



Eloá Carolina Nava Cardoso Palma

Biodonut: Um Sistema *Web Open Source* com visualização
2D e 3D para manipulação de clonagem em plasmídeos

Limeira
2014



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE TECNOLOGIA
MESTRADO EM TECNOLOGIA

Eloá Carolina Nava Cardoso Palma

Biodonut: Um Sistema *Web Open Source* com visualização 2D e 3D para manipulação de clonagem em plasmídeos

Dissertação de Mestrado apresentada ao programa de Pós-Graduação da Faculdade de Tecnologia da Universidade Estadual de Campinas para obtenção do título de Mestra em Tecnologia, área de concentração AA – Tecnologia e Inovação.

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL
DA DISSERTAÇÃO DEFENDIDA PELA ALUNA
ELOÁ CAROLINA NAVA CARDOSO PALMA E
ORIENTADA PELO PROF. DR. MARCO ANTONIO
GARGIA DE CARVALHO E CO-ORIENTADA PELO
PROF. DR. MARCO AURÉLIO TAKITA.

Orientador: Prof. Dr. Marco Antônio Garcia de Carvalho (FT/UNICAMP)
Co-Orientador: Prof. Dr. Marco Aurélio Takita (CENTRO DE CITRICULTURA/IAC)

Limeira
2014

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Tecnologia
Silvana Moreira da Silva Soares - CRB 8/3965

P18b Palma, Eloá Carolina Nava Cardoso, 1987-
Biodonut : um sistema web open source com com visualização 2D e 3D para manipulação de clonagem em plasmídeos / Eloá Carolina Nava Cardoso Palma. – Limeira, SP : [s.n.], 2014.

Orientador: Marco Antonio Garcia de Carvalho.
Coorientador: Marco Aurelio Takita.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Tecnologia.

1. Clonagem. 2. Plasmídeos. 3. *In silico*. 4. Visualização. 5. Modelagem 3D. I. Carvalho, Marco Antonio Garcia de, 1970-. II. Takita, Marco Aurelio. III. Universidade Estadual de Campinas. Faculdade de Tecnologia. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Biodonut : a web open source system with 2D and 3D visualization for plasmids cloning manipulation

Palavras-chave em inglês:

Cloning

Plasmids

In silico

3D Modelling

Visualization

Área de concentração: Tecnologia e Inovação

Titulação: Mestra em Tecnologia

Banca examinadora:

Marco Antonio Garcia de Carvalho [Orientador]

Carlos Norberto Fischer

Luiz Camolesi Júnior

Data de defesa: 09-09-2014

Programa de Pós-Graduação: Tecnologia

DISSERTAÇÃO DE MESTRADO EM TECNOLOGIA
ÁREA DE CONCENTRAÇÃO: TECNOLOGIA E INOVAÇÃO

BIODONUT: um sistema web open source com visualização 2D e 3D para manipulação de clonagem em plasmídeos

Eloá Carolina Nava Cardoso de Palma

A Banca Examinadora composta pelos membros abaixo aprovou esta Dissertação:



Prof. Dr. Marco Antonio Garcia de Carvalho
FT/UNICAMP
Presidente



Prof. Dr. Carlos Norberto Fischer
UNESP



Prof. Dr. Luiz Camolesi Júnior
FT/UNICAMP

Resumo

O surgimento dos sequenciadores automáticos de DNA por volta dos anos 1990 fez com que a quantidade de sequências e informações a serem analisadas aumentasse exponencialmente. Sendo assim, os recursos computacionais são cada vez mais exigidos em diversas tarefas de armazenamento desses dados. Para atender essas e outras necessidades, surgiu a Bioinformática, que engloba diversas áreas como Ciência da Computação, Matemática, Estatística, e a Biologia Molecular. Neste trabalho, desenvolvemos um *software* denominado Biodonut, que consiste em manipular plasmídeos utilizando visualização 2D e 3D. Biodonut é um produto de software de plataforma *web, open source* para que os pesquisadores possam manipular plasmídeos *in silico*, visualizando em formato 2D e 3D e possibilitando uma análise diferenciada das áreas julgadas importantes. A base de dados utilizada no software é o GenBank, por ser mundialmente conhecida e respeitada. Uma comparação com dois importantes *softwares* da área foi efetuada e é apresentada neste trabalho. Os resultados apontam que usar Biodonut é vantajoso em relação os outros dois *softwares*.

Palavras chaves: *clonagem, plasmídeos, in silico, visualização 3D.*

Abstract

The emergence of automated DNA sequencers circa 1990s, made the amount of sequences to be analyzed and information to increase exponentially. Thus, the computational resources is increasingly required in various tasks of storing and interpreting these data. Starting from such requirements Bioinformatics has emerged, encompassing diverse areas as Software Engineering, Mathematics, Statistics, Computer Science and Molecular Biology. In this work, we developed Biodonut system. The software is a system for manipulation of plasmids using 3D visualization. Biodonut is the result of a development of an open source web platform software in which researchers can manipulate plasmids *in silico*, viewing in 3D format, allowing a differentiated analysis of the areas judged important. The data base chosen was GenBank, to be globally known and respected. A comparison with two important software of the area was conducted and is presented in this work The results show that using Biodonut is advantageous over the other two softwares.

Keywords: cloning, plasmids, *in silico*, 3D visualization.

SUMÁRIO

RESUMO	VII
ABSTRACT	VIII
SUMÁRIO	IX
AGRADECIMENTOS	XI
LISTA DE FIGURAS	XII
1 INTRODUÇÃO	1
1.1 OBJETIVOS	3
1.2 CONTRIBUIÇÕES	4
1.3 ORGANIZAÇÃO DO TEXTO	4
2 FUNDAMENTAÇÃO TEÓRICA	5
2.1 FUNDAMENTOS BIOLÓGICOS	5
2.1.1 <i>Plasmídeos</i>	5
2.1.2 <i>A tecnologia do DNA recombinante</i>	6
2.2 FUNDAMENTOS COMPUTACIONAIS	9
2.2.1 <i>Interface humano computador</i>	9
2.2.2 <i>Visualização da informação</i>	10
2.2.3 <i>Visualização 3D</i>	11
2.3 TRABALHOS RELACIONADOS	14
2.3.1 <i>pDRAW32</i>	14
2.3.2 <i>PlasMapper</i>	222
3 MATERIAIS E MÉTODOS	25
3.1 DIAGRAMA DE CASO DE USO DO SISTEMA	25
3.2 BASES DE DADOS	26
3.2 RECURSOS COMPUTACIONAIS	27
4 EXPERIMENTOS E RESULTADOS	30
4.1 BIODONUT	30
4.2 VIZUALIZAÇÃO 3D	36
4.3 ESTUDO COMPARATIVO	38

5 CONCLUSÕES E TRABALHOS FUTUROS.....	44
5.1 CONCLUSÕES	44
5.2 TRABALHOS FUTUROS.....	45
REFERÊNCIAS BIBLIOGRÁFICAS.....	46

Agradecimentos

Primeiramente agradeço a Deus, por ter me dado esta oportunidade, pela saúde, pelo direcionamento, e por ter colocado pessoas tão importantes em minha vida. Agradeço à Ele por todas as conquistas e vitórias.

À minha querida mãe Rita, e aos meus amados avós Helena e Ângelo, que tanto me amaram desde pequena. Sempre me incentivaram com palavras e atitudes sábias, carinho, respeito e serenidade, cuidando de mim em todos os momentos da minha vida e estando sempre prontos para me acolher independente das circunstâncias.

Ao meu marido Luís Vinícius pelo incentivo, apoio, respeito confiança e amor. Você me ensinou a não desanimar, me mostrando que se um dia eu consegui, posso conseguir novamente, e se ainda não consegui, posso lutar para que tudo dê certo.

Aos meus filhos Luís Gabriel e Maria Clara, meus amores incondicionais, simplesmente por existirem.

Ao meu orientador Prof. Dr. Marco Antônio Garcia de Carvalho, e ao meu coorientador Prof. Dr. Marco Aurélio Takita, pela paciência, assistência, confiança e respeito. Gostaria que soubessem que possuem meu profundo respeito, admiração e gratidão. Que Deus ilumine muito vocês e suas famílias.

À todos os professores que contribuíram para minha formação, em especial ao Prof. Dr. Hélio Pedrini que muito me ensinou com sua atenção, serenidade, inteligência.

Aos meus grandes e importantes amigos de universidade Fátima, Laura, Priscila, Maciel, William, Juninho, Vlademir, José Yauri e Carlos Alves pela convivência e por todos os momentos passados durante todo esse período. Vocês sempre foram presentes, nos bons e maus momentos, me incentivando e apoiando.

À CAPES pelo apoio financeiro deste trabalho.

LISTA DE FIGURAS

FIGURA 1: BACTÉRIA COM DNA CROMOSSÔMICO E PLASMÍDEO	6
FIGURA 2: TÉCNICA DO DNA RECOMBINANTE.	7
FIGURA 3: ENZIMAS DE RESTRIÇÃO (ENDONUCLEASES DE RESTRIÇÃO).	8
FIGURA 4: IMAGEM 3D DO CRÂNIO HUMANO. FONTE: SOFTWARE INVESALIUS	12
FIGURA 5: MODELO DIGITAL DE RELEVO EM 3D. FONTE: ARCSCENE.....	13
FIGURA 6: EXEMPLO DA FIGURA GEOMÉTRICA TORUS	14
FIGURA 7: TELA INICIAL DO SOFTWARE PDRAW32.	15
FIGURA 8: CAIXA DE CLONAGEM DE NOVA SEQUÊNCIA DE PLASMÍDEO.	16
FIGURA 9: SEQUÊNCIA DE PLASMÍDEO LINEARIZADA	17
FIGURA 10: PLASMÍDEO YEP24, CIRCULARIZADO	18
FIGURA 11: CAIXA DE DIÁLOGO PARA SELEÇÃO DE ENZIMAS.	19
FIGURA 12: PLASMÍDEO YEP24 APÓS A MANIPULAÇÃO. O CÍRCULO VERMELHO INDICA O SEGMENTO INSERIDO NO PLASMÍDEO	20
FIGURA 13: GRÁFICO CIRCULAR DO PLASMÍDEO YEP24 DE TAMANHO 7769BP.	21
FIGURA 14: PÁGINA INICIAL DA FERRAMENTA PLASMAPPER.	23
FIGURA 15: GRÁFICO CIRCULAR DO PLASMÍDEO YEP24 GERADO PELA FERRAMENTA PLASMAPPER.	24
FIGURA 16: DIAGRAMA DE CASO DE USO DO SOFTWARE BIODONUT	25
FIGURA 17: DIAGRAMA DE FLUXO DE DESENVOLVIMENTO DO SOFTWARE BIODONUT.....	29
FIGURA 18: PÁGINA INICIAL DO BIODONUT.....	30
FIGURA 19: LISTAGEM DOS PLASMÍDEOS PRESENTES NO BIODONUT	31
FIGURA 20: TELA PARA VISUALIZAÇÃO DE PLASMÍDEO	32
FIGURA 21: TELA PARA MANIPULAÇÃO DE PLASMÍDEO.	33
FIGURA 22: REPRESENTAÇÃO GRÁFICA 2D DO PLASMÍDEO YEP24 ANTES E APÓS A MANIPULAÇÃO GERADA PELO BIODONUT:.....	35
FIGURA 23: VISUALIZAÇÃO 3D DO PLASMÍDEO YEP24 ANTES (A) E DEPOIS (B) DA MANIPULAÇÃO.	37

1 Introdução

Sistemas computacionais de informação estão cada vez mais presentes em diversos aspectos da vida humana, coletando e armazenando volumes de dados heterogêneos que crescem exponencialmente e de maneira distribuída. Essas características em conjunto dificultam consideravelmente a obtenção de informação relevante sobre estes dados. Nesse sentido, pesquisadores brasileiros apontam como um grande desafio de pesquisas em Ciência da Computação no Brasil até 2016 a integração de diferentes áreas da Computação para desenvolver aplicações que beneficiem o contexto sócio econômico cultural do país, provendo soluções para tratar, recuperar e disseminar informação relevante a partir desses volumes de dados (CARVALHO *et al.* 2006).

Esse objetivo é consoante com os objetivos da área de Visualização de Informação, que visa facilitar o processo de derivação e entendimento de informação a partir da análise visual de conjuntos de dados. Diferentes técnicas de Visualização de Informação utilizam recursos computacionais para representar esses dados de maneira gráfica e interativa, procurando otimizar o uso das capacidades visuais humanas para compreender fenômenos que não possuem em si mesmos uma representação espacial própria (CARD *et al.* 1999; CHEN 2002).

Bioinformática corresponde a aplicação das técnicas da informática, no sentido de análise da informação na área de estudo da biologia. Uma definição ampla é: (Bio)informática é o estudo da aplicação de técnicas computacionais e matemáticas à geração e gerenciamento de (bio)informação (ATTWOOD, GISEL, ERIKSSEN, BONGCAM-RUDLOFF, 2011). Assim, os computadores dão suporte na solução de problemas principalmente envolvendo a manipulação de grandes bancos de dados de genomas, sequências proteicas, entre outros. Adicionalmente, ela também engloba técnicas como a modelagem tridimensional de moléculas e demais sistemas biológicos. A bioinformática visa coletar, organizar, armazenar, recuperar e analisar dados biológicos e suas evoluções.

O estudo do sequenciamento genético foi o grande propulsor da bioinformática, surgindo da necessidade da integração de várias ciências dentre elas a computação que é fundamental para tal estudo. A importância da bioinformática para a ciência justifica-se na construção de base de

dados contendo informações dos genes e proteínas dos seres vivos, podendo convergir à identificação novos genes, medicamentos entre outras possíveis descobertas.

A bioinformática fortaleceu-se com o Projeto Genoma Humano (BORGES *et al*, 2013) pela necessidade de um software que pudesse armazenar e analisar as grandes quantidades de dados que eram gerados. Como os mapeamentos de genes geram um elevado número de informações, é necessário organizá-las em bancos de dados computacionais.

A maioria dos bancos de dados biológicos consiste em longas cadeias de caracteres para representar as bases do DNA (LEAL, WIECZOREK, 2003). Os Bancos de Dados Públicos de bioinformática envolvem sequência de nucleotídeos, aminoácidos ou estruturas de proteínas. Dentre os principais bancos de dados temos: GenBank, EBI (*European Bioinformatics Bank*) e DDBJ (*DNA Data Bank of Japan*). Estes bancos são membros do INSDC (*International Nucleotide Sequence Database Colaboration*) onde cada um deles permite a submissão individual de sequências de DNA e compartilham informações regularmente entre si .

A biotecnologia, ou engenharia genética ou ainda tecnologia do DNA recombinante, é a manipulação direta do material genético das células (DNA), para alterar ou introduzir características nos seres vivos. Assim, o uso da biotecnologia requer a identificação e isolamento de sequências de genes responsáveis por expressar uma característica de interesse. Para isso, são usados vetores de clonagem ¹como plasmídeos², onde a sequência de interesse é inserida. Esse processo produz um Organismo Geneticamente Modificado (OGM), também chamado de organismo transgênico, e sua característica adquirida passa a ser hereditária.

Uma aplicação de clonagem muito comum é a produção de insulina que é retirada de um gene humano e introduzida em bactérias. As bactérias então produzem grande quantidade do hormônio que é isolado e purificado, sendo utilizados por diabéticos (PIERCE, 2004).

Mapeamento de plasmídeo é uma das operações mais antigas e mais frequentemente realizado em bioinformática. O mapeamento gerado por computador é essencial para identificar, localizar e analisar as principais regiões em uma sequência de vetores (PIERCE, 2004), além de

¹ Clonagem é um mecanismo comum de reprodução de espécies de plantas ou bactérias.

² Plasmídeos são moléculas de DNA capazes de amplificar, em centenas de cópias, a informação genética que neles foi inserida. Existem diferentes tipos de vetores possuindo cada um particularidades que lhe são próprias.

encurtar o tempo gasto para análise do experimento e recursos financeiros, visto que os plasmídeos são comerciais.

A seguir, apresentam-se algumas aplicações mais conhecidas da técnica no DNA recombinante:

- Estudos de mecanismos de replicação e expressão de genes;
- Determinação da sequência de um gene (e proteína codificada);
- Manipulação genética a resistência a pestes em plantas para reduzir a dependência de pesticidas químicos;
- Fabricação de produtos farmacêuticos, como por exemplo a produção de insulina humana e hormônio de crescimento humano;
- Terapia gênica (transferência direta de genes em humanos para tratar uma doença);
- Mapeamento gênico (capacidade de determinar a presença de genes responsáveis por vários distúrbios humanos).

1.1 Objetivos

Este trabalho tem como objetivo desenvolver um sistema computacional denominado Biodonut, que possibilita a manipulação de clonagem usando plasmídeos *in silico*, bem como a simulação da operação de clonagem em plasmídeos com visualização 2D e 3D. A finalidade do software é permitir a análise e manipulação de sequências de DNA de uma forma simples e intuitiva, suprimindo as necessidades diárias de cientistas, e ajudá-los a manter o controle das novas sequências gênicas que produzem. Para isso, a implementação deste trabalho engloba visualização 2D e 3D dos plasmídeos, a fim de aumentar a interação do pesquisador com o sistema, além de permitir inserções de genes, deleções e edições nos plasmídeos.

1.2 Contribuições

A principal contribuição deste trabalho é agregar melhor visualização do grande volume de dados presente nos plasmídeos através da terceira dimensão. Além disso, a ferramenta desenvolvida é multiplataforma e *web*, auxiliando o processo de manipulação de clonagem *in silico* em diferentes laboratórios simultaneamente. Ainda há espaço para trabalhos futuros em diversas áreas da ciência da computação e da bioinformática abordadas no capítulo 5.

1.3 Organização do texto

Este texto está organizado em 5 capítulos. O capítulo 1 aborda o tema em questão em nível introdutório apresentando motivação e justificativa do trabalho, objetivos, contribuições bem como a organização do texto. O capítulo 2 expõe a fundamentação teórica necessária para entendimento do desenvolvimento deste trabalho. Nele encontramos fundamentos biológicos como conceitos sobre plasmídeos e tecnologia do DNA recombinante, além da exposição da fundamentação computacional utilizada para o desenvolvimento do sistema computacional Biodonut. Tais conceitos englobam interface humano computador, visualização da informação e visualização 3D. Ainda aponta alguns dos trabalhos relacionados existentes avaliados durante o desenvolvimento do sistema computacional Biodonut, como pDRAW32 e PlasMapper.

No capítulo 3 encontramos os materiais e métodos utilizados no desenvolvimento e recursos computacionais para conclusão do projeto, enquanto o capítulo 4 explora experimentos e resultados, bases de dados, *hardwares* utilizados para o desenvolvimento, bem como a apresentação do fluxo do sistema computacional Biodonut. Ainda neste capítulo são avaliados os resultados da visualização 3D. O capítulo 5 apresenta as conclusões e aponta possibilidades para trabalhos futuros. As referências bibliográficas utilizadas para a elaboração deste trabalho encontram-se ao final deste documento.

2 Fundamentação Teórica

Abordaremos agora, os fundamentos biológicos e computacionais relacionados ao tema. De forma geral, este capítulo auxilia no melhor entendimento dos conceitos utilizados neste trabalho.

2.1 Fundamentos Biológicos

Nesta seção são apresentados conceitos biológicos tais como plasmídeo e tecnologia do DNA recombinante.

2.1.1 Plasmídeos

Plasmídeo (ou vetor de clonagem) é um DNA extracromossômico, encontrado em alguns micro-organismos, sendo determinantes de resistência a antibióticos, produção de toxinas, além de possuir replicação própria. Os plasmídeos, por se apresentarem em um maior número de cópias, são usados como veículos de clonagem uma vez que permitem a amplificação do segmento do DNA neles clonado. Na Figura 1 podemos visualizar uma bactéria contendo seu DNA cromossômico e também seu DNA extracromossômico circular (plasmídeo). Na tecnologia de DNA recombinante, a resistência a antibióticos é de grande utilidade uma vez que pode determinar seleção, onde células transformadas com vetores contendo determinantes de resistência são capazes de crescer em meio contendo o antibiótico, enquanto que as células não transformadas acabam morrendo, facilitando o processo de clonagem. Um dos passos fundamentais no processo de clonagem molecular é o uso de enzima de restrição que produz extremidades compatíveis durante a clivagem³ do DNA a ser clonado (inserto) e a do DNA receptor (vetor) (LOPES et al., 2012).

³Clivagem equivale ao corte do DNA

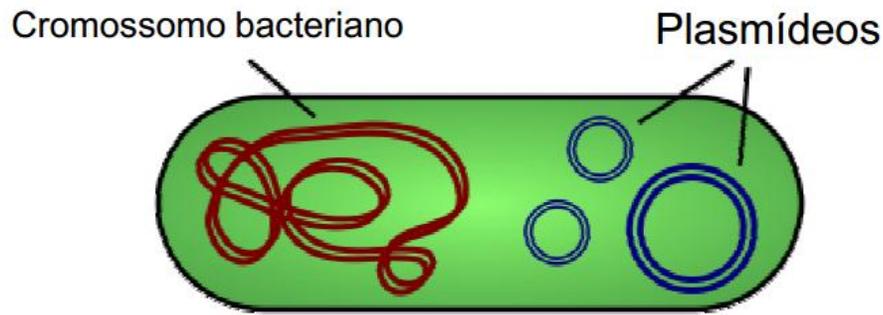


Figura 1: Bactéria com DNA cromossômico e plasmídeo

Fonte: MONDEGO, J. Notas da aula conceitos de biologia molecular.

2.1.2 A tecnologia do DNA recombinante

A Tecnologia do DNA Recombinante, também chamada engenharia genética, é um conjunto de técnicas para localizar, isolar, alterar e estudar segmentos de DNA. A engenharia genética pode ser definida como um conjunto de processos que permitem a manipulação do genoma de microrganismos vivos, com a consequente alteração das capacidades de cada espécie. Algumas aplicações da técnica do DNA recombinante podem ser encontradas em: teste de paternidade e identificação de criminosos, detecção de genes relacionados ao câncer, terapia gênica, produção de vacinas, obtenção de organismos transgênicos, entre outros (CANDEIAS, 1991).

O desenvolvimento desta nova tecnologia só foi possível graças a descoberta das enzimas de restrição. Este tipo de enzima atua como uma espécie de "tesoura biológica" que, após reconhecer determinada sequência do nucleotídeo, faz corte na ligação açúcar-fosfato da molécula de DNA, produzindo fragmentos. Elas são produzidas pelas próprias bactérias como forma de defesa contra vírus, onde clivam em diversos fragmentos o material genético dos vírus, impedindo sua reprodução na célula bacteriana. Por outro lado, a bactéria protege seu próprio DNA dessa degradação, modificando sua sequência de reconhecimento pela adição de grupos metila (metilação) (BURNS & BOTINO, 1991).

A Figura 2 ilustra a técnica do DNA recombinante, onde o gene humano da insulina foi ligado ao plasmídeo, gerando um plasmídeo recombinante. Logo, esta molécula híbrida deverá ser introduzida numa célula hospedeira (geralmente bactérias), para que o vetor possa sofrer

replicações (duplicações conforme a Figura 2) e conseqüentemente amplificar o número de cópias do inserto. A bactéria transformada será facilmente reconhecida pela aquisição de um novo

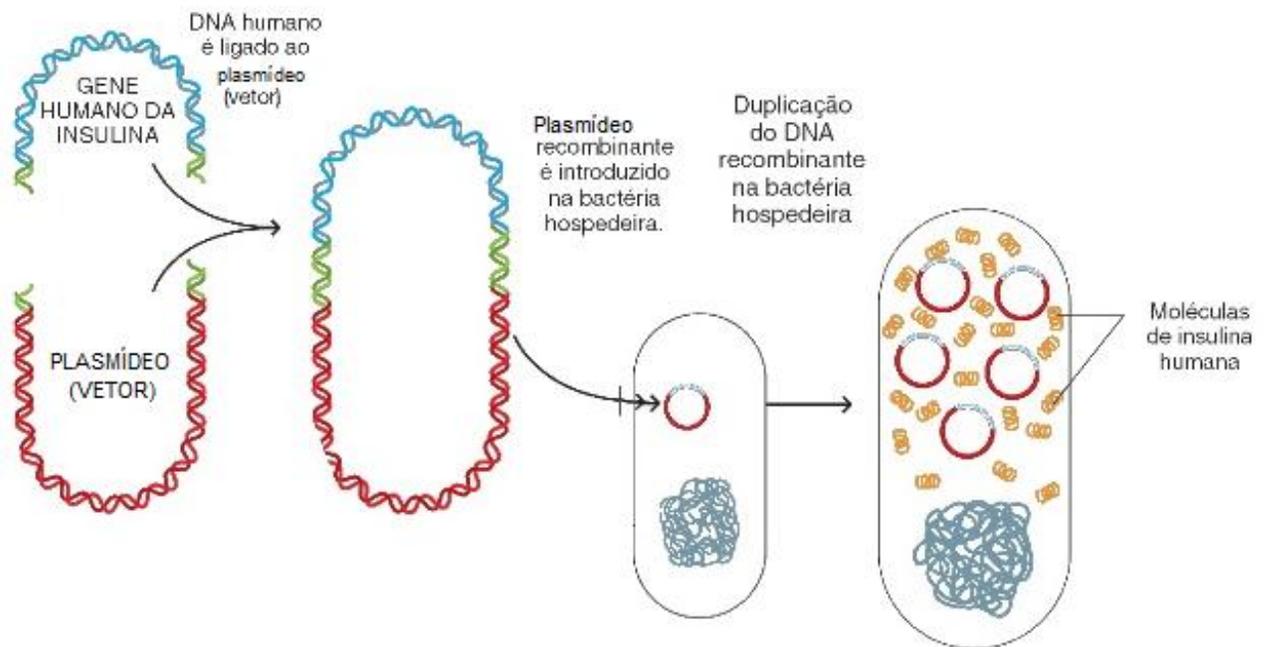


Figura 2: Técnica do DNA Recombinante.

Fonte: Tecnologia do DNA recombinante. *Biologia das populações*/José Mariano Amabis, Gilberto Rodrigues Martho. – 2. ed. – São Paulo: Moderna, 2004

fenótipo ⁴dado pelo plasmídeo, ou seja, pela capacidade de crescer em meios contendo antibiótico, por exemplo.

O interesse pelas enzimas aumentou ao constatar-se que elas poderiam ser usadas para fragmentar o DNA deixando extremidades de fitas simples de DNA que permitiam a ligação dos fragmentos. Além disto, DNA bacteriano poderia recombinar com DNA humano ou de qualquer outra espécie, abrindo a possibilidade de clonar genes humanos ou isolar proteínas de culturas bacterianas. Uma importante consequência da especificidade destas enzimas de restrição é que o

⁴ Fenótipo: característica física do indivíduo.

número de clivagens feito por cada uma delas no DNA de qualquer organismo é definido e permite o isolamento de fragmentos deste DNA. Portanto, cada enzima de restrição gera uma família única de fragmentos quando cliva uma molécula de DNA específica.

Uma enzima de restrição particular reconhece uma sequência única de bases. DNAs de origens diferentes sob a ação da mesma enzima de restrição produzem fragmentos com o mesmo conjunto de extremidades. Portanto, fragmentos de dois organismos diferentes (por exemplo, bactéria e humano) podem ser ligados. Além disto, se a ligação for "selada" com a enzima *DNA ligase*, depois do pareamento de bases, os fragmentos serão ligados permanentemente.

A Figura 3 mostra uma molécula de DNA de plasmídeo com apenas um sítio de clivagem para uma enzima de restrição. A mesma enzima é usada para cortar o DNA humano. Se os fragmentos de DNA humano são misturados com o DNA dos plasmídeos, permitindo a ligação entre eles, uma molécula de DNA do plasmídeo contendo DNA humano pode ser gerada. Este plasmídeo híbrido pode ser inserido numa bactéria e então o inserto será replicado como parte do plasmídeo.

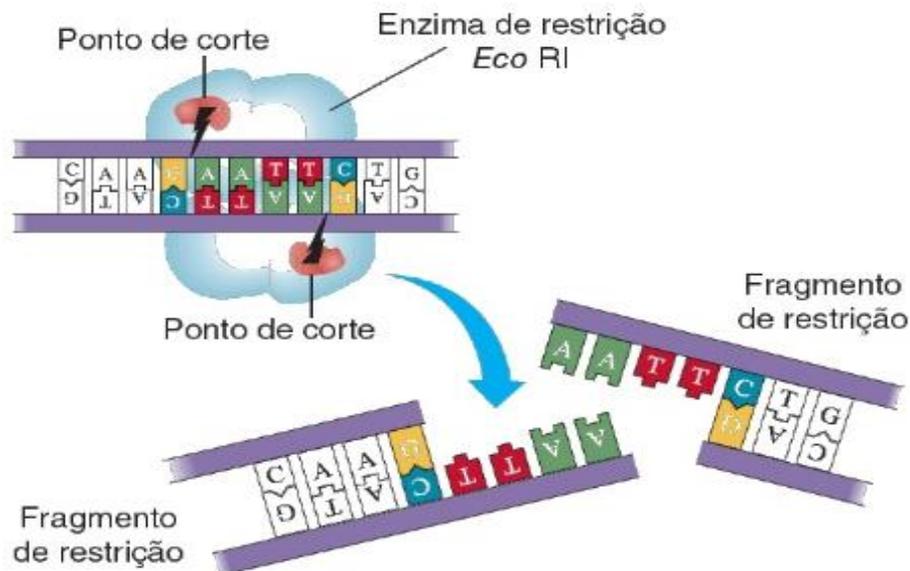


Figura 3: Enzimas de restrição (endonucleases de restrição).

Fonte: Tecnologia do DNA recombinante. Biologia das populações/José Mariano Amabis, Gilberto Rodrigues Martho. – 2. ed. – São Paulo: Moderna, 2004

2.2 Fundamentos computacionais

Nesta seção são apresentados conceitos computacionais como Visualização da informação e visualização 3D. Estes conceitos são considerados muito importantes para o entendimento do trabalho.

2.2.1 Interface humano computador

Para os experimentos e análise de resultados, foram utilizadas técnicas de IHC (acrônimo para Interface Humano Computador), que no inglês é encontrado sob a sigla HCI - *Human-Computer Interface*. O termo IHC possui afinidade com as questões de ‘interface com o usuário’ (Preece, 2005). A interface é o ambiente visual de comunicação entre o usuário e o sistema que dispõe de elementos básicos tais como: os menus, as janelas, as mensagens sonoras e diálogos. A interface deve dispor de condições para o seu uso e facilidades no aprendizado (GONÇALVES, 2001).

Segundo Jakob Nielsen (2007) a “usabilidade é um atributo de qualidade que avalia quão fácil uma interface é de usar”, ou “a medida de qualidade da experiência de um usuário ao interagir com um produto ou um sistema”. Em outras palavras a usabilidade está associada à utilização de métodos que contribuam com a facilidade de uso durante o processo de criação do produto (*website*, aplicação de software, tecnologia móvel, ou qualquer dispositivo operável por um usuário). A usabilidade está distribuída em diversos elementos, sendo tradicionalmente associada, segundo Jakob Nielsen, aos seguintes fatores:

- Facilidade de uso : o sistema deve ser fácil de assimilar pelo usuário, para que este possa aprender a trabalhar rapidamente;
- Eficácia: está relacionada à escolha e, depois de escolhido o que fazer, fazer esta coisa de forma produtiva leva à eficiência. A eficácia é o grau em que os resultados de uma ação que correspondem às necessidades e aos desejos do ambiente externo;
- Eficiência : o sistema deve ser eficiente para que o usuário, depois de o saber usar, possa atingir uma boa produtividade.

Tais fatores foram analisados e considerados nos experimentos e resultados do Biodonut e durante a análise comparativa com outros softwares com a mesma finalidade de Biodonut. Os experimentos e resultados baseados nestas métricas de IHC estão presentes no capítulo 4 deste documento.

2.2.2 Visualização da informação

Informações representadas graficamente geralmente são processadas de maneira quase que instantânea pela visão, em um processo mais superficial, paralelo (no sentido de obter várias informações simultaneamente), rápido e de capacidade elevada. Dessa forma, representar graficamente os dados a serem analisados é interessante do ponto de vista da obtenção de informação, pois faz com que não apenas mecanismos computacionais sejam usados para a análise de dados, mas também recursos da visão e da cognição humana. Nesse sentido, um estudo de Larkin e Simon (1987) estendido por Card et al. (1999) aponta diferentes aspectos pelos quais visualizações podem possibilitar uma ampliação cognitiva ou seja, uma maior facilidade no uso e aquisição do conhecimento, como:

- Aumentar os recursos de memória e de processamento disponíveis para usuários, através do uso direto dos recursos do sistema visual e da memória de trabalho externa e visual;
- Agrupar ou relacionar visualmente informações, compactando-as, exibindo uma visão geral ou mesmo mostrando detalhes sob demanda;
- Usar representações visuais para melhorar a detecção de padrões;
- Tornar óbvia a resposta de um problema através de uma representação visual;
- Usar mecanismos de atenção perceptiva para efetuar monitoramento de uma grande quantidade de eventos potenciais;
- Codificar informação numa mídia manipulável.

Para que essa ampliação cognitiva possa ocorrer, é importante definir de maneira apropriada o modo como as estruturas de dados são transformadas em estruturas visuais, processo

este chamado de mapeamento visual. Esse processo deve levar em conta quais as características dos dados a serem representados, das propriedades visuais que podem ser usadas para representar esses dados, da interface humano-computador envolvida no processo (no caso, a tela de exibição dos dados e os dispositivos de recebimento de interações do usuário), e da visão e cognição humana (Silva, 2006).

2.2.3 Visualização 3D

Nos últimos anos assistimos a substanciais avanços técnicos em computação gráfica e a uma proliferação de computadores significativamente mais poderosos com a possibilidade de gerar gráficos complexos 3D. Adicionalmente, também se assistiu a um grande aumento na acessibilidade a informação e na ligação a redes de comunicação. Estes fatores tornaram possível novas formas de apresentar visualmente e, portanto, comunicar informação.

Se comparado com o processo de visualização bidimensional, a visualização de objetos tridimensionais é mais complexa. Esta complexidade deve-se, basicamente, ao fato de que os dispositivos gráficos existentes são adequados à apresentação de imagens planas, bidimensionais. A visualização de informação em duas dimensões (2D) se utiliza de parâmetros visuais disponíveis em 2D (cor, tamanho, posição, forma) para representar características e propriedades dos dados. A visualização da informação em três dimensões acrescenta uma nova dimensão à representação dos dados, tornando assim possível uma utilização mais eficiente do espaço limitado disponível – o monitor do computador. Isto é particularmente importante para grandes quantidades de dados a serem visualizados.

A investigação em tecnologias de Realidade Virtual sugere que utilizar técnicas multimídia para codificar informação e apresentar esta informação num espaço perceptual 3D aumenta a quantidade de informação com a qual as pessoas são capazes de interagir (ROBERTSON *et al.*, 1991). Mais ainda, se os objetos forem colocados numa representação 3D, em vez de 2D, a complexidade perceptível da informação apresentada é menor (WARE e FRANCK, 1994).

Ware and Frank, 1994 concluíram o motivo de que modelar software em 3D poderá ser melhor do que em 2D, afirmando através de experiências com usuários que a quantidade de informação (na forma de gráficos) que pode ser absorvida utilizando 3D é mais ou menos 3 vezes

a quantidade em 2D. Isto indica que esses usuários conseguiram efetuar melhor as tarefas (que necessitavam uma boa compreensão da informação exibida) sobre a estrutura dos gráficos quando estes eram exibidos em 3D do que em 2D (Dwyer, 2001)

A visualização da informação em três dimensões também facilita a compreensão dos dados através da utilização de novos parâmetros visuais (material, luminosidade, transparência) e novas técnicas de interação (rotações geométricas 3D, “passeios” através dos dados) que convidam os usuários a explorar e manipular sistemas de informação grandes e complexos. Estes sistemas de informação podem ir desde catálogos de bibliotecas às cotações diárias das ações na bolsa ou aos arquivos de voos de uma companhia de aviação.

A Figura 4 apresenta um modelo virtual em três dimensões (3D) do crânio humano de pacientes em tratamento médico, o que auxilia no diagnóstico e o planejamento cirúrgico com imagens obtidas através de equipamentos de tomografia computadorizada ou ressonância magnética.

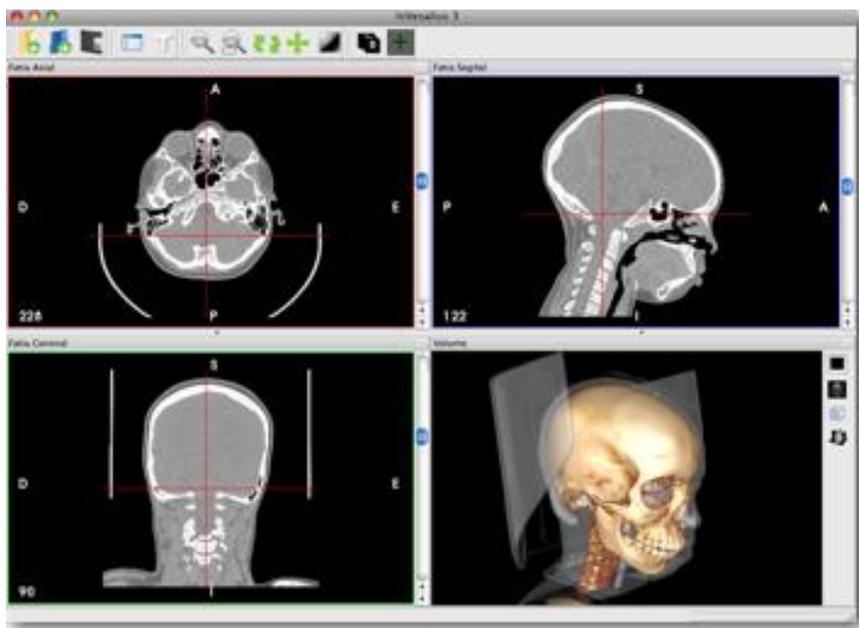


Figura 4: Imagem 3D do crânio humano.

Fonte: Software InVesalius

Na Figura 5, a visualização em 3D, atua sobre três dados dimensionais e de superfície.

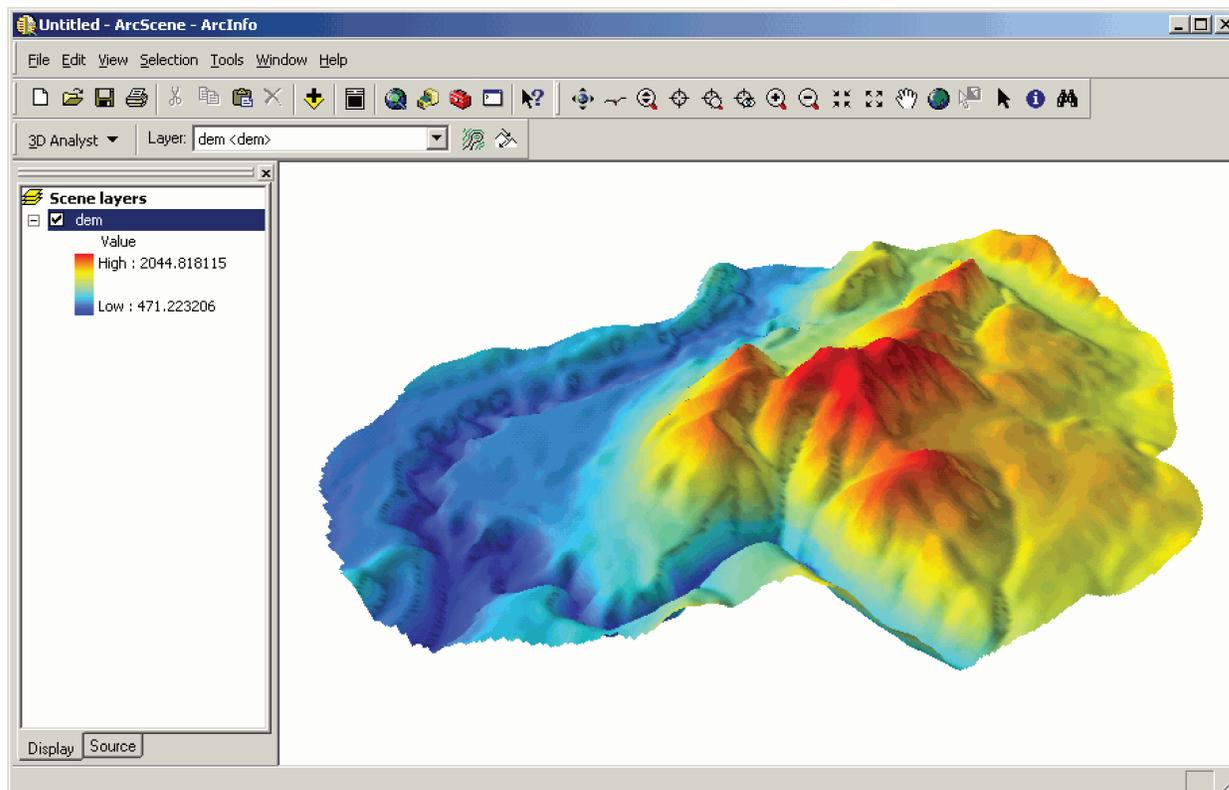


Figura 5: Modelo digital de relevo em 3D.

Fonte: ArcScene

Para o desenvolvimento da visualização 3D no sistema computacional Bionnut, foi utilizada a geometria Torus, ilustrada na Figura 6, por ser a figura geométrica mais próxima da realidade geométrica de um plasmídeo.

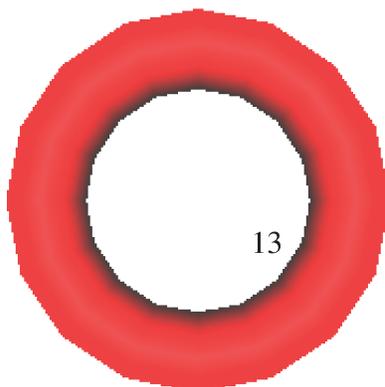


Figura 6: Exemplo da figura geométrica Torus

2.3 Trabalhos Relacionados

Para o desenvolvimento do *Biodonut*, algumas dos *softwares* mais utilizados no ambiente acadêmico e corporativo como *PlasMapper* e *pDRAW32* foram avaliados.

2.3.1 pDRAW32

pDRAW32⁵ é um software de análise de DNA, desenvolvido pela empresa AcaClone software, que permite a montagem de plasmídeos recombinantes *in silico*. Também analisa os sítios de restrição presentes numa sequência de DNA qualquer. Além disso, o pDRAW32 cria um gel virtual de agarose, onde aparecem as bandas de DNA correspondentes aos fragmentos gerados pela digestão do DNA com a enzima escolhida (ACACLONE, 2012).

Sua licença é do tipo *shareware*, porém muitas de suas funcionalidades podem ser usadas sem licença; no entanto, o licenciamento fornece ao usuário benefícios adicionais. Uma instalação licenciada do pDRAW32 não requer atualização regular, embora seja recomendável atualizar a mais recente versão. pDRAW32 permite realizar várias operações, tais como: anotações relativas ao DNA em estudo, clonar DNA, editar sequências, analisar sequências, selecionar enzima, exportar gráficos e texto, calcular homologias entre dois fragmentos de DNA entre outros. O pDRAW32 gera também gráficos plasmidiais conforme mostrado nas Figuras 10, 11 e 12, que representam o plasmídeo YEp24, com 7769 pares de bases, como visualizado pelo pDraw32, circularizado, nomeado, com seu comprimento identificado e com os sítios de restrição para o conjunto de enzimas do banco de dados de enzimas de restrição do programa já

⁵ pDRAW32 disponível no site da empresa que a desenvolveu: <http://www.acaclone.com>

assinalados.

A licença também garante que o usuário crie arquivos no formato pDRAW32. Sua interface inicial está ilustrada na Figura 7, enquanto que a Figura 8 mostra o procedimento para inserir o gene do plasmídeo que será usado na manipulação.

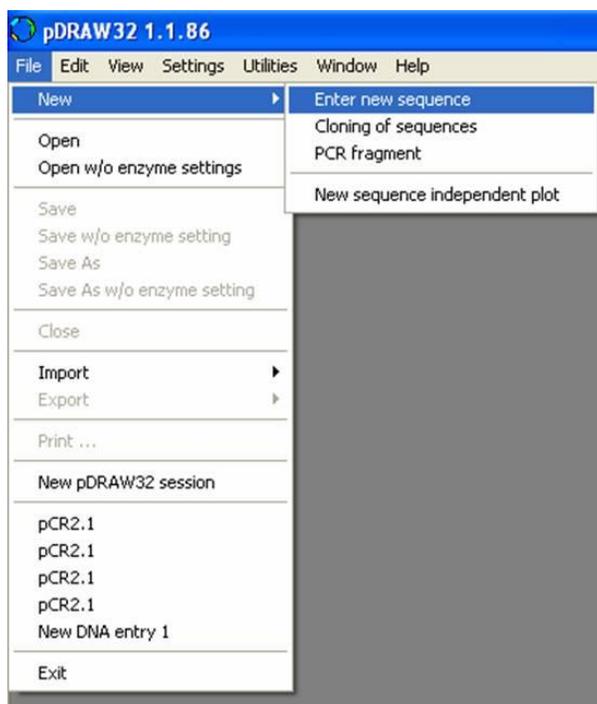


Figura 7: Tela inicial do software pDRAW32.

Fonte: Acaclone software



Figura 8:Caixa de clonagem de nova sequência de plasmídeo.

Fonte: Acaclone software

Na sequência do processo de manipulação de clonagem, a Figura 9 ilustra o DNA linear, correspondente à caixa de diálogo apresentada na Figura 8. Além disso, a Figura 9 ilustra as informações que o usuário insere para a clonagem realizada. Nesta caixa de diálogo é definida a circularização do plasmídeo.



Figura 9: Ssequência de plasmídeo linearizada .

A Figura 10 mostra o plasmídeo da Figura 9, porém circularizado , nomeado, com seu comprimento identificado e com os sítios de restrição para o conjunto de enzimas do banco de dados de enzimas de restrição.

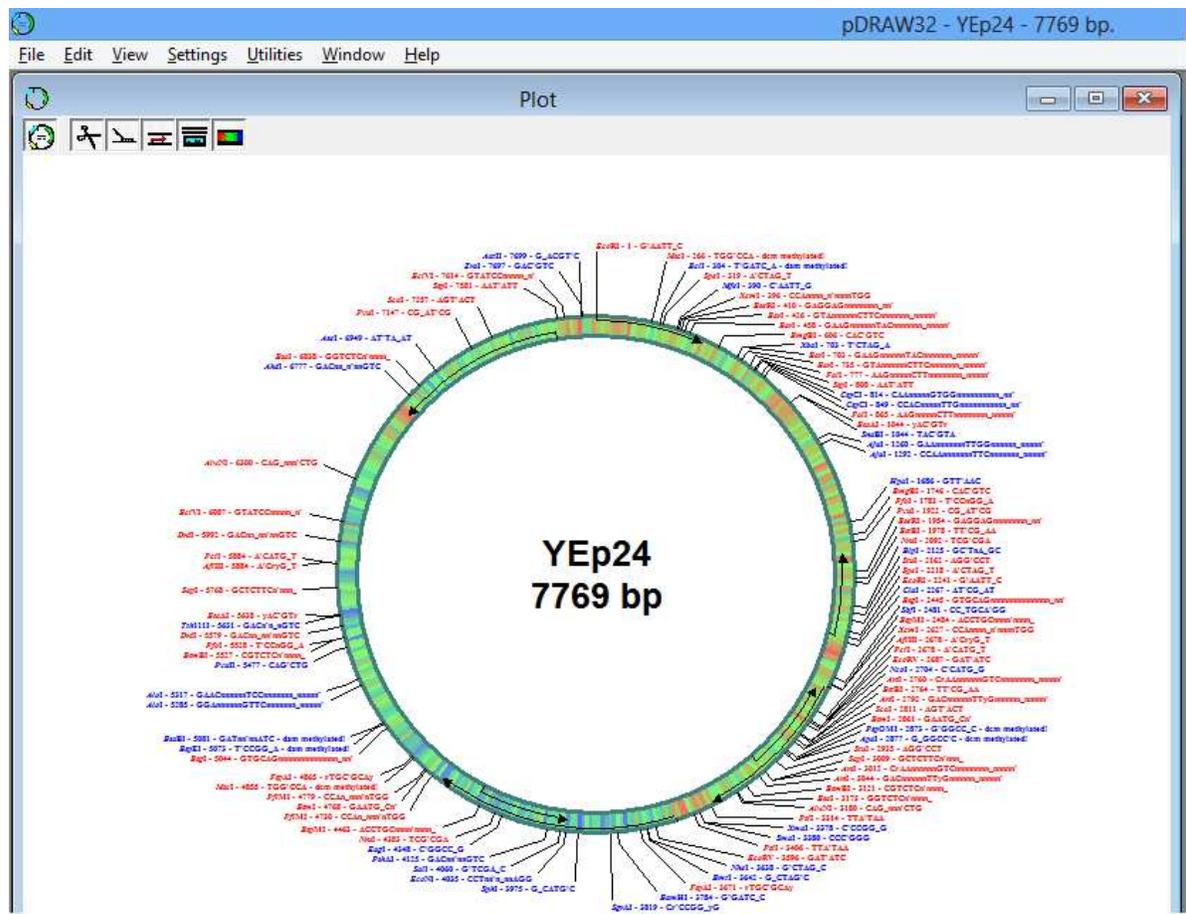


Figura 10: Plasmídeo YEp24, circularizado

Quando clonamos um inserto num vetor não podemos abrir o vetor em mais de 1 ponto. Então, podemos estudar que enzimas cortam o plasmídeo em apenas 1 ponto. Para tal usamos a caixa de diálogo disponibilizada em *Settings/ Enzyme selection* , mostrada na Figura 11 . Em seguida, na Figura 12, é ilustrado o plasmídeo após a manipulação. Tal manipulação é feita editando a sequência do plasmídeo, ou seja, na mesma caixa de diálogo da Figura 8.

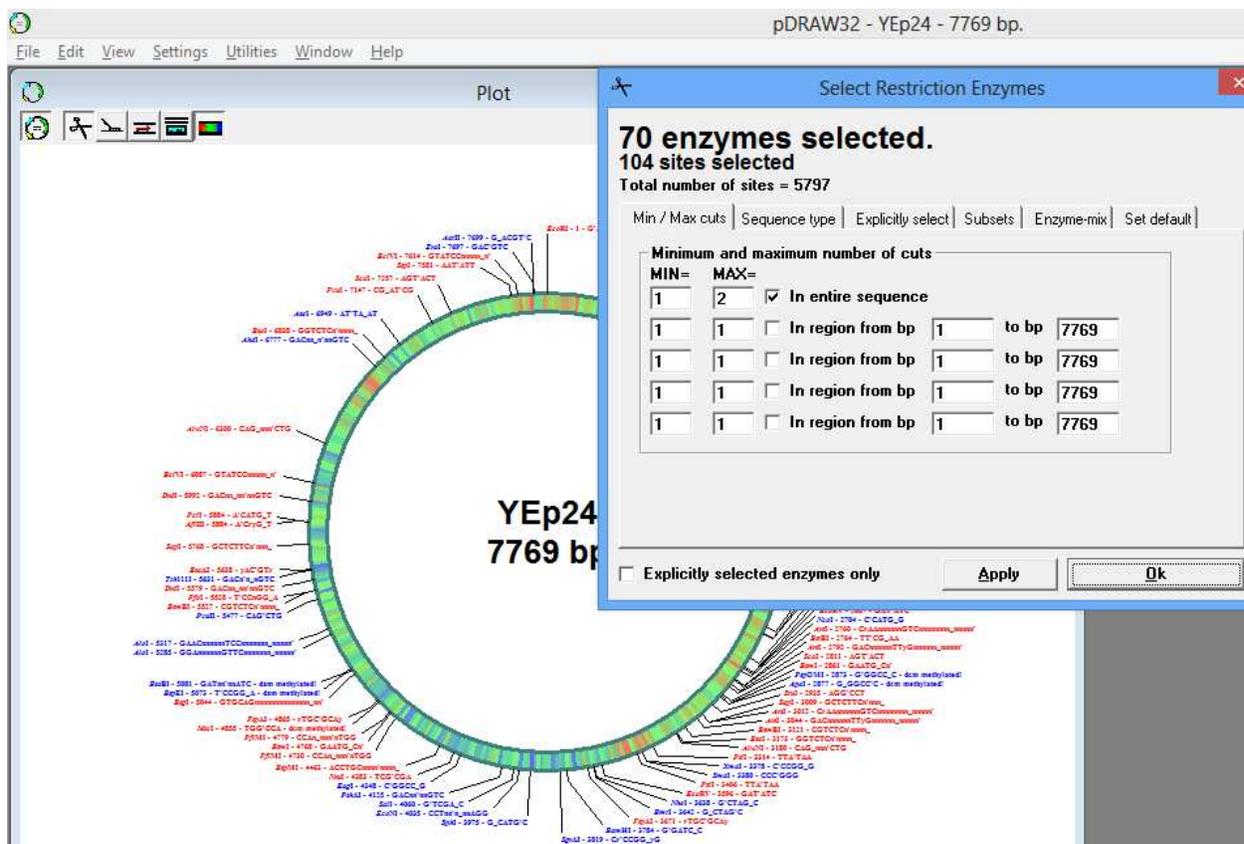


Figura 11: Caixa de diálogo para seleção de enzimas.

Na Figura 11, a caixa de diálogo para seleção em que o usuário pode escolher o número mínimo e máximo de cortes para a sequência toda ou para trechos dela (até 4 trechos). Observe que a caixa tem várias abas e a figura está mostrando apenas a primeira delas.

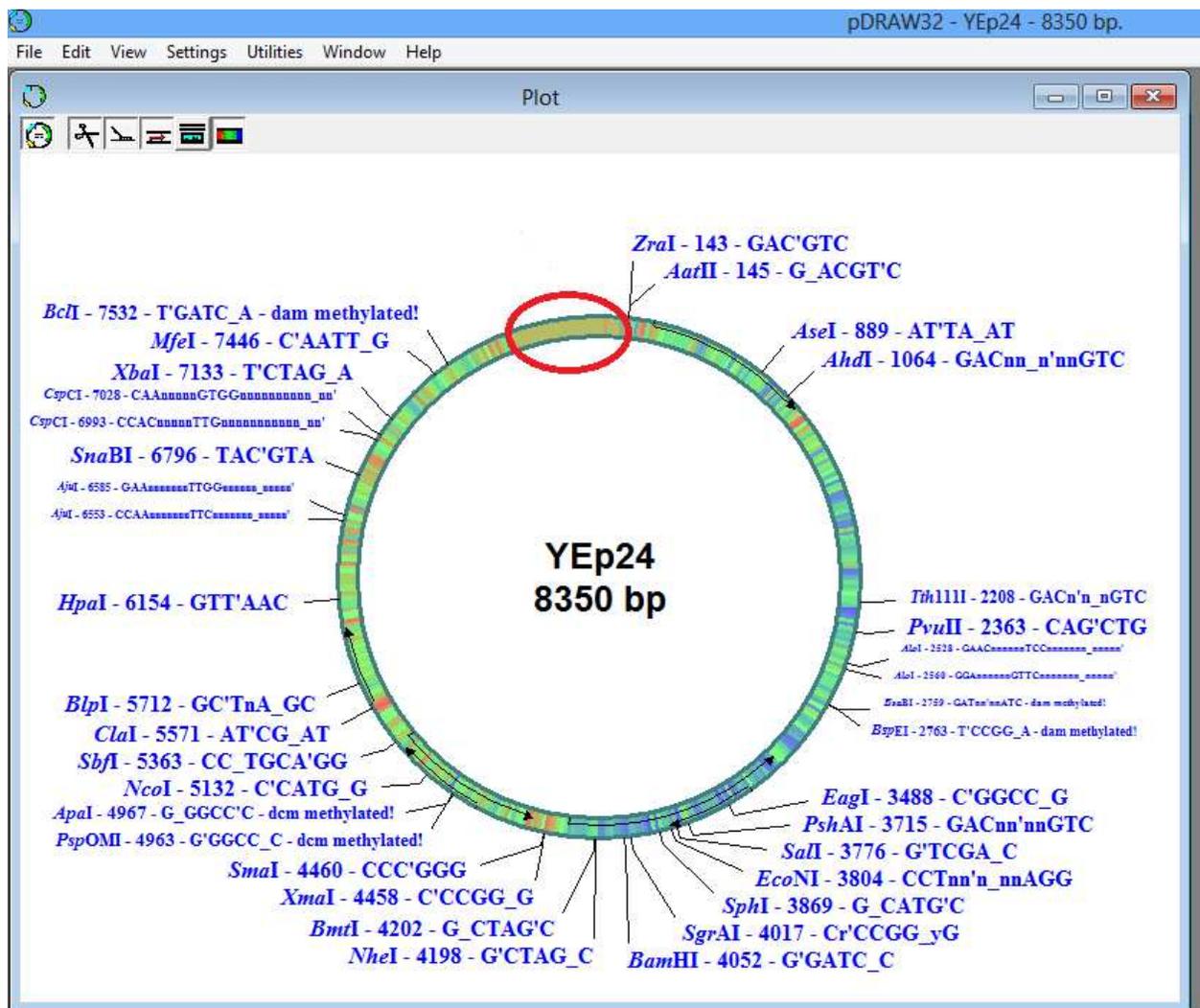


Figura 12: Plasmídeo YEp24 após a manipulação. o círculo vermelho indica o segmento inserido no plasmídeo

Utilizando opções mais avançadas, temos o mesmo plasmídeo apresentado na Figura 13, onde também são identificados locais de restrição para permitir a clivagem da sequência específica por endonuclease de restrição específica. Os sítios de restrição no vetor YEp24 de clonagem incluem *HindIII*, *EcoRI*, *BamHI*, *Sall*, *PvuI*, *PstI*, *ClaI*, etc.

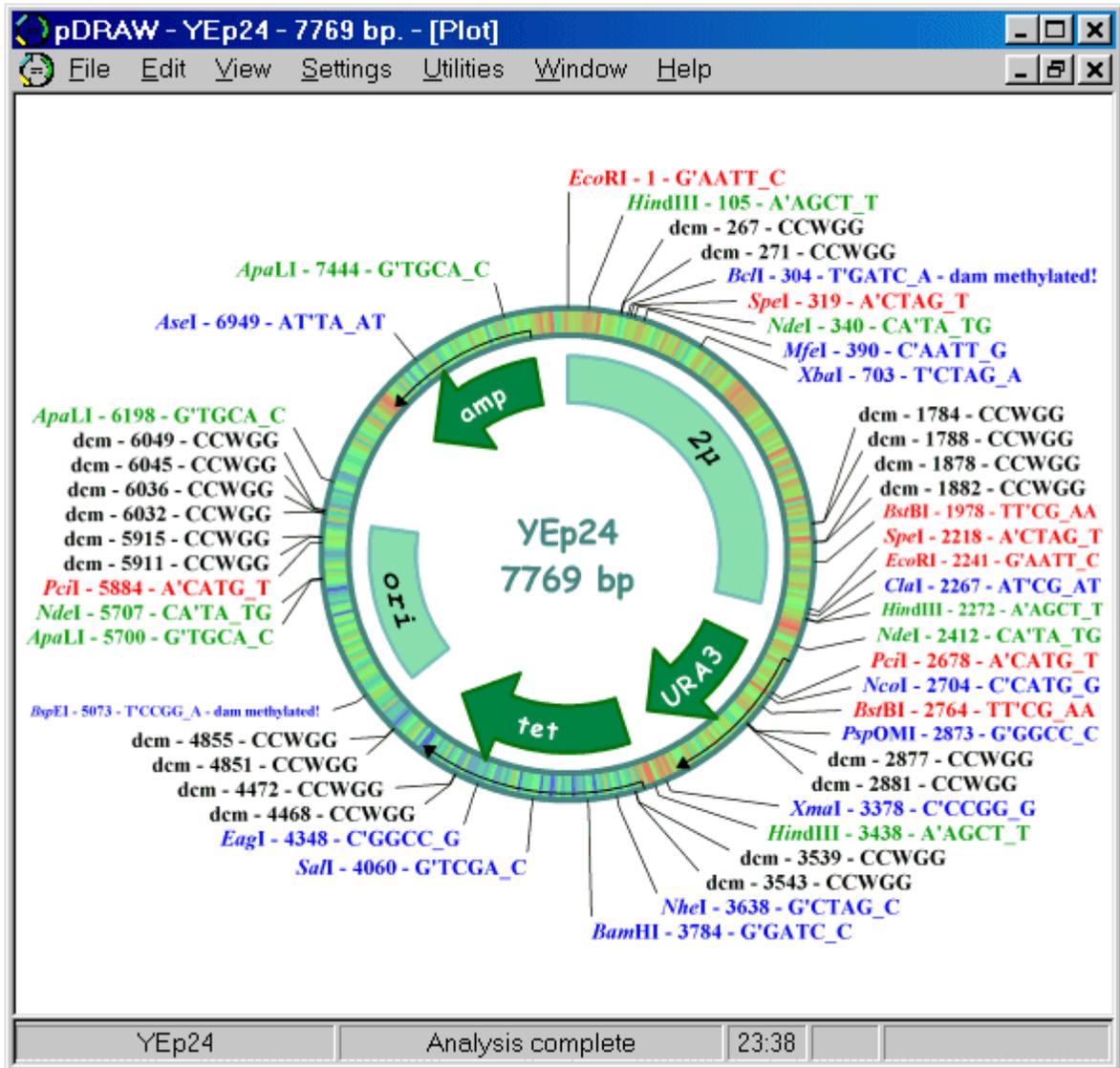


Figura 13: Gráfico circular do plasmídeo YEp24 de tamanho 7769bp.
Fonte: Acaclone software

2.3.2 PlasMapper

PlasMapper⁶ é um software de plataforma *web* de licença *freeware*, que gera e anota mapas de plasmídeos circulares de alta qualidade, desenvolvido pela Universidade de Alberta, no Canadá. Tomando a sequência de DNA do plasmídeo como entrada, PlasMapper utiliza o padrão de sequência de correspondência e alinhamento BLAST⁷ para identificar e anotar promotores, terminadores, locais de clonagem, sítios de restrição, genes repórter, genes marcadores selecionáveis, origens de replicação etc.

Em seguida, PlasMapper apresenta as características identificadas na forma textual com alta resolução e saída gráfica multicolorida. O software é composto de três partes: uma interface *web front-end* (gerado usando linguagem de programação Java), um *back-end* para renderização e correspondência de sequência (gerado usando as linguagens de programação Java e C) e um recurso de banco de dados.

Em sua página principal, o PlasMapper fornece uma interface gráfica do usuário que permite aos usuários tanto importar uma sequência de seu computador local quanto colar uma sequência em uma caixa de entrada para a geração de mapa do plasmídeo. Esta página também inclui muitas opções que permitem ao usuário personalizar a saída, conforme mostrado na Figura 14. Basicamente, a partir do momento em que o usuário escolhe o arquivo de plasmídeo ou insere a sequência do mesmo, o plasmídeo é plotado na tela, visto na Figura 15.

6 PlasMapper está disponível em: <http://wishart.biology.ualberta.ca/PlasMapper>

7 BLAST : do inglês *Basic Local Alignment Search Tool*. É um algoritmo para comparar informações de sequências biológicas primárias.



PlasMapper Version 2.0

The PlasMapper server automatically generates and annotates plasmid maps using only the plasmid DNA sequence as input. Plasmid sequences up to 20,000 bp may be annotated and displayed. Plasmid figures may be rendered in PNG, JPG, SVG or SVGZ format. PlasMapper supports an extensive array of display options.

[Download PlasMapper Source](#)

Please cite the following: [Xiaoli Dong, Paul Stothard, Ian J. Forsythe, and David S. Wishart "PlasMapper: a web server for drawing and auto-annotating plasmid maps" Nucleic Acids Res. 2004 Jul 1;32\(Web Server issue\):W660-4.](#)

For additional information on how to run PlasMapper, click

Select desired DNA sequence file (FASTA format only)

Nenhum arquivo selecionado

OR Select a plasmid based on supplier

OR paste the DNA sequence into the text window below (FASTA format only)

```
tgcataattctcttactgtcatgoccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaa ^
tagtgtatgoggcgaccgagttgctcttgccggcggtcaacacgggataataccgcgccacatagcagaactttaaaagt
gctcatcattggaaaaagcttcttggggcgaaaaactctcaaggatcttaaccgctgttgagatccagttcgatgtaacca
ctogtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgcc
gcaaaaaaggggaataagggcgacacggaaatggtgaatactcatactcttctctttttcaatattattgaagcatttatca
gggttattgtctcatgagcggatacatatgtgaatgtatttagaaaaataaacaataggggttccggcgacatttcccc
gaaaagtgcaccctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctt
ctcttcaa
```

Figura 14: Página inicial da ferramenta PlasMapper.
Fonte: <http://wishart.biology.ualberta.ca/PlasMapper/>

O software gera o gráfico do plasmídeo escolhido conforme Figura 15, gerada para o plasmídeo YEp24.

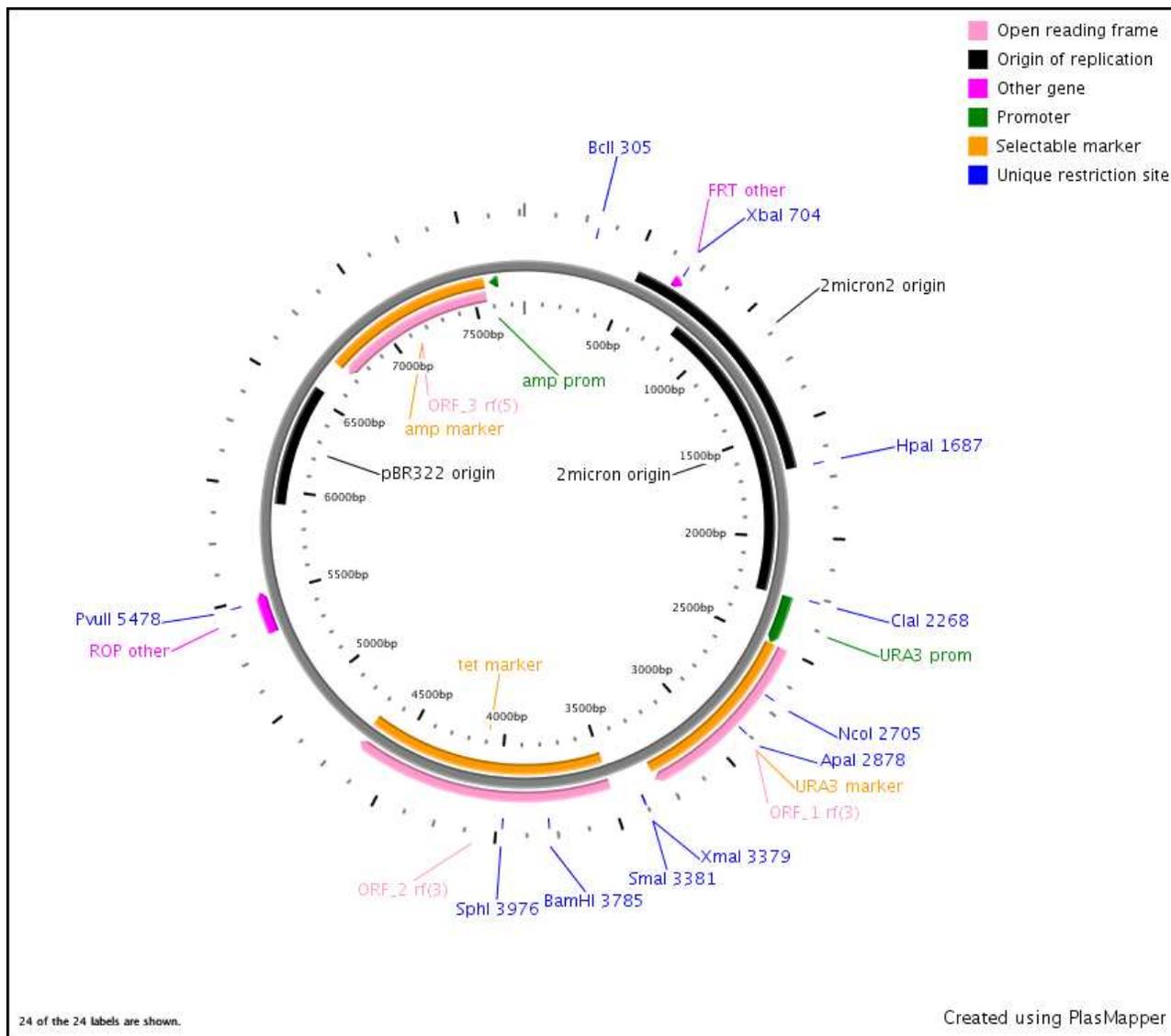


Figura 15: Gráfico circular do plasmídeo YEp24 gerado pela ferramenta PlasMapper.

Fonte:

http://wishart.biology.ualberta.ca/PlasMapper/jsp/displayPlasmidMap.jsp?fileName=plasMap204_1398200506374.png&fileFormat=png

3 Materiais e métodos

A aplicação Biodonut consiste em um software desenvolvido para ser executado em navegadores na Internet, foi concebida a partir das tecnologias, *frameworks*, ferramentas e componentes mencionados a seguir, como recursos computacionais.

3.1 Diagrama de caso de uso do sistema

Apresentamos nesta seção também o diagrama de caso de uso com o propósito de visualização das funcionalidades e fluxos do sistema, sob a ótica do usuário. Um diagrama de caso de uso deve descrever a funcionalidade específica que um sistema deve desempenhar ou exibir (PFLEEGER, 2004). Na Figura 16 apresentamos as funcionalidades do Biodonut, sendo que cada elipse representa uma ação realizada pelo ator seja usuário comum ou usuário administrativo. A linha pontilhada representa uma associação do tipo *extend*, ou seja, um comportamento opcional. A linha cheia representa uma associação obrigatória.

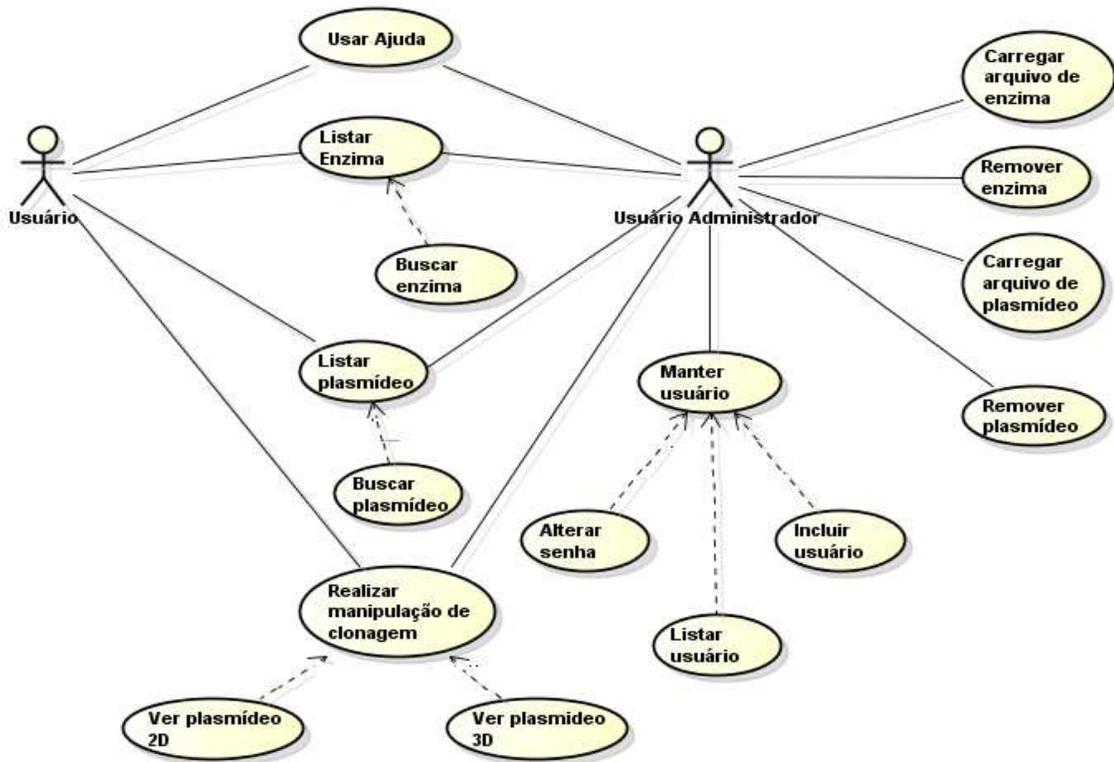


Figura 16: Diagrama de Caso de Uso do software Biodonut

3.2 Bases de dados

NCBI⁸ (*National Center for Biotechnology Information*) é mundialmente considerado como banco de dados central sobre informações genômicas. Diversos outros bancos de dados similares estão distribuídos em outros países, embora todos se comuniquem diariamente com o NCBI. Esta base de dados reúne trabalhos acadêmicos de todos os países, com artigos em diversos idiomas, além de ser pública e tradicional, criada em 1988, sendo mantida pelo NIH (*National Institutes of Health*). As quatro bases mais popularmente utilizadas do NCBI são:

- **GenBank** : reúne todo conhecimento público sobre a sequência de DNA que são enviadas por cientistas de todo o mundo e por grande parte dos centros envolvidos no projeto genoma. É o principal banco de dados do NCBI, e foi escolhido para o desenvolvimento deste trabalho por ser uma das bases mais importantes para pesquisa científica.
- **OMIM** : catálogo que contém milhares de genes e trabalhos sobre doenças genéticas e fenotípicas e serve como uma fonte para o Projeto Genoma Humano.
- **PubMed**: um dos mais famosos e conhecidos banco de dados científico no Brasil oferece acesso a mais de 11 milhões de citações, resumos e indexação de termos dos artigos em revistas de ciências biomédicas. Também inclui *links* para o texto completo revistas.
- **PubMed Central (PMC)** : banco de dados que possibilita a pesquisa bibliográfica em mais de 17 milhões de referências de artigos médicos publicados em cerca de 3.800 revistas científicas . O valor do PubMed Central, para além do seu papel como um arquivo, reside em o que pode ser feito quando os dados provenientes de diversas fontes são armazenadas em um formato comum, em um único repositório. PMC fornece atualmente livre e irrestrito acesso ao texto integral de diversas revistas das ciências da vida. O acesso é feito através do site do NCBI, disponível em <http://www.ncbi.nlm.nih.gov/>

⁸ NCB, disponível em: <<http://www.ncbi.nlm.nih.gov/>>

3.2 Recursos computacionais

Serão listados e explicados sucintamente aqui os recursos computacionais utilizados para o desenvolvimento da aplicação.

A IDE (*Integrated Development Environment*) escolhida foi Eclipse. A versão utilizada para a elaboração da aplicação Biodonut foi a 3.6, denominada Eclipse Helios. A plataforma *Java Standard Edition* (JSE), baseada em linguagem Java, foi escolhida como linguagem principal para a realização do trabalho em virtude de possibilitar portabilidade. Foi utilizado para o trabalho a versão 1.5 do Kit de Desenvolvimento de *Java Standard Edition* (Java SE SDK). Qualquer arquitetura computacional provida de JVM (do inglês *Java Virtual Machine*) instalada na versão 1.5 ou superior está apta para comportar a aplicação. Para poder construir e rodar aplicações distribuídas dentro da plataforma Java, é necessário que tais aplicações utilizem *browsers* como clientes através de protocolos de Internet (TCP/IP) ou outros protocolos proprietários de rede. A aplicação deve fazer uso da plataforma Java EE (*Enterprise Edition*). A versão utilizada foi o Java EE 6, que possui como características e especificações importantes o suporte a serviços *web*, injeção de dependências, uso facilitado de EJB (*Enterprise Java Beans*), entre outros.

Para o desenvolvimento da aplicação proposta, optou-se pela versão 7 do *apache* Tomcat, que é um servidor de aplicação Java. O critério para a escolha de um servidor de aplicação do tipo *webcontainer*, como o Tomcat, e não do tipo EJB container, como *Glassfish*, *Jboss* e *Weblogic* para executar e hospedar a aplicação Biodonut residiu nos seguintes fatos:

- Os componentes distribuídos rodam exclusivamente na Internet;
- Ausência de EJB (aplicação distribuída) simplifica o desenvolvimento e aperfeiçoa o desempenho no carregamento (*loading*) do conteúdo das páginas da aplicação;
- Menor tempo despendido ao se fazer a distribuição (*deploy*) da aplicação.

O *apache* Maven foi utilizado na versão 3.0.4. Com o uso do Maven, gerar os executáveis da aplicação na plataforma Java passa a ser uma tarefa de pouca complexidade, permitindo ao programador focar seu tempo em tarefas mais produtivas, e não em gerar o executável do software.

O mapeamento objeto-relacional foi feito com o uso de JPA, acrônimo do inglês *Java Persistence API*.

WebGL (do inglês *Web Graphics Library*) é uma API desenvolvida na linguagem de programação JavaScript, disponível a partir do novo elemento *canvas* (camadas) da HTML5, que oferece suporte para renderização de gráficos 2D e gráficos 3D. Por este motivo, este *framework* poderoso e inovador foi escolhido para a o desenvolvimento da renderização 3D do plasmídeo no contexto do Biodonut. No contexto WebGL, foi usada a biblioteca *Three.js* que é um *cross-browser* JavaScript biblioteca / API leve usado para criar e exibir animação de computação gráfica 3D em um navegador da Web através da linguagem JavaScript. Em função do uso de WebGL, fez-se necessário também uso da tecnologia HTML5.

A arquitetura MVC – (Modelo, Visualização e Controle) fornece uma maneira fácil de dividir as funcionalidades envolvidas na manutenção e apresentação de dados de uma aplicação. Por este motivo, foi usada no desenvolvimento do projeto aqui apresentado. Para favorecer a independência entre as camadas (injeção de dependências), foi usado o *framework* Spring.

O sistema gerenciador de banco de dados usado no Biodonut é o *MySQL* por possuir código aberto, sendo o eleito na grande maioria das ferramentas gratuitas para gerir suas bases de dados. O diagrama de fluxo de desenvolvimento é dado na Figura 17.

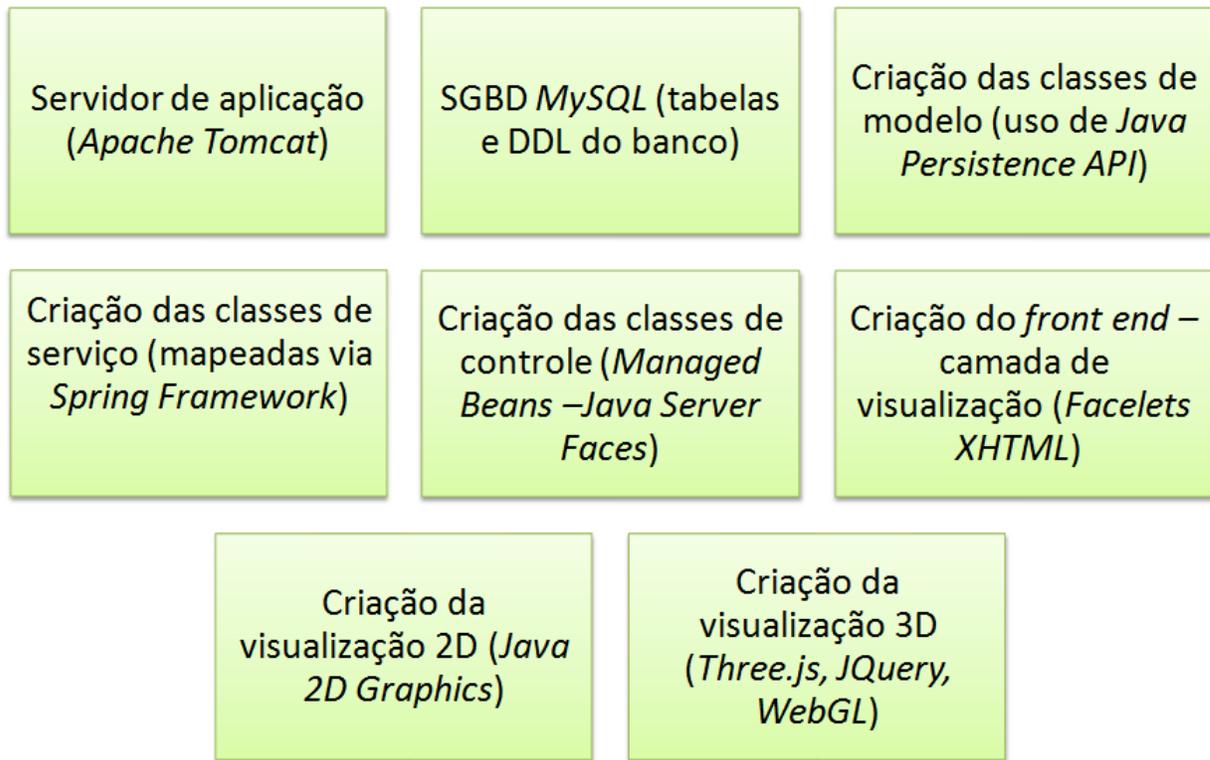


Figura 17: Diagrama de fluxo de desenvolvimento do software Biodonut

4 Experimentos e Resultados

Biodonut é um sistema de plataforma web, de licença *open source*, que possibilita simular a clonagem de plasmídeos feita em laboratório, para basicamente cadastrar plasmídeos e mapear suas regiões, considerando suas enzimas de restrição. Durante a validação do cadastro, o sistema separa cada parte da sequência de acordo com suas características, sendo: pontos de corte, DNA do plasmídeo e DNA externo (ou o que será inserido no plasmídeo). Os administradores e/ou usuários autenticados podem cadastrar plasmídeos e enzimas de corte que ficam disponíveis para acesso (consulta) público de forma que os usuários possam consultar toda a base de dados e possam utilizar estes dados para simular ensaios de clonagem.

4.1 Biodonut

O sistema Biodonut, disponível em biodonut.org, desenvolvido e aqui apresentado, apresenta a interface inicial ilustrada na Figura 18.

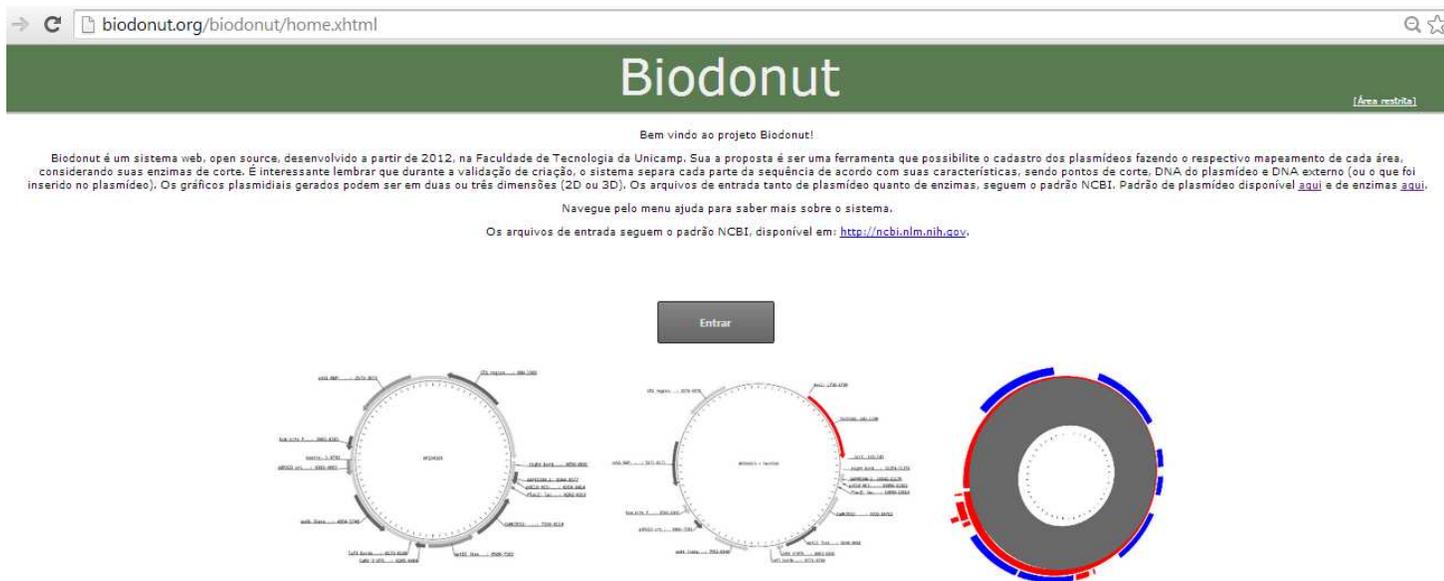


Figura 18: Página inicial do Biodonut

Ao clicar em "entrar", o usuário avança para a página com as funcionalidades que o software apresenta. A página de plasmídeo apresenta todos os plasmídeos cadastrados bem como

suas características. A coluna "ações" da tabela apresentada na Figura 19, possui dois subitens: Ver e manipular. Ao escolher a opção "ver", o usuário visualiza o plasmídeo graficamente em formato inicial 2D, podendo optar também pela sua visualização 3D, além de apresentar diversas características textuais do plasmídeo em questão, como mostrado na Figura 20.



Figura 19: Listagem dos plasmídeos presentes no Biodonut

→ C biodonut.org/biodonut/pages/plasmid/show.xhtml

Biodonut

[Área restrita]

Plasmídeo Enzima Ajuda ▾

Plasmídeo

Ver 3D Zoom

Gráfico do Plasmídeo

The circular map of plasmid AF234315 shows several key features:

- ori** (origin of replication) at 2573-3573
- ori** at 3093-4242
- source** at 1-8742
- ori** at 4393-6663
- ori** at 4954-5749
- ori** at 6173-6198
- ori** at 6265-6468
- ori** at 6906-7303
- ori** at 7334-8114
- ori** at 8056-8681
- ori** at 8344-8577
- ori** at 8758-8414
- ori** at 8262-8316
- ori** at 980-1080

Código Acesso
AF234315

Local
AF234315 8742 bp DNA circular SYN 24-APR-2000

Definição
Binary vector pCAMBIA-2300, complete sequence.

Versão
AF234315.1 GI:7638145

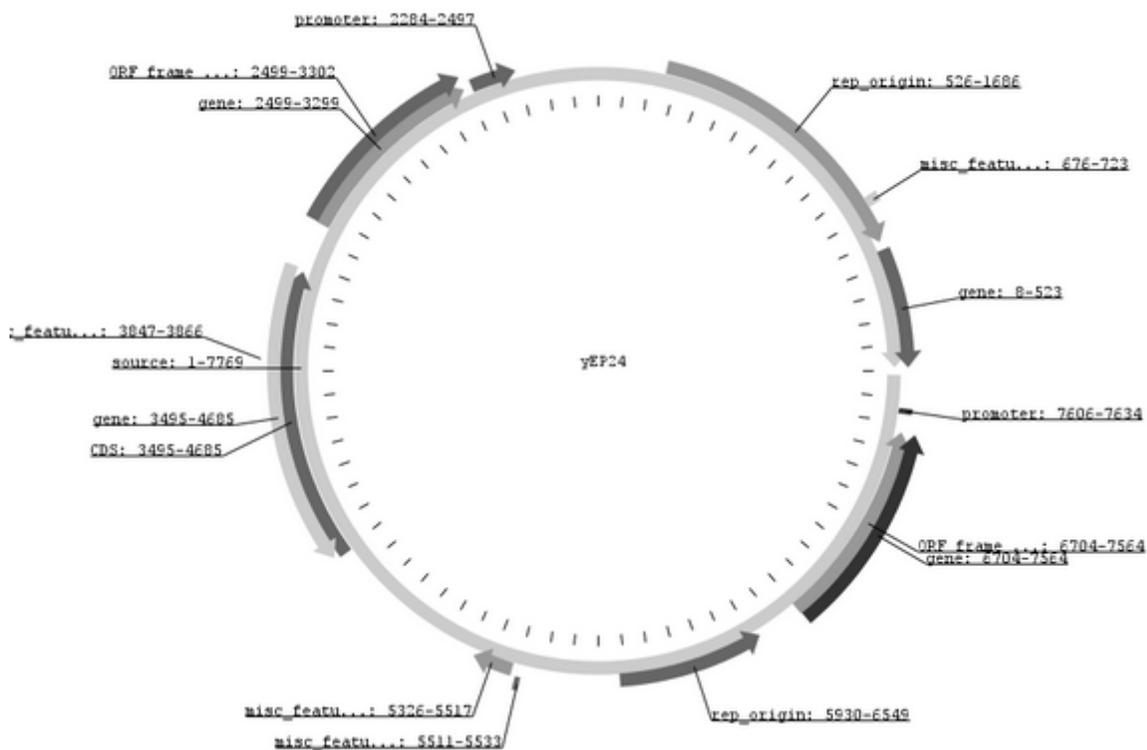
Palavras chave
.

Figura 20: Tela para visualização de plasmídeo

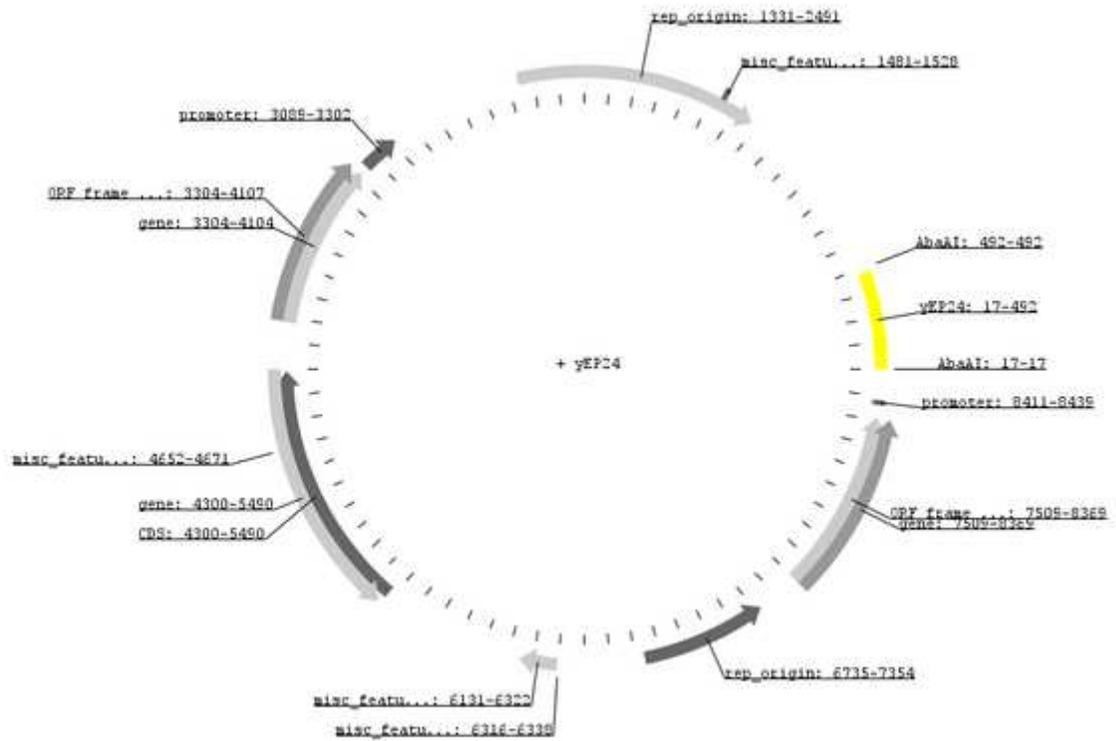
Na área aberta do *software* Biodonut, o usuário comum poderá: escolher o plasmídeo que está utilizando; identificar os pontos de corte (ou sítios de restrição) de início e fim para inclusão de nova sequência e realizar a manipulação de clonagem com visualização 2D e 3D. Na validação de criação, o sistema separa cada parte da sequência de acordo com suas características, sendo sítios de corte, DNA do plasmídeo e DNA externo (ou o que foi inserido no plasmídeo). O usuário deve informar o nome da sequência desejada e inserir no plasmídeo para que esta seja identificada tanto no plasmídeo quanto na listagem de características plasmidiais após a conclusão do ensaio. Este processo é visualizado na Figura 21, onde o usuário insere a sequência desejada e o sistema calcula quais enzimas de corte são compatíveis.

Para fazer essa análise, o software busca todas as enzimas de corte que fazem parte da

ou neutra (sem seta). Na Figura 22 (b) notamos que a sequência informada pelo usuário ficará destacada na cor vermelha (quando sentido anti-horário), azul (Quando sentido horário) ou amarela (quando neutra), e no centro do plasmídeo será exibido o nome do plasmídeo somado ao nome da sequência informada. As enzimas de corte utilizadas também serão identificadas.



(a)

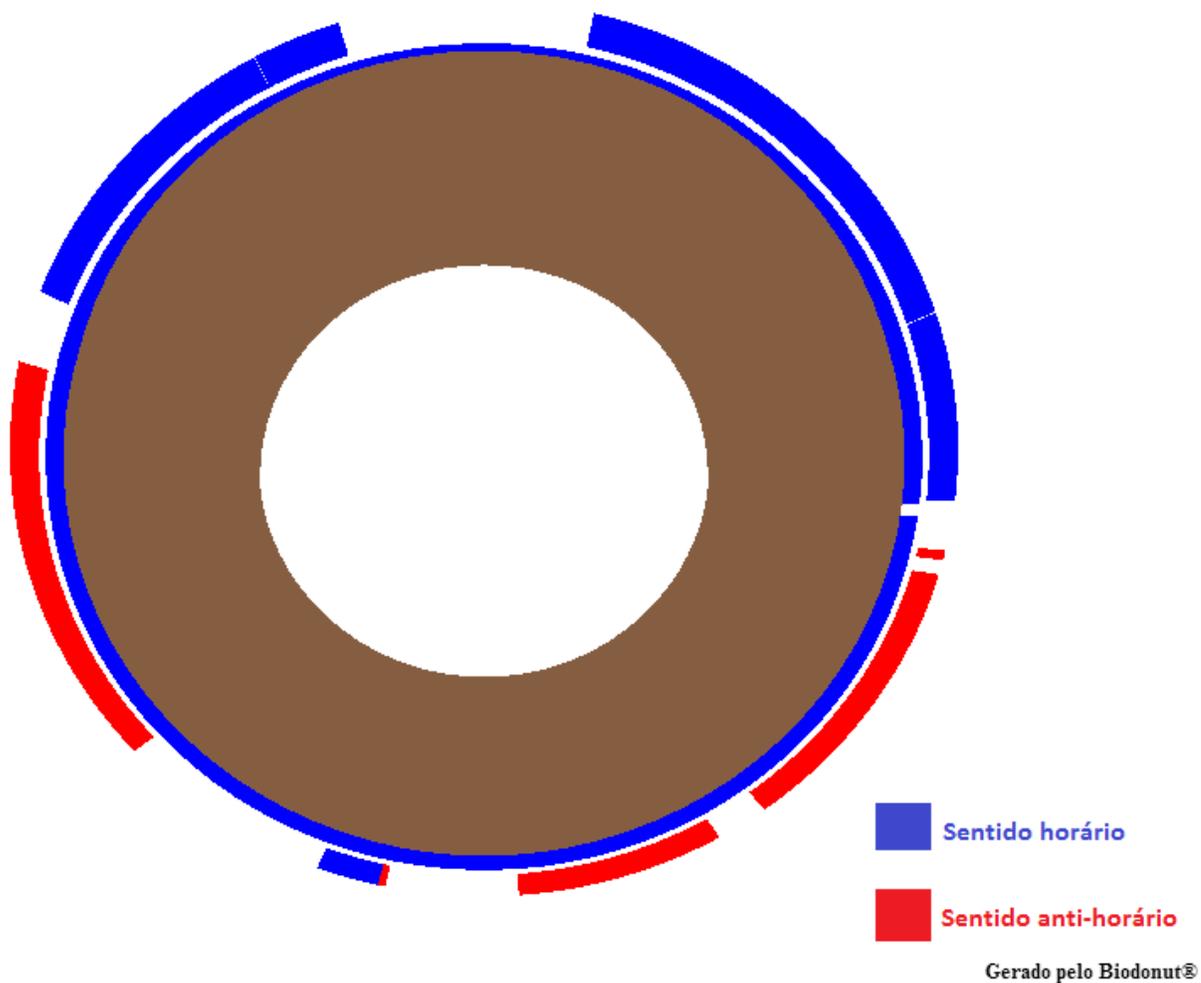


(b)

Figura 22: Representação gráfica 2D do plasmídeo YEp24 antes e após a manipulação gerada pelo Biodonut: (a) plasmídeo em original, sem alterações; (b) apresenta o plasmídeo modificado, onde após a manipulação virtual. Biodonut marca a inserção em conjunto com as enzimas utilizadas para a clonagem.

4.2 Vizualização 3D

A representação gráfica 3D do plasmídeo antes da manipulação é renderizada na tela do Bionnut , conforme vimos na Figura 23 (a). Nesta imagem, o sentido do inserto no plasmídeo YEp24 com sentido horário ou anti-horário são representados por cores azul (horário) e vermelha (anti-horário). que na Figura 23 (b) ilustramos o plasmídeo após o inserto. O plasmídeo é ilustrado pela forma geométrica do tipo Torus, conforme já citado, tendo na versão 3D suas features abstraídas. A orientação da sequência é dada pelas cores de acordo com a legenda.



(a)

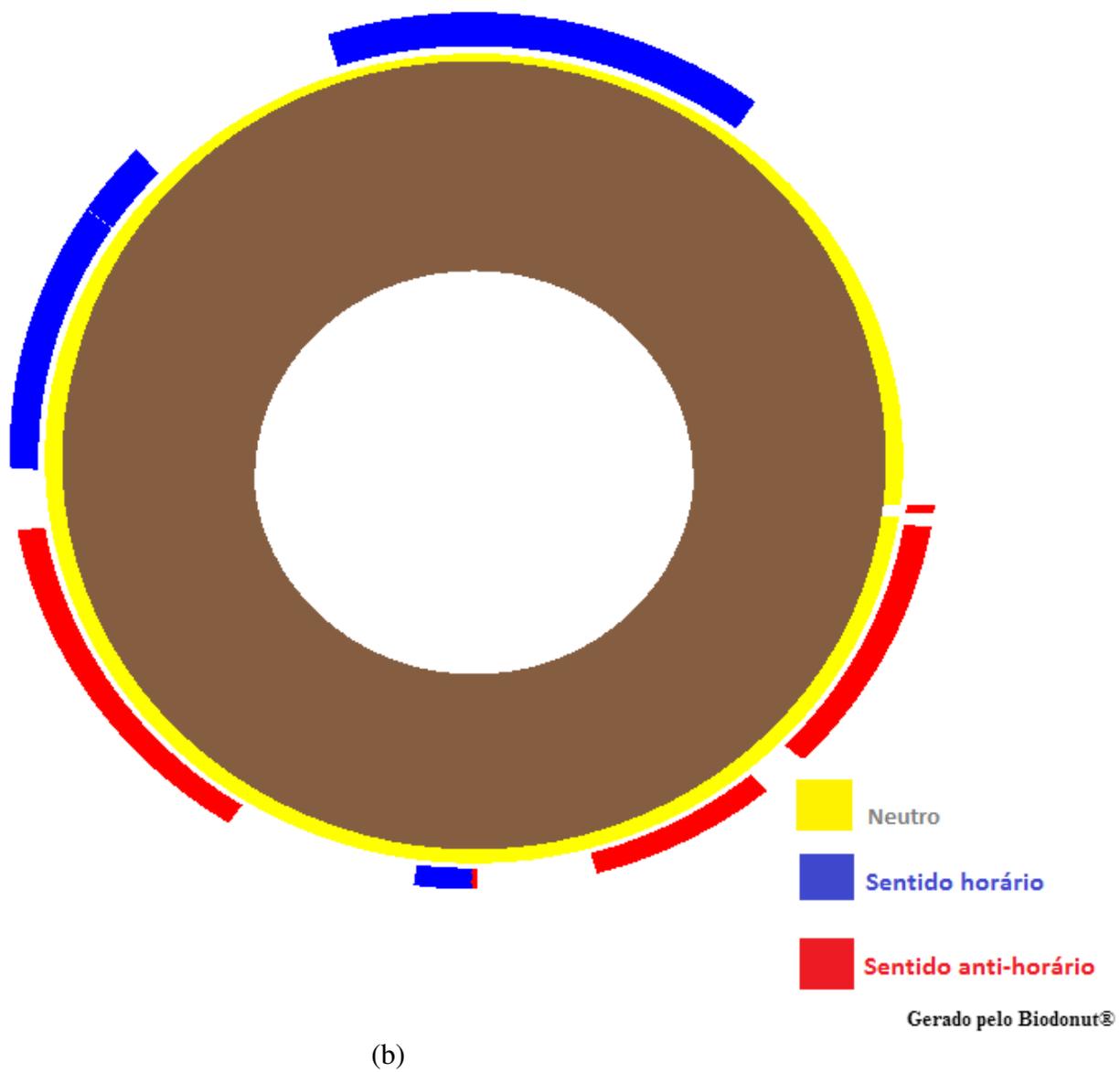


Figura 23: Visualização 3D do plasmídeo YEp24 antes (a) e depois (b) da manipulação.

4.3 Estudo comparativo

O objetivo desta seção é comparar os softwares apresentados sob a ótica de métricas de interface humano computador. Nas métricas de avaliação de software, foram utilizadas as seguintes etapas: instalação e processo de clonagem avaliando sempre a usabilidade do software, eficácia, eficiência e aprendizado. Como base de dados, o GenBank foi escolhido.

Até o momento não foi realizado nenhum estudo comparativo entre softwares de bioinformática geradores de gráfico de plasmídeos. Hu, Mellor, Wu e DeLisi (2004), compararam as ferramentas VisANT, Cytoscape e Osprey segundo os critérios de manipulação gráfica, entrada, saída, pesquisa, operações de rede, filtros, integração, licença e modo de execução. Tais ferramentas estão relacionadas à visualização de redes de interação molecular e vias biológicas para integrar essas redes com perfis de expressão gênica e outros dados. Este estudo foi considerado no desenvolvimento deste trabalho.

Foi também o estudo comparativo de Suderman e Hallet (2007) sobre as ferramentas para explorar visualmente redes biológicas. Este estudo pertence à grande área de bioinformática e algumas das ferramentas avaliadas foram também VisANT e Cytoscape. Neste trabalho, Suderman e Hallet forneceram uma ampla revisão das vantagens e desvantagens dos sistemas existentes com dois objetivos: ajudar os pesquisadores a identificar de forma eficiente as ferramentas existentes para visualização de dados a fim de descrever as metas necessárias e realistas para a próxima geração de ferramentas de visualização. Suderman e Hallet (2007) consideraram que a utilidade de qualquer ferramenta de visualização da rede biológica em última análise, depende dos recursos suportados. Assim, avaliaram algumas das características mais importantes, incluindo rotinas de layout da rede, a integração de análise para a visualização e a variedade de métodos de entrada do usuário, integração de diferentes fontes externas de dados biológicos.

Para o estudo comparativo entre as ferramentas Plasmapper, pDRAW e Biodonut, técnicas de usabilidade já apresentadas na seção 2 foram consideradas. As metas de usabilidade servem para entender as necessidades do usuário e buscar clareza do objetivo principal. Neste trabalho, iremos usar as seguintes metas de usabilidade como critérios comparativos dos softwares apresentados:

- Instalação: Existência ou não de instalação e quando há, qual o nível de complexidade no processo.
- Eficácia: Atingir o objetivo. Avaliação se o sistema serve ao propósito principal.
- Eficiência: Atingir o objetivo com o menor número de operações.
- Facilidade de uso: Nota atribuída ao sentimento do usuário quanto à facilidade de uso da ferramenta.

O processo de instalação foi incluído na avaliação, justamente por estarmos comparando softwares *web* e *desktop*. Uma nota final foi atribuída a cada um dos critérios através da soma dos valores associados mostrados nas Tabelas 1, 2, 3 e 4.

Tabela 1: Conceito comparativo segundo critério instalação

Instalação		
Conceito quantitativo	Conceito qualitativo	Definição
0	Extremamente complexa	Quando há necessidade de download de mais de um arquivo e o usuário ter que realizar aspectos de configuração de servidores de banco de dados e aplicação.
1-3	Complexa	Quando há necessidade de download de mais de um arquivo e o processo de instalação é pouco intuitivo.
4-6	Mediamente assistida	Quando não há necessidade de instalação de mais de um arquivo mas o processo de instalação é pouco intuitivo
7-9	Assistida	Quando há necessidade de download de apenas um arquivo e o processo é intuitivo.
10	Sem instalação	Não possui instalação.

Tabela 2: Conceito comparativo segundo critério eficácia

Eficácia		
Conceito quantitativo	Conceito qualitativo	Descrição
0	Extremamente Ineficaz	Não oferece funcionalidade de clonagem
1-3	Ineficaz	Oferece funcionalidade de clonagem sem opção de comparativos.
4-6	Mediamente eficaz	Oferece a funcionalidade de clonagem com opção de comparativos em interfaces diferentes.
7-9	Eficaz	Oferece comparativo antes e após a clonagem
10	Extremamente eficaz	Oferece comparativo antes e após a clonagem aliada a visualização 3D.

Tabela 3: Conceito comparativo segundo critério eficiência

Eficiência		
Conceito quantitativo	Conceito qualitativo	Descrição
0	Extremamente ineficiente	Etapas de clonagem $> = 20$
1-3	Ineficiente	$19 \leq$ Etapas de clonagem ≤ 15
4-6	Mediamente eficiente	$14 \leq$ Etapas de clonagem ≤ 10
7-9	Eficiente	$9 \leq$ Etapas de clonagem ≤ 7
10	Extremamente eficiente	Etapas de clonagem ≤ 6

Um grupo de 10 pessoas testou os softwares pDRAW, Plasmapper e Bionut para mensurar cada ferramenta conforme o fluxograma apresentado na Figura 26. Os testes foram realizados no Laboratório de Computação Visual IMAGELab da Faculdade de Tecnologia da Unicamp. Este grupo foi composto por:

- 1 professor universitário especialista em engenharia de software, rotulado como P1
- 1 cientista da computação, rotulado como P2
- 1 professor doutor em engenharia elétrica atuante, rotulado como P3
- 1 mestrando em tecnologia, rotulado como P4
- 3 bacharéis em ciências biológicas, rotulado como P5, P6 e P7 respectivamente
- 1 tecnólogo em processamento de dados, rotulado como P8
- 1 doutorando em engenharia biomédica, rotulado como P9
- 1 graduando em ciência da computação rotulado como P10

Tabela 4: Conceito comparativo segundo critério facilidade de uso

Conceito quantitativo	Conceito qualitativo	Descrição
0	Extremamente dedutivo	O próprio conceito quantitativo torna-se a nota.
1-3	Dedutivo	
4-6	Mediamente intuitivo	
7-9	Intuitivo	
10	Extremamente intuitivo	

As notas atribuídas por cada pessoa do grupo de testadores são apresentadas na Tabela 1. Cada testador é considerado na tabela como P_n , onde n representa um número de 1 até 10. Na tabela, podemos notar claramente que Bionut obteve nota superior às outras ferramentas por unanimidade. Notamos também que Bionut recebeu dos dez testadores a maior média dentre as três ferramentas no quesito facilidade de uso.

Tabela 5: Avaliação quantitativa atribuída pelo grupo de testadores para as ferramentas propostas

	P 1	P 2	P 3	P 4	P 5	P 6	P 7	P 8	P 9	P 10	Média
Biodonut	9	9	9,5	10	8,5	9	10	9,5	8,5	9	9,2
pDRAW	8	7	9	8,5	6	7	7	6,5	7,5	8	7,45
Plasmapper	8,5	7,5	8	9	8	8	7	7	7,5	8,5	7,9
Média/pessoa	8,5	7,8	8,8	9,2	7,5	8	8	7,7	7,8	8,5	

Para representar graficamente a Tabela 5, o Gráfico 1 foi criado. No gráfico, temos as notas atribuídas por cada pessoa participante do grupo, bem como a média aritmética para cada ferramenta, onde cada cor representa uma pessoa do grupo.

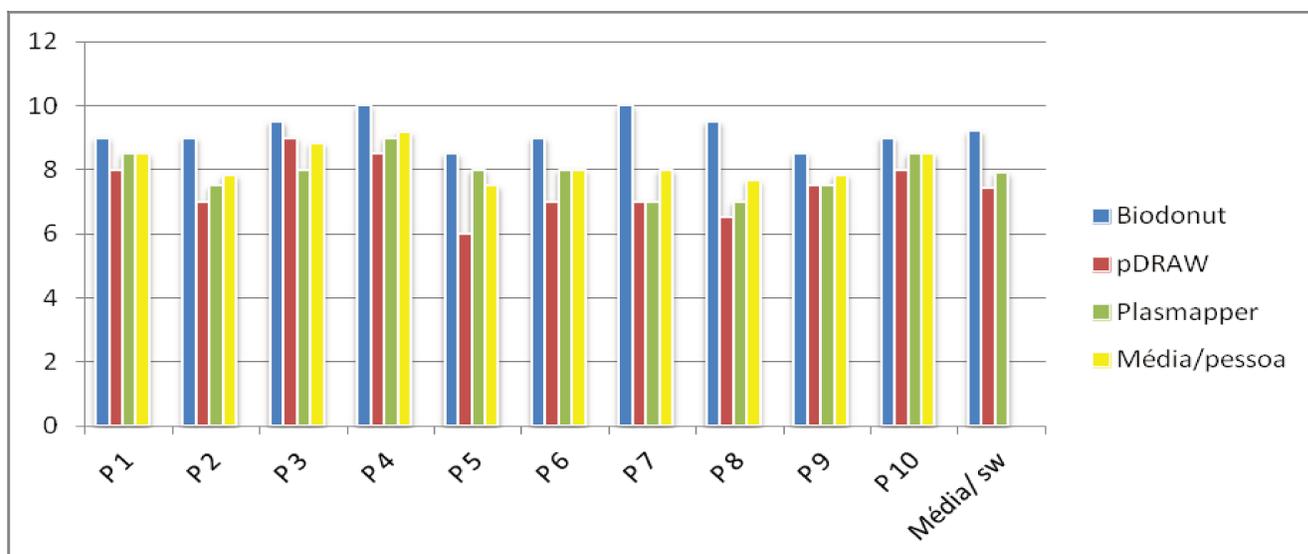


Gráfico 1: Histograma de critério de usabilidade - facilidade de uso

Apresentados os critérios estabelecidos para o estudo comparativo, os avaliadores classificaram as ferramentas como:

- pDRAW32 apresenta uma instalação complexa, sendo diferente de acordo com cada sistema operacional. Possui eficácia parcial pela ausência de um comparativo antes e depois do processo e sua eficiência é parcial em função da quantidade de etapas necessárias para atingir o objetivo. O aprendizado foi considerado mediano pois o

software pode induzir o usuário a dúvida em certas situações, pelo excesso de *menus* disponíveis na mesma tela. Uma interface mais simples/ objetiva seria uma forma de melhoria.

- Biodonut é um software de plataforma *web*, não necessitando de instalação. Possui alta eficácia e eficiência. o aprendizado foi considerado alto, pois o grupo de testadores não identificou complicações em seu uso, não deixando o usuário em dúvida quanto aos passos subsequentes.
- PlasMapper assim como Biodonut é um software de plataforma *web*, não necessitando de instalação. Sua eficácia foi considerada parcial sendo classificada com eficiência mediana, visto que não há uma tela comparativa dos plasmídeos antes e depois do inserto. O aprendizado foi considerado alto, pois não há complicações em seu uso, não deixando o usuário em dúvida quanto aos próximos passos.

5 Conclusões e trabalhos futuros

5.1 Conclusões

Este documento apresentou a proposta para o desenvolvimento do sistema Biodonut, que funciona em plataforma *web*, licença *open source* com visualização 2D e 3D para manipulação de clonagem em plasmídeos. Foram apresentados trabalhos relacionados como Plasmapper e pDRAW32, que embora já estejam mais consolidados, não possuem a opção de visualização 3D ou comparativo entre os plasmídeos antes e depois do processo de manipulação de clonagem. Realizados os testes, concluímos que o software Biodonut dentre os comparados destacou-se segundo critérios de usabilidade.

Feitos os testes com um grupo de usuários, constatou-se quantitativamente que sua interface é de simples entendimento garantindo que o usuário consiga concluir a manipulação de plasmídeo sem maiores dúvidas, de forma rápida e clara. Os demais softwares apresentados foram classificados como levemente menos ágeis, o que não desqualifica a importância de cada um deles, apenas os caracteriza com usabilidade menor que a ferramenta Biodonut.

A principal contribuição deste trabalho é a inclusão da funcionalidade de visualização 3D para esta classe de *software* aliada à plataforma *web*. Biodonut é uma ferramenta bastante promissora, podendo ser utilizada por usuários leigos em computação. A visualização 3D do *software* Biodonut foi desenvolvido utilizando uma tecnologia moderna, criada em 2013, denominada WebGL (*do inglês Web Graphics Library*) para renderizar o gráfico 3D. Para as demais funcionalidades, foram usadas tecnologias consolidadas como Java, JavaScript, HTML5, MySQL entre outras. Este trabalho é totalmente *open source*, abrindo possibilidade para entusiastas, pesquisadores e até mesmo empresas, possam contribuir com a evolução da ferramenta.

WebGL é baseado no OpenGL ES (OpenGL para sistemas embarcados) e fornece uma interface de programação de gráficos 3D. Ele usa o elemento canvas do HTML5 e é acessada por meio de interfaces DOM (do inglês *Document Object Model*). O gerenciamento automático de memória é fornecida como parte da linguagem JavaScript. Embora OpenGL seja uma especificação bastante consolidada, WebGL ainda não tem a mesma gama de funcionalidades

que OpenGL ES oferece. Além disso, não há tutoriais pontuais o bastante para uma evolutiva ágil de desenvolvimento para WebGL.

Houve uma dificuldade considerável em desenvolver um gráfico 3D mais dinâmico com WebGL. Talvez isso se deva ao fato de que OpenGL ES, base do WebGL teve uma versão considerável como estável pelo seu desenvolvedor, *Khronos Group*, a partir de 6 de agosto de 2012. Além disso, a *Khronos Group* declara WebGL com estado de desenvolvimento ainda não estável, sendo considerado como "em desenvolvimento".

A proposta inicial para o desenvolvimento deste trabalho era prover a visualização 3D de forma robusta e realista. Porém, o único meio computacional disponível até a presente data é o uso da biblioteca WebGL. Como a biblioteca em questão ainda não está estável em nível de desenvolvimento, o resultado obtido não foi o mesmo esperado. O trabalho de pesquisa foi realizado exaustivamente durante todo o processo de desenvolvimento do Biodonut. Ao longo do desenvolvimento deste trabalho, um artigo (PALMA et al.) e uma apresentação em congresso (PALMA et al.) foram realizadas, apresentando assim o andamento deste trabalho.

5.2 Trabalhos futuros

Vários temas são possíveis para trabalhos futuros que estão relacionados com o que foi estudado e desenvolvido durante o mestrado. O principal deles é o desenvolvimento da estabilidade da biblioteca WebGL com o intuito de desenvolver com maior exatidão a visualização 3D para os plasmídeos disponíveis na ferramenta Biodonut.

Outra opção de continuidade para este trabalho em termos funcionais, é incorporar outras bases de dados além do GenBank. Em termos arquiteturais, podemos considerar o banco de dados, por exemplo. Empregar um banco de dados NoSql permitindo mais escalabilidade e desempenho da aplicação. A aplicação ainda pode ser hospedada em *cloud computing*, o que não causaria grande impacto positivo pelo fato do software ter sido implementado no padrão MVC, mas certamente alavancaria algum benefício em termos de desempenho. A possibilidade de implementação para dispositivos móveis de sistema operacional *open source*, garantindo ubiquidade à aplicação. A portabilidade é uma boa aposta de necessidade computacional para as futuras implementações.

Referências Bibliográficas

ACACLONE SOFTWARE. **pDRAW32 DNA analysis software**, 2012. Disponível em <<http://www.aacclone.com>> Acessado em 23 dez. 2012.

ANGERS-LOUSTAU, A., RAINY, J. ; WARTIOVAARA, K. **PlasmaDNA: a free, cross-platform plasmid manipulation program for molecular biology laboratories**. BMC molecular biology.

APACHE MAVEN . Disponível em <<http://maven.apache.org/>> Acessado em 13 jan. 2013.

APACHE TOMCAT . Disponível em <<http://tomcat.apache.org/>> Acessado em 13 jan. 2013.

ARBEX, W.; COSTA, V.S.; SILVA, M.V.G.B. **Bioinformática como Ferramenta nas Pesquisas Atuais**. Disponível em: <<http://www.genmelhor.ufv.br/materiais/III%20egm/textobioinformatica.pdf>>. Acessado em 10 fev 2014.

ATTWOOD, T.K.; GISEL, A.; ERIKSSON, N.E.; BONGCAM-RUDLOFF, E. **Concepts, historical milestones and the central place of bioinformatics in modern biology: a European perspective**. Mahdavi MA, editor. Bioinformatics-Trends and Methodologies. Rijeka: Intech Online Publishers. 2011.

BORGES-OSÓRIO, Maria R.; ROBINSON, Wanyce M. **Genética Humana 3ed**. Artmed, 2013.

BURNS, G.W. & BOTTINO, P.J. **Genética**. Rio de Janeiro: Trad. CAMPOS, J.P. Editora Guanabara Koogan,1991.

CANDEIAS, J.A.N. **A engenharia genética**. Revista Saúde Pública, São Paulo, 1991 . Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89101991000100002&lng=en&nrm=iso>. Acessado em 21 Jan. 2013.

CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. **Readings in Information Visualization: Using Vision to Think**. Morgan Kaufman Publishers.1999.

CARVALHO, A. C. P. DE L. F. *et al.* **Grandes Desafios da Pesquisa em Computação no Brasil – 2006 – 2016**. 2006. Disponível em <<http://www.sbc.org.br/index.php?language=1&subject=8&content=downloads&id=231>> . Acessado em 18 de novembro de 2013.

CHEN, C. Editorial – **Information Visualization**. Palgrave Macmillan. 2002.

CONNOLLY, T.M., BEGG, C.E. **Database Systems – A Practical Approach to Design,Implementation, and Management**. Editora Addison-Wesley, 2002.

DATE, C.J. **Introduction to Database Systems**. Editora Addison-Wesley, 2003.

DÁVILA, A.; MENDES, P.N.; WAGNER, G.; Tschoeke, D.A. ; Cuadrat, R. R. ; LIBERMAN, F; Matos, Luciana ; SATAKE, Thiago ; OCAÑA, K; A C S ; TRIANA, O. ; GRISARD, Edmundo ; CAVALCANTI, M. C. ; CAMPOS, Maria Luiza Machado ; MATTOSO, Marta . **ProtozoaDB: dynamic visualization and exploration of protozoan genomes**. Nucleic Acids Research. 2007.

DE ROBERTIS, E. D. P.; DE ROBERTIS JR., E. M. F. **Bases da biologia celular e molecular**. Editora Guanabara Koogan, 2001

DONG, X.; STOTHARD, P.; FORSYTHE,I.J.; WISHART,D.S.: **PlasMapper: a web server for drawing and auto-annotating plasmid maps**. Nucleic Acids Res. 2004.

DWYER, T. **Three Dimensional UML using Force Directed Layout**. Proceedings of the Australian Symposium on Information Visualization (InVis.au'01), CRPIT 9, ACS. 2012.

ECLIPSE HELIOS. Disponível em <<http://www.eclipse.org/helios/>> Acessado em 13 jan. 2013.

ELMASRI, R., NAVATHE, S. **Fundamentals of Database Systems**. Editora Addison-Wesley, 2003.

FERREIRA, R. Watson & Crick. **A história da descoberta da estrutura do DNA**. Editora Odysseus, 2003.

GLASSFISH . Disponível em <<http://www.oracle.com/us/products/middleware/cloud-app-foundation/glassfish-server/overview/index.html>> Acessado em 13 jan. 2013.

GONÇALVES, Eliane S. B. **A Interação com o Usuário na Validação do Software Oficina de Relatório**. Dissertação de Mestrado (Mestrado em Engenharia de Produção). UFSC – Santa Catarina, 2001. Disponível em: <<http://teses.eps.ufsc.br/defesa/pdf/7047.pdf>> Acessado em 16 mai 2014.

HU, Z.; MELLOR, J.; WU, J.; DELISI, C. **VisANT: an online visualization and analysis tool for biological interaction data**. BMC Bioinformatics. 2004.

JAVA ENTERPRISE EDITION . Disponível em <<http://www.oracle.com/technetwork/java/javase/overview/index.html>> Acessado em 13 jan. 2013.

JAVA STANDARD EDITION. Disponível em<<http://www.oracle.com/technetwork/java/javase/documentation/index.html>> Acessado em 13 jan. 2013.

JAVA3D . Disponível em <http://java.sun.com/products/java-media/3D/>. 13 jan. 2013.

JBOSS . Disponível em <<http://www.jboss.org>> Acessado em 13 jan. 2013.

JPA . Disponível em <<http://www.oracle.com/technetwork/articles/javaee/jpa-137156.html>> 13 jan. 2013.

KARP, G. **Biologia celular e molecular, conceitos e experimentos**. Editora Manole Ltda, 2005.

KHRONOS GROUP. Disponível em <<https://www.khronos.org/>> 13 jan. 2013.

LARKIN, J.; SIMON, H. A. **Why a Diagram is (Sometimes) Worth Ten Thousand Words**. Cognitive Science.. 1987.

LEAL, E.; WIECZOREK, E. M. **Caminhos e Tendências do uso de Banco de Dados em Bioinformática**. Disponível em: <Http://www.wieczorek.com.br/publicacoes/artigo_IVencoinfo.pdf>. 2008. Acessado em 23 jul 2013.

LOPES D.S.A, PESSOA M H N, SANTOS R S, BARBOSA M S. **A produção de insulina artificial através da tecnologia do DNA recombinante para o tratamento de diabetes mellitus**.Revista da Universidade Vale do Rio Verde, Três Corações, 2012.

NIELSEN, J.; LORANGER, H. **Usabilidade na Web: Projetando Websites com Qualidade**. Rio de Janeiro: Campus. 2007

NIELSEN, J.; MOLICH, R. **Heuristic evaluation of user interfaces**. Empowering People - Chi'90 Conference Proceedings. New York: ACMPress. 1990

PALMA, E.C.N.C; BOMBONATO,M.E; PAIVA, W.R; TAKITA, M.A; CARVALHO, M.A.G. **BIODONUT: An Open Source Web-Based System For Simulating Cloning In Plasmids**, Campinas. 8th International Conference of the Brazilian Association for Bioinformatics and Computational Biology - X Meeting. 2012

PALMA, E.C.N.C; BOMBONATO,M.E; PAIVA, W.R; TAKITA, M.A; CARVALHO, M.A.G. **Biodonut: An Open Source Web-Based System For Simulating Cloning In Plasmids**, Revista Sodebrás, vol. 8, no. 85, pp 06-09. 2013.

PFLEEGER, S.L. **Engenharia de Software: Teoria e Prática**. Editora Pearson. 2004.

PIERCE, A. B. **Genética: Um Enfoque Conceitual**. Rio de Janeiro: Editora Guanabara Koogan, 2004.

PREECE, J. et al. **Human-Computer Interaction**. USA: Addison Wesley, 1994.

RAMAKRISHNAN, R.; GEHRKE, J. **Database Management Systems**. Editora McGrall Hill, 2003.

ROBERTSON, G. G., MACKINLAY, J. D., AND CARD, S. K. **Cone trees: Animated 3D visualizations of hierarchical information.** Communications of the ACM, 1993.

SILBERSCHATZ, A., KORTH, H. F., SUDARSHAN, S., **Sistema de Banco de Dados,** Editora Makron Books, 1999.

SILVA, C. G. **Exploração de bases de dados de ambientes de Educação a Distância por meio de ferramentas de consulta apoiadas por Visualização de Informação.** Tese de Doutorado. Instituto de Computação, Universidade Estadual de Campinas. 2006. Disponível em <<http://www.ic.unicamp.br/~celmar/tese>>. Acessado em 30 mai 2013.

SILVA, M. S. **Javascript: Guia do programador.** Editora Novatec, 2010.

SUDERMAN, M.; HALLETT, M. **Tools for visually exploring biological networks.** Bioinformatics. 2007.

UNIVERSITY OF HELSINKI. **PlasmaDNA: a free, cross-platform PLasmid MANipulation program.** 2007. Disponível em <<http://research.med.helsinki.fi/plasmadna/>> Acessado em 23 dez 2012.

WARE, C AND FRANK, G. **Viewing a graph in a virtual reality display is three times as good as a 2-D diagram.** IEEE Conference on Visual Languages, pages 182-183. 1994
WEBLOGIC . Disponível em
<<http://www.oracle.com/technetwork/middleware/weblogic/overview/index.html>> Acessado em 13 jan. 2013.