



Bioinformática

iPhy – Sistema de Apoio à filogenética para a análise de supermatrizes

Trabalho realizado por:

João Pedro Teixeira

|| Bioinformática

Objectivo do trabalho

- Análise de um artigo sobre um software de apoio à filogenética

Introdução

Técnicas

Resultados

Conclusões

Introdução

- Terminologias
- Motivação
- Implementação

Introdução

Terminologias

Taxon – Unidade taxonómica, que permite fazer a classificação de um dado sistema (pl. taxa);

Locus – Local fixo num cromossoma onde está localizado um determinado gene ou marcador genético (pl. loci);

Conteúdo AT – Percentagem de bases nitrogenadas de uma molécula DNA que são guanina(G) ou citosina(C);

Nematoides – Organismos de corpo aproximadamente cilíndrico que vivem em qualquer ambiente/ecossistema;

Filogenia – Hipóteses de relações evolutivas de um grupo de organismos, isto é, relações ancestrais entre espécies conhecidas;

Introdução

Motivação

- Estudo da filogenética envolvendo um grande número de taxa e loci, resolvem incertezas associadas a classificações;
- Usando supermatrizes na análise dos dados filogenéticos é possível oferecer uma maior resolução nos níveis de classificação;
- O agrupamento de muita informação de multigenes representa um investimento a nível de tempo e esforço;
- Os métodos e modelos filogenéticos ajudam a melhorar o conjunto de dados podendo ser re-analisados, usando novos modelos;
- iPhy, software user-friendly e cliente-servidor para o agrupamento e partilha de dados (alinhamentos) e produtos (árvores);

Introdução

Implementação

- O software iPhy foi desenvolvido em Java, usando várias bibliotecas e plugins fundamentais:
 1. BioJava => Execução de parsing de ficheiros;
 2. Gpars => Sincronização de sequências;
 3. ACEGI Grails Plugin => Gestão de utilizadores;
- Dividido em 2 características: Agregação de dados e Análise de subconjunto;

Técnicas

- Ferramentas
- Resultados de uma amostra

Técnicas

Ferramentas

✓ Importação de Sequências:

- Importação de Sequências de DNA do GenBank e FASTA;
- Todas as sequências são rastreadas;

✓ Subconjunto para análise:

- O iPhy seleciona a taxa baseada em alguns critérios (Níveis taxionómicos e Lista de loci);
- 3 critérios para a construção filogenética:
 - 1) N.º de Caracteres
 - 2) Composição da base de dados
 - 3) Taxa evolutiva

Técnicas

Ferramentas (cont.)

✓ Construção de sequências consensuais

- Geradas sempre que sequências múltiplas estejam disponíveis
- Derivação de sequências consensuais com parametros padrão
- Agregação de todas as sequências disponíveis num conjunto

✓ Anotações em árvores

- Por omissão, o iPhy nomeia os ficheiros de supermatrizes
- O iPhy oferece decorações de árvores para cada locus
- As árvores de arquivos podem ser carregadas com o iTol.

Resultados

Resultados

Agrupamento de conjunto de dados

- Extracções de sequências de neumatóides relevantes.
- Sequências usadas para construir BD no BLAST contendo proteínas.
- Diferentes loci envolvendo diferentes classificações tiveram score de identificação de genes para cada loci individual.
- No total, 24528 sequências geraram 81 loci.
- Usando construção consensual do iPhy resultou num conjunto de 8326 sequências consensuais

Resultados

Seleção e análise de dados

- Usado iPhy para construir subconjuntos de proteínas
- Cortes gerados segundo os 3 critérios
- Calculado o σ cada gene em cada subconjunto

Resultados (usando critérios de seleção):

- ✓ # de caracteres disponíveis \longrightarrow Desvio padrão + alto
- ✓ Conteúdo AT e taxa evolutiva baixa \longrightarrow Valores esperados

Enquanto estas análises são inconclusivas, mostram que novos loci e novas classificações podem ser adicionadas à análise de um grande conteúdo filogenético e múltiplos cortes podem ser facilmente exportados, analisados e visualizados.

Conclusões

- Interface de fácil utilização por parte de utilizadores;
- Agregação de grande quantidade de informação filogenética;
- Importação e exportação de dados para programas de apoio;
- Partilha de resultados através de URL;
- Aplicação cliente-servidor;
- Permite a dados incompletos o preenchimento através dos dados gerais;
- Não resolve o problema de identificação de sequências ortológicas;
- Problemas de identificação de sinónimos a grande quantidade de dados;
- Os ficheiros de formato GenBank não se encontram sempre correctos;

Obrigado pela
atenção!

