Differences in Bacterial Diversity, Composition and Function due to Long-Term Agriculture in Soils in the Eastern Free State of South Africa

Joel P. Dube 1, Angel Valverde 2, Joachim M. Steyn 1, Don A. Cowan 3 and Jacqueline E. van der Waals 1, * 1

Article

Abstract: Land-use change from natural to managed agricultural ecosystems significantly impacts soil bacterial diversity and function. The Eastern Free State (EFS) is one of the most productive agricultural regions in South Africa. However, no studies aiming to understand the changes in bacterial diversity, composition and function due to land-use change in this area have been conducted. This study investigated, using high-throughput 16S rRNA gene amplicon sequencing, the effects of long-term agriculture on bacterial diversity, composition and putative function in the EFS by comparing microbiomes from lands that have been under agronomic activity for over 50 years to those from uncultivated land. Results indicate that agriculture increased bacterial diversity. Soil chemical analysis showed that land-use shifted soils from being oligotrophic to copiotrophic, which changed bacterial communities from being Actinobacteria dominated to Proteobacteria dominated. Predictive functional analysis using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) suggested that agricultural soil was abundant in genes associated with plant fitness and plant growth promotion, while non-agricultural soil was abundant in genes related to organic matter degradation. Together, these results suggest that edaphic factors induced by long-term agriculture resulted in shifts in bacterial diversity and putative function in the EFS.

Keywords: bacterial diversity; bacterial function; high-throughput 16S rRNA gene sequencing

1. Introduction

The global human population is expected to reach 10 billion by 2050 [1] and it has been projected that crop production must double in the next 50 years to face growing food demand and bioenergy needs [2]. The South African population, currently estimated at 57.7 million [3], is increasing at a rate of 2% per year and is expected to reach 82 million by 2035 [4], therefore presenting a threat to food security. This has resulted in an increase in land-use for food production in areas such as the Eastern Free State (EFS). The agricultural production area in the EFS increased from 11,321,192 ha in 1993 [5] to 12,945,800 ha in 2017 [6].

The Eastern Free State is one of the greatest contributors to agriculture in South Africa and is part of the “granary” or “breadbasket” of South Africa [7]. The EFS region is characterized by good soils and reliable rainfall, making it one of the most productive regions in South Africa, producing a wide range of crops [8]. The area falls under the cold interior climatic zone with a cold semi-arid
climate and an aridity index of 0.2–0.5 [9]. The mean annual precipitation ranges 600–800 mm with an average daily temperature of 20–22 °C. To maximize yield in different crops, agronomic practices such as fertilization, conventional tillage and use of pesticides are common features of the Eastern Free State’s intensive farming regime [9]. However, there is concern regarding the effects of these agronomic practices on soil physical and chemical properties and health, particularly in relation to bacterial diversity and function.

Great attention is currently being given to adopting sustainable land-use practices that can reduce land degradation [10]. Globally, land under agricultural production has increased by 12%, and this has been associated with a decline in area under tropical forests and natural grasslands, especially in developing countries [11]. This reduction in natural habitats has largely been driven by agriculture for food production. Understanding the environmental effects of land-use conversion from natural habitats to intensively managed agricultural systems, specifically on bacterial communities that drive significant biogeochemical processes, is important for sustainable crop production. A few studies on soil microbiota have demonstrated the impact of agronomic practices on bacterial communities [12–15], made possible through advances in high-throughput sequencing [16,17]. These studies were conducted in temperate regions and in irrigated fields. However, no studies have been published on the impact of long-term agriculture on soil bacterial diversity and function in the semi-arid region of South-Africa where dry land cropping is practiced.

This study aimed to provide new insights into the long-term impact of agriculture on soil bacterial diversity and function using high-throughput sequencing. Specifically, the aim was to investigate differences in bacterial community structure and putative function in the EFS between uncultivated soil and soil under intensive agriculture, using 16S rRNA sequencing. To our knowledge, no work has been done to evaluate the effects of agronomic practices on the soil microbiota in the EFS. A previous ecological study in this region focused only on comparing conservation farming and no-tillage as sustainable practices [9].

2. Materials and Methods

2.1. Site Description and Sampling

Soil samples were collected in August 2016 from six locations (from 10 m to 21 km apart) in the EFS farming area (Figure 1). Soil samples were collected from five fields with at least 50 years of agricultural history and one location with its natural flora representing uncultivated, non-agricultural land. At the time of sample collection, no crops were present in the agricultural fields. The natural flora in the non-agricultural location comprised 20% trees, 70% grass and 10% shrubs. Sampling and processing equipment were sterilized with 10% sodium hypochlorite and rinsed in distilled water between each sampling location. The agricultural fields ranged in size from 6 to 10 ha and the non-agricultural field was approximately 0.5 ha. Forty samples were collected at 20 m spacings in each agricultural field and at 2 m spacings in the non-agricultural site in a classical W-shaped pattern. Distances were closer in the non-agricultural location because the total field area was smaller than the agricultural fields. From the 40 samples, composite samples were made by pooling every 10 samples, resulting in four composite samples for each agricultural field and non-agricultural site. Therefore, a total of 20 agricultural (four composite samples x five fields) and 20 non-agricultural soil samples were collected and sent for Illumina sequencing. All soil samples (approximately 1.7 kg) were collected using a 10 cm diameter soil auger at a depth of 0–15 cm. Soil samples were stored at −80 °C within 24 h of sampling until further processing.
Figure 1. Map of the study area showing the location of sampling areas in the Eastern Free State of South Africa. Black dots indicate the sampling locations. Map courtesy of Dr. Michelle Greve (University of Pretoria).

2.2. Soil Chemical Analyses

From each agricultural site, 10 soil samples, approximately 1.7 kg each and randomly collected from a sampling depth of 15 cm with the aid of a 10 cm diameter soil auger, were collected for chemical analysis. The 10 soil samples were combined to form a composite sample for chemical analysis for each sampling location. For the non-agricultural site, 10 soil cores, also approximately 1.7 kg each, were collected for soil chemical analysis at each subplot from a sampling depth of 15 cm with the aid of a 10 cm diameter soil auger. The 10 soil samples per non-agricultural subplot were also pooled to form a composite sample for chemical analysis. A total of ten soil samples (five agricultural and five non-agricultural) were submitted for soil chemical analyses. Soil chemical analyses were conducted in the analytical laboratory of the Department of Plant and Soil Sciences, University of Pretoria, South Africa, using standard protocols [18,19]. Basic soil parameters (pH, total C, P, N, Ca, Mg, Na and organic matter) were determined for each location. Soil pH was determined using the 1:1 water ratio method [20].

2.3. DNA Extraction

Total genomic DNA of each sample was extracted from 1 g of soil using the Mo Bio PowerSoil® DNA Isolation Kit (Carlsbad, CA, USA) according to the manufacturer’s recommendations. DNA was quantified using the Qubit 2.0 fluorometer (Invitrogen, Life Technologies, Carlsbad, CA, USA) and the quality was checked on a 1% agarose gel stained with Lonza Gel Stain, run at 90 V for 45 min and
visualized on a Gel Doc XR system (Bio-Rad, Pleasanton, CA, USA). The DNA was stored at −20 °C until further analysis.

2.4. 16S rRNA Gene Sequencing

Sequencing was performed at MRDNA (www.mrdnalab.com, Shallowater, TX, USA) using the Illumina MiSeq workflow following the manufacturer’s guidelines. The Illumina (2 × 250 bp) tailed polymerase chain reaction (PCR) approach was used to generate amplicon libraries. Amplicons from different samples were mixed in equal concentrations and purified using Agencourt Ampure beads (Agencourt Bioscience Corporation, USA). The V4 variable region of 16S rRNA gene was amplified using PCR primer pair 515F (5′-GTGYCAGCMGGCGGGRA-3′) and 806R (5′-GCCCGGCGCAGATTCMTTTRAG-3′) [21]. The forward primer was tagged with a barcode and used in a 28 cycle PCR (5 cycles used on PCR products) using the HotStarTaq Plus Master Mix Kit (Qiagen, USA) under the following conditions: 94 °C for 3 min, followed by 28 cycles of 94 °C for 30 s, 53 °C for 40 s and 72 °C for 1 min, after which a final elongation step at 72 °C for 5 min was performed. PCR products were confirmed on a 2% agarose gel and samples with 200 bp amplicons were purified using calibrated Ampure XP beads. The purified PCR products were used to prepare a DNA library by following the Illumina TruSeq DNA library preparation protocol.

2.5. 16S Amplicon Analysis

Sequence data were processed using the Quantitative Insights into Microbial Ecology (QIIME) software version 1.9.1 [21]. The sequences were joined, and barcodes removed, and sequences of less than 200 bp or with more than two ambiguous base calls were also removed. Quality trimming was performed for all sequences greater than 200 bp by removing all sequences with quality scores less than 25 or containing more than one mismatch to the sample-specific barcode or to the primer sequences. Chimeric sequences and sequences with homopolymer runs exceeding 6 bp were also removed. OTUs were selected at 3% divergence (97% similarity level) using USEARCH v6.1 [22]. All bacterial taxonomies were assigned to an OTU using the RDP Naïve Bayesian Classifier [23] with the SILVA-ARB database (release 123). Final OTUs were taxonomically classified using BLASTn against a curated database derived from GreenGenes, RDPII and NCBI (www.ncbi.nlm.nih.gov; [24]; http://rdp.cme.msu.edu) and all unassigned OTUs at phylum level were removed for downstream analyses. All samples were rarefied to the same number of sequences (183,613) per sample, and a total of 40 (20 agricultural and 20 non-agricultural) samples were used for downstream analysis.

2.6. Bacterial Community Analysis

Alpha diversity parameters (richness, Shannon and Simpson’s Diversity index) were calculated using the vegan R package [25,26]. The Simpson’s Index of Diversity (1-D) was calculated by the R-vegan package, where \( D = \sum p_i^2 \) and \( p_i \) is the proportional abundance of species (OTUs) \( i \). The observed differences in the Simpson index were not statistically significant. ACE and Chao1 parameters were calculated using Visual and Analysis of Microbial Population Structures (VAMPS) [27]. Analysis of variance (ANOVA) was carried out to determine significant differences in bacterial diversity and soil chemistry between agricultural and non-agricultural soil using the phia package [28]. Analysis of similarity (ANOSIM) was applied to determine significant differences in bacterial diversity and soil chemistry. Soil chemistry data were standardized and pair-wise distances computed based on Euclidean distances. Community data matrices were Hellinger-transformed, and the Bray–Curtis distance measure was used to generate a dissimilarity matrix. Weighted and unweighted UniFrac dissimilarities were also obtained [29]. Bacterial community structure and the environmental variables were visualized using Principal Component Analysis (PCA) plots generated using ClustVis [30]. PCA was used to explain the effect of environmental factors on variations observed in bacterial communities. Environmental variables that contributed to the variation in community composition were selected for further analysis. The Adonis function in Vegan was used to assess factors that
contributed to differences in bacterial community composition for each habitat or land use type. The average relative abundance and frequency of occurrence values of each phylum across each sample type were plotted as a sunburst pie chart in Krona [31], to infer putative ecological relevance. Rarefaction curves were also constructed in VAMPS [27].

2.7. Community Gene Function Prediction

Functional inferences were performed according to the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) [29]. Metagenomic inferences were done using the taxonomy generated from the GreenGenes database [24,32]. The OTU table was normalized based on the predicted 16S copy number for each organism in the OTU table using the normalize_by_copy_number.py script. Functional predictions of Kyoto Encyclopedia of Genes and Genomes (KEGG) Ortholog (KOs) were performed using the normalized OTU table as input. KOs were then categorized by function to collapse all KOs to the respective KEGG pathways. The contribution of each OTU to a given gene function was then estimated by multiplying OTU abundances by the set of gene abundances for each taxon. The resulting annotated table of predicted gene family counts for each sample was then visualized in Statistical Analysis of Metagenomic Profiles (STAMP) using Microbiome Helper. All predictions were subjected to a multiple test correction using the Benjamin–Hochberg FDR at a $q$-value filter of $<0.05$. Additionally, weighted nearest sequenced taxon index (NSTI) values were computed from metagenomic data using the -a option in the predict_metagenomes.py script [33]. NSTI scores are a quality control measure that is used to validate the accuracy of functional annotations by quantifying the availability of nearby genome representatives for each microbiome sample [34].

3. Results

3.1. Bacterial Communities from Agricultural and Non-Agricultural Soils

A total of 7,344,520 high-quality 16S rRNA gene amplicon sequences were recovered after pair-end alignments, quality filtering and removal of chimeric sequences and singletons. Sequences were assigned to 11,615 OTUs. The six most abundant bacterial phyla across all samples were Proteobacteria (39%), Actinobacteria (27%), Bacteroidetes (10%), Chloroflexi (8%), Firmicutes (6%) and Acidobacteria (4%) (Figure S1). Members of the Nitrospirae, Cyanobacteria, Planctomycetes, and Verrucomicrobia were detected in lower frequencies (<1%). The OTUs within the Proteobacteria phylum were mostly identified as members of the classes Gammaproteobacteria (37%), Alphaproteobacteria (27%), Betaproteobacteria (26%), and Deltaproteobacteria (10%). Actinobacteria were mainly represented by members of the class Actinobacteria (89%) and Thermoleophilia (8%), while Bacteroidetes were mostly represented by members of the classes Saprospirae (53%), Sphingobacteria (22%), Flavobacteria (11%), and Cytophagia (7%). Chloroflexi were mostly represented by members of the class Ktedonobacteria (35%), C0119 (23%) and Thermomicrobia (13%), while Firmicutes were mostly represented by members of the class Bacilli (96%) and Clostridia (4%). Acidobacteria were mostly represented by members of the class Acidobacteriia (20%), Solibacteres (18%), Chloracidobacteria (16%), Acidobacteria-6 (25%), iii-8 (13%) and DA 052 (4%). Overall, the genera Chitinophaga (5%), Enterobacteria (4%), Bacillus (3%), Dyella (3%), Streptomyces (3%) and Fulvimonas (3%) were the most abundant genera detected in all soil samples. Statistical analysis revealed highly significant differences ($p < 0.0001$, $R^2 = 0.489$) at genus level between agricultural and non-agricultural soil.

3.2. Compositional Differences in Bacterial Diversity Between Agricultural and Non-agricultural Soil

Significant compositional differences in bacterial microbiota were observed between agricultural and non-agricultural soils (Table 1) ($p < 0.001$, $R^2 = 0.241$). The relative abundance of Proteobacteria (40%), Bacteroidetes (12%), Chloroflexi (8%), and Firmicutes (7%) was higher in agricultural samples than in non-agricultural samples. Actinobacteria (40%) and Acidobacteria (6%) were more abundant in non-agricultural soil (Figure 1) than in agricultural soil that had 23% and 3%, respectively. Agricultural
samples exhibited more unique OTUs compared to non-agricultural samples. A total of 2352 unique OTUs (18.8%) were detected in agricultural soil compared to 910 unique OTUs (7.3%) detected in non-agricultural samples (Figure 2). There were 9263 shared OTUs between the two habitats which accounted for 74% of the OTUs.

Table 1. Total relative abundance of bacterial phyla detected in agricultural and non-agricultural soil samples from the EFS. P-values marked with an asterisk (*) indicate significant differences (p < 0.05).

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Agricultural</th>
<th>Non-Agricultural</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteobacteria</td>
<td>40%</td>
<td>34%</td>
<td>3.788 x 10^{-5} *</td>
</tr>
<tr>
<td>Actinobacteria</td>
<td>23%</td>
<td>40%</td>
<td>0.0323 *</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>12%</td>
<td>4%</td>
<td>0.0410 *</td>
</tr>
<tr>
<td>Chloroflexi</td>
<td>8%</td>
<td>6%</td>
<td>0.8106</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>7%</td>
<td>4%</td>
<td>0.4156</td>
</tr>
<tr>
<td>Acidobacteria</td>
<td>3%</td>
<td>6%</td>
<td>3.877 x 10^{-5} *</td>
</tr>
</tbody>
</table>

Figure 2. Unique and shared phyla in agricultural and non-agricultural soils.

3.3. Effects of Soil Chemistry on Bacterial Diversity

Bray–Curtis distances analysis showed significant effects (p < 0.001) of soil chemistry on soil bacterial community structure. The redundancy analysis eluded that N, P and K played a significant (p < 0.05) role in shaping bacterial community composition in agricultural soils, while soil organic matter (OM), Mg and Ca significantly influenced bacterial diversity in non-agricultural soils (Figure 3). Soil pH and Na did not have any significant effect on bacterial community structure (p > 0.05). The multivariate linear models revealed that soil chemistry explained 89.4% (PCO1 = 67.9% and PCO2 = 21.5%) of the total variation observed in the number of OTUs between the two habitats. Bray–Curtis analysis also showed a highly significant shift in bacterial diversity as land-use changed from natural flora to intensive agriculture (p < 0.001), with good sample separation between agricultural and non-agricultural soil samples (ANOSIM R = 0.81).
Figure 3. Distance-based redundancy analysis (db-RDA) biplot of a) bacterial communities and soil chemistry parameters. Only the environmental variables that significantly \( (p < 0.05) \) explained variability in microbial community structure are shown (arrows). The direction of the vectors indicates the direction of maximum change of that variable, whereas the length of the vector represents how well the parameter explains the distribution of the data.

3.4. Rarefaction Curves and Diversity Parameters

The OTU accumulation curves showed good sequence saturation at genus level for both agricultural and non-agricultural samples (Figure S2). All diversity measures (richness, ACE, Chao1, and Shannon index) were observed to be significantly higher in agricultural soil \( (p < 0.05) \) than in non-agricultural soil samples (Table 2). The Simpson index showed no significant differences between agricultural and non-agricultural soils.

Table 2. Differences in bacterial richness and diversity measures of the 16S rRNA gene libraries between the agricultural and non-agricultural soil samples. Standard deviations (SD) from the means are indicated in parentheses. Values with marked with an asterisk indicate the land use type that was significantly higher for each diversity index.

<table>
<thead>
<tr>
<th>Diversity Index</th>
<th>Agricultural</th>
<th>Non-Agricultural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richness</td>
<td>598 (2.85) *</td>
<td>504 (2.56)</td>
</tr>
<tr>
<td>ACE</td>
<td>46.4 (2.78) *</td>
<td>34.1 (3.20)</td>
</tr>
<tr>
<td>Chao1</td>
<td>44.5 (4.96) *</td>
<td>33.0 (4.09)</td>
</tr>
<tr>
<td>Shannon</td>
<td>2.80 (0.31) *</td>
<td>2.58 (0.12)</td>
</tr>
<tr>
<td>Simpson</td>
<td>0.76 (0.10)</td>
<td>0.80 (0.03)</td>
</tr>
</tbody>
</table>

A principal component analysis (PCA) using the sequence data revealed significant differences \( (p < 0.05) \) in bacterial community composition between agricultural and non-agricultural soils (Figure 4), with 22% (Principal Component 1 = 13.5% and Principal Component 2 = 8.5%) of the sequences explaining the observed variation. Soil samples from the respective habitats also showed separate clustering of agricultural samples from non-agricultural samples.
Figure 4. A principal component analysis (PCA) plot showing differences in structure of bacterial communities between agricultural (red circles) and non-agricultural (blue circles) soil at phylum level. Unit variance scaling is applied to rows; Singular value decomposition (SVD) with imputation was used to calculate principal components. X and Y axes show Principal Component 1 and Principal Component 2 that explain 13.5% and 8.5% of the total variance, respectively. Prediction ellipses are such that with a 0.95 probability, a new observation from the same group will fall inside the ellipse. N = 40 data points.

3.5. Functional Predictions

Functional predictions were generated using PICRUSt. A total of 6909 KEGG Orthology (KOs) groups were obtained, of which 6268 (90.7%) were assigned to functional categories, while 641 (9.3%) either did not have known functional roles or had uncharacterized functions and were filtered from the analysis. The two habitats shared 6104 (97.3%) KO groups, while 111 (1.8%) were unique to agricultural soil and 53 (0.8%) were unique to non-agricultural soil (Figure S3). A total of 328 active KEGG pathways were identified, although only 203 were statistically significant between agricultural and non-agricultural land. An extended error bar plot of the predicted functional profiles at the highest-level subsystems revealed that agricultural soil was significantly richer in KO groups for cellular processes, signaling, and environmental information processes, while non-agricultural soil was significantly richer in genetic information processing, metabolism and poorly characterized processes (Figure S4). Principal component analysis at level 1 illustrated that PC1 explained 58.8% while PC2 explained 20.4% of the variation observed in functional pathways between the two habitats (Figure 5). PICRUSt analyses also revealed significant differences (p < 0.0001) between average NSTI scores of agricultural soil (0.12 ± 0.031 standard deviation) and non-agricultural soil (0.20 ± 0.021 standard deviation), respectively. NSTI scores were thus lower in agricultural samples compared to non-agricultural samples.
Active features predicted in both habitats were amino acid metabolism, energy metabolism and membrane transport. The functional categories of translation, lipid metabolism, cell motility, nucleotide metabolism and metabolism of co-factors and enzymes were moderately abundant (Figure 6). Functional categories associated with plant growth promotion and plant fitness [35], including plant hormone balance, molecular communication, plant nutrition, plant adaptation to drought and disease suppression were compared between the two habitats (Figure S5). No significant differences were observed between the two habitats in any of the functions responsible for plant hormone balance. In terms of molecular communication, there were significant differences in secretion systems ($p < 0.0001$) and bacterial chemotaxis ($p < 0.0001$) between the two habitats. Agricultural soil exhibited higher relative abundance of genes coding for both secretion systems and bacterial chemotaxis functions compared to non-agricultural soil samples. For plant nutrition, nitrogen metabolism genes were significantly higher in agricultural soil ($p < 0.0001$) while calcium signaling was significantly higher in non-agricultural soil ($p < 0.0001$). For functions associated with plant adaptation to drought, lipopolysaccharide biosynthesis was observed to be significantly higher in agricultural samples ($p < 0.0001$) compared to non-agricultural samples. For disease suppression, sulfur metabolism and biosynthesis of siderophores were significantly higher in agricultural samples while streptomycin synthesis functions (although having a low percentage mean proportion) were significantly higher in non-agricultural soils ($p < 0.0001$).
Figure 6. Extended error bar plot showing the mean proportion (%) of significantly different predicted functional categories at level 3 found in soil bacterial communities. Points indicate the differences between agricultural and non-agricultural soil (blue and green bars, respectively), and the values on the right show the p-values that were derived from a White’s non-parametric t-test [36].

4. Discussion

Differences in the relative abundances of the microbes and their predicted functions were investigated for two habitats, agricultural and non-agricultural soils in the Eastern Free State of South Africa. In this study, the dominant bacterial phyla in agricultural and non-agricultural soils were Proteobacteria, Bacteroidetes, Chloroflexi, Acidobacteria and Firmicutes. These results agree with the findings of Janssen [37] and Delgado-Baquerizo et al. [38] who reported these six phyla among the nine most dominant and ubiquitous bacterial phyla in soil. The relative abundances of phyla found in this study also support the trends observed in the most dominant phyla in 21 clone libraries prepared from soil samples across Europe [37]. The overall high abundance of Proteobacteria in agricultural samples was reported by Dai et al. [39], who concluded that, although long-term nitrogen fertilization decreases bacterial diversity, it generally favors the abundance of Proteobacteria and Actinobacteria in agro-ecosystems across the globe. The high relative abundance of Proteobacteria in agricultural soil can account for the high predicted abundance of functions responsible for plant growth promotion. Many genera under the phylum Proteobacteria are classified as plant growth promoting rhizobacteria (PGPR), which facilitate nutrient availability and plant fitness [40]. Conversely, although recent studies have shown that Actinobacteria increase with nitrogen application in agricultural soils [39,41], Actinobacteria were observed to be more abundant in non-agricultural soil in this study.
These findings agree with the copiotroph/oligotroph hypothesis, as suggested by Fierer et al. [42], which postulates that oligotrophic bacteria are more abundant in soil with low nutrient levels. A previous study by Castro et al. [43] also confirms that members of the phylum Actinobacteria are predominant in oligotrophic soil. Additionally, the low relative abundance of cell division and cell cycle functions in non-agricultural soils further confirms its oligotrophic nature, as these functions were shown to be abundant in high-nutrient environments [44]. Overall, our comparative phylogenetic analyses suggest that the Actinobacteria are dominant in nutrient-deficient undisturbed soils, while Proteobacteria are dominant in nutrient-rich soils under intensive cultivation.

Several researchers have reported bacterial diversity loss due to agriculture, largely through conversion of natural habitats into intensely managed systems [45–48]. However, all the diversity indices calculated in this study indicated that land under intensive agricultural production for over 50 years had significantly higher bacterial diversity compared to the uncultivated land (Figure S6). Although most agricultural practices have been associated with soil microbial diversity loss, Altieri [49] found that some agricultural practices such as crop rotation have the potential to sustain soil biodiversity and enhance functional diversity. Crop rotation, as practiced in the EFS, could thus explain the observed increased diversity in the agricultural soils. Growing different crops in succession facilitates the supply of different nutrient sources in the form of root exudates and plant residue [50], which in turn may result in the selection or introduction of different types of microbes thereby continuously renewing biological activity in the soil [51]. In addition, management practices in agro-systems generate spatial and temporal changes in soil physical and chemical properties, creating rapidly fluctuating environments that provide a wide range of niches for microbial growth [15,52]. This selection process may gradually differentiate the soil microbiome of agricultural soil from the non-agricultural soil [53].

Our results also support previous views that soil microbes respond to differences in soil chemistry [54–59]. Soil chemical analyses indicated significant differences in soil chemistry due to cultivation. Lower levels of soil organic matter content in cultivated soil (compared to non-cultivated soils) have been widely reported [53,60–66]. The loss of soil organic matter in cultivated soils is attributed to management practices that involve the removal or burning of plant residues, and/or the accelerated breakdown, decomposition and mineralization of organic carbon [63,67,68]. In this study, uncultivated soil was also observed to be significantly higher (p < 0.0001) in soil organic matter and bacterial functions associated with high organic matter such as carbon fixation and methane metabolism. High levels of carbon fixation and methylotrophy functions correlate with high organic matter content in the soil [69,70]. In field experiments comparing the direct and residual effects of organic and inorganic fertilizers on microbial components [71], intensive use of fertilizers containing high levels of N, P and K have been associated with increased bacterial diversity in the soil. This is consistent with our findings, where N, P and K levels were the dominant drivers of bacterial diversity in cultivated soils.

Tools such as PICRUSt that predict function from metagenomes using phylogenetic marker genes can provide information on correlations between phylogeny and function. However, the reliability of such tools needs to be understood in assessing the validity of such predictions. Computing NSTI scores as a quality control measure has been widely used to validate metagenomic predictions from PICRUSt [72,73]. Although functional predictions from PICRUSt showed lower NSTI scores in agricultural soils (0.12 ± 0.031 s.d.) than in non-agricultural soils (0.20 ± 0.021 s.d.) in this study, these values indicate that most of the genomes from the soil samples have not yet been sequenced, suggesting poor sequence depth. Previous studies have shown that soil metagenomes have a moderate prediction accuracy (NSTI = 0.17) compared to human gut metagenomes that have a high prediction accuracy (NSTI < 0.03) [34]. This can be explained by the fact that much of the soil microbiome is still largely characterized and some of the functional capacities are unknown [74]. The discrepancy between results from the previous study and results from this study can also be explained by the differences in sequence read length, as the previous study used a 448 bp read length from the V3–V5 region to validate PICRUSt, compared to the 80 bp read length from the V6 region used in the present study [72].
A key finding in our study was the high relative abundance of predicted bacterial functions related to plant fitness, plant growth promotion and disease suppression (sulfur metabolism, lipopolysaccharide biosynthesis, and biosynthesis of siderophores) in agricultural soil compared to non-agricultural soil (Figure S5) [75,76]. Together with siderophore biosynthesis, lipopolysaccharide biosynthesis has been shown to be involved in induced systemic resistance in plants in response to a phytopathogen [77]. This is possibly related to the observed high relative abundances of bacteria that have been implicated in disease suppression (such as Bacilli, Burkholderia, Pseudomonas, Serratia, Streptomycyes, and Stenotrophomonas species (Figure S7)) and were significantly more abundant in agricultural soil than in non-agricultural soil ($p < 0.001$). However, the number of sequences for streptomycin biosynthesis, which is responsible for the production of streptomycin, an important suppressive antibiotic, was observed to be significantly higher in non-agricultural soil. This could be attributed to the observed high relative abundance of Actinomycetes in non-agricultural soil [78].

Our results indicate that the soil bacterial diversity and composition of cultivated soil in the Eastern Free State is different from that of uncultivated soil. The agricultural soil microbiome was predicted to be abundant in functions that promote plant growth and fitness compared to non-agricultural soil. Using PICRUSt, we were able to suggest functional differences that may be a product of agricultural practices. The findings from this study can serve as a baseline for future studies focusing on the impact specific groups of bacteria on crop productivity. The focus of this study was limited to bacterial diversity and functional inferences based on the 16S rRNA gene. Information on diversity of fungi and viruses would also be required to ensure a wholistic understanding of the soil microbiome in the Eastern Free State.

Supplementary Materials: The following are available online at http://www.mdpi.com/1424-2818/11/4/61/s1, Figure S1: Sunburst chart showing the total relative abundance of bacterial phyla detected in investigated samples, Figure S2: Rarefaction analyses of samples at phylum level, Figure S3: Unique and shared KEGG Orthologs (KOs) between the agricultural and non-agricultural soils, Figure S4: Extended error bar plot for the six active features (high-level subsystems) that had significant differences between agricultural and non-agricultural samples, Figure S5. Extended error bar plot showing the abundances of functions associated with plant interaction that had significant differences between agricultural and non-agricultural samples, Figure S6: Heat map showing differences in relative abundance of the 50 most abundant bacterial genera as revealed by ClustVis statistical analysis, Figure S7. Relative abundance of different genera involved in plant growth promotion and plant fitness between agricultural and non-agricultural soils.


Funding: This research was funded by Potatoes South Africa, grant specific unique reference number (UID) 105649.

Acknowledgments: A special thanks goes to Mr. Gert Bester in the Eastern Free State of South Africa for giving us access to his farm and assisting in sample collection. We also would like to acknowledge Dr. Michelle Greve for assisting us with the mapping software and drawing the map.

Conflicts of Interest: Authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References


15. Trivedi, P.; Delgado-Baquerizo, M.; Anderson, I.C.; Singh, B.K. Response of Soil Properties and Microbial Communities to Agriculture: Implications for Primary Productivity and Soil Health Indicators. Front. Plant Sci. 2016, 7, 990. [CrossRef]


22. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010, 26, 2460–2461. [CrossRef]
Diversity 2019, 11, 61


48. Dudley, N.; Alexander, S. Agriculture and biodiversity: A review. *Biodiversity* 2017, 18, 45–49. [CrossRef]


