

**Supplement figure 1.** Section of the Garzweiler open-pit mine (crossed out) and studied reclamation areas, modified from Pihlap *et al.* (2019).

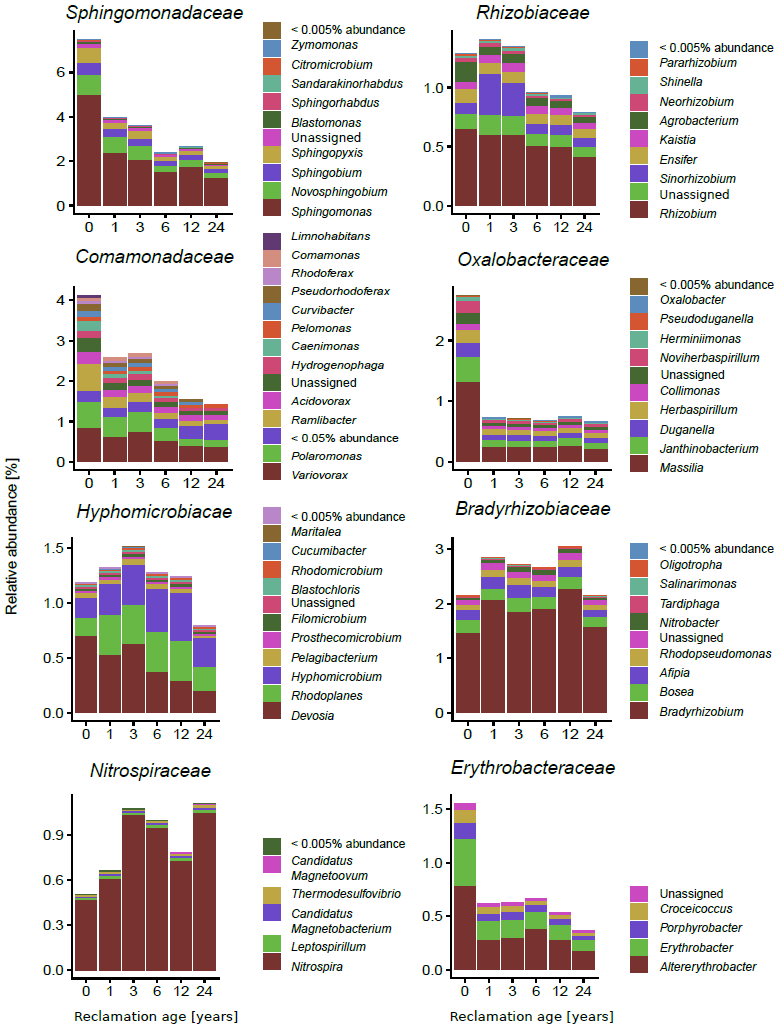
A screenshot of a cell phone

Description automatically generated

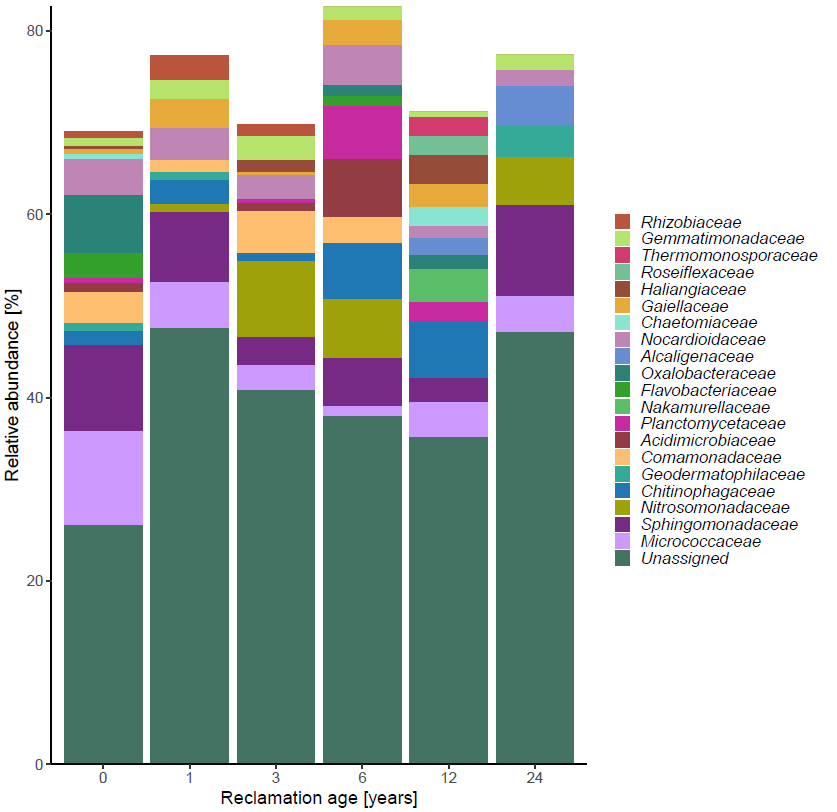
**Supplement figure 2.** Protein fraction of the extracellular substances calculated as equivalent of bovine serum albumin used as the standard, per gram of dry soil. Bars represent the means of triplicates with a standard deviation. Letters above bars represent pairwise comparisons adjusted with the Benjamini-Hochberg correction showing the significant differences between reclamation ages after a robust ANOVA test with trimmed means (p < 0.005).



**Supplement figure 3.** Sequencing summary showing the average number of raw and filtered reads of all metagenomes, the number of filtered reads per sample, and the average read length per sample.



**Supplement figure 4.** Distributionof reads assigned to genera within the most abundant families. Shown are mean abundances relativized to the number of bacterial reads.



**Supplement figure 5.** Distribution of the most abundant bacterial families based on the 16S rRNA gene (V4 region), as assigned by Ribotagger and SILVA database. Shown are mean abundances relativized to all reads assigned to the family level.

A screenshot of a cell phone

Description automatically generated

**Supplement figure 6.** Absolute gene abundance estimation. The mean value of 16S rRNA gene copy number in each reclamation age was divided by the highest mean value (found in RA24). Resulting ratios were multiplied with the mean relative abundance of each gene analysed from the metagenome to obtain a rough estimation of the absolute gene abundance in each reclamation age compared to RA24.

|  |  |  |  |
| --- | --- | --- | --- |
| **Gene** | **Protein** | **KO number** | **HMM ID** |
| *wza* | polysaccharide export outer membrane protein - Wza | K01991 | PF02563 |
| *algE* | alginate export outer membrane protein - AlgE | K16081 | PF13372 |
| *algJ* | alginate biosynthesis acetyltransferase - AlgJ | K19295 | PF16822 |
| *wcaB* | colanic acid biosynthesis acetyltransferase - WcaB | K03819 | TIGR04016 |
| *wcaF* | colanic acid biosynthesis acetyltransferase - WcaF | K03818 | TIGR04008 |
| *wcaK/amsJ* | colanic acid/amylovoran biosynthesis pyruvyl transferase - WcaK/AmsJ | K16710 | TIGR04006 |
| *kpsE* | capsular polysaccharide export system permease - KpsE | K10107 | TIGR01010 |
| *epsG* | exopolysaccharide biosynthesis transmembrane protein - EpsG | K19419 | PF14897 |
| *epsA* | exopolysaccharide biosynthesis tyrosine kinase modulator - EpsA | K19420 | TIGR01006 |
| *sacB* | levansucrase - SacB | K00692 | PF02435 |
| *wzt* | lipopolysaccharide transport system ATP-binding protein - Wzt | K09691 | PF14524 |
| *lptF* | LptBFGC lipopolysaccharide export complex permease - LptF | K07091 | TIGR04407 |
| *lptG* | LptBFGC lipopolysaccharide export complex permease - LptG | K11720 | TIGR04408, PF03739 |
| *lptC* | LptBFGC lipopolysaccharide export complex inner membrane protein - LptC | K11719 | TIGR04409, PF06835 |

**Supplement table 1.** Analysed genes related to exo- and lipopolysaccharide production, the proteins they code for, and their respective KO numbers and HMM IDs.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Reclamation age [years]** | **Coverage [%]** | **Diversity index** |
| Initial phase | 0 | 3.3a | 22.9a |
| 0 | 5a | 22.51a |
| 0 | 4a | 22.81a |
| 1 | 2.3ab | 25.35b |
| 1 | 3ab | 27b |
| 1 | 4.9ab | 25.86b |
| 3 | 1.7ab | 24.15a |
| 3 | 2.9ab | 22.79a |
| 3 | 3ab | 24.42a |
| Agricultural management phase | 6 | 1.5b | 22.59a |
| 6 | 1.5b | 22.54a |
| 6 | 1.6b | 22.35a |
| 12 | 0.08b | 22.49a |
| 12 | 1.4b | 22.11a |
| 12 | 1.2b | 22.72a |
| 24 | 1b | 22.6a |
| 24 | 2.5b | 22.27a |
| 24 | 2b | 22.25a |
|  |  |  |

**Supplement table 2.** Nonpareil coverage estimation and diversity index per replicate. Letters represent pairwise comparisons showing the significant differences between reclamation ages after a robust ANOVA test with trimmed means (p = 0.049, p = 0.014 for coverage and diversity, respectively).