# Towards a Semantic Foundation for Bioinformatics

Ross D. King

Department of Computer Science, Aberystwyth University, UK, rdk@aber.ac.uk

## 1 Abstract

With a two and half thousand year tradition logic is the best understood way of representing scientific knowledge. Only logic provides the semantic clarity necessary to ensure the comprehensibility, reproducibility, and free exchange of knowledge. The use of logic is also necessary to enable computers to play a full part in science [1]. The semantic web is transforming the dissemination of science by making for the first time making a large amount of scientific knowledge available expressed in logic.

Bioinformatics is one of the undoubted successes stories of the semantic web, with bioinformatic knowledge making up a large percentage of the scientific semantic web. Many of the problems that make semantic web reasoning difficult don't apply to bioinformatics: a ground truth of scientific knowledge exists, top level ontologies have been agreed (BFO), many other ontological standards exist, and the bioinformatic semantic web is large but not too large.

The use of bioinformatic software is essential to modern biology. However, there is a clear mismatch between the increasing use of the semantic web and logic, and the way bioinformatic systems utlilise and make inferences with this knowledge. This is because almost all computer based bioinformatic reasoning is done using *ad hoc* programs. From a formal point of view these programs are invariably making logical inferences: deductions, abductions, inductions, with perhaps a probabilistic element. However, what exactly these inferences exactly are is generally unclear.

The aim of my research is to make these inferences clear and to express them in logic, and make them executable across the semantic web.

For example, we argue that abductive inference is central to modern evolutionary based phylogenetics - clustering. This can be seen in evolutionary definition of a taxon (grouping of organisms): “that all members of a taxon are descendants of the nearest common ancestor (monophyly sensu stricto)” [2]. We express this in logic as:

$$\forall A \,.\, A \in taxon1 \Rightarrow (\exists Ancestor \,.\, \forall B \,.\, B \notin taxon1 \wedge ancestor(Ancestor, A) \wedge \neg ancestor(Ancestor, B)).$$

This clustering is based on the abductive inference of the existence of an ancestor organism not shared by any other taxon.

## References

1. King, R.D., Rowland, J., Oliver, S.G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L.N., Sparkes, A., Whelan, K.E., Clare, C. (2009) The Automation of Science. *Science*. **324**, 85-89.
2. Mayr, E. (1982). The Growth of Biological Thought: Diversity, Evolution, and Inheritance. Cambridge, Mass: Belknap Press.