

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA
MOLECULAR
(PPGBM)

Forças seletivas que atuam na dinâmica do alelo mutante do gene
ATXN2

LUCAS SCHENATTO DE SENA

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito para a obtenção do grau de Mestre.

Orientadora: Laura Bannach Jardim

Porto Alegre, março de 2018.

Financiamento

Esse projeto foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq- através da bolsa de estudo oferecida a Lucas Schenatto de Sena.

O financiamento para a aquisição de fotocópias dos questionários e demais documentos que se fizeram necessários ao longo dessa pesquisa foram fornecidos pelo FIPE- HCPA (Fundo de Pesquisa e Eventos- Hospital de Clinicas de Porto Alegre).

Houve, ainda, a liberação de uma linha telefônica do HCPA para investigação também fornecida pela FIPE- HCPA.

Agradecimentos:

Se eu vi mais longe, foi por estar sobre ombros de gigantes.

Isaac Newton

Todo trabalho intelectual e conseqüentemente científico é expressão de um processo coletivo e nesse modesto estudo para entender os mecanismos evolutivos envolvidos na SCA2 não seria diferente. Em primeiro lugar, estamos aqui, nas condições atuais, graças ao suor e sangue daqueles que ousaram fazer Ciência nas situações mais adversas, se chocando com a ordem vigente acalçada em um profundo obscurantismo religioso e/ou político.

Galileu Galilei foi condenado pela igreja na Idade Média por defender que a Terra girava em torno do sol e só se salvou da fogueira por fazer uma autocrítica e renegar suas ideias. Já Giordano Bruno, não capitulou e seguiu sustentando a ideia de que o Universo era infinito e morreu queimado. Pietro d'Abano, filósofo, matemático, físico e professor de Medicina e morreu em uma prisão da Inquisição acusado de bruxaria.

No entanto, a Ciência não sofreu apenas na mão da Igreja Católica da Idade Média. No Brasil, a ditadura militar impactou no encerramento ou prejuízo de 471 carreiras científicas e na morte de Ana Rosa Kucinski Silva, professora do instituto de Química da USP, com apenas 32 anos.

Hoje, pós democratização, a Ciência sofre com profundos ataques, sucessivos cortes de investimentos sob o governo de Temer e com o fortalecimento do obscurantismo pentecostal que atua com toda energia para combater a razão científica em detrimento da ignorância. Mas, apesar desse cenário adverso a comunidade científica resiste e vem travando importantes embates lutando em defesa de uma Ciência de excelência e que

contribua para a soberania do país. E por aqueles que lutaram e lutam que vão os meus primeiros agradecimentos.

Agora quero dedicar o restante do meu agradecimento ao grupo de Neurogênica no qual aprendi muito. Dentre toda equipe quero dar uma ênfase especial aos Professores: Dra Maria Luiza e Dr Jonas Saute, que fizeram importantes observações metodológicas a esse trabalho. Ao Dr Raphael Castilhos que me ensinou os aspectos principais para que essa pesquisa fosse realizada. Além do convívio cotidiano de minhas colegas Gabriela Bonzan, Anastacia Guimarães, Helena Fussier, Gabriel Furtado e Eduardo Mattos que ao longo desse projeto fizeram aportes importantes.

Aos pacientes e familiares pela colaboração para que esse projeto ocorresse e por apostar na Ciência como saída para resolver um grave dilema pessoal que suas famílias vem sofrendo com a doença.

Ao Elmo, secretário do PPGBM, uma pessoa extraordinária de prestatividade incomensurável, nos tranquilizando e ajudando em momentos em que tudo se desenhava para dar errado. Ao meu colega e amigo Thiago Corrêa, pela amizade estabelecida nessa caminhada e por se prestar a assistir todos os meus ensaios de qualquer apresentação que iria fazer. E aos demais amigos que fiz no PPGBM.

Ao meu filho e a companheira pelo incentivo e paciência que tiveram comigo nesses últimos dois anos e por serem um fator importante de estabilidade pessoal e fonte de motivação em situações bastante adversas, até mesmo quando eu acreditava que o que estava fazendo tinha pouco sentido eles foram a contra mola que me motivava novamente.

A minha mãe que teve uma atitude determinante para eu ingressar no programa e conseqüentemente concluir esse trabalho. Quando saiu o edital da seleção e constava na bibliografia “Introdução a Genética” do Griffiths, custava, em torno, de 300 reais. Eu estava em uma situação financeira apavorante.

Lembro que liguei para ela e falei que precisava de um livro. Ela que é empregada doméstica e não ganha uma fortuna, não titubeou e me deu o livro de presente, apesar dela ter todos os motivos para desconfiar de mais um “grande plano meu”.

E a professora Laura. A primeira coisa que eu esperava encontrar em um ambiente acadêmico era a vaidade e a arrogância. No entanto, tive sorte na escolha da orientadora. A postura que observei da Laura foi radicalmente oposta ao que esperava. Além de propiciar um ambiente acolhedor a alguém que vinha da completa marginalidade acadêmica, em nenhum momento usou de sua autoridade intelectual e hierárquica para impor seu ponto de vista em determinado rumo da pesquisa.

Essas suas características foram determinantes para o projeto tomar o rumo que tomou. Nosso projeto inicial se baseava no primeiro momento em análise da fertilidade em SCA2 “e ponto”. No entanto, o ambiente democrático, as críticas construtivas e sua preocupação permanente em formar seus alunos possibilitaram uma análise mais ampla a questão.

Sumário

	Página
1. Resumo	7
2. Abstract	11
3. Introdução	14
3.1 Ataxia espinocerebelar tipo 2	14
3.2 Seleção natural e adaptabilidade darwiniana	17
3.3 Pleiotropia antagônica	21
3.4 Fitness genético	22
3.5 A possibilidade do aumento do fitness nas doenças poliQ	23
4. Objetivo	26
5. Artigo	27
6 Discussão	45
6.1 Dinâmica do Alelo Expandido associado a SCA2	45
6.2 Múltiplas origens ou uma origem única	57
7. Conclusão	61
8. Perspectiva	63
9. Referência	64

Resumo

A ataxia espinocerebelar tipo 2 (SCA2) é causada por uma expansão em uma sequência repetitiva de códons CAG (CAG_{exp}) no exon 1 do gene ATXN2. O trato expandido de poliglutaminas está associado a um ganho de função tóxica, expressa predominantemente em neurônios. A SCA2 é uma condição neurodegenerativa autossômica dominante, com início, em média, entre os 30 e os 40 anos, manifesta-se por distúrbios neurológicos progressivos que levam os afetados à cadeira de rodas e à sobrevida reduzida. Embora propensa a forças seletivas negativas como a antecipação, a frequência da SCA2 parece ser estável nas populações onde foi observada, sugerindo a existência de mecanismos seletivos favoráveis simultâneos. Para averiguar essa hipótese, o presente estudo estimou o fitness, a segregação e o papel da antecipação em uma coorte de famílias com diagnóstico molecular de SCA2.

Dados de famílias com diagnóstico molecular realizado no Serviço de Genética Médica do Hospital de Clínicas e armazenados em banco de dados de acesso restrito aos investigadores serviram de base para identificar a população em estudo. Entre 2016 e 2017, sujeitos adultos e pertencentes a essas famílias foram convidados a participar desse estudo. Após apresentarem seu consentimento informado, esses sujeitos forneceram dados sobre todos os seus parentes ligados ao ramo familiar onde a SCA2 segregou, permitindo a construção de heredogramas completos. Os dados de ao menos dois informantes por família deveriam ser convergentes para serem estudados. As variáveis em estudo foram obtidas de todos os sujeitos: data de nascimento, sexo, relações de parentesco, ordem de nascimento, status sintomático e número de filhos. A data de falecimento e a idade de início do primeiro sintoma foram anotadas, quando adequado. O

fitness genético (ou o sucesso reprodutivo) dos portadores de SCA2 foi considerado igual a $w = W \text{ afetados} / W \text{ não afetados}$, onde W corresponde à mediana do número de filhos do grupo. Como o genótipo não estava disponível para a maioria dos membros das famílias SCA2, o fenótipo foi usado para classificar os sujeitos como afetados ou não afetados. Para reduzir os efeitos de potenciais vieses de memória e de aferição, dois critérios de inclusão foram utilizados: (1) ser mais velho do que um desvio-padrão acima da média da idade de início dos sintomas do grupo total; e (2) existir a informação completa da história reprodutiva tanto o sujeito como do seu genitor transmissor. A análise de segregação comparou o número de irmãos afetados com o dos não afetados; como critério de inclusão, novamente todos os irmãos estudados deveriam ser mais velhos do que um desvio padrão da média de idade de início da coorte completa.

Vinte e uma das trinta e uma famílias com diagnóstico de SCA2 foram recuperadas: os sujeitos de uma família negaram-se a participar do estudo, enquanto as demais dez famílias foram perdidas do follow-up. Os heredogramas estudados incluíram 1.017 pessoas, 164 das quais sintomáticas – e 84 delas ainda vivas. Como a idade de início média foi de 36.6 (14.9) anos, sujeitos afetados e não afetados com mais de 52 anos de idade foram utilizados para estimar padrões de fitness e segregação. Sessenta e sete sujeitos afetados e 97 sujeitos não afetados foram incluídos na análise do fitness e tiveram a mediana de 3 e 2 filhos respectivamente, o que resultou em um fitness genético igual a 1,5 ($p < 0,025$). A grande maioria dos filhos de afetados nasceu antes do início dos sintomas desses genitores. Finalmente, o fitness variou de acordo com a idade de início dos sujeitos, tornando-se menor do que o fitness dos sujeitos não afetados quando a idade de início dos sintomas era menor ou igual a 20 anos de idade.

Entre os sujeitos incluídos na análise de segregação, havia 137 não-afetados (59,6%) e 93 afetados (40,4%) e essa proporção foi significativamente diferente da esperada em uma segregação randômica de alelos ($p = 0,04$).

A idade de início foi inversamente proporcional ao tamanho do CAGexp em 49 indivíduos onde a informação molecular estava disponível ($\rho = -0.708$, $p < 0.0001$, Spearman); cada repetição CAG adicional associou-se a uma redução de 1,57 anos na idade de início ($p < 0.003$). A diferença média entre a idade de início dos sujeitos e a de seus genitores afetados foi de -10.77 anos, ou seja, a idade no início das gerações apontou para a antecipação como um fenômeno bastante frequente.

Portanto, nossos resultados apontaram que três forças seletivas podem afetar a recorrência da SCA2. A SCA2 teve um fitness genético aumentado nos portadores com idades de início medianas ou tardias. Em contraste, nós observamos uma distorção na segregação ou favorecendo os gametas com alelos normais – em mais de 95% das vezes, portadores de alelos com 22 repetições (CAG)22 – ou desfavorecendo os gametas com alelos expandidos. Nosso resultado, inclusive, sugere que a alta frequência do alelo (CAG)22 nos cromossomos humanos seja explicada por essa vantagem seletiva. Por fim, nossos dados sugerem que a antecipação é a ocorrência mais frequente nas transmissões entre afetados na comparação entre as idades de início nas gerações subsequentes. Essa última observação propõe que a transmissão do CAGexp instável tende a expandi-lo ainda mais; isso precisa ser confirmado pela observação de irmandades completamente genotipadas, em futuros estudos. Se ela se confirmar, a tendência à antecipação deverá se associar a um fitness reduzido nesses afetados. Em conclusão, deduzimos que embora o fitness dos afetados com idades de início médias seja maior do que a dos não afetados, a frequência da SCA2 não está aumentando na população, pois outras forças seletivas agem

em sentido contrário (seleção negativa): a segregação distorcida e que favorece a transmissão do alelo normal e a tendência à antecipação.

Abstract

Spinocerebellar ataxia type 2 (SCA2) is caused by an expansion in a repetitive sequence of CAG codons (CAG_{exp}) in exon 1 of the ATXN2 gene. The expanded polyglutamine tract is associated with a gain in toxic function, expressed predominantly in neurons. SCA2 is an autosomal dominant neurodegenerative condition, with onset between 30 and 40 years of age, manifestations due to progressive neurological disorders that lead to those affected in the wheelchair and reduced survival. Despite being a community of companies, as an anticipation, a frequency of SCA2, which is a place for the communities where it was observed, suggesting the existence of simultaneous favorable selective mechanisms. To ascertain this hypothesis, the present study, fitness estimate, a segregation and role of anticipation in a cohort of families with molecular diagnosis of SCA2.

Data from families with molecular diagnoses performed at the Medical Genetics Service of the Hospital de Clínicas and stored in databases of restricted access to the researchers served as a basis to identify the population under study. Between 2016 and 2017, adult subjects belonging to these families were invited to participate in this study. After presenting their informed consent, these subjects provided data on all their relatives related to the family branch where the SCA2 segregated, allowing the construction of complete heredograms. Data from at least two informants per family should be convergent to be studied. The study variables were obtained from all subjects: date of birth, gender, relationship, birth order, symptomatic status and number of children. The date of death and the age of onset of the first symptom were noted where appropriate. The genetic fitness (or reproductive success) of patients with SCA2 was considered equal to $w = W_{\text{affected}} / W_{\text{unaffected}}$, where W was the median number of children in the group.

As the genotype was not available for most members of the SCA2 families, the phenotype was used to classify the subjects as affected or unaffected. To reduce the effects of potential memory and bias biases, two inclusion criteria were used: (1) to be older than a standard deviation above the mean age of symptom onset of the total group; and (2) there is complete information on the reproductive history of both the subject and the transmitting parent. The segregation analysis compared the number of siblings affected with that of the unaffected; as inclusion criterion, again all siblings studied should be older than a standard deviation from the mean age of complete cohort start age.

Twenty-one of the 31 families diagnosed with SCA2 were recovered: subjects from one family refused to participate in the study, while the remaining 9 families were lost from follow-up. The heredograms studied included 1,017 people, 164 of whom were symptomatic - and 84 of them were still alive. As the mean age at onset was 36.6 (14.9) years, affected and unaffected subjects over 52 years of age were used to estimate fitness and segregation patterns. Sixty-seven affected subjects and 97 unaffected subjects were included in the fitness analysis and had a median of 3 and 2 children respectively, which resulted in a genetic fitness equal to 1.5 ($p < 0.025$). The vast majority of affected children were born before the onset of symptoms of these parents. Finally, the fitness varied according to the age of onset of subjects, becoming smaller than the fitness of the unaffected subjects when the age of onset of symptoms was less than or equal to 20 years of age.

Among the subjects included in the segregation analysis, there were 137 unaffected (59.6%) and 93 affected (40.4%), and this proportion was significantly different from that expected in a random segregation of alleles ($p = 0.04$).

The age of onset was inversely proportional to the size of CAGexp in 49 individuals where molecular information was available ($\rho = -0.708$, $p < 0.0001$,

Spearman); each additional CAG repeat was associated with a reduction of 1.57 years at the onset age ($p < 0.003$). The mean difference between the age of onset of the subjects and that of their affected parents was -10.77 years: that is, the age at the beginning of the generations pointed to the anticipation as a frequent phenomenon.

Therefore, our results indicated that three selective forces may affect the recurrence of SCA2. SCA2 had an increased genetic fitness in patients with median or late onset ages. In contrast, we observed a distortion in segregation or favoring gametes with normal alleles - more than 95% of the time, with alleles with 22 replicates (CAG) 22 - or disfavoring the gametes with expanded alleles. Our result even suggests that high frequency of the allele (CAG) 22 on human chromosomes is explained by this selective advantage. Finally, our data suggest that anticipation is the most frequent occurrence in transmissions between affected, in the comparison between the ages of onset in subsequent generations. This last observation proposes that the transmission of unstable CAG_{exp} tends to expand it further; this needs to be confirmed by the observation of completely genotyped brotherhoods in future studies. If it is confirmed, the tendency to anticipate should be associated with a reduced fitness in this affected. In conclusion, although the fitness of the affected individuals with the mean onset ages is increased, the frequency of ACS2 is not increasing in the population because other selective forces act in the opposite direction (negative selection): distorted segregation and favors the transmission of the allele normal and the tendency to anticipate.

3. Introdução

3.1 Ataxia espinocerebelar tipo 2

A ataxia espinocerebelar tipo 2 (SCA 2) é uma das condições autossômicas dominantes e neurodegenerativas causadas por expansões de sequências repetitivas CAG (CAGexp) traduzidas em tratos de poliglutamina (poliQ), dessa vez na matriz de leitura do gene *ATXN2*, no cromossomo 12q23-24.1 (Imbert et al 1996; Pulst et al 1996). Os alelos normais apresentam 12 a 31 repetições e os alelos mutantes apresentam 34 a 59. A faixa de 32 a 33 repetições é considerada uma zona de *overlap* ou de incerteza. A SCA2 parece ser a segunda ataxia espinocerebelar autossômica dominante (SCA) mais prevalente no mundo, sendo apenas ultrapassada pela Doença de Machado Joseph (DMJ/SCA3) (Schöls et al 2004; Sequeiros et al 2012).

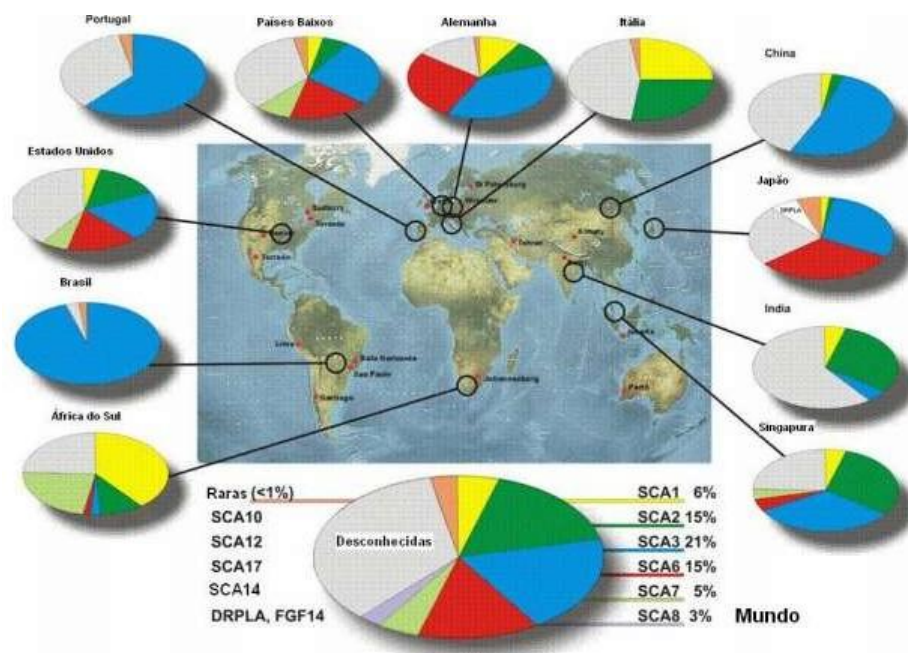


Figura 1: Prevalência das ataxias espinocerebelares em diferentes países e um panorama mundial. (modificada de www.geneclinics.org, 2007a)

Proporções semelhantes foram observadas nas populações brasileiras (de Castilhos et al 2014). Ademais, a Rede Neurogenética (www.redeneurogenetica.ufrgs.br) diagnosticou a SCA2 em pelo menos 83 pessoas (48 famílias).

A idade de início dos primeiros sintomas percebidos pelos doentes é, em média, entre 33 e 38 anos. O sintoma inicial mais comum é a ataxia de marcha (97%), seguido pela disartria (3%). No entanto, muitas outras manifestações neurológicas podem ser encontradas (Velázquez-Pérez et al 2011; Monte et al 2017). Podem aparecer lentidão dos movimentos sacádicos, hiporreflexia, tremor postural e mioclonia. Existem relatos de manifestações clínicas em crianças e jovens. Nesse caso, o número de repetições do CAG era bastante elevado, entre 130 e em alguns casos mais de 200. (Mao et al 2002).

O *ATXN2* produz a proteína ataxina-2, quando mutada (CAGexp), a ataxina-2 apresenta uma poliQ expandida que, por sua vez, está associada a um ganho de função tóxica e à neurodegeneração.

Há uma forte correlação entre o tamanho do CAGexp e a idade de início. Tal como em outras doenças de poliQs, a transmissão paterna está associada às maiores instabilidades meióticas e às mais nítidas antecipações. Isso significa que existe uma tendência, com o passar do tempo, de que os novos portadores da doença, em especial, filhos e filhas de homens portadores, manifestem sintomas mais precocemente do que em seus pais.

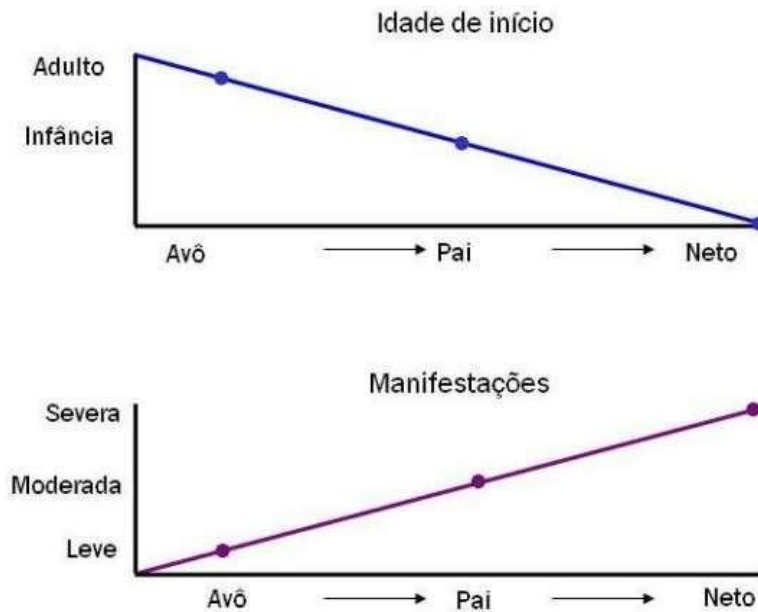


Figura 2: Fenômeno da antecipação, característica de algumas doenças genéticas, onde as sucessivas gerações tendem a apresentar os sintomas precocemente e/ou de maneira mais severa. (modificada de www.geneclinics.org, 2007b)

Até onde sabemos, a maior expansão já descrita foi a de um alelo de 202 CAGs, presente em um bebê que apresentava apnéia, hipotonia e disfagia e transmitido pelo pai, sintomático desde os 22 anos, e que tinha um alelo de 43 repetições (Babovic-Vuksanovic et al 1998). O início na infância tem sido algumas vezes descrito; entre nós, um portador de 67 repetições no CAGexp iniciou seus sintomas aos 3 anos (Pereira et al 2015). Seria intuitivo pensar que o fenômeno da antecipação estaria ligado à redução do pool de CAGexp no *ATXN2* na população, já que nas gerações subsequentes a doença ocorreria em idade progressivamente menor, com eventual suspensão de sua descendência.

Além do sistema nervoso central, outros tecidos contêm a ataxina-2 expressa. No entanto, efeitos da poliQ em outros sistemas ainda são desconhecidos. Para o foco do presente estudo, cabe mencionar que o *ATXN2* é expresso em tecidos reprodutivos como testículo epidídimo, placenta, útero, próstata. (The Human Atlas Protein)

3.2 Seleção natural e adaptabilidade darwiniana

Todos os organismos vivos estão constantemente sob forte influência da seleção natural. Essa é a principal força motriz que durante bilhões de anos vem forjando a complexidade e a beleza de todas as formas de vidas existentes em nosso planeta. Durante milhares de anos se acreditou que as extraordinárias estratégias de sobrevivência dos organismos vivos só poderiam ser explicadas sob um prisma de um criador que projetou cada forma de vida individualmente lhe dando mecanismos elegantes para sobreviver ao ambiente. Porém, a compreensão sobre a origem da vida e das espécies mudou radicalmente a partir da Teoria da Evolução proposta por Charles Darwin e de forma simultânea por Alfred Russel Wallace.

A Teoria da Evolução passou por importantes reformulações sob avanço da genética contribuindo com elementos que estavam inacessíveis a Darwin. Esse processo ficou conhecido como a síntese moderna ou neodarwinismo (Dawkins 1986). Já a compreensão da Biologia Molecular possibilitou um olhar mais acurado dos mecanismos de seleção natural e herdabilidade e evolução. Esses novos elementos, em nenhum momento, reduzem o êxito da formulação original de Darwin. Pelo contrário, reafirmam sua genialidade que, mesmo sem recursos científicos robustos, elaborou a espinha dorsal da teoria que unificou as ciências naturais, até então vistas como uma grande colcha de retalhos. Como afirmou o biólogo Theodosius Dobzhansky (1964): "*Nada na Biologia Faz Sentido Exceto à Luz da Evolução*".

Um dos elementos centrais da teoria da evolução está relacionado com o conceito de **adaptabilidade darwiniana**, que consiste na capacidade de um ser vivo sobreviver e se reproduzir. Nesse sentido, ela (a adaptabilidade) considera tanto a viabilidade de o organismo sobreviver como a sua capacidade de se reproduzir. Em última instância significa que se o indivíduo tiver uma extraordinária capacidade de se adaptar ao meio,

mas for incapaz de se reproduzir, ou ainda, por algum fator não conseguir deixar descendentes ele será considerado, do ponto de vista evolutivo, um organismo não adaptado

Do ponto de vista da genética de populações existe um conceito denominado adaptabilidade absoluta simbolizado pela letra W . De forma simplificada ela expressa a quantidade de descendentes que o organismo consegue deixar, ou seja, se um organismo tiver um descendente ele terá a $W=1$, mas se tiver dois descendentes terá $W=2$ e assim sucessivamente. No caso do indivíduo sem prole ele terá $W=0$, pois mesmo que ele apresente um fenótipo que lhe garanta um certo sucesso de sobrevivência os genes envolvidos nessas características não serão transmitidos. Já o conceito de adaptabilidade relativa w mensura a adaptabilidade de um indivíduo em relação a outro e, em geral, com o organismo com a W maior. Por exemplo, suponhamos que em uma população um organismo com maior fecundidade deixe em média 10 descendentes durante sua vida, nesse caso ele terá $W=10$. Caso tenha outro organismo com uma fecundidade menor, suponhamos $W=8$, esse terá sua adaptabilidade **relativa** w 0,8. Esse número surge da razão organismo com a maior w e o organismo que estamos analisando (Griffiths A. et al 2010)


Isso possibilita estimar qual é a dinâmica de um alelo ou conjunto de alelos. Se sua dinâmica é ser eliminado de uma dada população ou rumo para fixação em um período de tempo. Um organismo que não consegue se reproduzir ou que morre precocemente em relação aos demais de sua espécie (baixa capacidade de sobreviver) terá seus genes excluídos do *pool* genético pois terá um w muito baixo. Já um organismo que apresenta um w elevado irá deixar mais descendentes e os alelos envolvidos nesse processo,

responsáveis pela vantagem adaptativa, aumentarão sua frequência no conjunto da população e apresentarão a dinâmica de serem fixados.

Uma primeira impressão sobre as relações entre sobrevida e descendência poderia supor que quanto mais um organismo vive, maiores seriam suas possibilidades de deixar descendentes durante seu período reprodutivo. Isso parece ser óbvio e na natureza vemos inúmeros exemplos onde a redução na expectativa de vida irá apresentar de forma direta uma redução na fertilidade; afinal, indivíduos mortos não se reproduzem. Porém, a natureza apresenta aspectos aparentemente contraditórios e viver mais não é sinônimo de se reproduzir mais. A associação entre os dois aspectos é comum, mas não se trata de uma lei. Ao contrário, há muitos exemplos onde o sucesso reprodutivo está diretamente relacionado com uma redução da sobrevida.

Um dos exemplos mais marcantes entre sucesso reprodutivo e redução na sobrevida é encontrado entre algumas espécies de Louva Deus. Essas apresentam dimorfismo sexual, representado principalmente pelo menor tamanho do macho, que após a cópula apresenta grande probabilidade de ser devorado pela fêmea (Fea et al 2013). Assim, os indivíduos que por algumas características não conseguem se reproduzir apresentarão uma tendência a sobreviver por mais tempo do que aqueles que tiveram sucesso reprodutivo, pois estes correram o risco eminente de serem devorados.

Por isso, ao se analisar determinados fenótipos, - sob a ótica do senso comum- que apresentam forte impacto negativo na expectativa de vida, chegaríamos a conclusão que esses alelos seriam eliminados com o passar do tempo, por reduzir um dos elementos da adaptabilidade darwiniana: a capacidade de sobreviver. Porém, como vimos no exemplo do louva deus, maior tempo de sobrevivência não necessariamente significa maior aptidão reprodutiva.

Portanto, alguns fenótipos que dificultam a sobrevivência poderiam permanecer em uma população através de mecanismos compensatórios. A redução na frequência de um alelo, devida às desvantagens fenotípicas em uma determinada fase do organismo, poderia ser compensada, no pool populacional por diversos mecanismos como o aparecimento de mutações *de novo*, por efeito do acaso (deriva) ou por migrações (efeitos fundadores). No entanto, um dos mecanismos mais interessantes a se averiguar consiste no pleiotropia antagonista . 

3.3 Pleiotropia antagonista.

Um alelo pode conferir desvantagem em uma fase da vida ou a um certo grupo celular de um organismo e, ao mesmo tempo, vantagens para a adaptabilidade em outra fase da vida ou a outro grupo celular do organismo. Esse fenômeno é conhecido como pleiotropia antagonista. Uma determinada variante pode atuar como causadora de uma doença, mas também ser um efeito protetor contra outra; ou aumentar a fertilidade de um fenótipo patológico; ou ser vantajosa em algum momento da vida e desvantajosa em outro momento. Esse conceito foi desenvolvido pelo biólogo evolutivo, George C. Williams (Willian 1957) e foi identificado em inúmeros fenômenos que ocorrem na natureza.

Em seres humanos, uma série de características foram propostas como provavelmente associadas a pleiotropia antagonista. Por exemplo, já foi proposto que portadores (heterozigotos) da doença Tay Sachs poderiam apresentar maior resistência a tuberculose (Diamond J. 1988). O alelo APOE4 está associado a doenças coronarianas e principalmente à doença de Alzheimer, mas também confere uma resposta imunológica mais potente e um aumento na fertilidade aos seus portadores (van Exel 2017). Portadores sintomáticos da ataxia espinocerebelar tipo 3, também conhecida como doença de Machado-Joseph (SCA3/MJD), uma outra condição associada a presença de uma poliQ expandida, têm fitness aumentado, de acordo com evidências obtidas de um estudo realizado no âmbito deste PPG (Prestes et al 2008).

3.4 Fitness

O fitness de um fenótipo é uma medida simples de sucesso reprodutivo, onde aqueles que sobrevivem e produzem maior número de prole capaz de se reproduzir garantem a manutenção do *pool* gênico da sua população (Stine, 1989). Em 1997, Vogel e Motulsky definiram o fitness de um genótipo como a performance reprodutiva de um certo genótipo, ou seja, a eficiência em produzir descendência

A maior parte das doenças dominantes reduz as taxas médias de reprodução dos seus portadores. Isso acontece porque a mutação modifica o fenótipo do portador, reduzindo ou impedindo a sua reprodução – seja por impedi-la fisicamente, por reduzir a sobrevivência do portador ou por torná-lo sexualmente menos interessante para os seus parceiros. Em todo o caso, o portador de uma doença dominante que o afete de alguma dessas maneiras terá menos filhos do que o esperado. Haverá, portanto, uma perda da carga genética populacional (desse particular genótipo) em cada geração sucessiva, em uma proporção que dependerá da desvantagem seletiva dos envolvidos. Haverá uma desvantagem seletiva dos mesmos.

3.5 A possibilidade do aumento do fitness nas doenças por poliQ

Em 1996, Frontali e colaboradores usaram uma abordagem epidemiológica para analisar famílias europeias com doença de Huntington (HD) ou com ataxia espinocerebelar tipo 1 (SCA1). Ambas são condições autossômicas dominantes devido às poliQ expandidas, de início na vida adulta. Enquanto a HD se manifesta pela associação de hipercinesias com um quadro psiquiátrico demencial, a SCA1 é uma condição muito semelhante à SCA2 também na clínica, além de compartilhar características semelhantes no mecanismo de herança e na fisiopatologia. Esses autores encontraram um aumento da fecundidade entre portadores de ambas as condições. O fitness aumentado foi detectado entre os portadores de HD e de SCA1, especialmente se nascidos em gerações afastadas (no início do século XX) ou nos portadores de CAGexp relativamente intermediárias para cada condição, na faixa das 37-49 e das 47-54 CAGexp, respectivamente. A reprodução dos portadores de grandes expansões pareceu, ao contrário, estar bastante reduzida. Os autores propuseram um modelo de fitness diferencial, de acordo com a gravidade da CAGexp nas duas condições. O modelo prediria que a frequência de alelos CAGexp com expansões não muito grandes tenderia a se elevar nas gerações subsequentes por conta de um fitness aumentado dos seus portadores e que a frequência das grandes expansões só se manteria por conta de instabilidades e não de transmissões diretas – pois seus portadores teriam um fitness reduzido.

Em 2008, usando novamente uma abordagem epidemiológica, Prestes e colaboradores levantaram evidências bastante fortes de que a SCA3/MJD, outra doença neurodegenerativa por poliQ, teria um fitness aumentado. O estudo foi feito na população SCA3/MJD do Rio Grande do Sul, uma população notável em sua dimensão e devida a um efeito fundador açoriano (Souza et al 2015). Prestes et al descreveram que o número

médio de filhos das mulheres da população do RS e das mulheres com SCA3/MJD, em 2000, foi 1,90 e $2,93 \pm 2,3$ ($p < 0,001$). Se o *fitness* da população for considerado 1, o da SCA3/MJD foi de 1,54.

A fecundidade ou o *fitness* genético da SCA2 ainda não parecem ter sido estudados. Uma busca no Pubmed associando os termos “SCA2” e fertilidade localizou um estudo que descreveu os efeitos do knock-out da ataxina-2 em um camundongo transgênico, delineado para investigar as funções da proteína (Lastres-Becker et al 2008). Os animais homocigotos apresentavam fertilidade reduzida, aparentemente mediada por uma redução da expressão dos receptores da insulina.

Assim, existem evidências em favor de um aumento do *fitness* em três condições poliQ, ao menos em portadores dentro de uma certa faixa de CAGexp: a HD, a SCA1 e a SCA3/MJD. O mecanismo que media a vantagem na fecundidade associada à presença de poliQs ainda é especulativo. Em todo o caso, o aumento da fecundidade nessas condições seria mais um bom exemplo de pleiotropia antagonista.

Em todas essas condições, os sintomas dos portadores são progressivos, iniciando-se entre os 10 e os 70 anos – em média, entre os 30-40 anos. As expectativas de vida reduzem-se significativamente (Kieling et al 2007). Na medida em que as gerações se sucedem, o início dos sintomas vai se tornando mais precoce na vida. No entanto, somente em indivíduos onde a doença começa na infância ou na adolescência, o *fitness* ficaria prejudicado, levando a uma forte seleção negativa. Com a tendência do início cada vez mais precoce da doença, fruto da expansão da repetição CAG, e por se tratar de uma doença dominante, ou seja, os portadores da mutação irão manifestar a doença, seria correto perguntar: porque esses genes expandidos não são deletados na população por seleção natural?

Existem elementos concretos que fazem com que esses indivíduos apresentem uma redução na taxa de sobrevivência, ou seja, a seleção natural “atua” para excluir esses alelos e seria uma questão de tempo para que a seleção natural eliminasse o CAG expandido. Porém, o que é possível se observar, de forma empírica, é que a população de portadores de ambas doenças não apresenta tendência de queda; quiçá apresente o contrário (de Araújo 2016).

Ao que tudo indica, uma possível explicação para a manutenção desses alelos na população é a existência de fatores que produziram alguma seleção positiva. É possível que portadores de poliQs não tão extensas, com idade de início dentro da média, tenham algum tipo de vantagem para manter no *pool* genético os genes mutantes.

4. Objetivos

Principal

- Verificar se o “fitness” genético está aumentado em famílias brasileiras com SCA2
- Verificar se existe distorção na segregação do alelo expandido em relação ao não expandido
- Analisar o impacto da antecipação na redução do Fitness.

Secundários

- Comparar o número médio de filhos dos afetados com os dos não afetados, em sujeitos mais velhos do que um desvio padrão da idade de início, dentro das famílias. Relacionar o sexo dos portadores, a idade de início da doença e a CAGexp com o número de filhos.
- Verificar se o gênero do genitor afetado ou do próprio indivíduo tem impacto na distorção da segregação.
- Estabelecer a influência da antecipação com o passar das gerações e correlacionar com o número de filhos.

5. Artigo

Artigo científico a ser submetido à *Clinical Genetics*.

Selective forces related to spinocerebellar ataxia type 2.

Running title: Selective forces related to SCA2

Lucas Sena¹, Raphael Machado Castilhos², Eduardo Preusser Mattos^{1,3}, Gabriel Vasata Furtado^{1,3}, José Luiz Pedroso⁴, Orlando Barsottini⁴, Maria Marla Paiva de Amorim⁵, Clecio Godeiro⁵, Maria Luiza Saraiva Pereira^{1,3,6,7}, Laura Bannach Jardim^{1,3,6,8}

¹ Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul; ² Serviço de Neurologia, Hospital de Clínicas de Porto Alegre; ³ Laboratório de Identificação Genética, Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre; ⁴ Setor de Neurologia Geral e Ataxias. Disciplina de Neurologia Clínica da UNIFESP – Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, Brazil; ⁵ Universidade Federal do Rio Grande do Norte, Natal, Brazil; ⁶ Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre; ⁷ Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul; ⁸ Departamento de Medicina Interna, Universidade Federal do Rio Grande do Sul.

Corresponding author:

Prof. Laura Bannach Jardim

Medical Genetics Service

Hospital de Clinicas de Porto Alegre

Rua Ramiro Barcelos 2350

90035-003 Porto Alegre, Brazil

phone + 55 51 33598011

email: ljardim@hcpa.edu.br

Acknowledgements

The authors would like to thank the individuals who agreed to participate in this study. This work was supported by FIPE-HCPA – Fundo de Incentivo à Pesquisa do Hospital de Clínicas de Porto Alegre – Project GPPG HCPA 16-320. EPM, GVF, MLSP, and LBJ were supported by CNPq.

Abstract

Spinocerebellar ataxia type 2 (SCA2) is caused by unstable expanded CAG repeat tract (CAG_{exp}) at *ATXN2*. Although prone to selective forces such as anticipation, SCA2 frequency seems to be stable in populations. Our aim was to estimate fitness, segregation patterns, and the role of anticipation in SCA2. Adult subjects from families with molecular diagnosis provided data about all his/her relatives. Affected and unaffected sibs older than 52 years-old were used to estimate fitness and segregation patterns. Twenty

SCA2 families were studied, including 1,017 individuals (164 affected), born from 1840 to 2012. Among the 164 subjects included in the fitness analysis, the median number of children of the non-carriers and carriers were 2 and 3 ($p < 0.025$), respectively. Therefore, fitness of carriers was 1.5. Among subjects included in the segregation analysis, there were 137 non-carriers (59.6%) and 93 carriers (40.4%) ($p = 0.04$). Age at onset across generations pointed to anticipation as a frequent phenomenon. We raised evidence in favor of increased fitness related to the carrier state at *ATXN2*, and segregation distortion favoring the normal alleles. Since most of the normal alleles analysed carried 22 repeats,

we propose that this can be related to the high frequency this allele in human chromosomes.

Key words

anticipation; CAG repeats; genetic fitness; polyglutamine tract; segregation distortion; SCA2; spinocerebellar ataxia type 2; unstable transmission.

1. Introduction

Spinocerebellar ataxia type 2 (SCA2) is an autosomal dominant cerebellar ataxia caused by an unstable expansion of a CAG repeat tract (CAG_{exp}) at the ataxin-2 gene, *ATXN2* (1). Normal *ATXN2* CAG tracts range from 12 to 31 repeats. However frequency of alleles carrying 22 CAG repeats range from 75-90% in several populations (2,3,4). Ataxic phenotype occurs when the polymorphic tract is longer than 33 CAG repeats, and cases with 32-33 repeats fall in the gray zone for penetrance. The CAG expansion introduces an abnormally large polyglutamine (polyQ) in the protein, giving rise to toxic effects that lead to neuronal function and survival (5).

Very disabling, SCA2 is associated to a progressive cerebellar syndrome starting at 33-38 years of age, on average. As observed in other polyQ diseases, a strong negative correlation was observed between age at onset (AO) and length of CAG_{exp} (6). The expansion is unstable upon cellular divisions, and offsprings might carry shorter or longer expansions than CAG_{exp} of the transmitting parent. Therefore, the disease might start at a later or earlier age than in the parent. Although rare, very late onset (1) and even lack of family history (7) have been reported in subjects carrying short expanded alleles: contractions can be behind those cases. Anticipation has been much more frequently observed, with several reports of carriers starting the disease during childhood (8, 9,10).

Although a recent analysis stated that male and female transmissions were related to average increases of 4.8 ± 5 and 1.6 ± 2.0 CAG repeats in the offspring (9), sibships were not completely genotyped in the same report, and lack of information about asymptomatic children might have biased these data.

The overall impact of the intergenerational instability of the CAGexp transmission remains unknown. If further CAGexp expansions upon transmissions are really so frequent, a reduction in SCA2 frequency in populations would be expected, since severe anticipations would be associated to increased neurological deficits as well as to reduced fertility. However, there is no evidence pointing to a decline in SCA2 frequency. In contrast to this negative selection force, positive selective forces might be operating in SCA2 transmission simultaneously, as reported previously in other polyQ disorders, such as Huntington disease, spinocerebellar ataxia type 1, and spinocerebellar ataxia type 3/Machado-Joseph disease (12,13,14,15). Positive selective forces would counterbalance negative forces, maybe keeping expanded alleles in the genetic pool. In order to contribute with the understanding of these forces, our aim was to evaluate the genetic fitness – or the reproductive success – , the segregation patterns, and the anticipation phenomena in SCA2.

2.Methods

Adult subjects belonging to Brazilian families with molecular diagnosis of SCA2 performed in our Institution were invited to participate in the present survey, which was carried out from July, 2016 to July, 2017. Informants provided data about all his/her relatives, allowing the build out of complete family trees . Variables under study were investigated from all relatives belonging to the included families: date-of-birth, date-of-

death, gender, kinship relationships, order of birth, symptomatic status, and number of children. Age at onset of first symptom was obtained from the symptomatic subjects.

The genetic fitness (or the reproductive success) of SCA2 carriers was considered to be $\mathbf{W} = \frac{\mathbf{w\ affected}}{\mathbf{w\ unaffected}}$

Since the genotype was not available in the majority of family members under study, the phenotype was used to classify the relatives as affected or unaffected subjects.

In order to reduce the impact of potential ascertainment and memory bias, two inclusion criteria were considered for fitness analysis: (1) only subjects older than at least one standard deviation from the mean AO obtained in the entire cohort were considered; (2) the reproductive history of the individual included and of his/her affected parent should be entirely known.

Segregation analysis compared the number of affected with that of unaffected sibs, children of one affected parent. Again, all sibs should be older than at least one standard deviation from the mean AO of the entire cohort.

Group characteristics were compared with Mann-Whitney U or chi-square test, and a $p < 0.05$ was chosen to support assumptions made.

The study protocol was approved by the ethical committee of our institution (registered as CAAE: 56652416.2.0000.5327 at the Brazilian National platform, Plataforma Brasil). All patients gave informed consent to participate in the study.

3. Results

Thirty one Brazilian families received a molecular diagnosis of SCA2 from our Institution since 2001, and 21 agreed to participate in this study. One family declined to participate, and 9 were lost. Data from 1,017 relatives were retrieved: 164 were symptomatic subjects, 84 of them still alive. Molecular analyses were available for 50 individuals, reported elsewhere (16). Normal and expanded alleles showed an average \pm SD (range) of 23 \pm 3 (22-33) and 40 \pm 4 (34-53) CAG repeats. The (CAG)₂₂ allele corresponded to 78% of normal alleles. Since the mean (SD) age at onset of all 164 symptomatic individuals was 36.6(14.9) years of age, fitness and segregation analyses included only subjects older than 52 years of age at the time of the survey - or born before 1964. Symptomatic subjects were considered affected, while the remaining ones were considered unaffected.

Sixty-seven affected and 97 unaffected subjects were included in the fitness analysis. Their median number of children were 3 and 2, respectively ($p < 0.025$, Mann-Whitney U test). SCA2 fitness was therefore estimated as $w = 3/2 = 1.5$ (Figure 1A). The number of children of affected individuals showed an inverse correlation proportionally to age of onset of symptoms ($\rho = 0.49$, $p < 0.001$, Spearman) (Figure 1B). Fitness changed according to AO of the affected subject, turning lower than healthy sibs in subjects with AO earlier than 20 years.

Figure 1A

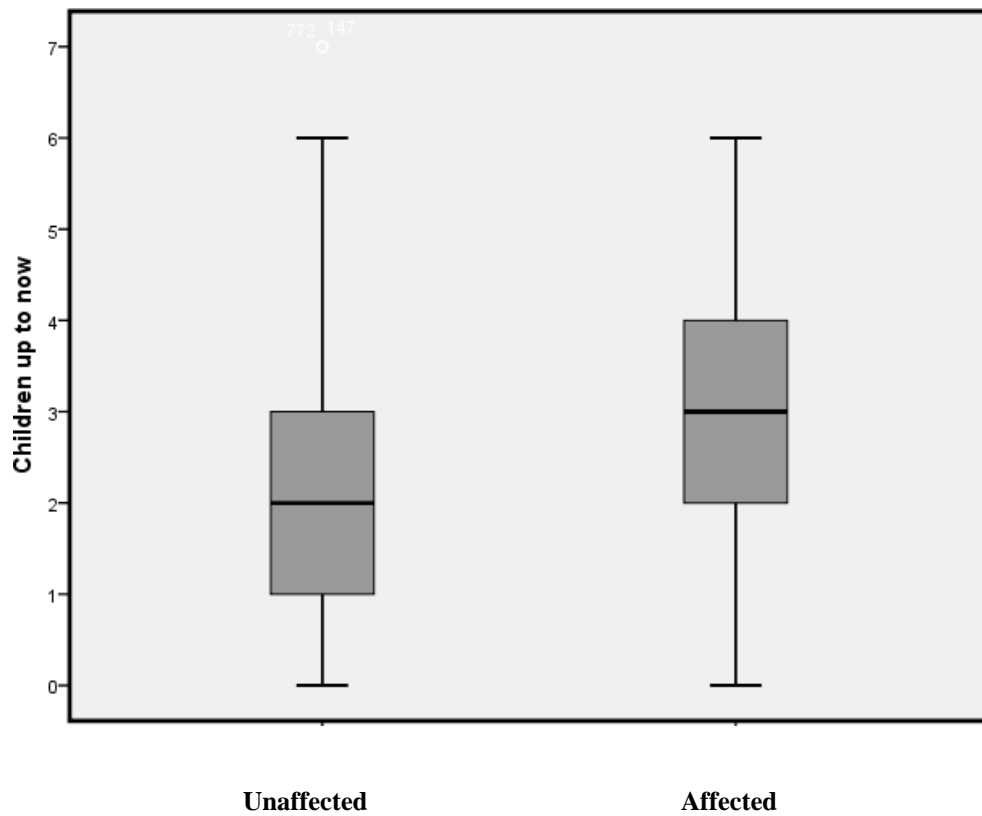


Figure 1A - Genetic fitness among SCA2 carriers. (A) Number of children of affected versus unaffected subjects from SCA2 families were 3 and 2, black line in the graph ($p = 0.025$, Mann-Whitney U test)

Figure 1B

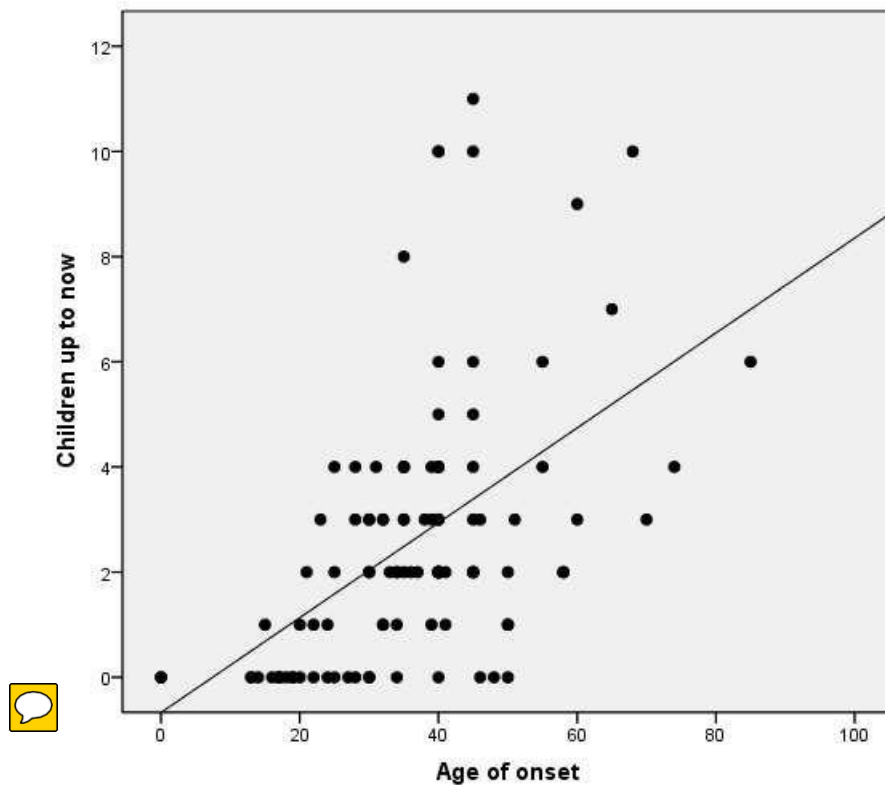



Figure-1B Correlation between age at onset of symptoms and number of children in all symptomatic carriers included in the present study ($\rho=0.49 < 0.001$, Spearman).

The mean age of the affected parents at birth of their children was 31.3 (9.1) years for men and 28.8 (8.8) years for women ($p=0.66$, Mann-Whitney U test), while AO were 37.07(15.7) SD and 35.93 (13.55) SD for affected fathers and mothers ($p=0.79$, Mann-Whitney U test), respectively. Great majority of affected children (92%) were born before onset of symptoms (Supplemental Figure 1).

Two hundred and thirty sibs, children of a symptomatic subject, were included in the analysis of segregation: 93 (40.4%) were affected while the remaining 137 (59.6%) were unaffected subjects ($\chi^2= 4.2476$, $p=0.04$). The effect of gender was explored and significant distortion was achieved among daughters (female offspring); however,

the number of affected sibs was always higher than 50%, and sample size might have limited the power of the other gender comparisons (Table 1).

Table 1

		Total	Affected subjects	Unaffected subjects	Chi-square 
All informative sibs		230	93 (40.4%)	137 (59.6%)	4.2476, p=0.004
Stratified according to the gender of the parent	Children of affected fathers	117	47(40.2%)	70(59.8%)	2.5096 p=0.011
	Children of affected mothers	113	45(39.8%)	68(60.2%)	5.038 p=0.025
Stratified according to the gender of both parent and child	Sons of affected parents	125	58(46.4%)	67(53.6)	5.6973 p=0.032
	Daughters of affected parents	105	37 (35.3%)	68 (64.7%)	9.152 p=0.002

The AO was inversely related to CAGexp obtained from 49 subjects ($\rho = -0.708$, $p < 0.0001$, Spearman); each additional CAG repeat was associated to a reduction of 1.57

years on AO ($p < 0.003$). AO was also inversely related to date of birth obtained from all 67 symptomatic subjects older than 52 years of age ($\rho = -0.382$, $p < 0.002$, Spearman) (Figure 2). The average difference between AO of children and parent (ΔAO) was -10.77.79.

Figure 2

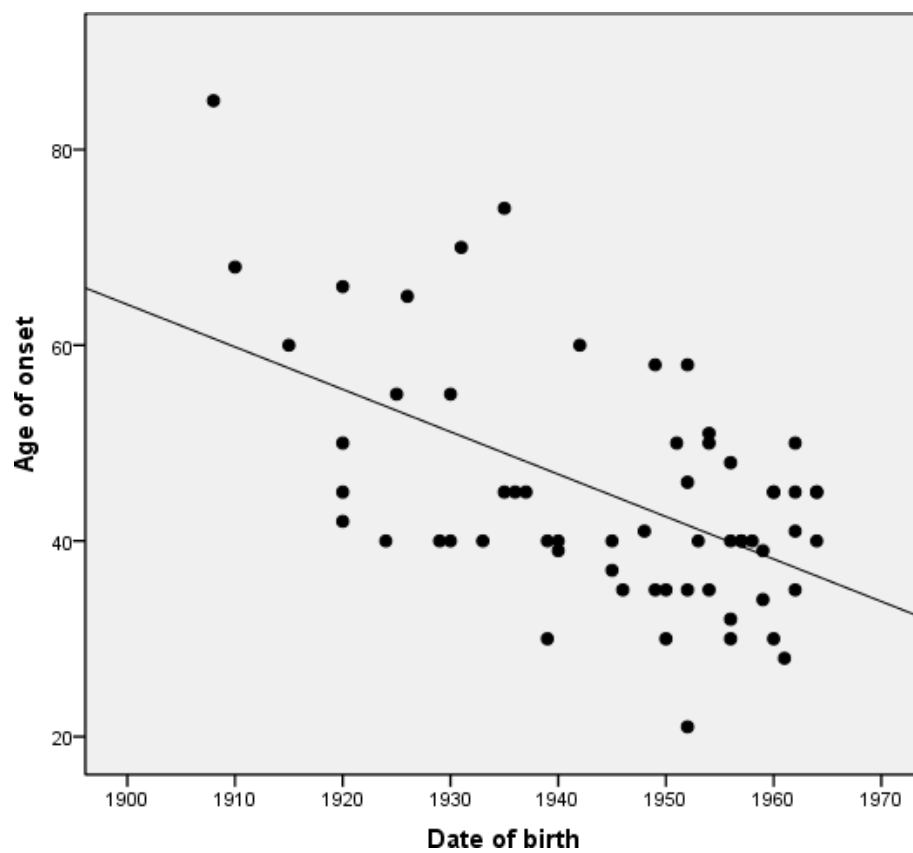


Figure 2 - Correlation between the year of birth and age of onset of symptoms ($\rho = -0.382$, $p < 0.002$, Spearman).

4. Discussion

Three selective forces potentially affecting SCA2 recurrence on next generations were studied through phenotype analysis. SCA2 was associated to an increase in reproductive success (fitness) for those carriers with average age of onset. In contrast, we observed a segregation distortion unfavoring gametes carrying the expanded allele or favoring gametes carrying the normal alleles. Finally, our data suggest that symptoms tended to start earlier than in former generations. If this last result is confirmed, the tendency for further anticipation would be associated to a reduced fitness in next generations. Although the increased fitness observed in the overall SCA2 population would contribute for an increase in its frequency, the other two forces - segregation distortion and anticipation - would work in the opposite direction, and it would tend to eliminate expanded alleles from the population.

Other polyQ diseases, such as HD, SCA1 and SCA3/MJD, have been associated to an increased fitness when compared carriers to non-carriers, featuring fitness of 1.25, 1.47, and 1.53, respectively (12,13). The effect was limited to carriers with average AO. Since polyQ diseases tend to be related to anticipation phenomena, the increasing number of children would be limited in subsequent generations, when the transmitted CAGexp would be too large, and the clinical picture associated would be more severe. It remains unclear the reason that polyQ proteins like huntingtin, ataxin-1, ataxin-3 and now ataxin-2 are related to increased fertility. Ataxin-2, for instance, is highly expressed in testes, ovaries, and other reproductive structures and tissues (16). Whatever the mechanism, the large fitness of polyQ diseases is a factor that contributes to the maintenance of these phenotypes/genotypes in the population.

The segregation distortion unfavoring the transmission of ATXN2 with a CAGexp would produce a negative selection, operating against SCA2 recurrence. This result was

in contrast to positive results obtained for SCA3/MJD (14) and, indeed, might have suffered from ascertainment mistakes in the present survey, based on the phenotype after 52 years of age. Leaving the possibility of a type I error aside, some considerations can be done. First, the distribution of normal CAG alleles at ATXN2 and ATXN3 are quite different in normal populations. There is large evidence showing that the ATXN2 allele with 22 repeats was/is strongly selected in humans (18,19,20). The allele with 22 CAG repeats is so frequent that would prevent a study on ATXN2 segregation in normal populations. Based on that, we speculate that segregation distortion, if present, would be more probably related to an advantage of the (CAG)₂₂ allele rather than to a disadvantage of the expanded alleles. Our data suggest that the advantage associated to the normal allele, more probably the (CAG)₂₂, is related to reproductive factors, mostly when this allele is transmitted to female offspring.

Anticipation is a phenomenon related to unstable transmissions of trinucleotide repeats, usually towards further expansions (21). In order to completely establish the frequency of unstable transmissions, evidence should be raised from studies that genotype all children of affected parents. This data was not available up to date. If CAG expansions are really so frequent as most reports suggest, recurrent anticipations would tend to reduce fertility towards zero.

We are aware that our study has some limitations. Our analyses were based on phenotypes due to the lack of total genotype information. This first issue was overcome by limiting the inclusion criteria to subjects older than 52 years of age. We are also aware that some sibs classified as unaffected could be in fact non-penetrant or very late onset SCA2 carriers. The chance of a type II error to occur in fitness analysis - i.e., to falsely infer the absence of an increased SCA2 fitness due to inclusion of carriers among the group of unaffected subjects - was reduced, although not completely avoided. Even at the

risk, we were able to show an increased SCA2 fitness; i.e., the results surpassed the risk of error. In contrast, results on segregation analysis could still be prone to a type I error, and therefore waits to be clarified by subsequent observations.

In conclusion, substantial evidence towards selective forces operating in SCA2 recurrence was raised by this study. Carriers of CAG_{exp} at ATXN2 had increased fitness, until early onset pictures, probably related to large expansions, supervened. Moreover, we suggested a potential mechanism for the selection of the (CAG)₂₂ allele in humans: segregation distortion and, therefore, reproductive advantage of the gamete carrying this normal allele. Finally, studies on the transmission of CAG_{exp} are urgently needed in order to clarify the role of unstable repeats over the preservation of SCA2 in the genetic pool of humans.

5. References

1. Pulst SM. Spinocerebellar Ataxia Type 2. GeneReviews® [Internet]. 1998. <https://www.ncbi.nlm.nih.gov/books/NBK1275/>
2. Laffita-Mesa, J.M., Bauer, P.O., Kourí, V. et al. Epigenetics DNA methylation in the core ataxin-2 gene promoter: novel physiological and pathological implications. *Hum Genet* (2012) 131: 625.
3. Montcel S.T, Durr A, Bauer P, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain*. 2014. Pages 2444–2455.
4. Geschwind DH, Perlman S, Figueroa CP, et al. The prevalence and wide clinical spectrum of the spinocerebellar ataxia type 2 trinucleotide repeat in patients with autosomal dominant cerebellar ataxia. *Am J Hum Genet*. 1997 Apr;60(4):842-50.
5. Matilla-Dueñas A, Sánchez I, Corral-Juan M, et al. Cellular and molecular pathways triggering neurodegeneration in the spinocerebellar ataxias. *Cerebellum*. 2010 Jun;9(2):148-66.
6. Auburger GW. Spinocerebellar ataxia type 2. *Handb Clin Neurol*. 2012;103:423–36.
7. Costanzi-Porrini S, Tessarolo D, Abbruzzese C, Liguori M, et al. An interrupted 34-CAG repeat SCA-2 allele in patients with sporadic spinocerebellar ataxia. *Neurology*. 2000 Jan 25;54(2):491-3

8. Babovic-Vuksanovic D, Snow K, Patterson MC, et al. Spinocerebellar ataxia type 2 (SCA 2) in an infant with extreme CAG repeat expansion. *Am J Med Genet.* 1998;79:383–7
9. Abdel-Aleem, A. & Zaki, M.S. Spinocerebellar ataxia type 2 (SCA2) in an Egyptian family presenting with polyphagia and marked CAG expansion in infancy. *J Neurol* (2008) 255: 413.
10. Di Fabio, R., Santorelli, F., Bertini, E. et al. Infantile childhood onset of spinocerebellar ataxia type 2. *Cerebellum* (2012) 11: 526.
11. Figueroa KP, Coon H, Santos N, et al. Genetic analysis of age at onset variation in spinocerebellar ataxia type 2. *Neurology: Genetics.* 2017;3(3):e155.
12. Frontali M, Sabbadini G, Novelletto A, et al. Genetic fitness in Huntington's Disease and Spinocerebellar Ataxia 1: a population genetics model for CAG repeat expansions. 1996
13. Prestes PR, Saraiva-Pereira ML, Silveira I, Sequeiros J, Jardim LB. Machado Joseph Disease Enhances Genetic Fitness: A Comparison Between Affected and Unaffected Women and Between MJD and the General Population. *Annals of Human Genetics* 2008; 72: 57–64.
14. da Silva Carvalho, G., Saute, J.A.M., Haas, C.B. et al. Cytokines in Machado Joseph disease/spinocerebellar ataxia 3. *Cerebellum* (2016) 15: 518.
15. Platonov FA, Tyryshkin K, Tikhonov DG, et al. Genetic fitness and selection intensity in a population affected with high-incidence spinocerebellar ataxia type 1. *Neurogenetics.* 2016;17(3):179-185.

16. Pereira FS, Monte TL, Locks-Coelho LD. Genes and Mitochondrial Polymorphism A10398G Did Not Modify Age at Onset in Spinocerebellar Ataxia Type 2 Patients from South America. *Cerebellum*. 2015 Dec;14(6):728-30.

17. Linn Fagerberg, Björn M. Hallström and Mathias Uhlén. Mol Cell Proteomics. 2014 Feb;13(2):397-406.

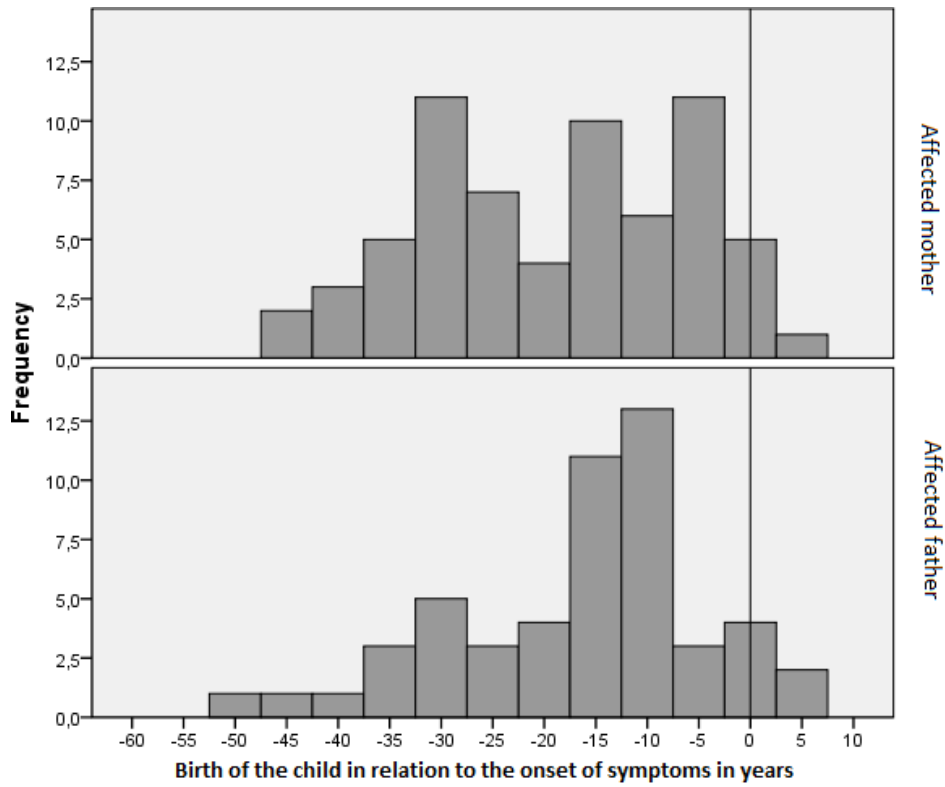
18. Diego Lorenzetti, Saeed Bohlega, Huda Y. Zoghbi. The expansion of the CAG repeat in ataxin-2 is a frequent cause of autosomal dominant spinocerebellar ataxia. *Neurology* Oct 1997, 49 (4) 1009-1013;

19. Mao R, Aylsworth AS, Potter N, Wilson WG, et al. Childhood-onset ataxia: testing for large CAG-repeats in SCA2 and SCA7. 2002 . *Am J Med Genet* 110:338–45.

20. Stefan-M. Pulst, Alex Nechiporuk, Tamilla Nechiporuk, et al. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nature Genetics*. 1996. pages 269–276. doi:10.1038/ng1196-269

21. McMurray CT. Mechanisms of trinucleotide repeat instability during human development. *Nature reviews Genetics*. 2010;11(11):786-799. doi:10.1038/nrg2828.

Supplemental Figure 1



Supplemental Figure 1- Histogram of the births of affected children in relation to the onset of parental symptoms. Where 0 represents the onset of symptoms, negative numbers years births before symptoms and positive numbers births after onset symptoms.

6. Discussão

O presente estudo levantou evidências originais sobre o papel de forças seletivas associadas à fecundidade e à distorção da segregação na SCA2. Através de uma abordagem epidemiológica e tradicional de análise de heredogramas, obtivemos informações novas sobre a dinâmica de manutenção do alelo expandido associado à SCA2. Os resultados concretos serão apresentados no artigo científico acima, a ser publicado em periódico de circulação internacional.

No entanto, eles suscitam muitas especulações adicionais. A seguir, apresentarei tanto algumas das minhas reflexões sobre o significado dos presentes achados, como perspectivas de continuidade da linha de investigação sobre as forças evolutivas que interagem com as doenças neurodegenerativas devidas às poliQs.

6.1 Dinâmica do Alelo Expandido associado à SCA2

Quatro forças seletivas parecem influenciar a dinâmica do alelo mutante, segundo a nossa observação:

1- **Os sintomas da doença.** São significativamente debilitantes e tendem a se intensificar com a progressão da doença reduzindo, assim, a capacidade de reprodução dos indivíduos até torná-la nula. No entanto, é bom lembrar que os indivíduos afetados ainda apresentam a capacidade de reprodução após os sintomas, pois 10% dos filhos de afetados nasceram após os pais iniciarem os sintomas (Supplementary Figure do artigo a ser submetido, e que está na p. 42).

2- **O aumento do fitness.** Indivíduos com o alelo expandido do gene ATXN2 apresentam uma mediana de 3 filhos e os não afetados tem uma mediana de 2 filhos ($p < 0.025$, Mann-Whitney U test). O fitness dos indivíduos afetados é de 1,5: ou seja, os afetados têm 50% mais filhos do que seus irmãos não afetados.

2- **A distorção na segregação:** Como observado, o alelo expandido é preferencialmente não transmitido ou apresenta uma distorção na segregação dos gametas. O gameta com alelo expandido tem uma desvantagem na fecundação em relação ao alelo normal. Isso é expresso pelo fato de 40.4% dos filhos de afetados apresentarem o alelo mutante, enquanto 60.6% dos filhos não apresentam a expansão(chi-square= 4.2476, p=0.04).

3- **Antecipação:** A transmissão do alelo expandido é instável, ou seja, existe uma tendência de o tamanho das repetições do alelo expandido transmitido ser diferente daquela encontrada em tecidos somáticos do progenitor afetado. Figueroa e colaboradores (2017) observaram uma tendência de aumento no tamanho do CAGexp transmitido quando o genitor era homem da ordem de 4.8 ± 5 CAGs a cada geração. Quando o alelo expandido foi herdado de uma mulher, um aumento de $1,6 \pm 2.0$ CAG foi observado a cada geração.

Existe uma correlação inversa entre o tamanho das expansões e a idade de início dos sintomas ($\rho = -0,619$, Spearman $p < 0.000$). Portanto, na medida em que ocorre um aumento no tamanho das expansões com o passar das gerações, existe uma tendência da doença se manifestar mais precocemente. Nosso trabalho demonstrou que o aumento de uma unidade de expansão CAG representou uma redução de 1,57 anos na idade de início da doença ($p < 0.003$).

Diante de tantas forças seletivas reais ou potenciais, seria necessário tentarmos responder à pergunta: Qual a dinâmica do alelo expandido ATXN2?

Existem quatro possibilidades para essa resposta.: O alelo pode estar sofrendo **seleção positiva**, aumentando sua frequência na população com a tendência de se fixar. **Seleção negativa**, apresentando um decréscimo de sua frequência até ser eliminado. Ter seleção positiva e uma determinada geração apresentar seleção negativa, ou seja, ter uma dinâmica de curva normal aumentando sua frequência até um ponto e posteriormente esse alelo apresentar um decréscimo em **sua frequência** ou estar estável.

Para respondermos essa questão poderíamos usar a clássica equação para fitness diferencial e distorção na segregação:

$$p' = \frac{p^2 \cdot w_{11} + 2k \cdot pq \cdot w_{12}}{\bar{w}}$$

Onde:

p' = Frequência do alelo p na geração subsequente

p^2 = Frequência do alelo p em homozigose

w_{11} = **Fitness** do alelo p em homozigose

k = Onde k representa o coeficiente de distorção

$2pq$ =Frequência genotípica do heterozigoto

w_{12} =fitness do heterozigoto

\bar{w} = fitness médio.

No entanto, o caso concreto que estamos querendo analisar apresenta uma facilidade e uma dificuldade que exigem uma adaptação da equação acima. Em primeiro lugar, estamos tratando de um alelo com uma frequência muito baixa, no qual a presença em homozigose não foi observada em nossas amostras. Além disso, por se tratar de uma doença dominante de penetrância praticamente completa, podemos analisar a dinâmica alélica ignorando os casos de homozigose e aplicar as forças seletivas encontradas diretamente sobre o alelo expandido.

Porém, um elemento complicador para nossa análise é a questão da antecipação, pois à medida em que a doença começa se manifestar mais precocemente com o passar das gerações, a janela reprodutiva dos afetados é encurtada. Assim precisamos fazer uma adaptação do fitness a cada geração representando o processo de antecipação e redução da janela reprodutiva. Nesse sentido a equação que melhor representa essa adaptação seria:

$$p' = p \cdot w \cdot coefexp \cdot 2 \cdot k$$

p' = frequência do alelo p (expandido) na geração subsequente

p = frequência do alelo p atual

w = fitness relativo

coeffexp = Coeficiente de expansão, que vai de 0 a 1, na qual irá impactar diretamente no fitness. Se os indivíduos tiverem $\text{coeffexp}=1$, o produto de 1 e w será w ($1 \cdot w = w$), ou seja, não influencia o fitness. Já se os indivíduos tiverem $\text{coeffexp}=0$ o produto de w e 0 será 0 ($0 \cdot w = 0$), ou seja, os indivíduos reduziram seu fitness a 0 e não conseguirão se reproduzir.

K = coeficiente de expansão

Ou seja:

Se um fenótipo apresentar um fitness de 1, não sofrer com a antecipação e não tiver distorção na segregação, segregar 0,5, é esperado que a sua frequência se mantenha estável (ignorando a deriva genética). Quando aplicamos esses valores hipotéticos à nossa equação em uma frequência $p=0,1$ podemos observar matematicamente a estabilidade do alelo na geração subsequente.

$$p' = (0,1) \cdot (1) \cdot (1) \cdot (2 \cdot 0,5)$$

$$p' = 0,1$$

Todavia como construir o coeficiente de expansão, de que forma ele pode representar a redução na oportunidade reprodutiva? A melhor forma em um primeiro momento é definir a amplitude da janela reprodutiva em anos desses indivíduos.

Em nossa amostra, a idade dos genitores quando tiveram seus filhos variou entre 15 e 55 anos. Idades que se aproximam da média do início da puberdade, 13 anos (Anderson 2003) e da menopausa de 49 a 53 anos (Takahashi 2015). Assim, teríamos uma janela reprodutiva de 40 anos, onde a dividiríamos em 10 e que cada parcela da divisão represente 0,1 no coeficiente de expansão.

Outro passo é prever qual a tendência de redução da idade de início com o passar das gerações. O estudo de Figueroa e cols (2017) apontou a tendência de aumento, no CAGexp, de 4,8 e 1,6 repetições CAG a cada geração para genitores homens e mulheres, respectivamente.

Como queremos analisar a antecipação independente do sexo do genitor, podemos dizer que a tendência de expansão do alelo expandido a cada geração é a média da taxa de aumento das repetições que ocorrem em genitores paternos e maternos, ou seja, um aumento de 3,2 CAG a cada geração.

No entanto, o quanto representa o aumento em 1 CAG na idade de início? Em nossas amostras, uma regressão linear entre a correlação de tamanho da CAG e idade de início mostrou que o aumento de uma unidade CAG representa a redução de 1,5 anos na idade de início da doença.

Nesse sentido, podemos prever que a cada geração a tendência é que tenhamos uma redução na idade de inícios dos sintomas de 4,8 (3,2 CAG p/ geração x 1,5 anos p/ CAG) anos a cada geração. Assim, a cada geração a janela reprodutiva é reduzida em 4,8 anos.

Assim, construímos uma tabela na qual correlaciona a geração com CAG, idade de inícios e coeficiente de expansão. Na mesma, estabelecemos como ponto de partida para geração 1: 34 repetições CAG, uma idade de início de 55 anos e coeficiente 1. E a cada geração um aumento na CAG e as respectivas reduções na idade de início e coeficiente de expansão. No entanto, fizemos uma correção de 10% para mais no coeficiente de expansão para representar a porcentagem de indivíduos que ainda nascem após os sintomas.

Tabela-1 Coeficiente de Expansão

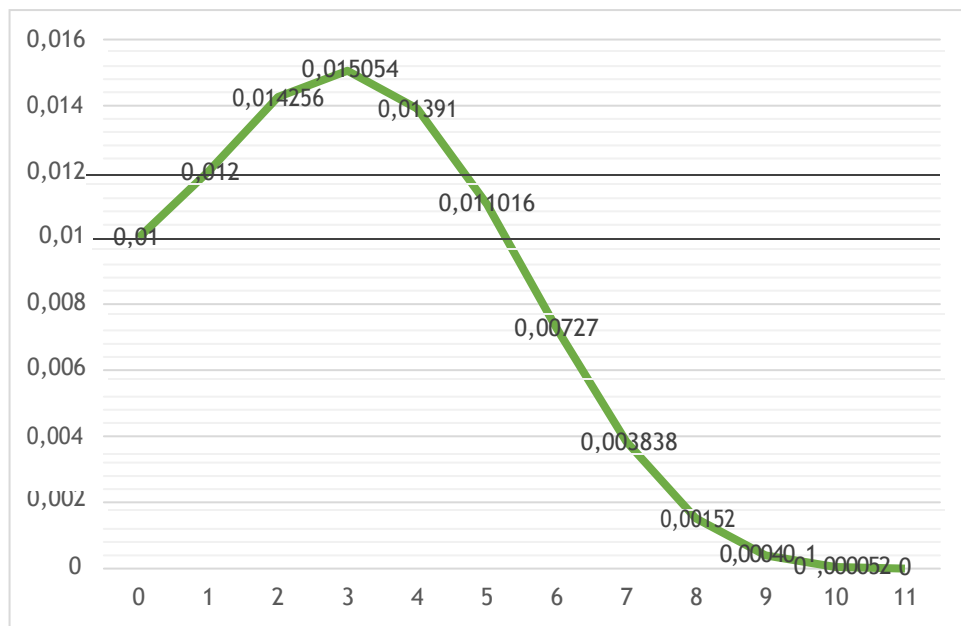
Geração	CAG	Idade de Início	Coeficiente de Expansão	Correção Coeficiente de Expansão
1	34	55	1	1,0
2	37,2	50,2	0,9	0,99
3	40,4	45,4	0,8	0,88
4	43,6	40,6	0,7	0,77
5	46,8	35,8	0,6	0,66
6	50	31	0,5	0,55
7	53,2	26,2	0,4	0,44
8	56,4	21,4	0,3	0,33
9	59,6	16,6	0,2	0,22
10	62,8	11,8	0,1	0,11
11	66	7	0	0

Como desconhecemos a frequência do alelo expandido na população e não estamos interessados em saber a frequência exata do alelo expandido (nesse

momento) nas gerações subsequentes, mas sim a sua dinâmica, iniciaremos a análise a partir de uma frequência hipotética. Essa frequência está longe ser próxima da realidade. No entanto, a sua escolha está baseada em facilidade da demonstração da equação. O valor escolhido para demonstrar a dinâmica será de 1% ou 0,01. Ou seja, se partíssemos de uma frequência 0,01 a cada geração (G_n) o quanto essa frequência iria variar?

Se colocarmos os dados de fitness, distorção de segregação e a redução na idade de início com o passar das gerações, observaríamos um aumento na prevalência da doença durante as três primeiras gerações. No entanto a medida que a antecipação se torna mais intensa observamos uma redução na prevalência até o alelo mutante ser eliminado da população.

Gráfico 1- Expectativa da dinâmica frequência alélica (y) por geração(x),



O que podemos concretamente estimar a partir dos dados existentes é que apesar da distorção da segregação em favor do alelo não expandindo, o fitness de 1.5 confere uma vantagem para o alelo mutante. A dinâmica do alelo mutante durante 3 gerações é aumentar sua frequência. No entanto, a antecipação progressivamente vai reduzindo o fitness com o passar das gerações, até o ponto de o alelo mutante ser eliminado da população.

6.2. Múltiplas origens ou uma única origem

Se o alelo ATXN2 expandido tende a desaparecer após seis ou mais gerações, cabe se supor que a SCA2 se mantém estável nas populações por conta de expansões *de novo*.

As repetições de trinucleotídeos são regiões de alta instabilidade em nosso genoma (McMurray 2010), ou seja, podem se expandir ou se contrair ao atravessar divisões celulares. A expressão disso nos alelos não expandidos de genes associados a poliglutaminopatias é que suas dimensões CAG são bastante polimórficas na população geral. No entanto, diferentemente das demais sequências repetitivas CAG em genes associados a poliQs, como as dos genes ATXN1, ATXN3, CACNA1A, ATXN7, HTT, etc, a repetição CAG do gene ATXN2 do alelo normal é bastante estável, pois em torno de 90% dos alelos humanos apresentam 22 repetições CAG ou (CAG)₂₂. (Sobczak 2014)

A primeira pergunta a ser respondida é se o alelo (CAG)₂₂ apresenta alguma diferença em relação aos demais alelos, além do número de repetição. O primeiro passo para responder essa questão é entender a estrutura do gene ATXN2. As repetições CAG se encontram no exon 1 desse gene e existem dois polimorfismos nucleares únicos (os SNPs, de *single nucleotide polymorphism*) intrônicos associados ao exon 1, rs695871 e rs695872, que podem ser marcadores para possíveis diferenças na repetição CAG do exon1 do gene ATXN2.

O SNP rs695871 pode apresentar uma base nitrogenada Citosina(C) ou Guanina(G), já o SNP rs695872 pode apresentar uma Timina(T) ou Citosina(C).

Dois estudos independentes mostraram que o SNP CC está associado a uma região CAG mais polimórfica no alelo não expandido (Shweta 2001, Ramos 2010). Além disso,

todos os indivíduos afetados estudados apresentavam o haplótipo CC. Em contraste, o haplótipo GT estava associado fortemente ao alelo (CAG)₂₂, além de não ser encontrado em indivíduos afetados.

No estudo realizado em uma corte indiana (Shweta 2001), os resultados demonstraram que 77,4% dos alelos não expandidos e que não apresentavam 22 repetições estavam associados ao haplótipo CC. Já o haplótipo GT estava associado ao alelo (CAG)₂₂ em 70,9% dos casos.

Além disso, o alelo (CAG)₂₂ associado ao haplótipo GT, na maioria dos casos, apresentava interrupções, com um padrão de repetições com 8 CAG, um CAA (que sintetiza glutamina), 4 CAG, um CAA e 8 CAG, totalizando as 22 glutaminas.

Um outro estudo com uma coorte brasileira, portuguesa e italiana verificou o mesmo padrão (Ramos 2010). O Haplótipo CC estava associado a uma região CAG mais polimórfica não expandida ou expandida, enquanto o haplótipo GT estava associado ao alelo (CAG)₂₂ com o padrão 8(CAG) CAA 4(CAG) CAA 8(CAG) ou a uma região não expandida menos polimórfica.

Fuli Yu (2005) encontrou esse padrão 8(CAG) CAA 4(CAG) CAA 8(CAG) em 73% em seu estudo em uma coorte europeia. Além disso, esse mesmo estudo observou que esse padrão de 22 repetições vem sofrendo seleção positiva.

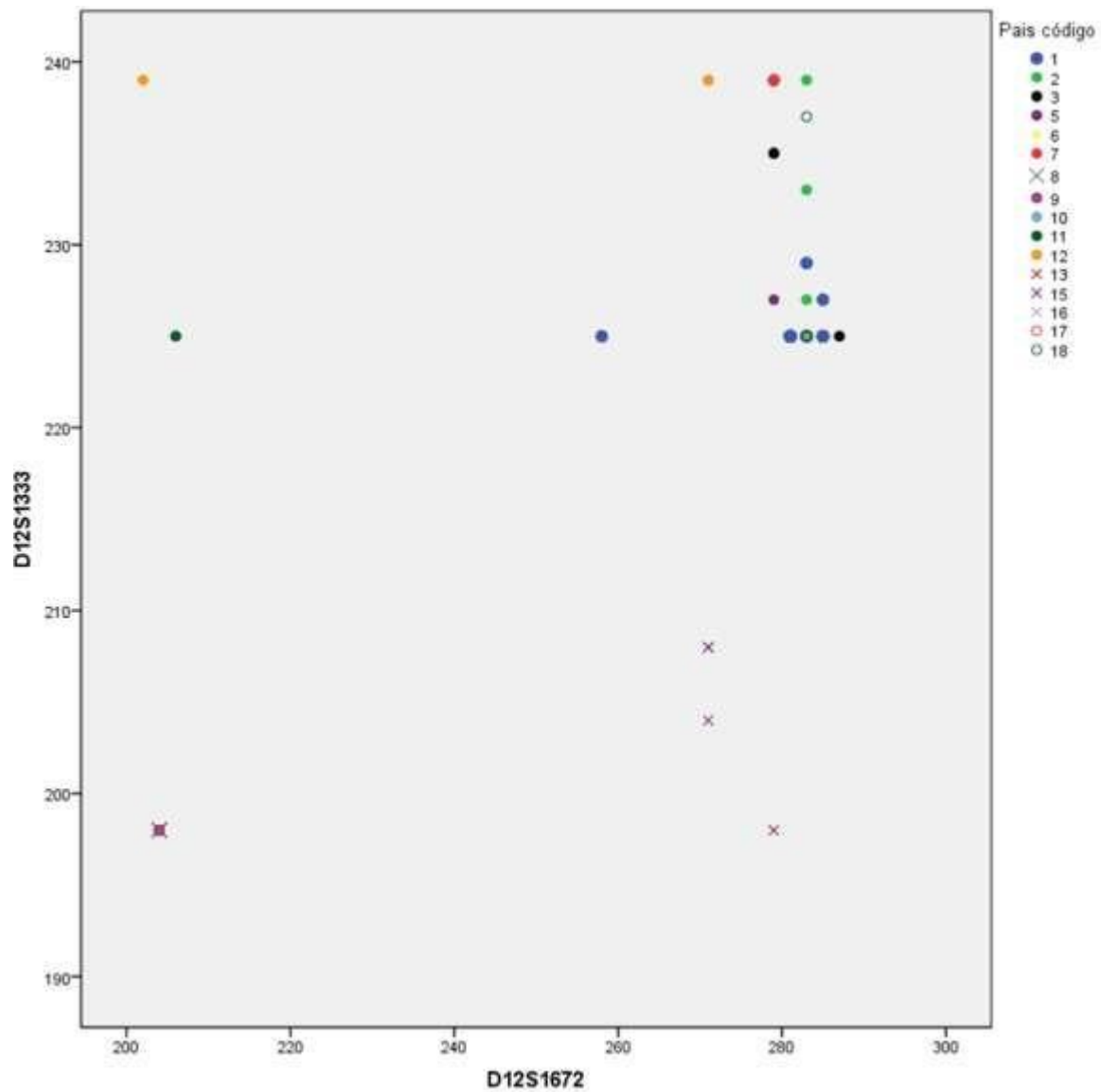
É possível se deduzir, a partir desses resultados, que exista uma região CAG estável no exon 1 no gene ATXN2 associada ao haplótipo GT, e uma outra região instável associada ao haplótipo CC. Essa região instável que está mais propensa a contrações ou expansões provavelmente seja a fonte de novas expansões causadoras da SCA2 e a

tendência é que teríamos múltiplas origens a partir dessa região instável comum a todas as linhagens de SCA2.

Buscamos na literatura estudos que analisaram outros haplótipos propostos, construídos a partir de *short tandem repeats* (STR) ou microssatélites em indivíduos com SCA2. Encontramos ao todo 4 estudos (Pang 1999, Didierjean 1999, Shweta 2001, Ramos 2010) que fizeram análise de STR das regiões D12S1333 e D12S1672 de indivíduos das seguintes nacionalidades: Portugal, Brasil, Italia, Austria, França, Cuba, Inglaterra Servia, Bélgica , Antilhas Francesa , Argélia, Marrocos e Japão.

Em seguida, construímos um gráfico de dispersão com o tamanho dos STRs das regiões D12S1333 e D12S1672 que estão em desequilíbrio de ligação. Caso os pontos estivessem associados sugeriria uma origem comum para SCA2 ou agrupado vários em *clusters* e fortalecendo a ideia de múltiplas origens. E o que observamos no gráfico foi a formação de clusters com relativa afinidade geográfica ou de processo de colonização. No entanto, essa metodologia reforça uma hipótese e não se trata de uma evidencia, de fato.

Gráfico-2 Dispersão tamanho em pb das regiões STR D12S1333 e D12S1672



Legenda país: 1-India; 2-Portugal; 3-Brasil; 5-Itália; 6-Austria; 7 e 10 França; 11-Bélgica; 12-Argélia; 13-Tunisia; 15- Índias Francesas; 16-Cuba; 17-Inglaterra; 18-Japão

7 Conclusões

Os portadores de SCA2 apresentam um fitness superior em relação aos não portadores. Essa é uma força seletiva que atua para manutenção do alelo expandido na população. No entanto, outra força atua contra o alelo mutante que é a distorção na segregação. Por fim, temos a antecipação, que com o passar das gerações reduz a idade reprodutiva dos indivíduos afetados e conseqüentemente seu fitness.

A tendência é que o alelo mutante seja eliminado da população com o passar das gerações. Assim, a sua manutenção só seria explicada pelo surgimento de novas linhagens de SCA2. A provável origem dessas novas linhagens é uma região instável associada ao haplotipos CC.

Por ser mais estável, o alelo (CAG)₂₂ não apresenta a tendência de sofrer expansões ou contrações com o passar das gerações. Os alelos expandidos, oriundos de uma região instável, apesar de uma vantagem inicial na fecundidade, após algumas gerações apresentam a tendência de ser eliminado da população.

Além disso, o alelo expandido apresenta desvantagem na segregação. Não sabemos ainda se essa desvantagem se dá perante a competição com qualquer alelo normal ou se ela é relativa apenas ao (CAG)₂₂. Nossa tendência, no entanto, é explicar que o alelo (CAG)₂₂ é o preferencialmente transmitido (evidências empíricas de seleção positiva reforçam nossa hipótese). Assim, a grande frequência do (CAG)₂₂ nas populações humanas poderia ser devida à distorção na segregação ao seu favor. Além disso, o (CAG)₂₂ não é eliminado por redução no

fitness com o passar das gerações e é mais estável tendo mais chance de ser transmitido com as mesmas 22 repetições. Uma evidência que colabora com nosso estudo é o fato de o alelo com 22 repetições com o padrão 8(CAG) CAA 4(CAG) CAA 8(CAG) estar sofrendo seleção positiva. (Fuly yu 2005)

As razões tanto da transmissão preferencial do alelo (CAG)₂₂, ou seja, a vantagem associada a esse alelo, como do fitness aumentado nos portadores de expansões CAG no ATXN2 permanecem misteriosas, pois as próprias funções da ataxina-2 permanecem em grande parte desconhecidas. Sabe-se que a ataxina-2 é uma proteína citoplásmica de 140-kDa, localizada no retículo endoplásmico rugoso, implicada na regulação de receptores de crescimento e no metabolismo do RNA e é só. Em todo o caso, a existência de um alelo muito frequente, o (CAG)₂₂, é um fenômeno específico da ataxina-2, tornando-a diferente das demais proteínas associadas às doenças de poliQs, e apontando para uma vantagem biológica associada a essa variante.

8. Perspectivas

Enquanto nossas evidências mais robustas são as relativas ao fitness, entendemos que os achados sobre distorção de segregação e sobre antecipação são ainda preliminares, pois são propensos a erros de tipo I e a vieses de observação. Por isso, pretendemos buscar evidências mais robustas para confirmar estes resultados, através da observação de irmandades e de transmissões completamente genotipadas nessa coorte.

Ainda assim, a dinâmica proposta para o alelo mutante pareceu bastante compatível com o padrão haplotípico descrito por autores que nos antecederam. Para corroborar a hipótese de que a SCA2 é uma condição que tende a reaparecer na população a partir de um pool de alelos instáveis, propomos o estudo dos haplótipos intragênicos ao ATXN2 na população brasileira normal e portadora de SCA2, construídos com os SNPs e STRs mencionados. Por nossa hipótese ser que a SCA2 apresenta múltiplas origens esperamos encontrar um padrão de SNPs e STRs significativamente distintos entre si.

9.Referências

- Anderson SE, Dallal GE, Must A (2003). «Relative weight and race influence average age at menarche: results from two nationally representative surveys of US girls studied 25 years apart». *Pediatrics*. 111 (4 Pt 1): 844–50
- Ashley JR Carter , Andrew Q Nguyen. Antagonistic pleiotropy as a widespread mechanism for the maintenance of polymorphic disease alleles. *BMC Medical Genetics*BMC series open, inclusive and trusted201112:160 DOI: 10.1186/1471-2350-12-160© Carter and Nguyen; licensee BioMed Central Ltd. 2011
- Babovic-Vuksanovic D, Snow K, Patterson MC, Michels VV. Spinocerebellar ataxia type 2 (SCA 2) in an infant with extreme CAG repeat expansion. *Am J Med Genet*. 1998;79:383–7
- Barry KL. 2015 Sexual deception in a cannibalistic mating system? Testing the Femme Fatale hypothesis. *Proc. R. Soc. B* 282: 20141428.
- Castilho AR, Mialhe FL, Barbosa TS, Puppim-Rontani RM. Influence of family environment on children’s oral health: a systematic review. *J Pediatr (Rio J)*. 2013;89:116–23.
- Cynthia T. McMurray.Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet*. 2010
- Dawkins R. *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design* 1986
- de Araújo MA, Raposo M, Kazachkova N, Vasconcelos J, Kay T, et al. (2016) Trends in the Epidemiology of Spinocerebellar Ataxia Type 3/ Machado-Joseph Disease in the Azores Islands, Portugal. *JSM Brain Sci* 1(1): 1001
- de Castilhos RM, Furtado GV, Gheno TC, Schaeffer P, Russo A, Barsottini O, et al. Spinocerebellar ataxias in Brazil—frequencies and modulating effects of related genes. *Cerebellum*. 2014;13:17–28.
- Diamond, J. M. Tay-Sachs carriers and tuberculosis resistance. (Letter) *Nature* 331: 666, 1988.

- Didierjean O, Cancel G, Stevanin G, et al. Linkage disequilibrium at the SCA2 locus. *Journal of Medical Genetics*. 1999;36(5):415-417.
- Dobzhansky T, ""Biology, Molecular and Organismic", *American Zoologist*, volume 4 1964.
- Duyao M1, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, Folstein S, Ross C, Franz M, Abbott M, et al. Trinucleotide repeat length instability and age of onset in Huntington's disease. 1993
- Fea MP, Stanley MC, Holwell GI. Fatal attraction: sexually cannibalistic invaders attract naive native mantids. 2013 *Biol Lett* 9: 0746.
- Figueroa KP, Coon H, Santos N, et al. Genetic analysis of age at onset variation in spinocerebellar ataxia type 2. *Neurology: Genetics*. 2017;3(3):e155.
- Frontali M1, Sabbadini G, Novelletto A, Jodice C, Naso F, Spadaro M, Giunti P, Jacopini AG, Veneziano L, Mantuano E, Malaspina P, Ulizzi L, Brice A, Durr A, Terrenato L. Genetic fitness in Huntington's Disease and Spinocerebellar Ataxia 1: a population genetics model for CAG repeat expansions. 1996
- Fuli Yu, Pardis C Sabeti, Paul Hardenbol, Qing Fu, Ben Fry, Xiuhua Lu, Sy Ghose, Richard Vega, Ag Perez, Shiran Pasternak, Suzanne M Leal, Thomas D Willis, David L Nelson, John Belmont, Richard A Gibbs. Positive Selection of a Pre-Expansion CAG Repeat of the Human SCA2 Gene. *Plos Genetic*. 2005
- Gabriele Nunes Souza , Nathalia Kersting , Ana Carolina Krum-Santos , Amanda Senna P. Santos , Gabriel Vasata Furtado, Daphne Pacheco , Thomaz Abramsson Gonçalves , Jonas Alex Morales Saute , Lavínia Schuler-Faccini; Eduardo Preusser Mattos , Maria Luiza Saraiva-Pereira , Laura Bannach Jardim. Spinocerebellar ataxia type 3/Machado-Joseph disease: segregation patterns and factors influencing instability of expanded CAG transmissions. 2015
- Griffiths, Anthony J. F. Lewontin, Richard C.; Carroll, Sean B.; Wessler, Susan R. *Introdução À Genética - 10ª Ed.* 2013
- Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier JM, Weber C, Mandel JL, Cancel G, Abbas N, Dürr A, Didierjean O, Stevanin G, Agid Y, Brice A. Cloning of

the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. 1996

Kieling C, Prestes PR, Saraiva-Pereira ML, Jardim LB. Survival estimates for patients with Machado-Joseph disease (SCA3). *Clin Genet*. 2007 Dec;72(6):543-5. Epub 2007 Sep 25.

Kenneth Wexler. Innateness and maturation in linguistic development. 2004

Langbehn DR, Hayden M, Paulsen JS, the PREDICT-HD Investigators of the Huntington Study Group. CAG-Repeat Length and the Age of Onset in Huntington Disease (HD): A Review and Validation Study of Statistical Approaches. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 2010;153B(2):397-408. doi:10.1002/ajmg.b.30992.

Lastres-Becker I1, Brodesser S, Lütjohann D, Azizov M, Buchmann J, Hintermann E, Sandhoff K, Schürmann A, Nowock J, Auburger G. Insulin receptor and lipid metabolism pathology in ataxin-2 knock-out mice.

Luis Velázquez-Pérez, Roberto Rodríguez-Labrada, Hans-Joachim Freund and Georg Auburger (2012) Spinocerebellar Ataxia Type 2, Spinocerebellar Ataxia, Dr. José Gazulla (Ed.), ISBN: 978-953-51-0542-8, InTech, Available from: <http://www.intechopen.com/books/spinocerebellar-ataxia/spinocerebellar-ataxia-type-2-sca2-2012>

Mao B1, Wu W, Davidson G, Marhold J, Li M, Mechler BM, Delius H, Hoppe D, Stannek P, Walter C, Glinka A, Niehrs C. Kremen proteins are Dickkopf receptors that regulate Wnt/beta-catenin signalling. 2002

McMurray CT. Mechanisms of trinucleotide repeat instability during human development. *Nature reviews Genetics*. 2010;11(11):786-799. doi:10.1038/nrg2828.

Morrison PJ, S Harding-Lester and A Bradley. Uptake of Huntington disease predictive testing in a complete population. 2011

Monte TL1, Pereira FS2, Reckziegel EDR3, Augustin MC3, Locks-Coelho LD3, Santos ASP3, Pedroso JL4, Barsottini O4, Vargas FR5, Saraiva-Pereira ML6, Jardim LB7; Rede Neurogenética. Neurological phenotypes in spinocerebellar ataxia type 2:

Role of mitochondrial polymorphism A10398G and other risk factors. *Parkinsonism Relat Disord.* 2017

Naylor N, Richardson SJ, McAllan BM. Boom and bust: a review of the physiology of the marsupial genus *Antechinus*. *J Comp Physiol B.* 2008 Jul;178(5):545-62. doi: 10.1007/s00360-007-0250-8. Epub 2008 Jan 22.

Pang J, Allotey R, Wadia N, Sasaki H, Bindoff L, Chamberlain S. A common disease haplotype segregating in spinocerebellar ataxia 2 (SCA2) pedigrees of diverse ethnic origin. *Eur J Hum Genet.* 1999.

Prestes PR, Saraiva-Pereira ML, Silveira I, Sequeiros J, Jardim LB. Machado Joseph Disease Enhances Genetic Fitness: A Comparison Between Affected and Unaffected Women and Between MJD and the General Population. *Annals of Human Genetics* 2008; 72: 57–64.

Pringsheim T1, Wiltshire K, Day L, Dykeman J, Steeves T, Jette N. The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis. 2012

Pulst, S. M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X. N., Lopes-Cendes, I. et al. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat. Genet.* 14, 269-276. 1996

Pulst SM. Spinocerebellar Ataxia Type 2. 1998 Oct 23 [Updated 2015 Nov 12]. In: Pagon RA, Adam MP, Ardinger HH, et al., editors. *GeneReviews*® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2016.

Schöls L, Bauer P, Schmidt T, Schulte T, Riess O Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. *Lancet Neurol* 3:291–304. (2004)

Sequeiros J, Martins S, Silveira I. Epidemiology and population genetics of degenerative ataxias. *Handb Clin Neurol* 103:227-251, 2012.

Stine OC, Pleasant N, Franz ML, Abbott MH, Folstein SE, Ross CA. Correlation between the onset age of Huntington's disease and length of the trinucleotide repeat in IT-15. *Hum Mol Genet.* 1993 Oct; 2(10):1547-9.

- Shweta Choudhry, Mitali Mukerji, Achal K. Srivastava, Satish Jain, Samir K. Brahmachari; CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms, *Human Molecular Genetics*, Volume 10, Issue 21, 2 October 2001, Pages 2437–2446,
- Sobczak K, Krzyzosiak WJ (2004) Patterns of CAG repeat interruptions in SCA1 and SCA2 genes in relation to repeat instability. *Human Mutation* 24: 236–247
- Sutherland GR, Richards RI. Simple tandem DNA repeats and human genetic disease. *Proceedings of the National Academy of Sciences of the United States of America*. 1995;92(9):3636-3641.
- Ramos EM, Martins S, Alonso I, Emmel VE, Saraiva-Pereira ML, Jardim LB, Coutinho P, Sequeiros J, Silveira I. 2010. Common Origin of Pure and Interrupted Repeat Expansions in Spinocerebellar Ataxia Type 2 (SCA2). *Am J Med Genet Part B* 153B:524–531.
- Richards R.I. and Sutherland,G.R. Dynamic mutation: possible mechanisms and significance in human disease. 1997. *Trends Biochem. Sci.*, 22, 432–436.
- Rosenblatt A, Liang KY, Zhou H, Abbott MH, Gourley LM, Margolis RL, Brandt J, Ross CA. The association of CAG repeat length with clinical progression in Huntington disease. *Neurology*. 2006;66:1016–20. [PubMed]
- Ross, C. A. et al. Huntington disease: natural history, biomarkers and prospects for therapeutics *Nature. Rev. Neurol.* 10, 204–216 (2014), published online 11 March 2014;
- Tamara Pringsheim MD, MSc1, Katie Wiltshire MD1, Lundy Day BSc1, Jonathan Dykeman BSc1, Thomas Steeves MD, MSc2 andNathalie Jette MD, MSc1.The incidence and prevalence of Huntington's disease: A systematic review and meta-analysis. DOI: 10.1002/mds.25075. 2012
- .
- Takahashi, TA; Johnson, KM (2015). «Menopause.». *The Medical clinics of North America*. 99 (3): 521–34. PMID 25841598. doi:10.1016/j.mcna.2015.01.006
- The U.S.–Venezuela Collaborative Research Project, Wexler NS. Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington’s disease age of

onset. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(10):3498-3503. doi:10.1073/pnas.0308679101.

Trott A, Maris AF, Miranda GB. Ataxias espinocerebelares causadas por expansão de poliglutamina: uma revisão.2009

Trut, Lyudmila . "Early Canid Domestication: The Farm-Fox Experiment . 1999

Warby SC, Graham RK, Hayden MR. Huntington Disease. 1998 Oct 23 [Updated 2014 Dec 11]. In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2016. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1305/>

Williams, G. C. 1957. Pleiotropy, Natural Selection, and the Evolution of Senescence. Evolution 11;4: 398-411

van Exel E, Koopman JJE, Bodegom Dv, Meij JJ, Knijff Pd, et al. (2017) Effect of APOE ε4 allele on survival and fertility in an adverse environment. PLOS ONE 12(7): e0179497.

Vincent Dion. Tissue specificity in DNA repair: lessons from trinucleotide repeat instability. 2014

Vogel F, Motulsky AG Human genetics. Problems and approaches, 3rd completely rev. ed. 1997