'Geoarchaeon NAG1' is a deeply-rooting lineage of the archaeal order Thermoproteales rather than a new phylum

Lionel Guy, Anja Spang, Jimmy H. Saw and Thijs J. G. Ettema*

Supplementary Material

Supplementary Methods (MS Word document): A more complete description of the methods used in the main article.

Supplementary Figure S1 (PDF): Individual ML phylogenies of the 33 r-proteins used in the concatenated alignment of Figure 1A. Euryarchaetoa and Nanohalarchaeota are shown in red, Nanoarchaeota and ARMAN strains are shown in green, Crenarchaeota in blue, Thaum- and Aigarchaeota in orange, Korarchaeota in pink, and bacteria in grey.

Supplementary Figure S2 (PDF): Individual ML phylogenies of the 25 non-r-proteins used in the concatenated alignment of Figure 1D. Colors as in Supplementary Figure S1.

Supplementary Figure S3 (PDF): Full phylogenies as depicted in Figure 1, i.e. without collapsed taxa. **(A)** ML inferred from 33 universal r-proteins. **(B)** ML inferred from the 50% most compositionally biased sites of 33 universal r-proteins. **(C)** ML inferred from the 50% least compositionally biased sites of 33 universal ribosomal proteins. **(D)** ML inferred from 25 conserved non-r-proteins, filtered for horizontal gene transfers. **(E)** ML inferred from a set of 57 conserved proteins (combination of 32 r-proteins and 25 others), filtered for horizontal gene transfers. **(F)** Bayesian phylogeny from concatenated alignments of the sequences of the 16S and 23S rRNA sequences. Colors as in Supplementary Figure S1.

30   **Supplementary Figure S4 (PDF):** Bayesian phylogeny (phylobayes, CAT-Poisson model)

31   inferred from 33 universal ribosomal proteins. Colors as in Supplementary Figure S1.

32

33   **Supplementary Figure S5 (PDF):** Assessment of the effect of removal of biased sites on

34   the support for placement of NAG1. The x-axis shows the percentage of sites included in the

35   alignment. The y-axis shows the number of times a particular bipartitions is supported in 100

36   non-parametric bootstraps. The three bipartitions tested are: NAG1 grouping with

37   Thermoproteales (Geo+Thermoprot, red); NAG1 separated from Crenarchaeota (Cren-Geo,

38   blue); and NAG1 grouping with Thaum- and Aigarchaeota (Geo+Thaum+Aig, green). The

39   figure shows the analysis for the 33 r-protein concatenated alignment, keeping the most **(A)**

40   or least **(B)** biased sites and for the 25 non-r-protein concatenated alignment, keeping the

41   most **(C)** or least **(D)** biased sites. Vertical dashed lines in A and B mark the alignments from

42   which phylogenies represented in Figure 1B and 1C, respectively, were inferred.

43

44   **Supplementary Figure S6 (PDF):** ML phylogeny of concatenated 16S and 23S genes.

45   Colors as in Supplementary Figure S1.

46

47   **Supplementary Figure S7 (PDF):** Maximum likelihood and Bayesian phylogenies without

48   the fast-evolving ARMAN, Nanohalo- and Nanoarchaeota lineages: **(A)** ML phylogeny of 33

49   r-proteins, as in Figure 1A and Supplementary Figure S2; **(B)** ML phylogeny of 25 conserved

50   non-r-proteins, as in Figure 1D and Supplementary Figure S5; **(C)** ML phylogeny of 57

51   conserved proteins, as in Figure 1E and Supplementary Figure S6; **(D)** Bayesian phylogeny

52   of concatenated 16S and 23S genes, as in Figure 1F and Supplementary Figure S7. Colors

53   as in Supplementary Figure S1.

54

55   **Supplementary Figure S8 (PDF):** Distribution of selected marker genes (Kozubal et al

56   2012, Spang et al 2010) in different archaeal lineages (distribution of r-proteins is presented

57   in Figure 2). Since, the distribution of repair genes might reflect similar environmental

58   adaptation rather than phylogenetic relatedness, these were omitted from this Figure.

59   Additionally, the pattern of repair genes in NAG1 has been shown to be similar to

60   Crenarchaeota (Kozubal et al 2012). Diamonds in blue fields indicate the presence of

61   homologs in most of the members of a group. In case only some members of a group

62   contain homologs, diamonds are shown in parentheses. Asterisks indicate presence in

63   SAGs related to NAG1 only.

64