

1    **Soil classification predicts differences in prokaryotic communities across a range**  
2    **of geographically distant soils once pH is accounted for**

3

4    Rachel Kaminsky,<sup>1</sup> Blandine Trouche,<sup>1</sup> and Sergio E. Morales<sup>1\*</sup>

5    <sup>1</sup> *Department of Microbiology and Immunology, Otago School of Medical Sciences,*

6    *University of Otago, Dunedin, New Zealand*

7

8    \*Corresponding author: [sergio.morales@otago.ac.nz](mailto:sergio.morales@otago.ac.nz) (S. E. Morales)

9

10

11

## 12    **Abstract**

13            Agricultural land is typically managed based on visible plant life at the  
14    expense of the belowground majority. However, microorganisms mediate processes  
15    sustaining plant life and the soil environment. To understand the role of microbes we  
16    first must understand what controls soil microbial community assembly. We assessed  
17    the distribution and composition of prokaryotic communities from soils representing  
18    four geographic regions on the South Island of New Zealand. These soils are under  
19    three different uses (dairy, sheep and beef, and high country farming) and are  
20    representative of major soil classification groups (brown, pallic, gley and recent). We  
21    hypothesized that pH would account for major community patterns based on 16S  
22    profiles, but that land use and location would be secondary modifiers. Community  
23    diversity and structure was linked to pH, coinciding with land use. Soil classification  
24    correlated with microbial community structure and evenness, but not richness in high  
25    country and sheep and beef communities. The impact of land use and pH remained  
26    significant at the regional scale, but soil classification provided support for  
27    community variability not explained by either of those factors. These results suggest  
28    that several edaphic properties must be examined at multiple spatial scales to robustly  
29    examine soil prokaryotic communities.

## 30 **Introduction**

31

32       Sustained population growth has placed a major strain on food production,  
 33 forcing the development of intensive land use practices that maximize yields<sup>1</sup>. This  
 34 includes use of heavy machinery and extensive applications of chemical amendments  
 35 such as fertilizers and herbicides. This intensification of agricultural production has  
 36 drastically altered soil conditions, causing physicochemical changes (e.g. compaction,  
 37 decreased organic matter and erosion)<sup>2,3,4</sup> that have led to well-documented losses in  
 38 biodiversity, including that of belowground microbial communities<sup>5,6,7</sup>. Microbes are  
 39 known to be important to maintaining ecosystem processes<sup>8,9</sup>. As a result,  
 40 understanding the consequences of these anthropogenic changes is essential for  
 41 sustained soil health.

42       Microorganisms are keystone species that contribute to soil health through  
 43 bioremediation of contaminants<sup>10,11,12</sup> and regulation of nutrient cycling<sup>13,14,15</sup>.  
 44 Despite this, the factors that control their distribution and composition are highly  
 45 contested. Many studies have shown that land use changes influence belowground  
 46 communities<sup>16,17,18</sup>, while pH is a consistent and dominant driver of microbial  
 47 assemblages on a continental scale and across a range of environments<sup>19,20,21,22</sup>.  
 48 However, other edaphic factors like C:N ratio<sup>23</sup> and soil texture<sup>24,25</sup> can affect  
 49 microbial communities. The confounding effects of specific soil factors draws  
 50 attention to a major gap in prediction and interpretation of microbial community  
 51 responses to land use change.

52       Despite the vast number of studies linking individual environmental factors to  
 53 changes in microbial community structure, the mechanisms underlying these  
 54 relationships have not been resolved. For example, though there is a widely reported

relationship between pH and microbial community structure, it is currently not clear whether pH itself is the most important factor, or if individual chemical and physical factors that contribute to pH are driving this variation<sup>19</sup>. Additionally, many studies concerning land use change focus on a single practice at a particular site<sup>24, 26, 27, 28</sup>. While such analyses provide insight into small-scale microbial community responses to land use intensification, information regarding the comparative responses of communities at multiple scales and across land use types is limited. Moreover, while microbial ecologists seek to capture any and all drivers of belowground communities, it is nearly impossible to measure all environmental factors in a given soil. Most studies evaluate physical factors in terms of soil texture, which is limited in its representation of the complexity of soil. Soil classification provides a more complete description of soils that takes into account the parent material, particle size and permeability, as well as major chemical traits<sup>29</sup>. This parameter also relates soil profiles to climactic and physicochemical features such as weathering, leaching, soil moisture, metal oxides and clay mineral content<sup>30</sup> and might provide additional resolution for characterizing prokaryotic communities.

To this end, our study used 16S rRNA gene profiles to investigate prokaryotic community composition and distribution in soils on both landscape and regional scales. We examined soils across a series of sites comprising three land use types and four geographic regions. We assess the relationship between prokaryotic communities in these soils with several abiotic factors including pH, land use and soil classification. We hypothesized that prokaryotic community structure would be primarily correlated to pH, while land use would have a secondary relationship with community structure. Furthermore, we hypothesized that soil classification—evaluated at the soil order and subgroup levels—would account for much of the

80 variation in prokaryotic communities not described by either land use or pH. Finally,  
81 we sought to understand how individual taxonomic groups responded to these factors.

82

## 83 **Results**

84

### 85 **Soil Characteristics**

86 We sampled soils under three land uses: dairy, sheep and beef, and high  
87 country. These uses differ in stock type as indicated by their names, but also in their  
88 management intensity (i.e. low country = highly managed soils with high stocking  
89 rates) as well as location (high country agriculture is carried out on high altitude  
90 pastures). Soil physicochemical characteristics varied across land uses, soil order and  
91 soil subgroup (Table S1). The sampled soils represented a range of pH values (5.1-  
92 6.3). High country soils had, on average, 1.08-fold lower pH than dairy and sheep and  
93 beef soils, which were similar in this respect. Soil classification varied within land  
94 uses, but most soils are classified within the brown and pallic soil orders, with a few  
95 dairy soils representing the recent and gley orders.

96

### 97 **Prokaryotic community structure varies with pH and land use**

98 We examined prokaryotic communities from sites representing three land uses  
99 and four geographic regions. A total of 115,445 OTUs (at 97% sequence similarity)  
100 were detected within 72 samples representing 24 sites. OTUs per sample ranged  
101 between 2,414 and 3,641. Prokaryotic alpha diversity was estimated across all  
102 samples and correlations with soil parameters were determined using linear  
103 regressions. Richness was correlated with land use (Figure 1A) (Kruskal-Wallis chi-  
104 squared = 11.3,  $p < 0.004$ ), with increasing richness from high-country sites to sheep

105 and beef sites. This trend was related to pH (Figure S1A) (regression  $R^2 = 0.23$ ,  $p <$   
 106 0.001) with richness increasing as pH became more neutral. Trends for the Shannon  
 107 diversity index were similar to those observed for richness with diversity being  
 108 correlated to both land use (Kruskal-Wallis chi-squared = 26.1,  $p < 0.001$ ) and pH  
 109 (Figure S1B) (regression  $R^2 = 0.48$ ,  $p < 0.001$ ). The remaining chemical data  
 110 measured in this study (Table S1) did not account for as much variability as pH and  
 111 land use.

112 Detrended correspondence analysis (DCA) confirmed trends observed using  
 113 alpha diversity, with both land use and pH linked to clustering of samples (Figure  
 114 1B). Samples from across the three land uses formed a gradient indicating that  
 115 differences in prokaryotic communities were primarily correlated with changes in pH  
 116 (Mantel  $R^2 = 0.63$ ,  $p < 0.001$ ). While three land uses are included in the study,  
 117 analysis of similarity (ANOSIM) testing indicated only two major categories: high  
 118 and low country soils (sheep and beef, and dairy) (Figure S2A, B) (ANOSIM  $R^2 =$   
 119 0.52,  $p < 0.001$ ). Hierarchical clustering of Bray-Curtis distances (Figure S3)  
 120 confirms the strength of high country and low country environments in explaining the  
 121 variance in prokaryotic communities (70% confidence). However, sub-clusters  
 122 representing individual replicates from a site within the high/low country split are  
 123 better supported using these methods (95% confidence), suggesting unaccounted for  
 124 factors that are linked to changes in community structure.

125

## 126 **Variation in community composition within land uses is explained by the** 127 **underlying soil classification.**

128 To assess relationships between soil properties and community variation, and  
 129 observed clustering of samples, within the three land uses data was subset by land use

130 and analyzed independently. Major differences in community structure within the  
 131 same land use were correlated with soil order, while soil subgroup resolved only a  
 132 few clusters (Figure 2). Soil subgroup has a significant effect on both the observed  
 133 species count (Kruskal-Wallis chi-squared = 32.4,  $p < 0.006$ ) and the Shannon  
 134 diversity index (Kruskal-Wallis chi-squared = 50.6,  $p < 0.001$ ) (Figure 2A).  
 135 Interestingly, samples grouped based on soil order (Figure 2B) do not have  
 136 significantly different richness values ( $p > 0.05$ ). However, soil order does correlate  
 137 weakly with Shannon diversity (Kruskal-Wallis chi-squared = 8.2,  $p < 0.05$ ).  
 138 DCA reveals that prokaryotic communities form distinct clusters based on soil  
 139 order (Figure 2C, D), though all land use sub-communities have statistically  
 140 significant relationships with both soil subgroup and soil order (ANOSIM  $p < 0.001$ ).  
 141 Soil order has a slightly stronger correlation with high country soils ( $R^2 = 0.91$ )  
 142 (Figure S4A), while sheep and beef communities ( $R^2 = 0.58$ ) (Figure S4C) have a  
 143 slightly stronger relationship with soil subgroup. Hierarchical clustering confirms  
 144 these results, where high country communities form two clusters (Figure S5), and  
 145 sheep and beef communities form two (Figure S6). On the other hand, dairy  
 146 communities do not separate according to soil classification, despite significant  
 147 correlations with soil order and subgroup ( $R^2 = 0.30, 0.67$ ) (Figure S7). These  
 148 communities remain stable across a wide geographic range, forming one large cluster  
 149 indicating that an unknown factor reduces variation in dairy soils.

150

# **151 Influences of pH and land use are stable across multiple spatial scales, but soil** **152 classification provides additional support**

153 To determine the impact of geographic scale on observed patterns (based on  
 154 pH, land use and soil classification), we individually examined the communities from

the four geographic regions (Figure S8 and S9). Prokaryotic community changes within regions confirm that pH and land use are the most significant predictors of community structure at multiple scales, while soil classification accounts for the remaining variation (Figure S10-13, Table S2). ). Interestingly, land use has the most significant relationships with regional communities where pH was the most significant variable at the multi-region scale.

161

# **Prokaryotic indicators of pH, land use and soil order**

Prokaryotic taxa (OTUs) significantly correlated ( $p < 0.001$ ) to changes in pH, land use, or soil order were identified using Spearman's correlations, the Wald test or the Kruskal-Wallis test respectively. The taxa were then mapped onto cladograms (Figure 3; taxa with correlations are provided in Supplementary Table S3).

Overall, we found 678 OTUs (0.6% of total OTUs) that were correlated with one or more edaphic properties. 34% of these OTUs correlated with pH, 27% correlated with land use and 40% correlated with soil order. The most represented phyla were the *Proteobacteria* (31% of significant OTUs), *Acidobacteria* (22%), *Actinobacteria* (17%), *Bacteroidetes* (6%) and *Planctomycetes* (5%). A consistent response to specific edaphic properties was not observed at the phylum level.

At the genus level, there was significant overlap between OTU's identified based on soils classification, pH and land use. Generally, high pH, low country soils, pallic, gley and recent soils shared correlated OTUs (e.g. *Adhaeribacter* and *Revranella*) while low pH, high country soils and brown soils had significantly correlated OTUs in common (e.g. *Bryobacter*, *Acidothumus*, *Koribacter*, *Telmatobacter*, *Mycobacterium* and *Candidatus Methyloacidiphilum*).

179           However, the relative abundances of several genera correlated with only one  
180   edaphic property. *Anaeromyxobacter*, *Singulisphaera* and *Rhodanobacter* had  
181   positive correlations with pH, while *Rhizobium*, *Variovorax* and *Flavobacterium* were  
182   negatively correlated to pH. High country soils were correlated with  
183   *Frigoribacterium*, *Jatrophihabitans* and *Massilia*, while low country soils had  
184   correlations with *Janibacter*, *Pseudonocardia* and *Pelobacter*. Lastly, *Rubrobacter*,  
185   *Defluviicoccus* and *Parasegetibacter* were most strongly correlated with brown soils  
186   while *Marmoricola*, *Nocardiodes* and *Gemmatimonas* had significant correlations  
187   with the other three soil orders.

188

## 189   **Discussion**

190

191           Results revealed that: prokaryotic assemblages differed significantly between  
192   land uses and across a pH gradient, however much of the variation within land uses  
193   and regions was better accounted for by soil order. Additionally, taxonomic profiles  
194   revealed that while overlap exists between OTUs identified as being correlated with  
195   pH, land use and soil classification, each parameter identified specific populations not  
196   correlated with either of the remaining two.

197           The studied soils harbored distinct prokaryotic communities, revealing  
198   consistent impacts of pH and, to a lesser extent, land use across spatial scales. Our  
199   results also confirm the notion that acidic soils support a smaller breadth of diversity.  
200   These results are in agreement with many previous studies that have established the  
201   role of pH and land use on prokaryotic communities<sup>19, 20, 21</sup>. It has been previously  
202   suggested that soil texture is an important predictor of prokaryotic community  
203   structure<sup>24, 31, 32</sup>. To build on this relationship, we evaluated the potential link between

204 soil classification (soil order and subgroup) and prokaryotic communities. This  
 205 allowed us to investigate the extent to which agricultural intensification impacts the  
 206 relationship between inherent soil properties, like soil texture, and prokaryotic  
 207 communities. The rationale was that soil classification provides a more thorough  
 208 representation of the soils' physical and chemical factors including those not  
 209 measured (e.g. metal oxides), as well as the geological origins of the soils.

210 We observed strong relationships between soil classification and prokaryotic  
 211 community diversity and structure. Brown soils had the lowest diversity, while pallic  
 212 soils had the highest. The low pH values of the sampled brown soils, combined with  
 213 the wet climate where some of the brown soils are commonly found<sup>30</sup>, results in low  
 214 nutrient levels compared to other NZ soils leading to conditions that select for a less  
 215 diverse community of microbes. In contrast, pallic soils have higher pH values and  
 216 are only weakly leached, retaining more nutrients allowing for a more diverse  
 217 community. While richness levels between the two soils were comparable, Shannon  
 218 diversity differed, indicating changes in evenness. As exemplified by the evenly high  
 219 levels of iron oxides in brown soils, depleting nutrient stocks and low pH lead to  
 220 uniform conditions favoring a smaller subset of taxa as shown in our study.

221 The analysis of sub-communities within each of the four regions suggests that  
 222 both land use and soil classification have strong relationships with prokaryotic  
 223 communities. Southland soils had the strongest relationship with land use, but soil  
 224 order resolved some differences between clusters along the second axis, where  
 225 communities from a recent soil clustered away from the brown soils. Recent soils are  
 226 unique in that they are weakly developed, meaning the soil has fewer horizons than  
 227 the moderately or well-developed soils comprising the other soil orders in this study<sup>33</sup>.  
 228 Prokaryotic communities from Otago soils were most strongly correlated with soil

229 subgroup. This is especially interesting, as in this region, one of the low country sites  
 230 grouped with the high country soils on the first DCA axis, but formed their own  
 231 cluster on the second axis. This cluster happens to contain communities from the only  
 232 brown soils in this particular region, providing further evidence for soil order as a  
 233 strong predictor of prokaryotic community structure. In Otago, the two pallic soils  
 234 clustered quite distantly from one another, explained by the distinction in soil  
 235 structure between laminar and fragic pallics; laminar soils have layers of clay in the  
 236 subsoil, while fragic soils are brittle, hard and contain a compacted pan in the  
 237 subsoil<sup>33</sup>.

238       Our finding that prokaryotic communities within land uses and regions  
 239 correlated with soil order indicates that soil classification is a good predictor for  
 240 prokaryotic communities that are geographically distant from one another. However,  
 241 we found that dairy communities do not separate clearly based on soil classification. It  
 242 is possible that the high stocking rates that are characteristic of dairy farms<sup>34, 35</sup> cause  
 243 heightened deposition of manure and urine, creating a new soil layer that is  
 244 fundamentally disconnected from the parent material. It has been shown previously  
 245 that dairying does impact soil ecosystems in ways that high country, and sheep and  
 246 beef management does not. For example, Barkle and colleagues observed that  
 247 application of dairy farm effluent (a mixture of water, urine and manure) onto pasture  
 248 leads to the accumulation of nutrients and increased prokaryotic biomass<sup>36</sup>. Haynes  
 249 and colleagues found similar results in camp areas (where livestock tends to  
 250 congregate) when compared to non-camp soils, which provides further insight into the  
 251 discrepancy in stocking rate as it affects prokaryotic communities<sup>37</sup>. As a result, the  
 252 inherent properties expected for soils subjected to dairy management wouldn't have a  
 253 relationship with prokaryotic communities. This also gives insight into pH, since soil

orders differ in this regard. While it is well established that soil pH is linked to prokaryotic communities on a continental scale, the factors that contribute to pH changes are unresolved<sup>19</sup>. We can hypothesize that the pH of sheep and beef, and high country soils is connected to inherent soil properties, represented by soil classification, while the pH of dairy soils has been modified by increased agricultural intensification, impacting prokaryotic communities accordingly. Furthermore, while we can be confident in the predictive power of soil order for other land uses, there is less resolution when using soil subgroup. Current methods (charting latitude and longitude onto LRIS soil maps) may not be precise enough to accurately classify soils at this level.

While we have established that pH, land use and soil order are good predictors of prokaryotic community structures, little is known about the mechanisms that account for these relationships. It is possible that pH, land use and soil order serve as integrative variables for multiple chemical and physical characteristics that individually impact prokaryotic communities. Our results suggest that land use, pH and soil order each exert direct pressure on certain prokaryotic taxa, but also contain some overlap in their taxonomic profiles, indicating that they may also integrate some of the same soil properties.

Members of both *Firmicutes* (*Bacillus*) and *Thaumarchaeota* (uncultured representative) are significantly represented in low country soils, but not at high pH levels. This is interesting, as many members of these phyla are thought to thrive at high pH levels<sup>38,39</sup>, suggesting that the members detected here have different life strategies that are selected for by land use. Additionally, DCA plotting showed that high country soils are strongly correlated with low pH, which is supported by their shared relationship with several *Acidobacteria* groups. However, there were several

279 members from the *Proteobacteria* (e.g. *Massilia*), *Actinobacteria* (e.g.  
280 *Frigoribacterium*), and *Chloroflexi* (e.g. *Ktedonobacter*) that were significantly  
281 represented in high country soils but not at low pH levels. Little is known about the  
282 ecophysiology of many of these genera. However, *Massilia* are copiotrophs, and are  
283 sensitive to nutrient availability. It is established that high country rangelands are  
284 subjected to less rigorous management regimes compared to their low country  
285 counterparts<sup>40</sup>. This management strategy may give rise to a nutrient profile that is  
286 preferable for the maintenance of *Massilia* populations<sup>41</sup>. Selection by land use is  
287 further evidenced by the strong correlation between high country soils and the  
288 verrucomicrobial phylotype Da101 and, contrastingly, a positive correlation with pH.  
289 As high country soils tend to have lower pH values, and *Verrucomicrobia* are thought  
290 to persist in low-nutrient environments<sup>42, 43</sup>, it can be inferred that the stable nutrient  
291 status of high country soils explains the abundance of this phylotype rather than pH.  
292 Other taxa, like *Gaiella* (originally isolated from an aquifer<sup>44</sup>) and *Nitrospira*, which  
293 are normally found in wet environments<sup>45</sup>, were most significantly correlated with  
294 gley soils. These soils are known to have high water tables<sup>30</sup>, which would likely  
295 provide preferable conditions for these microbes to thrive.

296 Our results confirm soil pH is the strongest predictor of community structure,  
297 diversity and composition across multiple spatial scales, but we also show strong  
298 relationships with land use and soil order. We propose that soil order may serve as an  
299 integrative factor that accounts for physical and chemical properties and can be used  
300 when direct assessment of specific edaphic factors is not possible. Further, the  
301 identification of specific OTUs correlated to more than one factor suggests that  
302 spurious correlations are highly likely and other factors besides pH might better  
303 explain observed patterns.

304

## 305 **Materials and methods**

306

### 307 **Soil sampling**

308       A total of 24 field sites across four regions on the south island of New Zealand  
 309 were sampled in this study (Figure 1). Sites were chosen to represent: the three main  
 310 land uses in New Zealand agriculture (dairy, sheep and beef, high country farming), a  
 311 wide range of edaphic parameters (Table S1), and four major regions of New Zealand  
 312 (North Canterbury, South Canterbury, Otago, Southland). Samples were collected at  
 313 the beginning of the growing season, between May 5 and May 16, 2014. Sites were  
 314 delineated in the field by twelve replicate plots (1m<sup>2</sup> each) within a gridded area  
 315 enclosed by a 6.5 by 5 m fence. Biological replicates from each site were collected by  
 316 sampling three random plots for a total of 72 samples in the study (24 sites x 3 plots at  
 317 each site). Each sample comprised a composite of four cores (7.5 cm depth and 2.5  
 318 cm diameter) that were taken 0.4 m apart diagonally across the 1m<sup>2</sup> plot. Cores were  
 319 screened prior to compositing to remove roots, worms and rocks. Samples were kept  
 320 on ice while in the field and stored at -20 degrees until returning to the lab for final  
 321 storage at -80 degrees.

322       Chemical analyses were performed by R.J. Hill Laboratories (Hamilton, NZ).  
 323 For soil pH determination, a 1:2 soil: water slurry was prepared followed by  
 324 potentiometric titration (CITE). Data for soil physical properties were obtained from  
 325 the New Zealand Land Resource Information Systems Portal  
 326 (<https://lris.scinfo.org.nz/>).

327

### 328 **DNA extraction and sequencing**

329 Genomic DNA was extracted from 0.25 g of soil using the Mo Bio PowerSoil-  
330 htp 96-well soil DNA isolation kit (Carlsbad, CA, USA) according to the  
331 manufacturer's instructions, but with a modification at the lysing step. Samples were  
332 placed on a Geno/Grinder homogenizer (SPEX Sample Prep, LLC, Metuchen, NJ,  
333 USA) for two rounds of fifteen seconds at 1750 strokes/minute. One extraction was  
334 performed on each sample. DNA concentration and purity was determined using a  
335 Nanodrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA).  
336 Absorbance was observed at 230, 260, 280 and 320 nm.

337 The V4 region of the 16S rRNA gene was amplified using the universal  
338 primer pair 515F (5'-NNNNNNNNGTGTGCCAGCMGCCGCGGTAA-3') and 806R  
339 (5'-GGACTACHVGGGTWTCTAAT-3') following the Earth Microbiome Project  
340 barcoded conditions<sup>46</sup>. Each sample was given a barcode sequence on the 5' end of  
341 the forward primer for multiplexed sequencing and loaded onto a single Illumina  
342 MiSeq 2 × 151 bp run (Illumina, Inc., CA, USA). Sequences were deposited at the  
343 Sequence Read Archive (NCBI) with the accession numbers: 5902515-5902586 under  
344 the BioProject ID: PRJNA348131.

345

### 346 **Sequence processing**

347 All sequences were initially processed using a QIIME 1.9.0 open-reference  
348 OTU-picking workflow<sup>47</sup>. In brief, raw sequences were first demultiplexed. Forward  
349 sequences were then clustered into OTUs (97% similarity) against the SILVA  
350 database release 119<sup>48</sup> using UCLUST<sup>49</sup>. Reads that failed to hit the reference  
351 database were clustered *de novo*. Taxonomy assignments were determined using  
352 BLAST<sup>50</sup> with a maximum e-value of 0.001 against the SILVA database. The  
353 resulting OTU table was then subsampled to an even depth of 12,000 sequences per

sample ten times followed by merging of the resulting ten OTU tables to reduce biases that arise from unequal library sizes. All data was then exported as a biom file.

## Statistical analyses

Sample counts were transformed by dividing the individual OTU abundances by the number of rarefactions (10) followed by rounding prior to downstream analysis using the phyloseq package<sup>51</sup> in R<sup>52, 53</sup>. Diversity estimates were determined using observed richness and the Shannon index, as calculated and plotted in phyloseq and ggplot2<sup>54</sup>. Regression analyses and Kruskal-Wallis tests were performed in R to assess the relationships between environmental variables and richness and diversity. Prokaryotic community differences were represented on a two-dimensional ordination plot using Detrended Correspondence Analysis (DCA) with the Bray-Curtis distance between samples in phyloseq and ggplot2. Analysis of Similarity (ANOSIM) was used to quantify the relationships between significant differences in community structure and categorical variables (land use and soil classification) within the vegan package<sup>55</sup>. The Mantel test was performed in vegan with 999 permutations to assess relationships between continuous variables (pH) and community structure. To identify consistent clustering patterns in the data, hierarchical clustering was performed in the pvclust package<sup>56</sup> using Ward's method and Bray-Curtis distances. To examine significant differences in the abundance and distribution of taxa between land uses, the data were transformed to relative abundance in phyloseq. The Wald chi-squared test was applied to the data using the DESeq2 package<sup>57</sup>. Spearman's rank correlations were used to test differences in taxa distributions along the pH gradient. The Kruskal-Wallis test was used to observe differences in OTU abundances of significance between the soil orders, and was performed in QIIME. Cladograms were

379 generated in GraPhlAn<sup>58</sup>. Mapping was done using GADM<sup>59</sup> in RStudio with  
380 packages: ggplot2, sp<sup>60, 61</sup>, raster<sup>62</sup>, rgdal<sup>63</sup> and ggsn<sup>64</sup>.

381

## 382 References

- 383 1. Green, R.E., Cornell, S.J., Scharlemann, J.P.W., Balmford, A. Farming and  
384 the fate of wild nature. *Science* **307**, 550-555 (2005).  
385
- 386 2. De Neve, S., Hofman, G. Influence of soil compaction on carbon and nitrogen  
387 mineralization of soil organic matter and crop residues. *Biol. Fertil. Soils*. **30**,  
388 544-549 (2000).  
389
- 390 3. Stoate, C. *et al.* Ecological impacts of arable intensification in *Europe*. *J.*  
391 *Environ. Manage.* **63**, 337-365 (2001).  
392
- 393 4. Quinton, J.N., Govers, G., Van Oost, K., Bardgett, R.D. The impact of  
394 agricultural soil erosion on biogeochemical cycling. *Nat. Geosci.* **3**, 311-314  
395 (2010).  
396
- 397 5. Buckley, D.H., Schmidt, T.M. The structure of microbial communities in soil  
398 and the lasting impact of cultivation. *Microb. Ecol.* **42**, 11-21 (2001).  
399
- 400 6. Flynn, D.F.B. *et al.* Loss of functional diversity under land use intensification  
401 across multiple taxa. *Ecol. Lett.* **12**, 22-33 (2009).  
402
- 403 7. Newbold, T. *et al.* Global effects of land use on local terrestrial biodiversity.  
404 *Nature* **520**, 45-50 (2015).  
405
- 406 8. Hooper, D.U. *et al.* Effects of biodiversity on ecosystem functioning: a  
407 consensus of current knowledge. *Ecol. Monogr.* **75**, 3-35 (2005).  
408
- 409 9. Bardgett, R.D., van der Putten, W.H. Belowground biodiversity and ecosystem  
410 functioning. *Nature* **515**, 505-511 (2014).  
411
- 412 10. Frankenberger, W.T., Arshad, M. Bioremediation of selenium-contaminated  
413 sediments and water. *Biofactors* **14**, 241-254 (2001).  
414
- 415 11. Le Borgne, S., Paniagua, D., Vazquez-Duhalt, R. Biodegradation of organic  
416 pollutants by halophilic bacteria and archaea. *J. Mol. Microbiol. Biotechnol.*  
417 **15**, 74-92 (2008).  
418
- 419 12. Kang, J.W. Removing environmental organic pollutants with bioremediation  
420 and phytoremediation. *Biotechnol. Lett.* **36**, 1129-1139 (2014).  
421

- 422 13. Ingham, R.E., Trofymow, J.A., Ingham, E.R., Coleman, D.C. Interactions of  
423 bacteria, fungi and their nematode grazers-effects on nutrient cycling and  
424 plant-growth. *Ecol. Monogr.* **55**, 119-140 (1985).  
425
- 426 14. Hamilton, E.W., Frank, D.A. Can plants stimulate soil microbes and their own  
427 nutrient supply? Evidence from a grazing tolerant grass. *Ecology* **82**, 2397-  
428 2402 (2001).  
429
- 430 15. Phillips, R.P., Finzi, A.C., Bernhardt, E.S. Enhanced root exudation induces  
431 microbial feedbacks to N cycling in a pine forest under long-term CO<sub>2</sub>  
432 fumigation. *Ecol. Lett.* **14**, 187-194 (2011).  
433
- 434 16. Bossio, D.A., Scow, K.M., Gunapala, N., Graham, K.J. Determinants of soil  
435 microbial communities: effects of agricultural management, season and soil  
436 type on phospholipid fatty acid profiles. *Microb. Ecol.* **36**, 1-12 (1998).  
437
- 438 17. Sala, O.E. *et al.* Biodiversity – global biodiversity scenarios for the year 2100.  
439 *Science* **287**, 1770–1774 (2000).  
440
- 441 18. Steenwerth, K., Jackson, L.E., Calderon, F.J., Stromberg, M.R., Scow, K.M.  
442 Soil microbial community composition and land use history in cultivated and  
443 grassland ecosystems of costal California. *Soil. Biol. Biochem.* **34**, 1599-1611  
444 (2002).  
445
- 446 19. Lauber, C.L., Hamady, M., Knight, R., Fierer, N. Pyrosequencing-based  
447 assessment of soil pH as a predictor of soil bacterial community structure at  
448 the continental scale. *Appl. Environ. Microbiol.* **75**, 5111-5120 (2009).  
449
- 450 20. Rousk, J. *et al.* Soil bacterial and fungal communities across a pH gradient in  
451 an arable soil. *ISME J.* **4**, 1340-1351 (2010).  
452
- 453 21. Siciliano, S.D. *et al.* Soil fertility is associated with fungal and bacterial  
454 richness whereas pH is associated with community composition in polar soil  
455 microbial communities. *Soil. Biol. Biochem.* **78**, 10-20 (2014).  
456
- 457 22. Samad, M.S. *et al.* Phylogenetic and functional potential links pH and N<sub>2</sub>O  
458 emissions in pasture soils. *Sci. Rep.* **6**, 35990; doi: 10.1038/srep35990 (2016).  
459
- 460 23. Kuramae, E.E. *et al.* Soil characteristics more strongly influence soil bacterial  
461 communities than land-use type. *FEMS. Microbiol. Ecol.* **79**, 12-24 (2012).  
462
- 463 24. Lauber, C.L., Strickland, M.S., Bradford, M.A., Fierer, N. The influence of  
464 soil properties on the structure of bacterial and fungal communities across  
465 land-use types. *Soil Biol. Biochem.* **40**, 2407-2415 (2008).  
466
- 467 25. Morales, S.E., Jha, N., Sagggar, S. Biogeography and biophysicochemical traits  
468 link N<sub>2</sub>O emissions, N<sub>2</sub>O emission potential and microbial communities  
469 across New Zealand pasture soils. *Soil Biol. Biochem.* **82**, 87–98 (2015).  
470

- 471 26. Kuramae, E.E., Gamper, H., van Veen, J., Kowalchuk, G. Soil and plant  
472 factors driving the community of soil-borne microorganisms across  
473 chronosequences of secondary succession of chalk grasslands with a neutral  
474 pH. *FEMS Microbiol. Ecol.* **77**, 285-294 (2011).  
475
- 476 27. Bartram, A.K. *et al.* Exploring links between pH and bacterial community  
477 composition in soils from the Craibstone Experimental Farm. *FEMS Microb.*  
478 *Ecol.* **87**, 403-415 (2014).  
479
- 480 28. Zhalnina, K. *et al.* Soil pH determines microbial diversity and composition in  
481 the Park Grass Experiment. *Microb. Ecol.* **69**, 395-406 (2015).  
482
- 483 29. Hewitt, A.E. Soil classification in New Zealand- Legacy and lessons. *Aust. J.*  
484 *Soil Res.* **30**, 843-854 (1992).  
485
- 486 30. Hewitt, A.E. Survey of New Zealand soil orders. In Dymond JR ed.  
487 Ecosystem services in New Zealand-conditions and trends. Manaaki Whenua  
488 Press, Lincoln, New Zealand (2013).  
489
- 490 31. Girvan, M.S., Bullimore, J., Pretty, J.N., Osborn, A.M., Ball, A.S. Soil type is  
491 the primary determinant of the composition of the total and active bacterial  
492 communities in arable soils. *Appl. Environ. Microbiol.* **69**, 1800–1809 (2003).  
493
- 494 32. Johnson M.J., Lee, K.Y., Scow, K.M. DNA fingerprinting reveals links among  
495 agricultural crops, soil properties, and the composition of soil microbial  
496 communities. *Geoderma* **114**, 279–303 (2003).  
497
- 498 33. Landcare Research. *Soil orders from the New Zealand soil classification*  
499 *(NZSC)*.  
500 [http://soils.landcareresearch.co.nz/contents/SoilNames\\_NZSoilClassification\\_](http://soils.landcareresearch.co.nz/contents/SoilNames_NZSoilClassification_SoilOrders.aspx?currentPage=SoilNames_NZSoilClassification_SoilOrders&menuItem=SoilNames)  
501 [SoilOrders.aspx?currentPage=SoilNames\\_NZSoilClassification\\_SoilOrders&](http://soils.landcareresearch.co.nz/contents/SoilNames_NZSoilClassification_SoilOrders.aspx?currentPage=SoilNames_NZSoilClassification_SoilOrders&menuItem=SoilNames)  
502 [menuItem=SoilNames](http://soils.landcareresearch.co.nz/contents/SoilNames_NZSoilClassification_SoilOrders.aspx?currentPage=SoilNames_NZSoilClassification_SoilOrders&menuItem=SoilNames) (2016).  
503
- 504 34. Monaghan, R.M., Paton, R.J., Smith, L.C., Drewry, J.J., Littlejohn, R.P. The  
505 impacts of nitrogen fertilisation and increased stocking rate on pasture yield,  
506 soil physical condition and nutrient losses in drainage from a cattle-grazed  
507 pasture. *New Zeal. J. Agr. Res.* **48**, 227-240 (2005).  
508
- 509 35. Clark D.A., Caradus, J.R., Monaghan, R.M., Sharp, P., Thorrold, B.S. Issues  
510 and options for future dairy farming in New Zealand. *New Zeal. J. Agr. Res.*  
511 **50**, 203-221 (2007).  
512
- 513 36. Barkle, G.F., Stenger, R., Singleton, P.L., Painter, D.J. Effect of regular  
514 irrigation with dairy farm effluent on soil organic matter and soil microbial  
515 biomass. *Aus. J. Soil Res.* **38**, 1087-1097 (2000).  
516
- 517 37. Haynes, R.J., Williams, P.H. Influence of stock camping behaviour on the soil  
518 microbiological and biochemical properties of grazed pastoral soils. *Biol. Fert.*  
519 *Soils* **28**, 253-258 (1999).

520  
521 38. Gordon, R.E., Haynes, W.C., Pang, C.H.N., Smith, N.R. The genus bacillus.  
522 *US Department of Agriculture handbook* **427**: 109-126 (1989).  
523  
524 39. Bates, S.T. *et al.* Examining the global distribution of dominant archaeal  
525 populations in soil. *ISME J.* **5**, 908-917 (2011).  
526  
527 40. Scott, D. Sustainability of New Zealand high-country pastures under  
528 contrasting development inputs. 7. Environmental gradients, plant species  
529 selection, and diversity. *New Zeal. J. Agr. Res.* **44**, 59-90 (2001).  
530  
531 41. Ofek, M., Hadar, Y., Minz, D. Ecology of root colonizing *Massilia*  
532 (Oxalobacteraceae). *PLoS ONE* **7**, e40117; doi: 10.1371/journal.pone.0040117  
533 (2012).  
534  
535 42. Fierer, N. *et al.* Reconstructing the microbial diversity and function of pre-  
536 agricultural tallgrass prairie soils in the United States. *Science* **342**, 621-624  
537 (2013).  
538  
539 43. Brewer, T.E., Handley, K.M., Carini, P., Gilbert, J.A., Fierer, N. Genome  
540 reduction in an abundant and ubiquitous soil bacterium ‘Candidatus  
541 *Udaeobacter copiosus*’. *Nat. Microbiol.* **2**, 16198;  
542 doi:10.1038/nmicrobiol.2016.198 (2016).  
543  
544 44. Albuquerque, L. *Gaiella occulta* gen. nov., sp. nov., a novel representative of  
545 a deep branching phylogenetic lineage within the class *Actinobacteria* and  
546 proposal of *Gaiellaceae* fam. nov. and *Gaiellales* ord. nov. *Sys. App.*  
547 *Microbiol.* **34**, 595-599 (2011).  
548  
549 45. Daims, H., Nielsen, J.L., Nielsen, P.H., Schleifer, K-H., Wagner, M. In situ  
550 characterization of *Nitrospira*-like nitrite-oxidizing bacteria in wastewater  
551 treatment plants. *Appl. Environ. Microbiol.* **67**, 5273-5284 (2001).  
552  
553 46. Caporaso, J.G. *et al.* Ultra-high-throughput microbial community analysis on  
554 the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621-1624 (2012).  
555  
556 47. Caporaso, J.G. *et al.* QIIME allows analysis of high-throughput community  
557 sequencing data. *Nature Methods* **7**, 335-336 (2010).  
558  
559 48. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved  
560 data processing and web-based tools. *Nucleic Acids Res.* **41**, D590-D596  
561 (2013).  
562  
563 49. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST.  
564 *Bioinformatics* **26**, 2460-2461 (2010).  
565  
566 50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment  
567 search tool. *J. Mol. Biol.* **215**, 403-410 (1990).  
568

51. McMurdie, P.J., Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217; doi: 10.1371/journal.pone.0061217 (2013).
52. R Development Core Team. R: A language and environment for statistical computing. <https://www.r-project.org/> (R Foundation for Statistical Computing, Vienna, Austria, 2008).
53. RStudio. RStudio: Integrated Development Environment for R (version 0.99.903). RStudio, Boston, MA, USA (2016).
54. Wickham, H. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York (2016).
55. Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.3-3. <http://CRAN.R-project.org/package=vegan> (2016).
56. Suzuki, R., Shimodaira, H. pvclust: hierarchical clustering with P-values via multiscale bootstrap resampling. R package version 2.0-0. <https://CRAN.R-project.org/package=pvclust> (2015).
57. Love, M.I., Huber, W., Anders, S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
58. Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**: e1029 doi: 10.7717/peerj.1029 (2015).
59. Global Administrative Areas. GADM database of Global Administrative Areas, version 2.8. [www.gadm.org](http://www.gadm.org) (2015).
60. Pebesma, E.J., Bivand, R.S. Classes and methods for spatial data in R. *R News* **5**: <http://cran.r-project.org/doc/Rnews/> (2005).
61. Bivand, R.S., Pebesma, E.J., Gomez-Rubio, V. Applied spatial data analysis with R, Second edition. Springer, New York (2013).
62. Hijmans, R.J. raster: Geographic Data Analysis and Modeling. R package version 2.5-8. <https://CRAN.R-project.org/package=raster> (2016).
63. Bivand, R.S., Keitt, T., Rowlingson, B. rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.2-4. <https://CRAN.R-project.org/package=rgdal> (2016).
64. Baquero, O.S. ggsn: North symbols and scale bars for maps created with 'ggplot2' or 'ggmap.' R package version 0.3.0. <https://CRAN.R-project.org/package=ggsn> (2016).

## Acknowledgements

618

619           We thank Mainland Minerals Ltd. for funding this research and assisting with  
620 soil sampling. We also thank Hill Laboratories and Soiltech for providing  
621 physicochemical analyses and Dr. Xochitl Morgan for her helpful comments on this  
622 manuscript. Rachel Kaminsky was funded through a Callaghan Innovation education  
623 fellowship (MMSOU1301).

624

#### 625 **Author Contributions**

626           S.E.M. designed the experiment. R.K collected and processed samples. R.K.,  
627 S.E.M and B.T. analyzed data. All authors were involved in the writing process.

628

#### 629 **Additional Information**

630

#### 631 **Accession codes**

632           Sequences were deposited at the Sequence Read Archive (NCBI) with the  
633 accession numbers: 5902515-5902586 under the BioProject ID: PRJNA348131.

634

#### 635 **Competing Financial Interests**

636           The authors declare a conflict of interest. This work was funded in part by a  
637 grant from Mainland Minerals Ltd, a fertilizer company in New Zealand.

638

## 639 **Figure legends**

640

641 **Figure 1** Relationships between bacterial communities under different land uses and  
642 pH. Changes in Alpha (Richness and Shannon Diversity) (A) and Beta (Detrended  
643 correspondence analysis based on Bray-Curtis dissimilarity) diversity metrics in  
644 response to land use and pH (B).

645

646 **Figure 2** Soil classification predicts prokaryotic community structuring within each  
647 land use. Comparison of diversity metrics for each soil subgroup (A) and (B) soil  
648 order. High country (C), Sheep and Beef (D), and dairy (E) soil communities  
649 evaluated using DCA ordination based on Bray-Curtis dissimilarity with color  
650 representing soil subgroup and shape representing soil order.

651

652 **Figure 3** Cladograms showing relationships between key taxa and edaphic properties.  
653 (A) OTUs (97% sequence similarity) significantly correlated with high or low country  
654 soils and are strongly correlated with changes in pH. Significance for land use  
655 preference was determined using the Ward method with a Z lower-limit of 6 and a *p*-  
656 value of <0.001. Correlation with pH was determined by a Spearman's correlation  
657 with a Rho lower-limit of 0.5/-0.5 and a *p*-value of <0.001. Light blue indicates a  
658 negative correlation with pH, and dark blue is positive (B) OTUs significantly  
659 correlated with specific soil orders. Significance was determined using the Kruskal-  
660 Wallis test with a chi-squared lower-limit of 27 and a *p*-value of <0.001. Brown soils  
661 are indicated by yellow, pallic by red, gley by green and recent by blue. A gradient of  
662 8 shades for each color was generated to indicate abundance, where white indicates an  
663 abundance of 0 and the darkest shades indicate an abundance >100.

664

665 **Figure 4** Map of sampling sites throughout the South Island of New Zealand. High  
 666 country, dairy, and sheep and beef sites are indicated by triangles, circles and squares,  
 667 respectively. 1-North Canterbury, 2-South Canterbury, 3-Otago, 4-Southland. The  
 668 map was generated using shapefiles from GADM (v. 2.8, <https://www.gadm.org>) in  
 669 RStudio (v. 0.99.903, <https://www.rstudio.com/>) using ggplot2 (v. 2.1.0), sp (v. 1.2-  
 670 3), raster (v. 2.5-8), rgdal (v. 1.2-4) and ggsn (v. 0.3.1).

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689  
690  
691  
692







