

# Logical and Probabilistic Reasoning for Genomic Rearrangement Detection



Keith Flanagan, Matthew Pocock, Robert Stevens<sup>1</sup>, Phillip Lord, Pete Lee, Anil Wipat  
School of Computing Science, University of Newcastle upon Tyne, United Kingdom

<sup>1</sup>Department of Computing Science, University of Manchester, United Kingdom  
keith.flanagan@ncl.ac.uk, anil.wipat@ncl.ac.uk



## Introduction

An organism's genome can be represented as a large string of characters. Particular functional regions of a genome encode the genes which contribute to an organism's characteristics. All genomes change over time, however bacterial genomes in particular are in a continual state of flux. Comparative genomics is the study of the processes and *edits* involved that turn one genome into another, with the aim of discovering functional and evolutionary information. The work presented here is concerned primarily with the automated pairwise comparisons of bacterial genomes. We have captured the domain specific knowledge of genome rearrangements in an OWL ontology. Here, we combine logical and Bayesian inference to classify genomic regions according to this ontology.

## Genome rearrangements

Pairs of genomes are compared. The comparison of these genomes results in a list of "edits" that would transform the **source** genome into the **target** genome. Bacterial genomes evolve over time using a multitude of possible rearrangement features. These would include:

**Inserted regions** – areas in the target genome that do not occur in the source genome

**Deleted regions** – areas in the source genome that no longer occur in the target genome

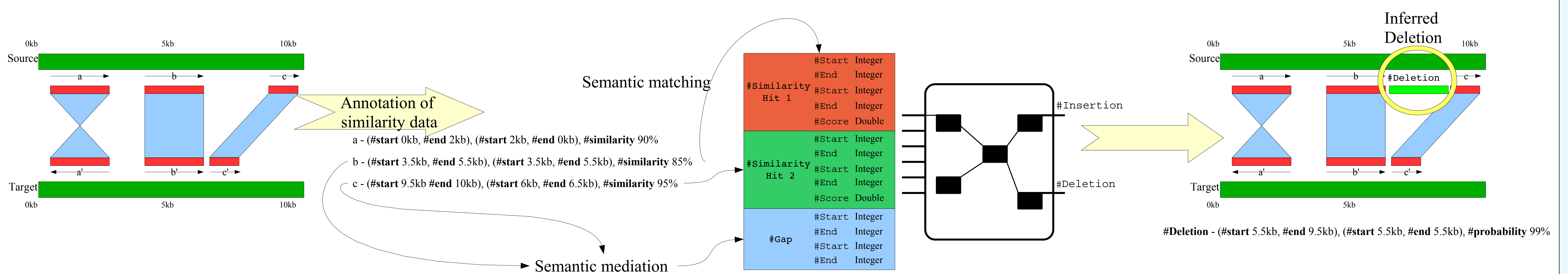
**Repeats** – areas in the source sequence that occur multiple times in the target sequence

We have built an ontology defining these rearrangements in OWL-DL.

## Why combine OWL with Bayesian networks?

- An expert biologist can identify genomic rearrangements from genomic comparison data, but this is time consuming and requires automation.
- OWL is sufficient to represent both genomic rearrangement and the data produced by pairwise comparison tools.
- However it is not sufficient to capture the probabilistic nature of the relationship between the two.
- Bayesian belief networks provide a convenient mechanism to describe the probabilistic rules needed to infer these relationships.
- We present here an architecture that combines these approaches.
- By selection of an appropriate Bayesian network and ontology we believe this architecture will be applicable to a wide range of problems.

## The process of combining OWL with Bayesian inference



- (1) Raw similarity data (e.g. Blast report)    (2) Annotated similarity data    (3) Two regions of similarity picked for analysis. Semantic matching and mediation are used to match input data to Bayesian network inputs    (4) Probabilistic inference using Bayesian network    (5) Classification results used to annotate pairwise comparison with OWL terms

## The roles of inference

- Logical reasoning used for:
  - Semantic **Matching** between comparison data and Bayesian network.
  - Semantic **Mediation** between comparison data and Bayesian network.
- Representation and secondary analysis of end results
- Probabilistic reasoning used for:
  - Determining belief in classification of real world data as a member of an OWL Class.

## Implications for the SW

- Combining Bayesian networks and OWL enables us to automate the annotation of genomes when neither technology alone would suffice.
- Probabilistically annotating real world data with OWL classes based its characteristics will be a common task for the Semantic Web.
- The methodology described here provides a generic solution for this type of problem.

## References

[Protege 2000] The Protege Project, <http://protege.stanford.edu>

[Netica] Netica Bayesian Network Software, <http://www.norsys.com>

[Altschul *et al.*, 1990] S. F. Altschul *et al.* Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403-410, October 1990.

[Mira *et al.*, 2001] Alex Mira *et al.* Deletional bias and the evolution of bacterial genomes. *Trends in Genetics*, 17(10):589-596, October 2001.

## Acknowledgements

This work is supported by: