ALLYSSON ALLAN DE FARIAS

A ancestralidade de populações do Nordeste brasileiro com elevadas frequências de casamentos consanguíneos e prevalência de doenças genéticas raras

The ancestry of Brazilian Northeast populations with high frequency of consanguineous marriage and prevalence of rare genetic diseases

São Paulo

2018

ALLYSSON ALLAN DE FARIAS

**A ancestralidade de populações do Nordeste brasileiro com
elevadas frequências de casamentos consanguíneos e
prevalência de doenças genéticas raras**

Tese apresentada ao Instituto de
Biociências da Universidade de
São Paulo como requerimento
parcial para obtenção do Título de
Doutor em Ciências, na área de
Genética e Biologia Evolutiva.

Orientador: Prof. Dr. Fernando
Kok

Coorientadora: Profa. Dra.
Silvana Santos

São Paulo

2018

FICHA CATALOGRÁFICA

Ficha catalográfica elaborada pelo Serviço de Biblioteca do Instituto de Biociências da USP,
com os dados fornecidos pelo autor no formulário:
<http://www.ib.usp.br/biblioteca/ficha-catalografica/ficha.php>

Farias, Allysson Allan de
A ancestralidade de populações do Nordeste brasileiro
com elevadas frequências de casamentos consanguíneos e
prevalência de doenças genéticas raras / Allysson Allan
de Farias; orientador Fernando Kok. -- São Paulo, 2018.

107 f.

Tese (Doutorado) - Instituto de Biociências da
Universidade de São Paulo, Departamento de Genética e
Biologia Evolutiva.

1. Ancestralidade. 2. Mutação. 3. Doenças genéticas de
herança recessiva. I. Kok, Fernando, orient. II. Título.

Bibliotecária responsável pela estrutura da catalogação da publicação:
Elisabete da Cruz Neves - CRB - 8/6228

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Orientador(a)

DEDICATÓRIA

Dedico a Silvana Santos,
Uirá Souto Melo e
Fernando Kok pela
parceria de intensos dez
anos de pesquisas,
formação pessoal e
atividades culturais.
Minhas referências de
paixão pela Ciência.

EPÍGRAFE

“- O que motiva fazer ciência: a catarse, o orgasmo, o estado de graça.”

(Prof. Dr. Paulo Otto em banca de qualificação de Filipe Faria Berçot, 2016)

AGRADECIMENTOS

Ao Prof. Dr. Fernando Kok por orientar o desenvolvimento desta tese dos momentos mais críticos aos mais prazerosos, e por me receber sagradamente às terças-feiras no Centro de Estudos do Genoma Humano. Inclui-se também nessa relação nossas conversas sobre demografia, história e cultura de comunidades do sertão Nordestino. À minha comadre Silvana Santos e ao compadre Uirá Souto Melo por acreditarem em minha carreira acadêmica e por compartilharem a leitura de artigos “publicados na semana passada”. Essa relação de compadrio é decana e perpassa a pesquisa, o ensino, a extensão e a vida pessoal.

Ao Prof. Dr. Rui Murrieta que me orientou do ano 2014 a 2015. Sob sua batuta pude me aprofundar em aspectos bio culturais e históricos de populações do mundo. Ao Prof. Dr. Walter Neves, que neste mesmo período me recebeu no Laboratório de Estudos Evolutivos Humanos (LEEH-USP), e sob sua caneta vermelha tive a oportunidade de aprofundar meus conhecimentos em evolução humana. Aos colegas do Laboratório: Fred, Leandro, Carol, Helbert, Pedro, Rodrigo, Carminha, Ivan, Max, Job e Marimari.

Ao Prof. Dr. Paulo Otto por quem tenho profunda admiração desde 2009, quando recebeu um grupo de entusiasmados graduandos da Universidade Estadual da Paraíba nas dependências do Instituto de Biociências da USP e em sua própria casa, isso me marcou bastante e me motivou a seguir na vida acadêmica. À Sra. Lilian Dluhosch, que com sua polidez organizou esse rito de passagem. De 2015 a 2017, tive a oportunidade de ter a permissão de fazer parte do laboratório mais eclético do Edifício André Dreyfus na sala de alunos do Prof. Paulo, que junto com o palmeirense Renan, o corintiano Gustavo, Guga e Esther pude compartilhar novidades, problemas de bioinformática, desabafos e cafeína. Devo aos quatro e a Ju Carnavalli, Leandro, Caçula, Vini, André, Alex, Rodrigo, Ju Borsoi, Quequel, Isa e Mari ótimos momentos de revigoração da saúde mental no café das 16 horas, no pastel e em momentos etílicos.

À Dra. Kelly Nunes por ter trilhado os rumos dessa tese em momentos de crise, e não ter deixado o barco à deriva. Ao Prof. Dr. Diogo Meyer pelas valiosas sugestões e por sempre estar disponível.

À Profa. Dra. Regina Mingroni pelo rigor nos conceitos em genética, pelos puxões de orelha em relação a prazos, organização da vida acadêmica, por aprender que o termo certo é “coralista” e não corista, e pela reviravolta de meu projeto de doutorado após a

qualificação. Aproveito o ensejo para agradecer aos Profs. Drs. Tábita Hunemeier, Sabine Eggers e Oswaldo Okamoto pelas contribuições dadas para enriquecer o conteúdo dessa tese.

Aos Profs. Angela Morgante e Eduardo Gorab pela profundidade em história da genética no mundo, pelos gritos e risadas pelos corredores e durante o café das 16 horas. Aos funcionários do IB-USP. Aos colegas da pós-graduação que se dispuseram a engrandecer minha carreira acadêmica, compartilhando momentos de catarse durante as disciplinas.

Ao meu pai Aderval, minha mãe Miriam e minha irmã Myslane, que entenderam que sair da zona de conforto foi necessário e libertador. Pelo apoio de todas as horas, inclusive nos momentos de insanidade e de perda. À Renata Kehdi e seu filho Heitor, que chegaram em maio de 2017 e desde então recheiam de diversos sentimentos meu cotidiano.

Nos bastidores, aos amigos e portos seguros! Da graduação: Riba, Aninha, Mama, Dani, Mags-Leleo-Alice-Abraão e Nandão. Da arqueologia: Dani, Hellbert, Nilo, Sarah, Pam e Rose. Da vida: Nah Friol e Vitão, Pam, Mérizona, Mércia, Inez, Léti, Juan, Thomaz, Taniel, Diego, André Cunha e Gefferson. Aos Soláricos: Curió, Êre, Japa, Baïaman, Cando, Emmanuel, Hilde, Bono, Ciceronha, Pinto, Somália, Ted, Vini, Boleta, Crispim e respectivas. Aos membros do NEGE: Profa. Dra. Roberta Marques, Prof. Dr. Mathias Weller e seus demais orientandos da Biologia e do mestrado em Saúde Pública. Aos amigos do Genoma incluindo agregados do time Kok: Ines, Thalita, Michel e Bruno. Aos paquistaneses: Irfan, Aamir, Usama e Saqib. À todos os pacientes, em especial os spoanzinhos de Serrinha dos Pintos, à Zilândia, e seus pais Dona Helo e Seu Lenildo.

A todos que fazem parte do Centro de Estudos do Genoma Humano pelo ambiente intelectual e pelo suporte financeiro. Agradeço a CAPES pela bolsa de doutorado.

ÍNDICE

AGRADECIMENTOS	5
LISTA DE FIGURAS	9
APRESENTAÇÃO	11
RESUMO	14
ABSTRACT	15
1. INTRODUÇÃO	16
1.1. FUNDAMENTAÇÃO E REVISÃO DA LITERATURA	20
1.1.1. ESTUDOS DE ANCESTRALIDADE	20
1.1.1.1. DNA MITOCONDRIAL	20
1.1.1.2. REGIÃO NÃO RECOMBINANTE DO CROMOSSOMO Y	23
1.1.1.3. DNA DOS AUTOSSOMOS	25
1.1.2. CONSANGUINIDADE	26
1.1.2.1. ÍNDICES DE FIXAÇÃO (Estatística F de Wright)	28
1.1.3. POPULAÇÕES DO NORDESTE BRASILEIRO	29
1.1.3.1. OS INDÍGENAS DO TERRITÓRIO ATUAL DO INTERIOR DOS ESTADOS DA PARAÍBA E DO RIO GRANDE DO NORTE	30
1.1.3.2. AS PRIMEIRAS INVASÕES E OCUPAÇÕES DA CAPITANIA DO RIO GRANDE DO SÉCULO XVI AO INÍCIO DO SÉCULO XVIII	32
1.1.3.3. OCUPAÇÃO DO SERTÃO NA SEGUNDA METADE DO SÉCULO XVIII E INÍCIO DO SÉCULO XIX	35
1.1.3.4. INFORMAÇÕES DEMOGRÁFICAS DO ATUAL ESTADO DO RIO GRANDE DO NORTE E HISTÓRIA DOS MUNICÍPIOS DO ALTO OESTE POTIGUAR	36
1.1.3.4.1. Rio Grande do Norte	36
1.1.3.4.2. Alto sertão potiguar	37
2. OBJETIVOS	41
3. MÉTODOS	41
3.1. CASUÍSTICA	41
3.2. ASPECTOS ÉTICOS	42
3.3. COLETA DE SANGUE E EXTRAÇÃO DE DNA GENÔMICO	42
3.4. ESTUDO I – ANCESTRALIDADE LOCAL E GLOBAL DE INDIVÍDUOS COM DOENÇAS GENÉTICAS	42
3.4.1. GENOTIPAGEM	43
3.4.2. FILTRAGEM	43
3.4.3. BANCOS DE DADOS POPULACIONAIS	44

3.4.4. PCA E ADMIXTURE	45
3.5. ESTUDO 2 – MATRILINHAGEM DE INDIVÍDUOS COM MUTAÇÃO EM <i>KLC2</i>	48
3.5.1. SEQUENCIAMENTO DA REGIÃO HVS-1 e HVS-2	48
5. CONSIDERAÇÕES FINAIS	49
6. REFERÊNCIAS	52

LISTA DE FIGURAS

- Figura 1.** Distribuição mundial dos haplogrupos de DNA mitocondrial (Fonte: KIVISILD, 2015). **23**
- Figura 2.** Distribuição mundial dos haplogrupos da região não-recombinante do cromossomo Y, baseando-se nas populações do banco de dados do projeto 1000 genomas. Imagem retirada de (JOBLING; TYLER-SMITH, 2017), em marrom claro Haplogrupo A, em marrom escuro R, em lilás O, e em cinza Q. **25**
- Figura 3.** Variação da porcentagem de uniões consanguíneas no mundo (BITTLES, 2015 - disponível em consang.net). **29**
- Figura 4.** Distribuição das extintas tribos indígenas do Rio Grande do Norte adaptado de Medeiros Filho (1984). **33**
- Figura 5.** Quatro frentes de colonização da capitania do Rio Grande no território do atual estado do Rio Grande do Norte, adaptado de Monteiro (2012). **35**
- Figura 6.** Microrregião alto oeste potiguar em cinza e municípios que apresentam afetados da síndrome SPOAN em preto: (a) São Miguel, (b) Coronel João Pessoa, (c) Doutor Severiano, (d) Encanto, (e) Pau dos Ferros, (f) Serrinha dos Pintos, (g) Martins. **39**

LISTA DE TABELAS

Tabela 1. Indivíduos afetados por doenças genéticas de herança autossômica recessiva. **45**

APRESENTAÇÃO

Este é um estudo de genética médica populacional e iniciou-se como parte de um projeto de doutorado intitulado “Estudo de evidências e modelos explicativos para herança em consulentes de serviços de aconselhamento genético” desenvolvido pela pesquisadora Silvana Santos (processo FAPESP 00/03660-1) entre 1999 e 2003. Durante esta pesquisa sobre explicações para os fenômenos hereditários foi identificada uma família com vários afetados por uma doença neurodegenerativa pela que tinha origem na região do alto oeste do estado do Rio Grande do Norte (RN). A investigação clínica e genética foi iniciada em 1999 e a descrição de uma nova doença genética ocorreu somente em 2005, quando ela foi denominada Síndrome SPOAN em virtude da combinação dos seus sintomas: paraplegia espástica, atrofia óptica e neuropatia.

A investigação foi ampliada a partir de 2006, com financiamento do Centro de Pesquisa, Inovação e Difusão (CEPID) (CEGH-CEL, processo FAPESP 13/08028-1); do Centro de Estudos do Genoma Humano (CEGH, processo FAPESP 98/14254-2), e do CNPq por meio do edital de Genética Clínica em 2006. Durante os anos de 2006 e 2008, houve um grande investimento para identificação de mais indivíduos com síndrome SPOAN a fim de completar os estudos de ligação necessários para a identificação do gene responsável pela doença. Por essa razão, a Profa. Silvana Santos fez a busca ativa de mais afetados em cinco municípios da região do alto oeste potiguar. Ao final de 2008, havia o registro de mais de 70 indivíduos acometidos pela doença em mais de dez municípios da região do alto oeste potiguar. Para realizar a investigação clínica desses pacientes, foram realizadas várias missões de pesquisa com a participação da equipe do Prof. Dr. Fernando Kok, neurologista do Hospital das Clínicas da Universidade de São Paulo, orientador deste projeto. E amplo apoio do CEGH e do laboratório da Profa. Dra. Mayana Zatz.

A investigação clínico-genética da síndrome SPOAN possibilitou a descrição clínica e estudo de ligação (MACEDO-SOUZA et al., 2005), o refinamento de seu locus (MACEDO-SOUZA et al., 2009), e mais recentemente a identificação de *KLC2* (MELO et al., 2015) como o responsável pela doença. Em 2008, foi feita a publicação de outra doença genética, denominada inicialmente como síndrome Santos (SANTOS et al., 2008), com seis afetados membros de uma família do município de Riacho de Santana, no Rio Grande do Norte, que apresentam baixa estatura devido a agenesia fíbula, oligodactilia e outras anomalias de membros. A identificação do gene responsável, *WNT7A* deu-se muito recentemente (ALVES et al., 2017). De 2008 a 2016, estudos de epidemiologia e genética comunitária foram realizados na Paraíba, envolvendo famílias com vários indivíduos acometidos por grave

deficiência intelectual (DI), permitiu a identificação de novas formas de DI de herança autossômica recessiva, associadas a *MED25* e *IMPA1* (FIGUEIREDO et al., 2015, 2016).

Essas investigações permitiram a coleta de material biológico e genealógico de mais de 100 indivíduos afetados e 300 de seus familiares que residem na região oeste dos estados do Rio Grande do Norte e Paraíba, no nordeste brasileiro. Foram identificados indivíduos afetados pela síndrome SPOAN nos municípios de São Miguel, Coronel João Pessoa, Doutor Severiano, Encanto, Pau dos Ferros, Serrinha dos Pintos, Martins, Riacho de Santana, no estado do Rio Grande do Norte; e em Lagoa, na vizinha Paraíba. Indivíduos com a síndrome Santos foram identificados no município de Riacho de Santana, RN; os pacientes com deficiência intelectual associada a *MED25* eram procedentes de Catolé do Rocha, PB e de *IMPA1* de Brejo dos Santos, PB. Todos esses municípios tiveram na sua formação contribuição de populações europeias, indígenas e africanas.

A identificação de novas doenças genéticas no nordeste brasileiro está associada com a tradição de casamentos consanguíneos. Nosso grupo de pesquisa retomou, por assim dizer, as pesquisas do Prof. Dr. Newton Freire Maia da metade do século XX. Com ajuda dos agentes comunitários de saúde, foram levantadas as frequências de uniões consanguíneas em municípios da Paraíba entre 2010 e 2012. Em média, 20% dos casamentos nestes municípios da Paraíba ocorrem entre pessoas aparentadas, sendo que esses valores podem chegar a quase 40%, aproximando-se das prevalências de alguns países do Oriente Médio. Esses dados deram origem a uma pergunta que vem sendo debatida há muito tempo no nosso grupo de pesquisa: por que o Nordeste conservou a tradição de casamentos consanguíneos?

Que fatores históricos e culturais explicam as elevadas frequências de casamentos consanguíneos no sertão do nordeste brasileiro? Essa foi a pergunta que deu origem ao meu projeto de doutorado, com o qual ingressei no Programa de Pós-Graduação em Genética e Biologia Evolutiva em fevereiro de 2014, com orientação do Prof. Dr. Rui Sérgio Sereni Murrieta (IB-USP) e a coorientação da Profa. Silvana Santos. O projeto propunha basicamente investigar a ancestralidade de populações do Nordeste onde havia alta consanguinidade e prevalência de doenças genéticas recessivas para testar a hipótese oferecida pelos historiadores. Os registros historiográficos apontam para uma forte influência de colonização Sefardita do sertão nordestino. Em virtude do trabalho ter um foco mais voltado para genética médica do que para a antropologia ou estudos de natureza biocultural, então a orientação do projeto passou para as mãos do Prof. Dr. Fernando Kok em abril de 2015.

Após dois anos, durante o exame de qualificação, a banca examinadora sugeriu mudar a pergunta e os métodos do projeto, que passaram a focar a ancestralidade de indivíduos afetados por doenças genéticas recessivas. A razão disso é que havia na literatura um debate sobre a existência de assinatura genéticas Sefarditas que pudesse ser investigada nessas populações. Por um lado, havia publicações descrevendo essas assinaturas em populações de Portugal e outras regiões no norte da África e Oriente Médio. Por outro lado, havia outros autores que defendiam a tese de que a miscigenação entre sefarditas e “cristãos velhos” foi muito intensa, não sendo possível compreender os sefarditas como grupo étnico que o diferencie de outros grupos. Então, minha pergunta passou a focar na ancestralidade das mutações e das famílias com doenças genéticas em vez das populações do Nordeste. Quando e onde surgiram as mutações que causam essas doenças raras no interior do Nordeste? Com as novas abordagens de inferência de ancestralidade local associadas aos estudos de idade de mutação, podemos tentar responder essa questão. Os estudos de ancestralidade global e local, e de matrilinearidade foram realizadas, portanto, ao longo de 2016 e 2017.

Esta tese foi organizada no formato de artigos. Na introdução, fizemos uma revisão de literatura sobre a genética de populações, ancestralidade, consanguinidade e história das populações do interior do Nordeste. Na sessão sobre materiais e métodos, descrevemos a casuística escolhida e os procedimentos técnicos. Os resultados consistem em um artigo científico sobre a ancestralidade local de quatro mutações que causam doenças genéticas de herança autossômica recessiva. O desfecho compõe as conclusões com perspectivas futuras e as limitações dessa tese visando aprofundar as questões no futuro para responder outras questões mais refinadas.

No período do doutoramento, tive oportunidade de colaborar com outros grupos de pesquisa, publicando um capítulo do livro “Assim caminhou a humanidade” organizado pelo Prof. Dr. Walter Neves (FARIAS et al., 2015), e quatro artigos, dois na *Clio* (Série Arqueológica) (FARIAS et al., 2016; MÜTZENBERG et al., 2015), um na *Fundamentos* (SULLASI et al., 2017) e um na *Annals of Human Genetics* (BARBOSA et al., 2017).

RESUMO

A idade da mutação e a ancestralidade local de segmentos cromossômicos que contém mutações associadas a doenças autossômicas recessivas em populações miscigenadas brasileiras continuam desconhecidas. Adicionalmente, a determinação dos níveis de endogamia continua grosseiramente estimados por abordagem genealógica e nada se sabe sobre a linhagem materna das famílias com indivíduos com doenças autossômicas recessivas.

A tese está dividida em dois estudos. O primeiro visou calcular os níveis de endogamia pela abordagem de grandes regiões em homozigose, estimar a idade da mutação e inferir a ancestralidade local para determinar a origem de cada segmento cromossômico contendo mutações em *KLC2*, *IMPA1*, *MED25* e *WNT7A*.

Foram genotipadas amostras de 19 pacientes com doenças autossômicas recessivas. Os dados genotipados passaram por distintos procedimentos de filtragem para estimar os níveis de endocruzamento por abordagem molecular, para calcular a idade das mutações por meio de haplótipos, para inferir a ancestralidade global utilizando ADMIXTURE e Análise de Componentes Principais, e de forma refinada foi inferida a ancestralidade local pelo software RFMix combinando os dados de nossas amostras com os dados do projeto 1000 genomas (*one thousand genome project - 1KGP*) e projeto de diversidade genômica Simons (*Simons genomic diversity project - SGDP*). Os resultados apresentados não diferiram bastante do esperado para populações endogâmicas e miscigenadas do nordeste brasileiro de acordo com a literatura. A mutação em *KLC2* mostrou-se mais antiga em relação às demais.

Em um segundo estudo, para os indivíduos com mutação em *KLC2* prosseguimos com a inferência dos haplogrupos de DNA mitocondrial para determinar a origem da linhagem materna das famílias. Onze pacientes com mutação em *KLC2* foram escolhidos para sequenciamento de Sanger da região controle do DNA mitocondrial para determinação da matrilinearidade. Foi encontrado um possível haplogrupo fundador que pode corroborar narrativas históricas de entrada de mães sefarditas no litoral do Nordeste junto com os holandeses.

Ambos trabalhos apresentam como novidade a contribuição para identificar a origem e entender como essas doenças são mantidas e dispersas em populações do Brasil e do mundo. Em especial, a matrilinearidade exibiu evidências de ligação entre as mães de famílias com indivíduos com mutação em *KLC2* e mães de judeus sefarditas portugueses.

ABSTRACT

The mutation age and local ancestry of chromosomal segments harbouring mutations associated with autosomal recessive (AR) disorders in Brazilian admixed populations remain unknown; additionally, inbreeding levels for these affected individuals continue to be roughly estimated by genealogical approach, and the maternal lineage of the families with affected individuals is unknown.

The thesis is divided into two studies. The first one aimed to calculate inbreeding levels by runs of homozygosity approach, estimating the age of the mutation and inferring the local ancestry to determine the origin of each chromosomal segment containing mutations in *KLC2*, *IMPA1*, *MED25* and *WNT7A*.

Samples from 19 patients with autosomal recessive diseases were genotyped. The genotyped data underwent different filtering procedures to estimate the inbreeding levels by molecular approach, to calculate the age of the mutations by means of haplotypes, to infer the global ancestry using ADMIXTURE and Principal Component Analysis, and to infer refined the local ancestry by the RFMix software combining the data from our samples with the one thousand genome project (1KGP) and the Simons genomic diversity project (SGDP). The results presented did not differ significantly from those expected for inbred and admixed populations of the Brazil Northeast according to the literature. The mutation in *KLC2* was older in relation to the others.

In a second study, for individuals with a *KLC2* mutation, we proceeded with the inference of mitochondrial DNA haplogroups to determine the origin of the families' maternal lineage. Eleven patients with mutation in *KLC2* were chosen for Sanger sequencing of the mitochondrial DNA control region for determination of matrilineage. It was found a possible founding haplogroup that can corroborate historical narratives of Sephardic mothers' influx along the coast of the Northeast with the Dutch.

Both papers present as a novelty the identification of the origin and the understanding of how these diseases are maintained and dispersed in populations of Brazil and the world. In particular, the matrilineage exhibited an evidence of kinship among mothers of families with individuals with mutation in *KLC2* and mothers of portuguese Sephardic jews.

1. INTRODUÇÃO

A restrição às relações sexuais entre pessoas aparentadas é um dos mais antigos tabus do Ocidente (BITTLES, 2015). Hoje, cerca de 10% das uniões de toda população mundial são consanguíneas (BITTLES; BLACK, 2010). Os casamentos entre primos em primeiro grau, por exemplo, em comunidades católicas são desencorajados, sendo necessário requerer uma autorização às autoridades eclesiásticas para a celebração do matrimônio (BITTLES, 2001). Por outro lado, nas sociedades islâmicas e entre judeus, especialmente no Oriente Médio, essas uniões são muito valorizadas e podem até ser predominantes em algumas comunidades (BITTLES, 2015). Os primos em primeiro grau compartilham $1/8$ de seus genes, herdados de um ancestral comum a ambos, e a progênie desse casal terá $1/16$ de seus *loci* gênicos idênticos por descendência, o que é usualmente estimado por meio do coeficiente de endocruzamento (F) (BITTLES, 1994, 2002, 2015). Esse tipo de união pode ocorrer por várias gerações nas comunidades que valorizam essas práticas, o que acarreta um efeito cumulativo que pode levar ao aumento da taxa de homozigose genômica, aumentando assim a prevalência de doenças genéticas (BITTLES, 2002, 2013).

No Brasil, as populações do Nordeste brasileiro apresentam elevadas taxas de consanguinidade. Em meados do século passado, a frequência dessas uniões no Nordeste (9%) era quinze vezes maior que no Sudeste (0,62%) (DA FONSECA; FREIRE-MAIA, 1970; FREIRE-MAIA, 1957, 1990). Recentemente, estudos abrangendo 44 municípios nordestinos mostraram que de 6% a 41% das uniões ainda ocorrem entre pessoas aparentadas, o que contribui para elevar a prevalência de pessoas com deficiência causadas por doenças genéticas (SANTOS et al., 2010; WELLER et al., 2012).. Alguns fatores sociodemográficos, como a densidade populacional, e educação, longevidade e renda (IDH), já foram examinados e refutados por Weller e cols (2012) na tentativa de se compreender o que precisamente explicaria os diferentes padrões de consanguinidade observados no Brasil. Os autores cogitaram a hipótese de que essas práticas poderiam estar associadas a fatores culturais e históricos (BITTLES; BLACK, 2010; KLITZ et al., 2010).

Na região Nordeste foram descritas novas doenças genéticas de herança recessiva. Uma delas, por exemplo, afeta mais de 70 indivíduos com ancestrais originários do interior do estado do Rio Grande do Norte e Paraíba, todos eles filhos de casais aparentados (MACEDO-SOUZA et al., 2005, 2009). Essa doença, chamada de Síndrome SPOAN, associa paraplegia espástica, neuropatia periférica e atrofia óptica. Ela tem herança autossômica recessiva e é causada por uma deleção na região não-codificante de *KLC2* (MELO et al., 2015). Além desse *cluster* encontrado no Nordeste, foram identificados mais

dois irmãos com esta doença no Rio Grande do Sul, uma paciente originária de Minas Gerais e finalmente, dois irmãos no Egito, todos frutos de uniões consanguíneas.

Outras duas doenças genéticas de herança recessiva que causam deficiência intelectual foram identificadas em duas famílias, uma de Brejo dos Santos e outra de Catolé do Rocha, no estado da Paraíba. A primeira apresenta nove afetados com deficiência intelectual associada a comportamentos agressivos e é causada por mutação bialélica em *IMPA1* (*Inositol Monophosphatase 1*), que codifica proteína responsável pela biotransformação do inositol monofosfato em inositol (FIGUEIREDO et al., 2016). No outro caso, foram identificados sete indivíduos que apresentavam dismorfias faciais e deficiência intelectual grave associadas a mutação bialélica em *MED25* (*Mediator Complex Subunit 25*) (FIGUEIREDO et al., 2015).

No município de Riacho de Santana, no estado do Rio Grande do Norte, que fica bem próximo de Catolé do Rocha e Brejo dos Santos na Paraíba, foi identificada uma família com seis indivíduos que apresentavam agenesia ou hipoplasia de fíbula com outras anomalias de membros causadas por mutação em *WNT7A* (*Wnt family member 7A*). Esta condição foi denominada síndrome Santos (ALVES et al., 2017; SANTOS et al., 2008).

Qual a origem e como se deu a dispersão dessas mutações em *KLC2*, *IMPA1*, *MED25* e *WNT7A*? Sabemos que a população brasileira é tri-híbrida, e que os europeus contribuíram significativamente para a formação das populações nordestinas (MOURA et al., 2015). Os europeus apresentam uma maior frequência de variantes deletérias (variação genética deletéria maior do que africanos, pois há um acúmulo de alelos danosos provavelmente resultado de um gargalo populacional sofrido pelos humanos durante o evento de migração para fora da África (LOHMUELLER et al., 2008). Mas, como se deu a colonização do Nordeste? Essas mutações são autóctones ou alóctones? Se essas variantes deletérias tiverem origem ameríndia, como explicar o caso de dois egípcios afetados com a síndrome SPOAN? Se forem européias, elas surgiram há quanto tempo? Há outras populações em risco de ter essas algumas destas doenças genéticas raras?

Nos últimos dois mil anos, a Europa, e particularmente a Península Ibérica se mostrou palco de interações e migrações de populações de origens geográficas diferentes e características culturais e religiosas bem particulares: católicos, judeus sefarditas e muçulmanos norte-africanos (ADAMS et al., 2008; CAMPBELL et al., 2012). Apenas quinze quilômetros do estreito de Gibraltar separam a Península Ibérica do Norte da África, e essa região foi marcada por migrações frequentes. Em 711, a Península foi invadida pelo exército Berbere sob o comando Árabe, o qual conquistou toda a região da Espanha e Portugal, permanecendo em seu controle por mais de cinco séculos até sua expulsão em 1492

(ADAMS et al., 2008). Com as Cruzadas, a partir de 1492 na Espanha e 1496 em Portugal, até o início do século XVII, os mouros foram expulsos ou convertidos pela Inquisição Católica apoiada pelos reinos de Castela e Aragão, principalmente. Estima-se que ao todo, cerca de 200.000 mouros foram exilados da região (ADAMS et al., 2008).

Durante esse período, na Península Ibérica, existiam também populações de judeus denominados sefarditas, derivado do nome *Sepharad*, que é Espanha em hebraico (ADAMS et al., 2008; NOGUEIRO et al., 2015a, 2015b; VELEZ et al., 2012). Há registros de sua entrada na península desde 70 d.C. com a queda da Judeia para o Império Romano e, posteriormente, com as migrações devido às perseguições de outros povos, como os visigodos. Com a ocupação árabe em 711, houve um período de convivência, sendo estimada uma população de 100.000 sefarditas na Espanha na metade do século VIII. No final do século XIV, uma onda de massacres afetou vários bairros judeus de cidades ibéricas, particularmente Barcelona e Girona (ADAMS et al., 2008). No final do século XV, estimou-se que havia cerca de 400.000 judeus na Espanha, dos quais mais de 160.000 foram expulsos, ocupando o entorno do Mar Mediterrâneo, enquanto os que ficaram foram forçados à conversão ao cristianismo e foram chamados de “judeus convertidos ou conversos” na Espanha e de cristãos-novos em Portugal (ADAMS et al., 2008).

Há algumas evidências históricas de que esses cristãos novos colonizaram o semiárido nordestino (DE ASSIS, 2010, 2012, 2013; FALBEL, 2008; FEITLER, 2011; MIELE, 2008; NOVINSKY, 1972), inclusive a Paraíba (NOVINSKY et al., 2015). Entretanto, esses registros são esparsos devido à necessidade daquelas famílias ocultarem sua identidade. As principais fontes de informação histórica são correspondências e processos do Tribunal da Santa Inquisição que foram arquivados pela Igreja Católica Romana. Naquela época, tais acusações eram fundamentadas em observações como o fato de não se ingerir carne de porco; não se misturar, em refeição, carne e leite; o descanso aos sábados – em vez dos domingos, como feito pela tradição cristã (ROITMAN; OTHERS, 2009). As práticas funerárias também eram observadas para fundamentar as acusações, pois os judeus usavam mortalhas ou cobriam as sepulturas com pedras (MIELE, 2008). Entretanto, essas evidências são bastante frágeis para se estabelecer, com mais precisão, a influência dos cristãos-novos sobre a estruturação genética e cultural das populações do Nordeste brasileiro.

Durante o processo de colonização dessa região, os grupos de católicos portugueses (cristãos-velhos) e cristãos-novos formavam uma população de 2900 indivíduos brancos na região da atual capital de Pernambuco durante o domínio holandês (1630-1654). Desse conjunto, 720 (~25%) eram judeus e, frequentemente, formavam famílias de

um homem, uma mulher e duas crianças para cada família, de acordo com informações obtidas de matrimônios da igreja católica e documentos arquivados na Kahal zur Israel, a primeira sinagoga das Américas (WOLFF; WOLFF, 1979). Os “cristãos velhos” e novos respeitavam os estatutos de sangue baseando-se em ordens da Inquisição. Os estatutos proibiam casamentos entre católicos e judeus portugueses; apesar dessa regra os homens cristãos-novos que vieram solteiros ao Brasil, muitas vezes se uniram com as nativas e cristãs-velhas a fim de ocultar sua identidade (NOVINSKY, 1972). Portanto, as práticas atuais de casamentos consanguíneos poderiam ser traços culturais deixados pela atuação da Inquisição de 300 a 400 anos atrás.

Ambos os grupos também encontraram os índios tupi no litoral e os tapuios cariris e tarairiús nos sertões, subdivididos em várias tribos tanto no Rio Grande do Norte (DOS SANTOS JÚNIOR, 2008) quanto na Paraíba (SANTOS, 2009). Este contato foi marcado por várias facetas, de alianças a guerras, de extermínio à miscigenação. Hoje sabemos que a colonização ocorreu provavelmente por duas rotas, conhecidas por “sertões de dentro” e “sertões de fora”. A primeira considera a trajetória do rio São Francisco partindo da foz, entre os estados de Alagoas e Sergipe, até os estados da Bahia e Pernambuco, o que facilitou a interiorização no sentido do Piauí e Ceará (DOS SANTOS JÚNIOR, 2008). A segunda rota teve início pelo litoral de Pernambuco seguindo o curso dos rios no sentido dos sertões da Paraíba e Rio Grande do Norte.

No litoral de Pernambuco aportaram diversos escravos africanos que, junto com os indígenas foram forçados a trabalhar no cultivo de cana de açúcar em toda a borda dos atuais estados do Rio Grande do Norte, Paraíba e Pernambuco. Cerca de 700 mil africanos desembarcaram durante os séculos 16 e 17, alcançando a um total de mais 5 milhões no final do século XIX, incluindo os nascidos no Brasil e os importados. Poucos deles, no entanto, migraram para o interior que foi formado basicamente por descendentes de europeus, geralmente homens, e mulheres índias (SCHAAN et al., 2017).

Vimos, portanto, que o Nordeste brasileiro tem uma rica história demográfica com múltiplas ocupações pouco documentadas. Adicionalmente, devido à tradição de casamentos consanguíneos e efeito de fundador, algumas doenças genéticas que causam diferentes formas de deficiência têm sua prevalência aumentada nessas populações, como é o caso da síndrome SPOAN. Neste trabalho, tomando a documentação histórica como ponto de partida para entender a colonização do interior do Nordeste, busquei explorar múltiplas possibilidades de análise e interpretação de dados moleculares, elucidando assim a história molecular de doenças autossômicas recessivas reunindo informação sobre as idades das mutações, miscigenação dos afetados, endogamia e matrilinearidade.

1.1. FUNDAMENTAÇÃO E REVISÃO DA LITERATURA

1.1.1. ESTUDOS DE ANCESTRALIDADE

A diversidade das populações humanas e padrões de migração têm sido estudados utilizando várias metodologias de diferentes campos do conhecimento humano como história, arqueologia e genética (ADAMS et al., 2008). A genética tem se debruçado sobre a compreensão de como os processos evolutivo e padrões de variação do DNA desenharam a história populacional dos humanos modernos (SCHRAIBER; AKEY, 2015). A emergência do uso técnicas de sequenciamento de genomas e exomas completos na análise de inferência de ancestralidade tem tornado os estudos mais robustos, detalhando os processos de miscigenação com melhores estimativas de datação, variação do tamanho populacional total ao longo de diversas gerações, e hibridização dos humanos anatomicamente modernos com neandertais e denisovanos (SCHRAIBER; AKEY, 2015).

A história de uma população é inferida a partir do uso de distintos marcadores moleculares. Os mais tradicionais são os estudos de matrilinearidade, por meio da determinação de haplótipos e haplogrupos do DNA mitocondrial e estudos de patrilinearidade, através da genotipagem de marcadores moleculares na região não recombinante do cromossomo Y, e de ambas as linhagens estudando-se o cromossomo X. Entretanto, mais recentemente, com o avanço das tecnologias de determinação de genótipos em SNPs autossômicos por meio de *arrays* e de sequenciamento de nova geração, utilizados em associação a ferramentas como análise estrutural (*structure-like analysis*), de componentes principais (PCA - *principal component analysis*), e inferência de ancestralidade local (LAI - *local ancestry inference*) houve enorme refinamento das análises para inferência de ancestralidade.

1.1.1.1. DNA MITOCONDRIAL

O DNA mitocondrial (DNAMt) tem herança exclusivamente materna, taxas elevadas de mutação, grande número de cópias por célula e ausência de recombinação, o que o torna ferramenta privilegiada para estudos da ancestralidade de linhagens maternas (KIVISILD, 2015), permitindo inclusive a inferência de ancestralidade a partir de DNA obtido de vestígios ósseos e dentários de populações pré-históricas (GONÇALVES et al., 2013). O DNAMt apresenta forma circular e compacta, com 16.569 pares de bases, e não apresenta histonas. Ele codifica 13 proteínas que fazem parte dos complexos de cadeia de transporte de elétrons e ATP sintase, e pela transcrição de 22 moléculas de RNA relacionadas ao

processo de tradução mitocondrial (van Oven e Kaiser, 2009). Em consequência das elevadas taxas de mutação, algumas regiões são altamente polimórficas. As combinações das variações de bases de nucleotídeos (SNPs) no DNAm constituem haplótipos que podem ser compartilhados por populações e etnias específicas, organizados em haplogrupos ou clados filogenéticos (WALLACE, 2007; YANG et al., 2010).

A partir de meados da década de 90, foi criada uma nomenclatura para organizar os diferentes haplogrupos que caracterizam a variação encontrada em linhagens humanas asiáticas e americanas, as quais receberam os rótulos de A a G. Para os europeus, foram designados os haplogrupos de H a K; e para os africanos, o grupo L. Essa nomenclatura foi usada para construção de um banco de dados filogenéticos com dados muito robustos sobre as diferentes populações humanas a partir da linhagem materna, como o Phylotree (www.phylotree.org). Os padrões de variação de DNAm foram associados com modelos de datação molecular e auxiliaram na compreensão da história evolutiva das populações humanas (Figura 1). Mais recentemente, com uso de técnicas de sequenciamento de nova geração, tornou-se possível o conhecimento de toda a sequência do DNAm de muitos indivíduos, ampliando o conhecimento sobre a estrutura genética dessas populações (KIVISILD, 2015).

Figura 1. Distribuição mundial dos haplogrupos de DNA mitocondrial (Fonte: KIVISILD, 2015).

Em Portugal, os estudos de DNAm^t evidenciaram que os haplogrupos mais comuns na região são o H (25,3%), o R (12,7%), o U (14%) e o T (11%). Estes grupos representam mais de 60% do total da diversidade genética matrilinear (MARQUES et al., 2015). Há também uma estratificação geográfica, sendo, por exemplo, o haplogrupo I exclusivo da região sul, com uma frequência de mais de 9% e o haplogrupo K é mais frequente na região central e norte (SANTOS et al., 2003). Na região da Andaluzia, foi observada alta frequência dos haplogrupos U6 e L, que mostram a influência de linhagens do Norte da África e da África subsaariana (BREHM et al., 2003; GONZÁLEZ et al., 2003; PEREIRA; PRATA; AMORIM, 2000; SANTOS et al., 2003).

Nogueiro e cols. (2014) amostraram 57 indivíduos autodenominados judeus, residentes em Bragança e 30 residentes em Belmonte, para determinar os haplogrupos mais frequentes. Buscaram verificar se havia haplogrupos específicos, utilizando como referência de comparação um banco de dados de 766 judeus (BEHAR et al., 2008) e 844 não judeus do oeste da Península Ibérica (PEREIRA; PRATA; AMORIM, 2000). Ao todo, foram identificados 32 haplogrupos diferentes, sendo identificadas, com significância estatística, como linhagens fundadoras os haplogrupos HV0b, N1, T2b11, T2e e U2e. Esses grupos responderam por quase 40% da diversidade genética da amostra. Em Belmonte, 93% das sequências de DNA analisadas de cristãos-novos pertenciam ao haplogrupo HV0b. Em Bragança, os haplogrupos N1 e U2e ocorreram em frequências maiores entre judeus quando comparados à população portuguesa. Devido à raridade do grupo N1, inferiu-se que sua origem deve ser anterior às primeiras diásporas judaicas (NOGUEIRO et al., 2015b). O haplogrupo T2e é considerado uma linhagem fundadora de uma comunidade tradicional sefardita búlgara (BEHAR et al., 2008), tendo sido também identificada na Turquia, norte do México e sul do Texas, em comunidades de descendentes de sefarditas (BEDFORD, 2012).

Alguns autores (BEDFORD; YACOB, 2015; MARCUS; EBEL; FRIEDMAN, 2015) têm criticado as generalizações feitas para toda a população de Portugal, ou até mesmo para as duas comunidades estudadas de Belmonte e Bragança, devido ao número reduzido de indivíduos genotipados por Nogueiro e cols. (2014, 2015). Os haplogrupos de DNAm^t mais frequentes em certas populações, diferentes das populações do entorno não são suficientes para explicar a complexa história de migração dos diferentes grupos de judeus atuais (asquenazitas, sefarditas, coxins, benes, caucasianos, iranianos, iraquianos, yemenitas, etíopios, uzbequistãos e marroquinos). Para sanar essas lacunas, têm sido empregadas as técnicas de sequenciamento de nova geração e a análise de marcadores autossômicos do tipo SNPs (analisados em *arrays*), assim como têm sido ampliadas as amostras de indivíduos genotipados.

No Brasil, há uma distribuição heterogênea dos haplogrupos para as diferentes regiões que forma um gradiente de contribuição com maior frequência de haplogrupos europeus no Sul (67.2%); africanos no Sudeste (47.1%); e nativo-americanos no Nordeste (47.5%) e Norte (59.2%) (SCHAAN et al., 2017). Em estudos mais antigos, os haplogrupos mais frequentes descritos no Nordeste eram de africanos (revisão em (SALZANO; SANS, 2014), entretanto as amostras foram colhidas em grandes capitais com particularidades históricas que tornam as generalizações sobre a ancestralidade enviesadas, por exemplo, na capital da Bahia, Salvador (ALVES-SILVA et al., 2000), onde houve maior entrada de mulheres africanas no Brasil.

1.1.1.2. REGIÃO NÃO RECOMBINANTE DO CROMOSSOMO Y

O cromossomo Y originou-se de um par de cromossomos autossomos homólogos, e seu processo de divergência ocorreu há 180 milhões de anos com o desenvolvimento do proto-cromossomo Y que progressivamente desligou a sua recombinação com o cromossomo X. Sem essa troca gênica, o tamanho do cromossomo degenerou de 150Mb para 60Mb (JOBILING; TYLER-SMITH, 2017). De herança patrilinear, o cromossomo Y possui uma região não recombinante com poucos genes (95%), sendo mais da metade dela constituída por sequências repetidas em tandem. Essa região é muito polimórfica, contendo milhares de SNPs e microssatélites (STRs) que podem ser usados para definir haplótipos e haplogrupos, e traçar a ancestralidade em linhagem paterna (JOBILING; TYLER-SMITH, 2003, 2017).

A distribuição dos haplogrupos do cromossomo Y no mundo é dada por um padrão geográfico dividido por continentes (Figura 2). Na África, houve a expansão com alta frequência do haplogrupo A0 de 190 a 110 mil anos, logo após foi substituído pelo haplogrupos B, C e E que estão em maior frequência nas populações atuais deste continente. O evento da saída de humanos modernos para fora da África promoveu uma dispersão do haplogrupo R1 em maior frequência nas populações europeias e do Oriente Médio. O haplogrupo O é prevalente em homens asiáticos; enquanto o haplogrupo Q1 é mais frequente em populações do Novo Mundo.

Figura 2. Distribuição mundial dos haplogrupos da região não-recombinante do cromossomo Y, baseando-se nas populações do banco de dados do projeto 1000 genomas. Imagem retirada de (JOBLING; TYLER-SMITH, 2017), em marrom claro Haplogrupo A, em marrom escuro R, em lilás O, e em cinza Q.

Em Portugal, Nogueiro e cols. - (NOGUEIRO et al., 2010) amostraram 57 homens autodesignados como judeus sefarditas da região de Trás-os-Montes aplicando 16 marcadores STRs e de SNPs, a fim de analisar a distribuição de haplogrupos de Y neste grupo e comparar com a população de não-judeus da mesma região e de Portugal. Nos judeus foram encontradas maiores frequências dos haplogrupos T-M70, J1(xJ1a)-M267 e J2-M172 e entre os não judeus o haplogrupo prevalente foi o R1b1b2. Já Adams e cols. (2008), amostrando 1.140 homens, analisaram marcadores binários e de microssatélites, estudando a contribuição de judeus sefarditas e árabes muçulmanos na formação das populações da Península Ibérica. As análises da estrutura genética dessas populações da Península Ibérica indicaram 10% de ancestralidade norte africano e 19% de sefardita.

Regueiro e cols. (2015) - (REGUEIRO et al., 2015) utilizaram 17 marcadores de microssatélites e uma amostra de 1.233 homens de oito regiões diferentes da Península Ibérica, Norte da África e Península Arábica para estudar a dispersão de haplogrupos mais frequentes em árabes e bérberes, identificando basicamente os subhaplogrupos J1a2b-P58 e E1b1b1b1a-M81. A contribuição de haplogrupos europeus ultrapassa os 80% de frequência para a formação do cromossomo Y de homens de todas as regiões brasileiras em comparação às contribuições nativo-americanas e africanas (OLIVEIRA et al., 2014; PALHA et al., 2012; RESQUE et al., 2016). No Nordeste, com dados obtidos de 49

indivíduos do estado de Pernambuco, verificou-se a contribuição da colonização holandesa pela presença do haplogrupo 2 em 18% dos indivíduos amostrados (CARVALHO-SILVA et al., 2001). Outros estudos, envolvendo 247 indivíduos em Alagoas e 107 no Ceará, corroboraram a estimativa de 80% de contribuição europeia para o cromossomo Y (DE AZEVEDO et al., 2009; RESQUE et al., 2016).

1.1.1.3. DNA DOS AUTOSSOMOS

A variação genética é originada por mutações. Elas podem ser da escala de um cromossomo inteiro a milhões de pares de bases, tendo como exemplos trissomias e monossomias; de milhões a dezenas de milhões de pares de bases como duplicações, inversões e deleções; de milhares a centenas de pares de base como os minissatélites, sequência Alu, inserções line-1, deleções e inserções; além da escala de poucos pares de base como microssatélites, e a mais simples e pequena escala de variação que compreende os polimorfismos de base única também chamados de SNP (*single nucleotide polymorphism*, em inglês). Os SNPs são substituições, inserções e deleções de uma ou mais bases entre duas sequências homólogas de DNA (JOBILING et al., 2013).

Os SNPs informativos de ancestralidade, localizados nos 22 pares de autossomos, permitem a análise da miscigenação de um indivíduo, estimando as parcelas de contribuição de populações diversas por meio da identificação de blocos ancestrais (HENN et al., 2010; TANG et al., 2006). Um bloco ancestral é um segmento do cromossomo que apresenta SNPs informativos de ancestralidade derivados de uma certa população. A análise do DNA autossômico, por meio da genotipagem em SNPs por *arrays* pode avaliar, por exemplo, mais de 800 mil SNPs. Para inferir a ancestralidade, verifica-se a frequência dos SNPs nas populações parentais em bancos de dados de referência como o painel de diversidade geográfica humana (HGDP) ou o catálogo profundo de variação humana conhecido como “1.000 genomas” (HENN et al., 2010; TANG et al., 2006). Essas frequências de SNPs variam de uma população parental para outra, permitindo diferenciá-las.

Os dados obtidos pelo array de SNPs permitem inferir a ancestralidade global de indivíduos. A ancestralidade global é a porcentagem do conjunto dos genomas de uma população que deriva de cada uma das populações parentais (SANTOS et al., 2015). Para inferência de ancestralidade global são utilizados dois tipos de análise. Uma análise de agrupamentos (*clusters*) que é representada por um método chamado *structure-like*, e outra análise por estatística multivariada que permite reduzir e dimensionar os dados em duas e três dimensões por autovetores (JOBILING et al., 2013). O método *structure-like* apresenta uma série de populações ancestrais hipotéticas chamadas de *k* e esse número pode variar

de duas a dez populações, cada uma com uma cor diferente (ALEXANDER; NOVEMBRE; LANGE, 2009; PRITCHARD; STEPHENS; DONNELLY, 2000).

Outra forma de analisar a ancestralidade global é utilizando o método multivariado por componentes principais (*Principal Component Analysis* - PCA), permitindo reduzir múltiplas dimensões em dois ou três autovetores também chamados de componentes principais (CPs). Os CPs são independentes e agregam a variação presente em dados brutos de frequências alélicas. A PCA revela as relações entre um conjunto de populações ou indivíduos a partir da distância genética (JOBILING et al., 2013).

Um exemplo da aplicação dessas abordagens é o estudo de Mychaleckyj e cols. (2017) (MYCHALECKYJ et al., 2017), que amostraram 2.010 indivíduos brasileiros recrutados em Fortaleza e Crato (CE), Picos (PI), Ouricuri (PE), Cajazeiras, Sousa e Patos (PB). Foram empregados cerca de 818 mil SNPs utilizando o array LAT Axiom da Affymetrix para inferência de ancestralidade global e subestruturação populacional (F_{st}). O estudo identificou SNPs conservados por seleção natural em haplótipos nativo-americanos que influenciam a imunidade, o metabolismo, e traços antropométricos de populações nativo-americanas e seus descendentes miscigenados que se isolaram e se diferenciaram há 8 mil anos no interior do Nordeste.

A análise de ancestralidade local difere da ancestralidade global por focalizar em SNPs de um segmento cromossômico, por exemplo, na região vizinha a de um gene que corresponde a uma doença monogênica em estudo. Recentemente, por esse método, inferimos a ancestralidade da região cromossômica onde foram localizadas as mutações de quatro diferentes doenças de herança autossômica recessiva (Farias *et al.*, manuscrito submetido).

1.1.2. CONSANGUINIDADE

A consanguinidade diz respeito ao casamento entre indivíduos aparentados. Devido à endogamia, duas pessoas podem compartilhar alelos idênticos herdados de um ancestral comum (BITTLES; BLACK, 2010). Acredita-se que na origem dos humanos modernos, grupos de caçadores-coletores podem ter praticado um nível substancial de cruzamento interno, intragrupo e intrafamiliar devido ao número reduzido de indivíduos nessas populações de cerca, que não chegavam a 10 mil indivíduos (BITTLES; BLACK, 2010). Esses padrões reprodutivos podem ter se repetido inúmeras vezes em ciclos, resultando em altos níveis de consanguinidade entre as primeiras populações humanas (BITTLES, 2008). Essas práticas se mantiveram em algumas comunidades, entretanto, na Idade Média e por influência de algumas religiões, houve restrições a essas uniões. Hoje, as práticas de

endogamia se mantêm principalmente nas populações muçulmanas do Oriente Médio e em regiões rurais e de baixa densidade demográfica. Em comunidades cristãs, como a Igreja Católica Romana, há restrição aos casamentos com exigência de licença matrimonial (BITTLES; BLACK, 2010).

Atualmente, estima-se uma prevalência de consanguinidade no mundo da ordem de 10%, sendo mais prevalente no Oriente Médio, como mostrado na Figura 3. No Nordeste brasileiro, as frequências de casamentos consanguíneos variam de 6% até 41% (WELLER et al., 2012). Os coeficientes de endocruzamento (F) dessas populações foram estimados por meio de registros de matrimônio da Igreja Católica, sendo estimados valores de 0.00365 no Nordeste, sendo este quatro vezes maior que o F observado no Sul de 0.00081 (DA FONSECA; FREIRE-MAIA, 1970; FREIRE-MAIA, 1957, 1990). Outros estudos envolvendo comunidades nordestinas estimaram um F médio de 0.006, que é alto também em relação a outras regiões do Brasil. O município de Serrinha dos Pintos (RN), por exemplo, apresentou um dos valores mais altos do Nordeste ($F = 0.009$).

Figura 3. Variação da porcentagem de uniões consanguíneas no mundo (Bittles, 2015 - disponível em consang.net).

1.1.2.1. ÍNDICES DE FIXAÇÃO (Estatística F de Wright)

Quando subpopulações estão completamente isoladas sem eventos de migração, todos os cruzamentos devem ocorrer entre os indivíduos de cada subpopulação causando subestruturação populacional (HARTL; CLARK, 2010). O cruzamento intrapopulacional significa que os indivíduos de cada subpopulação compartilharão alguns ancestrais comuns. Os cruzamentos que acontecem ao acaso em uma subpopulação são eventos que unem indivíduos que têm ancestrais comuns (HARTL; CLARK, 2010). Esses ancestrais comuns transmitem alelos idênticos por descendência que podem se encontrar nos descendentes desse cruzamento, inclusive em situações em que os membros de cada subpopulação escolhem seus parceiros aleatoriamente.

A análise da subestruturação compreende a estimativa por três indicadores, os quais são conhecidos como índices de fixação de Wright. O primeiro é o índice de fixação devido à subdivisão da população (F_{ST}); o segundo, o índice de fixação total (F_{IT}), devido tanto a diferenças alélicas entre subpopulações, como a endogamia dentro dessas mesmas subpopulações, e o índice de fixação devido à endogamia ocorrendo dentro das subpopulações (F_{IS}) (HOLSINGER; WEIR, 2009).

A probabilidade de um alelo ser herdado por descendência é estimada pelo um índice conhecido por “identidade por descendência” ou IBD. A probabilidade de que os alelos em um indivíduo sejam idênticos por descendência (IBD) é normalmente representado pelo coeficiente de endocruzamento (F). Quanto mais endocruzada for uma população, maior será a proporção de alelos compartilhados oriundos de ancestrais comuns, ou seja, quanto maior for o F maior é o IBD (HARTL; CLARK, 2010).

As estimativas do valor de F podem ser obtidas tanto de forma individual, por meio da análise de heredograma de uma família; como também para uma população por meio do coeficiente médio de endocruzamento. No caso do coeficiente médio, pode ser obtido da ponderação de todos os F familiares, estimados a partir da análise de todos os heredogramas de uma comunidade ou população; ou por meio das frequências populacionais, observada (h) e esperada ($2pq$), de heterozigotos quando há dois alelos de frequências estimadas p e q , segundo a fórmula $F = 1 - h/(2pq)$.

Para se calcular o coeficiente médio de endocruzamento por genealogias, há a metodologia clássica que é a contagem de passos genéticos contidos em cada genealogia (BEIGUELMAN, 2000). O passo genético é a passagem de um gene de uma geração a outra, correspondendo à probabilidade de um dos dois alelos de um loco diploide ser transmitido ao indivíduo da geração seguinte (taxa de segregação com valor $\frac{1}{2}$).

Outra forma de calcular o F é por meio da frequência de casais com mesmo sobrenome. O uso de sobrenomes foi, durante o final do século XX, utilizado para estimar o coeficiente de endocruzamento pelo princípio da isonomia em populações com informação genealógica incompleta ou ausente (CROW; MANGE, 1965). Várias críticas foram levantadas em situações onde o sobrenome é polifilético, ou seja, surgem sobrenomes sem uma origem comum (ROGERS, 1991). Para minimizar os erros, foram sugeridos os usos de marcadores uniparentais para verificar se há uma regra de passagem de sobrenome concomitantemente a passagem de haplogrupo de linhagem materna ou paterna (JOBBLING, 2001).

Atualmente, o coeficiente F tem sido obtido a partir de dados de grandes regiões em homozigose (*runs of homozygosity* – ROH). Nos casos de populações com endocruzamento recente, os ROHs tendem a ser longos devido à ausência de recombinação (KIRIN et al., 2010). A distribuição dos ROHs fornece informações sobre a história dos cruzamentos ou isolamentos e consanguinidade entre os ancestrais. A homozigose de regiões cromossômicas é importante, pois está relacionada à expressão de variantes recessivas e também é um importante fator de risco para condições genéticas raras (SZPIECH et al., 2013). As populações de nativos americanos da América do Sul, por exemplo, Karitiana, apresentam os maiores ROHs (quase um terço do genoma) em relação às populações de outros continentes (KIRIN et al., 2010; PEMBERTON et al., 2012). Os ROHs para populações humanas do Nordeste brasileiro ainda não foram estudados.

1.1.3. POPULAÇÕES DO NORDESTE BRASILEIRO

Nesta seção, serão compiladas as informações arqueológicas, históricas e demográficas sobre as populações do Nordeste com enfoque para a região do alto oeste do estado do Rio Grande do Norte, que historicamente fazia parte da capitania da Paraíba, a fim de contextualizar os processos de ocupação, migração e miscigenação já conhecidos para as populações estudadas.

O atual estado do Rio Grande do Norte foi criado em 1877. Antes disso, não havia delimitação clara de limites de fronteira entre os atuais estados da Paraíba e Ceará, pois ainda havia divisão por sesmarias. Hoje o estado possui 52.811,126 km², o que corresponde 0,62% do território brasileiro. A microrregião denominada Alto Oeste Potiguar fica na divisa com os Estados do Ceará e Paraíba e ocupa quase 10% do território estadual.

1.1.3.1. OS INDÍGENAS DO TERRITÓRIO ATUAL DO INTERIOR DOS ESTADOS DA PARAÍBA E DO RIO GRANDE DO NORTE

Os dados arqueológicos indicam presença de indígenas no território do atual estado do Rio Grande do Norte (RN) e Paraíba (PB) há, pelo menos, oito mil anos. Um dos achados foi o sítio arqueológico Pedra do Alexandre no município de Carnaúba dos Dantas - RN, que fica na região Seridó e é adjacente à região do Alto Oeste Potiguar (BUENO; DIAS; STEELE, 2013; MARTIN, 1997*). Os indígenas dessa região são subdivididos em dois grandes grupos, os Tupis e os Tapuias. A caracterização dos dois grupos era dada pela língua, entre os que falavam o Tupi e os que tinham outros dialetos e eram considerados Tapuias; ou pela região por eles habitada, os Tupis no litoral e os Tapuias no sertão (PUNTONI, 1997), que talvez não sejam distintos molecularmente por subestruturação. Os grupos tapuios eram considerados duplamente “bárbaros” devido a repulsa sentida pelo colonizador português à linguística complicada e aos valores culturais. Por outro lado, os tupis foram mais facilmente catequizados pelos cristãos, tendo uma língua fácil de ser entendida (PUNTONI, 1997).

Os registros historiográficos esparsos escritos por militares portugueses e holandeses indicam que o grupo linguístico tupi era representado pelos índios potiguares, que por sua vez eram divididos em diversas tribos, tais como Paiaguá, Jundiá, Guaraíra, Igapó e os Ibirapi. Os relatórios descrevem os tapuias eram divididos em dois grandes grupos, os cariris e os tarairiús, e existem interpretações diversas sobre suas características e diferenças (SANTOS, 2009). Os tarairiús compunham várias outras tribos como os Paiaçu, Jenipapo, Canindé, Sucuru, Pega, Panati, Caratiú (DOS SANTOS JÚNIOR, 2008). Existia outro grupo dos tapuias denominados de cariri e conhecidos como bultrins, que habitavam em limites fronteiriços com a Paraíba e o Ceará, englobando tribos chamadas de Corema, Icó Pequeno, Monxoró, Caboré e Caicó.

A ausência de informações que caracterizam as diferenças os Cariris e Tarairiús é uma norma entre as narrativas dos cronistas e historiadores. Não há uma relação de diferenciação de rituais, alimentação, trocas, tarefas, usos de ferramentas; há somente uma descrição superficial de relações entre os colonizadores portugueses, clero e indígenas com uma prevalência de narrativas de fuga, conflitos, alianças e dominação. Por essas tribos serem atualmente extintas, torna-se ainda mais difícil resgatar qualquer tipo de informação baseada em memória, talvez com o advento do estudo do DNA antigo possamos recuperar informações de esqueletos do período pré-contato para separá-los populacional e geograficamente.

A discriminação espacial da territorialidade das tribos tapuias é uma das características que ainda podem ser resgatadas pela genômica de populações atuais, embora nenhum estudo tenha sido realizado na região. A partir de registros historiográficos e de cronistas, os grupos tapuios tiveram atuação na microrregião atualmente conhecida como Alto Oeste, como os Paiacu, também chamados de Pacaju ou Baiacu, que habitavam a região do vale do Assu (Rio Grande do Norte) e Jaguaribe (Ceará), estendendo-se também toda circunscrição do rio Apodi-Mossoró que atravessa Pau dos Ferros, Martins, Patu no Rio Grande do Norte e Catolé do Rocha na Paraíba, com diversos outros registros de passagem e aldeamento em municípios como Aquiraz, Messejana e Baturité no Ceará (DE MEDEIROS FILHO, 1984; DOS SANTOS JÚNIOR, 2008).

Os indígenas Panati, que habitaram a região de Pau dos Ferros, Marcelino Vieira, Encanto, São Miguel e Doutor Severiano podem ter sido endêmicos da microrregião Alto Oeste Potiguar, embora existam registros deles em 1746, na cidade de Piancó - Paraíba, no aldeamento da missão São José, que dista por volta de 170 quilômetros de Pau dos Ferros-RN. Eles são os mais prováveis ancestrais indígenas das famílias dos municípios do Alto Oeste, devido à proximidade e provável endemismo. Já os Icós Pequenos habitaram a região atual de Marcelino Vieira, Pilões e Alexandria, toda a parte do Alto Oeste e parte do Seridó, com registros deles em 1724 na aldeia do rio do Peixe, no atual território de Sousa na Paraíba. Enquanto os Monxorós viveram nas margens do rio Mossoró-Apodi e conviviam com os Icós, Panatis e Paiacus, com registros desde o início do século XVI nessa região. Os Monxorós foram aldeados no começo de 1700s na serra dos dormentes no atual território do município de Portalegre - RN, entraram em conflito com os Paiacus em 1749 e depois foram levados para a serra da Cepilhada no município atual de Campo Grande- RN (DE MEDEIROS FILHO, 1984; DOS SANTOS JÚNIOR, 2008). Sabe-se que eles se movimentaram de forma ampla entre terras tapuias e tupis. A Figura 4 representa a distribuição dos indígenas no estado do Rio Grande do Norte.

Figura 4. Distribuição das extintas tribos indígenas do Rio Grande do Norte adaptado de Medeiros Filho (1984).

1.1.3.2. AS PRIMEIRAS INVASÕES E OCUPAÇÕES DA CAPITANIA DO RIO GRANDE DO SÉCULO XVI AO INÍCIO DO SÉCULO XVIII

O início da colonização da capitania do Rio Grande foi marcado por muita resistência e conflitos com os Tupis que se sentiram ameaçados com a instalação da colônia europeia. Os territórios conquistados pelos portugueses aos poucos foram divididos em capitanias hereditárias pela corte, a do Rio Grande, por exemplo, foi doada a João de Barros e Aires da Cunha, funcionários da coroa portuguesa. As capitanias eram definidas pelas cartas de doação e foral, a de doação indicava os limites localizando o território, e o foral era a relação dos direitos e deveres do capitão-mor. A expansão europeia com a colonização causou a extinção de tribos indígenas, quer seja pela exterminação ou pela incorporação e aculturação de muitos índios e tomada de seus territórios.

Não só portugueses ocuparam o território da capitania do Rio Grande. Em 1554, foram localizados 25 franceses nas terras dos índios potiguares, que negociavam a busca do pau-brasil com os nativos. Acabaram levando cerca de três mil “quintais” de pau-brasil cortado e pronto pelos portugueses. Abrigaram-se no sul do rio Potengi. Várias denúncias foram feitas devido a assaltos e “razias” dos intrusos. O francês não tinha exigência moral para o indígena nem pretendia fundar vilas, impor costumes e obrigar disciplina. No final do

século XVI com a Paraíba conquistada, restou o Rio Grande do Norte como núcleo irradiante das incursões francesas.

A ocupação e conquista do litoral do atual estado pelos portugueses se deu após o ano de 1599 quando conseguiram construir o Forte dos Reis com o intuito de expulsar os franceses e proteger a costa contra as invasões. Durante todo século XVI, os franceses negociaram a madeira junto aos indígenas a contragosto dos portugueses, que se intitulavam como donos das terras, sendo expulsos completamente no mesmo ano de construção do Forte.

Além dos franceses, os holandeses fizeram pequenas incursões para verificar as defesas portuguesas e os pontos fracos das bordas do Nordeste. Desembarcaram em Baía da Traição, comunidade que fazia parte da capitania do Rio Grande em 1625, e ficava na divisa com a capitania da Paraíba, voltando para a Holanda levando índios, que mais tarde serviriam para fazer aliança com os nativos. Cinco anos mais tarde conseguiram invadir o que atualmente é a cidade de Recife em Pernambuco, e em 1633 conseguiram conquistar o Forte dos Reis Magos na capitania do Rio Grande, chamando-o de Castelo de Keulen. Os holandeses governaram a capitania por 22 anos, um mês e quatorze dias (CASCUDO, 1968).

O período de dominação holandesa, segundo Câmara Cascudo (1968), foi relativamente libertário, pois os flamings, como eram chamados, acolheram judeus, mouros e protestantes que foram expulsos por Isabel de Castela e Fernando de Aragão, ambos os reinos unidos à inquisição católica (CASCUDO, 1968). Os judeus sefarditas, que significa judeus da Península Ibérica, fundaram a primeira sinagoga das Américas em Recife, a Kahal Zul Israel. Logo após expulsão dos holandeses em 1654, na chamada Insurreição Pernambucana, pouco se soube sobre o destino dos sefarditas, e aqueles convertidos conhecidos como cristãos-novos. Alguns registros historiográficos e evidências arqueológicas deste período podem ser encontrados em Portugal, na Holanda, e em Recife.

O processo de colonização dos portugueses teve, pelo menos, quatro frentes (Figura 5). Uma partindo do litoral pelo rio Potengi; outra subindo o Rio Piranhas-Assu com início no território da Paraíba, passando pelo Seridó; uma terceira pelo mesmo rio com início no sertão da Paraíba, e outra do Rio Piranhas-Assu em sentido ao rio Apodi-Mossoró no sertão do Rio Grande do Norte. As quatro frentes ocorreram no final do século XVII, com um aldeamento missionário em Apodi onde habitaram mais de cem casais indígenas e líderes cristãos (MONTEIRO, 2007).

Figura 5. Quatro frentes de colonização da capitania do Rio Grande no território do atual estado do Rio Grande do Norte, adaptado de Monteiro (2012).

A criação de gado foi a principal atividade econômica do interior que ao contrário das plantações de açúcar do litoral, fez com que as frentes de conquista avançassem a passos largos. A pecuária se estendeu também próximo às bordas no litoral ao norte da atual cidade de Natal, e fez com que houvesse o início da divisão das terras da capitania entre os próprios filhos dos donatários seguindo regras dadas pela coroa portuguesa.

No final do século XVII, por volta de 150 indivíduos recolonizaram o litoral da capitania do Rio Grande. Nessa época, a capitania encontrava-se ainda abalada pela guerra com os holandeses. Antes da decisão do capitão-mor, as bordas se encontravam praticamente desocupadas por colonos e por índios aliados dos holandeses, que acabaram fugindo por medo do contra-ataque português (LOPES, 2003).

Um dos maiores eventos de conflito armado do Brasil ocorreu no Rio Grande do Norte, sabe-se que foi entre indígenas e brancos e chamava-se Guerra dos Bárbaros. A guerra durou cerca de 40 anos, final do século XVII ao início do XVIII, entre portugueses com apoio dos paulistas que forneciam suprimentos alimentícios, e indígenas da maioria das tribos tapuias que eram aldeadas nos atuais territórios da Paraíba, Rio Grande do Norte e Ceará (MONTEIRO, 2007). Há relatos que o conflito tenha existido por causa do envio de

índios para Lisboa, e também devido ao aprisionamento de dois filhos de um chefe indígena da tribo Janduí (MONTEIRO, 2007).

A política de exterminação de índios era apoiada pelos governadores-mores e oficiais devido às vontades de tê-los como escravos com relatos de fuga e rebelião por parte dos índios, que não aceitavam tal condição. Os índios mais fortes e guerreiros eram escravizados e recrutados como soldados em troca do perdão, entretanto vários deles ao invés de agirem como defensores dos colonizadores saquearam as casas dos homens brancos, e agiram no sentido de expulsá-los fornecendo armas e poderio bélico aos índios (MONTEIRO, 2007). Em contraste com outros estados do Brasil, no estado do Rio Grande do Norte não existem territórios indígenas segundo mapa da Fundação Nacional do Índio de 2016.

Três grandes eventos de miscigenação ocorreram do século XVI ao final do século XVIII. No século XVI, muitos portugueses e poucos franceses tiveram filhos com índias Tupis e mantiveram-nos como escravos. Já no início século XVII, houve uma onda de migração e miscigenação que ocorreu no litoral com sentido ao sertão devido à entrada de holandeses e suas alianças com chefes tapuios e tupis, o mesmo evento de miscigenação aconteceu com portugueses com as índias tupis. O último momento foi da metade do século XVII até o início do século XVIII, com a expulsão dos holandeses, quando os portugueses puderam fazer incursões ao interior da capitania e deixar descendentes mestiços.

1.1.3.3. OCUPAÇÃO DO SERTÃO NA SEGUNDA METADE DO SÉCULO XVIII E INÍCIO DO SÉCULO XIX

Há um consenso entre os historiadores em torno da ocupação do sertão, datando-a em meados da metade do século XVIII, quando muitos sesmeiros começaram a povoar os milhares de quilômetros quadrados doados pela corte portuguesa. A organização e avanço se deram através da pecuária, onde os vaqueiros tiveram papel fundamental na manutenção e ocupação das sesmarias de grandes fazendeiros que lhes ofereciam em troca a quarta, que era o pagamento de um bezerro dentre quatro nascidos (MONTEIRO, 2007). Os vaqueiros eram recebidos pelos moradores de pequenos povoados que serviam como abrigos passageiros, muito deles tornaram-se vilas, a exemplo de Portalegre em 1761.

As primeiras fazendas instalavam-se nas margens dos rios e lagoas, e eram autossuficientes, não necessitavam de suprimentos alimentícios dos povoados das bordas. Cultivavam alimentos como feijão, mandioca, milho, caju. Ainda mantinham carne, peixes, queijos, rapadura e aguardente que eram produzidos em pequenos engenhos, tecidos de algodão, louça, tijolos e telhas de barro, objetos feitos de couro, dentre outros. Os caminhos

de gado orientaram a criação das estradas, que se destinavam ao escoamento dos bens produzidos nas fazendas para o litoral e Portugal (MONTEIRO, 2007).

A escravidão indígena passaria a ser proibida no Brasil a partir de 1755; as aldeias mais populosas se transformaram em vilas e municípios; as terras das missões deveriam ser distribuídas entre os indígenas; e os brancos poderiam se miscigenar com os indígenas, sem chamá-los de caboclos. Entretanto, nada disso aconteceu. Por exemplo, a vila de Portalegre recebeu mais de setenta famílias indígenas vindas de uma missão jesuíta em Apodi que os encarcerava (MONTEIRO, 2007).

Em 1805, os indígenas representavam cerca de 39,3% da população absoluta de Portalegre, enquanto os mulatos e negros constituíam 34,9% e os brancos 25,7% segundo a carta do capitão-mor do Rio Grande do Norte, José Francisco de Paula Cavalcante de Albuquerque ao príncipe regente D. João. Porém, mais de setenta índios foram fuzilados vinte anos depois devido aos conflitos ocasionados pelas regras novas da corte, e não adaptação dos brancos a essas regras. Poucos escravos negros ocuparam o sertão em comparação às capitanias litorâneas da Bahia e de Pernambuco, pois a maioria deles trabalhava nos engenhos da zona da mata próximos ao litoral. Os índios cumpriam o papel do escravo no sertão, e os poucos negros que existiam se miscigenaram com os nativos (MONTEIRO, 2007).

O início do século XIX foi marcado por um aumento de três vezes mais escravos negros no litoral em lavouras de algodão. No interior, os grandes proprietários forneciam pequenos lotes de terras aos lavradores que tinham um pequeno espaço para morarem e para produzirem a alimentação necessária para suas subsistências. Uma grande seca em 1816 foi responsável por perdas significativas de lavouras e criações, mas não por um grande declínio populacional. Em 1805, havia por volta de 43,482 indivíduos livres, 5,768 escravos; em 1835, 77,691 livres, e 10,240 escravos; e em 1844 por volta de 130,919 livres, e 18,153 escravos (MONTEIRO, 2007).

1.1.3.4. INFORMAÇÕES DEMOGRÁFICAS DO ATUAL ESTADO DO RIO GRANDE DO NORTE E HISTÓRIA DOS MUNICÍPIOS DO ALTO OESTE POTIGUAR

1.1.3.4.1. Rio Grande do Norte

O atual estado do Rio Grande do Norte foi chamado de província até 1877, e nele residiam 233.979 habitantes, sendo 220.959 (94,435%) consideradas pessoas livres e 13.020 (5,565%) eram escravos (NOBRE; DE MELO, 1971). O número de pessoas que

tinham nascido na região era de 232.982, enquanto o número de estrangeiros residentes era de 997, o que representa um fluxo migratório bem baixo na região costeira no final do século XIX, que era o ponto de chegada e encontro da principal cidade da província que se chamava Natal (NOBRE; DE MELO, 1971).

1.1.3.4.2. Alto sertão potiguar

A história dos municípios da região alto oeste potiguar (Figura 6) faz-se necessária para entender os eventos de evolução e depressão demográfica. Por um lado, manteve-se o registro histórico com as características econômicas, ecológicas, eventos de epidemia, por outro, focou-se nas características demográficas, embora sem confiança nessas informações devido ao possível viés das informações fornecidas pelos cronistas ou historiadores.

Figura 6. Microrregião alto oeste potiguar em cinza e municípios que apresentam afetados da síndrome SPOAN em preto: (a) São Miguel, (b) Coronel João Pessoa, (c) Doutor Severiano, (d) Encanto, (e) Pau dos Ferros, (f) Serrinha dos Pintos, (g) Martins.

Os primeiros portugueses ao subirem 650 metros de altitude da Serra do Apodi e olhar para o vasto território do entorno pensavam que estavam diante a “porta alegre do sertão” (DE MORAIS, 1998; NOBRE; DE MELO, 1971). O nome “Portalegre” também denomina uma vila no Alentejo em Portugal, onde famílias de municípios brasileiros como Crato no Ceará, Campo-Maior no Piauí, Extremoz e Arez no Rio Grande do Norte foram originadas (CASCUDO, 1968). Os habitantes eram reconhecidos como europeus, “brancos do país” e índios. A categoria “brancos do país” era dada aos brasileiros de origem europeia com conhecimento do parentesco brasileiro. Três grupos indígenas contribuíram para

formação da população da região: os Paiacu, das margens do Apodi, os Icós, que dominavam a região do rio do Peixe e os Panati, que habitavam a serra de seu nome (CASCUDO, 1968). A vila é uma das cinco primeiras criadas no Rio Grande, a partir da ordem régia de D. José I, que mandou suprimir os aldeamentos indígenas dirigidos por missionários religiosos, transformando-os em vilas, que conseguiram autonomia jurídica e administrativa. Ela foi criada através de alvará, datado de 6 de julho de 1755, mas, sua instalação deu-se no ano de 1761. Sendo considerado o terceiro município do Estado, é também uma das células-mater da formação municipalista, dando origem a vários outros municípios potiguares como Apodi (1835), Maioridade (1841, Martins), Pau dos Ferros (1856), Francisco Dantas, Riacho da Cruz, Rodolfo Fernandes, Salamandra e Viçosa, os cinco em 1963 (CASCUDO, 1968; DE MORAIS, 1998).

A origem a povoação de Martins, antiga serra do campo grande também chamada de Serra da Conceição, data do século XVIII. Existiam aldeamentos de Tapuios e recursos naturais em abundância para prática da agricultura (ONOFRE JR., 1966; DE MORAIS, 1998). Outro nome posterior foi terra “dos Picos”. O capitão Francisco Martins Roriz, que tinha domicílio na ribeira do Jaguaribe, construiu uma fazenda para pecuária e agricultura na serra, e por isso o povoado recebeu seu nome. Em 1742, conseguiu a posse das terras situadas entre os Rios Apodi e Umari, onde assumiu a condição de comandante da Serra, atuando na organização do povoamento. Em 1770, os moradores passaram a chamar a área de serra do capitão Martins, ou serra do Martins, que se desenvolveu em torno da lagoa (NOBRE; DE MELO, 1971).

Um dos primeiros registros do tamanho populacional absoluto de Martins compreende a informação de quatro mil indivíduos, segundo Aires de Casal em sua “Corografia Brasilica” de 1817, que foi antes do primeiro censo brasileiro de 1872. No ano de assinatura da lei áurea, ainda existiam 84 escravos. Em 10 de novembro de 1841, a serra do Martins foi elevada à condição de município com o nome de Maioridade, sendo seu território desmembrado de Portalegre. Maioridade foi uma homenagem ao Imperador D. Pedro, que teve sua maioridade antecipada naquele ano de 1841 para poder assumir o trono do Império, com apenas catorze anos e sete meses de idade. Algum tempo depois, em 30 de outubro de 1847, o município de Maioridade passou a se chamar de cidade da Imperatriz, em homenagem à Imperatriz do Brasil. Posteriormente, em fevereiro de 1890, veio o nome definitivo, Serra do Martins passou a se chamar oficialmente Martins (DE MORAIS, 1998; NOBRE; DE MELO, 1971). Dela foram desmembrados os seguintes municípios Patu (1890), Alexandria (1936), Umarizal (1958), Lucrécia, Antônio Martins e Frutuoso Gomes, os três em 1963 (CASCUDO, 1968).

A povoação de Alexandria começou a formar-se quando ainda era conhecida como fazenda barriguda. Em 1759, houve o registro do primeiro povoamento da fazenda que fica numa serra que tem o formato de barriga de mulher grávida. Em 1859 passou a ser distrito do município de Martins, exatamente 100 anos depois da chegada de seu primeiro habitante (DE MORAIS, 1998; NOBRE; DE MELO, 1971). Serviu como zona pastoril e agrícola, bastante utilizada pelos sertanejos de toda aquela região, o povoado começou a crescer e a repercutir. No ano de 1913, a Câmara Municipal de Martins mudou a denominação: de Barriguda, que passou a ser chamada de Alexandria. A nova denominação foi uma homenagem a Alexandrina Barreto Ferreira Chaves, rica moradora da região. O município foi oficialmente criado em 7 de novembro de 1930. Desmembrado do município de Martins, Alexandria passou, naquela oportunidade, a se chamar João Pessoa, porém, por pouco tempo. Em 1936, para se evitar a confusão com a capital da Paraíba e, ao mesmo tempo, confirmar a merecida homenagem a sua filha ilustre, voltou a seu nome definitivo, Alexandria.

Outro município desmembrado de Martins foi Antônio Martins, que se chamava povoamento de Boa Esperança em meados do século XIX. No final desse século, houve a construção de casas, um cemitério e uma capela que terminou de ser construída em 1901 (DE MORAIS, 1998; NOBRE; DE MELO, 1971). No dia 30 de dezembro de 1943, Boa Esperança mudou de nome, passando a se chamar Demétrio Lemos, numa homenagem ao Coronel Demétrio de Rego Lemos, natural de Martins. Tornando-se município no dia 26 de março de 1963, passando oficialmente a se chamar Antônio Martins, em homenagem ao médico e deputado, natural de Martins, que teve sua atuação política reconhecida pelos muitos serviços prestados à região Oeste do Estado.

O atual município de Marcelino Vieira foi berço dos índios Panatis, que deixaram presença marcada na história com a serra de mesmo nome. O povoamento do homem branco começou a acontecer através da chegada à serra de famílias vindas dos municípios de Martins e Portalegre, e dos estados vizinhos da Paraíba e Pernambuco, que se instalaram numa localidade chamada Passagem do Freijó, em busca de oportunidades de trabalho na lavoura e na criação de gado. A consolidação do povoado ocorreu a partir da instalação da capela de Vitória, onde hoje é Marcelino Vieira, a mais antiga do município de Pau dos Ferros, que foi construída de 1861 a 1863. O nome do povoamento mudou de “Passagem de Freijós” para “Vitória”, numa alusão a vitória da região contra a cólera (DE MORAIS, 1998; NOBRE; DE MELO, 1971). Com a expansão do povoamento e a construção de novas moradias, o pequeno povoado ganhou sua primeira escola de ensino primário no ano de 1884. No ano de 1890, por ordem do vereador Pedro Velho, Vitória foi elevada à

categoria de vila. Em 1943, a vila de Vitória passou a ser distrito de Panatis, em homenagem aos índios, seus primitivos habitantes. Em 24 de novembro de 1953, transformou o distrito de Panatis no município de Marcelino Vieira, tendo seu território desmembrado dos municípios de Alexandria e Pau dos Ferros. O nome Marcelino Vieira é uma homenagem ao agricultor e criador paraibano que se tornou liderança política.

O povoamento de Serrinha dos Pintos surgiu a partir do tenente-coronel Agostinho Pinto de Queiroz que estabeleceu sua moradia nas proximidades da Serra de Martins nos últimos anos do século XVIII. Ele tornou-se o pioneiro colonizador do território (NOBRE; DE MELO, 1971). O nome foi dado numa referência comparativa entre a serra de menor porte, onde fica a comunidade, e a grande Serra do Martins situada no entorno. O termo Serrinha vincula-se aos nomes dos descendentes do patriarca Agostinho Pinto de Queiroz, daí nasceu o nome da cidade (CASCUDO, 1968; DE MORAIS, 1998; NOBRE; DE MELO, 1971). Depois de um crescimento bastante lento e pautado por uma economia baseada na agricultura e pecuária. Serrinha dos Pintos ganhou sua primeira Igreja no ano de 1942, que exibem registros de matrimônios e batismos a partir desse ano. Em 29 de outubro de 1993, o município foi desmembrado de Martins, tornando-se parte do Rio Grande do Norte.

De mesmo nome que a Viçosa de Minas Gerais, a do Rio Grande do Norte surgiu a partir de criações de fazenda de gado, existente no território situado às margens do riacho da Forquilha em 1841. Entretanto, o povoamento de Viçosa só veio realmente a acontecer no início do século XX, com as terras localizadas nas proximidades do serrote da Forquilha, sendo ocupadas por famílias de agricultores vindas de várias partes da região oeste (CASCUDO, 1968; DE MORAIS, 1998; NOBRE; DE MELO, 1971). O nome do povoado faz referência à vila Viçosa em Portugal, antiga residência dos Duques de Bragança. Segundo registros dos mais antigos, o português Miguel Caldas Caldeira de Pina Castelo Branco, ao subir a serra, ficou impressionado com a beleza do vale, exposta na magnífica paisagem a sua frente, e teria dito a frase história: “outra Viçosa” (CASCUDO, 1968; DE MORAIS, 1998; NOBRE; DE MELO, 1971). Viçosa só alcançou sua autonomia política em 18 de dezembro de 1963, quando foi desmembrada de Portalegre, tornando-se município do Rio Grande do Norte.

As populações que hoje vivem na região do Alto Oeste Potiguar são caracterizadas por um alto grau de consanguinidade e relativo isolamento por várias gerações, o que aumenta a prevalência de doenças genéticas raras devido à homozigose genômica, deriva genética e efeito de fundador. Não foram ainda realizados estudos de ancestralidade para compreensão da estrutura genética dessas populações e estimativa dos valores do f molecular.

2. OBJETIVOS

2.1. OBJETIVO GERAL

- Realizar estudo de ancestralidade local, global e estimativa de consanguinidade para famílias do interior do Nordeste que apresentam indivíduos afetados por doenças recessivas e avaliar o tamanho e quantidade de regiões em homozigose (ROH) compartilhadas entre eles.

2.2. OBJETIVOS ESPECÍFICOS

- Inferir a ancestralidade global dos indivíduos acometidos pelas doenças genéticas descritas acima;
- Estimar o tamanho e quantidade de regiões em homozigose (ROH) em indivíduos de cada uma das famílias analisadas;
- Inferir a ancestralidade local onde se localizam as mutações nos genes *KLC2*, *MED25*, *IMPA1* e *WNT7A* que causam doenças de herança autossômica recessiva com vários indivíduos afetados na região nordeste do Brasil;
- Inferir a matrilinearidade dos indivíduos afetados pela síndrome SPOAN.

3. MÉTODOS

3.1. CASUÍSTICA

A inferência de ancestralidade foi realizada a partir de 19 amostras de indivíduos afetados pelas seguintes doenças genéticas: “Síndrome Santos” (mutação em *WNT7A* – três indivíduos de Riacho de Santana – Rio Grande do Norte); deficiência intelectual (mutação em *MED25*, três indivíduos de Catolé do Rocha - Paraíba); deficiência intelectual (mutação em *IMPA1*, três indivíduos de Brejo dos Santos - Paraíba), e síndrome SPOAN, representada por 10 indivíduos pertencentes a 7 diferentes núcleos familiares procedentes de Doutor Severiano, Serrinha dos Pintos, Martins, São Miguel e Pau dos Ferros, no Rio Grande do Norte; de Santo Ângelo, no Rio Grande do Sul; de Montes Claros, em Minas Gerais; e finalmente de Mit Ghamr, na Província de Ad Daqahliyah, no Egito (Tabela 1). As genealogias estão descritas nos trabalhos de (FIGUEIREDO et al., 2015, 2016; SANTOS; BIZZO, 2005; SANTOS et al., 2008).

Tabela 1. Indivíduos afetados por doenças genéticas de herança autossômica recessiva.

Doença (Gene)	Número de Indivíduos Selecionados para genotipagem	Município de origem
Síndrome SPOAN (<i>KLC2</i>)	4 (10)	Serrinha dos Pintos / Martins (RN)
Síndrome SPOAN (<i>KLC2</i>)	1 (10)	São Miguel (RN)
Síndrome SPOAN (<i>KLC2</i>)	1 (10)	Doutor Severiano (RN)
Síndrome SPOAN (<i>KLC2</i>)	1 (10)	Pau dos Ferros (RN)
Síndrome SPOAN (<i>KLC2</i>)	1 (10)	Santo Angelo (RS)
Síndrome SPOAN (<i>KLC2</i>)	1 (10)	Montes Claros (MG)
Síndrome SPOAN (<i>KLC2</i>)	1 (10)	Mit Ghamr (Egito)
Deficiência intelectual (<i>MED25</i>)	3	Catolé do Rocha (PB)
Deficiência intelectual (<i>IMPA1</i>)	3	Brejo dos Santos (PB)
Síndrome SANTOS (<i>WNT7A</i>)	3	Riacho de Santana (RN)

3.2. ASPECTOS ÉTICOS

A realização desse estudo foi aprovada pelo Comitê de Ética da Universidade de Estadual da Paraíba com o seguinte número 51704715.4.0000.5187 (CAEE). O termo de consentimento livre e esclarecido aprovado foi assinado por todos os participantes dessa pesquisa ou por seus responsáveis.

3.3. COLETA DE SANGUE E EXTRAÇÃO DE DNA GENÔMICO

As coletas de sangue periférico dos participantes foram realizadas nas dependências dos hospitais, policlínicas ou postos de saúde dos municípios selecionados para a pesquisa. Foram colhidos 8 ml de sangue de cada participante para extração de DNA. O DNA foi extraído segundo protocolo de rotina do Laboratório de Genética Humana com fenol e clorofórmio. O DNA foi extraído utilizando a plataforma *Autopure LS* da Gentra Systems.

As amostras de DNA foram quantificadas com o uso de um espectrofotômetro (Epoch, da BioTek) e uma alíquota foi diluída em água milli-Q autoclavada com o intuito de se obter DNA com concentração de, no mínimo, 100 ng/μL.

3.4. ESTUDO I – ANCESTRALIDADE LOCAL E GLOBAL DE INDIVÍDUOS COM DOENÇAS GENÉTICAS

Amostras de diferentes núcleos familiares foram escolhidas a partir de genealogias já publicadas sendo nove indivíduos com mutação em *KLC2*, três com mutação em *IMPA1*,

outros três com mutação em *MED25* e mais três com mutação no *WNT7A*. As amostras estão armazenadas no Centro de Estudos do Genoma Humano e Células-Tronco (CEGH-Cell) e foram aliquotadas com concentração de 100ng/μl para um volume de 20μl. Iniciando-se, assim, o procedimento de envio de amostras para genotipagem que foi realizada na sede da Affymetrix em Santa Clara (Califórnia - EUA) no ano de 2016.

3.4.1. GENOTIPAGEM

Na sede da Affymetrix, a preparação e marcação do DNA das 94 amostras foram realizadas seguindo as diretrizes do kit de ensaio “Axiom™ Reagent Kit 2.0” e “Axiom Array Gene Titan kit” (Affymetrix). Na primeira etapa, o DNA genômico foi desnaturado através de solução (*Axiom 2.0 Denat Soln 10X*). Os fragmentos de DNA foram amplificados com um mix de solução amplificadora e enzima, respectivamente *Axiom 2.0 Amp Soln* e *Axiom 2.0 Amp Enzyme*. Amostras amplificadas de cada digestão foram agrupadas, limpas e fragmentadas. A fragmentação também foi feita através de uma série de reações com enzimas, buffer 10X, solução de precipitação, diluente fornecidos pelo próprio kit (*Axiom™ Reagent Kit 2.0*). O DNA fragmentado foi utilizado para hibridação a uma lâmina (*arrays*).

A hibridação de lâminas foi realizada por meio da plataforma *Gene Titan MC* (Affymetrix), essa plataforma foi projetada para receber o *Axiom™ Genome-Wide LAT Array Plate* (Affymetrix). Cada *array* permite genotipar uma grande quantidade de marcadores, 813.551 SNPs autossomos, 4259 Indels, 123 SNPs mitocondriais, 234 SNPs no cromossomo Y, totalizando 817.810 marcadores.

As amostras de DNA fragmentado e marcado foram hibridizadas a uma lâmina (*array*). A hibridação, lavagem e varredura da lâmina foi realizada seguindo as instruções do protocolo de hibridização Affymetrix. Os dados brutos (*.cel e *.arr) que correspondem respectivamente ao arquivo que contém sinais de intensidade e seus respectivos desvios-padrão, e o arquivo que contém as informações detalhadas sobre a amostra. Ambos arquivos são obtidos da plataforma GeneTitan e enviados por serviço de armazenamento pela equipe de bioinformática da Affymetrix por FTP (*file transfer protocol*) pelos servidores da Amazon.

3.4.2. FILTRAGEM

A leitura e o processamento inicial dos dados foram realizados em linguagem R com as funções do pacote *GWASTools* (Gogarten et al., 2012) utilizando-se o arquivo de texto dado pela equipe de bioinformatas da Affymetrix com os valores dos parâmetros de erro do .cel de cada amostra. Os parâmetros de exclusão das amostras compreendem os limiares

abaixo de: 0.82 de *DQC* (*Dish Quality Control*), 95% de *Quality Control call rate* para indivíduos; 95% de *call rate*, 3.6 utilizando FLD (Fisher's Linear Discriminant), 0.1 de HetSO (*Heterozygous Cluster Strength Offset*) e 0.6 de HomRO (*Homozygote Ratio Offset*) para SNPs.

As amostras que passaram nesse filtro inicial serviram como input para o *software* "Console Command" (Affymetrix) instalado em um dos computadores do Laboratório de Genética Humana sob a coordenação da Profa. Dra. Regina Célia Mingroni-Netto no Instituto de Biociências da Universidade de São Paulo. O programa tem o objetivo de transformar os arquivos .cel e .arr em arquivos que possam ser gerenciados de forma mais fácil por programas como o PLINK 1.9 (PURCELL et al., 2007) que servem para, dentre outras funções, filtrar os dados e fornecer o controle de qualidade das amostras.

A filtragem de dados envolveu um primeiro passo realizado no programa PLINK 1.9 para retirada dos dados referentes aos cromossomos sexuais e DNA mitocondrial. Com ajuda do Dr. Renan Barbosa Lemes pude aplicar um programa de meu feitio em linguagem R utilizando os dados obtidos pelo PLINK para retirada das regiões pericentroméricas e periteloméricas, que são difíceis de serem genotipadas e ricas em sequências repetitivas que podem provocar vieses em análises de ancestralidade e consanguinidade. O método de filtragem compreende exclusão de 300 marcadores mais próximos das extremidades dos braços de todos os cromossomos e dos primeiros 2Mb de cada extremidade dos braços cromossômicos.

Outras filtrações realizadas para inferência de ancestralidade com o uso do PLINK 1.9 como retirada dos SNPs tri-alelicos, dos alelos alternativos diferentes, e dos alelos distintos entre os bancos de dados, só um alelo foi utilizado em caso de pares de alelos em desequilíbrio de ligação ($r^2 > 0.5$) e foram eliminados também alelos de menor frequência (5%). Depois de todas as filtrações o número total de SNPs foi de 489.151 e o número total de 18 indivíduos (os dados de um afetado por *IMPA1* não conseguiu atingir o limiar).

3.4.3. BANCOS DE DADOS POPULACIONAIS

Dois bancos de dados populacionais mundiais foram utilizados, o 1000 genomas (1000 GENOMES PROJECT CONSORTIUM et al., 2010) e projeto de diversidade genômica Simons - SGDP (MALLICK et al., 2016) como referências de populações europeias, africanas e nativo-americanas (incluindo miscigenadas e asiáticas) para inferir a ancestralidade dos indivíduos nordestinos escolhidos para o estudo da datação e inferência dos haplótipos que contêm mutações.

Informações mais detalhadas sobre as populações amostradas foram descritas nas versões atualizadas dos manuscritos originais para ambos os painéis de diversidade humanas (1000 GENOMES PROJECT CONSORTIUM et al., 2010; MALLICK et al., 2016). No total foram utilizadas cinco populações do banco de dados 1000 genomas ($n = 425$), 85 indivíduos da Península Ibérica - IBS, 85 indivíduos africanos Yoruba - YRI e 85 peruanos de Lima - PEL, que apresentam grande contribuição nativo-americana, e asiáticas, 85 japoneses (JPT) e 85 chineses Han (CHB).

O banco de dados SGDP forneceu 22 indivíduos africanos de diferentes populações: Bantu Herero ($n = 2$), Bantus do Quênia ($n = 2$), Biaka ($n = 2$), Esan ($n = 2$), Gambianos ($n = 2$), Mandenka ($n = 3$), Mbuti ($n = 4$), Mende ($n = 2$), Yoruba ($n = 3$). O mesmo número de europeus representados por Adygei ($n = 1$), albaniano ($n = 1$), bascos ($n = 2$), bergamascos ($n = 2$), búlgaros ($n = 2$), ingleses ($n = 2$), franceses ($n = 3$), húngaros ($n = 2$), polaco ($n = 1$), sardenhos ($n = 3$), espanhóis ($n = 2$) e toscanos ($n = 1$). Mais 22 indivíduos representando nativo-americanos Chane ($n = 1$), Karitiana ($n = 3$), maias ($n = 2$), Mixe ($n = 3$), Mixtec ($n = 2$), Piapoco ($n = 2$), Pima ($n = 2$), Quechua ($n = 3$), Surui ($n = 2$), Zapotec ($n = 2$). Ao todo foram incluídos 66 indivíduos desse banco de dados.

Por não existir uma população referência nativo-americana no 1KG, quatro diferentes análises foram realizadas com três diferentes populações deste banco de dados, e uma com diferentes populações que podem ser consideradas nativo-americanas do SGDP. As três análises com o 1KG compreendem (1) africanos, europeus e peruanos; (2) africanos, europeus e chineses; (3) africanos, europeus e japoneses; e utilizando o SGDP, (4) africanos, europeus e nativo-americanos.

3.4.4. PCA E ADMIXTURE

A inferência de ancestralidade foi feita a partir do algoritmo do programa ADMIXTURE (ALEXANDER; NOVEMBRE; LANGE, 2009), que estima a máxima verossimilhança de ancestralidades individuais de dados de genótipos em SNPs sem precisar de classificações a priori dos indivíduos nas populações. Essa abordagem identifica os grupos de indivíduos que apresentam frequências alélicas similares e os agrupam, os resultados obtidos permitiram verificar a contribuição de cada população ancestral chamada de K que é representada por uma cor distinta para as populações estudadas.

Primeiro, foi realizada uma caracterização dos indivíduos do estudo para suas estruturas genômicas utilizando o programa ADMIXTURE. A caracterização foi necessária para definir as populações parentais, identificar os componentes de miscigenação e definir

os parâmetros usados na análises para estimar a ancestralidade na região da mutação. Para caracterização foram utilizados os SNPs das populações ibéricas (IBS), africanas (YRI) e latino-americanas miscigenadas (PEL), asiáticas (JPT e CHB) e nativo-americanas do SGDP. Estas diferentes populações obedecem a parcela nativo-americana no modelo tri-híbrido de formação das populações brasileiras (europeu, africano e nativo-americano). Os resultados da análise foram dimensionados em um gráfico de barras verticais com três cores, uma para cada população ancestral, e um mosaico com as três cores para as populações miscigenadas.

Para as quatro análises foi utilizada também uma análise de componentes principais executada a partir do *script* do programa FastPCA (ABRAHAM; INOUE, 2014). A análise de componentes principais é um método estatístico descritivo para reduzir a dimensionalidade dos dados. Em genética, sumariza as relações dos genótipos entre os indivíduos em Componentes Principais não relacionados (CPs, ou autovetores).

Segundo, foram realizadas análises para definir a idade e a origem das mutações presentes em homozigose nos indivíduos afetados. A idade da mutação foi estimada por um método escolhido pela simplicidade de basear-se no compartilhamento do tamanho dos haplótipos ancestrais que contém a mutação entre os indivíduos afetados, que geralmente formam pequenas amostras. O indivíduo que carrega o haplótipo ancestral que contém a mutação é chamado de ancestral comum mais recente (ACMR), e o ACMR é aproximadamente igual ao valor da idade da mutação estimada em uma população.

Terceiro, os genótipos dos 18 indivíduos afetados por doenças genéticas de herança autossômico recessiva foram faseados com o programa SHAPEIT (DELANEAU; MARCHINI; ZAGURY, 2011), a referência utilizada para estimar os haplótipos foi de 1092 indivíduos da fase 3 do banco de dados 1KG. Após a obtenção dos haplótipos foi realizada a inferência da ancestralidade local utilizando o algoritmo do programa RFMix (MAPLES et al., 2013) que mostra a origem putativa do haplótipo compartilhado quando associado com uma idade de mutação estimada. Esses haplótipos compartilhados contém as mutações em homozigose em dado *locus* genômico conhecido. O RFMix é um método discriminativo que usa campos aleatórios condicionais baseados nos dados de referência dos dados do painel, 1KG no nosso caso, para cada segmento cromossômico representado por janelas de SNPs, adicionalmente o programa incorpora um algoritmo de maximização de expectativa para refinar a inferência de ancestralidade local. Os parâmetros escolhidos para rodar a análise compreendem a opção do faseamento populacional (PopPhased) e parametrização padrão.

Quarto, a proporção de endocruzamento pela estimativa de grandes regiões em homozigose por origem comum (ROH) fornece pistas sobre quão recentes são os

endocruzamento pela quantidade e tamanho das ROHs e quão endocruzadas são as populações pela soma do tamanho das ROHs obtidas. Com esse fim executamos o programa PLINK (PURCELL et al., 2007), aplicando-se o método descrito por Pemberton et al. (2012) com janelas com 100 SNPs e 1MB de tamanho. As análises de ROH, além de identificar a proporção de homozigose nos indivíduos ajudou a definir os parâmetros usados nas análises da idade da mutação.

Por último, análises estatísticas sumárias foram utilizadas para caracterizar os dados incluindo-se descrição das porcentagens de miscigenação para cada indivíduo amostrado, também foram realizadas análises estatísticas analíticas para estimar a idade da mutação pelo tamanho dos haplótipos e o uso do coeficiente de correlação para ancestralidade global calculada pelo programa ADMIXTURE correlacionando-a com a soma das ancestralidades locais pelo RFMix.

3.5. ESTUDO 2 – MATRILINHAGEM DE INDIVÍDUOS COM MUTAÇÃO EM *KLC2*

3.5.1. SEQUENCIAMENTO DA REGIÃO HVS-1 e HVS-2

O sequenciamento do DNA mitocondrial das regiões HVSI e HVSII permitiram caracterizar os haplogrupos presentes na amostra selecionada. A amplificação da região de controle do DNAmt foi realizada conforme protocolo descrito por Alves-Silva et al. (2000), com uso dos seguintes *primers*: F5'-TCAAAGCTTACACCAGTCTTGTAAC-3'; e R5'-CCTGAAGTAGGAACCAGATG-3' (HVSI); e para a região HVSII: F5'-GGTCTATCACCTATTAACCAC-3'; R5'-TTGAGGAGGTAAGCTACATA-3'.

As reações de amplificação foram realizadas em um volume final de 25 µL utilizando de 15-40 ng do DNA genômico, MgCl₂ 1,5 mM, 200 µM de cada dNTP, 0,16 µM de cada *primer* e 0,2 U de Taq DNA polimerase (Phoneutria). As amplificações foram realizadas em um termociclador gradiente Eppendorf[®]. Os produtos foram purificados por meio do sistema de enzimas hidrolíticas Exonuclease I e SAP (*Shrimp Alkaline Phosphatase*) (GE Healthcare) e sequenciados pelo método de Sanger a partir dos produtos de PCR, das fitas direta e reversa de cada amostra. As reações de sequenciamento foram realizadas através do kit de sequenciamento Big Dye Terminator Cycle Sequencing com protocolo sugerido pelo fabricante (Applied Biosystems). A análise dos fragmentos foi realizada no analisador genético ABI PRISM 310 e a análise de resultados foi feita por meio do software Sequencher 5.0.

5. CONSIDERAÇÕES FINAIS

Ao longo dos últimos dez anos, desde a descrição da síndrome Spooan em 2005, foi realizada busca ativa por mais indivíduos com mutação em *KLC2* no interior do Rio Grande do Norte e Paraíba, e foram divulgados, em congressos nacionais e internacionais, os achados dessa pesquisa. Com isso, foi possível identificar dois irmãos com a síndrome no Egito e outros dois no Rio Grande do Sul. A partir de 2014, foi iniciada a pesquisa para estimar a idade da mutação a partir de microssatélites de 70 indivíduos com síndrome SPOAN, e de inferir a ancestralidade local e global a partir do sequenciamento do genoma completo de um indivíduo com a mesma síndrome. A ideia era definir com mais detalhes quando se originou a mutação e onde isto teria acontecido para conhecermos as populações que poderiam ter mais afetados por essa síndrome.

Outra importante questão presente inicialmente no projeto de doutorado aprovado em 2014 dizia respeito à questão da consanguinidade no Nordeste. Como explicar a manutenção dessas práticas nas cidades do sertão nordestino, mas não em centenas de outras cidades rurais brasileiras? A nossa hipótese de trabalho partia da suposição baseada no registro histórico de que, talvez, a valorização dos casamentos consanguíneos tivesse relação com o processo de colonização dessa região por cristãos-novos ou judeus sefarditas convertidos ao cristianismo durante a Inquisição. Dadas as limitações técnicas pela falta de definição e separação de populações sefarditas e não-sefarditas, e pela falta de bancos de dados populacionais nativo-americanos, o projeto sofreu uma modificação definitiva no final de 2016, como relatado na apresentação da tese.

Realizamos uma análise pormenorizada da ancestralidade global e local, e estimativa das grandes regiões em homozigose de quatro doenças genéticas de herança autossômica recessiva. Os resultados dos estudos de ancestralidade global apontaram uma maior contribuição europeia para os indivíduos com doenças genéticas de herança autossômica recessiva do interior do Nordeste do Brasil. A maior contribuição europeia confirmou os registros históricos da entrada dos europeus na região litorânea e nos sertões do Nordeste do Brasil. De fato, as mutações em homozigose dos genes *MED25* e *IMPA1* apresentaram uma origem recente e seus respectivos haplótipos exibiram ancestralidade europeia. A mutação em *WNT7A* tem uma origem aproximada de 100 anos e foi o único com haplótipo de ancestralidade nativo-americana.

A mutação no gene *KLC2* causativa da síndrome Spooan revelou uma história de mais de 350 anos, remontando à época da entrada dos holandeses na região Nordeste do Brasil. O haplótipo dos indivíduos brasileiros e egípcios apresentou ancestralidade europeia,

sugerindo um ancestral comum na península ibérica. A dispersão dos alelos mutados em *KLC2* pelo interior do Nordeste pode ter sido impulsionada pela explosão demográfica que ocorreu entre os séculos XIX e XX associada à prática de casamentos consanguíneos. O tamanho e a quantidade de regiões em homozigose (ROH) mostrou que, conforme esperado para indivíduos com mutações recessivas em homozigose, os valores de Froh se aproximaram ao limiar de populações asiáticas e nativo-americanas. Não obstante, os valores de tamanho e quantidade cumulativos chegaram próximos de valores de populações do Oriente Médio.

No caso da síndrome Spoan, foram identificados mais de 70 indivíduos espalhados por vários municípios do sertão potiguar e da Paraíba. Da análise genealógica, foram identificadas menos de dez diferentes mulheres ancestrais dos afetados. Qual seria a ancestralidade dessas mulheres? Do registro histórico, a suposição inicial seria que elas fossem ameríndias. Entretanto, sabemos que houve também a entrada de judeus sefarditas no Nordeste do Brasil no período da colonização holandesa. De forma surpreendente, encontramos um haplogrupo ancestral entre os afetados que está em maior frequência nessas comunidades tradicionais de judeus sefarditas no *cluster* do nordeste brasileiro.

As outras três doenças como recentes, e sem presença de afetados fora do Brasil, não tiveram indivíduos estudados por esse marcador molecular porque foi iniciada uma pesquisa mais abrangente que pode dar boas pistas sobre a origem e ancestralidade das populações do sertão. No momento, estamos investigando a ancestralidade de todos os idosos com 80 anos ou mais fundadores de uma cidade do interior da Paraíba, Brejo dos Santos, mesma cidade onde foram encontrados os indivíduos com mutação em *MED25*. O estudo conseguiu recrutar todos os 179 idosos da cidade e 70 deles foram genotipados com uso de *SNP-array* com DNA. Os resultados estão sendo analisados para futura publicação, da qual sou colaborador.

Neste estudo, ainda foram cadastrados todos os idosos totalizando uma população de 731 indivíduos com 60 anos ou mais. Desse grupo, 532 foram entrevistados para fazer uma descrição de dados sociodemográficos e para investigar as suas concepções, atitudes e comportamentos em relação à consanguinidade. Esse trabalho revelou que há uma atitude muito positiva na comunidade em relação aos casamentos consanguíneos, e que a principal razão para casar-se com parente é a crença de que os casamentos são mais duradouros. Desse trabalho derivou uma publicação da qual sou coautor submetida ao *Journal of Biosocial Science*. Além disso, foram colhidos os nomes e sobrenomes de todos esses idosos e um estudo sobre o coeficiente de endocruzamento baseado em sobrenomes está sendo realizado pelo grupo do Prof. Dr. Paulo Alberto Otto.

Muitas perguntas presentes no projeto inicial de doutorado não foram respondidas e ainda permanecem em aberto para futuros estudos: existiria relação entre a ancestralidade dos indivíduos e a prática de casamentos consanguíneos? Qual a contribuição da ancestralidade paterna para as famílias desses indivíduos com mutações em homozigose? A consanguinidade manteria os mesmos limiares na população de fundadores de uma pequena cidade por meio do f molecular? As respostas para essas questões permitirão avançar o conhecimento sobre famílias e populações de uma região tão pobre em pesquisas genéticas e tão rica em problemas derivados de práticas de casamentos consanguíneos.

6. REFERÊNCIAS

1000 GENOMES PROJECT CONSORTIUM et al. A map of human genome variation from population-scale sequencing. **Nature**, v. 467, n. 7319, p. 1061–1073, 28 out. 2010.

ABRAHAM, G.; INOUE, M. Fast principal component analysis of large-scale genome-wide data. **PloS one**, v. 9, n. 4, p. e93766, 9 abr. 2014.

ADAMS, S. M. et al. The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. **American journal of human genetics**, v. 83, n. 6, p. 725–736, dez. 2008.

ALEXANDER, D. H.; NOVEMBRE, J.; LANGE, K. Fast model-based estimation of ancestry in unrelated individuals. **Genome research**, v. 19, n. 9, p. 1655–1664, set. 2009.

ALVES, L. U. et al. Santos syndrome is caused by mutation in the WNT7A gene. **Journal of human genetics**, v. 62, n. 12, p. 1073–1078, dez. 2017.

ALVES-SILVA, J. et al. The ancestry of Brazilian mtDNA lineages. **American journal of human genetics**, v. 67, n. 2, p. 444–461, ago. 2000.

BARBOSA, F. B. et al. Ancestry Informative Marker Panel to Estimate Population Stratification Using Genome-wide Human Array. **Annals of human genetics**, v. 81, n. 6, p. 225–233, 2017.

BEDFORD, F. L. Sephardic signature in haplogroup T mitochondrial DNA. **European journal of human genetics: EJHG**, v. 20, n. 4, p. 441–448, abr. 2012.

BEDFORD, F. L.; YACOBI, D. On two Jewish clades in mitochondrial DNA. **European journal of human genetics: EJHG**, v. 23, n. 8, p. 993–994, ago. 2015.

BEHAR, D. M. et al. Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. **PloS one**, v. 3, n. 4, p. e2062, 30 abr. 2008.

BEIGUELMAN, B. Human and medical genetics in Brazil. **Genetics and molecular biology**, v. 23, n. 2, p. 277–281, 2000.

BITTLES, A. Consanguinity and its relevance to clinical genetics. **Clinical genetics**, v. 60, n. 2, p. 89–98, ago. 2001.

BITTLES, A. H. The Role and Significance of Consanguinity as a Demographic Variable. **Population and development review**, v. 20, n. 3, p. 561–584, 1994.

BITTLES, A. H. Endogamy, consanguinity and community genetics. **Journal of genetics**, v. 81, n. 3, p. 91–98, dez. 2002.

BITTLES, A. H. A community genetics perspective on consanguineous marriage. **Community genetics**, v. 11, n. 6, p. 324–330, 5 ago. 2008.

BITTLES, A. H. Consanguineous marriages and congenital anomalies. **The Lancet**, v. 382, n. 9901, p. 1316–1317, 19 out. 2013.

BITTLES, A. H. The prevalence and outcomes of consanguineous marriage in contemporary

societies. **Cousin Marriages: Between Tradition, Genetic Risk and Cultural Change**, v. 28, p. 33, 2015.

BITTLES, A. H.; BLACK, M. L. Consanguinity, human evolution, and complex diseases. **of the National Academy of Sciences**, 2010.

BREHM, A. et al. Mitochondrial portraits of the Madeira and Açores archipelagos witness different genetic pools of its settlers. **Human genetics**, v. 114, n. 1, p. 77–86, dez. 2003.

BUENO, L.; DIAS, A. S.; STEELE, J. The Late Pleistocene/Early Holocene archaeological record in Brazil: A geo-referenced database. **Quaternary international: the journal of the International Union for Quaternary Research**, v. 301, p. 74–93, 8 jul. 2013.

CAMPBELL, C. L. et al. North African Jewish and non-Jewish populations form distinctive, orthogonal clusters. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 34, p. 13865–13870, 21 ago. 2012.

CARVALHO-SILVA, D. R. et al. The phylogeography of Brazilian Y-chromosome lineages. **American journal of human genetics**, v. 68, n. 1, p. 281–286, jan. 2001.

CASCUDO, L. DA C. Nomes da Terra: história, geografia e toponímia do Rio Grande do Norte. **Natal: Fundação José Augusto**, 1968.

CROW, J. F.; MANGE, A. P. Measurement of inbreeding from the frequency of marriages between persons of the same surname. **Eugenics quarterly**, v. 12, n. 4, p. 199–203, dez. 1965.

DA FONSECA, L. G.; FREIRE-MAIA, N. Further data on inbreeding levels in Brazilian populations. **Social biology**, v. 17, n. 4, p. 324–328, dez. 1970.

DE ASSIS, A. A. F. Criptojudaísmo no feminino. Uma análise da Resistência judaica na Bahia Quinhentista a partir das fontes da I Visitação do Santo Ofício ao Brasil. **Revista**, n. 9, p. 1–10, 2010.

DE ASSIS, A. A. F. Israel no Trópico? Mulheres criptojudias e identidades religiosas no Brasil colonial. **Cadernos de Língua e Literatura Hebraica**, 2012.

DE ASSIS, A. A. F. Um Israel possível na Bahia colonial: sobre mulheres e resistência judaica em tempos de perseguição. **Arquivo Maaravi: Revista Digital de Estudos Judaicos da UFMG**, v. 7, n. 12, p. 2–19, 30 mar. 2013.

DE AZEVEDO, D. A. et al. Analysis of Y chromosome SNPs in Alagoas, Northeastern Brazil. **Forensic Science International: Genetics Supplement Series**, v. 2, n. 1, p. 421–422, 1 dez. 2009.

DELANEAU, O.; MARCHINI, J.; ZAGURY, J.-F. A linear complexity phasing method for thousands of genomes. **Nature methods**, v. 9, n. 2, p. 179–181, 4 dez. 2011.

DE MEDEIROS FILHO, O. **índios do Açu e Seridó**. [s.l.] Senado Federal, Centro Grafico, 1984.

DE MORAIS, M. C. C. **Terras potiguares**. [s.l.] Dinâmica Editora, 1998.

DOS SANTOS JÚNIOR, V. **Os Índios Tapuias do Rio Grande do Norte**. [s.l.] Valdeci dos Santos Júnior, 2008.

FALBEL, N. **Judeus no Brasil: estudos e notas**. [s.l.] Editora Humanitas, 2008.

FARIAS, A. A. et al. **A origem dos humanos modernos**. Em: Neves, W e Murrieta, R. Assim caminhou a humanidade, ed. 1, Ed. Palas Athena, 2015.

FARIAS, A. A. et al. Ocupações e reocupações por grupos pretéritos no litoral cearense: os vestígios arqueológicos das dunas de Trairí, Ceará, Brasil. **Clio (série arqueológica)**, v. 31, n. 1, p. 53, 2016.

FEITLER, B. Four chapters in the history of crypto-Judaism in Brazil: the case of the northeastern New Christians (17th–21st centuries). **Jewish History**, v. 25, n. 2, p. 207–227, 1 maio 2011.

FIGUEIREDO, T. et al. Homozygous missense mutation in MED25 segregates with syndromic intellectual disability in a large consanguineous family. **Journal of medical genetics**, v. 52, n. 2, p. 123–127, fev. 2015.

FIGUEIREDO, T. et al. A homozygous loss-of-function mutation in inositol monophosphatase 1 (IMPA1) causes severe intellectual disability. **Molecular psychiatry**, v. 21, n. 8, p. 1125–1129, ago. 2016.

FREIRE-MAIA, N. Inbreeding in Brazil. **American journal of human genetics**, v. 9, n. 4, p. 284–298, dez. 1957.

FREIRE-MAIA, N. Genetic effects in Brazilian populations due to consanguineous marriages. **American journal of medical genetics**, v. 35, n. 1, p. 115–117, jan. 1990.

GONÇALVES, V. F. et al. Identification of Polynesian mtDNA haplogroups in remains of Botocudo Amerindians from Brazil. **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 16, p. 6465–6469, 16 abr. 2013.

GONZÁLEZ, A. M. et al. Mitochondrial DNA affinities at the Atlantic fringe of Europe. **American journal of physical anthropology**, v. 120, n. 4, p. 391–404, abr. 2003.

HARTL, D. L.; CLARK, A. G. **Princípios de Genética de Populações - 4.ed.** [s.l.] Artmed Editora, 2010.

HENN, B. M. et al. Fine-scale population structure and the era of next-generation sequencing. **Human molecular genetics**, v. 19, n. R2, p. R221–6, 15 out. 2010.

HOLSINGER, K. E.; WEIR, B. S. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). **Nature reviews. Genetics**, v. 10, n. 9, p. 639–650, set. 2009.

JOBLING, M. et al. **Human Evolutionary Genetics, Second Edition**. [s.l.] Garland Science, 2013.

JOBLING, M. A. In the name of the father: surnames and genetics. **Trends in genetics: TIG**, v. 17, n. 6, p. 353–357, 2001.

JOBLING, M. A.; TYLER-SMITH, C. The human Y chromosome: an evolutionary marker comes of age. **Nature reviews. Genetics**, v. 4, n. 8, p. 598–612, ago. 2003.

JOBLING, M. A.; TYLER-SMITH, C. Human Y-chromosome variation in the genome-sequencing era. **Nature reviews. Genetics**, v. 18, n. 8, p. 485–497, ago. 2017.

KIRIN, M. et al. Genomic runs of homozygosity record population history and consanguinity. **PloS one**, v. 5, n. 11, p. e13996, 15 nov. 2010.

KIVISILD, T. Maternal ancestry and population history from whole mitochondrial genomes. **Investigative genetics**, v. 6, p. 3, 10 mar. 2015.

KLITZ, W. et al. Genetic differentiation of Jewish populations. **Tissue antigens**, v. 76, n. 6, p. 442–458, dez. 2010.

LOHMUELLER, K. E. et al. Proportionally more deleterious genetic variation in European than in African populations. **Nature**, v. 451, n. 7181, p. 994–997, 21 fev. 2008.

LOPES, F. M. **Índios, colonos e missionários na colonização da capitania do Rio Grande do Norte**. [s.l.] Fundação Guimarães Duque, 2003. v. 1379

MACEDO-SOUZA, L. I. et al. Spastic paraplegia, optic atrophy, and neuropathy is linked to chromosome 11q13. **Annals of neurology**, v. 57, n. 5, p. 730–737, maio 2005.

MACEDO-SOUZA, L. I. et al. Spastic paraplegia, optic atrophy, and neuropathy: new observations, locus refinement, and exclusion of candidate genes. **Annals of human genetics**, v. 73, n. 3, p. 382–387, 2009.

MALLICK, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. **Nature**, v. 538, n. 7624, p. 201–206, 13 out. 2016.

MAPLES, B. K. et al. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. **American journal of human genetics**, v. 93, n. 2, p. 278–288, 8 ago. 2013.

MARCUS, A. W.; EBEL, E. R.; FRIEDMAN, D. A. Commentary: Portuguese crypto-Jews: the genetic heritage of a complex history. **Frontiers in genetics**, v. 6, p. 261, 7 ago. 2015.

MARQUES, S. L. et al. Portuguese mitochondrial DNA genetic diversity-An update and a phylogenetic revision. **Forensic science international. Genetics**, v. 15, p. 27–32, mar. 2015.

MARTIN, G. **Pré-história do Nordeste do Brasil**. [s.l.] Editora Universitária UFPE, 1997*.

MELO, U. S. et al. Overexpression of KLC2 due to a homozygous deletion in the non-coding region causes SPOAN syndrome. **Human molecular genetics**, v. 24, n. 24, p. 6877–6885, 15 dez. 2015.

MIELE, N. Velhos "cristãos-novos" no sertão paraibano. 2008.

MONTEIRO, D. M. Introdução à história do Rio Grande do Norte. 2007.

MOURA, R. R. DE et al. Meta-analysis of Brazilian genetic admixture and comparison with

other Latin America countries. **American journal of human biology: the official journal of the Human Biology Council**, v. 27, n. 5, p. 674–680, set. 2015.

MÜTZENBERG, D. C. S. et al. FUNDAMENTOS DA DIAGÊNSE ÓSSEA E SUAS FORMAS DE AVALIAÇÃO USANDO AS TÉCNICAS ESPECTROSCÓPICAS DE FITR-AT E DRX. **Clio**, v. 30, n. 2, p. 154, 2015.

MYCHALECKYJ, J. C. et al. Genome-Wide Analysis in Brazilians Reveals Highly Differentiated Native American Genome Regions. **Molecular biology and evolution**, v. 34, n. 3, p. 559–574, 1 mar. 2017.

NOBRE, M. F.; DE MELO, M. R. **Breve notícia sobre a província do Rio Grande do Norte: baseada nas leis, informações e fatos consi gnados na história antiga e moderna**. [s.l.] Pongetti, 1971.

NOGUEIRO, I. et al. Phylogeographic analysis of paternal lineages in NE Portuguese Jewish communities. **American journal of physical anthropology**, v. 141, n. 3, p. 373–381, mar. 2010.

NOGUEIRO, I. et al. Portuguese crypto-Jews: the genetic heritage of a complex history. **Frontiers in genetics**, v. 6, p. 12, 2 fev. 2015a.

NOGUEIRO, I. et al. Echoes from Sepharad: signatures on the maternal gene pool of crypto-Jewish descendants. **European journal of human genetics: EJHG**, v. 23, n. 5, p. 693–699, maio 2015b.

NOVINSKY, A. Cristãos novos na Bahia. 1972.

NOVINSKY, A. W. et al. **Os judeus que construíram o Brasil: fontes inéditas para uma nova visão da historia brasileira**. [s.l.] Planeta, 2015.

OLIVEIRA, A. M. et al. Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs. **Forensic science international. Genetics**, v. 13, p. 264–268, nov. 2014.

ONOFRE JR., M. **Martins - Sua Terra, Sua Gente**. Ensaio. Natal: Departamento Estadual de Imprensa, 1966.

PALHA, T. et al. Fourteen short tandem repeat loci Y chromosome haplotypes: Genetic analysis in populations from northern Brazil. **Forensic science international. Genetics**, v. 6, n. 3, p. 413–418, maio 2012.

PEMBERTON, T. J. et al. Genomic patterns of homozygosity in worldwide human populations. **American journal of human genetics**, v. 91, n. 2, p. 275–292, 10 ago. 2012.

PEREIRA, L.; PRATA, M. J.; AMORIM, A. Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. **Annals of human genetics**, v. 64, n. Pt 6, p. 491–506, nov. 2000.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v. 155, n. 2, p. 945–959, jun. 2000.

PUNTONI, P. Tupi ou não tupi. **Uma contribuição ao estudo da Etnohistória dos povos**

- indígenas no Brasil colônia. In: Ethnos, Revista Brasileira de Etnohistória. Universidade Federal de Pernambuco--Núcleo de Estudos Indigenistas: Ano**, v. 2, p. 05–19, 1997.
- PURCELL, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. **American journal of human genetics**, v. 81, n. 3, p. 559–575, set. 2007.
- REGUEIRO, M. et al. From Arabia to Iberia: A Y chromosome perspective. **Gene**, v. 564, n. 2, p. 141–152, 15 jun. 2015.
- RESQUE, R. et al. Male Lineages in Brazil: Intercontinental Admixture and Stratification of the European Background. **PloS one**, v. 11, n. 4, p. e0152573, 5 abr. 2016.
- ROGERS, A. R. Doubts about isonymy. **Human biology**, v. 63, n. 5, p. 663–668, out. 1991.
- ROITMAN, J. V.; OTHERS. **Us and Them: inter-cultural trade and the Sephardim, 1595-1640**. [s.l.] Department of History, Faculty of Humanities, Leiden University, 2009.
- SALZANO, F. M.; SANS, M. Interethnic admixture and the evolution of Latin American populations. **Genetics and molecular biology**, v. 37, n. 1 Suppl, p. 151–170, mar. 2014.
- SANTOS, C. et al. Genetic structure and origin of peopling in the Azores islands (Portugal): the view from mtDNA. **Annals of human genetics**, v. 67, n. Pt 5, p. 433–456, set. 2003.
- SANTOS, E. T. et al. Brasil e a Idiosincrasia da Miscigenação. 2015.
- SANTOS, J. S. Cariri e Tarairiú?: culturas tapuais nos sertões da Paraíba. 2009.
- SANTOS, S. et al. Inbreeding levels in Northeast Brazil: Strategies for the prospecting of new genetic disorders. **Genetics and molecular biology**, v. 33, n. 2, p. 220–223, abr. 2010.
- SANTOS, S.; BIZZO, N. From “new genetics” to everyday knowledge: Ideas about how genetic diseases are transmitted in two large Brazilian families. **Science education**, v. 89, n. 4, p. 564–576, jul. 2005.
- SANTOS, S. C. et al. A previously undescribed syndrome combining fibular agenesis/hypoplasia, oligodactylous clubfeet, anonychia/ungual hypoplasia, and other defects. **American journal of medical genetics. Part A**, v. 146A, n. 24, p. 3126–3131, 15 dez. 2008.
- SCHAAN, A. P. et al. mtDNA structure: the women who formed the Brazilian Northeast. **BMC evolutionary biology**, v. 17, n. 1, p. 185, 9 ago. 2017.
- SCHRAIBER, J. G.; AKEY, J. M. Methods and models for unravelling human evolutionary history. **Nature reviews. Genetics**, v. 16, n. 12, p. 727–740, dez. 2015.
- SULLASI, HSL et al. e. A Note on Diagenetic Parameters for Bone Remains from Pedra do Alexandre Paleoamerican Site Without Sample Destruction. **Fundamentos**, vol. XIX, p.. 74-85, 2017.
- SZPIECH, Z. A. et al. Long runs of homozygosity are enriched for deleterious variation. **American journal of human genetics**, v. 93, n. 1, p. 90–102, 11 jul. 2013.

TANG, H. et al. Reconstructing genetic ancestry blocks in admixed individuals. **American journal of human genetics**, v. 79, n. 1, p. 1–12, jul. 2006.

VELEZ, C. et al. The impact of Converso Jews on the genomes of modern Latin Americans. **Human genetics**, v. 131, n. 2, p. 251–263, fev. 2012.

WALLACE, D. C. Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. **Annual review of biochemistry**, v. 76, p. 781–821, 2007.

WELLER, M. et al. Consanguineous unions and the burden of disability: A population-based study in communities of Northeastern Brazil. **American journal of human biology: the official journal of the Human Biology Council**, v. 24, n. 6, p. 835–840, 2012.

WOLFF, E.; WOLFF, F. **A odisséia dos Judeus de Recife**. [s.l.] Centro de Estudos Judaicos, 1979.

YANG, N. N. et al. Contrasting patterns of nuclear and mtDNA diversity in Native American populations. **Annals of human genetics**, v. 74, n. 6, p. 525–538, nov. 2010.