From Pfam Alignments to Widgets

Presentation by,
Prasad Gunasekaran
Introduction to Pfam & alignments
Pfam

- Pfam - collection of protein domain families.

- Two classification of Pfam families - Pfam-A & Pfam-B.

- Pfam-A's are of high quality and are manually curated families.

- Each family represented by Hidden Markov Model (HMM) and multiple sequence alignments.

- Pfam-B's are of low quality and are generated automatically from the non-redundant clusters of the latest ADDA release.
Pfam family building procedure

- **Seed**
- **HMM model**
  - **hmmsearch**
- **Output**
  - **hmmalign**
  - **Align**
  - **Desc**
  - **Iteration**

Flowchart showing the process of Pfam family building, iterating from a seed to an output through a HMM model, with steps involving hmmsearch and hmmalign.
Pfam Alignments

- Two types of alignments: seed and full alignments.

- Jalview, Pfam viewer, HTML.

- Issues for large alignments
  - Jalview
    - Applet & Web start: Loading entire alignment causes memory issues.
  - HTML
    - Loads of markups & rendering time is high.
  - Pfam viewer
    - No issues
    - Drawbacks
      - Not user friendly.
      - Paged alignments.
Basic requirement for an alignment viewer

- User friendly.
- No memory issues.
- Page loading & rendering should be quick.
- Ability to handle massive alignments.
- Dynamic loading of alignments (LiveGrid).
LiveGrid

- JavaScript library provides a behavior for connecting an HTML element to a live data source via Ajax.

- Creates a scrollbar that becomes the live navigator for making Ajax data requests

- Connects it live to Ajax data responses.

- Automatically populates the response data.

- Implements data buffering and event registration strategies to improve performance.
Alignment widget

- Alignments are retrieved from selected das source using alignments command.

- LiveGrid library is used to populate the retrieved alignments.

- Consensus string is used to markup the alignments.

- Increase in time due to the parsing of the entire alignment response, if the optional xml tag `<alignment max=""/>` is omitted.

- Sources providing MSA could provide total number of rows in response.
Primitive *hybrid* DAS client
DAS client (widgets)

Data Integration

Alignment servers

Structure servers

Alignment widget

Structure widget

Feature widget

Interaction widget

Hybrid DAS client

Feature servers

Interaction servers
Demo continued...
Feature widget

- Features are retrieved from das sources for uniprot accession.
- Pfam domain graphics code is used to draw positional features on that sequence.
- List of other protein features sources can be selected.
- Alignment widget and feature widget are tagged to show the respective position in the alignment.
- Sources should provide the start and end of the segment. (eg: DS_446).

1.6 Spec:
<element name="SEGMENT"><attribute name="id"></attribute>
<attribute name="start"></attribute><attribute name="stop"></attribute></element>

1.5 Spec: <element name="SEGMENT">
<attribute name="id"></attribute>
<optional>
<group>
<attribute name="start"></attribute>
<attribute name="stop"></attribute></group>
</optional>
Future work...
Acknowledgement

- Rob
- John Tate
- Rest of Pfam team.
Questions?