

# Edited Highlights of the recent SO meeting

Karen Eilbeck

BDGP

September 2004

# Official discussion topics

- How to represent pseudogenes
- Polycistronic transcripts break the reasoner
- polyA tail is not part of the gene
- How to represent similarity/homology

# Unofficial discussion topic

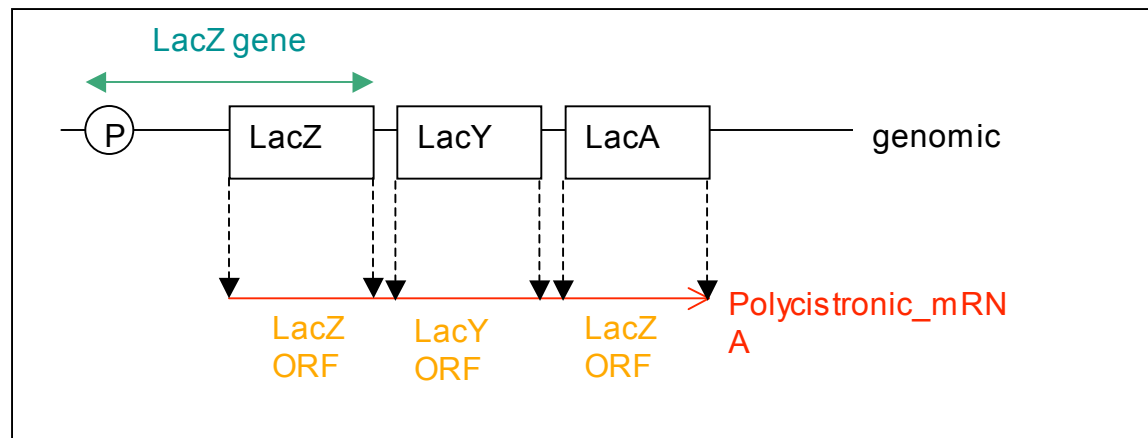
- What is a gene?
- Do we need to have the concept gene in SO?

# What part\_of means

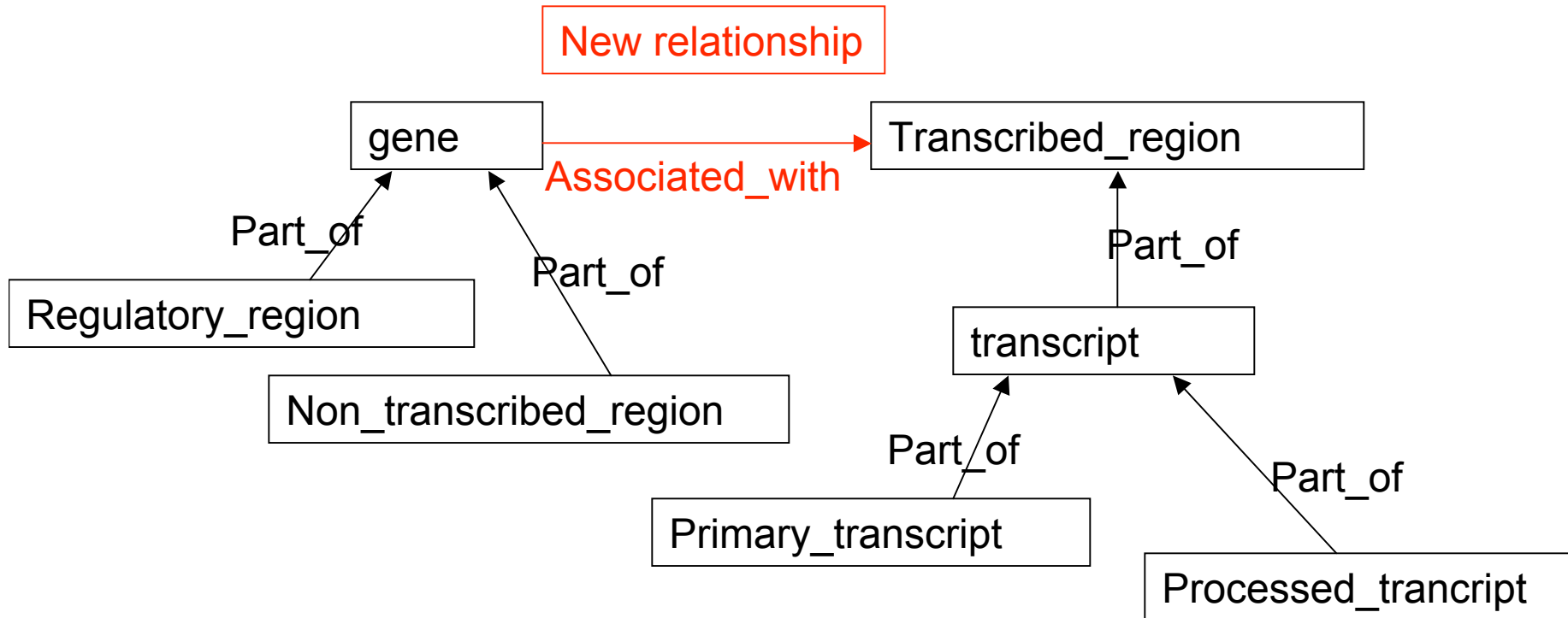
- The rules of being a part:
  - Nothing is a part of itself
  - If **A** is a proper part of **B** then the **B** is not a part of **A**
  - If **A** is a part of **B** and **B** is a part of **C** then **A** is a part of **C**
  - The relationship is **asymmetrical** and **transitive**
- The part\_of relation allows us to restrict the location of a term on a sequence.
  - *Exon* is part\_of *transcript*
  - So the coordinates of the *exon* must be within those of the *transcript*.
  - This is important if we want to be able to reason over parts in a productive way.

# Polycistronic transcripts

- A polycistronic transcript is not part\_of a gene
- A gene is not part\_of a polycistronic transcript



# Proposed solution to polycistronic transcript problem











# Another proposed solution

- Make an aspect in the ontology called Feature\_Collection.
- Gene inherits from feature collection.

# Transcript parts not part of genomic

- Cap and polyA tail are added to transcript but are not present in the original genomic sequence.
- SO the transitive relationship *cap is part of gene* is not true.
- This lead us to look at topological operators



A disjoint B	
A meets B	
A overlaps B	
A inside B	
A contains B	
A covers B	
A covered_by B	
A equals B	

# 'meets' solution

- polyA\_tail meets mRNA
- Cap meets mRNA
- This has also allowed us to align our definition of mRNA with GenBank/EMBL/DDBJ feature table
- "messenger RNA; includes 5'untranslated region (5'UTR), coding sequences (CDS, exon) and 3'untranslated region (3'UTR)"

# pseudogene

- SGD curators want to annotate pseudogenes in detail but were finding it difficult with the current structure

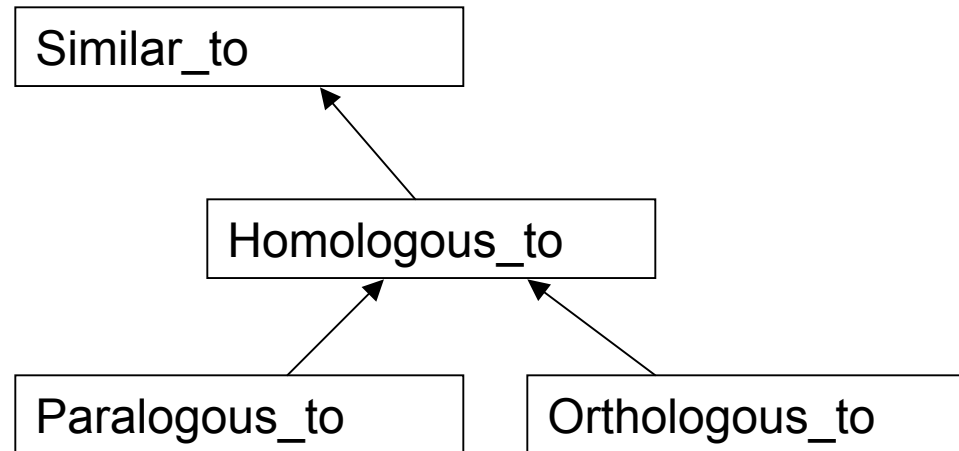
# Changes to the ontology:

- The definition of pseudogene must be changed. Remove the sentence “On occasion a pseudogene is functional as a consequence of being ‘captured’ by a non-paralogous gene, it is then known as a ‘captured\_pseudogene’.
- Remove the term captured\_pseudogene from being kind\_of pseudogene\_attribute.
- Add a term pseudogenic\_exon. It is\_a pseudogenic\_region. This is different to decayed\_exon as it will allow annotators to annotate pseudogenes to a deeper level.
- Create a new relationship called “non\_functional\_relative\_of” to allow us to annotate the relationship to the functional gene.

# Homology in SO

- Proposal to make concepts 'homologous\_region' , 'paralogous\_region' etc.
- Problem – homologous to what?
- Solution – create property relationships that link two regions together

# Similarity relationships



- These relationships are symmetric

# The changes to the ontology

- Some of these changes are pretty drastic.
- A test ontology has been created :
- `so-meeting.obo` in `SO cvs`