Fig. S1. Maximum likelihood phylogeny of full-length nosZ amino acid sequences, excluding signal peptide and C-terminal extensions, obtained from 138 genomes and 4 contigs showing the grouping of clades I and II. Symbols on tree tips specify major taxonomic group of source organisms, and scale indicates corrected amino-acid substitutions per site (LG+Γ+F). Genotypes with red names were used to test primer sets in the laboratory, and those followed by an asterisk indicate nosZ sequences obtained from contigs in which the signal peptide could be identified. Nodes with >70% bootstrap support (n=500) are indicated (●). The consensus of the clade II sequences at the forward and reverse primer sites are shown as SeqLogo graphs in the upper right, with the corresponding sequences for each genotype in the columns below. Positions with a dot indicate matches to the dominant (> 70% consensus) nucleotide among the clade II sequences at each primer site.