

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

LARISSA LUZ GOMES

APLICAÇÃO DE REDES NEURAS ARTIFICIAIS NA
CLASSIFICAÇÃO DE PADRÕES FILOGEOGRÁFICOS COM
BASE NA VARIABILIDADE GENÔMICA DO DNA
MITOCONDRIAL

DM 35/2007

UFPA / CT / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2007

UNIVERSIDADE FEDERAL DO PARÁ
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

LARISSA LUZ GOMES

APLICAÇÃO DE REDES NEURAS ARTIFICIAIS NA
CLASSIFICAÇÃO DE PADRÕES FILOGEOGRÁFICOS COM
BASE NA VARIABILIDADE GENÔMICA DO DNA
MITOCONDRIAL

Dissertação submetida à Banca
Examinadora do Programa de
Pós-Graduação em Engenharia
Elétrica da UFPA para a obtenção
do Grau de Mestre em
Engenharia Elétrica

UFPA / CT / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
200

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

APLICAÇÃO DE REDES NEURAIS ARTIFICIAIS NA CLASSIFICAÇÃO DE
PADRÕES FILOGEOGRÁFICOS COM BASE NA VARIABILIDADE GENÔMICA
DO DNA MITOCONDRIAL

AUTOR: LARISSA LUZ GOMES

DISSERTAÇÃO DE MESTRADO SUBMETIDA À AVALIAÇÃO DA BANCA
EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-
GRADUAÇÃO EM ENGENHARIA ELÉTRICA DA UNIVERSIDADE FEDERAL
DO PARÁ E JULGADA ADEQUADA PARA OBTENÇÃO DO GRAU DE
MESTRE EM ENGENHARIA ELÉTRICA NA ÁREA DE SISTEMAS DE
ENRGIA ELÉTRICA

APROVADA EM 20 / 12 / 2007

BANCA EXAMINADORA:

Prof. Dr. Roberto Célio Limão de Oliveira, Dr.
Orientador – UFPA

Prof. Dr. Ândrea Kelly Campos Ribeiro dos Santos, Dra.
Co – Orientadora – UFPA

Prof. Dr. Adriana Castro, Dra.
Membro da Banca – UFPA

Prof. Eloi Luiz Favero, Dr.
Membro da Banca – UFPA

Prof. Dr. Sidney Emanuel Batista dos Santos
Membro da Banca – UFPA

VISTO:

Prof. Dr. Evaldo Gonçalves Pelaes
(COORDENADOR DO PPGE/ITEC/UFPA)

Dedico esta dissertação aos meus pais,
Nonato e Silvana, e ao meu irmão Luís
Antônio, que são o meu modelo de vida
e meu conforto nos momentos difíceis.

Larissa.

AGRADECIMENTOS

Primeiramente a Deus, por ter me guiado pelos melhores caminhos e por ter me dado forças nos momentos difíceis, fazendo com que eu não desistisse nunca e assim conseguisse atingir meu objetivo.

Aos meus pais, meus grandes e únicos heróis e ídolos, a quem eu devo tudo que sou e tenho. Sempre estando ao meu lado, me apoiando, incentivando, encorajando, em todos os momentos. E **ao meu irmão** que sempre torceu por mim.

Ao meu orientador, Prof^o. Dr. Roberto Célio Limão de Oliveira, por ter acreditado em mim e principalmente no meu projeto e ter sido muito paciente, atencioso e solícito.

A minha Co-orientadora, Profa. Dra. Ândrea Kelly Campos Ribeiro dos Santos, pela atenção no decorrer deste trabalho, além de me acolher de forma carinhosa no Laboratório de Genética Humana e Médica.

Ao Wady Charone Júnior, que me incentivou a ingressar neste programa de mestrado e teve enorme contribuição na elaboração do primeiro projeto.

Ao meu grande amigo e Mestre Prof^o. Fabiano Cordeiro Moreira, pela paciência, boa vontade e conselhos técnicos do início ao fim deste trabalho.

Aos Mestres Biólogos e mais do que isso, Grandes Amigos, Bruno Maia Carvalho e Ana Cecília Feio dos Santos, os quais me ajudaram muito nas inúmeras dúvidas que surgiram na parte biológica do trabalho, do início ao final, assim como me acolheram e inseriram no grupo de bioinformática, mostraram também como o mundo das pesquisas científicas é muito bonito e rico para quem tem interesse e persistência em estudá-lo. Além de terem ensinado o valor de uma grande amizade, aquela que está sempre presente em todas as horas e momentos. Muito Obrigada!

A minha amiga Ana Régia que sempre me animou, incentivou e ajudou, desde os tempos de Cesupa e durante todo o mestrado, principalmente durante as difíceis horas de estudos para as inúmeras provas e trabalhos.

Ao amigo Jessé Lobato, o qual começou tudo me apresentando ao grupo de bioinformática possibilitando assim, a realização deste trabalho.

E a todos aqueles que me ajudaram direta ou indiretamente a realizar e concluir esta dissertação.

SUMÁRIO

LISTA DE FIGURAS	ix
LISTA DE TABELAS	xi
GLOSSÁRIO.....	xii
RESUMO	xiii
ABSTRACT	xiv
CAPÍTULO I	
1. INTRODUÇÃO.....	1
1.1. ORGANIAZAÇÃO DO TRABALHO.....	2
CAPÍTULO II	
2. ORIGEM E DISPERSÃO DO HOMEM MODERNO.....	4
2.1. POVOAMENTO DA AMÉRICA.....	5
2.2. POVOAMENTO DO BRASIL.....	6
2.3. POVOAMENTO DA AMAZÔNIA.....	7
2.4. ESTUDOS GENÉTICOS.....	7
2.4.1 A Molécula de DNA.....	7
2.4.1.1 Y-DNA.....	9
2.4.1.2 DNA Mitocondrial.....	9
2.4.2 Polimorfismo do DNA.....	11
2.4.3A variabilidade do DNA Mitocondrial nos Primeiros Grupos Étnicos.....	12
2.4.3.1 Polimorfismo do mtDNA em Populações Sul- Americanas.....	14
2.5. ANÁLISE DO mtDNA – BIOINFORMÁTICA.....	16
2.5.1 Técnicas Computacionais.....	16
CAPÍTULO III	
3. REDES NEURAIIS ARTIFICIAIS: VISÃO GERAL.....	18
3.1. FUNDAMENTOS BIOLÓGICOS.....	19
3.2. REDES NEURAIIS ARTIFICIAIS (RNA's).....	20
3.3. PRINCIPAIS CARACTERÍSTICAS DAS RNA's.....	21
3.4. MODELOS DE NEURÔNIOS ARTIFICIAIS.....	22

3.5. ARQUITETURA DE REDES NEURAIIS.....	25
3.6. PROCESSO DE APRENDIZAGEM DAS RNA's.....	26
3.7. ALGORITMO DE RETROPOGAÇÃO DO ERRO (BACKPROPAGATION).....	27
3.8. JUSTIFICATIVA E OBJETIVOS.....	30
CAPÍTULO IV	
4. MODELAGEM DA REDE NEURAL.....	32
4.1. ARQUITETURA E TREINAMENTO DA REDE NEURAL.....	32
4.2. RESULTADOS.....	34
CAPÍTULO V	
5. DISCUSSÃO.....	39
5.1. MODELAGEM DAS CAMADAS DE ENTRADA E INTERMEDIÁRIAS.....	39
5.2. CONFIGURAÇÕES E EFICIÊNCIA DO APRENDIZADO.....	40
5.3. CONSIDERAÇÕES FINAIS.....	41
5.4. TRABALHOS FUTUROS.....	42
REFERÊNCIAS BIBLIOGRÁFICAS.....	43
ANEXO 1.....	50
ANEXO 2.....	52
ANEXO 3.....	55
ANEXO 4.....	57

LISTA DE FIGURAS

FIGURA 2.1 – Mapa mostrando os movimentos migratórios do Homem Moderno.....	5
FIGURA 2.2 – Estrutura do DNA.....	8
FIGURA 2.3 - Figura esquemática do DNA mitocondrial.....	10
FIGURA 2.4 – Figura esquemática da herdabilidade uniparental do DNA mitocondrial e YDNA.....	11
FIGURA 2.5 - Mapa demonstrativo da dispersão de alguns haplogrupos mitocondriais pelo mundo.....	14
FIGURA 3.1 – Exemplo de um tecido neural do córtex cerebral.....	18
FIGURA 3.2 – Exemplo de uma sinapse biológica.....	19
FIGURA 3.3 – Exemplo de um neurônio biológico.....	20
FIGURA 3.4 – Modelo de Neurônio de McCulloch e Pitts.....	22
FIGURA 3.5 – Modelo de Neurônio atual.....	23
FIGURA 3.6 – Funções de Ativação do Neurônio Artificial.....	24
FIGURA 3.7 – Grafo mostrando um exemplo de uma rede de camada única, direta e totalmente conectada.....	25
FIGURA 3.8 – Grafo mostrando um exemplo de uma rede de múltiplas camadas, direta e totalmente conectada.....	26
FIGURA 3.9 – Exemplo de uma rede Mult-Layer Perceptron (MPL) com algoritmo de aprendizado do tipo <i>Backpropagation</i>	28
FIGURA 4.5 - Gráfico dos dados de saída de Ameríndio, antes do treino (cor azul), dados durante o treino (cor verde) e depois do treino, validação (cor vermelha).....	35
FIGURA 4.6 - Gráfico dos dados de saída de Africano, antes do treino (cor azul), dados durante o treino (cor verde) e depois do treino (cor vermelha).....	36
FIGURA 4.7 - Gráfico dos dados de saída de Europeu, antes do treino (cor azul), dados durante o treino (cor verde) e depois do treino (cor vermelha).....	36
FIGURA 4.8 - Gráfico dos dados de saída de Asiático, antes do treino (cor azul), dados durante o treino (cor verde) e depois do treino (cor vermelha).....	37

FIGURA 4.9 - Gráfico de treino das redes de Ameríndio e Africano, respectivamente.....37

FIGURA 4.10 - Gráfico de treino das redes de Europeu e Asiático, respectivamente...38

LISTA DE TABELAS

TABELA 2.1 - Estimativa de mistura interétnica em populações brasileiras por meio da análise do mtDNA.....	15
TABELA 4.1 - Descrição da estrutura utilizada na construção das RNA's.....	33

GLOSSÁRIO

Alelos: cada uma das várias formas alternativas do mesmo gene pode se apresentar.

Cromossomo: é uma longa sequência de DNA, que contém vários genes, e outras sequências de nucleótidos (nucleotídeos) com funções específicas nas células dos seres vivos.

Filogeografia: estudos dos princípios e processos que governam as distribuições geográficas de linhagens genealógicas, especialmente aquelas dentro e entre espécies relacionadas

Gene: segmento de um cromossoma a que corresponde um código distinto, uma informação para produzir uma determinada proteína ou controlar uma característica, por exemplo, a cor dos olhos.

Genoma: toda a informação hereditária de um organismo que está codificada em seu DNA.

Linhagens mitocondriais: são linhagens advindas do DNA mitocondrial. O DNA mitocondrial é um DNA que não se localiza no núcleo da célula, mas sim na mitocôndria. O DNA mitocondrial era passado para a prole unicamente através da mãe.

Lócus: local do cromossomo onde está localizado determinado gene ou outra sequência determinada

Mutações: são mudanças repentinas que ocorrem nos genes, ou seja, é o processo pelo qual um gene sofre uma mudança estrutural. As mutações distinguem-se das aberrações por serem alterações a nível de ponto, envolvendo a eliminação ou substituição de um ou poucos nucleotídeos da fita de DNA.

Nucleotídeos: são compostos ricos em energia e que auxiliam os processos metabólicos, principalmente as biossínteses, na maioria das células.

Polímero: são compostos químicos de elevada massa molecular relativa, resultantes de reações químicas de polimerização. Estes contêm os mesmos elementos nas mesmas proporções relativas, mas em maior quantidade absoluta. Os **polímeros** são macromoléculas formadas a partir de unidades estruturais menores (os monómeros).

Seqüências codificadoras: são seqüências de DNA que codificam uma determinada característica.

RESUMO

Historicamente, o processo de formação das populações da Amazônia, assim como de todo território brasileiro, envolveu três grupos étnicos principais: o ameríndio, o europeu e o africano. Como consequência, estas populações possuem em geral constituição miscigenada do ponto de vista social e biológico. Desde o final do século passado, estudos do DNA mitocondrial (mtDNA) tem sido desenvolvidos com o propósito de estimar a mistura interétnica presente nestas populações. Para isto, é de fundamental importância a classificação de uma determinada linhagem de mtDNA em um dos mais de 250 haplogrupos/subclados propostos na literatura. Com o objetivo de desenvolver um sistema automatizado, preciso e acurado de classificação de seqüências (linhagens) de mtDNA, o presente trabalho lançou mão da técnica de Redes Neurais Artificiais (RNA's) tendo como base os estudos de filogeografia. Para esta classificação, foram desenvolvidas quatro redes neurais artificiais diretas, com múltiplas camadas e algoritmo de aprendizagem de retropropagação. As entradas de cada rede equivalem às posições nucleotídicas polimórficas da região hipervariável do DNA mitocondrial, as quais retornam como saída a classificação específica de cada linhagem. Posterior ao treinamento, todas as redes apresentaram índices de acerto de 100%, demonstrando que a técnica de Rede Neural Artificial pode ser utilizada, com êxito, na classificação de padrões filogeográficos com base no DNA mitocondrial.

ABSTRACT

Historically, the process of formation of the populations of the Amazon, as well as from all Brazilian territory, involved three main ethnic groups: the Amerindian, European and African. As a result, these populations have in general admixed constitution the point of view of social and biological. Since the end of the last century, studies of mitochondrial DNA (mtDNA) has been developed for the purpose of estimating the mixture inter present in these populations. For this, it is of fundamental importance classification of a particular strain of mtDNA in one of more than 250 haplogroups/sub-clades proposed in the literature. With the goal of developing an automated system, precise and accurate classification of the sequences (strains) of mtDNA, this has worked hand of the art of Artificial Neural Networks (RNAs) on the basis of the studies of Philogeography. For this classification, four networks have been developed artificial neural direct, with multiple layers and the learning algorithm to backpropagation. The entries of each network equivalent positions at nucleotide polymorphic region's hipervariable of mitochondrial DNA, which returned as output classification specific to each lineage. Subsequent to the training, all the networks had indices of adjustment of 100%, demonstrating that the technique of Artificial Neural Network (ANN) can be used, with success, in the classification of standards Philogeography based on mitochondrial DNA.

CAPÍTULO I

1. INTRODUÇÃO

Estudos Genéticos têm produzido uma grande quantidade de dados biológicos relacionados às seqüências nucleotídica, SNPs e microssatélites (Feio-dos-Santos et al., 2006; Palha et al., 2007; Ribeiro-dos-Santos et al., 2007; Rodrigues et al., 2007). Estes resultados fornecem base para as diferentes investigações genéticas, dentre as quais se destacam os estudos bioantropológicos sobre os mecanismos de dispersão e formação das populações humanas do mundo.

Uma das metodologias mais utilizadas em pesquisas bioantropológicas é a investigação da região controle do mtDNA. Por meio dessas investigações é possível classificar a linhagem de um determinado indivíduo em um dos diferentes perfis mitocondriais, denominado haplogrupos, os quais variam entre as distintas populações humanas que habitam os continentes do planeta. Esta variação está relacionada com eventos biológicos, mutações, adquiridos durante os mecanismos de dispersão do homem moderno. Desta forma, populações de diferentes regiões geográficas do mundo possuem também estoques genéticos diferenciados em relação ao mtDNA.

Segundo Santos et al (1996) analisando as linhagens mitocondriais de uma população é possível entender a sua composição e estrutura genética, assim como entender seus processos históricos de formação e expansão.

Devido ao conteúdo da informação biológica contida no mtDNA, muitos estudos relacionados com a classificação e descrição da variabilidade da região controle de diferentes populações, foram desenvolvidos desde a década de 90 e direcionados tanto para o enfoque bioantropológico quanto para as áreas de genética forense. Como consequência tem-se atualmente um grande volume de seqüências nucleotídicas disponibilizadas em bancos de dados públicos e privados (Beleza et al., 2005 ; Friedlaender et al., 2005 , Behar et al . , 2007)

Para organizar e analisar este conjunto de informações biológicas observou-se a necessidade da utilização de técnicas computacionais com a finalidade de automatizar este processo, uma vez que este é atualmente realizado de forma manual e comparativa,

prejudicando a velocidade e a curacidade. Dentre estes processos, a classificação filogeográfica em haplogrupos mitocondriais é complexa e demorada devido ao grande número de informações (nucleotídeos variáveis) e classes (225 haplogrupos e sub-clados). Este processo ainda é feito em comparação às publicações científicas específicas.

Com o advento da Inteligência Computacional (IC) foram desenvolvidos sistemas que possibilitam a classificação e identificação de padrões de repetição. As Redes Neurais Artificiais (RNA's) vem sendo aplicadas em diversos problemas que envolvem classificação. Nestes casos, há necessidade de manipulação do conhecimento, visando possibilitar a construção de modelos a partir de exemplos previamente apresentados à rede (Haykin, 2001).

Diante da necessidade da classificação de linhagens do mtDNA e ainda, pelas crescentes aplicações das Redes Neurais Artificiais nos problemas que envolvem classificação de padrões, o presente trabalho integra estes conhecimentos de forma multidisciplinar, visto que reúne conhecimento e tecnologia de duas grandes áreas do conhecimento: a Computação e a Biologia. Esta união é conhecida como Bioinformática.

1.1. ORGANIZAÇÃO DO TRABALHO

O trabalho está organizado em sete capítulos como descrito a seguir:

O Capítulo I faz uma breve descrição do tema abordado nesta dissertação, evidenciando suas particularidades e áreas de abrangência assim como a mesma está organizada.

O Capítulo II apresenta uma visão dos vários conceitos biológicos que serão utilizados no decorrer do trabalho.

O Capítulo III apresenta as Redes Neurais Artificiais de forma a proporcionar uma visão geral a respeito dos principais aspectos que compõem a técnica, possibilitando assim um completo entendimento da sua utilização.

No Capítulo IV explica-se o como foram implementados os métodos e técnicas tratadas nos capítulos II e III. Além de exibir e discutir os resultados da classificação das linhagens mitocondriais por meio de Redes Neurais Artificiais.

O Capítulo V exibe as discussões obtidas a partir da análise dos resultados conseguidos no trabalho e apresenta as conclusões do mesmo, assim como perspectivas.

CAPÍTULO II

2. ORIEM E DISPERSÃO DO HOMEM MODERNO

A origem do homem e a forma como este colonizou vários territórios em todo o mundo constitui um grande desafio para vários segmentos da ciência. Pesquisas realizadas sobre o tema deram origem a algumas hipóteses. A teoria mais aceita na área biológica é a de Rebecca Cann intitulada: “A Hipótese Migracionista do Homem Moderno” (Cann et al., 1987). Segundo a autora, o homem moderno (*Homo sapiens*) surgiu no leste do continente africano, 200 mil anos antes do presente, e a partir daí dispersou-se subsequente em sucessivas levas migratórias. Foi desta forma que o homem ocupou inicialmente o continente africano e posteriormente conquistou os demais continentes do planeta.

A hipótese migracionista foi resultado de análises filogeográficas de linhagens mitocôndrias representativas de todo mundo. Esta análise foi possível porque o mtDNA acumulou um conjunto de mutações representativas, uma vez que a sua taxa de evolução é proporcional ao tempo em que os grupos populacionais humanos dispersaram-se pelo mundo. Essa relação entre o número de mutações e o tempo de divergência é chamada de relógio molecular (Cann et al, 1987)

Evidências arqueológicas e investigações antropológicas e genéticas ainda não convergiram para um único modelo de dispersão (Merriwether et al., 1991; Kastle and Smith, 2001; Neves et al., 2007). De uma maneira geral, podemos considerar que a dispersão ocorreu em três movimentos: o primeiro movimento conduziu o homem ao norte e oeste do continente africano, o segundo movimento conduziu o homem ao oriente médio, o terceiro movimento há cerca de 50 a 65 mil anos, levou o homem em direção ao oriente, à Arábia, ao sudeste da Ásia e Oceania e o quarto movimento deu seqüência aos anteriores. Novas levas migratórias partiram da África, já com conhecimento de tecnologias do paleolítico superior, chegando inclusive ao continente americano, espalhando-se por ele e nele povoando-o e vivendo como caçadores e coletores.

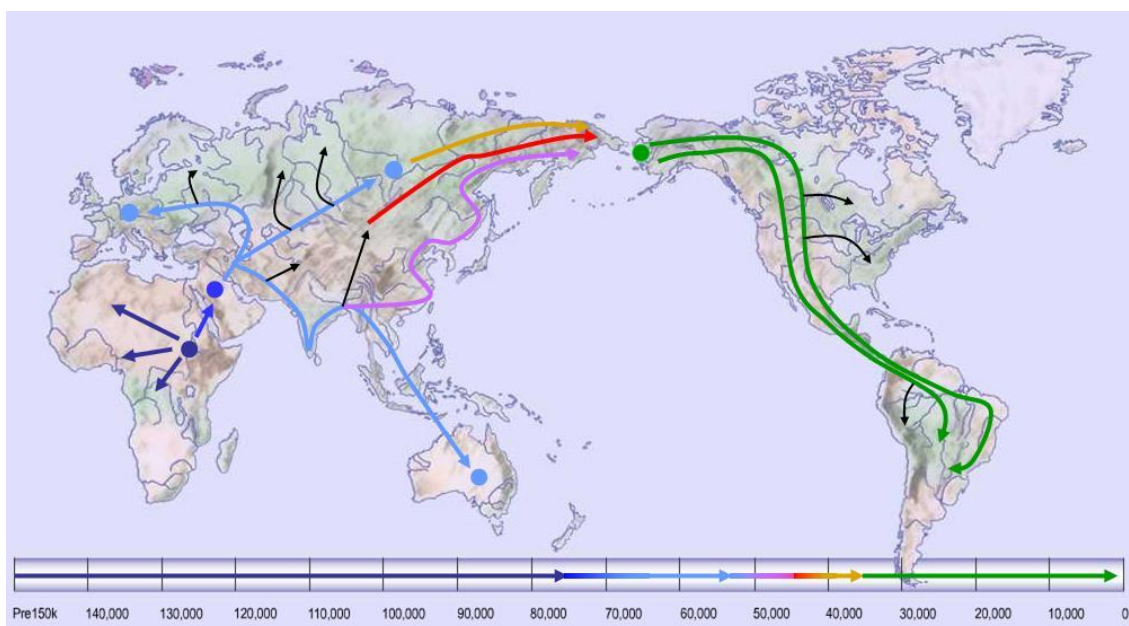


Figura 2.1 – Mapa mostrando os movimentos migratórios do Homem Moderno.

2.1. POVOAMENTO DA AMÉRICA

Neves (1989); Torroni et al. (1996) e Watson et al. (1999) e demonstram que migrações originárias do leste da Ásia, entre 18 mil e 45 mil anos atrás, trouxeram para o continente americano populações ancestrais com características antropométricas semelhantes aos povos orientais do leste asiático. Esta idéia é baseada na investigação de ossos encontrados em sítios arqueológicos da América do Sul e do Brasil, em que estes guardam fortes semelhanças com os indígenas atuais (André Prous, 1999). Desta forma é possível então explicar a semelhança entre populações asiáticas e populações nativas das Américas. Estes estudos são embasados na teoria que os primeiros habitantes do continente americano vieram da Ásia, devido à redução dos níveis dos oceanos o que possibilitou a travessia do estreito de Bering e a chegada ao continente americano.

Passado alguns séculos, mais especificamente no final do século XV, os europeus, chegaram às Américas e encontraram um continente habitado por milhões de pessoas, civilizações e cidades maiores que Madri ou Lisboa (Salzano e Bortolini,

2002). Mais tarde, principalmente no século XVIII, o conquistador europeu promoveu a vinda dos primeiros escravos africanos para as terras recém-descobertas. Nos quatro séculos que se seguiram ao desembarque de Colombo, cerca de doze milhões de africanos das regiões ao Sul do Saara foram transportados para as Américas. (Sérgio D.J. Pena; Maria Cátira Bortolini, 2004). Destes, quatro milhões foram trazidos para o Brasil, sendo a grande maioria deles originários da África central e ocidental.

2.2. POVOAMENTO DO BRASIL

A formação do povo brasileiro é resultado de cinco séculos de cruzamentos interétnicos, o que lhe proporcionou uma das populações mais heterogêneas do mundo, resultado principalmente da miscigenação entre povos de três continentes: europeus (colonizadores), escravos africanos (mão de obra) e ameríndios (colonizados) (Alves-Silva et al., 2000).

Uma intensa mistura social e biológica ocorreu entre colonizadores, colonizados e escravizados desde o início desse contato (Pena, 2002; Bortolini et al., 2004). Durante alguns séculos, vários termos surgiram ao longo dos anos para definir e classificar com base no fenótipo individual as diferentes ancestralidades: africana, europeia e ameríndia e suas combinações.

Mais recentemente, outros grupos humanos imigraram para o Brasil, provenientes de diferentes regiões do mundo, principalmente do Japão, Coreia, Síria e Líbano, o que proporcionou ao povo brasileiro, um acréscimo de diversidade à sua mistura interétnica. De acordo com Callegari-Jacques e Salzano (1999), 58% dos imigrantes que chegaram ao Brasil entre 1500 e 1972 eram europeus, 40% eram africanos e somente 2% eram asiáticos.

2.3. POVOAMENTO DA AMAZÔNIA

Estudos históricos ressaltam que populações primitivas chegaram na Amazônia há cerca de 12.000 anos atrás, vindos do norte e oeste do continente americano (Wallace et al., 1995; Santos et al., 1999).

Oficialmente, os primeiros conquistadores portugueses chegaram na região amazônica no início do século XVII, e se fixaram com a fundação da cidade de Belém em 1616 (Cruz, 1973). Por outro lado, os escravos africanos foram trazidos somente na segunda metade do século XVIII, em número considerável, cerca de 53.000 indivíduos, trazidos principalmente pela Companhia Geral do Grão Pará e Maranhão (Salles, 1988).

Análise de registros gráficos (ofícios, cartas e registros históricos) assim como os diferentes “marcadores étnicos”, tais como ritmos, culinária, língua, culturas e sociedades, permitem que a história de ocupações das regiões e territórios do mundo, assim como da Amazônia, possa ser recontada. Da mesma forma, a análise de eventos biológicos, registrados em unidades moleculares, como as proteínas e o Ácido Desoxirribonucléico (DNA), atualmente nos recontam a mesma história do ponto de vista biológico, pois guardam informações relevantes sobre cada grupo populacional.

2.4. ESTUDOS GENÉTICOS

2.4.1. A MOLÉCULA DE DNA

O DNA é uma molécula orgânica que contém a informação para o desenvolvimento e funcionamento de todos os organismos vivos. O seu principal papel é armazenar as informações necessárias para a codificação das proteínas, ácidos ribonucleicos (RNAs) e transmissão das características hereditárias.

Os segmentos de DNA que são responsáveis por carregar a informação genética, para síntese de moléculas de RNA e proteínas, são denominados genes. O

restante da seqüência de DNA tem importância estrutural ou está envolvido na regulação do uso da informação genética (Brown, 2002).

Do ponto de vista químico, o DNA (Figura 2.2) é um longo polímero de unidades simples (monômeros) denominados nucleotídeos unidos por ligações fosfodiéster, cujo cerne é formado por bases nitrogenadas, açúcar (neste caso desoxirribose) e fosfatos. As bases nitrogenadas são de dois tipos: as purinas, adenina (A) e a guanina (G), e as pirimidinas, timina (T) e citosina (C). É a seqüência dessas bases ao longo da molécula de DNA que carrega a informação genética, a qual especifica a seqüência linear dos aminoácidos das proteínas, por meio do código genético.

Do ponto de vista da herdabilidade do material genético temos em especial duas porções do DNA genômico que não sofrem recombinação e representam a patrilinearidade - Y-DNA e matrilinearidade - mtDNA (DNA mitocondrial).

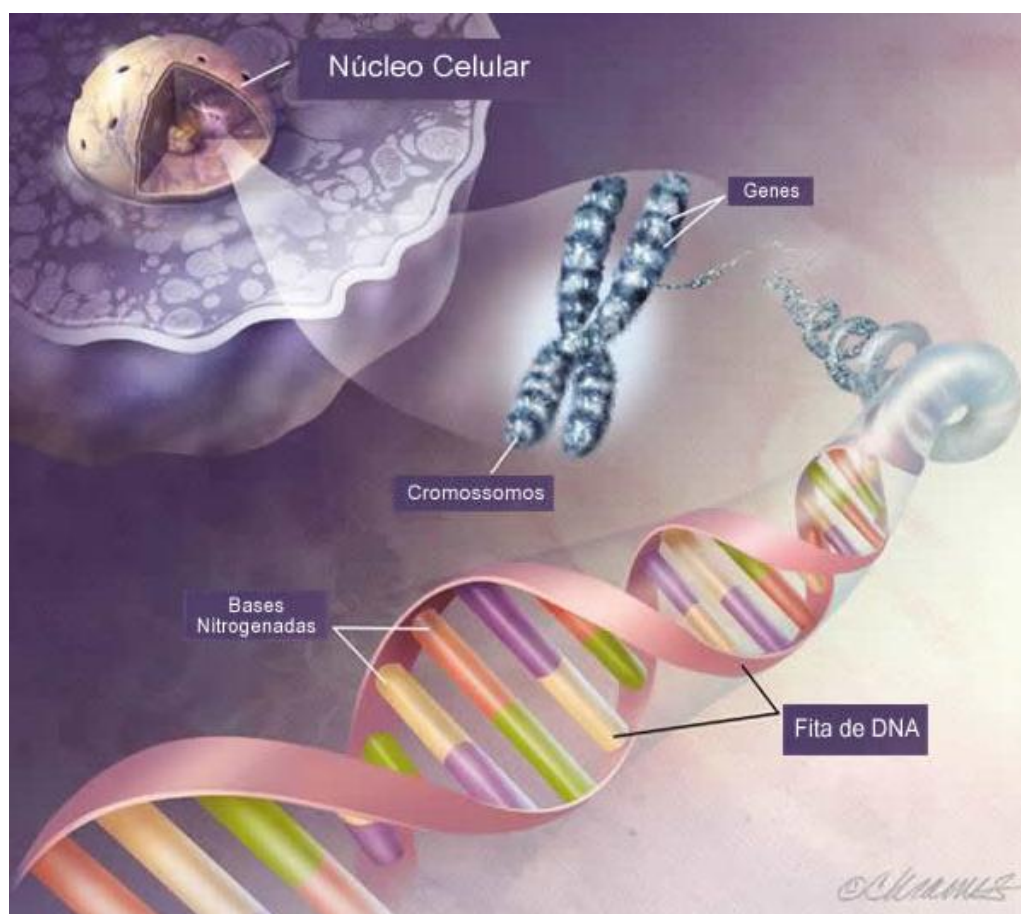


Figura 2.2 – Estrutura do DNA

2.4.1.1. Y-DNA

O material genético encontrado no cromossomo Y, responsável pelas características masculinas é chamado de Y-DNA. Na meiose, este cromossomo não sofre recombinação em mais de 90% de sua extensão, região chamada de NRY (non recombining portion of the Y) e é transmitido pelo pai somente a seus filhos homens. A única fonte de variação é proporcionada por eventos mutacionais que, com o tempo, se acumulam gradativamente (revisão em Hammer e Zegura, 1996).

2.4.1.2. DNA Mitocondrial

O DNA mitocondrial humano (mtDNA) é uma molécula pequena, circular, localizada no interior de mitocôndrias (organelas citoplasmáticas) responsáveis pela produção de energia na célula, composta por duas cadeias: uma interna, cadeia leve (rica em pirimidinas: C-T), e outra externa, cadeia pesada (rica em purinas: A-G).

O tamanho do genoma mitocondrial é de 16.569pb, os quais foram totalmente seqüenciados por Anderson et al. (1981) e revisados por Andrews et al. (1999). A maior parte, mais de 90% do genoma, é codificadora, ou seja, é usada como molde para síntese de proteínas mitocondriais e RNAs (transportadores e ribossomais). A taxa de mutação nesta região é aproximadamente cinco vezes maior do que a do DNA nuclear.

Em associação com as seqüências codificadoras, o mtDNA apresenta um pequeno número de nucleotídeos compondo seqüências em regiões intergênicas, dispersas pela molécula, e a região da Alça-D ou região hipervariável, localizada entre os genes tRNAPhe e tRNAPro, que possui a função biológica de controlar a replicação da molécula (Giles et al., 1980).

O modelo do mtDNA humano, baseado em Brown (2002), encontra-se representado na figura 2.3.

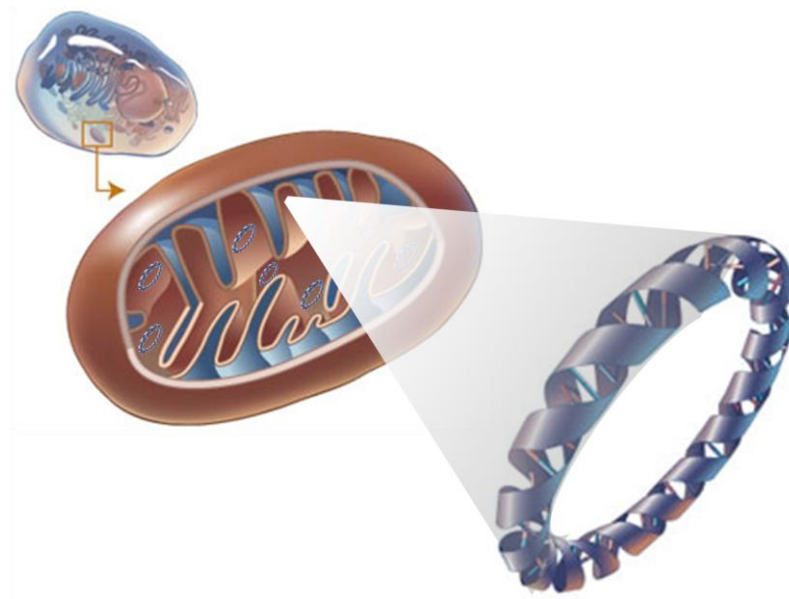


Figura 2.3 – Figura esquemática do DNA mitocondrial.

Algumas características particulares tornam o genoma mitocondrial uma excelente ferramenta para estudos de diversidade genética e evolução dos humanos. Entre elas citamos: o mtDNA é herdado matrilinearmente, ou seja, é transmitido somente das mães para filhos e filhas (Figura 2.4), não sofrendo recombinação com o material paterno, portanto, representa uma herança haplóide (Giles et al., 1980); elevado número de cópias por célula (10 a 100), facilitando o acesso ao genoma (Álvarez et al., 2001); a molécula apresenta uma rápida evolução. A taxa de mutação na região codificadora é aproximadamente cinco vezes maior do que a taxa de mutação do DNA nuclear (Cann et al., 1987; Wallace et al., 1987, 1995; Foster et al., 1996), e a região hipervariável evolui cinco vezes mais que o restante da molécula, portanto 25 vezes mais rápido que o DNA nuclear (Vigilant et al., 1991; Wallace, 1995; Pena et al., 2002); possui um ineficiente mecanismo de reparo de mutações (Saccone et al., 1999; Croteau et al., 1999).

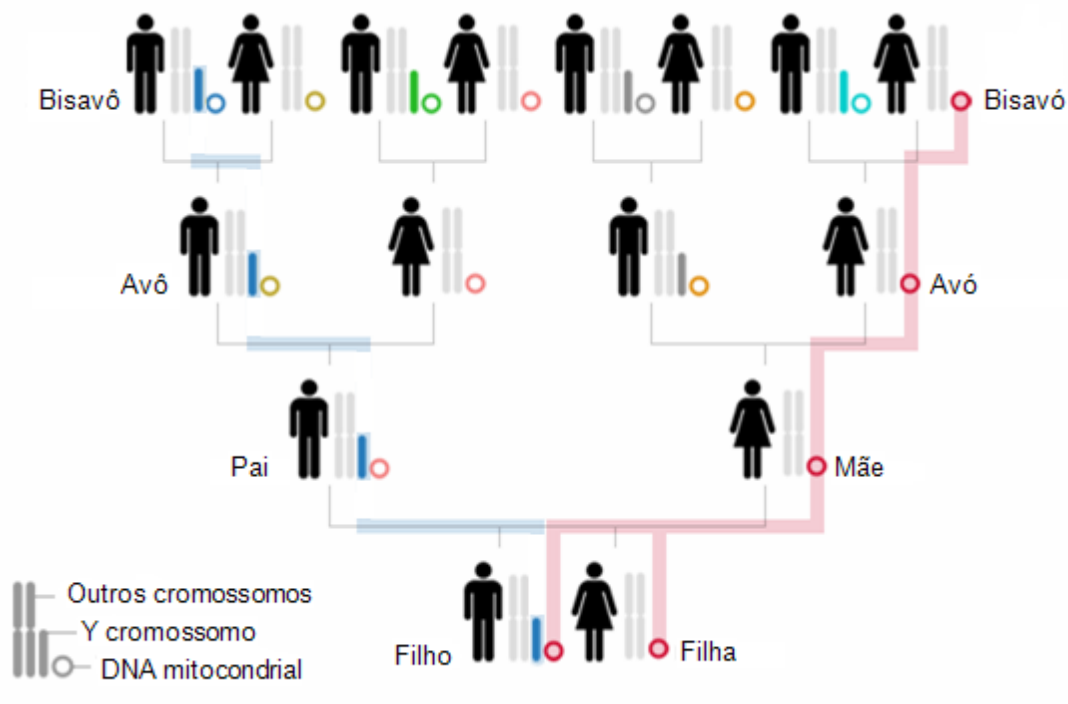


Figura 2.4 – Figura esquemática da herdabilidade uniparental do DNA mitocondrial e Y-DNA.

2.4.2. POLIMORFISMO DO DNA

Classicamente, os polimorfismos genéticos são definidos em função da presença de múltiplos (mais de um) alelos em um mesmo locus, em que dois desses alelos apresentam frequência igual ou superior a 1% em uma população (Brown, 2002).

Os polimorfismos de DNA são classificados em função de sua natureza molecular e da sua localização no genoma, possibilitando as mais diversas abordagens para o estudo de genética de populações. Os tipos mais frequentes podem ser classificados como:

- SNPs (polimorfismos de nucleotídeo único) caracterizados pela alteração de um único nucleotídeo na sequência de DNA;
- Inserções ALU, nas quais uma sequência de DNA apresenta o elemento ALU (uma curta sequência de DNA transponível);

- Inserções/Deleções, quando a sequência de DNA é caracterizada por um número de nucleotídeos maior que a sequência ancestral, no caso de inserção, ou a sequência nucleotídica sofre a perda de nucleotídeos, caracterizando uma deleção.

Em particular no mtDNA os polimorfismos encontrados são do tipo SNPs ou deleções e a variação na presença/ausência desses polimorfismos vão discriminar os diferentes haplogrupos.

2.4.3. A Variabilidade do DNA Mitocondrial nos Principais Grupos Étnicos

O mtDNA é uma das ferramentas que tem sido amplamente utilizada para o esclarecimento da dinâmica de formação e evolução de populações humanas (Cann et al., 1987; Merriwether et al., 1991, 1996; Stoneking, 1993; Richards et al., 1996; Ingman et al., 2000; Silva Jr. et al., 2002a, 2002b). A maior parte destes estudos, utilizou algumas abordagens para a detecção de polimorfismos nas sequências do mtDNA:

- UEPs (Uniques events polymorphisms) que são RFLPs (polimorfismos de tamanho de fragmento de restrição), deleções e inserções (Cann & Wilson, 1983; Hetzberg et al., 1989; Torroni et al., 1992; Chen et al., 1995; Soodyall et al., 1996; Watkins et al., 1999);
- Sequenciamento direto de segmentos hipervariáveis da região controle ou alça-D (Horai & Hayasaka, 1990; Vigilante et al., 1991; Torroni et al., 1993; Tajima et al., 2004, Kivisild et al., 2004).

As mutações que ocorrem no mtDNA são sempre herdadas em conjunto e constituem um haplótipo. Os haplótipos, por sua vez, são reunidos em diferentes haplogrupos, os quais variam entre os grupos étnicos presentes na atualidade. Desse modo, o que se observa é uma identidade, uma marca genética (*motiff*), distinta entre grupos populacionais, resultante de uma rápida e recente evolução na molécula de DNA mitocondrial (Pena, 2002).

Em populações ameríndias, a investigação dos polimorfismos no mtDNA em conjunto a análise de RFLPs e o seqüenciamento direto da alça-D, demonstraram a existência de cinco haplogrupos prevalentes: A, B, C, D e X (Ward et al., 1991; Torroni et al., 1992, 1994a; Horai et al., 1993; Santos et al., 1996; Smith et al., 1999; Foster et al., 1996; Reidla et al., 2003).

Os haplogrupos das populações de origem européia dividem-se em dois grandes grupos:

- Um que não apresenta o sítio de restrição para a enzima *Dde* I, no nucleotídeo 10.394 que reúne os haplogrupos H, T, U, V, W e X;
- Um segundo grupo que apresenta o sítio de restrição para a mesma enzima, onde encontra-se os haplogrupos I, J e K (Torroni et al., 1996).

Em povos de origem asiática, observa-se a presença dos haplogrupos M, N, G, F, Z, Y e R, específicos desta região continental, assim como, os haplogrupos A, B, C, D e X, que também estão representados em povos ameríndios do continente americano (Yao et al., 2002; Hernstadt et al., 2002).

Em populações do continente australiano (Polinésia, Melanésia, Micronésia e Australasia) além de compartilharem alguns haplogrupos asiáticos, essas populações possuem a prevalências de outros haplogrupos como P e Q e seus subclados.

As populações africanas participaram das primeiras análises de RFLP no mtDNA desenvolvidas em grupos humanos. Desde a última década, muitos dados de sítos polimórficos têm sido coletados em comunidades de diferentes regiões da África. Estudos de RFLPs, seqüenciamento das regiões hipervariáveis da alça-D e análises filogenéticas demonstraram que em todo o mundo a maior diversidade em relação ao mtDNA é encontrada nas populações africanas, assim como sugerem a origem e expansão do homem moderno neste continente (Cann et al., 1987; Vigilant et al., 1991; Chen et al., 1995; Rando et al., 1998; Salas et al., 2002)

Nas populações africanas observa-se a prevalência dos haplogrupos que variam de L0 a L5, assim como suas subclassificações (subclados) (Salas et al., 2002, 2004). Ver figura 2.5.

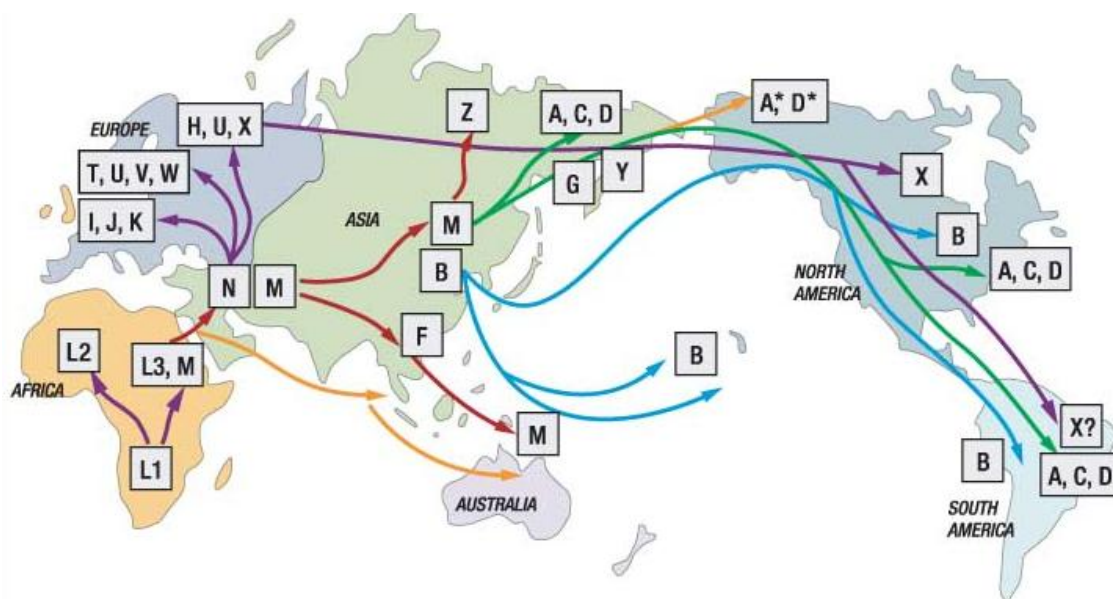


Figura 2.5 – Mapa demonstrativo da dispersão de alguns Haplogrupos mitocondriais pelo mundo.

2.4.3.1. Polimorfismo do mtDNA em Populações Sul-Americanas

A investigação de polimorfismos no mtDNA, por meio de biologia molecular, tem sido uma metodologia utilizada em vários países, para fazer-se inferências sobre a diversidade existente entre os diferentes grupos populacionais, assim como, seus mecanismos de formação e dispersão (Torroni et al., 1996; Rando et al., 1998; Passarino et al., 1998; Chen et al., 2000; Derbenerva et al., 2002; Salas et al., 2004).

No Brasil, pesquisas por biologia molecular já foram desenvolvidas com estes objetivos utilizando o DNA mitocondrial. Para exemplificar a miscigenação de diferentes grupos étnicos na composição das populações brasileiras, sejam elas miscigenadas, afro-descendentes ou euro-descendentes, este trabalho compilou resultados de estudos que utilizaram o mtDNA (Tabela 2.1)

Tabela 2.1 - Estimativa de mistura interétnica em populações brasileiras por meio da análise do mtDNA.

População	Frequência (%)			Referência
	Africano	Ameríndio	Europeu	
<i>Região Norte</i>	15	54	31	5
¹ Belém-PA	22,6	64,5	*	1
² Curiaú-AP	52,3	47,3	0	2
¹ Cametá-PA	40	60	0	3
¹ Santarém-PA	14,0	83,5	12,5	4
² Tamauari-PA*	42	54,8	0	6
² Mazagão-AP	39,4	57,6	3	6
² Marajó-PA	55,9	44,1	0	6
² Pitimandeu-PA	69	31	0	6
² Pacoval-PA	35	63	2	6
² Trombetas-PA	56,3	46,7	0	6
² Trombetas-PA	34	66	0	3
³ Santa-Flor-MA	19,5	75	0	7
<i>Região Nordeste</i>	44	22	34	5
¹ Salvador-BA	67	11	22	3
² Cajueiro-MA	70	30	0	3
<i>Região Sudeste</i>	34	33	31	5
¹ Ribeirão Preto-SP	95	0	5	3
<i>Região Sul</i>	12	22	66	5
¹ Porto Alegre-RS	93	7	0	3
¹ Ribeirão Preto-SP	95	0	5	3
² Paredão-RS	67,0	22,0	11,0	3
<i>Brasil</i>	28	33	39	5

¹População urbana; ²População Afro-descendente; ³População Euro-descendente
 1. Marinho (2004); 2. Ribeiro-dos-Santos et al. (2002); 3. Bortolini et al. (1999); 4. Feio-dos-Santos (2004); 5. Alves-Silva (20004); 6. Carvalho (2005); 7. Freitas et al. (2006).* Em Belém, Tamauari e Santa-Flor não foi observada a presença de linhagens de ancestralidade europeia. Todavia, existe um grupo de amostras classificadas como “outras” no qual podem estar contidas estas matrilineagens.

2.5. ANÁLISE DO MTDNA – BIOINFORMÁTICA

Os estudos filogeográficos do mtDNA têm produzido uma grande quantidade de informações biológicas, visto que, o estudo da região controle possibilita ao pesquisador trabalhar com aproximadamente 1.200 pontos variáveis, sendo cada ponto de grande importância para o entendimento da composição, estrutura genética, processos históricos de formação e expansão de uma determinada população.

Com o aperfeiçoamento das técnicas de biologia molecular, como o sequenciamento automático, têm-se uma maior facilidade e velocidade na obtenção de resultados, tanto em qualidade, quanto em quantidade. Desta forma, observou-se a necessidade da utilização de ferramentas e técnicas computacionais, de modo a automatizar o diagnóstico dos dados biológicos produzidos.

Visando a integração de conhecimentos de duas grandes áreas distintas, a Biologia e a Computação, originou-se uma nova área, multidisciplinar, conhecida como Bioinformática, a qual tem o objetivo de integração de conceitos biológicos e computacionais pra resolver questões biológicas, em suas variações e complexidades.

A Bioinformática pode ser descrita como a aquisição, manipulação, construção, edição, análise e armazenamento de dados biológicos, especificamente de ácidos nucleicos e proteínas. Análise em laboratório de dados biológicos é difícil e custosa, portanto, técnicas computacionais são essenciais (Ferreira, 2007). Hoje novas áreas foram criadas, e uma das mais importantes é a Biologia Computacional.

A Biologia Computacional é a área que estuda o desenvolvimento de modelos matemáticos e computacionais visando à compreensão e previsão do comportamento tanto de sistemas macromoleculares quanto de sistemas biológicos complexos (Labinfo - <http://www.labinfo.lncc.br/>).

2.5.1. Técnicas Computacionais

Várias técnicas computacionais são utilizadas pela Bioinformática, dentre elas destacam-se:

- Bancos de Dados: é utilizado para armazenar, recuperar, editar e manipular dados biológicos.
- Data Mining ou Mineração de Dados: consiste na busca de informação valiosa existente em grandes volumes de dados.
- Inteligência Computacional (IC): dentre as principais técnicas de IC, o Aprendizado de Máquina é bastante utilizada para desenvolver métodos capazes de extrair descrições de conceitos a partir de exemplos e tem sido muito bem sucedido quando aplicada a diversos problemas de Bioinformática.
- Redes Neurais Artificiais(RNA's): podem ser aplicadas a diversos problemas da Bioinformática, que envolvem reconhecimento de padrões. Os tipos de Redes Neurais Artificiais utilizados em Bioinformática são:
 - Perceptron: é muito utilizado em problemas de classificação simples;
 - Perceptron Multicamadas: utilizado em problemas biológicos de classificação mais complexa onde, geralmente, o número de unidades de saída é igual ao número de classes existentes;
 - Mapas de Kohonen: é utilizado quando o número de agrupamentos formados pelas saídas da rede irá refletir o número de classes existentes, o qual, em geral, deve ser conhecido a priori.

CAPÍTULO III

3. REDES NEURAIS ARTIFICIAIS: VISÃO GERAL

Segundo Renato Vieira e Mauro Rosenberg o ser humano é dotado de complexos circuitos neurais no cérebro que constam de variadas conexões entre seus neurônios, as sinapses. Estas interagem entre si e deste modo tornam possível um comportamento inteligente.

A partir da observação do funcionamento do cérebro humano, surgiu a idéia de que, conseguindo modelar computacionalmente as conexões neurais é possível também, criar comportamentos inteligentes em máquinas. Foi neste contexto que surgiram as Redes Neurais Artificiais (RNA's), as quais tiveram sua inspiração na natureza das redes de neurônios e sinapses biológicas (Azevedo, 1999), figuras 3.1 e 3.2 respectivamente.

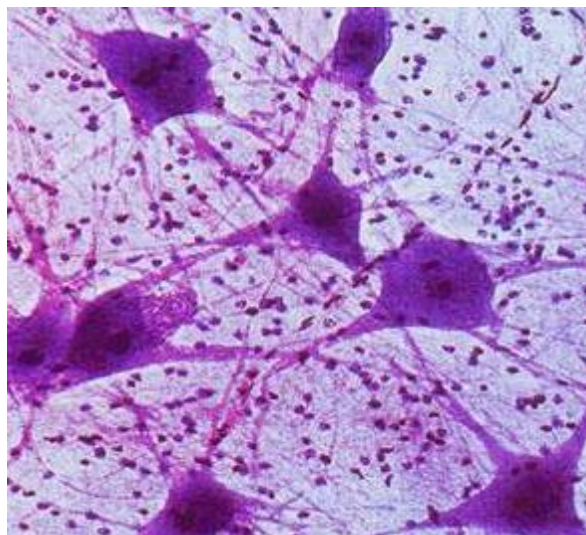


Figura 3.1 – Exemplo de um tecido neural do córtex cerebral

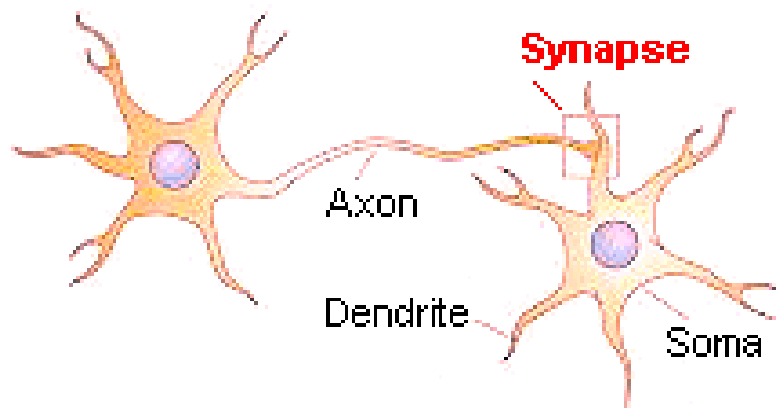


Figura 3.2 – Exemplo de uma sinapse biológica

3.1. FUNDAMENTOS BIOLÓGICOS

O cérebro é uma parte do sistema nervoso que concentra uma boa parte das funções que controlam o corpo. Ele é composto por vários tipos de células nervosas, a principal delas é o neurônio (Figura 3.3).

O neurônio é formado por um corpo celular chamado soma e diversas ramificações. As ramificações são conhecidas como dendritos, conduzem sinais das extremidades para o corpo celular. Existe também uma ramificação, geralmente única, chamada axônio, que transmite um sinal do corpo celular para as suas extremidades. As extremidades do axônio são conectadas com dendritos de outros neurônios pelas sinapses. Em muitos casos, um axônio é diretamente conectado com outros axônios ou com o corpo de outro neurônio. (BARRETO, 1999, p. 69).

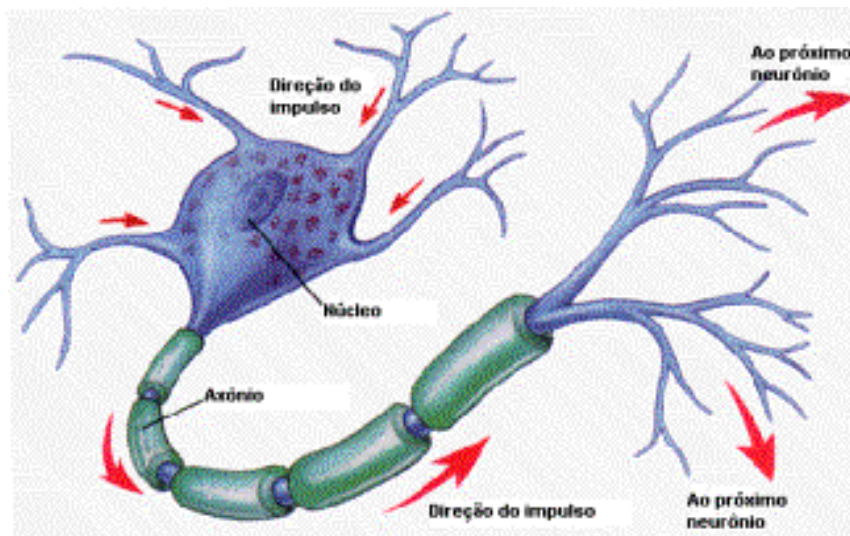


Figura 3.3 – Exemplo de um neurônio biológico

3.2. REDES NEURAIS ARTIFICIAIS (RNA's)

O interesse pelo estudo das Redes Neurais Artificiais (RNA's) deve-se a sua analogia com o cérebro humano. As RNA's possuem algumas características exclusivas do sistema neural, como o processamento paralelo de informação. A vantagem deste processamento esta na forma de armazenamento da informação, porque nos métodos tradicionais a informação fica armazenada em um endereço de memória, e em uma rede neural a informação é compartilhada por todas as unidades de processamento da rede, o que permite o processamento paralelo das informações. Com isso, as redes neurais artificiais têm a vantagem de analisar simultaneamente um número muito grande de variáveis.

“Uma rede neural artificial é um processador macivamente paralelamente distribuído constituído de unidades de processamento simples, que têm a propensão natural de armazenar conhecimento experimental e torná-lo disponível para o uso. Ela se assemelha ao cérebro humano em dois aspectos: 1) O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem. 2) Forças de conexão

entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.” (HAYKIN, 2001).

Segundo Rauber (1998), ao contrário de um computador com arquitetura de Von Neumann que é programado, a RNA é treinada através de exemplos. E através do algoritmo de aprendizagem que os dados são generalizados e o conhecimento é guardado nos pesos da rede. Assim para se construir uma RNA é necessário definir a arquitetura de rede que será utilizada e qual o melhor algoritmo de treinamento para a rede em questão.

As habilidades de generalização e aprendizagem das RNA's significam que estes modelos podem lidar com dados imprecisos e situações não apresentadas anteriormente à rede. Uma rede treinada, com entradas adequadas ao problema, tem a habilidade de generalizar quando se apresentam entradas próximas aquelas já conhecidas por ela.

3.3. PRINCIPAIS CARACTERÍSTICAS DAS RNA'S

As principais características das Redes Neurais Artificiais são:

- **Intenso Paralelismo** - uma das similaridades da rede neural artificial com o cérebro humano é que ambos funcionam como um sistema não linear. O paralelismo é uma característica muito importante, visto que, os sistemas reais são em sua grande maioria não lineares, sendo difíceis de serem modelados e controlados devido a sua complexidade. Desta forma, as RNA's possuem a capacidade de sintetizar mapeamentos complexos e não lineares a partir de treinamentos apropriados. Também devido ao paralelismo, a falha de alguns neurônios não causam efeitos significantes para a performance de todo o sistema, o que é chamado de tolerância a falhas.
- **Robustez** - a rede neural é tolerante a falhas, visto que, mesmo perdendo algumas conexões sinápticas, a mesma provavelmente continuará fornecendo resultados aceitáveis.

- **Aprendizado através de Exemplos** - a partir de um conjunto de exemplos, também chamado conjunto de treinamento, as redes neurais artificiais são capazes de armazenar informação, ou seja, de aprender.
- **Generalização** - é a capacidade que as redes neurais possuem, quando bem treinadas, de fornecer respostas satisfatórias mesmo para padrões de treinamento não incluídos durante o processo de aprendizagem da rede.

3.4. MODELOS DE NEURÔNIOS ARTIFICIAIS

A inspiração para a construção das Redes Neurais Artificiais veio dos neurônios biológicos e do sistema nervoso. O primeiro modelo de neurônio foi proposto por McCulloch e Pitts em 1943.

Segundo Barreto (1999), o modelo proposto por McCulloch e Pitts é também um modelo simples. Ele interpreta o funcionamento do neurônio biológico como sendo um circuito binário, portanto, o modelo proposto é binário e é representado pela figura 3.4.

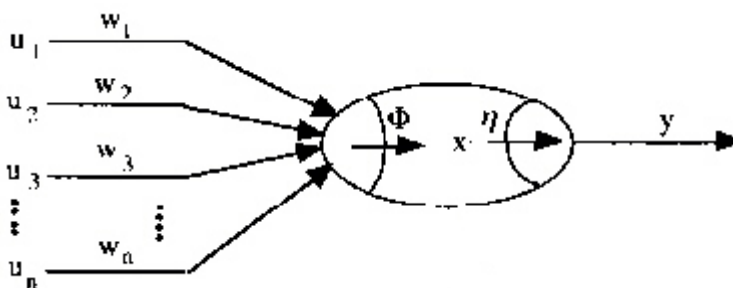


Figura 3.4 – Modelo de Neurônio de McCulloch e Pitts

Neste modelo, são identificados os elementos básicos que compõem o neurônio de McCulloch e Pitts. As entradas da rede são representadas por U , as conexões sinápticas são representadas pelos pesos W da rede, o Φ recebe os sinais de entrada multiplicados pelos pesos da rede, o η será o limiar do neurônio e o Y será a saída do neurônio, que por ser um modelo binário, só apresenta como saída os valores 0 ou 1.

O modelo de neurônio representado na figura 3.5 é o modelo utilizado atualmente.

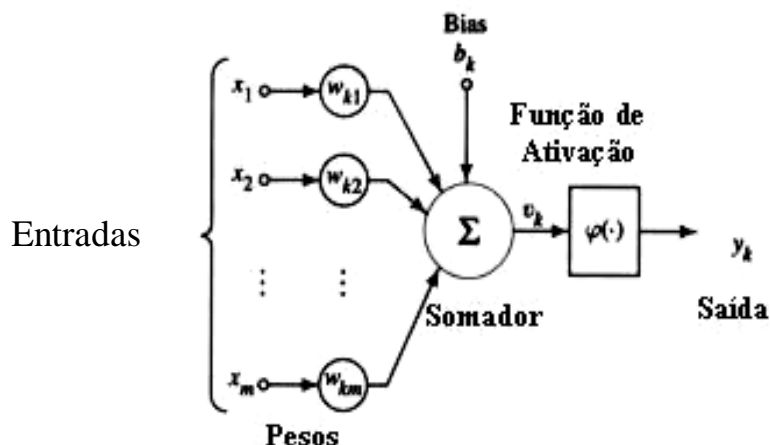


Figura 3.5 – Modelo de Neurônio atual.

Neste modelo são identificados cinco elementos básicos:

- As entradas da rede;
- As conexões sinápticas, representadas pelos pesos W da rede;
- O somador que recebe os sinais de entrada multiplicados pelos pesos da rede;
- O bias que é o responsável por acelerar a convergência da rede, a fim de deslocar a curva de ativação para a direita ou para a esquerda;
- A função de ativação que limita a amplitude do sinal de saída do neurônio. A combinação do somatório de todas as entradas (X) multiplicadas pelos seus respectivos pesos sinápticos (W) é denominada de **net** e em termos matemáticos é representado pela expressão:

$$net_i(t) = \sum_{j=1}^n w_{ij} \cdot x_j(t)$$

Segundo Tafner, Xerez e Rodrigues (1995), a função de ativação ou função de saída é uma decisão tomada pelo neurônio sobre o que fazer com o valor resultante do somatório das entradas ponderadas.

Existem vários tipos de função de ativação, sendo que as mais utilizadas são as funções: linear, tangente hiperbólica e sigmóide, como visto na figura 3.6.

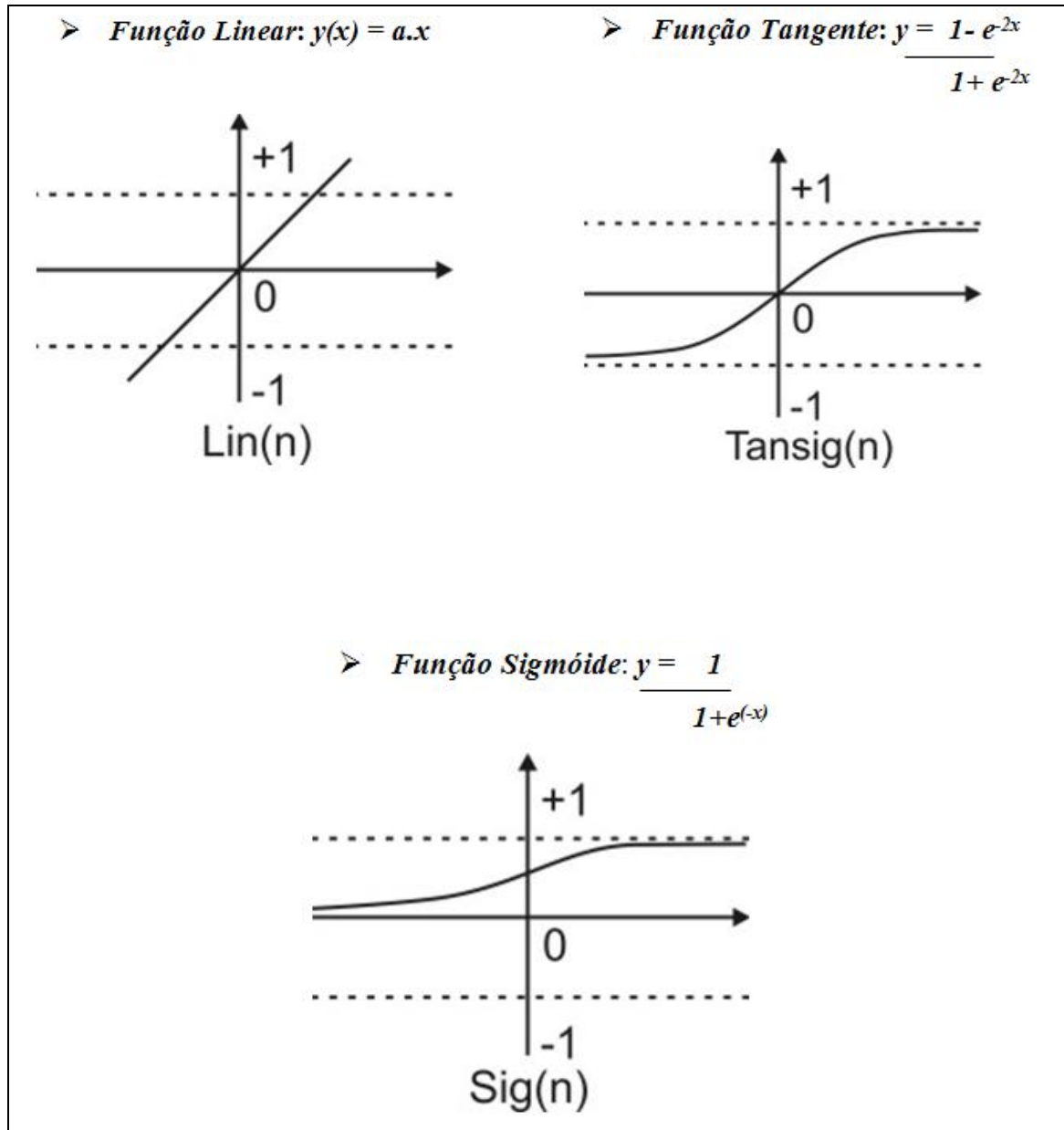


Figura 3.6 – Funções de Ativação do Neurônio Artificial.

3.5. ARQUITETURA DE REDES NEURAIIS

Deve-se ter uma grande preocupação no momento de se definir a arquitetura da RNA, visto que, “é a arquitetura que irá restringir o tipo de problema que a rede será apta a trabalhar” (Ferneda, 2006). A arquitetura da rede é determinada pelo número de camadas, pela topologia e pela conectividade.

De acordo com o número de camadas, segundo Ferneda (2006), as redes podem ser classificadas como sendo de *camada única* (Figura 3.7), quando possuem apenas uma camada, além da de entrada e redes de *múltiplas camadas* (Figura 3.8), quando possuem mais de uma camada além da de entrada. As camadas intermediárias podem ser chamadas de camadas escondidas (*hidden*).

Quanto à topologia, são classificadas em *diretas* ou *cíclicas*, são as redes que não possuem realimentação, ou seja, os neurônios que recebem sinais de excitação são os neurônios da camada de entrada, estas redes são as mais utilizadas. As redes *recorrentes* ou *cíclicas* são aquelas que contém ao menos um ciclo de realimentação, onde as saídas da rede também contribuem para a entrada da rede.

De acordo com a conectividade, as redes podem ser totalmente conectadas, “quando cada neurônio de uma camada está totalmente conectado na camada posterior”. E parcialmente conectada, “quando algumas conexões entre neurônios estão faltando” (FERNEDA, 2006).

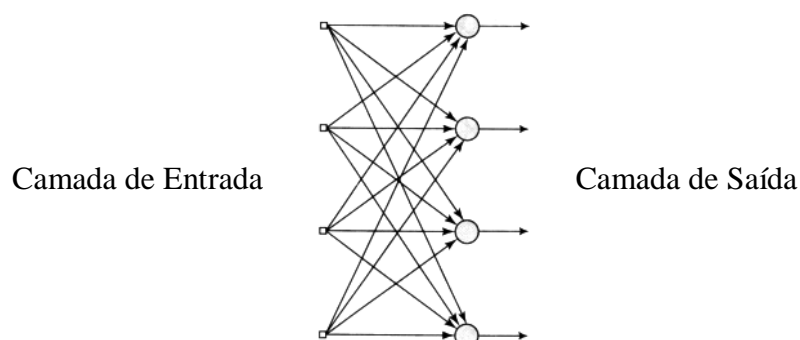


Figura 3.7 – Grafo mostrando um exemplo de uma rede de camada única, direta e totalmente conectada

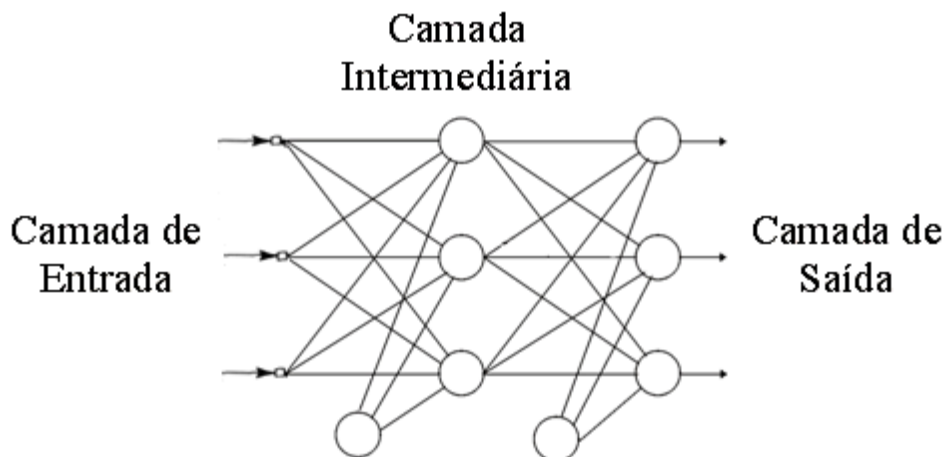


Figura 3.8 – Grafo mostrando um exemplo de uma rede de múltiplas camadas, direta e totalmente conectada

3.6. PROCESSO DE APRENDIZAGEM DAS RNA's

A principal propriedade das RNA é sua capacidade de aprender, a partir de um conjunto de exemplos fornecidos como padrões de treinamento para a rede. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma determinada classe de problemas. O processo de aprendizado acontece de duas formas: *o aprendizado supervisionado* e *o aprendizado não supervisionado*.

Segundo Tafner, Xerez e Rodrigues (1995), no aprendizado supervisionado a rede neural deverá possuir um conjunto de entradas e um conjunto com as saídas desejadas para cada entrada. Toda vez que for apresentado à rede o conjunto das entradas, deve ser verificado se a saída obtida confere com a saída desejada para àquela entrada.

Caso a resposta seja diferente, a rede deverá ajustar os pesos sinápticos de forma que armazene o conhecimento desejado. A iteratividade necessária para o treinamento da rede deverá ser repetida com todo o conjunto de treinamento, até que a taxa de acerto esteja dentro de uma faixa considerada satisfatória.

Outro tipo de aprendizado é o não-supervisionado, “para este treinamento não é necessário ter as saídas desejadas, pois serão utilizados apenas os valores de entrada. A rede trabalha essas entradas e se organiza de modo que acabe classificando-as, usando, para isso, os seus próprios critérios” (Tafner, Xerez e Rodrigues, 1995, p.65).

3.7. Algoritmo de Retropropagação do Erro (*Backpropagation*)

O algoritmo de *Backpropagation* ou *retropropagação do erro*, também é conhecido como *Regra Delta Generalizada*. Este algoritmo de aprendizado foi difundido por Rumelhart, Hinton e Williams (Rumelhart e McClelland, 1986) que faziam parte do grupo *PDP* ("*Parallel Distributed Processing*"), do MIT, na década de 80.

O algoritmo de *Backpropagation* é extremamente eficiente para o treinamento de redes de múltiplas camadas através da minimização do erro quadrático.

O algoritmo é utilizado em redes diretas com múltiplas camadas e com treinamento supervisionado. Estes tipos de redes também são conhecidos como Multi-Layer Perceptron (MLP) e diferente do *perceptron*, esta rede consegue classificar padrões não-linearmente separáveis, ele considera uma função de ativação não-linear, que serve para ampliar o potencial de classificação de um modelo, e as respostas assumem valores contínuos. Um exemplo de uma rede MLP com algoritmo de aprendizado *Backpropagation* encontra-se na figura 3.9.

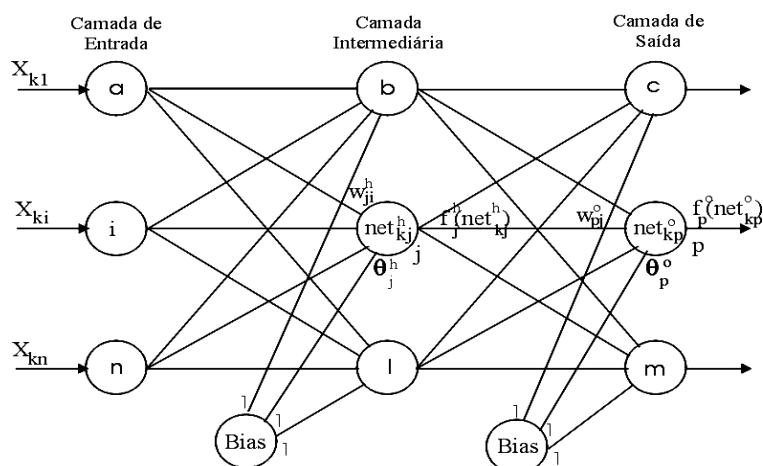


Figura 3.9 – Exemplo de uma rede Multi- Layer Perceptron (MLP) com algoritmo de aprendizado do tipo *Backpropagation*

O funcionamento do algoritmo de Backpropagation consiste no aprendizado da rede através de um conjunto pré-definido de pares formados pela entrada e saída desejada, para ajustar os pesos da rede e em ciclos de *propagação/adaptação*. O treinamento ocorre em duas fases: *forward* e *backward*.

A primeira fase (forward) o “vetor de treinamento é apresentado à rede e propagado por todos os nós até a camada de saída” (MOREIRA, 2004, slide 62). São comparados o valor de saída com o valor desejado, calculando-se o erro médio quadrático, o valor de ΔW é calculado e os pesos da camada de saída são atualizados. Nessa fase, são determinadas as saídas.

A segunda fase (backward) consiste na retropropagação, de forma que cada camada recalcula o valor de ΔW e atualiza seus pesos até a camada de entrada. Após atualizar todos os pesos da rede, um ciclo completo, o processo recomeça. Um ciclo completo é chamado de época e quando uma época é completada, a tendência é que o erro quadrático médio diminua, até alcançar um mínimo pré-estabelecido, momento no qual a rede converge. Nessa fase, as saídas, desejada e fornecida são comparadas e os pesos ajustados.

O algoritmo do *backpropagation* segue os seguintes passos:

- 1) **Aplica-se um vetor do conjunto de treinamento e este é propagado até a camada de saída**
 - um vetor é apresentado a rede;

- calcula-se, para cada neurônio da camada escondida, o net ;

$$net_{kj}^h = \sum_{j=1}^n w_{ji}^h \cdot x_{ki} + \theta_j^h$$

- a saída do neurônio da camada escondida será a entrada do neurônio da camada de saída, assim calcula-se o net e a saída da rede;

$$net_{kp}^o = \sum_{j=1}^l w_{pj}^o \cdot i_{kj} + \theta_p^o \quad o_{kp} = f_p^o (net_{kp}^o)$$

- 2) **Calcular o erro médio quadrático entre a saída obtida pela rede e a saída desejada no conjunto de treinamento**

$$E_k = \frac{1}{2} \sum_{p=1}^m (d_{kp} - o_{kp})^2$$

- 3) **Determinar o valor da mudança dos pesos da camada de saída**

- determina-se o ΔW e a direção da modificação dos pesos, será dada pelo gradiente de E_k com relação aos pesos W_{pj} .

$$\Delta_k W_{pj}^o = \eta (d_{kp} - o_{kp}) \cdot f_p^{o'} (net_{kj}^o) i_{kj}$$

Ressalta-se que o η é a taxa de aprendizado.

- 4) **Calcular os pesos da camada intermediária**

- calcula-se o gradiente do erro da camada de saída (E_k) referente aos pesos da camada intermediária (W_{ji}).

- 5) **Deve-se voltar ao passo 1 e escolher um novo vetor de entrada do conjunto de treinamento e repetir os passos de 1 a 5, somando o erro**

- 6) Após todos os vetores de treinamento terem sido apresentados, deve ser calculado o erro médio quadrático, se este for aceitável parar, senão, volta-se ao passo 1.**

O algoritmo de *backpropagation* propõe que o ajuste de pesos seja realizado através do método do gradiente descendente. Esse método garante que a rede caminhe na direção de redução de erros. Para problemas simples garante-se que esse método atingirá a solução com erro mínimo.

3.8. JUSTIFICATIVA E OBJETIVOS

A investigação da contribuição de diferentes grupos geográficos para a formação da sociedade brasileira vai muito além de contribuições meramente antropométricas, mas compreende a real identidade do povo brasileiro. Esta investigação resgata parte da história do Brasil, assim como, auxilia em um melhor entendimento da forma, tanto social, quanto biológica, de como esses povos contribuíram para a formação das populações.

Uma das formas de desenvolver estas investigações é fazer a análise do mtDNA. Conseqüentemente, hoje se tem produzido um grande volume de dados biológico referentes ao mtDNA, entre seqüências e SNPs, devido ao avanço de tecnologias na área da biologia molecular. Desta forma, desenvolver ferramentas computacionais que possam identificar padrões (classes) de linhagens de mtDNA, auxiliaria de forma direta na qualidade e velocidade dos estudos genéticos populacionais, uma vez que estas análises são praticadas até o momento manualmente, prejudicando a precisão ou mesmo causar classificações equivocadas.

Uma vez que a maioria das populações da região amazônica, assim como de todo território brasileiro, possuem constituição miscigenada do ponto de vista biológico, e as redes neurais são bastante eficientes como técnica para a classificação de padrões, o presente trabalho tem os seguintes objetivos:

- Desenvolver um sistema automatizado, preciso e acurado de classificação de seqüências (linhagens) de mtDNA por RNA's de acordo com pressupostos da filogenética;
- Desenvolver RNA's para classificar uma determinada linhagem de mtDNA em um dos três grupos filogeográficos principais que participaram da formação das populações brasileiras: ameríndio, europeu e africano;
- Desenvolver RNA's para um processo de classificação mais individualizado de uma linhagens de mtDNA nos principais haplogrupos de um determinado grupo filogeográfico.
- Desenvolver RNA's para uma classificação refinada de uma linhagem de mtDNA em um dos mais de 100 subclados propostos na literatura até o momento para os grupos filogeográficos ameríndio, europeu e africano.

CAPÍTULO IV

4. MODELAGEM DA REDE NEURAL

Para o desenvolvimento da rede neural foram analisados artigos científicos que propõe a classificação das linhagens mitocondriais das inúmeras populações dos diversos grupos geográficos humanos, uma vez que esta classificação é atualmente realizada pela análise e comparação destes artigos.

Para os haplogrupos Ameríndios foi utilizada a classificação proposta por Horai et al. (1993) e Torroni et al. (1994). Para os haplogrupos Africanos utilizou-se a classificação proposta nos artigos de Salas et al. (2002; 2004). Os artigos de Yao et al. (2002) e Tanaka et al. (2006) foram utilizados na classificação das linhagens asiáticas e finalmente, a classificação das linhagens européias foi baseada nos artigos de Richards et al. (2000) e Macaulay et al. (1999).

4.1. ARQUITETURA E TREINAMENTO DA REDE NEURAL

Em relação à criação das redes, foi realizado o mapeamento qualitativo e quantitativo das entradas a serem utilizadas para que possuam uma alta precisão e acurácia na classificação. Os dados de entrada referentes a região controle do mtDNA e SNPs variaram no intervalo de -1 a 1, como representado na Tabela 4.1.

Desta forma, foram criadas quatro redes neurais distintas, cada uma destinada à classificação de haplogrupos geográfico-específico: ameríndio, africano, asiático e europeu. Em todas as redes foi utilizada arquitetura direta, com múltiplas camadas e algoritmo de aprendizagem backpropagation. A função de ativação utilizada foi a tangente hiperbólica, pois deste modo, a resposta das redes pode variar, segundo os dados de entrada. Os parâmetros de treinamento empregados foram iguais para todas as redes:

- Taxa de aprendizagem: 0,1;
- Erro médio quadrático mínimo: 0,0001;
- Número de épocas: 300.000;

Para a realização dos treinos, testes e simulações foi utilizado o programa *Matlab* na versão 6.5 e utilizou-se amostras de mtDNA cadastradas e disponibilizadas no LGHMitdatabase (banco de dados nacional de seqüências de mtDNA, do Laboratório de Genética Humana e Médica da Universidade Federal do Pará, elaborado pela equipe de bioinformática do laboratório e disponibilizado no site: www.lghm.ufpa.br/bioinformatica).

As redes neurais desenvolvidas possuem os neurônios da camada de entrada representativos à variabilidade pertinente a classificação genética dos haplogrupos de ancestralidade grupo-específicas. A camada intermediária de cada rede neural apresentou sempre o mesmo número de neurônios da camada de entrada e a camada de saída possui o número de neurônios que indicam o número de haplogrupos/sub-clados, nos quais são classificadas as linhagens de cada grupo étnico. Ver tabela 4.1.

Tabela 4.1 - Descrição da estrutura utilizada na construção das RNAs

	Africanos	Ameríndios	Asiáticos	Europeus
Camada de entrada	88	16	113	66
Camada Intermediária	88	16	113	66
Camada de Saída	61	5	96	35

A parametrização das entradas da rede foi padronizada da seguinte forma: para referenciar a posição de um nucleotídeo variável utilizou-se os valores 1 ou -1; quando a informação biológica referente a uma determinada posição nucleotídica não pode ser obtida por ausência de testes utilizou-se o valor 0. A saída da rede foi definida com os valores 0 e 1, uma vez que 0 representa neurônios inativos, e 1 representa o neurônio ativo.

Em relação ao treinamento, cada classe foi exemplificada para a rede por meio do seu *motif*, logo o número de amostras utilizadas no treinamento de cada rede variou diretamente com a complexidade das classes que compõe cada rede.

Para buscar-se uma excelente precisão de classificação e testar a capacidade de generalização de uma RNA na classificação de seqüências de DNA, utilizou-se nos testes uma quantidade de amostras três vezes maior que o número usado no treinamento de cada rede.

4.2. RESULTADOS

Inicialmente, obteve-se resultados de somente três redes, sendo estas referentes a classificação de haplogrupos/subclados de ameríndios, africanos e europeus. Cada rede apresentou treinamento satisfatório em função do alcance do pequeno erro médio quadrático de 10^{-4} estipulado na construção de cada rede (Figuras 4.9 e 4.10).

Em relação aos testes, pôde-se observar o excelente desempenho das redes através dos gráficos que mostram os dados de saída obtidos antes do treino (em azul), dados de saída obtidos durante o treino (em verde) e os dados de saída obtidos após o treino da rede (em vermelho), para cada uma das redes neurais criadas (Figuras 4.5, 4.6, 4.7 e 4.8).

Como sistema de classificação de linhagens de mtDNA de ancestralidade ameríndia, africana e européia, as RNAs apresentaram resultados excelentes em relação a sua precisão e acurácia, pois obtiveram 100% de acerto tanto nos treinamentos, testes (ANEXOS 1, 2 e 3) e validação.

A priori, o objetivo do presente trabalho consiste no desenvolvimento de arquiteturas que sejam funcionais para a classificação destas linhagens. Como o resultado observado foi muito satisfatório, este trabalho estendeu sua investigação para linhagens mitocôndrias de ancestralidade asiática, uma vez que as linhagens ameríndias descendem deste grupo e povos do continente asiático também migraram para Brasil e a região amazônica para o trabalho na lavoura, nos séculos XIX e XX (Callegari-Jacques e Salzano, 1999).

Desta forma, foi criada uma quarta rede neural para a classificação de haplogrupos/subclados de ancestralidade asiática. Como o esperado, a rede também obteve 100% de acerto nos treinamentos, testes (ANEXO4) e validação.

De forma sumarizada, a Tabela 4.1 exibe o numero de amostras utilizadas no treinamento e no teste das redes de Ameríndios, Africano, Asiático e Europeu.

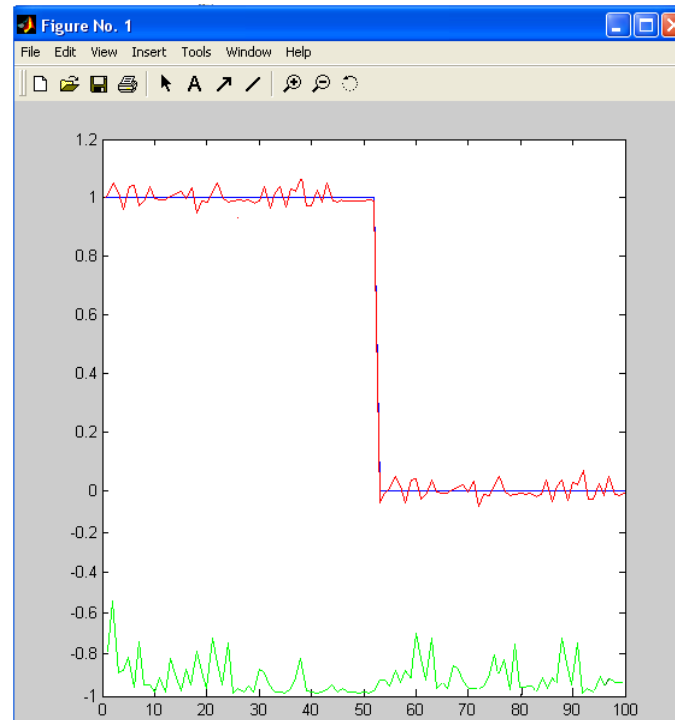


Figura 4.5 - Gráfico dos dados de saída de Ameríndio, antes do treino (cor azul), dados durante o treino (cor verde) e depois do treino, validação (cor vermelha).

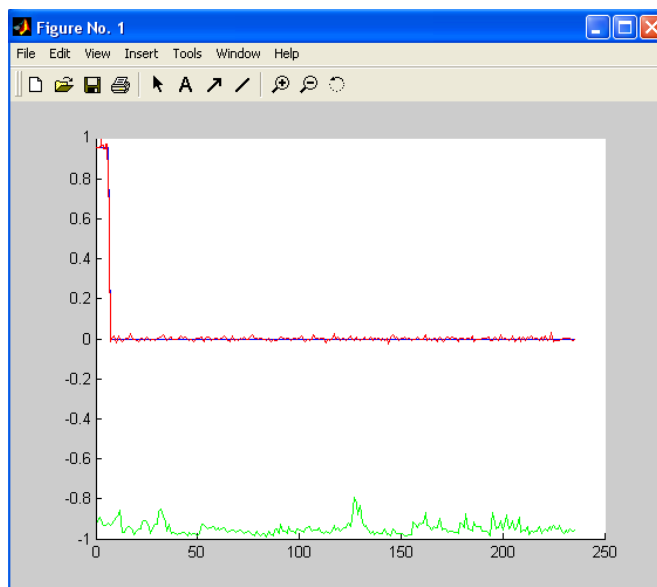


Figura 4.6 - Gráfico dos dados de saída de Africano, antes do treino (cor azul), dados durante o treino (cor verde) e depois do treino (cor vermelha).

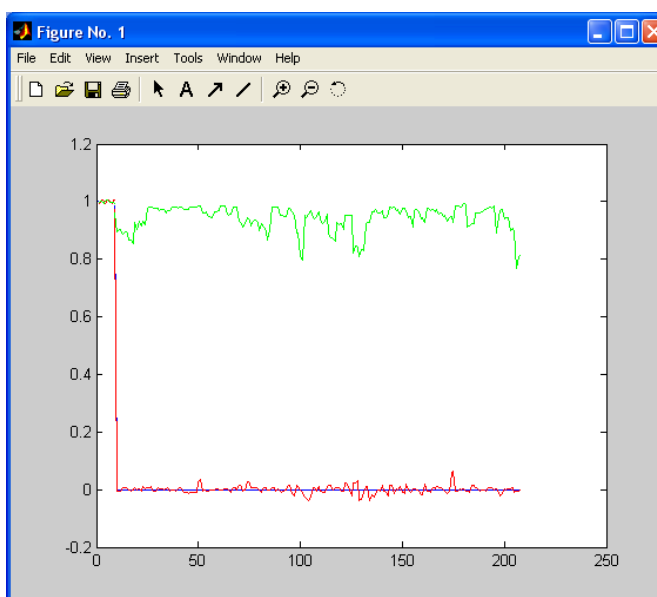


Figura 4.7 - Gráfico dos dados de saída de Europeu, antes do treino (cor azul), dados durante o treino (cor verde) e depois do treino (cor vermelha).

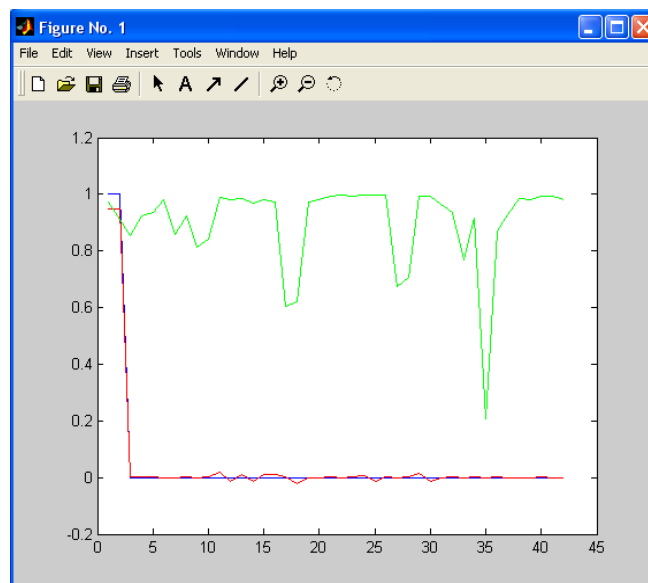


Figura 4.8 - Gráfico dos dados de saída de Asiático, antes do treino (cor azul), dados durante o treino (cor verde) e depois do treino (cor vermelha).

Abaixo encontram-se as figuras com os gráficos de treinamento das redes neurais criadas.

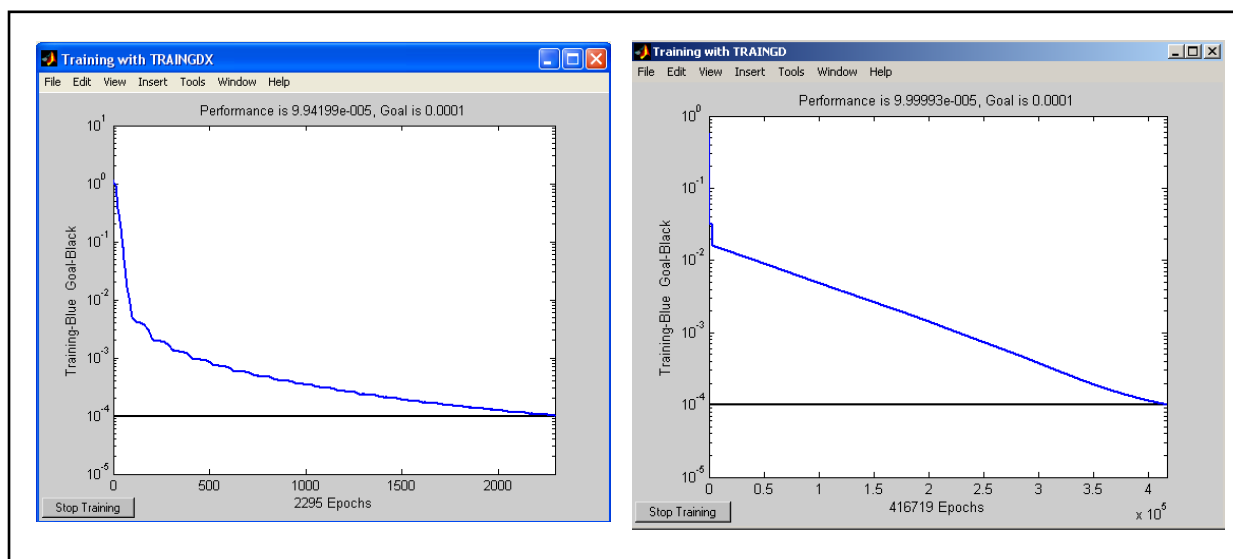


Figura 4.9 - Gráfico de treino das redes de Ameríndio e Africano, respectivamente.

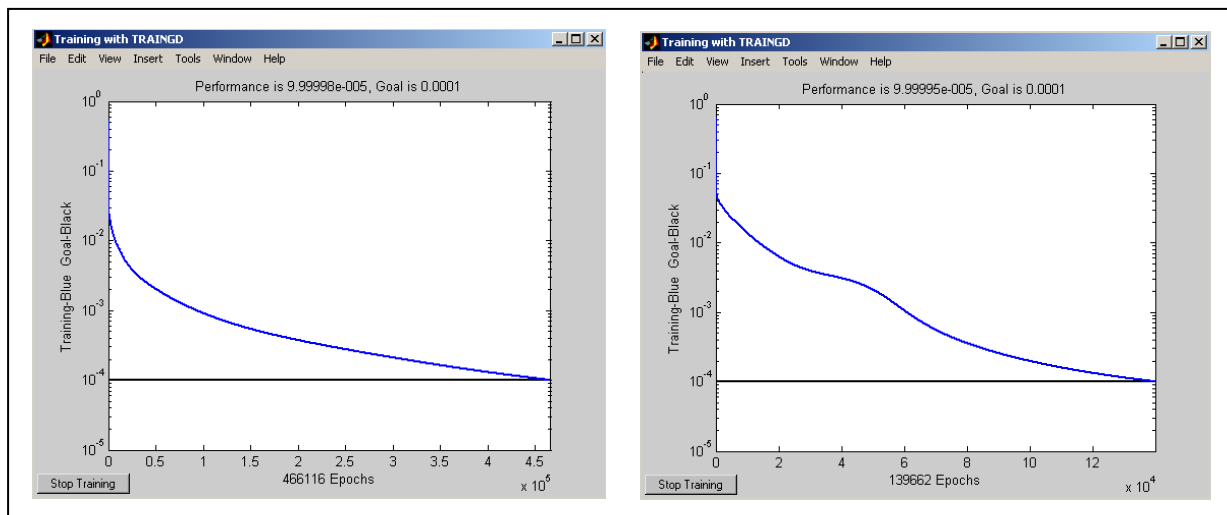


Figura 4.10 - Gráfico de treino das redes de Europeu e Asiático, respectivamente.

CAPÍTULO V

5. DISCUSSÃO

No desenvolvimento das RNA's, observou-se que alguns elementos, como os neurônios da camada de entrada e de saída, variaram entre as quatro diferentes redes. Essas variações proporcionaram algumas observações relevantes que marcaram de forma singular o processo de transformação de informações biológicas, contidas no mtDNA em forma de seqüências nucleotídicas, até a produção de um dado genético informativo, por meio da utilização de uma RNA.

5.1. MODELAGEM DAS CAMADAS DE ENTRADA E INTERMEDIÁRIA

Os trabalhos publicados na literatura que envolve reconhecimento de padrão em seqüências de DNA, por meio de RNA's, usam para a codificação da informação nucleotídica código binário (colocar referencia). Este código, como o Bin4 (colocar a referencia), gera para um nucleotídeo uma codificação de quatro caracteres entre 0 e 1 (p.e. A = 0 0 1 0). Entretanto esta codificação é direcionada para trabalhos com seqüência de nucleotídeos.

Entretanto este trabalho tem o objetivo de classificar linhagens de mtDNA em diferentes haplogrupos/subclados, trabalhando com toda a informação obtida da molécula de mtDNA, seja ela por meio de seqüenciamento genômico ou de trechos (Alça-D), ou mesmo da investigação pontual de alguns nucleotídeos (SNPs e RFLPs). Logo não se faz necessário codificar uma entrada com 16.569 neurônios e sim apenas com aquelas posições nucleotídicas variáveis dentro de cada grupo filogeográfico. Portanto, desenvolvemos uma codificação diferenciada (ver pág. 29). Isto diminuiu

sensivelmente o número de neurônios da camada de entrada de cada rede, uma vez que a maior apresenta apenas 113 (rede de asiático).

Segundo Vieira e Roisenberg, 2006, em relação à camada intermediária os testes realizados com a variação do seu tamanho, podem ser decisivos para que ocorra o aprendizado de uma rede ou fazê-lo com alto grau de pertinência. No presente trabalho cada rede foi treinada inicialmente com o mesmo número de neurônios da camada de entrada. Posteriormente foram feitos testes com 1,5 vezes e com o dobro de neurônios da camada de entrada, mantendo-se os demais parâmetros constantes. De posse dos resultados, não se observou variação significativa nos valores de pertinência, sendo então escolhida a primeira configuração visando empiricamente o menor processamento computacional. Portanto, para a classificação de linhagens de mtDNA, ou mesmo de qualquer outra seqüência de DNA, o número de neurônios da camada intermediária não interfere nos treinos e testes.

5.2. CONFIGURAÇÕES E EFICIÊNCIA DO APRENDIZADO

As configurações de cada rede variaram segundo a necessidade do número de neurônios da camada de entrada e de saída. Logo, quanto mais classes em um determinado grupo filogeográfico (Tabela 2.1), têm-se um maior número de amostras de treinamento (Tabela 4.1), assim como nucleotídeos variáveis característicos do grupo. Desta forma, as redes apresentam matrizes de entrada com tamanhos variáveis. Todavia, o tamanho da matriz de todas as redes não influenciou no treino e aprendizado das mesmas para qualquer nível de classificação exigido.

A limitação observada, relacionada ao tamanho da matriz de entrada, foi imposta pelo software utilizado, Matlab versão 6.5, visto que este possui uma limitação de tamanho de 65536 elementos, entre linhas e colunas, não suportando matrizes de tamanho superior, o que inviabilizou unir todas as redes em uma única.

5.3. CONSIDERAÇÕES FINAIS

Atualmente na região amazônica, assim como em todo o território brasileiro, tem-se a prevalência de haplogrupos/subclados mitocondriais de ancestralidade Ameríndia, africana e Européia. O desenvolvimento de RNAs como forma de automatizar e agilizar a classificação dessas linhagens é de grande importância para profissionais de diferentes áreas do conhecimento, visto que aumenta a precisão e acurácia desta classificação necessária principalmente para estudos bioantropológicos e médico-forenses.

Neste trabalho foram desenvolvidas quatro redes neurais, do tipo MLP, com algoritmo de aprendizagem do tipo Backpropagation Para realizar com precisão e acurácia a classificação automatizada de todos os 225 diferentes haplogrupos/subclados de linhagens mitocondriais descritas até o momento. Todas as redes possibilitaram a classificação de uma linhagem mitocondrial em três diferentes níveis filogenéticos: subclado, haplogrupo e grupo filogeográfico. Verificou-se ainda que as redes possuem a capacidade de associar e generalizar padrões, neste caso padrões que estão relacionados com eventos biológicos (mutações), mesmo com um número mínimo de amostras representativas de cada classe analisada. Desta forma, demonstra-se que Redes Neurais Artificiais podem ser utilizadas na classificação de padrões genéticos propostos na literatura para o mtDNA.

Vale ressaltar também, que o presente trabalho contribuiu e teve sua importância reconhecida pelas respectivas áreas por meio da publicação de artigo e resumo sendo eles:

- LARISSA LUZ-GOMES, ANA CECÍLIA FEIO-DOS-SANTOS, BRUNO M. CARVALHO, JESSÉ BARROS-LOBATO, JE-FERSON M. GOMES, ANA RÉGIA D. M. NEVES, ÂNDREA K. C. RIBEIRO-DOS-SANTOS, ROBERTO CÉLIO LIMÃO DE OLIVEIRA. Aplicação de Redes Neurais Artificiais na Classificação dos Haplogrupos Ameríndio e Africano a partir da Análise do DNA Mitocondrial. VIII CBRN – Congresso Brasileiro de Redes Neurais. Florianópolis – SC, 2007.

- LUZ-GOMES, L., FEIO-DOS-SANTOS, A.C., CARVALHO, B.M., BARROS-LOBATO, J., RIBEIRO-DOS-SANTOS, A.K.C., LIMÃO-DE-OLIVEIRA RC. Uso de Redes Neurais Artificiais no Reconhecimento de Perfil Genético Humano com base na Variabilidade Genômica do DNA Mitocondrial. I CBGF – Congresso Brasileiro de Genética Forense. Belém – PA, 2007.

5.4. TRABALHOS FUTUROS

Como perspectiva, este trabalho será aplicado no desenvolvimento de uma ferramenta computacional, denominada MITCLASS, a qual será implementada como um aplicativo web no site do LGHMitDatabase da UFPA para que o usuário possa fazer a classificação de forma automática e on-line de qualquer linhagem mitocondrial, visando facilitar a classificação das seqüências dos usuários e desta forma ampliar os conhecimentos biológicos por meio da análise do mtDNA.

Outro projeto será o desenvolvimento de uma rede neural artificial não supervisionada, a qual terá como objetivo principal reavaliar as mais de 250 classes, dentre haplogrupos e subclados, propostas pelas investigações filogeográficas.

REFERÊNCIAS BIBLIOGRÁFICAS

ÁLVAREZ, J.C., ENTRALA, C. LORENTE, J.A., FERNANDÉS-ROSADO, F.J., MARTÍNEZ-SPÍN, E., RODRÍGUES, E. V., E. **Análisis de ADN. Mitochondrial. Forénsica**, 1: 40-59, 2001.

ALVES-SILVA, J., SANTOS, M.S., GUMARÃES, P.E.M., FERREIRA A.C.S., BANDELTI, H.J. PENA, S.D.J., PRADO, V.F. **The Ancestry of Brazilian mtDNA Lineages**. American Journal of Human Genetics, 67: 444-461, 2000.

ANDERSON, S., BANKIER, A.T., BARRELL, B.G., DE BRUIJN, M.H.L., COULSON, A.R., DROUIN, J., EPERON, I.C., NIERLICH, D.P., ROSE, B.A., SANGER, F., SCHREIER, P.H., SMITH, A.J.H., STADEN, R., YOUNG, I.G. **Sequence and Organization of the Human Mitochondrial Genome**. Nature, 290: 457-467, 1981.

ANDREWS, R.M., KUBACKA, I., CHINNERY, P.F., LIGHTOWLER, R.N.TURNBULL, D.M., HOWELL, N. **Reanalysis and Revision of the Cambridge Reference Sequence for the Human Mitochondrial DNA**. Nature Genetics, 23: 147, 1999.

BARRETO, J. M. **Inteligência Artificial no limiar do século XXI**. 2ed. Florianópolis: Duplic, 1999. 296p.

BEHAR, D. M., ROSSET, S.; BLUE-SMITH, J.; BALANOVSKY, O.; TZUR, S.; COMAS, D.; MITCHELL, R. J.; QUINTANA-MURCI, L.; TYLER-SMITH, C.; WELLS, R. S. **The Genographic Consortium PLoS Genetics**. The Genographic Project Public Participation, Mitochondrial DNA Database. June 2007, Volume 3, Issue 6, e104.

BELEZA, S.; GUSMÃO, L; AMORIM, A.; CARRACEDO, A; SALAS, A. **The genetic legacy of western Bantu migrations**. Hum Genet (2005). Published online: 1 June 2005.

BROWN, M.D. HOSSEINI, S.H., TORRONI, A. BANDELT, H.J., ALLEN, J.C., SCHURR, T.G., SCOZZARI, R., CRUCIANI, F., WALLACE, D.C. **mtDNA haplogroup X: An Ancient Link Between Europe/Western Asia and North America?** American Journal of Human Genetics, 63: 1852-1861, 1998.

CALLEGARI-JACQUES, M.S. & SALZANO, F.M. **Brazilian Indian Non-Indian Interactions and their Effects**. Ciência & Cultura, 51: 166-174, 1999.

CANN, R.L., STONEKING, M., WILSON, A.C. **Mitochondrial DNA and Human Evolution**. Nature, 325: 31-36, 1987.

CAVALLI-SFORZA, L.L.; MENOZI, P.; PIAZZA, A. **The History and Geography of Human Genes**. Princeton, New Jersey, Princeton University Press, 1994. 518p

CAVALLI-SFORZA, L.L. (1997) **Genes, peoples, and languages**. Proceeding of the National Academy of Science USA, 94:7719 – 7724

CERNY V, SALAS A, HAJEK M, ZALOUDKOVA M, BRDICKA R. **A Bidirectional Corridor in the Sahel-Sudan Belt and the Distinctive Features of the Chad Basin Populations: A History Revealed by the Mitochondrial DNA Genome**. Ann Hum Genet. 2007 Jan 17.

CROTEAU, D.L., STIERUM, R.H., BOHR, V.A., **Mitochondrial DNA Repair Pathways**. Mutation Research, 434: 149-159, 1999.

CRUZ, E. **História do Pará. Belém**. Belém, Governo do Estado do Pará, 1973.

- FEIO-DOS-SANTOS, A.C., CARVALHO, B. M., SANTOS, S. E. B. e RIBEIRO-DOS-SANTOS, Â. K. C. **Nucleotide variability of HV-I in admixed population of the Brazilian Amazon Region.** Forensic Science International, Volume 164, Issues 2-3, 20 December 2006, Pages 276-277.
- FERNEDA, Edberto. **Redes neurais e sua aplicação em sistemas de recuperação de informação.** Ciência da Informação. Vol.35, Brasília, Jan./Abr. 2006.
- FERREIRA, Ronaldo Rodrigues. **Introdução a Bioinformática.** Centro de Biologia Genômica e Molecular da Universidade Federal do Rio Grande do Sul. Disponível em:<<http://www.inf.ufrgs.br/~rrferreira/bioinf/Apresentacoes/introducaoBioinf.pdf>> . Acesso em 10 de outubro de 2007.
- FRIEDLAENDER, J.; SCHURR, T.; GENTZ, F.; KOKI, G.; FRIEDLAENDER, F.; HORVAT, G.; BABB, P.; CERCHIO, S.; KAESTLE, F.; SCHANFIELD, M.; DEKA, R.; YANAGIHARA, R.; MERRIWETHERK, A. . **Expanding Southwest Pacific Mitochondrial Haplogroups P and Q.** Mol. Biol. Evol. Molecular biology and evolution 22(6):1506–1517. 2005.
- GILES, R.E., BLANC, H., CANN, H.M., WALLACE, D.C. **Inheritance of Human Mitochondrial DNA.** Proceeding of the National Academy of Science USA, 77: 6715-6719, 1980.
- HAYKIN, S, **Neural Network: A Comprehensive Foundation.** MacMillan College Publishing Co. New York , 1994.
- HAYKIN, S. **Redes Neurais - Princípios e Práticas.** 2ed. Tradução: Paulo Martins Engel. Porto Alegre: Bookman. 2001.
- KAESTLE, F. A. & SMITH, D.G. **Ancient Mitochondrial DNA Evidence for Prehistoric Population Movement: The Numic Expansion.** American Journal of Physical Anthropology, 115: 1-12, 2001.

KIVISILD, T., REIDLA, M.; METSPALU, E.; ROSA, A.; BREHM, A.; PENNARUN, E.; PARIK, J.; GEBERHIWOT, T.; USANGA, E.; VILLEMS, R. **Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears.** American Journal of Human Genetics, 75: 752 – 770. 2004.

KOVÁCS, Z. L. **Redes neurais artificiais: fundamentos e aplicações.** 2ed. rev. aum. São Paulo: Collegium Cognitio, 1996.

LABINFO - <http://www.labinfo.lncc.br/>. Acessado em: 20 de novembro de 2007.

MACA-MEYER, N., GONZÁLES, A. M., LARRUGA, J.M., FLORES, C., CABRERA, V.M. **Major Genomic Mitochondrial lineages delineate early human expansions.** BMC Genetics, 2:13, 2001.

MERRIWETHER DA, CLARK AG, BALLINGER SW, SCHURR TG, SOODYALL H, JENKINS T, SHERRY ST, WALLACE DC. **The structure of human mitochondrial DNA variation.** J Mol Evol. 1991 Dec;33(6):543-55. PMID: 1685753 [PubMed - indexed for MEDLINE].

PALHA, T. J. B. F., RODRIGUES, E. M. R. e SANTOS, S. E. B. **Y-chromosomal STR haplotypes in a population from the Amazon region.** Brazil. Forensic Science International, Volume 166, Issues 2-3, 2 March 2007, Pages 233-239.

PENA, S.D.J. **Retrato Molecular do Brasil, Versão 2001. In: Homo Brasilis: aspectos genéticos, lingüísticos e históricos e socioantropológicos da formação do povo brasileiro.** Pena, S.D.J. Ribeirão Preto, São Paulo. FUNPEC, 2002. p. 11 - 28.

PANTE-DE-SOUZA, G.; MOUSINHO-RIBEIRO, R.C.; SANTOS, E.J.M; ZAGO, M.A.; GUERREIRO, J.F. **Origin of the hemoglobin S gene in a northern Brazilian population: the combined effects of slave trade and internal migration.** Genetics and Molecular Biology, 21:427 – 430, 1998.

PROPUS, A. **Povoamento das Américas: Um debate sem fim.** Ciência Hoje, vol.25, nº149, maio 1999.

REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações.** Manoele, 2003. 525p.

RIBEIRO-DOS-SANTOS, A. K. C., CARVALHO, B. M., FEIO-DOS-SANTOS, A. C. e SANTOS, S. E. B. **Nucleotide variability of HV-I in Afro-descendants populations of the Brazilian Amazon Region.** Forensic Science International, Volume 167, Issue 1, 22 March 2007, Pages 77-80.

RODRIGUES, EMR; PALHA TJBF and SANTOS SEB. **Allele frequencies data and statistic parameters for 13 STR loci in a population of the Brazilian Amazon Region.** Forensic Science International, Volume 168, Issues 2-3, 24 May 2007, Pages 244-247.

SACCONE , C., DE GIORGI, C., GISSI, C., PESOLE, G., REYES, A. **evolutionary Genomics in Metazoa: the Mitochondrial DNA as a Model System.** Gene, 238: 195-209, 1999.

SALAS, A., RICHARDS, M., DE LA FE, T., LAREU, M.V., SOBRINO, B., SÁNCHEZ-DIZ, P., MACAULAY, V., CARRACEDO, A. **The making of the African mtDNA Landscape.** American Journal of Human Genetics, 71: 1082-1111, 2002.

SALAS, A., RICHARDS, M., LAREU, M.V., SCOZZARI, R., COPPA, A., TORRONI, A., MACAULAY, V., CARRACEDO, A. **The African Diaspora: Mitochondrial DNA and the Atlantic Slave Trade.** American Journal of Human Genetics, 74: 000-000, 2004.

SANTOS, S.E.B., RIBEIRO-DOS-SANTOS, A.K.C., MEYER, D., ZAGO, M.A.
Multiple Founder haplotypes of mitochondrial DNA in Amerindians revealed by RFLP and sequencing. *Annals of Human Genetics*, 60: 305-319, 1996.

SANTOS, S.E.B., RODRIGUES, J.D., RIBEIRO-DOS-SANTOS, A.K.C., ZAGO, A.
Differential Contribution of Indigenous Men and Women to the Formation of an Urban Population in the Amazon Region as Revealed by mtDNA and y-DNA. *American Journal Physical Anthropology*, 109: 175-180, 1999.

SMITH, D.G., MALHI, R.S., ESHLEMAM, J, LORENZ, J.G., KAESTLE, F.A.
Distribution of mtDNA Haplogroup X Among Native North Americans. *American Journal of Physical Anthropology*, 110: 271-284, 1999.

TAFNER, M. A.; XEREZ, M.; R. F., Ilson W. **Redes neurais artificiais: introdução e princípios de neurocomputação.** Blumenau: Eko, 1995. 199p.

TORRONI, A., SCHURR, T.G., CABELL, M.F., BROWN, W.M., NELL, J.V., LARSEN, M., SMITH, D.G., VULLO, C.M. & WALLACE, D.C. **Asian Affinities and Continental Radiation of the four Founding Native Amerindians mtDNAs.** *American Journal of Human Genetics*, 53: 563-590. 1993.

UNDERHILL, P.A.; SHEN, P.; LIN, A.A.; JIN, L. PASSARINO, G.; YANG, W.H.; KAUFFMAN, E.; BONNE-TAMIR, B.; BERTRANPETIT, J.; FRANCALACCI, P.; IBRAHIM, M.; JENKINS, T.; KIDD, J.R.; MEHDI, S.Q.; SEIELSTAD, M.T.; WELLS, R.S.; PIAZZA, A.; DAVIS, R.W.; FELDMAN, M.W.; CAVALLI-SFORZA, L.L.; OEFNER, P.J. **Y Chromosome sequence variation and the history of the human population.** *Nature Genetics*, 26: 358-361 (2000).

VIEIRA , Renato Corrêa; ROISENBERG, Mauro. **Redes Neurais Artificiais: Um Breve Tutorial.** Laboratório de Conexão e Ciências Cognitivas (L3C), Universidade Federal de Santa Catarina (UFSC) , Florianópolis. Disponível em: <<http://www.inf.ufsc.br/marcelot/academico/posgraduacao/rna/TutorialRNA.pdf> >. Acesso em: 20 de setembro 2006.

VIGILANT, L., STONEKING, M., HARPENDING, H., HAWKES, K., WILSON, A.C. **African Populations and the Evolution of Human Mitochondrial DNA.** *Science*, 253: 1503-1507, 1991.

WALLACE, D. C. **Mitochondrial DNA Variation in Human Evolution, Degenerative Disease, and Aging.** *American Journal of Human Genetics*, 57: 201-223, 1995.

ANEXO 1

Resultados selecionados dentre as amostras testadas que possuem ancestralidade Ameríndia.

Haplogrupo A

	1	2	3	4	5	6	7	8
A	0.98393	0.98429	0.98489	0.98393	0.98443	0.98227	0.98221	0.98399
B	-0.0007872	-0.0081816	0.009005	-0.0007872	-0.0017873	0.010066	-0.0051474	-0.0035212
C	-0.0086083	0.0052043	-0.00011555	-0.0086083	-0.0026456	-0.0081797	0.0039226	-0.016035
D	0.00017232	0.0093927	-0.018429	0.00017232	0.034099	-0.039323	-0.042871	0.028418

Haplogrupo B

	53	54	55	56	57	58	59	60
A	-0.0012163	-0.033182	-0.035476	0.071651	-0.0046798	-0.0049554	-0.003792	0.0033128
B	0.95895	0.9606	0.96244	0.94189	0.96088	0.96519	0.95335	0.96188
C	-0.054186	0.043315	0.04553	0.02635	-0.0012871	-0.05645	0.0060743	-0.0083481
D	0.066793	-0.047513	-0.030032	0.082738	-0.048294	-0.0092959	0.0062379	-0.017473

ANEXO 1

Resultados selecionados dentre as amostras testadas que possuem ancestralidade Ameríndia.

Haplogrupo C

	69	70	71	72	73	74	75	76
A	0.016049	-0.032057	0.006137	-0.013539	-0.0097164	-0.026155	0.0076255	0.016983
B	-0.02146	-0,33825	-0.025405	0.0029498	0.0054443	-0.00268	-0.0023081	-0.012199
C	0.96598	0.95698	0.96094	0.94974	0.97807	0.97592	0.97418	0.97243
D	-0.026949	-0.0047747	-0.025445	-0.019277	0.02324	-0.021948	0.022858	-0.059925

Haplogrupo D

	88	89	90	91	92	93	94	95
A	0.018654	-0.016782	-0.028742	0.0065931	0.028849	-0.019993	-0.011263	-0.0055086
B	0.005786	0.0087423	0.0076344	-0.012401	-0.00087055	-0.0052861	-0.0015733	-0.016685
C	-0.024467	-0.049962	-0.0060341	-0.0064226	0.23455	-0.035287	-0.051634	-0.0107
D	0.92974	0.93989	0.94887	0.93265	0.90236	0.94183	0.94307	0.9491

ANEXO 2

Resultados selecionados dentre as amostras testadas que possuem ancestralidade Africana.

	1	2	10	11	13	14	16	17
L0	0.97797	0.9856	-0.000549	-0.00053333	0.00080964	0.00091874	0.0011271	0.00070139
L0a	0.0087595	-0.0034374	0.92765	0.92751	0.001917	0.0022492	0.00053293	0.0015225
L0a1	0.0059185	-0.011289	-0.00054711	0.0017753	0.93945	0.9396	0.0024486	0.0031561
L0a1a	-0.0091107	0.014123	0.0013018	0.0016442	0.0040041	0.0032878	0.94922	0.94911
L0a2	0.022784	-0.0092582	0.00257	0.0021234	0.0017454	0.0024356	0.0013252	0.0015618
L0d	0.011452	0.0015329	-0.0013726	-0,067567	-0.00099563	-0.00020739	-0.00019894	0.00083847
L0d1	0.0027686	-0.0038618	0.0026997	0.0021897	-0.00064378	-0.00069887	-0.0011127	-0.00046086
L0d2	-0.0015326	0.0029339	0.001051	0.00040672	0.00054085	0.00072264	-0.00072728	-0.00076494
L0f	0.034191	-0.027537	0.00097747	0.0014537	-0.0015059	-0.0010938	0.00093639	0.0010373
L0k	-0.009916	0.016783	0.0020728	0.00019567	0.0019605	0.00094809	-0.000508	-0.0022133
L0k1	0.031092	-0.0088153	0.00018789	-0.00038847	0.00067002	0.00035592	0.00029271	-0.00033949
L1	0.054094	-0.038916	-0.0021731	-0.00036233	-0.00037493	0.00056716	-0.00085074	0.00082342
L1B	-0.015032	0.0099089	-0.00075641	-0.00018464	0.0004391	0.00083092	-0.0010873	-0.00017702
L1B1	0.029798	-0.022875	-0.00089774	-0.00018778	-0.0011099	-0.00056213	0.0018496	0.0027895
L1C	0.019672	-0.029143	0.0018581	0.00030216	0.00030508	-0.0008315	-0.0024441	-0.00070498
L1C1	-0.014612	0.003366	-0.0012337	0,46527	-0.0024191	-0.0014182	0.0016275	0.0022345
L1C2	0.023806	-0.013201	0.0017041	0.0016143	-0.0013595	-0.0012939	-0.0013552	-0.00036524
L1C3	-0.041285	0.049623	-0.00068162	-0.0010581	0.0018615	0.0015901	0.00015766	-0.0020716
L2a	0.020369	-0.012304	-0.0022574	0.00023505	-0.0036326	-0.0011995	-0.0026866	-0.00031619
L2a1	-0.0055251	0.00020888	-0.0016092	0.0008426	-0.0012364	0.00088866	0.00238	0.0029281

ANEXO 2

Resultados selecionados dentre as amostras testadas que possuem ancestralidade Africana.

	46	47	48	49	50	51	52	53
L0	-0.00042359	0.0002744	0.0041241	-0.0030663	-0.0015776	0.0035555	-0.0022567	-0.0019064
L0a	-0.0014858	0.0012258	-0.00043146	-0.002562	0.0016353	0,98327	0.00068189	0.0023922
L0a1	0,23775	0.00035764	-0.00016633	-0.0021565	-0.00037329	-0.0020918	-0.00061959	-0.0010763
L0a1a	0.0027105	0.0015859	0.00070736	0.0022712	-0.001775	-0.0017383	0.00033436	0.00012304
L0a2	0.0025033	0.00234	0,47405	0.00015519	0,08162	-0.0060738	-0.0006679	-0,86575
L0d	0,84629	0,0086242	-0.0053204	0.0039173	0.0027303	-0.00042372	0.0032214	0.0017992
L0d1	0.00075546	0.00054694	-0.0023766	0.0015169	0.0015214	-0.0022775	-0.00050426	0.00059289
L0d2	-0.00040736	-0.00067927	0.0049296	-0.0021759	-0.0023099	0.0013681	-0.002349	-0.0032584
L0f	-0.0043116	-0.0025135	0.0027635	-0.00262	-0.00042762	0.0023242	-0.0038162	-0.0021696
L0k	0.0025839	0.0013979	-0.0042073	0.0041972	0.0024849	-0.0026018	-0.00064563	-0.0008765
L0k1	0.00072115	0.0015042	-0.00065358	0.0010993	0.0014917	-0.0034899	-0.00047898	0.00017322
L1	-0.0023311	0.00066735	0.0061494	-0.0044621	-0.0024642	0.0018035	-0.0011886	-0,67821
L1B	0.9452	0.94534	0.94592	-0.00027436	0.003119	0.004116	0.0013648	0.0029282
L1B1	0.0064703	0.007457	0.011347	0.92645	0.92654	0.92751	-0.0039909	-0.0035854
L1C	0.0028557	0.00084588	0.0018344	-0.0019355	-0.0029312	0.00046548	0.91102	0.91075
L1C1	-0.00038211	-0.00010107	-0.0018803	0.00039744	0.00058137	-0.00045835	0.0016924	0.00068258
L1C1A	-0.0006421	0.00034414	-0,57771	0.00036111	0.0022737	-0.00041893	-0.00084633	0.0005711
L1C3	0.00095112	0,43539	-0.0014981	0.00051313	-0.0013661	-0.00087627	0.0049683	0.0042097
L2a	-0.0014105	0.0010239	0.0058501	-0.0055256	-0.0048068	0.0048604	-0.0011944	-0.00079307
L2a1	0.0008378	0.0018323	0.00055195	-0.00048251	0.0016111	-0.0048616	-0.00038144	-0.001133
L2a1a	0.00031867	-0.00023643	0.00095167	0,32349	-0.0024505	-0.0012265	0.00057564	0.00076555
L2a1b	-0.0013384	0.00023518	-0.0032376	-0.00042069	0.0013878	-0.001363	-0.0016003	-0.00048591

ANEXO 2

Resultados selecionados dentre as amostras testadas que possuem ancestralidade Africana.

	55	56	58	59	61	62	64	65
L1C1	0.93337	0.93312	0.0026071	0.0012512	0.0025941	0.00032278	0.0022812	0.0012976
L1C1A	0.00072876	0.0027173	0.94082	0.94095	0.0020553	0.0027831	0.0010504	0.00099364
L1C1A1	0.0015767	-0.00065876	0.0049645	0.0032559	0.96346	0.96337	0.0016606	-0.00063651
L1C2	0.0012439	0.0015038	0.00053622	0.0013061	0.00039124	0.0011908	0.95673	0.95687
L1C3	0.00093217	-0.00090146	0.0025155	0.00048767	-0.0019337	-0.0031962	0.0022643	0.0002806
L2a	0.0012539	0.00045475	0.0011089	0.00078761	-0.0042984	-0.0036076	0.0014559	0.0014231
L2a1	0.0026727	0.0025272	0.0017861	0.0013529	0.0025622	0.001984	-0.0029437	-0.00075245
L2a1a	-0.0009936	-0.00103	0.0007958	0,83893	0.0042717	0.0012775	-0.0011606	-0.0010242
L2a1b	-0.0012411	0.0011732	-0.00079548	0.0007242	-0.0036649	-0.0016396	0.0020215	0.0031532
L2a1c	-0.0009769	-0.00076766	-0.00046526	0,01595	0.0020949	0.0022654	0.0014222	0.00089584
L2a1d	-0.0010108	-0.00078661	0.0029294	0.0027142	-0.0034682	-0.0029947	-0.0022047	-0.0010643
L2b	0.00048497	0,62897	-0.00018701	-0.0003484	0.0056359	0.0041911	-0.00012383	0.00052465
L2b1	-0.0017573	-0.0020968	0.00097675	-0,082612	-0.0044021	-0.0059979	0.00039906	-0.0011939
L2c	-0.0013508	0.00032091	-0.0027392	-0.0011543	-0.0040582	-0.00061139	-0.0023975	-0.00062446
L2c1	0.0035905	0.0038902	-0.0012741	0.0001048	0.0021929	0.003936	0.001817	0.0025097
L2c2	-0.00027555	-0.00056675	-0.0017804	-0.0024745	0.0033066	0.001927	-0.0013133	-0.0026395
L2d2	-0.0015049	-0.00073632	0.00077448	0.0012726	-0.00097436	-0.0013023	-0.0025065	-0.0018137
L3b	0.00089984	0.0019422	-0.0027688	-0.0023927	0.00034324	0.0023752	-0.001177	0,50741
L3b1	0.0018936	-0.00055845	0.0036078	0.0016652	0.00069013	-0.001327	0.0025058	0.0011244
L3b2	0.00088983	0.0007003	0.00011788	0.00041298	0.0033209	0.0025483	0.00099683	-0.00071981
L3d	-0.00028313	0.0011984	0.0020653	0.0033236	-0.003137	-0.0012988	-0.0020828	-0.002108

ANEXO 3

Resultados selecionados dentre as amostras testadas que possuem ancestralidade Européia.

	1	2	7	8	13	14	37	38	52
PRE-HV	0.92669	0.95976	-0.015148	0.002441	-0.016175	-0.0014562	0.0071069	-0.0040382	0.0080672
HV1	-0.0063148	0.014783	0.97263	0.9698	0.0087743	-0.0019925	0.00063384	0.0044909	0.0067056
V	0.017687	0.0053703	-0.0051192	0.00014032	0.98036	0.97887	0.009543	0.010973	0.0037043
K	0.0066676	0.0048712	-0.0068048	0.018475	-0.0056756	-0.0076989	0.97341	0.97692	-0.00034272
U1	0.0045725	-0.0054443	-0.00082623	0.0044431	0.0038782	-0.0077033	0.0031986	0.011047	0.94199
U2	-0.0010874	-0.0030782	0.0065014	-0.0018458	-0.006647	0.01426	-0.003708	-0.0067395	-0.00092488
U3b	0.0070043	-0.011448	-0.0074765	0.019344	0.0014826	0.00054931	0.0010145	-0.00025943	-0.0041196
U4	-0.013886	0.0030804	-0.0029899	0.0098443	-0.0070486	-0.0082943	-0.0013609	-0.0026616	-0.0032942
U5	0.0033539	0.0028609	0.0034295	0.0085647	0.0051098	-0.0039252	0.011696	-0.0045862	-0.019425
U5a	0.016297	0.0072429	-0.0083392	0.015935	0.015176	-0.0057307	0.00095863	-0.0068375	-0.017735
U5a1	-0.0030166	-0.0036858	0.006157	-0.0085814	0.0011026	0.001861	-0.00012778	0.0057099	0.0074652
U5a1a	-0.013264	-0.00076581	-0.011389	0.0030483	0.0055396	-0.0080163	0.001835	0.0026406	-0.014298
U5b	0.0057197	0.0061484	-0.0024354	-0.00017203	-0.019625	0.017104	-0.00012865	0.0073231	0.011658
U7	0.0052821	0.011642	-0.00067552	0.012047	0.0027169	0.01139	-0.0073405	0.0096495	-0.0021266
J	0.0064415	-0.0014477	-0.00045126	-0.0057325	-0.00092118	0.0039853	0.014329	-0.00098299	0.0028076
J1	0.0073587	0.0020302	0.0063688	-0.0028934	0.00065604	-0.0029661	-0.0098851	0.0081011	0.0046563
J1a	0.0011563	0.0066006	0.0054094	0.00318	0.0070431	-0.00097296	0.003747	0.0046802	0.002825
J1b	0.0028959	0.014709	0.010197	-0.0052121	0.0028896	0.010444	0.0044861	0.011218	-0.0059508
J1b1	-0.0060123	-0.0075285	-0.014672	-0.00023365	-0.00528	-0.00039454	0.0013329	0.0032453	-0.0035855

	125	131	135	143	148	152	156	170	179
U7	0.95316	0.0044199	-0.0097745	0.002024	-0.001013	0.0068121	0.0070072	-0.0012164	0.00078121
J	0.014225	0.94751	-0.00065721	0.0074741	0.0066373	0.00056169	0.0056216	-0.0041792	-0.0010471
J1	-0.0070952	-0.0038669	0.96768	0.00091596	-0.01587	0.0036235	0.0026625	-0.014458	0.0026778
J1a	0.0034782	-0.023274	-0.003687	0.92767	0.0062007	-0.0021647	0.017835	-0.0070797	-0.0011074
J1b	-0.0034179	-0.012122	-0.024912	0.0095318	0.95615	0.0023805	0.0026291	-0.0063621	0.0037914
J1b1	-0.021782	-0.02176	-0.013099	-0.0013168	-0.010098	0.96823	-0.0015509	-0.010249	-0.000532
T	-0.0022246	0.0064606	0.011692	-0.00024911	-0.0015797	0.011526	0.97595	-0.019306	-0.00047385
T1	-0.030542	0.009195	0.0079861	-0.0092006	-0.015624	0.010236	-0.011617	0.98986	0.0023521
T2	0.027293	0.00473	-0.0063277	-0.0017209	-0.0004202	0.0053327	0.003335	-0.00037939	0.97616
T3	-0.011486	0.0075811	-0.011759	0.010674	0.0011751	-0.0021938	-0.016601	-0.010727	0.00265
T4	-0.012062	-0.001769	-0.0056208	-0.0010454	0.0031021	0.012671	0.015293	-0.0061202	0.010568
T5	-0.013711	0.00081926	0.0022861	-0.0092443	0.66081	-0.0044242	0.015965	-0.00673	0.0018104
I	0.01432	0.0093838	-0.013673	0.002855	-0.00094055	-0.0014353	0.0046658	-0.00018759	0.0069289
W	0.015088	-0.0010221	-0.0034405	-0.011753	-0.0039915	0.0076752	0.0047652	-0.00062384	-0.003044
X	-0.017137	0.0075517	0.011196	0.00099307	0.0006355	-0.013095	-0.00084	-0.00773	0.0011536

ANEXO 4

Resultados selecionados dentre as amostras testadas que possuem ancestralidade Asiática.

	3	6	13	17	23	27	28	29
D2	0.90228	-0.041331	-0.00046032	-0.012222	-0.007183	0.0017761	-0.0057806	-0.0029753
D4a	-0.019828	0.86511	-0.010834	0.022855	-0.0052564	-0.0028303	-0.0058469	0.016611
D4b1/D4c	-0.0090175	-0.013839	0.011148	-0.0039228	-0.020688	-0.0029024	-0.0095396	0.018713
D4d	0.011227	0.017954	-0.021977	0.011332	-0.019149	-0.0057192	-0.0019269	0.0008594
D4h	0.00075557	-0.0058415	0.94867	-0.0021906	-0.005238	-0.0048798	0.0001045	-0.0017342
D4i	0.014855	-0.039264	-0.023165	0.88231	0.0041536	-0.0072758	-0.0085509	0.0077412
D4j	0.017265	-0.050757	0.0056489	0.0065633	0.028897	0.0043125	0.0018561	-0.0063406
D4l	-0.01519	0.024248	0.00023607	0.0099512	0.93321	0.0022282	0.0069336	-0.0076849
D4n	0.021376	-0.012572	0.017944	-0.0015652	0.012157	0.92773	0.92718	0.9298
D5a1	-0.014206	0.081814	0.0021894	-0.0068288	0.023588	0.0094091	0.0058354	-0.0062769
D5a2/D5c2	-0.0060706	0.0042473	0.010226	-0.032808	-0.0085026	0.43629	-0.0043077	-0.00088547
D5b2	0.013071	0.017883	-0.018324	0.0034243	-0.011811	-0.0078308	-0.0034781	0.0072633
D5c1	0.0013009	0.066647	0.0071879	-0.0051054	0.013627	0.011623	-0.0033584	-0.014412
G1a2	0.004878	0.010438	0.020076	-0.01152	-0.0054411	0.012383	0.0019618	-0.012143
G2	-0.0066625	0.0060984	-0.0040477	-0.0025144	0.0065184	0.011339	-0.0028557	-0.0081223
G2a1	0.0068673	0.029877	0.010879	-0.0055251	-0.027186	0.0076066	0.0083322	-0.0067271
G2a1a	0.010432	0.047458	-0.02817	0.017684	0.0060303	-0.016739	-0.0076732	0.020495
G2a1b	-0.014966	-0.041049	0.019342	-0.0086383	0.010567	0.018235	0.0091695	-0.019805
M7a2	0.027399	-0.034127	-0.032676	0.0086909	-0.0029091	-0.020479	-0.0078832	0.0086324
M7b	0.0026123	-0.010734	0.0074501	0.004986	-0.0077522	0.00025633	0.0038696	0.0008227
M7c	0.014609	0.0013867	-0.0010957	-0.0093753	-0.014951	-0.019283	-0.0105	0.023645

ANEXO 4

Resultados selecionados dentre as amostras testadas que possuem ancestralidade Asiática.

	81	85	91	99	105	114	117	121
M7b	0.93994	0.010238	-0.0068607	-0.0049965	-0.0023662	-0.0067403	-0.0070454	0.01441
M7c	-0.00095523	0.94972	0.00073879	-0.011177	0.0080427	-0.0037729	-0.012982	0.0076071
M8a	-0.010607	-0.011583	0.82658	-0.0073175	-0.012436	0.0026412	0.013471	0.0064357
M8a1	0.00025441	-0.05128	0.028071	0.0049573	-0.016584	0.0040018	-0.025485	-0.0081032
M8a2	0.00090821	-0.0076937	0.030665	0.90661	-0.0066333	-0.010404	-0.024543	0.0091101
M9a	-0.0032751	0.0039004	-0.0089865	0.0039998	0.87278	0.016832	0.012428	0.0056497
M9a1	0.0041842	0.015572	-0.0039781	-0.0080385	0.010084	0.0065361	-0.0069411	-0.0050355
M9a2a	0.0047932	0.019621	0.039545	-0.0059506	0.010862	-0.0057474	0.0038002	-0.0062666
M9a2b	-0.0067157	-0.0066285	-0.020009	-0.012494	0.0046712	0.95002	-0.002016	-0.012886
M10a	0.021069	0.015813	-0.0085881	-0.013843	0.0084904	-0.011968	0.95298	-0.00021619
M10b	0.0056439	0.019203	-0.01958	0.0154	-0.019864	0.013389	-0.0085682	0.92897

ANEXO 4

Resultados selecionados dentre as amostras testadas que possuem ancestralidade Asiática.

	133	136	143	146	151	153	161	169
Z/Z2	0.89512	0.0093528	-0.007672	-0.015209	0.013297	-0.0076347	-0.006559	-0.00014211
Z1	0.0018276	0.92269	-0.016355	-0.0066824	-0.016159	0.0054259	0.0089661	0.0081603
A1b	0.013019	0.023127	0.93687	-0.018717	0.048451	0.017923	-0.0021465	-0.017558
A1c	-0.01571	-0.021507	0.0061074	0.9438	0.035993	-0.0094761	-0.0043167	0.030631
A/A2	0.0093209	-0.012927	-0.010018	-0.01194	0.93157	-0.075409	0.0023654	-0.012834
N9a	-0.0041231	0.0056964	-0.016031	-0.0012234	-0.027534	0.67708	-0.023227	-0.018636
N9a1	-0.0031205	-0.0017431	-0.010901	-0.013188	0.044796	0.096934	0.018024	0.0068559
N9a2	0.00063518	-0.011265	0.0089932	-0.000311	0.01181	0.11236	0.0051955	-0.024281
N9a2(b)	0.0032927	-0.023262	0.0037076	0.020438	0.00092382	-0.046055	0.90084	-0.0093013
N9a2(-c)	0.0086564	0.0055032	0.024307	0.012766	-0.039149	-0.030805	0.011492	-0.011921
N9b(1a)	0.0039939	0.0015563	0.0087888	-0.001019	-0.015413	-0.0069315	-0.0039116	0.94375
Y	0.01417	0.0011716	0.001902	-0.0068658	-0.002024	0.029144	-0.0007739	0.011915
Y1	0.0016056	-0.0062992	0.011515	0.013327	-0.028003	-0.011037	-0.016259	0.0203
Y2	0.01366	-0.018395	0.0098636	-0.0027687	-0.0037756	-0.012248	-0.0011724	-0.027414