

Rui Nascimento <sup>(1, 2)</sup>, Heloísa Afonso Costa <sup>(1, 3)</sup>, Eugénia Cunha <sup>(1, 4)</sup>, Francisco Corte Real <sup>(1, 5)</sup>, António Amorim <sup>(1, 6)</sup>

<sup>1</sup>Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal; <sup>2</sup>Escola de Ciências da Vida e Ambiente da Universidade de Trás-os-Montes e Alto Douro, Portugal; <sup>3</sup>Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal; <sup>4</sup>Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Portugal; <sup>5</sup>Faculdade de Medicina da Universidade de Coimbra, Portugal; <sup>6</sup>Faculdade de Ciências da Universidade de Lisboa, Portugal

## Introdução

Na genética forense, o estudo do cromossoma Y é essencial em casos de misturas de ADN masculino-feminino, como em agressões sexuais, permitindo determinar o número de contribuintes masculinos. As linhagens paternas podem ser analisadas através do estudo de Y-STR ou Y-SNP. Os Y-STR definem linhagens paternas próximas com pouca informação biogeográfica, enquanto os Y-SNP definem linhagens distantes com informações de ancestralidade mais precisas.

A sequenciação de nova geração (NGS) possibilita a análise de um grande número de SNPs em simultâneo, como tal, emergiu nos últimos anos como uma alternativa às tecnologias existentes. A aplicação das tecnologias de NGS ao estudo do cromossoma Y, proporcionam uma inferência de haplogrupo Y de alta resolução. Contudo, a grande quantidade de dados gerada pelas as tecnologias NGS podem ser de difícil análise, inviabilizando a sua aplicabilidade em casos reais.

Este trabalho tem como objetivo apresentar o método desenvolvido no Serviço de Genética e Biologia Forenses da Delegação do Sul (SGBF-S) para a determinação dos haplogrupos do cromossoma Y de uma forma simplificada, com recurso ao Torrent Server e ao software Yleaf para determinação do haplogrupo e determinação da informação biogeográfica da linhagem paterna.

## Métodos

As amostras estudadas foram colhidas no SGBF-S a voluntários após a assinatura de consentimento informado. Após extração e purificação do ADN, a preparação de bibliotecas foi feita recorrendo ao painel on AmpliSeq HID Y-SNP Research Panel v1 que inclui 859 Y-SNPs específicos de linhagem paterna, o que permite a diferenciação entre 640 Y-haplogrupos. A sequenciação foi realizada no Ion GeneStudio™ S5 System recorrendo a Ion 530™ Chips. A análise primária da sequência detetada foi realizada no Torrent Server por comparação ao genoma de referência humano, hg19. Posteriormente a análise foi feita em ambiente Linux, os ficheiros BAM (Figura 1) de cada amostra, foram retirados do Torrent Server e foram analisados utilizando o software Yleaf (Figura 2) que assiste na determinação do Haplótipo e Haplogrupo do cromossoma Y. A informação biogeográfica foi obtida recorrendo à base de dados da Haplogroup Ytree disponibilizada pelo Yfull (Figura 3).

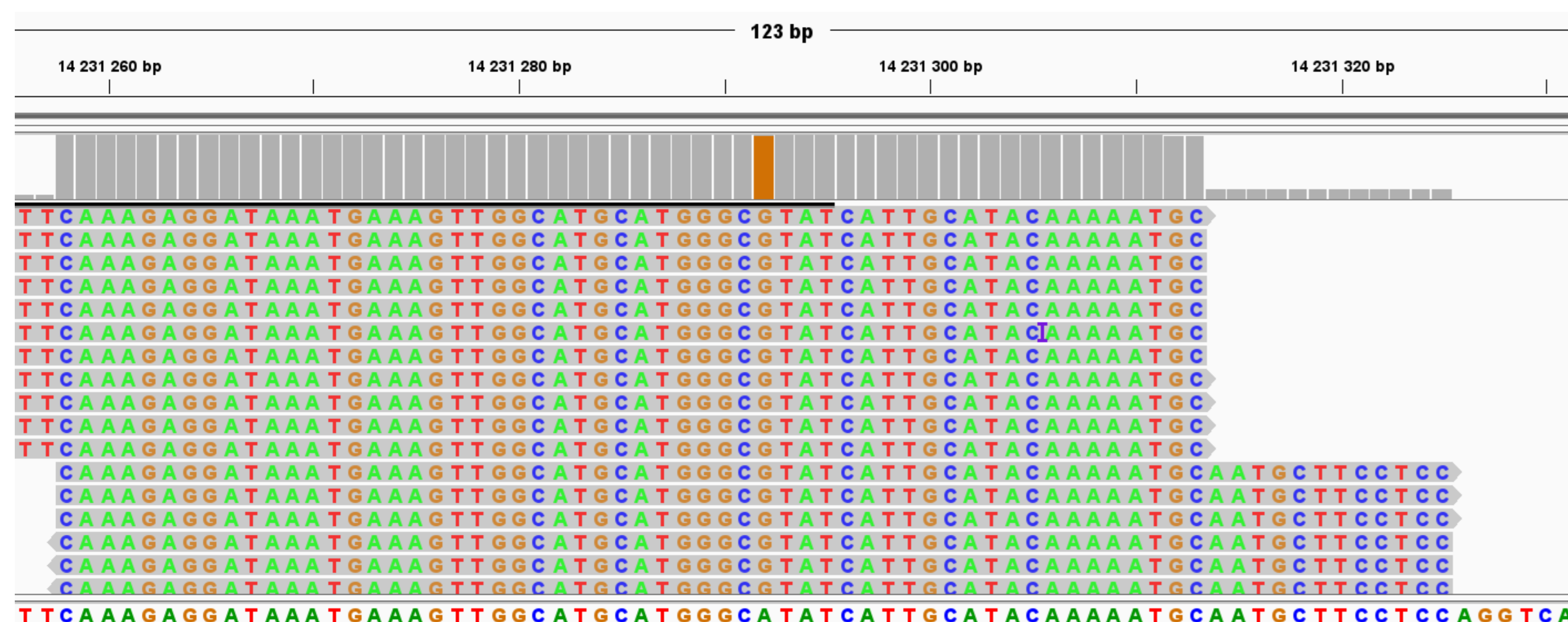


Figura 1 – Visualização de "Reads" de um ficheiro .bam no software IGV, numa amostra amplificada com o painel AmpliSeq HID Y-SNP Research Panel v1

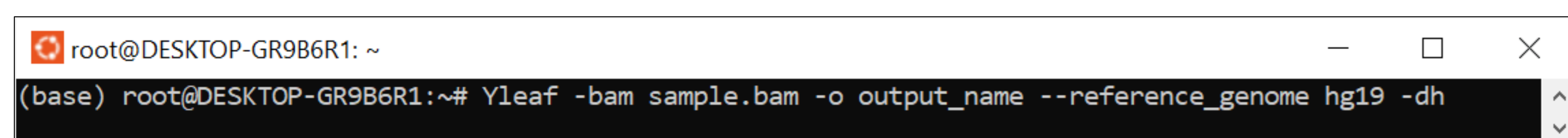


Figura 2 – Exemplo de código para executar o software Yleaf com um ficheiro .bam

## Conclusão

O trabalho efetuado e os resultados obtidos, permite implementar um método de análise simples e eficiente dos dados obtidos por métodos de sequenciação NGS para o estudo de Y-SNPs. Obtendo um haplótipo e um haplogrupo de alta resolução, superando assim, todas as limitações das tecnologias anteriores.

## Bibliografia

1. YTree. <https://www.yfull.com/tree/>. (acedido em Agosto 2024)
2. Kayser, M. Forensic use of Y-chromosome DNA: a general overview. Hum Genet 136, 621–635 (2017).
3. Ralf, A. et al. Forensic Y-SNP analysis beyond SNaPshot: High-resolution Y-chromosomal haplogrouping from low quality and quantity DNA using Ion AmpliSeq and targeted massively parallel sequencing. Forensic Science International: Genetics 41, 93–106 (2019).
4. Børsting, C. & Morling, N. Next generation sequencing and its applications in forensic genetics. Forensic Science International: Genetics 18, 78–89 (2015).
5. Sobrino, B., Brión, M. & Carracedo, A. SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Science International 154, 181–194 (2005).
6. Ralf, A., Montiel González, D., Zhong, K. & Kayser, M. Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. Molecular Biology and Evolution 35, 1291–1294 (2018).

## Resultados

Os resultados obtidos pelo software Yleaf vêm sob a forma de tabelas, representadas em infra. A tabela 1 apresenta a predição final do haplogrupo, tal como a predição de haplogrupo tendo em consideração o último marcador encontrado e os resultados dos controlos de qualidade que asseguram uma boa predição do algoritmo. A tabela 2 apresenta um excerto da tabela de resultados para todos os SNPs estudados. Nesta tabela estão disponíveis várias informações sobre os SNPs estudados, tais como o nome do marcador, a sua posição, o haplogrupo correspondente em caso de mutação, a mutação esperada, o genótipo determinado e os parâmetros de qualidade.

Tabela 1 – Resultado obtido no ficheiro .hg, após correr o software Yleaf

Sample_name	Hg	Hg_marker	Total_reads	Valid_markers	QC-score	QC-1	QC-2	QC-3
HID-YSNP-Sample	R-P312*(xR-DF19, R-U152,...)	S116	758560	1285	1.0	1.0	1.0	1.0

Tabela 2 – Resultado obtido no ficheiro .out, após correr o software Yleaf

chr	pos	marker_name	haplogroup	mutation	reads	called_perc	called_base	state	depth
chrY	14231291	L1118	A0-T	T->C	1083	99	C	D	1
chrY	21739790	L1145	A0-T	C->T	1193	99	T	D	1
chrY	17947672	V168	A1	G->A	1215	99	A	D	2
chrY	18759708	L1053	A1	C->A	718	100	A	D	2
chrY	7589303	V221	A1b	G->T	537	100	T	D	3
chrY	6932831	L413	BT	G->A	330	100	A	D	4
chrY	23244049	M9379	BT	G->C	1386	99	C	D	4
chrY	14197867	PF2587	CF	G->A	1423	99	A	D	6
chrY	14813991	M168	CT	C->T	1282	99	T	D	5
chrY	22744945	M294	CT	C->T	1338	100	T	D	5
chrY	21917313	PF2746	F	C->T	1304	99	T	D	7
chrY	8589031	M3658	GHIJK	C->T	633	99	T	D	8
chrY	7202703	F929	HIJK	C->T	1034	99	T	D	9
chrY	7173143	L16	IJK	G->A	965	99	A	D	10
chrY	15472863	P131	K	C->T	672	100	T	D	11

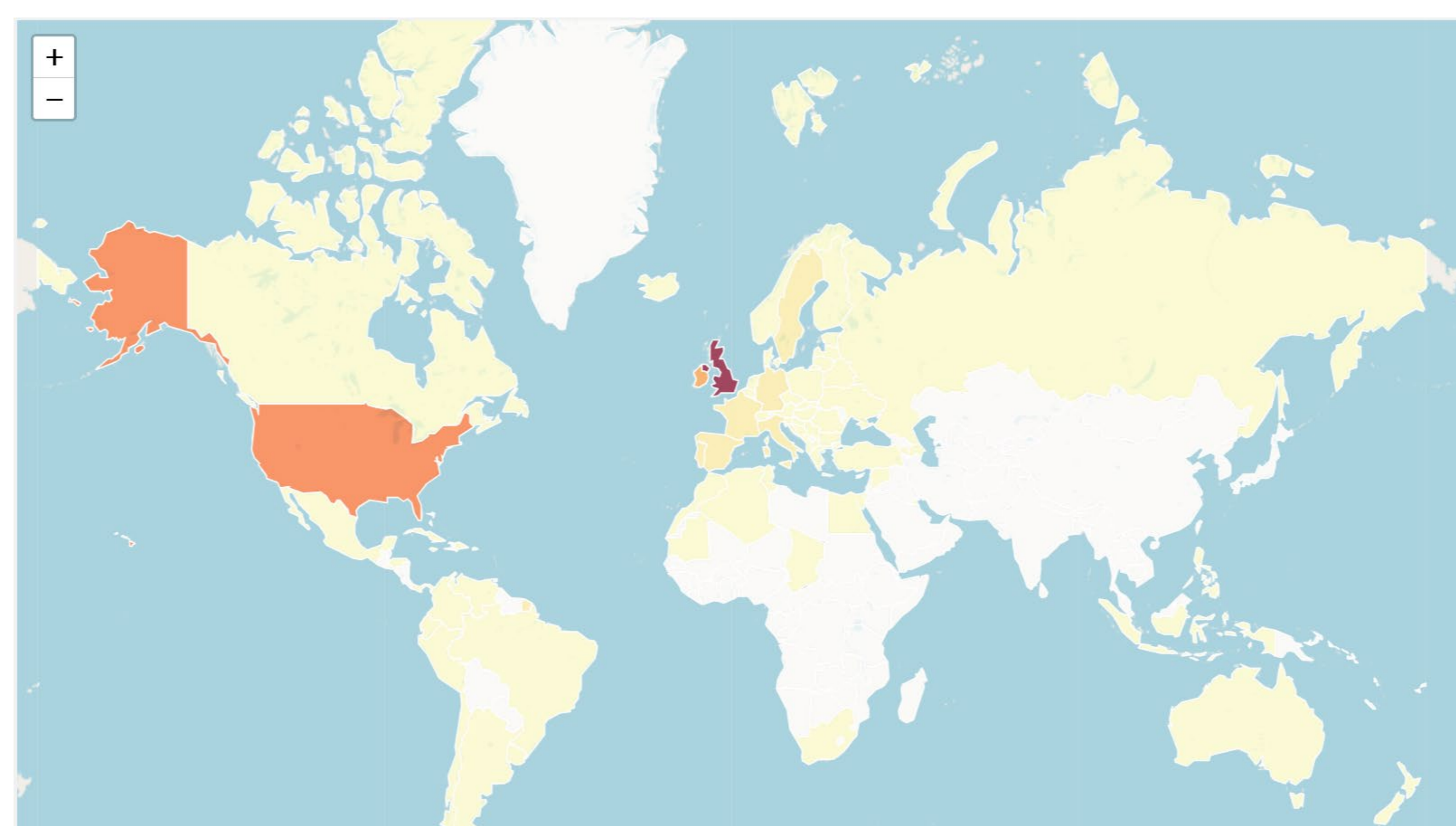


Figura 3 – Visualização da distribuição geográfica de um haplogrupo no website Yfull, após a determinação do mesmo com o Yleaf.