

Guia Prático de Genealogia Genética

Uma breve introdução

Por Aaron Salles Torres*

*Aaron Salles Torres é diplomado em Cinema, Vídeo e Novas Mídias pela School of the Art Institute of Chicago e possui um MBA com ênfase em Marketing pela Loyola University Chicago. É escritor, cineasta, e um dos pioneiros entusiastas da Genealogia Genética no Brasil. Administra os seguintes projetos genéticos na Family Tree DNA: I2b1/M223, I1 Ibérico e R1b1a2 Ibérico, além de vários projetos genealógicos, incluindo os das famílias Braga, Costa, Fernandes, Ferreira, Freitas, Gomes, Gonzalez, Martinez, Mendes, Oliveira, Pires, Ribeiro, Rocha, Rodrigues, Salles, Silveira, Souza e Torres.

Que genealogista nunca encontrou um obstáculo aparentemente intransponível na pesquisa de uma linha de sua família, um mistério que a pesquisa documental continua a deixar insolúvel? A mesma curiosidade que leva o genealogista a se debruçar sobre montanha após montanha de documentos antigos para organizar o quebra-cabeças de sua história familiar, essa curiosidade existencial também o leva a buscar novas formas de conquistar as maiores barreiras a sua pesquisa. Assim nasceu a Genealogia Genética, com o intuito de solucionar as mais complexas questões familiares utilizando-se da maior e mais duradoura das enciclopédias: o DNA.

A Genealogia Genética consiste da aplicação de conceitos e métodos da biologia e genética à pesquisa histórico-documental praticada pela genealogia tradicional, complementando dados ou cobrindo lacunas documentais que, de outra forma, permaneceriam em aberto. O conceito básico pode ser explicado da seguinte maneira: dois indivíduos que possuem um ancestral em comum compartilharão passagens de DNA idênticas, que herdaram desse ancestral¹. É a descoberta dessas passagens comuns de DNA que se dedicam as empresas que realizam os testes de genealogia genética, adotando padrões uniformes de forma que os resultados possam ser comparados entre si, em várias plataformas. Com seus exames em mãos,

¹ Em seu uso prático, entretanto, este conceito complica-se um pouco, pois quanto maior a distância - em gerações - de um ancestral comum, menor é a passagem contínua de DNA herdada dele que será preservada em seus herdeiros, como será explicado a seguir.

genealogistas então optam por se utilizar da internet para conectar suas enciclopédias particulares de DNA, criando uma imensa base de dados onde é evidente o entrelaçamento originado pelos ancestrais comuns dos indivíduos testados. Uma coisa é certa: documentos podem faltar, por serem perdidos ou destruídos; entretanto, a passagem de um indivíduo pela terra sempre será marcada pelo rastro de DNA deixado em seus descendentes - essa marca genética nunca será apagada.

Conceitos

Nosso DNA² possui três componentes, distintos na forma em que são passados de pais para filhos, que são úteis aos genealogistas: o DNA autossômico, o cromossomo Y (sexual) e o DNA mitocondrial.

Uma célula humana é composta de várias estruturas, entre as quais o núcleo – onde se concentram os 23 pares de cromossomos humanos, 22 somáticos e 1 sexual – e a mitocôndria, responsável pela respiração celular, como pode ser visto na figura a seguir:

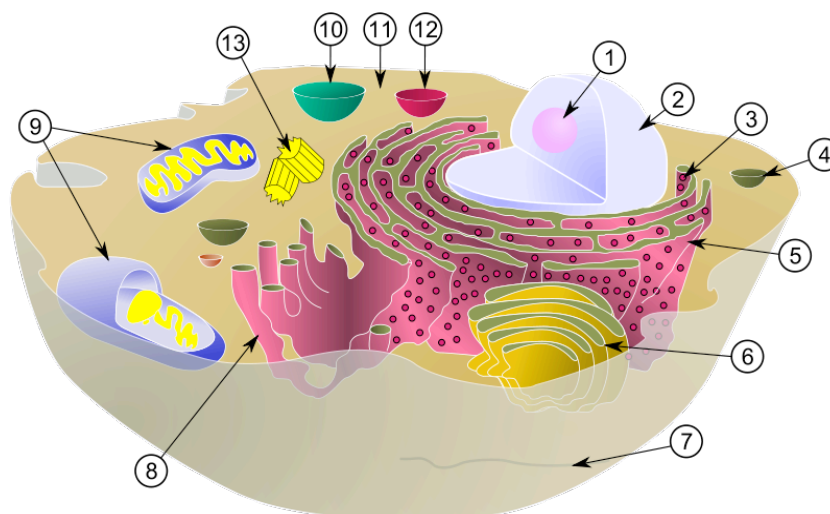


Figura 1 – estrutura de uma célula animal típica (Wikipédia): 1. Nucléolo; 2. Núcleo celular; 3. Ribossomos; 4. Vesículas; 5. Ergastoplasma ou Retículo endoplasmático rugoso

² O ácido desoxirribonucleico (DNA) “é um composto orgânico cujas moléculas contêm as instruções genéticas que coordenam o desenvolvimento e funcionamento de todos os seres vivos e alguns vírus.” Fonte: “Ácido Desoxirribonucleico.” *Wikipédia, a Enciclopédia Livre*. Web. 01 Mar. 2011. <<http://pt.wikipedia.org/wiki/DNA>>.

(RER); 6. Complexo de Golgi; 7. Microtúbulos; 8. Retículo endoplasmático liso (REL); 9. Mitocôndrias; 10. Vacúolo; 11. Citoplasma; 12. Lisossomas; 13. Centríolos.

Concentrados no núcleo, os cromossomos humanos podem ser representados como no seguinte cariótipo:

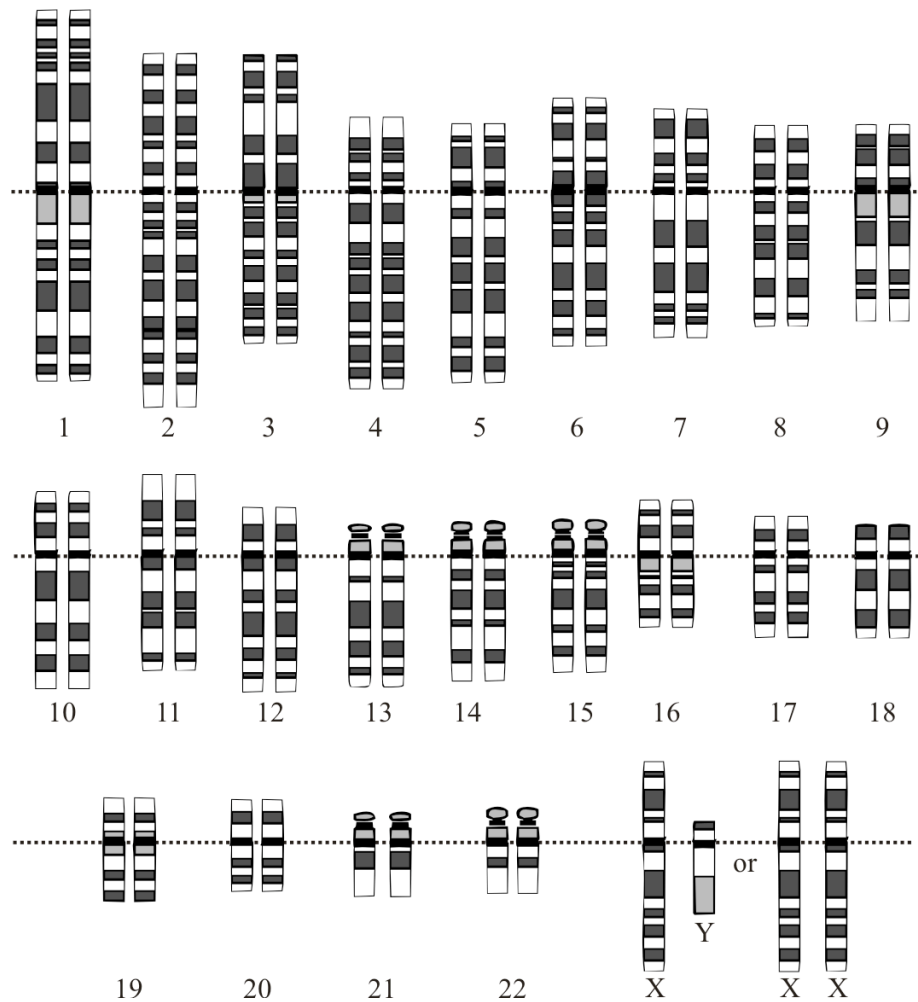


Figura 2 – cariótipo com os cromossomos humanos (Wikipédia): 22 pares de cromossomos somáticos, em que um cromossomo de cada par é herdado do pai e outro, da mãe (números de 1 a 22); e 1 par de cromossomos sexuais (XY no caso de um indivíduo do sexo masculino, ou XX no caso de um indivíduo do sexo feminino). No caso de um homem, o cromossomo Y é sempre herdado do pai. No caso de uma mulher, herda-se um cromossomo X da mãe e um cromossomo X do pai.

DNA Autossômico

O DNA autossômico refere-se aos 22 pares de cromossomos somáticos encontrados no núcleo celular. São assim chamados porque são encontrados em igual número de cópias em indivíduos do sexo masculino e feminino, e herdados em igual

quantidade tanto do pai, quanto da mãe – um cromossomo em cada par é recebido de cada um. Assim, um indivíduo possui $\frac{1}{2}$ do DNA de um pai e, teoricamente, $\frac{1}{4}$ do genoma de um avô, $\frac{1}{8}$ do material genético de um bisavô, $\frac{1}{16}$ do DNA de um trisavô e assim consecutivamente. Entretanto, a realidade não é tão exata. Na transmissão de cromossomos somáticos de pai para filho atua um dos mais importantes dispositivos à preservação das espécies: a recombinação cromossômica.

Durante a produção de gametas³, os cromossomos somáticos são “embaralhados”, criando diversidade genética essencial à vida. O processo de recombinação cromossômica ocorre durante a etapa Prófase I, a mais longa da meiose, em que o pareamento de cromossomos homólogos permite a troca aleatória de material genético entre os mesmos, criando um novo cromossomo que é único em sua configuração:

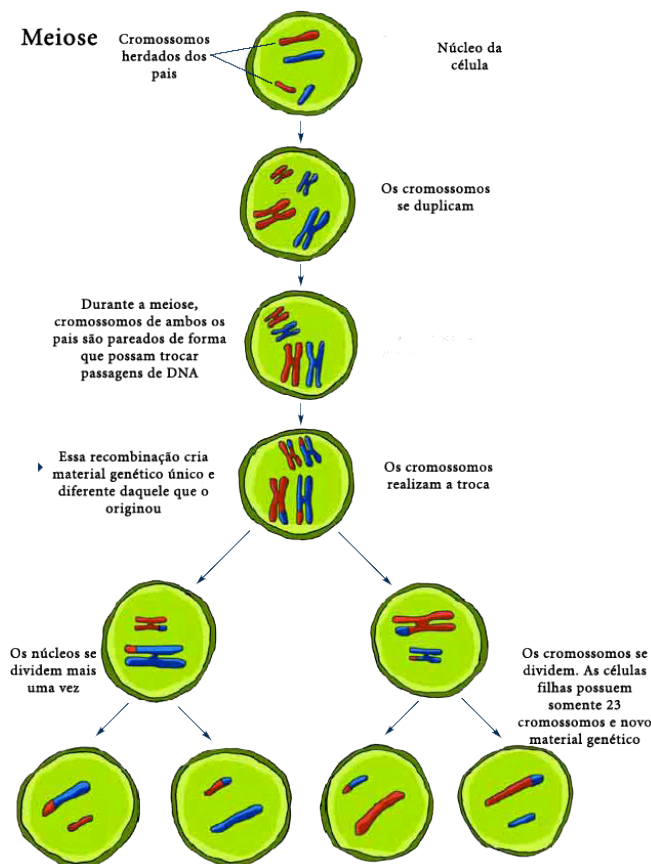


Figura 3 – meiose (wereyouwondering.com): durante a etapa Prófase I ocorre a permuta (recombinação cromossômica).

³ Gametas são células reprodutivas haplóides, ou n (23 cromossomos), que se unem na fecundação para formar o zigoto, 2n (46 cromossomos).

É necessário notar que a permuta de material genético entre cromossomos afeta não somente os cromossomos somáticos, como também os cromossomos sexuais X e Y – este último em menores proporções e somente em locais específicos, passagens estas que têm seu uso descartado para os fins discutidos aqui. Para o genealogista, a relevância da recombinação cromossômica se torna evidente com o passar das gerações, como pode ser notado no quadro seguinte:

Herança de Cromossomos Somáticos

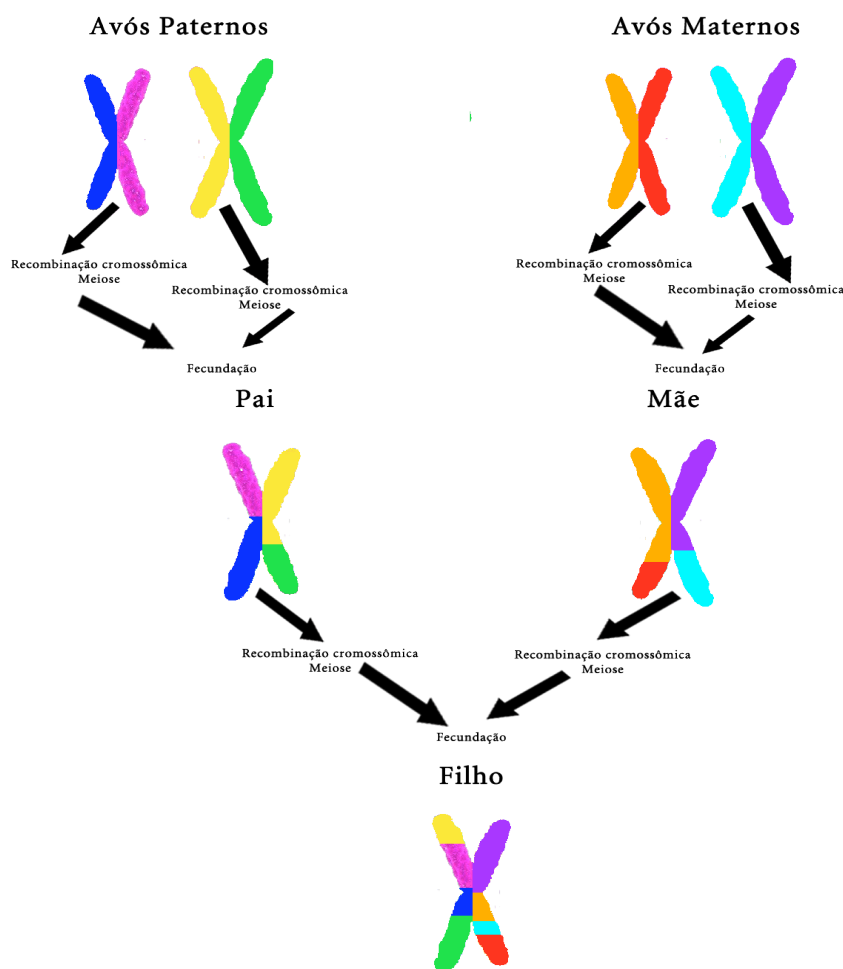


Figura 4 – herança de cromossomos somáticos, levando-se em consideração a recombinação cromossômica.

Neste exemplo, fica claro que o filho do casal possui uma maior quantidade de material genético da avó materna (indivíduo número 4) que do avô materno (indivíduo número 3, da esquerda para a direita), ressaltando a maneira aleatória

como dados genéticos são herdados por descendentes. Ao aplicarmos este conceito a um número maior de gerações, conclui-se que o DNA transmitido por um ancestral remoto pode ser perdido em um indivíduo contemporâneo, ou reduzido a passagens mínimas que são difíceis de serem reconhecidas. De fato, a dificuldade de se identificar passagens recebidas de um ancestral distante em específico evidencia a necessidade de cooperação na genealogia, pois é somente através do cruzamento dos dados de diversos descendentes de um dado ancestral longínquo que se torna possível estabelecer a distribuição de sua carga genética nos dias atuais. Métodos para a utilização do DNA autossômico em genealogia serão discutidos adiante.

Cromossomo Y

O cromossomo Y é um cromossomo responsável pela determinação do sexo masculino na maioria dos mamíferos, incluindo os seres humanos. Assim sendo, é considerado um cromossomo sexual e somente pode ser transmitido de pai para filho, não estando presente em indivíduos do sexo feminino. Por estar presente em uma única cópia (o outro cromossomo sexual é o X, formando XY em machos), o cromossomo Y limita a recombinação cromossômica com seu parceiro X a passagens específicas, o que significa que a maior parte de sua estrutura é preservada. Assim, o cromossomo Y é passado intacto de pai para filho. Conclui-se portanto que, a excluir mutações⁴ ocorridas em seu interior durante a produção do gameta que origina cada geração seguinte, o cromossomo Y encontrado em um indivíduo contemporâneo é, em sua maior parte, o mesmo encontrado em um ancestral pela linha paterna direta

⁴ Mutações são variações na sequência de pares de bases de um indivíduo. As mutações que ocorrem durante a formação das gametas são passadas de pais para filhos. Mutações não são sempre negativas, pois geram maior diversidade genética, o que é importante à sobrevivência das espécies. Diferentes áreas do DNA possuem diferentes tendências a mutação. A propensão de uma certa passagem do DNA a mutação constitui a taxa de mutação dessa sequência.

que viveu há 1.000 ou 4.000 anos. Por esse motivo, o cromossomo Y é uma ferramenta preciosa à investigação desse tipo de linha.

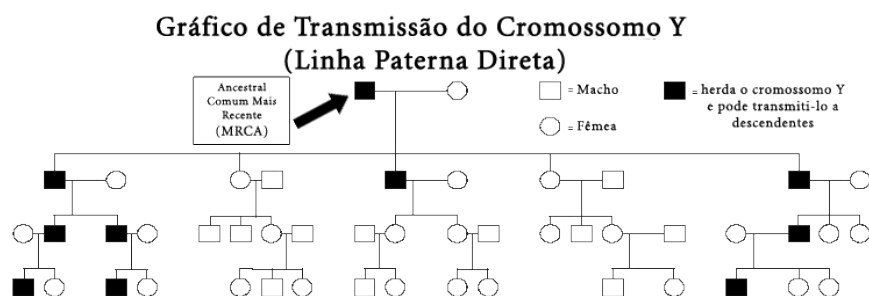


Figura 5 – herança do cromossomo Y: somente os descendentes masculinos pela linha paterna direta herdam o cromossomo Y do Ancestral Comum Mais Recente (MRCA) e podem transmiti-lo a seus descendentes.

Em sua aplicação à Genealogia Genética, o cromossomo Y é discutido em termos de passagens específicas⁵. Por exemplo, o DYS391 é um segmento do cromossomo Y caracterizado pela repetição do código AGAT. O número de repetições pode variar de 9 a 17. Assim, quando um laboratório de genealogia genética confere o valor 9 ao DYS391 de um certo indivíduo, a configuração de seu DYS391 é a seguinte: AGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT. Caso esse mesmo indivíduo possua um parente pela linha paterna direta que apresente o valor 10 no DYS391, esse parente pode ter possuído uma mutação que aumentou seu número de repetições do código AGAT de 9 para 10, configurando seu marcador DYS391 assim: AGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT.

Somente a testagem de outros parentes pela linha paterna direta, ou o contraste com

⁵ O DNA é formado por quatro bases (adenina – A, citosina - C, timina – T, e guanina - G), que se sucedem em cadeia e se unem a outros ingredientes para formar a dupla hélice que conhecemos. Segmentos específicos podem ser reconhecidos devido à repetição de códigos. Por exemplo, um segmento pode ser caracterizado pelas bases A, C, T e G, organizadas da seguinte maneira: GATTACA , e repetidas 5 vezes, formando: GATTACAGATTACAGATTACAGATTACAGATTACA. Esse segmento (“microsatélite”) diferencia-se do microsatélite anterior, que pode ser TTATTATTATTATTATTATTA e do posterior, que pode ser CAGTTCAGTTCAGTTCAGTT. Uma vez que o DNA humano foi sequenciado, tornou-se possível estabelecer a localização precisa desses microsatélites (*Short Tandem Repeats*, ou STR na sigla em inglês) em seus devidos cromossomos. Os microsatélites (STRs) do cromossomo Y recebem o termo DYS (ou DNA Y-chromosome Segment, na sigla em inglês) e um número, indicando sua posição nesse cromossomo – por exemplo, o DYS464. Também há microsatélites com nomes específicos, como YCAII, Y-GATA-H4 e CDY, todos no cromossomo Y. Uma vez que esses segmentos tenham recebido sua terminologia e numeração, também podem ser chamados de “marcadores” – por exemplo, o marcador DYS464 tem sua posição e configuração marcada no cromossomo Y.

outros indivíduos de um grupo próximo, pode determinar qual dos dois indivíduos apresentou a mutação desde que as linhagens deles se separaram de seu ancestral paterno direto mais recente (*Most Recent Common Ancestor*, ou MRCA na sigla em inglês). A comparação de dados poderia demonstrar que o valor ancestral provável para o marcador DYS391 fosse 9, por exemplo, indicando que o segundo indivíduo foi mesmo quem apresentou a mutação - criando mais uma repetição do código AGAT.

Para os fins da Genealogia Genética, 67 marcadores são geralmente testados para se estabelecer, com confiança, o número de mutações apresentada por cada um dos descendentes, assim como a distância cronológica entre eles e o Ancestral Comum Mais Recente e os prováveis valores desse ancestral em todos os marcadores testados. Resultados individuais são apresentados da seguinte maneira:

PANEL 1 (1-12)												
Locus	1	2	3	4	5	6	7	8	9	10	11	12
DYS#	393	390	19*	391	385a	385b	426	388	439	389-1	392	389-2
Alleles	13	22	14	10	13	14	11	14	11	12	11	28

PANEL 2 (13-25)													
Locus	13	14	15	16	17	18	19	20	21	22	23	24	25
DYS#	458	459a	459b	455	454	447	437	448	449	464a**	464b**	464c**	464d**
Alleles	15	8	9	8	11	23	16	20	28	12	14	15	15

PANEL 3 (26-37)												
Locus	26	27	28	29	30	31	32	33	34	35	36	37
DYS#	460	GATA H4	YCA II a	YCA II b	456	607	576	570	CDY a	CDY b	442	438
Alleles	10	10	21	21	15	14	16	20	34	38	12	10

PANEL 4 (38 - 47)												
Locus	38	39	40	41	42	43	44	45	46	47		
DYS#	531	578	395S1a	395S1b	590	537	641	472	406S1	511		
Alleles	11	8	15	15	8	11	10	8	9	9		

PANEL 4 (48 - 60)													
Locus	48	49	50	51	52	53	54	55	56	57	58	59	60
DYS#	425	413a	413b	557	594	436	490	534	450	444	481	520	446
Alleles	12	24	25	16	10	12	12	17	8	13	26	20	13

PANEL 4 (61 - 67)												
Locus	61	62	63	64	65	66	67					
DYS#	617	568	487	572	640	492	565					
Alleles	13	11	12	11	11	12	11					

*Also known as DYS 394

Figura 6 – resultados de 67 marcadores de indivíduo pertencente ao haplogrupo II. O conjunto de resultados de um indivíduo é chamados de haplotipo (*haplotype*, em inglês).

Além dos marcadores utilizados para a distinção entre indivíduos descendentes de um ancestral comum mais recente, utilizam-se testes para identificar mutações únicas em pares de bases. Tais variações, chamadas de Polimorfismo de Nucleotídeo Simples (*Single-Nucleotide Polymorphism*, ou SNP na sigla em inglês), ocorrem em frequência muito menor que mutações em microssatélites e, uma vez que tenham ocorrido, não se revertem, sendo passadas para todos os descendentes masculinos diretos do indivíduo afetado. Por exemplo, uma mutação pode causar a substituição de citosina por timina no segmento AAGCCTA, criando AAGCTTA. Tal mutação cria um SNP e passa a diferenciar um indivíduo e seus descendentes de toda a espécie humana a partir de então, e pode ser representada assim:

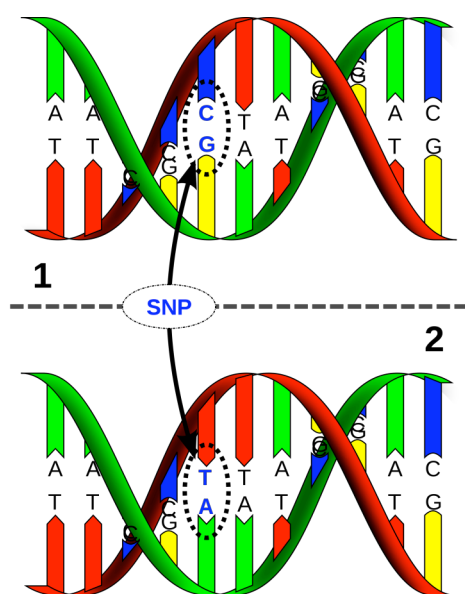


Figura 7 – exemplo de mutação de SNP (Wikipédia): um indivíduo passa a apresentar T em vez de C no segmento AAGCCTA, criando AAGCTTA. Essa mutação será carregada por seus descendentes, enquanto o restante da espécie humana continuará a apresentar AAGCCTA. A mutação em C leva a uma troca da base correspondente, que passa de G para A – pois C somente pode se ligar a G e T somente pode se ligar a A no DNA.

A importância das mutações que originam SNP's reside no fato de criarem “famílias” de cromossomo Y, uma vez que, ao serem transmitidas pelas gerações através dos milênios, criam um grande conjunto de indivíduos que possuem SNP's e também valores de marcadores em comum. Essas famílias são chamadas de haplogrupos. Por exemplo, uma mutação na posição M253 originou uma subdivisão do haplogrupo I*(marcado pelo SNP M170) e a fundação do haplogrupo I1 entre

15.000 e 20.000 anos atrás. O gráfico a seguir demonstra as subdivisões do haplogrupo I, sendo as mais maiores o haplogrupo I1 e o haplogrupo I2:

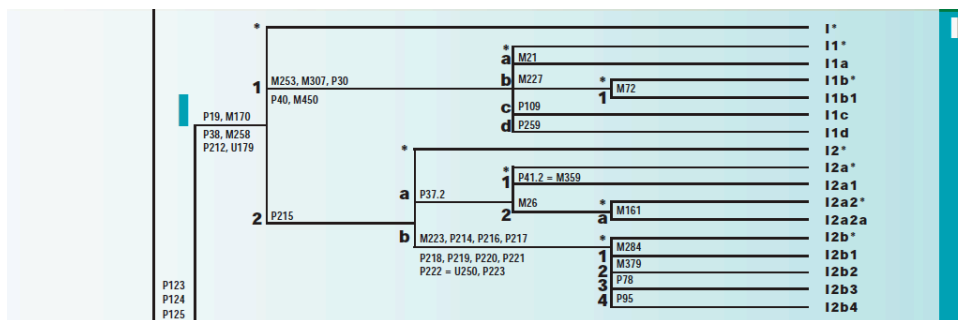


Figura 8 – subdivisões do haplogrupo I (Family Tree DNA): o haplogrupo I1 é marcado pelo SNP M253, entre outros, e o haplogrupo I2 é marcado pelo SNP P215. As subdivisões desses haplogrupos são marcadas por SNP específicos, como o haplogrupo I2b1, originado pelo SNP M284. Um indivíduo pertencente ao haplogrupo I2b1 possui o SNP M223 (seu ancestral na árvore) mas não possui o SNP M379 (seu “irmão”), que marca o haplogrupo I2b2. Da mesma maneira, um indivíduo do haplogrupo I2b2 possui o SNP M223 (seu ancestral), mas não possui o SNP M284 (seu “irmão”).

O haplogrupo I se insere da seguinte maneira na árvore dos haplogrupos:

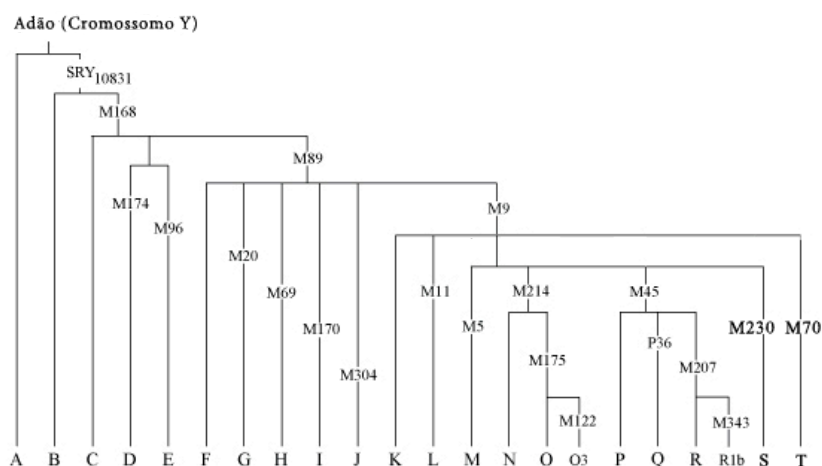


Figura 9 – árvore dos haplogrupos (ou *haplotree*, em inglês): cada subdivisão é gerada pelo aparecimento de um novo SNP. No topo da árvore, encontra-se o chamado “Adão do cromossomo Y”. Esse “Adão” é o ancestral comum mais recente de todos os haplogrupos existentes hoje, e não necessariamente o primeiro homem a andar sobre a face da Terra.

Para o genealogista genético, é essencial ter conhecimento do haplogrupo (ou sub-haplogrupo) exato a que pertence a linhagem estudada. Isso porque, mesmo que dois indivíduos possuam resultados idênticos em seus haplotipos (marcadores testados), o que indicaria um ancestral comum recente pela linha paterna direta, essa semelhança pode ser mera coincidência caso os indivíduos pertençam a haplogrupos diferentes. Por exemplo, caso dois indivíduos com os mesmos resultados em 67

marcadores testados pertençam um ao haplogrupo I1 e o outro, ao haplogrupo I2b1, a análise isolada dos marcadores poderia levar à incorreta conclusão de que os dois possuem um ancestral comum pela linha paterna direta no máximo 3 gerações atrás. No entanto, considerando-se que o haplogrupo I1 se separou do haplogrupo I* por volta de 20.000 anos atrás, e que o haplogrupo I2 se separou do haplogrupo I* por volta de 17.000 atrás, o fato é que o último ancestral comum pela linha paterna direta desses dois indivíduos viveu há ao menos 20.000 anos. Ancestrais comuns recentes erroneamente atribuídos a dois indivíduos pertencentes a haplogrupos diferentes são chamados de “ancestrais fantasmas”. Coincidências em resultados de marcadores são comuns e costumeiramente levam genealogistas a erros em suas investigações. A correta verificação dos haplogrupos é a única maneira de evitar esse problema.

Outra informação contida na determinação de haplogrupo de um certo indivíduo é a posição geográfica e cronológica de seu ancestral remoto. Cruzando informações a respeito da localização geográfica de indivíduos pertencentes a um dado haplogrupo com a variação de resultados de marcadores em localidades específicas, o que indica diversidade genética, é possível ter-se uma ideia da localização geográfica e época de origem desse haplogrupo. Por exemplo, o indivíduo que apresentou o SNP M253 pela primeira vez e foi o fundador do haplogrupo I1 teria vivido na Península Ibérica, de acordo com o Genographic Project. A variação genética de seus descendentes data a existência desse indivíduo entre 16.000 e 20.000 anos atrás, durante o Último Máximo Glacial, como citado anteriormente.

DNA Mitocondrial

A mitocôndria é uma organela responsável pela respiração celular, passada de uma mãe para seus filhos. Uma vez que a mitocôndria encontra-se no óvulo – as

mitocôndrias do espermatozóide são destruídas quando da fertilização—, ela não pode ser transmitida pela linha masculina. Por esse motivo, e também por a mitocôndria não sofrer recombinações genéticas com outras organelas no interior da célula, a mesma pode ser utilizada para a investigação da linha direta feminina assim como o cromossomo Y é usado para a pesquisa da linha paterna direta. Diferentemente do cromossomo Y, no entanto, a presença da mitocôndria independe do sexo. Assim, um indivíduo do sexo masculino poderá testar sua mitocôndria para obter maiores informações sobre a linhagem de sua mãe – embora, como acima explicado, esse homem não possa transmitir a mitocôndria herdada de sua mãe para seus filhos.

O primeiro sequenciamento do DNA mitocondrial foi realizado pela Cambridge University em 1981 e constituiu o pontapé inicial ao Projeto Genoma Humano. O resultado foi o chamado *Cambridge Reference Sequence (CRS)*, na sigla em inglês), que então passou a ser utilizado como padrão para o sequenciamento do DNA mitocondrial de outros indivíduos. Nesse modelo, somente as diferenças entre o indivíduo testado e o CRS são anotadas. Tais diferenças provêm de mutações, que podem estar presentes em um pequeno número de indivíduos (descendentes de uma ancestral relativamente recente) ou em um grande número de indivíduos, que descendem de uma ancestral remota e formam, assim, um haplogrupo de DNA mitocondrial. Resultados individuais são apresentados da seguinte maneira:

Haplogroup - A

HVR1 differences from CRS			HVR2 differences from CRS		
16111T	16189C	16223T	64T	73G	146C
16290T	16319A	16362C	153G	185A	235G
16468C	16519C		263G	309.1C	309.2C
			315.1C	338T	522-
			523-		

Figura 10 – resultados de DNA mitocondrial de um indivíduo pertencente ao haplogrupo mitocondrial A (Family Tree DNA): os valores anotados correspondem às diferenças entre o DNA mitocondrial testado e o CRS. Por exemplo, 16111T indica que na posição 16111 da mitocôndria o indivíduo possui uma base (timina, ou T) diferente da base apresentada pela CRS. Essa diferença indica uma mutação. Algumas das diferenças indicam

que o indivíduo pertence ao haplogrupo A⁶. Outras são apresentadas somente por descendentes de uma ancestral comum mais recente, e não constituem um haplogrupo.

Mais tarde, aprendeu-se que o indivíduo testado pela Cambridge University pertencia ao haplogrupo denominado H2a2a. A árvore dos haplogrupos de DNA mitocondrial possui a seguinte configuração:

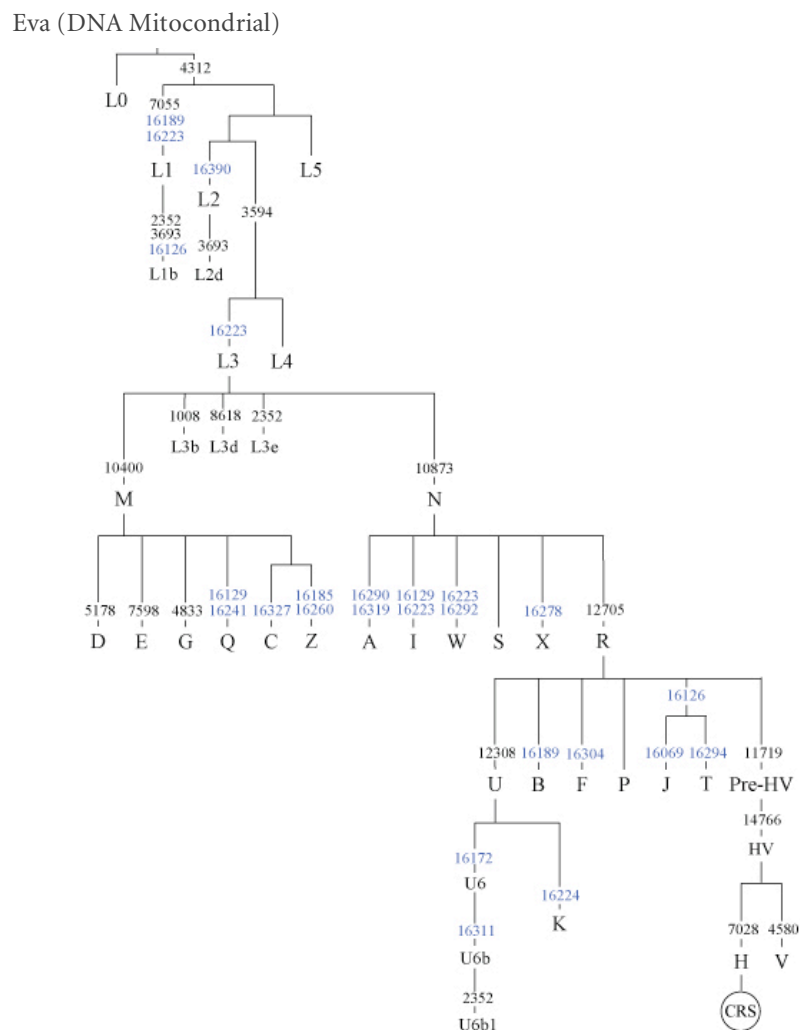


Figura 11 – árvore dos haplogrupos mitocondriais (GeneBase): o haplogrupo A do indivíduo representado na figura 10, acima, foi originado por uma mulher pertencente ao haplogrupo N. O haplogrupo H, utilizado no *CRS*, também pode ser visto. A ancestral comum mais recente de todas as mitocôndrias humanas hoje em existência chama-se “Eva do DNA mitocondrial”.

Uma das particularidades da aplicação do DNA mitocondrial na pesquisa genealógica tem a ver com a taxa de mutação dessa parte de nosso genoma, que é muito mais baixa que a taxa de mutação do cromossomo Y, por exemplo. Para o

⁶ As diferenças que definem o haplogrupo mitocondrial A são: 152C, 235G, 663G, 1736G, 4248C, 4824G, 8794T, 16290T e 16319A. Fonte: <https://www.familytreedna.com/mtDNA-Haplogroup-Mutations.aspx>.

genealogista, a implicação é que exames de DNA mitocondrial não possuem a mesma exatidão cronológica que os do cromossomo Y. Resultados idênticos nos painéis geralmente utilizados para a testagem do DNA mitocondrial (HVR1 e HVR2) indicam uma probabilidade de 50% de uma ancestral comum ter vivido até 700 anos atrás. Dessa forma, são necessários exames mais profundos, um sequenciamento total do DNA mitocondrial, para se determinar a época exata da existência desse ancestral.

Assim como ocorre com o cromossomo Y, entretanto, a determinação do haplogrupo mitocondrial é uma rica fonte de informações sobre as origens remotas da linha materna direta. O haplogrupo A citado acima, por exemplo, originou-se entre 30.000 e 50.000 anos atrás no leste da Ásia, e nos dias atuais é encontrado com maior frequência entre os nativos das Américas. De fato, no Brasil, 33% da população considerada “branca” possui DNA mitocondrial ameríndio (haplogroups A, B, C, D e X)- o DNA mitocondrial africano representa outros 28% do total, e o europeu, os 39% restantesⁱ. Essa frequência contrasta com o cromossomo Y encontrado nos homens “brancos” no país, que é em sua grande maioria de origem europeia – 98%, para ser mais exato.

Sob o ponto de vista da saúde, o sequenciamento do DNA mitocondrial também oferece um vasto potencial, uma vez que determinadas mutações na mitocôndria possuem ligação direta com doenças genéticasⁱⁱ. Por esse motivo, exames como este estarão cada vez mais em voga através da medicina molecular, que tende a ser a medicina do futuro.

Métodos

Um dos avanços tecnológicos mais importantes também para os genealogistas nas últimas décadas foi a internet, que permitiu a fácil compartilhamento de dados e

mais rápido progresso nas pesquisas através do trabalho conjunto. A chave para o sucesso na genealogia genética também reside na partilha de bases de dados, mas de DNA. Ao realizar seus exames de DNA, um genealogista pode optar por ser avisado pelo laboratório sobre resultados similares conforme outros indivíduos são testados e seus dados entram no sistema. Há também bases de dados públicas que permitem que indivíduos testados por empresas diferentes comparem seus resultados com os de outros usuários. Uma vez que haplotipos semelhantes sejam encontrados, é então possível contactar os outros pesquisadores e contrastar registros documentais, possibilitando progresso mútuo. Sem o compartilhamento de dados, não há genealogia genética, pois somente as origens remotas e questões médicas podem ser deduzidas sem trabalho em conjunto com outros genealogistas. Os métodos discutidos por este artigo partem da suposição de que a informação é compartilhada.

Que Laboratórios Utilizar

Ao realizar uma pesquisa em um motor de buscas por laboratórios de genealogia genética, um genealogista poderá se surpreender com a grande quantidade de resultados - de fato, há inúmeras empresas que oferecem esses serviços. Entretanto, poucas possuem bases de dados grandes o suficiente, ou o correto foco em genealogia, de forma a oferecer um produto realmente atraente ao genealogista. No momento presente, as duas empresas mais recomendadas aos fins da genealogia são a Family Tree DNA e a 23andMe.

A empresa Family Tree DNA (FT DNA) nasceu em 1999, com a missão de aplicar a genética à genealogia. No ano de 2009, a empresa já havia testado mais de 500.000 kits – em parceria, entre outros, com o Genographic Project. A empresa tornou-se famosa por seus testes de cromossomo Y e DNA mitocondrial, criando

padrões posteriormente copiados por outras empresas. Recentemente, a FT DNA passou a oferecer também testes de DNA autossômico, e pretende permitir a “importação” a sua base de dados de resultados de testes deste tipo que foram realizados por outras empresas. Por sua vez, a 23andMe foi fundada em 2006 com o objetivo de permitir a seus clientes uma maior compreensão de seu genoma. A empresa sempre focou na saúde e no futuro da medicina molecular através do oferecimento de exames de SNP do DNA autossômico, tendo sido adotada por genealogistas desejosos por obter informações sobre seus diversos ancestrais por linhas que não as diretas, paterna e materna.

Ainda hoje, a Family Tree DNA mantém a dominância nos testes de cromossomo Y e DNA mitocondrial, oferecendo produtos eficazes e facilmente aplicáveis à genealogia. Seu avanço nos testes autossômicos promete roubar clientes da 23andMe, principalmente os genealogistas, uma vez que a grande maioria do público desta última não pratica a genealogia e, assim, não possui informações úteis às pesquisas familiares – ou o desejo de participar das mesmas. O tamanho da base de dados de DNA autossômico da Family Tree DNA deve, assim, rapidamente ultrapassar o de sua competidora, dando ainda mais motivos para que genealogistas utilizem-se dela para todos os seus exames. A 23andMe, entretanto, deve manter sua liderança na testagem genética para fins de medicina molecular.

Testes de DNA Autossômico

O principal propósito da aplicação da testagem do DNA autossômico à genealogia é encontrar passagens herdadas de um ancestral obscuro, passagens estas que o genealogista descobre possuir em comum com outro pesquisador, que pode possuir maiores informações sobre o dado ancestral ou passar a colaborar na pesquisa

a respeito do mesmo. Abaixo encontra-se um exemplo real das passagens autossômicas em comum entre um indivíduo, seu pai, seu avô e seu bisavô:

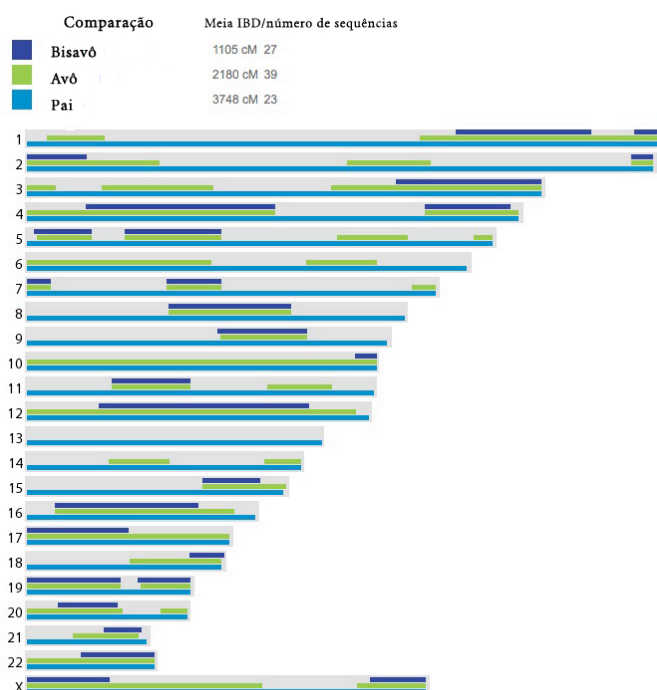


Figura 12 – seqüências autossômicas em comum entre um filho, seu pai, seu avô e seu bisavô (yourgeneticgenealogist.com): é possível notar que o filho possui exatamente 50% de seu genoma em comum com seu pai (em azul claro), ou seja: 1 cromossomo de cada par. Nesses cromossomos, a quantidade em comum com o avô paterno (em verde abacate) é menor que a quantidade em comum com o pai. A quantidade em comum com o bisavô paterno, em azul escuro, é menor ainda. A quantidade de DNA herdada de um dado ancestral é menor quanto mais distantes as gerações.

A unidade utilizada para medir distância em um cromossomo é o centiMorgan (cM). Para a empresa Family Tree DNA, a existência de um segmento comum entre 5 e 10 centiMorgans sugere parentesco entre dois indivíduos. Já qualquer segmento maior que 10 centiMorgans significa certeza de um ancestral em comumⁱⁱⁱ. Já a empresa 23andMe considera somente segmentos maiores que 7cM. Uma vez que um segmento comum desse tamanho ou maior seja encontrado entre dois indivíduos, a empresa também aponta os segmentos de, no mínimo, 5cM entre esses indivíduos. Segundo a empresa 23andMe, a quantidade de DNA compartilhada entre um indivíduo e seus primos é a seguinte: primo em primeiro-grau – 12.5%; primo em segundo-grau – 3.13%; primo em terceiro-grau – 0.78%; primo em quarto-grau –

0.20%; primo em quinto-grau – 0.05%; primo em sexto-grau – 0.01%; primo acima do sétimo-grau – menor que 0.01%. Isso se deve à maneira aleatória como o DNA autossômico é passado de pais para filhos e à recombinação. Veja nos quadros abaixo exemplos reais de DNA compartilhado entre um indivíduo e membros de sua família:



Figura 13 – sequências autossômicas em comum entre um indivíduo e seu tio-bisavô (irmão da avó paterna de seu pai).



Figura 14 – sequências autossômicas em comum entre um indivíduo e sua tia-avó (irmã da mãe de sua mãe).



Figura 15 – sequências autossômicas em comum entre um indivíduo e sua avó paterna. As passagens em cinza foram, então, herdadas do avô paterno. As figuras 13 e 15 podem ser sobrepostas, pois encontram-se na mesma metade do par de cromossomos.

A maneira aleatória como os cromossomos somáticos são transmitidos de pais para seus filhos é responsável por uma das maiores dificuldades na utilização do DNA autossômico na genealogia genética: a correta identificação do ancestral comum do qual dois indivíduos herdaram a mesma sequência de DNA. Somente através da pesquisa documental isso se faz possível, a partir da comparação entre as árvores genealógicas inteiras dos dois indivíduos para a seleção de linhas candidatas. É necessário que ambos os pesquisadores tenham conhecimento de suas árvores, assim como interesse em compartilhar as informações.

Uma vez que uma sequência em comum tenha sido encontrada com outro indivíduo, também é possível utilizar-se o que se sabe sobre o próprio DNA de forma a encontrar o ancestral comum que originou a similaridade com o outro pesquisador. Tomando por exemplo o indivíduo das figuras 13, 14 e 15, acima, na dúvida sobre a origem de um parentesco, ele pode descartar o que é conhecido. Digamos que tampouco a avó paterna, o bisavô paterno ou a tia-avó materna possuam passagens em comum com o recém-descoberto parente. Isso nos deixa duas possibilidades: o parentesco com o outro pesquisador pode provir do bisavô pela linha paterna direta (que casou-se com a irmã do tio-bisavô paterno representado na figura 13) ou do avô materno (que casou-se com a irmã da tia-avó materna representada na figura 14). Aqui fazemos uma inferência: certamente herdamos um pouco mais de DNA de um avô ou bisavô que o que possuímos em comum com o irmão ou irmã deles; entretanto, é muito baixa a probabilidade de, justamente nessas minúsculas passagens a mais, possuímos uma passagem em comum com outro indivíduo, sem que os resultados dos próprios tios apontem qualquer parentesco. De forma a simplificar a questão, podemos sobrepor as figuras 13 e 15, que representam a mesma metade dos pares de cromossomos que foi herdada pelo lado paterno. Dá-se o seguinte gráfico:

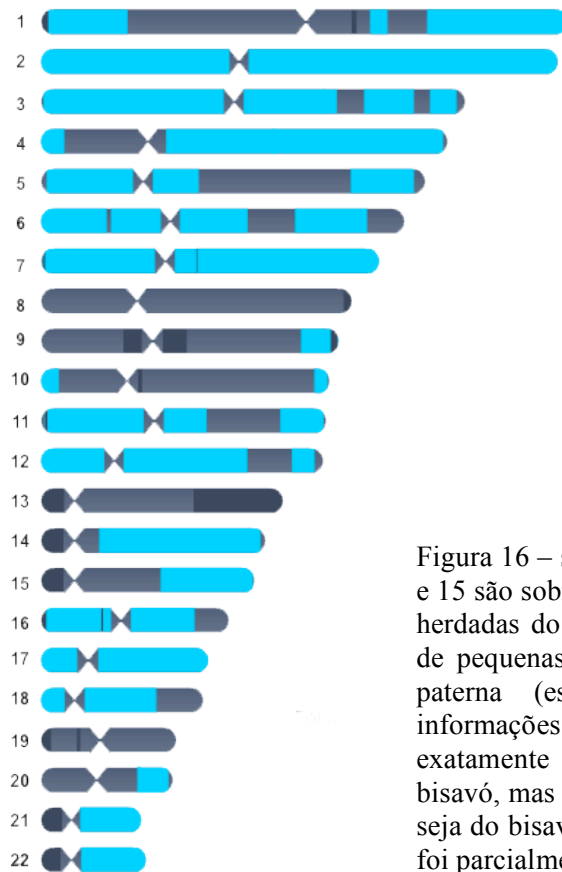


Figura 16 – sequências autossômicas das figuras 13 e 15 são sobrepostas, revelando passagens em cinza herdadas do bisavô pela linha paterna direta, além de pequenas partes adicionais herdadas da bisavó paterna (esposa do dito bisavô): com as informações em mãos, não é possível distinguir exatamente as passagens adicionais herdadas da bisavó, mas presume-se que a maior parte em cinza seja do bisavô, pois a herança genética da bisavó já foi parcialmente representada por seu irmão.

Poderíamos, então, sobrepor as passagens em comum com o outro pesquisador a este gráfico. Caso as mesmas se encaixem perfeitamente nas passagens em cinza herdadas do bisavô paterno direto, podemos investigar esse lado da árvore genealógica em busca do ancestral comum. Caso as passagens não se sobreponham com os blocos herdados desse bisavô, podemos, então, investigar a árvore genealógica do avô materno – e as passagens do novo parente se encaixarão, então, nos blocos em cinza da figura 14 acima. Em qualquer caso, a colaboração por parte do outro pesquisador é essencial para que se possa obter maiores informações sobre o ancestral específico que originou o parentesco.

Testes de DNA Mitocondrial

Diferentemente da testagem do DNA autossômico, ou mesmo de testes do cromossomo Y, a investigação genealógica através do DNA mitocondrial não deixa

tanto espaço para a interpretação. Comparado com o restante de nosso genoma, na mitocôndria ocorrem ainda menos mutações, o que significa que não é fácil diferenciar entre parentes próximos ou afastados quando possuem resultados idênticos. Dois indivíduos com os mesmos valores no painel básico, chamado de HVR1, possuem somente 50% de chances de possuírem uma ancestral direta comum nos últimos 1.300 anos. Caso essa semelhança nos resultados se confirme no painel de alta resolução (HVR2), as chances passam a ser de 50% de que a ancestral comum na linha materna direta dos dois indivíduos tenha vivido nos últimos 700 anos. Por outro lado, resultados díspares não deixam sombra de dúvida: sua ancestral comum viveu há milênios.

Por motivos como estes – e também pelo machismo que permeia nossa sociedade-, poucos pesquisadores se aventuram a se guiar pelo DNA mitocondrial em suas pesquisas genealógicas. Assim, muitos acabam se contentando com testes de haplogrupo que trazem informações sobre as origens remotas da linha materna - se a mesma possui raízes nas Américas ou na Europa, etc... Entretanto, não se pode esquecer que uma pioneira da genealogia genética – a doutora Ann Turner – começou suas investigações na área ao contrastar o seu DNA mitocondrial com o de uma prima. Também não se pode ignorar o papel da mitocôndria em nossa saúde, uma vez que mutações nessa organela são a causa de inúmeras doenças genéticas – descobrir se somos portadores ou não dessas doenças pode ser essencial às decisões que tomaremos no futuro. Ademais, recentemente tornou-se acessível aos consumidores o mapeamento inteiro do DNA mitocondrial – que pode ser muito útil tanto à genealogia, quanto à medicina.

O exame, disponibilizado pela empresa FT DNA sob o nome *mtDNA Full Sequence*, coloca a testagem do DNA mitocondrial à altura do cromossomo Y ou do

DNA autossômico em sua aplicação na genética. De fato, resultados idênticos neste teste significa que há 91% de chances de a ancestral comum mais recente de dois indivíduos ter vivido nos últimos 600 anos. Trata-se de uma cronologia bastante razoável, especialmente ao considerarmos que será possível chegar a essa ancestral por meio de, no mínimo, dois caminhos documentais na pesquisa tradicional.

Testes de Cromossomo Y

O padrão criado pelo time da Family Tree DNA para testes de cromossomo Y se tornou modelo para toda a indústria. A empresa oferece quatro painéis de testes de marcadores – 12, 25, 37 e 67-, além do teste de *Deep Clade* (haplogrupo), formulados de forma a se obter taxas médias de mutação que são específicas para cada painel. Isso é possível porque é sabido que determinados marcadores possuem taxas de mutação maiores que outros. Por exemplo, o marcador DYS388 possui taxa de mutação de apenas 0.00022. O DYS724 (ou CDY), por outro lado, possui mutações muito mais frequentes, com uma taxa de 0.03531 – ou seja, uma mutação a cada 28 gerações. Assim, pode-se agrupar marcadores de forma que as taxas de mutação de cada painel sirvam a uma função própria na investigação genealógica^{iv}. (Os marcadores e sua distribuição nos painéis da Family Tree DNA podem ser vistos na figura 6, acima).

12 marcadores

O teste de 12 marcadores da FT DNA foi criado de forma a incluir microssatélites com menor taxa de mutação. A taxa média de variação desses marcadores é de .002 (0.00187 ± 0.00028 , de acordo com John Chandler^v). Ao multiplicarmos essa taxa por 12 marcadores, temos uma taxa de .024 para o total do painel. Para obtermos a quantidade de gerações para cada mutação, dividimos 1

mutação por .024 e obtemos 41.6, ou seja, aproximadamente 41 gerações para que haja uma mutação neste painel. Uma vez que, historicamente, a idade média para procriação pode ser considerada como sendo de 25 anos, multiplicamos 41 por 25 e obtemos 1.025 anos – ou seja, uma única diferença nesse painel representa, aproximadamente, uma distância cronológica de 1025 anos entre o nascimento dos dois indivíduos em questão e seu ancestral comum mais recente.

Como é conhecido, a adoção de sobrenomes na Europa passou a se popularizar a partir do ano 1.300 d.C.. Assim, na tradição anglo-germânica, mais rígida na manutenção de sobrenomes nas linhas paternas que a tradição luso-brasileira, caso dois indivíduos possuam o mesmo sobrenome e possuam resultados idênticos nos primeiros 12 marcadores, é bem provável que descendam do primeiro indivíduo que adotou esse sobrenome. No Brasil e em Portugal, em que uma maior rigidez quanto a sobrenomes não foi adotada até mais recentemente, é necessário uma semelhança também em outros painéis de forma que se possa tirar conclusões sobre a relação entre sobrenomes e ancestrais comuns sem a ajuda de registros documentais.

Por esses motivos, o painel de 12 marcadores deve ser sempre utilizado juntamente a outros painéis na investigação genético-genealógica. A única circunstância em que o uso isolado do painel de 12 marcadores é aconselhável é em projetos de família em que vários indivíduos já tiveram seus 67 marcadores testados, sendo conhecido o comportamento genético dos indivíduos descendentes de um dado ancestral e estando os registros documentais em completa ordem. Neste caso, o painel de 12 marcadores serve apenas como uma confirmação do que já é conhecido.

25 marcadores

O painel de 25 marcadores adiciona 13 marcadores aos 12 previamente testados. Esses treze novos marcadores possuem uma taxa de mutação média de

0.0029^{vi} (ou 0.00278 ± 0.00042 , de acordo com John Chandler). A taxa de mutação média destes 25 marcadores passa a ser, então, 0.0027. Ao multiplicarmos esse valor por 25, obtemos 0.0675, ou seja, uma mutação a cada 15 gerações, aproximadamente. Esse valor nos indica que resultados idênticos entre dois indivíduos em todos os 25 marcadores nos leva a um ancestral comum nascido aproximadamente 375 anos antes dos dois indivíduos comparados, ou seja, provavelmente no século XVI ou XVII. Na tradição anglo-saxônica, muito certamente isso significa que, possuindo dois indivíduos o mesmo sobrenome, eles então descendem do mesmo ancestral paterno direto – uma vez que são raros os casos de mudanças de sobrenome nas linhagens paternas após essa época. Na tradição luso-brasileira, em que trocas de sobrenome ou heranças por vias maternas são comuns, o painel de 25 marcadores permite, ao menos, um avanço a uma época que é mais fácil de ser verificada por meio de registros documentais.

37 marcadores

Após mudanças implementadas no ano de 2010, 37 é o número mínimo de marcadores disponível à venda pela FT DNA a indivíduos que não estão devidamente “enraizados” em projetos de família através de registros documentais. A medida foi tomada de forma a evitar interpretações errôneas de resultados por pesquisadores novatos na área de genealogia genética. Os 12 marcadores adicionados no painel de 37 marcadores possuem taxa de mutação alta, a 0.0071, trazendo a taxa de mutação média do painel a 0.00492 ± 0.00074 , segundo John Chandler. Resultados idênticos neste painel indicam um ancestral comum pela linha paterna direta que pode ter vivido há cinco gerações. De fato, a alta taxa de mutação neste painel funciona como uma peneira bastante exigente, deixando passar somente parentes relativamente próximos. Pode-se dizer que, dado o número de marcadores, o painel de 37

marcadores é aquele em que mais se encontra diferenças entre haplotipos relacionados.

Por tais motivos, esse painel é extremamente útil a genealogistas portugueses e brasileiros, uma vez que as leis que regem o uso de sobrenomes em Portugal e Brasil somente se tornaram mais rígidas no século XIX ou início do século XX, dificultando as trocas frequentes que costumavam marcar os nomes de família nesses países. Assim, indivíduos com resultados idênticos neste painel e que possuem o mesmo sobrenome, têm grandes chances de possuir um ancestral comum mais recente (MRCA) que utilizava esse nome. Mesmo que seus sobrenomes sejam diferentes, a proximidade genética significa que é mais fácil encontrar o ancestral comum através da pesquisa documental conjunta, pois o período cronológico é curto e mais preciso.

67 marcadores

A criação do painel de 67 marcadores teve como função tornar a aplicação da genética à genealogia uma tarefa ainda mais exata. Aos 37 marcadores anteriores, são adicionados 30 marcadores com taxas de mutação média ou baixa. A taxa média de mutação dos marcadores no painel é de 0.003332 de acordo com Kaye N. Ballantyne^{vii}, excluindo-se os marcadores YCAIIa, YCAIIb, DYS607, CDYa, CDYb, DYS395S1a, DYS395S1b, DYS406S1, DYS413a e DYS413b. Isso significa um valor de .189918 no total de marcadores, ou seja, resultados idênticos em todos os microssatélites confirmam um ancestral comum pela linha paterna direta até cinco gerações atrás. Estas projeções são seguras e conservadoras, pois na prática é comum encontrar indivíduos que possuem um ancestral comum há somente quatro ou cinco gerações e que, mesmo assim, apresentam uma ou duas mutações no painel de 67 marcadores.

Nos resultados, pode-se notar que o indivíduo na linha 4 é a mesma pessoa cujos resultados se encontram na figura 6, acima. O indivíduo na segunda linha compartilha com ele um ancestral comum mais recente nascido em meados do século XIX – mas, mesmo assim, possuem uma diferença no marcador CDY, em que a linhagem do segundo indivíduo apresentou uma mutação após a separação. Os indivíduos nas linhas 1, 3 e 5 separaram-se dos indivíduos na linha 2 e 4 relativamente cedo. Pouco depois, separou-se o indivíduo 1 dos indivíduos da linhagem de 3 e 5. A proximidade genealógica dos indivíduos nas linhas 3 e 5 é provavelmente parecida à existente entre os indivíduos nas linhas 2 e 4, e suas linhagens devem ter se separado no final do século XVIII ou início do século XIX. Estima-se que o ancestral comum mais recente de todos os indivíduos acima tenha vivido por volta dos anos 1300 a 1500 no noroeste da Península Ibérica. Algumas das linhas de pesquisa histórico-documentais já estão próximas desse período, e para os pesquisadores será uma questão de tempo até que esse ancestral seja encontrado.

Conclusão

A Genealogia Genética é um campo em constante desenvolvimento, assim como as ciências biológicas que a guiam. Descobertas no ramo da genética são feitas a cada dia e recebem imensas contribuições de genealogistas de todo o mundo, que fornecem dados, casos para estudo, e seus próprios *insights*. A genealogia também possui muito a ganhar com a genética, que é capaz de cobrir as imensas lacunas documentais que são comuns na história de países como o Brasil, em que grande parte dos colonizadores veio de outros países, deixando suas histórias para trás. A genética é igualmente útil à pesquisa de linhagens não europeias – ameríndias ou africanas –, que muitas vezes nunca foram devidamente documentadas mas que podem ter suas

histórias desvendadas através da desconstrução biológica das camadas de miscigenação. Assim, a Genealogia Genética é acessível e democrática, possuindo também o enorme papel de confirmar árvores traçadas no passado, muitas vezes sem o devido cuidado ou embasamento documental. A Genealogia Genética é a genealogia do futuro, futuro este que já chegou.

Referências

-
- ⁱ Alves-Silva, Juliana, Magda Da Silva Santos, Pedro E. M. Guimarães, Alessandro C. S. Ferreira, Hans-Jürgen Bandelt, Sérgio D. J. Pena, and Vania Ferreira Prado. "The Ancestry of Brazilian MtDNA Lineages." *The American Journal of Human Genetics*. Web. 07 Mar. 2011. <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1287189/>>.
- ⁱⁱ "Types of Mitochondrial Disease." *United Mitochondrial Disease Foundation*. Web. 07 Mar. 2011. <http://www.umdf.org/site/c.otJVJ7MMIqE/b.5692881/k.4B7B/Types_of_Mitochondrial_Disease.htm>.
- ⁱⁱⁱ Fonte: Family Tree DNA <https://www.familytreedna.com/faq/answers/default.aspx?faqid=17#608>
- ^{iv} Kerchner, Jr., P.E., Charles F. "DNA Mutation Rates - An Overview and Discussion." *Index Page*. Web. 08 Mar. 2011. <<http://www.kerchner.com/dnamutationrates.htm>>.
- ^v Chandler, John F. "Estimating Per-Locus Mutation Rates." *Journal of Genetic Genealogy*. 2006. Web. 08 Mar. 2011. <<http://www.jogg.info/22/index22.htm>>.
- ^{vi} Athey, Whit. "Editor's Corner: Mutation Rates – Who's Got the Right Values?" *Journal of Genetic Genealogy*. Web. 08 Mar. 2011. <<http://www.jogg.info/32/editorial.htm>>.
- ^{vii} Ballantyne, Kaye N. "AJHG - Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications." *Cell*. The American Journal of Human Genetics, 02 Sept. 2010. Web. 08 Mar. 2011. <[http://www.cell.com/AJHG/abstract/S0002-9297\(10\)00419-2](http://www.cell.com/AJHG/abstract/S0002-9297(10)00419-2)>.