

Supplementary Figure S1

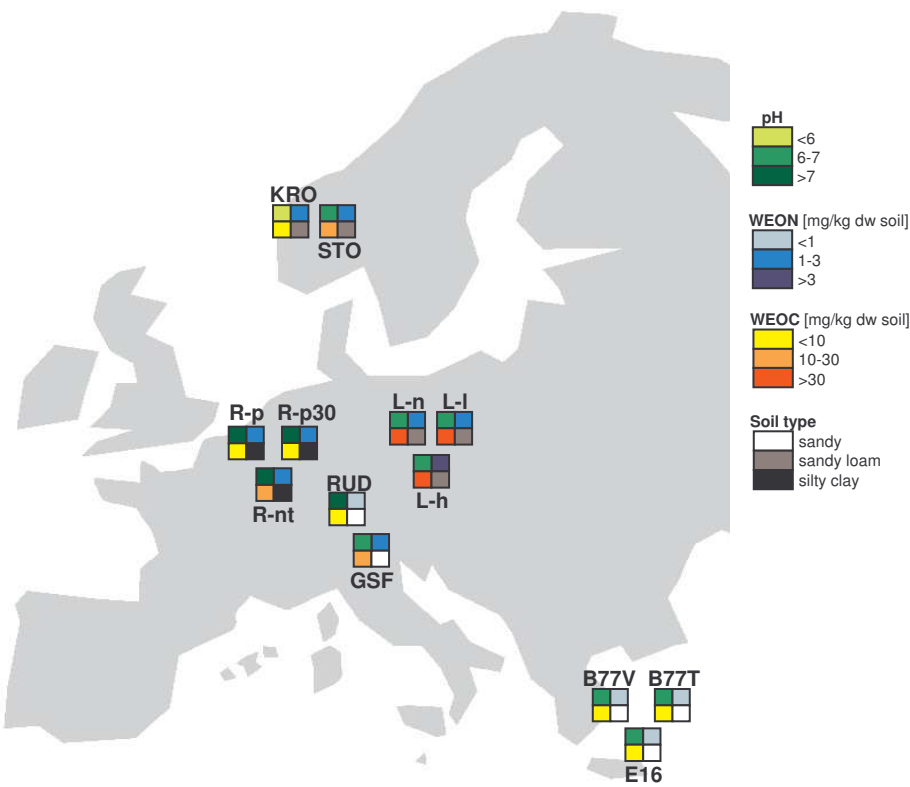


Figure S1: (Geo)Graphical representation of soil samples used in this study, including some of their chemical characteristics as given in Table S1.

Supplementary Figure S2

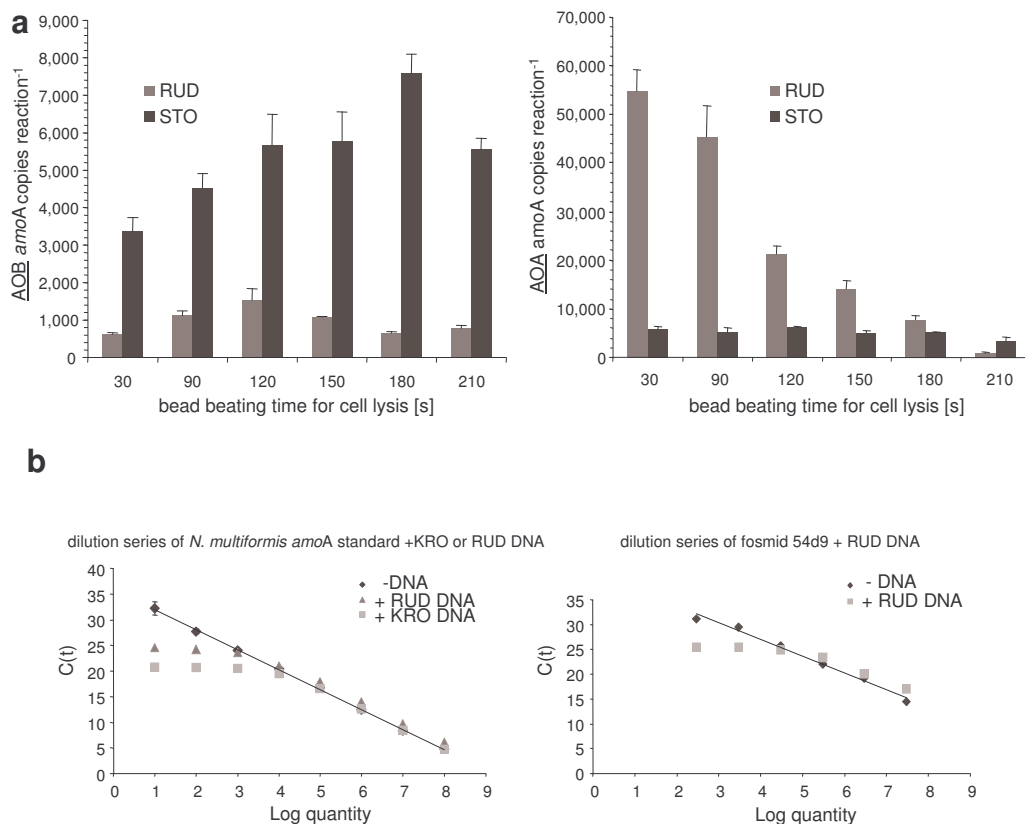


Figure S2: Controls for *amoA* qPCR. **a:** influence of different bead beating times during DNA cell lysis on AOB and AOA *amoA* copy numbers in two different soils using the Griffiths method²⁶. DNA from RUD (light grey) yielded most AOB *amoA* copies at 120 s beat beating time while copy numbers of STO derived DNA (dark grey) were highest at 180 s (left graph). Different results were obtained for AOA *amoA*, for which highest values were measured at 30 s for RUD, while STO revealed constant copy numbers over treatment time (right graph). **b:** influence of inhibitory substances on *amoA* copies in the two soils that were most heavily contaminated by humic acids. Ten ng (4 µl) of soil DNA preparation from KRO or RUD, respectively, was added to diluted standard DNA, in this case from *N. multiformis*. While no inhibition was observed for sample KRO, some inhibition was observed in heavily contaminated soil RUD, reducing the determined *amoA* copy number by up to 50%, (between log 5 and log 8 C(t)-values are slightly higher where soil DNA was added to standard). Inhibition tests on AOA *amoA* qPCR are shown with RUD soil (light grey) and fosmid 54d9 (dark grey) as a standard dilution series. C(t)-values indicated similar inhibition as in AOB *amoA* qPCR. C(t) = cycle threshold in qPCR, log quantity = log of template copy numbers.

Supplementary Figure S3

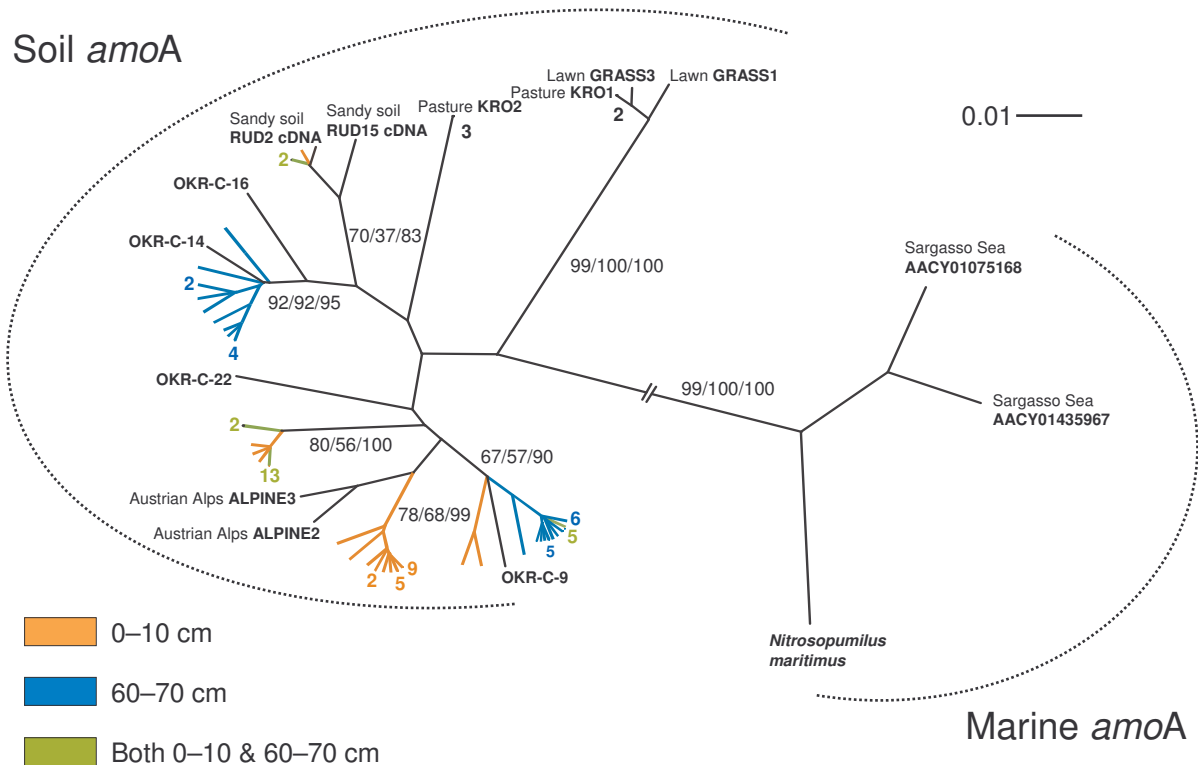


Figure S3: Phylogeny of 72 derived crenarchaeal *amoA* sequences recovered from two distinct depths (0 to 10 cm and 60 to 70 cm) of a soil profile sampled at Rotböll, Germany. Sequences from other surveys included those from Rotböll, Germany⁸ (RUD: accession numbers DQ534808-DQ534814); Alpine soil, Austria (R: DQ534698-DQ534699); Public lawn, Scotland (GRASS: DQ534702, DQ534704); Pasture, Norway (KRO: DQ534815, DQ534816), Tennessee soil, USA⁵⁰ (OKR: DQ148872, DQ148876, DQ148877 and DQ148882), Sargasso Sea¹⁰ (AACY01075168 and AACY01435967) and *Nitrosopumilus maritimus*¹¹ (DQ085098). Values at tips indicate the number of identical sequences represented by the individual branch [orange: all from 0-10 cm; blue: all from 60-70 cm; green: from both 0-10 and 60-70 cm (and +/- from other surveys); black: from other surveys only]. Maximum likelihood distances were calculated from 169 aligned amino acid positions and values at major nodes represent bootstrap support from protein distance/protein maximum-parsimony/DNA distance analyses (expressed as a percentage of 1000 replicates). LogDet DNA distances were calculated from an initial alignment of 507 nucleotides after all 3rd codon positions were excluded. Scale bar indicates 0.01 changes per site.

Supplementary Figure S4

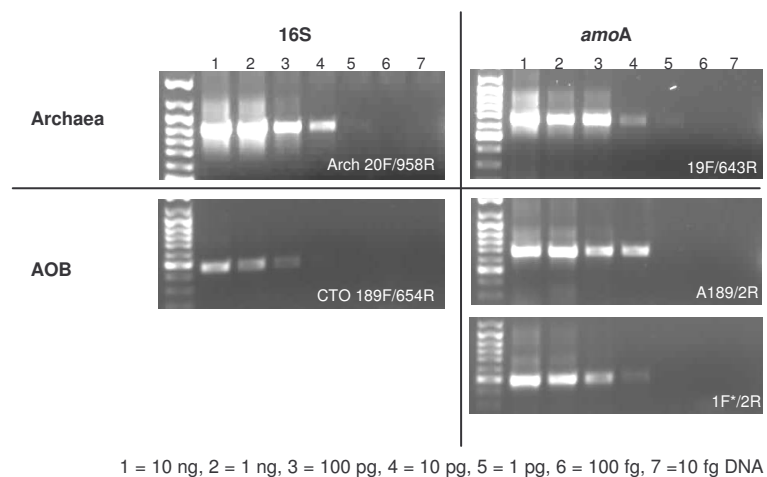


Figure S4: Typical results of an MPN PCR experiment with different dilutions of DNA extracted from the L-h top soil. Of the five different primer pairs used, two targeted 16S rRNA genes and three targeted *amoA* genes (indicated bottom right of each panel).

Supplementary Figure S5

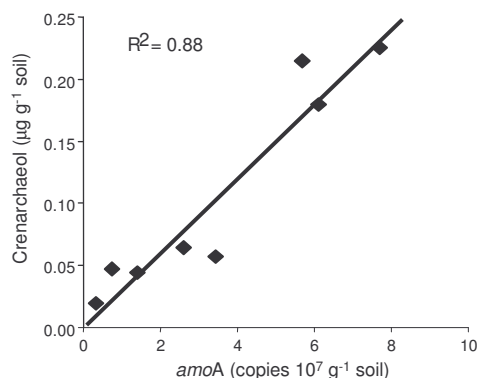


Figure S5: Bivariate plot of crenarchaeol abundance (μg g⁻¹ dry soil) versus AOA *amoA* copies (10⁷ g⁻¹ dry soil). The high coefficient of correlation ($R^2 = 0.88$; $n = 8$) indicates a common origin of crenarchaeol from ammonia oxidizers. GDGT and crenarchaeol concentrations determined for samples from managed grassland KRO and STO were excluded, due to deviations in archaeal lipid composition, which was strongly dominated by non-isoprenoidal GDGT (see chromatogram in Fig. 2b). In managed grasslands caldarchaeol comprised >75% and crenarchaeol <10% of isoprenoidal GDGT, whereas in the pristine grasslands and arable soils compositions were <60% caldarchaeol and >30% crenarchaeol (see also Supplementary Table S1).

Supplementary Figure S6

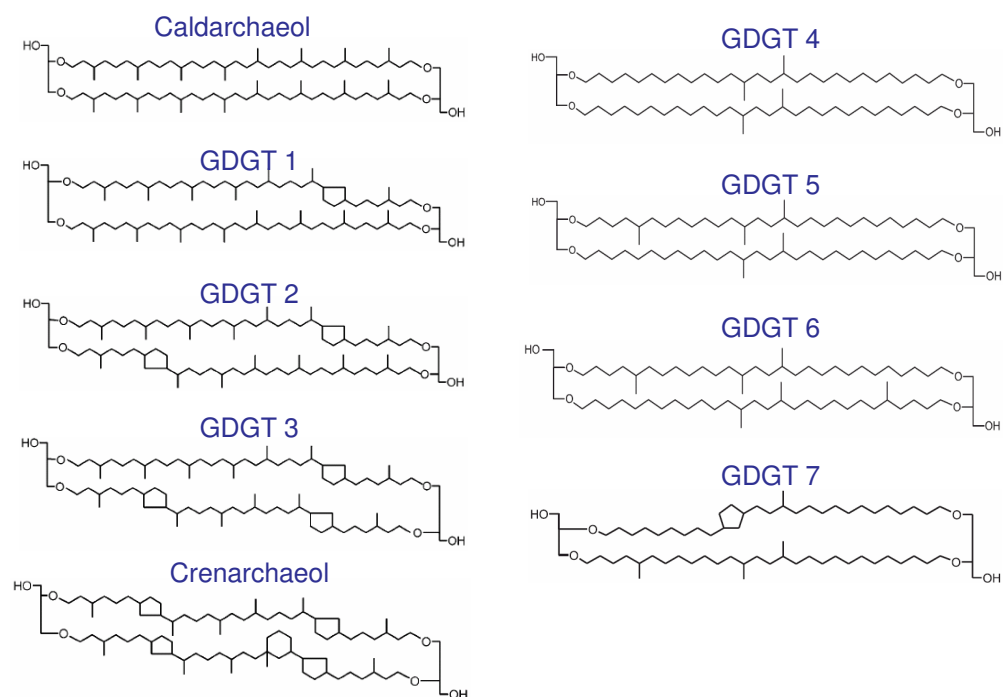


Figure S6: Structures of glyceroldialkylglycerol tetraethers encountered in grassland and arable soil. Names and numbers refer to labels in chromatograms (Fig. 2).