



Emprego do *in silico* de genes de bactérias patogênicas para cultivos agrícolas

Vagne de Melo Oliveira¹

¹Médico Veterinário – UFRPE e Doutorando em Ciências Biológicas – UFPE. E-mail: vagne_melo@hotmail.com

Resumo: Bioinformática é uma prática recente que está melhorando a cada dia. As inovações tecnológicas estão permitindo a criação de software que auxiliam na compreensão da taxonomia clássica. A utilização de estudos *in silico*, sem interferência humana, para avaliar o grau de compatibilidade de diferentes genes e, assim, melhorar os métodos e técnicas empregadas nas diversas áreas de ciências agrícolas. Este estudo teve como objetivo utilizar ferramentas *online* de bioinformática para verificar compatibilidade entre genes que afetam cultivos agrícolas. Questões metodológicas incluíram o uso de ferramentas como o NCBI, CAP3, ORF, BLAST e PRIMER3. Os resultados mostraram compatibilidade com genes de Nitrosomonas, uma importante proteobactéria nitrificante para o ciclo do nitrogênio. Bioinformática análise *in silico* de genomas múltiplos torna-se uma abordagem rápida, prática e objetiva para a identificação de genes e definição de relações taxonômicas.

Palavras-chave: bioinformática, microbiologia, genes agrícolas.

1. INTRODUÇÃO

As Ciências Agrárias formam uma área multidisciplinar de estudos e pesquisas, envolvendo vários campos do conhecimento (Engenharias Agrícola, Ambiental, de Pesca e Florestal, Medicina Veterinária, Agronomia, Zootecnia, entre outros). É uma área que visa o aumento da produção agropecuária, o aperfeiçoamento técnico e melhorias no manejo e preservação dos recursos naturais e englobam, atualmente, alguns dos campos mais promissores em termos de pesquisas tecnológicas, dentre os quais a engenharia genética. Neste sentido, as ciências agrárias estão cada vez mais se inserindo e absorvendo as inovações tecnológicas existentes atualmente, tais como o emprego da bioinformática (Oliveira e França, 2010).

O uso da bioinformática trouxe uma mudança na maneira que as invenções biológicas podem ser protegidos por leis de patentes. Usando abordagens desenvolvidas nas áreas de ciência da computação, requerentes de patentes agora procuram proteger certos aspectos de suas invenções, que incluem software, métodos de fazer negócios e usos da informação bem como os mais tradicionais produtos e processos biotecnológicos (Maschio e Kowalski, 2001). Experimentos *in silico* são aqueles em que o comportamento dos indivíduos envolvidos é descrito com modelos computacionais, sendo composto exclusivamente por modelos numéricos onde nenhuma interação humana é permitida. Este trabalho objetivou avaliar através de software de bioinformática *online* genes de bactérias que afetam cultivos agrícolas.

2. MATERIAL E MÉTODOS

O estudo *in silico* é realizado por meio da utilização de ferramentas computacionais, disponíveis através de softwares ou mesmo *online*. Em nosso estudo, foram utilizadas ferramentas *online* (NCBI, CAP3, ORF, BLAST e PRIMER3) para comprovação da viabilidade deste tipo de artifício nos diversos campos das ciências agrárias, atreladas a biotecnologia. Para identificação dos nove genes *in silico*, descritos na tabela 1, foi realizado, inicialmente, uma busca no banco de dados do Genbank localizado no NCBI (National Center of Biotechnology Information) para identificação das sequências de genes disponíveis. Em seguida, foi realizada uma seleção de sequências codificadoras (CDS) pertencentes a essas famílias de genes. Nesta etapa foi possível analisar, comparativamente, a quantidade de dados disponível no banco de dados do NCBI referentes ao número de famílias de genes descritas e ao número de CDS disponíveis para cada gênero. Posteriormente, as sequências foram clusterizadas, utilizando-se o programa Sequence Assembly Program - CAP3 (Huang & Madan, 1999;



CAP3, 2011). Em seguida, foi feita a identificação das respectivas Open Reading Frame (ORFs) com o programa ORFinder (ORFINDER, 2011), com a identificação das regiões 3'-UTR e 5'-UTR de cada uma das sequências. As ORFs foram submetidas ao BLASTN para a identificação e seleção de sequências correlacionadas aos genes (BLAST, 2011). Nessa etapa, as ORFs foram comparadas a outras sequências depositadas no banco de dados do NCBI database. Finalizando, para desenho dos iniciadores foi utilizado o programa PRIMER3 (Rozen & Skaletsky, 2000).

3. RESULTADOS E DISCUSSÃO

O número de acesso ou *accession number* é o identificador do registro da sequência depositada no *GenBank*, que combina letras e números, e que pertence então à coleção de sequências do banco de dados. Normalmente, este identificador compreende a combinação de uma letra seguida de cinco dígitos ou duas letras e seis dígitos (Amaral et al., 2007). Ele representa o relatório completo da sequência e não somente a sequência em si. A tabela 1 lista nove números de acessos com suas respectivas sequências gênicas utilizadas neste trabalho, enquanto a figura 1 ilustra a localização das sequências selecionadas.

Tabela 1 – Sítios de inserção usados em bioinformática para Uncultured bacterium partial 16S rRNA gene

Número de acesso	Sequência	CAP3
>gi 343887224 emb FR875177.1	GGAGGAAAGTAGGGGATCGCAAGA C...	>Contig1 GGAGGAAA
>gi 343887223 emb FR875176.1	GGAGGAAAGTAGGGGATCGCAAGA C...	GCAGGGGA TCGCAAGA
>gi 343887222 emb FR875175.1	GGAGGAAAGTAGGGGATCGCAAGA C...	CCTTGCGC TAAAGGAG
>gi 343887221 emb FR875174.1	GGGAGAAAGTAGGGGATCGCAAGA C...	CGGCCGAT GTCTGATT
>gi 343887220 emb FR875173.1	GGGAGAAAGCAGGGGATCGCAAGA C...	AGTT AGTTGGTG
>gi 343887219 emb FR875172.1	GGAGAAAAGCAGGGGATCGCAAGA C...	GGGTAAAG GCCTACCA
>gi 343887218 emb FR875171.1	GGAGAAAAGCAGGGGATCGCAAGA C...	AGGCAACG ATCAGTAG
>gi 343887217 emb FR875170.1	GGAGGAAAGCAGGGGATCGCAAGA C...	CTGGTCTG AGAGGACG
>gi 343887216 emb FR875169.1	GGAGAAAAGCAGGGGATCGACAAA AC...	

View 1 GenBank Redraw 100 SixFrames Frame from to Length
-1 3..115 114

Length: 37 aa

Accept Alternative Initiation Codons

```

116 cgtcctctcagaccagctactgatcgttgcccttggttaggccttta
   R P L R P A T D R C L G R P L
71 ccccccaactaactaatcagacatcgccgctcctttagcgcaa
   P H Q L T N Q T S A A P L A Q
26 ggtcttgcgatccccctgcttct 3
   G L A I P C F P

```

Figura 1 – Localização da sequência >|cl|Sequence 1 ORF:3..116 Frame -1, com 37 aminoácidos.

A utilização das informações dos bancos de DNA e proteínas é de grande utilidade para a elucidação da função e expressão de novos genes. Os dados disponíveis podem ser utilizados de duas formas, direta - implica em identificar sequências similares, ou com propriedades comuns, e estabelecer relações entre elas (por exemplo, filogenia molecular); e indireta - é feita através da predição de regiões e sequências de genes relacionados, em espécies desconhecidas, através de alinhamento local (BLAST) (Malone et al., 2006). Os resultados desta pesquisa *in silico online* indicaram compatibilidade dos genes pesquisados com a nitrosomomas, uma proteobactéria nitrificante importante no ciclo de nitrogênio, transformando amoníaco em nitrito. Não houve outra detecção de compatibilidade, sobretudo com bactérias patogênicas, em relação aos gêneses selecionados.

4. CONCLUSÕES

De acordo com os dados obtidos, não houve relação dos genes selecionados com bactérias patogênicas. O que não significa que a ferramenta online é inválida. O trabalho é sugestivo para novas combinações de genes a fim de conhecer diferentes graus de compatibilidade.



5. REFERÊNCIAS

AMARAL, AM; REIS, MS; SILVA, FR. Programa **BLAST**: guia prático de utilização Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF: Embrapa Recursos Genéticos e Biotecnologia, 24 p., 2007.

BLAST - Assembled RefSeq Genomes. 2011. Disponível em <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. Acessado em 26 de Agosto 2011.

CAP3 - Sequence Assembly Program CAP3. Disponível em: <http://pbil.univ-lyon1.fr/cap3.php> Acesso em 2 Junho de 2011.

HUANG, X; MADAN, A. CAP3: A DNA Sequence Assembly Program. **Genome Research**, v. 9, p. 868–877, 1999.

MALONE, G; ZIMMER, PD; MENEGHELLO, GE; BINNECK, E; PESKE, ST. Prospecção de genes em bibliotecas de cDNA. **Revista Brasileira de Agrociência**, v.12, p.07-13, 2006.

MASCHIO, T; KOWALSKI, T. Bioinformatics – a patenting view. **Trends in Biotechnology**, v.19, p. 334-339, 2001.

OLIVEIRA, VM; FRANÇA, RC. 2010. Cursos de licenciatura em ciências agrárias/agrícolas: levantamento geográfico, áreas de atuação e perfil profissional. **Revista de Educação Agrícola Superior**, v.25, p.13-17, 2010.

ORFINDER, Open Reading Frame Finder. 2011. Disponível: <http://www.ncbi.nlm.nih.gov/projects/gorf/>. Acessado em 26 de Agosto 2011.

ROZEN, S; SKALETSKY, H. Primer3 on the WWW for general users and for biologist programmers. In: KRAWETZ, S.; MISENER, S. (eds). **Bioinformatics Methods and Protocols: Methods in Molecular Biology**. Humana Press, Totowa, p. 365-386, 2000.