BioStag

Chris Mungall

Berkeley Drosophila Genome Project

cjm@fruitfly.org

http://stag.sourceforge.net/

What is BioStag?

- **Stag**: Structured Tags
  - Simplified subset of XML
  - Autoobjects
- **DBIx::DBStag**: SQL Query **Templates** and Relational **Nesting**
  - accessable from command line, perl API or HTTP
- **BioStag**: Useful ready-made ad-hoc configurable queries for bioinformatics databases
  - middleware
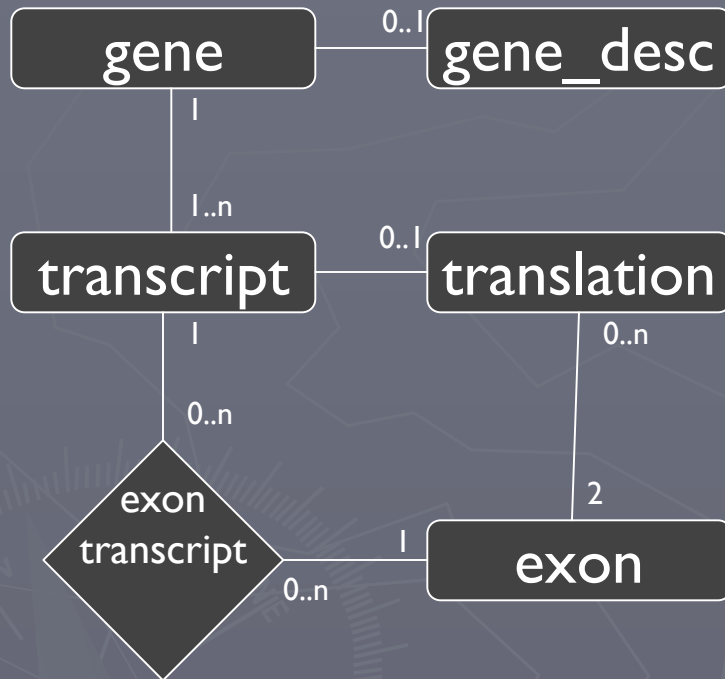 - end-user complex data mining queries

SQL Template

```
SELECT *
FROM gene
  INNER JOIN gene_desc USING (gene_id)
  INNER JOIN transcript USING (gene_id)
WHERE
  [gene_desc.description => &desc& ]
  [gene.seq_region_end  >= &rangemin& ]
  [gene.seq_region_start <= &rangemax& ]
USE NESTING (gene(gene_desc)(transcript))
//      --- metadata below this point ---
schema: enscore
schema-version: 20
desc: basic example gene range query
example_input: rangemin=1
```

BioStag Template Library

chado   [12]

GO   [17] ─────────

enscore [6] ─────────

biosql   [3]

gadfly   [26]

TOTAL [64]

go-term

go-geneproduct

go-transitive-correlations

go-common_ancestors

enscore-genemodel

go-descendent_term

enscore-intron

go-peptide_motify-by-term

enscore-spliced_3prime_utr

…

enscore-contigdna

…

Many databases
Many db specific APIs

Relational Nesting

gene —0..1— gene_desc

gene | 1

gene | 1..n

transcript —0..1— translation

transcript | 1

transcript | 0..n

translation | 0..n

exon transcript —1— exon | 2

exon transcript | 0..n

```
<gene>
  <gene_desc>….</>
  <transcript>
    <exon>…</>
    <exon>…</>
    <translation>..</>
  </transcript>
</gene>
```

SELECT * FROM gene * gene_desc * transcript *
              exon_transcript * exon * translation
WHERE […]
USE NESTING (gene(gene_desc)
                (transcript(exon)(translation)))

Everything else…

- Templates reside in biostag cvs
- DBI based
  - works on mysql and Pg databases
- Autogenerated web interface
- Generic XML->DB mapping
- TODO: Multi0DB queries, XSLTs, SOAP & Autogenerated WSDL
- ISSUES: speed, schema evolution

Acknowledgements

Brad Marshall
ShengQiang Shu
Chris Smith
Simon Prochnik
Jon Tupy
Josh Kaminker
Karen Eilbeck

George Hartzell
Nomi Harris
Mark Yandell
Sima Misra
Suzanna Lewis
Gerry Rubin