



## Invited Talk Sarah Hunter:

### Classification of uncharacterised sequences; challenges and considerations

In recent years there has been a quantum increase in the affordability and thus availability of DNA sequencing machines. Existing large sequencing centres are now able to operate at a much higher capacity and laboratories which previously were unable to carry out any sequencing at all due to high costs, now have that capability.

New sequencing projects abound, such as those capturing variation between humans (1000 genomes project); those sequencing microbial communities found in and on our bodies (MetaHIT and HMP) and those surveying the communities in the environment (Tara Oceans). A major consequence of this is that large volumes of new sequences are flooding into biological databases and typically, very little is known about their function.

Resources such as InterPro use predictive models to characterise these sequences and to suggest possible functions, however, developing and searching the models with millions of sequences comes at a computational cost. Sarah will describe the challenges posed by this 'data tsunami' and how we hope to mitigate some of the issues we face.

**Date:** Thursday 28th October .

**Place:** Hotel Son Don Pablo. Sala Toledo.