Towards understanding the first genome sequence of a crenarchaeon by genome annotation using clusters of orthologous groups of proteins (COGs)

ABSTRACT
Special-purpose databases based on phylogenetic analysis and carefully curated with known and predicted protein functions offer a significant improvement in genome annotation. A differential genome display approach is used to investigate common and unique features of gene repertoires, and may reveal unexpected connections that could indicate functional similarities between distant organisms and lateral gene exchange.

INTRODUCTION
'Crenarchaeon' is in the plural, as it is a group of organisms that includes all living organisms, including plant and animal species.

Carnivores and herbivores

Crenarchaeon is a herbivore, but it is not a carnivore, as they are not 'carnivorous' in the sense of eating other animals. The term carnivorous is derived from the Latin words 'carnis' meaning 'to eat' and'vegeta' meaning 'animal'.

The term carnivorous is used to describe organisms that are generally herbivorous, but may occasionally include a carnivorous diet. For example, the carnivorous bacterium Escherichia coli Despite the importance of functional annotation, genome annotation is not an entirely new concept in the genomics field. It requires reliable assignment of gene function at the appropriate level of specificity to transfer information from experimentally characterized genes to their uncharacterized homologations through the use of sequenced database searches. The absence of detailed documentation and structure in general sequencing databases severely limits their usefulness as resources. Rather than relying on a 'Catch-22' approach, one can use specialized databases like KEGG and WIT to organize protein sequences into categories such as their structural, functional, or phylogenetic categories, which could be more effective for genome annotation. Phylogenetic classification of proteins encoded in fully sequenced genomes has been attempted through the creation of a database called Clusters of Orthologous Groups (COGs) using functional analysis and annotation. The COG system is organized based on combining evolutionary analyses and detailed sequence and structure analysis with the transfer of information between gene families and ancestral duplications. We describe the application of the COGs in systematic systematic analysis and evolutionary analysis of two newly sequenced archaeal genomes, namely those of Pyrococcus abyssi and crenarchaemia (assisted): these are used to compare the utility of this method of annotation on two types of genome-wise, where one is closely related to another genome already included in the system, and the other represents 'the group of previously unshared features of Archaeleota'; here we examine the working methods that are using the approach to analyse each large

CONCLUSION
Remarkable conclusions An annotation to a new genome is likely to be as

effective as the database(s) used to compile it, given that it relies on the phylogenetic principle of protein classification (clustering by (probable) orthology), and extensive effort has been invested in the functional characterization and classification of the COGs, resulting in many useful functional predictions being obtained through routine annotation protocols. Furthermore, the structure of this database enables us to identify unexpected genetic information and reveal evolutionary affinities between organisms.