

DETECÇÃO DE GENES E PROTEINAS

Machado, André Candido; Andrade, Ândreas Adelino; Russo, Bruno Alves; Filho, Walmir

Marques (1); LUZ, Francisco Donizeti Vieira (2)

(1) Acadêmicos do Curso de Bacharelado em Ciência da Computação da UNIFENAS; (2)

Orientador.

Resumo

O objetivo principal desse projeto é desenvolver uma ferramenta que identifique uma possível anomalia em uma sequência genética específica, auxiliando o profissional que irá trabalhar com esse tipo de informação.

Em laboratório é obtido a sequência FASTA. A sequência FASTA é uma sequência de DNA simplificada. A análise é feita utilizando uma amostra de DNA previamente cadastrada na base de dados. Utilizando tal amostra como padrão, é feita uma comparação entre a amostra padrão e a sequência que se deseja identificar, e assim dizer se houve ou não alguma alteração. Depois da comparação feita, é devolvido um relatório dizendo se houve ou não alteração. Se a resposta for sim, é identificada a posição e a possível consequência de tal alteração no indivíduo.

Abstract

The main objective of this project is to develop a tool to identify a possible abnormality in a specific gene sequence, assisting the professional that will work with this type of information. In laboratory FASTA sequence is obtained. The FASTA sequence is a DNA sequence simplified. The analysis is done using a DNA sample previously registered in the database. Using this as a standard sample, a comparison is made between the standard sample and the sequence you want to identify, and so tell whether or not there was any change. After the comparison made is returned report saying whether or not change. If the answer is yes, the position is identified and the possible consequence of such change in the individual.

1 INTRODUÇÃO

No início do século vinte geneticistas e químicos ainda se questionavam sobre a natureza química do material genético. Das pesquisas desenvolvidas, surgiu a conclusão de que o DNA era a molécula que armazenava a informação genética. Posteriormente, com a descoberta do código genético e do fluxo da informação biológica dos ácidos nucleicos para as proteínas, tais polímeros passaram a constituir os principais objetos de estudo de uma nova ciência, a Biologia Molecular. Logo surgiram métodos de sequenciamento desses polímeros, principalmente do DNA, e desde então, mais de 18 bilhões dessas sequências foram produzidas e estão disponíveis nos bancos de dados públicos.

A proposta deste trabalho é desenvolver uma ferramenta computacional capaz de analisar sequências genéticas obtidas em laboratório e auxiliar o profissional responsável a identificar possíveis alterações nessas sequências. Para a realização dessa análise utiliza-se um formato mundial de sequência genética, chamado de Fasta. Essa sequência será obtida pela análise celular de um recém-nascido e rigorosamente comparada a um gene padrão. O resultado dessa comparação será analisado e, partir dele, será emitido um relatório contendo os resultados da análise.

1.1 Objetivo

O objetivo deste trabalho é desenvolver uma ferramenta computacional que seja capaz de analisar uma amostra genética específica, e a partir da obtenção de seu sequenciamento, compará-la com outra sequência de mesma natureza pré-definida em um banco de dados, para então identificar possíveis anomalias e que consequências futuras poderão ocorrer no organismo da pessoa se houveram divergências na sua amostra genética.

2 REFERENCIAL TEÓRICO

2.1 Definição SGBD

O Sistema de Gerenciamento de Banco de Dados (SGBD) é a parte mais importantes de um projeto. É o conjunto de informações e programas para acesso de dados.

Um Sistema de Gerenciamento de Banco de Dados (SGBD) é um conjunto de informações e programas inter-relacionados para acessar esses dados. A coleção de dados, normalmente chamada de banco de dados, contém informações relevantes a uma empresa. O principal objetivo de um SGBD é fornecer uma maneira de recuperar informações de banco de dados que seja tanto conveniente quanto eficiente. (SILBERSCHATZ; KORTH; SUDARSHAN, 2006, p. 1)

Segundo Silberschatz, Korth e Sudarshan (2006), os Sistemas de banco de dados estão estruturados para receber um volume muito extenso de informações. O gerenciamento de informações está associado às estruturas de armazenamento de dados. É preciso garantir a segurança dos dados armazenados e impedir tentativas de acesso não autorizado.

2.2 Componentes SGBD

O Sistema de Banco de Dados envolve quatro componentes principais: dados, hardware, software e usuários.

2.3 Definição de Genética

“Genética é o ramo da Biologia que estuda o mecanismo de transmissão dos caracteres de uma espécie passados de uma geração para outra; é a ciência da hereditariedade” (PAULINO, 2000, p. 292).

“As características de uma espécie são condicionados pelos genes, estruturas hereditárias presentes nos cromossomos do núcleo da célula” (PAULINO, 2000, p. 292).

De acordo com Paulino (2000), o gene é definido como um conjunto de DNA cromossômico capaz de determinar a síntese de uma proteína. O tipo de proteína a ser formada depende do código estabelecido pela sequência de bases que o gene (DNA) possui.

2.4 Código Genético

Segundo Kessler (2012), o código genético, que é a informação contida no DNA, está registrado na sequência de suas bases na cadeia (timina sempre ligada à adenina, e citosina sempre com guanina). A sequência indica outra sequência, a de aminoácidos (substâncias que constituem as proteínas). Para formar as proteínas, o DNA precisa de outro tipo de RNA, o RNA mensageiro, no processo chamado transcrição. O código genético, na forma de unidades conhecidas como genes, está no DNA, no núcleo das células. Já a fabricação de proteínas se localiza no citoplasma celular em estruturas específicas, os ribossomos, para onde se dirige o RNA mensageiro. Na transcrição, apenas os genes relacionados à proteína que se quer produzir são copiados na forma de RNA mensageiro.

Conforme Ribeiro (2012), o DNA necessita de certa coesão em sua linguagem gramatical nas bases que constituem os dois filamentos da molécula. Este é o código, com poucos elementos, mas com diversas combinações, o responsável por formar os genes dos cromossomos de todas as espécies viventes.

Um segmento menor do DNA contém informações para identificação de um gene, de acordo com a FIG. 1.

Quando os genes são ativados, inicia-se o mecanismo de transcrição (formação de moléculas de RNA: ribossômico, transportador e mensageiro), prosseguindo com o processo de tradução (síntese de proteínas), forma pela qual o DNA coordena todo o metabolismo celular. (RIBEIRO, 2012, <http://www.mundoeducacao.com.br/biologia/codigo-genetico.htm>)

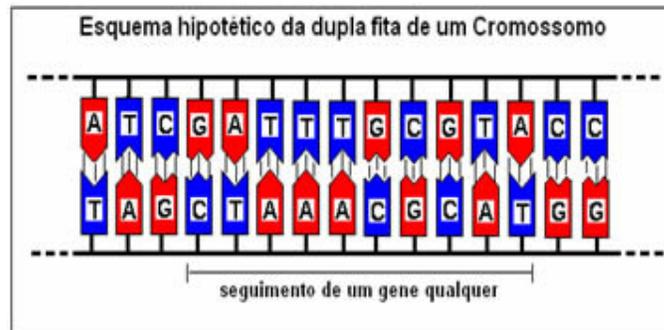


FIGURA 1 – Esquema hipotético da dupla fita de um Cromossomo.
 Fonte: <http://www.mundoeducacao.com.br/biologia/codigo-genetico.htm>

De acordo com Ribeiro (2012), o código biológico e os fatores ambientais, determinam as características de um indivíduo: a estatura, a cor da pele e dos olhos, o tipo de cabelo e até algumas doenças congênitas ou predisposições.

2.5 Bioinformática

Segundo Alves (2012), a bioinformática tem o objetivo de mapear as sequências genéticas. Após seu desenvolvimento, os biólogos moleculares passaram a utilizar métodos estatísticos capazes de analisar grandes quantidades de dados biológicos.

2.6 Projeto Genoma

O Projeto Genoma Humano surgiu em 1990, com a finalidade de identificar, no prazo de até o ano de 2005, cada um dos 100 mil genes através de um processo chamado mapeamento genético humano. Esse mapeamento consiste em registrar cada um dos genes do cromossomo, determinar a ordem dos nucleotídeos e sua função. O ponto positivo está no fato da identificação da cura e da causa de muitas doenças, o que será de grande benefício para a humanidade. (Algo Sobre, 2012)

O ponto negativo é o uso indevido do Projeto que pode fazer com que as pessoas percam sua individualidade tornem-se vulneráveis, devido ao fato de

que por um simples exame possa-se detectar uma má reprodução da célula e um futuro câncer, que dificultará sua admissão no emprego. (Algo Sobre, 2012)

Atualmente, já foram mapeados 97% do código genético humano. Os genes (pedaços de moléculas de DNA) são apenas rascunho de como se “fabrica” um ser vivo. Eles contêm a matéria e as proteínas, mas não todas as instruções de como monta-las de modo que o resultado final seja um bebe humano saudável. (Algo Sobre, 2012).

2.7 Bancos de dados Genômicos

Segundo Santos e Ortega (2012), devido a essa imensa quantidade de dados gerados em inúmeros laboratórios de todo o mundo, faz-se necessário organizá-los de maneira acessível, de modo a evitar redundância na pesquisa científica e possibilitar a análise por um maior número possível de cientistas.

“O NCBI, ou Centro Nacional para Informação Biotecnológica dos EUA, é considerado o banco de dados central sobre informações genômicas.” (SANTOS e ORTEGA, 2012, p. 5).

De acordo com Santos e Ortega (2012), no NCBI, o banco de dados RefSeq reúne a mais representativa sequência de um DNA transcrita, editada e inspecionada por um curador. É, frequentemente, o melhor banco de dados para se evitar a redundância natural num universo com tantas informações. A existência destes bancos de dados, ditos secundários, têm sido tão importantes quanto preservar os dados originais no GenBank.

2.8 O bioinformata

Segundo Tadeu (apud Meidanis, 2004), os bioinformatas ainda formam um pequeno grupo de pesquisadores, cientistas e profissionais de computação, cujo alvo de estudos vão de sequenciamentos genéticos a substâncias de plantas e moléculas para elaboração de novos remédios. O profissional da área tem de conhecer a teoria básica da biologia computacional e saber trabalhar

com ferramentas de bioinformática, para comparar sequências de genes e montar fragmentos de DNA.

2.9 Bioinformática no Brasil

De acordo com Santos e Ortega (2012), no Brasil, o Laboratório de Bioinformática da Unicamp é pioneiro nesta área, desenvolvendo e aplicando várias ferramentas à pesquisa genômica. Este laboratório foi responsável pela montagem, no computador, do genoma do primeiro organismo sequenciado no País em 2000, a bactéria *Xyllela fastidiosa* (Simpson et al. 2000), causadora da doença do amarelinho-da-laranja.

2.10 Mutações do DNA

Segundo Regateiro (2007), as mutações definem-se como alterações permanentes provocadas na sequência de DNA. Diversos fatores estão ligados às mutações genéticas. “A probabilidade de ocorrência de uma mutação num gene depende de fatores intrínsecos (mutações espontâneas) e de fatores extrínsecos (mutações induzidas). As mutações induzidas ocorrem, normalmente, com uma frequência muito maior do que as mutações espontâneas.” (Regateiro; 2007, p. 47).

2.10.1 Mutações Pontuais

De acordo com Regateiro(2007), as mutações pontuais podem resultar da substituição de uma base por outra, ou da inserção ou deleção de uma única base.

As mutações pontuais que resultam da substituição de uma única base podem funcionar como mutações sinônimas, mutações “missence” ou mutações “nosense”. As mutações sinônimas, ou silenciosas, são responsáveis por cerca de 25% das mutações pontuais. São mutações pontuais que resultam da substituição de uma base nucleotídica, em posição que não altera a codificação para o aminoácido em questão, embora altere o codão. (Regateiro; 2007, p. 50)

2.11 Sequência FASTA

Em bioinformática, o formato FASTA é um formato baseado em texto para representar tanto sequências de nucleotídeos quanto sequências de peptídeos, no qual os nucleotídeos ou aminoácidos são representados usando códigos de uma única letra. O formato também permite sequências de nomes e comentários precedendo as sequências. O formato se origina do FASTA, pacote de software, mas agora se tornou um padrão na área de bioinformática. A simplicidade do formato FASTA torna mais fácil manipular e analisar sequências usando ferramentas de processamento de texto e linguagens de script como Python, Ruby, e Perl.

3. MATERIAL E MÉTODOS

3.1 Recursos

Humanos: Alunos autores do projeto, professor orientador e professor co-orientador.

Financeiros: Não será preciso o uso de recursos financeiros para a realização deste trabalho, pois, os alunos autores já possuem todo o material necessário.

3.2 Material

- Hardware:

- Processador Pentium Dual-Core 2.20GHz;
- Memória RAM de 3 GB DDR3;
- HD SATA de 320 GB;

- Softwares:

- Windows 7 Ultimate Copyright 2009 Microsoft Corporation
- Embarcadero RAD Studio XE2
- Firebird 2.0 Server Manager

Foi utilizado para a realização do projeto a plataforma Delphi na versão XE2 onde não deixa nada a desejar diante de outros compiladores de alto nível. Esse compilador dispõe de ferramentas tão boas ou melhores que outros compiladores extremamente populares.

Ele gera um executável (.exe) e não necessita instalar mas nenhum outro software para funcionar.

Além disso, a Embarcadero RAD Studio prove de outras estruturas como XML, HTML e C++ além de compatibilidade para desenvolver softwares para iOS e dispositivos móveis.

3.3 Métodos

Para a realização deste trabalho foi adotado o método exploratório de literaturas obtendo um resultado que foi discutido, seguido de uma conclusão.

O resultado será um software desenvolvido pelos autores desse trabalho sobre orientação do corpo docente desta universidade.

4. DESENVOLVIMENTO

4.1 Introdução

Basicamente o projeto abre um arquivo, armazenado em um banco de dados, que contém as informações de um paciente, inclusive seu código genético. Este código será analisado em comparação com outro código semelhante (a sequência de uma proteína ou gene, por exemplo) e identifica, a partir desta comparação, se o paciente está com determinada doença ou venha a desenvolvê-la, se esta não for tratada a partir do diagnóstico.

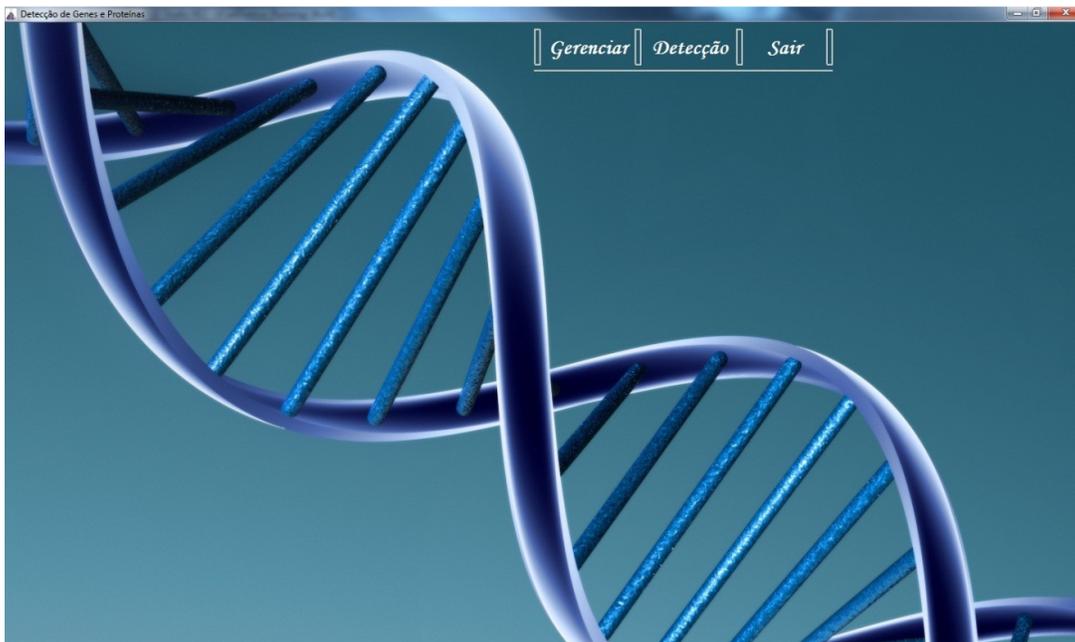


FIGURA 2 – Tela principal

4.2 Funções

Como na maioria dos softwares que requisitam consulta a um banco de dados, é necessária uma série de etapas para o funcionamento adequado de inserções, alterações, consultas e exclusões de registros para que todos os requisitos sejam satisfeitos e gerem um resultado adequado.

4.2.1 Conexão com o banco de dados

Primeiramente o software é conectado a um banco de dados, no caso usado o Firebird 2.0. No Delphi há um modelo de dados que funciona exclusivamente para armazenar os componentes de comunicação com o banco, chamado de Data Module. Dentro dele são colocados os componentes principais que serão usados: o TSQLConnetion , que faz a comunicação do Delphi com o Firebird, o TSQLQuery , que faz as rotinas de alteração do banco (inclusão, alteração e exclusão), o TSimpleDataSet , destinado as operações de consulta e o TDataSource . O TDataSource funciona da seguinte forma: o TSimpleDataSet liga no banco, o TDataSource liga no TSimpleDataSet e a TDBGrid (componente de visualização de dados em execução) é ligada no TDataSource onde são emitidas as informações na tela.

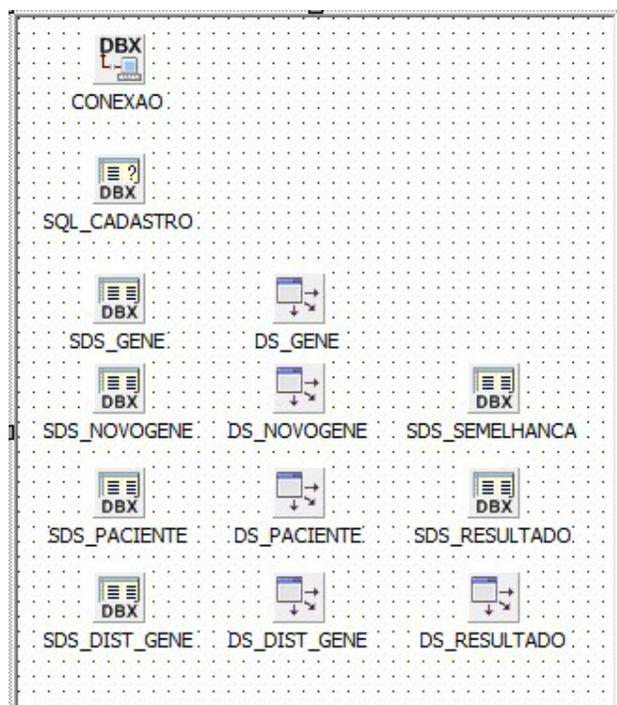


FIGURA 3 – Data Module

4.2.2 Cadastros e Consultas

São cadastrados primeiramente os genes com suas respectivas doenças e descrições.

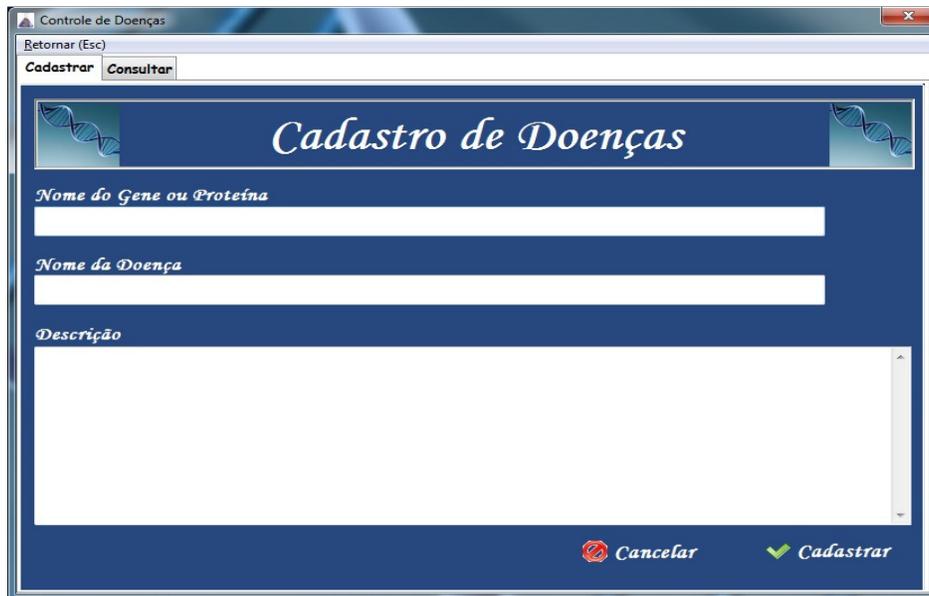


FIGURA 4 – Cadastro de doenças.

Para a realização de consultas das doenças já cadastradas, existe a aba Consultar.



FIGURA 5 – Consulta de doenças.

Em seguida são cadastradas as anomalias que podem ocorrer se houver alguma deformidade nos caracteres do código original do gene. Essas anomalias são retiradas da tabela universal do Blast (site que contém dados atualizados e informações sobre diferenças de genes).

Controle de Anomalia

Retornar (Esc)

Anomalias

Original Anomalia Posição

Doença

✗ Excluir
 ✓ Salvar

POSICAO	CORRETO	IRREGULAR	NOME DA DOENÇA
12 H	Q		DEJERINE-SOTTAS SYNDROME (DSS)
16 L	P		DEJERINE-SOTTAS SYNDROME (DSS)
16 L	P		CHARCOT-MARIE-TOOTH DISEASE TYPE 1A (CM
19 L	P		DEJERINE-SOTTAS SYNDROME (DSS)
22 S	F		CHARCOT-MARIE-TOOTH DISEASE TYPE 1A (CM
22 S	F		HEREDITARY NEUROPATHY WITH LIABILITY TO F
23 T	R		CHARCOT-MARIE-TOOTH DISEASE TYPE 1E (CM
28 W	R		CHARCOT-MARIE-TOOTH DISEASE TYPE 1E (CM

FIGURA 6 – Cadastro e consulta de anomalias.

Por fim é cadastrado o paciente, onde, juntamente das informações pessoais, será cadastrado os gene que ele possui (feito por exame laboratorial).

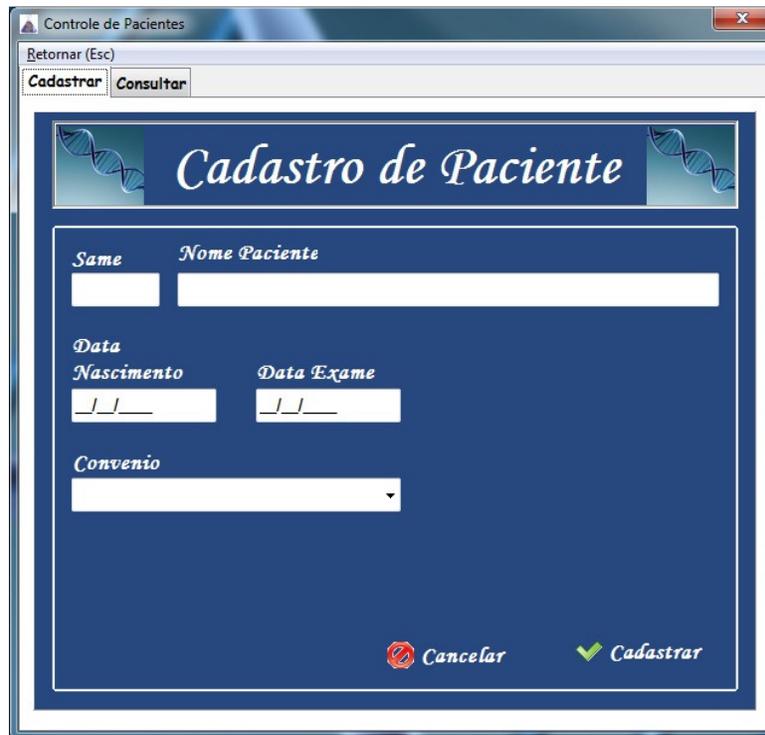


FIGURA 7 – Cadastro de pacientes.

Existe outra aba para as consultas dos pacientes cadastrados.

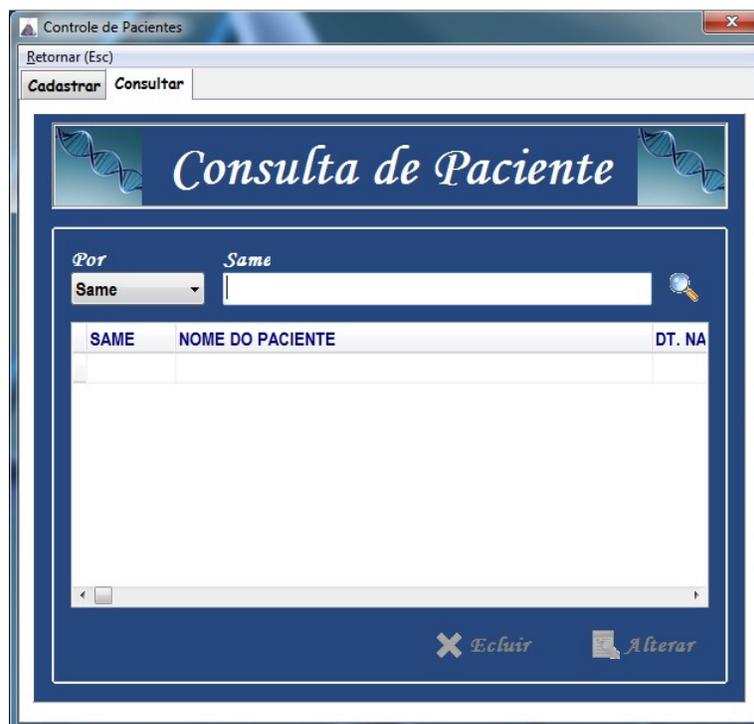


FIGURA 8 – Consulta de pacientes.

Há possibilidade de realizar consultas para encontrar nomes específicos dos genes ou proteínas e também caracteres adversos àquele gene ou proteína, ou ainda pesquisar os pacientes e suas doenças.

4.2.3 Identificação da doença

Hoje em dia, muitos laudos médicos são feitos a partir de tentativa e erro como na maioria das profissões. Na medicina não pode ser bem assim. Precisa-se de um diagnóstico muito mais concreto, pois se trata de vidas.

Na tela de detecção existe a possibilidade de identificação de doença a partir de um gene ou proteína.



FIGURA 9 – Tela de detecção.

Com a implantação deste software há a possibilidade de filtrar esse tipo de problema.

Primeiro seleciona o paciente, em seguida o nome do gene. Logo abaixo existem dois campos onde é colocada a sequência genética correta e no outro a sequência retirada do paciente.

Clicando no botão analisar, o software retorna do banco dados a doença daquele gene, naquela posição e a comparação dos caracteres.

The screenshot shows a software window titled "Detecção" with a menu bar containing "Retornar (Esc)" and "Imprimir". The main interface has a dark blue background with the word "Detecção" in a large, white, serif font at the top center. Below the title, there are two dropdown menus: "Paciente" with the value "ANDRÉ CANDIDO MACHADO" and "Gene ou Proteína" with the value "PMP22".

Below the dropdowns are two buttons: "Abrir código da sequência original" and "Abrir código da sequência em análise". Under each button is a text area containing the following amino acid sequence: MLLLLSIIVLHVAVLVLLFVSTIVSQWIVGNHGATDLWQNCSTSSSGNVHHC FSSSPNEWLQSVQATMILSIIFSILSLFFCQLFTLTGGGRFYITGIFQILAGL CVMSAAAIYTVRHPEWHLNSDYSYGFAYILAWVAFPLALLSGVIYVILRKRE.

At the bottom center is a button labeled "Analisar" with a small icon of a microscope. Below this button, the patient name "ANDRÉ CANDIDO MACHADO" and the gene "PMP22" are displayed. A large text area contains the following text:

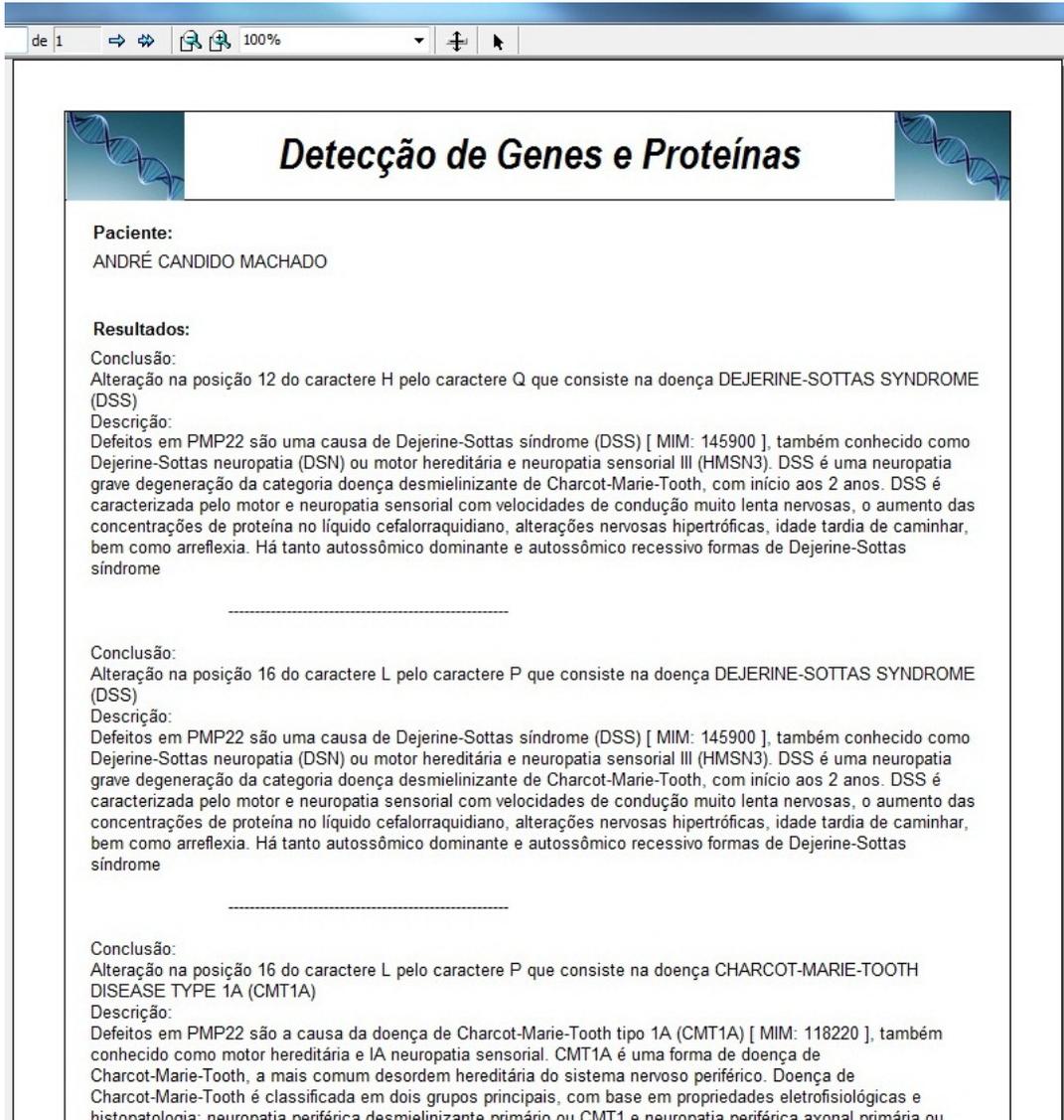
Conclusão:
Alteração na posição 12 do caractere H pelo caractere Q que consiste na doença DEJERINE-SOTTAS SYNDROME (DSS)

Descrição:
Defeitos em PMP22 são uma causa de Dejerine-Sottas síndrome (DSS) [MIM: 145900], também conhecido como Dejerine-Sottas neuropatia (DSN) ou motor hereditária e neuropatia sensorial III (HMSN3). DSS é uma neuropatia grave degeneração da categoria doença desmielinizante de Charcot-Marie-Tooth, com início aos 2 anos. DSS é caracterizada pelo motor e neuropatia sensorial com velocidades de condução muito lenta nervosas, o aumento das concentrações de proteína no líquido cefalorraquidiano, alterações nervosas hipertróficas, idade tardia de caminhar, bem como arreflexia. Há tanto autossômico dominante e autossômico recessivo formas|síndrome de Dejerine-Sottas

To the right of the text area is a large blue circle containing the text "100%" in yellow.

FIGURA 10 – Tela de detecção preenchida e retornando o resultado.

Sabendo qual o gene e quais suas diferenças, é gerado um relatório com o diagnóstico mostrando as comparações e suas respectivas doenças.



The image shows a web browser window with a report titled "Detecção de Genes e Proteínas". The report is structured as follows:

Paciente:
ANDRÉ CANDIDO MACHADO

Resultados:

Conclusão:
Alteração na posição 12 do caractere H pelo caractere Q que consiste na doença DEJERINE-SOTTAS SYNDROME (DSS)

Descrição:
Defeitos em PMP22 são uma causa de Dejerine-Sottas síndrome (DSS) [MIM: 145900], também conhecido como Dejerine-Sottas neuropatia (DSN) ou motor hereditária e neuropatia sensorial III (HMSN3). DSS é uma neuropatia grave degeneração da categoria doença desmielinizante de Charcot-Marie-Tooth, com início aos 2 anos. DSS é caracterizada pelo motor e neuropatia sensorial com velocidades de condução muito lenta nervosas, o aumento das concentrações de proteína no líquido cefalorraquidiano, alterações nervosas hipertróficas, idade tardia de caminhar, bem como arreflexia. Há tanto autossômico dominante e autossômico recessivo formas de Dejerine-Sottas síndrome

Conclusão:
Alteração na posição 16 do caractere L pelo caractere P que consiste na doença DEJERINE-SOTTAS SYNDROME (DSS)

Descrição:
Defeitos em PMP22 são uma causa de Dejerine-Sottas síndrome (DSS) [MIM: 145900], também conhecido como Dejerine-Sottas neuropatia (DSN) ou motor hereditária e neuropatia sensorial III (HMSN3). DSS é uma neuropatia grave degeneração da categoria doença desmielinizante de Charcot-Marie-Tooth, com início aos 2 anos. DSS é caracterizada pelo motor e neuropatia sensorial com velocidades de condução muito lenta nervosas, o aumento das concentrações de proteína no líquido cefalorraquidiano, alterações nervosas hipertróficas, idade tardia de caminhar, bem como arreflexia. Há tanto autossômico dominante e autossômico recessivo formas de Dejerine-Sottas síndrome

Conclusão:
Alteração na posição 16 do caractere L pelo caractere P que consiste na doença CHARCOT-MARIE-TOOTH DISEASE TYPE 1A (CMT1A)

Descrição:
Defeitos em PMP22 são a causa da doença de Charcot-Marie-Tooth tipo 1A (CMT1A) [MIM: 118220], também conhecido como motor hereditária e 1A neuropatia sensorial. CMT1A é uma forma de doença de Charcot-Marie-Tooth, a mais comum desordem hereditária do sistema nervoso periférico. Doença de Charcot-Marie-Tooth é classificada em dois grupos principais, com base em propriedades eletrofisiológicas e histopatologia: neuropatia periférica desmielinizante primário ou CMT1 e neuropatia periférica axonal primária ou

FIGURA 11 – Relatório.

Em um futuro não tão distante, isso ocorrerá em todos os centros médicos do mundo. O intuito é prever qual ou quais doenças uma pessoa pode vir a ter, possibilitando fazer um tratamento preventivo, afim de que a pessoa não a desenvolva.

“Dar nome a uma doença é apressar-lhe os avanços.” (Sthendal)

Na detecção de doenças, é comparado o gene ou proteína com código correto (de acordo com a tabela BLAST em formato FASTA) com outro gene ou proteína do paciente. Se nesta comparação houver alguma diferença, o software identifica o ponto exato onde foi encontrada e verifica no banco de dados a doença predominante dessa diferença de caracteres. Existem casos em que mais de uma diferença é encontrada.

5 RESULTADOS E DISCUSSÃO

O problema em questão era como as análises laboratoriais de sequências genéticas eram feitas. Profissionais ligados a essa área como biomédicos, bioquímicos, e os próprios bioinformatas, convivem com esse tipo de situação frequentemente.

Existem várias ferramentas que auxiliam nesse tipo trabalho, e o uso delas é imprescindível para agilizar os procedimentos de análise e comparação provenientes das sequências genéticas. O processo que é feito à mão demanda muito tempo e atenção, pois, como o ser humano é sujeito a falhas, muitas das vezes é necessário repetir todo o procedimento, a fim de se obter mais segurança nos resultados.

O grande problema é que muita das vezes, essas ferramentas não atendem cem por cento dos requisitos, pois, como existem vários tipos de necessidades em se tratando de manipulação de sequências genéticas, surgiu a ideia da construção de um software que fosse capaz de “absorver” gradativamente a essas demandas.

Partindo dessa ideia, desenvolvemos uma ferramenta que funciona de forma pontual, ou seja, ela contém em um banco de dados a definição de uma proteína importante no sistema nervoso do ser humano, que foi sugerida por nosso co-orientador. Ela serviu para demonstrar que o programa é capaz de analisar uma sequência genética ou proteica, por se tratar de um arquivo FASTA, e de maneira muito simples, gerar um relatório para o profissional que o estiver utilizando. Neste relatório é respondida a seguinte pergunta: Houve algum tipo de divergência da sequência analisada em comparação a sequência padrão do banco de dados? Se a resposta é sim e se ela for reconhecida pela biologia, também é gerado um resumo da mutação a partir de artigos científicos relacionados a ela, caso contrário, apenas indica-se um tipo de mutação não reconhecido. Também é possível identificar exatamente onde ocorreu a mutação dentro da sequência.

Este software poderá ser adequado às necessidades de um determinado laboratório, por exemplo. Fazendo-se um levantamento dos genes mais

trabalhados, é possível acrescentar à base de dados do programa as informações desses genes, e com posterior implementação, o software será capaz, também, de trabalhar com as novas sequências.

6 CONCLUSÃO

O software foi concluído com êxito. A análise de sequências genéticas é feita em um tempo muito menor daquele registrado quando a análise foi feita manualmente. O software retorna um relatório com a posição da alteração e se tal alteração é conhecida ou não, e a verificação é realizada através de uma comparação entre um padrão previamente cadastrado em sua base de dados, com a amostra que se deseja obter o resultado. Com isso é possível aumentar o número de mutações que o software identifica apenas cadastrando padrões para novas doenças, usando uma interface amigável e simples de utilização. Quanto mais doenças forem sendo cadastradas, mais eficientes serão os resultados obtido pelo software.

REFERÊNCIAS

SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S.. Sistema de Banco de Dados. Tradução da 5ª Edição. Editora Campus - 2006, 5v.

KESSLER, Cynara C.. DNA e RNA. Descrição física do meio eletrônico. Disponível em: < <http://www.algosobre.com.br/biologia/dna-e-rna.html>>. Acesso em: 17 maio 2012

SANTOS, Fabrício R.; ORTEGA, José Miguel. Bioinformática aplicada à Genômica. Descrição física do meio eletrônico. Disponível em: <<ftp://www.ufv.br/DBG/material%20curso%20bioinfo/Leitura%20Complementar/artigos/Bioinform%20E1tica%20aplicada%20%E0%20Gen%F4mica.pdf>>. Acesso em: 17 maio 2012