



**Fig. S3. A genome-wide HGT workflow for identification of all HGT-derived coding and noncoding elements. Panels a and b** depicts the workflows for identifying coding and noncoding HGTs using the transcriptome and the genome, respectively. HGT analyses were mainly focused on the predicted protein sequences from a comprehensive transcriptome (**a**) but also included the search of genome assembly for noncoding elements and pseudogenes (**b**). The identification of HGT-derived protein coding genes relied on two approaches: 1) a phylogenomic approach with large-scale gene trees and 2) a BLAST-based approach with the union of results of two methods (HGTector (Zhu et al., 2014) and a modified approach we call HGTpropor). Candidate HGTs identified from tree and BLAST were then cross-validated. Identification of the noncoding part of the genome (including repeats and pseudogenes) started with BLASTN search of all genomic scaffolds, followed by candidate prediction using HGTpropor and final validation (details can be found in Materials & Methods). (**c**) compares the candidates of high-confidence HGT-derived protein coding genes by BLAST-based and phylogenomics approaches. Orange ones represent coding HGT OrthoFinder groups specific to BLAST-based approach, green shared by both approaches, and yellow specific to phylogenomics approach. Numbers in the parenthesis represent the number of events.

## References

Zhu, Q., Kosoy, M. & Dittmar, K. HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers. *BMC Genomics* 15, 717, doi:10.1186/1471-2164-15-717 (2014).