

Memorial e Currículo Lattes

Augusto Fernandes Vellozo

8 de março de 2010

Memorial e Currículo Lattes apresentado como parte das exigências do Concurso Público de Provas e Títulos para o Magistério de Ensino Superior da Universidade Federal de São Paulo no campus de São José dos Campos.

Bacharel em Ciência da Computação na Unicamp e Doutor em Ciências na área de concentração de Ciência da Computação na USP, atualmente no terceiro ano de pós-doutorado na Universidade Lyon I - França na área de Biologia Computacional.

Sumário

1	Informações pessoais	4
2	Retrospectiva profissional e acadêmica	5
3	Formação acadêmica	9
3.1	Pós-Doutorado - 07/2009 a atual	9
3.2	Pós-Doutorado - 07/2008 a 07/2009	9
3.3	Pós-Doutorado - 07/2007 a 07/2008	9
3.4	Doutorado em Ciência da Computação- 03/2001 a 04/2007	9
3.5	Especialização em Informática e Sistemas de Informação - 03/1997 a 07/1998	10
3.6	Especialização em Administração de marketing - 03/1995 a 07/1996	10
3.7	Graduação em Ciência da Computação - 03/1985 a 07/1989	10
4	Produção bibliográfica	11
4.1	Genome Sequence of the Pea Aphid <i>Acyrtosiphon pisum</i> . PLoS Biology - 2010	11
4.2	Alignment with non-overlapping inversions in $O(n^3)$ -time. WABI, 2006 . .	11
4.3	Alignment with non-overlapping inversions in $O(n^3 \log n)$ -time. GRACO, 2005	12
5	Produção técnica acadêmica	13
5.1	<i>Software</i> - Cycads - Cyc Annotation Database system. 2009.	13
5.2	<i>Base de dados</i> - AcypiCyc - <i>Acyrtosiphon pisum</i> Cyc database. 2009. . .	13
6	Projetos de pesquisa	14
6.1	MetNet - Systems level analysis of the Pea Aphid - 2008 a atual	14
6.2	NEMO - Network Motifs - 2009 a atual	14
6.3	Computational Biology, Algorithmics, Combinatorial Optimization - 2006 a 2007	15
6.4	Algorithmic Questions on Molecular Biology - 2004 a 2005	15
7	Participação em eventos	17
7.1	Metabolic Pathways Analysis-2009 - Leiden - Holanda	17
7.2	6th ISS Congress-2009 - Madison - EUA	17
7.3	II Pea Aphid Genome Annotation Workshop-2009 - Barcelona - Espanha .	17
7.4	8th International Symposium on Aphids-2009 - Catania - Itália	17
7.5	Réseau Français de Biologie Adaptative des Pucerons-2009 - Paris - França	17
7.6	I Pea Aphid Genome Annotation Workshop-2008 - Princeton - EUA	17
7.7	IPG (Integrative Post-Genomics)-2008 - Lyon - França	18
7.8	BioCyc and Pathway Tools tutorial-2008 - Londres - Inglaterra	18
7.9	Journées Ouvertes Biologie Informatique Mathématiques-JOBIM-2008 - Lille - França	18
8	Experiência didática	19
8.1	Instituto Brasileiro de Tecnologia Avançada, IBTA, Brasil - 01/2002 a 07/2003	19

9	Experiência profissional em empresas	20
9.1	Universo OnLine Ltda.-UOL, Brasil - 04/2000 a 01/2004	20
9.2	GlobalSys-Bradesco, Brasil - 01/1999 a 04/2000	20
9.3	Banco Finasa, Brasil - 06/1997 a 01/1999	20
9.4	Vellozo Informática, Brasil - 09/1995 a 05/1997	20
9.5	ABR Sistemas, Brasil - 03/1992 a 08/1995	20
9.6	Quantum Computadores, Brasil - 10/1989 a 10/1991	20
10	Considerações finais	21
11	Currículo Lattes	22

1 Informações pessoais

Nome: Augusto Fernandes Vellozo

Nome em citações bibliográficas: Augusto F Vellozo, Augusto Vellozo

Filiação: Esutáquio Fernandes Vellozo e Ignêz Dacal Vellozo

Estado civil: casado

Nascimento: 12/04/1967

Naturalidade: Avaré/SP - Brasil

Nacionalidade: Brasileira e Portuguesa

RG: 16.187.494 - SSP - SP

CPF: 074.309.138-81

Endereço residencial: 10 Rue Paul Cambon - Villeurbanne - França

CEP: 69100

Telefone: (+33) 426766582

Endereço para correspondência no Brasil: Rua Mato Grosso, 30 - Santa Terezinha

CEP: 18040-020 - Sorocaba - SP

Telefones para contato no Brasil: (15) 3221.6841 (Adriana) ou (11) 2092-8834 (Luíza)

E-mail: augusto@vellozo.org

2 Retrospectiva profissional e acadêmica

Nesta retrospectiva, tentarei relatar de forma narrativa os fatos que marcaram minha vida profissional e acadêmica.

Desde o Segundo Grau tenho facilidade com a área de exatas. No primeiro ano do segundo grau participei das Olimpíadas de Matemática e, depois de passar pelas seletivas da escola e da cidade de Sorocaba, fui classificado entre os 10 primeiros (9º lugar) no estado de São Paulo, o que foi um resultado surpreendente dado que eu estudava em uma escola pública cujo currículo era defasado em relação às escolas particulares.

Lembro-me que no fim do segundo ano do segundo grau meu professor de matemática (Sr. Edimir) veio conversar comigo e com um outro aluno (Carlos, que também ia bem em matemática) depois da aula. Ele queria nos orientar sobre os caminhos a seguir. Lembro dele dizer sobre o nosso potencial e que deveríamos fazer um terceiro ano preparatório num cursinho para entrarmos numa boa universidade. Meu colega Carlos disse que iria para o Objetivo e eu disse a ele que meu pai não tinha condição de pagar um terceiro ano num cursinho e que iria tentar continuar o terceiro ano no colégio público e caso não passasse numa boa universidade iria fazer um cursinho. Este professor conversou com os outros professores do colégio e dois outros professores (Dona Wanda de História e o Sr. Zé Eduardo de Química) meus que trabalhavam também no Anglo e eles conseguiram uma bolsa para eu fazer o terceiro ano no Anglo.

Dentre as opções da área de exatas, escolhi o curso de Ciência da Computação pois, depois de conversar com alguns professores, achei que era um curso onde a lógica era muito importante e que tinha um bom futuro profissional. Escolhi prestar o vestibular para a Unicamp pois este era indicado como o melhor curso de Computação naquela época.

Prestei também o vestibular do Barro Branco, pois um vizinho da minha mãe que era Sargento da Polícia Militar, convenceu minha mãe que esta era a melhor opção para mim. Eu não queria ser militar, mas como o vestibular do Barro Branco era um mês antes da FUVEST (o vestibular da Unicamp era pela FUVEST naquela época) resolvi fazê-lo para agradar minha mãe (apesar de já ter dito a ela que minha preferência era pela Unicamp) e para me preparar para a FUVEST. Lembro-me inclusive que um dia antes de sair a lista de resultados do Barro Branco fui chamado a comparecer perante o comandante da polícia militar da região de Sorocaba, o qual me contou que eu havia sido um dos aprovados no Barro Branco e que gostaria de me dar os parabéns e boa sorte na minha carreira militar. Em seguida me deu algumas orientações de como eu deveria me comportar na academia. Depois daquela conversa tinha certeza que não iria para a carreira militar, mesmo que não passasse na Unicamp.

Com 17 anos passei então no segundo curso mais concorrido do país na época: Ciência da Computação na Unicamp.

Mudei-me para Campinas e lá fiz o curso de bacharelado em Ciência da Computação. Ao final do primeiro ano fui convidado a fazer a modalidade científico-industrial. Naquela época o curso tinha, a partir do segundo ano, duas modalidades: a comercial e a científico-industrial. A modalidade científico-industrial era limitada a 20 alunos que tivessem um bom coeficiente de rendimento no fim do primeiro ano. Como minhas notas eram boas, fui convidado a fazer essa modalidade.

Terminei a graduação sem qualquer reprovação e consegui terminar o curso em quatro anos e meio, sendo que o previsto pela currículo da Unicamp eram 5 anos.

Depois de formado, um professor convidou-me para trabalhar numa empresa (Quantum Computadores) onde era gerente da área de pesquisa e desenvolvimento. Trabalhei lá durante

dois anos no desenvolvimento e manutenção de BIOS para PCs 386 e 486. A área de pesquisa e desenvolvimento da empresa era em Campinas e tinha 6 funcionários (2 engenheiros elétricos para hardware, 3 bacharéis em Ciência da Computação para o software básico e 1 gerente). Naqueles anos a importação de computadores era proibida no Brasil e todos os computadores, assim como os softwares básicos, deveriam ser desenvolvidos e fabricados no Brasil.

Com o fim da reserva de mercado a empresa fechou a área de pesquisa e desenvolvimento e fui transferido para São Paulo e trabalhei quase dois meses lá dando suporte para a área comercial. Não gostei das novas funções nas quais vinha trabalhando e resolvi sair da Quantum, voltar para Sorocaba e abrir em sociedade com um colega uma empresa (ABR Sistemas) para a comercialização de computadores e desenvolvimento de softwares. A empresa prosperou e chegou a ter 12 funcionários, sendo uma das maiores e mais conhecidas na comercialização de computadores em Sorocaba.

Sentindo falta do meio acadêmico e desejando diversificar um pouco mais meus conhecimentos, comecei a fazer uma pós-graduação lato-sensu em Administração de Marketing na UNISO (Universidade de Sorocaba).

Depois de quatro anos competindo com o mercado paralelo de venda de computadores, resolvi vender a empresa para o meu sócio e abrir uma outra empresa (Vellozo Informática) só para o desenvolvimento de softwares. Consegui durante dois anos desenvolver softwares para pequenas empresas e prestar serviços para algumas grandes indústrias de Sorocaba. Neste período também terminei minha pós-graduação em Administração de Marketing e comecei uma outra pós-graduação em Sistema de Informação na UNISO também.

Nesta época já estava namorando e pretendendo me casar. Para isto achei que precisava de algum trabalho que me desse mais segurança para montar uma família.

Mande alguns currículos para verificar se conseguia algo que se adequasse mais aos meus planos de vida no momento.

Consegui então um emprego na equipe de desenvolvimento do "Home Banking" do Fina em São Paulo. Naquela época (1997 e 1998) a internet ainda estava começando no Brasil e as pessoas utilizavam os serviços do banco em casa através de um programa fornecido pelo banco (*Home Banking*) e que se comunicava com os mainframes do banco através do modem discado. Trabalhei lá durante quase dois anos no desenvolvimento e manutenção do *Home Banking* do banco. Nesta época também finalizei a minha segunda pós-graduação na UNISO.

Ainda trabalhando com bancos, consegui um emprego com um bom salário numa empresa de consultoria para o Bradesco em Alphaville onde trabalhei por 1 ano e meio.

Com o mercado de internet num momento muito bom no Brasil, no ano de 2000, fui trabalhar no BOL/UOL em São Paulo. A princípio fui trabalhar no BOL, na equipe para desenvolvimento de aplicativos de email. O BOL foi montado por Victor Ribeiro que fez mestrado na UFMG. Victor montou a equipe técnica do Bol com muitas pessoas da UFMG, sendo muitas destas com mestrado. Logo Victor passou a ser também o diretor de tecnologia de todo o UOL que fica na Av. Faria Lima, não muito longe da USP. Fui transferido junto com a equipe do BOL para a Av Faria Lima. Como existiam muitas pessoas do meio acadêmico na equipe de desenvolvimento e como a USP era relativamente perto do UOL, começaram a surgir pessoas no nosso setor interessadas em fazer um mestrado na USP. Eu, já há 2 anos e meio longe da universidade, era um deles e junto com outros dois colegas do UOL fizemos a prova para entrar no mestrado (PosComp) do IME-USP em Ciência da Computação. Passei em quarto lugar no PosComp e assim comecei a fazer o mestrado no IME-USP. Um dos colegas de trabalho não foi muito bem na prova e depois de entrar em

contato com alguns professores conseguiu fazer algumas disciplinas de verão para habilitá-lo a começar o mestrado. Como fui muito bem na prova para entrar no mestrado, fui matriculado diretamente e me deram a opção de bolsa. Porém como era recém-casado e estava construindo minha casa, não podia sair do UOL para ficar somente com a bolsa de mestrado. Desta forma, comecei o mestrado sem bolsa e trabalhando no UOL. No fim do primeiro ano consegui terminar 6 disciplinas, o que é esperado para quem faz o mestrado sem trabalhar. Com o bom desempenho do primeiro ano, novamente foi-me oferecida a bolsa de mestrado e novamente a recusei, pois optei em continuar no UOL.

No fim do primeiro ano optei por fazer o meu mestrado na área de Biologia Computacional, pois era uma área em expansão e onde se usava bastante teoria da computação. Dentre os professores que trabalhavam com Biologia Computacional, escolhi trabalhar com Alair Pereira do Lago.

No início do segundo ano de mestrado surgiu a oportunidade de dar aulas num Instituto de Graduação que estava começando em São Paulo (IBTA). Como já pensava em seguir a carreira acadêmica, resolvi experimentar para ver se iria me adaptar bem e gostar de dar aulas. Desta forma, no segundo ano e início do terceiro de mestrado, dei aulas no IBTA.

No terceiro ano de mestrado (2003), depois de verificar que gostava de dar aulas e ter um bom desempenho nas disciplinas do mestrado, resolvi que iria seguir na carreira acadêmica. Porém, neste ano nasceu o meu primeiro filho (Bruno) e optei por manter ainda naquele ano o emprego no UOL e parar com as aulas no IBTA.

No mestrado estudava algoritmos em árvores de sufixos.

Os dois colegas do UOL que começaram comigo o mestrado não conseguiram terminá-lo. Um deles desistiu no fim do segundo ano, após não conseguir ser aprovado em algumas disciplinas e ficar estressado com o acúmulo de trabalho (UOL e mestrado). O outro colega foi dispensado do UOL no início do segundo ano do mestrado e teve que refazer o PosComp e começar outra vez o mestrado após ser reprovado duas vezes seguidas na disciplina de Análise de Algoritmos, a qual é obrigatória.

No fim daquele ano (terceiro ano de mestrado), já com o filho bem de saúde e a casa construída, resolvi me dedicar totalmente à carreira acadêmica e achei que era o momento de sair do UOL. Quando comuniquei ao meu orientador sobre as minhas decisões, ele me sugeriu que comesse um doutorado direto e pegasse uma bolsa institucional de doutorado. Segui o seu conselho e comecei em 2004 o meu doutorado. Foi um início de doutorado muito difícil, pois minha mãe e meu pai estavam doentes e faleceram no fim de 2003 e início de 2004 respectivamente.

No doutorado comecei a trabalhar com alinhamento de sequências com inversões. Este trabalho continuava o trabalho [10] do meu orientador (Alair Pereira do Lago) no pós-doutorado que ele havia feito nos EUA. Trabalhei também com Carlos Eduardo Rodrigues Alves que foi o primeiro autor num artigo [1] que fizemos sobre um algoritmo de alinhamento com inversões não sobrepostas em tempo $O(n^3 \log n)$. Neste ano de 2004 trabalhei na implementação e nos testes deste algoritmo.

No ano seguinte (2005) consegui desenvolver um novo algoritmo com tempo de processamento melhor para o mesmo problema de alinhamento de sequências com inversões não sobrepostas [23] e trabalhei nos testes deste algoritmo.

Em 2006 fiz um trabalho novo para o alinhamento de sequências, desta vez considerando as duplicações. Com este trabalho ficou clara a evolução que tive no doutorado, pois comecei com um problema já existente (alinhamento de sequências com inversões não sobrepostas), aperfeiçoando o melhor método até então conhecido e depois comecei independentemente o estudo de um novo problema (alinhamento de sequências com duplicações).

No fim de 2006 e início de 2007 escrevi e defendi minha tese: Alinhamento de sequências com rearranjos [22].

Em Julho de 2007 comecei o meu pós-doutorado em Lyon-França com Marie-France Sagot na universidade Lyon I. Marie-France é uma pessoa excepcional e me ajudou bastante na difícil adaptação que tive com a esposa e os dois filhos num outro país. A experiência neste pós-doutorado está sendo muito gratificante, tanto na parte profissional como na parte pessoal.

Neste pós-doutorado conheci muitos pesquisadores de vários países diferentes e pude participar de vários *workshops* e conferências na Europa. Isto pôde alavancar bastante minha parte profissional no meio acadêmico, aumentando muito minha rede social de pesquisadores e com isto criando muitas possibilidades de colaboração para o futuro. Além disto, pude aprender sobre muitos assuntos diferentes, inclusive me adentrando um pouco mais na parte prática da Bioinformática, pois no meu doutorado estive muito restrito a parte teórica deste assunto. Ou seja, no meu doutorado estudei muito a teoria computacional dos problemas e no pós-doutorado aprendi e trabalhei mais com a parte prática de alguns problemas na área de Bioinformática.

Nos primeiros 6 meses do meu pós-doutorado, além de todos os problemas de adaptação, trabalhei com a continuação da minha tese e na implementação dos algoritmos para o alinhamento de sequências com duplicações.

Depois disto entrei no projeto genoma do inseto *Acyrtosiphon pisum* [6]. Neste projeto trabalhei com a parte da anotação e construção da rede metabólica deste organismo. Para isto desenvolvi o software CycADS [25]. CycADS foi desenvolvido para possibilitar a construção da rede metabólica de um organismo qualquer. A rede metabólica é construída a partir de várias fontes distintas das anotações ECs e/ou GOs das proteínas. Para a atribuição das reações relacionadas a cada EC e/ou GO e a visualização da rede metabólica do organismo, CycADS utiliza um outro software (PathwayTools [16]).

Ainda no assunto de rede metabólicas, estou trabalhando atualmente num algoritmo para a inferência de motivos coloridos [18] numa rede metabólica. Como não conhecemos nenhum software que faça a inferência de motivos coloridos, atualmente estou fazendo alguns testes com outros softwares que fazem algo próximo a isto.

3 Formação acadêmica

3.1 Pós-Doutorado - 07/2009 a atual

Université Claude Bernarde Lyon I, Lyon, França.

Laboratoire Biométrie et Biologie Évolutive - UMR CNRS 5558 - LBBE

Equipe BAOBAB - coordenadora: Marie-France Sagot

<http://lbbe.univ-lyon1.fr/-Equipe-Baobab.html>

Na tentativa de compreender melhor a rede metabólica de um organismo, muitos estudos e análises estão sendo realizados atualmente. Uma das maneiras de se realizar estes estudos e análises é através da procura e identificação de padrões (ou *motifs*) [18] na rede metabólica.

Neste pós-doutorado estou trabalhando com análise de redes metabólicas e de interação proteína-proteína (PPI) através de *motifs*.

Além disto tive contato também com outras técnicas de análise de redes metabólicas, como modos elementares [20] e os conjuntos minimais de precursores de um conjunto de produtos alvo [7].

3.2 Pós-Doutorado - 07/2008 a 07/2009

Institut National des Sciences Appliquees de Lyon, INSA-Lyon, Lyon, França.

Laboratoire Biologie Fonctionnelle Insectes et Interactions - UMR203 - BF2I

Equipe SymTrophique - coordenador: Hubert Charles

<http://criport1.insa-lyon.fr/bf2i/index.php?Rub=31>

Dentro do projeto do sequenciamento do genoma do inseto *Acyrtosiphon pisum* [6], desenvolvemos o aplicativo CycADS [25] que gera uma pontuação (ou número de evidências da anotação) para cada anotação funcional do organismo.

Com este software conseguimos construir a base de dados *Acypicyc* [24] e através dela a possibilidade da análise da rede metabólica do recém sequenciado organismo *Acyrtosiphon pisum*.

3.3 Pós-Doutorado - 07/2007 a 07/2008

Universite Claude Bernarde Lyon I, Lyon, França.

Laboratoire Biométrie et Biologie Évolutive - UMR CNRS 5558 - LBBE

Equipe BAOBAB - coordenadora: Marie-France Sagot

<http://lbbe.univ-lyon1.fr/-Equipe-Baobab.html>

Neste pós-doutorado continuei os meus estudos e trabalhos do doutorado na área de alinhamento de sequências com duplicações.

Iniciei neste período no projeto do sequenciamento do genoma do inseto *Acyrtosiphon pisum* [6].

3.4 Doutorado em Ciência da Computação- 03/2001 a 04/2007

Universidade de São Paulo, USP, Brasil.

Orientador: Alair Pereira do Lago.

Título da tese: Alinhamento de seqüências com rearranjos [22].

Resumo da tese: Uma das tarefas mais básicas em bioinformática é a comparação de sequências feita por algoritmos de alinhamento, que modelam as alterações evolutivas nas sequências biológicas através de mutações como inserção, remoção e substituições de símbolos.

Este trabalho trata de generalizações nos algoritmos de alinhamento que levam em consideração outras mutações conhecidas como rearranjos, mais especificamente, inversões, duplicações em *tandem* e duplicações por transposição.

Alinhamento com inversões não tem um algoritmo polinomial conhecido e uma simplificação para o problema que considera somente inversões não sobrepostas foi proposta em 1992 por Schöniger e Waterman [21]. Em 2003, trabalhos independentes propuseram algoritmos com tempo¹ $O(n^4)$ [10, 13, 11] para alinhar duas sequências com inversões não sobrepostas. Desenvolvemos dois algoritmos que resolvem este mesmo problema: um com tempo de execução $O(n^3 \log n)$ [2] e outro que, sob algumas condições no sistema de pontuação, tem tempo de execução $O(n^3)$ [23], ambos em memória $O(n^2)$.

Em 1997, Benson propôs um modelo de alinhamento que reconhecesse também as duplicações em *tandem*. Ele propôs dois algoritmos exatos para alinhar duas sequências com duplicações em *tandem*: um em tempo $O(n^5)$ e memória $O(n^2)$, e outro em tempo $O(n^4)$ e memória $O(n^3)$. Propomos um algoritmo para alinhar duas sequências com duplicações em *tandem* em tempo $O(n^3)$ e memória $O(n^2)$. Propomos também um algoritmo para alinhar duas sequências com *transposons* (um tipo mais geral que a duplicação em *tandem*), em tempo $O(n^3)$ e memória $O(n^2)$.

3.5 Especialização em Informática e Sistemas de Informação - 03/1997 a 07/1998

Universidade de Sorocaba, UNISO, Brasil.

Carga Horária: 384h

Esta especialização foi útil para retomar os estudos em computação e também o contato com os professores que faziam doutorado, pós-doutorado ou docência na USP ou UNICAMP.

Pude também estabelecer contato com várias pessoas que trabalhavam na área de informática em várias empresas da região de Sorocaba e São Paulo.

3.6 Especialização em Administração de marketing - 03/1995 a 07/1996

Universidade de Sorocaba, UNISO, Brasil.

Carga Horária: 362h

Nesta especialização pude retomar o contato com o meio acadêmico, pois já faziam quase 6 anos que não fazia um curso numa Universidade.

Também foi muito válido para aprender várias técnicas de administração e marketing, podendo com isto entender melhor o meio empresarial.

3.7 Graduação em Ciência da Computação - 03/1985 a 07/1989

Universidade Estadual de Campinas, UNICAMP, Brasil.

Foi neste curso que tive o primeiro contato com o meio acadêmico e onde pude obter as bases para os meus estudos futuros.

Foi certamente muito importante na minha carreira acadêmica.

¹Neste caso, n denota o tamanho máximo dentre as duas sequências a serem alinhadas.

4 Produção bibliográfica

4.1 Genome Sequence of the Pea Aphid *Acyrtosiphon pisum*. PLoS Biology - 2010

Abstract [6]: Aphids are important agricultural pests and also biological models for studies of insect-plant interactions, symbiosis, virus vectoring, and the developmental causes of extreme phenotypic plasticity. Here we present the 464 Mb draft genome assembly of the pea aphid *Acyrtosiphon pisum*. This first published whole genome sequence of a basal hemimetabolous insect provides an outgroup to the multiple published genomes of holometabolous insects. Pea aphids are host-plant specialists, they can reproduce both sexually and asexually, and they have coevolved with an obligate bacterial symbiont. Here we highlight findings from whole genome analysis that may be related to these unusual biological features. These findings include discovery of extensive gene duplication in more than 2000 gene families as well as loss of evolutionarily conserved genes. Gene family expansions relative to other published genomes include genes involved in chromatin modification, miRNA synthesis, and sugar transport. Gene losses include genes central to the IMD immune pathway, selenoprotein utilization, purine salvage, and the entire urea cycle. The pea aphid genome reveals that only a limited number of genes have been acquired from bacteria; thus the reduced gene count of *Buchnera* does not reflect gene transfer to the host genome. The inventory of metabolic genes in the pea aphid genome suggests that there is extensive metabolite exchange between the aphid and *Buchnera*, including sharing of amino acid biosynthesis between the aphid and *Buchnera*. The pea aphid genome provides a foundation for post-genomic studies of fundamental biological questions and applied agricultural problems.

4.2 Alignment with non-overlapping inversions in $O(n^3)$ -time. WABI, 2006

Abstract [23]: Alignments of sequences are widely used for biological sequence comparisons. Only biological events like mutations, insertions and deletions are usually modeled and other biological events like inversions are not automatically detected by the usual alignment algorithms.

Alignment with inversions does not have a known polynomial algorithm and a simplification to the problem that considers only non-overlapping inversions were proposed by Schöniger and Waterman [21] in 1992 as well as a corresponding $O(n^6)$ solution². An improvement to an algorithm with $O(n^3 \log n)$ -time complexity was announced in an extended abstract [2] and, in this present paper, we give an algorithm that solves this simplified problem in $O(n^3)$ -time and $O(n^2)$ -space in the more general framework of an edit graph.

Inversions have recently [3, 8, 14, 17] been discovered to be very important in Comparative Genomics and Scherer et al. in 2005 [12] experimentally verified inversions that were found to be polymorphic in the human genome. Moreover, 10% of the 1,576 putative inversions reported overlap RefSeq genes in the human genome. We believe our new algorithms may open the possibility to more detailed studies of inversions on DNA sequences using exact optimization algorithms and we hope this may be particularly interesting if applied to regions around known rearrangements boundaries. Scherer report 29 such cases and prioritize them as candidates for biological and evolutionary studies.

²In this case, n denotes the maximal length of the two aligned sequences.

4.3 Alignment with non-overlapping inversions in $O(n^3 \log n)$ -time. GRACO, 2005

Abstract [1]: Alignment of sequences is widely used for biological sequence comparisons, and only biological events like mutations, insertions and deletions are usually modeled. Other biological events like inversions are not automatically detected by the usual alignment algorithms.

Alignment with inversions has not a known polynomial algorithm and a simplification to the problem that considers only non-overlapping inversions were proposed by Schöniger and Waterman [21] in 1992 as well as a corresponding $O(n^6)$ solution³. Recent works [13, 10, 9, 11] improved these results to $O(n^4)$ -time and $O(n^2)$ -space complexity. In this present extended abstract, we announce an algorithm that solves this simplified problem in $O(n^3 \log n)$ -time and $O(n^2)$ -space.

³In this paper, n denotes the maximal length of the two aligned sequences.

5 Produção técnica acadêmica

5.1 *Software* - Cycads - Cyc Annotation Database system. 2009.

<http://code.cycadsys.org/> Resumo obtido do artigo que está em preparação para ser publicado.

Abstract:

Background In recent years, more and more organisms are having their genome sequenced, but their annotation remains a lengthy and long lasting process. The BioCyc databases offer a framework for the integrated analysis of metabolic networks. The Pathway tool software suite offers automation to build a database starting from an annotated genome, but all annotations have to be integrated in a summary file and/or extracted from a GenBank file.

Results To allow the easy creation and updates of a BioCyc database starting from multiple genome annotation resources available over time, we developed an *ad hoc* data management system that we named CycADS (Cyc Annotation Database System). CycADS is centred on a specific database model and on a set of Java programs to import and export relevant information. Data from GenBank and other annotation sources (including for example: KAAS [19], PRIAM [4], BLAST2GO [5] and PhylomeDB [15]) are collected in the database and later extracted to generate a complete annotation file to build a BioCyc database using the PathoLogic program. The CycADS pipeline for annotation management was used to build the AcypiCyc database for the pea aphid (*Acyrtosiphon pisum*) whose genome was recently sequenced. The AcypiCyc database page includes also, for comparative analysis, two other metabolic reconstruction BioCyc databases generated using CycADS: TricaCyc for *Tribolium castaneum* and DromeCyc for *Drosophila melanogaster*.

Conclusions CycADS offers, thanks to its flexible design, a powerful software tool for the generation and regular update of metabolic BioCyc databases and it is particularly suited for newly sequenced genome metabolism gene annotation and network reconstruction.

5.2 *Base de dados* - AcypiCyc - Acyrthosiphon pisum Cyc database. 2009.

Resumo obtido do *site* da base de dados (<http://acypicyc.cycadsys.org>).

Abstract: The AcypiCyc database has been generated at the Université de Lyon in a collaboration between the UMR203 INRA-INSA de Lyon, Biologie fonctionnelle des insectes et interactions (BF2I) and the Baobab team at UMR5558 LBBE CNRS-UCBL1; both groups are member of the INRIA Bamboo team-project. The database is hosted at the Pôle Rhône-Alpes de Bioinformatique (PRABI). We are working on a version that includes KAAS [19], PRIAM [4], BLAST2GO [5] and PhylomeDB [15] annotations. We will add more annotations soon. At present, you will be able to compare *Acyrtosiphon pisum* metabolism to *Buchnera aphidicola* APS REVISED (generated in Lyon inside the SymbioCyc database).

6 Projetos de pesquisa

6.1 MetNet - Systems level analysis of the Pea Aphid - 2008 a atual

<http://www.metnet4sysbio.org/>

Desenvolvido Université Claude Bernanrd - Lyon I em parceria com a University of York - Inglaterra.

A seguir um resumo em inglês retirado da página do projeto.

From a metabolic perspective, an animal comprises a set of metabolic compartments (including organs) connected by the inter-compartment transfer of metabolites. We seek to construct and analyse multi-compartment models of metabolism, by transforming a single global metabolic network into a set of compartment networks of overlapping composition using flux balance analysis. In parallel, we will develop and apply two computational approaches for network analysis: an exploration of metabolic network motifs, and elementary modes to characterize and compare networks. Large inter-compartment metabolic differences facilitating model development will be achieved in our system comprising two distinct genomes within a single animal: the pea aphid *Acyrtosiphon pisum* and its symbiotic bacteria *Buchnera* housed in a special organ, the bacteriome. The first model comprises two compartments, describing how the insect and bacterial networks mediate the transformation of precisely known diet components to yield biomass production (and waste); and subsequent models are of increasing complexity, involving the subdivision of the insect network into overlapping, but distinct, networks of the contributing compartments/organs. Model construction will be informed by empirical data collected on the aphid transcriptome, *Buchnera* proteome and embryo nutrient budgets within this project, and the model outputs will be validated against whole-insect metabolic and nutritional datasets obtained previously from the investigators' labs. Algorithms and models developed in this project will be maintained in a central Repository linked to pre-existing databases, suitable for data-sharing within the project and dissemination to the wider community. The outputs will include metabolic network data for a range of insects with well-developed genomic resources.

6.2 NEMO - Network Motifs - 2009 a atual

<http://nemo.ssbgroup.fr/>

Desenvolvido na Université Claude Bernanrd - Lyon I em parceria com a AgroParis-Tech/INRA, Paris e CNRS/Université d'Evry /INRA.

A seguir um resumo em inglês retirado da página do projeto.

Recent technical advances have allowed the collection of data which are structured like networks. Many scientific fields are concerned by this network revolution such as physics, sociology and biology. The mathematical analysis of such networks aims at uncovering structures which can help in the understanding of the system. Global and local strategies can be considered and our project focuses on the latter. Studying substructures, called network motifs, is of particular interest in biology, since these building blocks constitute functional units which combine to ensure cell regulatory processes.

From a mathematical point of view, the first focus of this project is the study of the statistical distribution of motif counts in random graphs to detect the most interesting motifs of a given network. We will propose relevant definitions of network motifs and efficient algorithms to enumerate their occurrences. We will then propose random graph models fitting the global properties of observed biological networks and propose statistical methods to estimate their parameters. We will finally derive the exact or approximate distribution of the

motif counts in these models. A public software will be developed to make all these results available to biologists. As motifs can be considered as the building blocks of networks, we will develop, in a second phase, network comparison procedures based on motif counts. We will use motifs to confront networks to other sources, e.g. gene expression data: Gaussian graphical models will provide us a natural framework to achieve this confrontation.

Members of the NeMo project are statisticians, discrete mathematicians and computer scientists whose task force will be gathered to answer theoretical questions in algorithmics, probability and statistics and to develop efficient computer software for biologists. As many biological processes are governed by interacting elements, we will study different biological networks, such as protein-protein interaction networks, regulatory networks and metabolic networks. Appropriate mathematical tools will be required in each case.

The project is organized into 5 tasks:

1. Definition and enumeration of network motifs
2. Realistic random graph models for biological networks
3. Assessing motif exceptionality
4. Network comparison
5. From network motifs to data integration

6.3 Computational Biology, Algorithmics, Combinatorial Optimization - 2006 a 2007

Desenvolvido no departamento de Ciência da Computação-IME-USP em parceria com a Université Claude Bernard Lyon I.

A seguir um resumo em inglês retirado do projeto de pesquisa enviado a CAPES (http://servicos.capes.gov.br/arquivos/avaliacao/estudos/dados1/2006/33002010/002/2006_002_33002010176P0_ProjPesq.pdf).

The research proposed will focus mainly on computational biology. The technological advances in biological related areas of recent years has allowed a huge amount of information to be produced. It is one of the main current challenges for biologists to be able to extract knowledge from all this data, and a challenge also for the computer scientists and mathematicians to participate with the biologists in this process. The analysis of this data involves, among many other issues, the formalization of problems and the search for efficient algorithms to solve them. This is a continuous process in the sense that the solutions produced by the algorithms will usually lead to new research that has for objective refining the models, and will therefore result in new problems and/or the development of increasingly more sophisticated algorithms. The main goal of the association between the HELIX and USP teams is to work on such types of questions: the formalization of problems coming from biology and the search for efficient algorithms to solve these problems.

6.4 Algorithmic Questions on Molecular Biology - 2004 a 2005

<http://www.ime.usp.br/~cris/inria/>

Desenvolvido no departamento de Ciência da Computação-IME-USP em parceria com a Université Claude Bernard Lyon I.

Foi através deste projeto que visitei pela primeira vez a Universidade Claude Bernard em Lyon-França em Dezembro de 2005.

A seguir um resumo em inglês retirado da página do projeto.

The research proposed here will mainly focus on computational biology. The technological advances in biology related areas allowed a huge amount of information to be extracted. It is one of the main current challenges for biologists to be able to extract knowledge from all this data and a challenge also for the computer scientists to help the biologists in this process. The analysis of this data involves, among many other issues, the formalization of problems and the search for efficient algorithms to solve these problems. This is a continuous process in the sense that the solutions produced by the algorithms will usually lead a refinement of the models and problem formulations, and will therefore require new algorithms. The main goal of this project is to work on these type of questions: the formalization of problems coming from biology and the search for efficient algorithms for these problems.

7 Participação em eventos

7.1 Metabolic Pathways Analysis-2009 - Leiden - Holanda

<http://www.lorentzcenter.nl/lc/web/2009/376/info.php3?wsid=376>

Excelente congresso com vários renomados pesquisadores que trabalham na área de análise de redes metabólicas.

Participei como ouvinte das palestras.

7.2 6th ISS Congress-2009 - Madison - EUA

<http://www.union.wisc.edu/symbiosis>

Poster: AcypiCyc (Acyrtosiphon pisum Cyc database): a model database to visualize and study the metabolic network underlying the pea aphid-Buchnera symbiosis.

O poster foi apresentado por meu colega e co-autor Stefano Colella.

7.3 II Pea Aphid Genome Annotation Workshop-2009 - Barcelona - Espanha

<http://aphid.cgenomics.org>

Poster: The Acyrthosiphon pisum Cyc database (AcypiCyc) created using a novel BioCyc Annotation Database System (CycADS): a useful tool to explore and study the pea aphid metabolism.

O poster foi apresentado por mim e meu colega e co-autor Stefano Colella.

7.4 8th International Symposium on Aphids-2009 - Catania - Itália

<http://www.8isasicily.net>

Poster: AcypiCyc (Acyrtosiphon pisum Cyc database) and CycADS (Cyc Annotation Database System): moving from genome sequence annotation to metabolic network analyses.

O poster foi apresentado por meu colega e co-autor Yvan Rahbé.

7.5 Réseau Français de Biologie Adaptative des Pucerons-2009 - Paris - França

Poster: AcypiCyc (Acyrtosiphon pisum Cyc database) and CycADS (Cyc Annotation Database System): moving from genome sequence annotation to metabolic network analyses.

O poster foi apresentado por meu colega e co-autor Stefano Colella.

7.6 I Pea Aphid Genome Annotation Workshop-2008 - Princeton - EUA

<http://www.as.miami.edu/AphidWorkshop/>

Poster: Annotation of metabolism genes: towards the implementation of an Acyrthosiphon pisum Cyc database (ApsCyc) to perform metabolic network analyses.

O poster foi apresentado por meu colega e co-autor Stefano Colella.

7.7 IPG (Integrative Post-Genomics)-2008 - Lyon - França

<http://bsmc.insa-lyon.fr/ipg/ipg08/contributedabstracts.html>

Poster: Development of the Acyrthosiphon pisum Cyc database (ApsCyc): from genome sequence to metabolic network analyses.

O poster foi apresentado por mim e meu colega e co-autor Stefano Colella.

7.8 BioCyc and Pathway Tools tutorial-2008 - Londres - Inglaterra

<http://brg.ai.sri.com/ptools/tutorial/2008-09-London.html>

Tutorial para entender melhor o funcionamento e utilização do aplicativo PathwayTools para a visualização de várias características da rede metabólica de um organismo.

Participei como ouvinte das palestras.

7.9 Journées Ouvertes Biologie Informatique Mathématiques-JOBIM-2008 - Lille - França

<http://www2.lifl.fr/jobim2008/>

Evento que reúne vários pesquisadores, principalmente da França, da área de bioinformática.

Participei como ouvinte das palestras.

8 Experiência didática

8.1 Instituto Brasileiro de Tecnologia Avançada, IBTA, Brasil - 01/2002 a 07/2003

Como professor titular II (pois já tinha especialização) ministrei aulas de Fundamentos de Software (prática e teoria) e Programação orientada a objetos (prática e teoria) para as turmas do primeiro ano dos cursos de Banco de Dados e Administração de Redes, da seguinte forma:

- Prática em Fundamentos de Software para a turma do 1o. semestre do curso de Banco de dados
- Teoria em Fundamentos de Software para a turma do 1o. semestre do curso de Banco de dados
- Prática em Fundamentos de Software para a turma do 1o. semestre do curso de Administração de redes
- Teoria em Fundamentos de Software para a turma do 1o. semestre do curso de Administração de redes
- Prática em Programação orientada a objetos para a turma do 2o. semestre do curso de Banco de dados
- Teoria em Programação orientada a objetos para a turma do 2o. semestre do curso de Banco de dados
- Prática em Programação orientada a objetos para a turma do 2o. semestre do curso de Administração de redes
- Teoria em Programação orientada a objetos para a turma do 2o. semestre do curso de Administração de redes

9 Experiência profissional em empresas

9.1 Universo OnLine Ltda.-UOL, Brasil - 04/2000 a 01/2004

Função: Analista programador e gerente de projetos

Departamento: desenvolvimento de software.

No início (8 meses) trabalhei como consultor terceirizado da empresa. Depois fui contratado como funcionário CLT.

Desenvolvi aplicativos para as áreas de Email e Cadastro de clientes da empresa.

Gerenciei projetos para a área de Atendimento ao Cliente (SAC).

9.2 GlobalSys-Bradesco, Brasil - 01/1999 a 04/2000

Função: Analista programador e suporte

Departamento: empresa terceirizada (Globalsys).

Trabalhei como consultor terceirizado da empresa.

Desenvolvi e dei suporte a aplicativos para uso em áreas internas do Bradesco.

9.3 Banco Finasa, Brasil - 06/1997 a 01/1999

Função: Analista programador

Departamento: microinformática.

Desenvolvi e dei suporte para os aplicativos de *Home Banking* do FINASA.

9.4 Vellozo Informática, Brasil - 09/1995 a 05/1997

Função: Consultor e sócio proprietário

Desenvolvi aplicativos e dei suporte e consultoria para empresas em Sorocaba e região.

9.5 ABR Sistemas, Brasil - 03/1992 a 08/1995

Função: diretor e sócio proprietário

Criei, gerenciei e administrei esta empresa para a comercialização de *software* e computadores em Sorocaba e região.

9.6 Quantum Computadores, Brasil - 10/1989 a 10/1991

Função: Analista programador de software básico

Departamento: Pesquisa e desenvolvimento.

Desenvolvimento da BIOS para um micro-computador PC 486.

Suporte e manutenção da BIOS de um micro-computador PC 386.

10 Considerações finais

Gostaria de finalizar dizendo que apesar de já estar há quase 3 anos no exterior, gostaria de fixar residência novamente no Brasil. Gostaria que meus filhos recebessem a cultura brasileira e fossem criados no Brasil. Acredito que estes anos no exterior foram muito gratificantes, mas já foram suficientes.

Gostaria de ressaltar também que apesar de ter uma carreira acadêmica muito boa, com doutorado no IME-USP e pós-doutorado na França, tive também experiências profissionais em empresas por muito tempo no Brasil. Acredito que esta minha experiência profissional em várias empresas pode ajudar a propiciar uma boa visão deste ramo aos alunos.

Além disto, devido a este meu passado profissional, tive contato com empresários e profissionais influentes (diretoria e vice-presidência) de empresas e indústrias no Brasil (UOL, Google e Andrews). Acredito que com isto possa ajudar a concretizar parcerias da universidade com empresas e alavancar projetos que podem se auto-sustentar através do financiamento destas empresas. Desta forma as empresas podem conseguir resultados expressivos e inovatórios, assim como a universidade pode conseguir aprimorar o conhecimento dos alunos e aproximá-los de problemas reais e do ambiente e estrutura das empresas.

11 Currículo Lattes



Augusto Fernandes Vellozo

possui graduação em Ciência da Computação pela Universidade Estadual de Campinas (1989), doutorado em Ciência da Computação pela Universidade de São Paulo (2007) e pós-doutorado em bioinformática na Université Lyon I (2007-2010). Tem experiência na área de Ciência da Computação, com ênfase em Teoria da Computação, atuando principalmente nos seguintes temas: alinhamento de sequências, biologia computacional, programação dinâmica, alinhamento com rearranjos, redes metabólicas, motifs e bioinformática.

(Texto informado pelo autor)

Última atualização em 08/03/2010


Endereço para acessar este CV:
<http://lattes.cnpq.br/2024974104338199>

Dados Pessoais

Nome	Augusto Fernandes Vellozo
Filiação	Eustáquio Fernandes Vellozo e Ignêz Dacal Vellozo
Nascimento	12/04/1967 - Avaré/SP - Brasil
Carteira de Identidade	16187494 SSP - SP - 10/09/1981
CPF	07430913881
Endereço residencial	10 rue Paul Cambon - Villeurbanne 69100, - França Telefone: 33 426766582
Endereço profissional	Universite Claude Bernarde Lyon I 43 bd du 11 novembre 1918 - Bât. Grégor Mendel - Villeurbanne 69622, - França
Endereço eletrônico	e-mail para contato : augustovmail-cnpq@yahoo.com.br e-mail alternativo : avellozo@gmail.com

Formação Acadêmica/Titulação

- | | |
|--------------------|---|
| 2009 | Pós-Doutorado.
Universite Claude Bernarde Lyon I, LYON I, Lyon, França |
| 2008 - 2009 | Pós-Doutorado.
Institut National des Sciences Appliquees de Lyon, INSAL*, Lyon, França |
| 2007 - 2008 | Pós-Doutorado.
Universite Claude Bernarde Lyon I, LYON I, Lyon, França |
| 2001 - 2007 | Doutorado em Ciencia da Computacao.
Universidade de São Paulo, USP, Sao Paulo, Brasil
Título: Alinhamento de sequências com rearranjos, Ano de obtenção: 2007

Orientador: Alair Pereira do Lago 
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior |
| 1997 - 1998 | Especialização em Informática e Sistemas de Informação.
Universidade de Sorocaba, UNISO, Sorocaba, Brasil
Título: Banco de dados orientado a objetos |

- 1995 - 1996** Especialização em Administração de marketing.
Universidade de Sorocaba, UNISO, Sorocaba, Brasil
Título: Marketing em pequenas empresas
- 1985 - 1989** Graduação em Ciência da Computação.
Universidade Estadual de Campinas, UNICAMP, Campinas, Brasil

Atuação profissional

1. Universite Claude Bernarde Lyon I - LYON I

Vínculo institucional

- 2009 - Atual** Vínculo: Celetista formal , Enquadramento funcional: Pós-doutorado , Carga horária: 40, Regime: Dedicação Exclusiva
- 2007 - 2008** Vínculo: Celetista formal , Enquadramento funcional: Pós-doutorado , Carga horária: 40, Regime: Dedicação Exclusiva

Atividades

- 07/2009 - Atual** Projetos de pesquisa, LBBE - UMR CNRS 5558 Laboratoire de Biométrie et Biologie évolutive

Participação em projetos:
MetNet - Systems level analysis of the Pea Aphid , NEMO - Network Motifs

2. Institut National des Sciences Appliquees de Lyon - INSAL*

Vínculo institucional

- 2008 - 2009** Vínculo: Celetista formal , Enquadramento funcional: Pós-doutorado , Carga horária: 40, Regime: Dedicação Exclusiva

Atividades

- 07/2008 - 07/2009** Projetos de pesquisa, Biologie Fonctionnelle Insectes et Interactions - BF2I

Participação em projetos:
MetNet - Systems level analysis of the Pea Aphid

3. Universidade de São Paulo - USP

Vínculo institucional

- 2004 - 2007** Vínculo: Outro (Aluno) , Enquadramento funcional: Aluno Doutorado

Atividades

- 2006 - 2007** Projetos de pesquisa, Instituto de Matemática e Estatística, Departamento de Ciência da Computação

Participação em projetos:
Computational Biology, Algorithmics, Combinatorial Optimization

- 2004 - 2007** Projetos de pesquisa, Instituto de Matemática e Estatística, Departamento de Ciência da Computação

Participação em projetos:
Algorithmic Questions on Molecular Biology

4. Instituto Brasileiro de Tecnologia Avançada - IBTA

**Vínculo
institucional**

2002 - 2003 Vínculo: Celetista , Enquadramento funcional: Professor titular II , Carga horária: 12, Regime: Parcial

Atividades

02/2002 - 06/2003 Graduação, Administração de redes

*Disciplinas Ministradas:
Programação orientada a objetos , Fundamentos de software*

02/2002 - 06/2003 Graduação, Banco de Dados

*Disciplinas Ministradas:
Fundamentos de software , Programação orientada a objetos*

5. Universo OnLine Ltda. - UOL**Vínculo
institucional**

2000 - 2004 Vínculo: Celetista , Enquadramento funcional: Analista programador , Carga horária: 40, Regime: Integral

Atividades

04/2000 - 01/2004 Serviço Técnico Especializado, Departamento de desenvolvimento de software

*Especificação:
Desenvolvimento de software*

6. Banco Bradesco - BRADESCO**Vínculo institucional**

1999 - 2000 Vínculo: Colaborador , Enquadramento funcional: Analista programador , Carga horária: 40, Regime: Dedicção Exclusiva

7. Banco Finasa - FINASA**Vínculo institucional**

1997 - 1999 Vínculo: Celetista , Enquadramento funcional: Analista programador , Carga horária: 40, Regime: Dedicção Exclusiva

8. Vellozo Informática - VELLOZO**Vínculo institucional**

1995 - 1997 Vínculo: Proprietário , Enquadramento funcional: Analista programador , Carga horária: 40, Regime: Dedicção Exclusiva

9. ABR Sistemas - ABR**Vínculo
institucional**

1992 - 1995 Vínculo: Sócio , Enquadramento funcional: Diretor , Carga horária: 40, Regime: Dedicção Exclusiva

10. Quantum Computadores - QUANTUM**Vínculo institucional**

1989 - 1991 Vínculo: Celetista , Enquadramento funcional: Analista programador , Carga horária: 40, Regime: Dedicção Exclusiva

Projetos**2009 - 2010** MetNet - Systems level analysis of the Pea Aphid

Descrição: From a metabolic perspective, an animal comprises a set of metabolic compartments (including organs) connected by the inter-compartment transfer of metabolites. We seek to construct and analyse multi-compartment models of metabolism, by transforming a single global metabolic network into a set of compartment networks of overlapping composition using flux balance analysis. In parallel, we will develop and apply two computational approaches for network analysis: an exploration of metabolic network motifs, and elementary modes to characterize and compare networks. Large inter-compartment metabolic differences facilitating model development will be achieved in our system comprising two distinct genomes within a single animal: the pea aphid *Acyrtosiphon pisum* and its symbiotic bacteria *Buchnera* housed in a special organ, the bacteriome. The first model comprises two compartments, describing how the insect and bacterial networks mediate the transformation of precisely known diet components to yield biomass production (and waste); and subsequent models are of increasing complexity, involving the subdivision of the insect network into overlapping, but distinct, networks of the contributing compartments/organs. Model construction will be informed by empirical data collected on the aphid transcriptome, *Buchnera* proteome and embryo nutrient budgets within this project, and the model outputs will be validated against whole-insect metabolic and nutritional datasets obtained previously from the investigators' labs. Algorithms and models developed in this project will be maintained in a central Repository linked to pre-existing databases, suitable for data-sharing within the project and dissemination to the wider community. The outputs will include metabolic network data for a range of insects with well-developed genomic resources.

Situação: Em Andamento Natureza: Pesquisa

Integrantes: Augusto Fernandes Vellozo; Marie France Sagot (Responsável)

Financiador(es): Agence Nationale de la Recherche-ANR

2009 - 2010 NEMO - Network Motifs

Descrição: The project is organized into 5 tasks: 1. Definition and enumeration of network motifs 2. Realistic random graph models for biological networks 3. Assessing motif exceptionality 4. Network comparison 5. From network motifs to data integration

Situação: Em Andamento Natureza: Pesquisa

Integrantes: Augusto Fernandes Vellozo; Marie France Sagot (Responsável)

Financiador(es):

2008 - 2009 MetNet - Systems level analysis of the Pea Aphid

Descrição: From a metabolic perspective, an animal comprises a set of metabolic compartments (including organs) connected by the inter-compartment transfer of metabolites. We seek to construct and analyse multi-compartment models of metabolism, by transforming a single global metabolic network into a set of compartment networks of overlapping composition using flux balance analysis. In parallel, we will develop and apply two computational approaches for network analysis: an exploration of metabolic network motifs, and elementary modes to characterize and compare networks. Large inter-compartment metabolic differences facilitating model development will be achieved in our system comprising two distinct genomes within a single animal: the pea aphid *Acyrtosiphon pisum* and its symbiotic bacteria *Buchnera* housed in a special organ, the bacteriome. The first model comprises two compartments, describing how the insect and bacterial networks mediate the transformation of precisely known diet components to yield biomass production (and waste); and subsequent models are of increasing complexity, involving the subdivision of the insect network into overlapping, but distinct, networks of the contributing compartments/organs. Model construction will be informed by empirical data collected on the aphid transcriptome, *Buchnera* proteome and embryo nutrient budgets within this project, and the model outputs will be validated against whole-insect metabolic and nutritional datasets obtained previously from the investigators' labs. Algorithms and models developed in this project will be maintained in a central Repository linked to pre-existing databases, suitable for data-sharing within the project and dissemination to the wider community. The outputs will include metabolic network data for a range of insects with well-developed genomic resources.

Situação: Concluído Natureza: Pesquisa

Integrantes: Augusto Fernandes Vellozo; Marie France Sagot (Responsável)

Financiador(es): Agence Nationale de la Recherche-ANR

2006 - 2007 Computational Biology, Algorithmics, Combinatorial Optimization

Descrição: The research proposed will focus mainly on computational biology. The technological advances in biological related areas of recent years has allowed a huge amount of information to be produced. It is one of the main current challenges for biologists to be able to extract knowledge from all this data, and a challenge also for the computer scientists and mathematicians to participate with the biologists in this process. The analysis of this data involves, among many other issues, the formalization of problems and the search for efficient algorithms to solve them. This is a continuous process in the sense that the solutions produced by the algorithms will usually lead to new research that has for objective refining the models, and will therefore result in new problems and/or the development of increasingly more sophisticated algorithms. The main goal of the association between the HELIX and USP teams is to work on such types of questions: the formalization of problems coming from biology and the search for efficient algorithms to solve these problems.

Situação: Concluído Natureza: Pesquisa

Alunos envolvidos: Graduação (0); Especialização (0); Mestrado acadêmico (0); Mestrado profissionalizante (0); Doutorado (2);

Integrantes: Augusto Fernandes Vellozo; Yoshiko Wakabayashi (Responsável)

Financiador(es): Fundação de Amparo à Pesquisa do Estado de São Paulo-FAPESP

2004 - 2005 Algorithmic Questions on Molecular Biology

Descrição: The research proposed here will mainly focus on computational biology. The technological advances in biology related areas allowed a huge amount of information to be extracted. It is one of the main current challenges for biologists to be able to extract knowledge from all this data and a challenge also for the computer scientists to help the biologists in this process. The analysis of this data involves, among many other issues, the formalization of problems and the search for efficient algorithms to solve these problems. This is a continuous process in the sense that the solutions produced by the algorithms will usually lead a refinement of the models and problem formulations, and will therefore require new algorithms. The main goal of this project is to work on these type of questions: the formalization of problems coming from biology and the search for efficient algorithms for these problems. <http://www.ime.usp.br/~cris/inria/>

Situação: Concluído Natureza: Pesquisa

Alunos envolvidos: Graduação (2); Especialização (0); Mestrado acadêmico (0); Mestrado profissionalizante (2); Doutorado (2);

Integrantes: Augusto Fernandes Vellozo; Carlos Eduardo Ferreira (Responsável)

Financiador(es): Fundação de Amparo à Pesquisa do Estado de São Paulo-FAPESP

Áreas de atuação

1. Biologia Computacional
2. Teoria da Computação
3. Sistemas de Computação
4. Metodologia e Técnicas da Computação


Idiomas

Inglês Compreende Razoavelmente , Fala Razoavelmente, Escreve Razoavelmente, Lê Bem


Francês Compreende Bem , Fala Razoavelmente, Escreve Razoavelmente, Lê Bem

Português Compreende Bem , Fala Bem, Escreve Bem, Lê Bem


Produção em C, T & A**Produção bibliográfica****Artigos completos publicados em periódicos**

1. ★  The International Aphid Genomics Consortium, Augusto Vellozo
Genome Sequence of the Pea Aphid *Acyrtosiphon pisum*. PLoS Biology. , v.8, p.e1000313 - , 2010.
Referências adicionais : Inglês. Meio de divulgação: Meio digital, Home page: [<http://dx.doi.org/10.1371%2Fjournal.pbio.1000313>]
[doi:10.1371/journal.pbio.1000313]

Trabalhos publicados em anais de eventos (completo)

1. ★  Augusto F. Vellozo, Carlos e Alves, Alair Pereira do Lago
Alignment with non-overlapping inversions in $O(n^3 \log n)$ -time In: WABI, 2006, Zurique.
Algorithms in Bioinformatics. Berlin / Heidelberg: Springer, 2006. v.4175. p.186 - 196
Referências adicionais : Suíça/Inglês. Meio de divulgação: Vários, Home page: [<http://www.springerlink.com/content/w1gw8u371006g603/>]

Trabalhos publicados em anais de eventos (resumo expandido)

1. ★  Carlos e Alves, Alair Pereira do Lago, Augusto F. Vellozo
Alignment with non-overlapping inversions in $O(n^3 \log n)$ -time In: GRACO, 2005, Angra dos Reis.
Electronic Notes in Discrete Mathematics - Proceedings of GRACO2005. Amsterdam: Elsevier, 2005. v.19. p.365 - 371
Referências adicionais : Holanda/Inglês. Meio de divulgação: Vários, Home page: [<http://www.sciencedirect.com/science/article/B75gv-4g8mt0t-1p/215ca35c3828cd2c20967326055dbf2f7>]

Produção Técnica**Softwares sem registro ou patente**

1. ★ Augusto Vellozo, Stefano Colella
Cycads - Cyc Annotation Database system, 2009

Demais produções técnicas

1. ★ Augusto Vellozo, Stefano Colella
AcypiCyc - Acyrthosiphon pisum Cyc database, 2009. (Outra produção técnica)
Referências adicionais : França/Inglês. Meio de divulgação: Meio digital
<http://acypicyc.cycadsys.org> <http://acypicyc.cycadsys.org> Abstract: The AcypiCyc database has been generated at the Université de Lyon in a collaboration between the UMR203 INRA-INSA de Lyon, Biologie fonctionnelle des insectes et interactions (BF2I) and the Baobab team at UMR5558 LBBE CNRS-UCBL1; both groups are member of the INRIA Bamboo team-project. The database is hosted at the Pôle Rhône-Alpes de Bioinformatique (PRABI). We are working on a version that includes KAAS, PRIAM, BLAST2GO and PhylomeDB annotations. We will add more annotations soon. At present, you will be able to compare \emph{Acyrthosiphon pisum} metabolism to \emph{Buchnera aphidicola} APS REVISED (generated in Lyon inside the SymbioCyc database).

Eventos**Participação em eventos**

1. Apresentação de Poster / Painel no(a) **6th ISS Congress.**, 2009. (Congresso)
AcypiCyc (Acyrthosiphon pisum Cyc database): a model database to visualize and study the metabolic network underlying the pea aphid-Buchnera symbiosis.
2. Apresentação de Poster / Painel no(a) **Réseau Français de Biologie Adaptative des Pucerons**, 2009. (Simpósio)
AcypiCyc (Acyrthosiphon pisum Cyc database) and CycADS (Cyc Annotation Database System): moving from genome sequence annotation to metabolic network analyses..
3. Apresentação de Poster / Painel no(a) **8th International Symposium on Aphids**, 2009. (Simpósio)
AcypiCyc (Acyrthosiphon pisum Cyc database) and CycADS (Cyc Annotation Database System): moving from genome sequence annotation to metabolic network analyses..
4. Apresentação de Poster / Painel no(a) **II Pea Aphid Genome Annotation Workshop**, 2009. (Simpósio)
The Acyrthosiphon pisum Cyc database (AcypiCyc) created using a novel BioCyc Annotation Database System (CycADS): a useful tool to explore and study the pea aphid metabolism..
5. **Metabolic Pathways Analysis 2009**, 2009. (Simpósio)
.
6. Apresentação de Poster / Painel no(a) **Pea Aphid Genome Annotation Workshop I**, 2008. (Simpósio)
Annotation of metabolism genes: towards the implementation of an Acyrthosiphon pisum Cyc database (ApsCyc) to perform metabolic network analyses.
7. Apresentação de Poster / Painel no(a) **IPG (Integrative Post-Genomics)**, 2008. (Simpósio)
Development of the Acyrthosiphon pisum Cyc database (ApsCyc): from genome sequence to metabolic network analyses.

8. **BioCyc and Pathway Tools tutorial**, 2008. (Oficina)

.

9. **JOBIM**, 2008. (Congresso)

.

Página gerada pelo Sistema Currículo Lattes em 08/03/2010 às 10:28:15.

Referências

- [1] C. E. R. Alves, A. P. do Lago, and A. F. Vellozo. Alignment with non-overlapping inversions in $O(n^3 \log n)$ -time. In *Proceedings of GRACO2005*, volume 19 of *Electron. Notes Discrete Math.*, pages 365–371 (electronic), Amsterdam, 2005. Elsevier.
- [2] C. E. R. Alves, A. P. do Lago, and A. F. Vellozo. Alignment with non-overlapping inversions in $O(n^3 \log n)$ -time. *Electronic Notes in Discrete Mathematics*, 19:365–371, Jun 2005.
- [3] Cerdeño-Tárraga, Patrick, Crossman, Blakely, Abratt, Lennard, Poxton, Duerden, Harris, Quail, Barron, Clark, Corton, Doggett, Holden, Larke, Line, Lord, Norbertczak, Ormond, Price, Rabbino-witsch, Woodward, Barrell, and Parkhill. Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science*, 307(5714):1463–1465, Mar 2005.
- [4] C. Claudel-Renard, C. Chevalet, T. Faraut, and D. Kahn. Enzyme-specific profiles for genome annotation: PRIAM. *Nucl. Acids Res.*, 31(22):6633–6639, 2003.
- [5] A. Conesa, S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon, and M. Robles. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18):3674–6, 2005.
- [6] A. T. I. A. G. Consortium. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol*, 8(2):e1000313, 02 2010.
- [7] L. Cottret, P. Vieira Milreu, V. Acuña, A. Marchetti-Spaccamela, F. Viduani Martinez, M.-F. Sagot, and L. Stougie. Enumerating precursor sets of target metabolites in a metabolic network. In *WABI '08: Proceedings of the 8th international workshop on Algorithms in Bioinformatics*, pages 233–244, Berlin, Heidelberg, 2008. Springer-Verlag.
- [8] Cáceres, Ranz, Barbadilla, Long, and Ruiz. Generation of a widespread *Drosophila* inversion by a transposable element. *Science*, 285(5426):415–418, Jul 1999.
- [9] A. P. do Lago, C. A. Kulikowski, E. Linton, J. Messing, and I. Muchnik. Comparative genomics: simultaneous identification of conserved regions and their rearrangements through global optimization. In *The Second University of Sao Paulo/Rutgers University Biotechnology Conference*, Rutgers University Inn and Conference Center, New Brunswick, NJ, August 2001.
- [10] A. P. do Lago, I. Muchnik, and C. Kulikowski. An $O(n^4)$ algorithm for alignment with non-overlapping inversions. In *Second Brazilian Workshop on Bioinformatics, WOB 2003*, Macaé, RJ, Brazil, 2003. <http://www.ime.usp.br/~alair/wob03.pdf>.
- [11] A. P. do Lago, I. Muchnik, and C. Kulikowski. A sparse dynamic programming algorithm for alignment with non-overlapping inversions. *Theor. Inform. Appl.*, 39(1):175–189, 2005.
- [12] Feuk, MacDonald, Tang, Carson, Li, Rao, Khaja, and Scherer. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *PLoS Genet*, 1(4):e56, Oct 2005.

- [13] Y. Gao, J. Wu, R. Niewiadowski, Y. Wang, Z.-Z. Chen, and G. Lin. A space efficient algorithm for sequence alignment with inversions. In *Computing and Combinatorics, 9th Annual International Conference, COCOON 2003*, volume 2697 of *Lecture Notes in Computer Science*, pages 57–67. Springer-Verlag, 2003.
- [14] Graham and Olmstead. Evolutionary significance of an unusual chloroplast DNA inversion found in two basal angiosperm lineages. *Curr Genet*, 37(3):183–188, Mar 2000.
- [15] J. Huerta-Cepas, A. Bueno, J. Dopazo, and T. Gabaldon. PhylomeDB: a database for genome-wide collections of gene phylogenies. *Nucl. Acids Res.*, page gkm899, 2007.
- [16] P. D. Karp, S. M. Paley, M. Krummenacker, M. Latendresse, J. M. Dale, T. J. Lee, P. Kaipa, F. Gilham, A. Spaulding, L. Popescu, T. Altman, I. Paulsen, I. M. Keseler, and R. Caspi. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform*, 11(1):40–79, 2010.
- [17] Kuwahara, Yamashita, Hirakawa, Nakayama, Toh, Okada, Kuhara, Hattori, Hayashi, and Ohnishi. Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. *Proceedings of the National Academy of Sciences U S A*, 101(41):14919–14924, Oct 2004.
- [18] V. Lacroix, C. G. Fernandes, and M.-F. Sagot. Motif search in graphs: Application to metabolic networks. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 3(4):360–368, 2006.
- [19] Y. Moriya, M. Itoh, S. Okuda, A. C. Yoshizawa, and M. Kanehisa. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*, 35(Web Server issue):W182–5, 2007.
- [20] S. Schuster and C. Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems (JBS)*, 2(2):165–182, 1994.
- [21] M. Schöniger and M. S. Waterman. A local algorithm for DNA sequence alignment with inversions. *Bulletin of Mathematical Biology*, 54(4):521–536, Jul 1992.
- [22] A. F. Vellozo. *Alinhamento de sequências com rearranjos*. PhD thesis, USP- Universidade de São Paulo, Abril 2007.
- [23] A. F. Vellozo, C. E. R. Alves, and A. P. do Lago. Alignment with non-overlapping inversions in $o(n^3)$ -time. In *6th Workshop on Algorithms in Bioinformatics*. Springer, 2006. Lecture Notes in Bioinformatics 4175.
- [24] A. F. Vellozo and S. Colella. Acypicyc - *Acyrtosiphon pisum* cyc database. 2009. <http://acypicyc.cycadsys.org>.
- [25] A. F. Vellozo and S. Colella. Cycads - cyc annotation database system. 2009. <http://code.cycadsys.org>.