



ESCOLA DE APERFEIÇOAMENTO DE OFICIAIS  
ESCOLA DE FORMAÇÃO COMPLEMENTAR DO EXÉRCITO



Cap QCO Info Pablo Alessandro Barbosa Viana

**COMPUTAÇÃO APLICADA À DEFESA QUÍMICA,  
BIOLÓGICA, RADIOLÓGICA E NUCLEAR: AVALIAÇÃO DE  
FERRAMENTAS COMPUTACIONAIS EMPREGADAS NA  
BIOPROSPECÇÃO**

**Manaus  
2019**

**Cap QCO Info PABLO ALESSANDRO BARBOSA VIANA**

**COMPUTAÇÃO APLICADA À DEFESA QUÍMICA, BIOLÓGICA, RADIOLÓGICA E  
NUCLEAR: AVALIAÇÃO DE FERRAMENTAS COMPUTACIONAIS  
EMPREGADAS NA BIOPROSPECÇÃO**

Trabalho de Conclusão do Curso apresentado à Escola de Formação Complementar do Exército / Escola de Aperfeiçoamento de Oficiais como requisito parcial para a obtenção do Grau de Especialização em Ciências Militares.

**Orientador: Maj Marcelo Antônio do Nascimento**

**Manaus  
2019**

Cap QCO Info PABLO ALESSANDRO BARBOSA VIANA

**COMPUTAÇÃO APLICADA À DEFESA QUÍMICA, BIOLÓGICA, RADIOLÓGICA E  
NUCLEAR: AVALIAÇÃO DE FERRAMENTAS COMPUTACIONAIS  
EMPREGADAS NA BIOPROSPECÇÃO**

Trabalho de Conclusão do Curso  
apresentado à Escola de Formação  
Complementar do Exército / Escola de  
Aperfeiçoamento de Oficiais como  
requisito parcial para a obtenção do Grau  
de Especialização em Ciências  
Militares.

Aprovado em \_\_\_/\_\_\_/\_\_\_\_\_

COMISSÃO DE AVALIAÇÃO

---

Maj QCO Info **Marcelo** Antônio do Nascimento

---

TC Com Rodrigo Lestinho **Ávila**

# COMPUTAÇÃO APLICADA À DEFESA QUÍMICA, BIOLÓGICA, RADIOLÓGICA E NUCLEAR: AVALIAÇÃO DE FERRAMENTAS COMPUTACIONAIS EMPREGADAS NA BIOPROSPECÇÃO

Pablo Alessandro Barbosa Viana

## RESUMO

A Defesa Química Biológica, Radiológica e Nuclear (DQBRN) vem assumindo importância cada vez maior no cenário mundial. Tornando-se um tema estratégico e de crescente repercussão em diversos setores do Estado e da sociedade. Este trabalho visa demonstrar como é possível utilizar a Bioprospecção e ferramentas da Bioinformática para obtenção de antígenos com potencial ação na DQBRN. Os processos discutidos aqui são comuns a diversos ramos da Biologia Molecular e Engenharia Genética e tem aplicações para as mais variadas áreas. É apresentado o estado da arte em ferramentas de Bioinformática em especial para realização do sequenciamento genético, anotação de genomas e determinação da estrutura das proteínas, porém, é importante salientar que este ramo está em franco desenvolvimento e novas ferramentas são criadas a cada dia.

**Palavras-chave:** DQBRN, defesa, guerra biológica, bioprospecção, bioinformática.

## ABSTRACT

Chemical, Biological, Radiological and Nuclear Defense (CBRND) is becoming increasingly important in the world scenario. It's a strategic theme of growing repercussion in many sectors of the state and society. This paper aims to demonstrate how to use Bioprospection and Bioinformatics tools to obtain antigens with potential action in CBRN. The processes discussed here are common to many branches of Molecular Biology and Genetic Engineering and have applications for varied areas. It's presented the state of the art in Bioinformatics tools, especially, for performing genetic sequencing, genome annotation and protein structure determination, however, it is important to note that these subjects are in full development and new tools are being created every day.

**Keywords:** CBRN, defense, biological warfare, bioprospection, bioinformatics.

## SUMÁRIO

1	INTRODUÇÃO.....	6
1.1	PROBLEMA.....	7
1.2	OBJETIVO.....	7
1.2.1	Objetivo Geral.....	7
1.2.2	Objetivos Específicos.....	7
1.3	QUESTÕES DE ESTUDO.....	8
1.4	JUSTIFICATIVA.....	8
2	REVISÃO DA LITERATURA.....	8
2.1	SEQUENCIAMENTO E MONTAGEM DE GENOMAS.....	9
2.2	ANOTAÇÃO GENÔMICA.....	10
2.3	DETERMINAÇÃO DA ESTRUTURA DE PROTEÍNAS.....	10
3	METODOLOGIA.....	11
4	RESULTADOS E DISCUSSÃO.....	12
5	CONCLUSÃO.....	19
	REFERÊNCIAS.....	20

# COMPUTAÇÃO APLICADA À DEFESA QUÍMICA, BIOLÓGICA, RADIOLÓGICA E NUCLEAR: AVALIAÇÃO DE FERRAMENTAS COMPUTACIONAIS EMPREGADAS NA BIOPROSPECÇÃO

## 1 INTRODUÇÃO

De acordo com Taylor (1992), “os agentes de guerra química e biológica constituem as classes de armas não convencionais de mais baixo custo, de mais difícil detecção e controle e, além das armas nucleares, são as únicas capazes de causar uma destruição de vidas sem precedentes na história da humanidade”. (Pinto, Lima e Franca, 2015)

Ao longo da história foram observados alguns eventos sejam acidentais ou propositais (utilizados como armas) em que os agentes QBRN causaram grandes catástrofes. Como a utilização de agentes químicos na Primeira Guerra Mundial (1914-1918), as bombas nucleares de Hiroshima e Nagasaki na Segunda Guerra Mundial (1945), o acidente nuclear de Chernobyl (1986), o incidente com Césio-137 em Goiânia (1987), o incidente com reatores nucleares em Fukushima (2011), além de diversos casos de bioterrorismo em especial utilizando antraz e ricina. (de Vasconcelos, 2018)

A Defesa Química Biológica, Radiológica e Nuclear (DQBRN) vem assumindo importância cada vez maior no cenário mundial. Tornando-se um tema estratégico e de crescente repercussão em diversos setores do Estado e da sociedade.

Visando modernizar-se nas novas vertentes da guerra, o Exército Brasileiro (EB) concebeu, em 2012, a reestruturação do Sistema de Defesa Química, Biológica, Radiológica e Nuclear do Exército (SisDQBRNEx). Com o objetivo de prever ações para adequação e modernização de meios, bem como a qualificação de recursos humanos, ampliando a capacidade do EB atuar com eficácia e eficiência em atividades de DQBRN, integrando-se com outros órgãos no contexto da Defesa Civil e atendendo a compromissos do Brasil com organismos internacionais. (de Vasconcelos, 2018)

A principal finalidade do SisDQBRNEx é dotar a Força Terrestre de um instrumento para responder a uma ameaça e/ou desastre QBRN. Sendo sua missão ser capaz de reagir a um incidente QBRN e estar habilitado a realizar o rápido reconhecimento e identificação de agentes, a demarcação e predição de áreas

contaminadas, a coleta de amostras, as medidas de descontaminação de emergência e a diminuição de danos. (de Vasconcelos, 2018)

## 1.1 PROBLEMA

Para executar estas atividades é necessário a utilização de instrumentos, equipamentos de proteção individual (EPIs) e antígenos químicos que podem ser utilizados para descontaminação ou desintoxicação. A criação desse ferramental requer estudos para compreensão das propriedades físicas e químicas dos agentes QBRN e de novas substâncias que possam ser utilizadas como sensores, bioacumuladores, biorremediadores ou antígenos, que possuam alguma atividade que possa ser utilizada para DQBRN.

## 1.2 OBJETIVO

A seguir serão apresentados os objetivos gerais e específicos deste estudo, estabelecendo a abordagem que será realizada neste trabalho para responder às questões de estudo.

### 1.2.1 Objetivo Geral

O presente estudo pretende analisar os principais problemas computacionais existentes nos processos de identificação de genes e as ferramentas computacionais utilizadas neste processo, visando a bioprospecção de proteínas com potencial uso na Defesa Química, Biológica, Radiológica e Nuclear (DQBRN).

### 1.2.2 Objetivos Específicos

Com a finalidade de delimitar e alcançar o desfecho esperado para o objetivo geral, levantou-se objetivos específicos que irão conduzir na consecução do objetivo deste estudo, os quais são transcritos abaixo:

1) Definir Bioprospecção e descrever como a Computação está envolvida neste processo.

2) Identificar os principais problemas da Bioinformática e as principais ferramentas computacionais utilizadas.

3) Discorrer sobre como a utilização dessas ferramentas está inserida no processo de Bioprospecção e apresentar uma possível utilização do presente estudo na DQBRN.

### 1.3 QUESTÕES DE ESTUDO

Neste contexto, podem ser formuladas as seguintes questões de estudo:

- 1) Como identificar agentes que possam ser utilizados como sensores, bioacumuladores, biorremediadores ou antígenos para DQBRN?
- 2) Quais ferramentas computacionais podem ser utilizadas nestes processos?
- 3) Como funcionam estas ferramentas e quais são os principais problemas resolvidos por elas?

### 1.4 JUSTIFICATIVA

O projeto se justifica pela necessidade de dotar a estrutura de DQBRN do Exército de meios que o capacitem a atuar no amplo espectro dos conflitos e de apoio a órgãos governamentais, fomentando pesquisa e desenvolvimento em DQBRN, de acordo com estudos prospectivos e as prioridades científico-tecnológicas decorrentes da Estratégia Nacional de Defesa e incentivando a interação do SisDQBRNEx com instituições científicas para fins de capacitação, pesquisas, desenvolvimento, ou para a solução de eventuais emergências ou crises. (COTER, 2014)

Com o presente estudo pretende-se contribuir para linhas de pesquisa do EB no âmbito da DQBRN, buscando a aplicação de técnicas atuais de Biologia Molecular, Engenharia Genética, Bioinformática e Biologia Computacional. Ajudando no entendimento das principais técnicas de bioprospecção, sequenciamento genético, montagem de genomas, determinação de estrutura de proteínas e modelagem molecular.

## 2 REVISÃO DA LITERATURA

Segundo Laird (2002), o termo bioprospecção é definido como qualquer pesquisa da biodiversidade com a intenção de coletar, amostrar e caracterizar organismos vivos com o objetivo de procurar e apresentar potenciais substâncias bioquímicas e recursos genéticos, presentes neles, que possam ser comercialmente úteis. Porém, atualmente, esta definição pode ser extrapolada, levando-se em conta a descobertas e o entendimento dos processos metabólicos intrínsecos dos

organismos e as infindáveis interações a nível molecular, dos mediadores químicos, existentes no relacionamento interespecies que tem grande potencial para serem reproduzidos e utilizados em processos biotecnológicos.

Por conta desta ampla funcionalidade faz-se necessário uma definição mais completa que inclua também a descoberta, descrição e potencial utilização da biodiversidade e seus processos, e como estes se relacionam com o ambiente, pois desse relacionamento ocorre a expressão de seu metabolismo, em parte, na forma de substâncias químicas, que atuam em diferentes níveis, como resultado do longo processo de evolução biológica.

Os diferentes microrganismos presentes em meio natural tentam explorar todos seus processos fisiológicos e metabólicos que possam lhe conferir vantagens competitivas para sua sobrevivência em detrimento a outros microrganismos. Eles devem utilizar os nutrientes comuns mais rapidamente, utilizar nutrientes que outros não conseguem metabolizar e ainda se possível produzir metabólitos que inibem o crescimento dos concorrentes. (Tortora et al., 2006).

É compreendendo esta realidade da competitividade biológica e da luta pela sobrevivência que utilizamos a bioprospecção para encontrar as vantagens adquiridas, por exemplo, por microrganismos que vivem em ambientes contaminados por resíduos químicos que seriam tóxicos ao homem e outros animais ou que vivem em área com elevada radioatividade. Estes organismos adaptaram-se evolutivamente a estes ambientes inóspitos e a compreensão dos fatores que levaram a isto pode permitir sua utilização pelo homem.

## 2.1 SEQUENCIAMENTO E MONTAGEM DE GENOMAS

O advento da tecnologia de Sequenciamento de DNA (Sanger, Nicklen e Coulson, 1977) em conjunto com a Bioinformática modificou radicalmente o cenário biológico científico, ajudando a caracterizar os mais diversos organismos. Atualmente, a Biologia Molecular e a Bioinformática possuem uma íntima relação, sendo praticamente impossível desenvolver pesquisas em ambas as áreas sem conhecer um pouco da outra. (Gauthier et al., 2018)

Desde o início do século XXI, o desenvolvimento de novas tecnologias de sequenciamento de DNA mais rápidas, mais baratas e mais precisas impulsionaram expressivamente o aumento da capacidade de geração de dados genômicos.

Porém, até mesmo o sequenciamento do genoma de organismos simples como as bactérias é da ordem de alguns milhões de nucleotídeos, esta imensa quantidade de dados é intratável manualmente e exige a criação de ferramentas automatizadas (programas de computadores) para que seja possível montá-los e analisá-los. (Lopes, 2015 e Gauthier et al., 2018)

## 2.2 ANOTAÇÃO GENÔMICA

Após o sequenciamento e montagem do genoma, para que seja possível extrair informações biológicas relevantes, é realizado um processo chamado de anotação genômica. Este processo irá inferir estruturas e funções de elementos do DNA, identificando os genes codificadores de proteínas, RNA's não codificantes, regiões regulatórias, sequências repetitivas, entre outras estruturas. (del Angel et al., 2018)

Apesar da anotação genômica envolver a caracterização de diversos elementos, uma maior atenção é dada à identificação dos genes codificadores de proteínas. Não que os outros elementos sejam menos importantes, mas principalmente porque os métodos utilizados para caracterizá-los requerem análises mais especializadas, enquanto o processo de determinar corretamente a localização e estrutura dos genes codificadores de proteínas é um processo bem compreendido e descrito, com vários algoritmos definidos e desenvolvidos nas últimas décadas. (del Angel et al., 2018)

## 2.3 DETERMINAÇÃO DA ESTRUTURA DE PROTEÍNAS

Segundo Lodish et al. (2014), as proteínas são as mais abundantes e funcionalmente versáteis macromoléculas celulares. Elas são moléculas dinâmicas cujas funções dependem de modo quase invariável de suas interações com outras moléculas e essas interações são afetadas de maneiras fisiologicamente importantes, por mudanças sutis ou súbitas na conformação espacial da proteína.

As interações moleculares são de extrema importância para a função proteica e muitas delas envolvem interações com uma grande variedade de moléculas diferentes, onde a maioria dessas interações é transitória. (Lodish et al., 2014)

Nosso sistema imunológico nada mais é que um grupo de células, tecidos e órgãos que atuam na defesa do organismo contra o ataque de invasores externos. Estes invasores podem ser microrganismos (bactérias, fungos, protozoários ou vírus) ou substâncias tóxicas (advindas do ambiente). (Lodish et al., 2014)

Tudo que é estranho que o corpo não reconhece são denominados antígeno. Os antígenos são combatidos por substâncias produzidas pelo sistema imune, de natureza proteica denominadas anticorpos, que reagem de forma específica com os antígenos possuindo a função de destruição celular, neutralização tanto das toxinas quanto do próprio antígeno e de impedir a multiplicação dos microrganismos invasores. (Lodish et al., 2014)

Como foi abordado, a ativação das proteínas está intimamente relacionada com a estrutura tridimensional da mesma, que por sua vez está diretamente ligada a composição de aminoácidos de cada proteína. O problema de sequenciar e determinar a estrutura tridimensional das proteínas foi o primeiro a ser analisado computacionalmente e remonta às origens da Bioinformática, no início dos anos 1960s. (Hagen, 2000)

Analisando estas informações já é possível compreender a importância da aplicação de técnicas de modelagem da estrutura de proteínas no combate de agentes QBRN. Pois, a partir da identificação dos antígenos responsáveis por uma ação danosa podemos, por engenharia reversa, estudar e propor, através de modelagem computacional, possíveis anticorpos complementares para combatê-lo. Da estrutura tridimensional inferir a composição de aminoácidos e desta a sequência de DNA codificante que por engenharia genética pode ser produzido e testado.

### **3 METODOLOGIA**

A trajetória desenvolvida pela presente pesquisa teve seu início na revisão teórica do assunto, através da consulta bibliográfica a manuais doutrinários, documentos e trabalhos científicos (artigos, teses e dissertações), a qual prosseguiu até a fase de análise e discussão destas informações.

A pesquisa foi realizada através de um estudo bibliográfico que, para sua conclusão, teve por método a leitura de material de pesquisa, bem como sua revisão

integrativa. A seleção das fontes de pesquisa foi baseada em publicações de autores renomados no meio acadêmico e em artigos veiculados em periódicos indexados pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

O delineamento da pesquisa contemplou as fases de levantamento e seleção da bibliografia, coleta de dados, crítica dos dados, leitura analítica e fichamento das fontes, argumentação e discussão dos resultados.

O presente trabalho limitou-se à citação e explicação de métodos computacionais desenvolvidos e publicados em artigos e textos científicos, em especial os atinentes a ferramentas computacionais disponíveis gratuitamente online e amplamente utilizadas pela comunidade científica no desenvolvimento de análises de Biologia Molecular, Engenharia Genética, Bioinformática, e demais áreas correlatas.

#### **4 RESULTADOS E DISCUSSÃO**

Um método muito utilizado para bioprospecção é o cultivo microbiano em meio sólido. Este meio pode ser preparado com características que se desejam observar, como utilizando possíveis toxinas de agentes químicos ou biológicos, onde serão inoculadas amostras coletadas do ambiente e colocados em estufas para estimular o crescimento microbiano. Após aguardar 24-48h de crescimento, as colônias são isoladas e então podem ser analisadas individualmente.

Com a colônia isolada podem ser realizados testes de resistência e antagonismo que irão identificar as bactérias com potencial para serem utilizadas. A partir desta seleção o passo mais comum a ser realizado é a extração e análise do material genético, o ácido desoxirribonucleico ou DNA.

A análise do DNA é iniciada através da utilização de um processo de sequenciamento. O sequenciamento identificará a sequência de bases nitrogenadas, adenina (A), citosina (C), timina (T) e guanina (G), existente na cadeia de DNA. Esta sequência é diretamente responsável por determinar as proteínas que podem ser produzidas e os componentes regulatórios de sua expressão, ou seja, da ativação ou inibição dos processos de produção das proteínas, conhecido como tradução.

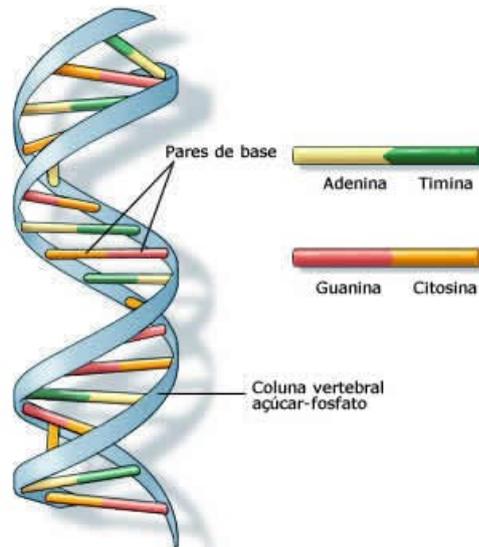


Figura 1: Estrutura tridimensional da cadeia de DNA, com esquema de suas bases nitrogenadas.  
**Fonte: <https://blogdoenem.com.br/dna-biologia-enem/>**

Contudo, o processo de sequenciamento é extremamente complexo e pode ser realizado por diferentes técnicas que possuem em comum o fato de que a leitura do DNA é feita em pequenos trechos, de 100 a 1000 bases, enquanto que a cadeia completa de DNA varia de alguns milhões a bilhões de bases. Para realizar a montagem do DNA completo, são necessários, então, métodos computacionais que consigam automatizar este processo.

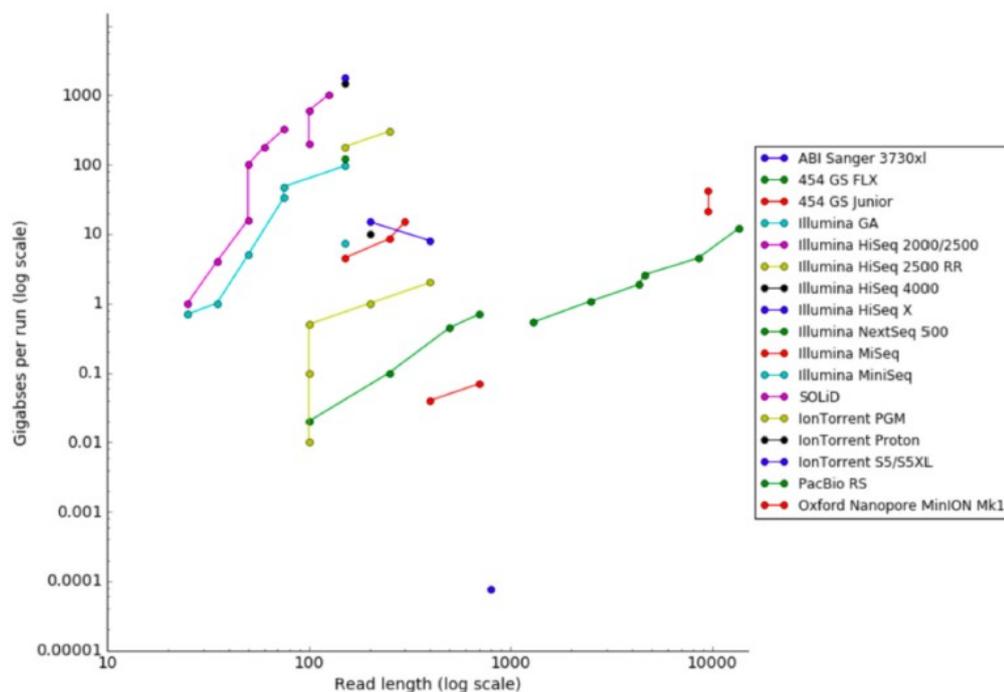


Figura 2: Comparativo das principais tecnologias de sequenciamento, plotando a relação throughput de dados por execução pelo tamanho médio dos fragmentos sequenciados.  
**Fonte: del Angel et al. (2018)**

As principais técnicas de montagem de DNA utilizam-se do seguinte artifício, são feitas inúmeras cópias do mesmo DNA, que através de processos físicos ou químicos, são quebrados, em posições aleatórias, em fragmentos menores com tamanhos capazes de serem sequenciados. Isto fará com que existam porções comuns nos fragmentos sequenciados, permitindo a tentativa de inferir a continuidade da cadeia.

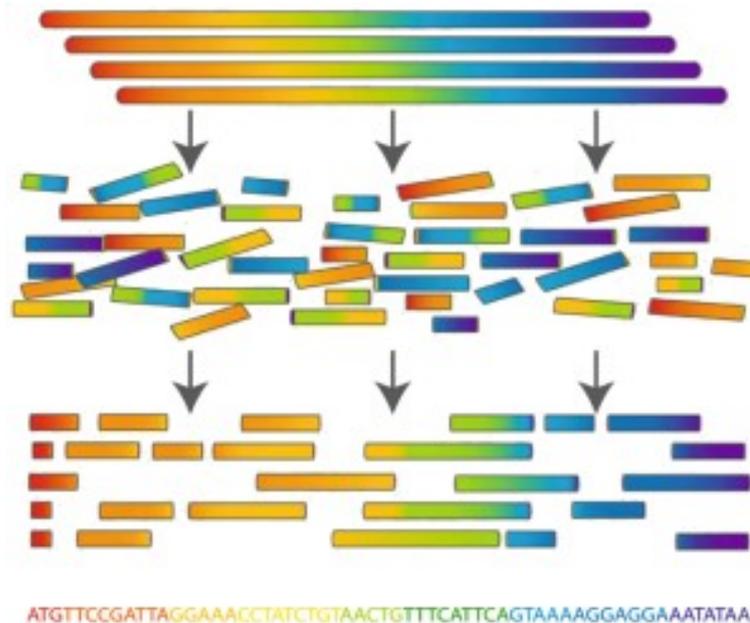


Figura 3: Esquema da montagem do sequenciamento do DNA.

Fonte: <https://dodona.ugent.be/en/exercises/1535795563/>

Apesar da montagem a princípio não ser precisa, a grande quantidade de cópias fará com que a mesma região seja analisada com diferentes fragmentos, o que é chamado de cobertura. Assim, regiões com alta cobertura tem alta probabilidade de expressar a sequência real e o sistema é capaz de analisando estas características descartar regiões de baixa cobertura e manter as regiões de alta cobertura.

É comum, e até mesmo esperado, que o processo de sequenciamento de DNA não gere a sequência completa do DNA em sua primeira execução. Porém, serão descobertos grandes fragmentos chamados de *contigs*, que cobrem grande parte da sequência total. Um segundo passo de sequenciamento pode ser realizado, utilizando os *contigs* já descobertos o que facilitará o preenchimento das lacunas, para obter o genoma completo.

Não existe uma ferramenta universal que realize todas as etapas de montagem do genoma, ainda mais, porque conforme exemplificado na figura 2

existem inúmeras tecnologias de sequenciamento, cada uma com suas características peculiares, que exige um conhecimento técnico do equipamento e do processo para que o procedimento seja realizado.

Assim, existem algumas iniciativas *open source* que buscam agrupar estas ferramentas em *pipelines* de execução para facilitar o uso pelos pesquisadores, um desses *pipelines* segue esquematizado na figura 4, abaixo.

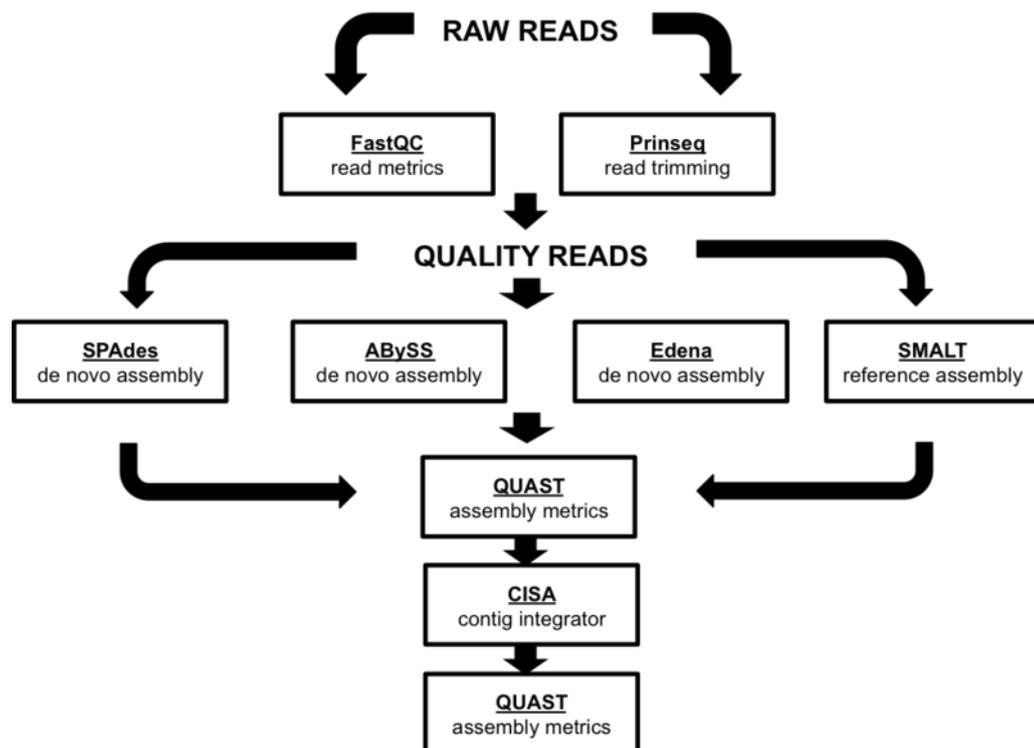


Figura 4: Pipeline de ferramentas para montagem do sequenciamento do DNA.  
 Fonte: [http://compgenomics2015.biology.gatech.edu/index.php/Genome\\_Assembly\\_Group](http://compgenomics2015.biology.gatech.edu/index.php/Genome_Assembly_Group)

A primeira linha descrita na figura 4 refere-se aos processos de pré-processamento. O **FastQC** é uma ferramenta que busca realizar um controle de qualidade dos fragmentos (*reads*) brutos, ou seja, que acabaram de ser recebidos do equipamento sequenciador, utilizando diversas métricas que ajudarão na filtragem de *reads* possivelmente problemáticos. O **Printseq** realiza um processo conhecido como *trimmagem* (corte), de adaptadores de cada fragmento, isto porque, diversas tecnologias de sequenciamento utilizam-se de adaptadores para realização do processo de clonagem (cópia) dos fragmentos, como estas sequências não fazem parte da cadeia original devem ser removidas antes de iniciar o processo de montagem propriamente dito.

A segunda linha do *pipeline* refere-se a ferramentas de montagem a partir dos

fragmentos preprocessados. O **SPAdes**, **AbySS** e **Edena** são ferramentas para montagem “*de novo*”, este tipo de montagem é puramente algorítmica utilizando variadas técnicas para determinar as maiores sequências de alta cobertura, os *contigs*. Estas 3 ferramentas utilizam métodos distintos para realizar esta mesma atividade, dependendo do tipo de amostra, podem ser utilizadas em paralelo para confirmar e complementar a montagem que está sendo realizada. Porém, comumente, para realizar o sequenciamento de longas cadeias de DNA ou de genomas completos utiliza-se uma sequência base, que é chamada de referência, de um organismo próximo evolutivamente que possui o DNA sequenciado, pois sabemos que existem poucas mudanças entre eles. Isto permite que a sequência de referência seja utilizada como um modelo inicial sobre o qual os fragmentos serão montados. O **SMALT** é uma ferramenta que utiliza esta técnica de montagem a partir de uma sequência de referência.

As últimas ferramentas do pipeline são o **QUAST**, que realiza a análise das sequências para medir a qualidade da montagem, retornando diversas estatísticas como número de *contigs*, tamanho dos *contigs*, tamanho total, N50 e L50 (parâmetros de qualidade); e o **CISA** que é utilizado para juntar *contigs* e obter sequências maiores.

Como o mundo do desenvolvimento de programas *open source* é muito aberto e altamente variável, já que qualquer pessoa pode utilizar um programa pronto e modificar apenas pequenas partes para adaptá-lo às suas necessidades, torna-se difícil manter registro das variadas ferramentas. Para ajudar neste sentido, é muito utilizado academicamente uma ferramenta conhecida como **Projeto Galaxy**. O Galaxy aproveita-se do fato de que a maioria das ferramentas de Bioinformática são utilizadas apenas através da linha de comando para criar uma interface comum que pode ser implementada para cada programa, unificando diversas ferramentas em um mesmo servidor web, facilitando seu acesso, utilização e compartilhamento. Assim, além dos diversos programas preexistentes, podem ser adicionados programas novos de maneira customizada, desde que sejam implementadas as adaptações para que ele possa ser chamado através da interface web do Galaxy. A figura 5, ilustra a interface do Projeto Galaxy que encontra-se disponível online no site da ferramenta. Porém, cada centro de pesquisa pode montar seu próprio *cluster* de processamento e incluir suas ferramentas localmente.

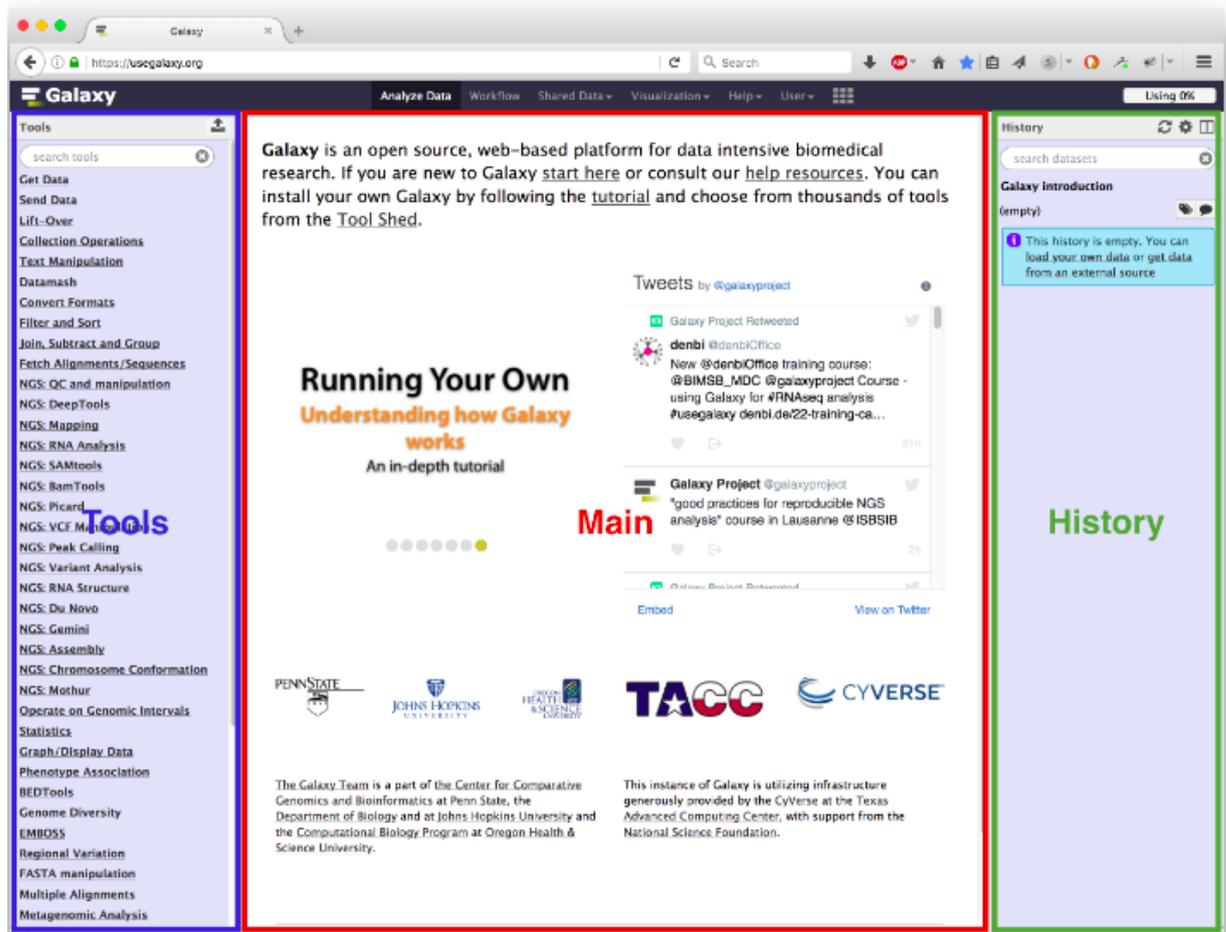


Figura 5: Estrutura do site do Projeto Galaxy, evidenciando o grupo de ferramentas instaladas à esquerda.

Fonte: <http://usegalaxy.org>

Com o genoma sequenciado, ainda há a necessidade de adicionar semântica aos dados adquiridos, isto é feito através do processo de anotação genômica, que irá determinar os genes e respectivas proteínas produzidas por este genoma. O processo de anotação é realizado através da comparação das sequências com grandes bancos de dados de genes anotados manualmente, ou seja, que foram curados e conferidos por biólogos antes de serem inseridos no sistema. Isto porque, apesar de se entender com bastante precisão a codificação genética dos genes, existem muitos elementos do DNA que não são codificantes o que torna necessário a confirmação biológica antes de garantir a existência de um gene. Porém, como os bancos de dados curados já estão bastante populados esta base de dados é confiável e consegue identificar grande parte dos genes existentes.

Uma ferramenta bastante utilizada neste processo é chamada de *RAST*, *Rapid Annotation Subsystem Technology*, traduzindo, Anotação Rápida usando Tecnologia de Subistema. Ele é um servidor de anotação online, gratuito e

totalmente automatizado para genomas completos, ou quase completos, de arqueas e bactérias. O RAST foi inicialmente idealizado para o uso do *National Microbial Pathogen Data Resource* (NMPDR), mas rapidamente tornou-se globalmente utilizado. Neste servidor o usuário entra com o genoma que deseja analisar, ou um conjunto de *contigs*, e recebe acesso ao genoma anotado em um ambiente para análise e comparação com outras centenas de genomas preexistentes no sistema.

Após determinar a função dos genes anotados, o RAST realiza uma reconstrução metabólica inicial, combinando genes do novo genoma a papéis funcionais de subsistemas predefinidos. Estes subsistemas estão agrupados em categorias refletindo os diversos módulos da maquinaria celular. Apesar de atualmente, na maioria das vezes, não ser possível estimar completamente os subsistemas de todos os genes, o conteúdo genômico que é conectado aos subsistemas provê uma detalhada estrutura para análise biológica, conforme evidenciado na figura 6.

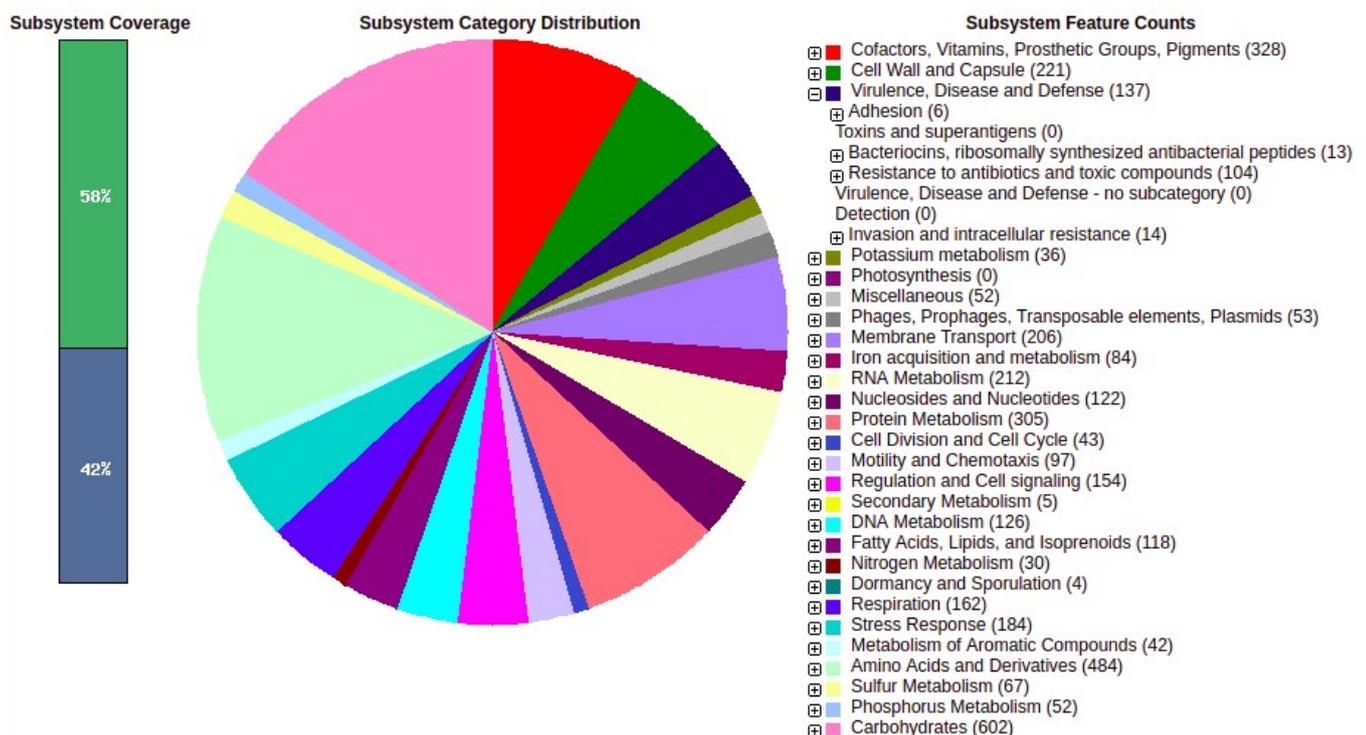


Figura 6: RAST: Genes conectados a subsistemas e sua distribuição em diferentes categorias.

Fonte: <http://rast.theseed.org/>

Esta informação é extremamente relevante para, por exemplo, determinar as proteínas envolvidas nos processos que se deseja estudar, no contexto de DQBRN destaca-se o subsistema de “Virulência, Doenças e Defesa” do qual podem ser

extraídas algumas proteínas proeminentes para estudo e utilização.

Para analisar a atuação de uma proteína é possível realizar, preliminarmente, a simulação computacional das interações moleculares, chamadas de *docking*, entre proteína e o seu ligante, visando compreender sua atuação e até propor modificações que possam tornar a proteína mais efetiva para ação em um determinado ligante. Os programas que realizam esta simulação baseiam-se na estrutura tridimensional destas moléculas e nas capacidades físico-químicas de seus componentes.

## 5 CONCLUSÃO

O compartilhamento de dados e ferramentas tem sido o grande motor das descobertas biológicas e científicas desta Área. Uma grande evidência deste fato é a existência dos diversos bancos de dados compartilhados, publicamente na internet, pelos grandes centros de pesquisas, os principais destes, o *NCBI Nucleotide Database* e o *RCSB Protein Data Bank*, possuem milhões de entradas que aumentam a cada dia, visto que tornou-se padrão requerer a submissão dos dados de uma pesquisa nestes tipos de depósitos para que uma publicação seja aceita. Ajudando não só a disseminação do resultado da pesquisa específica, mas incentivando e facilitando, amplamente, o desenvolvimento de novas pesquisas, permitindo inclusive que diversos trabalhos de Bioinformática sejam realizados sem a necessidade de qualquer experimento laboratorial, apenas pela consulta dessas bases públicas.

Do exposto, evidenciamos neste estudo a importância da computação no desenvolvimento do conhecimento biológico e como os inúmeros desafios apresentados podem ser resolvidos utilizando ferramentas computacionais. Apesar da existência destas ferramentas, tratar um volume cada vez maior de dados requer a pesquisa de algoritmos mais eficientes e mais precisos e o desenvolvimento de novas ferramentas que auxiliem os pesquisadores em suas atividades.

## REFERÊNCIAS

- SANGER, F.; NICKLEN, S.; COULSON, A. R. **DNA Sequencing with chain-terminating inhibitors**. Proc. Acad. Sci. USA. 1977.
- TAYLOR, C. L.; TAYLOR JR., L. B. **Chemical and Biological Warfare**. Nova Iorque: Franklin Watts, 1992
- HAGEN, J. B. **The origins of bioinformatics**. Nature Reviews, Genetics. Macmillan Magazines Ltd. Volume 1. 2000.
- LAIRD, S. A. **Biodiversity and Traditional Knowledge: Equitable Partnerships in Practice**. Earthscan Publications Ltd. London & Sterling (USA), pg. 41. 2002.
- de ARAÚJO, N. D.; de FARIAS, R. P.; PEREIRA, P. B.; de FIGUEIRÊDO, F. M.; de MORAIS, A. M. B.; SALDANHA, L. C.; GABRIEL, J. E. **A Era Da Bioinformática: Seu Potencial E Suas Implicações Para As Ciências Da Saúde**. PUC-PR. Estud Biol. 2008.
- TORTORA, G. J.; FUNKE, B. R.; CASE, C. L. **Microbiologia**. 10ª edição, Artmed, Porto Alegre. 2012.
- LODISH, H.; BERK, A.; KAISER, C. A.; KRIEGER, M.; BRETSCHER, A.; PLOEGH, H.; AMON, A. **Biologia Celular e Molecular**. 7ª Edição. 2014
- COTER. **A Evolução da Defesa Química, Biológica, Radiológica e Nuclear do Exército Brasileiro**. Revista Operacional: Defesa e Segurança. Acessado em: 12 de julho de 2019. Disponível em: <<https://www.revistaoperacional.com.br/2014/exercito/a-evolucao-da-defesa-quimica-biologica-radiologica-e-nuclear-do-exercito-brasileiro/>>. 2014
- PINTO, A. C.; LIMA, A. L. S.; FRANÇA, T. C. C. **Defesa Química**. Editora E-papers. Rio de Janeiro, 2015.
- LOPES, R. B. M. **Montagem e análise do genoma parcial de Bacillus thuringiensis BAC3151 endofítico das folhas do feijoeiro comum (Phaseolus vulgaris)**. Dissertação de Mestrado - Universidade Federal de Viçosa. 2015.
- de VASCONCELOS, A. M. C. **As Operações de Defesa Química, Biológica, Radiológica e Nuclear nos Grandes Eventos**. Doutrina Militar terrestre em revista. 2018
- GAUTHIER, J.; VINCENT, A. T.; CHARETTE, S. J.; DEROME, N. **A brief history of bioinformatics**. *Briefings in Bioinformatics*, 2018, 1-16. Oxford Press. 2018.
- del ANGEL, V. D.; et al. **Ten steps to get started in Genome Assembly and Annotation**. F1000Reserach 2018, 7(ELIXIR):148. 2018.