environmental microbiology

Environmental Microbiology (2015) doi:10.1111/1462-2920.12986

Determinants of bacterial communities in Canadian agroforestry systems

Samiran Banerjee,[1,3]* Mark Baah-Acheamfour,[1] Cameron N. Carlyle,[2] Andrew Bissett,[3] Alan E. Richardson,[3] Tariq Siddique,[1] Edward W. Bork[2] and Scott X. Chang[1]**

[1]*Department of Renewable Resources, University of Alberta, 442 Earth Science Building, Edmonton, Alberta T6G 2E3, Canada.*

[2]*Department of Agricultural, Food and Nutritional Science, University of Alberta, 410 Agriculture/Forestry Centre, Edmonton, Alberta T6G 2H1, Canada.*

[3]*CSIRO Agriculture Flagship, Crace, ACT 2911, Australia.*

Summary

Land-use change is one of the most important factors influencing soil microbial communities, which play a pivotal role in most biogeochemical and ecological processes. Using agroforestry systems as a model, this study examined the effects of land uses and edaphic properties on bacterial communities in three agroforestry types covering a 270 km soil-climate gradient in Alberta, Canada. Our results demonstrate that land-use patterns exert stronger effects on soil bacterial communities than soil zones in these agroforestry systems. Plots with trees in agroforestry systems promoted greater bacterial abundance and to some extent species richness, which was associated with more nutrient-rich soil resources. While *Acidobacteria*, *Actinobacteria* and *Alphaproteobacteria* were the dominant bacterial phyla and subphyla across land uses, *Arthrobacter*, *Acidobacteria_Gp16*, *Burkholderia*, *Rhodanobacter* and *Rhizobium* were the keystone taxa in these agroforestry systems. Soil pH and carbon contents emerged as the major determinants of bacterial community characteristics. We found non-random co-occurrence and modular patterns of soil bacterial communities, and these patterns were controlled by edaphic factors and not their taxonomy. Overall, this study highlights the drivers and co-occurrence patterns of soil microbial communities in agroforestry systems.

Introduction

Forty per cent of Canada's land area is covered by forests, with approximately 5% (50 million ha) currently being used for agroforestry systems (Thevathasan and Gordon, 2004). Agroforestry is a complementary system of land management, whereby trees and/or shrubs are grown alongside herblands with crops and/or livestock to promote diversity and ecological sustainability while offering social, economic and environmental benefits (AAFC, 2014). Among the various types of agroforestry systems prevalent in Canada, shelterbelt, hedgerow and silvopasture systems are predominant in western Canada (AAFC, 2014). Among the benefits rendered by agroforestry systems is their role in carbon sequestration and mitigation of greenhouse gas emissions, which have received considerable research interest in the last two decades (Thevathasan and Gordon, 2004; Baah-Acheamfour *et al.*, 2014). Soil microbial communities play a central role in most biogeochemical and ecological processes (Bardgett and van der Putten, 2014). Thus, information on microbial community composition, diversity and their determinants is critical to understanding and managing these processes in agroforestry systems. However, studies examining soil microbial communities and their drivers in agroforestry systems are limited and often the findings are contradictory and inconclusive (Saggar *et al.*, 2001; Lacombe *et al.*, 2009; Rivest *et al.*, 2013). For example, while Lacombe and colleagues (2009) and Rivest and colleagues (2013) found higher microbial biomass in agroforestry systems, Saggar and colleagues (2001) reported the opposite.

Land-use change is one of the most significant factors influencing belowground diversity, with effects predicted to be greater than that of climate change by 2100 (Sala *et al.*, 2000; Lauber *et al.*, 2008). Agroforestry systems co-occurring with other land uses can serve as models to address how soil microbial communities respond to land-use change and how we might best manage these responses. Agroforestry systems differ in their plant cover and species composition, factors which directly affect the quality and quantity of organic matter input with

For correspondence. *E-mail samiran.banerjee@csiro.au; Tel. +61 2 6246 4863; Fax +61-2-6246-5000. **E-mail scott.chang@ales.ualberta.ca; Tel. +1 780 792 6375; Fax +1-780492-1767.
Received 31 March, 2015; accepted 12 July, 2015.

Table 1. Soil and microbial properties measured at 0–10 cm depth in 54 soil samples obtained from three types of agroforestry systems across a 270 km soil-climate gradient in Alberta, Canada.

	Shelterbelt		Hedgerow		Silvopasture	
Soil properties	*Agricultural*	*Shelterbelt*	*Agricultural*	*Hedgerow*	*Grassland*	*Woodland*
Bact. 16S rRNA*#$	6.61 (0.08)[a]	6.91 (0.07)[abc]	6.67 (0.12)[a]	7.17 (0.13)[bc]	6.85 (0.13)[ab]	7.22 (0.20)[c]
pH	5.74 (0.24)[a]	6.00 (0.17)[a]	5.44 (0.13)[a]	5.67 (0.19)[a]	6.15 (0.37)[a]	5.87 (0.30)[a]
WHC (%)	74.23 (3.55)[ab]	72.59 (7.99)[ab]	62.47 (6.00)[a]	104.62 (16.32)[ab]	64.06 (12.29)[ab]	115.08 (25.34)[b]
Moisture (%)	26.48 (2.72)[a]	18.47 (1.98)[a]	22.30 (2.85)[a]	31.75 (8.24)[a]	22.25 (6.75)[a]	38.64 (12.74)[a]
NH_4-N	1.18 (0.26)[a]	1.66 (0.39)[a]	1.05 (0.16)[a]	3.28 (1.00[ab]	1.78 (0.81)[a]	4.58 (1.58)[b]
NO_3-N	6.89 (0.80)[a]	3.82 (1.17)[a]	5.73 (0.92)[a]	3.20 (1.75)[a]	5.28 (1.71)[a]	6.74 (3.06)[a]
TN	0.39 (0.03)[a]	0.46 (0.06)[a]	0.38 (0.08)[a]	0.62 (0.13)[a]	0.39 (0.10)[a]	0.70 (0.19)[a]
TOC	4.64 (0.38)[a]	5.87 (0.71)[ab]	4.55 (1.07)[a]	8.61 (2.02)[ab]	5.40 (1.63)[ab]	10.05 (2.61)[b]
DOC	2.05 (0.17)[ab]	5.31 (1.17)[c]	1.70 (0.14)[a]	4.29 (0.54)[bc]	2.39 (0.50)[abc]	5.09 (1.20)[c]
DON	0.85 (0.08)[a]	0.86 (0.24)[a]	0.73 (0.11)[a]	0.64 (0.13)[a]	0.60 (0.16)[a]	1.03 (0.24)[a]
TC	102.65 (8.73)[ab]	265.53 (58.63)[b]	85.14 (7.21)[a]	214.17 (27.20)[ab]	119.53 (25.20)[ab]	254.38 (60.13)[ab]

*Unit of bacterial 16S rRNA is $\log_{10}$ copies per gram of dry soil and all other soil properties (with exception of soil pH and moisture %) are mg kg^{-1} soil.
#Each value is the mean of nine replicates with standard errors in parentheses.
$Within each row different letters show statistical significance at $P < 0.05$.
Each agroforestry system consisted of a forested plot (shelterbelt or hedgerow or woodland) and an adjacent herbland plot (agricultural land or grassland). One-way ANOVA with Duncan's post hoc test was performed on six land-use types ($n = 6$).

subsequent effects on soil properties. While the linkage between plant and microbial diversity has been proposed for a long time (Wardle *et al*., 2004; van der Heijden *et al*., 2008), a recent study showed that plant diversity is positively associated with microbial beta but not alpha diversity across biomes (Prober *et al*., 2015). This suggests that treed plots in agroforestry systems with greater plant species composition may harbour more diverse microbial communities. Furthermore, little is known about the specificity of soil bacterial taxa in agroforestry systems. For example, it is possible that some taxa have greater habitat breadth (generalists) across agroforestry systems, while others are restricted to particular systems (specialists) (Barberán *et al*., 2012).

Linking microbial distribution patterns (abundance, community composition and diversity) to their drivers remains a central goal in microbial ecology. Identifying determinants of bacterial communities in agroforestry systems requires looking beyond mere sample-level comparison and exploring how edaphic factors are related with microbial modules. Microbial co-occurrence patterns can disentangle complex microbial communities and delineate the underlying ecological processes (Fuhrman, 2009). Network analysis of microbial co-occurrence patterns has been found insightful in recent years, with studies emerging from a wide range of soil environments (Barberán *et al*., 2012; Bissett *et al*., 2014; de Menezes *et al*., 2014; Hartmann *et al*., 2014; Lupatini *et al*., 2014). Unlike reductionist approaches, network analysis can yield a holistic view of biotic and abiotic interactions occurring in an environment (Bissett *et al*., 2013). By finding non-random co-occurrence in complex microbial communities, networks can also identify keystone taxa i.e. microorganisms that have the largest influence on community structure and possible function regardless of their population size (Lupatini *et al*., 2014; Vick-Majors *et al*., 2014).

In this study, we used 454 pyrosequencing of bacterial 16S ribosomal RNA (rRNA) genes and network analysis to assess drivers of soil bacterial communities in agroforestry systems along a 270 km north–south soil-climate gradient in Alberta, Canada. We focused on both broad microbial patterns and explicit microbial interactions at a finer scale. Specifically, we examined: (i) whether agroforestry systems promoted bacterial abundance and diversity across soil zones in Alberta, (ii) if the factors modulating soil bacterial communities were consistent across agroforestry systems, (iii) whether microbial co-occurrence was common in these systems and (iv) which microbial members were the keystone taxa in agroforestry systems.

Results

Bacterial abundance and soil properties

Bacterial abundance and most soil properties varied significantly between treed plots and herblands (Table 1). Soil bacterial abundance as measured by 16S rRNA gene copy number exceeded 10^6 copies per gram of dry soil in all soils, with significantly ($df = 5$; $P < 0.05$) higher abundance in hedgerows and woodlands compared with the neighbouring herblands. Although these soils were mildly acidic and the overall pH did not vary significantly, considerable variation (4.27 in a southern agricultural plot to 7.29 in a northern grassland plot) was observed among the samples and across the three soil zones (Table 1; Table S1). Soil moisture and NH_4^+ contents were higher ($df = 5$; $P < 0.05$) in hedgerows and woodlands than

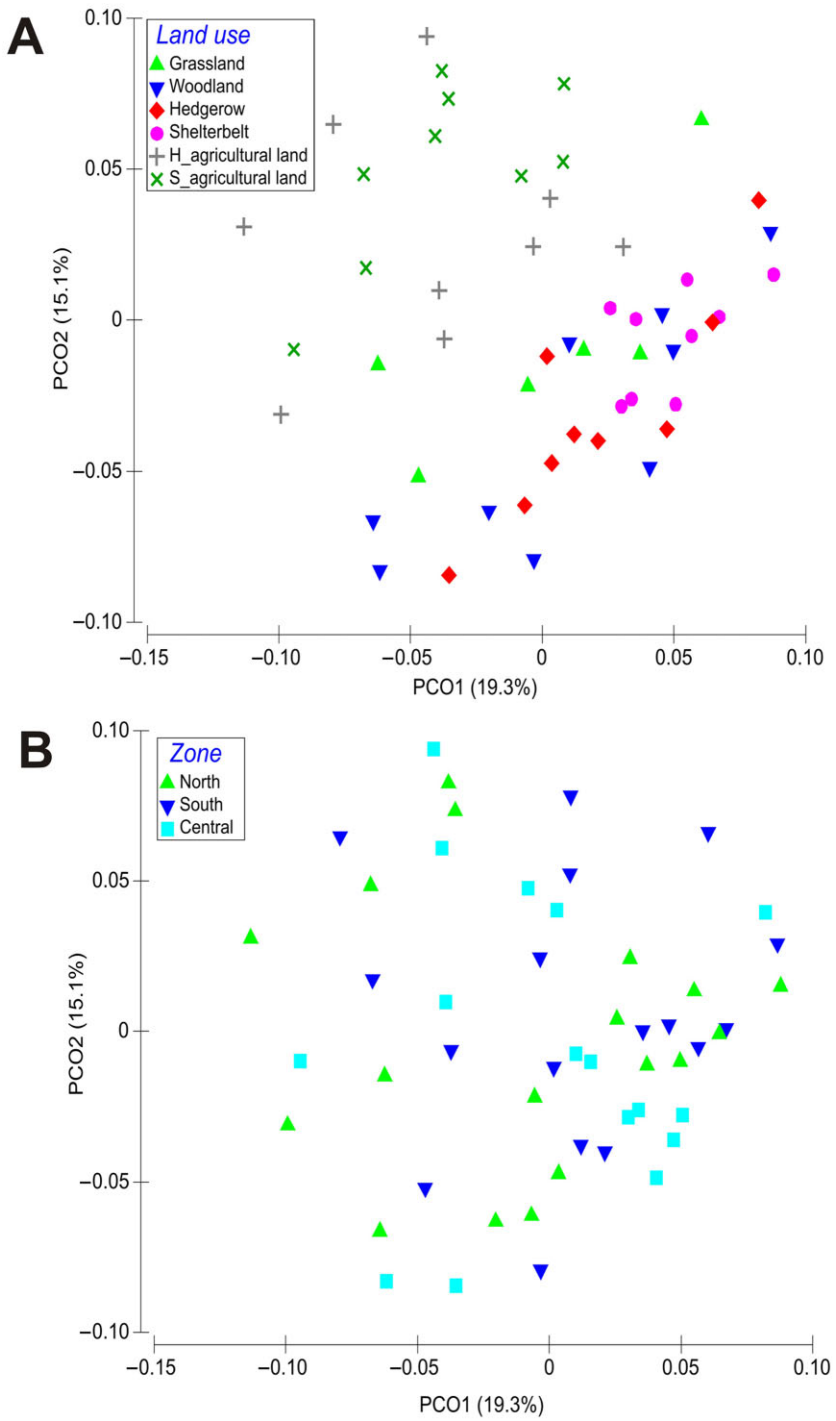

Fig. 1. Principal coordinate analysis (PCoA) plots of weighted UniFrac distances showing patterns of beta diversity in bacterial communities across land-use types (A) and soil-climate zones (B). Points situated closer together on the plot represent similar communities. Three agroforestry systems consisted of plots with trees (shelterbelt or hedgerow or woodland) and adjacent herbland plots (agricultural land or grassland). S_Agricultural land and H_Agricultural land represent agricultural soils adjacent to shelterbelts and hedgerows respectively.

adjacent plots. Different measures of soil carbon (C) content were consistently greater in plots with trees, with dissolved organic carbon (DOC) and total carbon (TC) contents almost two- to threefold higher than herblands. Only a few properties were significantly different when the three soil zones (north, central and south) were compared separately (Table S1).

Community structure and diversity

Principal coordinate analysis with weighted UniFrac distance matrix indicated that bacterial communities were somewhat more distinct across land-use types than across soil-climate zones (Fig. 1). While agricultural plots were similar to each other, they were quite different from

Table 2. Alpha diversity indices of bacterial communities at 0–10 cm depth in 51 soil samples obtained from three types of agroforestry systems across a 270 km soil-climate gradient in Alberta, Canada.

	Shelterbelt		Hedgerow		Silvopasture	
Diversity indices	*Agricultural*	*Shelterbelt*	*Agricultural*	*Hedgerow*	*Grassland*	*Woodland*
SOBS[#$]	592 (18.0)[a]	582 (12.8)[a]	568 (25.1)[a]	570 (19.1)[a]	570 (34.0)[a]	593 (18.0)[a]
ACE	1203 (128)[a]	1230 (120)[a]	1178 (144)[a]	1227 (154)[a]	1022 (198)[a]	1324 (218)[a]
Chao1	976 (59.2)[a]	960 (42.8)[a]	940 (66.4)[a]	960 (69.4)[a]	888 (101)[a]	1021 (96.1)[a]
Shannon	5.89 (0.05)[b]	5.74 (0.06)[ab]	5.76 (0.06)[ab]	5.68 (0.05)[a]	5.80 (0.06)[ab]	5.73 (0.05)[ab]
invSimspon	227 (17.6)[c]	145 (17.2)[ab]	178 (22.4)[bc]	111 (14.1)[a]	179 (15.8)[bc]	124 (20.7)[ab]

[#]Sequences were clustered at 0.03 difference.
[$]Within each row different letters show statistical significance at $P < 0.05$.
Each agroforestry system comprised a forested plot (shelterbelt or hedgerow or woodland) and an adjacent herbland plot (agricultural land or grassland).

plots with trees i.e. shelterbelts, hedgerows and woodlands. Bacterial communities in grasslands were located between that in agricultural and plots with trees (Fig. 1A). On the other hand, bacterial communities across soil zones were intermixed and not different from each other (Fig. 1B). The effect of land use was more evident in canonical analysis of principal coordinate plots (Fig. S1). The canonical analysis of principal coordinates (CAP) revealed that bacterial communities formed distinct clusters separating (Fig. S1A) herblands from plots with trees and the three soil climatic zones (Fig. S1B). CAP yielded two canonical axes with squared canonical correlations of $\delta_1^2 = 0.950$ and $\delta_1^2 = 0.821$ ($P = 0.0001$) for land uses and $\delta_1^2 = 0.912$ and $\delta_1^2 = 0.770$ ($P = 0.002$) for soil zones. These findings were reinforced by permutational multivariate analysis of variance (PERMANOVA) results that showed that the effect of land use was significant (df = 5; $P < 0.001$) compared with the effect of soil zone (df = 2; $P < 0.112$). There was no significant (df = 10; $P > 0.05$) interactive effect of agriculture types and soil zones. Similarly, analysis of similarity (ANOSIM) showed Global R of 0.482 ($P < 0.001$) and 0.054 ($P = 0.24$) for effects of land uses and soil zones respectively, highlighting the significance of land-use types, but not soil zones. Although not statistically significant, plots with trees showed greater species richness with ACE and Chao1 displaying their highest values in woodland soils (Table 2). Diversity indices (Shannon and InvSimpson) showed significant ($P < 0.05$) differences across land uses. Hedgerow and woodland soils had the lowest Shannon and InvSimpson values, whereas agricultural soils showed significantly higher bacterial diversity.

Relative abundance of major phyla and genera

The most abundant phyla across the three agroforestry systems were *Proteobacteria* (37–40%), *Acidobacteria* (14–19%) and *Actinobacteria* (14–26%) (Fig. 2A). Among the classes of *Proteobacteria*, *Alphaproteobacteria* (20–29%) were three to five times more abundant than other classes. Agricultural and grassland soils had lower abundance of *Actinobacteria*, but higher *Gammaproteobacteria* and *Bacteriodetes*. No distinct influence of soil zones was found on major bacterial phyla (data not shown). The most abundant genera across all agroforestry systems were *Gemmatimonas*, *Pseudomonas* and *Arthrobacter* (Fig. 2B). Plots with trees had significantly (df = 5; $P < 0.05$) higher abundance of *Solirubrobacter*, *Nocardioides*, *Streptomyces* and *Verrucomicrobia*, whereas agricultural soils showed significantly (df = 5; $P < 0.05$) higher abundance of *Flavobacterium*, *Nitrosospira* and *Gemmatimonas*. *Nitrospira* was also abundant in agricultural and woodland soils. On the other hand, grassland soils had high abundance of *Bradyrhizobium* and a twofold higher abundance of *Arthrobacter* than all other systems. Abundance of *Pseudomonas* was not significantly (df = 5; $P > 0.05$) different among land uses and soil zones, with the exception of two southern plots where it was 50–70 times higher in a shelterbelt and an agricultural soil. SIMPER analysis showed high similarity (64–75%) in bacterial community composition across land uses (Table S2). The top five most influential operational taxonomic units (OTUs) that contributed 6–10% of overall similarity included *Bradyrhizobiaceae*, *Pseudomonas*, *Acidobacteria_GP4*, *Propionibacteriaceae*, *Gemmatimonas* and *Arthrobacter*.

Drivers of bacterial communities in agroforestry systems

Environmental variables explained 58.5% of variance (Rho = 0.585) in bacterial communities according to BEST analysis. Indeed, soil properties displayed strong correlations with bacterial 16S rRNA gene copy number and the correlations were more consistent in the plots with trees than neighbouring herbland plots (Table 3). For example, bacterial abundance was significantly correlated to a few soil properties such as moisture and carbon contents in agricultural soils, whereas hedgerows and woodlands showed strong ($P < 0.01$) correlations with most soil attributes. Richness estimators, but not diversity indices, were correlated to edaphic factors in general

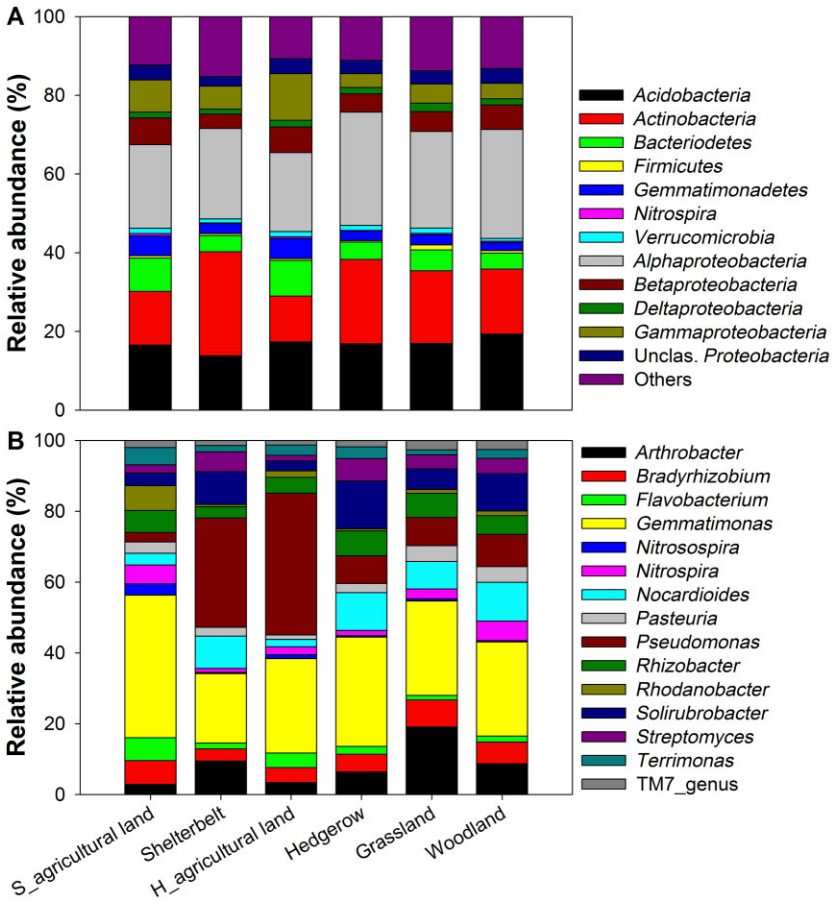

Fig. 2. Stacked bar chart showing relative abundance of the (A) major bacterial phyla and dominant classes of *Proteobacteria* and (B) 15 most abundant bacterial genera in three agroforestry systems consisting of plots with trees (shelterbelt or hedgerow or woodland) and adjacent herbland plots (agricultural land or grassland). S_Agricultural land and H_Agricultural land represent agricultural soils adjacent to shelterbelts and hedgerows respectively.

(Table S3). Soil water holding capacity (WHC), total nitrogen (TN), total organic carbon (TOC) and dissolved organic nitrogen (DON) exhibited significant correlations with number of observed species (SOBS) and Chao1 while ACE was only correlated to WHC. Major bacterial phyla were significantly correlated to most studied soil attributes, with the strongest correlations with pH, WHC, NH_4, TOC and TN contents (Table S4). For example, *Burkholderia* and *Microvirga* had strong and consistent correlations with moisture, NH_4^+ and soil C contents, and

Table 3. Pearson correlations between bacterial 16S rRNA abundance and soil properties at 0–10 cm depth in 54 soil samples obtained from three types of agroforestry systems across a 270 km soil-climate gradient in Alberta, Canada.

	Shelterbelt		Hedgerow		Silvopasture	
Soil properties	*Agricultural*	*Shelterbelt*	*Agricultural*	*Hedgerow*	*Grassland*	*Woodland*
pH	0.243#	–0.347	0.573	0.891**	0.354	0.658*
WHC (%)	0.957**	0.661*	0.705*	0.745*	0.599	0.979**
Moisture (%)	0.774*	0.757*	0.708*	0.931**	0.758*	0.966**
NH_4-N	0.580	0.596	0.742*	0.866**	0.730*	0.861**
NO_3-N	–0.136	0.677*	0.404	0.899**	0.808**	0.450
TN	0.897**	0.821**	0.524	0.843**	0.678*	0.980**
TOC	0.840**	0.807**	0.424	0.828**	0.762*	0.984**
DOC	0.046	0.590	0.713*	0.692*	0.690*	0.915**
DON	–0.001	0.688*	0.545	0.855**	0.589	0.801**
TC	0.046	0.590	0.713*	0.692*	0.690*	0.915**

#Each value is the correlation of nine replicates.
*Significant correlation at $P < 0.05$
**Significant correlation $P < 0.01$.
Each agroforestry system consisted of a forested plot (shelterbelt or hedgerow or woodland) and an adjacent herbland plot (agricultural land or grassland).

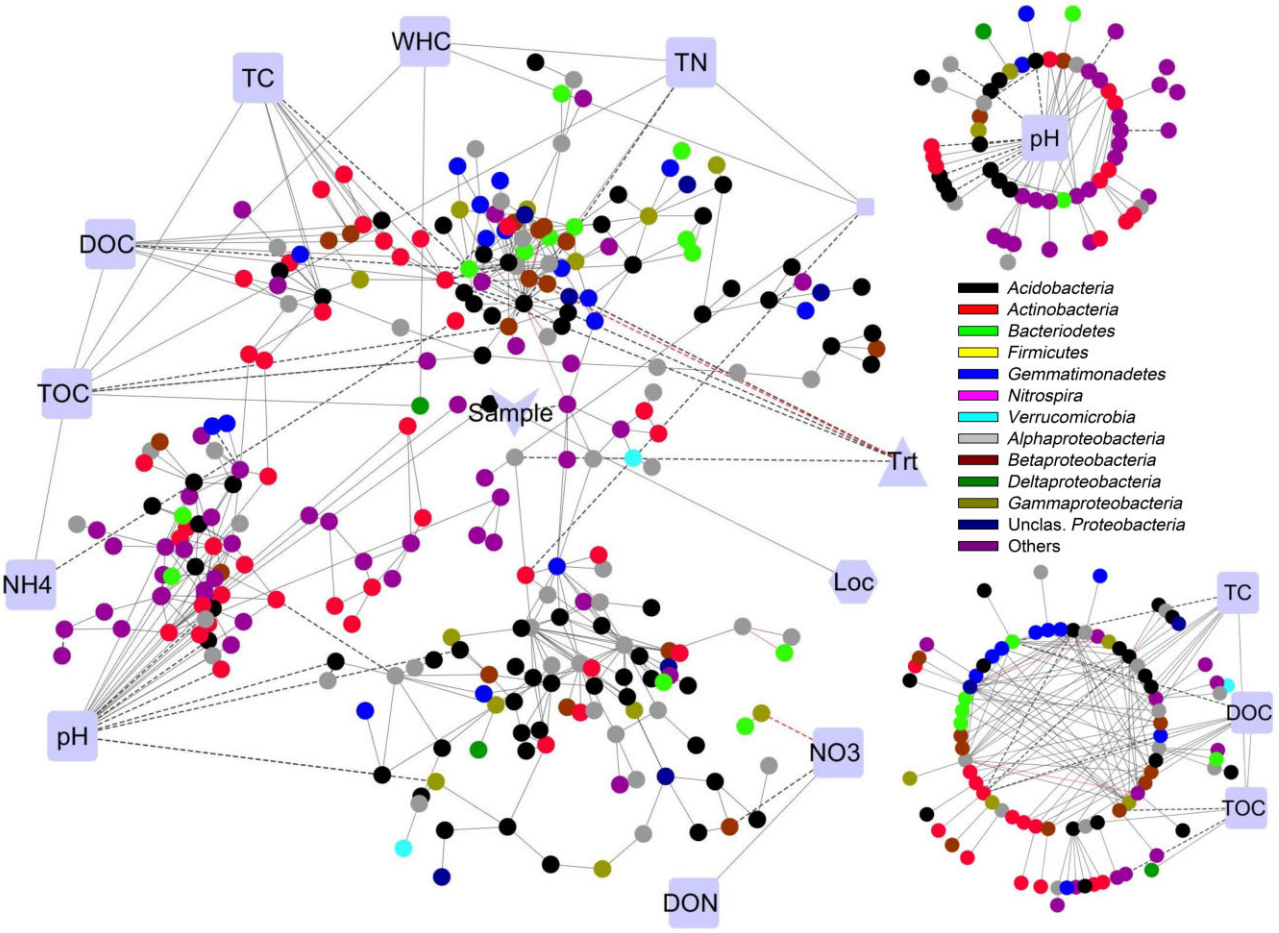

Fig. 3. Network analysis showing connectedness between bacterial communities and edaphic factors. Blue solid lines represent significantly strong (>0.6) positive linear relationship, black dashed lines represent strong (< –0.6) negative linear relationship whereas significant non-linear relationships are represented by red lines. Variables in blue boxes show various soil attributes and dummy variables of treatments (Trt) and locations (Loc). Soil pH and carbon contents showed maximum connectedness and thus their associations are also shown separately.

Nocardioides and *Solirubrobacter* were significantly ($P < 0.01$) correlated to soil pH, DOC and TC contents (Table S5). *Arthrobacter* and *Bradyrhizobium* were correlated to soil pH, while *Nitrospira* was correlated to nitrate and DON contents.

Networks and connectedness

The network comprised 336 nodes and 500 edges, with an average number of neighbours of 2.96 and characteristic path length of 6.55 (Fig. 3). The network had an overall diameter of 19 edges and an average clustering coefficient of 0.211. Soil pH and soil C content were associated with the maximum number of nodes. While TOC, DOC and TC were mostly associated with nodes belonging to *Acidobacteria*, *Actinobacteria* and *Betaproteobacteria*, pH was associated with a wide range of nodes. Based on betweenness centrality scores (data not shown), the OTUs identified as keystone species belonged to *Arthrobacter*, *Acidobacteria_Gp16*, *Burkholderia*, *Rhodanobacter* and an unclassified member of the *Rhizobiales*. MCODE analysis returned eight significant clusters with network scores ranging between 4.9 and 2.6 (Fig. S4). Treatment, DOC and TC produced the rank 1 cluster, which had 48 nodes and 115 edges. The rank 2 cluster consisted of only bacterial interactions while pH constituted the rank 3 cluster with 17 nodes and 28 edges.

Discussion

Land use but not soil zone affects bacterial communities

Microbial communities in agroforestry systems are largely unexplored, with a number of previous studies reporting contrasting results. Our results show that land-use patterns exert stronger effects on soil bacterial communities than soil zones across the selected geographical gradient

in Alberta. The distinction between biotic communities was more pronounced in different land-use types than in three soil zones, which could be ascribed to the differences in edaphic factors i.e. most soil properties were significantly different in the six land-use types. This is consistent with previous studies showing that land-use alterations exerted long-lasting effects on soil physical and chemical properties, which in turn affect microbial abundance and community composition (Lauber *et al*., 2008; Bissett *et al*., 2014). Typically, tree heterogeneity increases from shelterbelts to woodlands, potentially resulting in more abundant soil resources (e.g. C, N and moisture), which might have contributed to differences in 16S rRNA gene copy number and community structure. Indeed, gradient in clustering of bacterial communities was observed from shelterbelts to woodlands. On the other hand, despite belonging to three soil zones, biotic communities were similar in agricultural soils with all samples showing strong clustering and being distinct from other land-use types. This reinforces the notion of land-use impact on soil biotic communities and contrasts with Girvan and colleagues (2003) who found that soil type was the primary driver of microbial communities in arable soils. Nonetheless, it should be noted that although three different soil types were common in our geographical gradient, each zone was dominated by at least two soil types (Soil Classification Working Group, 1998). At the same time, climatic conditions such as number of frost-free days, annual precipitation and mean annual air temperature did not differ substantially across the zones.

Agroforestry promotes bacterial abundance, but not diversity

The present study shows that plots with trees support greater bacterial abundance than adjacent herbland plots, with all three land-use types with trees having significantly higher 16S rRNA gene copies. The higher soil bacterial abundance can be attributed to nutrient-rich soil resources in plots with trees. For example, consistent with our previous study (Baah-Acheamfour *et al*., 2014), we found that soils in plots with trees had two to three times higher soil organic carbon, moisture, DOC and NH_4-N than adjacent agricultural soils. Higher microbial abundance with increasing number of trees was also observed in previous studies (Zak *et al*., 2003). Thus, our results contradict the negative impact of agroforestry on microbial communities observed by Saggar and colleagues (2001) and demonstrate that agroforestry promotes microbial abundance. Although not statistically significant, ACE and Chao1 estimators were higher in soils with trees, suggesting that these soils offer edaphic conditions that harbour less-abundant or rare species (Hughes *et al*., 2001). Nonetheless, plots with trees had significantly lower diversity compared with herblands, consistent with a recent study that shows that plant diversity does not necessarily promote microbial alpha diversity (Prober *et al*., 2015). Overall, our results demonstrate that plots with trees in agroforestry systems harbour abundant and species-rich bacterial communities in this region of Canada.

Dominant bacterial groups and keystone taxa

Acidobacteria, *Actinobacteria* and *Alphaproteobacteria* were the dominant bacterial groups in agroforestry systems of Alberta. These groups are abundant in most ecosystems and their dominance was also observed in previous studies (Fierer *et al*., 2007; Lauber *et al*., 2009). While this study reports the general characteristics of bacterial communities in western Canadian agri-systems, it also reveals some key patterns. For example, agricultural soils showed higher abundance of copiotrophic (i.e. preferring labile soil nutrients) *Bacteriodetes*, whereas soil organic matter (SOM)-degrading *Actinobacteria* were more common in soils in treed plots with high SOM content (Fierer *et al*., 2007). The consistent abundance of recalcitrant SOM-degrading *Verrucomicrobia* found in this study is congruent with a recent study (Fierer *et al*., 2013), which shows the abundance of this group in North American prairie soils. Similar nutritional patterns were also evident when bacterial genera were considered: cellobiose-degrading *Nocardioides* and *Solirubrobacter* were abundant in SOM-rich soils with trees while *Nitrosospira* and *Nitrospira* were more prevalent in mineral N rich agriculture and woodland soils. On the other hand, grasslands with greater abundance of leguminous plants and often high phosphorus content showed a greater number of symbiotic nitrogen fixing *Bradyrhizobium* and phosphate solubilizing *Arthrobacter*. SIMPER analysis showed *Bradyrhizobiaceae*, *Pseudomonas*, *Acidobacteria_GP4*, *Propionibacteriaceae*, *Gemmatimonas* and *Arthrobacter* were the most influential OTUs in agroforestry systems, and this is consistent with their betweenness centrality scores in network analysis. Indeed *Arthrobacter*, *Acidobacter* Gp16, *Burkholderia*, *Rhodanobacter* and an unclassified member of *Rhizobiales* emerged as keystone taxa in the agroforestry systems we investigated. Thus, the present study highlights the usefulness of betweenness centrality in disentangling the most influential taxa (Martin González *et al*., 2010; Lupatini *et al*., 2014). This score discerns the modules that are most important in maintaining connectivity in an ecological network, and thus can be regarded as keystone species in that system (Vick-Majors *et al*., 2014). The fact that the keystone taxa identified through betweenness centrality were also identified by SIMPER analysis reinforces the importance of these taxa in maintaining soil ecological processes in agroforestry systems.

Soil pH and carbon content as drivers of bacterial communities

This study aimed to identify the factors that modulate bacterial spatial distribution in agroforestry systems and whether these communities are determined by broader (consistent across systems) or local (restricted to a particular soil or land-use type) edaphic factors. Our results demonstrate consistent correlation between soil resources and microbial community characteristics (e.g. overall abundance, species richness and important genera). The fact that correlations among biotic communities and soil resources are more consistent in plots with trees than adjacent herbland plots reveals that soil bacterial communities with trees operate in a somewhat predictable manner. Our results also show broader control on biotic communities across agroforestry systems. For example, soil pH, TOC, TN, moisture contents were consistently associated with bacterial abundance with the latter two also correlated with species richness. The importance of these factors in shaping bacterial community structure was also reported previously (Lauber *et al*., 2009; Banerjee and Siciliano, 2012). Interestingly, bacterial diversity varied significantly across six land-use types; however, richness was significantly correlated to soil attributes, but not diversity. This lack of correlation is not uncommon and our results accord with a recent study showing species richness, but not diversity, is correlated with edaphic factors (Fierer *et al*., 2013). It should also be noted that ACE and Chao1 estimators were different across land-use types. For example, woodlands had higher species richness (ACE = 1324; Chao1 = 1021) than grassland (ACE = 1022; Chao1 = 888) or hedgerow-associated agricultural soils (ACE = 1178; Chao1 = 940). At the same time our results also indicate that less abundant species (as indicated by ACE and Chao1; Hughes *et al*., 2001) might have played a key role in these soils. However, the fact that Shannon's and InvSimpson consider evenness and are sensitive to changes in rare species abundance might also have contributed to lack of correlation (Tuomisto, 2012).

The edaphic determinants of bacterial communities became clearer when the most abundant bacterial phyla and genera were considered. Soil pH and carbon contents exhibited consistent correlations with most bacterial groups. SOM-degrading *Solirubrobacter* was significantly correlated with TOC and TC while *Acidobacteria* was associated with soil pH. Thus, our results agree with Jones and colleagues (2009) and Naether and colleagues (2012) who showed soil pH as a major driver shaping *Acidobacterial* populations. While various soil C contents together constituted important clusters, pH alone determined third and sixth ranked clusters in the network and was associated with a wide range of nodes belonging to various major phyla. It should be noted that although the difference was not statistically significant, soil pH did vary among the samples and across land uses in three soil zones. For example, shelterbelts in the north zone (6.51) was higher than the central (5.69) or south (5.79), whereas southern woodlands had a much higher (6.57) pH than central (5.23) woodlands. In the north zone, pH in shelterbelt-associated agricultural soils (6.12) was higher than hedgerow-associated agricultural soils (5.18). We speculate that differences in soil pH between samples were effective in influencing bacterial community composition. Thus, consistent with previous reports (Lauber *et al*., 2008; 2009), our results demonstrate the importance of soil pH as a major edaphic determinant shaping soil bacterial communities. It should be noted that low-abundance microbial communities, i.e. rare biosphere, also play various ecological functions in the environment (Sogin *et al*., 2006; Lynch and Neufeld, 2015). Alongside the edaphic factors, the rare biosphere of soil could also have contributed to the observed differences.

Connectedness and modular structure of bacterial communities

This is the first detailed study of soil microbial communities in Canadian agroforestry systems and our results demonstrate non-random co-occurrence and connectedness of bacterial communities across such systems. This highlights the role of deterministic ecological processes and possible historical effects in shaping these communities. Our goal was to explore the overall co-occurrence patterns to gain a holistic view of microbial communities. We found a modular structure of bacterial communities in agroforestry systems, indicating the importance of studying microbial modules in soil ecosystems to identify microbial associations across sample sets and to improve understanding of their relationships with abiotic factors that are not clear when communities are studied in arbitrarily compartmentalized groups (Bissett *et al*., 2013; de Menezes *et al*., 2014). These modules do not necessarily follow taxonomic classification i.e. major bacterial phyla interact with each other independent of their taxonomy. This supports the notion that bacterial community assembly is determined by environmentally driven functional characteristics and not phylogeny (Burke *et al*., 2011). Non-random associations are determined by microbial physiological requirement and thus are more likely to be driven by environmental factors. Indeed, soil pH and C content exhibited maximum number of associations and emerged as the determinants of soil bacterial communities in the agroforestry systems we studied.

Conclusions

Understanding the patterns and determinants of microbial communities is a central goal of microbial ecology. Despite the importance and benefits of agroforestry systems, a fundamental component of these systems remains to be explored. Our results show that plots with trees in agroforestry systems promote soil bacterial abundance and species richness, and there is a degree of predictability in the control of microbial communities in agroforestry systems. Land-use type exerts strong effects on bacterial abundance and community composition, and this impact mediates through change in edaphic factors. *Acidobacteria*, *Actinobacteria* and *Alphaproteobacteria* are the dominant bacterial groups, while *Arthrobacter*, *Acidobacter* Gp16, *Burkholderia*, *Rhodanobacter* and *Rhizobium* are the keystone taxa in agri-systems of Alberta. Soil pH and carbon contents are the principal drivers of bacterial community characteristics. Finally, we demonstrate non-random co-occurrence and environmentally driven modular patterns of microbial communities. While this is the first study to examine soil microbial communities in agroforestry systems, it also unravels some fundamental microbial ecological patterns that are important in all soil systems.

Experimental procedures

Study area

The study area was established along a 270 km north–south soil-climate gradient (54° 35′ N to 52° 28′ N) across 10 counties in central Alberta, Canada (Fig. S1). The gradient also encompassed a 226 km east–west area. The northern part of the study area receives 115–125 frost-free days per annum, while the southern area receives 125–145 frost-free days (Environment Canada). The annual precipitation varies from 463 mm in the north to 448 mm in the south, whereas the mean annual air temperatures in the north and south are 1.9 and 2.4°C respectively. The sampling area has three embedded soil zones: north zone with Gray Luvisol and Dark Gray Chernozem, central zone with Black Chernozem and Dark Gray Chernozem, and south zone with Black Chernozem and Gray Luvisol (Soil Classification Working Group, 1998). Three common agroforestry systems were selected for the study: shelterbelts, hedgerows and silvopastures. In each of these systems, two plots were established for this study, plots with trees (woodland or hedgerow or shelterbelt) and without trees (herbland plots i.e. agricultural land or grassland). Shelterbelts and hedgerows were strips (3–5 m wide) consisting of trees and shrubs planted or naturally grown at the edges of agricultural fields (S_agricultural land and H_agricultural land, respectively). Shelterbelts consisted of one or two rows of trees comprising 20- to 50-year-old coniferous and deciduous trees dominated by *Picea glauca* Moench. Hedgerows were 40- to 100-year-old deciduous stands dominated by some mix of *Populus tremuloides* Michx, *Betula papyrifera* Marsh. and *Populus balsamifera* L. In both systems, the adjacent agricultural plots were used for production of annual crops such as *Hordeum vulgare* L., *Triticum aestivum* L. or *Brassica napus* L. in rotation. Silvopastures consisted of two plots: a mosaic of variably aged *P. tremuloides* woodland and a grassland. Both these plots are subjected to intermittent grazing. Woodlands were similar in species composition to hedgerows. In general, the number and types of trees were greater in woodlands compared with shelterbelts and hedgerows.

Sampling design

A total of 27 paired sites that included three replicates of each agroforestry system within each of the three soil zones, were sampled in this study. A split-plot design was used to sample each site, with one plot established in the treed areas and another in the adjacent herbland (Fig. S2). A 30 m transect was laid out in each plot 10 m (generally one tree-height) from the forest edge to alleviate the influence of trees. Soil samples were collected between September and October 2012 after crop harvest and after ranchers had removed their livestock. To account for spatial variability, 10 soil samples (0–10 cm soil depth, 3 cm diameter) were collected at 3 m intervals along each transect and pooled to obtain a composite sample. Thus, a total of 54 samples were obtained from 27 paired sites. The core sampler was cleaned with 70% ethyl alcohol between plots. After removal, samples were kept on ice and transported to the lab where they were subsampled for various physical, chemical and molecular analyses.

Soil physical and chemical characterization

Gravimetric soil moisture content was determined by weighing and oven drying 20 g of fresh soil at 105°C to constant weight. Soil pH was measured in a 1:2 soil solution (0.01 M $CaCl_2$) with a digital pH meter at 20°C (Omega Eng. Inc., Stamford, CT, USA). Plant available NH_4-N and NO_3-N were determined using an autoanalyser (SmartChem Discrete Wet Chemistry Analyzer, Westco Scientific Limited, Brookfield, CT, USA) after extraction using 2 M KCl solution at a ratio of 1:5 (soil:KCl). To estimate soil WHC, a known mass of oven-dried soil (105°C for 48 h) was placed in a porous funnel and saturated with water. The sample was then placed in a humid enclosure and allowed to drain for 24 h and reweighed. DOC and dissolved total nitrogen contents were determined with fresh soil extracts (same as NH_4^+ and NO_3^- content analysis) using a TOC-VCPN analyser (Shimadzu Scientific Instruments, Columbia, MD, USA). DON was measured by subtracting the mineral nitrogen content (sum of exchangeable NH_4^+ and NO_3^-) from dissolved total nitrogen content. TC or TOC and TN were determined using an ECS 4010 Elemental Analyzer System (Costech International Strumatzione, Florence, Italy, 2003).

Deoxyribonucleic acid (DNA) extraction and quantitative polymerase chain reaction (PCR)

DNA was extracted from 0.25 g soil using the Power soil DNA isolation kit (MoBio, Carlsbad, CA, USA) according to the manufacturer's instructions. Concentrations of purified DNA were determined by nanodrop spectrophotometer

(NanoDrop, Wilmington, DE, USA). The number of bacterial 16S rRNA gene copies present in the purified DNA was determined by quantitative real-time PCR (qPCR) using the Qiagen QuantiTect™ SYBR® Green PCR Master Mix (Qiagen Inc., Ontario, Canada) and an ABI Step-One real-time PCR machine (Applied Biosystems, Foster City, CA, USA). A 174 bp fragment of bacterial 16S rRNA gene was amplified using 341F (CCT ACG GGA GGC AGC AG) and 515R (ATT CCG CGG CTG GCA) primers as described by López-Gutiérrez and colleagues (2004). Details of thermal cycling conditions and quality assessment are given in the Supporting information.

454 pyrosequencing of bacterial 16S rRNA

Bacterial tag-encoded FLX *amplicon pyrosequencing* was performed on 54 samples at Mr DNA Lab (http://www.mrdnalab.com, MR DNA, Shallowater, TX, USA) using primers 28F and 519R as described in Dowd and colleagues (2008). Pyrosequencing flowgrams were converted to sequence reads and analysed using the MOTHUR software package (Schloss *et al*., 2009). Sequence reads were depleted of barcodes and primers. Sequences less than 200 bp, sequences with ambiguous base calls, and homopolymer runs exceeding 6 bp were then removed. Sequences were also deionized using *shhh.flows* command, the MOTHUR implementation of the PyroNoise. Chimeras were detected and removed using *chimera.uchime*. Reads were aligned to the SILVA reference database v102 (Pruesse *et al*., 2007) with the mother NAST aligner, and unmatched sequences were removed. Singletons were removed, OTUs defined by clustering at 97% similarity, and classified according to SILVA v102 using the Naïve Bayesian classifier as implemented in MOTHUR (Wang *et al*., 2007). Three of our 54 soil DNA samples failed to yield sequences: two grasslands in the central zone and one grassland in the south. From the remaining 51 samples, 646 981 reads and 9463 OTUs were obtained, and each group contained between 1560 and 25 440 reads. Subsampling was performed at the 1560 read level for each sample before estimating rarefaction curves and alpha diversity indices.

Statistical analyses

Bacterial 16S rRNA abundance and environmental data were transformed before performing univariate Pearson's correlation, analysis of variance and Duncan's post hoc tests in IBM® SPSS® Statistics version 20 (IBM, Armonk, NY, USA). Euclidian resemblance matrix of environmental data and Bray–Curtis similarity matrix of bacterial OTU abundance were generated using PRIMER-E v6 (PRIMER-E, Plymouth, UK). We performed BEST analysis (Clarke, 1993) to examine the correlation between Bray–Curtis similarity matrix and the Euclidian resemblance matrix. ANOSIM and PERMANOVA (Clarke, 1993; Anderson, 2001) were also carried out on the Bray–Curtis matrix with 10^5 permutation tests to study the variation in bacterial distribution due to six land uses and three soil-climate zones. SIMPER analysis (Clarke, 1993) was conducted to explore the similarity/dissimilarity between bacterial communities in land uses and to disentangle the most important OTUs that were responsible for the observed differences. Canonical analysis of principal coordinates (Anderson and Willis, 2003) was performed with 10^5 permutations to examine bacterial community clustering in different land uses and soil-climate regions. Bacterial beta diversity patterns across land-use types and soil-climate zones were assessed using weighted UniFrac distance matrix (Lozupone and Knight, 2005).

Network analysis

To examine co-occurrence networks in bacterial communities, network analysis was conducted on soil properties and pyrosequencing data using the maximal information coefficient in MINE software (Reshef *et al*., 2011). Relationships between edaphic factors and biotic variables that were significant at a false discovery rate (FDR) (Benjamini and Yekutieli, 2001) of 10% were then visualized in CYTOSCAPE version 3.2.0 (Shannon *et al*., 2003). To minimize pairwise comparisons and manage the FDR, network analysis was performed on the top 500 most important OTUs identified through SIMPER, which resulted in 1300 OTUs in total. From the pairwise comparisons, top 500 interactions were selected, which generated 336 nodes in the network. To assess non-random pattern in the resultant network, we compared our network against its randomized version using *Randomnetworks* plugin in CYTOSCAPE v2.6. Structural attributes of this network such as clustering coefficient, diameter, mean shortest path, number of shortest paths and connected components etc. were significantly different than that of the random network with equal nodes and edges. *NetworkAnalyzer* tool was used to calculate network topology parameters. Modular structure and groups of highly interconnected nodes were analysed using the MCODE application with standard parameters [degree cut-off of 2, node score cut-off of 0.2, K-core of 2 and maximum depth from seed of 100 (Bader and Hogue, 2002)]. OTUs with highest betweenness centrality scores were considered as keystone species (Martin González *et al*., 2010; Vick-Majors *et al*., 2014).

Acknowledgements

The authors would like to acknowledge financial support from Agriculture and Agri-Food Canada (AAFC), funded under the Agricultural Greenhouse Gases Program (AGGP). We thank Qiting Chen, Aggy Hery and Isobel Fodor for participating in field data collection. Special thanks also go to the landowners in Alberta who allowed us access to their farmlands. We thank the editor and two anonymous reviewers for their constructive comments and insightful suggestions.

References

Agriculture and Agri-Food Canada (2014) The advantages of agroforestry. [WWW document]. URL http://www4.agr.gc.ca/resources/prod/doc/terr/pdf/agroforest_qc_final_e.pdf.

Anderson, M. (2001) A new method for non-parametric multivariate analysis of variance. *Aust Ecol* **26:** 32–46.

Anderson, M.J., and Willis, T. (2003) Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* **84:** 511–525.

Baah-Acheamfour, M., Carlyle, C.N., Bork, E.W., and Chang, S.X. (2014) Trees increase soil carbon and its stability in three agroforestry systems in central Alberta, Canada. *For Ecol Manage* **328:** 131–139.

Bader, G.D., and Hogue, C.W.V. (2002) Analyzing yeast protein–protein interaction data obtained from different sources. *Nat Biotechnol* **20:** 991–997.

Banerjee, S., and Siciliano, S.D. (2012) Spatially tripartite interactions of denitrifiers in arctic ecosystems: activities, functional groups and soil resources. *Environ Microbiol* **14:** 2601–2613.

Barberán, A., Bates, S.T., Casamayor, E.O., and Fierer, N. (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J* **6:** 343–351.

Bardgett, R.D., and van der Putten, W.H. (2014) Belowground biodiversity and ecosystem functioning. *Nature* **515:** 505–511.

Benjamini, Y., and Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* **29:** 1165–1188.

Bissett, A., Brown, M.V., Siciliano, S.D., and Thrall, P.H. (2013) Microbial community responses to anthropogenically induced environmental change: towards a systems approach. *Ecol Lett* **16:** 128–139.

Bissett, A., Abell, G.C.J., Brown, M., Thrall, P.H., Bodrossy, L., Smith, M.C., *et al*. (2014) Land-use and management practices affect soil ammonia oxidiser community structure, activity and connectedness. *Soil Biol Biochem* **78:** 138–148.

Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011) Bacterial community assembly based on functional genes rather than species. *Proc Natl Acad Sci USA* **108:** 14288–14293.

Clarke, K.R. (1993) Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* **18:** 117–143.

De Menezes, A.B., Prendergast-Miller, M.T., Richardson, A.E., Toscas, P., Farrell, M., Macdonald, L.M., *et al*. (2014) Network analysis reveals that bacteria and fungi form modules that correlate independently with soil parameters. *Environ Microbiol* **61:** 1–37.

Dowd, S.E., *et al*. (2008) Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol* **8:** 125. doi:10.1186/1471-2180-8-125.

Fierer, N., Bradford, M.A., and Jackson, R.B. (2007) Toward an ecological classification of soil bacteria. *Ecology* **88:** 1354–1364.

Fierer, N., Ladau, J., Clemente, J.C., Leff, J.W., Owens, S.M., Pollard, K.S., *et al*. (2013) Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. *Science* **342:** 621–624.

Fuhrman, J.A. (2009) Microbial community structure and its functional implications. *Nature* **459:** 193–199.

Girvan, M.S., Bullimore, J., Pretty, J.N., Mark, A., Ball, A.S., and Osborn, A.M. (2003) Soil type is the primary determinant of the composition of the total and bacterial communities in arable soils. *Appl Environ Microbiol* **69:** 1800–1809.

Hartmann, M., Niklaus, P.A., Zimmermann, S., Schmutz, S., Kremer, J., Abarenkov, K., *et al*. (2014) Resistance and resilience of the forest soil microbiome to logging-associated compaction. *ISME J* **8:** 226–244.

Hughes, J.B., Hellmann, J.J., Ricketts, T.H., and Bohannan, B.J.M. (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* **67:** 4399–4406.

Jangid, K., Williams, M.A., Franzluebbers, A.J., Schmidt, T.M., Coleman, D.C., and Whitman, W.B. (2011) Land-use history has a stronger impact on soil microbial community composition than aboveground vegetation and soil properties. *Soil Biol Biochem* **43:** 2184–2193.

Jones, R.T., Robeson, M.S., Lauber, C.L., Hamady, M., Knight, R., and Fierer, N. (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J* **3:** 442–453.

Lacombe, S., Bradley, R.L., Hamel, C., and Beaulieu, C. (2009) Do tree-based intercropping systems increase the diversity and stability of soil microbial communities? *Agric Ecosyst Environ* **131:** 25–31.

Lauber, C.L., Hamady, M., Knight, R., and Fierer, N. (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* **75:** 5111–5120.

Lauber, C.L., Strickland, M.S., Bradford, M.A., and Fierer, N. (2008) The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biol Biochem* **40:** 2407–2415.

López-Gutiérrez, J.C., Henry, S., Hallet, S., Martin-Laurent, F., Catroux, G., and Philippot, L. (2004) Quantification of a novel group of nitrate-reducing bacteria in the environment by real-time PCR. *J Microbiol Methods* **57:** 399–407.

Lozupone, C., and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71:** 8228–8235.

Lupatini, M., Suleiman, A.K.A., Jacques, R.J.S., Antoniolli, Z.I., de Siqueira Ferreira, A., Kuramae, E.E., and Roesch, L.F.W. (2014) Network topology reveals high connectance levels and few key microbial genera within soils. *Front Environ Sci* **2:** 1–11.

Lynch, M.D.J., and Neufeld, J.D. (2015) Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* **13:** 217–229.

Martin González, A.M., Dalsgaard, B., and Olesen, J.M. (2010) Centrality measures and the importance of generalist species in pollination networks. *Ecol Complex* **7:** 36–43.

Naether, A., Foesel, B.U., Naegele, V., Wüst, P.K., Weinert, J., Bonkowski, M., *et al*. (2012) Environmental factors affect acidobacterial communities below the subgroup level in grassland and forest soils. *Appl Environ Microbiol* **78:** 7398–7406.

Prober, S.M., Leff, J.W., Bates, S.T., Borer, E.T., Firn, J., Harpole, W.S., *et al*. (2015) Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecol Lett* **18:** 85–95.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., *et al*. (2011) Detecting novel associations in large data sets. *Science* **334:** 1518–1524.

Rivest, D., Lorente, M., Olivier, A., and Messier, C. (2013) Soil biochemical properties and microbial resilience in agroforestry systems: effects on wheat growth under controlled drought and flooding conditions. *Sci Total Environ* **463–464:** 51–60.

Saggar, S., Hedley, C.B., and Salt, G.J. (2001) Soil microbial biomass, metabolic quotient, and carbon and nitrogen mineralisation in 25-year-old *Pinus radiata* agroforestry regimes. *Aust J Soil Res* **39:** 491–504.

Sala, O.E., Iii, F.S.C., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., *et al*. (2000) Global biodiversity scenarios for the year 2100. *Science* **287:** 1770–1775.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., *et al*. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13:** 2498–2504.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., *et al*. (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103:** 12115–12120.

Soil Classification Working Group (1998) *The Canadian System of Soil Classfication*. Agriculture and Agri-Food Canada Publication 1646. National Research Council of Canada, Ottawa, Canada: NRC Research Press.

Thevathasan, N.V., and Gordon, A.M. (2004) Ecology of tree intercropping systems in the North temperate region: experiences from southern Ontario, Canada. *Agrofor Syst* **61–62:** 257–268.

Tourna, M., Freitag, T.E., Nicol, G.W., and Prosser, J.I. (2008) Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environ Microbiol* **10:** 1357–1364.

Tuomisto, H. (2012) An updated consumer's guide to evenness and related indices. *Oikos* **121:** 1203–1218.

Van der Heijden, M.G.A., Bardgett, R.D., and van Straalen, N.M. (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol Lett* **11:** 296–310.

Vick-Majors, T.J., Priscu, J.C., and Amaral-Zettler, L.A. (2014) Modular community structure suggests metabolic plasticity during the transition to polar night in ice-covered Antarctic lakes. *ISME J* **8:** 778–789.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73:** 5261–5267.

Wardle, D.A., Bardgett, R.D., Klironomos, J.N., Setälä, H., van der Putten, W.H., and Wall, D.H. (2004) Ecological linkages between aboveground and belowground biota. *Science* **304:** 1629–1633.

Zak, D.R., Holmes, W.E., White, D.C., Peacock, A.D., Ecology, S., and Aug, N. (2003) Plant diversity, soil microbial communities, and ecosystem function: are there any links? *Ecology* **84:** 2042–2050.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Study area across 270 km north–south and 226 east–west geographical gradient in Alberta, Canada. The 270 km north–south gradient comprised three soil zones (north, central and south).

Fig. S2. Interactive diagram showing split-plot design of sampling and number of different agroforestry systems in three soil-climate zones in Alberta, Canada.

Fig. S3. Canonical analysis of principal coordinates (CAP) of Bray–Curtis similarity matrix with 10^5 permutations showing clustering of bacterial communities in (A) six land-use types and (B) three soil zones. Three agroforestry systems consisting of plots with trees (shelterbelt or hedgerow or woodland) and adjacent herbland plots (agricultural land and grassland). S_Agricultural land and H_Agricultural land represent agricultural lands adjacent to shelterbelts and hedgerows respectively. Each value is the average of three replicates.

Fig. S4. Results of MCODE analysis showing eight sub-networks with highest node scores. Networks and connectedness were visualized in PHYTOSCAPE 3.2.0 software. MCODE app was used to analyse networks with maximum scores.

Table S1. Soil and microbial properties measured at 0–10 cm depth in 54 samples obtained from three types of agroforestry systems across three soil-climate zones (north, central and south) in Alberta, Canada. Each agroforestry system consisted of a forested plot (shelterbelts or hedgerow or woodland) and an adjacent herbland plot (agricultural land or grassland). One-way ANOVA with Duncan's post hoc test was performed across three soil-climate zones within each land use ($n = 3$).

Table S2. Results of SIMPER analyses indicating the contribution of specific operational taxonomic units (OTUs) to observed similarities between land-use types.

Table S3. Overall Pearson correlations among soil properties and bacterial species richness estimators in 54 samples obtained from 27 agroforestry systems across a 270 km soil-climate gradient in Alberta, Canada.

Table S4. Overall Pearson correlations among the major phyla and soil properties in 54 samples obtained from 27 agroforestry systems across a 270 km soil-climate gradient in Alberta, Canada.

Table S5. Overall Pearson correlations among the most abundant genera and soil properties in 54 samples obtained from 27 agroforestry systems across a 270 km soil-climate gradient in Alberta, Canada.