pangenome and reconstructed phylogenetic trees of those genes as above.

Synteny Analysis and Horizontally Transferred Genes

Synteny blocks in archaeal fosmids were defined as arrays of one of more contiguous genes of same origin class (archaeal core, lineage-specific core, early HT-genes, late HT-genes, and others, including the remaining predicted genes without homologs in the database). Each gene or synteny block is flanked by blocks of one or two different origins. Because we consider five possible origin gene classes for Thaumarchaeota or GII/III-Euryarchaeota, there are 15 (5 + 4 + 3 + 2 + 1) different possibilities for any synteny block (or gene) to be bounded. Bounding-couple occurrence was compiled for each synteny block in Thaumarchaeota and GII/III-Euryarchaeota fosmids separately and the corresponding data matrix subjected to hierarchical clustering analysis (Eisen 1998) using the MeV package (Saeed et al. 2003).

Codon Usage and Codon Adaptation Index Analysis of HT-Genes

The codon adaptation index (CAI) for a total of 26,678 genes acquired through HGT by Thaumarchaeota and GII/III-Euryarchaeota was calculated as follows: 1,244 ribosomal protein genes (459 from GII-Euryarchaeota and 785 from Thaumarchaeota fosmids) were first selected as a reference pool of highly expressed genes, either together or in groups of similar origin, and their codon usage table calculated using the cusp program from the EMBOSS suite, version 6.5.7 (Rice et al. 2000). The CAI for all genes was then calculated with the three sets of ribosomal genes codon usage tables (Thaumarchaeota, Euryarchaeota, or Thaum + Euryarchaeota) serving as reference with the CAI program (EMBOSS). Codon usage values were then submitted to PCA analysis (Raychaudhuri et al. 2000) using MeV (Saeed et al. 2003).

Results

Metagenomic Fosmid Sequences and Functional Classification of Genes in Archaeal Pangenomes

We obtained complete sequences of 545 and 452 fosmid clones from metagenomic libraries of deep-Mediterranean plankton (Ionian and Adriatic Seas at, respectively, 3,000 and 1,000 m depth) clearly affiliated with, respectively, Thaumarchaeota and GII/III-Euryarchaeota. Phylogenetic ascription was initially based on the phylogeny of genes located at both insert ends (Brochier-Armanet et al. 2011) and, subsequently, confirmed or corrected based on the phylogeny of all the genes that they contained (see below). Only high-quality, full-fosmid sequences showing no indication of potential chimerism (e.g., frameshifts, truncated genes, or unmixed distribution of archaeal and bacterial genes in two fosmid regions) were retained for this study. Details about the genomic sequences generated are given in table 1. Because Thaumarchaeota and GII/III-Euryarchaeota seem to have a single copy of rRNA genes (this is the case in all sequenced genomes of Thaumarchaeota as well as the Thermoplasmatales and *A. boonei*, the closest relatives of GII/III-Euryarchaeota) (Moreira et al. 2004), the number of archaeal genomes sequenced could be estimated at, respectively, 14 and 9, based on the number of rRNA gene copies identified. These values were in good agreement with estimates obtained from a collection of 40 additional genes typically found in single copy in prokaryotic genomes (Creevey et al. 2011), 16.5 and 9.3, respectively (supplementary fig. S1, Supplementary Material online, and table 1). The identification of similar gene counts for all those single-copy genes additionally suggests that those archaeal genome equivalents had complete (or nearly so) coverage in our libraries in terms of gene content. However, the assembly of individual genomes was not possible due to the within-group archaeal diversity captured by the fosmids (see below; fig. 1). To try to bin fosmids within different phylogenetic groups, we analyzed tetranucleotide and pentanucleotide frequencies, which are often used for the assignment of genome fragments to distinct groups (Teeling, Meyerdierks, et al. 2004). A PCA of tetranucleotide frequencies showed a clear separation of Thaumarchaeota and Euryarchaeota fosmids (fig. 2). Thaumarchaeota fosmids formed a tight cluster, and different subgroups were not distinguishable. Euryarchaeota fosmids formed a much more dispersed cloud, with a small cluster of fosmids loosely segregating from the main cloud. However, contrary to our initial expectations, this smaller cluster does not correspond to GIII-Euryarchaeota, because fosmids containing 16S rRNA genes of GIII-Euryarchaeota fell in the two clouds (mostly in the bigger cloud). The eccentricity of those clones is so far unclear; they might contain genomic islands with biased GC content/codon usage or differentially

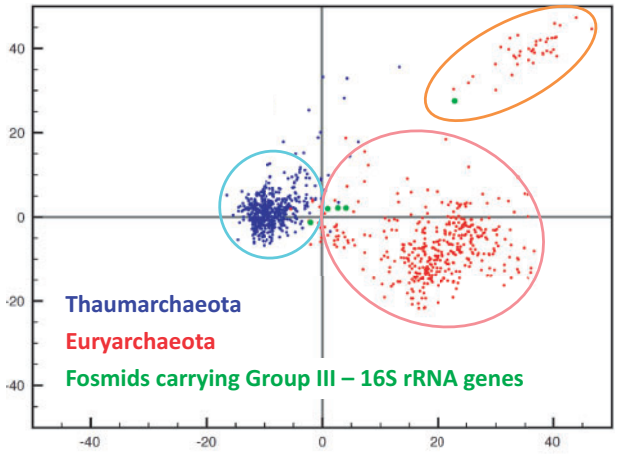

Fig. 2.—PCA of tetranucleotide frequencies in sequenced fosmids for Thaumarchaeota (blue) and GII/III-Euryarchaeota (red).