**Methods**

HGT among the 1312 seaweed-associated MAGs was detected with MetaCHIP v1.10.4 at phylum, class, order, family and genus levels. MetaCHIP uses best-hit analysis between defined phylogenetic/ taxonomic groups to identify candidate genes for HGT, which are subsequently further validated using phylogenetic analysis by the reconciliation of species and gene trees. This gives high confidence in the prediction and allows to determine the direction of gene transfer. HGTs detected at all taxonomic levels were then combined and dereplicated. Horizontally transferred COG functions were considered to be enriched if their relative proportions in the HGT dataset were above the 75% percentile of their relative abundance across the MAGs.

**Results**

MetaCHIP identified 2402 HGTs and 2183 of them were annotated to 809 unique COG functions. The relative proportion of COG functional categories (COG-FC) of these genes were analysed, and COG-FCs were considered to be enriched in the HGT if their proportion was above the 75% percentile of the relative abundance across all input MAGs.

The results revealed that genes with functions involved in energy production and conversion (C), as well as the transportation and metabolism of amino acid (E), carbohydrate (G), lipid (I) and secondary metabolites (Q) are more likely to be subjected to HGT (Figure).

The results revealed that most COG-FCs responsible for metabolism were enriched in HGTs (Figure) such as, C (energy production and conversion), E (amino acid transport and metabolism), G (carbohydrate transport and metabolism), I (lipid transport and metabolism) and Q (secondary metabolites biosynthesis, transport and catabolism). The enrichment of carbohydrate transport and metabolism (G) was mainly due to 271 HGT involving genes.

Chart, box and whisker chart

Description automatically generated

**Figure.** Relative proportion of COG functional categories and predicted HGTs from the seaweed-associated MAGs. The boxes in the plot represent the relative abundances of COG-FCs and are bound by the 25–75% quartile proportions with the thick line being the median value. Q1, Q3 and IQR refer to the 25%, 75% and interquartile range, respectively. The upper whisker refers to the largest observation less than or equal to upper Q3 + 1.5 × IQR, while the lower whisker refers to the smallest observation greater than or equal to Q1 − 1.5 × IQR. Symbols show the relative abundance of COG-FCs subject to HGT and are highlighted, if enriched (red triangle), depleted (blue triangle) or neither (black square). The letters on the x-axis indicate COG categories: C (energy production and conversion), D (cell cycle control, cell division and chromosome partitioning), E (amino acid transport and metabolism), F (nucleotide transport and metabolism), G (carbohydrate transport and metabolism), H (coenzyme transport and metabolism), I (lipid transport and metabolism), J (translation, ribosomal structure and biogenesis), K (transcription), L (replication, recombination and repair), M (cell wall/membrane/envelope biogenesis), N (cell motility), O (posttranslational modification, protein turnover and chaperones), P (inorganic ion transport and metabolism), Q (secondary metabolites biosynthesis, transport and catabolism), R (general function prediction only), S (function unknown), T (Signal transduction mechanisms), U (Intracellular trafficking, secretion, and vesicular transport), V (Defence mechanisms), W (Extracellular structures), X (Mobilome: prophages, transposons) and Z (Cytoskeleton).