



UNIVERSIDADE D  
COIMBRA

Susana Filipa Martins Godinho

**VALIDAÇÃO DO EQUIPAMENTO**  
***APPLIED BIOSYSTEMS™ SEQSTUDIO™***  
***GENETIC ANALYZER***

Dissertação realizada no âmbito do Mestrado em Medicina Legal e Ciências Forenses, orientada pelo Professor Doutor Francisco Corte-Real Gonçalves e pela Mestre Lisa Andrade Sampaio e apresentada à Faculdade de Medicina da Universidade de Coimbra

Setembro de 2022





FMUC FACULDADE DE MEDICINA  
UNIVERSIDADE DE COIMBRA

Susana Filipa Martins Godinho

# Validação do equipamento *Applied Biosystems™ SeqStudio™ Genetic Analyzer*



Dissertação realizada no âmbito do Mestrado em Medicina Legal e Ciências Forenses, orientada pelo Professor Doutor Francisco Corte-Real Gonçalves e pela Mestre Lisa Andrade Sampaio e apresentada à Faculdade de Medicina da Universidade de Coimbra

Setembro de 2022

Imagem de sub-capa:  
Equipamento *SeqStudio™ Genetic Analyzer*. Adaptado de (Applied Biosystems, 2019c)

***A vida sem ciência é uma espécie de morte.***



- Sócrates (470 - 399 a.C.)



## AGRADECIMENTOS

---

Ao Professor Doutor Francisco Corte-Real Gonçalves, Presidente do Instituto Nacional de Medicina Legal e Ciências Forenses (INMLCF) e orientador, por me indicar o Serviço de Genética e Biologia Forenses (SGBF) como entidade acolhedora para a realização da presente dissertação e pela apreciação científica da mesma.

À Mestre Lisa Sampaio, coorientadora, por me ter concedido a oportunidade de realizar este trabalho no Serviço da Delegação do Centro (SGBF-C), pela amizade e simpatia com que me recebeu e pela disponibilidade e paciência com que sempre me auxiliou.

À Doutora Nair Gouveia, responsável pelo equipamento *SeqStudio*<sup>TM</sup>, que me acompanhou desde o primeiro dia na apresentação das instalações e da dinâmica no Serviço, por todo o tempo dispensado para me auxiliar na concretização prática laboratorial do presente estudo, e por partilhar comigo dicas de trabalho e de resolução de imprevistos laboratoriais, enquanto gerava uma nova vida dentro de si.

À Dra. Virgínia Lopes, pelo seu incansável apoio ao dispensar inúmeras vezes do seu tempo no trabalho de rotina para me conceder documentos relevantes e esclarecer conceitos teórico-práticos importantes para a elaboração desta Tese.

A todos os restantes colaboradores do SGBF-C, em especial ao Dr. Armando Serra, Dra. Maria João Porto, Dra. Ana Margarida Bento e Dra. Vanessa Bogas, pelo contributo indireto que prestaram, por todos os conhecimentos que me transmitiram e cuja cooperação foi crucial na minha ambientação à rotina do Serviço.

À minha família e amigos próximos, pelo apoio e incentivo para que nunca desistisse, por mais desfavoráveis que estivessem as minhas condições pessoais ou de saúde.

Às colegas que realizaram igualmente as suas Teses no mesmo Serviço - Francisca Lopes, Joana Ferreira e Márcia Madureira - com as quais acabei inevitavelmente por desenvolver grande relação de amizade e companheirismo, e às quais desejo uma excelente defesa de tese.

A Deus que sempre me ajudou nos momentos de aflição.

Obrigada a todos!





## RESUMO

---

As atividades desempenhadas pela Genética Forense centram-se na determinação e análise de perfis de ADN, particularmente de zonas do genoma nuclear humano altamente polimórficas designadas *Short Tandem Repeats* (STRs), pelo grande potencial que as mesmas apresentam em distinguir e identificar indivíduos. Os perfis genéticos são atualmente obtidos sob a forma de eletroferogramas (EFs), através da separação e deteção fluorescente desses fragmentos específicos por eletroforese capilar (EC), em sequenciadores automáticos. Uma vez que os perfis são frequentemente usados como prova em tribunal, podendo deles derivar decisões jurídicas sobre os indivíduos envolvidos em processos judiciais, é de extrema importância garantir a veracidade dos resultados obtidos e a confiança na aptidão instrumental dos equipamentos que os geraram. Por essa razão, é necessário que os sequenciadores automáticos utilizados nos laboratórios de genética forense estejam devidamente validados, assegurando assim, que a sua utilização permitirá uma correta resolução de questões legais.

Com o presente trabalho pretendeu-se validar internamente o novo sequenciador automático *Applied Biosystems™ SeqStudio™ Genetic Analyzer* recentemente adquirido pelo SGBF-C, contribuindo para a sua futura introdução na rotina laboratorial do serviço. Para proceder à validação interna do *SeqStudio™*, foram estudados os seus limiares de sensibilidade, e realizados estudos comparativos entre ele e o sequenciador em atual utilização no SGBF-C e já completamente validado - *Applied Biosystems™ 3500 Genetic Analyzer*. Para isso, recorreu-se a amostras de sangue e saliva da rotina do SGBF-C. O ADN das amostras foi extraído através de *Prep-n-Go™* e *PrepFiler Express™*, a sua quantificação foi realizada com recurso ao *kit Quantifiler™ Trio*, e a sua amplificação com o *kit GlobalFiler™*. As amostras foram aplicadas nas condições padrão de eletroforese, a 10s de injeção, e os EFs obtidos foram analisados no *software GeneMapper™ ID-X*.

Este estudo de validação permitiu determinar os Limiares Analítico, de *Stutter*, Estocástico e de Desequilíbrio de Heterozigóticos, como parâmetros a estabelecer no SGBF-C, para garantir corretas conclusões aquando da análise e interpretação dos perfis gerados pelo *SeqStudio™* para amostras amplificadas com *GlobalFiler™*. Comprovou-se a capacidade do *SeqStudio™* em produzir perfis equivalentes aos do *3500*, e demonstrou-se que o mesmo se encontra apto para executar a sua função de EC, nas condições laboratoriais do SGBF-C, para fins de identificação humana. Este trabalho possibilitou também caracterizar as principais vantagens e limitações do *SeqStudio™* face ao *3500*.

**Palavras-Chave:** Eletroforese Capilar; *GlobalFiler™*; Limiares de sensibilidade; *SeqStudio™ Genetic Analyzer*; Validação interna



## ABSTRACT

---

The activities performed by Forensic Genetics focus on the determination and analysis of DNA profiles, particularly of highly polymorphic areas of the human nuclear genome called Short Tandem Repeats (STRs), due to their enormous potential to distinguish and identify individuals. Genetic profiles are currently obtained in the form of electropherograms, through the separation and fluorescent detection of these specific fragments by capillary electrophoresis, in automated DNA sequencers. Since profiles are often used as evidence in court and can lead to legal decisions over individuals involved in forensic cases, it is extremely important to guarantee the veracity of the results obtained and confidence in the instrumental aptitude of the equipment that generated them. For this reason, the automated sequencers used in forensic genetics laboratories must be properly validated, thus ensuring that their use will allow a correct resolution of legal issues.

The present work aimed to internally validate the new *Applied Biosystems™ SeqStudio™ Genetic Analyzer* automated sequencer recently acquired by the SGBF-C, thus contributing to its future introduction in the laboratory routine.

To perform the internal validation of *SeqStudio™*, its sensitivity thresholds were studied, and comparative studies were conducted between it and the sequencer currently used in the SGBF-C and already fully validated - *Applied Biosystems™ 3500 Genetic Analyzer*. For this, blood and saliva samples from the SGBF-C routine were used. The samples' DNA was extracted using *Prep-n-Go™* and *PrepFiler Express™*, its quantification was performed using the *Quantifiller™ Trio kit*, and its amplification with the *GlobalFiler™ kit*. The samples were applied under standard electrophoresis conditions, at the injection time of 10s, and the electropherograms obtained were analyzed under the *GeneMapper™ ID-X software*.

This validation study allowed to determine the Analytical, Stutter, Stochastic and Heterozygotic Imbalance Thresholds, as parameters to be established in the SGBF-C, to ensure correct conclusions when analyzing and interpreting the profiles generated by *SeqStudio™* for samples amplified with *GlobalFiler™*. The ability of *SeqStudio™* to produce profiles equivalent to those of the *3500* was proved, and it was demonstrated that it is able to perform its EC function, under the laboratory conditions of the SGBF-C, for human identification purposes. This work also made it possible to characterize the main advantages and limitations of *SeqStudio™* compared to the *3500*.

**Keywords:** Capillary Electrophoresis; *GlobalFiler™*; Sensitivity thresholds; *SeqStudio™ Genetic Analyzer*; Internal validation



## ÍNDICE

---

<b>Agradecimentos</b> .....	i
<b>Resumo</b> .....	iii
<b>Abstract</b> .....	v
<b>Índice de Figuras</b> .....	ix
<b>Índice de Tabelas</b> .....	xii
<b>Lista de Siglas e Abreviaturas</b> .....	xv
<b>1. Introdução</b> .....	1
1.1 Princípios de Genética Forense.....	1
1.2. Introdução ao SGBF .....	7
1.3. Tipo de amostras em contexto forense .....	8
1.4. Percurso laboratorial das amostras para determinação de perfis genéticos.....	10
1.4.1. Extração de ADN .....	10
1.4.2. Quantificação do ADN extraído.....	11
1.4.3. Amplificação de STRs por PCR .....	13
1.4.4. Separação e Detecção do produto amplificado.....	16
1.5. Análise e interpretação dos perfis genéticos obtidos .....	21
1.5.1. Ruído de fundo .....	21
1.5.2. <i>Stutters</i> .....	22
1.5.3. Picos partidos .....	24
1.5.4. <i>Pull-ups</i> e <i>Spikes</i> .....	25
1.5.5. <i>Drop-in</i> .....	27
1.5.6. Alelos <i>Off-Ladder</i> .....	27
1.5.7. Desequilíbrio Heterozigótico .....	28
1.5.8. <i>Drop-out</i> .....	29
1.5.9. Degradação .....	32
1.6. Sequenciadores Automáticos – <i>Seqstudio™ Genetic Analyzer (AB™)</i> .....	34
1.7. Contextualização da Validação.....	37
<b>2. Justificação e Objetivos</b> .....	40
2.1. Objetivo Geral.....	40
2.2. Objetivos Específicos.....	40

<b>3. Materiais e Métodos</b> .....	42
3.1. Amostras biológicas.....	42
3.2. Colheita e suporte das amostras .....	42
3.3. Práticas de prevenção da contaminação .....	42
3.4. Metodologias de Extração .....	44
3.4.1. Extração com <i>Prep-n-Go™</i> .....	44
3.4.2. Extração com <i>PrepFiler Express™</i> .....	44
3.5. Metodologia de Quantificação – <i>Kit Quantifiler™ Trio</i> .....	45
3.6. Metodologia de Amplificação – <i>Kit Globalfiler™</i> .....	46
3.7. Metodologia de Separação e Detecção do produto amplificado – Aplicação nos Sequenciadores Automáticos <i>3500</i> e <i>Seqstudio™</i> .....	49
3.8. Parâmetros estudados.....	50
3.8.1. Limiares de sensibilidade.....	50
I – Limiar Analítico (LA) .....	50
II – Limiar de <i>Stutter</i> (LS) .....	51
III – Limiar Estocástico (LE).....	54
IV – Limiar de Desequilíbrio de Heterozigóticos (LD) .....	55
3.8.2. Estudos Comparativos <i>Seqstudio™</i> vs. <i>3500</i> .....	56
<b>4. Resultados e Discussão</b> .....	57
4.1. Limiar Analítico (LA) .....	57
4.2. Limiar de <i>Stutter</i> (LS).....	59
4.3. Limiar Estocástico (LE) .....	59
4.4. Limiar de Desequilíbrio de Heterozigóticos (LD) .....	64
4.5. Estudos Comparativos <i>Seqstudio™</i> vs. <i>3500</i> .....	68
4.6. Vantagens e Limitações <i>Seqstudio™</i> vs. <i>3500</i> .....	75
<b>5. Conclusões</b> .....	77
<b>6. Limitações do Estudo e Perspetivas Futuras</b> .....	79
<b>7. Referências Bibliográficas</b> .....	81
<b>8. Anexos</b> .....	85

## ÍNDICE DE FIGURAS

---

<b>Figura 1.</b> Estrutura e constituição da molécula de ADN. Adaptado de (Griffiths, 2012).....	1
<b>Figura 2.</b> Organização do ADN e da sua frequência estimada no genoma humano. Adaptado de (Goodwin <i>et al.</i> , 2011) .....	3
<b>Figura 3.</b> Estrutura de um STR. Estão representados dois alelos possíveis para um mesmo <i>locus</i> (D8S1179). Os alelos são designados de acordo com o número de cópias do motivo de repetição específico do <i>locus</i> (TCTA). No alelo representado superiormente, o motivo ocorre 8 vezes, adotando, portanto, a denominação de “Alelo 8” e no representado inferiormente, ocorre 10 vezes, sendo por isso chamado de “Alelo 10”. Um indivíduo portador destes alelos, apresenta um genótipo heterozigótico de 8,10 neste <i>locus</i> . Adaptado de (Goodwin <i>et al.</i> , 2011).....	4
<b>Figura 4.</b> <i>Core loci</i> que compõem o atual sistema CODIS, e sua distribuição cromossômica. A amarelo estão representados os <i>loci</i> autossômicos e a azul, a Amelogenina (AMEL), <i>locus</i> existente nos cromossomas sexuais X (AMELX) e Y (AMELY) e utilizado como marcador de determinação de género sexual. Os 12 <i>loci</i> autossômicos FGA, TH01, vWA, D3S1358, D8S1179, D18S51, D21S11, D1S1656, D2S441, D10S1248, D12S391 e D22S1045, são partilhados pelo ESS. Adaptado de (Butler & Li, 2014; Udogadi <i>et al.</i> , 2020).....	6
<b>Figura 5.</b> Representação esquemática da tecnologia de qPCR baseada na utilização de sondas <i>TaqMan™</i> . Adaptado de (Goodwin <i>et al.</i> , 2011).....	12
<b>Figura 6.</b> Representação esquemática da amplificação de uma sequência de interesse por PCR. Adaptado de (Khan Academy, 2017) .....	14
<b>Figura 7.</b> Marcadores que o <i>kit GlobalFiler™</i> permite amplificar em <i>multiplex</i> e seus respectivos intervalos de tamanhos alélicos e marcações fluorescentes. Adaptado de (Applied Biosystems, 2019b; Butler & Gittelsohn, 2015).....	15
<b>Figura 8.</b> Ilustração esquemática da separação de fragmentos por eletroforese capilar, e sua deteção fluorescente, num sequenciador automático. Adaptado de (Butler, 2011) .....	17
<b>Figura 9.</b> Painel de cor de um eletroferograma após análise por <i>GeneMapper™ ID-X (AB™)</i> . Os picos detetados em RFUs estão devidamente identificados, com o alelo a que correspondem e o respetivo tamanho de fragmento determinado (bp).....	18
<b>Figura 10.</b> Percurso laboratorial para análise de fragmentos vs. sequenciação de Sanger. Na análise de fragmentos, são usados <i>primers</i> fluorescentes, enquanto na sequenciação de	

Sanger a fluorescência é usada em ddNTPs. O resultado da análise de fragmentos culmina na atribuição alélica dos marcadores em análise, enquanto na sequenciação obtém-se a ordem de bases nucleotídicas dos fragmentos de interesse. Adaptado de (Thermo Fisher Scientific, 2020).....20

**Figura 11.** Exemplo que demonstra a importância do estabelecimento de um LA. Um pico com altura inferior ao LA não pode ser considerado com segurança como sendo efetivamente um sinal analítico, pois poderá corresponder a ruído de fundo. Um pico com altura superior ao LA pode ser considerado como sinal analítico de confiança, i.e., como um produto efetivamente resultante da amplificação. Adaptado de (Butler, 2010) .....22

**Figura 12.** Exemplo de picos *stutter* - num *locus* heterozigótico. O *stutter* à esquerda corresponde apenas a 6.0% da altura do alelo real (15); o *stutter* da direita corresponde apenas a 7.8% da altura do alelo verdadeiro (18). Os picos relativos a esses artefactos não surgem identificados como alelos, porque neste EF foi já aplicado um filtro de *stutters*. Adaptado de (Butler, 2010) .....24

**Figura 13.** Picos partidos (-A) vs. picos correspondentes aos alelos reais (+A). Adaptado de (Butler, 2010) .....25

**Figura 14.** Picos verdes de baixa intensidade imediatamente abaixo dos picos azuis de elevada intensidade, são *pull-ups* (a); O pico que atravessa todos os painéis de cor, na mesma posição, é um *spike* (b). Adaptado de (Butler, 2010) .....26

**Figura 15.** Esquema representativo de possíveis resultados de amplificação: **a)** alelos com picos totalmente equilibrados - amplificação ideal; **b)** ligeiro desnível entre picos, com alelos bem balanceados – típico de amostras seguras; **c)** desequilíbrio heterozigótico – típico de amostras com baixa quantidade de ADN ou de mistura; **d)** *drop-out* alélico. Adaptado de (Butler, 2010) .....30

**Figura 16.** Diferença interpretativa entre um pico que se situe abaixo do LE e acima do LE. Se num determinado *locus*, surgir um pico abaixo do LE estabelecido (como o representado à esquerda), é possível ter ocorrido *drop-out* do seu eventual alelo irmão, não se conseguindo afirmar com certeza, a homozigotia no *locus*. Mas se surgir um pico acima do LE (como o representado à direita), é razoável admitir que não terá ocorrido *drop-out* e, portanto, o *locus* pode ser considerado com segurança, como homozigótico. Adaptado de (Crespillo *et al.*, 2012).....31

**Figura 17.** Perfil genético característico de uma amostra com ADN degradado. Adaptado de (Goodwin *et al.*, 2011) .....32



**Figura 18. Componentes do cartucho *SeqStudio™ Genetic Analyzer Cartridge v2*** (em cima): 1 – Polímero universal (POP-1™); 2 - Sistema de distribuição de polímero (bomba polímero para os capilares); 3 – Reservatório contendo o tampão do ânodo; 4 – Janela de detecção ótica; 5 – Sulco do cartucho para sua inserção no instrumento; 6 – Pontas dos 4 capilares (extremidade do cátodo), encontrando-se a restante matriz protegida no interior do cartucho; **Componentes da frente e interior do instrumento *SeqStudio™*** (em baixo): 7 – Cartucho; 8 – *Autosampler*; 9 – Tampão do cátodo; 10 – Suporte para a placa ou tubos das amostras; 11 – Ecrã tátil; 12 – porta USB. Adaptado de (Applied Biosystems, 2019g).....34

**Figura 19.** Componentes e configuração do *3500 Genetic Analyzer*. Adaptado de (Applied Biosystems, 2015).....35

**Figura 20.** Perfil gerado pelo *SeqStudio™*, para a amostra anonimizada como 2021-4168-1.C2.A1.....69

**Figura 21.** Perfil gerado pelo *3500*, para a amostra anonimizada como 2021-4168-1.C2.A1. ....70

**Figura 22.** Alelo 15 detetado no marcador D3S1358 pelo *SeqStudio™* (à esquerda) vs. pelo *3500*, para a amostra anonimizada como 2021-1958-1.C1.X1.A1. No EF do *SeqStudio™*, o GM atribuiu um sinalizador de qualidade laranja devido à intensidade de sinal ser demasiado elevada e próxima do limite de saturação do detetor.....71

**Figura 23.** Picos identificados alelicamente pelo GM, no marcador D2S441, no EF gerado pelo *SeqStudio™* (à esquerda) vs. pelo *3500*, para a amostra anonimizada como 2022-1826-1.C1.A1. O pico identificado como “alelo 10” pelo GM, corresponderá a um *stutter* - do alelo 11, que não foi filtrado. ....73

## ÍNDICE DE TABELAS

---

<b>Tabela 1.</b> Tipo de polímero e de capilares utilizados no 3500 e no <i>SeqStudio™</i> . Adaptado de (Applied Biosystems, 2019a).....	35
<b>Tabela 2.</b> Componentes e respectivos volumes adicionados a cada poço para a reação de quantificação. ....	46
<b>Tabela 3.</b> Componentes e respectivos volumes adicionados a cada poço para a reação de amplificação de amostras seguras de saliva. O mesmo se aplicou às amostras seguras de sangue que atravessaram pela fase de extração e quantificação e que foram seguidamente amplificadas como amostras seguras.....	47
<b>Tabela 4.</b> Componentes e respectivos volumes ou quantidades adicionados a cada poço para a reação de amplificação direta de amostras seguras de sangue. ....	47
<b>Tabela 5.</b> Componentes e respectivos volumes adicionados a cada poço para a reação de amplificação de amostras seguras de sangue tratadas como amostras problema (volume final de amplificação 25 µL/poço).....	48
<b>Tabela 6.</b> Programa de amplificação “ <i>GlobalFiler – Seguras – 27C</i> ” selecionado no termociclador para amostras seguras.....	48
<b>Tabela 7.</b> Programa de amplificação “ <i>GlobalFiler – Vestigiais – 29C</i> ” selecionado no termociclador para amostras seguras tratadas como amostras problema. ....	48
<b>Tabela 8.</b> Componentes e respectivos volumes adicionados a cada poço para aplicação das amostras em sequenciador automático. ....	49
<b>Tabela 9.</b> Valores de LS estabelecidos para os marcadores autossômicos do <i>kit GlobalFiler™</i> segundo os estudos de validação de desenvolvimento realizados pelo fabricante, aplicáveis como filtro de <i>stutter</i> no 3500 e no <i>SeqStudio™</i> (Applied Biosystems, 2019f; RV-SGBF-C-006, 2020).....	53
<b>Tabela 10.</b> Valores de LA obtidos em cada painel de cor, utilizando o <i>kit GlobalFiler™</i> . Na coluna à direita estão os maiores valores de LA observados para cada uma das 3 fórmulas de cálculo utilizadas. ....	57
<b>Tabela 11.</b> Genótipos identificados em cada marcador contabilizado para os cálculos do LE, em cada uma das 6 diluições da Amostra A tratadas como amostras seguras (12,5 µL de volume de amplificação e 27 ciclos de PCR). Indicação dos resultados da quantificação da	

solução inicial de ADN extraído da Amostra A e das diluições a partir dela preparadas, bem como as concentrações esperadas para essas mesmas diluições.....60

**Tabela 12.** Genótipos identificados em cada marcador contabilizado para os cálculos do LE, em cada uma das 6 diluições da Amostra B tratadas como amostras seguras (12,5 µL de volume de amplificação e 27 ciclos de PCR). Indicação dos resultados da quantificação da solução inicial de ADN extraído da Amostra B e das diluições a partir dela preparadas, bem como as concentrações esperadas para essas mesmas diluições.....60

**Tabela 13.** Genótipos identificados em cada marcador contabilizado para os cálculos do LE, em cada uma das 6 diluições da Amostra A tratadas como amostras problema (25 µL de volume de amplificação e 29 ciclos de PCR). Indicação dos resultados da quantificação da solução inicial de ADN extraído da Amostra A e das diluições a partir dela preparadas, bem como as concentrações esperadas para essas mesmas diluições.....61

**Tabela 14.** Genótipos identificados em cada marcador contabilizado para os cálculos do LE, em cada uma das 6 diluições da Amostra B tratadas como amostras problema (25 µL de volume de amplificação e 29 ciclos de PCR). Indicação dos resultados da quantificação da solução inicial de ADN extraído da Amostra B e das diluições a partir dela preparadas, bem como as concentrações esperadas para essas mesmas diluições.....61

**Tabela 15.** Valores de LE obtidos para a Amostra A e B nas condições laboratoriais de amostras seguras (12,5µL, 27 ciclos) e problema (25µL, 29 ciclos). .....62

**Tabela 16.** Valores de LD obtidos para cada marcador autossômico, contabilizando todos os picos alélicos dos sistemas heterozigóticos encontrados nos perfis (à esquerda) vs. contabilizando apenas os picos alélicos que diferiam em mais do que 1 unidade de repetição (à direita).....64



## LISTA DE SIGLAS E ABREVIATURAS

---

**A** - Adenina

**AB™** - *Applied Biosystems™*

**ADN** - **Á**cido **D**esoxirribonucleico

**ADNmt** - **ADN** mitocondrial

**ADNnu** - **ADN** nuclear

**AMEL** - Gene homólogo da **A**melogenina

**Automate Express™** - *Applied Biosystems™ AutoMate Express™ Forensic DNA Extraction System*

**BTA** - **B**one, **T**eeth and **A**dhesive

**C** - Citosina

**CODIS** - *Combined DNA Index System*

**CR** - **C**riminalística **B**iológica

**ddNTPs** - Didesoxinucleótidos trifosfatados

**dNTPs** - **D**esoxinucleótidos **t**rifosfatados

**DTT** - **D**itiotreitol

**EC** - **E**letoforese **C**apilar

**EF** - **E**letoferograma

**ENFSI** - *European Network of Forensic Science Institutes* - Rede Europeia de Institutos de Ciências Forenses

**ESS** - *European Standard Set* - Painel Europeu de marcadores genéticos

**FBI** - *Federal Bureau of Investigation*

**GHEP- ISFG** - *Grupo de Habla Española y Portuguesa de la International Society for Forensic Genetics* - Grupo de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense

**GlobalFiler™** - *Applied Biosystems™ GlobalFiler™ PCR Amplification Kit*

**GM** - *Applied Biosystems™ GeneMapper™ ID-X Software*

**G** - Guanina

**h** - horas

**HID** - *Human Identification Analysis* - Análise de fragmentos para Identificação Humana

**H<sub>2</sub>O** - Água

**i.e.** - isto é

**II** - Identificação Genética Individual

**INMLCF** - Instituto Nacional de Medicina Legal e Ciências Forenses, I.P.

**INTCF-B** - *Instituto Nacional de Toxicologia y Ciencias Forenses de Barcelona* – Instituto Nacional de Toxicologia e Ciências Forenses de Barcelona

**IP** - Investigação Biológica de Parentescos

**IPAC** - Instituto Português de Acreditação

**IUPAC** - *International Union of Pure and Applied Chemistry* - União Internacional de Química Pura e Aplicada

**kV** - quilovolt

**LA** - Limiar Analítico

**LD** - Limiar de Desequilíbrio de Heterozigóticos

**LE** - Limiar Estocástico

**LS** - Limiar de *Stutter*

**min** - minutos

**mL** - mililitros

**mM** – milimolar

**ng** - nanograma

**OL** - *Off-Ladder*

**OS** - *Off-Scale*

**P** - Grupo Fosfato

**pb** - Pares de bases

**PCR** - *Polymerase Chain Reaction* - Reação em Cadeia da Polimerase

**pH** - Potencial de Hidrogénio

**POP** - *Performance Optimized Polymer* – Polímero de Performance Otimizada

**Prepfiler Express™** - *Applied Biosystems™ PrepFiler Express™ Forensic DNA Extraction Kit*

**Q** - *quencher*

**qPCR** - **PCR** quantitativa em tempo real (ou **RT PCR** - **Real-Time PCR**)

**Quantifiler™ Trio** - *Applied Biosystems™ Quantifiler™ Trio DNA Quantification*

**R** - fluoróforo repórter

**RFU** - *Relative Fluorescent Units* - Unidades Relativas de Fluorescência

**rpm** - rotações por minuto

**s** - segundos

**SeqStudio™** - *Applied Biosystems™ SeqStudio™ Genetic Analyzer*

**SGBF-C** - Serviço de **Genética** e **Biologia Forenses** da Delegação do **Centro** do INMLCF

**STRs** - *Short Tandem Repeats*

**SWGAM** - *Scientific Working Group on DNA Analysis Methods*

**T** - Timina

**Taq** - *Thermus aquaticus*

**TE** - Tris-EDTA (Cloridrato de Tris-(hidroximetil)-aminometano + Ácido etilenodiamino tetraacético)

**UV** - radiação Ultravioleta

**vs.** - *versus*

**Y-STRs** - *loci STR* do cromossoma **Y**

**3500** - *Applied Biosystems™ 3500 Genetic Analyzer*

**°C** - Graus Celsius

**μL** - Microlitro

**%** - Percentagem







O ADN humano localiza-se essencialmente no núcleo das células, sob a forma de cromossomas - **ADN nuclear** ou cromossómico (**ADNnu**); havendo, contudo, uma pequena porção localizada extra-nuclearmente, nas mitocôndrias – **ADN mitocondrial (ADNmt)** – organelos celulares responsáveis pela respiração celular e produção de energia. O **genoma humano** compreende todo esse conjunto de ADN existente em cada célula humana.

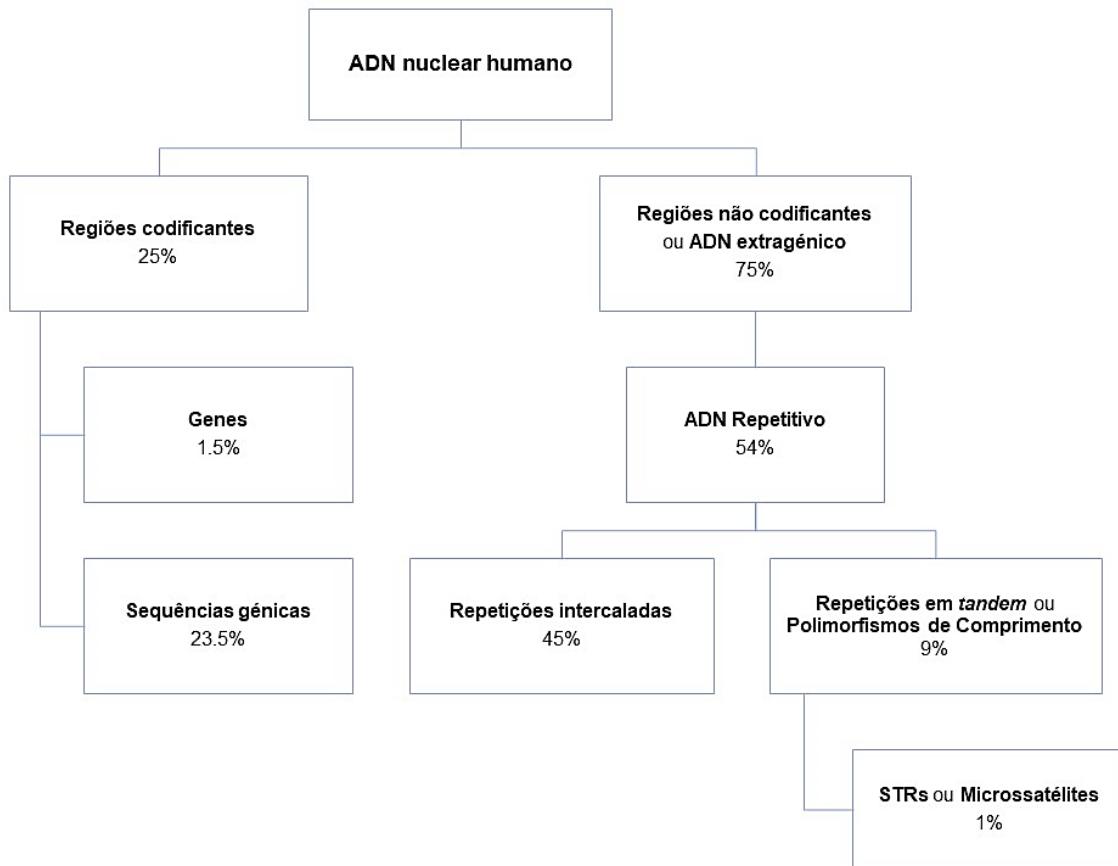
Relativamente ao ADNnu humano, o mesmo encontra-se organizado, nas células somáticas, em 22 pares de **cromossomas autossómicos** e 1 par de **cromossomas sexuais** (XX na mulher e XY no homem), totalizando um cariótipo de 46 cromossomas. Por conterem cromossomas organizados aos pares, as células somáticas são designadas como diplóides. Os cromossomas de cada par - **cromossomas homólogos** - apresentam a mesma forma, tamanho e estrutura genética, sendo que um é herdado do progenitor masculino e o outro do progenitor feminino (Butler, 2010; Corte-Real & Vieira, 2015).

O ADNnu humano é composto por duas categorias diferentes de ADN, no que concerne à sua estrutura e função (**Figura 2**): regiões codificantes e não codificantes.

As **regiões codificantes** constituem 25% do genoma, e nesta categoria incluem-se os **genes** – porções de ADN que codificam sequências de aminoácidos para a produção proteica – que representam apenas 1.5% do ADN genómico total; e **sequências génicas** responsáveis por regular a expressão dos genes, que representam os restantes 23.5%.

Já as **regiões não codificantes** (ou **ADN extragénico**) constituem 75% do genoma, e são em grande parte (54%) compostas por **ADN repetitivo**. Dessa subcategoria, 9% corresponde a **ADN repetido em tandem**, isto é, ADN cuja sequência é formada por um padrão de nucleótidos variável que se repete - **unidade (ou motivo) de repetição** - várias vezes, e cujas repetições estão dispostas diretamente adjacentes umas às outras. O número de cópias da unidade de repetição é muito variável de indivíduo para indivíduo, podendo originar sequências de diversos comprimentos, doravante denominadas como alelos. Por conseguirem assumir diversas formas alélicas em termos de comprimento, as sequências de ADN repetido em *tandem* são caracterizadas como sendo altamente polimórficas, e por isso são igualmente designadas de **Polimorfismos de Comprimento**. Os Polimorfismos de Comprimento podem ainda ser divididos em vários tipos, dos quais se destacam, pela sua já extensivamente descrita utilidade biomolecular, os **Short Tandem Repeats (STRs) ou Microssatélites** (1%), cujas unidades de repetição são, como o nome indica, muito curtas, possuindo entre 2 e 7 pares de bases (**pb**), e cujo número de repetições origina tipicamente alelos que variam entre 50 e 300 pb de comprimento. Os STRs que ocorrem com maior

frequência no genoma humano, são os tetranucleotídicos (ou tetraméricos), ou seja, os que possuem motivos de repetição de 4 pbs (**Figura 3**) (Butler, 2010; Goodwin *et al.*, 2011).



**Figura 2.** Organização do ADNnu e da sua frequência estimada no genoma humano.

Adaptado de (Goodwin *et al.*, 2011)

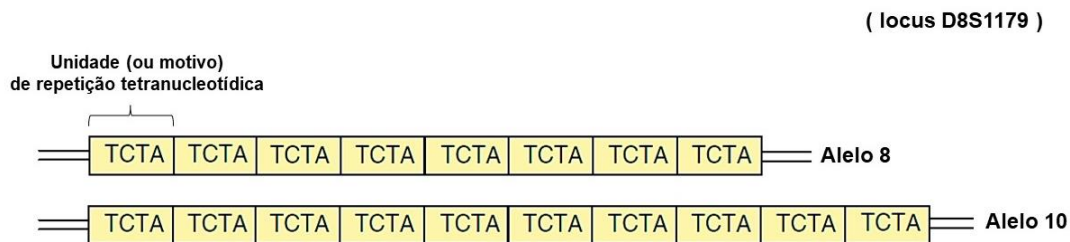
Estima-se que a maior parte do ADNnu humano (99.7%) seja igual entre indivíduos, e que apenas 0.3% difira significativamente, sendo essa pequena fração responsável por nos tornar indivíduos únicos (Butler, 2010). Surge aqui a importância do ADN repetido em *tandem*, que por apresentar elevado grau de polimorfismo alélico e grande variabilidade entre indivíduos de uma mesma população, constitui um aspeto individualizante, pois a probabilidade de dois indivíduos possuírem exatamente todos os mesmos alelos, é extremamente baixa. Evidencia-se, portanto, a notável utilidade dos Polimorfismos de Comprimento como **marcadores genéticos** em situações que envolvam a discriminação e identificação genética de indivíduos, como é o caso das perícias biológicas forenses.

No atual contexto forense, a nível internacional, os métodos de análise de ADN recorrem aos STRs como marcadores genéticos de eleição, dadas as vantagens que estes apresentam relativamente a outros polimorfismos genéticos, nomeadamente o seu maior

poder discriminatório entre indivíduos e também o seu tamanho mais reduzido e menor sujeição à degradação, o que facilita a sua posterior deteção e análise (Goodwin *et al.*, 2011).

A posição onde se encontra alocado um STR num par de cromossomas homólogos em particular é denominado **locus** (plural: **loci**) e as alternativas possíveis dentro de um mesmo **locus** designam-se, como anteriormente referido, **alelos**. Cada alelo é designado de acordo com o número de vezes em que a sua unidade de repetição ocorre num dado cromossoma (**Figura 3**). Este método abreviado e numérico de designar os alelos presentes numa amostra, facilita a comparação de resultados entre várias amostras.

Já um **genótipo** é a caracterização dos alelos presentes num determinado **locus**. Por exemplo, se um indivíduo possuir para um mesmo **locus**, um alelo com 8 repetições e o outro com 10 repetições, é caracterizado como tendo um genótipo de 8,10 para aquele **locus**. No caso referido, o alelo herdado de cada progenitor é diferente, e como tal, o genótipo é descrito como **heterozigótico** (**Figura 3**). Contudo, caso os alelos fossem iguais, seria descrito como **homozigótico** (Butler, 2010).



**Figura 3.** Estrutura de um STR. Estão representados dois alelos possíveis para um mesmo **locus** (D8S1179). Os alelos são designados de acordo com o número de cópias do motivo de repetição específico do **locus** (TCTA). No alelo representado superiormente, o motivo ocorre 8 vezes, adotando, portanto, a denominação de “Alelo 8” e no representado inferiormente, ocorre 10 vezes, sendo por isso chamado de “Alelo 10”. Um indivíduo portador destes alelos, apresenta um genótipo heterozigótico de 8,10 neste **locus**. Adaptado de (Goodwin *et al.*, 2011)

Quando se determinam os genótipos de um indivíduo para um conjunto de marcadores STRs é gerado um **perfil genético individual**. Essa análise simultânea de múltiplos **loci** STR é importante na determinação de perfis genéticos de sujeitos envolvidos em questões legais, pois aumenta o poder de discriminação entre indivíduos, o que reduz a possibilidade de correspondências genotípicas (*'matches'*) aleatórias e erróneas entre pessoas diferentes, visando assim uma maior fiabilidade nas conclusões obtidas e uma correta resolução de casos

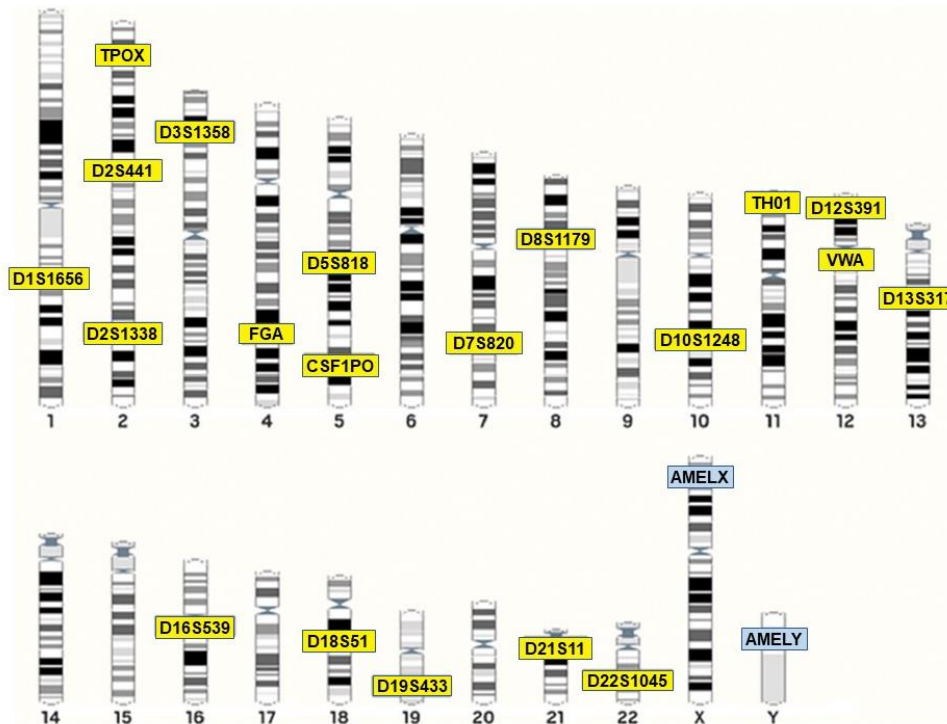
jurídicos. Ao processo de determinação de perfis genéticos é atribuído o nome de **genotipagem** (ou **tipagem**) de ADN (do inglês: *'DNA fingerprinting'*, *'DNA profiling* ou *'DNA typing'*) (Butler, 2010).

Nos laboratórios de Genética Forense atuais, os perfis são determinados por tipagem preferencial de marcadores STRs autossómicos. Contudo, é também possível recorrer à análise de outros polimorfismos, nomeadamente STRs específicos dos cromossomas sexuais (**X-STRs** e **Y-STRs**) e **ADNmt**, que apesar de menos discriminantes, são úteis quando há necessidade de informações genéticas adicionais às autossómicas ou quando a habitual tipagem de STRs falha em proporcionar resultados (Corte-Real & Vieira, 2015).

Visando a standardização de metodologias, para facilitar a comparação de resultados entre diferentes laboratórios e a construção de bases de dados genéticos - sistema com registos eletrónicos de perfis de ADN - várias instituições selecionaram um conjunto de marcadores STRs ("**core loci**") a serem analisados de forma padronizada nos laboratórios de genética forense (Pinheiro, 2010).

Assim, o *Federal Bureau of Investigation* (FBI) dos Estados Unidos da América, selecionou, para que integrasse a sua base de dados nacional **CODIS** (*Combined DNA Index System*), o seguinte conjunto de 20 *loci* de STRs autossómicos: CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433 e D22S1045, juntamente com o gene homólogo da Amelogenina (**AMEL**) - um *locus* de determinação sexual que existe em ambos os cromossomas sexuais, X (AMELX) e Y (AMELY) – como mostra a **Figura 4** (Hares, 2012, 2015; Udogadi *et al.*, 2020).

A nível europeu, foi também estabelecido, pela Rede Europeia de Institutos de Ciências Forenses (**ENFSI**), um painel de STRs, designado por *European Standard Set* (**ESS**), como conjunto de *loci* adequado para a definição de perfis genéticos a serem inseridos em todas as bases de dados nacionais da Europa. O ESS conta atualmente com os seguintes 12 *loci* autossómicos: FGA, TH01, vWA, D3S1358, D8S1179, D18S51, D21S11, D1S1656, D2S441, D10S1248, D12S391 e D22S1045, que são comuns aos do sistema CODIS (**Figura 4**) (Butler, 2011).



**Figura 4.** *Core loci* que compõem o atual sistema CODIS, e sua distribuição cromossômica. A amarelo estão representados os *loci* autossômicos e a azul, a Amelogenina (AMEL), *locus* existente nos cromossomas sexuais X (AMELX) e Y (AMELY) e utilizado como marcador de determinação de género sexual. Os 12 *loci* autossômicos FGA, TH01, vWA, D3S1358, D8S1179, D18S51, D21S11, D1S1656, D2S441, D10S1248, D12S391 e D22S1045, são partilhados pelo ESS. Adaptado de (Butler & Li, 2014; Udogadi *et al.*, 2020)

Os *core loci* do CODIS e do ESS tornaram-se amplamente aceites a nível internacional, sendo adotados pelos laboratórios forenses como critério de análise mínimo para determinação de perfis genéticos e sua inserção nas bases de dados.

A inserção destes conjuntos de *loci* STR em bases de dados forenses nacionais facilita a partilha e comparação de perfis genéticos entre países, promovendo assim, a cooperação judiciária internacional (Machado & Moniz, 2014).

## 1.2. INTRODUÇÃO AO SGBF

---

Ao Serviço de Genética e Biologia Forenses (SGBF) do Instituto de Medicina Legal e Ciências Forenses, I.P. (INMLCF) compete assegurar, a nível nacional, a realização de perícias e exames de identificação genética de interesse público por solicitação de tribunais e/ou autoridades policiais, podendo também ser realizadas a pedido de particulares.

A atividade pericial do SGBF envolve a **investigação biológica de parentescos (IP)**, maioritariamente de paternidades; a **identificação genética individual (II)** de cadáveres e de restos cadavéricos ou esqueléticos no âmbito de investigações de pessoas desaparecidas ou identificação de vítimas de desastres em massa; a identificação genética de vestígios biológicos colhidos no âmbito de investigações criminais (**Criminalística Biológica, CR**); e também a identificação de perfis genéticos para inserção na **Base de Dados nacional de Perfis de ADN** (Portaria n.º 19/2013, de 21 de Janeiro, 2013; INMLCF, 2020).

Qualquer que seja a casuística forense em questão, o SGBF (e qualquer outro laboratório de Genética Forense) depende da determinação de perfis genéticos na sua rotina de trabalho, recorrendo aos resultados da sua análise como meio de prova científica em tribunal para auxiliar a resolução jurídica de processos criminais e/ou civis que envolvam amostras biológicas.

### 1.3. TIPO DE AMOSTRAS EM CONTEXTO FORENSE

---

Para obter um perfil genético passível de ser utilizado para fins de identificação humana forense, é necessária a realização de uma série de etapas que vão desde a recolha de uma amostra biológica até à análise e interpretação do perfil gerado.

Em Genética Forense as amostras biológicas podem classificar-se quanto à sua proveniência e intuito com que são analisadas, sendo categorizadas, pelo SGBF-C, em: amostras de referência, amostras problema e amostras seguras.

**Amostras de referência** são amostras de origem conhecida, que servem como o nome indica, de referência a um indivíduo específico, para determinação do seu perfil genético e cuja tipagem ou consulta na base de dados, é efetuada para fins de comparação com outros perfis (Butler, 2011; INMLCF, 2020; PE-SGBF-C-002, 2018). Sem este tipo de amostra não é possível estabelecer relações genéticas, nem realizar qualquer identificação (Corte-Real & Vieira, 2015). As amostras de referência são, por norma, de saliva e/ou sangue e podem ser, ou não, consideradas como seguras (ver definição abaixo) consoante a qualidade e quantidade do seu conteúdo genético seja elevada ou não (PE-SGBF-C-002, 2018).

**Amostras problema** são amostras, sob investigação, cuja identificação se desconhece e se pretende estabelecer através de comparação com referências. São amostras que podem ser recolhidas de objetos, indivíduos vivos e/ou cadáveres não identificados (INMLCF, 2020; PE-SGBF-C-002, 2018), como sangue, saliva, cabelos, fluidos genitais, tecidos moles, ossos, dentes, unhas, entre outros (Corte-Real & Vieira, 2015). A quantidade e qualidade de ADN nelas presente, aquando da sua recolha ou receção, é uma incógnita. Por essa razão, assume-se sempre que esse tipo de amostras poderá apresentar quantidades muito reduzidas de ADN, encontrar-se em diferentes estados de degradação e contaminação, e incluir ADN de mais do que um indivíduo (misturas), podendo gerar perfis genéticos incompletos e insuficientemente esclarecedores (daí o nome “problema”) (PE-SGBF-C-002, 2018). Esta situação aplica-se principalmente quando as amostras provêm de locais de crime ou de vítimas de crime (amostras de vestígios ou vestigiais) (Pinheiro, 2010). As amostras problema, por serem muitas vezes residuais, exigem um maior cuidado na sua manipulação e processamento laboratorial, por forma a evitar contaminações e desperdícios desnecessários de amostra.



**Amostras seguras** são amostras de origem conhecida, colhidas diretamente a indivíduos vivos ou cadáveres frescos - cadáveres recentes sem que qualquer processo de decomposição tenha sido iniciado – por norma, de saliva e/ou sangue, para determinação do seu perfil genético individual. Por serem recolhidas e armazenadas em condições ideais, contrariamente às amostras problema, estas amostras apresentam geralmente uma quantidade e qualidade de ADN suficientes para obter um perfil genético completo. O nome “seguras” provém dessa segurança que é à partida assegurada. Pelo facto de serem provenientes de um único indivíduo, todas as amostras seguras são consideradas como amostras de referência, pois servem de referência ao indivíduo de quem foram recolhidas (PE-SGBF-C-002, 2018).

No SGBF-C, as amostras seguras e as amostras problema são processadas separadamente, em tempo e espaço, por forma a evitar contaminações no percurso laboratorial das mesmas, e com recurso a diferentes metodologias consoante a sua melhor adaptação ao tipo de amostra em questão.

## 1.4. PERCURSO LABORATORIAL DAS AMOSTRAS PARA DETERMINAÇÃO DE PERFIS GENÉTICOS

---

Após a colheita e preservação de uma amostra biológica, a sua análise forense percorre a seguinte ordem de etapas laboratoriais: extração de ADN; quantificação do ADN extraído; amplificação dos fragmentos de interesse (STRs); separação e deteção dos fragmentos amplificados. Deste processo resulta um eletroferograma contendo o perfil genético da amostra, que é analisado e interpretado cuidadosamente para que a partir dele, seja elaborado um relatório pericial esclarecedor à entidade requisitante (Butler, 2011).

### 1.4.1. EXTRAÇÃO DE ADN

Uma amostra biológica, contém sempre várias substâncias além de ADN. Para que seja possível analisar o ADN, é necessário isolá-lo num processo designado de extração. A extração contempla a **lise** das células de forma a libertar o ADN nelas contido, seguida da separação do ADN do restante material celular, e por fim a **purificação** da solução através de lavagens que removem o material celular não relevante e eventuais inibidores da PCR (uma técnica mais à frente aplicada) (Butler, 2011).

No SGBF-C, são aplicadas diferentes metodologias de extração consoante o tipo de amostra a analisar.

Amostras seguras de saliva e sangue são submetidas a uma técnica de extração rápida com a solução-tampão **Prep-n-Go™** (*Applied Biosystems™*), seguindo imediatamente para a fase de amplificação (Gomes *et al.*, 2017; IT-SGBF-C-017, 2021). Amostras seguras de sangue podem, contudo, sofrer amplificação direta, não necessitando de extração. Pelo facto de possuírem quantidade e qualidade satisfatórias de ADN, não se justifica proceder à quantificação de amostras seguras (PE-SGBF-C-001, 2020).

Já as amostras problema, em razão da quantidade genética vestigial e eventuais inibidores nelas contidas, requerem uma extração mais eficaz e lavagens de purificação intensas, na tentativa de obter uma solução final de ADN com a máxima quantidade e pureza possíveis, para que seja exequível a sua posterior amplificação e seja gerado um perfil genético minimamente interpretável (Balsa *et al.*, 2011; Thermo Fisher Scientific, 2021). Por isso, o SGBF-C recorre ao *kit* de extração: **PrepFiler Express™ Forensic DNA Extraction Kit** (AB™), no caso de amostras comuns como sangue, saliva e outros fluidos biológicos; e

ao **PrepFiler Express BTA™ Forensic DNA Extraction Kit (AB™)** no caso de amostras mais desafiantes como ossos, dentes e amostras em substratos adesivos como pontas de cigarro, fita-cola e pastilhas elásticas (**B** = *Bone*; **T** = *Tooth*; **A** = *Adhesive*) (Applied Biosystems, 2017; IT-SGBF-C-014, 2022). A metodologia de extração com estes dois *kits* é realizada maioritariamente de forma automática no equipamento robotizado **AutoMate Express™ Forensic DNA Extraction System (AB™)** (Applied Biosystems, 2019e). Seguidamente à extração, as amostras problema prosseguem obrigatoriamente para uma fase de quantificação, onde os resultados determinam se as mesmas são favoráveis ao prosseguimento para amplificação ou não. Amostras seguras também são extraídas com o *kit PrepFiler Express™* quando, por alguma razão, é de interesse quantificá-las.

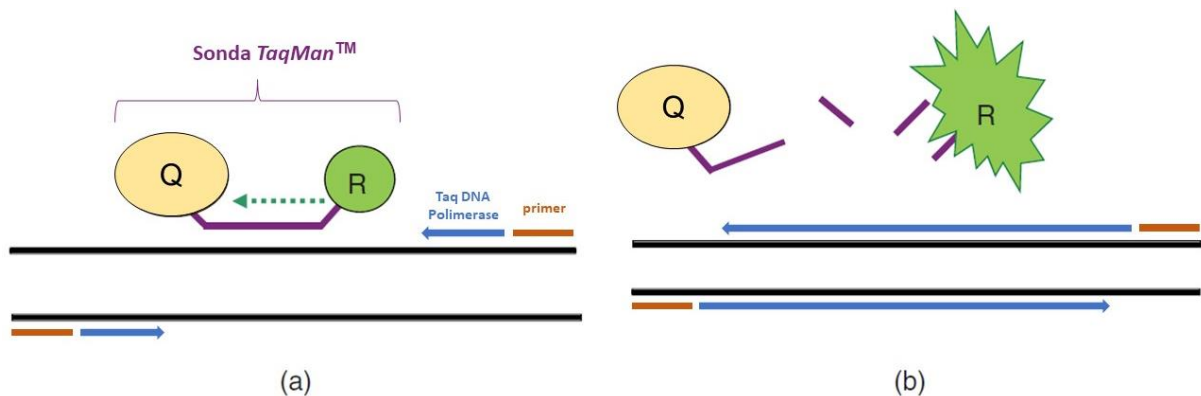
#### 1.4.2. QUANTIFICAÇÃO DO ADN EXTRAÍDO

O ADN recuperado da extração de amostras problema corresponde, muitas vezes, não só ao ADN humano de interesse, mas também a eventual ADN ‘contaminante’ de fontes externas como bactérias, fungos, plantas, animais, etc. (Butler, 2011). Além disso, o ADN de interesse recuperado encontra-se frequentemente em quantidades reduzidas e/ou com algum nível de degradação associado, o que pode dificultar a obtenção de perfis genéticos. Para que possam gerar perfis completos e interpretáveis, é tipicamente recomendado que as amostras problema prossigam para a fase de amplificação, contendo idealmente entre **0.5–2.5 ng** de ADN humano total (Gill *et al.*, 2020).

Por conseguinte, previamente à amplificação por PCR de amostras problema, é importante garantir a existência de ADN especificamente humano e determinar a quantidade total desse ADN amplificável. Esta avaliação pré-PCR é realizada através de uma fase laboratorial denominada quantificação.

A técnica de quantificação mais amplamente utilizada é a PCR quantitativa em tempo real (**qPCR**) e sua abordagem mais comum é baseada na utilização de **sondas** nucleotídicas **TaqMan™** numa reação semelhante à da PCR, explicada no subtema abaixo (**1.4.3 Amplificação de STRs por PCR**). Resumidamente, estas sondas são marcadas com um fluoróforo (ou fluorocromo) repórter (**R**) – molécula que emite fluorescência – e um *quencher* (**Q**) – molécula que absorve a fluorescência emitida pelo fluoróforo – e, na presença de ADN amplificável, hibridizam com determinadas sequências de interesse do ADN humano delimitadas por *primers*. Inicialmente, a proximidade física entre R e Q leva à supressão da fluorescência do R (**Figura 5 (a)**), devido à transferência de energia entre ambos. Ao longo

da reação de qPCR, a enzima *Taq* ADN Polimerase vai amplificando o ADN e, atingindo o local da sonda, hidrolisa-a por ação da sua atividade exonucleotídica. Essa clivagem da sonda provoca a separação espacial entre R e Q, resultando numa emissão detetável de fluorescência pelo R (**Figura 5 (b)**), que é medida em tempo real (ao fim de cada ciclo de PCR) por um sistema ótico, e que é proporcional à quantidade de produto amplificado (Goodwin *et al.*, 2011).



**Figura 5.** Representação esquemática da tecnologia de qPCR baseada na utilização de sondas *TaqMan™*. Adaptado de (Goodwin *et al.*, 2011)

A intensidade do sinal de fluorescência detetada durante a qPCR é por fim traduzida numa estimativa da concentração (**ng/μl**) absoluta de ADN humano total presente na amostra em análise, através da comparação com curvas de calibração. Essas curvas de referência são previamente elaboradas com recurso a amostras padrão que contêm as sequências a amplificar durante a qPCR em concentrações conhecidas e, correlacionam, através das suas fórmulas matemáticas, os valores aproximados das concentrações de ADN amostrais às respetivas intensidades de fluorescência (Butler, 2011).

Os resultados de quantificação obtidos permitem concluir se existe material de interesse suficiente no extrato de ADN para se proceder à análise de STRs e orientam o perito acerca do volume (μl) do extrato com que deve trabalhar na fase subsequente de amplificação para que a sua quantidade de ADN esteja na janela ótima de funcionamento da PCR (0.5 – 2.5 ng).

No SGBF-C, a etapa de quantificação é atualmente realizada com recurso ao *kit* comercial **Quantifiler™ Trio DNA Quantification** (AB™), que permite quantificar simultaneamente a concentração de ADN humano total e a de ADN humano exclusivamente masculino amplificáveis numa amostra biológica, utilizando uma tecnologia de quantificação

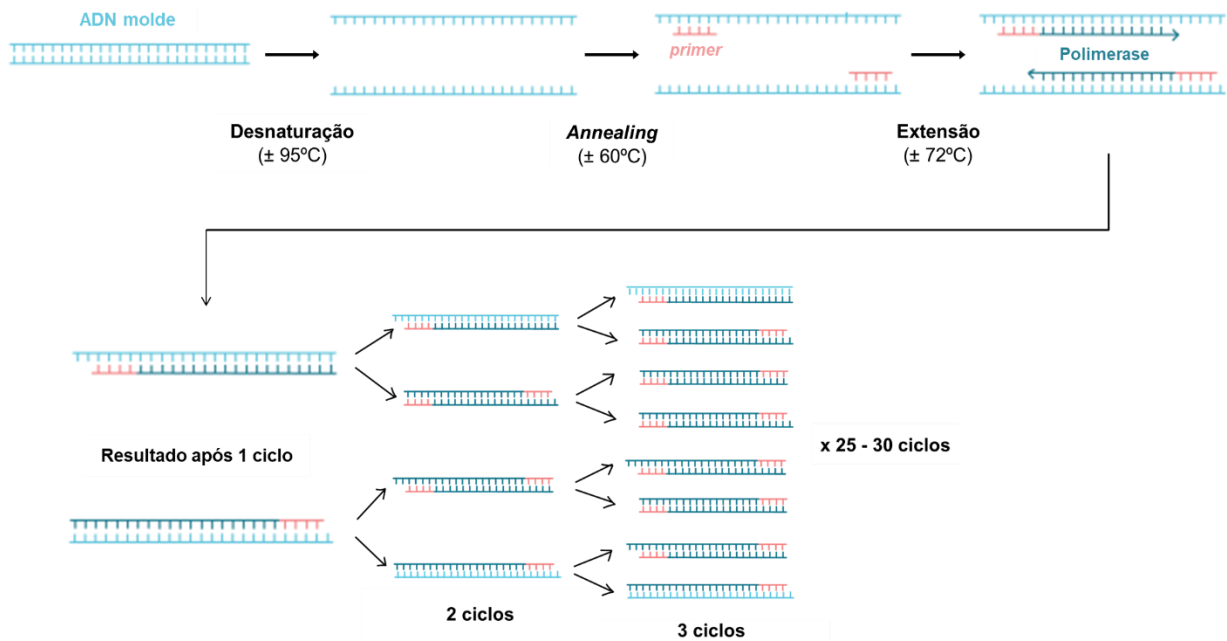
por qPCR com sondas *TaqMan™* no equipamento **7500 Real-Time PCR Instrument (AB™)** em associação com o software de análise dos resultados *HID Real-Time PCR Analysis Software v.1.2. (AB™)* (Applied Biosystems, 2018; IT-SGBF-C-015, 2020).

### 1.4.3. AMPLIFICAÇÃO DE STRs POR PCR

Para que na fase seguinte, seja possível a análise dos STRs de interesse presentes numa amostra biológica, é necessário multiplicá-los seletivamente de entre todo o ADN existente, até que atinjam uma quantidade suficientemente detetável em sequenciadores automáticos. Para isso, qualquer amostra forense tem de ser submetida a uma reação de PCR.

A **Reação em Cadeia da Polimerase (PCR)**, do inglês *Polymerase Chain Reaction*) é um processo enzimático através do qual uma sequência específica de ADN (ADN alvo) é replicada várias vezes, originando milhões de cópias dessa mesma região, num curto espaço de tempo, por aplicação repetida de ciclos térmicos. A técnica de PCR envolve a presença de: **ADN alvo** de uma amostra biológica, que será usado como molde para a sua amplificação; iniciadores de cadeia (**primers**) específicos que delimitam a região de interesse a ser copiada; nucleótidos livres - desoxinucleótidos trifosfatados (**dNTPs**) - necessários para a produção das novas cadeias (dATP, dCTP, dGTP, dTTP); e uma **ADN Polimerase** termoestável - geralmente a *Taq* Polimerase (extraída de bactérias *Thermus aquaticus*) – que adiciona os dNTPs orientando-se pela localização dos *primers*.

A amplificação por PCR compreende a aplicação de 25 - 30 ciclos de aquecimento e arrefecimento, produzidos num termociclador, e cada ciclo atravessa três fases distintas, conforme esquematizado na **Figura 6**: 1) a fase de **Desnaturação**, onde a temperatura da reação é aumentada até  $\pm 95^{\circ}\text{C}$ , quebrando as ligações de hidrogénio entre os pares de bases do ADN em cadeia dupla, causando a sua separação em cadeias simples; 2) a fase de **Annealing ou Hibridização** em que a temperatura é reduzida até  $\pm 60^{\circ}\text{C}$  promovendo a ligação dos *primers* às sequências que lhes são complementares no ADN molde de cadeia simples, delimitando a região a ser amplificada; 3) e a fase de **Extensão**, em que a temperatura é novamente aumentada, desta vez até atingir a temperatura ótima de atuação da Polimerase (que no caso da *Taq* Polimerase é de  $\pm 72^{\circ}\text{C}$ ), permitindo a elongação das novas cadeias por incorporação dos dNTPs complementares à cadeia de interesse (Butler, 2011).



**Figura 6.** Representação esquemática da amplificação de uma sequência de interesse por PCR.

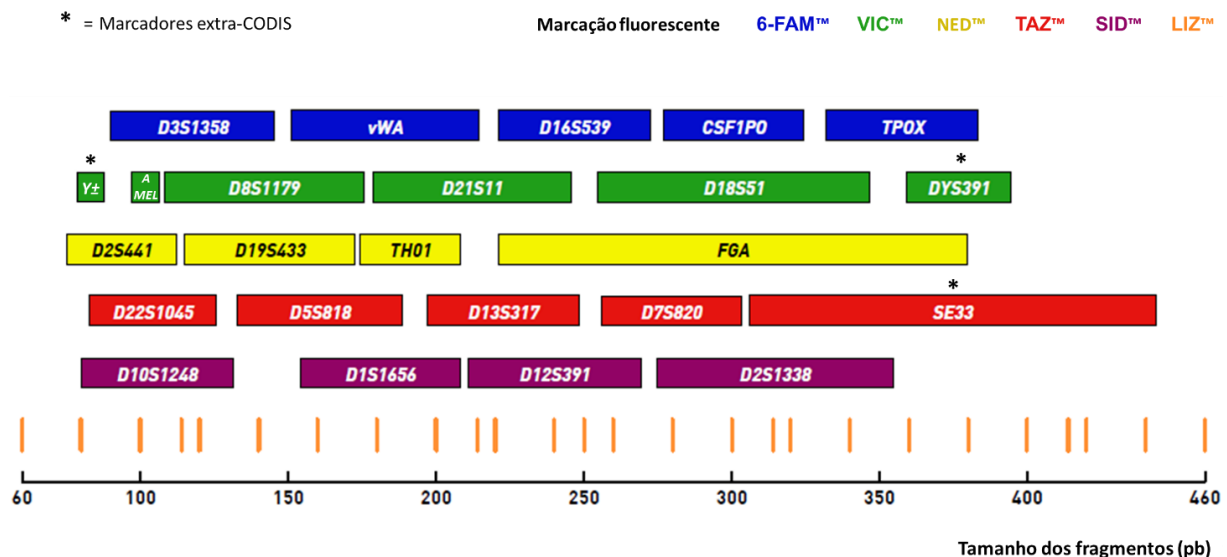
Adaptado de (Khan Academy, 2017)

A PCR pode amplificar, não só uma sequência de interesse, conforme descrito, mas também várias regiões alvo diferentes em simultâneo, designando-se nesse caso por **PCR em multiplex**, método este amplamente utilizado em Genética Forense pelo seu interesse em analisar múltiplos *loci* STR para determinação de perfis.

Atualmente existem disponíveis vários *kits* comerciais para amplificação de STRs em *multiplex*, de forma a abranger os *core loci* exigidos para tipagem pelas entidades forenses, sendo as principais empresas que os desenvolvem, a **Applied Biosystems** da corporação *Thermo Fisher Scientific (AB™)* e a **Promega Corporation**. Esses *kits* contêm já todos os reagentes necessários para a reação de PCR ocorrer, o que agiliza imenso o trabalho dos peritos, incluindo um conjunto de *primers* que além de serem específicos para cada marcador STR de interesse, também são marcados com fluorocromos de diferentes cores permitindo assim, a distinção entre *loci* de tamanhos (pb) semelhantes, na fase subsequente de separação e deteção do produto amplificado (secção 1.4.4.). Desta forma, o eletroferograma que se obtém no final, surge organizado por painéis de cor, de acordo com o sistema de marcação fluorescente utilizado (Butler, 2011).

No SGBF-C, todas as amostras são atualmente amplificadas com recurso aos *kits multiplex* de STRs maioritariamente autossómicos: **GlobalFiler™ PCR Amplification kit** (AB™) e/ou **PowerPlex Fusion® 6C System** (*Promega*). Na eventual necessidade de recorrer à análise de *loci* suplementares podem ser usados outros *kits* como o **Yfiler™ Plus PCR Amplification Kit** (AB™), que permite a amplificação e subsequente tipagem de Y-STRs (MQ-SGBF-C, 2022).

No presente trabalho, foi utilizado o *kit GlobalFiler™* (**Figura 7**) que permite a co-amplificação e posterior deteção fluorescente de 24 marcadores, nomeadamente todos os *core loci* pertencentes ao CODIS e adicionalmente o *locus* STR autossómico SE33, o DYS391 (Y-STR) e o *locus* de inserção/deleção específico do cromossoma Y (Y-InDel). Este sistema *multiplex* utiliza uma marcação de *primers* por 5 fluorocromos: 6-FAM™ que emite cor azul, VIC™ (verde), NED™ (amarelo), TAZ™ (vermelho) e SID™ (roxo) (Applied Biosystems, 2019f).



**Figura 7.** Marcadores que o *kit GlobalFiler™* permite amplificar em *multiplex* e seus respetivos intervalos de tamanhos alélicos e marcações fluorescentes. Adaptado de (Applied Biosystems, 2019b; Butler & Gittelson, 2015)

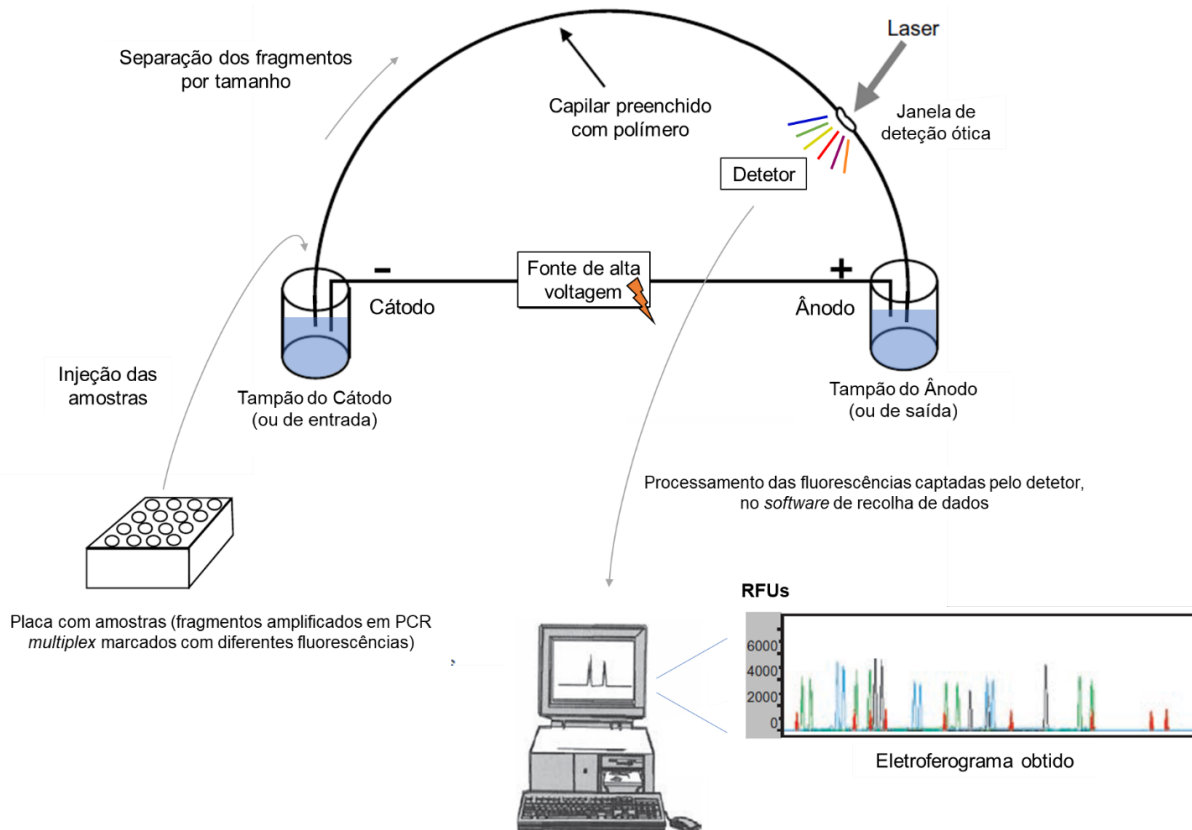
#### 1.4.4. SEPARAÇÃO E DETECÇÃO DO PRODUTO AMPLIFICADO

Após a amplificação dos fragmentos de interesse, por PCR, estes têm de ser separados e detetados. Este procedimento é efetuado em **sequenciadores automáticos**.

Os fragmentos amplificados começam por ser separados através de **eletroforese capilar (EC)** de acordo com o seu tamanho e são depois detetados por fluorescência. Na EC, as amostras são primeiramente injetadas em **capilares** revestidos internamente por um **polímero** fluído desnaturante, que funciona como matriz de separação, oferecendo resistência à passagem das cadeias simples de ADN. Depois, as extremidades dos capilares são imersas em soluções tampão associadas a eléctrodos, sendo que a extremidade com as amostras fica no tampão com o eléctrodo negativo (**cátodo**) e a outra extremidade, com o positivo (**ânodo**). Então, por aplicação de um **campo elétrico** de alta tensão, as cadeias de ADN – que são carregadas negativamente pelos seus grupos fosfato - migram através do polímero, do polo negativo (cátodo) para o polo positivo (ânodo). Durante a migração, os fragmentos mais curtos (com menos pb e, portanto, com menor peso molecular) deslocam-se mais rapidamente através dos poros do polímero relativamente aos fragmentos maiores, sendo por isso, detetados primeiro na etapa seguinte. Antes de alcançarem o ânodo, os fragmentos de ADN passam por uma zona em que o revestimento do capilar deixa de ser opaco (**janela de deteção ótica**) onde são atingidos por um **laser de árgon**. Isto faz com que os fluorocromos dos seus *primers* sejam excitados e emitam fluorescência nas diferentes cores que lhes correspondem. As fluorescências emitidas são captadas por um dispositivo de deteção ótico e convertidas em sinais elétricos, que são proporcionais à quantidade de luz (intensidade de fluorescência) detetada, i.e., proporcionais ao número de fluorocromos (e, portanto, à quantidade de fragmentos amplificados) que atravessaram o laser em cada momento. Esses sinais de intensidade de fluorescência são traduzidos em tempo real, por um *software* de recolha de dados associado ao sequenciador automático, em **unidades relativas de fluorescência (RFUs)**, dando origem aos picos que vemos num **eletroferograma (EF)** (**Figura 8**) (Butler, 2011; Kim *et al.*, 2018).

Um eletroferograma consiste então num registo gráfico que relaciona a intensidade de fluorescência detetada em cada fragmento, em RFU (no eixo do y) com o tempo da sua deteção pelo laser (eixo do x). O EF apresenta assim, os picos por ordem de tamanho dos fragmentos, sendo que os primeiros picos detetados (à esquerda) correspondem aos marcadores mais pequenos e à direita, aos maiores (**Figura 9**) (Butler, 2015).





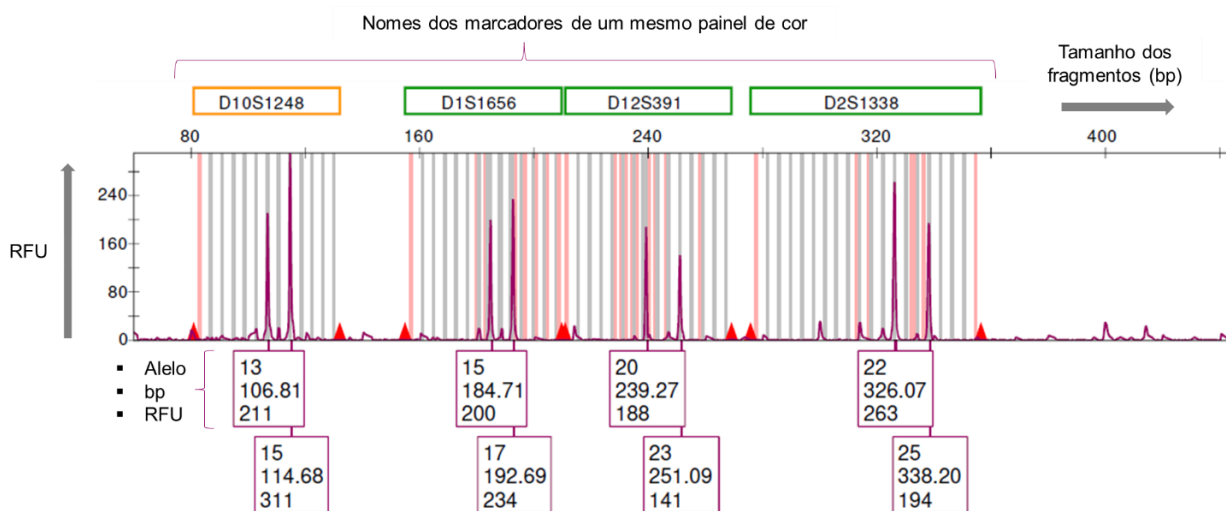
**Figura 8.** Ilustração esquemática da separação de fragmentos por eletroforese capilar, e sua detecção fluorescente, num sequenciador automático. Adaptado de (Butler, 2011)

Os resultados obtidos sob a forma de eletroferograma são depois exportados para um *software* de análise – geralmente o **GeneMapper™ ID-X** (AB™) - que automaticamente: determina o tamanho (pb) de cada fragmento detetado por comparação com um *standard* interno; organiza os fragmentos por pb e divide-os em painéis de cor consoante a cor da fluorescência emitida por cada um; e realiza a atribuição alélica aos fragmentos das amostras por comparação com um *ladder* alélico (**Figura 9**). Esta análise facilita a leitura do eletroferograma e obtém por fim, o perfil genético de cada amostra (Butler, 2015).

Um **standard interno** consiste numa mistura artificial de fragmentos de ADN de tamanho conhecido (em pb) marcados com uma fluorescência diferente da dos restantes fragmentos a analisar. Esse “padrão interno de tamanhos”, sendo pipetado e corrido juntamente com cada amostra na placa, funciona então como uma régua, permitindo, por comparação, dimensionar os fragmentos da amostra. Um *standard* interno muito utilizado no SGBF-C é o **GeneScan™ 600 LIZ™ Size Standard v2.0** (AB™), que contém um conjunto de 36 fragmentos em cadeia simples entre os 20 e os 600 pb, todos eles marcados com o fluoróforo LIZ™ (laranja). Este *standard* permite assim dimensionar uma grande diversidade

de marcadores, nos quais se encontram abrangidos, pelo intervalo de tamanhos, os do *kit GlobalFiler™* (60 - 460 pb) (**Figura 7**) (Applied Biosystems, 2022).

Já o **ladder alélico** corresponde a uma mistura artificial de fragmentos que contém os alelos mais frequentes, na população humana, de cada marcador a analisar. O *ladder* é corrido em poços à parte das amostras, e por comparação com cada amostra, permite designar os alelos presentes em cada marcador das amostras, realizando assim, a genotipagem da amostra (Butler, 2011). Para a designação alélica de fragmentos amplificados com o *kit GlobalFiler™*, utiliza-se o **GlobalFiler™ Allelic Ladder** que vem incluído no próprio *kit* (Applied Biosystems, 2019f).



**Figura 9.** Painel de cor de um eletroferograma após análise por *GeneMapper™ ID-X* (AB™). Os picos detetados em RFUs estão devidamente identificados, com o alelo a que correspondem e o respetivo tamanho de fragmento determinado (bp).

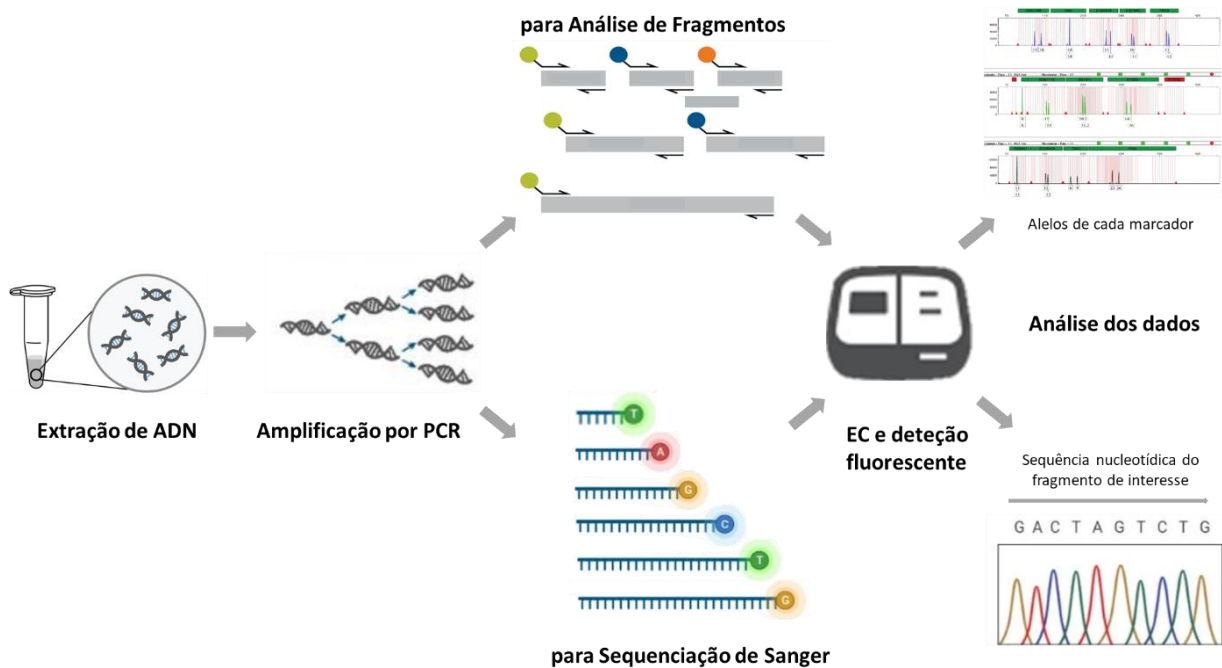
Antes da análise dos EFs pelo *GeneMapper™*, cada laboratório deve definir limiares que permitam a correta genotipagem pelo *software*, e auxiliem o perito na interpretação dos perfis obtidos e nas conclusões deles retirados. Tais limiares serão abordados na secção seguinte (**1.5. Análise e interpretação dos perfis genéticos obtidos**).

É de notar que os sequenciadores automáticos permitem, não só efetuar análise de fragmentos conforme descrito, mas também sequenciar fragmentos por sequenciação de Sanger, ambos através da tecnologia de EC com deteção automática de fluorescência.

A sequenciação consiste no processo de determinar a sequência exata de nucleótidos em fragmentos de interesse. Para isso, as amostras atravessam um percurso laboratorial relativamente semelhante ao relatado até agora, mas adicionalmente, segundo a metodologia de Sanger, na fase de amplificação são acrescentados didesoxinucleótidos trifosfatados (**ddNTPs**) que se comportam como terminadores de cadeia, pois ao serem incorporados pela ADN polimerase impedem a adição de mais nucleótidos. Esses ddNTPs são marcados com fluorescências - ao invés dos *primers* - de 4 cores diferentes, uma para cada nucleótido (ddATP, ddCTP, ddGTP, ddTTP), permitindo distinguir, na fase de detecção fluorescente, a base nucleotídica que foi incorporada por último em cada fragmento. Assim, fragmentos de diferentes comprimentos, diferindo entre si por uma base, são separados por tamanho em EC, num polímero de maior resolução, e ao serem atingidos pelo laser, são gerados sinais, que são convertidos pelo *software* de recolha de dados, num eletroferograma. Um *software* de análise depois associa automaticamente os picos à ordem de bases nucleotídicas que passaram pelo laser, consoante o comprimento de onda (cor) detetado, determinando assim, a sequência dos fragmentos em análise (**Figura 10**) e analisa variações nos polimorfismos de sequência para efeitos de identificação e discriminação de indivíduos (Applied Biosystems, 2016; Butler, 2011).

Os laboratórios de genética forense, não recorrem habitualmente à análise de sequência, contrariamente à análise de fragmentos. Contudo, quando uma amostra problema contém pouco ADNnu ou o mesmo encontra-se extremamente degradado, impossibilitando a obtenção de resultados satisfatórios através da normal tipagem de STRs, justifica-se tipar ADNmt por sequenciação como alternativa. Apesar dos seus polimorfismos não serem tão discriminatórios na identificação humana, o ADNmt existe em grande número de cópias (cerca de 1000 a 10000) por célula, e a sua conformação circular torna-o menos suscetível à degradação, permitindo facilmente a obtenção de resultados (Amorim *et al.*, 2019; Pajnič, 2020).

No SGBF-C, a sequenciação de ADNmt é realizada exclusivamente para obtenção de informação genética a partir de ossos, dentes ou restos cadavéricos em elevado estado de degradação cuja tipagem de STRs não gerou resultados e/ou de cabelos/pelos sem raiz, recolhidos em cenas de crime das quais não se tem nenhum outro tipo de amostra disponível para analisar pela via normal (PO-SGBF-N-010, 2020). Pelo facto do ADNmt ser herdado exclusivamente pela via materna, muito excecionalmente, também pode ser importante a sua análise para a definição de linhagens maternas em processos de identificação de parentescos biológicos e em estudos de ancestralidades (Corte-Real & Vieira, 2015; Pinheiro, 2010).



**Figura 10.** Percurso laboratorial para análise de fragmentos vs. sequenciação de Sanger. Na análise de fragmentos, são usados *primers* fluorescentes, enquanto na sequenciação de Sanger a fluorescência é usada em ddNTPs. O resultado da análise de fragmentos culmina na atribuição alélica dos marcadores em análise, enquanto na sequenciação obtém-se a ordem de bases nucleotídicas dos fragmentos de interesse. Adaptado de (Thermo Fisher Scientific, 2020)

Atualmente, o SGBF-C tem em utilização um único sequenciador automático, o *3500 Genetic Analyzer (AB™)*, tendo recentemente adquirido um outro, o *SeqStudio™ Genetic Analyzer (AB™)*, ambos apresentados em maior detalhe na secção **1.6. Sequenciadores Automáticos – Seqstudio™ Genetic Analyzer (AB™)**.

## 1.5. ANÁLISE E INTERPRETAÇÃO DOS PERFIS GENÉTICOS OBTIDOS

---

Concluídas todas as etapas que levam à obtenção de um perfil genético de STRs, é necessário proceder à sua análise e interpretação, para que a partir dele seja elaborado um relatório pericial esclarecedor à entidade requisitante (Corte-Real & Vieira, 2015). Contudo, essa análise constitui um desafio para os peritos, pois verificam-se, frequentemente nos eletroferogramas, a ocorrência de certos fenómenos – **artefactos** - que dificultam a correta genotipagem das amostras, principalmente nas amostras problema. Esses fenómenos têm de ser identificados e discriminados do verdadeiro perfil presente, com a devida cautela, pois sobre as conclusões finais da interpretação dos perfis recaem decisões e/ou sanções jurídicas nos indivíduos envolvidos, das quais não se pretende uma incorreta identificação individual ou de parentesco, ilibação do agressor ou condenação de um inocente (Butler, 2010, 2015).

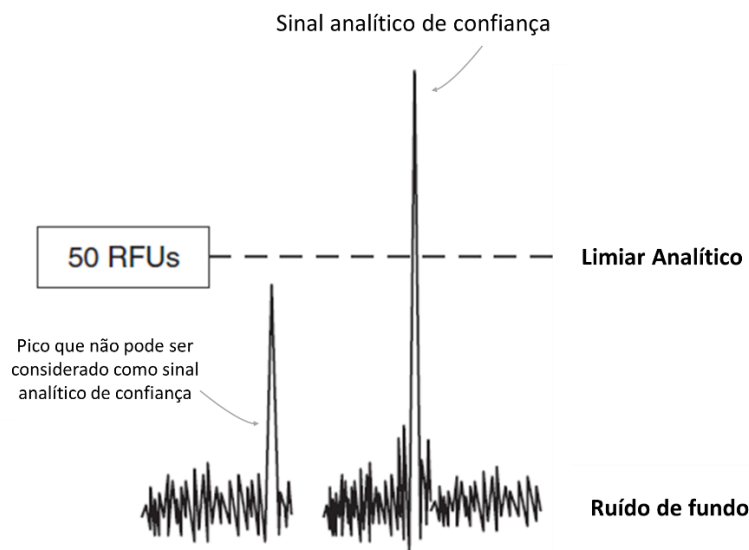
Os principais artefactos a ter atenção durante a análise de perfis são apresentados abaixo, bem como as situações específicas em que eles ocorrem e as possíveis soluções para contorná-los.

### 1.5.1. RUÍDO DE FUNDO

Após qualquer corrida de EC, os EFs apresentam picos inespecíficos que não correspondem a material amplificado, que são produzidos pelo próprio funcionamento elétrico do instrumento de EC e pelos reagentes utilizados. Esse artefacto instrumental denomina-se **ruído de fundo**. Para não gerar confusão na atribuição alélica ou na interpretação do perfil, é necessário estabelecer um valor limite que permita distinguir entre sinal correspondente a ruído e sinal efetivamente analítico. Para isso, os laboratórios de genética forense, estabelecem um Limiar de Detecção Mínimo de picos (ou **Limiar Analítico - LA**), um valor fixo em RFU que, nas condições do laboratório, garante com confiança suficiente que qualquer pico com altura acima desse valor, é de facto um produto da amplificação por PCR e que qualquer pico detetado abaixo do mesmo, não pode ser distinguido de ruído de fundo com segurança, não sendo, portanto, passível de ser interpretado no EF (**Figura 11**). É de notar que um pico acima do LA pode não corresponder necessariamente a um verdadeiro alelo, podendo tratar-se de um outro tipo de artefacto que origine picos extra (como *stutters*, picos partidos, *pull-ups*, *spikes* e *drop-ins*, mais abaixo explicados), ou até mesmo de um pico de ruído elevado que, ocasionalmente, ultrapassou o LA fixado.

Os fabricantes dos equipamentos de EC recomendam por defeito o estabelecimento de um LA de **50 RFU**, contudo cada laboratório deve fazer os seus próprios cálculos baseando-se na distribuição de valores de ruído observados nas suas condições específicas de trabalho.

O valor de LA determinado é então utilizado no *software* de análise (doravante referido como **GM**) por forma a excluir automaticamente do EF os picos abaixo desse valor e a apresentar apenas como sinal analítico de confiança, os picos acima do LA. Assim, o perito pode então proceder à interpretação dos picos detetados, distinguindo-os entre verdadeiros alelos e artefactos (Butler, 2010, 2015; SWGDAM, 2017).



**Figura 11.** Exemplo que demonstra a importância do estabelecimento de um LA. Um pico com altura inferior ao LA não pode ser considerado com segurança como sendo efetivamente um sinal analítico, pois poderá corresponder a ruído de fundo. Um pico com altura superior ao LA pode ser considerado como sinal analítico de confiança, i.e., como um produto efetivamente resultante da amplificação. Adaptado de (Butler, 2010)

### 1.5.2. STUTTERS

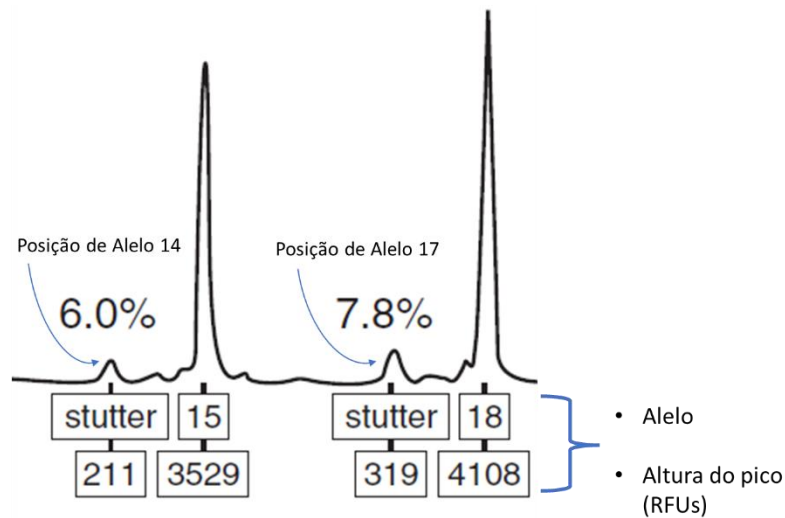
Durante a PCR, é comum haver a formação de fragmentos que sejam uma unidade de repetição maior ou menor que o verdadeiro alelo presente na amostra, devido a um deslizamento da ADN polimerase durante a fase de extensão. Esses fragmentos originam picos adicionais no EF denominados **stutter +** e **stutter -**, respetivamente, sendo os **stutter -** os mais comumente observados e os que levantam mais questões por frequentemente apresentarem picos acima do LA.

A formação de picos *stutter* não é uniforme, sendo a sua ocorrência mais pronunciada, quanto menores forem as unidades de repetição. Por isso, STRs di- e trinucleotídicos têm maior percentagem de formação de *stutter* associada, comparativamente aos STRs tetra- e pentanucleotídicos, sendo esta uma das razões para a utilização preferencial destes últimos em genética forense. A formação de *stutters* aumenta também quando se utilizam quantidades de ADN elevadas para amplificação ou quando o número de ciclos de PCR aplicados é superior ao recomendado (Butler, 2010; Goodwin *et al.*, 2011).

Se não forem desconsiderados do EF, a presença deste tipo de artefactos leva a uma incorreta genotipagem pelo GM, que fará atribuição alélica aos *stutters*, podendo levar à interpretação errada dos genótipos envolvidos no perfil da amostra (Crespillo *et al.*, 2012). Em amostras seguras, os *stutters* não costumam gerar problemas interpretativos pois são facilmente identificáveis por apresentarem uma altura de pico (RFUs) consideravelmente menor em comparação com o pico do alelo verdadeiro - geralmente não correspondem a mais do que **15%** da altura do verdadeiro alelo - e por se encontrarem numa posição imediatamente antes (**Figura 12**) ou depois do pico do alelo real. A presença de *stutters* introduz mais dificuldades interpretativas nas amostras problema, pois: em amostras com quantidades vestigiais de ADN e/ou ADN degradado, os picos verdadeiramente alélicos, em razão da sua escassa presença e/ou integridade na amostra, surgem com baixa intensidade de sinal (RFU) aparecendo no EF com alturas na mesma ordem de grandeza dos produtos meramente *stutter*, tornando difícil a distinção entre os alelos e esses artefactos; e em amostras de mistura com desproporção nas contribuições quantitativas de ADN entre indivíduos, picos na verdade pertencentes a um contribuinte minoritário podem ser erroneamente considerados como picos de artefacto quando se encontram na posição de *stutter* (posição com menos ou mais 1 unidade de repetição) de um alelo do contribuinte maioritário, correndo o risco de serem descartados do EF pelo perito.

Para auxiliar o perito na distinção entre picos verdadeiramente alélicos e picos meramente *stutter* nos EFs, os laboratórios calculam, para cada marcador genético, um **Limiar de Stutter (LS)** - um valor estimado em % relativamente aos alelos principais, que determina o valor acima do qual é razoável admitir que um pico observado naquele marcador específico será um alelo real e não apenas um pico *stutter*. Os laboratórios podem, em alternativa, adotar um LS global de 15%, contudo a abordagem anterior é preferível, uma vez que cada *locus* tem uma tendência diferente para produzir *stutters*. Atualmente, os fabricantes dos *kits* de amplificação já fazem esse estudo exaustivo que conduz aos LS de cada *locus*, incluindo nos seus cálculos *stutters* de todos os alelos possíveis em cada marcador, e, portanto, os laboratórios podem usar esses valores recomendados.

Os valores de LS determinados são então aplicados como “**filtro de stutters**” nos parâmetros de análise do GM, por forma a que o *software* genotipe apenas os picos que se encontrem acima do LS do marcador em questão, e que ignore automaticamente os picos que se encontrem abaixo desse valor. Assim, o perito tem o seu trabalho de interpretação do perfil facilitado. É de notar que picos acima do LS podem não corresponder necessariamente a alelos, mas a outros artefactos como picos partidos, *pull-ups*, *spikes* e *drop-ins* (mais abaixo mencionados) ou até mesmo produtos *stutter* ocasionalmente elevados perante o LS estabelecido para o marcador em questão (Butler, 2015; Gill *et al.*, 2020; SWGDAM, 2017).



**Figura 12.** Exemplo de picos *stutter* - num *locus* heterozigótico. O *stutter* à esquerda corresponde apenas a 6.0% da altura do alelo real (15); o *stutter* da direita corresponde apenas a 7.8% da altura do alelo verdadeiro (18). Os picos relativos a esses artefactos não surgem identificados como alelos, porque neste EF foi já aplicado um filtro de *stutters*. Adaptado de (Butler, 2010)

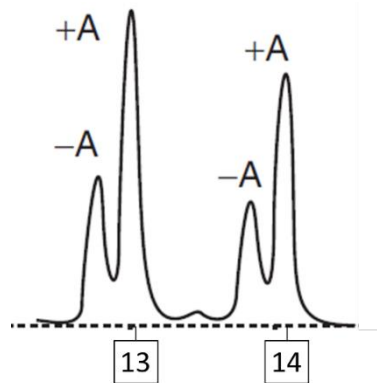
### 1.5.3. PICOS PARTIDOS

Na sua imperfeição, as polimerases usadas na PCR apresentam também uma tendência para acrescentar um nucleótido, geralmente uma adenosina (A), no final de cada fragmento, durante a fase de extensão. Esta adição não baseada no ADN molde resulta em fragmentos com um par de bases a mais do que a sequência original (+A), mas cujos picos nos EFs correspondem aos alelos reais. No entanto, se houver fragmentos amplificados que não tenham sofrido esta adição (-A) – extensão incompleta - gera-se um **pico partido**, à esquerda do pico +A (**Figura 13**).



Estes picos adicionais (-A) podem levar à incorreta atribuição alélica pelo GM, no entanto, não podem ser interpretados como verdadeiros alelos. Como tal, devem ser eliminados manualmente pelo perito, que os deve reconhecer facilmente pelo seu formato, que não forma sequer um pico bem definido, mas sim uma “barriga” à esquerda do pico do alelo principal; por apresentarem uma altura menor relativamente ao alelo maior; e por serem visíveis ao longo de todo o EF indicando que os fragmentos não foram completamente amplificados.

Os picos partidos ocorrem geralmente quando existe excesso de ADN na reação de amplificação, podendo este efeito ser minimizado aumentando o tempo de extensão para 45 a 60 min para que a polimerase tenha mais tempo para fazer a adenilação de todos os fragmentos ou em alternativa pode proceder-se à diluição da amostra e sua reanálise. É de notar que, se o tempo de extensão for demasiado prolongado, pode ocorrer a adição de mais do que uma adenosina (geralmente adição dupla) no final dos produtos de PCR, resultando em picos partidos à direita do pico +A (Butler, 2010; Goodwin *et al.*, 2011).



**Figura 13.** Picos partidos (-A) vs. picos correspondentes aos alelos reais (+A).

Adaptado de (Butler, 2010)

#### 1.5.4. *PULL-UPS* e *SPIKES*

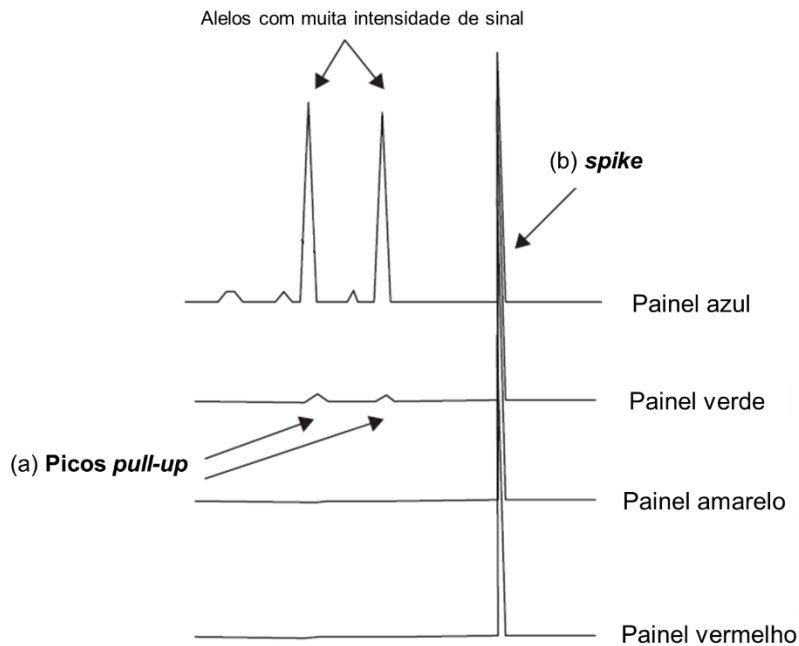
Durante a fase de deteção do produto amplificado é utilizada uma matriz, para calibração espectral, que elimina a sobreposição de cores de fluorescências que é gerada no final da eletroforese, permitindo assim distingui-las e dar origem aos diferentes painéis de cor que vemos num EF. Quando esta matriz não se encontra bem calibrada, e na presença de um pico de grande intensidade num determinado painel de cor, esse pico é como que puxado para outros painéis, surgindo então picos adicionais de baixa intensidade e na mesma posição do pico saturado (**Figura 14 (a)**). Este fenómeno designa-se por **passagem de cor** e os picos

que dela resultam são designados de ***pull-ups***. As passagens de cor também podem ocorrer sem que a matriz esteja descalibrada, quando existe excesso de ADN em análise, cuja intensidade de fluorescência demasiado forte satura o detetor (***picos off-scale, OS***), originando um perfil sobrecarregado com tendência a *pull-ups*.

Os *pull-ups*, não correspondendo a verdadeiros alelos, devem ser descartados do EF para evitar genotipagens incorretas, sendo facilmente identificáveis por surgirem na mesma posição de picos demasiado altos, no(s) painel(eis) de cor imediatamente adjacente(s) ou até mesmo em todos os painéis do EF.

Outro tipo de artefacto são disparos aleatórios de fluorescência (***spikes***) que podem ocorrer devido a alterações mínimas de voltagem durante a corrida de EC ou pela passagem de “lixo” pelo laser, como cristais de polímero ou bolhas de ar. A morfologia destes picos permite distingui-los de um verdadeiro alelo pois são mais altos, pontiagudos e estreitos e geralmente aparecem em todos os painéis de cor do EF, no mesmo sítio e com a mesma intensidade, o que facilita o seu descarte do EF (**Figura 14 (b)**) (Butler, 2010, 2015).

Os *softwares* de análise atuais incorporam algoritmos computacionais que identificam automaticamente estes artefactos instrumentais (*pull-up* e *spikes*) facilitando o trabalho de interpretação do perito (Applied Biosystems, 2012, 2019d).



**Figura 14.** Picos verdes de baixa intensidade imediatamente abaixo dos picos azuis de elevada intensidade, são *pull-ups* (a); O pico que atravessa todos os painéis de cor, na mesma posição, é um *spike* (b). Adaptado de (Butler, 2010)

### 1.5.5. **DROP-IN**

As técnicas utilizadas para obtenção de perfis genéticos são cada vez mais sensíveis aumentando assim a probabilidade de obtenção de um perfil completo, no entanto, aumenta igualmente o surgimento de alelos extra (**drop-in**), decorrentes de contaminações esporádicas. Esses picos contaminantes levantam problemas particularmente em amostras com pouca quantidade de ADN, pois podem apresentar alturas na mesma ordem dos do verdadeiro perfil. Uma amostra pode ser contaminada por ADN presente no ambiente que a rodeava aquando da sua recolha, por ADN de amostras diferentes aquando do seu processamento (contaminação cruzada) ou pelo próprio ADN do colaborador.

O fenómeno de *drop-in* geralmente não é reprodutível e pode ser detetado analisando uma mesma amostra várias vezes, sendo muito baixa a probabilidade de se obter o mesmo alelo extra em duas análises independentes. Por esta razão, é implementada na rotina laboratorial: a amplificação das amostras extraídas em duplicado (sempre que possível) e dessa forma, um alelo só é considerado como efetivamente pertencente ao perfil, se estiver presente em ambos os EFs; a inclusão de controlos negativos, pois se revelarem os mesmos alelos extra que surgiram na amostra, esses alelos podem ser desconsiderados do perfil, sendo eliminados manualmente do EF, no GM; a adoção de práticas de prevenção da contaminação que garantam um ambiente de trabalho estéril e que evitem a contaminação cruzada (mais extensivamente abordadas na secção **3.3. Práticas de prevenção da contaminação**); e a inserção do perfil genético dos colaboradores na base de dados do laboratório, para que se possa fazer a exclusão de eventuais contaminações pelo operador (Butler, 2010; Gill *et al.*, 2020).

### 1.5.6. **ALELOS OFF-LADDER**

Na população humana surgem, por vezes, alelos STR raros que contêm uma unidade de repetição incompleta - **alelos microvariantes** (ou não consensuais) - diferindo assim dos alelos mais comuns (consensuais) para os marcadores analisados, por apenas alguns pares de bases, a mais ou a menos. Essas variações na sua sequência resultam geralmente de inserções ou deleções de nucleótidos.

Alelos microvariantes ocorrem regularmente em *loci* STR mais polimórficos. Um exemplo de microvariante que ocorre com muita frequência é o alelo 9.3 no *locus* TH01, que contém 9 repetições completas do motivo de repetição AATG mais uma repetição parcial de

3 bases (ATG) na 7.<sup>a</sup> unidade, diferindo assim do alelo 10 por uma única base de adenina a menos, e do alelo 9 por 3 bases a mais.

Pelo facto de não possuírem o mesmo tamanho (pb) que os alelos mais comuns presentes no *ladder* alélico de referência, os alelos microvariantes geralmente não são designados numericamente (o GM não lhes faz atribuição alélica), surgindo no EF com a designação **OL (Off-Ladder)**. Essa situação não se aplica no caso de microvariantes frequentes (como é o caso do exemplo acima indicado), que já vêm incluídos na faixa alélica dos *ladders*.

Também pode acontecer, ainda que raramente, o surgimento de OLs não necessariamente correspondentes a microvariantes, mas sim a novos alelos com unidades de repetição completas, contudo em número pouco frequente na população, e por isso também não inseridos nos *ladders*.

O surgimento de alelos OL geralmente não constitui problema aquando da interpretação dos EFs, podendo facilmente ser confirmados como microvariantes ou novos alelos, ou descartados como artefactos resultantes por exemplo de passagens de cor (*pull-ups*) ou de “lixo” que passou pelo laser, por comparação com os EFs de duplicados da amplificação ou de reinjeções (Butler, 2010, 2015).

### 1.5.7. DESEQUILÍBRIO HETEROZIGÓTICO

Um *locus* homozigótico possui 2 alelos de igual tamanho, indistinguíveis ao nível da EC pelo facto de migrarem ao mesmo tempo, surgindo, por isso, associado a esse *locus*, um único pico no EF. Já um *locus* heterozigótico possui 2 alelos de diferente tamanho, distinguíveis na EC pelos diferentes tempos de migração, surgindo então 2 picos ao nível do EF.

Idealmente, seria expectável que os 2 picos de um *locus* heterozigótico apresentassem exatamente a mesma altura (RFUs), uma vez que ambos os alelos estão presentes em igual quantidade na amostra inicial e que a PCR os amplifica, à partida, de modo uniforme (**Figura 15 a**). Contudo, isso raramente acontece, havendo sempre um pico maior do que outro. Esse desnível resulta de uma amplificação desigual entre os alelos causada por **efeitos estocásticos** – fenómenos aleatórios que introduzem variabilidade nos resultados podendo ocorrer aquando da extração, havendo mais recuperação de um alelo do que de outro; na

pipetagem das amostras, havendo a amostragem preferencial de um dos alelos em detrimento do outro; e/ou durante a PCR na qual, um dos alelos, por acaso, é mais replicado pela ADN Polimerase nos primeiros ciclos, sendo por isso preferencialmente amplificado nos ciclos seguintes.

Em amostras de um único contribuinte e com quantidade adequada de ADN, os alelos de *locus* heterozigóticos, apresentam picos que, apesar de não exatamente iguais, são bem balanceados (**Figura 15 b**) (Butler, 2010). Segundo a bibliografia, **alelos balanceados** (ou equilibrados) apresentam, de maneira generalista, uma **altura (RFU) do pico menor > 60% da altura do pico maior**, ou seja, apresentam um **desnível** entre alturas **< 40%**. Quando se verificam **desníveis  $\geq 40\%$** , ocorre o que se designa de **desequilíbrio heterozigótico (Figura 15 c)**, o que em amostras problema, introduz dificuldades de interpretação dos EFs, pois pode ser indicativo de um perfil de mistura, em que os picos desproporcionalmente mais baixos sejam pertencentes a um contribuinte minoritário. Os valores de proporção mencionados podem, no entanto, variar consoante o *locus* em questão, e devem ser determinados por cada laboratório, para cada marcador utilizado em rotina, por forma a auxiliar o perito a distinguir com maior exatidão um desnível entre picos considerado “normal” (e portanto, a presença de um perfil singular) e um desequilíbrio que sugira a presença de um perfil de mistura. Para isso, os laboratórios determinam o chamado **Limiar de Desequilíbrio de Heterozigóticos (LD)** – valor, estimado em percentagem (%), para cada marcador genético, abaixo do qual é razoável admitir que a existência de 2 alelos poderá corresponder a uma mistura de pelo menos 2 intervenientes, e não de 1 só indivíduo heterozigótico. O estabelecimento deste limiar é essencial para concluir, com confiança, a presença ou não de um perfil de mistura, e para direcionar a análise e interpretação do perfil (Butler, 2015; Crespillo *et al.*, 2012; Gill *et al.*, 2020; SWGDAM, 2010, 2017).

#### **1.5.8. DROP-OUT**

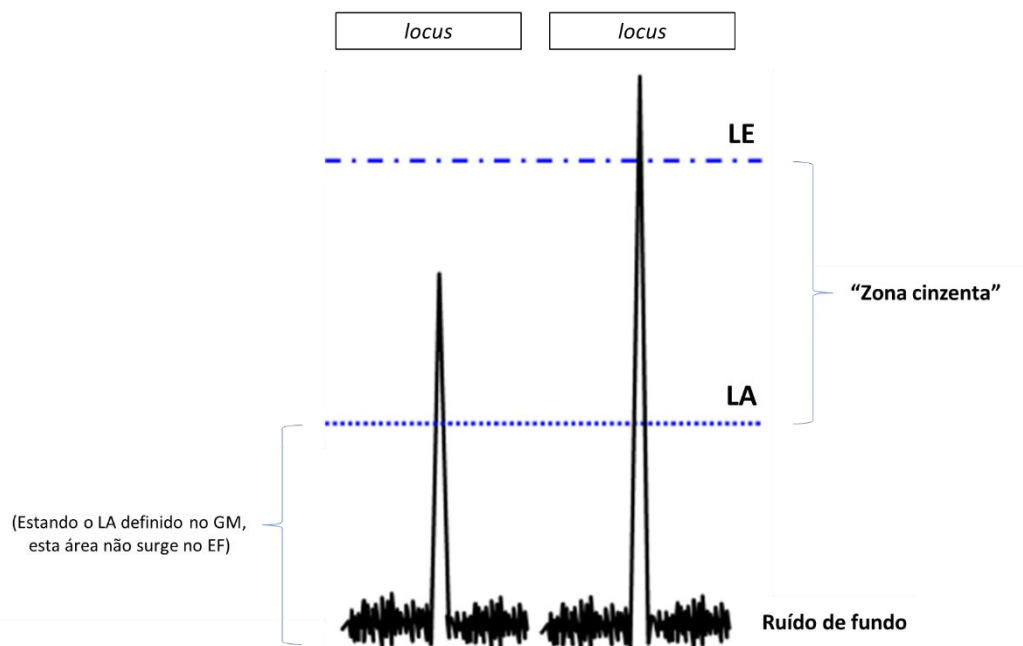
Em amostras com quantidades vestigiais de ADN, a amplificação desigual dos alelos de *loci* heterozigóticos, devida a flutuações estocásticas, ocorre de maneira mais acentuada, verificando-se frequentemente situações de **desequilíbrio heterozigótico (Figura 15 c)**). Por vezes, a disparidade entre as amplificações é tão grande que se pode verificar um **drop-out alélico (Figura 15 d)**, isto é, o desaparecimento de um dos alelos do *locus* no EF - quer por não ter amplificado de todo, ou pela amplificação sofrida ter sido de tal forma diminuta que originou um sinal indistinguível do ruído de fundo (abaixo do LA) - conduzindo à produção de perfis parciais.

O desequilíbrio heterozigótico e o *drop-out* podem ocorrer, não só devido a eventos estocásticos, mas também, embora raramente, devido à existência de uma mutação na zona de *annealing* dos *primers* num dos alelos.



**Figura 15.** Esquema representativo de possíveis resultados de amplificação: **a)** alelos com picos totalmente equilibrados - amplificação ideal; **b)** ligeiro desnível entre picos, com alelos bem balanceados – típico de amostras seguras; **c)** desequilíbrio heterozigótico – típico de amostras com baixa quantidade de ADN ou de mistura; **d)** *drop-out* alélico. Adaptado de (Butler, 2010)

Em amostras problema que surjam no laboratório, o fenómeno de *drop-out* pode originar uma interpretação errada de um perfil como homozigótico num *locus* que na verdade é heterozigótico (Butler, 2015; Goodwin *et al.*, 2011). Para que o perito consiga interpretar o perfil com confiança suficiente relativamente a possíveis efeitos estocásticos em amostras com pouca quantidade de ADN, o laboratório deve estabelecer um **Limiar Estocástico (LE)**. Esse limiar consiste no valor (em RFU) acima do qual se pode admitir que não ocorreu perda alélica (*drop-out*) em heterozigóticos, pelo que pode ser assumida, com total segurança, que a presença de 1 só alelo, num determinado *locus*, é compatível com homozigotia, não havendo a possibilidade do genótipo estar apenas parcialmente presente naquele marcador (Crespillo *et al.*, 2012; SWGDAM, 2010, 2017). A área entre o LA e o LE (“zona cinzenta”) corresponde assim a uma região influenciada por efeitos estocásticos. Informação genética que nela se encontre transmite, portanto, alguma incerteza, não sendo totalmente esclarecedora quanto há possível perda alélica (**Figura 16**). Tipicamente são estabelecidos valores de LE **entre 150 e 200 RFU**, mas os valores variam consoante a sensibilidade do equipamento de EC usado e os reagentes aplicados, devendo, portanto, cada laboratório determinar experimentalmente o seu próprio LE nas suas condições laboratoriais específicas (Gill *et al.*, 2009; SWGDAM, 2010, 2017).



**Figura 16.** Diferença interpretativa entre um pico que se situe abaixo do LE e acima do LE. Se num determinado *locus*, surgir um pico abaixo do LE estabelecido (como o representado à esquerda), é possível ter ocorrido *drop-out* do seu eventual alelo irmão, não se conseguindo afirmar com certeza, a homozigotia no *locus*. Mas se surgir um pico acima do LE (como o representado à direita), é razoável admitir que não terá ocorrido *drop-out* e, portanto, o *locus* pode ser considerado com segurança, como homozigótico. Adaptado de (Crespillo *et al.*, 2012)

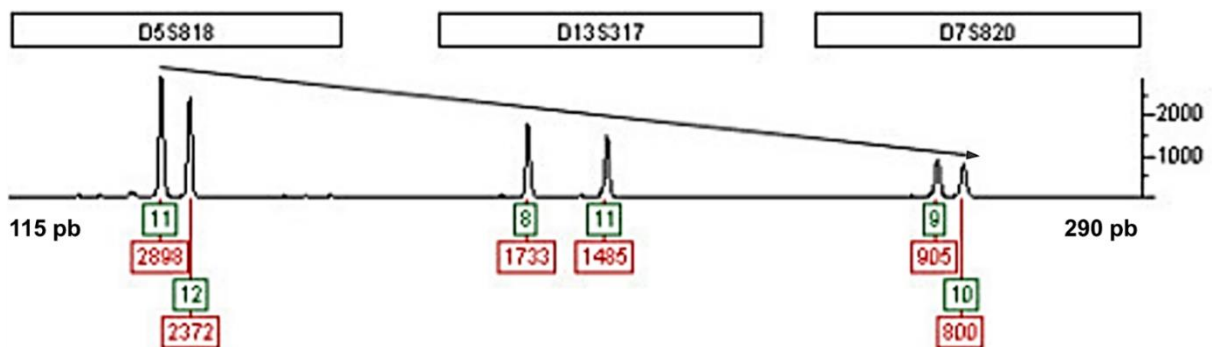
Além de recorrer a um LE para melhor aferir se o perfil em questão estará completo ou incompleto, por vezes, os laboratórios recorrem também à otimização das condições de eletroforese, quando na presença de amostras problema com pouco ADN. No SGBF-C, essa otimização centra-se na alteração do **tempo de injeção**. O serviço analisa sempre todas as suas amostras com **10s** de injeção (PE-SGBF-C-001, 2020; PE-SGBF-C-002, 2018), tempo esse considerado *standard* por produzir a melhor resolução alélica nos EFs segundo os estudos efetuados pelos fabricantes dos equipamentos de EC (Applied Biosystems, 2014; Butler, 2011, 2015). No entanto, quando após análise de amostras vestigiais nas condições padrão de eletroforese, surgem EFs pouco definidos (picos com baixa intensidade de sinal não sendo suficientemente esclarecedores na distinção entre verdadeiros alelos vs. artefactos e/ou que geram dúvidas quanto à confirmação de homozigotia em marcadores cujo sinal se encontra abaixo do LE) ou perfis aparentemente incompletos para os quais se pretende perceber se há algum alelo mascarado, é prática comum no SGBF-C reinjetar as amostras a **40s** (PE-SGBF-C-002, 2018). A reanálise de amostras com um tempo de injeção mais prolongado permite, assim, “aspirar” mais amostra, e o maior número de fragmentos a migrar

origina maiores alturas de picos (RFU), o que gera conseqüentemente um sinal mais forte e um EF melhor definido, permitindo então, o esclarecimento de certos perfis (Applied Biosystems, 2014; Butler, 2011, 2015).

### 1.5.9. DEGRADAÇÃO

O ADN presente em amostras recolhidas de cenas de crime e de restos mortais, muitas vezes, encontra-se degradado devido aos próprios processos de digestão enzimática celular e à exposição a fatores ambientais como humidade, radiação solar e variações de temperatura, que o alteram estruturalmente e o fragmentam.

Alelos maiores sofrem tendencialmente mais degradação, do que alelos mais pequenos. Assim, a sua amplificação aquando da PCR ocorre também de maneira desigual, havendo amplificação preferencial dos fragmentos mais pequenos. Isto traduz-se, ao nível do EF, numa diminuição da altura dos picos (RFU) à medida que o tamanho dos marcadores (pb) aumenta, i.e., da esquerda para a direita, conforme ilustra a **Figura 17** (Gill *et al.*, 2020; Goodwin *et al.*, 2011). Perfis genéticos de amostras degradadas são muitas vezes incompletos devido a fenómenos de *drop-out* que se verificam nos *loci* maiores. Essa presença de informação genética parcial complica a interpretação do perfil, correndo-se o risco de assumir certos *loci* como homocigóticos, que na verdade são heterocigóticos. Surge aqui novamente a importância do estabelecimento de um LE e da eventual aplicação de um maior tempo de injeção.



**Figura 17.** Perfil genético característico de uma amostra com ADN degradado.

Adaptado de (Goodwin *et al.*, 2011)



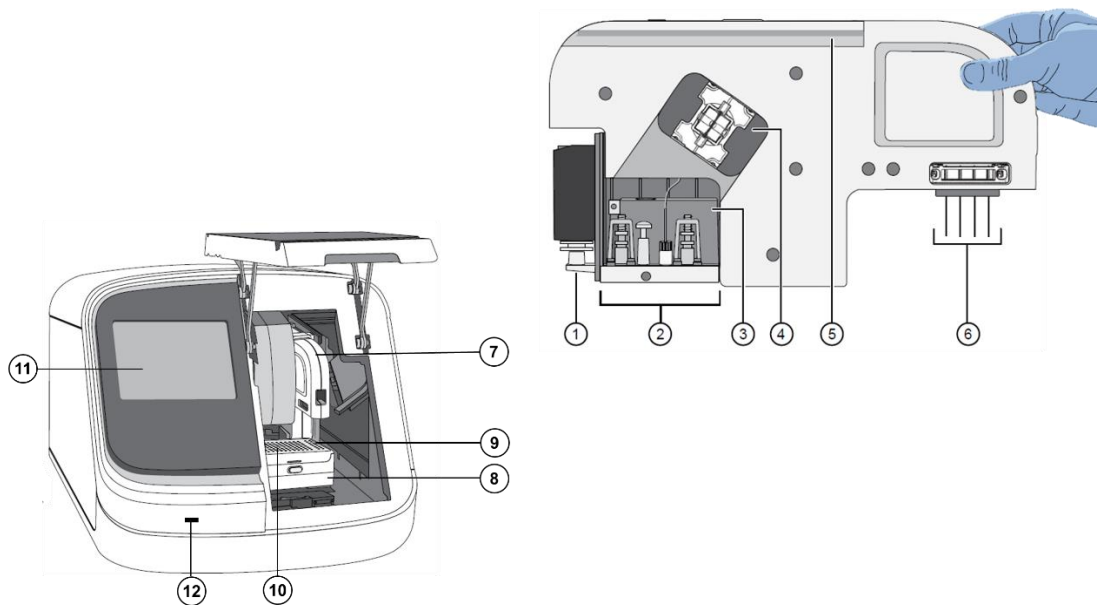
Conforme referido ao longo de toda a secção **1.5.**, para um correto entendimento dos EFs obtidos - e particularmente dos das amostras problema (degradadas, com pouco ADN e, eventualmente, de mistura) que são as que levantam mais dúvidas - é essencial estabelecer limiares como critérios fixos de análise, edição e interpretação de perfis. O cálculo desses valores é feito através de estudos de validação interna (que serão mais à frente abordados na secção **1.7. Contextualização da Validação**) que são realizados individualmente por cada laboratório uma vez que cada um possui as suas condições laboratoriais específicas.

Estando um perfil devidamente genotipado e interpretado com base nesses parâmetros, há confiança suficiente na qualidade do mesmo, sendo este, por fim, validado. Estando o perfil validado, o perito pode proceder à sua comparação com perfis de outras amostras - geralmente as de referência - avaliando se são coincidentes (inclusão) ou discrepantes (exclusão); efetuar valorizações probabilísticas; e emitir por fim o relatório pericial com as conclusões obtidas (Butler, 2015).

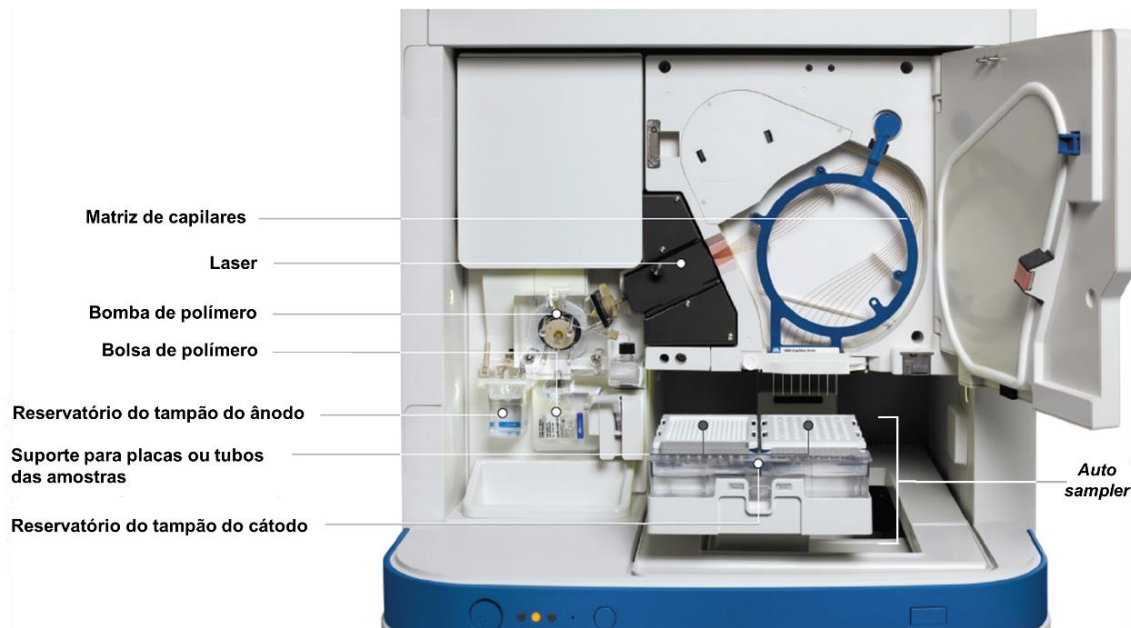
## 1.6. SEQUENCIADORES AUTOMÁTICOS – *SEQSTUDIO™ GENETIC ANALYZER (AB™)*

O SGBF-C tem em utilização, desde 2010, um único sequenciador automático, o **3500 Genetic Analyzer (AB™)**, considerado o *standard* na eletroforese capilar até à data. Porém, devido à elevada demanda das suas perícias, viu-se justificada a aquisição de um outro sequenciador, mais recente, que pudesse agilizar o trabalho do serviço, tendo sido escolhido o **SeqStudio™ Genetic Analyzer (AB™)**, pelas vantagens a seguir apresentadas.

O *SeqStudio™* apresenta um formato inovador e prático de funcionamento, que reúne a maioria dos consumíveis necessários à EC numa única peça – o cartucho *SeqStudio™ Genetic Analyzer Cartridge v2* – nomeadamente a matriz de capilares; o polímero; um sistema de distribuição de polímero; e o tampão de ânodo, conforme ilustrado na **Figura 18 (em cima)**. Esta configuração “*all-in-one*” aliada ao seu encaixe direto no instrumento, reduz o contacto manual do utilizador com os componentes da EC, evitando assim possíveis contaminações e minimizando o risco de danificar os capilares que são tipicamente frágeis, contrastando com a configuração do *3500* no qual a inserção de cada componente é realizada individualmente, tornando o processo mais demorado e mais exposto a contaminações, erros de colocação e quebra de capilares (**Figura 19**).





**Figura 18. Componentes do cartucho *SeqStudio™ Genetic Analyzer Cartridge v2* (em cima):** 1 – Polímero universal (POP-1™); 2 - Sistema de distribuição de polímero (bomba polímero para os capilares); 3 – Reservatório contendo o tampão do ânodo; 4 – Janela de deteção ótica; 5 – Sulco do cartucho para sua inserção no instrumento; 6 – Pontas dos 4 capilares (extremidade do cátodo), encontrando-se a restante matriz protegida no interior do cartucho; **Componentes da frente e interior do instrumento *SeqStudio™* (em baixo):** 7 – Cartucho; 8 – *Autosampler*; 9 – Tampão do cátodo; 10 – Suporte para a placa ou tubos das amostras; 11 – Ecrã tátil; 12 – porta USB. Adaptado de (Applied Biosystems, 2019g)



**Figura 19.** Componentes e configuração do 3500 Genetic Analyzer. Adaptado de (Applied Biosystems, 2015)

Além disso, o polímero do cartucho (POP-1™) e o tamanho dos seus capilares são de utilização universal, estando otimizados para realizar análise de fragmentos e sequenciação de Sanger numa mesma corrida, o que não é possível no 3500, havendo neste último, a necessidade de trocar o tipo de polímero e a matriz de capilares consoante o módulo de corrida a executar (**Tabela 1**). Por último, apesar de não tão importante, destaca-se também a existência de um ecrã tátil interativo e de fácil utilização que permite inserir as características das amostras a correr e definir os parâmetros da corrida diretamente no instrumento, sem a obrigatoriedade de o ligar a um computador propositadamente para esse fim (**Figura 18, em baixo**) (Applied Biosystems, 2010, 2019g, 2021).

**Tabela 1.** Tipo de polímero e de capilares utilizados no 3500 e no SeqStudio™. Adaptado de (Applied Biosystems, 2019a).

		Sequenciador automático	
		3500	SeqStudio
	<b>Módulo de corrida</b>		
<b>Tipo de Polímero</b>	Análise de Fragmentos	POP-4	POP-1
	Sequenciação de Sanger	POP-6 ou POP-7	
<b>Comprimento dos capilares (cm)</b>	Análise de Fragmentos	36	28
	Sequenciação de Sanger	50	

Tendo em vista a facilidade, rapidez e conveniência associadas ao manuseio do *SeqStudio™*, este foi o sequenciador adquirido em 2021 para colmatar a necessidade de resposta aos casos forenses que dão entrada no SGBF-C. Contudo, o equipamento ainda carece de validação interna para entrar em funcionamento na rotina laboratorial do serviço, para a qual o presente trabalho pretende contribuir.

## 1.7. CONTEXTUALIZAÇÃO DA VALIDAÇÃO

---

A confiança nos resultados obtidos de uma análise de ADN é um aspeto de extrema importância em Genética Forense, pelo facto de importantes decisões jurídicas poderem deles derivar. Por isso, é recomendada a **acreditação** de todos os serviços forenses que realizam perfis de ADN, segundo a norma de referência internacional **NP EN ISO/IEC 17025** - “*Requisitos gerais de competência para laboratórios de ensaio e calibração*” (Houck, 2015). Esta norma estabelece diretrizes destinadas a garantir a realização de boas práticas laboratoriais e a adoção de sistemas de gestão e qualidade, cujo cumprimento permite comprovar que determinado laboratório opera com competência técnica e que gera resultados exatos e corretamente interpretados nas atividades que executa. Um laboratório acreditado segundo a norma supramencionada promove assim a confiança no seu trabalho e, portanto, a aceitação e equivalência dos resultados por ele obtidos tanto a nível nacional como internacional, o que por sua vez facilita a cooperação com outros laboratórios igualmente acreditados (NP EN ISO/IEC 17025, 2018; Pinheiro, 2010).

Os laboratórios afetos ao SGBF do INMLCF encontram-se acreditados pelo Instituto Português de Acreditação (**IPAC**) no âmbito da atividade laboral que desenvolvem, de acordo com a norma NP EN ISO/IEC 17025, assegurando assim a qualidade e integridade no serviço prestado, desde os seus procedimentos implementados à emissão dos resultados de tipagem, sendo estes credibilizados em território nacional e internacional (INMLCF, 2014).

Um dos requisitos técnicos descritos na norma supramencionada implica que todos os métodos (procedimentos, técnicas e equipamentos) de utilização recorrente no serviço estejam devidamente validados (NP EN ISO/IEC 17025, 2018).

A **validação** consiste em avaliar o funcionamento de um determinado método para que seja definido o intervalo de condições nas quais ele gera resultados ou medições analíticas fiáveis, com o objetivo de comprovar que o mesmo é adequado ao fim a que se destina, operando nesse intervalo de confiabilidade validado. Métodos validados garantem assim confiança nos seus resultados e interpretações, podendo, portanto, ser implementados com segurança na rotina de um laboratório. O processo de validação e de interpretação dos seus resultados é sempre guiado por um protocolo que segue **guidelines** pré-estabelecidas por uma entidade competente e internacionalmente reconhecida na área laboratorial em questão, e todos os estudos de validação realizados nos laboratórios devem ser registados de forma a constituírem evidência documentada de que o método conduz a resultados de qualidade (Butler, 2015; Pinheiro, 2010).

No âmbito da genética forense, existem dois tipos de validação possíveis: validação de desenvolvimento e validação interna. A **validação de desenvolvimento** é geralmente realizada por fabricantes comerciais, envolvendo estudos extensivos que comprovam que uma nova tecnologia por eles desenvolvida é passível de ser instituída na comunidade científica. Já uma **validação interna**, envolve estudos menos extensivos que os anteriores, sendo efetuada individualmente por laboratórios antes da introdução de um novo método (já previamente validado pelo fabricante) na sua rotina, por forma a demonstrar que o mesmo funciona adequadamente, naquelas condições laboratoriais específicas e para que sejam estabelecidos os parâmetros de garantia de qualidade e diretrizes de interpretação de resultados adequados (ENFSI, 2014; FBI, 2020; SWGDAM, 2016).

Surge assim, no âmbito do cumprimento dos requisitos de qualidade impostos pela norma NP EN ISO/IEC 17025, a necessidade de validar internamente o sequenciador automático recentemente adquirido pelo SGBF-C - *SeqStudio™ Genetic Analyzer (AB™)* – para que este possa ser posteriormente implementado na rotina laboratorial do serviço para separação e deteção de fragmentos, com confiança nos resultados de tipagem obtidos.

No SGBF-C, existe um documento designado “*Procedimento Geral – Validação de Ensaios*” (PG-SGBF-C-001, 2020) onde estão estabelecidos os requisitos mínimos para a validação interna de um novo método de análise de ADN, segundo as recomendações do *Scientific Working Group on DNA Analysis Methods (SWGDAM)* - um grupo científico responsável por estabelecer *guidelines* de validação de métodos de análise de ADN e de interpretação dos resultados de tipagem particularmente direcionadas aos laboratórios de genética forense (SWGDAM, 2016, 2017).

Segundo essas diretrizes, para a validação interna de um novo equipamento de EC, é necessária a **determinação dos limiares de sensibilidade** do instrumento, nomeadamente o **Limiar Analítico (LA)**, **Limiar Estocástico (LE)**, **Limiar de Stutter (LS)** e **Limiar de Desequilíbrio de Heterozigóticos (LD)**. O cálculo desses limiares é importante para que os seus valores sejam estabelecidos como parâmetros fixos de análise e de interpretação dos EFs gerados pelo equipamento, visando assim uma exclusão de artefactos, uma correta genotipagem dos perfis resultantes e uma conclusão fundamentada relativa à presença de um perfil singular ou de mistura, i.e., visando a validação dos resultados obtidos. Só assim se garante a qualidade dos perfis obtidos tornando-os então passíveis de serem comparados com perfis de outras amostras e utilizados como prova para a elaboração de relatórios periciais. Para uma completa validação, o laboratório deve calcular os limiares de análise do novo sequenciador para cada *kit* comercial de amplificação *multiplex* utilizado na sua rotina.

Além da determinação dos limiares mencionados, é também recomendado, na presença de outro(s) sequenciador(es) automático(s) em utilização no laboratório, a confirmação da reprodutibilidade dos resultados (perfis) através da realização de **estudos comparativos** entre o novo sequenciador e o(s) pré-existente(s) e já devidamente validados. Os estudos comparativos devem demonstrar que os perfis obtidos com o novo equipamento são equivalentes, concordantes e com a mesma, ou até melhor, qualidade (ENFSI, 2010; SWGDAM, 2016, 2017).

Por fim, é também necessária, a participação do SGBF-C em exercícios de colaboração para comparação de resultados entre diferentes laboratórios acreditados para o mesmo âmbito – **Comparações interlaboratoriais**. Nesses ensaios, o laboratório recebe amostras biológicas a partir das quais deve determinar o perfil genético correspondente, aplicando os mesmos procedimentos que utiliza nas amostras de rotina. Este exercício de controlo de qualidade, através da obtenção de resultados consenso entre os vários participantes, permite ao SGBF-C garantir a obtenção de resultados corretos, relativamente aos procedimentos laboratoriais utilizados na rotina laboratorial, bem como garantir os resultados das validações realizadas (Fernández *et al.*, 2015; NP EN ISO/IEC 17025, 2018; PG-SGBF-C-001, 2020). O SGBF-C participa anualmente, desde 2000, em exercícios deste cariz, organizados pela equipa do Grupo de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense (**GHEP-ISFG**) e agora, que tem na sua posse um novo equipamento de EC (*SeqStudio™*) deve incluí-lo nos exercícios, por forma a comprovar que o mesmo gera efetivamente resultados corretos e que os limiares para ele estabelecidos estão bem aferidos.

## 2. JUSTIFICAÇÃO E OBJETIVOS

---

As atividades desempenhadas pela Genética Forense centram-se na determinação e interpretação de perfis de ADN. Os perfis genéticos são atualmente obtidos através da separação e deteção fluorescente em tempo real de fragmentos, maioritariamente STRs, por eletroforese capilar em sequenciadores automáticos. Assim, este tipo de equipamento desempenha um papel relevante e indispensável nos laboratórios de genética forense.

Uma vez que dos perfis obtidos, enquanto meio de prova, podem derivar importantes decisões jurídicas, e devendo os laboratórios forenses cumprir os requisitos de qualidade impostos pela norma de referência internacional NP EN ISO/IEC 17025, é imperativo que a aptidão instrumental dos sequenciadores automáticos utilizados e os resultados por eles produzidos estejam devidamente validados, para que se garanta a fiabilidade na sua utilização em contexto de rotina laboratorial.

Surge então neste âmbito, a necessidade de validar o sequenciador automático recentemente adquirido pelo SGBF-C - *Applied Biosystems™ SeqStudio™ Genetic Analyzer* – por forma a demonstrar internamente que a sua utilização é adequada ao fim a que se destina.

### 2.1. OBJETIVO GERAL

Com o presente trabalho, pretende-se documentar os estudos de validação interna realizados no equipamento de eletroforese capilar *SeqStudio™*, descrevendo os procedimentos aplicados e os parâmetros a partir deles estabelecidos para a análise e interpretação dos perfis gerados, contribuindo assim para a posterior implementação do instrumento na rotina laboratorial do serviço, com confiança nos resultados de tipagem obtidos.

### 2.2. OBJETIVOS ESPECÍFICOS

1. Determinar e/ou estabelecer os limiares de sensibilidade do *SeqStudio™*, nomeadamente o Limiar Analítico, Limiar de *Stutter*, Limiar Estocástico e Limiar de Desequilíbrio de Heterozigóticos, para um dos *kits* de amplificação mais usados no SGBF-C - o *GlobaFiler™* – para a sua definição no serviço como parâmetros de análise e interpretação dos perfis gerados pelo equipamento;



2. Efetuar estudos comparativos entre o *SeqStudio™* e o equipamento de EC atualmente em utilização no SGBF-C e que se encontra já devidamente validado – o *Applied Biosystems™ 3500 Genetic Analyzer* - para comprovar a reprodutibilidade dos perfis gerados;
3. Verificar a aptidão do equipamento *SeqStudio™* para executar as funções de EC no âmbito da identificação humana que é realizada na rotina do SGBF-C;
4. Analisar as vantagens que o *SeqStudio™* apresenta em comparação com o 3500, e averiguar eventuais limitações do mesmo.

### 3. MATERIAIS E MÉTODOS

---

A componente prática do presente trabalho foi realizada nas instalações do SGBF-C do INMLCF.

#### 3.1. AMOSTRAS BIOLÓGICAS

---

Foram utilizadas amostras seguras de sangue e de saliva da rotina laboratorial do SGBF-C. Todas elas foram devidamente anonimizadas através de códigos, no decorrer de todos os procedimentos laboratoriais e tratamento de resultados, para que nenhuma informação sobre a sua origem - como dados pessoais dos intervenientes e o tipo de processo (IP, II, CR) a que pertenciam - fosse exposta. As amostras não foram utilizadas para qualquer outro fim além do previsto.

#### 3.2. COLHEITA E SUPORTE DAS AMOSTRAS

---

As amostras de sangue utilizadas foram colhidas através de punção capilar com lancetas descartáveis estéreis *Accu-Check® Safe-T Pro Plus* e impregnadas em cartões FTA® *Whatman® Human ID Bloodstain Card* sob a forma de mancha.

As amostras de saliva utilizadas foram colhidas por raspagem da mucosa bucal com o auxílio de zaragatoas bucais estéreis *Whatman® FTA® Omni Swab*.

Depois de secas à temperatura ambiente, as amostras em ambos os suportes, foram acondicionadas em envelopes no arquivo de amostras seguras, que se encontra na sala *Extração de Amostras Seguras*, à temperatura ambiente.

#### 3.3. PRÁTICAS DE PREVENÇÃO DA CONTAMINAÇÃO

---

No decurso da análise de amostras forenses, é relevante a adoção de estratégias que minimizem contaminações indesejadas que possam afetar os resultados obtidos e dificultar a sua interpretação.

Um parâmetro de prevenção instituído no SGBF-C é a realização das diferentes etapas de análise em **laboratórios separados**. Assim, o SGBF-C conta com duas salas de extração e duas salas de pré-PCR, sendo uma exclusivamente dedicada a amostras seguras, e outra a amostras problema; uma sala para quantificação; outra para amplificação; e finalmente, uma para separação e detecção do produto amplificado. As amostras seguem sempre uma **linha unidirecional** pelos laboratórios no sentido sala de extração → sala de separação e detecção do produto amplificado, não podendo retroceder (Corte-Real & Vieira, 2015).

Também uma medida preventiva inicial é realizada pelo operador antes de começar qualquer procedimento. Cada sala a utilizar é descontaminada por **irradiação UV** durante 10-20 min, por forma a que todas as superfícies de trabalho que serão usadas e que estarão em contacto com a amostra (bancadas, *DNA Workstations*, materiais, equipamentos e batas) estejam livres de ADN alheio. Terminado o procedimento, efetua-se a **limpeza de** todas essas **superfícies**, geralmente com *DNA AWAY™ Surface Decontaminant (Thermo Scientific™)* e repete-se a irradiação com UV.

Durante o processamento das amostras, o operador utiliza **acessórios de proteção**, nomeadamente bata, luvas de *latex*, máscara e rede de cabelo, todos eles **descartáveis** a cada procedimento concluído e **exclusivos de cada sala**, não sendo permitida a circulação entre salas diferentes com a mesma roupa de trabalho. Os equipamentos, materiais e reagentes são também exclusivos de cada laboratório específico, não podendo ser transportados para outras salas. O material usado em cada sala, no processamento de amostras, é preferencialmente descartável e estéril (como pinças e tesouras), ou na impossibilidade de tal, são esterilizados em autoclave.

Através destas práticas, minimiza-se a possibilidade de contaminação entre amostras diferentes e entre reagentes, materiais e instrumentos dos diferentes procedimentos realizados.

Além disso, para efeitos de **controlo de qualidade interno**, é imperativo adicionar sempre, em todas as fases de análise: **controles negativos** (ou brancos de amostra) para que se possa descartar (ou não) a hipótese de ter ocorrido contaminação nalgum dos procedimentos; e **controles positivos** para confirmar se o método está a funcionar corretamente e, portanto, a produzir resultados fiáveis nas restantes amostras (Butler, 2011).

Por fim, são registados todos os materiais, reagentes, equipamentos e técnicas a que se recorreram bem como os responsáveis pelo seu manuseio, data e hora do mesmo, nas folhas de registo de cada procedimento. Esse registo, garante a rastreabilidade das amostras e auxilia no controlo de reagentes e consumíveis.

### 3.4. METODOLOGIAS DE EXTRAÇÃO

---

#### 3.4.1. Extração com *Prep-n-Go™*

As amostras seguras de saliva contidas em zaragoas bucais foram alvo de extração com o tampão *Prep-n-Go™*, na sala para *Extração de Amostras Seguras*.

De cada zaragoa foi cortado um pequeno fragmento ('dente'), com o auxílio de tesoura cirúrgica, e o mesmo foi colocado no interior de um tubo *Eppendorf®* de 1.5 mL. A cada tubo foram adicionados 100µL de tampão *Prep-n-Go™* + 100µL de H<sub>2</sub>O *nucleasse-free*. Os tubos foram agitados em *vortex* para uma melhor mistura dos reagentes em torno da amostra, e colocados de seguida num banho térmico a 95°C por 5 min num *Eppendorf® ThermoMixer®* com agitação a 400 rpm, para promover a lise celular, a desnaturação das proteínas celulares e a libertação do ADN.

No final, o produto extraído ficou à temperatura ambiente durante cerca de 15 min antes de prosseguir para amplificação. Nos casos em que as amostras não prosseguiram para amplificação no próprio dia, foram guardadas a temperatura de congelamento -25°C ± 10°C na arca designada para o armazenamento de extrações de amostras seguras (IT-SGBF-C-017, 2021).

#### 3.4.2. Extração com *PrepFiler Express™*

Amostras seguras de sangue que tiveram de atravessar uma fase de quantificação, foram submetidas à metodologia de extração com o *kit PrepFiler Express™* automatizada no instrumento *Automate Express™*, na sala para *Extração de Amostras Problema*, seguindo o protocolo recomendado pelo fabricante (Applied Biosystems, 2017).

De cada mancha de sangue foi retirada uma pequena circunferência (*punch*) com auxílio de um alicate "*single hole paper puncher*" e a mesma foi inserida numa coluna *LySep™ Filter Column* encaixada num *sample tube*. Elaborou-se a solução de lise - 500µL *PrepFiler™*

*Lysis Buffer* + 5µL DTT (1M) por amostra - num tubo *Falcon*® e transferiram-se 500µL dessa mistura para cada coluna. As amostras foram incubadas em *ThermoMixer*®, a 70°C durante 40 min, com agitação 750 rpm. Durante o período de incubação, preparou-se o *robot* de extração *Automate Express™* de acordo com as instruções do manual do equipamento (Applied Biosystems, 2019e).

Terminada a incubação, as amostras foram centrifugadas por 2 min a 10000 rpm para que o lisado descesse da coluna para o fundo do tubo, enquanto o suporte da amostra (*punch*) permanecia retido na coluna. Os tubos contendo o lisado foram colocados no *robot*, selecionou-se o volume de eluição 50µL e iniciou-se a corrida de extração, que tem duração de aproximadamente 30 min.

Durante esse processo automático, o lisado começa por ser misturado com partículas magnéticas para que o ADN presente se ligue a elas; depois as partículas magnéticas que retêm o ADN são atraídas a um íman sendo separadas do restante lisado, e purificadas com tampões de lavagem para remoção de eventuais inibidores da PCR presentes na solução; por fim, o ADN, já concentrado e purificado, é eluído (i.e., solta-se das partículas magnéticas) com tampão de eluição.

Findando a extração automatizada, os tubos de eluição contendo a amostra extraída, foram retirados do equipamento e armazenados na arca designada para o armazenamento de ADN extraído de amostras problema à temperatura de congelamento  $-25^{\circ}\text{C} \pm 10^{\circ}\text{C}$ .

### 3.5. METODOLOGIA DE QUANTIFICAÇÃO – *Kit Quantifiler™ Trio*

---

A quantificação de ADN foi realizada com o *kit Quantifiler™ Trio DNA Quantification (AB™)*, de acordo com as recomendações do fabricante (Applied Biosystems, 2018), na sala de *pré-PCR de Amostras Problema*.

Elaborou-se uma mistura com os reagentes *Quantifiler™ Trio PCR Reaction Mix* e *Quantifiler™ Trio Primer Mix* pertencentes ao *kit*, de acordo com os valores indicados na **Tabela 2**. Foram dispensados 18µL dessa mistura por cada poço a ser utilizado, numa placa *MicroAmp™ Optical 96-Well Reaction Plate (AB™)*. Foram seguidamente adicionados 2µL de cada amostra (soluções de ADN extraído contidas nos tubos de eluição) e dos controlos, aos poços respetivos, previamente definidos num esquema de placa feito a partir do *HID Real-Time PCR Analysis Software v.1.2*. Perfez-se assim, um volume final de 20µL por poço.

**Tabela 2.** Componentes e respectivos volumes adicionados a cada poço para a reação de quantificação.

Componentes da reação	Volume (µL) por poço a usar
<i>Quantifiler™ THP PCR Reaction Mix</i>	10
<i>Quantifiler™ Trio Primer Mix</i>	8
Solução de ADN extraído	2
<b>Volume final (µL) por poço</b>	<b>20</b>

A placa foi selada com uma película adesiva transparente *MicroAmp™ Optical Adhesive Film*, tendo o cuidado de evitar qualquer rugosidade que pudesse interferir com a medição da fluorescência no passo seguinte. Centrifugou-se levemente a placa por 20s a 3000 rpm para retirar eventuais bolhas e por fim a mesma foi introduzida no equipamento *7500 Real-Time PCR Instrument*. na sala de *PCR | RT PCR*. A corrida de quantificação teve a duração de 1h e os resultados foram analisados pelo *HID Real-Time PCR Analysis Software v.1.2.*, que a partir das curvas de calibração, estimou a quantidade de ADN presente em cada poço (Applied Biosystems, 2018; IT-SGBF-C-015, 2020).

### 3.6. METODOLOGIA DE AMPLIFICAÇÃO – *Kit GlobalFiler™*

A amplificação de ADN de todas as amostras foi realizada com o *kit GlobalFiler™ PCR Amplification (AB™)*, segundo as especificações descritas pelo fabricante (Applied Biosystems, 2019f).

Para todas as amostras seguras de saliva, preparou-se, na sala de *pré-PCR de Amostras Seguras*, uma mistura constituída pelos reagentes *GlobalFiler™ Master Mix* e *GlobalFiler™ Primer Mix* pertencentes ao *kit*, de acordo com os valores indicados na **Tabela 3**. Foram dispensados 5 µL dessa mistura por cada poço a ser utilizado, numa placa *MicroAmp™ Optical 96-Well Reaction Plate (AB™)*. Foi adicionado 1 µL de cada amostra (solução de ADN extraído) ou controlo ao poço correspondente, segundo o mapa de placa criado, e 6.5 µL de Tampão *Low TE* (10mM Tris-HCL, 0.1mM EDTA, pH = 8.0) de forma a perfazer o volume final definido para a amplificação de amostras seguras (12.5 µL). Para as amostras seguras de sangue submetidas a amplificação direta, o procedimento foi idêntico, mas ao invés de 1 µL de produto extraído, adicionou-se 1 ou 2 *punchs*, consoante a mancha estivesse bem ou pouco impregnada, respetivamente, mantendo sensivelmente o mesmo volume final de 12.5 µL (**Tabela 4**).

**Tabela 3.** Componentes e respetivos volumes adicionados a cada poço para a reação de amplificação de amostras seguras de saliva. O mesmo se aplicou às amostras seguras de sangue que atravessaram pela fase de extração e quantificação e que foram seguidamente amplificadas como amostras seguras.

Componentes da reação	Volume (µL) por poço a usar
<i>GlobalFiler™ Master Mix</i>	3.75
<i>GlobalFiler™ Primer Mix</i>	1.25
Solução de ADN extraído	1
TE	6.5
<b>Volume final (µL) por poço</b>	<b>12.5</b>

**Tabela 4.** Componentes e respetivos volumes ou quantidades adicionados a cada poço para a reação de amplificação direta de amostras seguras de sangue.

Componentes da reação	Volume (µL) ou quantidade por poço a usar
<i>GlobalFiler™ Master Mix</i>	3.75
<i>GlobalFiler™ Primer Mix</i>	1.25
Mancha de sangue	1 ou 2 <i>punches</i>
TE	6.5
<b>Volume final (µL) por poço</b>	<b>12.5</b>

Já para as amostras seguras de sangue que sofreram extração e quantificação:

- e cujo processamento seguiu as condições laboratoriais típicas das amostras seguras, foi aplicado o mesmo procedimento de amplificação relativo à **Tabela 3**;
- e cujo processamento seguiu as condições laboratoriais de amostras problema, foi preparada, na sala de *pré-PCR de Amostras Problema*, a mistura de reagentes consoante os valores presentes na **Tabela 5**. Dessa mistura, foram dispensados 10 µL por cada poço da placa a usar. A cada poço foi adicionado 1 µL de cada amostra ou controlo, de acordo com o mapa de placa, e 14 µL de TE de maneira a perfazer o volume final definido para a amplificação de amostras problema (25 µL).

**Tabela 5.** Componentes e respectivos volumes adicionados a cada poço para a reação de amplificação de amostras seguras de sangue tratadas como amostras problema (volume final de amplificação 25 µL/poço).

Componentes da reação	Volume (µL) por poço a usar
<i>GlobalFiler™ Master Mix</i>	7.5
<i>GlobalFiler™ Primer Mix</i>	2.5
Solução de ADN extraído	1
TE	14
<b>Volume final (µL) por poço</b>	<b>25</b>

As placas foram seladas com *MicroAmp™ Optical 8-Cap Strips (AB™)* e centrifugadas levemente por 20s a 3000 rpm. Por fim, colocaram-se as placas num termociclador *Veriti™ 96-Well Thermal Cycler (AB™)*, na sala de *PCR | RT PCR* no qual decorreu a reação de amplificação por PCR. O programa de amplificação selecionado no termociclador para as amostras seguras de sangue e saliva foi o “*GlobalFiler – Seguras – 27C*” descrito na **Tabela 6**, e para as amostras seguras de sangue que foram tratadas como amostras problema foi o “*GlobalFiler – Vestigiais – 29C*”, descrito na **Tabela 7**.

**Tabela 6.** Programa de amplificação “*GlobalFiler – Seguras – 27C*” selecionado no termociclador para amostras seguras.

27 ciclos				
Incubação inicial	Desnaturação	Annealing	Extensão Final	Hold
95°C, 1 min	94°C, 10 s	59°C, 90 s	60°C, 60 min	4°C, ∞

**Tabela 7.** Programa de amplificação “*GlobalFiler – Vestigiais – 29C*” selecionado no termociclador para amostras seguras tratadas como amostras problema.

29 ciclos				
Incubação inicial	Desnaturação	Annealing	Extensão Final	Hold
95°C, 1 min	94°C, 10 s	59°C, 90 s	60°C, 10 min	4°C, ∞



### 3.7. METODOLOGIA DE SEPARAÇÃO E DETEÇÃO DO PRODUTO AMPLIFICADO – Aplicação nos sequenciadores automáticos 3500 e *SeqStudio™*

Terminada a amplificação, prepararam-se as amostras para aplicação em sequenciador automático. Na sala de *Sequenciação e Separação de Fragmentos*, foi então preparada uma mistura contendo *Hi-Di™ Formamide* (AB™) e o *standard* interno *GeneScan™ 600 LIZ™ Size Standard v2.0* (AB™), de acordo com os valores indicados na **Tabela 8**. Distribuíram-se 10 µL dessa mistura pelos poços a usar numa nova placa *MicroAmp™ Optical 96-Well Reaction Plate* (AB™). Foi adicionado 1 µL do produto amplificado, controlo ou *GlobalFiler™ Allelic Ladder* nos respetivos poços, respeitando o mapa de placa previamente elaborado. Findado o preenchimento da placa de aplicação, a mesma foi selada com uma *septa 96-Well Septa* (AB™) e centrifugada brevemente para que o conteúdo ficasse bem misturado e reunido no fundo dos poços. A placa foi depois colocada num termociclador *GeneAmp® PCR System 2700* (AB™) a 95°C por 3 min para desnaturar as amostras e imediatamente inserida no *autosampler* do sequenciador, para iniciar a corrida de eletroforese capilar (Applied Biosystems, 2019f; PO-SGBF-C-012, 2021).

**Tabela 8.** Componentes e respetivos volumes adicionados a cada poço para aplicação das amostras em sequenciador automático.

Componentes da reação	Volume (µL) por poço a usar
<i>Hi-Di™ Formamide</i>	9.5
<i>GeneScan™ 600 LIZ™ Size Standard v2.0</i>	0.5
Amostra (produto amplificado)	1
<b>Volume final (µL) por poço</b>	<b>11</b>

A separação e deteção dos fragmentos amplificados foi realizada nos sequenciadores automáticos *SeqStudio™ Genetic Analyzer* (AB™) e *3500 Genetic Analyzer* (AB™).

No *SeqStudio™*, as configurações das placas e os parâmetros de corrida foram selecionados diretamente no ecrã tátil do equipamento com o *SeqStudio™ Data Collection Software v.1.2*; ou através do computador associado ao sequenciador, no programa *SeqStudio™ Plate Manager*, e importando depois o *plate setup* criado para o equipamento através de *pen USB*. Já no *3500*, essas propriedades foram definidas no computador associado ao sequenciador através do *3500 Series Data Collection Software v2*.

Em ambos os sequenciadores, foi selecionada a opção de módulo de corrida *HID Analysis* (temperatura dos capilares a 60°C; voltagem de injeção: 1.2kV; tempo de injeção: 10s; voltagem de corrida: 11 kV).

Para que as corridas decorressem através dos componentes adequados ao tipo de análise de fragmentos *HID*, as mesmas foram realizadas, no *SeqStudio™*, utilizando o cartucho *SeqStudio™ Genetic Analyzer Cartridge v2* que contém o polímero universal POP-1™ e capilares de 28 cm, enquanto no *3500*, foram realizadas utilizando o polímero POP-4™ e capilares de 36cm.

O programa de controlo automático da corrida e de recolha de dados utilizado no *SeqStudio™* foi o *SeqStudio™ Data Collection Software v.1.2.*, já no *3500* foi o *3500 Series Data Collection Software v2.*

Terminadas as corridas, os eletroferogramas obtidos em ambos os sequenciadores, foram analisados no *GeneMapper™ ID-X Software v.1.6.*

## 3.8. PARÂMETROS ESTUDADOS

---

### 3.8.1. LIMIARES DE SENSIBILIDADE

#### I – LIMIAZ ANALÍTICO (LA)

Para determinar este limiar, foram elaborados 32 negativos de amplificação contendo apenas os reagentes do *kit* de amplificação *GlobalFiler™*, sem adição de ADN (3.75 µL *GlobalFiler™ Master Mix* + 1.25 µL *GlobalFiler™ Primer Mix* + 7.5 µL tampão TE). Estes brancos atravessaram depois por um programa de PCR de 27 ciclos, sendo aplicado, por fim, 1 µL de cada um, no *SeqStudio™*, juntamente com a mistura de aplicação (9.5 µL *Hi-Di™ Formamide* + 0.5 µL *GeneScan™ 600 LIZ™ Size Standard v2.0*). Concluída uma primeira injeção de cada negativo, foi feita uma reinjeção de cada um, totalizando assim, 64 amostras.

Após a obtenção dos eletroferogramas, com módulo de análise a 1 RFU no GM para todos os canais de cor, foi registado em folha de cálculo Excel, o valor em RFU do pico mais alto do ruído de fundo observado, em cada canal de cor (excetuando o canal do *standard interno*) para todos os negativos. Com base nos valores obtidos, foram calculados os LA de acordo com 3 métodos descritos em bibliografia, para cada canal de cor, nomeadamente:

- A abordagem clássica definida pela *International Union of Pure and Applied Chemistry* (IUPAC, 1976), baseada na fórmula proposta por Kaiser no âmbito da química analítica (Bregu *et al.*, 2013; Kaiser, 1970):

$$LA = \bar{x} + 3\sigma$$

- $\bar{x}$  = média dos valores de ruído mais elevados observados para cada amostra
- $\sigma$  = desvio padrão associado aos valores de ruído observados

- A fórmula de validação proposta pelo SWGDAM (SWGDAM, 2010):

$$LA = 2 \times (Hp - p)$$

- **Hp** = altura do pico de ruído mais alto (RFU)
- **p** = altura do pico de ruído mais baixo (RFU)

- A abordagem de validação proposta pelo Serviço de Biologia do Instituto Nacional de Toxicologia e Ciências Forenses de Barcelona (**INTCF-B**), presente nas recomendações do GHEP-ISFG (Crespillo *et al.*, 2012):

$$LA = Q3 + 3 \times (IQR)$$

- **IQR** = amplitude interquartil =  $Q3 - Q1$
- **Q3** = valor do 3º quartil ; **Q1** = valor do 1º quartil
- Para determinar os quartis ordenaram-se os valores de RFU por ordem crescente

## II – LIMIAR DE STUTTER (LS)

Para determinar este limiar, devem ser avaliados, segundo as *guidelines* (Crespillo *et al.*, 2012; SWGDAM, 2010), pelo menos 150 EFs de amostras seguras de rotina, de indivíduos diferentes, amplificadas com o *kit* de amplificação em questão - no caso, o *GlobalFiler™* - na tentativa de observar todos os alelos abrangidos pelo *ladder* alélico e, portanto, todos os *stutters* possíveis. É também aconselhado como ideal incluir nos cálculos do LS, pelo menos 5 observações de cada alelo, para que os valores de LS obtidos sejam mais significativos e mais próximos da realidade. Para efeitos de cálculo são incluídos tanto sistemas

homozigóticos como heterozigóticos de marcadores autossômicos, sendo que nestes últimos, apenas devem ser considerados aqueles que possuam alelos com mais do que 1 unidade de repetição de diferença, para evitar efeitos aditivos provocados pelos eventuais *stutter* - e + do alelo imediatamente seguinte ou anterior.

Os EFs são obtidos utilizando já como módulo de análise no GM, o valor de LA previamente determinado, e eliminando manualmente picos partidos e artefactos OL.

Seguidamente, são registados em folhas de cálculo Excel - uma para cada marcador autossômico - todos os alelos (contabilizáveis) observados e as alturas dos seus picos (RFUs) em cada amostra, bem como os RFUs dos seus produtos *stutters* correspondentes. É depois calculado o valor de *stutter*, para cada alelo, em cada amostra, utilizando a seguinte fórmula:

$$\text{Proporção de } \textit{stutter} \text{ alelo (\%)} = (\text{RFU}_{\textit{stutter}} / \text{RFU}_{\text{alelo}}) \times 100$$

Calcula-se então a média dos valores de *stutter* obtidos no passo anterior e os respetivos desvios padrão em cada marcador genético, para que se possa determinar o valor final de LS para cada *locus*, aplicando a fórmula final de cálculo:

$$\text{LS}_{\text{marcador}} (\%) = \bar{x} + 3\sigma$$

- $\bar{x}$  = média dos valores de *stutter* obtidos para os alelos encontrados no marcador autossômico em questão
- $\sigma$  = desvio padrão associado aos valores acima mencionados

Conforme referido, para que o cálculo do LS seja o mais representativo possível de cada marcador, seria desejável incluir no cálculo, *stutters* para todos os alelos frequentes na população para o *locus* em questão e abrangidos pelo *ladder* alélico. Contudo, segundo a experiência do SGBF-C em estudos de validação interna prévios de equipamentos de EC, tem vindo a notar-se que, mesmo analisando um número significativo de amostras (cerca de 150), não tem sido possível observar todos os alelos de cada *locus*, pelo que os LS obtidos nos cálculos do serviço, além de se demonstrarem sempre discrepantes e tendencialmente abaixo (**ANEXO I**) dos valores recomendados pelo fabricante, não representarão tão fidedignamente o comportamento de *stutter* de cada marcador comparativamente aos valores determinados pelo fabricante, cuja amostragem terá sido mais abrangente, envolvendo a totalidade dos alelos.

Assumindo que os cálculos realizados pelos fabricantes dos *kits* de amplificação nos estudos de validação de desenvolvimento serão mais rigorosos e, portanto, mais representativos da realidade, o SGBF-C tem vindo a adotar, como filtro de *stutter* no GM, os valores de LS por eles recomendados (Applied Biosystems, 2019f), aplicando-os atualmente no 3500 (**Tabela 9** – valores da coluna “3500”) (RV-SGBF-C-006, 2020).

Posto isto, para a obtenção de todos os EFs do *SeqStudio™* no presente trabalho (excetuando os EFs utilizados para determinar o LA), foram introduzidos como módulo de análise no GM, os valores de LS estabelecidos pela *Applied Biosystems* para cada marcador autossómico do *kit GlobalFiler™*, que se encontram abaixo indicados na coluna “*SeqStudio™*” da **Tabela 9**.

**Tabela 9.** Valores de LS estabelecidos para os marcadores autossómicos do *kit GlobalFiler™* segundo os estudos de validação de desenvolvimento realizados pelo fabricante, aplicáveis como filtro de *stutter* no 3500 e no *SeqStudio™* (Applied Biosystems, 2019f; RV-SGBF-C-006, 2020).

<i>GlobalFiler™</i>		
Sistema	LS fabricante (%)	
	3500	<i>SeqStudio™</i>
D3S1358	12.45	10.98
vWA	12.33	10.73
D16S539	10.17	9.48
CSF1PO	11.40	8.77
TPOX	5.43	5.55
D8S1179	10.20	9.60
D21S11	11.42	10.45
D18S51	13.47	12.42
D2S441	8.75	8.10
D19S433	10.58	9.97
TH01	5.24	4.45
FGA	11.96	11.55
D22S1045	17.30	16.26
D5S818	10.84	9.16
D13S317	9.98	9.19
D7S820	10.21	8.32
SE33	14.42	14.49
D10S1248	12.50	11.46
D1S1656	13.08	12.21
D12S391	15.08	13.66
D2S1338	13.14	11.73

De modo a comprovar a efetiva adequação de tais valores de LS (coluna “*SeqStudio™*” da **Tabela 9**) a um filtro de *stutter* futuramente aplicável no GM aquando da análise de amostras amplificadas com *GlobalFiler™* e corridas no *SeqStudio™*, nas condições laboratoriais do SGBF-C, compararam-se os perfis das mesmas amostras obtidos tanto no *3500* como no *SeqStudio™*, aquando da realização dos estudos comparativos (explicados na secção **3.8.2.**), verificando se foram descartados os mesmos *stutters* e realizadas as mesmas atribuições alélicas.

### III – LIMIAR ESTOCÁSTICO (LE)

Para determinar este limiar, recorreu-se a 2 amostras seguras de sangue (A e B) seleccionadas pelo elevado número de heterozigotias que apresentavam (os seus perfis tinham já sido previamente determinados por colaboradores, no *3500*, em contexto de rotina, sendo o seu genótipo conhecido *a priori*).

As 2 amostras foram extraídas com *PrepFiler Express™* e seguidamente quantificadas com *Quantifiler™ Trio*. A quantificação foi realizada para confirmar se as soluções de ADN extraído da amostra A e B possuíam uma concentração elevada de ADN total (de aproximadamente 2.5 ng/μL) passível de ser submetida às diluições pré-PCR seguidamente mencionadas; e para que, com base na concentração de ADN total nelas presente, se pudessem calcular os volumes (μL) a alíquotar dessas soluções iniciais e de TE a pipetar, para a preparação de cada uma das diluições.

Foram realizadas 6 diluições seriadas de 1:2 em ambas as amostras, para se obterem concentrações de, aproximadamente, 1,250 ng/μL, 0,625 ng/μL, 0,313 ng/μL, 0,156 ng/μL, 0,078 ng/μL e 0,039 ng/μL, por forma a garantir que em algum momento, surgissem situações de *drop-out* alélico. Cada diluição preparada (A1, A2, A3, A4, A5, A6 e B1, B2, B3, B4, B5, B6) foi também quantificada com *Quantifiler™ Trio* para confirmar a proximidade da concentração “real” de ADN total presente vs. a esperada.

Cada diluição (A1-6 e B1-6) foi amplificada com *GlobalFiler™* em duplicado (2 réplicas de cada diluição – rep1 e rep2) tanto com o volume de amplificação e número de ciclos de PCR definidos para amostras seguras (12,5μL, 27 ciclos) como para amostras problema (25μL, 29 ciclos), em placas diferentes. Por fim todas as amostras foram aplicadas no *SeqStudio™*.

Os EFs foram obtidos utilizando já como módulo de análise no GM, o valor de LA determinado experimentalmente para o *SeqStudio™* e o LS do fabricante, e deles foram depois manualmente eliminados picos partidos, artefactos OL, picos *stutter* que não foram filtrados por estarem ocasionalmente elevados e eventuais picos contaminantes.

Para o cálculo do LE seguiu-se a abordagem da (SWGDM, 2010, 2017) que, relativamente à maior diluição sem ocorrência de *drop-out* alélico, considera o valor médio, em RFU, dos alelos maiores (em intensidade de sinal, RFU) dos sistemas heterozigóticos onde ocorreu *drop-out* na diluição seguinte. Esta abordagem foi aplicada a ambas amostras (A e B), determinando-se um valor de LE para cada uma, e para cada condição laboratorial aplicada (procedimento de amostras seguras: 12,5µL, 27 ciclos; procedimento de amostras problema: 25µL, 29 ciclos) para averiguar eventuais diferenças nos efeitos estocásticos sofridos. De salientar que sistemas heterozigóticos são possíveis unicamente nos marcadores autossómicos e na AMEL, sendo estes os *loci* contabilizados na análise dos EFs para determinação do LE.

#### **IV – LIMIAR DE DESEQUILÍBRIO DE HETEROZIGÓTIOS (LD)**

Para determinar este limiar, foram corridas no *SeqStudio™*, 112 amostras seguras de indivíduos diferentes, amplificadas com *GlobalFiler™*, das quais 29 eram de sangue, 82 de saliva e 1 correspondia ao controlo positivo da amplificação *DNA Control 007* incluído no *kit GlobalFiler™*.

Os EFs foram obtidos utilizando já como módulo no GM, o valor de LA determinado experimentalmente para o *SeqStudio™* e o LS do fabricante, e deles foram depois manualmente eliminados picos partidos, artefactos OL e eventuais picos contaminantes.

Seguidamente, foram registados em folhas de cálculo Excel - uma para cada marcador autossómico - todos os alelos observados nos *loci* heterozigóticos dos perfis obtidos e as alturas dos seus picos (RFUs) em cada amostra. Paralelamente, realizou-se o mesmo procedimento, mas contabilizando apenas os alelos em sistemas heterozigóticos que diferiam em mais do que 1 unidade de repetição, para evitar efeitos aditivos provocados pelos eventuais *stutter* - e + do alelo imediatamente seguinte ou anterior.

Foi depois calculado o valor, em percentagem, da proporção entre as alturas de cada par alélico, de todos os sistemas heterozigóticos encontrados nas amostras, utilizando a seguinte fórmula:

$$\text{Proporção de altura (\%)} = (\text{RFU}_{\text{pico menor}} / \text{RFU}_{\text{pico maior}}) \times 100$$

Por fim, foi calculado o valor final de LD para cada marcador genético com base nas médias dos valores de proporção de altura determinados para o mesmo marcador em todas as amostras, bem como os respetivos desvios padrão, seguindo a fórmula abaixo:

$$\text{LD}_{\text{marcador}} (\%) = \bar{x} - 3\sigma$$

- $\bar{x}$  = média dos valores de proporção de altura calculados para o marcador em questão
- $\sigma$  = desvio padrão associado aos valores acima mencionados

### 3.8.2. ESTUDOS COMPARATIVOS *SeqStudio™* vs. *3500*

As 112 amostras utilizadas para a determinação do LD (secção 3.8.1. – IV), foram igualmente aplicadas para fins comparativos, tendo sido corridas tanto no *SeqStudio™* como no *3500*, nas mesmas condições de eletroforese.

Os EFs foram obtidos utilizando como módulos de análise no GM, os limiares de sensibilidade correspondentes a cada equipamento - para o *SeqStudio™*, os limiares determinados no presente trabalho; e para o *3500*, os limiares estabelecidos pelo SGBF-C para utilização com o *GlobalFiler™*, na sua última validação (RV-SGBF-C-006, 2020): LA = 70, LE = 200, LS do fabricante presentes na **Tabela 9** – valores da coluna “3500” e LD = 50%.

Dos EFs obtidos foram eliminados manualmente picos partidos, artefactos OL e eventuais picos contaminantes.

Procedeu-se, por fim, à comparação dos perfis gerados pelos dois equipamentos de EC, para cada amostra, avaliando a correspondência (ou não) genotípica obtida.



## 4. RESULTADOS E DISCUSSÃO

### 4.1. LIMIAR ANALÍTICO (LA)

Das 64 amostras analisadas, observou-se uma que apresentava picos de ruído com valores atípicos, na ordem dos 500 e dos 1000 RFU, em todos os canais de cor, provavelmente resultantes da eventual presença de sujidade ou cristais de polímero que terão gerado disparos anómalos de fluorescência. Para evitar, a introdução de valores erróneos nos cálculos do LA, essa amostra não foi contabilizada nos cálculos. Sendo assim, para a determinação do LA foram analisados os EFs de 63 negativos de amplificação. No **ANEXO II**, encontra-se o perfil de ruído obtido para um dos canais de cor de uma das amostras, como exemplificativo do tipo de EFs analisados nesta secção.

A **Tabela 10** apresenta os valores de LA obtidos para cada canal de cor do *kit GlobalFiler™*, por cada um dos 3 métodos de cálculo utilizados. Segundo as *guidelines* do SWGDAM, o laboratório pode optar por um único valor de LA aplicável a todos os painéis de cor ou por um valor de LA para cada painel. Por questões de praticidade, o SGBF-C opta sempre por um LA aplicável a todo o EF. Pretendendo-se definir um limiar único para utilizar em todos os canais de cor, foi registado o valor mais elevado de LA observado em cada fórmula de cálculo (“LA<sub>final</sub>”).

**Tabela 10.** Valores de LA obtidos em cada painel de cor, utilizando o *kit GlobalFiler™*. Na coluna à direita estão os maiores valores de LA observados para cada uma das 3 fórmulas de cálculo utilizadas.

<i>GlobalFiler™</i>	AZUL 6-FAM™	VERDE VIC™	AMARELO NED™	VERMELHO TAZ™	ROXO SID™	LA <sub>final</sub> (RFU)
LA (IUPAC + Kaiser)	51,59	78,64	42,51	28,03	48,50	79
LA (SWGDAM)	92	142	194	110	96	194
LA (INTCF-B + GHEP-ISFG)	76	102	17	15	60	102

Após a análise dos resultados, verificou-se uma maior proximidade e concordância entre os valores obtidos para os diferentes painéis de cor pela fórmula de cálculo da IUPAC + Kaiser.

Tendo também em conta que um LA demasiado elevado aumenta o risco de descartar picos verdadeiramente alélicos que se encontrem com pouco sinal (por ex. devido a pertencerem a um contribuinte minoritário ou por simplesmente a amostra possuir pouco ADN), confundindo-os com ruído (falsos negativos), considerou-se que seria melhor optar por estabelecer um valor de LA mais conservador. O LA<sub>final</sub> mais baixo observado foi obtido segundo a abordagem IUPAC + Kaiser (79 RFU).

Além disso, o valor obtido por essa mesma abordagem, é o mais próximo do valor recomendado por defeito pelos fabricantes (50 RFU). Tendo em vista que o valor de LA<sub>final</sub> de 79 RFU, mesmo sendo o mais baixo observado, é superior ao tipicamente recomendado, não se colocam dúvidas sobre se o seu valor será demasiado baixo ao ponto de incluir demasiados artefactos e ruído de fundo ocasionalmente elevado (falsos positivos), que possam tornar a análise e interpretação dos picos pelo perito demasiado maçadora.

Sendo suficientemente elevado para filtrar picos de ruído, mas não demasiado elevado ao ponto de excluir informação alélica dos EFs. um LA de 79 RFU parece ser, portanto, adequado para distinguir com segurança, nas condições laboratoriais do SGBF-C, entre picos efetivamente produtos da PCR (alélicos ou não) e meros picos de ruído de fundo produzidos pelo próprio *SeqStudio™* e pelos reagentes aplicados.

Foi assim decido que se utilizaria o valor de LA resultante da abordagem IUPAC + Kaiser (79 RFU). Uma vez que a definição deste valor é flexível, o mesmo foi arredondado para 80 RFU.

Este valor (**LA = 80 RFU**) deverá então ser introduzido como módulo de análise no GM para todos os canais de cor, aquando da análise de EFs de amostras amplificadas com *GlobalFiler™* e corridas no *SeqStudio™*. Assim, aplicando este valor, serão apenas exibidos nos EFs, os picos detetados acima de 80 RFU; os restantes picos ( $\leq 80$  RFU) serão automaticamente descartados pelo GM, devido à elevada probabilidade de corresponderem a ruído de fundo produzido pelo próprio instrumento e pelos reagentes utilizados, evitando-se, desta forma, a introdução desnecessária de dúvidas na atribuição genotípica. Este limiar é um dos parâmetros mais importantes para a validação uma vez que todo o procedimento de validação foi realizado tendo como base este valor.

## 4.2. LIMAR DE STUTTER (LS)

Analisando os valores de LS estipulados pelo fabricante, presentes na **Tabela 9**, observou-se que o marcador D22S1045 possui o maior valor de LS, podendo concluir-se que os seus alelos produzirão produtos *stutter* mais elevados em RFU, que os restantes marcadores, em relação ao alelo principal. Tendo em conta que os alelos do D22S1045 possuem unidades de repetição trinucleotídicas, e que os restantes marcadores da tabela possuem motivos de repetição tetranucleotídicos (NIST, 2015), confirma-se que a ocorrência de *stutters* é efetivamente mais pronunciada, quanto menores forem as unidades de repetição, conforme referido na introdução e amplamente descrito em bibliografia. Comprovada a maior percentagem de formação de *stutter* em STRs trinucleotídicos comparativamente aos tetranucleotídicos, faz sentido que os fabricantes dos *kits* de amplificação para fins forenses optem pela utilização preferencial destes últimos, na tentativa de gerar perfis que induzam menores dúvidas interpretativas.

Além da observação acima, comprovou-se também, com base na comparação dos perfis das mesmas amostras obtidos no 3500 e no *SeqStudio™*, que os valores de LS recomendados pela *Applied Biosystems* para os marcadores autossómicos do *kit GlobalFiler™* no *SeqStudio™* (coluna “*SeqStudio™*” da **Tabela 9**) são efetivamente adequados para aplicar no GM como filtro de *stutter* no novo equipamento. Essa constatação encontra-se detalhadamente explicada na secção de resultados e discussão **4.5. Estudos Comparativos *SeqStudio™* vs. 3500.**

## 4.3. LIMAR ESTOCÁSTICO (LE)

Após extensiva análise dos EFs gerados pelo *SeqStudio™*, para cada diluição da Amostra A e B, elaboraram-se as tabelas abaixo para facilitar a visualização das situações de *drop-out*, e dos marcadores e concentrações de ADN<sub>total</sub> em que as mesmas ocorreram.

**Tabela 11.** Genótipos identificados em cada marcador contabilizado para os cálculos do LE, em cada uma das 6 diluições da Amostra A tratadas como amostras seguras (12,5 µL de volume de amplificação e 27 ciclos de PCR). Indicação dos resultados da quantificação da solução inicial de ADN extraído da Amostra A e das diluições a partir dela preparadas, bem como as concentrações esperadas para essas mesmas diluições.

Resultado quantificação Amostra A = 1,5139 ng/µl

12.5 µL/27 ciclos, 10s	Diluição 1		Diluição 2		Diluição 3		Diluição 4		Diluição 5		Diluição 6	
	A1 rep1	A1 rep2	A2 rep1	A2 rep2	A3 rep1	A3 rep2	A4 rep1	A4 rep2	A5 rep1	A5 rep2	A6 rep1	A6 rep2
ADN total (ng/µl) esperada vs. real	1,250 vs. 0,7328		0,625 vs. 0,4226		0,313 vs. 0,5416		0,156 vs. 0,0894		0,078 vs. 0,0462		0,039 vs. 0,0234	
Amelogenina	---	X	X	X	X	X	X	X	X	X	X	X
D3S1358	---	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	---	15, 17	15, 17	17	15
vWA	---	14, 17	14, 17	14, 17	14, 17	14, 17	14, 17	---	14, 17	14, 17	17	14, 17
D16S539	---	9	9	9	9	9	9	9	9	9	9	9
CSF1PO	---	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	---	10, 12	10, 12	10	10
TPOX	---	8, 11	8, 11	8, 11	8, 11	8, 11	8, 11	---	8, 11	8	11	---
D8S1179	---	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	13	13	13, 14	14	13, 14
D21S11	---	31, 32.2	31, 32.2	31, 32.2	31, 32.2	31, 32.2	31, 32.2	---	31, 32.2	31, 32.2	32.2	---
D18S51	---	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	---	---	17	19	17
D2S441	---	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	---	10, 11	10, 11	---	---
D19S433	---	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	---	13	12	---	---
TH01	---	6, 9.3	6, 9.3	6, 9.3	6, 9.3	6, 9.3	6, 9.3	---	6, 9.3	6	---	---
FGA	---	20, 22	20, 22	20, 22	20, 22	20, 22	20, 22	---	20	20	---	22
D22S1045	---	12, 17	12, 17	12, 17	12, 17	12, 17	12, 17	---	12, 17	12, 17	12	---
D5S818	---	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	---	12, 13	12	---	12
D13S317	---	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	---	13, 14	13, 14	---	---
D7S820	---	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	---	10, 12	10, 12	---	---
SE33	---	17, 28.2	17, 28.2	17, 28.2	17, 28.2	17, 28.2	17, 28.2	---	17	17, 28.2	---	---
D10S1248	---	13, 15	13, 15	13, 15	13, 15	13, 15	13, 15	---	13, 15	13, 15	13	13
D1S1656	---	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	17	15, 17	15, 17	17	---
D12S391	---	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20	20, 23	20, 23	---	---
D2S1338	---	22, 25	22, 25	22, 25	22, 25	22, 25	22, 25	---	22, 25	22, 25	22, 25	22
Alelos deletados	---	42	42	42	42	42	42	5	36	36	14	13

**Tabela 12.** Genótipos identificados em cada marcador contabilizado para os cálculos do LE, em cada uma das 6 diluições da Amostra B tratadas como amostras seguras (12,5 µL de volume de amplificação e 27 ciclos de PCR). Indicação dos resultados da quantificação da solução inicial de ADN extraído da Amostra B e das diluições a partir dela preparadas, bem como as concentrações esperadas para essas mesmas diluições.

Resultado quantificação Amostra B = 2,3871 ng/µl

12.5 µL/27 ciclos, 10s	Diluição 1		Diluição 2		Diluição 3		Diluição 4		Diluição 5		Diluição 6	
	B1 rep1	B1 rep2	B2 rep1	B2 rep2	B3 rep1	B3 rep2	B4 rep1	B4 rep2	B5 rep1	B5 rep2	B6 rep1	B6 rep2
ADN total (ng/µl) esperada vs. real	1,250 vs. 1,0227		0,625 vs. 0,4272		0,313 vs. 0,2090		0,156 vs. 0,1253		0,078 vs. 0,0489		0,039 vs. 0,0261	
Amelogenina	XY	---	XY	---	XY	XY	XY	XY	---	XY	Y	X
D3S1358	15, 16	---	15, 16	---	15, 16	15, 16	15, 16	15, 16	---	15, 16	15, 16	---
vWA	16, 17	---	16, 17	---	16, 17	16, 17	16, 17	16, 17	---	16, 17	17	17
D16S539	10, 11	---	10, 11	---	10, 11	10, 11	10, 11	10, 11	---	10, 11	11	10
CSF1PO	11, 12	---	11, 12	---	11, 12	11, 12	11, 12	11, 12	---	11	11, 12	12
TPOX	9, 12	---	9, 12	---	9, 12	9, 12	9, 12	9, 12	---	12	12	---
D8S1179	8, 12	---	8, 12	---	8, 12	8, 12	8, 12	8, 12	---	8, 12	---	12
D21S11	28, 31.2	---	28, 31.2	---	28, 31.2	28, 31.2	28, 31.2	28, 31.2	---	28, 31.2	---	28, 31.2
D18S51	12, 15	---	12, 15	---	12, 15	12, 15	12, 15	12, 15	---	12	12	12, 15
D2S441	10, 11	---	10, 11	---	10, 11	10, 11	10, 11	10, 11	---	---	---	---
D19S433	13, 15	---	13, 15	---	13, 15	13, 15	13, 15	13, 15	---	13	---	---
TH01	9, 9.3	---	9, 9.3	---	9, 9.3	9, 9.3	9, 9.3	9, 9.3	---	---	9.3	---
FGA	21, 22	---	21, 22	---	21, 22	21, 22	21, 22	21, 22	---	21	---	---
D22S1045	15, 16	---	15, 16	---	15, 16	15, 16	15, 16	15, 16	---	15	16	15, 16
D5S818	12, 13	---	12, 13	---	12, 13	12, 13	12, 13	12, 13	---	12, 13	12	12
D13S317	8, 11	---	8, 11	---	8, 11	8, 11	8, 11	8, 11	---	8, 11	---	8, 11
D7S820	9, 10	---	9, 10	---	9, 10	9, 10	9, 10	9, 10	---	9	9, 10	---
SE33	19, 23.2	---	19, 23.2	---	19, 23.2	19, 23.2	19, 23.2	19, 23.2	---	19, 23.2	---	23.2
D10S1248	14, 15	---	14, 15	---	14, 15	14, 15	14, 15	14, 15	---	14, 15	14, 15	14, 15
D1S1656	12, 13	---	12, 13	---	12, 13	12, 13	12, 13	12, 13	---	12, 13	12	12
D12S391	18, 24	---	18, 24	---	18, 24	18, 24	18, 24	18, 24	---	18, 24	24	18, 24
D2S1338	16, 19	---	16, 19	---	16, 19	16, 19	16, 19	16, 19	---	19	16, 19	16, 19
Alelos deletados	44	---	44	---	44	44	44	44	---	32	20	22

**Tabela 13.** Genótipos identificados em cada marcador contabilizado para os cálculos do LE, em cada uma das 6 diluições da Amostra A tratadas como amostras problema (25 µL de volume de amplificação e 29 ciclos de PCR). Indicação dos resultados da quantificação da solução inicial de ADN extraído da Amostra A e das diluições a partir dela preparadas, bem como as concentrações esperadas para essas mesmas diluições.

Resultado quantificação Amostra A = 1,5139 ng/µl

25 µL/29 ciclos, 10s	Diluição 1		Diluição 2		Diluição 3		Diluição 4		Diluição 5		Diluição 6	
	A1 rep1	A1 rep2	A2 rep1	A2 rep2	A3 rep1	A3 rep2	A4 rep1	A4 rep2	A5 rep1	A5 rep2	A6 rep1	A6 rep2
ADN total (ng/µl) esperada vs. real	1,250 vs. 0,7328		0,625 vs. 0,4226		0,313 vs. 0,5416		0,156 vs. 0,0894		0,078 vs. 0,0462		0,039 vs. 0,0234	
Amelogenina	X	X	X	---	X	X	X	X	X	X	---	X
D3S1358	15, 17	15, 17	15, 17	---	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	---	15, 17
vWA	14, 17	14, 17	14, 17	---	14, 17	14, 17	14, 17	14, 17	14, 17	14, 17	---	14, 17
D16S539	9	9	9	---	9	9	9	9	9	9	---	9
CSF1PO	10, 12	10, 12	10, 12	---	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	---	10, 12
TPOX	8, 11	8, 11	8, 11	---	8, 11	8, 11	8, 11	8, 11	8, 11	8, 11	---	8, 11
D8S1179	13, 14	13, 14	13, 14	---	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	---	13
D21S11	31, 32.2	31, 32.2	31, 32.2	---	31, 32.2	31, 32.2	31, 32.2	31, 32.2	31, 32.2	31, 32.2	---	31, 32.2
D18S51	17, 19	17, 19	17, 19	---	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	---	19
D2S441	10, 11	10, 11	10, 11	---	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	---	10, 11
D19S433	12, 13	12, 13	12, 13	---	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	---	12, 13
TH01	6, 9.3	6, 9.3	6, 9.3	---	6, 9.3	6, 9.3	6, 9.3	6, 9.3	6, 9.3	6, 9.3	---	6, 9.3
FGA	20, 22	20, 22	20, 22	---	20, 22	20, 22	20, 22	20, 22	20, 22	20, 22	---	20, 22
D22S1045	12, 17	12, 17	12, 17	---	12, 17	12, 17	12, 17	12, 17	12	12, 17	---	12, 17
D5S818	12, 13	12, 13	12, 13	---	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	---	12, 13
D13S317	13, 14	13, 14	13, 14	---	13, 14	13, 14	13, 14	13, 14	13, 14	13	---	13, 14
D7S820	10, 12	10, 12	10, 12	---	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	---	---
SE33	17, 28.2	17, 28.2	17, 28.2	---	17, 28.2	17, 28.2	17, 28.2	17, 28.2	17, 28.2	17	---	17, 28.2
D10S1248	13, 15	13, 15	13, 15	---	13, 15	13, 15	13, 15	13, 15	13, 15	13, 15	---	13, 15
D1S1656	15, 17	15, 17	15, 17	---	15, 17	15, 17	15, 17	15, 17	15, 17	15	---	15
D12S391	20, 23	20, 23	20, 23	---	20, 23	20, 23	20, 23	20, 23	---	20, 23	---	20
D2S1338	22, 25	22, 25	22, 25	---	22, 25	22, 25	22, 25	22, 25	22, 25	22	---	22, 25
Alelos detetados	42	42	42	---	42	42	42	42	39	37	0	42

**Tabela 14.** Genótipos identificados em cada marcador contabilizado para os cálculos do LE, em cada uma das 6 diluições da Amostra B tratadas como amostras problema (25 µL de volume de amplificação e 29 ciclos de PCR). Indicação dos resultados da quantificação da solução inicial de ADN extraído da Amostra B e das diluições a partir dela preparadas, bem como as concentrações esperadas para essas mesmas diluições.

Resultado quantificação Amostra B = 2,3871 ng/µl

25 µL/29 ciclos, 10s	Diluição 1		Diluição 2		Diluição 3		Diluição 4		Diluição 5		Diluição 6	
	B1 rep1	B1 rep2	B2 rep1	B2 rep2	B3 rep1	B3 rep2	B4 rep1	B4 rep2	B5 rep1	B5 rep2	B6 rep1	B6 rep2
ADN total (ng/µl) esperada vs. real	1,250 vs. 1,0227		0,625 vs. 0,4272		0,313 vs. 0,2090		0,156 vs. 0,1253		0,078 vs. 0,0489		0,039 vs. 0,0261	
Amelogenina	XY	---	XY	---	XY	XY	XY	XY	XY	XY	XY	XY
D3S1358	15, 16	---	15, 16	---	15, 16	15, 16	15, 16	15, 16	15, 16	15, 16	15, 16	15, 16
vWA	16, 17	---	16, 17	---	16, 17	16, 17	16, 17	16, 17	16, 17	16, 17	16, 17	16
D16S539	10, 11	---	10, 11	---	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	---	10
CSF1PO	11, 12	---	11, 12	---	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11
TPOX	9, 12	---	9, 12	---	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12
D8S1179	8, 12	---	8, 12	---	8, 12	8, 12	8, 12	8, 12	8, 12	8, 12	8, 12	8, 12
D21S11	28, 31.2	---	28, 31.2	---	28, 31.2	28, 31.2	28, 31.2	28, 31.2	28, 31.2	28, 31.2	28, 31.2	28, 31.2
D18S51	12, 15	---	12, 15	---	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15	15
D2S441	10, 11	---	10, 11	---	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11
D19S433	13, 15	---	13, 15	---	13, 15	13, 15	13, 15	13, 15	13, 15	13, 15	13, 15	13, 15
TH01	9, 9.3	---	9, 9.3	---	9, 9.3	9, 9.3	9, 9.3	9, 9.3	9, 9.3	9, 9.3	9, 9.3	9, 9.3
FGA	21, 22	---	21, 22	---	21, 22	21, 22	21, 22	21, 22	21, 22	21, 22	21, 22	22
D22S1045	15, 16	---	15, 16	---	15, 16	15, 16	15, 16	15, 16	15, 16	15, 16	15, 16	15, 16
D5S818	12, 13	---	12, 13	---	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13
D13S317	8, 11	---	8, 11	---	8, 11	8, 11	8, 11	8, 11	8, 11	8, 11	8, 11	11
D7S820	9, 10	---	9, 10	---	9, 10	9, 10	9, 10	9, 10	9	9, 10	9, 10	9, 10
SE33	19, 23.2	---	19, 23.2	---	19, 23.2	19, 23.2	19, 23.2	19, 23.2	19, 23.2	19, 23.2	19, 23.2	19, 23.2
D10S1248	14, 15	---	14, 15	---	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15
D1S1656	12, 13	---	12, 13	---	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13
D12S391	18, 24	---	18, 24	---	18, 24	18, 24	18, 24	18, 24	18, 24	18, 24	18, 24	24
D2S1338	16, 19	---	16, 19	---	16, 19	16, 19	16, 19	16, 19	16, 19	16, 19	16, 19	16, 19
Alelos detetados	44	---	44	---	44	44	44	44	43	44	41	37

O facto de em algumas diluições iniciais não ter sido detetado qualquer alelo, deve-se a erros de pipetagem. Contudo, houve sempre um duplicado da mesma diluição contendo todos os alelos, compensando a falta de informação genética da outra réplica e garantindo a manutenção de um perfil completo.

Na diluição A4 rep2 tratada como amostra segura, apesar de terem sido identificados poucos alelos, conclui-se que esta situação se deverá também a um erro de pipetagem tendo em conta que no seu duplicado (A4 rep1) surgiam ainda todos os alelos e que vários dos supostos alelos desaparecidos nesta amostra ressurgiram depois na diluição seguinte (A5).

Assim, com base na análise das tabelas acima apresentadas, facilmente se percebe que situações de *drop-out* começaram a ocorrer a partir da Diluição 5, em ambas as amostras A e B tratadas tanto como seguras como problema, o que demonstra que os efeitos estocásticos se fazem sentir de modo semelhante em ambas as tipologias de processamento e que começam a surgir sensivelmente a partir de concentrações de ADN total de 0,078 ng/μL. Tendo esta evidência em consideração e consultando os valores de concentração obtidos na fase de quantificação, conclui-se que o *SeqStudio™* é capaz de gerar perfis completos para amostras contendo até uma concentração mínima de ADN total de sensivelmente 0.0894ng/μL, concentração essa bem menor que a tipicamente recomendada para obter perfis completos (0.5 – 2.5 ng/μL).

Através das tabelas elaboradas (**Tabelas 11, 12, 13 e 14**), e por forma a seguir a abordagem da SWGDAM, consultaram-se os EFs das diluições maiores sem ocorrência de *drop-out* (Diluições 4), registaram-se as maiores intensidades de sinal (RFU) detetadas nos marcadores heterozigóticos onde ocorreu *drop-out* na diluição seguinte e calculou-se como LE, o valor médio, em RFU, para cada amostra, e para cada tipologia de processamento laboratorial. Obtiveram-se assim os valores de LE abaixo apresentados na **Tabela 15**.

**Tabela 15.** Valores de LE obtidos para a Amostra A e B nas condições laboratoriais de amostras seguras (12,5μL, 27 ciclos) e problema (25μL, 29 ciclos).

<i>GlobalFiler™</i>			
Vol. Amplificação, nr. ciclos PCR	Média (RFU) Amostra A	Média (RFU) Amostra B	Média (RFU) por tratamento laboratorial
12,5μL, 27 ciclos	418	252	335
25μL, 29 ciclos	375	332	354
<b>Média total (RFU)</b>	344		

Comparando o valor de LE correspondente ao processamento laboratorial tipicamente aplicado a amostras seguras (335 RFU) com o correspondente ao do tratamento aplicado nas problema (354 RFU), não se notabiliza uma discrepância significativa que justifique o estabelecimento de valores de LE individuais para cada tipo de condição laboratorial a que se recorre no SGBF-C. Além disso, o SGBF-C tem vindo a adotar um único valor de LE, para efeitos de análise e interpretação de perfis de amostras amplificadas com o mesmo *kit* e no mesmo equipamento de EC. Posto isto, o valor de LE a ser estabelecido no SGBF-C para amostras amplificadas com *GlobalFiler™* e corridas no *SeqStudio™*, deverá ser a média total dos LE calculados, 344 RFU. Sendo a definição deste valor flexível e por forma a estabelecer um valor que seja numericamente mais fácil de instituir no processo de análise e interpretação de perfis, o mesmo foi arredondado para 340 RFU. Apesar dos valores tipicamente definidos para LE rondarem, de acordo com a bibliografia, os 150 e os 200 RFU, o *SeqStudio™*, por se tratar de um equipamento recente (2017), será à partida mais sensível que os equipamentos de EC até aí existentes, o que justificará o estabelecimento de um LE superior.

Este valor (**LE = 340 RFU**) deverá assim:

- ser aplicado como módulo de análise no GM para todos os canais de cor dos EFs aquando da análise de amostras amplificadas com *GlobalFiler™* e corridas no *SeqStudio™*. Desta forma, sempre que o GM detetar picos com intensidade de sinal  $\leq$  340 RFU, num determinado marcador, acusará um sinalizador de qualidade para alertar o operador quanto à possibilidade daquele *locus* ter sofrido acentuados efeitos estocásticos que tenham levado ao *drop-out* de um eventual alelo irmão;
- ser utilizado como critério de interpretação nos perfis gerados pelo *SeqStudio™* para amostras amplificadas com *GlobalFiler™*, particularmente em situações onde os efeitos estocásticos são mais significativos (amostras com ADN vestigial e/ou misturas com contribuintes minoritários), para auxiliar o perito a determinar, com confiança fundamentada, o genótipo completo ou parcial das mesmas. Assim, caso um pico alélico, num determinado sistema heterozigótico, possua  $>$  340 RFU poderá admitir-se com segurança suficiente que não ocorreu *drop-out* de um eventual alelo irmão, pois acima desse valor uma perda alélica será improvável, pelo que a presença de um só alelo, no *locus*, representará, seguramente, um genótipo homozigótico. Caso o pico alélico detetado possua  $\leq$  340 RFU (encontrando-se na “zona cinzenta”) o perito deve ter em atenção que um eventual alelo irmão poderá ter sofrido *drop-out*, pelo que o genótipo observado no *locus* em questão poderá estar incompleto, ficando apenas parcialmente concluído para aquele marcador. É de notar que em perfis genéticos

singulares (derivados de amostras seguras), existindo 2 alelos abaixo do LE num determinado *locus*, o seu genótipo para esse marcador poderá ser garantidamente assumido como heterozigótico, o que não é possível nos perfis de amostras problemas.

#### 4.4. LIMIAR DE DESEQUILÍBRIO DE HETEROZIGÓTICOS (LD)

Os valores de LD final calculados para cada marcador genético são apresentados na **Tabela 16**.

**Tabela 16.** Valores de LD obtidos para cada marcador autossômico, contabilizando todos os picos alélicos dos sistemas heterozigóticos encontrados nos perfis (à esquerda) vs. contabilizando apenas os picos alélicos que diferiam em mais do que 1 unidade de repetição (à direita).

<b>GlobalFiler™</b>		
<b>Sistema</b>	<b>LD (%)</b>	
	<b>Todos os pares alélicos</b>	<b>Apenas pares alélicos com mais do que 1 unidade de repetição entre si</b>
<b>D3S1358</b>	75.32	73.42
<b>vWA</b>	71.41	69.47
<b>D16S539</b>	70.94	68.83
<b>CSF1PO</b>	71.24	65.80
<b>TPOX</b>	73.01	73.48
<b>D8S1179</b>	73.78	73.11
<b>D21S11</b>	73.72	74.73
<b>D18S51</b>	71.35	70.33
<b>D2S441</b>	76.40	76.93
<b>D19S433</b>	76.45	74.06
<b>TH01</b>	78.89	78.32
<b>FGA</b>	78.12	77.04
<b>D22S1045</b>	69.52	65.98
<b>D5S818</b>	67.66	76.29
<b>D13S317</b>	76.88	76.33
<b>D7S820</b>	76.36	75.13
<b>SE33</b>	66.92	65.69
<b>D10S1248</b>	70.43	68.19
<b>D1S1656</b>	73.58	73.90
<b>D12S391</b>	71.68	69.19
<b>D2S1338</b>	69.96	70.29



Os valores obtidos, em ambas as situações, são consistentes com o descrito em bibliografia para amostras seguras (um único contribuinte e quantidade adequada de ADN), uma vez que se observam valores sempre superiores a 60%, transparecendo um normal desbalanceamento entre alelos num mesmo *locus*.

Numa avaliação mais profunda, observa-se que o valor mais elevado de LD surge associado ao marcador TH01, o que faz sentido tendo em conta que o seu intervalo de possibilidades alélicas (i.e., de polimorfismos) é o menor dos marcadores autossómicos amplificados pelo *GlobalFiler™* (ver **Figura 7**), acabando os alelos irmãos nele presentes por ser frequentemente muito próximos em termos de comprimento e, portanto, com maior semelhança de alturas entre os seus picos (maior equilíbrio alélico), uma vez que a Polimerase os consegue amplificar de modo semelhante.

De maneira oposta, verifica-se que o valor mais baixo de LD surge associado ao marcador SE33, o que também faz sentido tendo em conta tratar-se de um dos marcadores mais polimórfico do *GlobalFiler™* (ver **Figura 7**), acabando por se observar maior variedade e distanciamento em termos de comprimento, entre alelos irmãos. Ora quanto mais afastados eles forem, a Polimerase conseguirá mais eficientemente amplificar um alelo (o menor) em detrimento do outro, provocando uma maior discrepância entre as alturas dos seus picos nos EFs (menor balanceamento alélico).

Averigua-se também que os valores de LD calculados contabilizando as intensidades de sinal de todos os alelos irmãos (incluindo os numericamente seguidos), apesar de relativamente semelhantes, são tendencialmente superiores aos dos que foram calculados levando apenas em consideração as proporções entre os RFUs de alelos irmãos distanciados por mais do que 1 unidade de repetição. Esta apuração faz sentido uma vez que, alelos irmãos que difiram entre si por 1 unidade de repetição, estão sujeitos a apresentar valores de RFU artificialmente elevados pelo eventual *stutter* + do alelo imediatamente anterior e/ou pelo eventual *stutter* - do alelo imediatamente seguinte, o que se traduzirá em valores de LD falsamente superiores, ao incluir nos cálculos estes sistemas heterozigóticos. Assim sendo, consideram-se mais corretos os valores de LD obtidos contabilizando apenas pares de picos alélicos com mais do que 1 unidade de repetição de diferença (coluna direita da **Tabela 16**).

Segundo as recomendações da (SWGAM, 2010, 2017) cada laboratório pode optar tanto por aplicar no seu *software* de análise de EFs e como critério de interpretação de perfis, valores de LD individualmente para cada *locus*, ou alternativamente, um único valor de LD aplicável a todos os *loci*. Por questões de praticidade, e para diminuir a morosidade na análise

de perfis, o SGBF-C tem vindo a optar pelo estabelecimento de um valor singular de LD. Posto isto, foi decidido que se utilizaria o valor de LD mínimo obtido (o correspondente ao marcador SE33 na **Tabela 16**), uma vez que foi esse o valor que traduziu o desbalanceamento máximo possível entre picos alélicos de um *locus* heterozigótico, que poderá ser aceite e expectável como “normal”, mesmo sendo alvo de efeitos estocásticos, em amostras de um único interveniente.

Tendo em vista que o valor calculado contabilizando apenas os alelos distanciados por mais do que 1 unidade de repetição será o mais correto, estabeleceu-se como LD final 65.69%. Sendo a definição deste valor flexível e para facilitar o processo de interpretação de perfis, o mesmo foi arredondado para 65%.

Este valor (**LD = 65%**) deverá assim:

- ser aplicado como módulo de análise no GM para todos os marcadores autossómicos, aquando da análise de EFs de amostras amplificadas com *GlobalFiler™* e corridas no *SeqStudio™*. Desta forma, sempre que o GM detetar uma proporção de altura entre picos alegadamente alélicos  $\leq 65\%$  (ou um desnível  $\geq 35\%$ ), num determinado marcador, acusará um sinalizador de qualidade para alertar o operador quanto à possibilidade dos alelos detetados pertencerem a diferentes contribuintes, ao invés de a um só indivíduo heterozigótico. O perito deve depois proceder à avaliação do perfil, tendo em atenção que nem sempre quando se verificam picos com proporção de alturas  $\leq$  LD estabelecido, se tratará automaticamente de um perfil de mistura, pois essa desproporção anormal entre picos pode também ser devida à presença de ADN em quantidades vestigiais na amostra (sujeita a efeitos estocásticos mais acentuados), ADN altamente degradado (com amplificação preferencial dos alelos menores) ou a eventuais problemas associados à fase de PCR como presença de inibidores ou fraca ligação de *primers*;
- ser utilizado como critério de interpretação de perfis gerados pelo *SeqStudio™* para amostras amplificadas com *GlobalFiler™* por forma a auxiliar o perito a concluir, com confiança - caso as informações circunstanciais da situação forense o justificarem - a presença (ou não) de um perfil de mistura com contribuições desproporcionais entre indivíduos e a inferir o genótipo do contribuinte maioritário vs. o do contribuinte minoritário. É de notar, que em caso de mistura com uma contribuição quantitativa semelhante entre intervenientes, o LD definido já não permitirá uma distinção entre

eles, pois a semelhança nas intensidades de sinal dos picos, permite várias combinações alélicas porque todas elas apresentarão proporção entre picos > 65%.

É de salientar que o LD=65% será apenas aplicável a picos alélicos que ultrapassam o valor definido para o LE (no caso, LE = 340 RFU), pois abaixo desse valor não é possível afirmar com segurança se num certo *locus*, um determinado pico alélico terá (ou não) perdido o seu alelo irmão e, portanto, não se conseguirá aferir com confiança o genótipo do(s) indivíduo(s) para esse marcador.

#### 4.5. ESTUDOS COMPARATIVOS *SeqStudio™* vs. 3500

Em ambos os equipamentos, o resultado de tipagem do controlo positivo da amplificação *DNA Control 007* incluído no *kit GlobalFiler™* para fins de controlo interno da qualidade, mostrou-se igual ao alegado como correto pelo fabricante (**ANEXO III**). Isto prova, não só que a etapa de amplificação funcionou eficientemente, como também demonstra a produção, em ambos os sequenciadores, de genotipagens fiáveis nas restantes amostras (o *ladder* alélico utilizado estaria funcional).

Após comparação dos perfis gerados pelo *SeqStudio™* com os do 3500, para as mesmas amostras, verificou-se que todos eles obtiveram as mesmas atribuições alélicas em todos os marcadores, não havendo, em nenhum deles, qualquer designação alélica de picos discrepante ou inexistente. Obtiveram-se, portanto, perfis concordantes e completos, para todas as amostras, bem como uma boa resolução de picos conforme esperado, tendo em conta que foram corridas unicamente amostras seguras.

Confirmada a total correspondência genotípica entre os perfis do *SeqStudio™* e do 3500, comprova-se a capacidade do *SeqStudio™* em obter resultados de tipagem equivalentes aos do equipamento de EC atualmente em utilização no SGBF-C. Assumindo os perfis gerados pelo 3500 como corretos visto tratar-se de um equipamento já devidamente validado, e tendo em vista a correta genotipagem do controlo positivo por ambos os equipamentos e a reprodutibilidade dos perfis genéticos evidenciada, verifica-se o correto funcionamento do *SeqStudio™*, sendo este, como tal, igualmente capaz de gerar resultados de tipagem confiáveis, válidos e verdadeiros.

Verificou-se também que os EFs gerados pelo *SeqStudio™* apresentam picos com intensidade de sinal (RFU) consistentemente superior relativamente à que surge nos EFs do 3500 para os mesmos fragmentos, o que prova que o *SeqStudio™* apresenta maior sensibilidade de deteção de fragmentos. Esta constatação é também corroborada pelo facto dos limiares de sensibilidade determinados no presente trabalho para o *SeqStudio™*, nomeadamente o LA = 80 e o LE = 340, se mostrarem superiores aos pré-validados para o 3500 (LA = 70 e LE = 200).

Abaixo, nas **Figuras 20 e 21**, são apresentados, como exemplo da concordância genotípica observada, da qualidade dos perfis obtidos e da maior intensidade de sinal associada ao *SeqStudio™*, os perfis gerados pelo *SeqStudio™* e pelo 3500, para uma mesma amostra.

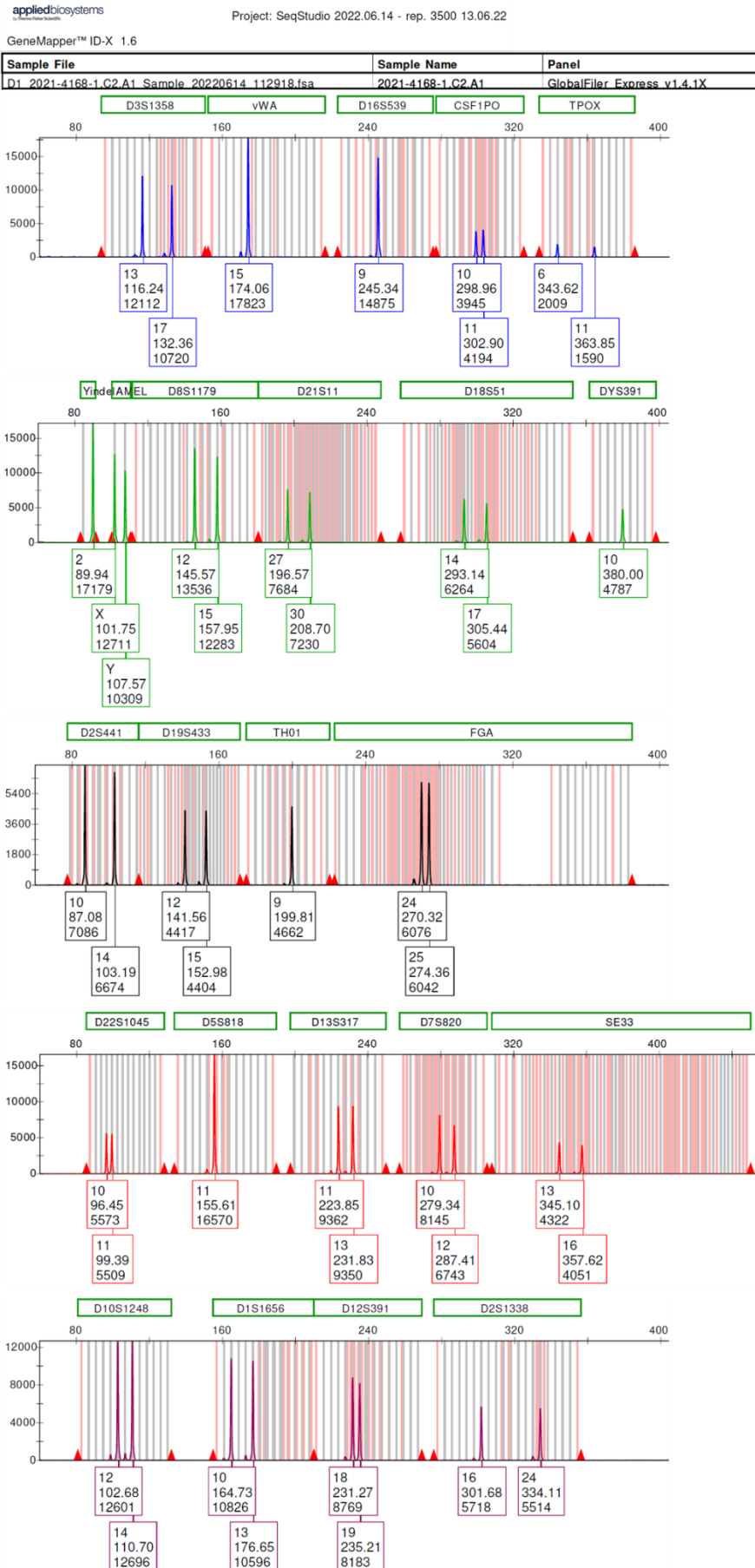


Figura 20. Perfil gerado pelo *SeqStudio™*, para a amostra anonimizada como 2021-4168-1.C2.A1.

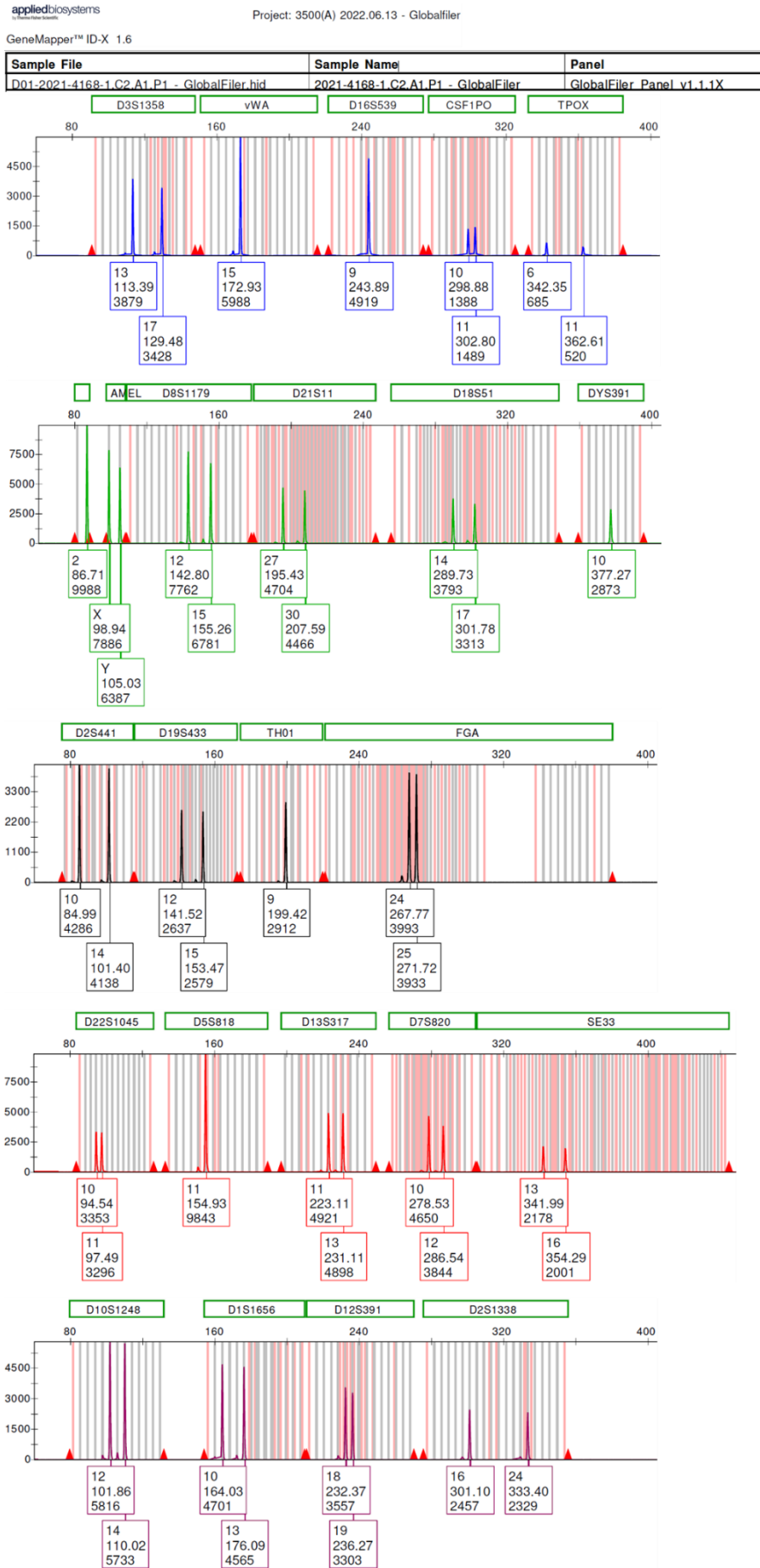
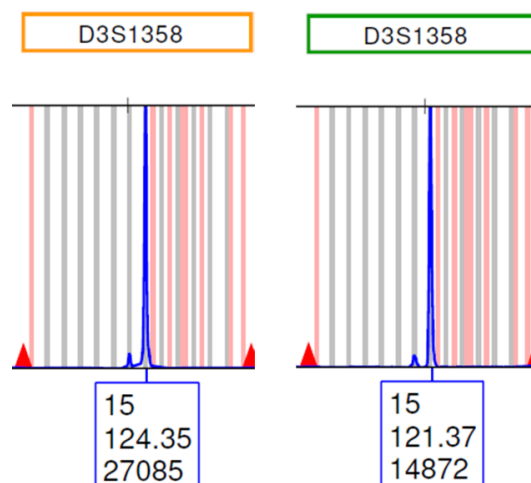


Figura 21. Perfil gerado pelo 3500, para a amostra anonimizada como 2021-4168-1.C2.A1.

Os retângulos coloridos que envolvem o nome dos marcadores nos perfis, representam um sistema de avaliação de qualidade genotípica atribuída automaticamente pelo GM a cada *locus*, depois das filtragens efetuadas pelos módulos de análise, com base em parâmetros como excesso ou ausência de alelos, formato dos picos, existência de OL's, picos OS e *spikes* e presença de desequilíbrio heterozigótico face ao LD estabelecido. A cada um destes parâmetros, o GM atribui um determinado valor de penalização (definido pelo fabricante), que se reflete depois em sinalizadores de qualidade genotípica (*quality flags*) de cor verde em caso de qualidade considerada aceitável; amarela em situações de qualidade intermédia; vermelha em caso de qualidade não aceitável; e cinzenta caso a qualidade atribuída tenha sido intermédia ou não aceitável, mas que após revisão daquele polimorfismo pelo perito, foi realizada alguma alteração (exemplo: descarte de um pico que recebeu atribuição alélica, por se considerar que corresponde a um artefacto) (**Figura 23** – sinalizador à esquerda).

A maior sensibilidade do equipamento *SeqStudio™* aliada à utilização de amostras com elevada quantidade de ADN (seguras), levou por vezes à saturação do seu detetor ( $\approx 25\ 000 - 30\ 000$  RFU) (Thermo Fisher Scientific, 2019) com picos OS por excesso de produto amplificado a atravessar o laser, o que fez com que o GM atribuisse a essas situações um sinalizador de qualidade amarelo ou vermelho (**Figura 22**). Este facto poderia justificar eventuais futuras adaptações de volume de amostras seguras a aplicar no equipamento ou diluições das mesmas. Contudo, tratando-se de uma situação que ocorrerá unicamente em amostras seguras - que raramente suscitam dúvidas relativamente ao seu genótipo e qualidade de perfil - o surgimento destes sinalizadores não constitui problema significativo.



**Figura 22.** Alelo 15 detetado no marcador D3S1358 pelo *SeqStudio™* (à esquerda) vs. pelo 3500, para a amostra anonimizada como 2021-1958-1.C1.X1.A1. No EF do *SeqStudio™*, o GM atribuiu um sinalizador de qualidade laranja devido à intensidade de sinal ser demasiado elevada e próxima do limite de saturação do detetor.

Excetuando as ocasiões de saturação fluorescente e de picos partidos (de normal ocorrência devido à elevada quantidade de ADN) e a presença ocasional de picos OL's, os restantes sinalizadores de qualidade atribuídos aos EFs do *SeqStudio™*, foram maioritariamente verdes, transparecendo uma boa qualidade dos perfis por ele gerados.

Através da realização dos estudos comparativos, conclui-se também uma aparente adequação dos limiares estabelecidos para o *SeqStudio™*, no presente trabalho, como módulo de análise no GM e de interpretação de perfis em amostras amplificadas com *GlobalFiler™*.

### **Discussão relativa ao LA atribuído ao *SeqStudio™***

Não foi detetada qualquer ausência alélica, pelo que o LA definido (80 RFU) não estará demasiado elevado ao ponto de eliminar picos verdadeiramente alélicos. Não foram excessivamente detetados picos inespecíficos, pelo que o ruído aparenta ser bem filtrado através do LA estabelecido.

### **Discussão relativa ao LE atribuído ao *SeqStudio™***

Tendo em conta que se obtiveram perfis completos para todas as amostras e que todos os picos verdadeiramente alélicos presentes nos perfis gerados pelo *SeqStudio™* apresentaram intensidades de sinal sempre acima do LE definido (340 RFU), pode admitir-se que esse valor permite afirmar com certeza que acima dele não ocorreu *drop-out* alélico num *locus* heterozigótico, e que, portanto, a presença de um único alelo, acima de 340 RFU, comprova a homozigotia daquele *locus*, não se colocando a hipótese do perfil estar parcialmente exposto naquele marcador.

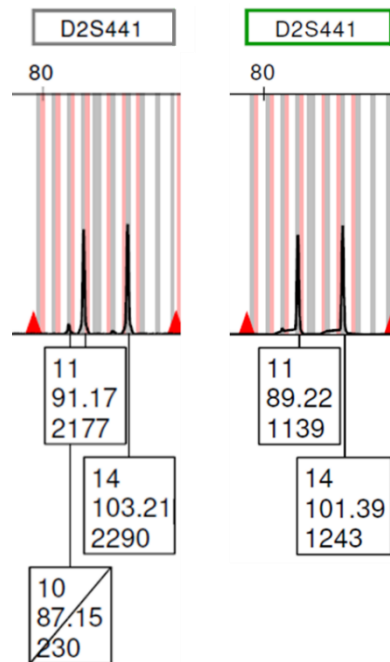
### **Discussão relativa aos LS aplicados no *SeqStudio™***

Em continuidade com a discussão da secção **4.2. Limiar de Stutter (LS)**, os estudos comparativos permitiram obter conclusões relativas à utilização dos LS recomendados pela *Applied Biosystems*, como filtro de *stutter* no *SeqStudio™* (coluna "*SeqStudio™*" da **Tabela 9**).

Aquando da análise dos EFs obtidos, notou-se a presença de um terceiro pico identificado como alelo pelo GM, no marcador D2S441 de uma amostra (**Figura 23**) corrida



no *SeqStudio™*. Sendo que todas as amostras utilizadas nos estudos provinham de um único contribuinte, apenas se poderiam identificar, no máximo, 2 alelos em cada *locus* (descartando-se a possibilidade de situações de trissomia). Visto que o pico se encontrava em posição *stutter*, possuía uma altura claramente discrepante relativamente aos outros dois alelos, incompatibilizando a possibilidade de constituir par alélico com qualquer um deles e descartando a possibilidade de trissomia no *locus*, constatou-se que o pico corresponderia a um *stutter* - que não foi eficientemente filtrado pelo filtro de *stutter* introduzido no GM. Sendo a proporção de *stutter* do pico em questão relativamente ao alelo real (230 RFU / 2177 RFU)  $\times 100 = 10.56\%$ , e o LS estabelecido para o marcador D2S441, apenas 8.10%, faz sentido que o GM o tenha identificado alélicamente uma vez que se encontrava acima do LS estabelecido. No EF da mesma amostra corrida no 3500 não ocorreu esta atribuição alélica errónea (o filtro de *stutter* aplicado para este equipamento é adequado e está devidamente validado como tal). Tendo em vista que este pico meramente *stutter* poderia erroneamente ser atribuído a um alelo de um contribuinte minoritário, caso não houvesse confiança suficiente de que a amostra em questão era proveniente um único contribuinte, talvez seja importante determinar o LS para o marcador D2S441, de modo experimental, no *SeqStudio™*, em amostras amplificadas com *GlobalFiler™*, para melhor estabelecê-lo face às condições laboratoriais do SGBF-C. Contudo, por se ter tratado de uma situação única e isolada (não ocorreu de forma recorrente nas restantes 111 amostras, no mesmo marcador, nem em qualquer outro *locus*), podendo ter sido fruto eventualmente de uma contaminação esporádica (situação de *drop-in*), não se considerou um problema significativo.



**Figura 23.** Picos identificados alélicamente pelo GM, no marcador D2S441, no EF gerado pelo *SeqStudio™* (à esquerda) vs. pelo 3500, para a amostra anonimizada como 2022-1826-1.C1.A1. O pico identificado como “alelo 10” pelo GM, corresponderá a um *stutter* - do alelo 11, que não foi filtrado.

Assim, tendo em vista que as atribuições alélicas foram amplamente equivalentes entre os perfis gerados no 3500 e no *SeqStudio™*, terão sido eliminados os mesmos *stutters*, apesar dos valores de LS utilizados como módulo de análise no GM serem diferentes entre os sequenciadores, para cada marcador genético. Confirma-se então, que a utilização dos valores de LS recomendados pelo fabricante indicados na coluna “*SeqStudio™*” da **Tabela 9** geram uma filtragem de *stutters* correta e de confiança. Posto isto, os valores tabelados concluem-se adequados como filtro de *stutter* para o novo equipamento, nas condições laboratoriais do SGBF-C, podendo, portanto, ser futuramente aplicados no *software* aquando da análise de EFs de amostras de rotina amplificadas com *GlobalFiler™* e corridas no *SeqStudio™*.

Desta forma, quando forem detetados pelo *SeqStudio™*, picos em posição *stutter* (i.e., com menos ou mais 1 unidade de repetição que um pico garantidamente alélico, no mesmo *locus*) cuja proporção de altura relativamente ao alelo seguinte seja inferior ou igual ao LS estabelecido para aquele marcador, estes serão considerados como artefactos *stutter* e sofrerão filtragem pelo GM sendo automaticamente descartados do perfil. Assim, nos EFs resultantes de corridas no *SeqStudio™*, apenas surgirão alelicamente identificados, pelo *software* de análise, picos em posição *stutter* que possuam uma proporção de altura relativamente ao alelo real, superior ao LS estabelecido para o *locus* em questão. O estabelecimento destes valores de LS como módulo de análise no GM, será particularmente relevante no auxílio dos peritos na interpretação de misturas, uma vez que ao desconsiderar picos como potencialmente *stutter* dos EFs, a genotipagem do perfil fica facilitada, diminuindo o risco de contabilizar erroneamente artefactos *stutter* como alelos de um contribuinte minoritário.

Vale lembrar que muitas vezes, nas amostras problema, verdadeiros alelos presentes em quantidades vestigiais, altamente degradados ou pertencentes a contribuintes minoritários apresentam picos com intensidade de sinal semelhante à produzida pelos seus *stutters*, sendo impossibilitada a distinção entre o alelo e o artefacto. Por isso, quando na presença de picos em posição *stutter* que receberam atribuição alélica pelo GM, em EFs de amostras problema - nomeadamente vestigiais, degradadas ou potencialmente de mistura - o perito não deve interpretá-los de imediato como garantidamente alélicos, pois podem corresponder a *stutters* ocasionalmente elevados face à pouca quantidade e/ou qualidade alélica. Nestes casos, para que os genótipos das amostras problema sejam corretamente concluídos, o perito deve proceder à avaliação cuidada do perfil, se possível, por comparação com perfis de duplicados e/ou amostras de referência.

## **Discussão relativa ao funcionamento do *SeqStudio™* para *HID***

Os estudos comparativos, bem como todas as situações em que se recorreu à utilização do *SeqStudio™*, permitiram verificar que o equipamento cumpre as funções descritas e previstas pelo fabricante, concluindo-se assim a aptidão do *SeqStudio™* para executar as funções de EC no âmbito da identificação humana para fins forenses, sendo, portanto, à semelhança do *3500*, adequada a sua utilização para esse efeito, na rotina do SGBF-C.

### **4.6. VANTAGENS E LIMITAÇÕES *SeqStudio™* vs. *3500***

Comprovado o correto funcionamento do *SeqStudio™*, resta abordar as vantagens e eventuais limitações identificadas em comparação com o equipamento de EC em atual utilização no SGBF-C.

O manuseamento do *SeqStudio™* durante os procedimentos experimentais do presente trabalho permitiu reconhecer como vantajosas as características do equipamento mencionadas na introdução (na secção **1.6.**), perante o *3500*, nomeadamente a configuração “*all-in-one*” do seu cartucho e a sua plataforma tátil incorporada no próprio instrumento, que tornam a sua utilização mais fácil, prática, intuitiva e menos propícia a contaminações e danos de manuseio.

Além disso, a maior sensibilidade de deteção do *SeqStudio™* face ao *3500*, constatada nos estudos comparativos (secção **4.5.**), constitui vantagem na sua utilização futura, particularmente na análise de amostras problema *com* ADN em quantidades vestigiais e/ou muito degradadas, e de potenciais misturas. Isto porque, enquanto fragmentos que no *3500* surgirão associados a baixos valores de RFU não sendo identificados alelicamente pelo GM devido à filtragem automática *stutters* ou do próprio LA, ou cujos picos surgem no EF mas com alturas que deixam dúvidas quanto à sua classificação como alelo *vs.* artefacto, no *SeqStudio™* há maior probabilidade desses fragmentos surgirem com atribuições alélicas associadas e de uma interpretação mais facilitada dos picos como pertencentes a contribuintes minoritários.

Por fim, apesar de não se ter recorrido à sua função de sequenciação de Sanger, o facto do *SeqStudio™* permitir a sua realização simultânea com análise de fragmentos numa

mesma placa, é vantajoso, uma vez que não obriga à mudança de polímero nem de capilares consoante o tipo de análise a efetuar, como acontece no 3500, com o qual, aquando da necessidade de análises a ADNmt, a rotina de STRs do SGBF-C tem de parar temporariamente. Assim, o *SeqStudio™* poderá, a longo prazo, tornar desnecessária essa estagnação e constrangimentos no funcionamento do serviço quando se efetuarem procedimentos de sequenciação.

Como limitações do *SeqStudio™* destaca-se o facto de possuir uma matriz de somente 4 capilares, enquanto que o 3500 possui uma matriz de 8 capilares. No entanto, o tempo aproximado de corrida *HID* em cada capilar é sensivelmente o mesmo, em ambos os equipamentos: 30 - 40 min para cada amostra (Applied Biosystems, 2010, 2019g, 2021). Ou seja, para correr 1 coluna inteira (8 amostras) de uma placa de 96 poços, apesar do *SeqStudio™* ter de efetuar 2 injeções enquanto o 3500 efetua apenas uma, o tempo que ambos os equipamentos demoram a correr as 8 amostras é praticamente o mesmo. Assim, a matriz diminuída do *SeqStudio™* acaba por nem constituir uma verdadeira desvantagem face à do 3500.

Posto isto, apenas se pode apontar como verdadeira limitação do *SeqStudio™*, o facto de permitir a análise de uma única placa de 96 poços de cada vez, enquanto que o 3500 permite a análise de duas dessas placas. Neste aspeto, o 3500 mostra-se capaz de correr mais amostras (o dobro) que o *SeqStudio™* numa única utilização.

## 5. CONCLUSÕES

---

Com a realização do presente trabalho foi possível validar internamente a função de eletroforese capilar do novo sequenciador automático *SeqStudio™ Genetic Analyzer* recentemente adquirido pelo SGBF-C, para análise de amostras seguras e problema amplificadas com o *kit GlobalFiler™*, nas condições laboratoriais específicas do serviço, para fins de identificação humana forense. No âmbito da validação efetuada foram determinados, os limiares de sensibilidade abaixo indicados, como parâmetros de análise de eletroferogramas a introduzir no *GeneMapper™* e como critérios a estabelecer no serviço para interpretação dos perfis gerados pelo equipamento, por forma a assegurar corretas atribuições alélicas e, portanto, a garantir resultados de qualidade:

- Um **Limiar Analítico** de **80 RFU** aplicável a todo o EF;
- Um **Limiar Estocástico** de **340 RFU** aplicável a todos os sistemas heterozigóticos;
- O conjunto de **Limiares de Stutter (%)** recomendado pelo fabricante para cada marcador autossómico (valores da coluna “*SeqStudio™*” da **Tabela 9**);
- Um **Limiar de Desequilíbrio de Heterozigóticos** único para todos os marcadores autossómicos, de **65%**.

A elaboração deste trabalho permitiu também, através de estudos comparativos, demonstrar que o *SeqStudio™* se apresenta capaz de produzir atribuições genotípicas equivalentes às geradas pelo equipamento de EC em atual utilização no SGBF-C e já devidamente validado - *3500 Genetic Analyzer* - para as mesmas amostras, comprovando-se assim a **reprodutibilidade dos perfis** pelo novo instrumento.

Todo o manuseio do *SeqStudio™* durante a elaboração do trabalho possibilitou igualmente constatar que o equipamento possui evidentes **vantagens face ao 3500**, das quais se destacam: a sua configuração “*all-in-one*” que reúne a maioria dos consumíveis necessários à EC num único cartucho; o facto de recorrer a um polímero e comprimento de capilares universal para diferentes análises; e o seu ecrã tátil que dispensa a ligação do instrumento a um computador, que tornam a sua utilização bem mais conveniente, facilitada, rápida e inovadora. Salieta-se também a maior sensibilidade do *SeqStudio™* concluída aquando da realização dos estudos comparativos e da própria comparação dos limiares de sensibilidade dos equipamentos, característica que será particularmente útil na sua utilização futura na análise de amostras problema com ADN vestigial, degradado e/ou de potenciais misturas com contribuintes minoritários. O *SeqStudio™* demonstrou uma única **limitação face**

**ao 3500**, nomeadamente a impossibilidade de introduzir simultaneamente para análise, mais do que uma placa de 96 poços, numa mesma utilização.

O presente trabalho constitui, para o SGBF-C, um extenso registo dos estudos de validação interna realizados no *SeqStudio™*, que evidencia, sob forma documentada, as condições em que o instrumento conduz à obtenção de perfis válidos, verdadeiros e de qualidade em amostras amplificadas com *GlobalFiler™*, podendo, por isso, ser doravante implementado para esse fim na rotina do SGBF-C, com confiança nos resultados obtidos, através da aplicação dos parâmetros estabelecidos. Este documento assegura, assim, que a (futura) emissão de relatórios periciais com base nos resultados de tipagem gerados pelo *SeqStudio™* está bem fundamentada. A presente dissertação, poderá servir, portanto, como prova do cumprimento dos requisitos de qualidade impostos pela norma NP EN ISO/IEC 17025 no âmbito da acreditação do laboratório pelo IPAC, garantindo que a utilização do *SeqStudio™* no SGBF-C mantém a qualidade e integridade do serviço prestado.

Por fim, tendo em vista, que interessa às autoridades judiciais, a rapidez na resolução de questões legais para uma eficiente resposta da Justiça, e que das vantagens do *SeqStudio™* se destaca a maior facilidade e rapidez no seu manuseio face ao 3500, o equipamento validado surge então como uma ferramenta ideal na resolução de casos forenses que envolvam análises genéticas. Posto isto, a introdução do *SeqStudio™* na rotina laboratorial do SGBF-C, constituirá, sem dúvida, uma mais-valia, não só para os peritos, que terão o seu trabalho facilitado na fase de EC, como também, em plano geral, contribuirá para uma Justiça com melhor capacidade de resposta em casos que envolvam a perícia prestada pelo SGBF-C.

## 6. LIMITAÇÕES DO ESTUDO E PERSPETIVAS FUTURAS

---

É de salientar que os limiares de sensibilidade definidos na presente validação como parâmetros fixos de análise e de interpretação dos EFs gerados pelo *SeqStudio™* não constituem, de todo, valores estanques e definitivos. Com a utilização recorrente do equipamento, após a sua introdução na rotina laboratorial, o mesmo será inevitavelmente exposto a alterações funcionais, como por exemplo desgaste do laser, e a eventuais mudanças nas condições ambientais envolventes, como por exemplo a temperatura e humidade do laboratório. Todos esses fatores, a longo prazo, poderão alterar o intervalo de condições em que o equipamento consegue fiavelmente efetuar medições analíticas e gerar resultados de confiança. Como tal, e para a manutenção da qualidade nos resultados de tipagem obtidos pelo *SeqStudio™* ao longo do tempo, estudos de validação interna semelhantes ao aqui descrito, deverão ser periodicamente concretizados no SGBF-C, de modo a reavaliar se os limiares em vigor se mantêm (ou não) adequados para uma correta determinação e interpretação dos perfis obtidos pelo equipamento, e a determinar novos parâmetros caso se evidenciem drásticas variações nos valores de confiabilidade outrora validados.

Para uma completa validação interna do equipamento *SeqStudio™*, é fundamental determinar os limiares de sensibilidade e efetuar estudos comparativos com todos os *kits* de amplificação a que o SGBF-C recorre, uma vez que se tenciona utilizá-los futuramente no equipamento em questão. Posto isto, será necessário, efetuar os mesmos procedimentos práticos deste trabalho, com os *kits PowerPlex Fusion® 6C System (Promega)* e *Yfiler™ Plus PCR Amplification Kit (AB™)*, sendo que para este último, não se tratando de um sistema heterozigótico, mas sim haplotípico, não se determinaria o LE nem o LD.

Além disso, os resultados da validação apresentados são apenas referentes ao tempo de injeção padrão (10s). Tendo em conta que o SGBF-C, recorre, por vezes, à reinjeção de amostras a 40s, para aumentar a intensidade de sinal nos eletroferogramas, seria de interesse avaliar adicionalmente a sensibilidade do *SeqStudio™* para esse tempo de injeção, por determinar um LE nessas condições, conforme recomendam as *guidelines* do SWGDAM. Essa avaliação teria como objetivo perceber se a discrepância observada entre o LE determinado a 40s e o LE a 10s (340 RFU) justificaria (ou não) a utilização de um valor de LE diferente na análise de amostras que tenham sido reaplicadas com 40s (SWGDAM, 2017).

Conforme referido na secção 1.7. *Contextualização da Validação*, outra forma de garantir a qualidade e fiabilidade nos perfis obtidos, no âmbito do cumprimento da norma NP

EN ISO/IEC 17025, consiste na participação em exercícios de comparações interlaboratoriais de resultados. Para que este requisito se cumpra, o SGBF-C deve, portanto, incluir o *SeqStudio™* nos exercícios de controlo de qualidade interlaboratorial organizados pelo GHEP-ISFG nos quais participa regularmente, correndo as amostras enviadas pela entidade no novo equipamento e confirmar a consensualidade dos perfis obtidos com os restantes laboratórios participantes. A inclusão do *SeqStudio™* nestes exercícios será apenas possível, assim que se tiverem estabelecido todos os limiares de análise e interpretação, para todos os *kits* em utilização no SGBF-C, no equipamento.

Apesar de não contemplada no âmbito da acreditação, por não se tratar de uma análise frequente no SGBF-C, poderá também ser avaliada a performance do *SeqStudio™* na sua função de Sequenciação de Sanger com amostras de ADNmt.



## 7. REFERÊNCIAS BIBLIOGRÁFICAS

---

- Amorim, A., Fernandes, T., & Taveira, N. (2019). Mitochondrial DNA in human identification: a review. *PeerJ*, 7. <https://doi.org/10.7717/peerj.7314>
- Applied Biosystems. (2010). *Applied Biosystems 3500/3500xL Genetic Analyzer - User Guide (Pub. No. 4401661 Rev. C)*.
- Applied Biosystems. (2012). *GeneMapper® ID-X Software Version 1.4 - User Bulletin (Pub. No. 4477684 Rev. B)*. <https://www.thermofisher.com/order/catalog/product/4479707>
- Applied Biosystems. (2014). *DNA Fragment Analysis by Capillary Electrophoresis - User Guide (Pub. No. 4474504 Rev. B)*. <https://assets.thermofisher.com/TFS-Assets/LSG/manuals/4474504.pdf>
- Applied Biosystems. (2015). *3500 and 3500xL Series Genetic Analyzers - Specification Sheet*.
- Applied Biosystems. (2016). *DNA Sequencing by Capillary Electrophoresis Chemistry Guide (3rd ed.)*. [thermofisher.com/sangersequencing](https://www.thermofisher.com/sangersequencing)
- Applied Biosystems. (2017). *PrepFiler Express™ and PrepFiler Express BTA™ Forensic DNA Extraction Kits for use with: AutoMate Express™ Forensic DNA Extraction System - User Guide (Pub. No. 4442699 Rev. D)*.
- Applied Biosystems. (2018). *Quantifiler™ HP and Trio DNA Quantification Kits - User Guide (Pub. No. 4485354 Rev. H)*. <https://www.thermofisher.com/order/catalog/product/4482910>
- Applied Biosystems. (2019a). *DNA fragment analysis by capillary electrophoresis - Go beyond sequencing*. <https://www.thermofisher.com/pt/en/home/life-science/sequencing/fragment-analysis/fragment-analysis-fundamentals.html?CID=fl-faworkflow>
- Applied Biosystems. (2019b). *GlobalFiler PCR amplification kits - GlobalFiler, GlobalFiler IQC, and GlobalFiler Express kits*. [thermofisher.com/globalfiler](https://www.thermofisher.com/globalfiler)
- Applied Biosystems. (2019c). *SeqStudio Genetic Analyzer for Human Identification - Answers you can trust-in just one click*. [thermofisher.com/hid-seqstudio](https://www.thermofisher.com/hid-seqstudio)
- Applied Biosystems. (2019d). *GeneMapper™ ID-X Software v1.6 - User Bulletin (Pub. No. 100073905 Rev. B)*.
- Applied Biosystems. (2019e). *AutoMate Express™ Instrument - User Guide (Pub. No. 4441982 Rev. G)*.
- Applied Biosystems. (2019f). *GlobalFiler™ and GlobalFiler™ IQC PCR Amplification Kits - User Guide (Pub. No. 4477604 Rev. F)*.
- Applied Biosystems. (2019g). *SeqStudio™ Genetic Analyzer Instrument and Software - User Guide (Pub. No. MAN0018646 Rev. A.0)*.
- Applied Biosystems. (2021). *Applied Biosystems™ 3500/3500xL GeneticAnalyzer - User Guide (Pub. No. 100031809 Rev. E)*.
- Applied Biosystems. (2022). *GeneScan™ 600 LIZ™ Size Standard v2.0 Product Information Sheet*. <https://www.thermofisher.com/order/catalog/product/4408399>
- Balsa, F., Bogas, V., Cunha, P., Brito, P., Serra, A., Lopes, V., Carvalho, M., Andrade, L., Bento, A. M., São Bento, M., Corte-Real, F., & Anjos, M. J. (2011). Preliminary validation of PrepFiler Express™ Extraction kit in AutoMate Express DNA Extraction System. *Forensic Science International: Genetics Supplement Series*, 3(1). <https://doi.org/10.1016/j.fsigss.2011.09.050>

- Bregu, J., Conklin, D., Coronado, E., Terrill, M., Cotton, R. W., & Grgicak, C. M. (2013). Analytical Thresholds and Sensitivity: Establishing RFU Thresholds for Forensic DNA Analysis. *Journal of Forensic Sciences*, 58(1), 120–129. <https://doi.org/10.1111/1556-4029.12008>
- Butler, E., & Li, R. (2014). Genetic Markers for Sex Identification in Forensic DNA Analysis. *Journal of Forensic Investigation*, 02(03).
- Butler, J. M. (2010). *Fundamentals of Forensic DNA Typing* (3rd ed.). Academic Press, Elsevier.
- Butler, J. M. (2011). *Advanced Topics in Forensic DNA Typing: Methodology*. Academic Press, Elsevier.
- Butler, J. M. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation* (3rd ed.). Academic Press, Elsevier.
- Butler, J. M., & Gittelson, S. N. (2015). *Basic STR Interpretation Workshop*. <http://www.cstl.nist.gov/strbase/training.htm>
- Corte-Real, F., & Vieira, D. N. (2015). Princípios de Genética Forense. In *Princípios de genética forense*. Imprensa da Universidade de Coimbra. <https://doi.org/10.14195/978-989-26-0957-7>
- Crespillo, M., Luque, J. A., Paredes, M. R., & Barrio, P. A. (2012). *Recomendaciones de la Comisión de Trabajo de Mezclas del GHEP-ISFG (GHEPMIX) - Criterios mínimos recomendados para la Aceptación y Evaluación de Perfiles Mezclas*.
- ENFSI. (2010). *Recommended Minimum Criteria for the Validation of Various Aspects of the DNA Profiling Process Aim*. <https://enfsi.eu/about-enfsi/structure/working-groups/documents-page/documents/>
- ENFSI. (2014). *Guidelines for the single laboratory Validation of Instrumental and Human Based Methods in Forensic Science*. <https://enfsi.eu/about-enfsi/structure/working-groups/documents-page/documents/>
- FBI. (2020). *Quality Assurance Standards (QAS) for Forensic DNA Testing Laboratories*.
- Fernández, K., Gómez, J., García-Hirschfeld, J., Cubillo, E., de la Torre, C. S., & Vallejo, G. (2015). Accreditation of the GHEP-ISFG proficiency test: One step forward to assure and improve quality. *Forensic Science International: Genetics Supplement Series*, 5, e515–e517. <https://doi.org/10.1016/j.fsigss.2015.09.204>
- Gill, P., Bleka, Ø., Hansson, O., Benschop, C., & Haned, H. (2020). *Forensic Practitioner's Guide to the interpretation of Complex DNA Profiles* (1st ed.). Academic Press, Elsevier.
- Gill, P., Puch-Solis, R., & Curran, J. (2009). The low-template-DNA (stochastic) threshold-Its determination relative to risk analysis for national DNA databases. *Forensic Science International: Genetics*, 3(2), 104–111. <https://doi.org/10.1016/j.fsigen.2008.11.009>
- Gomes, C., Martínez-Gómez, J., Díez-Juárez, L., Díaz-Sánchez, S., Palomo-Díez, S., Arroyo-Pardo, E., Cano-López, M., & Fernández-Serrano, J. (2017). Prep-n-Go™: A new and fast extraction method for forensic blood samples. *Forensic Science International: Genetics Supplement Series*, 6, e265–e266. <https://doi.org/10.1016/j.fsigss.2017.09.089>
- Goodwin, W., Linacre, A., & Hadi, S. (2011). *An Introduction to Forensic Genetics* (2nd ed.). Wiley-Blackwell.
- Griffiths, A. J. F. (2012, July 5). *DNA sequencing*. *Encyclopedia Britannica*. <https://www.britannica.com/science/DNA-sequencing>
- Hares, D. R. (2012). Expanding the CODIS core loci in the United States. In *Forensic Science International: Genetics* (Vol. 6, Issue 1). <https://doi.org/10.1016/j.fsigen.2011.04.012>

- Hares, D. R. (2015). Selection and implementation of expanded CODIS core loci in the United States. In *Forensic Science International: Genetics* (Vol. 17, pp. 33–34). Elsevier Ireland Ltd. <https://doi.org/10.1016/j.fsigen.2015.03.006>
- Houck, M. M. (2015). *Advanced Forensic Science Series - Forensic Biology* (1st ed.). Academic Press, Elsevier.
- INMLCF. (2014). *Serviço de Genética e Biologia Forenses*. Sobre o INMLCF - Serviços Técnicos. [https://www.inmlcf.mj.pt/index.php?option=com\\_content&view=article&id=130:servico-de-genetica-e-biologia-forenses&catid=37&Itemid=292](https://www.inmlcf.mj.pt/index.php?option=com_content&view=article&id=130:servico-de-genetica-e-biologia-forenses&catid=37&Itemid=292)
- INMLCF. (2020). *Relatório de Atividades | 2020*. [https://www.inmlcf.mj.pt/index.php?option=com\\_content&view=article&id=169&catid=40&Itemid=295](https://www.inmlcf.mj.pt/index.php?option=com_content&view=article&id=169&catid=40&Itemid=295)
- IT-SGBF-C-014. (2022). *Instrução Técnica “Extração de ADN de amostras-problema de sangue, saliva, sémen e/ou outros fluídos biológicos não determinados (mancha e/ou zaragatoa)” Rev06*.
- IT-SGBF-C-015. (2020). *Instrução Técnica “Preparação de amostras para quantificação em tempo real (QUAN002)” Rev05*.
- IT-SGBF-C-017. (2021). *Instrução Técnica “Extração de ADN de amostras de sangue e saliva (Prep-n-Go™)” Rev07*.
- IUPAC. (1976). Nomenclature, Symbols, Units and their Usage in Spectrochemical Analysis - II. Data Interpretation. *Pure and Applied Chemistry*, 45(2), 99–103. <https://doi.org/http://dx.doi.org/10.1351/pac197645020099>
- Kaiser, H. (1970). Report for analytical chemists: Part II. Quantitation in Elemental Analysis. *Analytical Chemistry*, 42(4), 26A-59A. <https://doi.org/https://doi.org/10.1021/ac60286a027>
- Khan Academy. (2017). *Reação em cadeia da polimerase (PCR)*. <https://pt.khanacademy.org/science/biology/biotech-dna-technology/dna-sequencing-pcr-electrophoresis/a/polymerase-chain-reaction-pcr>
- Kim, Y. T., Heo, H. Y., & Seo, T. S. (2018). Advanced Short Tandem Repeat Genotyping for Forensic Human Identification. In H. N. Chang (Ed.), *Emerging Areas in Bioengineering*. Wiley-VCH Verlag GmbH & Co. KGaA.
- Machado, H., & Moniz, H. (2014). *Bases de dados genéticos forenses: Tecnologias de controlo e ordem social*. <https://www.researchgate.net/publication/285401152>
- MQ-SGBF-C. (2022). *Manual da Qualidade do Serviço de Genética e Biologia Forenses (Rev14)*.
- NIST. (2015). *National Institute of Standards and Technology - Overview of STR Fact Sheets*. [https://strbase.nist.gov/str\\_fact.htm](https://strbase.nist.gov/str_fact.htm)
- NP EN ISO/IEC 17025, Guia para a aplicação da NP EN ISO/IEC 17025:2018 (2018). [www.ipac.pt](http://www.ipac.pt)
- Pajnič, I. Z. (2020). Mitochondrial DNA in forensic analyses. *Slovenian Medical Journal*, 89(1–2), 55–72. <https://doi.org/10.6016/ZdravVestn.2932>
- PE-SGBF-C-001. (2020). *Procedimento de Ensaio “Determinação de perfil genético de ADN por PCR / Eletroforese Capilar a partir de amostras de referência de sangue e saliva*.
- PE-SGBF-C-002. (2018). *Procedimento de Ensaio “Determinação de Perfil Genético de ADN por PCR/Eletroforese Capilar a partir de amostras-problema de sangue, saliva, sémen e/ou outros fluídos biológicos não determinados (em mancha e/ou zaragatoa)”*.
- PG-SGBF-C-001. (2020). *Procedimento Geral “Validação de Ensaios.”*

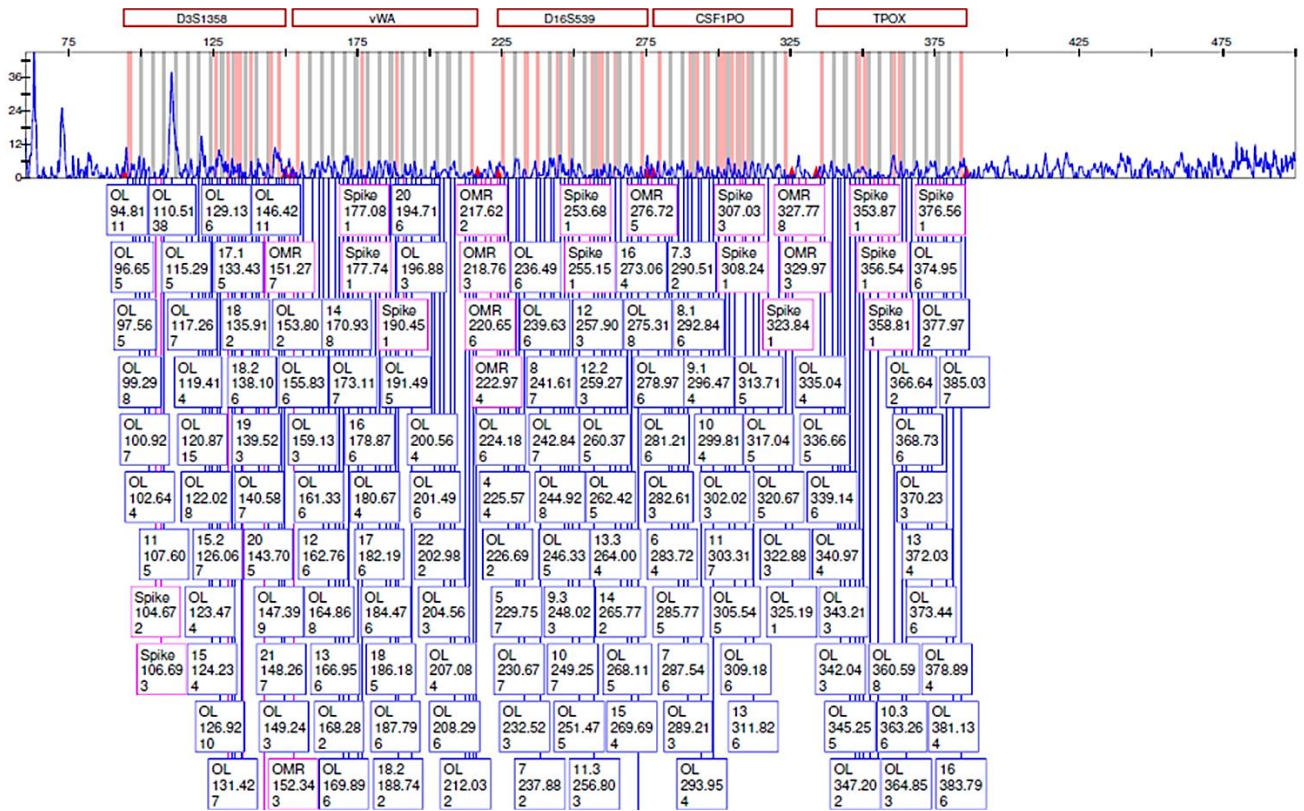
- Pinheiro, M. (2010). *Genética Forense: perspectivas da identificação genética*. edições Universidade Fernando Pessoa.
- Portaria n.º 19/2013, de 21 de janeiro, Diário da República 1.ª série — N.º 14 — 21 de janeiro de 2013. Artigo 5º da Portaria n.º 19/2013 - Estatutos do Instituto Nacional de Medicina Legal e Ciências Forenses, I. P. - Serviço de Genética e Biologia Forenses (2013). <https://dre.pt/dre/detalhe/portaria/19-2013-257062>
- PO-SGBF-C-012. (2021). *Procedimento Operacional “Amplificação de ADN por PCR com GlobalFiler™ PCR Amplification Kit” Rev06*.
- PO-SGBF-N-010. (2020). *Procedimento Operacional “Análise do mtDNA por Sequenciação Directa.”*
- RV-SGBF-C-006. (2020). *Relatório de Validação “Validação de perfis genéticos” Rev02*.
- SWGDM. (2010). *SWGDM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories*. [http://media.wix.com/ugd/4344b0\\_61b46a0e1a4c41ccb65f719a533b8e29.pdf](http://media.wix.com/ugd/4344b0_61b46a0e1a4c41ccb65f719a533b8e29.pdf)
- SWGDM. (2016). *Validation Guidelines for DNA Analysis Methods*. <https://www.swgdam.org/publications>
- SWGDM. (2017). *Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories*. <https://www.swgdam.org/publications>
- Thermo Fisher Scientific. (2019). *SeqStudio™ Genetic Analyzer Troubleshooting (Rev. 1.1)*.
- Thermo Fisher Scientific. (2020). *Combine Sanger sequencing and fragment analysis in the same instrument run*. <https://www.thermofisher.com/pt/en/home/life-science/sequencing/sanger-sequencing/sanger-sequencing-technology-accessories/seqstudio-genetic-analyzer.html>
- Thermo Fisher Scientific. (2021). *PrepFiler Express and AutoMate Express DNA extraction systems - Product Bulletin*. <https://www.thermofisher.com/order/catalog/product/4441352>
- Udogadi, N. S., Abdullahi, M. K., Bukola, A. T., Imose, O. P., & Esewi, A. D. (2020). Forensic dna profiling: Autosomal short tandem repeat as a prominent marker in crime investigation. In *Malaysian Journal of Medical Sciences* (Vol. 27, Issue 4, pp. 22–35). Penerbit Universiti Sains Malaysia. <https://doi.org/10.21315/mjms2020.27.4.3>
- Watson, J. D., & Crick, F. H. (1953). The structure of DNA. *Cold Spring Harbor Symposia on Quantitative Biology*, 18, 123–131. <https://doi.org/10.1101/SQB.1953.018.01.020>

## 8. ANEXOS

**ANEXO I** – Tabela contendo os valores de LS recomendados pelo fabricante do *GlobalFiler™* para o 3500 vs. os valores obtidos experimentalmente pelo SGBF-C na última validação interna realizada ao equipamento (RV-SGBF-C-006, 2020). Os valores obtidos pelo SGBF-C são tendencialmente inferiores aos da *Applied Biosystems™* em todos os marcadores, o que pode ser explicado, segundo o relatório de validação atrás referenciado, por “*não ter sido possível observar no estudo, nomeadamente, os alelos de maiores dimensões de cada marcador*”. Não sendo possível incluir nos cálculos, os alelos de maior comprimento (cujos *stutters* são maiores, em termos de RFU), os valores de LS do serviço surgem naturalmente menores, não conseguindo tão bem efetuar uma filtragem de *stutters*, pois alguns desses artefactos acabarão por sofrer atribuição alélica quando ultrapassam o valor estipulado de LS para o marcador em questão. É por esta razão que o SGBF-C opta pelos valores de LS do fabricante, cuja amostragem é sempre mais abrangente e, portanto, mais próxima da realidade.

<b>GlobalFiler™- 3500</b>		
<b>Sistema</b>	<b>LS fabricante (%)</b>	<b>LS SGBF-C (%)</b>
<b>D3S1358</b>	12.45	10.15
<b>vWA</b>	12.33	9.82
<b>D16S539</b>	10.17	8.57
<b>CSF1PO</b>	11.40	8.40
<b>TPOX</b>	5.43	4.71
<b>D8S1179</b>	10.20	8.35
<b>D21S11</b>	11.42	9.16
<b>D18S51</b>	13.47	11.81
<b>D2S441</b>	8.75	6.82
<b>D19S433</b>	10.58	8.53
<b>TH01</b>	5.24	3.98
<b>FGA</b>	11.96	9.88
<b>D22S1045</b>	17.30	15.47
<b>D5S818</b>	10.84	8.24
<b>D13S317</b>	9.98	8.50
<b>D7S820</b>	10.21	7.43
<b>SE33</b>	14.42	12.53
<b>D10S1248</b>	12.50	10.38
<b>D1S1656</b>	13.08	10.99
<b>D12S391</b>	15.08	12.84
<b>D2S1338</b>	13.14	10.35

**ANEXO II** - Picos de ruído de fundo detetados no painel azul do EF de uma das 63 amostras contabilizadas para os cálculos do LA. O maior pico de ruído observado, neste exemplo, é de 38 RFU, e o menor é de 1 RFU.



**ANEXO III** – Perfil do controlo positivo *DNA Control 007* incluído no kit *GlobalFiler™*, segundo o fabricante. Adaptado de (Applied Biosystems, 2019f)

D3S1358	vWA	D16S539	CSF1PO	TPOX	
15,16	14,16	9,10	11,12	8	
Yindel	AMEL	D8S1179	D21S11	D18S51	DYS391
2	X,Y	12,13	28,31	12,15	11
D2S441	D19S433	TH01	FGA		
14,15	14,15	7,9,3	24,26		
D22S1045	D5S818	D13S317	D7S820	SE33	
11,16	11	11	7,12	17,25.2	
D10S1248	D1S1656	D12S391	D2S1338		
12,15	13,16	18,19	20,23		