

BioAgents: Um Sistema Multiagente para Anotação Manual em Projetos de Sequenciamento de Genomas

Richardson Silva Lima¹, Célia Ghedini Ralha¹, Maria Emília Machado T. Walter¹, Hugo Wruck Schneider¹, Anderson Gray F. Pereira¹, Marcelo Macedo Brígido²

¹Departamento de Ciência da Computação, Instituto de Ciências Exatas
Universidade de Brasília, Campus Universitário Darcy Ribeiro
Caixa Postal 4466, Brasília-Brasil, CEP 70.910-900

²Instituto de Biologia, Universidade de Brasília
Campus Universitário Darcy Ribeiro, Brasília-Brasil, CEP 70.910-900

{rlima,ghedini}@cic.unb.br, {mariaemilia,brigido}@unb.br

{0332658,0336416}@aluno.unb.br

Abstract. *Genome sequencing projects identify biological sequences of organisms and their functions. The discovery of biological functions constitutes the annotation phase, that is divided into automatic and manual. Automatic annotation has the objective of inferring functions to the project sequences, using databases containing biological sequences and previously determined functions. Manual annotation is done by biologists, that decide the functions using their biological knowledge. This work presents BioAgents, a system that uses the Multiagent paradigm to support manual annotation. BioAgents provides interaction of different agents using the automatic annotation outputs, and suggests manual annotations that must be validated by biologists.*

Resumo. *Projetos de sequenciamento de genomas identificam seqüências biológicas de organismos e suas funções. A descoberta das funções biológicas constitui a fase de anotação, dividida em automática e manual. A anotação automática visa atribuir funções às seqüências do projeto, usando bancos de dados de seqüências biológicas com funções previamente determinadas. A anotação manual é feita por biólogos, que decidem as funções usando seu conhecimento biológico. Este trabalho apresenta o BioAgents, um sistema que utiliza o paradigma Multiagente para apoiar a anotação manual. BioAgents provê a interação entre diferentes agentes usando os resultados da anotação automática, e sugere anotações manuais que deverão ser validadas pelos biólogos.*

1. Introdução

Em 1953, Watson e Crick propuseram uma estrutura molecular para o DNA [Watson and Crick 1953]. Desde essa época, a comunidade científica vem dispendendo grandes esforços com o objetivo de compreender melhor a estrutura e o funcionamento da biologia molecular dos seres vivos. No início da década de 1990, foi iniciado o Projeto Genoma Humano, que visava mapear e sequenciar, por completo, o genoma humano. Este projeto foi concluído em 2001 [Venter et al. 2001, Lander et al. 2001], e apresentou o genoma humano com 3 bilhões de bases e aproximadamente 30.000 genes.

O Projeto Genoma Humano e inúmeros outros projetos de seqüenciamento de genomas surgidos em todo o mundo propiciaram grandes e rápidos avanços em técnicas da Biologia Molecular e Bioinformática [Liolios et al. 2006]. Assim, desde a década de 1990, podemos observar um crescimento exponencial no volume de dados gerados pelos diversos projetos de seqüenciamento de genomas. Em relação ao gerenciamento e análise destes dados, a área de Computação tem desenvolvido técnicas e *softwares* que apoiam o esforço dos biólogos no armazenamento e análise dos dados gerados nestes projetos. O sistema computacional que apoia estes projetos é denominado de *pipeline* ou *workflow* [Lemos 2004]. Um *pipeline* é dividido em três fases: submissão, montagem e anotação. A fase de submissão visa receber as seqüências geradas nos laboratórios de Biologia Molecular, transformando-as em cadeias de caracteres e armazenando-as em bancos de dados. A fase de montagem visa agrupar seqüências que potencialmente tenham vindo da mesma região do *DNA*. Cada grupo com mais de uma seqüência recebe o nome de *contig* e tem uma seqüência consenso que representa o grupo. Seqüências não agrupadas recebem o nome de *singlet*.

A fase de anotação tem o objetivo de inferir as funções biológicas das seqüências resultantes da montagem, utilizando funções conhecidas de seqüências similares disponibilizadas em bancos de dados biológicos. Esta fase é dividida em duas etapas: automática e manual. A anotação automática compara as seqüências geradas no projeto com seqüências de bancos de dados privados e/ou públicos (como o *GenBank* [Benson et al. 2006]). Métodos de comparação aproximada de seqüências¹ (como *BLAST* [Altschul et al. 1990] e *FASTA* [Pearson and Lipman 1988]) são utilizados para inferir funções das seqüências estudadas. Estas inferências são feitas comparando com seqüências semelhantes que tiveram suas funções previamente determinadas. Na anotação manual, os biólogos utilizam as informações da anotação automática, bem como seus conhecimentos, para determinar a função que deve ser associada à seqüência analisada.

Neste trabalho serão apresentados uma arquitetura e um protótipo de Sistemas Multiagente [Wooldridge 2002, Weiss 2000], denominado *BioAgents*, que visa auxiliar os biólogos na tarefa de anotação manual em projetos de seqüenciamento de genomas [Lima et al. 2005]. A escolha da abordagem Multiagente deve-se principalmente ao fato da aplicação apresentar características específicas adequadas ao uso desta tecnologia, a saber: (i) utiliza bancos de dados heterogêneos e descentralizados, (ii) constitui um ambiente dinâmico (por exemplo, novos tipos de dados e fontes de dados com constantes alterações), (iii) o processo de anotação pode ser realizado de forma independente por vários biólogos. A arquitetura apresentada foi implementada através de um protótipo que utiliza a plataforma de desenvolvimento de agentes *JADE* [Bellifemine et al. 2003], integrada ao motor de inferência *JESS* [Friedman-Hill 2003].

O protótipo foi utilizado em um estudo de caso que utiliza os dados do Projeto Genoma Funcional e Diferencial do fungo *Paracoccidioides brasiliensis* (Pb) [Felipe et al. 2005]. Esse projeto foi executado pela Rede Genoma Centro-Oeste, que integra instituições de ensino e pesquisa em Biologia Molecular do Distrito Federal, Goiás, Mato Grosso e Mato Grosso do Sul. As sugestões geradas automaticamente foram va-

¹Dizemos que duas seqüências são similares quando partes delas são "aproximadamente iguais", isto é, quando as duas seqüências têm exatamente os mesmos caracteres, com poucas exceções de caracteres diferentes, ou inserções e remoções de caracteres de uma das seqüências em relação à outra.

lidadas através de comparações de resultados gerados pelo *BioAgents* com as anotações manuais previamente realizadas no Projeto Genoma Pb.

Este trabalho está dividido em cinco seções. Na seção 2 são mostrados alguns trabalhos correlatos. Na seção 3 é apresentada a arquitetura Multiagente e descrito o protótipo implementado. Na seção 4 o estudo de caso é apresentado, sendo feita uma breve discussão dos resultados. Na seção 5 concluímos e apresentamos trabalhos futuros.

2. Trabalhos Correlatos

Vários trabalhos na área de Bioinformática utilizam técnicas de Inteligência Artificial, através do uso de abordagens distintas como a de Sistemas Multiagente (SMA), Mineração de Dados e/ou Aprendizagem de Máquina. Essas abordagens têm sido aplicadas em diferentes processos envolvidos no *pipeline* de execução, incluindo desde a comparação e análise de genomas até a inferência das funções dos genes dos organismos. Porém, não encontramos na literatura trabalhos que apliquem a abordagem de SMA para o processo de anotação manual. Apresentamos então trabalhos relacionados ao processo de anotação.

O sistema *BioMAS* utiliza a abordagem de SMA para anotação automática do vírus da herpes [Decker et al. 2001]. O foco do trabalho está na extração da informação contida nos bancos de dados públicos e no processo de anotação automática.

O *Electronic Annotation-EAnnot* é uma ferramenta originalmente desenvolvida para a anotação manual do genoma humano [Ding et al. 2004]. O software combina ferramentas para extrair e analisar grandes volumes de dados em bancos públicos, gerando anotações automáticas e predições de genes de forma rápida. *EAnnot* usa informações contidas em *messenger RNA-mRNA*, *Expressed Sequence Tags-ESTs* e alinhamentos de proteínas, além de identificar pseudogenes, entre outras características.

O software *Ambiente para Anotação Automática e Comparação de Genomas-A3C* [Santos and Bazzan 2004] é baseado em uma arquitetura de SMA e tem como propósito a integração de tarefas relacionadas a anotação denominada pelos autores como nível 1 e a comparação de genomas considerada como nível 2. O nível 1 é composto por ferramentas para a anotação automática de proteínas; enquanto o nível 2 é composto por algoritmos para comparação de genomas que visam a extração de informações úteis aos resultados do nível 1. O objetivo do A3C é descobrir a relação entre diversos organismos, obtendo então informações específicas sobre um dado genoma através do conhecimento sobre outros genomas que já se encontram seqüenciados.

A ferramenta denominada *Agent-based environment for automatic annotation of Genomes-ATUCG* é baseada em uma arquitetura de agentes, tendo como objetivo reduzir o trabalho manual dos biólogos através da re-anotação [Nascimento and Bazzan 2005]. No processo de re-anotação as informações adquiridas das seqüências originalmente anotadas são revisadas e comparadas com novos modelos e dados para se obter características e informações sobre as seqüências e refazer a anotação manual, caso seja necessário.

3. A Arquitetura Multiagente e o Protótipo do *BioAgents*

Como dito anteriormente, o *BioAgents* visa auxiliar os biólogos no processo de anotação manual. O processo de anotação manual é executado pelos biólogos basicamente: analisando as saídas das ferramentas executadas durante o processo de anotação automática, e

interpretando estes resultados, de acordo com seu conhecimento biológico, para inferir as funções e categorias funcionais das seqüências a serem anotadas. O *BioAgents* se propõe a simular esta tarefa dos biólogos.

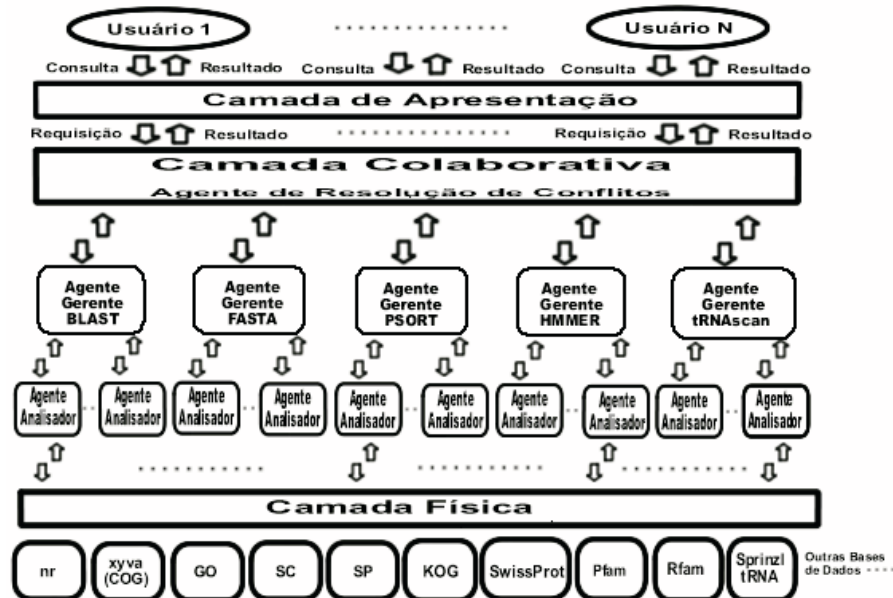


Figura 1. A arquitetura em três camadas do sistema *BioAgents*.

A Figura 1 representa a arquitetura SMA do *BioAgents*, que é composta por três camadas:

- A *Camada de Apresentação* é responsável por receber as requisições submetidas ao sistema e retornar o resultado do processamento ao usuário. A requisição consiste na submissão de seqüências a serem analisadas. Na atual implementação, as ferramentas *BLAST* e *FASTA* e os bancos de dados utilizados, foram apenas informados para o sistema. Os arquivos de saída já processados constituíram a entrada para os *Agentes Analisadores* (ANL). Estes arquivos de saída contêm os resultados das comparações efetuadas pelo *BLAST* e *FASTA*, tendo sido obtidos na etapa de anotação automática.
- A *Camada Colaborativa* é responsável pela consolidação dos resultados provenientes das análises feitas sobre os bancos de dados da *Camada Física* e por retorná-los à *Camada de Apresentação*. A *Camada de Colaboração* é composta pelo *Agente de Resolução de Conflitos* (RC), pelos *Agentes Gerentes* (GR) e pelos ANLs.
 - O agente RC tem o objetivo de submeter as requisições enviadas pela *Camada de Apresentação* aos agentes GR especializados. Após receber os resultados dos agentes GR, decide a sugestão mais apropriada para ser enviada à *Camada de Apresentação*. No estudo de caso realizado, foram utilizados os agentes GR e ANL *BLAST* e *FASTA*.
 - Os agentes GR recebem mensagens do agente RC com solicitações de acordo com sua especialidade. Um particular agente GR verifica quais são os bancos de dados e saídas dos programas que devem ter sido previamente executados na anotação automática. O agente GR aloca os agentes ANL

para fazer a análise individual dessas saídas juntamente com os bancos de dados. O agente GR aguarda as sugestões de todos os agentes ANL, consolidando-as através do uso das regras de produção previamente definidas. Como cada agente GR é especializado em um programa, ele pode avaliar e consolidar os resultados retornados pelos agentes ANL.

- Cada agente ANL executa um arquivo de saída gerado por uma ferramenta específica. Quando é criado por solicitação de um agente GR, cada agente ANL utiliza um *parser* específico para extrair informações do arquivo de saída, gerando uma estrutura contendo dados específicos da ferramenta. O resultado desse processamento com a sugestão é retornada ao agente GR solicitante.
- A *Camada Física* é responsável pelos bancos de dados utilizadas pelo *BioAgents*. Em nosso estudo de caso foram utilizadas as seguintes fontes de dados: *nr-GenBank* (<http://www.ncbi.nlm.nih.gov/Genbank/>); *Gene Ontology* (GO) (<http://www.geneontology.org/>); *Clusters of Orthologous Groups of proteins* (COG) (<http://www.ncbi.nlm.nih.gov/COG/>) e os bancos de dados dos fungos *Saccharomyces cerevisiae* (SC) e *Schizosaccharomyces pombe* (SP).

3.1. O Protótipo

Para implementar a arquitetura SMA proposta, utilizamos a linguagem *Java* (<http://java.sun.com>) no ambiente de desenvolvimento *Eclipse SDK*, versão 3.1.2 (<http://www.eclipse.org>). Como *framework* de desenvolvimento de agentes, utilizamos o *Java Agent DEvelopment Framework-JADE* versão 3.4.1 (<http://jade.tilab.com>). Na Figura 2, o *Analysis Agent* é uma interface de inicialização do *BioAgents*.

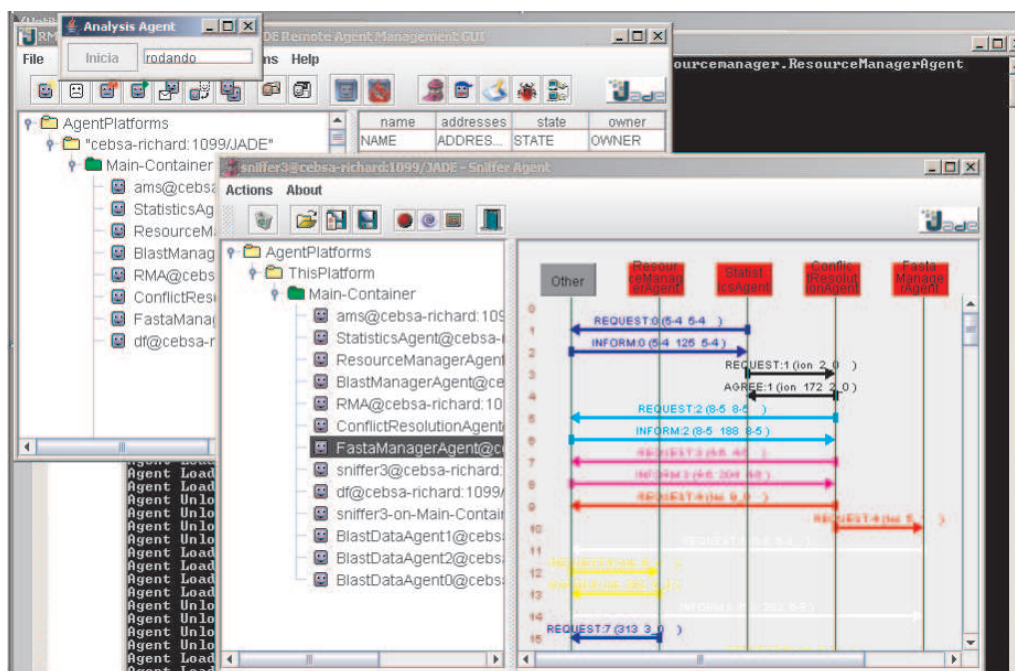


Figura 2. Screenshot da tela de execução e do sniffer dos agentes do *BioAgents* no framework *JADE*.

A utilização do *JADE* deve-se a diversos fatores, a saber: (i) ser distribuído como software livre sob licença *LGPL*; (ii) a linguagem de programação suportada ser *Java*, possibilitando boa portabilidade; (iii) as especificações de *JADE* serem compatíveis com o padrão *The Foundation of Intelligent Physical Agents-FIPA*², oferecendo uma biblioteca de classes de protocolos de interação padronizados e prontas para serem instanciadas ou estendidas; (iv) não apresentar necessidade de implementar a plataforma de agentes, as funcionalidades e a ontologia de gerenciamento de agentes, nem os mecanismos de transporte e *parsing* de mensagens; (v) oferecer um transporte eficiente de mensagens entre os agentes pelo uso da linguagem *FIPA Agent Communication Language - FIPA ACL* (<http://www.fipa.org/repository/aclspecs.html>); (vi) possuir suporte a usuários, tendo uma grande comunidade ativa de desenvolvedores e uma vasta documentação disponível para consulta.

Os *parsers* utilizados pelos agentes ANL para a manipulação dos arquivos de saída foram implementados a partir da adaptação de algumas bibliotecas do *framework BioJava* versão 1.4. O *BioJava* fornece objetos para manipulação de seqüências biológicas e *parsers* para arquivos de seqüências, dentre outras funcionalidades (http://biojava.org/wiki/Main_Page).

Como motor de inferência para o desenvolvimento do protótipo utilizamos o *Java Expert System Shell-JESS* versão 6.1 (<http://www.jessrules.com/jess/index.shtml>) [Friedman-Hill 2003]. O *JESS* é utilizado para construir bancos de conhecimento e obter inferências a partir de padrões pré-estabelecidos. O *JESS* foi especialmente desenvolvido para ser integrado à linguagem *Java*, o que permite a criação de *softwares Java* com capacidade de resolução de problemas usando conhecimento vindo das regras de produção implementadas no *JESS*. Estas regras representam o conhecimento explícito utilizado pelos biólogos na tarefa de anotação manual estando relacionadas ao conhecimento tácito utilizado durante o processo de sugestões de anotação.

4. Estudo de Caso

O estudo de caso realizado neste trabalho consistiu em utilizar o *BioAgents* com os dados do Projeto Genoma Pb, visando propor anotação a partir dos resultados *BLAST* e *FASTA* deste projeto, para comparar as anotações sugeridas com as anotações manuais previamente concluídas pelos biólogos. Os dados analisados foram os arquivos de saída do programa *BLAST* executado sobre os bancos *nr*, *COG* e *GO*; os arquivos de saída do programa *FASTA* com os bancos de dados dos fungos *Saccharomyces cerevisiae* e *Schizosaccharomyces pombe*, bem como os arquivos de anotações manuais do Projeto Genoma Pb.

Para avaliar os arquivos de saída do *BLAST* e *FASTA*, o *BioAgents* analisou dois parâmetros, o *expectation-value (e-value)* e o *score*. Estes dois parâmetros são produzidos pelo *BLAST* e pelo *FASTA* e expressam o grau de similaridade entre cada seqüência gerada no projeto e cada seqüência já existente em um banco de dados. Ambos os programas produzem *alinhamentos* entre duas seqüências, que expressam o grau de similaridade entre elas. Quanto menor o *e-value* maior a semelhança entre duas seqüências, e quanto

²FIPA é uma organização que segue o padrão internacional de especificação da *Institute of Electrical and Electronics Engineers - IEEE* para o desenvolvimento de tecnologias baseadas em agentes inteligentes de software (<http://www.fipa.org>).

maior o *score* mais próximas são as seqüências. A inferência de função é feita assumindo que quanto maior a proximidade entre duas seqüências, maior a chance de possuírem a mesma função biológica.

```
(defglobal ?*maxValue* = 1.0E-5) ; (1)

(defmodule Eval)

(defrule Exists_Value_Above_Limit
  "Activate GoodValueAnalysis module if there is at least one good eval"
  (exists (BlastHit
    (HitValue ?evalue &lt;= ?evalue ?*maxValue*)) (2)
  )
  =>
  (focus GoodValueAnalysis)
  (run)
)

(defmodule GoodValueAnalysis)

(defrule Best_Value
  "comment"
  (ValueAnalysis(Evalue ?evalue1) (Score ?score1))
  ?fact <- (ValueAnalysis(Evalue ?evalue2 &gt;= ?evalue2 ?evalue1))
  (Score ?score2 &lt; ?score2 ?score1)) (3)
  =>
  (retract ?fact)
)
```

Figura 3. Conjunto de regras Jess para análise de saídas BLAST e FASTA.

A Figura 3 ilustra a sintaxe de duas regras com uso do JESS. Ressaltamos que estas regras foram testadas com os agentes GR e ANL, usando os programas BLAST e FASTA. As regras descritas nesta figura capturam o seguinte conhecimento biológico:

- Verificar a existência de alinhamentos cujo *e-value* seja menor ou igual a 10^{-5} (valor estabelecido pelos biólogos no Projeto Genoma Pb);
- Dentre os alinhamentos que atendem à restrição anterior, selecionar o menor *e-value*;
- Caso existam dois *e-values* iguais, selecionar o de maior *score*.

Como resultado da aplicação do *BioAgents*, foram analisados 6.107 seqüências do Projeto Genoma Pb (Tabela 1). Deste total, 3.774 genes foram anotados manualmente por biólogos, e 2.333 não foram anotados. Na Tabela 1 podemos observar um tempo de execução longo, motivado pelo fato dos dados do Projeto Genoma Pb serem compostos por arquivos do tipo texto com tamanho de aproximadamente 1.5 GB.

Tabela 1. Resultados do BioAgents utilizando dados do Projeto Genoma Pb.

Quantidade de genes	6.107
Quantidade de genes anotados manualmente	3.774
Quantidade de anotações sugeridas pelo <i>BioAgents</i>	3.502
Quantidade de anotações acertadas pelo <i>BioAgents</i> / Quantidade de anotações sugeridas(% de acerto)	1.547/3.502 44.17%
Quantidade de anotações sugeridas para genes não anotados manualmente/total de genes não anotados	336/2.333
Tempo de execução do sistema (hh:mm)	01:30

Note que 3.502 anotações foram sugeridas pelo *BioAgents*, sendo que 1.547 foram sugestões corretas quando comparadas com as anotações manuais do Projeto Genoma Pb,

o que corresponde a um índice de acerto de 44.17%. Note também que das 1.955 sugestões não corretas quando comparadas com as anotações manuais do Projeto Genoma Pb, 336 foram sugestões do sistema a genes não anotados, o que corresponde a 9.59% (336/3.502), e 1.619 foram sugestões diferentes das anotadas pelos biólogos, correspondendo a 46.23% (1.619/3.502). Conforme avaliação dos biólogos, os resultados são bons e podem ainda ser melhores à medida que for expandida a base de conhecimento dos agentes.

Com base nos resultados deste estudo de caso, julgamos que o *BioAgents* pode realmente auxiliar os biólogos na fase de anotação manual em projetos de sequenciamento de genomas.

5. Conclusões e Trabalhos Futuros

Neste trabalho, apresentamos uma arquitetura, baseada no paradigma Multiagente, e o protótipo do sistema *BioAgents* para apoiar o processo de anotação manual feita por biólogos em projetos de sequenciamento de genomas. Esta aplicação possui ambiente heterogêneo e dinâmico, pois utiliza diferentes bancos de dados, descentralizados, sendo os dados constantemente alterados. Assim, esta aplicação é adequada para ser solucionada utilizando a abordagem Multiagente. No *BioAgents* os agentes são especializados em tarefas distintas, de tal forma que podem atuar de forma independente, utilizando regras específicas. Esta arquitetura foi implementada utilizando o *framework JADE*, e as regras da base de conhecimento foram desenvolvidas no *JESS*.

Realizamos um estudo de caso com os dados de anotação manual do Projeto Genoma Pb. Usando poucas regras de produção, tivemos um índice de acerto de 44.17%, computado a partir do número de sugestões corretas do *BioAgents* quando comparadas com as anotações manuais do Projeto Genoma Pb. Além disso, o projeto sugeriu 336 anotações para sequências não anotadas, consideradas corretas pelos biólogos que analisaram os dados.

Trabalhos futuros incluem a implementação com execução distribuída dos agentes, para reduzir o tempo de execução do *BioAgents*. Poderia ser desenvolvida uma *interface Web* para a *Camada de Apresentação*, provendo o acesso público aos pesquisadores que utilizassem o sistema. Pretendemos também utilizar o *BioAgents* no Projeto Genoma *Anaplasma* que em breve estará na fase de anotação manual (<http://dna.biomol.unb.br/ANA/>). O aprimoramento do conhecimento dos agentes GR e ANL também é necessário para possibilitar uma maior acurácia nas sugestões das anotações manuais. Isto poderia ser feito incluindo novos métodos e bases de dados (como detecção de *RNAs* não-codificadores (*ncRNAs*), identificação de *RNAs* de transferência (*tRNAs*) e identificação de homologias em famílias de proteínas - *HMMER/Pfam*).

Referências

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, pages 403–410.
- Bellifemine, F., Caire, G., Poggi, A., and Rimassa, G. (2003). Jade - a white paper. White Paper 3, TILAB - Telecom Italia Lab.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2006). Genbank. *Oxford Journals, Nucleic Acids Research*, 34:D16–D20.

- Decker, K., Zheng, X., and Schmidt, C. (2001). A multi-agent system for automated genomic annotation. In *AGENTS '01: Proceedings of the 5th international conference on Autonomous agents*, New York, NY, USA. ACM Press.
- Ding, L., Sabo, A., Berkowicz, N., Meyer, R. R., Shotland, Y., Johnson, M. R., Pepin, K. H., Wilson, R. K., and Spieth, J. (2004). Eannot: A genome annotation tool using experimental evidence. *Genome Research*, 14(12):2503–2509.
- Felipe, M. S. S., Andrade, R. V., Arraes, F. B. M., Nicola, A. M., and et al (2005). Transcriptional profiles of the human pathogenic fungus *paracoccidioides brasiliensis* in mycelium and yeast cells. *Journal of Biological Chemistry (JBC)*, 280(26):24706–24714.
- Friedman-Hill, E. (2003). *Jess in Action: Rule-Based Systems in Java*. Manning Publications Co, Greenwich, CT.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., and et al (2001). Initial sequencing and analysis of the human genome. *Nature*, 409:860–921.
- Lemos, M. (2004). *Workflow para bioinformática*. PhD thesis, Pontifícia Universidade Católica do Rio de Janeiro (Puc-Rio).
- Lima, R. S., Ralha, C. G., Walter, M. E. M. T., and Brígido, M. M. (2005). A multiagent system to help manual annotation on genome sequencing projects. Proceedings of the IGWD '05 - International Workshop on Genomic Databases and Problem- Rio de Janeiro, Brazil, November 2005. Disponível em: <http://www.biowebdb.org/iwgd05/proceedings/multiagent-system.pdf>. Acesso em: Fevereiro de 2007.
- Liolios, K., Tavernarakis, N., Hugenholtz, P., and Kyrpides, N. C. (2006). The genomes on line database (gold) v.2: a monitor of genome projects worldwide. *Oxford Journals, Nucleic Acids Research*, 34:D332–D334.
- Nascimento, L. V. and Bazzan, A. L. (2005). An agent-based system for re-annotation of genomes. *Genetics and Molecular Research*, 4(3).
- Pearson, W. R. and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the USA*, 85:2444–2448.
- Santos, C. T. and Bazzan, A. L. C. (2004). Using the A3C system for annotation of keywords - a case study. *III Brazilian Workshop on Bioinformatics (WOB)*. Brasília, DF.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., and et al (2001). The sequence of the human project. *Science*, 291(16):1304–1351.
- Watson, J. O. and Crick, F. H. C. (1953). Molecular structure of nucleic acids- a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738.
- Weiss, G., editor (2000). *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. The MIT Press, Cambridge, Massachusetts.
- Wooldridge, M. (2002). *An Introduction to Multiagent Systems*. John Wiley & Sons, LTD, England.