BioPerl

Ketan Mane

IV-Lab @ SLIS, IU

BioPerl

- Perl and now BioPerl -- Why ???
- Availability
- Advantages for Bioinformatics

BioPerl

- Perl – Scripting language

  – Powerful text parser
  – Includes set of regular expression & matching tools
  – Good string manipulation tool
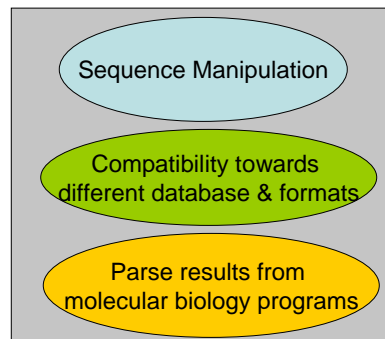 – Portable to web using CGI scripts

BioPerl

- BioPerl = Bio + Perl

  – Mainly object-oriented approach
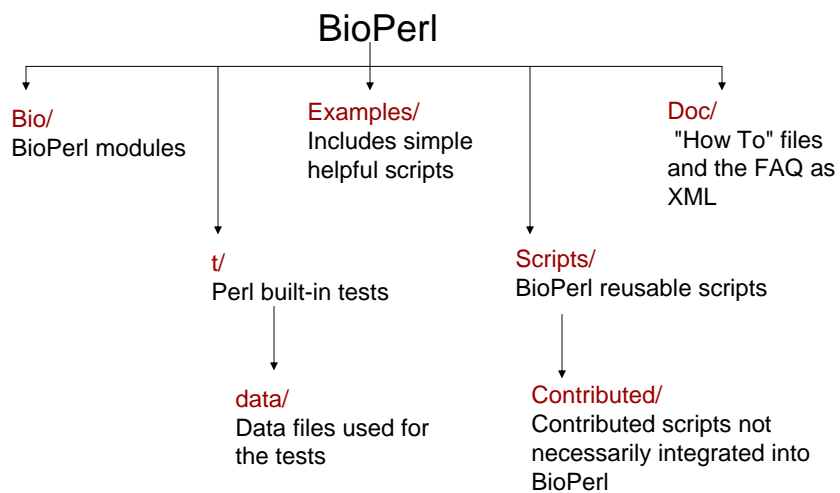 – Bioinformatics specific perl-modules available

  Sequence Manipulation

  Compatibility towards different database & formats

  Parse results from molecular biology programs

BioPerl

- Main url source:
  - http://bioperl.org/

  - Includes information on:
    - Projects.
    - Tutorials.
    - …. Tons of other related information

  - Bioperl releases are also mirrored by the Comprehensive Perl Archive Network (CPAN). **...**

BioPerl

- Directory Structure

BioPerl

Bio/
BioPerl modules

Examples/
Includes simple helpful scripts

Doc/
"How To" files and the FAQ as XML

t/
Perl built-in tests

Scripts/
BioPerl reusable scripts

data/
Data files used for the tests

Contributed/
Contributed scripts not necessarily integrated into BioPerl

BioPerl

- Sequence Manipulation

  - Module ⇔ Bio::SeqIO
  - Input = Sequence, File-format (eg: FASTA)

  - Sample program for file format conversion:

    ```
    use Bio::SeqIO;
    $in = Bio::SeqIO->new(-file => "inputfilename", -format => 'Fasta');
    $out = Bio::SeqIO->new(-file => ">outputfilename", -format => 'EMBL');
    while ( my $seq = $in->next_seq() ) {
            $out->write_seq($seq);

      }
    ```

BioPerl

- Sequence Manipulation:

- To acquire basic sequence statistics –
  - molecular weights
  - residue
  - codon frequencies

  Modules
  - SeqStats
  - SeqWord

BioPerl

- Database Access:
  - Supports sequence data retrieval from different databases: Genbank, RefSeq, Swissprot, EMBL etc..

  - Sample program
    ```
    $gb = new Bio::DB::GenBank();
    $seq1 = $gb->get_Seq_by_id('MUSIGHBA1');
    $seq2 = $gb->get_Seq_by_acc('AF303112');
    ```

BioPerl

- Sequence Similarity Operations:

  - Availability of Sequence Alignment Programs
    - Local alignment program: Smith Waterman
    - Module : Bio::Tools::pSW

    - Sample Code:
      ```
      $factory = new Bio::Tools::pSW( '-matrix' =>
                              'BLOSUM62', '-gap' => 12, '-ext' => 2);
      ```

BioPerl

- Parse information from mol. biology programs
  - Eg: Blast

  - Remote execution of BLAST programs with RemoteBlast.pm bioperl module
  - From blast report, the parse able information includes hit count, best local sequence and their scores and other …
  - SearchIO.pm – robust module, so preferred over Search.pm

BioPerl

Thanks !!!