2323837	9	197 100 000 000 000	14,29468662	BX
2369906	9	41 600 000 000	10,61909333	BY
2382152	9	103 900 000 000	11,01661555	BZ
2437446	9	29 780 000 000	10,47392469	CA
2451684	9	11 800 000 000 000	13,07188201	/
1167088	9	239 400 000 000	11,37912415	CB
2512734	9	670 700 000 000 000	14,82652831	G
2512735	9	670 700 000 000 000	14,82652831	G
2512810	9	1 815 000 000 000	12,25887663	CC
2514047	9	46 020 000 000 000	13,66294661	A
2521070	9	392 000 000 000 000	14,59328607	CD
2369932	9	14 660 000 000 000	13,16613397	CE
2568885	9	6 090 000 000 000	12,78461729	CF
2568932	9	670 700 000 000 000	14,82652831	G
2568933	9	670 700 000 000 000	14,82652831	G
2568966	9	17 510 000 000 000	13,24328615	G
2570896	9	670 700 000 000 000	14,82652831	G
2614934	9	670 700 000 000 000	14,82652831	G
2614935	9	670 700 000 000 000	14,82652831	G
2642343	9	87 640 000 000 000	13,94270237	CG
2642349	9	87 640 000 000 000	13,94270237	BW
2642816	9	70 130 000 000	10,84590384	CH
2698871	9	670 700 000 000 000	14,82652831	G
2838231	9	670 700 000 000 000	14,82652831	G
2862651	9	1 417 000 000 000	12,15136985	V
2910572	9	670 700 000 000 000	14,82652831	G
3047679	9	6 748 000 000 000 000	15,82917507	CI
3250702	9	40 830 000 000	10,61097938	CJ
3254780	9	1 022 000 000 000 000	15,0094509	U
3268817	9	35 290 000 000 000	13,54765166	V
3268819	9	2 287 000 000 000	12,35926616	V
3356840	9	253 400 000 000	11,40380661	CK
3374046	9	830 600 000 000	11,91939193	V

3374077	9	35 290 000 000 000	13,54765166	V
3376123	9	870 300 000 000	11,93966898	CL
3201256	9	183 300 000 000	11,26316246	CM
3389950	9	9 091 000 000	9,958611658	CN
3409199	9	4 880 000 000 000	12,68841982	CO
3419876	9	196 000 000 000	11,29225607	W
177582	9	117 300 000 000	11,06929801	CP
3097786	10	744 600 000 000 000	14,87192303	CQ
2558049	10	340 500 000 000 000	14,53211712	CR
2780777	10	1 570 000 000 000 000	15,19589965	CS
2570886	10	76 450 000 000 000 000	16,88337749	G
2568898	10	76 450 000 000 000 000	16,88337749	G
2688786	10	17 410 000 000 000	13,24079877	CT
2803910	10	231 700 000 000 000	14,36492603	CU
2838236	10	76 450 000 000 000 000	16,88337749	G
2512688	10	76 450 000 000 000 000	16,88337749	G
3249381	10	708 700 000 000 000	14,85046243	CV
3356841	10	57 500 000 000 000	13,75966784	CK
2634766	10	3 382 000 000 000	12,5291736	CW
2614939	10	76 450 000 000 000 000	16,88337749	G
2512736	10	76 450 000 000 000 000	16,88337749	G
2512715	10	76 450 000 000 000 000	16,88337749	G
2433841	10	20 850 000 000 000 000	16,31910606	CX
2372378	10	4 143 000 000 000 000	15,61731493	CY
3413216	10	119 300 000 000	11,07664044	CZ
2856179	10	1 587 000 000 000 000	15,20057693	DA
2678820	10	3 882 000 000 000	12,58905553	L
3253229	10	52 830 000 000 000	13,72288061	BP
2692873	10	1 199 000 000 000 000	15,07881918	DB
2910559	10	76 450 000 000 000 000	16,88337749	G
2399823	10	282 100 000 000 000	14,45040309	AK
2698854	10	76 450 000 000 000 000	16,88337749	G
2864369	10	4 417 000 000 000	12,6451274	DC
2838215	10	76 450 000 000 000 000	16,88337749	G

2437429	10	4 281 000 000 000 000	15,63154523	/
2514050	10	1 315 000 000 000 000	15,11892575	DD
2334913	10	161 200 000 000 000	14,20736504	DE
2568893	10	76 450 000 000 000 000	16,88337749	G
2849712	10	46 380 000 000 000	13,66633074	DF
2399839	10	34 640 000 000 000	13,53957788	DG
3247272	10	825 100 000 000	11,91650659	DH
2279909	10	9 337 000 000 000	12,97020736	DI
2512708	10	76 450 000 000 000 000	16,88337749	G
3152139	10	79 620 000 000 000	13,90102217	DJ
2910556	10	76 450 000 000 000 000	16,88337749	G
2512690	10	76 450 000 000 000 000	16,88337749	G
2614942	10	76 450 000 000 000 000	16,88337749	G
2568909	10	76 450 000 000 000 000	16,88337749	G
3237969	10	9 625 000 000 000	12,98340074	DK
2698850	10	76 450 000 000 000 000	16,88337749	G
3438240	10	46 730 000 000 000	13,66959578	V
3237970	10	545 900 000 000 000 000	17,73711309	DL
2910567	10	76 450 000 000 000 000	16,88337749	G
2910571	10	76 450 000 000 000 000	16,88337749	G
2568913	10	76 450 000 000 000 000	16,88337749	G
2572979	10	10 250 000 000 000	13,01072387	DM
2326367	10	654 900 000 000	11,81617499	DN
2570899	10	76 450 000 000 000 000	16,88337749	G
2779282	10	11 260 000 000 000	13,05153839	R
2634763	10	1 096 000 000 000 000	15,03981055	/
2698853	10	76 450 000 000 000 000	16,88337749	G
2189334	10	1 259 000 000 000 000	15,10002573	BN
2692871	10	27 710 000 000 000	13,44263653	DO
2614927	10	76 450 000 000 000 000	16,88337749	G
2698889	10	1 127 000 000 000 000	15,05192392	G
2910569	10	76 450 000 000 000 000	16,88337749	G
2515733	10	11 720 000 000 000	13,06892761	DP
2568900	10	76 450 000 000 000 000	16,88337749	G

2555484	10	27 710 000 000 000	13,44263653	DO
3268813	10	2 647 000 000 000 000	15,42275394	/
2512768	10	1 127 000 000 000 000	15,05192392	G
2570887	10	76 450 000 000 000 000	16,88337749	G
2433846	10	11 180 000 000 000	13,0484418	DQ
2555483	10	54 260 000 000 000	13,73447979	DR
2512742	10	76 450 000 000 000 000	16,88337749	G
2910547	10	76 450 000 000 000 000	16,88337749	G
2390878	10	74 020 000 000 000	13,86934908	DS
2910551	10	76 450 000 000 000 000	16,88337749	G
2556816	10	17 410 000 000 000	13,24079877	CT
2698873	10	76 450 000 000 000 000	16,88337749	G
2512741	10	76 450 000 000 000 000	16,88337749	G
2568902	10	76 450 000 000 000 000	16,88337749	G
2369900	10	34 640 000 000 000	13,53957788	DG
3244433	10	1 267 000 000 000	12,10277661	DT
2570885	10	76 450 000 000 000 000	16,88337749	G
2572978	10	4 129 000 000 000	12,61584488	DU
2838219	10	76 450 000 000 000 000	16,88337749	G
2568917	10	76 450 000 000 000 000	16,88337749	G
2512699	10	76 450 000 000 000 000	16,88337749	G
2570876	10	76 450 000 000 000 000	16,88337749	G
2514045	10	1 315 000 000 000 000	15,11892575	DD
2838211	10	76 450 000 000 000 000	16,88337749	G
2568935	10	76 450 000 000 000 000	16,88337749	G
3419878	10	18 580 000 000 000	13,26904571	DV
2568937	10	76 450 000 000 000 000	16,88337749	G
2512743	10	76 450 000 000 000 000	16,88337749	G
2568931	10	76 450 000 000 000 000	16,88337749	G
2558098	10	72 410 000 000 000	13,85979855	/
3456995	10	2 573 000 000 000 000	15,41043979	/
2568901	10	76 450 000 000 000 000	16,88337749	G
3090278	10	134 200 000 000	11,12775252	DW
2570895	10	76 450 000 000 000 000	16,88337749	G

2432940	10	5 872 000 000 000 000	15,76878605	DX
2323838	10	1 566 000 000 000 000	15,19479176	DY
2512706	10	76 450 000 000 000 000	16,88337749	G
2838297	10	46 730 000 000 000	13,66959578	V
2698867	10	76 450 000 000 000 000	16,88337749	G
