

Introduction

In mid September, following the Genome Informatics meeting in Hinxton, a SO workshop was held focusing on plasmids, phages, transposons and other mobile elements. This meeting primarily addressed several needs of the prokaryote community. As SO has developed, it has had a 'Eukaryotic slant' which is not surprising considering that the main contributors have been the eukaryotic model organism databases such as yeast, fly, worm, and mouse. This bias needed to be addressed to allow the prokaryotic community to feel comfortable using SO terms.

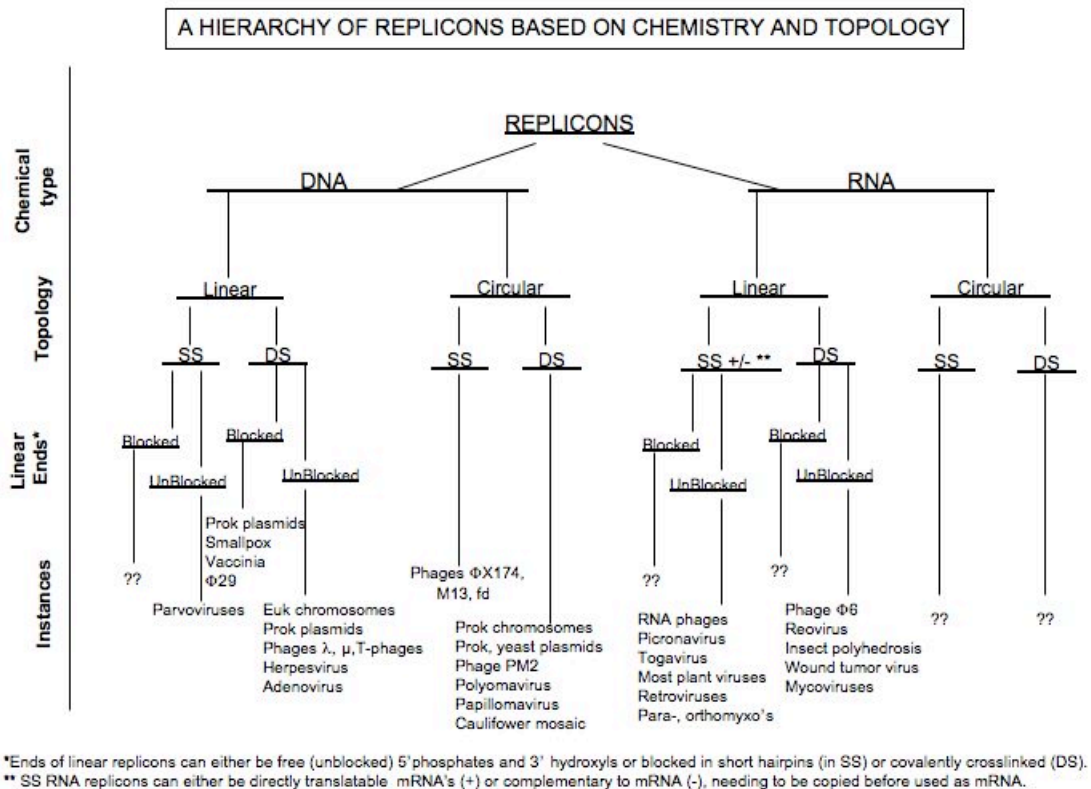
Several key people to spend two days working through some of the problems. These people were Anne Summers and Jessica Kissinger from the university of Georgia, and Ariane Toussaint from ULB in Brussels. Nick Thompson and Christiane Hertz-Fowler dropped by for relevant sessions. The SO people in attendance were Michael Ashburner, Suzi Lewis and Karen Eilbeck.

The issues

The replicon issue.

The idea of chromosome in SO alienate the prokaryotic and viral communities. We need to think in terms of replicons.

Anne presented a very convincing argument for extending SO to include all kinds of replicons.



The definition of chromosome was broadened to cover all kinds of replicons:

Structural unit composed of a nucleic acid molecule, which controls its own replication through the interaction of specific proteins at one or more origins of replication.

Subtypes of **chromosome** were created to describe chromosomes from the different walks of life. The criteria for making these divisions are polymer (DNA/RNA), strand (double/single) and topology (circular/linear). For example, `linear_double_stranded_DNA_chromosome` is a kind of chromosome.

These new chromosome terms have intersection definitions as they are cross-products with the qualities that differentiate them.

Recombination terms.

The recombination terms in SO were not complete and needed more organization. Subtypes were specified for `site_specific_recombination_target_region`.

- ☐ ← **i** `integration_excision_site`
 - ← **i** `attB_site`
 - ← **i** `attC_site`
 - ← **i** `attI_site`
 - ← **i** `attL_site`
 - ← **i** `attP_site`
 - ← **i** `attR_site`
- ☐ ← **i** `inversion_site`
 - ← **i** `FRT_site`
- ☐ ← **i** `resolution_site`
 - ← **i** `dif_site`
 - ← **i** `loxP_site`

Origins

The origin of transfer was expanded and the terms presented in a more uniform manner.

- ☐ ← **i** `origin_of_replication`
 - ← **i** `amplification_origin`
 - ← **i** `ARS`
 - ← **i** `D_loop`
 - ← **i** `oriC`
 - ← **i** `oriT`
 - ← **i** `oriV`

Regulatory regions

Enhancers, bacterial terminators and attenuators can be part of an mRNA

Transposable_element

Transposable element was moved from being a kind of repeat to being a kind of region.

Integrated_virus

This term was renamed proviral_region

Kinds of transcript

The definitions were clarified to allow prokaryotic annotations to use the processed_transcript terms. It was agreed that the terms names were unfortunate, but we did not change them as it would cause more problems for existing annotations.

Processed_transcript:

A transcript which has undergone the necessary modifications for its function. In eukaryotes this includes, for example, processing of introns, cleavage, base modification, and modifications to the 5' and/or the 3' ends, other than addition of bases. In bacteria functional mRNAs are usually not modified.

Primary_transcript:

A transcript that in its initial state requires modification to be functional.

Focus on trypanosome genomes

As an aside to the main proceedings, Jessica presented the problems encountered when trying to annotate a trypanosome genome. This included complicated kinetoplast genomes with mini and maxi circles and extensive mRNA editing.

Her slides are available at http://www.sequenceontology.org/meetings/rna_editing.pdf

The changes made to the ontology include extending the kinds of kinetoplast chromosomes and genes to include minicircle and maxicircles. Also anchor_binding_site, was added as a part of a transcript, (the terms editing_block, editing_domain, pre_editing_region and unedited_region already existed.) The subtypes of edit_operation define the possible edits. guide_RNA has a part anchor_region and a part template_region with a synonym information_region.

Remaining issues:

Plasmid is still a kind of *reagent* in the ontology. There is a conflict here between those who use plasmids as a means for performing a task, and those who are interested in the biology of plasmids.

Genomic_islands We did not have time to add anything more to the part of the ontology dealing with genomic islands. The terms needs to be better defined in the ontology.

Obsoleted terms

SO:0000660 dna_invertase_target_sequence

SO:0000187 repeat_family