



# WP5 LiceBase Introductory Course

## Part 6 Annotation and Curation Guide

Michael Dondrup  
Bergen 03.10.2014  
Oslo 6.10.2014



# Genome curation

- Genome annotation can be improved in different ways
- adding functional annotation to existing genes
- reporting errors
- adding novel genes and sequences

# Adding annotations

- Annotators can Edit features: -> Edit
- Add synonyms
- Add descriptions: Description field: choose 'Note' for most properties
- Adding annotation is based on sufficient evidence, as seen in part 3 and 4
- You should also run a new Blast vs. Swissprot and NR (NCBI Blast) before annotation, to see if something has changed
- In some cases use the term: "Putative ...."
- You should annotate genes you know well
- We don't have a naming scheme for genes!

# Reporting error

- Reporting an erratum can be done using the form:  
Example
- Most cases need a valid sequence id
- Use the 'list errata' link to look for reports
- avoid duplicate reports
- evidence files can be attached

# What if my sequence is missing?

- What you can do depends on sequence type (DNA, AA), and where it comes from
- Try hard to find your gene first
- AA: blastp (known genes), then tblastn to find novel genes
- DNA: blastn (landmark location), blastx (to find annotated genes), tblastx (novel genes, anything remotely similar)
- More sensitive tools are available (HMMER, PSI-Blast, etc.), but not in LiceBase (yet), ask us.
- Check if the gene can be found from its automatic annotation (e.g. via KEGG)

# What if my sequence is missing?

- If the sequence can be found, check if its location overlaps other features-> if not:
- Add a new feature in Tripal
- Paste 'Residues'
- Save feature and run Blast
- Blast provides a simple 'auto-locator' function to assign an assembly location to the sequence
- Be careful with auto-location: high identity (>99%) and query coverage required, check uniqueness
- Doesn't handle splicing
- Location can only be automatically assigned once
- Check evidence including RNA-seq coverage, ESTs and strand
- **Always file an annotation error, linking to the new sequence**

# More complex problems

- If the change is more complex we need to make alterations manually
- Examples: add/change exons, introns, changes in the assembly structure or base sequence (one gene split over contigs, etc.)
- We do not have a strategy for this scenario: who decides what is corrected and how, without breaking the assembly
- Anyway, reports will help

# Advanced stuff to be covered in the next course

- Custom reports can be created by any annotator in principle, but this is very complex (Drupal views)
- Uploading your own analysis results to Tripal (e.g. new prediction method)
- Dealing with GFF files
- Using the downloadable files
- Advanced GBrowse functions: e.g. making custom tracks, visualizing own data on the genome
- Workflow for efficient RNAi experiment design
- Primer design