Supplemental material

Figure S1 Loss of ancient F due to parasitic lifestyle and reductive evolution of archaeal ancestors. **A.** Cumulative loss in the 53 most ancestral F as a function of their age (*nd*, defined as in Fig.1B) reveals the proteome reduction tendency of organisms with parasitic (P) and obligate parasitic (OP) lifestyles, compared to the free-living (FL). **B.** A phylogenomic tree describing the evolution of the 53 most ancestral F after a heuristic maximum parsimony search with tree-bisection-reconnection (TBR) branch swapping and 100 replicates of random addition sequence. Two trees of 11,518 steps (CI=0.258, RI=0.630; g1=-0.512; PTP test, P=0.001) were obtained. F are labeled according to SCOP nomenclature. Bootstrap support values >50% are shown above nodes. Branches labeled with different shades indicate the percentage of genomes that share these architectures. F labeled in black are omnipresent and most of them appear at the base of the tree, as they are the most ancient in the protein world. **C.** Summary of F loss in species. F shared by >90% genomes are missing almost exclusively in parasitic organisms with reduced genomes. Most of FL organisms that lost F are Archaea.

Figure S2 Evolution of biological function along the six phases of the architectural chronology. A. The plots describe how FSF corresponding to coarse-grained functional sub-categories for each superkingdom are used in different phases of the evolutionary timeline. B. Relative acquisition of FSF within individual functional categories. Bars represent the percentage of FSF within a function that is used by a superkingdom, and should be compared horizontally. Four uninformative 'not annotated' FSF (d.58.45 and e.30.1 of phase V, and a.125.1 and d.46.1 of phase VI) were not included in the analysis.

Figure S3 Phylogenomic trees of proteomes described in Figure 3 shown with taxon labels.

Figure S4 Evolutionary mechanisms that affect distribution of architectures in proteomes (*f*): a model. FSF *a* (labeled in red) was discovered at the base of a tree of six proteomes (labeled P1 through P6). Due to its early appearance, FSF *a* distributed widely by sorting in lineages but was lost in the lineage leading to P6 (f=5/6=0.83). Architectural loss

diminished *f*. FSF *b* (blue) was discovered later in evolution and its distribution was confined to only proteomes P4 and P5. Its late appearance decreased *f* (*f*=2/6=0.33). However, an event of HGT extended the distribution of FSF *b* to proteome P3 increasing *f* (*f*=3/6=0.5). The early discovery of FSF combination *cd* (green and purple) distributed it throughout the clade (*f*=6/6=1). However, FSF *cd* components appeared later in evolution when the combination suffered fissions in several lineages. Despite their late appearance, high fission rates were responsible for wide and patchy distributions of FSF *c* (green) and FSF *d* (purple), increasing in both cases *f* (*f*=4/6=0.66).

Figure S5 Phylogenomic analysis strategy. The flowchart describes the use of information embedded in models of molecular structure and entire protein complements (proteomes) to generate hypotheses (phylogenomic trees) about groups of architectures (e.g.. folds F1, F2, …) and proteomes (P1, P2, …) and derivation of architectural chronologies and universal 'trees of life' that describe timelines of structural and organismal diversification from an hypothetical ancestor (a), respectively. This involves a structural census defined by advanced HMMs that assign domain structure to genomic sequences, normalization of data, and phylogenetic analysis. Analyses of architectural distribution in proteomes enable the reconstruction of architectural chronologies. Functional FSF annotations enable the evolutionary study of biological function. This information is later used to select subsets of characters for phylogenomic reconstruction of refined universal trees.