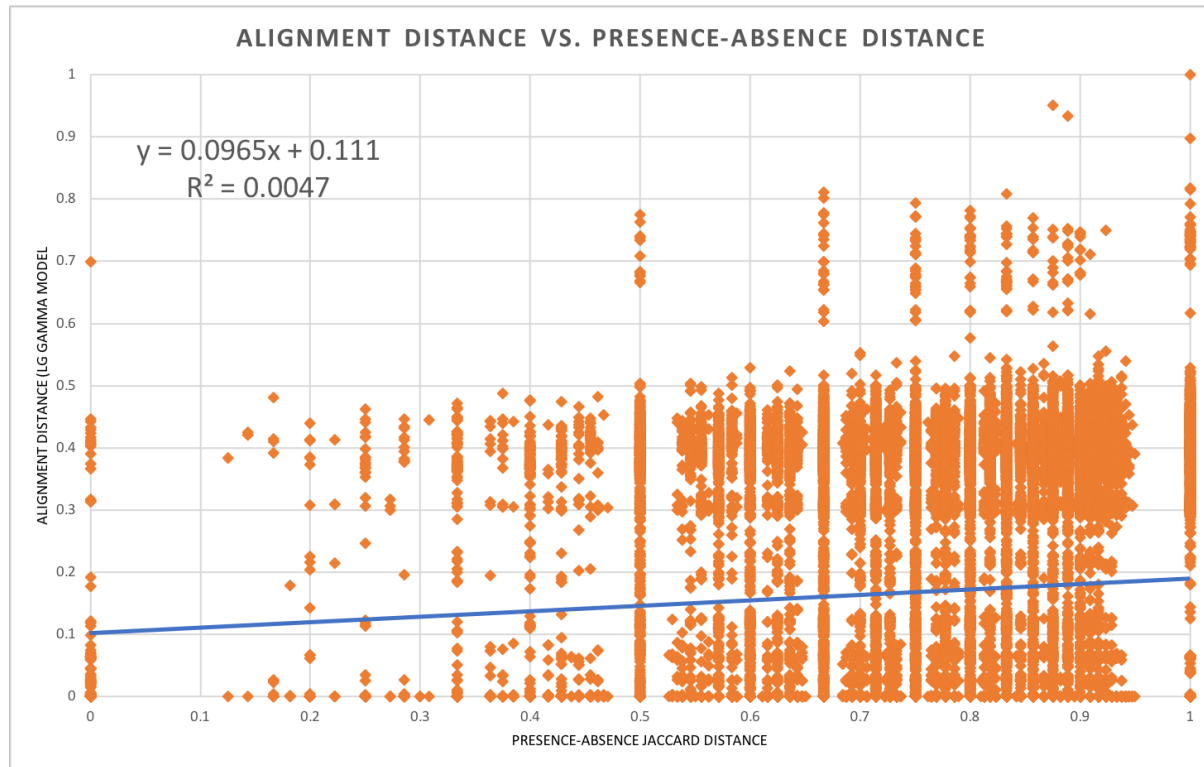
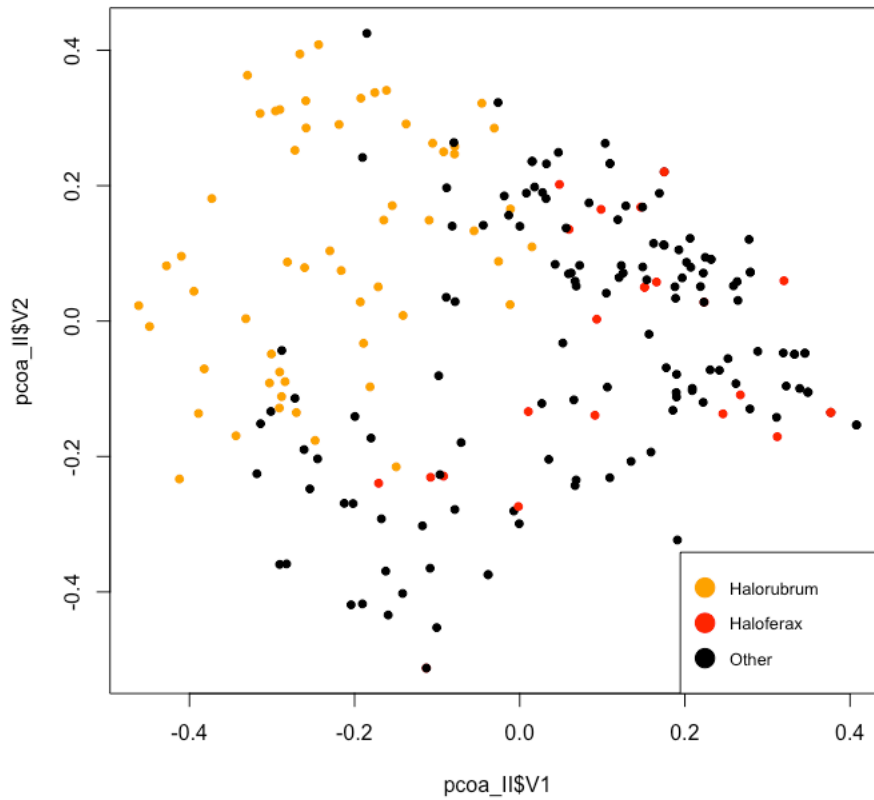


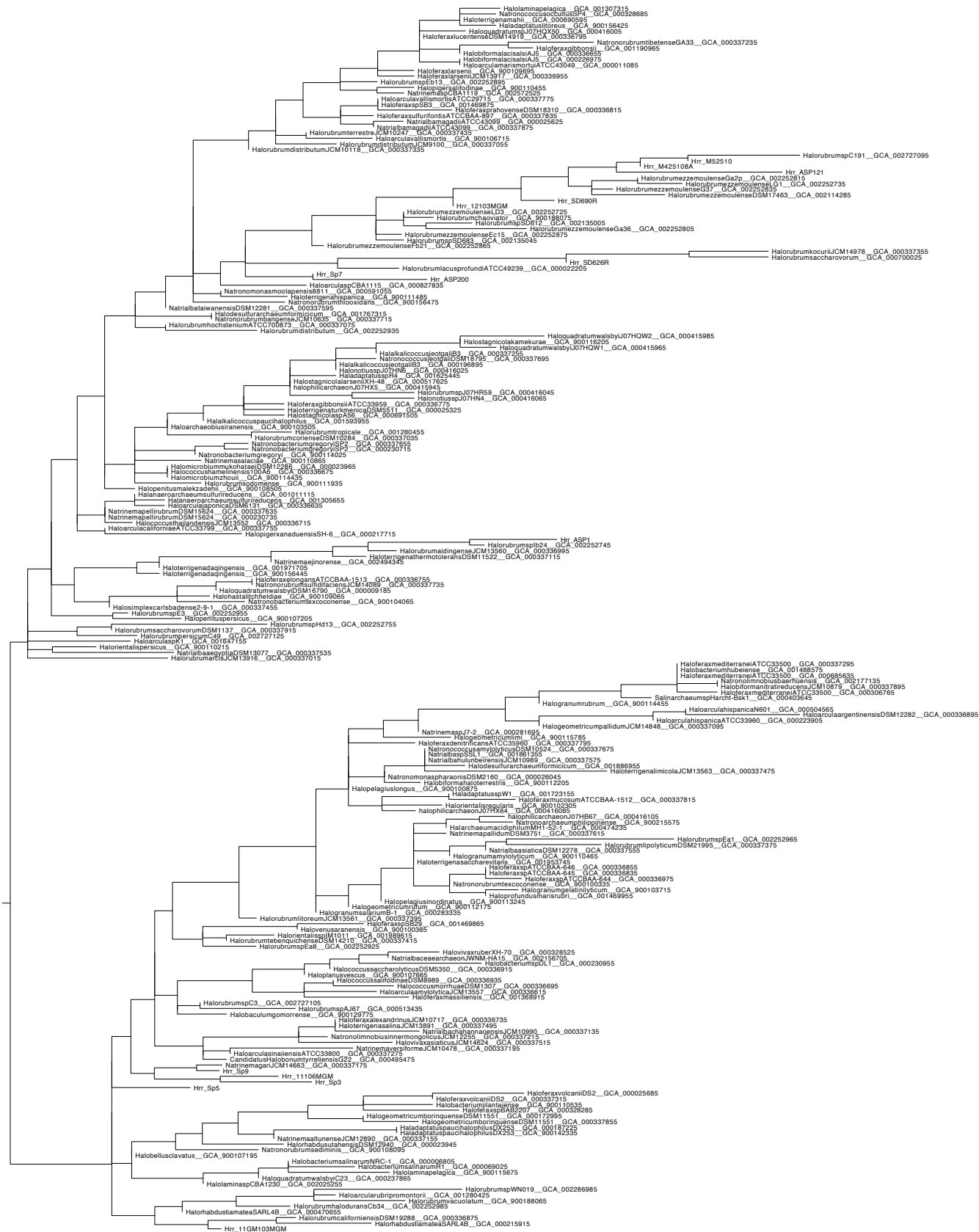
**Figure S1.** Workflow of RMS-candidate gene search strategy. Red circles represent input data and tools used to operate on data. Blue rectangles represent use and modification of the data during the process. Orange rectangles represent outputs at significant intermediate points in the process.



**Figure S2.** Plot of alignment distance as a function of presence-absence distance. Alignment distance was calculated from an LG+Gamma model and presence-absence using the Jaccard distance. If presence-absence pattern were a function of vertical descent, then the best fit line should broadly follow a line through the origin (0/0). The correlation between the two measures is negligible ( $r$ -squared = 0.0047).



**Figure S3.** PCoA plot of the distances between the RMS presence-absence profiles of the 217 analyzed Halobacterial genomes. The two best sampled groups, *Halorubrum* (orange) and *Haloferax* (red), are colored separately. The overlap of ranges between the three groups illustrates that few RM system genes are unique to a given group and a large overlap in inventories between groups.



0.09

**Figure S4.** Maximum-likelihood phylogeny of cHG presence-absence matrix. An F81 model with empirical base-frequency was employed. The root was placed arbitrarily at the midpoint on account of the topology not bearing meaningful resemblance to the reference topology (cf. Figure 2).

[illegible]

**Figure S5.** Bootstrap support values of the presence-absence phylogeny mapped onto the ribosomal protein reference tree. All nodes with values of 10% or greater are shown. Only few groupings of organisms belonging to the same species or genus are moderately supported by both, the presence-absence phylogeny and the ribosomal proteins. The lack of mutual support between ribosomal reference tree and the RM system derived phylogeny confirms the discord between the datasets (ribosomal proteins and RM system related proteins).