Supplementary methods

Field sites

Glenrock and Bogo woodlands were relatively open and Glenrock had scattered individuals *of A. dealbata*, whilst at Bogo there were no acacias present. At Glenrock and Bogo patches of native Australian grasses were more frequent than at Talmo, whilst still not being extensive, and some exotic grasses were also present at the Bogo woodland plot. The sites were chosen within 15 km of each other to minimise variability in microclimate. The distances between pairs of pasture and woodland plots were: Talmo, 160 m; Glenrock, 200 m; and Bogo, 440 m. At each site, woodland plots were located at slightly higher elevation than pastures and flux of fertilisers and minerals to the woodlands was therefore minimised. The close proximity between plots also minimised differences in parent soil type. Each plot (100 x 100 m) was gridded by 25 m intervals, and samples were taken at all grid intersections (see appendix Fig. S1). For three of the 25 m^2 subplots within each plot, additional samples were taken at 12.5 m intervals.

Soil biogeochemistry analysis

Soil organic C fractions (particulate-, humus- and resistant-organic carbon (POC, HOC and ROC) and clay composition were determined using diffuse reflectance mid-infra red (MIR) spectra (Nicolet 6700 FTIR spectrometer equipped with a KBr beam-splitter and a DTGS detector, Thermo Fisher Scientific Inc, MA, USA) acquired over 8000 – 4000 cm^{-1} (resolution of 8 cm^{-1}). Organic C fractions were determined as described in Baldock *et al*. (1) and clay content was estimated using the prediction algorithms of Janik and Skjemstad (2). The level of confidence in the predicted data was deemed appropriate after assessment of the associated error statistics which were below the respective thresholds (inlier/outlier ratio <1

for Baldock *et al.* (1); F-statistic <3 for Janik and Skjemstad (2) indicating that the overall spectral characteristics of the soils were represented within the wider calibration dataset. However, it is acknowledged that the calibration datasets are largely representative of Australian agricultural soils and not woodland soils.

GH48 primer design

The primer pair (GH48_F1: 5'-RRCATBTACGGBATGCACTGGCT-3' and GH48_R1: 5'-VCCGCCCCABGMGTARTACC-3') amplified a region of the GH48 catalytic region varying from 483 bp to 513 bp. Primer specificity was tested using MFE-primer 2.0 (3), using a wide range of in-silico PCR conditions (Tm from 47 to 80°C, product size from 400 to 600 bp and a PPC cut off of 30%) which allowed the detection of non-specific amplification products even if these are unlikely under the PCR conditions used in the laboratory. We further confirmed specificity by cloning and sequencing PCR products.

The inclusion of GH48 sequences from *Cellulomonas* spp., *Jonesia denitrificans* and *Xylanimonas cellulosilytica* led to a high number of degenerate bases in the primers and increased mismatches between primers and target sequences. Therefore the sequences from those species were removed from the alignment and a second primer pair was designed specifically for these species. The second primer pair (GH48_F1_cell: 5'-AYGTCGACAACRTSTACGGMTWCG-3' and GH48_R1_cell: 5' CCGCCCCASGCSWWRTACC-3') was also used for cloning and T-RFLP analysis, however the diversity recovered with both primer pairs was similar and in-silico analysis of primer specificity indicated that primer pair 1 (GH48_F1 and GH48_R1) would also amplify the GH48 gene from *Jonesia denitrificans* and *Xylanimonas cellulosilytica*.

GH48 qPCR primer design

The primer sequences were qPCR_GH48_F8: 5'-GCCADGHTBGGCGACTACCT-3' and qPCR_GH48_R5: 5'- CGCCCCABGMSWWGTACCA-3'. The pair designed here has similar lengths (19 or 20 bp), melting temperatures (55-60°C) and amplify a region of 150 bp. In-silico specificity was assessed using MFEprimer 2.0 (Tm from 47 to 80°C, product size from 100 to 300 bp and a PPC cut off of 30%) (3). Amplification was tested using DNA from soil, GH48 clones obtained with the PCR primers described above (GH48_F1 and GH48_R1) and DNA from a local soil actinomycete isolate. Following amplification of soil DNA, visual inspection of PCR products revealed the presence of expected sizes, and sequencing of 8 soil clones showed that all high quality sequence clones obtained had top BLASTn hits to actinobacterial GH48 sequences.

Supplementary methods references

1. **Baldock JA, Hawke B, Sanderman J, Macdonald LM.** 2013. Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. Soil Res **51:**577-583.
2. **Janik LJ, Skjemstad JO.** 1995. Characterization and analysis of soils using midinfrared partial least-squares 2. Correlations with some laboratory data. Australian Journal of Soil Research **33:**637-650.
3. **Qu W, Zhou Y, Zhang Y, Lu Y, Wang X, Zhao D, Yang Y, Zhang C.** 2012. MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity. Nucleic Acids Res **40:**W205-W208.

Supplementary tables

Table S1: Results of in-silico analysis of GH48 primer specificity using MFEprimer. Primer specificity using a wide range of in-silico PCR conditions (Tm from 47 to 80°C, product size from 100 to 300 bp (qPCR primers) and 400 to 600 bp (standard PCR primers) and a PPC cut off of 30%).

qPCR_GH48_F8 - qPCR_GH48_R5		
Sequence name	highest score	PPC	Species name
CP000481	96.4	*Acidothermus cellulolyticus*
CP006259	96.4	*Streptomyces collinus*
FR845719	96.4	*Streptomyces venezuelae*
CP000667	96.4	*Salinispora tropica* CNB-440
CP000850	96.4	*Salinispora arenicola* CNS-205
CP001630	96.4	*Actinosynnema mirum*
CP001814	96.4	*Streptosporangium roseum*
CP002162	96.4	*Micromonospora aurantiaca*
CP002399	96.4	*Micromonospora* sp. L5
L38827	96.4	*Cellulomonas fimi*
CP002047	96.4	*Streptomyces bingchenggensis*
CP003720	96.4	*Streptomyces hygroscopicus*
CP003170	96.4	*Actinoplanes* sp. SE50/110
AP012319	96.4	*Actinoplanes missouriensis*
CP000088	95	*Thermobifida fusca*
CP001964	95	*Cellulomonas flavigena*
AP010968	86.4	*Kitasatospora setae*
AL939128	78.3	*Streptomyces coelicolor*
CP002993	78.3	*Streptomyces* sp. SirexAA-E
FN554889	78.3	*Streptomyces scabiei*
CP001874	78.3	*Thermobispora bispora*
CP002040	78.3	*Nocardiopsis dassonvillei*
CP003729	78.3	*Amycolatopsis mediterranei*
CP002665	76.1	*Cellvibrio gilvus*

GH48_F1-GH48_R1		
Sequence name	highest score	PPC	*Species name*
CP000481	90.1	*Acidothermus cellulolyticus*
AP012319	90.1	*Actinoplanes missouriensis*
CP005929	90.1	*Actinoplanes* sp. N902-109
CP006272	90.1	*Actinoplanes friuliensis*
CP001630	90.1	*Actinosynnema mirum*

CP001700	90.1	*Catenulispora acidiphila*
NC_007645.1	90.1	*Hahella chejuensis*
CP001706	90.1	*Jonesia denitrificans*
CP000850	90.1	*Salinispora arenicola* CNS-205
CP000667	90.1	*Salinispora tropica* CNB-440
BA000030	90.1	*Streptomyces avermitilis*
AL939128	90.1	*Streptomyces coelicolor*
FN554889	90.1	*Streptomyces scabiei*
CP001814	90.1	*Streptosporangium roseum*
CP000088	90.1	*Thermobifida fusca*
CP001874	90.1	*Thermobispora bispora*
CP001821	90.1	*Xylanimonas cellulosilytica*
CP002040	87	*Nocardiopsis dassonvillei*
CP003170	79.6	*Actinoplanes* sp. SE50/110
HE971709	79.6	*Streptomyces davawensis*
CP009124	79.6	*Streptomyces lividans*
CP009438	79.6	*Streptomyces glaucescens*
CP003729	79.6	*Amycolatopsis mediterranei*
AP010968	79.6	*Kitasatospora setae*
CP002162	79.6	*Micromonospora aurantiaca*
CP002399	79.6	*Micromonospora* sp. L5
CP002638	79.6	*Verrucosispora maris*

GH48_F1_cell-GH48_R1_cell

Sequence name	highest score	PPC	*Species name*
BA000030	83.6		*Streptomyces avermitilis*
CP005929	83.6		*Actinoplanes* sp. N902-109
LK022848	83.6		*Streptomyces iranensis*
CP009124	83.6		*Streptomyces lividans*
CP009438	83.6		*Streptomyces glaucescens*
CP006567	83.6		*Streptomyces rapamycinicus*
CP006259	83.6		*Streptomyces collinus*
FR845719	83.6		*Streptomyces venezuelae*
CP001630	83.6		*Actinosynnema mirum*
CP001706	83.6		*Jonesia denitrificans*
CP001821	83.6		*Xylanimonas cellulosilytica*
FN554889	83.6		*Streptomyces scabiei* 87.22
CP001964	83.6		*Cellulomonas flavigena*
CP002638	83.6		*Verrucosispora maris*
L38827	83.6		*Cellulomonas fimi*
CP002993	83.6		*Streptomyces* sp. SirexAA-E
AP010968	83.6		*Kitasatospora setae*
CP002047	83.6		*Streptomyces bingchenggensis*
CP003720	83.6		*Streptomyces hygroscopicus*
CP006272	82.2		*Actinoplanes friuliensis*
CP000481	79.2		*Acidothermus cellulolyticus*

HE971709	79.2	*Streptomyces davawensis*
CP001874	79.2	*Thermobispora bispora*
CP002399	79.2	*Micromonospora* sp. L5
CP003170	79.2	*Actinoplanes* sp. SE50/110
CP002665	58.6	*Cellvibrio gilvus*

Table S2: Summary of actinobacterial GH48 genes amplified by each primer pair according to MFEprimer analysis.

Species	qPCR_GH48_F8 - qPCR_GH48_R5	GH48_F1-GH48_R1	GH48_F1_cell-GH48_R1_cell
CP001630 *Actinosynnema mirum*	x	x	x
CP002040 *Nocardiopsis dassonvillei*	x	x	
CP000481 *Acidothermus cellulolyticus*	x	x	
CP000850 *Salinispora arenicola*	x	x	
CP000667 *Salinispora tropica*	x	x	
CP001874 *Thermobispora bispora*	x	x	x
CP001821 *Xylanimonas cellulosilytica*		x	x
AP012319 *Actinoplanes missouriensis*	x	x	
CP005929 *Actinoplanes* sp. N902-109		x	x
CP003170 *Actinoplanes* sp. SE50/110	x	x	x
CP006272 *Actinoplanes friuliensis*		x	x
CP001700 *Catenulispora acidiphila*		x	
L38827 *Cellulomonas fimi*	x		x
CP001964 *Cellulomonas flavigena*	x		x
CP001706 *Jonesia denitrificans*		x	x
AP010968 *Kitasatospora setae*	x	x	x
CP002399 *Micromonospora* sp. L5	x	x	x
CP002162 *Micromonospora aurantiaca*	x	x	x
AL939128 *Streptomyces coelicolor*	x	x	
HE971709 *Streptomyces davawensis*		x	x
BA000030 *Streptomyces avermitilis*		x	x
CP003720 *Streptomyces hygroscopicus*	x		x
CP002047 *Streptomyces bingchenggensis*	x		x
CP003990 *Streptomyces* sp. PAMC26508			
CP002475 *Streptomyces pratensis*			
LK022848 *Streptomyces iranensis*			x
CP009124 *Streptomyces lividans*		x	x
CP009438 *Streptomyces glaucescens*		x	x
CP006567 *Streptomyces rapamycinicus*			x
CP006259 *Streptomyces collinus*	x		x
FR845719 *Streptomyces venezuelae*	x		x
CP002993 *Streptomyces* sp. SirexAA-	x		x
FN554889 *Streptomyces scabiei* (C9ZEQ0)	x	x	x
FN554889 *Streptomyces scabiei* (C9Z9L6)			
CP001814 *Streptosporangium roseum*	x	x	

CP000088	*Thermobifida fusca*	x	x	
CP002638	*Verrucosispora maris*		x	x
CP003729	*Amycolatopsis mediterranei*	x	x	
CP002665	*Cellvibrio gilvus*	x		x

Table S3: Percentage of GH48 clones showing the presence of conserved residues with known function in cellobiohydrolase function. Residues numbers are according to the Cel48F gene from *Clostridium cellulolyticum* H10.

Residue	Clones with residue (%)	Function	Presence in *T. fusca* GH48 protein
N178	100	Hydrogen bonding	yes
F180	57%	Hydrophobic stacking interactions	no
Q181	100	Hydrogen bonding	yes
G183	100	Hydrogen bonding	yes
Q185	47%	Calcium coordination	no
E186	100	Hydrogen bonding	yes
E190	71%	Calcium coordination	yes
Q222	100	Hydrogen bonding	yes
T226	88.7%	Hydrogen bonding	yes
K274	98.7%	Hydrogen bonding	yes
Y275	96%	Hydrogen bonding	yes
W298	98.7%	Hydrogen bonding	yes

Supplementary figures

Figure S1

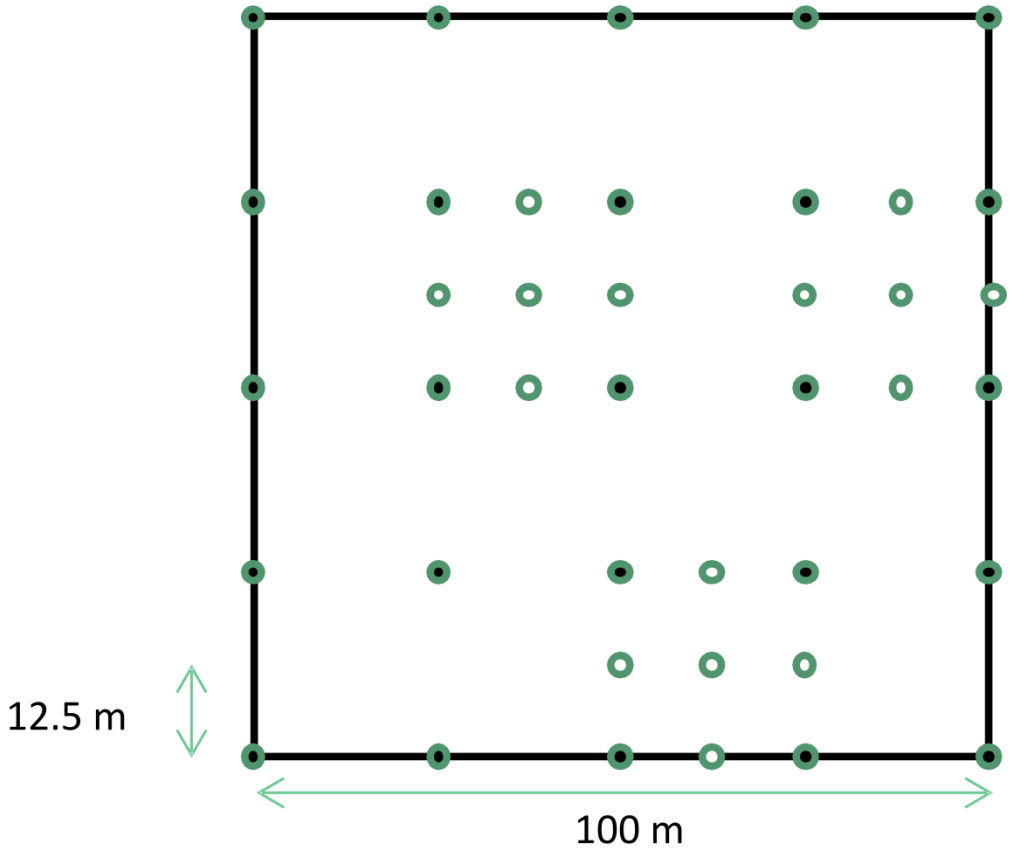

Figure S1. Sampling grid used in this study. The 1 hectare plots were gridded by 25 m intervals, and samples were taken at all grid intersections . For three of the 25 m^2 subplots within each plot, additional samples were taken at 12.5 m intervals.

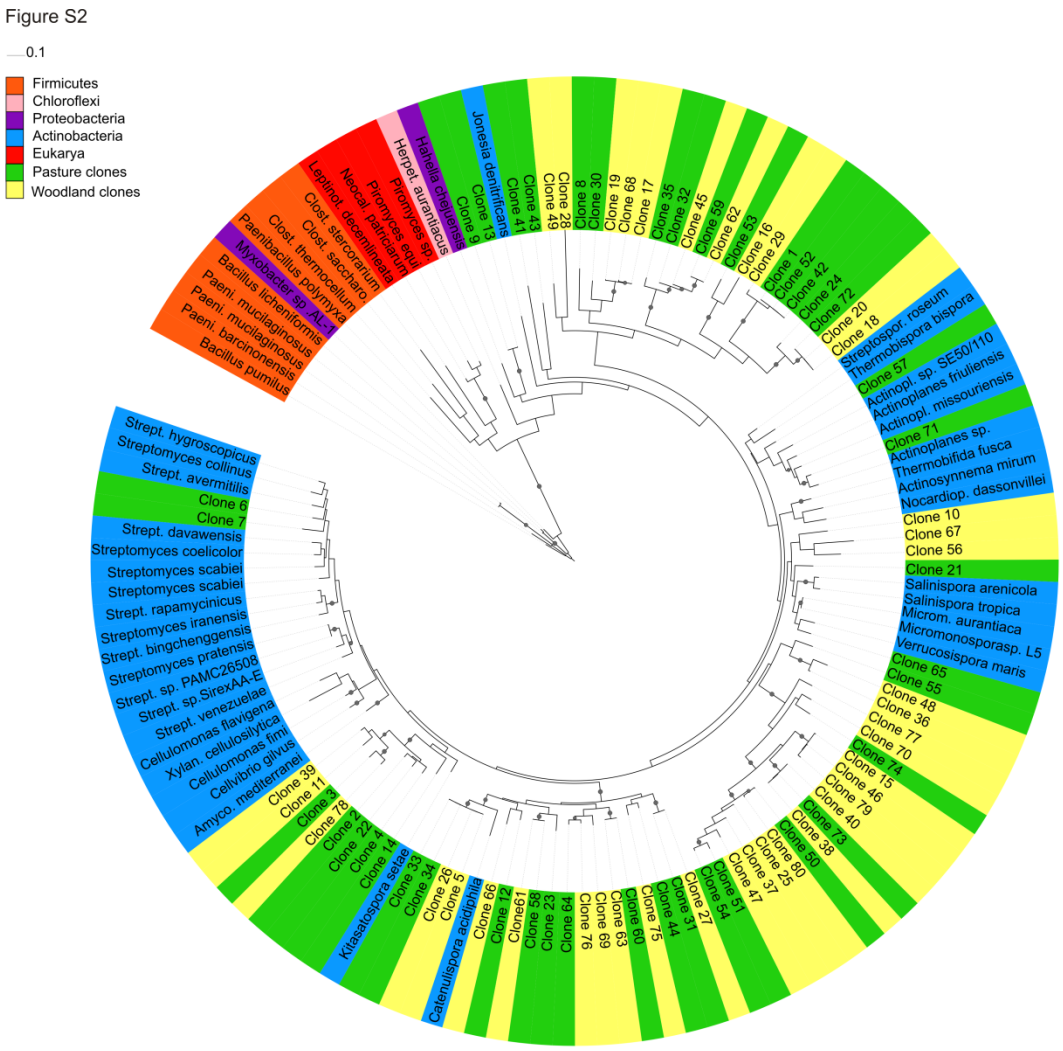

Figure S2. Maximum likelihood tree (PhyML) constructed with GH48 sequences from soil clones and cultured strains from Actinobacteria, Firmicutes, Neocallimastigales (anaerobic fungi), Proteobacteria, Chloroflexi and Insecta. Nodes in tree branches indicate bootstrap support > 0.8. Sequences from *Bacillus* spp. and *Paenibacillus* spp. were used as outgroups. Sequence accessions are indicated following strain name. Colours indicate sequence taxonomy or soil clone provenance. TW, Talmo woodland; GW, Glenrock woodland; BW, Bogo woodland; TP, Talmo pasture; GP, Glenrock pasture; BP, Bogo pasture.