Linking Evolution to Genomics Using Phenotype Ontologies

Hilmar Lapp, NESCent

Scientific proposition

"Link evolutionary character changes to (developmentally) responsible genes using phenotypic mutants."

Genomics <-> Evolution

- **Genomic Biology**
  - Mutants of model organisms on a large scale
    - Mutation, KO strains, site-directed, or reverse genetics
    - Genotype is known a-priori (except for reverse genetics)
  - Phenotypes and traits to be measured and described
- **Evolutionary Biology**
  - Cataloging specimen on a large scale
  - Many taxa, some of which are extinct
  - Characters to be measured and described

Evolutionary Characters

- Phylogenetic inference uses character evolution
  - Character x Taxon character matrix
  - Molecular evolution is just a special case
- Characters are the manifestation of the genome
  - Which genes are responsible for the evolution of a character?
  - 'Which genes make us human?'

The Link: Phenotype

- Phenotype connects Genomics and Evolutionary Biology
  - Do the same gene(s) underlie an evolving character as underlies the same mutant phenotype?
- Language problem
  - Phenotype and character annotation use different languages, though.
  - In fact, both use proverbial free text.
  - Machines can't align free text.

Computable Annotation

- Ontologies standardize naming
  - Both words *and* semantics
- How to describe character change, or phenotype, in a standard way?
  - Standard for the character: Entity Ontology (E)
    - Anatomy, stage, biological process, cell type, etc
  - Standard for the change: Quality Ontology (Q)
    - Phenotype And Trait Ontology (PATO)
  - EQ annotation: E x Q
  - Some qualities are relational: EQ/E (E x Q_r x E)

E	A	V
head	size	small
eye	size	small
heart	structure	edematous
ventral mandibular arch	thickness	thick
swim bladder inflation	process_quality	arrested

Domain Knowledge

- Entity and quality ontologies have been driven by genomic MOs and mutants
- Need ontologies suitable for evolutionary character annotation
  - Cross-species Anatomy Ontology
  - Evolutionary character changes in PATO
  - Taxonomy Ontology

Connecting Ontologies

- Ontologies need to be extended to capture the evolutionary biology domain
  - We have tools to do this (e.g., OBO-Edit)
  - We know how to store this
    - GO data model, Chado cv module, BioSQL model
- However, how to constrain anatomy by taxon?
  - This is knowledge, so should be triple to be computable?
  - Relationship type? Tool to do this?
  - Is this a cross-product? Not really (or so I think)
  - To which ontology does this belong?

Anatomical Homology

- Morphology has a long history
  - Different names for homologous but diverged anatomical parts
  - Need a homology table that connects one anatomical term to another and identifies the evidence.
  - How do we map this into the Chado data model?
    - Aside from the lack of a tool to do this …

Storing EQ/E Annotation

- EQ/E annotation has so far been driven by genomic MODs
- Need to introduce evolutionary biology concepts & use-cases
    - Genotype (mostly) won't apply
    - No (or non-obvious) concept of 'specimen'
    - Need reference to annotator
    - How to tie EQ/E statements together as a 'description'
    - Conflicting descriptions conceivable

Pulling It All Together

- Canned queries:
  - Given a set of evolutionary character changes, what are genes associated with matching mutant phenotypes?
- What is the platform that will support this?
  - Recursive query language would be ideal
  - What is the break-down point for joins?
    - Normalized model and de-normalized warehouse?

Acknowledgments

- Paula Mabee, U. of South Dakota
- Todd Vision, NESCent
- Monte Westerfield, U. Oregon/ZFIN