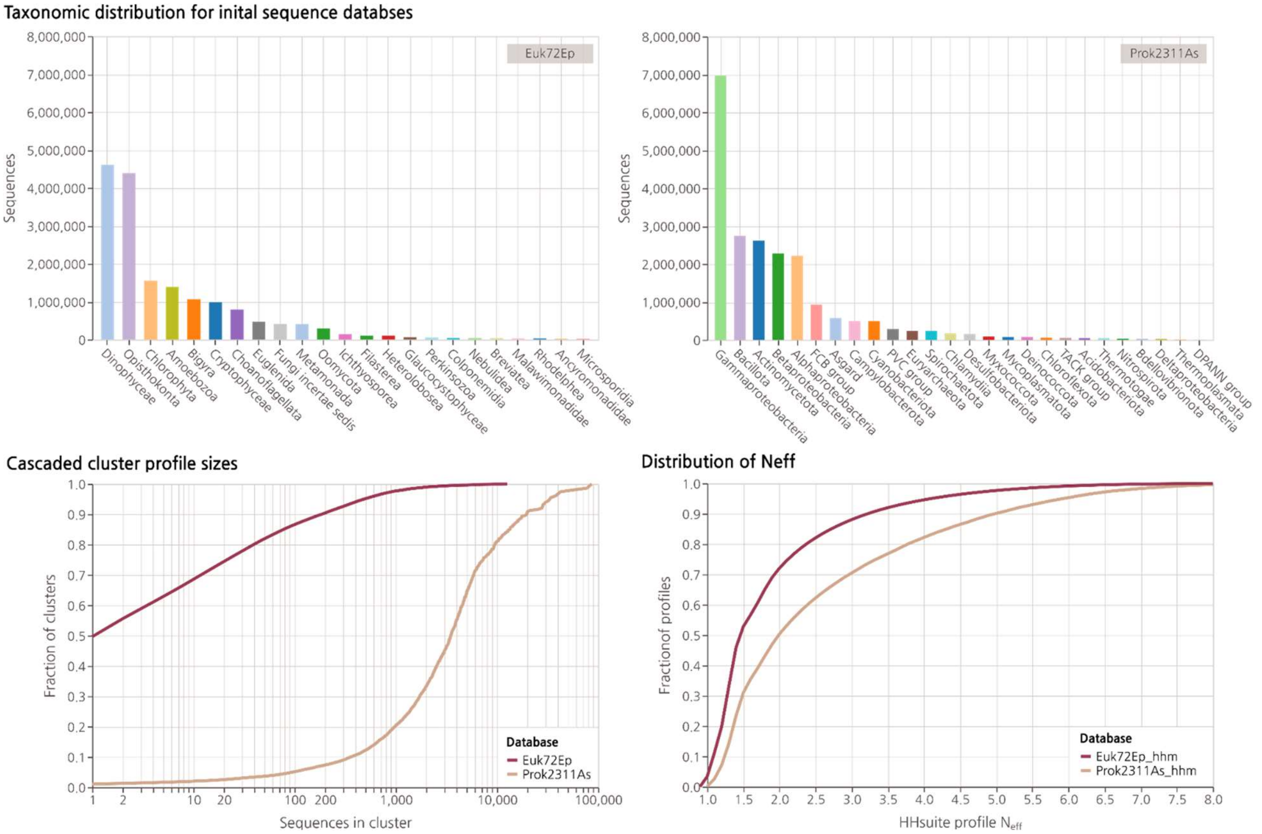
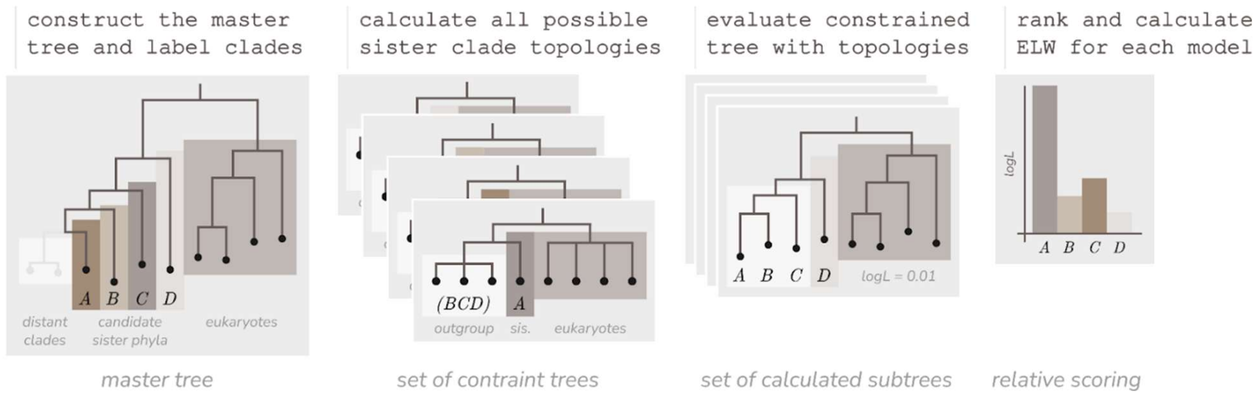
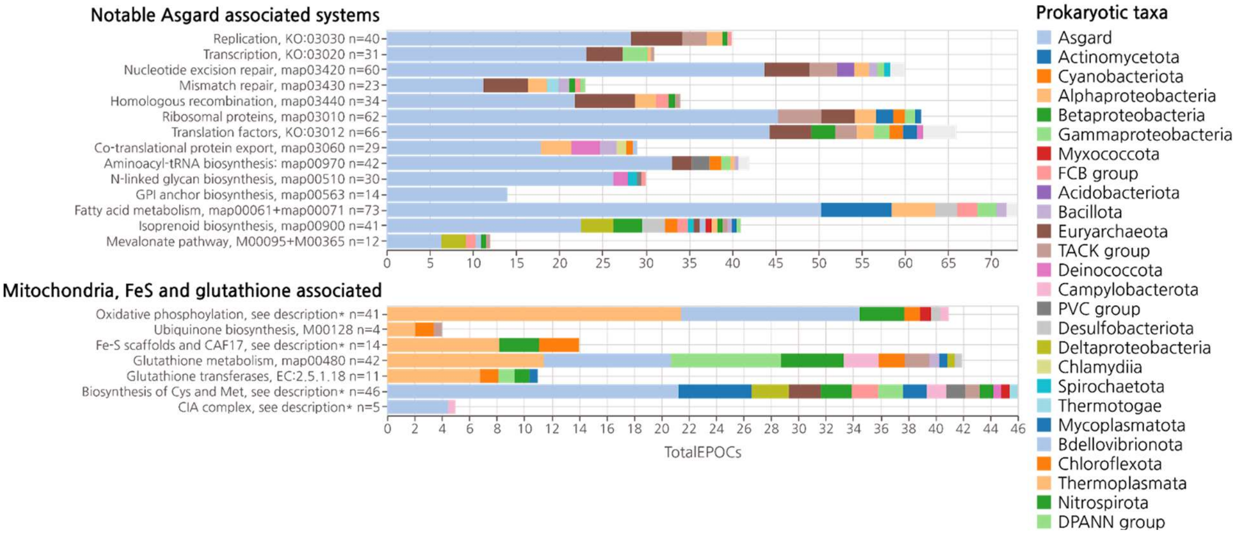
Extended Data



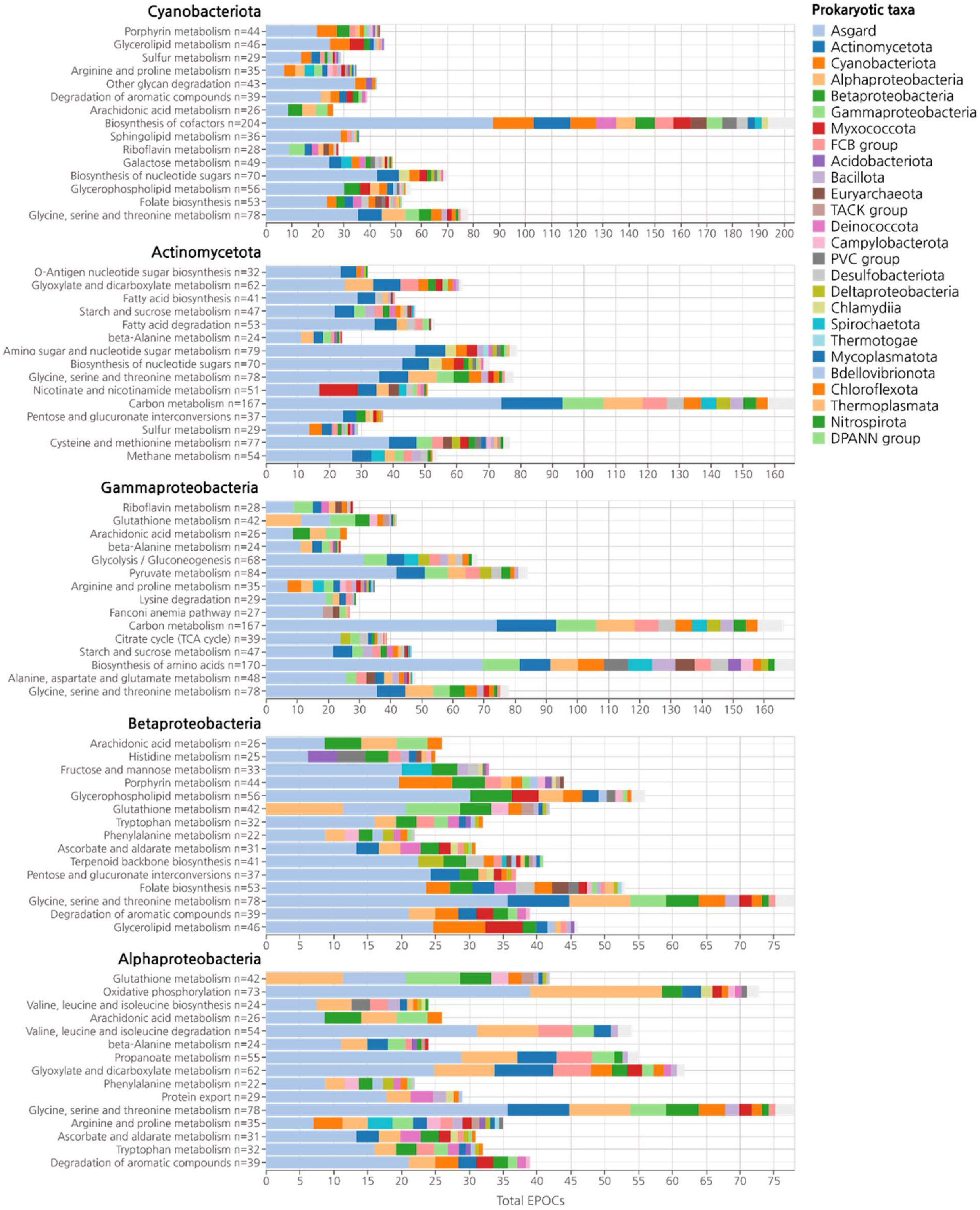
Extended Data Figure 1. Global database statistics. Distribution of taxonomic labels present within Euk72Ep and Prok2111As. Final cluster size distribution following cascaded mmseqs profile/sequence clustering. Distribution of N_{eff} values for resulting profile databases.



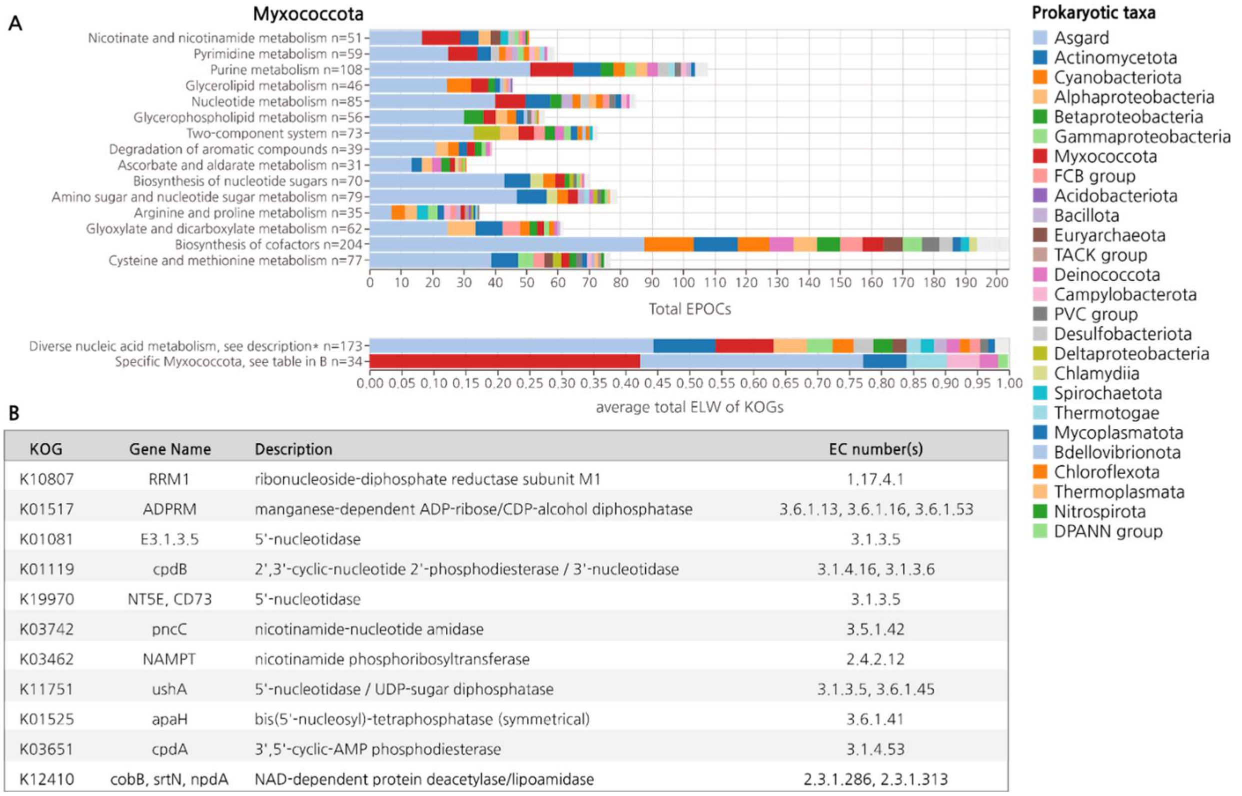
Extended Data Figure 2. Evolutionary hypothesis testing constraint trees and Expected Likelihood Weights. Overview of ELW calculation procedure. One master tree is constructed from all sequences within a EPOC and the closest n sister clades to the eukaryotic outgroup identified. A set of constraint trees are generated, each enforcing three clades, one eukaryotic, one sample sister and one outgroup with remaining prokaryotic sequences, is present. The set of subtrees are then constructed, each forced to conform to a single guide tree. LogLikelihood values for resulting trees are compared and evaluated using IQtree -z to produce confidence sets and Expected Likelihood Weight estimations.



Extended Data Figure 3. ELW breakdown of additional maps and pathways with prominent Asgard and Alphaproteobacterial contributions. Average ELW for all EPOCs within relevant pathways. Averages of KOG sets annotated by map number (ex map0000X), module number (ex. M0000X), or Enzyme commission number (EC:X). Oxidative phosphorylation (map00190) also contains entries for V-type ATPases which are excluded here. Fe-S scaffolds are not members of any map the manually curated set includes KOGs: K22063, K22072, K22068, K22073, K22070, K22071, K04043, K04044 and K04082. The CIA complex is likewise not grouped within BRITE and the manual set contains KOGs: K24730, K26403 and K15075. Finally, cysteine and methionine synthesis are spread across several modules and consist of the union of M00021, M00338, M00609, M00017, M00034 and M00035

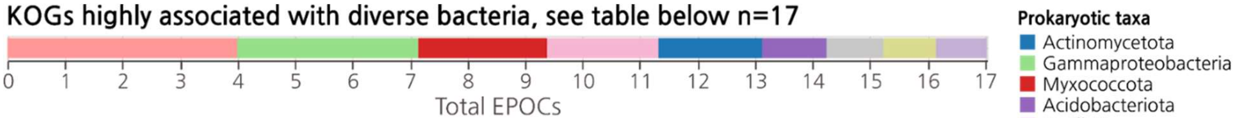


Extended Data Figure 4. Strongest per taxa association of diverse bacteria to eukaryotes. Top 5 non-alphaproteobacterial associated taxa with their corresponding to 15 highest ELW pathways.



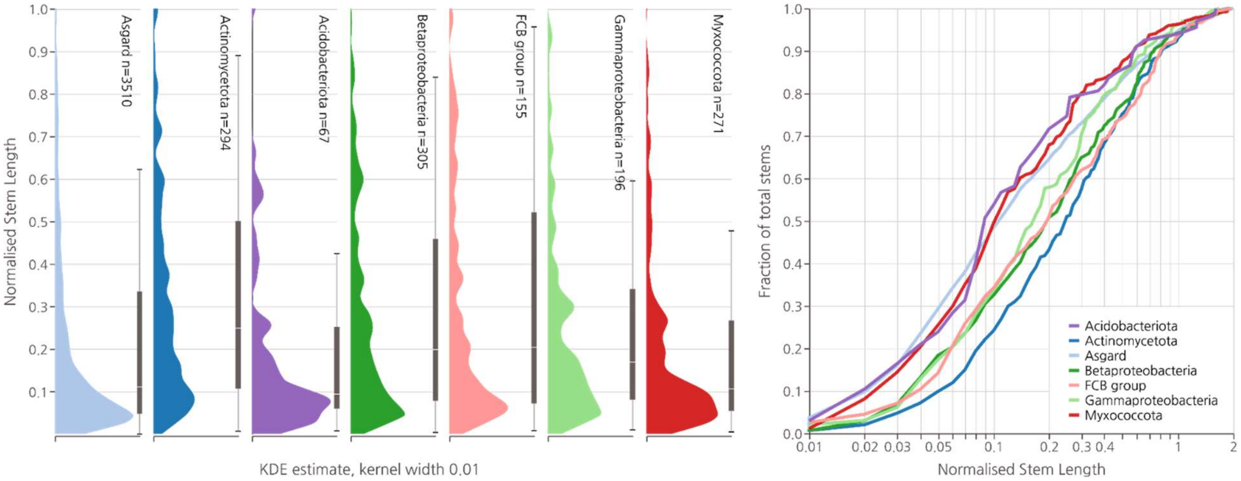
KOG	Gene Name	Description	EC number(s)
K10807	RRM1	ribonucleoside-diphosphate reductase subunit M1	1.17.4.1
K01517	ADPRM	manganese-dependent ADP-ribose/CDP-alcohol diphosphatase	3.6.1.13, 3.6.1.16, 3.6.1.53
K01081	E3.1.3.5	5'-nucleotidase	3.1.3.5
K01119	cpdB	2',3'-cyclic-nucleotide 2'-phosphodiesterase / 3'-nucleotidase	3.1.4.16, 3.1.3.6
K19970	NT5E, CD73	5'-nucleotidase	3.1.3.5
K03742	pncC	nicotinamide-nucleotide amidase	3.5.1.42
K03462	NAMPT	nicotinamide phosphoribosyltransferase	2.4.2.12
K11751	ushA	5'-nucleotidase / UDP-sugar diphosphatase	3.1.3.5, 3.6.1.45
K01525	apaH	bis(5'-nucleosyl)-tetraphosphatase (symmetrical)	3.6.1.41
K03651	cpdA	3',5'-cyclic-AMP phosphodiesterase	3.1.4.53
K12410	cobB, srtN, npdA	NAD-dependent protein deacetylase/lipoamidase	2.3.1.286, 2.3.1.313

Extended Data Figure 5. Sparse Myxococcotan association with eukaryotes. A) Top, ELW breakdown of highly enriched Myxococcotan pathways showing primary association with nucleotide synthesis. Bottom, curated set of Myxococcotan enriched EPOCs containing general nucleic acid metabolism and modification pathways corresponding to KEGG map00760, map01232, map00240 and map00230 with notable presence of Myxococcota. Further selection of only those KOGs within the above pathways which include Myxococcota. B) List of KOGs with Myxococcotan presence within the curated set of nucleotide synthesis specifically highlighting phosphatases and phosphotransferases.



KOG	Gene Name	Description	EC number(s)
K02564	nagB, GNPDA	glucosamine-6-phosphate deaminase	EC:3.5.99.6
K12862	PLRG1, PRL1, PRP46	pleiotropic regulator 1	
K07198	PRKAA, AMPK	5'-AMP-activated protein kinase, catalytic alpha subunit	EC:2.7.11.31
K10570	ERCC8, CKN1, CSA	DNA excision repair protein ERCC-8	
K06972	PITRM1, PreP, CYM1	presequence protease	EC:3.4.24.-
K05967	K05967	uncharacterized protein	
K03781	katE, CAT, catB, srpA	catalase	EC:1.11.1.6
K04688	RPS6KB	ribosomal protein S6 kinase beta	EC:2.7.11.1
K01908	ACSS3, prpE	propionyl-CoA synthetase	EC:6.2.1.17
K13137	STRAP, UNRIP	serine-threonine kinase receptor-associated protein	
K06013	STE24	STE24 endopeptidase	EC:3.4.24.84
K14085	ALDH7A1	aldehyde dehydrogenase family 7 member A1	EC:1.2.1.31, 1.2.1.8, 1.2.1.3
K00597	MTRR	methionine synthase reductase	EC:1.16.1.8
K00023	phbB	acetoacetyl-CoA reductase	EC:1.1.1.36
K14430	PHO87_91	phosphate transporter	
K08956	AFG3	AFG3 family protein	EC:3.4.24.-
K00261	GLUD1_2, gdhA	glutamate dehydrogenase (NAD(P)+)	EC:1.4.1.3

Extended Data Figure 6. KOGs with strong bacterial associations. KOGs associated with EPOCs with stringent criteria. EPOCs must contain an eukaryotic outgroup with at least 15 unique clades encompassing LECA, have a prokaryotic sister clade with an ELW of > 0.7 which is not derived from Asgards, Euryarchaeota, the TACK group, Cyanobacteriota or Alphaproteobacteria and contain more than 20 sequences. The 32 EPOCs which meet these criteria are highlighted and their global ELW displayed.



Extended Data Figure 7: **Stem length distribution for diverse bacteria compared to Asgard archaea**. Distributions of normalized stem lengths of diverse bacteria compared to Asgards. Includes all eukaryotic clade stem lengths between 0.05 and 1.0.