

### 43. More than 140 genes have been acquired through horizontal gene transfer from Bacteria to rumen ciliates.

Guénola Ricarda, Neil R. McEwan<sup>b</sup>, Johannes H.P. Hacksteinc and Martijn A. Huynena.

*a* Center for Molecular and Biomolecular Informatics, Nijmegen Center for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

*b* Rowett Research Institute, Aberdeen, AB21 9SB, Scotland

*c* Department of Evolutionary Microbiology, Radboud University Nijmegen, Nijmegen, The Netherlands

The horizontal transfer of genes into Ciliates from Bacteria which live in close proximity to them within the rumen (the foregut of ruminants) was studied using ciliate Expressed Sequence Tags (ESTs). About 4000 ESTs were isolated from the two main types of rumen Ciliates: Entodiniomorphs (e.g. *Polyplastron multivesiculatum*, *Epidinium ecaudatum*, *Eudiplodinium maggii*, *Entodinium caudatum*, *Entodinium simplex*, *Diploplastron affine* and *Metadinium medium*) and Vestibuliferida, previously called Holotrichs (e.g. *Isotricha prostoma*, *Isotricha intestinalis* and *Dasytricha ruminantium*). A comparison of the sequences with the completely sequenced genomes, followed by large-scale construction of phylogenies, identified 148 ciliate genes that specifically cluster with genes from the Bacteria. Of these genes, 34 cluster with genes from the Firmicutes, a phylum of Bacteria that is well represented in the rumen. This phylogenetic clustering, coupled with the absence of close relatives of these genes in Tetrahymena, indicates that they have recently been acquired via Horizontal Gene Transfer (HGT). Among these HGT candidates, we found an over-representation (>70%) of genes involved in the anaerobic breakdown of complex carbohydrates, a rich food source in the rumen. We propose that the acquisition of these genes has facilitated the Ciliates' colonization of the rumen.

INTRODUCTION: Horizontal Gene Transfer (HGT), which consists of the transfer of genetic material between species, has not yet been extensively studied in the Eukaryotes. The transfer of 16 bacterial genes to Nematodes (Scholl et al. 2003) and of 96 genes to *Entamoeba histolytica* (Loftus et al. 2005) are the only example where HGT from Bacteria to Eukaryotes was investigated in a large scale.

As part of the European projects ERCULE (European Rumen Ciliate Culture Collection) and CIMES (Ciliates as Monitors for Environmental Safety), we have obtained a large set of cDNAs (around 4000 sequences) from anaerobic rumen Ciliates. Rumen Ciliates are interesting organisms as they are unicellular eukaryotes, living in the vicinity of Bacteria, which they use as food, in the gut of the ruminants. A large-scale analysis of the ciliate sequences was performed in order to find evidence of HGT. The results show that more than 140 ciliates sequences seem to have been acquired from Bacteria through HGT.

METHODS: First, we clustered sequences on the evidence of their belonging to the same gene (identity > 97% over more than 100bp). We then performed a comparison of our ciliates sequences against 148 predicted proteomes from the complete genomes. We examined the species distribution of the Best Hit for each ciliate cluster. The sequences having their Best Hit against a bacterial sequence were further examined by phylogenetic analysis. For each of these sequences, we retrieved its most similar proteins, we aligned them with the ciliate sequence using MUSCLE and we used a Maximum Likelihood method as implemented in MrBayes to calculate the corresponding tree.

To decide if our ciliate sequence was a HGT candidate, we examined the species composition of the second smallest partition containing our sequence. As the trees are not rooted we then assumed the root to be in the bigger partition.

Within our set of predicted proteomes, *Plasmodium falciparum* which belongs to the apicomplexa, is

the phylogenetically closest organism to our ciliates. As no ciliate genome has been published, we decided to include in our trees the nucleotide data from the most extensively sequenced ciliate, namely *Tetrahymena thermophila*, where these data were available.

Contamination was checked a priori, by removing sequences with a strong hit against a known protein at the nucleotide level and, a posteriori, by performing a Principal Component Analysis (PCA) on the codon usage of the HGT candidates together with homologous bacterial sequences.

**RESULTS:** Of 362 clusters that had a Bacterial sequence as their Best Hit, 224 trees were constructed for ciliate sequences having enough homologous sequences to allow such construction. Among them, 63 also include a *Tetrahymena* translated amino acid sequence. The results show that 133 ciliate sequences were found within a bacterial partition, 2 within an archaeal partition, and 13 within Bacteria and Archaea. These 148 sequences were thus considered as HGT candidates. Apart from these, we found 68 ciliates sequences within a partition containing at least one eukaryotic sequence and ten within a partition containing only Eukaryotes. Among the 148 HGT candidates, metabolic functions are strongly over-represented, compared to the functions presented in the whole dataset. Among them the largest fraction, like cellulases and xylanases, is involved, in the degradation of carbohydrates and more particularly of plant cell wall.

The HGT candidates were examined for evidence of signs of aberrant codon usage reflecting a non-complete adaptation to the codon usage of the Ciliate host. The HGT candidates' codon usage is indistinguishable from that of the non-HGT candidates and not relatively similar to that of a Firmicute like *C. acetobutylicum*, indicating complete adaptation to the Ciliates' codon usage.

**CONCLUSION:** As we have shown here, rumen Ciliates appear to have acquired several genes from the Bacteria by Horizontal Gene Transfer. The transferred genes that we found are mainly involved in the degradation of complex carbohydrates, showing an adaptation to the carbohydrate-rich environment in which these Ciliates live.

The number of HGTs we found in the rumen ciliate sequences is considerably larger than the number of HGTs found in Nematodes (16 for the whole genome). The difference may be explained by the fact that in multicellular organisms, DNA insertion has to take place in the germ cells to lead to any stable incorporation. Whereas HGT seems to be relatively rare in multicellular organisms, even in parasitic ones, it seems to play an important role in the evolution of Ciliates.

**DISCUSSION:** Although we found a large number of HGT candidates, the introduction of foreign DNA into a eukaryotic genome does not appear to be trivial. Indeed, the over-representation of metabolic enzymes we found and more particularly of fibrolytic enzymes (cellulases/xylanases), shows that, in order to be acquired, the new gene should bring a selective advantage to the recipient organism.