A Bioinformática é uma nova ciência que tem por objectivo desenvolver soluções eficientes para o armazenamento, comunicação e análise de bioinformação.

A bioinformação corresponde aos dados biológicos em sentido lato, mas actualmente centra-se sobretudo em informação do ADN, ARN, proteínas e análise de expressão génica em microarrays.

Workshop em Bioinformática

Na era pós-genómica é fundamental o contributo da bioinformática para desvendar os segredos funcionais que a bioinformação codifica.

Programa do WB'2015

Quarta-feira, dia 4 de novembro de 2015
Auditório Pequeno da ESA/IPB:

09:00 - Sessão de Abertura, com a presença do Exmo. Diretor da Escola Superior Agrária de Bragança - Professor Doutor Albino Bento

09:15 - *O que é e para que serve a bioinformática?*
Sérgio Deusdado, ESA/IPB e CIMO

10:00 - *"Docking" Molecular na elucidação do mecanismo de ação de derivados de thieno[3,2-b]pyridinas como inibidores da tirosina cinase VEGFR2*
Ricardo C. Calhelha, Isabel C.F.R. Ferreira,
Maria-João R.P. Queiroz, Rui M.V Abreu, ESA/IPB e CIMO, CQ-UM

10:45 - Cofee break

11:00 - *Desempenho de ferramentas in silico: Avaliação de variantes de mutações missense do gene UGT1A1*
Carina Rodrigues, ESSa/IPB,
Alice Santos-Silva, Elísio Costa, Elsa Bronze-da-Rocha,
UCIBIO/REQUIMTE, Fac. de Farmácia, Univ. do Porto

11:45 - *Simulação Computacional do Contágio da Gripe Pessoa a Pessoa*
Ana Araújo, Carlos Balsa e João Paulo Almeida, ESTIG/IPB

12:30 - ALMOÇO

Sessões Hands On no Laboratório de Informática do CIESA:

14:30 - *Utilização de métodos informáticos para desenvolvimento de novos fármacos*
Rui M. V. Abreu, ESA/IPB e CIMO

16:00 - *Análise bioinformática da estrutura e função da informação biológica*
Altino Choupina e Sérgio Deusdado, ESA/IPB e CIMO

Quinta-feira, dia 5 de novembro de 2015
Auditório Pequeno da ESA/IPB:

09:15- *NGS de Genomas e Exomas Humanos*
Hugo J.C. Froufe e Conceição Egas, GENOINSEQ, BiocantPark

09:45 - *Hvar: Base de Dados de Variações Humanas*
Joana V. S. Sousa, Hugo J. C. Froufe, Conceição Egas e Paulo J. Novais, GENOINSEQ, BiocantPark e Dep. Informática, Univ. do Minho

10:15- *seqQI: Indicador de qualidade para RNA-Seq.*
Abel. E. F. Sousa, Hugo. J. C. Froufe, Conceição Egas e Rui Mendes
GENOINSEQ, BiocantPark e Dep. Informática, Univ. do Minho

10:45 - Cofee break

11:00 - *Utilização de redes de Bayes na descoberta de tendências de dados*
Pedro Bastos, ESA/IPB e CIMO

11:45 - *Desempenho comparativo do BLAST e do mpiBLAST*
José Rufino, ESTIG/IPB

12:30 - ALMOÇO

14:30 - *Análise da expressão genética com silenciamento de genes por iRNA*
Rodrigo Costa, Univ. de Salamanca e Altino Choupina, ESA/IPB e CIMO

Sessão Hands On no Laboratório de Informática do CIESA:

15:15 - *Anotação e Análise de Variantes Humanas*
Hugo J.C. Froufe e Conceição Egas, GENOINSEQ, BiocantPark

16:45 - *Encerramento*

ipb INSTITUTO POLITÉCNICO DE BRAGANÇA Escola Superior Agrária

ipb INSTITUTO POLITÉCNICO DE BRAGANÇA Escola Superior Agrária

www.esa.ipb.pt/bioinformatica

Comissão Organizadora:
Sérgio Deusdado - sergiod@ipb.pt
Altino Choupina - albracho@ipb.pt
Lurdes Jorge - lurdesjo@ipb.pt
Rui Abreu - ruiabreu@ipb.pt

INSTITUTO POLITÉCNICO DE BRAGANÇA Escola Superior Agrária

INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior Agrária

6ª Edição

Workshop em Bioinformática

LIVRO DE RESUMOS

4 e 5 de novembro de 2015

Escola Superior Agrária de Bragança

Apoios:

INSTITUTO POLITÉCNICO DE BRAGANÇA Escola Superior Agrária

CIMO Centro de Investigação de Montanha

Desempenho Comparativo do BLAST e do mpiBLAST

José Rufino

Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Bragança, Bragança, Portugal.

Em Bioinformática são frequentes problemas cujo tratamento necessita de considerável poder de processamento/cálculo e/ou grande capacidade de armazenamento de dados e elevada largura de banda no acesso aos mesmos (de forma não comprometer a eficiência do seu processamento). Um exemplo deste tipo de problemas é a busca de regiões de similaridade em sequências de amino-ácidos de proteínas, ou em sequências de nucleótidos de DNA, por comparação com uma dada sequência fornecida (query sequence). Neste âmbito, a ferramenta computacional porventura mais conhecida e usada é o BLAST (Basic Local Alignment Search Tool) [1]. Donde, qualquer incremento no desempenho desta ferramenta tem impacto considerável (desde logo positivo) na atividade de quem a utiliza regularmente (seja para investigação, seja para fins comerciais).

Precisamente, desde que o BLAST foi inicialmente introduzido, foram surgindo diversas versões, com desempenho melhorado, nomeadamente através da aplicação de técnicas de paralelização às várias fases do algoritmo (e.g., partição e distribuição das bases de dados a pesquisar, segmentação das queries, etc.), capazes de tirar partido de diferentes ambientes computacionais de execução paralela, como: máquinas multi-core (BLAST+ [2]), clusters de nós multi-core (mpiBLAST[3]) e, mais recentemente, co-processadores aceleradores como GPUs[4] ou FPGAs. É também possível usar as ferramentas da família BLAST através de um interface/sítio WEB[5], que permite, de forma expedita, a pesquisa de uma variedade de bases de dados conhecidas (e em permanente atualização), com tempos de resposta suficientemente pequenos para a maioria dos utilizadores, graças aos recursos computacionais de elevado desempenho que sustentam o seu backend. Ainda assim, esta forma de utilização do BLAST poderá não ser a melhor opção em algumas situações, como por exemplo quando as bases de dados a pesquisar ainda não são de domínio público, ou, sendo-o, não estão disponíveis no referido sítio WEB. Adicionalmente, a utilização do referido sítio como ferramenta de trabalho regular pressupõe a sua disponibilidade permanente (dependente de terceiros) e uma largura de banda de qualidade suficiente, do lado do cliente, para uma interacção eficiente com o mesmo. Por estas razões, poderá ter interesse (ou ser mesmo necessário) implantar uma infra-estrutura BLAST local, capaz de albergar as bases de dados pertinentes e de suportar a sua pesquisa da forma mais eficiente possível, tudo isto levando em conta eventuais constrangimentos financeiros que limitam o tipo de hardware usado na implementação dessa infra-estrutura.

Neste contexto, foi realizado um estudo comparativo de diversas versões do BLAST, numa infra-estrutura de computação paralela do IPB, baseada em componentes commodity: um cluster de 8 nós (virtuais, sob VMWare ESXi) de computação (com CPU i7-4790K 4GHz, 32GB RAM e 128GB SSD) e um nó dotado de uma GPU (CPU i7-2600 3.8GHz, 32GB RAM, 128 GB SSD, 1 TB HD, NVIDIA GTX 580). Assim, o foco principal incidiu na avaliação do desempenho do BLAST original e do mpiBLAST, dado que são fornecidos de base na distribuição Linux em que assenta

o cluster [6]. Complementarmente, avaliou-se também o BLAST+ e o gpuBLAST no nó dotado de GPU. A avaliação contemplou diversas configurações de recursos, incluindo diferentes números de nós utilizados e diferentes plataformas de armazenamento das bases de dados (HD, SSD, NFS). As bases de dados pesquisadas correspondem a um subconjunto representativo das disponíveis no sítio WEB do BLAST, cobrindo uma variedade de dimensões (desde algumas dezenas de MBytes, até à centena de GBytes) e contendo quer sequências de amino-ácidos (env_nr e nr), quer de nucleótidos (drosohp.nt, env_nt, mito.nt, nt e patnt). Para as pesquisas foram usadas sequências arbitrárias de 568 letras em formato FASTA, e adoptadas as opções por omissão dos vários aplicativos BLAST. Salvo menção em contrário, os tempos de execução considerados nas comparações e no cálculo de speedups são relativos à primeira execução de uma pesquisa, não sendo assim beneficiados por qualquer efeito de cache; esta opção assume um cenário real em que não é habitual que uma mesma query seja executada várias vezes seguidas (embora possa ser re-executada, mais tarde).

As principais conclusões do estudo comparativo realizado foram as seguintes:

- é necessário acautelar, à priori, recursos de armazenamento com capacidade suficiente para albergar as bases de dados nas suas várias versões (originais/compactadas, descompactadas e formatadas); no nosso cenário de teste a coexistência de todas estas versões consumiu 600GBytes;
- o tempo de preparação (formatação) das bases de dados para posterior pesquisa pode ser considerável; no nosso cenário experimental, a formatação das bases de dados mais pesadas (nr, env_nt e nt) demorou entre 30m a 40m (para o BLAST), e entre 45m a 55m (para o mpiBLAST);
- embora economicamente mais onerosos, a utilização de discos de estado sólido, em alternativa a discos rígidos tradicionais, permite melhorar o tempo da formatação das bases de dados; no entanto, os benefícios registados (à volta de 9%) ficam bastante aquém do inicialmente esperado;
- o tempo de execução do BLAST é fortemente penalizado quando as bases de dados são acedidas através da rede, via NFS; neste caso, nem sequer compensa usar vários cores; quando as bases de dados são locais e estão em SSD, o tempo de execução melhora bastante, em especial com a utilização de vários cores; neste caso, com 4 cores, o speedup chega a atingir 3.5 (sendo o ideal 4) para a pesquisa de BDs de proteínas, mas não passa de 1.8 para a pesquisa de BDs de nucleótidos;
- o tempo de execução do mpiBLAST é muito prejudicado quando os fragmentos das bases de dados ainda não se encontram nos nós do cluster, tendo que ser distribuídos previamente à pesquisa propriamente dita; após a distribuição, a repetição das mesmas queries beneficia de speedups de 14 a 70; porém, como a mesma base de dados poderá ser usada para responder a diferentes queries, então não é necessário repetir a mesma query para amortizar o esforço de distribuição;
- no cenário de teste, a utilização do mpiBLAST com 32+2 cores, face ao BLAST com 4 cores, traduz-se em speedups que, conforme a base de dados pesquisada (e previamente distribuída), variam entre 2 a 5, valores aquém do máximo teórico de 8.5 (34/4), mas ainda assim demonstradores de que, havendo essa possibilidade, compensa realizar as pesquisas em cluster;

- a comparação do BLAST com o BLAST+ (que tal como o BLAST tem capacidade de explorar vários cores) e com o gpuBLAST, realizada no nó com GPU (representativo de uma workstation típica), permite aferir qual a melhor opção no caso de não serem possíveis pesquisas em cluster; as observações realizadas indicam que não há diferenças significativas entre o BLAST e o BLAST+; adicionalmente, o desempenho do gpuBLAST foi sempre pior (aproximadmente em 50%) que o do BLAST e BLAST+, o que pode encontrar explicação na longevidade do modelo da GPU usada;
- finalmente, a comparação da melhor opção no nosso cenário de teste, representada pelo uso do mpiBLAST, com o recurso a pesquisa online, no site do BLAST[5], revela que o mpiBLAST apresenta um desempenho bastante competitivo com o BLAST online, chegando a ser claramente superior se se considerarem os tempos do mpiBLAST tirando partido de efeitos de cache; esta assunção acaba por se justa, já que BLAST online também rentabiliza o mesmo tipo de efeitos; no entanto, com tempos de pequisa tão reduzidos (< 30s), só é defensável a utilização do mpiBLAST numa infra-estrutura local se o objetivo for a pesquisa de Bds não pesquisáveis via BLAS+ online;

1) Altschul, Stephen; Gish, Warren; Miller, Webb; Myers, Eugene; Lipman, David. "Basic local alignment search tool". Journal of Molecular Biology 215 (3): 403–410. 1990.
2) Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. "BLAST+: architecture and applications". BMC Bioinformatics 10:421. 2008.
3) Darling, A.; Carey, L.; Feng, W. "The Design, Implementation, and Evaluation of mpiBLAST". 4th International Conference on Linux Clusters: The HPC Revolution 2003 in conjunction with ClusterWorld Conference & Expo, June 2003.
4) Vouzis, Panagiotis D.; Sahinidis, Nikolaos V. "GPU-BLAST: using graphics processors to accelerate protein sequence alignment". Vol. 27, no. 2, pages 182-188, Bioinformatics, 2011 (Open Access).
5) BLAST: Basic Local Alignment Search Tool - http://blast.ncbi.nlm.nih.gov/
6) ROCKS Bio Roll - http://central6.rocksclusters.org/roll-documentation/bio/6.1/

Análise da expressão genética com silenciamento de genes por iRNA.

Rodrigo Costa[1] e Altino Choupina[2]

1) Universidade de Salamanca
2) Instituto Politécnico de Bragança, Campus de Santa Apolónia, Bragança, Portugal, CIMO-Centro de Investigação de Montanha, Campus de Santa Apolónia, Bragança, Portugal.

Introdução

Uma das maiores descobertas do ENCODE Pilot Project foi que "quase o genoma inteiro pode ser representado em transcritos primários que extensivamente se sobrepõem e incluem muitas regiões não-codificantes de proteínas. Um RNA não-codificante (ncRNA, em inglês) é qualquer molécula de RNA que não é traduzida em proteína. Moléculas de alguns destes RNA's não codificantes inibem a expressão génica causando a destruição de moléculas de mRNA específicos. Estes ncRNA's incluem os microRNA (miRNA) e small interfering RNA (siRNA).

Os miRNA são produtos da transcrição de genes presentes em muitos eucariotas com cerca de 22 nucleotídeos, capazes de se parear e formar estruturas do tipo hairpin. siRNA são derivados de longas moléculas de RNA dupla fita de origem exógena (como aquelas provenientes de vírus de RNA).

A enzima Dicer corta RNA de dupla fita, de modo a formar siRNA ou miRNA. Estes RNAs processados são incorporados no complexo RISC, o qual tem como alvo moléculas de RNA mensageiro, onde atuam impedindo o processo de tradução.

Quando o pareamento entre a fita guia e o mRNA envolve diversas bases, gerando um pareamento efetivo, este mRNA será degradado pela ação catalítica de uma das subunidades de RISC: a enzima denominada Argonauta. Quando o pareamento entre a fita guia e o mRNA alvo ocorre de maneira parcial, RISC não promove a clivagem do mRNA, mas atua inibindo o processo de tradução deste. Nesta condição, o mRNA desestabilizado pode ser conduzido aos chamados corpos de processamento (corpos-P), estruturas citosólicas responsáveis pela degradação de mRNA.

Funções Biológicas - Imunidade

RNA de interferência é uma parte vital da resposta imune a vírus ou outro material genético estranho, especialmente em plantas onde se pode também prevenir a auto-propagação de transposões. O silenciamento de genes induzido em plantas poderia espalhar-se por toda a planta mesmo através de enxertia.

No nosso laboratório obtivemos resultados satisfatórios no silenciamento do gene NPP1 em Phytophthora cinnamomi por meio da introdução de um intrão entre fragmentos sentido e anti sentido do gene alvo na construção hairpin-RNA.

Na nossa apresentação explicaremos a utilização das ferramentas da bioinformática no silenciamento do gene NPP1 por iRNA.