UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCIÊNCIAS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

ATAXIAS ESPINOCEREBELARES TIPO 2 E TIPO 3: UMA PERSPECTIVA EVOLUTIVA

LUCAS SCHENATTO DE SENA

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de **Doutor em Genética e Biologia Molecular.**

Orientadora: Profª. Drª. Laura Bannach Jardim

Coorientadora: Profª. Drª. Maria Luiza Saraiva-Pereira

Porto Alegre, agosto de 2022.

O presente trabalho foi desenvolvido pelas seguintes instituições:

● Centros de Pesquisa Clínica e Experimental, Hospital de clínicas de Porto Alegre (HCPA);

● Serviço de Genética Médica do HCPA;

Agências financiadoras:

● Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

● Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

● Fundo de incentivo à pesquisa e Eventos do Hospital de clínicas de Porto Alegre (FIPE-HCPA)

““Ignorance more frequently begets confidence than does knowledge: it is those who know little, not those who know much, who so positively assert that this or that problem will never be solved by science.”

Charles Darwin(1809-1882)

Agradecimentos

“Sejamos realistas, exijamos o impossível”
Paris, Maio de 68

O trajeto até aqui é precedido de uma história e algumas pessoas determinantes para que esse momento acontecesse, quero agradecê-las profundamente por isso.

Durante 8 anos trabalhei nos Correios entregando cartas ou descarregando caminhões quando decidi dar uma guinada em minha vida e fazer licenciatura em biologia, me inscrevi em uma faculdade EAD pois era a modalidade que meu corpo aguentaria. Prestes a terminar o curso domei uma segunda decisão, faria mestrado na UFRGS, não sabia exatamente do que, mas que esse passo seria importante para “maquiar” minha graduação EAD. Comecei a pesquisar os cursos e quase todos processos seletivos tinham um padrão, muitas pessoas passavam nas provas e o desempate era no currículo (inexistente no meu caso) e/ou na entrevista, mas tinha uma exceção o PPGBM, sempre sobrava vagas pois a maioria dos inscritos rodavam na prova.

Então “bastava” passar na prova, o que era uma tarefa complexa, isso porque meus conhecimentos em genética/biologia molecular não eram os dos melhores, afinal os “aprendi” 3 aulas de 1 hora, esse era o primeiro grande desafio. Aqui entra a primeira pessoa determinante para superá-lo, minha mãe. Saio dos correios para estudar, pego um estágio e minha renda drasticamente é reduzida. Tinha que sustentar a casa, filho, a vida se converteu em um turbilhão, e minha mãe se tornou um importante polo de apoio material e moral. Ela confiou no meu projeto, e fez todo esforço possível para que eu pudesse realizar esse desafio.

Nunca vou esquecer o dia que fui na casa dela e pedi para ela comprar o livro texto para estudar para seleção do mestrado, ela não tinha noção do quão caro um livro poderia ser, mas, apesar do salário baixo que recebia como empregada doméstica, me acompanhou até a loja e o comprou. Foi um período intenso debruçado naquele livro, intercalado com idas à casa da minha mãe para pegar comida e levar para casa. Muito obrigado mãe, por confiar em mim, sem esse apoio nada disso ocorreria. Vejo o brilho nos teus olhos quando tu falas que teu filho faz doutorado na UFRGS, sinto uma alegria indescritível te deixar orgulhosa e te prometo deixar ainda mais.

Após passar no processo seletivo, entra a segunda pessoa fundamental para realização deste trabalho, minha orientadora Laura Jardim, tive uma sorte absurda em ter te escolhido como orientadora. Lembro-me da primeira vez que falei com você, em dezembro de 2015. Fui ao HCPA para saber se ela gostaria de ser minha orientadora, dizer que não tinha nenhuma ideia do que gostaria de pesquisar, mas que gostava de Evolução. Esperava encontrar um ambiente pouco receptivo, mas a Laura fez com que ele se tornasse muito acolhedor.

Laura, você deve ter tido um desafio e tanto em orientar alguém com uma “trajetória acadêmica” como a minha, que tinha que aprender tudo do zero e queria estudar evolução, que não era a principal linha de investigação do nosso grupo, obrigado pela dedicação. Sou grato por três fatores fundamentais no papel que cumpriste na minha formação e execução desse trabalho. Primeiro, pelo equilíbrio complexo de colocar numa mesma equação: exigência, incentivo, cobrança, apoio, ensinar; me acompanhar e me dar autonomia. A combinação de todos esses elementos só foi possível devido ao fato de ser a pesquisadora, a professora e principalmente o ser humano que és.

Segundo, a me ajudar na construção de uma linha de investigação a partir de uma pergunta aparentemente simples (mas que na essência é complexa), que se tornou a força motriz de todo esse processo, muito obrigado por ter me ensinado a essência da ciência: a ser movido pelas perguntas. Terceiro, por ter aperfeiçoado minha forma de pensar me ajudando a entender a importância da organização das informações, dados, etc. Além de me fazer estabelecer os processos de uma maneira mais sequencial. Muito obrigado por tudo isso, Laura.

Quero agradecer a minha coorientadora, Maria Luiza, pelo aporte e ensinamentos principalmente nas questões metodológicas e laboratoriais, pelo papel que desempenhou nas redações de nossas publicações científicas, atentando e me corrigindo em muitos elementos e detalhes que a mim passavam despercebidos. Luiza, você é um exemplo de superação, ser cientista e professora universitária em uma sociedade excludente, sendo mulher, negra e oriunda de escola pública, faz com que seja uma fonte de inspiração.

A todos meus colegas do grupo de pesquisa em especial a Jordânia Pinheiro que teve um papel fundamental nos manuscritos 1 e 2, revisando cada artigo e resultados das revisões sistemáticas comigo, Gabriel Furtado, que teve um aporte determinante na parte laboratorial, auxiliando na obtenção dos dados para o nosso manuscrito 4. Além dos meus colegas de grupo, agradeço ao colega Renan Lemes, do Instituto de Biociências da USP. por me ajudar e ensinar a realizar as simulações do manuscrito 3.

Agradeço a minha companheira Jade por estar ao meu lado em um dos momentos complexos de minha vida e ser um importante elemento de apoio nos momentos mais críticos. De topar embarcar na minha vida, nos meus sonhos e minhas loucuras, que agora são nossos, e viver a cada momento de forma intensa. Faço das palavras de Carl Sagan para sua esposa Ann Druyan as minhas para você: *“Diante da vastidão do tempo e da imensidão do universo, é um imenso prazer para mim dividir um planeta e uma época com você.”*

Por fim, agradeço aquele que de uma forma subjetiva vem cumprindo um papel determinante nos últimos 12 anos, meu filho Leonardo. Te agradeço profundamente por ser meu incentivo a nunca recuar.

SUMÁRIO

RESUMO 9

ABSTRACT 11

CAPÍTULO I. INTRODUÇÃO E OBJETIVOS 13

1. INTRODUÇÃO 13

1.1 Poliglutaminopatias 13

1.1.1 SCA3/MJD 14

1.1.1.1 Descrição e epidemiologia 14

1.1.1.2 Da neurotoxicidade às Manifestações Clínicas 16

1.1.1.3 O gene ATXN3 17

1.1.1.3.1 Aspectos gerais 17

1.1.1.3.2 O polimorfismo no tamanho da CAGexp e sua relação com a idade de início 18

1.1.1.3.3 Fatores que podem influenciar a variação da frequência alélica do ATXN3 19

1.1.1.3.4 Haplótipos ancestrais 20

1.1.2 Ataxia espinocerebelar tipo 2 20

1.1.2.1 Descrição e epidemiologia 21

1.1.2.2 Da neurotoxicidade as Manifestações Clínicas: 21

1.1.2.3 O gene ATXN2 22

1.1.2.3.1 Aspectos gerais: 22

1.1.2.3.2 O polimorfismo nas repetições CAG sua relação com a idade de início 23

1.1.2.3.3 Fatores que podem influenciar a variação da frequência alélica do ATXN2 24

1.1.2.3.4 Haplótipos ancestrais 24

1.2 Evolução e doenças poliQs 25

1.2.1 Antecipação, uma força seletiva negativa para manutenção dos alelos expandidos. 26

1.2.2 Equilíbrio seleção mutação 26

1.2.3 Pleiotropia Antagonica 27

1.2.4 Abordagens evolutivas nas poliQs 28

2 Objetivo 30

2.1 Objetivo Geral 30

2.2 Objetivos específicos: 30

CAPÍTULO II. FORÇAS SELETIVAS ASSOCIADAS AOS ALELOS EXPANDIDOS DO ATXN3 31

Manuscrito 1: Selective forces acting on spinocerebellar ataxia type 3/Machado-Joseph disease recurrency: A systematic review and meta-analysis. 31

CAPÍTULO III. FORÇAS SELETIVAS ASSOCIADAS AOS ALELOS EXPANDIDOS DO ATXN2 44

Manuscrito 2: “Spinocerebellar ataxia type 2 from an evolutionary perspective: Systematic review and meta-analysis” 44

CAPÍTULO IV: DINÂMICA DOS ALELOS EXPANDIDOS DO ATXN2 E ATXN3 55

Manuscrito 3: A model for the dynamics of expanded CAG repeat alleles: ATXN2 and ATXN3 as prototypes. 55

CAPÍTULO V. HAPLÓTIPOS ANCESTRAIS DAS LINHAGENS SCA2 NO BRASIL, NO PERU E NO URUGUAI 95

CAPÍTULO VI. CONCLUSÕES, LIMITES E PERSPECTIVAS 123

REFERÊNCIAS BIBLIOGRÁFICAS GERAIS: 133

RESUMO

As Ataxias Espinocerebelares tipo 2 (SCA2) e tipo 3 ou doença de Machado Joseph (SCA3/DMJ) fazem parte de um grupo de doenças autossômicas dominantes neurodegenerativas conhecidas como poliglutaminopatias, causadas por expansões de sequências repetitivas traduzidas do trinucleotídeo CAG (CAGexp) que por sua vez codificam tratos expandidos e neurotóxicos de glutamina nas proteínas associadas. Na SCA2 e SCA3/DMJ, as CAGexp estão localizadas respectivamente nos genes *ATXN2* e *ATXN3* e apresentam instabilidades em sua transmissão, tendendo a aumentar de tamanho a cada geração. A idade de início dos sintomas, essencialmente motores, apresenta uma correlação inversa com a CAGexp. Dessa forma, há uma tendência dos sintomas começarem mais cedo na prole do que no genitor afetado, fenômeno conhecido como antecipação. O objetivo deste trabalho foi ajudar a compreender como doenças neurodegenerativas autossômicas dominantes, que tendem a se manifestar mais cedo a cada geração, não são eliminadas pela seleção natural, através do estudo da dinâmica dos seus alelos expandidos. Inicialmente, realizamos duas revisões sistemáticas - uma para SCA3 e outra para SCA2 - nas quais consolidamos as evidências até hoje obtidas sobre antecipação, instabilidade, fitness e segregação de alelos. Quando possível, metanalizamos as forças evolutivas que pudessem influenciar a frequência alélica dos CAGexp no *ATXN3* e no *ATXN2.* Na sequência, propusemos um modelo matemático baseado em uma equação clássica da genética de populações, mas voltado à previsão da dinâmica de alelos dominantes CAGexp. Alimentamos este modelo com os dados obtidos das revisões sistemáticas e metanálises sobre o *ATXN3* e o *ATXN2* e realizamos simulações computacionais para prever em quantas gerações os alelos expandidos se manteriam em mil linhagens por gene, e em simulações que durassem até 650 gerações. Os resultados das simulações apontaram que o *ATXN3* e o *ATXN2* expandidos apresentam dinâmicas distintas. Várias linhagens com o alelo expandido no *ATXN3* se mantiveram por 650 gerações, denotando que a combinação do fitness elevado e da distorção da segregação favorecendo o alelo mutante na SCA3/DMJ - dados observacionais metanalizados - parecem compensar a seleção negativa produzida pela antecipação. Nossos resultados convergem com os resultados descritos na

literatura, que apontam que a SCA3/DMJ apresenta poucas origens ancestrais antigas, sem descrição de mutações *de novo*. Em contraste, as simulações realizadas para o *ATXN2* resultaram em extinção das linhagens em uma mediana (variação) de 10 (1 a 149) gerações. Se os dados observacionais até hoje obtidos - e usados nas nossas simulações - para fitness, distorção da segregação e antecipação da SCA2 estão corretos, a dinâmica que o nosso modelo descreveu para o *ATXN2* expandido obriga-nos a prever que mutações *de novo* sejam um fator que explique a manutenção da SCA2 nas populações. Para sustentar essa hipótese, seria necessário se demonstrar que as famílias SCA2 tenham origens ancestrais distintas. Estudos sobre haplótipos ancestrais são escassos na literatura. Os mais robustos genotiparam o rs695871 e apontaram para a existência de um haplótipo comum às famílias estudadas. Para responder a essa questão, realizamos então o estudo de 47 famílias SCA2 do Brasil, do Peru e do Uruguai, nas quais determinamos haplótipos intragênicos construídos com os marcadores rs9300319, rs695871, rs593226, D12S1333 e D12S1672 ligados ao CAGexp no *ATXN2.* Encontramos seis haplótipos ancestrais nas famílias SCA2. Os haplótipos rs9300319-rs695871-rs593226 mais comuns foram o T-C-G e o C-C-G; digna de nota foi também a descrição de três haplótipos contendo o alelo G no SNP rs695871: C-G-A, T-G-A e T-G-G. Nossos achados aumentaram sobremaneira a variedade de origens da SCA2 e vieram ao encontro das inferências do nosso modelo matemático, que previam a necessidade de haver múltiplas origens para sustentar a presença da SCA2 nas populações. Assim, concluímos que o modelo sobre a dinâmica dos CAGexp pareceu ter previsto eventos confirmados pela observação do que acontece com a SCA2 e a SCA3/MJD. Embora esse modelo ainda precise ser testado para outras poliglutaminopatias, nós propomos que ele deverá dar conta de muitos fenômenos que ajudam a explicar a existência dessas doenças nas populações.

ABSTRACT

Spinocerebellar ataxias type 2 (SCA2) and type 3 or Machado Joseph's disease (SCA3/MJD) belong to a group of autosomal dominant neurodegenerative diseases known as polyglutaminopathies (polyQ diseases). PolyQ diseases are caused by expansions of translated repetitive sequences of the CAG trinucleotide (CAGexp) that in turn encode expanded and neurotoxic polyglutamine tracts in associated proteins. CAGexp are located in the *ATXN2* and *ATXN3* genes in SCA2 and SCA3/DMJ, respectively, and are unstable upon transmission, tending to increase in size with each generation. The age of onset of symptoms, mainly motor, has an inverse correlation with CAGexp. There is a tendency for symptoms to start earlier in the offspring than in the affected parent, a phenomenon known as anticipation. The aim of the present thesis was to help understand how autosomal dominant neurodegenerative diseases, which tend to manifest earlier in each generation, are not eliminated by natural selection, through the study of the dynamics of their expanded alleles. Initially, we carried out two systematic reviews - one for SCA3/MJD and one for SCA2 - in which we consolidated the evidence obtained so far on anticipation, instability, fitness and allele segregation. When possible, we meta-analyzed the evolutionary forces that could influence the allele frequency of CAGexp in *ATXN3* and *ATXN2*. Next, we proposed a mathematical model based on a classical equation from population genetics, but aimed at predicting the dynamics of dominant CAGexp alleles. We fed this model with data obtained from systematic reviews and meta-analyses on *ATXN3* and *ATXN2* and performed computer simulations to predict in how many generations the expanded alleles would remain at 1,000 lineages per gene, and in simulations lasting up to 650 generations. The simulation results showed that the expanded *ATXN3* and *ATXN2* present different dynamics. Several lineages carrying the expanded *ATXN3* allele were maintained for 650 generations, indicating that the combination of high fitness and segregation distortion favoring the mutant allele in SCA3/DMJ – meta-analyzed observational data - seems to compensate for the negative selection produced by anticipation. Our results converge with the results described in the literature, which indicate that SCA3/DMJ has few ancient ancestral origins, with no description of *de novo*

mutations. In contrast, the simulations performed for the expanded *ATXN2* resulted in extinction of all lineages in a median (range) of 10 (1 to 149) generations. If the observational data obtained so far - and used in our simulations - for fitness, segregation distortion and anticipation of SCA2 are correct, the dynamics that our model described for the expanded *ATXN2* forces us to predict that *de novo* mutations are obligatory to maintain SCA2 in populations. To support this hypothesis, it would be necessary to demonstrate that SCA2 families have distinct ancestral origins. Studies on ancestral haplotypes are scarce in the literature. The most robust of them have genotyped rs695871 and pointed to the existence of one ancestral haplotype, common to the families studied. To answer this question, we carried out a study of 47 SCA2 families from Brazil, Peru and Uruguay, in which we determined intragenic haplotypes constructed with the markers rs9300319, rs695871, rs593226, D12S1333 and D12S1672 linked to CAGexp in *ATXN2*. We found six ancestral haplotypes in the SCA2 families. The most common rs9300319-rs695871-rs593226 haplotypes were T-C-G and C-C-G; noteworthy was also the description of three haplotypes containing the G allele in SNP rs695871: C-G-A, T-G-A and T-G-G. Our findings greatly increased the variety of SCA2 origins and met the inferences of our mathematical model, which predicted the need for multiple origins to support the presence of SCA2 in populations. Thus, we conclude that the model on the dynamics of CAGexp seemed to have predicted events confirmed by the observation of what happens with SCA2 and SCA3/MJD. Although this model still needs to be tested for other polyQ diseases, we propose that it should account for the majority of phenomena that explain the existence of these diseases in populations.

CAPÍTULO I. INTRODUÇÃO E OBJETIVOS

1. INTRODUÇÃO

1.1 Poliglutaminopatias

As Poliglutaminopatias ou doenças de poliQs são um conjunto de doenças neurodegenerativas autossômicas dominantes, causadas por expansões de repetições CAG (CAGexp), nos seus respectivos genes, que passam a codificar tratos de poliglutamina (poliQ) nas respectivas proteínas traduzidas. Estão incluídas nas doenças de poliQs a doença de Huntington (HD), a Ataxia espinocerebelar tipo 1 (SCA1), tipo 2 (SCA2), tipo 3 (também conhecida como doença de Machado-Joseph, SCA3/MJD), tipo 6 (SCA6), tipo 7 (SCA7), tipo 17 (SCA17), atrofia dentato-rubro-palido-luisiana (DRPLA) e atrofia muscular espinobulbar (SBMA, também conhecida como doença de Kennedy) (Lieberman et al. 2019).

Evidências experimentais robustas sustentam que os tratos de poliglutamina (poliQ) expandidos adquirem um caráter neurotóxico, sendo um dos mais famosos exemplos de mutações patogênicas por ganho de função tóxica. A neurotoxicidade parece se dar por distintos mecanismos, incluindo modificações pós-transcricionais e translacionais, e mecanismos autofágicos (Adegbuyiroa et al. 2017; Bunting et al. 2021). A neurodegeneração causada pela proteína neurotóxica, leva a sintomas predominantemente motores (Fan et al. 2014).

Nessas doenças, o início dos sintomas geralmente se dá na vida adulta, com uma relação inversa entre idade de início e tamanho das repetições CAG. Ademais, as CAGexp caracterizam-se por serem repetições instáveis. Na maior das doenças de poliQs, as CAGexp tendem a se expandir ao atravessar meioses e, consequentemente, a se associarem a reduções na idade de início dos sintomas na geração seguinte (Stoyas & Spada. 2018).

As doenças PolyQ são raras, sendo a prevalência médias de 1–10 casos por 100.000 pessoas (Margulis BA et al. 2013), sendo a HD a mais prevalente na Europa (Bettencourt et al. 2016). Mundialmente, dentre as ataxias espinocerebelares as mais prevalentes são a SCA3/MJD seguida da SCA2 (Ruano et al. 2014). As razões pelas quais a SCA3/MJD e SCA2 são as mais prevalentes dentre as ataxias dominantes não foram elucidadas até o momento.

1.1.1 SCA3/MJD

1.1.1.1 Descrição e epidemiologia

Nos anos 1970, duas grandes famílias estadunidenses de origem açoriana foram descritas de forma isolada em três publicações. Membros da família Machado, proveniente da Ilha de São Miguel (Nakano et al. 1972; Woods et al. 1972), e da família Joseph, proveniente da Ilha de Flores (Rosenberg et al. 1976) apresentavam manifestações neurológicas autossômicas dominantes desde o início da vida adulta, como falta de equilíbrio e dificuldade na coordenação motora. Com base nesses relatos, uma série de estudos observacionais estabeleceram critérios clínicos para o diagnóstico da doença de Machado Joseph (Lima & Coutinho 1980). No Brasil, o primeiro relato de indivíduos com essa doença ocorreu em 1988, em São Paulo (Radvany et al. 1988).

Em, 1993 o locus associado à doença foi descoberto no cromossomo 14 (14q24.3-q32) (Takiyama et al. 1993), e em 1994 uma região de repetições CAG nesse locus foi associada a indivíduos com a doença de Machado Joseph (Kawaguchi et al. 1994). Essa sequência repetitiva CAG está localizada no gene hoje chamado de *ATXN3 e* responsável pela síntese da proteína Ataxina 3 *(*Paulson et al. 1997*)*. A doença de Machado Joseph, então, foi a terceira ataxia espinocerebelar autossômica dominante a ter seu gene descoberto, e por isso passou a ser definida como SCA3/MJD.

Os primeiros casos de SCA3/MJD gaúchos foram identificados em 1997 pela Profa. Laura Bannach Jardim, em colaboração com os Profs. Guy Rouleau (Universidade MacGill, de Montreal) e Jorge Sequeiros (Universidade do Porto). Suas características clínicas e genéticas têm sido relatadas sob variadas perspectivas desde então. Ambulatórios e laboratório assistenciais bem como instalações de pesquisa voltados à SCA3/MJD foram organizados no Hospital de Clínicas e no nosso PPG, com a participação da Profa. Maria Luiza Saraiva Pereira.

Não existem estudos que estimem a prevalência global da SCA3/MJD, porém dentre as ataxias espinocerebelares ela é a mais comum, sendo responsável por 15% a 45% dos casos (Paulson. 2007; Schöls et al. 2004). No entanto, sua frequência entre as SCAs apresenta heterogeneidade ao redor do mundo. É possível dividi-la em regiões com frequência mais elevadas: Portugal (58%) (Vale et al., 2010), Singapura (53%) (Zhao et al. 2002), China (49%) (Jiang et al. 2005), Holanda (44%) (van de Warrenburg et al. 2002) e Alemanha (42%) (Schols et al. 1997); e regiões com frequências menores: Canadá (24%) (Kraft et al. 2005), EUA (21%) (Moseley et al. 1998), México (12%) (Alonso et al. 2007), Austrália (12%) (Storey et al. 2000), Índia (3%) (Faruq et al. 2009), África do Sul (4%) (Bryer et al. 2003) e Itália (1%) (Brusco et al. 2004).

O arquipélago dos Açores é a região do planeta com a maior prevalência estimada mundial com 39:100.000. Como a população dos Açores é pequena, de aproximadamente 237.000 pessoas, a população de sujeitos SCA3/MJD sintomáticos gira em torno de 92 pessoas, naquelas ilhas. É de se notar que nessa região a prevalência da SCA3/DMJ tem aumentado nas últimas duas décadas (de Araújo et al. 2016).

No Brasil, nosso grupo descreveu que a SCA3/DMJ é responsável por 59.6% casos de ataxias espinocerebelares. Por sua vez, nos estados do Rio Grande do Sul (RS) e Santa Catarina essa proporção é de 78.4% (de Castilhos et al. 2014). No RS, a prevalência estimada da SCA3/DMJ é de 7:100.000 (Rodríguez-Labrada et al. 2020), a partir de 770 sujeitos sintomáticos identificados em busca ativa, em 2019. Essa frequência elevada de SCA3/MJD tem sido atribuída, pelo nosso grupo, a um efeito fundador gerado pela migração de mulheres e homens originários dos

Açores entre 1750 e 1770, estabelecidos no RS para ocupar o espaço geográfico recém conquistado pela coroa portuguesa (Jardim et al. 2001).

1.1.1.2 Da neurotoxicidade às Manifestações Clínicas

A proteína correspondente ao gene *ATXN3* é denominada ataxina-3, e sua expressão já foi detectada em todos os tecidos investigados, tanto no período embrionário como na vida adulta (Ichikawa et al. 2001; Riess et al. 2008). De fato, a região promotora do *ATXN3* contém características de um gene *housekeeping*, com um padrão de expressão global tanto em células neuronais, como em células não-neuronais. A expressão de ataxina-3 é maior mas não é restrita às áreas do sistema nervoso central mais caracteristicamente afetadas pela DMJ/SCA3 (Riess et al. 2008). Finalmente, ainda que existam algumas isoformas de ataxina-3 de significado incerto e que são detectadas exclusivamente em pacientes com DMJ/SCA3 (Bettencourt et al. 2010), não há diferenças nos níveis de expressão de ataxina-3 entre pacientes e indivíduos saudáveis.

O trato de glutamina expandido na ataxina-3 mutante parece induzir o estresse celular devido à sua tendência de formar agregados proteicos (Dantuma & Herzog, 2020). De fato, há bastante tempo se especula que proteínas com tratos poliQ expandidos possam afetar os proteassomos devido à toxicidade das espécies amiloidogênicas, ou ao sequestro de subunidades proteassomais em agregados (Schmidt et al. 2002; Dantuma & Bott. 2014). Esse mecanismo caracteriza, como já dissemos, um ganho de função tóxico, levando à disfunção e morte neuronal (Matilla-Dueñas et al. 2014).

De fato, a morte de células neuronais do cerebelo, núcleos do tronco cerebral e gânglios da base, entre outras regiões neuronais, tem sido documentada em todos os estudos post-mortem. Alterações neuroanatômicas características como atrofia do tronco cerebral, cerebelo, núcleo rubro, núcleo subtalâmico, globo pálido e alguns núcleos talâmicos são vistas nos exames de imagem (Durr et al.1996; Rüb et al. 2013; McLoughlin et al. 2019).

As principais manifestações clínicas causadas pela neurodegeneração dessas regiões encefálicas são a ataxia de marcha, ataxia de membros, disartria, espacidade, oftalmoparesia, disartria e marcha espástica. (Saute & Jardim, 2015). À medida que a neurodegeneração progride, os sintomas vão se tornando mais intensos (Schulz et al. 2010). Uma das escalas clínicas mais utilizadas para mensurar a progressão da SCA3/MJD é a Scale for Assessment and Rating of Ataxia (SARA) que avalia a marcha, o equilíbrio, a fala e a coordenação dos membros, através de tarefas que mensuram a coordenação do individuo, gerando um score que varia de 0 (sem ataxia) a 40 (comprometimento extremo) (Schmitz-Hübsch et al. 2006). A progressão anual da SARA foi de 1.41 (0.97 - 1.84) pontos na SCA3/MJD. (Diallo et al. 2021).

Além de ser uma doença debilitante, os portadores da SCA3/DMJ apresentam uma redução na expectativa de vida. Nosso grupo estimou que a sobrevida média foi de 63,96 anos [intervalo de confiança de 95% (IC), 62,09-65,83] para o grupo afetado e 78,61 anos (IC 95%, 74,75-82,47) para o grupo não afetado ($p < 0,001$) (Kieling et al. 2007).

1.1.1.3 O gene *ATXN3*

1.1.1.3.1 Aspectos gerais

O *ATXN3* apresenta cerca de 48,2 kb e se divide em 13 éxons, e a sequência repetitiva CAG causadora da doença está localizada do exon 10 (Bettencourt et al. 2010). A menor CAGexp descrita como associada a sintomas neurológicos continha 51 repetições (Gu et al. 2004). Como já dissemos, os portadores de CAGexp não apresentaram diferença em relação aos não portadores no nível de expressão de mRNA do ATXN3 (Nishiyama et al. 1996) ou da proteína ataxina-3 (Paulson et al. 1997a; Tait et al. 1998).

Como também já dissemos, a ataxina-3 é sintetizada em todos tecidos estudados, tanto no período embrionário como na vida adulta (Ichikawa et al. 2001; Riess et al. 2008). A proteína selvagem apresenta funções de enzima de deubiquitinação (DUB) atuando na via de ubiquitina proteossomo. Apesar de interessante, a função da ataxina-3 não está necessariamente envolvida na patogênese da SCA3/MJD. Por exemplo, o CAGexp não abole a função DUB da ataxina-3. Ao contrário, estudos *in vitro* indicam que a ataxina-3 mutante possui uma atividade DUB maior do que a versão selvagem da proteína (Burnett et al. 2003; Todi et al. 2010). Tampouco camundongos nocaute para os dois alelos de ATXN3 apresentaram efeitos adversos no desenvolvimento embrionário ou diferenças morfológicas marcantes em relação a camundongos selvagens (Schmitt et al. 2007).

1.1.1.3.2 O polimorfismo no tamanho da CAGexp e sua relação com a idade de início

No *ATXN3* não expandido, a repetição CAG apresentou uma variância de 31.80 e heterozigosidade de 0.851, que representam os valores mais altos dentre as doenças de poliQs, já o tamanho médio das repetições foi de 22.72 CAGs (Andrés et al. 2003). Em um estudo que analisou 27.090 alelos de uma coorte europeia, o tamanho das repetições no alelo normal variou entre 11 e 44 repetições CAG. Além disso, ficou evidente o GAP existente entre alelos normais e expandidos (Gardiner et al. 2019).

O tamanho da CAGexp é a principal fonte moduladora da idade de início dos sintomas, e as CAGexp muito grandes podem fazer com que os sintomas se iniciem na infância. Em nossa coorte, 8 casos (2.2 %) se iniciaram antes dos 12 anos de idade. Enquanto a média (variação) das CAGexp vistas nos sujeitos com início após os 12 anos foi de 75 (65-87) nos oito sujeitos com início na infância a mesma foi de 84 (80-91) repetições (Donis et al. 2016).

Uma metanálise realizada pelo nosso grupo analisou a correlação do tamanho das repetições CAG e a idade de início de 2099 indivíduos SCA3/MJD de 12 coortes diferentes. Segundo a mesma, o tamanho da CAGexp explicaria 55,2%

(95% CI 50,8 to 59,0; p<0,001) da variação da idade de início (de Mattos et al. 2019).

A transmissão do CAGexp é instável, o que aumenta o polimorfismo dentro da faixa de repetições expandidas. Apesar dessa instabilidade poder gerar contrações ou expansões, as expansões são predominantes (Souza et al. 2016). O aumento médio das CAGexp a cada geração foi de 2,22 (Maruyama et al. 1995), 1,66 (Souza et al. 2016), 0,86 (Martins et al. 2008) e 0,59 (Cancel et al. 1995), em distintos estudos.

Essa tendência de aumento das repetições CAG a cada geração, combinado com a idade de início ter correlação inversa ao tamanho das repetições CAG, faz com que a idade de início dos sintomas tenha a tendência de começar mais precocemente a cada geração, fenômeno conhecido como antecipação. De fato, a média de idade de início dos sintomas apresentou uma redução a cada geração de 7,32 (Souza et al. 2016), 7,8 (Song et al. 1997) e 9,2 anos (Takayama et al. 1995).

1.1.1.3.3 Fatores que podem influenciar a variação da frequência alélica do *ATXN3*

O aumento do fitness dos portadores da SCA3/MJD é um fator que pode influenciar a variação da frequência alélica no *ATXN3.* Um estudo realizado pelo nosso grupo comparou o número de filhos de mulheres portadoras e não portadoras da mesma família e com mais de 45 anos. Como as portadoras apresentavam uma média de 3.89 filhos, enquanto suas irmãs não portadoras tinham em média 2,68 filhos (p=0.0037, test T), o fitness das portadoras SCA3/MJD foi estimado ser de 1,45 (Prestes et al 2008). Esse fenômeno contribuiria para aumentar a frequência alélica do *ATXN3* ou reduzir/retardar a chance dele ser eliminado.

A distorção na segregação dos alelos também é um fator que influencia a frequência alélica na população. No *ATXN3,* um estudo documentou haver

distorção da segregação a favor do alelo expandido em espermatozoides de seis indivíduos com SCA3/DMJ - 629 espermatozoides continham o alelo expandido e 407 tinham o alelo sem expansão (χ^2 = 47,57, p<0,0001) (Takiyama et al 1997). A distorção a favor do alelo com expansões também foi observada nos filhos de portadores da SCA3/DMJ, na nossa coorte do RS. Em 72 transmissões alélicas de irmandades completamente genotipadas, 48 (67%) indivíduos herdaram o alelo expandido, enquanto 24 (33%) herdaram o alelo sem expansão (X^2=4.14, p<0.031) (Souza et al. 2016).

A distorção na segregação também foi observada entre alelos maiores e menores de alelos não expandidos. Um estudo com famílias sem SCA3 observou que 185 pessoas herdaram o *ATXN3* não expandido de menor tamanho enquanto 140 herdaram o ATXN3 não expandido de maior tamanho (X^2=6,231, p=0,013) (Bettencourt et al. 2008).

1.1.1.3.4 Haplótipos ancestrais

Estudos de haplótipos têm inferindo as origens e datações do *ATXN3* expandido. Os primeiros estudos apontaram para dois haplótipos principais, o A-C-A e o G-G-C (Gaspar et al. 2001). Estudos subsequentes ampliaram os marcadores moleculares, combinando SNPs e STR e informaram que o haplótipo A-C-A teria uma origem asiática datada em torno de 5774 ± 1116 anos. Já o haplótipo G-G-C teria uma origem europeia com datação de 1416 ± 434 anos (Martins et al. 2007). Em 2019 um estudo realizado em uma população chinesa, indicou que na verdade a linhagem A-C-A expandida era mais complexa e composta de dois ramos sucessivos, datados de 16,335 ± 1,966 e 11,837 ± 1,871 anos. Assim, de uma forma geral, os estudos de haplótipos associados a SCA3/MJD têm apontado para poucas e antigas origens.

1.1.2 Ataxia espinocerebelar tipo 2

1.1.2.1 Descrição e epidemiologia

Embora o assunto seja controverso, costuma-se dizer que as primeiras descrições da SCA2 ocorreram em 1971 na Índia. Wadia e Swami relataram a associação de uma ataxia dominante com sacadas oculares lentas - achado hoje reconhecido como característico - em 37 pacientes de 12 famílias. Na época, o quadro ficou conhecido como ataxia de tipo Wadia–Swami (Wadia & Swami. 1971). Em 1990 foi descrito um outro grupo, agora de 500 indivíduos provavelmente descendentes do mesmo antecessor espanhol, vivendo na província de Holguín, em Cuba, e que apresentavam características clínicas similares às da ataxia de Wadia–Swami (Orozco et al. 1990). No ano de 1995, o locus causador da SCA2 foi identificado no cromossomo 12q23-q24.1 (Hernandez et al. 1995). Em 1996, foi descoberta que a sequência repetitiva CAG neste locus estava associada à SCA2. O tamanho das CAGexp variaram de 35 a 59 repetições em dois estudos praticamente simultâneos (Sanpei et al.1996; Pulst et al.1996).

A região de maior prevalência é a província de Holguín, no extremo oriental de Cuba. Em Holguin, a prevalência da doença atinge 47,9:100.000 habitantes, com uma frequência estimada do *ATXN2* expandido na região de 188,6:100.000. Essa alta prevalência também foi atribuída a um efeito fundador da doença na população (Rodríguez-Labrada et al. 2020).

Outros países onde a frequência relativa da SCA2 em relação às demais ataxias dominantes é elevada são o México (Alonso et al. 2007), a África do Sul (Bryer et al. 2003), a Índia (Faruq et al. 2009), a Itália (Brusco et al. 2007) e a Venezuela (Paradisi et al. 2015). No Brasil, a SCA2 foi a segunda ataxia espinocerebelar mais frequentemente diagnosticada. Mesmo assim, representou apenas 49 dos 518 atáxicos, ou 9,4 % dos casos (de Castilhos et al. 2014).

1.1.2.2 Da neurotoxicidade as Manifestações Clínicas:

A ataxina-2 é uma proteína de 145 kDa altamente conservada entre as espécies. Fortes evidências sugerem que a ataxina-2 participa do metabolismo

citoplasmático do RNA, através da participação na montagem de grânulos de estresse, e que ela teria um papel no controle da tradução em geral (Ostrowski et al. 2017). Como na ataxina-3, o mecanismo exato pelo qual a expansão poliQ na ataxina-2 causa neurodegeneração permanece pouco compreendido. Mas os perfis de expressão alterados vistos nos camundongos transgênicos com um *ATXN2* expandido são completamente diversos (ausentes) dos encontrados em camundongos com *ATXN2* normal, sugerindo novamente que o mecanismo que determina a doença seja o de um ganho de função tóxica (Pflieger et al. 2017).

Qualquer que seja o modo de ação, a presença da proteína expandida resulta em disfunção e morte neuronal em algum momento na vida adulta, especialmente no cerebelo, tronco encefálico, medula espinhal e córtex cerebral, que conduzem os mecanismos da síndrome cerebelar progressiva, incluindo a -características cerebelares, que caracterizam clinicamente a doença (Auburger et al. 2012; Velázquez-Pérez et al. 2017). Estudos apontam a formação de agregados polyQ em regiões do tronco cerebral de indivíduos com SCA2, além disso os agregados citoplasmáticos se correlacionaram com a gravidade da doença (Seidel et al. 2017).

Os principais achados clínicos e sua frequência encontrados nos indivíduos com SCA2 foram: síndrome cerebelar progressiva, marcha atáxica, disartria cerebelar, dismetria, disdiadococinesia associada a movimentos sacádicos lentos, além de neuropatias periféricas fasciculações, dores musculares contraturas, distúrbios do sono e disfagia (Magaña et al. 2012). Na escala SARA a média da progressão da doença foi de 1,40 (1,19 -- 1,61) pontos por ano (Diallo et al. 2021).

1.1.2.3 O gene *ATXN2*

1.1.2.3.1 Aspectos gerais:

O *ATXN2* é formado por 25 exons (4,500 bp) e sintetiza a proteína ataxina-2, com 140kDa (Sahba et al. 1998). O gene contém uma sequência

repetitiva CAG que pode ser pura, ou interrompida por trincas CAA, e que é considerada normal até 32 ou 33 repetições CAG. Este limite tem sido determinado pelo estudo de diversas populações. Nosso grupo e outros já descreveram alelos de 33 repetições segregando em pessoas assintomáticas (Cancel et al. 1997; Socal et al. 2009). Finalmente, não é claro se a presença de uma interrupção CAA confere ou retira patogenicidade a esses alelos incertos de 32-33 CAGn. Alelos francamente patológicos seriam aqueles com 34 ou mais repetições (de Castilhos et al. 2014). Alelos normais, mas com repetições maiores que 27 estão associados a um aumento no risco da Esclerose Lateral Amiotrófica (Lee et al. 2011; Elden et al. 2010). Na faixa da normalidade - a que vai de 13 a 33 repetições CAG -, o alelo mais comum é o com 22 repetições (CAG_{22}), presente em 90% dos indivíduos normais (Mao et al. 2002). Assim, ao contrário do *ATXN3* não expandido, o *ATXN2* não expandido apresenta a menor variância e heterozigosidade nas repetições CAG, respectivamente de 1,21 e 0,253. Essa redução de polimorfismo nas repetições CAG do alelo normal do *ATXN2* tem sido relacionada à presença de interrupções CAA entre as repetições CAG do CAG_{22}, num padrão (CAG)8-CAA-(CAG)4-CAA-(CAG)8 (Choudhry et al. 2001).

1.1.2.3.2 O polimorfismo nas repetições CAG sua relação com a idade de início

A idade de início dos sintomas apresenta uma correlação inversa com o tamanho do CAGexp, com um r^2 estimado entre 0,46 (Chattopadhyay et al. 2003) e 0,69 (Figueroa et al. 2017). Um conjunto de casos com idade de início dos sintomas com menos de 1 ano foram descritos com repetições CAG podendo chegar a 500 repetições CAG (Singh et al. 2014).

Como era esperado, as CAGexp do *ATXN2* são instáveis durante as transmissões meióticas, as expansões foram os fenômenos mais descritos, ocorrendo em 89,03% dos casos, enquanto as contratações ocorreram em 10,97% (Velázquez Pérez et al. 2009). A média de aumento das CAGexp variou em distintos estudos entre 2,08 (Almaguer-Mederos et al. 2018), 2,2 (Cancel et al. 1997) e 5,73 (Riess et al. 1997). Assim como na SCA3/MJD, a antecipação é comum, com uma

média de redução na idade de início na geração subsequente entre 11,41 (Sena et al. 2019); 18,8 (Fila et al. 1999) e 21,28 (Giuffrida et al. 1999).

1.1.2.3.3 Fatores que podem influenciar a variação da frequência alélica do *ATXN2*

Da mesma forma que na SCA3/MJD, a SCA2 apresentou um fitness diferencial em seus portadores em relação aos não portadores. No meu mestrado, comparei a fertilidade entre afetados e não afetados com mais de 54 anos e pertencentes às mesmas famílias. Os indivíduos afetados e os não afetados apresentaram medianas de filhos respectivamente de 3 e de 2 filhos (p=0,025, Mann-Whitney U test), os portadores da SCA2 apresentando portanto um fitness de 1,5 (Sena et al. 2019).

No mesmo estudo, se demonstrou haver distorção na segregação contra o alelo *ATXN2* expandido, na coorte averiguada. Acompanhei 139 transmissões de mães e pais com SCA2: 56 (40,4%) de suas filhas e filhos herdaram o alelo expandido e 83 (59,6%) herdaram o alelo sem expansão (qui-quadrado=5,245, p=0,022) (Sena et al. 2019).

1.1.2.3.4 Haplótipos ancestrais

Uma série de estudos foram realizados usando ora STRs ora apenas SNPs e em algumas situações combinando ambos. No entanto, nenhum desses estudos de haplótipos na SCA2 em escala de colaboração internacional apontaram para o número de origens e sua datação.

Estudos que apontam para uma ancestralidade em comum se basearam em STRs para fazer essa inferência. Com base no D12S1672 e o D12S1333 sugeriu-se que numa coorte japonesa existem duas origens SCA2 (Mizushima et al 1999). Já outro estudo usando apenas o D12S1672 sugeriu que famílias SCA2 Cubanas, Inglesas, Indianas e Italianas ou apresentam uma origem

comum, ou que exista uma região do gene predisposta à expansão (Pang J et al 1999).

Outros estudos apresentam maior polimorfismo entre os STRs, o pode sugerir a existência de múltiplas origens para SCA2 (Didierjean et al,). Um fator que converge com essa sugestão é a descrição de mutação *de novo* em um indivíduo SCA2 com 22/35 repetições CAG no qual seus pais assintomáticos apresentavam 22/22 e 22/32 repetições CAG (Futuyma et al 1998) e da descrição de casos esporádicos (Alonso et al 2007).

Outros três estudos com famílias SCA2 da Índia, Brasil, Portugal e Itália analisaram os rs695871 e rs695872, onde o haplótipo C-C foi encontrado em 100% das famílias SCA2 (Choudhry et al. 2001; Ramos et al. 2010; Sonakar et al. 2021). Para esses SNPs, o haplótipo GT não foi encontrado em nenhum caso de SCA2 e esteve associado a repetições CAG normais menos polimórficas; já o haplótipo CC apresentou uma maior variação no número de repetições CAG não expandidas. No entanto, apesar de estar associado a repetições mais estáveis, alelos normais grandes, com 30-32 CAG, também foram encontradas no haplótipo GT (Choudhry et al. 2001; Ramos et al. 2010).

1.2 Evolução e doenças poliQs

Uma definição mais clássica de evolução é a mudança na composição genética de populações. Nesse sentido, para muitos biólogos evolucionistas, os processos que modificam a composição genética de uma população seriam a seleção natural, deriva, fluxo gênico e mutação. No entanto, a teoria evolutiva vem se transformando ao longo do tempo e incorporando múltiplos elementos a tornado uma teoria mais abrangente (Araújo. 2001). No entanto, nos atentaremos a seguir nas forças seletivas que parecem estar associadas com a dinâmica dos alelos CAGexp.

1.2.1 Antecipação, uma força seletiva negativa para manutenção dos alelos expandidos.

Assim como em outras doenças PoliQ, a antecipação é um fenômeno encontrado na SCA2 e SCA3. A tendência dos sintomas se manifestarem mais cedo a cada geração gera um desdobramento prático que impacta negativamente na manutenção dos alelos expandidos do *ATXN2* e *ATXN3.* Isso porque, quanto mais precoce os sintomas se iniciam, menor será a janela reprodutiva de seus portadores, reduzindo a chance da transmissão dos alelos com expansão.

Esse inferência pode ser provada empiricamente, pois apesar dos portadores da SCA3 e SCA2 terem uma fertilidade maior que os não portadores, os indivíduos que manifestam os sintomas mais precocemente tendem a terem menos filhos do que os que manifestam os sintomas mais tardiamente (Prestes et al. 2008, Sena et al. 2019).

Assim podemos inferir que existe uma tendência intrínseca para eliminação dos alelos expandidos do *ATXN3* e *ATXN2.* Isso porque, alelos que reduzem a aptidão de um organismo tendem a ser eliminados por seleção negativa ou também chamada de seleção purificadora (Loewe L. 2008). Dessa forma, os alelos expandidos do *ATXN3* e *ATXN2* são exemplos de *Darwinian puzzle*, nos fazendo refletir de por que esses alelos permanecem na população? Assim forças evolutivas compensadoras possam vir a explicar esse processo.

1.2.2 Equilíbrio seleção mutação

Um dos fenômenos que podem levar alelos deletérios a ter sua frequência estável é o processo conhecido como equilíbrio seleção mutação, no qual novas mutações repõem os alelos que são eliminados pela seleção natural. Para alelos dominantes esse equilíbrio se estabelece a partir da equação abaixo(Hartl & Clark 2007):

$$\widehat{q} \approx \sqrt{\frac{\mu}{hs}}$$

Onde:
$\widehat{q}$ = frequência alélica em equilíbrio
μ = taxa de mutação
s = coeficiente de seleção
h = coeficiente de dominância

Em algumas doenças a sua manutenção pode ser explicada por uma taxa mutacional similar/próxima a taxa de seleção, como por exemplo, a Atrofia muscular espinhal (Wirth et al. 1997) e a Distrofia Muscular de Duchenne (Lee et al. 2013).

1.2.3 Pleiotropia Antagonica

A pleiotropia antagônica é o processo no qual um gene controla mais de uma característica, sendo que pelo menos uma exerce um aumento na aptidão do organismo e outra uma piora no valor adaptativo. Esse conceito foi proposto pela primeira vez pelo biólogo evolucionista George C. Williams em 1957 (Williams. 1957).

Inicialmente a Pleiotropia Antagônica teve como objetivo explicar a senescência, propondo que os alelos de genes que levariam a maiores chances de sobrevivência e/ou reprodução nas etapas iniciais das vidas estivessem associadas a características deletérias em um segundo momento, levando à perda de homeostase e autodestruição (Rose. 1991; Mitteldorf. 2019). Essa mesma hipótese foi aventada para explicar a ocorrência frequente de doenças neurodegenerativas na espécie humana e, entre elas, as doenças de poliQs.

Uma associação favorável entre eventos biológicos e presença de CAGexp tem sido repetidas vezes descrita: a redução na frequência de câncer entre seus portadores (Sørensen et al. 1999; Ji et al 2012; Ji et al. 2012; Souza et al. 2017; McNulty et al. 2018). Embora essa vantagem aparente conferida pela presença de

uma CAGexp não nos interesse diretamente, por provavelmente não ter repercussão na sobrevida de descendentes, ela é um argumento interessante a sustentar o pleiotropismo antagonístico das CAGexp.

O que é de fato de interesse para uma abordagem evolutiva é que a SCA3/MJD e a SCA2, entre outras doenças poliQs, parecem se associar a fitness elevados, como vimos nas seções 1.1.1.3.3 e 1.1.2.3.3, acima. Na seção 1.1.1.3.3, vimos também que o alelo *ATXN3* expandido é segregado preferencialmente. Fecundidade elevada e segregação preferencial seriam duas características biológicas favoráveis, manifestas em períodos vitais e em funções vitais não neurológicas, e antagonistas às características desfavoráveis dos CAGexp nos neurônios.

Dessa forma, para o *ATXN2* e *ATXN3* expandidos, além dos aspectos que prejudicam a aptidão dos portadores como os sintomas da doença e a antecipação, tempos também elementos associados a melhora da aptidão como o aumento da fertilidade. Além disso, a distorção na segregação contribui de forma distinta na variação da frequência alélica para o *ATXN3* e *ATXN2* expandido. Por esses aspectos, nos encontramos diante de dois exemplos de Pleiotropia antagônica.

1.2.4 Abordagens evolutivas nas poliQs

O fato da antecipação ser um fenômeno presente nas doenças de poliQs impôs a reflexão sobre as razões da sua permanência nas populações, e fez com que abordagens evolutivas tenham sido usadas para entender a sua manutenção. Para SCA1, uma abordagem evolutiva foi usada para compreender o aumento da prevalência na República de Sakha, na Rússia (Platonov et al. 2016). Na doença de Huntington, além dos estudos sobre a fertilidade dos portadores (Shokeir at al. 1975, Walker et al. 1983, Morrison et al. 1995, Pridmore & Adams. 1991, Frontali et al. 1996), uma perspectiva mais abrangente foi utilizada no artigo "*A Darwinian approach to Huntington's disease: subtle health benefits of a neurological disorder" (*Eskenazi et al. 2007).

Abordagens evolutivas também têm sido utilizadas para compreensão da manutenção e diversidade fenotípica associadas a SCA3 (Lima et al. 2001), e que interpretam a manutenção do *ATXN3* expandido em decorrência de seu caráter pleiotrópico antagônico, devido às múltiplas forças seletivas associadas ele (Saute & Jardim. 2015).

Meu mestrado também tentou auxiliar no entendimento sobre a ocorrência de uma doença de poliQs, a despeito das suas intensas antecipações, estudando a fertilidade e a distorção na segregação da SCA2 (Sena et al. 2019).

Apesar dessas abordagens, nenhuma delas combinou diferentes evidências e mensurou o seu peso das distintas forças seletivas que pudessem apontar uma dinâmica da frequência alélica dos alelos expandidos em uma perspectiva mais geral.

2 Objetivo

2.1 Objetivo Geral

Estudar como se dá a dinâmica da transmissão da SCA2 e da SCA3/MJD, tal que, apesar de deletérias aos afetados e às suas proles, essas doenças continuem aparentemente se mantendo nas populações.

2.2 Objetivos específicos:

1) Revisar e metanalizar as forças seletivas associadas ao *ATXN3* expandido.
2) Revisar e metanalizar as forças seletivas associadas ao *ATXN2* expandido.
3) Construir um modelo matemático que abranja as forças seletivas associadas ao *ATXN2exp* e *ATXN3exp.*
4) Calcular a tendência de variação da frequência alélica via simulações computacionais, combinado o modelo matemático e as forças seletivas associadas ao *ATXN2exp* e *ATXN3exp.*
5) Estimar o número de gerações que o *ATXN2exp* e *ATXN3exp* permaneceram na população a partir de um evento fundador.
6) Relacionar os resultados das simulações da variação da frequência alélica e gerações de permanência do *ATXN2exp* e *ATXN3exp* com os dados empíricos.
7) Descrever os haplótipos ancestrais associados à SCA2 em famílias brasileiras, uruguaias e peruanas, e averiguar se os resultados virão ao encontro das dinâmicas propostas pelo modelo matemático.

CAPÍTULO II. FORÇAS SELETIVAS ASSOCIADAS AOS ALELOS EXPANDIDOS DO ATXN3

Manuscrito 1: ***Selective forces acting on spinocerebellar ataxia type 3/Machado-Joseph disease recurrency: A systematic review and meta-analysis*****.**

Artigo publicado no periódico

Clinical Genetics

(DOI: 10.1111/cge.13888)

Received: 28 July 2020 | Revised: 25 October 2020 | Accepted: 17 November 2020

DOI: 10.1111/cge.13888

REVIEW

CLINICAL GENETICS WILEY

Selective forces acting on spinocerebellar ataxia type 3/Machado–Joseph disease recurrency: A systematic review and meta-analysis

Lucas Schenatto Sena[1,2] | Jordânia dos Santos Pinheiro[2,3] | Maria Luiza Saraiva-Pereira[1,2,4,5,6] | Laura Bannach Jardim[1,2,4,6,7]

[1]Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[2]Centros de Pesquisa Clínica e Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

[3]Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[4]Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

[5]Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[6]Instituto de Genética Médica Populacional, Porto Alegre, Brazil

[7]Departamento de Medicina Interna, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Correspondence
Laura Bannach Jardim, Centros de Pesquisa Clínica e Experimental, Hospital de Clínicas de Porto Alegre, Rua Ramiro Barcelos 2350, 90035-003 Porto Alegre, Brazil.
Email: ljardim@hcpa.edu.br

Funding information
Fundo de Incentivo à Pesquisa do Hospital de Clínicas de Porto Alegre, Grant/Award Number: 2019-0169; Conselho Nacional de Desenvolvimento Científico e Tecnológico; Fundação de Amparo à Pesquisa do Rio Grande do Sul

Abstract

Spinocerebellar ataxia type 3/Machado–Joseph disease (SCA3/MJD) is a dominant neurodegenerative disease caused by the expansion of a CAG repeat tract in *ATXN3*. Anticipation and worsening of clinical picture in subsequent generations were repeatedly reported, but there is no indication that SCA3/MJD frequency is changing. Thus, we performed a systematic review and meta-analysis on phenomena with potential effect on SCA3/MJD recurrency in populations: instability of CAG repeat transmissions, anticipation, fitness, and segregation of alleles. Transmission of the mutant allele was associated with an increase of 1.23 CAG repeats in the next generation, and the average change in age at onset showed an anticipation of 7.75 years per generation; but biased recruitments cannot be ruled out. Affected SCA3/MJD individuals had 45% more children than related controls. Transmissions from SCA3/MJD carriers showed that the expanded allele was segregated in 64% of their children. In contrast, transmissions from normal subjects showed that the minor allele was segregated in 54%. The present meta-analysis concluded that there is a segregation distortion favoring the expanded allele, among children of carriers. Therefore, further studies on transmissions and anticipation phenomena as well as more observations about fertility are required to clarify these selective forces over SCA3/MJD.

KEYWORDS

allele segregation, ancestral haplotypes, anticipation, fertility, Machado–Joseph disease, meta-analysis, spinocerebellar ataxia type 3, unstable transmissions

1 | INTRODUCTION

Spinocerebellar ataxia type 3, also known as Machado–Joseph disease (SCA3/MJD), is an autosomal dominant neurodegenerative disease caused by the expansion of a CAG repeat tract in the *ATXN3*.[1,2] Normal alleles range from 11 to 44 repeats. Pathological alleles usually have 60 or more CAG repeats, although the shortest CAG repeat associated with a neurological picture in a SCA3/MJD family was 51 repeats long.[3,4] The expanded repeat (CAGexp) is in the open reading frame, and produces an augmented polyglutamine stretch into the translated protein, ataxin-3. PolyQ-expanded ataxin-3 is prone to aggregation. The presence of this polyQ-expanded ataxin-3, or of peptides produced after its proteolysis, is associated with neuronal toxicity and degeneration.[5]

The most important and well-known consequence of CAGexp at *ATXN3* is the late onset of predominantly neurological disabilities. Affected subjects and their relatives note the first symptoms, usually gait ataxia, at 36–40 years-old, on average. However, disease can

actually start from 4 to over 72 years of age.[6] The main reason for this wide variation in age at onset (AO) relies on length of CAG expansion, which explains on average 55.2% of AO variation.[7] Only small size effects over the remaining variability have been postulated to be associated with other genes, such as *ATXN2*, *FAN1*, *IL6*, *APOE*, *TRIM29*, and *RAG*.[7-9] Besides gait problems, other symptoms appear progressively, including for instance speech and swallowing difficulties, diplopia, sensory losses, pyramidal and extrapyramidal findings. The CAGexp length was associated with velocity of disease progression. Survival following onset of symptoms is reduced to an average of 23 years.[6]

Anticipation of AO and worsening of clinical picture in subsequent generations have been frequently documented and were in part due to the instability of the CAGexp length during meiosis, more frequently detected as further expansions.[4] Although never directly studied, anticipation might be associated with shortage of the reproductive period, and to lower reproduction rates of SCA3/MJD carriers.

Despite this tendency to produce worse functional disabilities as generations progresses, SCA3/MJD has the peculiarity of being among the most common forms of hereditary ataxia worldwide.[10] In contrast to the relatively large number of symptomatic carriers found in many different populations, haplotype studies detected few ancestral haplotypes. A recent study dated the most frequent of them, known as the ACA haplotype, to be over 16 000-years-old.[11] To complete a real Darwinian puzzle, *de novo* mutations have never been reported in SCA3/MJD.

How to explain that severe neurological disabilities associated with anticipation can occur in a disease with quite a few and very old ancestral origins? Some authors proposed that the maintenance of disadvantageous alleles in the population could be due to compensatory evolutionary forces.[12,13]

A detailed review may help to clarify whether there is evidence about positive selective forces, and might help to predict epidemiological changes of SCA3/MJD in generations to come. Thus, our purpose was to perform a systematic review and meta-analysis on non-neurological outcomes related to the CAGexp at *ATXN3* with potential effect over SCA3/MJD recurrence in populations as follows: (a) instability of the expanded repeat when crossing meiosis; (b) changes in AO among different generations; (c) differences in reproductive rates between carriers and noncarriers; and (d) meiotic segregation. In order to test if there is subsidiary evidence in line with these previous main outcomes, (e) a systematic review about ancestral haplotypes was also included.

2 | METHODS

A detailed methodology protocol for this study was registered (CRD42020170173) prior to data extraction, and is available at https://www.crd.york.ac.uk/PROSPERO/.

The term "carrier" was used here to indicate subjects, symptomatic or not, with the heterozygous state for SCA3/MJD, that is, with the presence of one *ATXN3* allele with ≥51 CAG repeats in her/his genotype. The term "fitness" was used to mean the reproductive success of a phenotype, more specifically, the ratio between the mean number of children of carriers over the mean number of children of controls. The term "emeritus" was used to indicate that a person probably reached the end of reproductive age, by being older than an age determined by each study.

2.1 | Literature search and data extraction

MEDLINE (PubMed) was searched from January 1995 to December 2019 for reports on four main outcomes as well as to the fifth subsidiary question under study: (a) instability of the expanded repeat when crossing meiosis (contractions and/or expansions); (b) differences in AO between different generations, named as "anticipation"; (c) reproductive rates (or success) of the carriers compared to noncarriers, called "genetic fitness"; (d) meiotic segregation—if Mendelian or distorted; and (e) ancestral haplotypes. Five searches were performed, one for each outcome under study. All five searches started with terms: ("Machado Joseph" OR "SCA3" OR "Spinocerebellar ataxia type 3" OR Machado-Joseph OR "SCA 3" or "MJD/SCA3" OR "MJD" OR "Spinocerebellar ataxia type-3"). Following that, search for (a) continued with "instability" OR "contraction" OR "further expansion" OR "meiosis" OR "transmission" OR "parent–child". Search for (b) continued with "age at onset" OR "AO" OR "age of onset" OR "anticipation". Search for (c) continued with "fitness" OR "number of children" OR "children" OR "reproduct*". Search for (d), with "segregation" OR "meiotic drift". And, finally, search for (e), with "haplotype" or "ancestral origin" or "mutational origins" or "common ancestor".

Peer-reviewed articles and meeting abstracts in English language were included, and references were checked to assure maximal coverage. Two reviewers (L. S. S. and J. S. P.) assessed and extracted data into evidence tables independently. Any disagreement regarding eligibility was discussed with two other reviewers (M. L. S-P. and L. B. J.). Results of any systematic review would be described if found in one search subject.

2.2 | Population, exposure, comparators, outcomes, and inclusion and exclusion criteria

Populations from diverse geographical origins comprised the population under study. Being a carrier of a CAGexp at *ATXN3* and length of CAGexp were the main exposure considered for meta-analysis. The outcomes were (1) instability of the expanded repeat when crossing meiosis (contractions and/or expansions); (2) differences in AO between different generations (anticipation); (3) genetic fitness; (4) meiotic segregation; (5) ancestral haplotypes identified so far.

Inclusion criterium for all searches was the confirmed molecular diagnosis of SCA3/MJD in symptomatic and/or asymptomatic heterozygotes. In addition, a case–control design was among the inclusion

criteria for outcome (3). If multiple publications from the same study group and/or institutions were detected, only the most updated and complete data set was included in order to avoid overrepresentation.

2.3 | Risk of bias assessment and quality control

Risk of bias was assessed according to the search questions. Inclusion of all offspring of a given carrier, and balance between parental sexes were used to assess risk of bias on studies of instability of the expanded repeat when crossing meiosis. Inclusion of all offspring of a given carrier, exclusion of the present generation (to avoid bias toward cases of younger age at onset), and balance between parental sexes were the criteria used to assess bias on studies about differences in AO between different generations. In the search on genetic fitness, risk of bias was assessed by the type of controls, that is, general population versus relatives, determination of the noncarrier status by genotype or emeritus phenotypes, number of generations under study, and inclusion of all family members. The noninclusion of all individuals from a given sibship and the genotype attribution based on phenotype were considered as risks of bias on segregation studies. Finally, if a high heterogeneity of results was detected between studies, a careful reanalysis was performed in order to detect any additional risk of bias in the discrepant study. Heterogeneity of results was considered low if results differed from 0% to 50%, medium, from 50% to 75%, or high, if above 75%. High heterogeneity was considered acceptable in studies related to ancestral haplotypes only.

Quality control for the instability of the expanded repeat was performed by comparing results of recent versus older reports, or by differences in laboratory techniques, when reported. Quality control for ancestral haplotypes included the presence of a control group from the original populations. All potential biases and quality parameters were summarized in the Results section, if appropriate.

2.4 | Analysis and data synthesis

Descriptive measures of central tendency and dispersion were estimated according to the data distribution. A meta-analysis was performed using the R Program version 3.6.2 (2019-12-12), package meta version 4.9–7, when two or more studies on one of the outcomes reached inclusion criteria. Random effects model was chosen in order to avoid effects related to differences between inclusion criteria, sample sizes, and/or variances of the studies selected. Summary statistics from aggregated databases (ADs) were planned to be used for all outcomes under study, except for ancestral haplotypes. A confidence interval of 95% was chosen to attribute significance to results.

3 | RESULTS

A total of 366 papers were obtained after searching databases. Some papers were found in more than one of the five searches: 62 were selected twice and 27 were selected three times (Figure 1). After analysis, eligibility was reached in 13 articles on instability of the expanded repeat, eight on anticipation, one on fertility, nine on segregation distortion, and 13 on ancestral haplotypes. Six papers were selected for two outcomes, and one paper was selected for three outcomes under analysis. Therefore, 32 papers were selected in our systematic review in total. Studies obtained from our five searches and reasons for inclusions or exclusions can be found in Supplemental material Data S1.

3.1 | Instability

Twelve studies described potential instabilities of CAG*exp* length when transmitted from affected parents to their children (Figure 1 (A))[14-26] (Supplemental Material Data S2).

Transmission of the mutant allele, regardless of sex of the affected parent, was associated to an increase of 1.23 CAG repeats in the next generation (Figure 2(A)). The average change of the CAGexp was of 0.79 CAG (Figure 2(B)) and 2.11 CAG repeats in maternal and paternal transmissions, respectively (Figure 2(C)).

None of the studies clearly stated that all offspring of a given carrier were included. None of them stated to have made efforts to recruit balanced samples of maternal or paternal transmissions: the number of maternal transmissions was always bigger than the paternal ones.

3.2 | Anticipation

Seven papers studied differences between AO of children and their affected parents (Figure 1(B))[14,16,18,22,24,25,27] (Supplemental Material Data S3). All of them used the term "age at onset" as the time when the first symptom was noted by the patient and/or his/her relatives, without specifying the symptom (usually gait ataxia but not always). None of them stated that all children of the studied carriers were included; none of them excluded the present generation to avoid bias toward cases of younger age at onset. The number of maternal transmissions was larger than those of the paternal transmissions in the studies (n = 6) that stratified results according to parental sexes.

The average change between generations was toward an anticipation of 7.75 years per generation (Figure 2(D)). The anticipation was 7.13 in case of maternal transmission (Figure 2(E)), and 8.02 years in case of paternal transmission (Figure 2(F)).

3.3 | Fitness

Two studies reported on "fitness" in SCA3/MJD, but only one truly compared the number of children of SCA3/MJD carriers with those of controls, presented a clearcut criteria to determine the noncarrier status (emeritus phenotypes) (Figure 1(C)), and stated that the researchers tried to include all family members.[28] This study used the

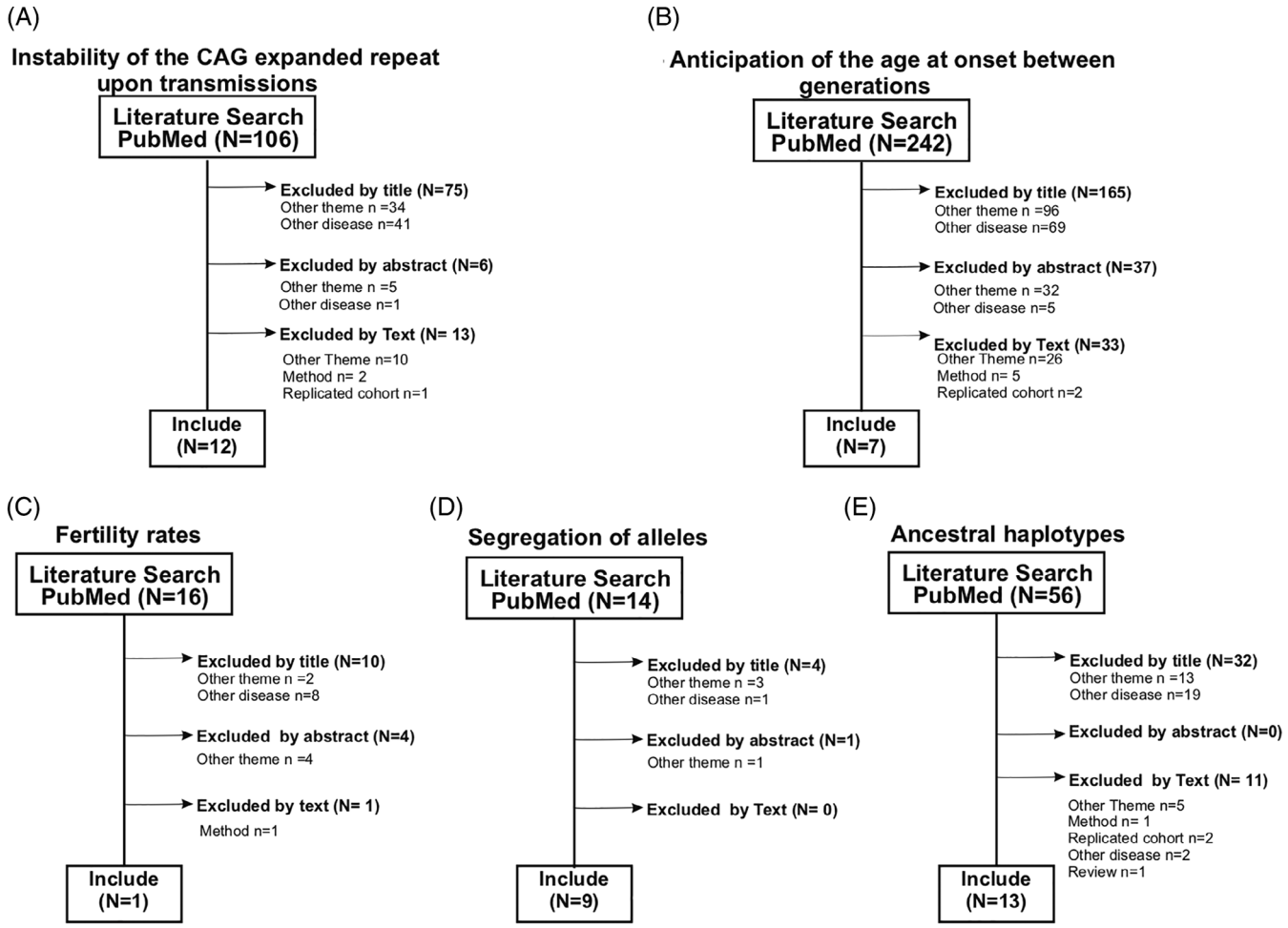

FIGURE 1 PRISMA flowchart of studies selected for the following meta-analyses. (A) Instability of the CAG expanded repeat upon transmissions; (B) Anticipation of the age at onset between generations; (C) Fertility rates; (D) Segregation of alleles; (E) Ancestral haplotypes

census of a general population by the year 2000 as the main control group and, therefore, partially avoided to include the most recent SCA3/MJD diagnoses into analysis. The number of children of 222 SCA3/MJD symptomatic women was reported, using both the general female population and nonsymptomatic female relatives older than 45 years (the cutoff chosen by these authors for the emeritus phenotypes) as controls. Populational data described just the number of children of mothers to avoid overrepresentation; due to that, only SCA3/MJD women were included in the group of cases. The number of children of the general population and of SCA3/MJD subjects were 1.90 and 2.93 ± 2.3, respectively (p = 0.0037). The fitness estimated among emeritus subjects from SCA3/MJD families and from the general population was not so different (3.52 vs. 3.89). When comparisons among emeritus subjects between SCA3/MJD and related controls were done, significant differences reappeared: the mean number of children of unaffected and affected women older than 45 years were 2.68 and 3.89, respectively (p = 0.0037). Therefore, emeritus fitness of the affected SCA3/MJD individuals was 1.45 when compared to their unaffected relatives.

3.4 | Segregation distortion

Literature search retrieved nine articles that met the inclusion criteria proposed (Figure 1(D)). Three distinct approaches to study segregation were found: description of the total number of children of SCA3/MJD carriers that inherited and that did not inherited the mutant allele; comparison between the number of expanded versus nonexpanded alleles in sperm samples collected from SCA3/MJD subjects; and report of the total number of children of normal subjects that inherited the shorter versus the longer normal CAG alleles, aiming to analyze the potential distortion among normal alleles (Supplemental Material Data S4).

Three studies analyzed the segregation distortion by comparing the genotypes of all children of some SCA3/MJD carriers to the expected Mendelian segregation.[14,29,30] A fourth study used a combination of genotypic and phenotypic information to characterize the sibships.[31] A total of 783 allelic transmissions of the *ATXN3* gene from SCA3/MJD carriers were followed. At first analysis, the expanded allele was segregated in 59% (CI 51%–66%) of children, therefore different from the 50% expected from a Mendelian segregation. However, the results from Bettencourt et al. 2008[32] were considerably

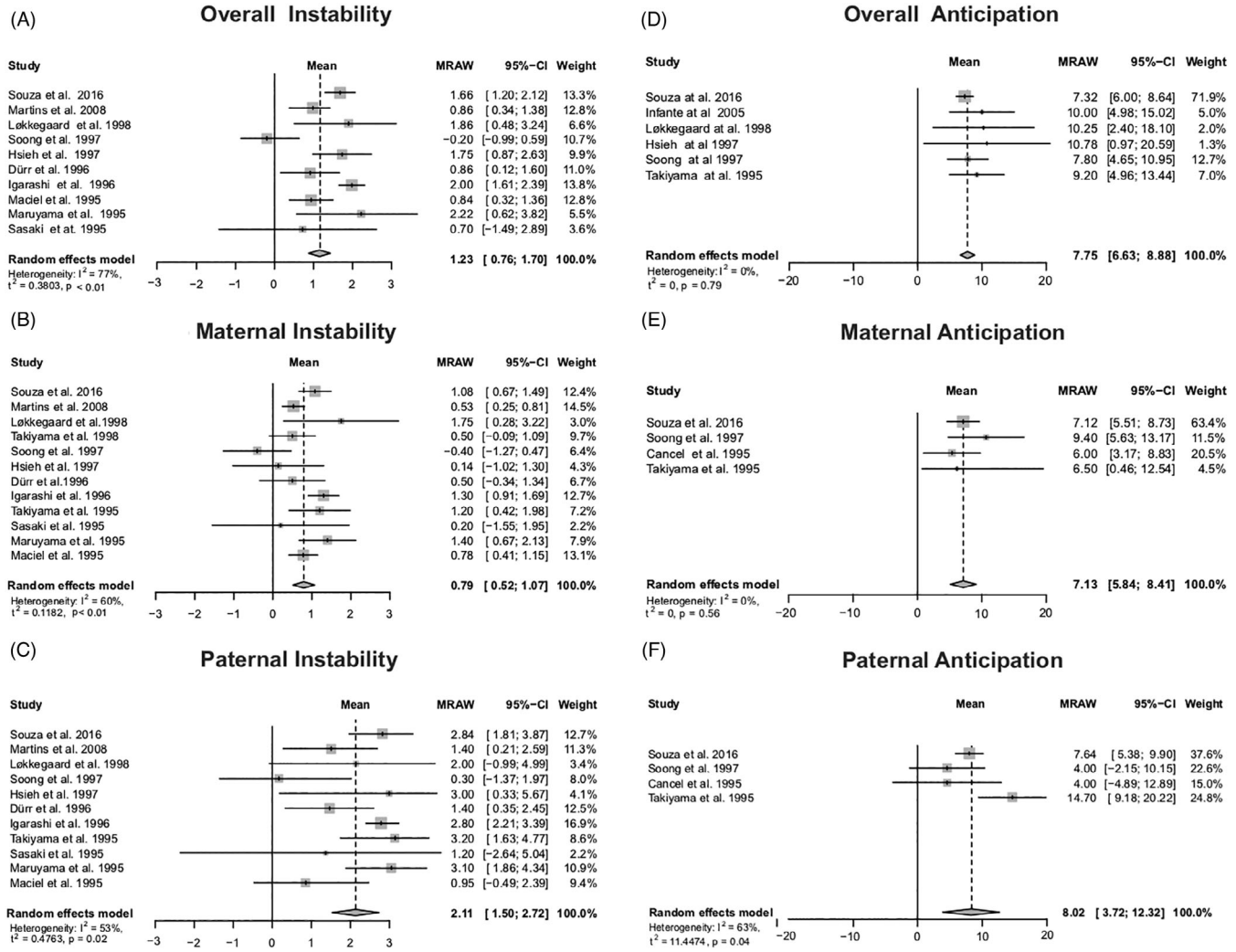

FIGURE 2 Forest plots of instabilities of the CAG expanded repeat number when crossing generations and of anticipation of the age at onset. (A) Overall instability of the CAG expanded repeat; (B) Instabilities observed when women transmitted the expansion; (C) Instabilities observed when men transmitted the expansion; (D) overall anticipation of the age at onset; (E) Anticipation observed when women transmitted the disease; (F) Anticipation observed when men transmitted the disease. Results with 95% confidence interval were calculated using a random-effects meta-analysis. I^2 quantifies heterogeneity; p for heterogeneity. r^2 magnitude of the heterogeneity between studies from random-effect model; RE: random-effect

heterogeneous when compared to ~~the~~ other results (heterogeneity of 78%, Supplemental Material Data S5). After further review, an inconsistency was found between the total number of subjects analyzed and the number of symptomatic and asymptomatic individuals included in the analysis. Since a significant risk of attribution bias was found, this article was excluded. The final meta-analysis then showed that the expanded allele was segregated in 64% (CI 59–68) (Figure 3 (A)). In maternal lineages, the expanded allele was segregated in 68% (CI 60–75) (Figure 3(B)), and in paternal transmissions, it was segregated in 60% (CI 52–68) (Figure 3(C)).

The proportion of spermatozoids carrying an expanded allele in the sperm of SCA3/MJD carriers was Mendelian (50%) in one study and non-Mendelian (60%, $p < 0.05$) in the second study retrieved from the literature.[33,34] Considering these two reports were the only ones related to sperm cells analysis and their results were considerably different, they were excluded from the meta-analysis.

Three papers studied the segregation between shorter and longer nonexpanded CAG alleles of normal individuals.[32,35,36] In the meta-analysis, the segregation of shorter alleles was established to be 54% (CI 51%–57%), a proportion significantly different from the 50% expected from a Mendelian segregation (Figure 3(D)). When the transmissions were stratified according to sex of the transmitting parent, this segregation distortion favoring the shorter allele was maintained in 638 maternal transmissions, but was not detected in 604 paternal transmissions under study (Figure 3(E,F)).

3.5 | Haplotypes

Thirteen studies were published in the literature on this subject (Figure 1(E)). The haplotypes described here covered at least the classical three SNPs, as proposed by the first paper that described SCA3/

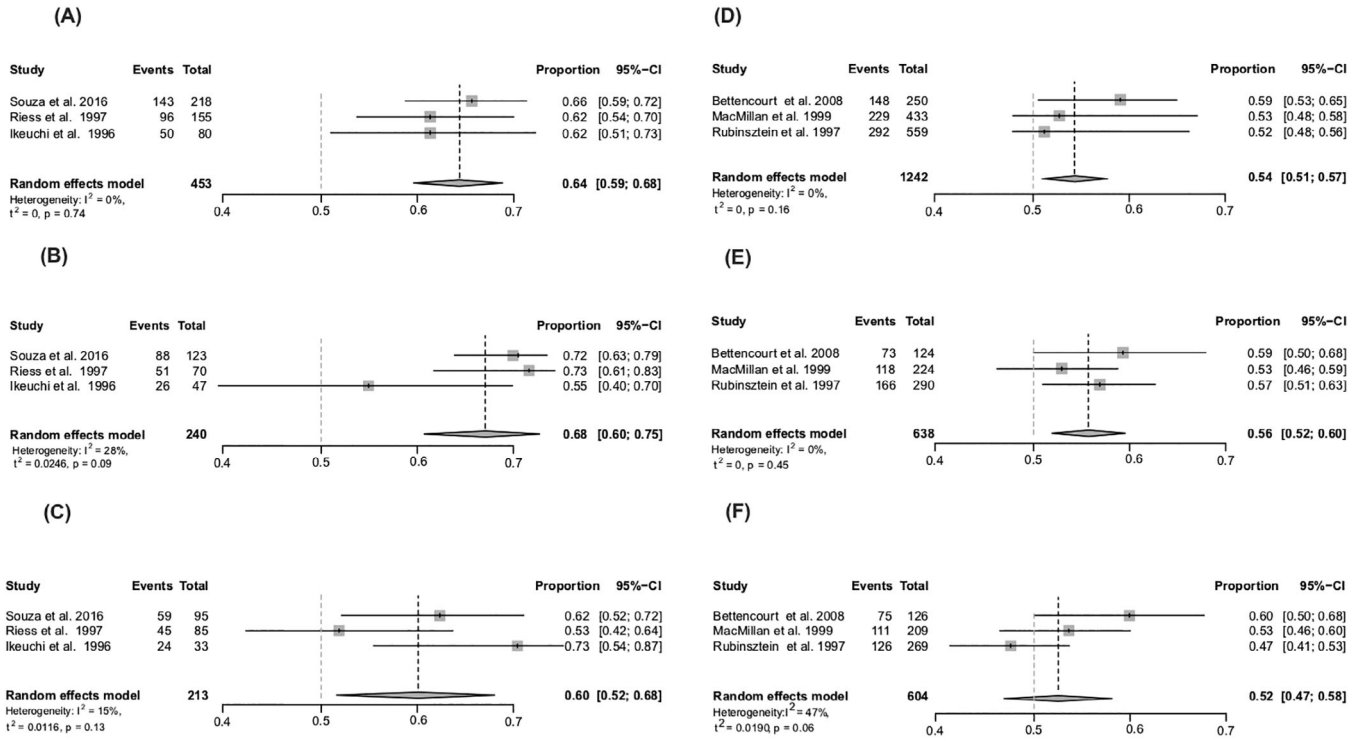

FIGURE 3 Forest plots of allele segregation. (A) Overall segregation between the normal and the expanded allele, when one parent was a spinocerebellar ataxia type 3/Machado–Joseph disease (SCA3/MJD) carrier; (B) Segregation between the normal and the expanded allele, when the mother was a SCA3/MJD carrier; (C) Segregation between the normal and the expanded allele, when the father was a SCA3/MJD carrier; (D) Overall segregation between the normal short and the normal large allele, when the parent was a normal control; (E) Segregation between the normal short and the normal large allele, when the parent was a female control; (F) Segregation between the normal short and the normal large allele, when the parent was a male control. I^2 quantifies heterogeneity; p for heterogeneity; r^2 magnitude of the heterogeneity between studies from random-effect model; RE, random-effect. Segregation distortions with 95% confidence interval were calculated using a random-effects meta-analysis

MJD ancestral haplotypes: $A^{669}TG/G^{669}TG$ (rs1048755), $C^{987}GG/G^{987}GG$ (rs12895357), and TAA^{1118}/TAC^{1118} (rs7158733).[37] From these 13 publications, five included both SCA3/MJD index cases as well as controls,[11,37-40] six included SCA3/MJD subjects only,[41-46] and two studied the general population only.[47,48] Besides those three classical SNPs (rs1048755, rs12895357, and rs7158733), these studies included different combinations of 31 additional intragenic markers (other SNPs and the CAG repeat tract) and/or four short tandem repeats (STRs) close to *ATXN3*. This wide range of markers makes it difficult to homogenize data for analysis. General results of haplotypes based on the three classical SNPs, from these 13 studies, are summarized in Table 1. Haplotype proportions obtained in SCA3/MJD families were presented near the proportions obtained by studies performed in the general population from the same geographical origin.

Six haplotypes built with the classical SNPs (rs1048755, rs12895357, and rs7158733), and probably identifying ancestral lineages, were detected in SCA3/MJD families: ACA (219 families from all papers, around the world), AGA (40 families from China, Morocco, United States, and Yemenite Israel), GGC (33 families from Portugal, Spain, and United States), GGA (one family from China and one from French Guiana), GCC (one family from Portugal), and GCA (one family from China). However, these haplotypes based on three SNPs are more probably oversimplifications. For instance, the most common haplotype, ACA, was recently subdivided into two lineages of descents in a study that included Chinese SCA3/MJD families; the oldest one being haplotype A^{669}(rs1048755)-A(rs56268847)-C^{987}(rs12895357)- A^{1118}(rs7158733), or haplotype A, and the most recent being A^{669}-(rs1048755)-G(rs56268847)-C^{987}-(rs12895357)-A^{1118}(rs7158733), or haplotype B.[11]

Genetic distances among populations were estimated by four studies, considering the most parsimonious number of single mutation steps between alleles built with SNPs and STRs. According to these analyses and to most recent data, ACA, AGA, and AGC haplotypes would have Asian origins, while GGC would have originated in the Portuguese population. Age estimations, in mean (SD), were 16 335 (1966) years for the ACA (the oldest form, called haplotype A), 11 837 (1871) years for the AGA, and 1291 ± 553 years for the GGC.[11,43]

4 | DISCUSSION

The coexistence of few and old ancestral strains of a mutation with the tendency for important anticipations of age at onset has been one

TABLE 1 General results of 13 studies on ancestral haplotypes at *ATXN3*, when haplotypes were built with the three classical SNPs A^{669}TG/G^{669}TG (rs1048755), C^{987}GG/G^{987}GG (rs12895357), and TAA^{1118}/TAC^{1118} (rs7158733). Results obtained in SCA3/MJD families were presented near the results obtained in controls from the same geographical origin

	Number of SCA3 families	Number of controls	Haplotypes found using the three classical SNPs												
			ACA		AGA		AGC		GGC		GCC		Others		
Country			SCA3	Controls	SCA3	Controls	SCA3	Controls	SCA3	Controls	SCA3	Controls	SCA3	Controls	Study
India	7	824	100%	28.6%		26%		0%		35%		10%	0%	0%	Chattopadhyay et al 2003[38]
	4		100%		0%		0%		0%		0%		0%		Martins et al 2007[43]
China		100		49%		6%				43%				2%	Martins et al 2006[48]
	30	218	100%	86%	0%	2,4%	0%	10%	0%	1.6%	0%	0%	0%	0%	Gan et al 2015[39]
	51	319	45%	29.1%	49%	14.4%	2%	3.4%	0%	42.3%	0%	4.7%	0%	0%	Li et al 2019[11]
Japan	27		81.5%		0%								18.5%		Martins et al 2007[43]
Czech Republic		204		6%		5.4%				73.5%		15%		0%	Bauer et al 2006[47]
France	29		86%										14%		Martins et al 2007[43]
Netherlands	22		100%		0%		0%		0%		0%		0%		Verbeek et al 2004[41]
Germany	14		92%										8%		Martins et al 2007[43]
Portugal mainland		432		25%		0.5%		0%		72%		2%		0%	Martins et al 2006[48]
	104		32%						56.7%		2%		10%		Martins et al 2007[43]
Flores Island	10		90%		0%		0%		10%		0%		0%		Gaspar et al 2001[37]
São Miguel	12		0%		0%		0%		100%		0%		0%		Gaspar et al 2001[37]
Australia	2		100%		0%		0%		0%		0%		0%		Martins et al 2012[44]
Israel	6	100	0%	26%	100%	1%	0%	0%	0%	73%	0%	0%	0%	0%	Sharony et al 2019[40]
Morocco	1		0%		100%		0%		0%		0%		0%		Gaspar et al 2001[37]
Mozambique and Angola		76		14.5%		36.8%		0%		46%		0%		2%	Martins et al 2006[48]
Niger	5		100%		0%		0%		0%		0%		0%		Ogun et al 2015[46]
French Guyana	1		0%		0%		0%		0%		0%		0%		Gaspar et al 2001[37]
United States	11		54%		18%				27%				0%		Gaspar et al 2001[37]
	14		36%		43%				21.5%				0%		Subramony et al 2002[42]
	23		83%						17%				0%		Martins et al 2007[43]
World	370	2276	55%	28.3%	10.8%	13.6%	0.2%	1%	20.8%	45.6%	0.5%	6%	5.4%	0.1%	

of the most intriguing queries for anyone studying SCA3/MJD. The contradiction between the two phenomena denotes that some positive selective force exists and acts on this condition. Bearing that in mind, we performed a systematic review of the evidence gathered on forces that influence the maintenance or removal of the mutant *ATXN3* alleles from the population pool to date. Studies on instability of the CAGexp and on changes in the AO across generations concluded in favor of successive expansions and anticipations across generations. However, we detected potential bias favoring the recruitment of affected children with further expansions over those with contractions. Fitness of SCA3/MJD was properly analyzed and revealed increased fertility rates; but fitness was studied once only. Segregation of gametes was the approach with more robust results: meta-analysis showed that segregation favors the transmission of the mutant over the normal *ATXN3* allele. Therefore, we propose that increased fitness and meiotic drive are among the positive forces that compensate the negative selection produced by anticipation, all related to the expanded repeat at *ATXN3*, allowing the potentially long survival of these mutations in the human population.

Thirteen studies on CAGexp instabilities concluded in favor of successive increases of the CAGexp length across generations. However, one can assume that the strength of this line of evidence remains limited, since none of these reports stated the analysis of all children of carriers. Bias in favor of pairing more severe offspring with less severe parents was not avoided. The time span of these studies per se can limit observations.[20] Observers will probably be unable to report, for instance, a pair of a 40-years-old onset parent and a 60-years-old onset child. This hypothetical parent will be probably deceased by the time of the child's recruitment or diagnostic. Detection of such a potential CAGexp contraction could not be simultaneous in parent and child, but would depend on institutional reports covering 30–40 chronological years, a timespan not yet available since the discovery of the *ATXN3* mutation.

Although rarely obtained in SCA3/MJD, results of studies on somatic mosaicism in polyglutaminopathies could be seen as an indirect support for speculation on the way instability during meiotic cell divisions may affect the repeats. Post-mortem studies in affected patients by SCA3/MJD, and especially in those with Huntington's disease and SCA1, revealed the presence of expanded CAG repeat instabilities in multiple tissues.[49-51] The result of instability in most tissues was toward expansions, although the tendency for the predominance of contractions has been seen in ovaries and testicles of HD transgenic mice.[52] In humans, the study of HD cases was more focused on assessing the expansion and not the instability.[51] Even so, this study described that ~~the~~ gonads presented low expansion rates, with the exception of a subject with juvenile HD and a large original expansion. We expect that these researchers will report in-depth analysis of CAG instability (contractions, stability, and further expansions) observed in the gonads of these post-mortem studies. Meanwhile, long-term cohorts, including more than one generation and, more essentially, including all transmissions, will be another good tool to clarify if instability of the CAGexp is indeed a negative or neutral force related to the maintenance of mutation in the population.

In contrast, differences between instabilities inherited from paternal and maternal transmissions seem to reflect a real biological phenomenon. Proposed mechanisms have been summarized elsewhere.[53] The overall effect over progeny of SCA3/MJD carriers is not well estimated, as none of the studies clearly stated that efforts were done to recruit balanced samples of transmitting mothers and fathers.

The results related to changes in AO between parents and children seemed to have been more prone to bias than those related to CAGexp instabilities. The risk of this specific bias has been acknowledged early after the discovery of *ATXN3* mutation.[20] There are at least two different levels of bias: those due to different AO measures among studies, and those due to recruitment biases. AO was defined in all articles as age at the first symptom as noted by the patient and/or his/her relatives. Cultural and family differences can possibly affect different perceptions of AO among different SCA3/MJD populations studied. However, the meta-analysis of factors that influence AO in SCA3/MJD[7] showed that the mean as well as variability of AO was similar in 10 different populations studied (more than 2000 subjects). If memory biases occurred, they must have reached different populations in a similar manner. On the other hand, reasons for potential recruitment bias here were similar to those detected in studies about instabilities. Papers did not state that all children of the included parents were analyzed; none of them clearly excluded the present generation to avoid bias toward cases at younger age at onset; none of them balanced maternal and paternal transmissions. Among these sources of bias, maybe the most important was the absence of an age cutoff for the inclusion of contemporary subjects. As many mutation carriers could be still asymptomatic during the observation, anticipation might have been overestimated.

There is another reason to make us skeptical about the anticipation meta-analyzed here: the order of magnitude. The CAGexp was related to 55% of variability in AO.[7] If numbers estimated here are correct, then to each CAG repeat added due to an unstable transmission, the change in AO would be 3.5 years (or 55% of the 7.75 years minus AO divided per 1.23 added to CAG in the expanded repeat, per average transmission). This value is much more than the larger CAGexp effect estimated in a large meta-analysis with more than 2000 individual participant data: for each additional CAG in the expanded repeat, a reduction of 2.58 years in the AO would be expected.[7]

The fact remains that, despite of the apparent real differences in instabilities inherited from affected mothers or fathers, the average change in the direction of a strong anticipation of 7.75 years per generation did not present substantial differences related to sex of the transmitting parent, and there is no explanation for this similarity. Since the CAGexp explains 55% of variability in AO[7] and since the variability in the inherited CAGexp was wider in male than in female transmissions, then other factors influencing anticipation would be more present in female than in male transmissions.

Trinucleotide repeat expansions related to neurodegenerative disorders are more frequently studied due to the negative impact on health of people. Ascertaining whether or not there are positive effects associated with the presence of CAG expansions appears to

be counterintuitive and even futile, and perhaps that is the reason for few studies conducted from this perspective. The effects of expansion on fertility and meiotic segregation fall into this field of potentially positive effects from the point of view of natural selection. Only one study compared the number of children of SCA3/MJD carriers to those of noncarriers, a previous publication of our group.[28] Emeritus SCA3/MJD carriers have 45% more children than related noncarriers. This result cannot be generalized to other SCA3/MJD cohorts, since it was observed just in the Rio Grande do Sul cohort. However, evidence of similar increased fitness was also obtained in other polyQ diseases, such as Huntington disease, SCA1, and SCA2,[54,55] suggesting that other SCA3/MJD cohorts can present similar findings.

Meiotic drive or segregation distortion is the phenomenon in which the presence of one allele (D) at a locus reduces the chances of gametes carrying the alternative allele (d).[56] The segregation analysis by the inclusion of all children of a given affected carrier was performed four times in the literature. In three reports, the noncarrier status was acceptably established. Results included in this meta-analysis point to a favorable selection of the expanded repeat at some point between meiosis and childbirth. In contrast, different results obtained by the only two studies on sperm cells prevented them to be meta-analyzed and are difficult to explain. We speculate whether such results would be due to diverse cell conditions such as temperature and time between collection and analysis.

A collateral but very interesting result was the one related to segregation distortion among normal persons carrying normal CAG length alleles in the general population. Indeed, the shorter CAG repeat allele is more transmitted than the larger normal one in controls. Thus, during *ATXN3* allele segregations, there is a selection of both extreme length CAG alleles: the expanded one, when present, and the shortest one, when there is no expanded allele, in a classic case of disruptive selection. This phenomenon might be related to the virtual empty range of CAG repeats between 45 and 50 repeats. Normal subjects present up to 44 repeats, while SCA3/MJD carriers present one allele with 51 or more CAG repeats. This range of six repeats between normal and expanded alleles was never populated with observations – at least they were never formally described. At the same time, this range (45–50 repeats) could theoretically be called "intermediate alleles" (AI, in the literature). AI are usually conceptualized as nonpathogenic alleles that are prone to instabilities when crossing meiosis, and therefore as source of *de novo* mutations. The lack of carriers of 45–50 CAG repeats at *ATXN3* might be associated with the fact that *de novo* mutations were never reported in SCA3/MJD.

In fact, ancestral origins of SCA3/MJD seem to be few, the oldest one dating of 16 000 years back, or around 640 generations or more.[11] These older lineages originated in the ancestral Chinese populations. The ancestral haplotype ACA, through one or two haplotypic variants A and B,[11] is the oldest and most common worldwide and seems to have spread through migrations to different parts of the contemporary world. Presence of so old ancestral haplotypes would be hard to be matched with a predominance of negative selection.

We are aware of difficulties related with the study of natural selection in humans. It is well known that natural selection is the weakest on traits manifested after the majority of an organism's reproduction is complete. Even taking this into account, many monogenic diseases, including those related to neurodegenerative diseases, are present at higher frequencies than would be expected from the elimination carried out by natural selection over time. One possible explanation would be the antagonistic pleiotropy of a gene, that is, pleiotropic functions acting antagonistically in different cells, tissues, or periods of time. A number of empirical examples of antagonistic pleiotropy associated with deleterious genes in humans have already been described or proposed, suggesting that the phenomenon may be much more frequent than expected.[13] Antagonistic pleiotropy associated with exonic CAG expansions can produce events as diverse as the selective advantages studied here, or as the lower predisposition to cancer described in patients with SCA3/MJD and in other polyQ diseases[57-61] as well as neurodegeneration.

The overall effect of these forces over the epidemiology of SCA3/MJD will take time to be demonstrated, as many generations need to be monitored. So far, only one study has followed changes in the prevalence of SCA3/MJD over 44 years in a geographic region, the Azores.[62] The disease prevalence increased from 23/100 000 in 1981 to 39/100 000 in 2015. At this point, it is hard to say whether this was due to recruitment biases or if it was a true effect.

In addition, the balance between negative and positive selective forces is also likely to vary according to the severity of the phenotypes of parents with SCA3/MJD. The various phenotypes associated with the disease should change these evolutionary vectors at each phenotype aggravation. One study on the classical SCA3/MJD subphenotypes[63] detected that subjects with type 1 had fewer children than subjects with types 2 and 3.[64] Type 1 is characterized by early AO, longer CAG repeat expansions, and presence of intense spasticity and dystonia. However, the reduction in fertility disappeared from type 1 subjects with AO similar to AO of type 2 and 3 individuals, suggesting that the reduction in fertility was more probably related to the early AO than to the neurological manifestations per se, in that cohort. In fact, this result was similar to the results reported in the study that addressed fitness[29]: cases with later AO had more children. Adding the effects of two different evolutionary forces (fertility on one hand and segregation distortion on the other), we can say that the reproductive success would ensure that the meiotic drive takes effect, that is, more affected than nonaffected children were generated. All of this happens in an opposite direction among cases with early AO (less children, less effect of the meiotic drive). However, there is no doubt that the characterization of the phenotype only through AO is limited, especially in a condition such as SCA3/MJD. More papers relating fertility to the neurological severity such as this former one[64] are required, because the severity of the disease manifested in a subject should greatly influence his/her reproductive capacity. For instance, sicker people will have less chance of adapting to social life, and less chance of taking care of their children.

If the effect of selective forces is relevant to the epidemiology of genetic diseases, it is essential that clues about positive versus negative selection be further studied in human populations. Whether positive and negative selective forces are at balance in SCA3/MJD or if one

of them prevails, remains to be established. If they are at balance, the frequency of disease would not change substantially in time, provided that other bottleneck effects did not occur. If negative forces predominate, frequency of SCA3/MJD would be reduced in human populations. We are convinced that at least the segregation of alleles is a positive force that increases SCA3/MJD frequency. SCA3/MJD fitness is possibly but not surely increased. If these forces predominate, then SCA3/MJD frequency would increase in the future years.

In conclusion, the present meta-analysis gave satisfactory results on segregation distortion, while other outcomes under study should be better addressed in the future. Better studies on unstable transmissions are required, with genotyping of all sibships to avoid bias in favor of more severe cases. Better studies on anticipation are also necessary, where all offspring of a parent should be genotyped and/or a minimum age should be a criterion to include children with late AO and, therefore, without anticipation. More studies are also required on genetic fitness in SCA3/MJD, so that the increased fertility might be validated. If these selective forces would be better clarified, the impact on prevalence prediction and on information about recurrence will undoubtedly assist health policies and help the genetic counseling of carriers of SCA3/MJD.

ACKNOWLEDGEMENTS

We would like to thank Vanessa Leotti and Vania Hirakata for statistical support. This study was supported by Fundo de Incentivo à Pesquisa do Hospital de Clínicas de Porto Alegre (FIPE-HCPA), grant number 2019-0169. Jordânia dos Santos Pinheiro was supported by Fundação de Amparo à Pesquisa do Rio Grande do Sul (FAPERGS). Lucas Schenatto Sena, Maria Luiza Saraiva-Pereira and Laura Bannach Jardim were supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/cge.13888.

DATA AVAILABILITY STATEMENT

All data are available as Supplemental Materials.

ORCID

Laura Bannach Jardim https://orcid.org/0000-0001-6907-5068

REFERENCES

1. Takiyama Y, Nishizawa M, Tanaka H, et al. The gene for Machado-Joseph disease maps to human chromosome 14q. *Nat Genet*. 1993;4:300-304.
2. Kawaguchi Y, Okamoto T, Taniwaki M, et al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32. *Nat Genet*. 1994;8:221-228.
3. Gu W, Ma H, Wang K, et al. The shortest expanded allele of the MJD1 gene in a Chinese MJD kindred with autonomic dysfunction. *Eur Neurol*. 2004;52:107-111.
4. Paulson H, Shakkottai V. (2020). In: Adam MP, Ardinger HH, Pagon RA, et al., eds. Spinocerebellar Ataxia Type 3. 1998 Oct 10 [Updated June 4, 2020]. GeneReviews®. Seattle (WA): University of Washington, Seattle; 1993–2020. https://www.ncbi.nlm.nih.gov/books/NBK1196/.
5. Matilla-Dueñas A, Ashizawa T, Brice A, et al. Consensus paper: pathological mechanisms underlying neurodegeneration in spinocerebellar ataxias. *Cerebellum*. 2014;13:269-302.
6. Saute JAM, Jardim LB. Machado Joseph disease: clinical and genetic aspects, and current treatment. *Expert Opin Orphan Drugs*. 2015;3:517-535.
7. De Mattos EP, Musskopf MK, Leotti VB, Saraiva-Pereira ML, Jardim LB. Genetic risk factors for modulation of age at onset in Machado-Joseph disease/spinocerebellar ataxia type 3: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry*. 2019;90(2):203-210.
8. Akçimen F, Martins S, Liao C, et al. Genome-wide association study identifies genetic factors that modify age at onset in Machado-Joseph disease. *Aging*. 2020;12(6):4742-4756.
9. Mergener R, Furtado GV, de Mattos EP, Leotti VB, Jardim LB, Saraiva-Pereira ML. Variation in DNA repair system gene as an additional modifier of age at onset in spinocerebellar ataxia type 3/Machado–Joseph disease. *Neuromolecular Med*. 2020;22(1):133-138.
10. Ruano L, Melo C, Silva MC, Coutinho P. The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies. *Neuroepidemiology*. 2014;42(3):174-183.
11. Li T, Martins S, Peng Y, et al. Is the high frequency of Machado-Joseph disease in China due to new mutational origins? *Front Genet*. 2019;9:740.
12. Williams GC. Pleiotropy, natural selection, and the evolution of senescence. *Evolution*. 1957;11(4):398-411.
13. Carter AJ, Nguyen AQ. Antagonistic pleiotropy as a widespread mechanism for the maintenance of polymorphic disease alleles. *BMC Med Genet*. 2011;12:160.
14. Souza GN, Kersting N, Krum-Santos AC, et al. Spinocerebellar ataxia type 3/Machado–joseph disease: segregation patterns and factors influencing instability of expanded CAG transmissions. *Clin Genet*. 2016;90(2):134-140.
15. Martins S, Coutinho P, Silveira I, et al. Cis-acting factors promoting the CAG intergenerational instability in Machado–Joseph disease. *Am J Med Genet Part B*. 2008;147:439-446.
16. Løkkegaard T, Nielsen JE, Hasholt L, et al. Machado–Joseph disease in three Scandinavian families. *J Neurol Sci*. 1998;156(2):152-157.
17. Takiyama Y, Shimazaki H, Morita M, et al. Maternal anticipation in Machado-Joseph disease (MJD): some maternal factors independent of the number of CAG repeat units may play a role in genetic anticipation in a Japanese MJD family. *J Neurol Sci*. 1998;155(2):141-145.
18. Soong BW, Cheng CH, Liu RS, Shan DE. Machado-joseph disease: clinical, molecular, and metabolic characterization in Chinese kindreds. *Ann Neurol*. 1997;41(4):446-452.
19. Maruyama H, Nakamura S, Matsuyama Z, et al. Molecular features of the CAG repeats and clinical manifestation of Machado–Joseph disease. *Hum Mol Genet*. 1995;4(5):807-812.
20. Maciel P, Gaspar C, DeStefano AL, et al. Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *Am J Hum Genet*. 1995;57(1):54-61.
21. Durr A, Stevanin G, Cancel G, et al. Spinocerebellar ataxia 3 and Machado-Joseph disease: clinical, molecular, and neuropathological features. *Ann Neurol*. 1996;39(4):490-499.
22. Cancel G, Abbas N, Stevanin G, et al. Marked phenotypic heterogeneity associated with expansion of a CAG repeat sequence at the spinocerebellar ataxia 3/Machado-Joseph disease locus. *Am J Hum Genet*. 1995;57(4):809-816.
23. Igarashi S, Takiyama Y, Cancel G, et al. Intergenerational instability of the CAG repeat of the gene for Machado-Joseph disease (MJD1) is

affected by the genotype of the normal chromosome: implications for the molecular mechanisms of the instability of the CAG repeat. *Hum Mol Genet*. 1996;5(7):923-932.
24. Takiyama Y, Igasrashi S, Rogaeva EA, et al. Evidence for intergenerational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. *Hum Mol Genet*. 1995;4(7):1137-1146.
25. Hsieh M, Tsai HF, Lu TM, Yang CY, Wu HM, Li SY. Studies of the CAG repeat in the Machado-Joseph disease gene in Taiwan. *Hum Genet*. 1997;100(2):155-162.
26. Sasaki H, Wakisaka A, Fukazawa T, et al. CAG repeat expansion of Machado-Joseph disease in the Japanese: analysis of the repeat instability for parental transmission, and correlation with disease phenotype. *J Neurol Sci*. 1995;133(1–2):128-133.
27. Infante J, Combarros O, Volpini V, Corral J, Llorca J, Berciano J. Autosomal dominant cerebellar ataxias in Spain: molecular and clinical correlations, prevalence estimation and survival analysis. *Acta Neurol Scand*. 2005;111(6):391-399.
28. 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. *Ann Hum Genet*. 2008;72:57-64.
29. Bettencourt C, Santos C, Kay T, Vasconcelos J, Lima M. Analysis of segregation patterns in Machado–Joseph disease pedigrees. *J Hum Genet*. 2008a;53(10):920-923.
30. Ikeuchi T, Igarashi S, Takiyama Y, et al. Non-Mendelian transmission in dentatorubral-pallidoluysian atrophy and Machado-Joseph disease: the mutant allele is preferentially transmitted in male meiosis. *Am J Hum Genet*. 1996;58(4):730-733.
31. Riess O, Epplen JT, Amoiridis G, Przuntek H, Schols L. Transmission distortion of the mutant alleles in spinocerebellar ataxia. *Hum Genet*. 1997;99:282-284.
32. Bettencourt C, Fialho RN, Santos C, et al. Segregation distortion of wild-type alleles at the Machado-Joseph disease locus: a study in normal families from the Azores islands (Portugal). *J Hum Genet*. 2008b; 53(4):333-339.
33. Grewal RP, Cancel G, Leeflang EP, et al. French Machado-Joseph disease patients do not exhibit gametic segregation distortion: a sperm typing analysis. *Hum Mol Genet*. 1999;8(9):1779-1784.
34. Takiyama Y, Sakoe K, Soutome M, et al. Single sperm analysis of the CAG repeats in the gene for Machado-Joseph disease (MJD1): evidence for non-Mendelian transmission of the MJD1 gene and for the effect of the intragenic C GG/G GG polymorphism on the intergenerational instability. *Hum Mol Genet*. 1997;6(7):1063-1068.
35. Rubinsztein DC, Leggo J. Non-Mendelian transmission at the Machado-Joseph disease locus in normal females: preferential transmission of alleles with smaller CAG repeats. *J Med Genet*. 1997;34(3): 234-236.
36. MacMillan JC, Voisey J, Healey SC, Martin NG. Mendelian segregation of normal CAG trinucleotide repeat alleles at three autosomal loci. *J Med Genet*. 1999;36(3):258-259.
37. Gaspar C, Lopes-Cendes I, Hayes E, et al. Ancestral origins of the Machado-Joseph disease mutation: a worldwide haplotype study. *Am J Hum Genet*. 2001;68(2):523-528.
38. Chattopadhyay B, Basu P, Gangopadhyay PK, et al. Variation of CAG repeats and two intragenic polymorphisms at SCA3 locus among Machado–Joseph disease/SCA3 patients and diverse normal populations from eastern India. *Acta Neurol Scand*. 2003;108(6):407-414.
39. Gan SR, Ni W, Dong Y, Wang N, Wu ZY. Population genetics and new insight into range of CAG repeats of spinocerebellar ataxia type 3 in the Han Chinese population. *PLoS One*. 2015;10(8):134405.
40. Sharony R, Martins S, Costa IP, et al. Yemenite-Jewish families with Machado–Joseph disease (MJD/SCA3) share a recent common ancestor. *Eur J Hum Genet*. 2019;27(11):1731-1737.
41. Verbeek DS, Piersma SJ, Hennekam EF, Ippel EF, Pearson PL, Sinke RJ. Haplotype study in Dutch SCA3 and SCA6 families: evidence for common founder mutations. *Eur J Hum Genet*. 2004;12(6):441-446.
42. Subramony SH, Hernandez D, Adam A, et al. Ethnic differences in the expression of neurodegenerative disease: Machado-Joseph disease in Africans and Caucasians. *Mov Disord*. 2002;17(5):1068-1071.
43. Martins S, Calafell F, Gaspar C, et al. Asian origin for the worldwide-spread mutational event in Machado-Joseph disease. *Arch Neurol*. 2007;64(10):1502-1508.
44. Martins S, Soong B, Wong VCN, et al. Mutational origin of Machado-Joseph disease in the Australian aboriginal communities of Groote Eylandt and Yirrkala. *Arch Neurol*. 2012;69(6):746-751.
45. Mittal U, Srivastava AK, Jain S, Jain S, Mukerji M. Founder haplotype for Machado-Joseph disease in the Indian population: novel insights from history and polymorphism studies. *Arch Neurol*. 2005;62(4):637-640.
46. Ogun S, Martins S, Adebayo P, et al. Machado–Joseph disease in a Nigerian family: mutational origin and review of the literature. *Eur J Hum Genet*. 2015;23:271-273.
47. Bauer PO, Zumrova A, Matoska V, et al. Absence of spinocerebellar ataxia type 3/Machado–Joseph disease within ataxic patients in the Czech population. *Eur J Neurol*. 2005;12(11):851-857.
48. Martins S, Calafell F, Wong VC, Sequeiros J, Amorim A. A multistep mutation mechanism drives the evolution of the CAG repeat at MJD/SCA3 locus. *Eur J Hum Genet*. 2006;14(8):932-940.
49. Tanaka F, Sobue G, Doyu M, et al. Differential pattern in tissue-specific somatic mosaicism of expanded CAG trinucleotide repeat in dentatorubral-pallidoluysian atrophy, Machado-Joseph disease, and X-linked recessive spinal and bulbar muscular atrophy. *J Neurol Sci*. 1996;135:43-50.
50. Ito Y, Tanaka F, Yamamoto M, et al. Somatic mosaicism of the expanded CAG trinucleotide repeat in mRNAs for the responsible gene of Machado-Joseph disease (MJD), dentatorubral-pallidoluysian atrophy (DRPLA), and spinal and bulbar muscular atrophy (SBMA). *Neurochem Res*. 1998;23:25-32.
51. Mouro Pinto R, Arning L, Giordano JV, et al. Patterns of CAG repeat instability in the central nervous system and periphery in Huntington's disease and in spinocerebellar ataxia type 1. *Hum Mol Genet*. 2020;29:2551-2567. https://doi.org/10.1093/hmg/ddaa139.
52. Lee JM, Zhang J, Su AI, et al. A novel approach to investigate tissue-specific trinucleotide repeat instability. *BMC Syst Biol*. 2010;4:29. https://doi.org/10.1186/1752-0509-4-29.
53. Martins S, Pearson CE, Coutinho P, et al. Modifiers of (CAG)n instability in Machado–Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes. *Hum Genet*. 2014;133:1311-1318.
54. 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 expansion. *Ann Hum Genet*. 1996;60 (5):423-435.
55. Sena LS, Castilhos RM, Mattos EP, et al. Selective forces related to spinocerebellar ataxia type 2. *Cerebellum*. 2019;18(2):188-194.
56. Charlesworth B. Population genetics. In: Levin SA, ed. *Encyclopedia of Biodiversity*. 2nd ed. Amsterdam, Netherlands: Elsevier; 2013:182-198.
57. Sørensen SA, Fenger K, Olsen JH. Significantly lower incidence of cancer among patients with Huntington disease: an apoptotic effect of an expanded polyglutamine tract? *Cancer*. 1999;86:1342-1346.
58. Ji J, Sundquist K, Sundquist J. Cancer incidence in patients with polyglutamine diseases: a population-based study in Sweden. *Lancet Oncol*. 2012;13:642-648.
59. Turner MR, Goldacre R, Goldacre MJ. Reduced cancer incidence in Huntington's disease: record linkage study clue to an evolutionary trade-off? *Clin Genet*. 2013;83:588-590.
60. Souza GN, Kersting N, Gonçalves TA, et al. Cancer in Machado–Joseph disease patients—low frequency as a cause of death. *Cancer Genet*. 2017;212-213:19-23.

61. Coarelli G, Diallo A, Thion MS, et al. Low cancer prevalence in polyglutamine expansion diseases. *Neurology*. 2017;88(12):1114-1119.
62. Araújo MA, Raposo M, Kazachkova N, Vasconcelos J, Kay T, Lima M. Trends in the epidemiology of spinocerebellar ataxia type 3/-Machado-Joseph disease in the Azores Islands, Portugal. *JSM Brain Sci*. 2016;1(1):1001.
63. Lima L, Coutinho P. Clinical criteria for diagnosis of Machado-Joseph disease: report of a non-Azorena Portuguese family. *Neurology*. 1980; 30(3):319-322.
64. Lima M, Smith MT, Silva C, Abade A, Mayer FM, Coutinho P. Natural selection at the MJD locus: phenotypic diversity, survival and fertility among Machado-Joseph disease patients from the Azores. *J Biosoc Sci*. 2001;33(3):361-373.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Sena LS, dos Santos Pinheiro J, Saraiva-Pereira ML, Jardim LB. Selective forces acting on spinocerebellar ataxia type 3/Machado–Joseph disease recurrency: A systematic review and meta-analysis. *Clinical Genetics*. 2020;1–12. https://doi.org/10.1111/cge.13888

CAPÍTULO III. FORÇAS SELETIVAS ASSOCIADAS AOS ALELOS EXPANDIDOS DO *ATXN2*

Manuscrito 2: "Spinocerebellar ataxia type 2 from an evolutionary perspective: Systematic review and meta-analysis"

Artigo publicado no periódico

Clinical Genetics

(doi: 10.1111/cge.13978)

Received: 2 March 2021 | Revised: 19 April 2021 | Accepted: 5 May 2021

DOI: 10.1111/cge.13978

ORIGINAL ARTICLE

CLINICAL GENETICS WILEY

Spinocerebellar ataxia type 2 from an evolutionary perspective: Systematic review and meta-analysis

Lucas Schenatto Sena[1,2] | Jordânia dos Santos Pinheiro[2,3] | Ali Hasan[1,2] | Maria Luiza Saraiva-Pereira[1,2,4,5,6] | Laura Bannach Jardim[1,2,5,6,7]

[1]Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[2]Centros de Pesquisa Clínica e Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

[3]Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[4]Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[5]Instituto de Genética Médica Populacional, Porto Alegre, Brazil

[6]Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

[7]Departamento de Medicina Interna, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Correspondence
Laura Bannach Jardim, DMI FAMED UFRGS, and Medical Genetics Service Hospital de Clínicas de Porto Alegre Rua Ramiro Barcelos 2350 90035-003 Porto Alegre, Brazil.
Email: ljardim@hcpa.edu.br

Funding information
Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Numbers: 141781/2018-1, 303577/2016-19, 313132/2018-6; Fundação do Amparo à Pesquisa do Rio Grande do Sul (FAPERGS), Grant/Award Number: 126984/2019-0; Fundo de Incentivo à Pesquisa do Hospital de Clínicas de Porto Alegre (FIPE-HCPA), Grant/Award Number: 2019-0169

Abstract

Dominant diseases due to expanded CAG repeat tracts, such as spinocerebellar ataxia type 2 (SCA2), are prone to anticipation and worsening of clinical picture in subsequent generations. There is insufficient data about selective forces acting on the maintenance of these diseases in populations. We made a systematic review and meta-analysis on the effect of the CAG length over age at onset, instability of transmissions, anticipation, *de novo* or sporadic cases, fitness, segregation of alleles, and ancestral haplotypes. The correlation between CAG expanded and age at onset was $r^2 = 0.577$, and transmission of the mutant allele was associated with an increase of 2.42 CAG repeats in the next generation and an anticipation of 14.62 years per generation, on average. One *de novo* and 18 sporadic cases were detected. Affected SCA2 individuals seem to have more children than controls. The expanded allele was less segregated than the 22-repeat allele in children of SCA2 subjects. Several ancestral SCA2 haplotypes were published. Data suggest that SCA2 lineages may tend to disappear eventually, due to strong anticipation phenomena. Whether or not the novel cases come from common haplotypes associated with a predisposition to further expansions is a question that needs to be addressed by future studies.

KEYWORDS
allele segregation, ancestral haplotypes, anticipation, fertility, meta-analysis, spinocerebellar ataxia type 2, unstable transmissions

1 | INTRODUCTION

Spinocerebellar ataxia type 2 (SCA2) is an autosomal dominant neurodegenerative disease caused by the expansion of a CAG repeat in *ATXN2*. Normal CAG tracts in *ATXN2* range from 12 to 32–33 repeats, being the most frequent of them the 22 repeats long allele that represent 75%–90% of alleles in several populations.[1] The reason for that remains unknown. Tracts with more than 33 repeats are pathogenic, with autosomal dominant expression. The most accepted pathogenetic mechanism relates ataxin-2 with the expanded

polyQ—the entire protein or peptides produced after its proteolysis—to intracellular aggregation, neuronal toxicity, and degeneration.[2] Some lines of evidence link the role of ataxin-2 in RNA and protein homeostasis to events in SCA2 pathogenesis; even in this context, toxic gain-of-function rather than loss-of-function mechanisms seems to be the key driver of the disease.[3]

SCA2 onset starts around the mid-30s, and includes gait ataxia and dysarthria, associated with slow saccadic eye movements.[1,4] The age at onset (AO) of first symptoms correlates with CAG expansion length.[5-7] There is a progressive worsening of the clinical picture as the individual gets older, with successive involvement of motor neurons, basal ganglia, and peripheral afferent pathways, among others.[8] Several clinical scales focused on neurological manifestations documented the speed of this deterioration. A recent meta-analysis estimated that the Scale of Assessment and Rating of Ataxia (SARA) worsened on average 1.40/40 (1.19–1.61) points per year after onset of symptoms in SCA2.[9] Earlier AO were associated with faster progression, meaning that the sooner SCA2 starts, the more severe and faster the disease progresses.[10] Survival is reduced to 68 [95% CI: 65–70] years of age, usually after a wheelchair period.[11] Longer expansions are one of the strongest contributing risk factors for early death.[12]

The expanded CAG repeat is unstable upon transmissions.[6] Although the disease might start, in theory, at a later or earlier age than in the transmitting parent, further expansions and, consequently, anticipation of AO in the offspring have been more frequently observed, with several reports of carriers starting the disease during childhood.[13,14] Since early onset can affect the reproductive chances of the carriers, anticipation is a selective force that would contribute to elimination of expanded *ATXN2* lineages from populations.

However, SCA2 is one of the most common SCAs worldwide. The contradiction between anticipation and the relative frequency of the disease raises the question of possible selective forces acting over SCA2. Thus, our aim was to perform a systematic review and meta-analysis on some outcomes related to the CAG repeat expansion at *ATXN2* with potential effect over SCA2 recurrence. In order to achieve that, the following aspects were reviewed: (1) the correlation between AO and length of the expanded CAG repeat; (2) the instability of the expanded repeat when crossing meiosis; (3) changes in AO among different generations; (4) *de novo* or sporadic cases; (5) differences in reproductive rates between carriers and non-carriers; (6) meiotic segregation; and (7) ancestral haplotypes.

2 | METHODS

The methodology protocol for this study was registered at Prospero platform (https://www.crd.york.ac.uk/PROSPERO/) under the number CRD42020182293, and prior to data extraction (https://www.crd.york.ac.uk/PROSPERO/).

This review, the term "carrier" means an individual, symptomatic or not, heterozygous for one *ATXN2* allele with more than 33 CAG repeats in her/his genotype. The term "fitness" stands for the reproductive success of a phenotype, and was estimated by the ratio between the mean number of children of carriers over the mean number of children of controls. The term "emeritus" was the term used for a person that probably reached the end of her/his reproductive age, by being older than an arbitrated age by a given study. The term "*de novo*" stands for subjects or alleles (in which a CAG tract with more than 33 repeats was inherited from parents with documented alleles equal or shorter than 33 repeats. The term "sporadic case" refers to symptomatic SCA2 subjects without a family history of ataxia, whose parents were no more available for genotyping.

2.1 | Literature search and data extraction

MEDLINE (PubMed) was searched from November 1996 to July 2020 for reports on the following main outcomes: (1) correlation between AO and the expanded CAG repeat length; (2) instability of the expanded repeat when crossing meiosis (contractions and/or expansions); (3) differences in AO between different generations, usually named as "anticipation"; (4) *de novo* or sporadic cases; (5) reproductive rates of the carriers compared to those of non-carriers, also called "genetic fitness"; (6) meiotic segregation; and (7) ancestral haplotypes. Each research question was answered by one search; therefore, seven searches were performed in total. Each search started with terms: ("SCA2" OR "Spinocerebellar ataxia type 2" or "SCA 2"). After them, search (1) continued with terms AND ("age of onset" OR 'age-of-onset' OR 'age at onset' OR 'age-at-onset'). Search (2) continued with: AND ("instability" OR "contraction" OR "further expansion" OR "meiosis" OR "transmission" OR "parent–child"). Search (3) proceeded with: AND ("age at onset" OR "AO" OR "age of onset" OR "anticipation"). (4) AND ("*De novo* mutation" OR "sporadic cases") Search (5) continued with: AND ("fitness" OR "number of children" OR "children" OR "reproduct*"). Search (6) added: AND ("segregation" OR "meiotic drift"). And search (7) added: AND ("haplotype" or "ancestral origin" or "mutational origins" or "common ancestor").

Peer-reviewed articles and meeting abstracts in English were included, and references were checked to guarantee maximal coverage. Three reviewers (LSS, JSP, and AH) independently assessed and extracted data into evidence tables. Any disagreement regarding eligibility was discussed with the other two reviewers (MLSP and LBJ).

2.2 | Population, exposure, comparators, outcomes, and inclusion and exclusion criteria

Studies performed in 10 or more SCA2 subjects were included. The main exposures considered for meta-analysis were carrier status for SCA2 and length of the expanded CAG repeat. Outcomes under study were (1) correlation between age of onset and CAG length; (2) instability of the expanded repeat when crossing meiosis (contractions and/or expansions); (3) differences in AO between different generations (anticipation); (4) *de novo* mutation or sporadic cases; (5) genetic fitness; (6) meiotic segregation; and (7) the ancestral haplotypes described up to date.

The molecular confirmation of SCA2 in symptomatic and/or asymptomatic heterozygotes was a requirement to include studies in all searches. For the search question related to fitness, the case–control design was also an inclusion criterium. The most updated and complete data set was selected if multiple publications from the same study group and/or institutions were detected, in order to avoid gross overrepresentation.

2.3 | Risk of bias assessment and quality control

A priori, three potential causes of heterogeneity between studies that correlated AO with CAG repeat length (search 1) were identified. Bias related to incomplete ascertainment of subjects was partially controlled by the exclusion of studies with less than 10 subjects and with non-significant correlations. Different definitions of the onset of symptoms were controlled and analyzed if appropriate. Finally, since detailed laboratory methodology was lacking in most studies, differences in precision of the CAG measurements due to diverse laboratory protocols were controlled by comparing studies performed before and after the median year between the first and the last studies included in the meta-analysis.

Ascertainment bias was also an important risk for search (2): this risk was assessed by controlling if inclusion of all offspring of a given carrier and balance between parental sexes were respected. Inclusion of all offspring of a given carrier, exclusion of the present generation (to avoid bias towards younger AO), and balance between parental sexes were the criteria used to assess bias on studies retrieved by search 3. A priori, *de novo* mutations or sporadic cases (search 4) are expected to get case reports only or reports among series of general ataxic subjects without a dominant inheritance; both under-detection - including cases with more access to diagnosis – and erroneous allocation due to unknown parenthood are expected to occur and go undetectable. Studies on genetic fitness (search 5) could be prone to allocation and chronological biases; these risks were assessed by the type of controls considered as non-carriers - general population versus relatives -, number of generations under study, and inclusion of all family members. Studies on segregation of alleles (search 6) were considered free from ascertainment bias if all individuals from a given sibship were included and genotyped. Heterogeneity of results was considered low if results differed from 0% to 50%, medium, from 50% to 75%, or high, if above 75%.[15] High heterogeneity of results between studies was considered suggestive of the presence of bias, or methodological discrepancies between studies. If present, re-analysis was performed in order to detect studies with potential problems.

Participation in quality control programs were valued and mentioned, if available. Quality of studies on ancestral haplotypes was also evaluated by the existence of a control group from the original populations. All potential biases and quality parameters were summarized in the Results section (Section 3), if appropriate.

2.4 | Analysis and data synthesis

Descriptive measures of central tendency and dispersion were retrieved from original studies, and described according to the data distribution. Aggregate measures of the expanded CAG repeat length, AO, number of children, and number of segregated gametes for the entire cohorts were obtained to perform the meta-analyses using R Program version 3.6.2 (2019–12-12), package meta version 4.9-7, if two or more studies reached inclusion criteria. Random effects model was chosen in order to avoid effects related to differences between inclusion criteria, sample sizes and/or variances of the studies selected. Summary statistics from aggregated databases were planned to be used for all outcomes under study, with exception of the ancestral haplotypes. A confidence interval of 95% was chosen to attribute significance to results.

3 | RESULTS

A total of 339 papers were obtained after searching databases. Some papers were found in more than one of the seven searches: 47 were selected twice, 16 were selected three times, and four were selected four times. None of the papers mentioned that their institutions participated in quality control programs. After analysis, eligibility was reached in 21 papers on correlation between CAG and AO of the symptoms, six papers and one unpublished data on instability of the expanded repeat, eight on anticipation, 10 articles on *de novo* mutation or sporadic cases, one on fertility, one on segregation distortion, and 11 on ancestral haplotypes. Studies obtained from our seven searches as well as reasons for inclusions or exclusions can be found in Supplemental materials 1 and 2.

3.1 | Correlation between age at onset and CAG length

Twenty-one studies[4,5,16-34] reported on correlations between AO and length of expanded CAG repeats found in SCA2 cohorts (Supplemental Materials 1 and 3).

Ten papers defined AO as the age when the first symptom was noticed by the subject or relatives, three papers defined AO as the onset of gait ataxia, and 11 did not present AO definition. In the unique paper that described both criteria, AO were exactly the same for the first symptom and for gait ataxia of each individual analyzed.[28] Since most SCA2 patients develop gait ataxia as the first symptom, we considered that both concepts were equivalent, allowing all data to be analyzed together. The global linear correlation coefficient between CAG expanded and AO was $r = 0.75$ [0.71-0.79] ($r^2 = 0.577$) in 1406 individuals, meaning that, on average, 57.7% of the AO variability in SCA2 worldwide is determined by the causative mutation. The mean age at which symptoms started was 33.85 (32.14–35.56) years and the mean length of the expanded CAG was 40.62 (39.86–41.37) repeats. Figure 1 summarizes the chronological information from papers included as well as the global results of this meta-analysis. In addition, Figure 2 compares CAG repeat length dispersions with those of the AO for those studies that reported this data.

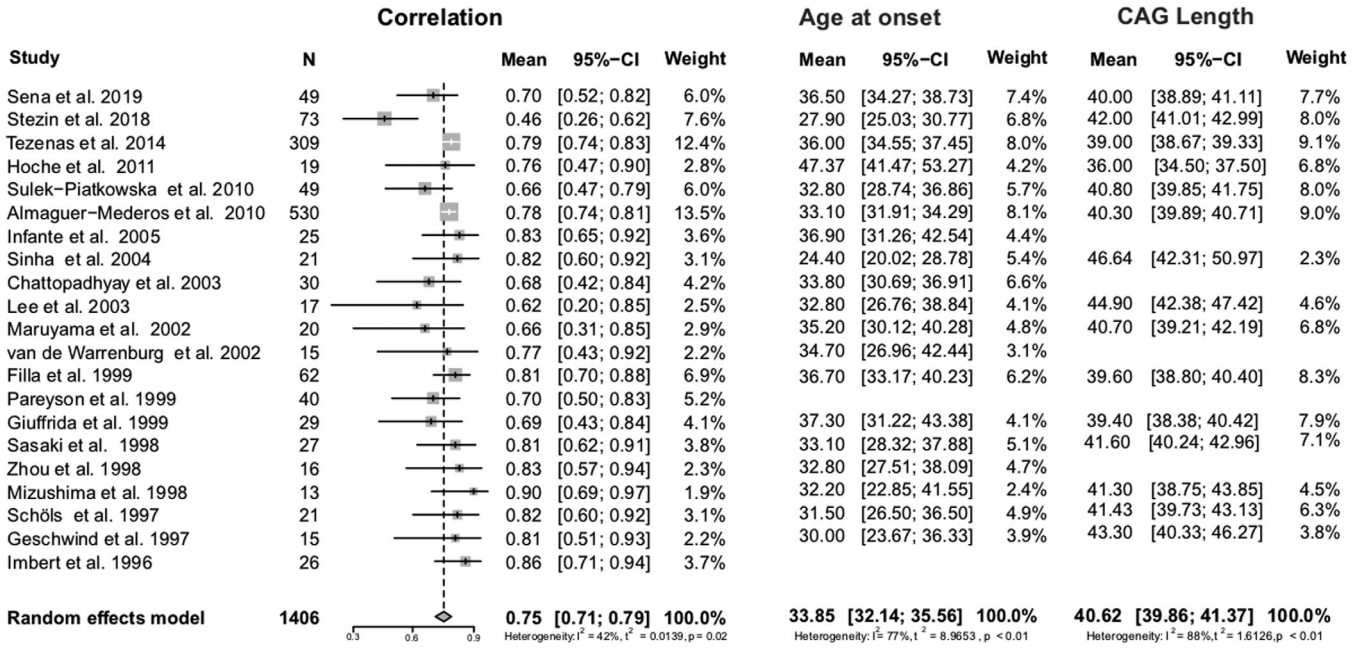

FIGURE 1 Forest plots of the correlation between age at onset (AO) and length of expanded CAG repeats found in SCA2 cohorts. Results with 95% confidence interval were calculated using a random-effects meta-analysis. RE: random-effect; I^2 quantifies heterogeneity; t^2 is magnitude of the between-study variance in our meta-analysis, p for heterogeneity

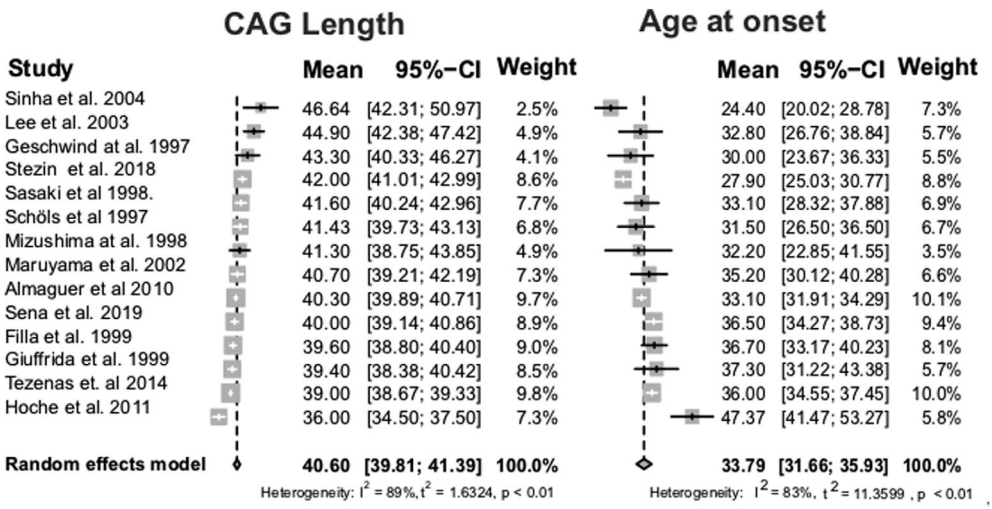

FIGURE 2 Forest plots of the CAG repeat length and age at onset (AO) dispersions for studies that reported these data. The studies were listed in descending order according to the average length of the CAG expansion, compared to their respective ages at onset. Results with 95% confidence interval were calculated using a random-effects meta-analysis. RE: random-effect; I^2 quantifies heterogeneity; t^2 is magnitude of the between-study variance in our meta-analysis, p for heterogeneity

3.2 | Instability

Seven papers[26,34-38] were retrieved to describe potential instabilities of CAG expanded length when transmitted from affected parents to their children (Supplemental Materials 1 and 4). Transmission of the mutant allele, regardless of sex of the affected parent, was associated to an increase of 2.42 CAG repeats in the next generation (Figure 3 (A)). The average increase was larger in paternal (3.84 CAG) than in maternal (1.19 CAG) transmissions (Figure 3(B),(C)).

None of these studies clearly stated that all offspring of a given carrier were included. The papers did not mention that efforts were made to recruit balanced samples of transmitting mothers and fathers. Since the number of maternal transmissions was larger than the paternal ones in all papers included, we assume that samples were not balanced, and may have missed a substantial number of paternal transmissions.

3.3 | Anticipation

Eight papers studied differences between AO of children and their affected parents[16,26,28,33,39-42] (Supplemental Materials 1 and 5). The average AO change between generations was towards an anticipation of 14.62 years per generation. However, the heterogeneity between these results was 78%. After excluding the most discrepant data, heterogeneity between papers was acceptable while the average

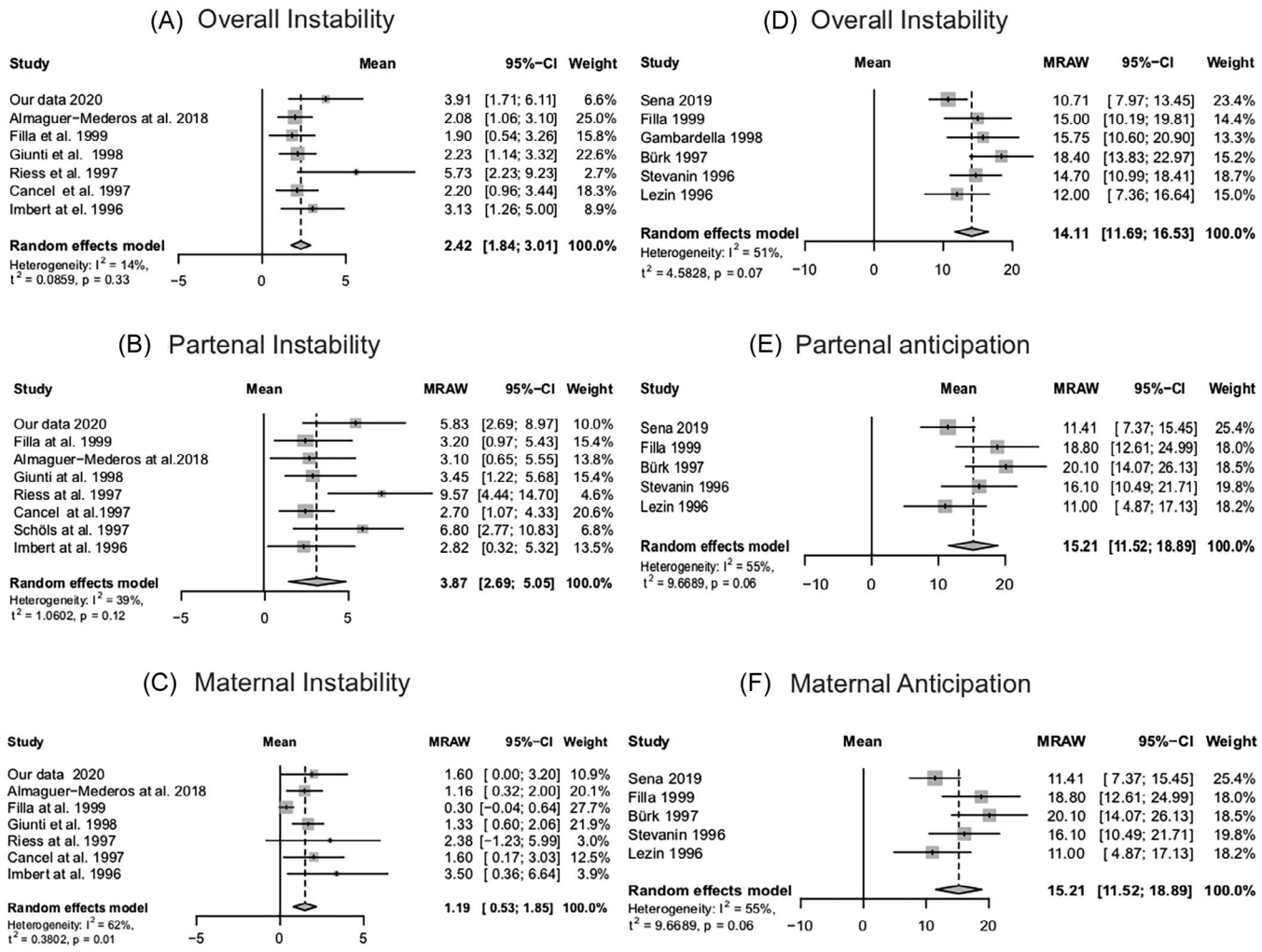

FIGURE 3 Forest plots of instabilities of the CAG expanded repeat length when crossing generations and of anticipation of the AO. (A) Overall instability of the CAG expanded repeat; (B) instabilities observed when paternal transmitted the expansion; (C) instabilities observed in maternal transmitted the expansion; (D) overall anticipation of the age at onset; (E) anticipation observed in paternal transmitted the disease; (F) anticipation observed in maternal transmitted the disease. Results with 95% confidence interval were calculated using a random-effects meta-analysis. RE: random-effect; I^2 quantifies heterogeneity; t^2 is magnitude of the between-study variance in our meta-analysis, p for heterogeneity

anticipation was of 14.11 years per generation (Figure 3(D); see Supplemental Material 6 for the overall results). Paternal and maternal transmissions were associated with anticipations of 15.21 and 11.69 years, respectively (Figure 3(G),(F)).

3.4 | *De novo* expansions and sporadic cases

One *de novo* mutation and 18 sporadic cases were detected from ten papers retrieved from the literature[32,43-51] (Supplemental Materials 1 and 7).

Three papers described *de novo* expansions. However, only one fitted our strict criteria: Futamura et al.[50] described a symptomatic individual with 22/35 CAG repeats at *ATXN2* and onset of symptoms at 36 years of age; his asymptomatic mother and father carried 22/22 and 22/32 repeats, respectively. In a time when the pathogenic cutoffs of the CAG repeats were still to be determined, a seminal paper about instabilities in SCA2 described a symptomatic individual with 41 repeats at *ATXN2* whose deceased and non-ataxic father was discovered to carry 34 repeats[32]—a length in the pathogenic range, nowadays. Finally, Laffita-Mesa et al.[43] described a family with recurrent cases of amyotrophic lateral sclerosis (ALS). One ALS case carrying 22/35 CAG repeats at *ATXN2* was the son of a woman with 22/25 CAG repeats, who were presented as the transmitting parent. However, the lack of data about the father's genotype prevented a definitive characterization of that ALS subject as a *de novo* case, for the present review.

Six papers were characterized as case series of sporadic or isolated cases of ataxia of adult onset - ie, ataxic subjects without a family history of ataxia and without molecular information about their parents[44-46,48,49,51]; however, one study was excluded due to uncertainty about presence of family history.[51] One additional paper was a

TABLE 1 Some biological characteristics of SCA2 and SCA3/MJD that might act or express underlying selective pressures over these diseases

	SCA2	SCA3/MJD	References
Correlation between CAGexp and AO (r^2)	0.577	0.552	Mattos et al. 2018; present data
Did the correlation between CAGexp and AO vary across ethnic groups?	No	Yes	Mattos et al. 2018; present data
CAGexp instability upon meiosis Mean [95% CI]	2.42 [1.84;3.01]	1.23 [0.761.70]	Sena et al. 2021; present data
Anticipation Mean [95% CI]	14.62 [11.59;17.65]	7.75 [6. 63-8.88]	Sena et al. 2021; present data
de novo expansions reported in the literature	Yes	No	Sena et al. 2021; present data
Ancestral haplotypes identified so far	Undetermined	At least seven	Sena et al. 2021; present data
Fitness	1.50	1.45	Prestes et al. 2008; Sena et al 2019
Segregation of the expanded allele	40.4%	64.0%	Sena et al. 2019 and 2021

Abbreviations: AO, age at onset of symptoms; CAGexp, the expanded CAG repeat; CI, confidence interval.

case series on sporadic Parkinson disease patients.[47] The number of sporadic SCA2 subjects per sporadic ataxic subjects studied were 1/20 in India, [48] 1/39 in Korea,[49] 6/123 in Mexico,[46] 1/15 in Japan,[45] 9/237 in China,[44] and 1/242 PD in Taiwan.[47]

3.5 | Fitness

One study[16] estimated fitness of SCA2 carriers (Supplemental material 1). Number of children of symptomatic and asymptomatic relatives older than 65.7 years (2 SD from the mean AO of the 164 subjects from that cohort) were compared. The median number of children of the non-carriers and carriers included in the reproductive success analysis were 2 and 3 ($p < 0.025$), respectively; fitness of carriers was 1.5.

3.6 | Segregation distortion

Literature search on segregation distortion included one paper[16] only (Supplemental material 1). The non-carrier status was defined as the absence of ataxic symptoms in subjects older than 65.7 years (2 SD from the mean AO of the 164 subjects from that cohort). One hundred thirty-nine sibs, children of symptomatic subjects, met this criteria and were included in the segregation analysis: 56 (40.4%) were affected, while the remaining 83 (59.6%) were unaffected. The $(CAG)_{22}$ allele corresponded to 78% of the 51 normal alleles analyzed. Segregation distortion favoring the normal allele was more evident in women: among 62 daughters of affected parents, 21 (33.86%) were symptomatic (carriers) and 41 (66.14%) were healthy (non-carriers) ($p = 0.011$).

3.7 | Haplotypes

Eleven papers[21,36,52-60] (Supplemental Materials 1 and 8) studied *ATXN2* haplotypes associated with SCA2. Three SNPs and 18 STRs were used to construct haplotypes in different combinations across the literature. SCA2 alleles were universally associated with C allele at rs695871 and C allele at rs695872.[54,55] The most used STRs were D12S1332, D12S1672, D12S1333, but different codings were used between papers, and in most of them the primers employed were not described. Considering that a common set of SNPs and STRs markers has not been chosen by the authors of original papers, meta-analysis was not possible.

4 | DISCUSSION

There are still many weaknesses affecting the knowledge of forces that interact in the maintenance of CAG expansions in *ATXN2* in human populations, such as recruitment bias, lack of standards in haplotype studies, and scarcity of studies. Nevertheless, the documented data is sufficient to state that peculiar selective pressures act on the expanded allele at *ATXN2*. On the one hand, further expansions of the expanded tract with each transmission were prevalent, although contractions may not have been observed due to recruitment bias. The absolutely unusual occurrence of *de novo* SCA2 cases suggests that contractions prior to these cases are in fact extremely rare events. If the general trend is to anticipate the onset of symptoms, then extinction of SCA2 lineages would be expected. Though, the variety of ancestral haplotypes points to the existence of multiple ancestral origins. These two different groups of data suggest that SCA2, as a neurological phenotype, can appear, disappear, and then reappear in populations. The study of ancestral haplotypes may clarify, in the future, whether sporadic cases represent novel lineages or the return of positive phenotypes in previous contracted lineages. Finally, studies on the effects of *ATXN2* CAG expansion on reproduction—a sphere independent of neuronal damage associated with SCA2—are still very rare. Evidence that fitness and allele segregation are influenced by the presence of CAG expansion requires confirmation.

We have followed, on purpose, the same approach as the systematic review carried out on another polyQ ataxia, SCA3/MJD.[61] Our intention was to compare the evolutionary destinies of different polyQ diseases by following the same systematic questions. Similarities would indicate common mechanisms, linked to the polyQ tracts themselves, while differences would point out to the role of the original protein (Table 1).

The correlations between the expanded CAG at *ATXN2* and AO varied widely between studies, with r between 0.46 and 0.86 (Figure 1). However, there are many studies with a small number of patients, with probable recruitment bias. In fact, the data suggest more similarities than differences between different ethnic origins: there is an almost perfect reciprocity between variations in CAG repeats lengths versus variations in AO, in the case series represented in Figure 2. In other words, the distributions of AO correspond to those predicted by the expanded CAG distributions, regardless of the origin of subjects. The same is not true for SCA3/MJD (Table 1). In this disease, the effect of CAG expanded allele on AO depends upon the origin of population,[62] which suggests the occurrence of *ATXN3* internal modifiers that vary according to different geographical areas and likely founding effects. In contrast, Figure 2 did not detect a SCA2 population where the effect of expanded CAG on the SCA2 AO is different from that of other SCA2 populations. Therefore, there is no suggestion that other variations within *ATXN2*, present in one geographic group but not in others, would be influencing AO. The remaining 42.3% of AO variation in SCA2, independent of the CAG expanded length, should probably be sought in other genetic or environmental factors.

Recruitment biases were not clearly avoided in studies on instability and anticipation in SCA2, as observed previously in SCA3/MJD.[61] However, it is important to note that the main recruitment problems would tend to operate in opposite directions. The potential bias of not including offspring with shorter expansions (still asymptomatic) would distort instabilities towards large expansions. The bias of including more data on transmitting mothers than fathers would reduce the average instability observed. In any case, the instabilities of the expanded CAG in SCA2 were on average much more intense than those detected in SCA3/MJD (Table 1).

Among 19 sporadic SCA2 cases, only one was categorically documented as carrying a *de novo* expansion, originating from a large normal allele. In other sporadic cases, the late onset probably prevented the family from being accurately investigated. Intermediate alleles or normal alleles prone to instability would be one of the most important factors explaining the maintenance of the expanded allele in the population.[63] In any case, this finding diverges from the total absence of reports of novel SCA3/MJD cases (Table 1).

SCA7 is another polyQ SCA in which there are clear records of relevant instabilities. Five *de novo* SCA7 cases[64-66] and 28 SCA7 cases with onset at childhood, six of them inherited from asymptomatic transmitting parents[67] have been reported. SCA7 also resembles SCA2 in the length of pathogenic repeats. Expanded CAG tracts are more than 33 and 38 repeats long in the genes that cause SCA2 and SCA7, respectively. But the literature suggests that the sources of the novel cases - confirmed or suspected (sporadic) - may be different in SCA2 and SCA7. Contractions in the expanded CAG may have occurred in pedigrees, several generations before reappearance of SCA2 as a sporadic case. The heredogram presented by Lafita-Mesa et al seems to illustrate this phenomenon.[43] In contrast, although there is some debate about common ancestral haplotypes,[68] the cases of new SCA7 described to date appear to have been associated with a variety of independent haplotypes. In summary, we cannot rule out that molecular characteristics—intragenic or in the protein—common to SCAs 2 and 7 may explain both the range in which repetitive sequences are pathogenic as well as expansion instabilities associated either with severe anticipations or with sporadic cases of these diseases. On the contrary, the population dynamics of SCA2 and SCA7 differs from that of SCA3/MJD, where the expanded CAG is longer, there are very few ancestral lineages, and no reported novel cases to date.

Studies on the fitness of polyQs have shown mixed results. Fitness in SCA1 was estimated to be increased[69] or unchanged[70] while increased fitness was reported in SCA3/MJD[71] and in SCA2.[16] Studies on HD have described increased fitness[69] in one study and unchanged fitness in another,[72] but the latter was performed prior to HD gene discovery. In other words, there are very few studies, or almost a single study on each polyQ disease, on this topic. Further observational studies would be essential to assist the comprehension whether polyQ tracts, regardless of the protein they are inserted in, would have a general effect on fertility.

Our study confirmed that there is a slight meiotic drift favoring the normal *ATXN2* allele. This finding supports the hypothesis that this gene may indeed influence gametogenesis or early intrauterine life. Interestingly, this effect was the opposite of that observed in SCA3/MJD, where expanded alleles are favored in segregation.[61] The different directions that the segregation distortion takes place between these two diseases may mean that the polyQs in ataxin-2 and ataxin-3 would not be the real segregation modifiers (Table 1). They could be tags only, being linked to variants that are actually functional in ataxin-2 and ataxin-3. In SCA2, the distorted segregation may not be exactly in favor of normal alleles but may be in favor of the 22-repeats allele, the most common normal and almost universally present in the genotypes of SCA cases. This finding converges with a study that showed that the 22-repeats allele in the following sequence $(CAG)_8CAA(CAG)_4CAA(CAG)_8$ is undergoing positive selection.[73] Of note, the CAA interruptions seem to provide stability to CAG repeat.[55] Thus, stability plus segregation distortion might help explain why this 22-repeats allele is so common in human populations.

Finally, we need to point out that the search performed was limited to PubMed as well as to papers published in English. Those limitations may have had an impact on the included studies, which can be overcome by future, complementary meta-analyzes on the topic. However, the entire database is available to the community, as supplementary material, and data from other databases or from papers published in other languages may then be included.

According to our analysis of the present data, it is likely that for each polyQ disease, different biological conditions would lead to

different evolutionary pressures. After all, selective pressures on SCA2 were shown to be quite diverse from those observed in SCA3/MJD (Table 1). In this sense, SCA2 and SCA3/MJD can serve as two different evolutionary models for polyQs. SCA2 could represent a group of polyQs caused by pathogenic expansions from 34-40 CAG repeats that are characterized by severe anticipations with lineages extinctions, and novel cases with multiple ancestral origins, such as SCA7. Further and more comprehensive studies on instabilities, fitness, segregation of alleles, and ancestral haplotypes should shed light upon selective pressures of SCA2.

ACKNOWLEDGEMENTS

This study was supported by Fundo de Incentivo à Pesquisa do Hospital de Clínicas de Porto Alegre (FIPE-HCPA), grant number 2019-0169. JSP was supported by Fundação de Amparo à Pesquisa do Rio Grande do Sul (FAPERGS), grant number 126984/2019-0. LSS, MLSP and LBJ were supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, grant numbers 141781/2018-1, 313132/2018-6 and 303577/2016-19, respectively.

CONFLICT OF INTEREST

Authors report there are no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/cge.13978.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available as Supplemental Materials related to the present paper.

ORCID

Laura Bannach Jardim https://orcid.org/0000-0001-6907-5068

REFERENCES

1. Pulst SM. Spinocerebellar ataxia type 2. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews® [Internet]*. Seattle (WA): University of Washington, Seattle; 1993-2020.
2. Matilla-Duenãs A, Ashizawa T, Brice A, et al. Consensus paper: pathological mechanisms underlying neurodegeneration in spinocerebellar ataxias. *Cerebellum*. 2014;13:269-302.
3. Paulson HL, Shakkottai VG, Clark HB, Orr HT. Polyglutamine spinocerebellar ataxias - from genes to potential treatments. *Nat Rev Neurosci*. 2017;18:613-626.
4. Magaña JJ, Velázquez-Pérez L, Cisneros B. Spinocerebellar ataxia type 2: clinical presentation, molecular mechanisms, and therapeutic perspectives. *Mol Neurobiol*. 2013;47:90-104.
5. Tezenas du Montcel S, Durr A, Bauer P, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain*. 2014;137:2444-2455.
6. Figueroa KP, Coon H, Santos N, et al. Genetic analysis of age at onset variation in spinocerebellar ataxia type 2. *Neurol Genet*. 2017;3:e155.
7. Pereira FS, Monte TL, Locks-Coelho LD, et al. ATXN3, ATXN7, CACNA1A, and RAI1 genes and mitochondrial polymorphism A10398G did not modify age at onset in Spinocerebellar ataxia type 2 patients from South America. *Cerebellum*. 2015;14:728-730.
8. Monte TL, Reckziegel EDR, Augustin MC, et al. The progression rate of spinocerebellar ataxia type 2 changes with stage of disease. *Orphanet J Rare Dis*. 2018;13:20.
9. Diallo A, Jacobi H, Tezenas du Montcel S, Klockgether T. Natural history of most common spinocerebellar ataxia: a systematic review and meta-analysis. *J Neurol*. 2020; Epub ahead of print.
10. Jacobi H, du Montcel ST, Bauer P, et al. Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6: a longitudinal cohort study. *Lancet Neurol*. 2015;14:1101-1108.
11. Monin ML, Tezenas du Montcel S, Marelli C, et al. Survival and severity in dominant cerebellar ataxias. *Ann Clin Transl Neurol*. 2015;2:202-207.
12. Diallo A, Jacobi H, Cook A, et al. Survival in patients with spinocerebellar ataxia types 1, 2, 3, and 6 (EUROSCA): a longitudinal cohort study. *Lancet Neurol*. 2018;17:327-334.
13. 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;79:383-387.
14. Di Fabio R, Santorelli F, Bertini E, et al. Infantile childhood onset of spinocerebellar ataxia type 2. *Cerebellum*. 2012;11:526-530.
15. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-560.
16. Sena LS, Castilhos RM, Mattos EP, et al. Selective forces related to Spinocerebellar ataxia type 2. *Cerebellum*. 2019;18:188-194.
17. Stezin A, Venkatesh SD, Thennarasu K, et al. Non-ataxic manifestations of Spinocerebellar ataxia-2, their determinants and predictors. *J Neurol Sci*. 2018;394:14-18.
18. Hoche F, Balikó L, den Dunnen W, et al. Spinocerebellar ataxia type 2 (SCA2): identification of early brain degeneration in one monozygous twin in the initial disease stage. *Cerebellum*. 2011;10:245-253.
19. Sułek-Piatkowska A, Zdzienicka E, Raczyńska-Rakowicz M, et al. The occurrence of spinocerebellar ataxias caused by dynamic mutations in polish patients. *Neurol Neurochir Pol*. 2010;44:238-245.
20. Infante J, Combarros O, Volpini V, Corral J, Llorca J, Berciano J. Autosomal dominant cerebellar ataxias in Spain: molecular and clinical correlations, prevalence estimation and survival analysis. *Acta Neurol Scand*. 2005;111:391-399.
21. Sinha KK, Worth PF, Jha DK, et al. Autosomal dominant cerebellar ataxia: SCA2 is the most frequent mutation in eastern India. *J Neurol Neurosurg Psychiatry*. 2004;75:448-452.
22. Chattopadhyay B, Ghosh S, Gangopadhyay PK, et al. Modulation of age at onset in Huntington's disease and spinocerebellar ataxia type 2 patients originated from eastern India. *Neurosci Lett*. 2003;345:93-96.
23. Lee WY, Jin DK, Oh MR, et al. Frequency analysis and clinical characterization of spinocerebellar ataxia types 1, 2, 3, 6, and 7 in Korean patients. *Arch Neurol*. 2003;60:858-863.
24. Maruyama H, Izumi Y, Morino H, et al. Difference in disease-free survival curve and regional distribution according to subtype of spinocerebellar ataxia: a study of 1,286 Japanese patients. *Am J Med Genet*. 2002;114:578-583.
25. van de Warrenburg BP, Sinke RJ, Verschuuren-Bemelmans CC, et al. Spinocerebellar ataxias in The Netherlands: prevalence and age at onset variance analysis. *Neurology*. 2002;58:702-708.
26. Filla A, De Michele G, Santoro L, et al. Spinocerebellar ataxia type 2 in southern Italy: a clinical and molecular study of 30 families. *J Neurol*. 1999;246:467-471.
27. Pareyson D, Gellera C, Castellotti B, et al. Clinical and molecular studies of 73 Italian families with autosomal dominant cerebellar ataxia type I: SCA1 and SCA2 are the most common genotypes. *J Neurol*. 1999;246:389-393.

28. Giuffrida S, Lanza S, Restivo DA, et al. Clinical and molecular analysis of 11 Sicilian SCA2 families: influence of gender on age at onset. *Eur J Neurol*. 1999;6:301-307.
29. Sasaki H, Wakisaka A, Sanpei K, et al. Phenotype variation correlates with CAG repeat length in SCA2—a study of 28 Japanese patients. *J Neurol Sci*. 1998;159:202-208.
30. Zhou YX, Wang GX, Tang BS, et al. Spinocerebellar ataxia type 2 in China: molecular analysis and genotype-phenotype correlation in nine families. *Neurology*. 1998;51:595-598.
31. Mizushima K, Watanabe M, Abe K, et al. Analysis of spinocerebellar ataxia type 2 in Gunma prefecture in Japan: CAG trinucleotide expansion and clinical characteristics. *J Neurol Sci*. 1998;156:180-185.
32. Schöls L, Gispert S, Vorgerd M, et al. Spinocerebellar ataxia type 2: genotype and phenotype in German Kindreds. *Arch Neurol*. 1997;54:1073-1080.
33. Geschwind DH, Perlman S, Figueroa CP, Treiman LJ, Pulst SM. 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;60:842-850.
34. Imbert G, Saudou F, Yvert G, et al. Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat Genet*. 1996;14:285-291.
35. Almaguer-Mederos LE, Mesa JML, González-Zaldívar Y, et al. Factors associated with ATXN2 CAG/CAA repeat intergenerational instability in Spinocerebellar ataxia type 2. *Clin Genet*. 2018;94:346-350.
36. Giunti P, Sabbadini G, Sweeney MG, et al. The role of the SCA2 trinucleotide repeat expansion in 89 autosomal dominant cerebellar ataxia families. Frequency, clinical and genetic correlates. *Brain*. 1998;121:459-467.
37. Riess O, Laccone FA, Gispert S, et al. SCA2 trinucleotide expansion in German SCA patients. *Neurogenetics*. 1997;1:59-64.
38. Cancel G, Dürr A, Didierjean O, et al. Molecular and clinical correlations in spinocerebellar ataxia 2: a study of 32 families. *Hum Mol Genet*. 1997;6:709-715.
39. Bürk K, Stevanin G, Didierjean O, et al. Clinical and genetic analysis of three German kindreds with autosomal dominant cerebellar ataxia type I linked to the SCA2 locus. *J Neurol*. 1997;244:256-261.
40. Gambardella A, Annesi G, Bono F, et al. CAG repeat length and clinical features in three Italian families with spinocerebellar ataxia type 2 (SCA2): early impairment of Wisconsin card sorting test and saccade velocity. *J Neurol*. 1998;245:647-652.
41. Stevanin G, Trottier Y, Cancel G, et al. Screening for proteins with polyglutamine expansions in autosomal dominant cerebellar ataxias. *Hum Mol Genet*. 1996;5:1887-1892.
42. Lezin A, Cancel G, Stevanin G, et al. Autosomal dominant cerebellar ataxia type I in Martinique (French West Indies): genetic analysis of three unrelated SCA2 families. *Hum Genet*. 1996;97:671-676.
43. Laffita-Mesa JM, Rodríguez Pupo JM, Moreno Sera R, et al. De novo mutations in ataxin-2 gene and ALS risk. *PLoS One*. 2013;8:e70560.
44. Wang J, Shen L, Lei L, et al. Spinocerebellar ataxias in mainland China: an updated genetic analysis among a large cohort of familial and sporadic cases. *Zhong Nan Da Xue Xue Bao Yi Xue Ban*. 2011;36:482-489.
45. Yamamoto-Watanabe Y, Watanabe M, Hikichi M, et al. Prevalence of autosomal dominant cerebellar ataxia in Aomori, the northernmost prefecture of Honshu, Japan. *Intern Med*. 2010;49:2409-2414.
46. Alonso E, Martínez-Ruano L, De Biase I, et al. Distinct distribution of autosomal dominant spinocerebellar ataxia in the Mexican population. *Mov Disord*. 2007;22:1050-1053.
47. Shan DE, Liu RS, Sun CM, Lee SJ, Liao KK, Soong BW. Presence of spinocerebellar ataxia type 2 gene mutation in a patient with apparently sporadic Parkinson's disease: clinical implications. *Mov Disord*. 2004;19:1357-1360.
48. Chakravarty A, Mukherjee SC. Primary degenerative cerebellar ataxias in ethnic Bengalees in West Bengal: some observations. *Neurol India*. 2003;51:227-234.
49. Kim JY, Park SS, Joo SI, Kim JM, Jeon BS. Molecular analysis of spinocerebellar ataxias in Koreans: frequencies and reference ranges of SCA1, SCA2, SCA3, SCA6, and SCA7. *Mol Cells*. 2001;12:336-341.
50. Futamura N, Matsumura R, Fujimoto Y, Horikawa H, Suzumura A, Takayanagi T. CAG repeat expansions in patients with sporadic cerebellar ataxia. *Acta Neurol Scand*. 1998;98:55-59.
51. Leggo J, Dalton A, Morrison PJ, et al. Analysis of spinocerebellar ataxia types 1, 2, 3, and 6, dentatorubral-pallidoluysian atrophy, and Friedreich's ataxia genes in spinocerebellar ataxia patients in the UK. *J Med Genet*. 1997;34:982-985.
52. Wang C, Xu Y, Feng X, et al. Linkage analysis and whole-exome sequencing exclude extra mutations responsible for the parkinsonian phenotype of spinocerebellar ataxia-2. *Neurobiol Aging*. 2015;36:545.
53. Laffita-Mesa JM, Velázquez-Pérez LC, Santos Falcón N, et al. Unexpanded and intermediate CAG polymorphisms at the SCA2 locus (ATXN2) in the Cuban population: evidence about the origin of expanded SCA2 alleles. *Eur J Hum Genet*. 2012;20:41-49.
54. Ramos EM, Martins S, Alonso I, et al. Common origin of pure and interrupted repeat expansions in spinocerebellar ataxia type 2 (SCA2). *Am J Med Genet B Neuropsychiatr Genet*. 2010;153B:524-531.
55. Choudhry S, Mukerji M, Srivastava AK, Jain S, Brahmachari SK. CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. *Hum Mol Genet*. 2001;10:2437-2446.
56. Saleem Q, Choudhry S, Mukerji M, et al. Molecular analysis of autosomal dominant hereditary ataxias in the Indian population: high frequency of SCA2 and evidence for a common founder mutation. *Hum Genet*. 2000;106:179-187.
57. 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;7:841-845.
58. Didierjean O, Cancel G, Stevanin G, et al. Linkage disequilibrium at the SCA2 locus. *J Med Genet*. 1999;36:415-417.
59. Mizushima K, Watanabe M, Kondo I, et al. Analysis of spinocerebellar ataxia type 2 gene and haplotype analysis: (CCG)1-2 polymorphism and contribution to founder effect. *J Med Genet*. 1999;36:112-114.
60. Nechiporuk A, Lopes-Cendes I, Nechiporuk T, et al. Genetic mapping of the spinocerebellar ataxia type 2 gene on human chromosome 12. *Neurology*. 1996;46:1731-1735.
61. Sena LS, Pinheiro JS, Saraiva-Pereira ML, Jardim LB. Selective forces acting on spinocerebellar ataxia type 3/Machado-Joseph disease recurrency: a systematic review and meta-analysis. *Clin Genet*. 2021;99:347-358.
62. de Mattos EP, Leotti VB, Soong BW, et al. Age at onset prediction in spinocerebellar ataxia type 3 changes according to population of origin. *Eur J Neurol*. 2019;26:113-120.
63. Gardiner SL, Boogaard MW, Trompet S, et al. Prevalence of carriers of intermediate and pathological polyglutamine disease-associated alleles among large population-based cohorts. *JAMA Neurol*. 2019;76:650-656.
64. Stevanin G, David G, Dürr A, et al. Multiple origins of the spinocerebellar ataxia 7 (SCA7) mutation revealed by linkage disequilibrium studies with closely flanking markers, including an intragenic polymorphism (G3145TG/A3145TG). *Eur J Hum Genet*. 1999;7:889-896.
65. Mittal U, Roy S, Jain S, Srivastava AK, Mukerji M. Post-zygotic de novo trinucleotide repeat expansion at spinocerebellar ataxia type 7 locus: evidence from an Indian family. *J Hum Genet*. 2005;50:155-157.
66. Italiano D, Tarantino P, De Marco EV, et al. Spinocerebellar ataxia type 7: report of a new Italian family. *Intern Med*. 2012;51:2953-2955.
67. Bah MG, Rodriguez D, Cazeneuve C, et al. Deciphering the natural history of SCA7 in children. *Eur J Neurol*. 2020;27:2267-2276.

68. Faruq M, Magaña JJ, Suroliya V, et al. Complete association of an intronic SNP rs6798742 with origin of spinocerebellar ataxia type 7-CAG expansion loci in the Indian and Mexican population. *Ann Hum Genet*. 2017;81:197-204.
69. 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. *Ann Hum Genet*. 1996;60:423-435.
70. 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;17:179-185.
71. 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. *Ann Hum Genet*. 2008;72:57-64.
72. Mastromauro CA, Meissen GJ, Cupples LA, Kiely DK, Berkman B, Myers RH. Estimation of fertility and fitness in Huntington disease in New England. *Am J Med Genet*. 1989;33:248-254.
73. Yu F, Sabeti PC, Hardenbol P, et al. Positive selection of a pre-expansion CAG repeat of the human SCA2 gene. *PLoS Genet*. 2005;1:e41.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Sena LS, dos Santos Pinheiro J, Hasan A, Saraiva-Pereira ML, Jardim LB. Spinocerebellar ataxia type 2 from an evolutionary perspective: Systematic review and meta-analysis. *Clinical Genetics*. 2021;1–10. https://doi.org/10.1111/cge.13978

CAPÍTULO IV: DINÂMICA DOS ALELOS EXPANDIDOS DO *ATXN2* E *ATXN3*

Manuscrito 3: A model for the dynamics of expanded CAG repeat alleles: *ATXN2* and *ATXN3* as prototypes.

Artigo preparado para ser submetido no periódico

Genetics

CAPÍTULO V. HAPLÓTIPOS ANCESTRAIS DAS LINHAGENS SCA2 NO BRASIL, NO PERU E NO URUGUAI

Manuscrito: "*Ancestral Haplotypes of Spinocerebellar Ataxia Type 2 in Brazil, Peru and Uruguay*"

Artigo em preparação para ser submetido no periódico

Frontiers in Genetics

CAPÍTULO VI. CONCLUSÕES, LIMITES E PERSPECTIVAS

A abordagem evolutiva apresentada nesta tese buscou compreender a manutenção dos alelos expandidos nos genes *ATXN2* e *ATXN3.* Revisei de forma sistemática as evidências disponíveis sobre as forças seletivas que atuam sobre os alelos expandidos desses *loci*, quando possível agregando os resultados através de metanálises. Depois, propus uma adaptação de uma equação clássica da genética de populações para estimar a dinâmica de alelos CAGexp. As simulações realizadas utilizando os resultados das metanálises prévias estimaram permanências de linhagens compatíveis com os dados empíricos obtidos de estudos observacionais feitos ao redor do mundo sobre a SCA2 e a SCA3/MJD. A exceção que permanecia a descoberto era a previsão, a partir do meu modelo, de que a SCA2 tivesse múltiplas origens. A detecção de ao menos seis haplótipos SCA2 nas famílias sul-americanas estudadas na presente tese não só expandiu o número de origens ancestrais identificadas até hoje para essa condição, como confirmou as previsões do meu modelo. Houve, portanto, uma convergência entre as forças seletivas associadas a cada gene, os prognósticos obtidos nas simulações e os nossos resultados nos estudos de haplótipos SCA2 e com os da literatura na SCA3/MJD.

A tese foi estruturada para abordar antes a SCA3/MJD e depois a SCA2, pois as evidências empíricas existentes para a SCA3/MJD eram mais numerosas e robustas do que as existentes para a SCA2. Agora eu modificarei a ordem do debate. Começarei com a SCA2, pois foi nela que justamente nossas observações empíricas trouxeram novidades substanciais.

De uma forma geral, podemos inferir que a dinâmica da SCA2 passa por linhagens que, apesar do aumento de fertilidade do seus portadores, serão extintas devido à distorção da segregação e à intensidade da antecipação a cada geração. Todas as 1000 simulações realizadas com *ATXN2* expandidos foram extintas em uma mediana de 10 gerações (figura 3B, Capítulo IV). Segundo esses resultados, o modelo previu que a aparente manutenção da doença nas populações dependeria de expansões *de novo* geradas a partir de sequências CAG normais

instáveis. Nossos resultados obtidos no capítulo V a partir dos alelos não expandidos estavam de acordo com essa previsão teórica. O tamanho das repetições CAG normais foi maior no grupo com o haplótipo G-C-C, com uma média (sd) de 27,45(5,317), comparado dos haplótipos A-G-C e G-C-T que tiveram respectivamente tamanhos médios de 22,08 (0,577) e 22,44(1,933). Os múltiplos haplótipos ancestrais para a SCA2 que encontramos nas famílias provenientes da América do Sul convergiram com essa afirmação.

Além disso, mecanismos evolutivos também podem ajudar a entender a alta prevalência do CAG_{22} no *ATXN2* na população. Vimos que alelos não expandidos (majoritariamente 22 CAG) teriam vantagem na segregação alélica contra os alelos expandidos. Isso combinado com evidências de que o padrão de interrupções CAA nos alelos com CAG_{22} $(CAG)_8CAA$ $(CAG)_4CAA(CAG)_8$ parece lhe conferir estabilidade, seriam forças importantes não só a manter alelos com 22 repetições a seguir com o mesmo tamanho através das gerações, como a o tornarem tão comuns.

Não obstante, essa inferência **deixa três lacunas importantes** que precisam ser investigadas e que mencionarei a seguir.

No que tange à distorção na segregação, não está claro se a distorção ocorre a favor de qualquer alelo não expandido ou especificamente do alelo com 22 repetições. Estudos de distorções na segregação em não portadores de SCA2, assim como os realizados na SCA3 (Rubinsztein et al. 1997, MacMillan et al. 1999, Bettencourt et al. 2008) e DRPLA (MacMillan et al. 1999) podem responder essa questão. No entanto, a elevada frequência do alelo com 22 repetições faz com que a heterozigosidade nesse alelo seja de 0,253, assim, recrutar indivíduos para esse estudo é o passo mais complexo, porém necessário para responder essa questão.

Apesar de inferirmos que as linhagens SCA2 serão extintas e que sua detecção nas populações seja explicada por mutações *de novo*, uma segunda lacuna se evidencia: não sabemos se de fato o *ATXN2* está em equilíbrio mutação seleção.

O equilíbrio mutação-seleção foi utilizado para explicar a manutenção da HD na população (Reed & Neel. 1959). Naquela década, a ausência dos conhecimentos moleculares causadores de doenças (nem o gene causador tinha sido descoberto) dificultavam que as explicações sobre a manutenção da doença fossem mais sólidas. No entanto, décadas depois a descoberta de haplogrupos associados a repetições CAG intermediárias instáveis convergiram com a explicação de que mutações de novo no *HTT* contribuíriam para explicar a frequência da HD.

Alguns fatores possibilitaram que os resultados do estudo de 1959 fossem obtidos. A HD é uma doença caracterizada desde 1872 a partir de aspectos clínicos muito peculiares (Huntington. 1872); o fenótipo per se possibilitou existirem estudos que analisassem a variação de sua frequência através das gerações. Em segundo lugar, a HD é a mais prevalente das doenças por poliQs, e especialmente na Europa e Estados Unidos ela tem recebido grande atenção científica. E em terceiro lugar, e muito importantemente, hoje se sabe com certeza de que muitos (10%, no hemisfério norte) dos casos da HD são gerados a partir de mutação *de novo* (Caron et al. 2020).

No nosso caso, para responder se o *ATXN2* está em equilíbrio mutação seleção, precisaríamos saber pelo menos um dos seguintes elementos: (1) a variação da frequência alélica do *ATXN2* expandido através das gerações. Porém, não existem estudos de prevalência longitudinais na SCA2. E (2) a taxa de mutação da CAGexp no *ATXN2*. Em estudos de datação dos haplótipos do *ATXN3*, a taxa mutacional dos short trinucleotide repeats (STRs) marcadores foi inferida como sendo de $6.13X10^{-4}$ mutações por meiose (Martins et al. 2007). No entanto, não podemos usar essa inferência para generalizar uma taxa de mutação que nos aponte de quanto em quanto tempo ocorreram mutações *de novo* no *ATXN2.* Primeiro, pelo fato das taxas mutacionais se elevarem nos STRs maiores (Brinkmann et al. 1998). Teríamos taxas mutacionais distintas nas repetições CAG do *ATXN2* para alelos normais, intermediários e expandidos; tal como na HD, onde o *HTT*exp tende a ser mais instável quão maior ele mesmo for (quão maior o CAGexp for) (Wheeler et al. 2007). O segundo fator se refere ao *ATXN2* normal ter as repetições CAG com a menor variância entre os genes associados às doenças

poliQs, sugerindo que as repetições CAG no *ATXN2* tendem a ser conservadas possivelmente pelo padrão de interrupções CAA nas repetições CAG. Assim, estudos observacionais sobre a transmissão alélica do *ATXN2* são necessários para estimarmos uma taxa de mutações *de novo* na SCA2.

Já as simulações feitas para a SCA3/MJD apontaram para resultados bastante distintos dos obtidos nas simulações para SCA2. Das 1000 linhagens simuladas a partir do *ATXN3* expandido, 35,1% não foram extintas até a 650 geração. Além disso, o número de gerações em que as linhagens foram eliminadas foi de 26 gerações (figura 5B, Capítulo IV). Esses resultados sugerem que linhagens SCA3/MJD possam se manter por longos períodos e convergem com as datações de haplótipos do *ATXN3* expandido.

Os resultados da modelagem feita para a SCA3/MJD podem ser explicados pela existência de duas forças seletivas favoráveis ao alelo *ATXN3* expandido, além de uma antecipação menos intensa, quando comparados ao alelo *ATXN2* expandido: fitness e segregação não mendeliana. Características específicas dos dois genes *ATXN2 e ATXN3* e/ou de regiões circunvizinhas, ou de suas funções fisiológicas, devem estar por trás dessas diferenças substanciais nas dinâmicas previstas. A convergência com a datação proposta para os haplótipos ancestrais SCA3/MJD e a não descrição de mutações *de novo* no *ATXN3* reforçam a hipótese da manutenção do *ATXN3* expandido se dar por forças seletivas compensatórias.

A não descrição de mutações *de novo* na SCA3/MJD pode estar em grande parte associada à raridade de alelos não expandidos grandes, caracterizada pelo *gap* entre os maiores alelos não expandidos e os alelos causadores da doença. São alelos intermediários a principal fonte de mutações *de novo* em outra importante doença de poliQs, a HD (Myers et al. 1993; Semaka et al. 2013). Um dos aspectos que podem explicar esse *gap* alélico seria o de que entre indivíduos não portadores de SCA3/MJD a transmissão alélica favorece o alelo menor (ver capítulo II figura 3D), enquanto nos portadores da SCA3/MJD o alelo expandido é transmitido preferencialmente. Esse fenômeno faz com que o tamanho das repetições CAG no *ATXN3* sofra uma seleção disruptiva. Um processo similar foi encontrado numa outra doença de poliQs, a DRPLA, doença para a qual não existem relatos de mutação *de*

novo, e em cujo gene *ATN1* o padrão de repetições CAG (Gardiner et al. 2019) e de segregação alélica são similares ao encontrado no *ATXN3* (Ikeuchi et al. 1996).

No entanto, seria a seleção disruptiva e, consequentemente, a raridade de alelos intermediários os únicos fatores que explicariam a ausência de mutação *de novo* na SCA3/MJD? Para responder a essa questão, a caracterização molecular dos raros alelos aparentemente intermediários descritos até o momento precisaria ser estudada - como por exemplo, expressividade do fenótipo versus penetrância, taxa de instabilidade, padrão de transmissão, interrupções internas da repetição e eventual estabilidade. Porém, é de se prever que a raridade de portadores desses alelos inviabilize estudos desse tipo.

Em outras duas doenças de poliQs, o aumento de fertilidade também foi descrito entre seus portadores: na SCA1 (Frontali et al. 1996) e na HD (Shokeir at al. 1975; Walker et al. 1983; Morrison et al. 1995; Pridmore & Adams 1991; Frontali et al. 1996). No entanto, os mecanismos biológicos que conferem aumento na fertilidade dos portadores das doenças poliQs ainda é um completo mistério. Na HD, por exemplo, foi levantada a hipótese de os portadores apresentarem um comportamento sexual promíscuo e/ou hipersexualizado (Dewhurst & McKnight 1970) para explicar o aumento na fertilidade dos portadores. Não obstante, essa hipótese perdeu força, uma vez que estudos mais recentes apontam que os portadores HD têm como disfunção sexual mais comum a hiposexualidade (Craufurd et al. 2001; Szymuś et al. 2020). Além disso, mesmo que a promiscuidade em si fosse uma característica comportamental dos seus portadores, ela não explicaria um aumento de fertilidade. Estudos em mamíferos apontam que o comportamento sexual promíscuo não está diretamente associado a um aumento de fertilidade (Wolff et al. 2004).

Outros aspectos conjunturais, podem explicar a variação de fertilidade em humanos, como por exemplo aspectos culturais, renda e escolaridade, entre outros. Assim, o aumento da fertilidade na SCA2 e SCA3/MJD poderia ser explicado por elementos de cunho cultural. Entretanto, estudos de fertilidade realizados na SCA2 e SCA3/MJD compararam a fertilidade entre irmãos afetados e não afetados, reduzindo assim a possibilidade de viés cultural. Dessa forma, trabalhamos com a

hipótese de modificações fisiológicas do processo reprodutivo dos seus portadores. Um aspecto que reforça isso foi o estudo em camundongo HD transgênico em que machos portadores da mutação tiveram mais filhotes que os não portadores (Morton et al. 2019).

O campo da fertilidade (ou fecundidade) também está envolvido entre as perguntas que permanecem abertas a partir das nossas evidências. Por que portadores de SCA2 e de SCA3/MJD têm mais filhos que os não portadores? E por que os alelos no *ATXN2* e *ATXN3* apresentam distorção na segregação? Respostas a essas perguntas podem ser relevantes; podem inclusive vir a contribuir para o aperfeiçoamento das terapias reprodutivas existentes.

Até o momento, ative-me a comentar sobre as respostas que minha tese trouxe aos objetivos propostos, ou seja, comentei sobre as hipóteses que me pareceram ser melhor sustentadas. Agora eu tentarei dirigir a atenção para as lacunas do conhecimento que precisam ser respondidas para melhor sustentar meus resultados originais. Esses resultados originais são de fato o modelo sobre a dinâmica dos CAGexp e a existência de múltiplas origens ancestrais da SCA2.

As simulações computacionais foram realizadas a partir da combinação das forças seletivas sistematizadas nos capítulos 2 e 3 da tese com o modelo matemático proposto no capítulo 4. Elas pareceram ser suficientes para apontar a dinâmica geral dos alelos expandidos do *ATXN2* e *ATXN3*. Mas serão de fato as únicas forças necessárias e suficientes? Alguns ajustes precisam ser realizados para os resultados se tornarem ainda mais próximos da realidade.

O primeiro ajuste passa por ampliar os estudos de forças seletivas para mais coortes, conferindo mais solidez aos dados das forças seletivas. Dentre os resultados obtidos nas metanálises, o valor da antecipação é o que exige maior atenção. Isso porque na grande maioria dos estudos não delimitou uma idade mínima para inclusão dos sujeitos e pode ter produzido um excesso de resultados a favor da antecipação.

A solução para reduzir esse viés foi inferir a antecipação a partir do produto entre a instabilidade média e o valor da regressão linear da variação da

idade de início com o tamanho da CAG. Para SCA3/MJD, essa regressão foi realizada a partir dos dados brutos da metanálise com distintas coortes mundiais (de Mattos et al. 2018). Já para SCA2, a regressão foi realizada com indivíduos portadores de nossa coorte. A realização de um estudo em escala mundial da relação entre variação da CAGexp e a idade de início dos sintomas na SCA2 traria subsídios mais robustos para confirmar nossos cálculos.

Um mérito que o modelo teve foi o de incorporar a variação na taxa de fertilidade por idade, que possibilitou dar pesos distintos a cada redução da vida reprodutiva em decorrência da antecipação nas taxas de fertilidade, gerando assim o coeficiente de antecipação. No entanto, ajustes podem torná-lo mais realista, como por exemplo, incorporar a taxa de filhos de portadores que nascem após o início dos sintomas. Apesar de interessante, esse ajuste não deverá modificar a essência dos resultados, pois além de minoritários, esses nascimentos ocorrem em um marco temporal próximo ao início dos sintomas.

É importante também refletirmos que, se quisermos aumentar a aplicação do nosso modelo matemático para olhar a dinâmica alélica tanto no passado como também no futuro, precisaremos levar em consideração muitos elementos ainda desconhecidos ou mesmo ainda impensados. Por exemplo, a mudança do comportamento reprodutivo e a variação da expectativa de vida através do tempo.

Esses elementos ainda "desconhecidos" podem tornar as forças seletivas negativas associadas aos *ATXN2 e ATXN3* expandidos mais ou menos intensas. Por exemplo, haverá um relaxamento na intensidade das forças seletivas negativas associadas ao início dos sintomas e da antecipação se a população tiver filhos mais cedo, pois a janela reprodutiva dos portadores aumenta; bem como se a expectativa de vida for menor, pois o tempo adicional que os não portadores têm em sua janela em relação aos portadores será diminuída. E apesar da heterogeneidade e da escassez dos dados até hoje observados, podemos inferir que existe uma tendência da humanidade ter filhos mais tardiamente e viver mais, na medida que o tempo passa. Por exemplo, a expectativa de vida na Europa em 1770 era de 34.3 anos passando para 78.6 em 2019, mundialmente nesse mesmo período a

expectativa de vida passou de 28.7 para 72.6 anos(Roses. 2019). Já a idade média da primeira gestação das mulheres nos EUA em 1935 era de 20,8 anos (Kirmeyer & Hamilton 2001) e em 2020 essa média aumentou para 27,1 anos (https://www.cdc.gov/nchs/fastats/births.ht). Em mulheres dinamarquesas, a idade média do primeiro filho foi de 23,1 anos em 1960, 28,9 em 2004 e 29 anos em 2014 (Blomberg Jensen et al 2015). Em 2020, a idade média da primeira gestação nas europeias foi de 29.5 anos (Eurostat).

Dessa forma, existe uma intensificação das forças seletivas contrárias à manutenção da SCA2 e da SCA3/MJD na população em decorrência das mudanças culturais e tecnológicas.

Nosso modelo abre algumas perspectivas e aponta desafios a serem vencidos na sequência. Primeiro de tudo, acredito que o nosso modelo de uma forma geral pode ser aplicado às demais doenças de poliQs, em decorrência das suas características comuns. Pesquisadores interessados em prever a dinâmica de outros alelos com CAGexp e munidos de dados empíricos adequados poderão prever a dinâmica desses outros alelos. O alelo CAGexp até hoje mais estudado é o do *HTT*, e, portanto, prevejo que o modelo possa ser aplicado de imediato à HD.

O fato de desconsiderarmos em nosso modelo os portadores homozigotos não deve ter impacto relevante nos prognósticos dos alelos expandidos, já que todos os CAGexp têm frequências muito baixas e os casos de homozigose são excepcionalmente raros. Porém, em condições muito específicas como populações pequenas, baixo fluxo gênico, e que a frequência do CAGexp é elevada em decorrência de efeitos fundadores, as chances de homozigose para o CAGexp se elevam. Consequentemente, desconsiderá-la gera um viés a favor da força seletiva associada à distorção na segregação. Dessa forma, analisar a dinâmica do CAGexp nessas condições exigiria ajustes no nosso modelo.

Confrontar as previsões do modelo com estudos de prevalência repetidos ao longo do tempo em diferentes geografias será importante para inferir de uma forma mais precisa quão próximo o modelo está da realidade em relação a variação da frequência alélica dos alelos expandidos. Nosso modelo e também as

evidências até agora publicadas sobre aumentos de prevalências de doenças poliQs nos Açores (de Araújo et al. 2016) e da população da República de Sakha, na Rússia (Platonov et al. 2016) devem servir de estímulo para futuros estudos epidemiológicos que incluam comparações longitudinais.

A ampliação de marcadores moleculares para a construção de haplótipos ancestrais SCA2 é, finalmente, uma das perspectivas concretas que imagino estará entre as mais imediatas a serem questionadas por novos estudos. Nossos resultados nessa esfera são muito provocativos. Eles rearticulam, ou melhor, apresentam um novo cânone de poliformismos informativos para haplótipos do *ATXN2*: os marcadores s695871, rs593226 e rs9300319. Com eles, pela primeira vez se apresentou um grupo robusto de mais de um haplótipo ancestral. É previsível que de fato haja mais de seis origens, como as que conseguimos identificar. Ademais, a ampliação de marcadores também possibilitará datarmos os haplótipos SCA2 e verificar se o número de gerações datadas são convergentes com as simuladas.

Por fim, trabalhar com hipóteses científicas, procurar respondê-las e acabar por descobrir que as respostas na verdade se constituem em outras perguntas é uma experiência subjetiva relatada por inúmeras pessoas. Estou entre elas. Interroguei-me algumas vezes sobre a natureza e os limites do conhecimento humano, eu como um sujeito indagativo e as CAGexp e a evolução como seres externos e independentes a serem observados. Em especial frente aos sucessos da minha empreitada e que pareceram depois ilusórios, porque dependiam de novas demonstrações.

Esses eventos evocaram para mim, e com particular força, uma aparente repetição de conflitos - entre tese e antítese e de novo tese, entre afirmação e negação e de novo, afirmação. As pessoas que se aproximam do estudo das CAGexp vêm quase sempre do estudo das doenças poliQs e da experiência vivida com seus portadores. Vemos as CAGexp como traços deletérios. Como podem eles também ser traços vantajosos? Descrevê-los apenas como “deletérios e vantajosos” não é ainda suficiente para sairmos do impasse aparente, no entanto é um passo lógico importante, pois admite que elementos antagônicos

podem compor uma mesma unidade, nos libertando da dicotomia inerente de nossa forma de pensar. Esse processo é uma das três leis gerais da dialética denominada *“lei da interpenetração dos contrários”* (Engels. 1978).

Essa unidade de elementos antagônicos não se reduz a si mesma. É do choque dessa contradição que deriva outra lei da dialética chamada de “lei negação da negação” (Engels. 1978), onde *o* choque dos contrários é responsável pelo movimento, fazendo com que um dos elementos negue o outro e por sua vez é negado por um nível superior de desenvolvimento, gerando o novo que ainda preserva alguma coisa de ambos os termos negados, processo que muitas vezes é descrito como o esquema triádico de *tese, antítese e síntese.*

Para mim, reunir elementos antagonistas no plano dos fenômenos da experiência humana - adoecer, ter filhos, transmitir doenças à prole, ver o início dos sintomas se modificar através das gerações e combiná-los em um modelo matemático, foi um processo constante de negação da negação; ou, em outras palavras, analisar a SCA2 e SCA3 em uma perspectiva dialética. Espero que as dúvidas obtidas a partir das respostas sejam os elementos que contribuam para a busca de novas respostas.

Em conclusão, de forma concreta, identificamos que os fenômenos até hoje descritos sobre prevalências, antiguidade de origens ancestrais, ocorrência ou não de portadores de expansões de *novo*, fitness e segregação compõem um todo que de fato é coerente, no caso da SCA2 e da SCA3/MJD. Ademais, concluímos sobre a necessidade de ampliação dos dados empíricos associados às forças seletivas associadas aos alelos expandidos do *ATXN2* e *ATXN3* e sobre a necessidade de futuros ajustes no nosso modelo matemático. Concluímos também que o estudo das origens ancestrais da SCA2 merece a atenção a mais marcadores, pois nossas evidências apontaram claramente para múltiplas origens. No entanto, foi possível neste estudo responder ao seu objetivo principal: Como a SCA2 e SCA3, mesmo sendo doenças debilitantes, se mantém na população através das gerações.

REFERÊNCIAS BIBLIOGRÁFICAS GERAIS:

Adegbuyiroa A, Sedighia F, Pilkington AW, Groovera S, Legleitera J. (2017).Proteins containing expanded polyglutamine tracts and neurodegenerative disease. Biochemistry. March 07; 56(9): 1199–1217. doi:10.1021/acs.biochem.6b00936.

Almaguer-Mederos LE, Mesa JML, González-Zaldívar Y, Almaguer-Gotay D, Cuello-Almarales D, Aguilera-Rodríguez R, Falcón NS, Gispert S, Auburger G, Velázquez-Pérez L.(2018). Factors associated with ATXN2 CAG/CAA repeat intergenerational instability in Spinocerebellar ataxia type 2. Clin Genet. Oct;94(3-4):346-350. doi: 10.1111/cge.13380. Epub 2018 Jun 29. PMID: 29756284.

Alonso E, Martínez-Ruano L, De Biase I, Mader C, Ochoa A, Yescas P, Gutiérrez R, White M, Ruano L, Fragoso-Benítez M, Ashizawa T, Bidichandani SI, Rasmussen A. (2007). Distinct distribution of autosomal dominant spinocerebellar ataxia in the Mexican population. Mov Disord. May 15;22(7):1050-3. doi: 10.1002/mds.21470. PMID: 17427938.

Andrés AM, Lao O, Soldevila M, Calafell F, Bertranpetit J. (2003). Dynamics of CAG repeat loci revealed by the analysis of their variability. Hum Mutat. Jan;21(1):61-70. doi: 10.1002/humu.10151. PMID: 12497632.

Araújo, Aldo mellender de. (2021). de Darwin ao século XXI: Uma breve revisão da jornada histórico-epistemológica das ideias sobre evolução. conjectura: filosofia e educação (ucs), v. 26, p. 68-84,

Auburger GW.(2012). Spinocerebellar ataxia type 2. Handb Clin Neurol 103:423–36. 10.1016/B978-0-444-51892-7.00026-7

Bettencourt C, Fialho RN, Santos C, Montiel R, Bruges-Armas J, Maciel P, Lima M.(2008). Segregation distortion of wild-type alleles at the Machado-Joseph disease locus: a study in normal families from the Azores islands (Portugal). J Hum Genet.;53(4):333-339. doi: 10.1007/s10038-008-0261-7. Epub 2008 Feb 20. PMID: 18286225.

Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C, Stevanin G, Koutsis G, Karadima G, Panas M, Yescas-Gómez P, García-Velázquez LE, Alonso-Vilatela ME, Lima M, Raposo M, Traynor B, Sweeney M, Wood N, Giunti P; SPATAX Network, Durr A, Holmans P, Houlden H, Tabrizi SJ, Jones L. (2016). DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. Ann Neurol. Jun;79(6):983-90. doi: 10.1002/ana.24656. Epub 2016 May 6. PMID: 27044000; PMCID: PMC4914895.

Bettencourt C, Santos C, Montiel R, Costa Mdo C, Cruz-Morales P, Santos LR, Simões N, Kay T, Vasconcelos J, Maciel P, Lima M. (2010). Increased transcript diversity: novel splicing variants of Machado-Joseph disease gene (ATXN3). Neurogenetics. May;11(2):193-202. doi: 10.1007/s10048-009-0216-y. Epub 2009 Aug 28. PMID: 19714377.

Blomberg Jensen M, Priskorn L, Jensen TK, Juul A, Skakkebaek NE. (2015). Temporal Trends in Fertility Rates: A Nationwide Registry Based Study from 1901 to 2014. PLoS One. 2015;10(12):e0143722. doi:10.1371/journal.pone.0143722

Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet. 1998 Jun;62(6):1408-15. doi: 10.1086/301869. PMID: 9585597; PMCID: PMC1377148.

Brusco A, Gellera C, Cagnoli C, Saluto A, Castucci A, Michielotto C, et al. (2004). Molecular genetics of hereditary spinocerebellar ataxia: mutation analysis of spinocerebellar ataxia genes and CAG/CTG repeat expansion detection in 225 Italian families. Arch Neurol. ;61:727–33.

Bryer A, Krause A, Bill P, Davids V, Bryant D, Butler J, et al. (2003).The hereditary adult-onset ataxias in South Africa. J Neurol Sci. ;216:47–54. - PubMed - DOI

Bunting EL, Hamilton J, Tabrizi SJ. Polyglutamine diseases. (2021). Curr Opin Neurobiol. Sep 3;72:39-47. doi: 10.1016/j.conb.2021.07.001. Online ahead of print.PMID: 3448

Burnett B, Li F and Pittman RN (2003). The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitylated proteins and has ubiquitin protease activity. Hum Mol Genet 12:3195–3205.

Cancel G, Abbas N, Stevanin G, Dürr A, Chneiweiss H, Néri C, Duyckaerts C, Penet C, Cann HM, Agid Y, et al. (1995). Marked phenotypic heterogeneity associated with expansion of a CAG repeat sequence at the spinocerebellar ataxia 3/Machado-Joseph disease locus. Am J Hum Genet. Oct;57(4):809-16. PMID: 7573040; PMCID: PMC1801502.

Cancel G, Dürr A, Didierjean O, Imbert G, Bürk K, Lezin A, Belal S, Benomar A, Abada-Bendib M, Vial C, Guimarães J, Chneiweiss H, Stevanin G, Yvert G, Abbas N, Saudou F, Lebre AS, Yahyaoui M, Hentati F, Vernant JC, Klockgether T, Mandel JL, Agid Y, Brice A. (1997). Molecular and clinical correlations in spinocerebellar ataxia 2: a study of 32 families. Hum Mol Genet. May;6(5):709-15. doi: 10.1093/hmg/6.5.709. PMID: 9158145.

Caron NS, Wright GEB, Hayden MR. Huntington Disease. (1998). Oct 23 [Updated 2020 Jun 11]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews®

[Internet]. Seattle (WA): University of Washington, Seattle; 1993-2022. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1305/

Chattopadhyay B, Ghosh S, Gangopadhyay PK, Das SK, Roy T, Sinha KK, Jha DK, Mukherjee SC, Chakraborty A, Singhal BS, Bhattacharya AK, Bhattacharyya NP. Modulation of age at onset in Huntington's disease and spinocerebellar ataxia type 2 patients originated from eastern India. (2003). Neurosci Lett. Jul 17;345(2):93-6. doi: 10.1016/s0304-3940(03)00436-1. PMID: 12821179.

Choudhry S, Mukerji M, Srivastava AK, Jain S, Brahmachari SK. CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. (2001). Hum Mol Genet. Oct 1;10(21):2437-46. doi: 10.1093/hmg/10.21.2437. PMID: 11689490.

Costa Mdo C, Paulson HL. Toward understanding Machado-Joseph disease(2012). Prog Neurobiol. May;97(2):239-57. doi: 10.1016/j.pneurobio.2011.11.006. Epub 2011 Nov 23. PMID: 22133674; PMCID: PMC3306771.

Craufurd D, Thompson JC, Snowden JS. (2001). Behavioral changes in Huntington Disease. Neuropsychiatry Neuropsychol Behav Neurol. Oct-Dec;14(4):219-26. PMID: 11725215.

Dantuma NP and Bott LC (2014) The ubiquitin-proteasome system in neurodegenerative diseases: precipitating factor, yet part of the solution. Front Mol Neurosci 7:1–18.

Dantuma, N. P., & Herzog, L. K. (2020). Machado-Joseph Disease: A Stress Combating Deubiquitylating Enzyme Changing Sides. Advances in experimental medicine and biology, 1233, 237–260. https://doi.org/10.1007/978-3-030-38266-7_10.

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, Pedroso JL, Salarini DZ, Vargas FR, de Lima MA, Godeiro C, Santana-da-Silva LC, Toralles MB, Santos S, van der Linden H Jr, Wanderley HY, de Medeiros PF, Pereira ET, Ribeiro E, Saraiva-Pereira ML, Jardim LB; Rede Neurogenetica. (2014). Spinocerebellar ataxias in Brazil--frequencies and modulating effects of related genes. Cerebellum. Feb;13(1):17-28. doi: 10.1007/s12311-013-0510-y. PMID: 23943520.

de Mattos EP, Kolbe Musskopf M, Bielefeldt Leotti V, Saraiva-Pereira ML, Jardim LB(2019). Genetic risk factors for modulation of age at onset in Machado-Joseph disease/spinocerebellar ataxia type 3: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry. Feb;90(2):203-210. doi: 10.1136/jnnp-2018-319200. Epub 2018 Oct 18. PMID: 30337442.

Dewhurst K, Oliver JE, McKnight AL. (1970). Socio-psychiatric consequences of Huntington's disease. Brit J Psychiat;116:255–8.

Diallo A, Jacobi H, Tezenas du Montcel S, Klockgether T. (2021) Natural history of most common spinocerebellar ataxia: a systematic review and meta-analysis. J Neurol. Aug;268(8):2749-2756. doi: 10.1007/s00415-020-09815-2. Epub 2020 Apr 7. PMID: 32266540.

Didierjean O, Cancel G, Stevanin G, Dürr A, Bürk K, Benomar A, Lezin A, Belal S, Abada-Bendid M, Klockgether T, Brice A. (1999). Linkage disequilibrium at the SCA2 locus. J Med Genet. May;36(5):415-7. PMID: 10353790; PMCID: PMC1734371.

Donis KC, Saute JA, Krum-Santos AC, Furtado GV, Mattos EP, Saraiva-Pereira ML, Torman VL, Jardim LB. (2016). Spinocerebellar ataxia type 3/Machado-Joseph disease starting before adolescence. Neurogenetics 17, 107–113 https://doi.org/10.1007/s10048-016-0473-5

Dürr A, Stevanin G, Cancel G, Duyckaerts C, Abbas N, Didierjean O, Chneiweiss H, Benomar A, Lyon-Caen O, Julien J, Serdaru M, Penet C, Agid Y, Brice A. (1996). Spinocerebellar ataxia 3 and Machado-Joseph disease: clinical, molecular, and neuropathological features. Ann Neurol. Apr;39(4):490-9. doi: 10.1002/ana.410390411. PMID: 8619527.

Elden AC, Kim HJ, Hart MP, Chen-Plotkin AS, Johnson BS, Fang X, Armakola M, Geser F, Greene R, Lu MM, Padmanabhan A, Clay-Falcone D, McCluskey L, Elman L, Juhr D, Gruber PJ, Rüb U, Auburger G, Trojanowski JQ, Lee VM, Van Deerlin VM, Bonini NM, Gitler AD.(2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature 466:1069–1075. doi:10.1038/nature09320

Engels, Friedrich. Dialética da Natureza. (1978). Editorial Presença.

Eskenazi BR, Wilson-Rich NS, Starks PT. (2007). A Darwinian approach to Huntington's disease: subtle health benefits of a neurological disorder. Med Hypotheses. 69(6):1183-9. doi: 10.1016/j.mehy.2007.02.046. Epub 2007 Aug 8. PMID: 17689877.

Fan HC, Ho LI, Chi CS, Chen SJ, Peng GS, Chan TM, Lin SZ, Harn HJ. (2014). Polyglutamine (PolyQ) diseases: genetics to treatments. Cell Transplant. ;23(4-5):441-58. doi: 10.3727/096368914X678454. PMID: 24816443.

Faruq M, Scaria V, Singh I, Tyagi S, Srivastava AK, Mukerji M. (2009). SCA-LSD: a repeat-oriented locus-specific variation database for genotype to phenotype correlations in spinocerebellar ataxias. Hum Mutat 30:1037-1042.

Figueroa KP, Coon H, Santos N, Velazquez L, Mederos LA, Pulst SM. (2017). Genetic analysis of age at onset variation in spinocerebellar ataxia type 2. Neurol

Genet. May 15;3(3):e155. doi: 10.1212/NXG.0000000000000155. PMID: 28534046; PMCID: PMC5432368.

Filla A, De Michele G, Santoro L, Calabrese O, Castaldo I, Giuffrida S, Restivo D, Serlenga L, Condorelli DF, Bonuccelli U, Scala R, Coppola G, Caruso G, Cocozza S. (1999). Spinocerebellar ataxia type 2 in southern Italy: a clinical and molecular study of 30 families. J Neurol. Jun;246(6):467-71. doi: 10.1007/s004150050385. PMID: 10431773.

Frontali M, 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. (1996). Genetic fitness in Huntington's Disease and Spinocerebellar Ataxia 1: a population genetics model for CAG repeat expansions. Ann Hum Genet. Sep;60(5):423-35. doi: 10.1111/j.1469-1809.1996.tb00440.x. PMID: 8912795.

Gardiner SL, Boogaard MW, Trompet S, de Mutsert R, Rosendaal FR, Gussekloo J, Jukema JW, Roos RAC, Aziz NA. (2019). Prevalence of Carriers of Intermediate and Pathological Polyglutamine Disease-Associated Alleles Among Large Population-Based Cohorts. JAMA Neurol. Jun 1;76(6):650-656. doi: 10.1001/jamaneurol.2019.0423. PMID: 30933216; PMCID: PMC6563569.

Gaspar C, Lopes-Cendes I, Hayes S, Goto J, Arvidsson K, Dias A, Silveira I, Maciel P, Coutinho P, Lima M, Zhou YX, Soong BW, Watanabe M, Giunti P, Stevanin G, Riess O, Sasaki H, Hsieh M, Nicholson GA, Brunt E, Higgins JJ, Lauritzen M, Tranebjaerg L, Volpini V, Wood N, Ranum L, Tsuji S, Brice A, Sequeiros J, Rouleau GA. (2001). Ancestral origins of the Machado-Joseph disease mutation: a worldwide haplotype study. American journal of human genetics, 68(2), 523–528. https://doi.org/10.1086/318184

Giuffrida S, Lanza S, Restivo DA, Saponara R, Valvo SC, Le Pira F, Trovato Salinaro A, Spinella F, Nicoletti A, Condorelli DF. Clinical and molecular analysis of 11 Sicilian SCA2 families: influence of gender on age at onset. (1999). Eur J Neurol. May;6(3):301-7. doi: 10.1046/j.1468-1331.1999.630301.x. PMID: 10210910.

Gu W, Ma H, Wang K, et al. The shortest expanded allele of the MJD1 gene in a Chinese MJD kindred with autonomic dysfunction. Eur Neurol 2004;52:107-11

Hartl, D. L., & Clark, A. G. (2007). Principles of population genetics (4th ed.). Oxford University Press.

Hernández A, Magariño C, Gispert S, Santos N, Lunkes A, Orozco G, Heredero L, Beckmann J, Auburger G. (1995). Genetic mapping of the spinocerebellar ataxia 2 (SCA2) locus on chromosome 12q23-q24.1. Genomics. Jan 20;25(2):433-5. doi: 10.1016/0888-7543(95)80043-l. PMID: 7789976.

Huntington G. (1872). On chorea. Med Surg Rep ; 26: 317–21

Ichikawa Y, Goto J, Hattori M, Toyoda A, Ishii K, Jeong SY, Hashida H, Masuda N, Ogata K, Kasai F, Hirai M, Maciel P, Rouleau GA, Sakaki Y, Kanazawa I. (2001). The genomic structure and expression of MJD, the Machado-Joseph disease gene. J Hum Genet. ;46(7):413-22. doi: 10.1007/s100380170060. PMID: 11450850.

Ikeuchi T, Igarashi S, Takiyama Y, Onodera O, Oyake M, Takano H, Koide R, Tanaka H, Tsuji S. (1996). Non-Mendelian transmission in dentatorubral-pallidoluysian atrophy and Machado-Joseph disease: the mutant allele is preferentially transmitted in male meiosis. Am J Hum Genet.

Jardim LB, Silveira I, Pereira ML, Ferro A, Alonso I, do Céu Moreira M, Mendonça P, Ferreirinha F, Sequeiros J, Giugliani R. (2001). A survey of spinocerebellar ataxia in South Brazil - 66 new cases with Machado-Joseph disease, SCA7, SCA8, or unidentified disease-causing mutations. J Neurol. ;248(10):870–6.

Ji J, Sundquist K, Sundquist J.(2012). Cancer incidence in patients with polyglutamine diseases: a population-based study in Sweden. Lancet Oncol. Jun;13(6):642-8. doi: 10.1016/S1470-2045(12)70132-8.

Jiang, H., Tang, B., Xia, K., Zhou, Y., Xu, B., Zhao, G., Li, H., Shen, L., Pan, Q., & Cai, F. (2005). Spinocerebellar ataxia type 6 in Mainland China: molecular and clinical features in four families. *Journal of the neurological sciences*, *236*(1-2), 25–29. https://doi.org/10.1016/j.jns.2005.04.009

Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., & Akiguchi, I. (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nature genetics, 8(3), 221–228. https://doi.org/10.1038/ng1194-221

Kieling C, Prestes PR, Saraiva-Pereira ML, Jardim LB. (2007). Survival estimates for patients with Machado-Joseph disease (SCA3). Clin Genet. Dec;72(6):543-5. doi: 10.1111/j.1399-0004.2007.00910.x.

Kirmeyer SE, Hamilton BE. Childbearing differences among three generations of U.S. women. (2011). NCHS data brief, no 68. Hyattsville, MD: National Center for Health Statistics.

Kraft S, Furtado S, Ranawaya R, Parboosingh J, Bleoo S, McElligott K, Bridge P, Spacey S, Das S, Suchowersky O. (2005). Adult onset spinocerebellar ataxia in a Canadian movement disorders clinic. Can J Neurol Sci. Nov;32(4):450-8. doi: 10.1017/s0317167100004431. PMID: 16408574.

Lee, T., Li, Y. R., Ingre, C., Weber, M., Grehl, T., Gredal, O., de Carvalho, M., Meyer, T., Tysnes, O. B., Auburger, G., Gispert, S., Bonini, N. M., Andersen, P. M., & Gitler, A. D. (2011). Ataxin-2 intermediate-length polyglutamine expansions in European ALS patients. *Human molecular genetics*, *20*(9), 1697–1700. https://doi.org/10.1093/hmg/ddr045

Lee T, Takeshima Y, Kusunoki N, Awano H, Yagi M, Matsuo M, Iijima K. (2014). Differences in carrier frequency between mothers of Duchenne and Becker muscular dystrophy patients. J Hum Genet. Jan;59(1):46-50. doi: 10.1038/jhg.2013.119. Epub 2013 Nov 14. PMID: 24225992; PMCID: PMC3970902.

Lieberman AP, Shakkottai VG, Albin RL (2019). Polyglutamine Repeats in Neurodegenerative Diseases . Annu Rev Pathol. 14:1-27. doi:10.1146/annurev-pathmechdis-012418-012857

Lima L, Coutinho P. Clinical criteria for diagnosis of Machado-Joseph disease: report of a non-Azorena Portuguese family(1980). Neurology. Mar;30(3):319-22. doi: 10.1212/wnl.30.3.319. PMID: 7189034.

Lima M, Smith MT, Silva C, Abade A, Mayer FM, Coutinho P. (2001). Natural selection at the MJD locus: phenotypic diversity, survival and fertility among Machado-Joseph Disease patients from the Azores. J Biosoc Sci. Jul;33(3):361-73. doi: 10.1017/s0021932001003613. PMID: 11446399.

Loewe, L. (2008) Negative selection. Nature Education 1(1):59

Bonanomi, M., Mazzucchelli, S., D'Urzo, A., Nardini, M., Konarev, P. V., Invernizzi, G., Svergun, D. I., Vanoni, M., Regonesi, M. E., & Tortora, P. (2014). Interactions of ataxin-3 with its molecular partners in the protein machinery that sorts protein aggregates to the aggresome. *The international journal of biochemistry & cell biology*, *51*, 58–64. https://doi.org/10.1016/j.biocel.2014.03.015

Magaña JJ, Velázquez-Pérez L, Cisneros B. Spinocerebellar ataxia type 2: clinical presentation, molecular mechanisms, and therapeutic perspectives. (2013). .Mol Neurobiol. Feb;47(1):90-104. doi: 10.1007/s12035-012-8348-8. Epub 2012 Sep 21. PMID: 22996397.

Mao, R., Aylsworth, A. S., Potter, N., Wilson, W. G., Breningstall, G., Wick, M. J., Babovic-Vuksanovic, D., Nance, M., Patterson, M. C., Gomez, C. M., & Snow, K. (2002). Childhood-onset ataxia: testing for large CAG-repeats in SCA2 and SCA7. *American journal of medical genetics*, *110*(4), 338–345. https://doi.org/10.1002/ajmg.10467

Margulis BA, Vigont V, Lazarev VF, Kaznacheyeva EV, Guzhova IV. (2013). Pharmacological protein targets in polyglutamine diseases: mutant polypeptides and their interactors. FEBS Lett. Jun 27;587(13):1997-2007. doi: 10.1016/j.febslet.2013.05.022. Epub 2013 May 15. PMID: 23684638.

Martins S, Calafell F, Gaspar C, Wong VC, Silveira I, Nicholson GA, Brunt ER, Tranebjaerg L, Stevanin G, Hsieh M, Soong BW, Loureiro L, Dürr A, Tsuji S, Watanabe M, Jardim LB, Giunti P, Riess O, Ranum LP, Brice A, Rouleau GA, Coutinho P, Amorim A, Sequeiros J. (2007). Asian origin for the worldwide-spread mutational event in Machado-Joseph disease. Arch Neurol. Oct;64(10):1502-8.

Martins S, Coutinho P, Silveira I, Giunti P, Jardim LB, Calafell F, Sequeiros J, Amorim A. (2008). Cis-acting factors promoting the CAG intergenerational instability in Machado-Joseph disease. Am J Med Genet B Neuropsychiatr Genet. Jun 5;147B(4):439-46. doi: 10.1002/ajmg.b.30624. PMID: 17948873.

Maruyama H, Nakamura S, Matsuyama Z, Sakai T, Doyu M, Sobue G, Seto M, Tsujihata M, Oh-i T, Nishio T, et al. Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. (1995). Hum Mol Genet. May;4(5):807-12. doi: 10.1093/hmg/4.5.807. PMID: 7633439

Matilla-Dueñas A, Ashizawa T, Brice A, et al. (2014). Consensus paper: pathological mechanisms underlying neurodegeneration in spinocerebellar ataxias. Cerebellum ;13:269–302.

Max Roser, Esteban Ortiz-Ospina and Hannah Ritchie (2013). last revised in October 2019.- ""Life Expectancy"". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/life-expectancy' [Online Resource]

McLoughlin HS, Moore LR, Paulson HL. Pathogenesis of SCA3 and implications for other polyglutamine diseases(2020). Neurobiol Dis. Feb;134:104635. doi: 10.1016/j.nbd.2019.104635. Epub 2019 Oct 24. PMID: 31669734; PMCID: PMC6980715.

McNulty P, Pilcher R, Ramesh R, Necuiniate R, Hughes A, Farewell D, Holmans P, Jones L; REGISTRY Investigators of the European Huntington's Disease Network.(2018). Reduced Cancer Incidence in Huntington's Disease: Analysis in the Registry Study. J Huntingtons Dis. ;7(3):209-222. doi: 10.3233/JHD-170263. PMID: 30103338.

Mitteldorf J. What Is Antagonistic Pleiotropy? Biochemistry (Mosc). (2019). Dec;84(12):1458-1468. doi: 10.1134/S0006297919120058. PMID: 31870250.

Mizushima K, Watanabe M, Kondo I, Okamoto K, Shizuka M, Abe K, Aoki M, Shoji M.(1999). Analysis of spinocerebellar ataxia type 2 gene and haplotype analysis: (CCG)1-2 polymorphism and contribution to founder effect. J Med Genet. Feb;36(2):112-4. PMID: 10051008; PMCID: PMC1734293.

Morrison PJ, Johnston WP, Nevin NC. (1995). The epidemiology of Huntington's disease in Northern Ireland. J Med Genet. Jul;32(7):524-30. doi: 10.1136/jmg.32.7.524. PMID: 7562964; PMCID: PMC1050544.

Morton AJ, Skillings EA, Wood NI, Zheng Z. (2019). Antagonistic pleiotropy in mice carrying a CAG repeat expansion in the range causing Huntington's disease. Sci Rep. ;9(1):37. Published 2019 Jan 10. doi:10.1038/s41598-018-37102-8

Moseley ML, Benzow KA, Schut LJ, et al. (1998). Incidence of dominant spinocerebellar and Friedreich triplet repeats among 361 ataxia families. Neurology 51: 1666–1671.

Myers RH, MacDonald ME, Koroshetz WJ, Duyao MP, Ambrose CM, Taylor SA, Barnes G, Srinidhi J, Lin CS, Whaley WL, et al. (1993). De novo expansion of a (CAG)n repeat in sporadic Huntington's disease. Nat Genet. Oct;5(2):168-73. doi: 10.1038/ng1093-168. PMID: 8252042.

Nakano KK, Dawson DM and Spence A (1972). Machado disease. A hereditary ataxia in Portuguese emigrants to Massachusetts. Neurology 22:49–55.
Nishiyama K, Murayama S, Goto J, Watanabe M, Hashida H, Katayama S, Nomura Y, Nakamura S and Kanazawa I (1996) Regional and cellular expression of the Machado-Joseph disease gene in brains of normal and affected individuals. AnnNeurol 40:776–781.

Orozco Diaz G, Nodarse Fleites A, Cordovés Sagaz R, Auburger G. (1990). Autosomal dominant cerebellar ataxia: clinical analysis of 263 patients from a homogeneous population in Holguín, Cuba. Neurology. Sep;40(9):1369-75. doi: 10.1212/wnl.40.9.1369. PMID: 2392220.

Ostrowski LA, Hall AC, Mekhail K. Ataxin-2: From RNA Control to Human Health and Disease. Genes (Basel). (2017). Jun 5;8(6):157. doi: 10.3390/genes8060157. PMID: 28587229; PMCID: PMC5485521.

Pang J, Allotey R, Wadia N, Sasaki H, Bindoff L, Chamberlain S. (1999). A common disease haplotype segregating in spinocerebellar ataxia 2 (SCA2) pedigrees of diverse ethnic origin. Eur J Hum Genet. Oct-Nov;7(7):841-5. doi: 10.1038/sj.ejhg.5200372. PMID: 10573020.

Paradisi I, Ikonomu V, Arias S. (2016). Spinocerebellar ataxias in Venezuela: genetic epidemiology and their most likely ethnic descent. J Hum Genet. Mar;61(3):215-22. doi: 10.1038/jhg.2015.131. Epub 2015 Nov 5. PMID: 26538302.

Paulson HL. (2007). Dominantly inherited ataxias: lessons learned from Machado-Joseph disease/spinocerebellar ataxia type 3. Semin Neurol. Apr;27(2):133-42. doi: 10.1055/s-2007-971172. PMID: 17390258.

Paulson HL, Das SS, Crino PB, Perez MK, Patel SC, Gotsdiner D, Fischbeck KH, Pittman RN. (1997). Machado-Joseph disease gene product is a cytoplasmic protein widely expressed in brain. Ann Neurol. Apr;41(4):453-62. doi: 10.1002/ana.410410408. PMID: 9124802.

Pflieger LT, Dansithong W, Paul S, Scoles DR, Figueroa KP, Meera P, Otis TS, Facelli JC, Pulst SM. (2017). Gene co-expression network analysis for identifying modules and functionally enriched pathways in SCA2. Hum Mol Genet. Aug 15;26(16):3069-3080. doi: 10.1093/hmg/ddx191. PMID: 28525545; PMCID: PMC5886232.

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. doi:10.1007/s10048-016-0481-5

Prestes PR, Saraiva-Pereira ML, Silveira I, Sequeiros J, Jardim LB. (2008). Machado-Joseph disease enhances genetic fitness: a comparison between affected and unaffected women and between MJD and the general population. Ann Hum Genet. Jan;72(Pt 1):57-64. doi: 10.1111/j.1469-1809.2007.00388.x. Epub 2007 Aug 7. PMID: 17683516.

Pridmore SA, Adams GC. The fertility of HD-affected individuals in Tasmania. (1991). Aust N Z J Psychiatry. Jun;25(2):262-4. doi: 10.1080/00048679109077743. PMID: 1831608.

Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, Lopes-Cendes I, Pearlman S, Starkman S, Orozco-Diaz G, Lunkes A, DeJong P, Rouleau GA, Auburger G, Korenberg JR, Figueroa C, Sahba S. (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet. Nov;14(3):269-76. doi: 10.1038/ng1196-269. PMID: 8896555

Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, Lopes-Cendes I, Pearlman S, Starkman S, Orozco-Diaz G, Lunkes A, DeJong P, Rouleau GA, Auburger G, Korenberg JR, Figueroa C, Sahba S. (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet. ;14:269–76

Radvany J, Avila JO, Gabbai AA, Bacheschi LA (1988). [Machado-Joseph disease: the first two reported families from Brazil]. Abstract (Portuguese). Arq Neuro-Psiquiat: 46(S):152

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 B Neuropsychiatr Genet. Mar 5;153B(2):524-531. doi: 10.1002/ajmg.b.31013. PMID: 19676102.

Reed TE, Neel JV. (1959). Huntington's chorea in Michigan. 2. Selection and mutation. Am J Hum Genet. Jun;11(2 Part 1):107-36. PMID: 13661148; PMCID: PMC1931980.

Riess O, Laccone FA, Gispert S, Schöls L, Zühlke C, Vieira-Saecker AM, Herlt S, Wessel K, Epplen JT, Weber BH, Kreuz F, Chahrokh-Zadeh S, Meindl A, Lunkes A, Aguiar J, Macek M Jr, Krebsová A, Macek M Sr, Bürk K, Tinschert S, Schreyer I, Pulst SM, Auburger G. (1997). SCA2 trinucleotide expansion in German SCA patients. Neurogenetics. May;1(1):59-64. doi: 10.1007/s100480050009. PMID: 10735276.

Riess O, Rüb U, Pastore A, Bauer P, Schöls L. SCA3: neurological features, pathogenesis and animal models. (2008). Cerebellum.; 7(2):125-37. doi: 10.1007/s12311-008-0013-4. PMID: 18418689.

Rodríguez-Labrada R, Martins AC, Magaña JJ, Vazquez-Mojena Y, Medrano-Montero J, Fernandez-Ruíz J, Cisneros B, Teive H, McFarland KN, Saraiva-Pereira ML, Cerecedo-Zapata CM, Gomez CM, Ashizawa T, Velázquez-Pérez L, Jardim LB; PanAmerican Hereditary Ataxia Network. (2020). Founder Effects of Spinocerebellar Ataxias in the American Continents and the Caribbean. Cerebellum. Jun;19(3):446-458. doi: 10.1007/s12311-020-01109-7. PMID: 32086717.

Rose, M. (1991). Evolutionary Biology of Aging, Oxford University Press, Oxford, England.

Rosenberg RN, Nyhan WL, Bay C and Shore P (1976). Autosomal dominant striatonigral degeneration. A clinical, pathologic, and biochemical study of a new genetic disorder. Neurology 26:703–714.

Ruano L, Melo C, Silva MC, Coutinho P. (2014). The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies. Neuroepidemiology. ;42(3):174-83. doi: 10.1159/000358801. Epub 2014 Mar 5. PMID: 24603320.

Rüb U, Schöls L, Paulson H, Auburger G, Kermer P, Jen JC, Seidel K, Korf HW, Deller T. (2013). Clinical features, neurogenetics and neuropathology of the polyglutamine spinocerebellar ataxias type 1, 2, 3, 6 and 7. Prog Neurobiol. May;104:38-66. doi: 10.1016/j.pneurobio.2013.01.001. Epub 2013 Feb 21. PMID: 23438480.

Rubinsztein DC, Leggo J. (1997). Non-Mendelian transmission at the Machado-Joseph disease locus in normal females: preferential transmission of alleles with smaller CAG repeats. J Med Genet. Mar;34(3):234-6. doi: 10.1136/jmg.34.3.234. PMID: 9132496; PMCID: PMC1050899.

Sahba S, Nechiporuk A, Figueroa KP, Nechiporuk T, Pulst SM (1998). Genomic structure of the human gene for spinocerebellar ataxia 2 (SCA2) on chromosome 12q24.1. Genomics 47:359–64. doi:10.1006/geno.1997.5131

Sanpei K, Takano H, Igarashi S, Sato T, Oyake M, Sasaki H, Wakisaka A, Tashiro K, Ishida Y, Ikeuchi T, Koide R, Saito M, Sato A, Tanaka T, Hanyu S, Takiyama Y, Nishizawa M, Shimizu N, Nomura Y, Segawa M, Iwabuchi K, Eguchi I, Tanaka H, Takahashi H, Tsuji S. (1996). Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. Nat Genet. ;14:277–84

Saute JAM, Jardim LB (2015). Machado Joseph disease: clinical and genetic aspects, and current treatment. Expert Opin Orphan Drugs 3:517–535

Schmidt, T., Lindenberg, K. S., Krebs, A., Schöls, L., Laccone, F., Herms, J., Rechsteiner, M., Riess, O., & Landwehrmeyer, G. B. (2002). Protein surveillance machinery in brains with spinocerebellar ataxia type 3: redistribution and differential recruitment of 26S proteasome subunits and chaperones to neuronal intranuclear inclusions. *Annals of neurology*, *51*(3), 302–310. https://doi.org/10.1002/ana.10101

Schmitt I, Linden M, Khazneh H, Evert BO, Breuer P, Klockgether T and Wuellner U. (2007). Inactivation of the mouse Atxn3 (ataxin-3) gene increases protein ubiquitination. Biochem Biophys Res Commun 362:734–739.

Schmitz-Hübsch T, du Montcel ST, Baliko L, Berciano J, Boesch S, Depondt C, Giunti P, Globas C, Infante J, Kang J-S et al. (2006). Scale for the assessment and rating of ataxia: development of a new clinical scale. Neurology 66:1717–1720.

Schöls, L., Amoiridis, G., Büttner, T., Przuntek, H., Epplen, J. T., & Riess, O. (1997). Autosomal dominant cerebellar ataxia: phenotypic differences in genetically defined subtypes?. *Annals of neurology*, *42*(6), 924–932. https://doi.org/10.1002/ana.410420615

Schöls, L., Bauer, P., Schmidt, T., Schulte, T., & Riess, O. (2004). Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. *The Lancet. Neurology*, *3*(5), 291–304. https://doi.org/10.1016/S1474-4422(04)00737-9

Schulz JB, Borkert J, Wolf S, Schmitz-Hübsch T, Rakowicz M, Mariotti C, Schöls L, Timmann D, van de Warrenburg B, Dürr A, Pandolfo M, Kang JS, Mandly AG, Nägele T, Grisoli M, Boguslawska R, Bauer P, Klockgether T, Hauser TK. (2010). Visualization, quantification and correlation of brain atrophy with clinical symptoms in spinocerebellar ataxia types 1, 3 and 6. Neuroimage. (2010). Jan 1;49(1):158-68. doi: 10.1016/j.neuroimage.2009.07.027. Epub 2009 Jul 22. Erratum in: Neuroimage. 2010 May 1;50(4):1712. Schoels, Ludger [corrected to Schöls, Ludger]. PMID: 19631275.

Seidel K, Siswanto S, Fredrich M, Bouzrou M, denDunnen WFA, Ozerden I, Korf HW, Melegh B, de Vries JJ, Brunt ER, Auburger G, Rüb U. (2017). On the distribution of intranuclear and cytoplasmic aggregates in the brainstem of patients with spinocerebellar ataxia type 2 and 3. Brain Pathol.;27:345–55.

Semaka A, Kay C, Doty C, Collins JA, Bijlsma EK, Richards F, Goldberg YP, Hayden MR. (2013). CAG size-specific risk estimates for intermediate allele repeat instability in Huntington disease. J Med Genet. Oct;50(10):696-703. doi: 10.1136/jmedgenet-2013-101796. Epub 2013 Jul 29. PMID: 23896435.

Shokeir MH. (1975). Investigations on Huntington's disease in the Canadian Prairies. I. Prevalence. Clin Genet. Apr;7(4):345-8. doi: 10.1111/j.1399-0004.1975.tb00340.x. PMID: 123837.

Singh A, Faruq M, Mukerji M, Dwivedi MK, Pruthi S, Kapoor S. (2014). Infantile onset spinocerebellar ataxia 2 (SCA2): a clinical report with review of previous cases. J Child Neurol. Jan;29(1):139-44. doi: 10.1177/0883073813509015. Epub 2013 Dec 2. PMID: 24300164.

Socal MP, Emmel VE, Rieder CR, Hilbig A, Saraiva-Pereira ML, Jardim LB. (2009). Intrafamilial variability of Parkinson phenotype in SCAs: novel cases due to SCA2 and SCA3 expansions. Parkinsonism Relat Disord. Jun;15(5):374-8. doi: 10.1016/j.parkreldis.2008.09.005. Epub 2008 Nov 5. PMID: 18990604.

Sonakar AK, Shamim U, Srivastava MP, Faruq M, Srivastava AK. (2021). SCA2 in the Indian population: Unified haplotype and variable phenotypic patterns in a large case series. Parkinsonism Relat Disord. Aug;89:139-145. doi: 10.1016/j.parkreldis.2021.07.011. Epub 2021 Jul 14. PMID: 34298214.

Soong B, Cheng C, Liu R, Shan D. Machado-Joseph disease: clinical, molecular, and metabolic characterization in Chinese kindreds (1997). Ann Neurol. Apr;41(4):446-52. doi: 10.1002/ana.410410407. PMID: 9124801.

Sørensen SA, Fenger K, Olsen JH. (1999). Significantly lower incidence of cancer among patients with Huntington disease: An apoptotic effect of an expanded polyglutamine tract? Cancer. Oct 1;86(7):1342-6. PMID: 10506723.

Souza GN, Kersting N, Gonçalves TA, Pacheco DLO, Saraiva-Pereira ML, Camey SA, Saute JAM, Jardim LB. (2017). Cancer in Machado-Joseph disease patients-low frequency as a cause of death. Cancer Genet. Apr;212-213:19-23. doi: 10.1016/j.cancergen.2017.03.008. Epub 2017 Mar 30. PMID: 28449807.

Souza GN, Kersting N, Krum-Santos AC, Santos AS, Furtado GV, Pacheco D, Gonçalves TA, Saute JA, Schuler-Faccini L, Mattos EP, Saraiva-Pereira ML, Jardim LB. (2016). Spinocerebellar ataxia type 3/Machado-Joseph disease: segregation patterns and factors influencing instability of expanded CAG transmissions. Clin Genet. Aug;90(2):134-40. doi: 10.1111/cge.12719.

Storey E, du Sart D, Shaw JH, et al. (2000). Frequency of spinocerebellar ataxia types 1, 2, 3, 6, and 7 in Australian patients with spinocerebellar ataxia. Am J Med Genet 95:351-357.

Stoyas CA, La Spada AR. (2018). The CAG-polyglutamine repeat diseases: a clinical, molecular, genetic, and pathophysiologic nosology. Handb Clin Neurol. ;147:143-170. doi: 10.1016/B978-0-444-63233-3.00011-7. PMID: 29325609.

Szymuś K, Bystrzyński A, Kwaśniak-Butowska M, Konkel A, Leśnicka A, Nowacka M, Sławek J. (2020). Sexual dysfunction in Huntington's Disease - a systematic review. Neurol Neurochir Pol. ;54(4):305-311. doi: 10.5603/JNNS.a2020.0025. Epub 2020 Apr 3. PMID: 32242915.

Tait D, Riccio M, Sittler A, Scherzinger E, Santi S, Ognibene A, Maraldi NM, Lehrach H and Wanker EE (1998). Ataxin-3 is transported into the nucleus and associates with the nuclear matrix. Hum Mol Genet 7:991–997.

Takiyama Y, Igarashi S, Rogaeva EA, Endo K, Rogaev EI, Tanaka H, Sherrington R, Sanpei K, Liang Y, Saito M, et al. (1995). Evidence for inter-generational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. Hum Mol Genet. Jul;4(7):1137-46. doi: 10.1093/hmg/4.7.1137. PMID: 8528200.

Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, Shimazaki H, Soutome M, Endo K, Ohta S, et al. (1993). The gene for Machado-Joseph disease maps to human chromosome 14q. Nat Genet. Jul;4(3):300-4. doi: 10.1038/ng0793-300. PMID: 8358439.

Takiyama Y, Sakoe K, Soutome M, Namekawa M, Ogawa T, Nakano I, Igarashi S, Oyake M, Tanaka H, Tsuji S, Nishizawa M. Single sperm analysis of the CAG repeats in the gene for Machado-Joseph disease (MJD1): evidence for non-Mendelian transmission of the MJD1 gene and for the effect of the intragenic CGG/GGG polymorphism on the intergenerational instability. Hum Mol Genet. (1997). Jul;6(7):1063-8. doi: 10.1093/hmg/6.7.1063. PMID: 9215676.

Tezenas du Montcel S, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brussino A, Forlani S, Rakowicz M, Schöls L, Mariotti C, van de Warrenburg BP, Orsi L, Giunti P, Filla A, Szymanski S, Klockgether T, Berciano J, Pandolfo M, Boesch S, Melegh B, Timmann D, Mandich P, Camuzat A; Clinical Research Consortium for Spinocerebellar Ataxia (CRC-SCA); EUROSCA network, Goto J, Ashizawa T, Cazeneuve C, Tsuji S, Pulst SM, Brusco A, Riess O, Brice A, Stevanin G. (2014). Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. Brain. Sep;137(Pt 9):2444-55. doi: 10.1093/brain/awu174. Epub 2014 Jun 26.

Turner MR, Goldacre R, Goldacre MJ. Reduced cancer incidence in Huntington's disease: record linkage study clue to an evolutionary trade-off? Clin Genet. 2013 Jun;83(6):588-90. doi: 10.1111/cge.12010. Epub 2012 Sep 28. PMID: 23017147.

Vale, J., Bugalho, P., Silveira, I., Sequeiros, J., Guimarães, J., & Coutinho, P. (2010). Autosomal dominant cerebellar ataxia: frequency analysis and clinical characterization of 45 families from Portugal. *European journal of neurology*, *17*(1), 124–128. https://doi.org/10.1111/j.1468-1331.2009.02757.x

Vallés L, Estrada GL, Bastecherrea SL. Algunas formas de heredoataxia en una región de Cuba. Rev Neurol (Cubana). (1978). ;27:163–76.

van Alfen, N., Sinke, R. J., Zwarts, M. J., Gabreëls-Festen, A., Praamstra, P., Kremer, B. P., & Horstink, M. W. (2001). Intermediate CAG repeat lengths (53,54) for MJD/SCA3 are associated with an abnormal phenotype. *Annals of neurology*, *49*(6), 805–807. https://doi.org/10.1002/ana.1089

van de Warrenburg, B. P., Sinke, R. J., Verschuuren-Bemelmans, C. C., Scheffer, H., Brunt, E. R., Ippel, P. F., Maat-Kievit, J. A., Dooijes, D., Notermans, N. C., Lindhout, D., Knoers, N. V., & Kremer, H. P. (2002). Spinocerebellar ataxias in the Netherlands: prevalence and age at onset variance analysis. *Neurology*, *58*(5), 702–708. https://doi.org/10.1212/wnl.58.5.702

Velázquez Pérez L, Cruz GS, Santos Falcón N, Enrique Almaguer Mederos L, Escalona Batallan K, Rodríguez Labrada R et al (2009) Molecular epidemiology of spinocerebellar ataxias in Cuba: insights into SCA2 founder effect in Holguín. Neurosci Lett 454:157–60. doi:10.1016/j.neulet.2009.03.015

Velázquez-Pérez, L., Rodríguez-Labrada, R., García-Rodríguez, J. C., Almaguer-Mederos, L. E., Cruz-Mariño, T., & Laffita-Mesa, J. M. (2011). A comprehensive review of spinocerebellar ataxia type 2 in Cuba. *Cerebellum (London, England)*, *10*(2), 184–198. https://doi.org/10.1007/s12311-011-0265-2

Wadia NH, Swami RK (1971) A new form of heredo-familiar spinocerebellar degeneration with slow eye movements (nine families). Brain 94:359–74

Walker DA, Harper PS, Newcombe RG, Davies K. (1983). Huntington's chorea in South Wales: mutation, fertility, and genetic fitness. J Med Genet. Feb;20(1):12-7. doi: 10.1136/jmg.20.1.12. PMID: 6221100; PMCID: PMC1048979.

Wheeler VC, Persichetti F, McNeil SM, Mysore JS, Mysore SS, MacDonald ME, Myers RH, Gusella JF, Wexler NS; US-Venezuela Collaborative Research Group. (2007). Factors associated with HD CAG repeat instability in Huntington disease. J Med Genet.. Nov;44(11):695-701. doi: 10.1136/jmg.2007.050930. Epub 2007 Jul 27. PMID: 17660463; PMCID: PMC2705129.

Williams, G. (1957). Pleiotropy, natural selection, and the evolution of senescence, Evolution, 11, 398–411.

Wirth B, Schmidt T, Hahnen E, Rudnik-Schöneborn S, Krawczak M, Müller-Myhsok B, Schönling J, Zerres K. (1997). De novo rearrangements found in 2% of index patients with spinal muscular atrophy: mutational mechanisms, parental origin, mutation rate, and implications for genetic counseling. Am J Hum Genet. Nov;61(5):1102-11. doi: 10.1086/301608. PMID: 9345102; PMCID: PMC1716038.

Wirth, B; Schmidt, T; Hahnen, E; Rudnik-Schöneborn, S; Krawczak, M; Müller-Myhsok, B; Schönling, J; Zerres, K (1997). De Novo Rearrangements Found in 2% of Index Patients with Spinal Muscular Atrophy: Mutational Mechanisms, Parental Origin, Mutation Rate, and Implications for Genetic Counseling"". The American Journal of Human Genetics. 61 (5):

Wolff JO, Macdonald DW. Promiscuous females protect their offspring. Trends Ecol Evol. (2004). Mar;19(3):127-34. doi: 10.1016/j.tree.2003.12.009. PMID: 16701243."

Woods BT, Schaumburg HH. (1972). Nigro-spino-dentatal degeneration with nuclear oph- thalmoplegia: a unique and partially treatable clinico-pathological entity. J Neurol Sci.;17:149-166.

Zhao Y, Tan EK, Law HY, et al. (2002). Prevalence and ethnic differences of autosomal-dominant cerebellar ataxia in Singapore. Clin Genet 62:478-481.