colloquium exactarum

IMPLANTAÇÃO DE METODOLOGIAS PARA ANÁLISE COMPARATIVA DE SEQUÊNCIAS GENÔMICAS RELACIONADAS AO CÂNCER DE PELE

IMPLEMENTATION OF METHODOLOGIES FOR COMPARATIVE ANALYSIS OF GENOMIC SEQUENCES RELATED TO SKIN CANCER

Antonio Alves dos Santos Neto[1], Antonio Fluminhan[2]

[1] Universidade do Oeste Paulista – UNOESTE, Mestrado em Meio Ambiente e Desenvolvimento Regional – MMADRE,Presidente Prudente - SP. E-mail: aasneto@hotmail.com

RESUMO - A exposição aos raios ultravioletas tem provocado o aumento da incidência do câncer de pele, geralmente associados a alterações na estrutura do DNA. Bancos de dados genômicos permitem o acesso público às sequencias de DNA que podem ser comparadas e analisadas, tais como o gene P53, que está relacionado ao tipo de câncer de pele com maiork
incidência. Esta pesquisa visa implantar metodologias de análise comparativa de sequências de nucleotídeos para a avaliação de impactos ambientais. Estão sendo realizadas investigações para averiguar as mais frequentes mutações no gene em bancos de dados de populações oriundas de países com diferenças contrastantes. A sequência original do gene será comparada com sequências mutantes descritas em pacientes com câncer de pele, através de uma ferramenta de bioinformática. Estão sendo realizados estudos comparativos, de modo a estabelecer correlações entre as mutações mais frequentes e fatores ambientais que possam estar relacionados com a origem das mutações.
Palavras-chave: bioinformática; mutações; banco de dados genômicos; câncer de pele; gene P53.

ABSTRACT - Exposure to ultraviolet rays has caused the increased incidence of skin cancer, usually associated with changes in the DNA structure. Genome databases allow public access to DNA sequences that can be compared and analyzed, such as the P53 gene, related to the type of skin cancer with the highest incidence. This research aims to establish methodologies for comparative analysis of nucleotide sequences for the assessment of environmental impacts. Investigations aiming to determine the most frequent mutations in the gene databases populations from countries with contrasting differences are being held. The original gene sequence is compared

Recebido em: 10/08/2014_
Aprovado em: 28/08/2014

with mutant sequences described in patients with skin cancer through a bioinformatics tool. Comparative studies will be conducted in order to establish correlations between the most frequent mutations and environmental factors that may be related to the origin of the mutations.

Keywords: bioinformatics; mutations; genomic databank; skin cancer; p53 gene.

1 INTRODUÇÃO

Segundo Santos (2007) as dramáticas mudanças ambientais observadas na atualidade são, na maioria das vezes, causadas por gerações e gerações de populações humanas que desconheciam o equilíbrio homem/ambiente e construíram um modelo de desenvolvimento predatório. O impacto da espécie humana sobre o meio ambiente tem sido comparado, por alguns cientistas, às grandes catástrofes do passado geológico da Terra. Em função do aumento da população surgiu a necessidade de se produzir cada vez mais, submetendo o meio ambiente a uma agressão que está provocando o declínio cada vez mais acelerado de sua qualidade e de sua capacidade para sustentar a vida. O uso de combustíveis fósseis tem aumentado a concentração de dióxido de carbono (CO_2) na atmosfera, dando lugar a um aumento da temperatura global na terra.

Em função do crescimento populacional, da industrialização acelerada e a escassez de alimentos o homem moderno criou substancias artificiais que tem contribuído com a diminuição da camada de ozônio aumentando a penetração da radiação altamente prejudicial para os seres vivos.

Segundo Dessler (2000) a camada de ozônio é uma região da atmosfera terrestre localizada em torno da terra entre 20 e 30 km de altura. É um gás composto de 3 átomos de oxigênio atômico. Esta camada tem capacidade de impedir que a radiação ultravioleta do tipo B, prejudicial à vida atinja a superfície da terra. A radiação solar é a energia que emana do sol e está distribuída em diversas ondas: desde o infravermelho até o ultravioleta (UV). Os raios UV são divididos em UV-C totalmente absorvido na atmosfera terrestre; o UV-A que não é absorvido pela atmosfera e o UV-B que é absorvido pela camada de ozônio.

A principal consequência da destruição da camada de ozônio tem sido o grande aumento da incidência do câncer de pele, já que os raios UV são mutagênicos (DESSLER, 2000). A exposição excessiva a radiação ultravioleta (UV), em especial ao ultravioleta tipo B (UVB), tem sido associada ao aumento do risco para o desenvolvimento dos canceres cutâneos, pois pode causar mutações genicas no ácido desoxirribonucleico (DNA) dos queratinócitos, sendo que a falha no reparo dessas alterações

genicas pode levar a crescimento celular desordenado e posteriormente a formação de tumor. Além disso, a radiação UV tem grande efeito sobre o sistema imune cutâneo, induzindo a um estado de imunossupressão local que impede a rejeição do tumor neoformado.

Algumas das alterações na estrutura do DNA têm sido atribuídas a exposição excessiva aos raios ultravioletas (UV), em especial ao ultravioleta tipo B (UVB), e como consequência direta o aumento da incidência de canceres cutâneos. (MARTINEZ, 2006). Geralmente quando ocorrem danos no DNA, desencadeiam-se nas células mecanismos bioquímicos capazes de reparar tais lesões. Para que este processo ocorra é necessário que a célula pare o ciclo celular a fim de não perpetuar a mutação. Essa função conhecida como: apoptose (morte celular programada) está diretamente atribuída a proteína TP53, normalmente referida como a "guardiã do genoma". Em geral esta proteína está relacionada ao tipo de câncer cutâneo menos evasivo mais com maior incidência conhecido como não melanoma.

Diante desta realidade tem surgido a necessidade de se criar metodologias para a avaliação das consequências dos impactos ambientais, decorrentes de atividades humanas ou provocados por fenômenos naturais, em especial ao nível biológico. Dados biológicos armazenados em banco de dados com acesso público, tais como sequências de DNA e de proteínas tem facilitado o trabalho de muitos cientistas, principalmente nos estudos de comparações de sequências de nucleotídeos para diversas finalidades, tais como: identificação de mutações ou polimorfismos genéticos, estudos evolutivos e filogenéticos. A disponibilidade dessas informações e a facilidade de acesso a internet, revolucionou a forma como as pesquisas, desta área, são realizadas na atualidade reduzindo significativamente o tempo destinado aos trabalhos.

Esta pesquisa tem por objetivo a implantação de metodologias no Laboratório de Citogenômica e Bioinformática da UNOESTE visando à análise comparativa de sequências de nucleotídeos em genes marcadores de impactos ambientais. Para tanto, estão sendo realizadas pesquisas com o objetivo de averiguar as mais frequentes mutações do gene P53 no banco de dados da Europa EBI, dos EUA NCBI e do Japão DDBJ que são os mais importantes bancos de dados e que estão disponíveis na internet com acesso público. Serão avaliadas as mutações gênicas mais comuns em populações humanas oriundas de países com diferenças contrastantes em relação aos fatores: grupo étnico, localização geográfica em respeito à latitude e valores de IDH, entre outros.

2 METODOLOGIA

O presente trabalho concentrou-se em realizar consultas nos três principais bancos de dados já mencionados, por meio da internet, localizados em portais de acesso público. As consultas se restringiram as variações do gene P53 encontradas em tecidos da epiderme retirados de pacientes diagnosticados com câncer de pele.

Para tanto, está sendo utilizado um notebook com processador intel Core I5, memória RAM de 4 GigaByte e com 500 GigaByte de capacidade de armazenamento interno. Este equipamento está conectado com a internet com velocidade superior a 1 MegaBit por segundo e utilizando o navegador de internet Google Chrome. Futuramente este notebook será substituído por um microcomputador que está em processo de compra e que será instalado definitivamente no Laboratório de Citogenômica e Bioinformática da Unoeste, com a finalidade de gerenciar, processar e armazenar as informações de materiais fornecidos por pesquisadores, e assim criar bases de dados próprios.

Os genes mutantes armazenados nos bancos de dados são representados pela sequencia de bases nitrogenadas identificadas pelas letras A (adenina), C (citosina), G (guanina), T (timina) e indexados por uma descrição que identifica o códon em que ocorreu a mutação e o tipo de alteração que pode ser a troca, inserção ou deleção da base.

Basicamente um códon é uma sequência de três bases nitrogenadas que codificam um determinado aminoácido ou que indicam o ponto de início ou fim de tradução da cadeia de RNAm. Isto significa que cada conjunto de três bases consecutivas é responsável pela codificação de um aminoácido.

3 COLETA E INTERPRETAÇÃO DOS DADOS GENÔMICOS

As informações dos genes mutantes foram copiadas dos bancos de dados e armazenados em um arquivo com formato texto e indexados pelo nome da mutação. A ferramenta Microsoft Word foi utilizada para estruturar as sequencias de cada gene mutante, afim de, facilitar posteriormente a localização e a recuperação dos respectivos dados de cada mutante. Cada gene é representado por quase trinta e três mil caracteres que estão dispostos sequencialmente e no formato maiúsculo.

No banco de dados do instituto americano NCBI foram localizadas, em julho de 2014, cento e noventa e uma mutação para gene P53, mas somente cinquenta e cinco possuíam a sequencia de aminoácidos disponíveis para serem copiadas. As comparações se concentraram apenas nos genes que possuíam a sequencia disponível.

Paralelamente ao processo de cópia das sequencias localizadas nos bancos de dados: americano, europeu e japonês está sendo feito estudos para escolha do software que será

utilizado como ferramenta de análise em busca de similaridade entre as sequencias mutantes e comparadas com a sequencia original. Estão sendo estudadas as ferramentas BLAST e CLUSTALW.

O software BLAST recebe este nome por ser uma Basic Local Alignment Search Tool (Ferramenta Básica de Busca de Alinhamentos Locais). Como o nome indica, trata-se de uma ferramenta de busca de similaridade entre sequencias biológicas (DNA ou aminoácidos). O Algoritmo foi desenvolvido na década de noventa nos Estados Unidos. Prioriza o alinhamento de locais específicos da sequencia, com respaldo estatístico, em lugar de realizar alinhamentos globais.

O CLUSTALW é um programa para alinhamento múltiplo de sequencias usado para a análise de proteínas e nucleotídeos. A ferramenta dispõe de um ambiente integrado para leitura de arquivos em vários formatos, alinhamento de sequencias, alinhamento via de profiles e analise de resultados.

Tanto BLAST como o CLUSTALW estão disponíveis gratuitamente na internet e estão habilitados para funcionarem no ambiente web, ou seja, não há necessidade de instalação do software em um computador, podem ser executados diretamente pelo navegador de web.

As sequências dos genes P53 serão comparadas, através de uma das ferramentas escolhidas, com sequências de variantes (mutantes) descritos em pacientes acometidos com câncer de pele, e armazenadas em bancos genômicos, de acesso público, localizados em diversos países, implicando em diferentes situações de impacto ambiental e com diferenças contrastantes em relação aos fatores: grupo étnico, localização geográfica em respeito à latitude e valores de IDH, entre outros. Serão realizados estudos comparativos de modo a estabelecer possíveis correlações entre as mutações mais freqüentes e a origem das referidas mutações.

A implantação destas metodologias viabilizará a realização de pesquisas com foco no estudo de exposição a agentes causadores de danos genéticos, de modo a disponibilizar técnicas mais avançadas para avaliação do grau de mutagenicidade apresentado por uma determinada ação antrópica ou natural.

4 RESULTADOS

No banco de dados do Instituto Americano NCBI foram encontradas em julho de 2014 cento e noventa e uma mutação para gene P53, dentre estas, foram copiadas cinquenta e cinco mutações que possuíam as sequencias de disponíveis. Com essas informações foi possível identificar que as mutações estão concentradas entre os códons trinta e um e um mil e cento e setenta e cinco. A figura 1 exibe a distribuição da frequência das mutações distribuídas nos códons.

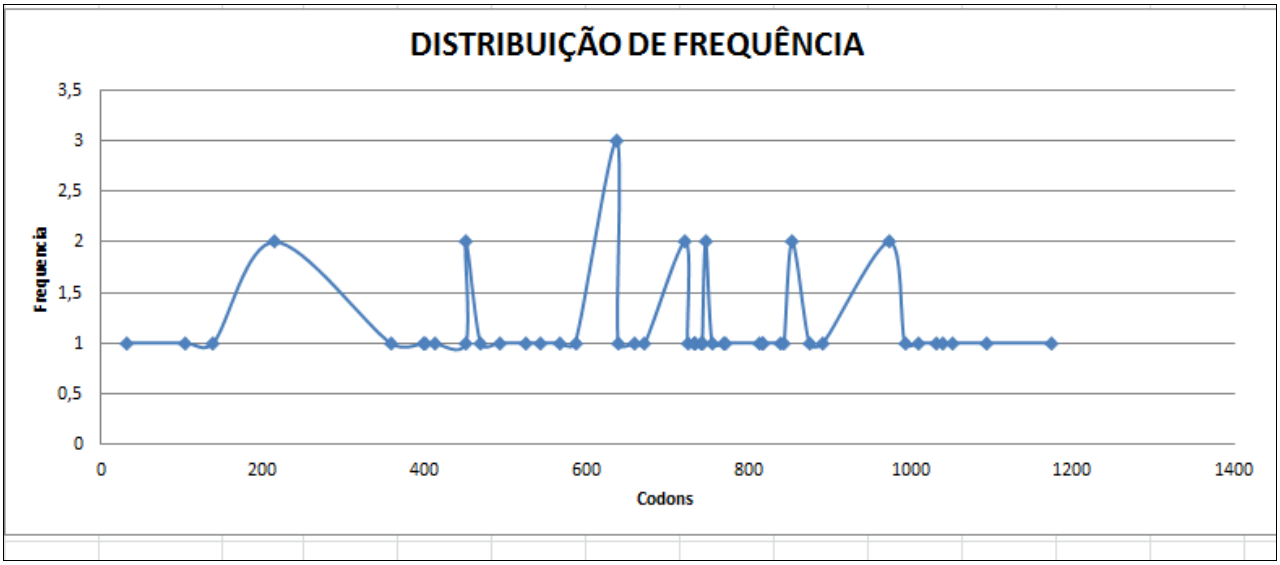


Figura 1. Distribuição das frequências de mutações encontradas em códons do gene p53, conforme dados disponibilizados pelo Gene Databank do NCBI (National Center for Biotechnological Information), EUA.

Na figura 2 são apresentadas as mutações encontradas no gene p53, com suas respectivas frequências, devidamente classificadas conforme padrão de ocorrência nas sequências depositadas no Gene Databank do NCBI (National Center for Biotechnological Information), EUA. Só foram consideradas sequencias com mutações que ocorreram substituições de aminoácidos, as mutações com inserção ou deleção não possuíam as sequencias de aminoácidos disponíveis.

MUTAÇÃO	FREQ
C > T	14
G > T	10
G > A	9
T > C	5
A > T	4
C > A	3
G > C	3
A > C	2
A > G	2
C > G	2
T > A	1
TOTAL	55

LEGENDA	
A	ADENINA
C	CITOSINA
G	GUANINA
T	TIMINA

Figura 1. Frequência das mutações mais comumente encontradas no gene p53, conforme dados disponibilizados pelo pelo Gene Databank do NCBI (National Center for Biotechnological Information), EUA.

Com as informações compiliadas do banco de dados do Instituto de pesquisas americano (NCBI), foi possível também identificar geograficamente (a) laboratórios onde os tecidos dos pacientes diagnosticados com câncer de pele foram coletados, tal como apresentado na figura 3. Neste momento da pesquisa, ainda não foram coletadas informações pessoais sobre os pacientes. Entretanto, já está disponível um considerável volume de informações sobre o sequenciamento do material coletado.



Figura 3. Mapa dos Estados Unidos mostrando a localização dos laboratórios onde foram estudadas as mutações para o gene p53, descritas no presente estudo. Nota-se uma relativa concentração de dados provenientes das regiões costeiras do Atlântico e do Pacífico, e praticamente inexistência de dados provenientes do Meio-Oeste e da região Sul dos EUA.

5 DISCUSSÃO

O uso da Informática nos campos de estudo da vida é essencial para a interação entre as ciências, propiciando assim um melhor desenvolvimento humano e tecnológico. Nos últimos anos, a Biologia vem se apropriando com avidez das ferramentas proporcionadas pela Informática. Assim, o computador tornou-se peça chave nas pesquisas que vem sendo realizadas em áreas especializadas como a Biologia Molecular e o armazenamento de dados biológicos em bancos de dados públicos tem se tornado cada vez mais comum. A “febre” genômica direcionada para a (identificação e o

mapeamento dos milhares de genes nas células de inúmeros organismos vivos), que se alastrou pelo mundo, em especial o projeto Genoma Humano, fez com que a Informática Biológica se desenvolvesse rapidamente (SABBATINI, 1993).

Dados biológicos armazenados em banco de dados com acesso público, tais como sequências de DNA e de proteínas tem facilitado o trabalho de muitos cientistas, principalmente nos estudos de comparações de sequências de nucleotídeos para diversas finalidades, tais como: identificação de mutações ou polimorfismos genéticos, estudos evolutivos e filogenéticos. A disponibilidade dessas informações e a facilidade de acesso a internet, revolucionou a forma como as pesquisas, desta área, são realizadas na atualidade reduzindo significativamente o tempo destinado aos trabalhos de pesquisa relacionadas a sequenciamento genético.

Os resultados permitem afirmar que há a viabilidade de se estabelecer uma estratégia de análise genômica através das ferramentas propostas. Estudos semelhantes serão realizados nos bancos de dados localizados na Europa e no Japão.

6 CONCLUSÃO

A internet possibilita que bancos de dados públicos de sequências de genoma ofereçam serviços por meio de uma interface uniforme para uma comunidade mundial de pesquisadores. Com um navegador, um biólogo molecular pode, então, comparar uma sequência de DNA desconhecida com a coleção completa de sequências de DNA públicas. Na sequencia, pode disponibilizar seus resultados por meio da rede e contribuir com demais pesquisadores, tornando o volume de informações ainda maior.

A implantação destas metodologias viabilizará a realização de pesquisas com foco no estudo de exposição a agentes causadores de danos genéticos, de modo a disponibilizar técnicas avançadas para avaliação do grau de mutagenicidade apresentado por uma determinada ação antrópica ou natural.

Na UNOESTE, alguns grupos já produzem, por iniciativas isoladas, trabalhos de iniciação científica e monografias relacionadas com a Bioinformática. Na Faculdade de Informática de Presidente Prudente – FIPP, já foram desenvolvidos trabalhos na área de bioinformática. Atualmente não há trabalhos em andamento nos cursos da FIPP voltados para a Bioinformática. Uma avaliação realista das potencialidades da Universidade poderá contribuir para o caráter multidisciplinar destas atividades, aumentando a relevância e a produtividade nas análises Bioinformáticas. Neste contexto, a Unoeste vem se preparando para uma inserção significativa na chamada era pós-genômica, de transformação da informação genômica em conhecimento científico e tecnológico.

REFERÊNCIAS

DESSLER, A. **The chemistry and physics of stratospheric ozone.** Londres: Academic Press, 2000.

MARTINEZ, M. A. R. **Genética molecular aplicada ao câncer cutâneo não melanoma.** São Paulo: USP, 2006.

SABBATINI, R. M. E. Informática e biologia: a nova aliança. In: SIMPÓSIO DE APLICAÇÕES DA INFORMÁTICA EM BIOLOGIA 1. **Anais...** Campinas: Unicamp, 1993.

SANTOS, E. T. A. **Educação ambiental na escola: conscientização da necessidade de proteção da camada de ozônio.** Santa Maria: UFSM, 2007.