Rhizosphere shotgun metagenomic analyses fail to show differences between ancestral and modern wheat genotypes grown under low fertiliser inputs

Liliana Quiza[1], Julien Tremblay[2], Charles W. Greer[2], Sean M. Hemmingsen[3], Marc St-Arnaud[4], Curtis J. Pozniak[5], Etienne Yergeau[1]*

[1]Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique, QC, Canada

[2] Energy, Mining, and Environment, National Research Council Canada, Montréal, QC, Canada

[3]Aquatic and Crop Resource Development, National Research Council Canada, Saskatoon, SK, Canada

[4]Institut de recherche en biologie végétale, Université de Montréal and Jardin botanique de Montréal, Montréal, QC, Canada

[5]Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada

*Correspondence: Etienne.Yergeau@inrs.ca

Abstract

It is thought that modern wheat genotypes have lost their capacity to associate with soil microbes that would help them acquire nutrients from the soil. To test this hypothesis, ten ancestral and modern wheat genotypes were seeded in a field experiment under low fertilization conditions. The rhizosphere soil was collected, its DNA extracted and submitted to shotgun metagenomic sequencing. In contrast to our hypothesis, there was no significant difference in the global rhizosphere metagenomes of the different genotypes, and this held true when focusing the analyses on specific taxonomic or functional categories of genes. Some genes were significantly more abundant in the rhizosphere of one genotype or another, but they comprised only a small portion of the total genes identified and did not affect the global rhizosphere metagenomes. Our study shows for the first time that the rhizosphere metagenome of wheat is stable across a wide variety of genotypes when growing under nutrient poor conditions.

Introduction

Wheat breeding programs have been developed to increase productivity and agronomic performance under optimal fertilization and have been implemented without considering how this would alter the root-associated microbiome (Siciliano *et al.*, 1998, Germida & Siciliano, 2001, Paterson *et al.*, 2007). If breeding occurs in a high-input environment that obviates the need for plant or microbial-based ecosystem services that can provide nutrients from the soil, then the traits supporting beneficial rhizosphere interactions may be lost during selection (Emmett *et al.*, 2018). In this context, plant traits related to plant-microbe interactions, like exudation and nutrient trade-offs, can be drastically affected. Indeed, by changing exudation patterns and the nutrient trade-offs with microorganisms, which in turn shape the structure and activities of microbial communities, breeding programs might have adversely influenced plant growth when nutrients are scarce (Aira *et al.*, 2010, Bakker *et al.*, 2012, Bulgarelli *et al.*, 2012, Lundberg *et al.*, 2012).

Although domestication has caused a strong decrease in the genetic diversity of crops (Pérez-Jaramillo *et al.*, 2016), there are multiple reports that showed that different closely related plant genotypes harbor contrasting microbial communities. Among others, such patterns were found for maize (Schmidt *et al.*, 2016), beans (Mendes *et al.*, 2017), willows (Bell *et al.*, 2014, Yergeau *et al.*, 2018) and wheat (Germida & Siciliano, 2001, Okubara *et al.*, 2004, Nelson A. G. *et al.*, 2011, Donn *et al.*, 2015, Mahoney *et al.*, 2017, Azarbad *et al.*, 2018, Azarbad *et al.*, 2020, Valente *et al.*, 2020, Yergeau *et al.*, 2020). These differences are often explained by variations in root physiology e.g. rhizodeposition (Mahoney *et al.*, 2017) or morphology, e.g. root biomass and structure (Okubara *et al.*, 2004, Venter *et al.*, 2004, Nelson *et al.*, 2011, Azarbad *et al.*, 2020). Germida & Siciliano (2001), and Okubara et al. (2004) reported differences between wheat genotypes in the abundance, diversity and root colonization capacity of *Pseudomonas fluorescens*, *Bacillus* spp., *Aureobacter* spp. and *Salmonella* spp. Mahoney *et al.* (2017) found that 24 out of their 1305 most abundant operational taxonomical units (OTUs) varied in frequency in the rhizospheres of different wheat genotypes, some of which could provide beneficial services to the plants such as promoting plant growth (hormones production), and plant and soil health (antifungal and antibiotic metabolites) (Mahoney *et al.*, 2017, Valente *et al.*, 2020).

Although there seems to be differences between the microbial communities of different wheat genotypes, it is difficult to conclude if these shifts result in changes in functions, especially those related to plant nutrition. Interestingly Yergeau *et al.*, (2020) showed a significant difference in the abundance of genes encoding for the archaeal ammonia monooxygenase and for the nitrite reductase between two wheat genotypes, suggesting a difference in the competition for inorganic nitrogen, which was correlated to grain quality and yield. Additionally, Azarbad *et al.* (2018) reported that soil processes such as CO_2 production and H_2 oxidation were influenced by wheat genotype, in interaction with soil water stress history and soil water content.

In this study, we sought to test the hypothesis that modern and ancestral wheat genotypes associate with functionally different microbial communities in their rhizosphere under low fertilization conditions. In order to do so, we performed shotgun metagenomic sequencing on DNA extracted from the rhizosphere of 10 different wheat genotypes (modern and ancestral) grown under low fertilization and assessed the changes in the abundance of genes related to nutrient acquisition and cycling.

Materials and methods

Experimental design

A field experiment was conducted in 2013 at the Nassar Crop Research Farm of the University of Saskatchewan, Saskatoon, Canada. This farm has been managed for more than 50 years to conduct experiments under low fertilization conditions in breeding programs and genotype selection and is commonly used to evaluate wheat line performance. We selected 10 wheat genotypes (year of introduction): Red Fife (1845), Marquis (1911), CDC Teal (1991), AC Barrie (1994), Lillian (2003), CDC Kernen (2009) and CDC Stanley (2009) belonging to the Canada Western Red Spring (CWRS) class (*Triticum aestivum* or bread wheat) and Pelissier (1929), Strongfield (2004) and CDC Verona (2008) belonging to the Canada Western Amber Durum (CWAD) class (*T. turgidum L.* ssp. *durum* or durum wheat) (https://grainscanada.gc.ca/en/grain-quality/grain-grading/wheat-classes.html). The experiment was arranged in a randomized block design with three blocks, each consisting of ten 6.2 m^2 plots to which the cultivars were randomly assigned. Each plot contained eight rows spaced at intervals of 20 cm. On May 25, 2013, all plots were seeded at 320 seeds m^{-2}, which is typical of western Canadian wheat production systems. To minimize the effect of the seed source on plant performance, all cultivars were grown from seed in a common field in a previous year (under low fertilization) and harvested to be used in the present experiment. To obtain a good establishment of the plants in early spring, 15 kg ha^{-1} of 11-55-0 (% N -% P_2O_5 -% K) fertilizer was added at seeding. This application of fertilizer was required to minimize any erroneous effect on productivity measurements due to poor seedling establishment.

Wheat yields and harvest index measurements

Time to maturity (days needed to reach maturity, Zadoks 90 growth stage), and height of each cultivar were measured in each plot. At maturity, the grain yield was evaluated on four rows in each plot, whereas the harvest index was assessed from collecting all aboveground biomass from a 0.5 m^2 quadrat and measuring total biomass weight (dry weight basis). The grains were separated from vegetative tissue to estimate the harvest index as grain weight (g)/total biomass weight (g) x 100. Test weight (TWT: bulk density weight per volume measurement) and thousand kernel weight (TKW, weight of a 1000 seeds) were also measured.

Rhizosphere soil sampling, DNA extraction and metagenomics sequencing

Rhizosphere soil samples were collected on July 2, 2013. Five to eight plants were uprooted from three 13 by 13 cm regions within each plot. Each plant was vigorously shaken, and excess bulk soil was removed by hand from the roots of each plant, until only tightly adhering soil remained. The roots of the 5 to 8 different plants were cut from the rest of the plant and pooled in 200 mL of sterile phosphate buffered saline. After shaking at 150 rpm at 22°C for 25 minutes the roots were removed, and the rhizosphere soil was recovered from the PBS by centrifugation at >2000 × g for 5 minutes. For each of the 30 samples (10 cultivars x 3 blocks), total DNA was extracted from 250 mg of rhizosphere soil using the Power Soil DNA kit (MoBio Laboratory, CA, USA). Libraries for metagenomic analyses were then generated using the Nextera XT DNA Library Prep kit (Illumina) according to the protocol described in the Illumina Nextera® XT DNA Prep Reference Guide (Part # 15031942 Rev. C). Each DNA sample was quantified by fluorescence detection (TECAN safire, Austria) using the kit Quanti-it™ PicoGreen (Invitrogen, ltd., UK) and libraries were pooled in equal volumes before sequencing. The libraries were submitted for sequencing on five lanes of Illumina HiSeq 2000 with a 2 × 100 configuration at the

Centre d'expertise et de service Génome Québec (Montreal, QC, Canada). Raw data sets are available in the NCBI Sequence Read Archive (SRA) under the BioProject accession PRJNA643787.

Bioinformatics

Sequencing raw data (116 Gb) were processed as previously described (Tremblay *et al.*, 2017). The number of reads per sample after each key steps of the analyses is detailed in Supplementary Table 1. Briefly, sequencing adapters were removed from each read and bases at the end of reads having a quality score <30 were cut off (Trimmomatic v0.32) (Bolger *et al.*, 2014) and scanned for sequencing adapters contaminants reads using DUK (unpublished - http://duk.sourceforge.net/) to generate quality controlled (QC) reads. Each QC-passed read from each sample was assembled into a large metagenome assembly using the Ray software v2.3.1 (Boisvert *et al.*, 2012) with a kmer size of 31. Gene prediction on the obtained contigs was performed by calling genes on each assembled contig using MetageneMark v1.0 (Tang & Borodovsky, 2015). Genes were annotated following the JGI's guidelines (Huntemann *et al.*, 2016) using six different databases: 1) RPSBLAST (v2.2.29+) (Camacho *et al.*, 2009) against COG database (v3.11); 2) RPSBLAST (v2.2.29+) against KOG database (v3.11); 3) HMMSCAN (v3.1b1) (Eddy, 2011) against PFAM-A v27.0 database (Finn *et al.*, 2014); 4) TIGRFAM database v15.0; 5) BLASTP (v2.2.29+) against KEGG database (v71.0); and 6) BLASTN (v2.2.29+) against NCBI's nucleotide (nt) database (Li & Durbin, 2009). Contigs (and not genes) sequences were also blasted against NCBI's nt database for taxonomic assignment. For each of these database comparisons, the best hit having at least an e-value ≥ 0.01 was kept for each query. QC-passed reads were mapped (BWA mem v0.7.10) (Li & Durbin, 2010) against contigs to assess the quality of metagenome assembly and to obtain contig abundance profiles. Alignment files in bam format were sorted by read coordinates using SAMtools v1.1 (http://www.htslib.org/doc/samtools-1.0.html) and only properly aligned read pairs were kept for downstream steps. Each bam file (containing properly aligned paired-reads only) was analyzed for coverage of called genes and contigs using bedtools (v2.17.0) (Quinlan & Hall, 2010) using a custom bed file representing gene coordinates on each contig. Only paired-reads both overlapping their contigs or genes were considered for gene counts. Coverage profiles of each sample were merged to generate an abundance matrix (rows = contig, columns = samples) for which a corresponding CPM (Counts Per Million) abundance matrix (edgeR v3.10.2) (Robinson *et al.*, 2010) was generated as well.

Taxonomy of each contig was assigned using the NCBI taxonomy database (Benson *et al.*, 2009, Sayers *et al.*, 2011) (ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz). GenInfo Identifier (GIs) resulting from BLASTN against nt was used to retrieve full taxonomic lineage (when available) from the NCBI taxonomy database. Taxonomic lineages were integrated to the contig abundance of read counts matrix to generate an OTU table format file (with contigs replacing OTUs as rows). Taxonomic summaries were performed using a combination of in-house Perl, R scripts and Qiime v.1.9.0 (Caporaso *et al.*, 2010).

Statistical analysis

All statistical analyses were performed in R v.3.3.1 (R Development Core Team, 2010). Differences between genotypes for yield, harvest indices and the relative abundance of various gene categories were tested using one-way ANOVA and *a posteriori* comparisons using Tukey's HSD test, using the "aov" and "HSD.test" functions from the "agricolae" package (Felipe de Mendiburu, 2016). Decreases in yield related to the effect of the low fertilization treatment were assessed by comparing the yields of the genotypes against expected yields under optimal fertilisation conditions, as stated in official genotype descriptions (Table 1). The similarity between samples due to the relative abundance of genes, contigs and bins data was visualized by principal coordinate analysis (“cmdscale” function) based on Bray-Curtis dissimilarity matrices (“vegdist”) function of the “vegan” package (Oksanen *et al.*, 2013). The effects of genotypes and blocks on the community composition was tested by permutational multivariate analysis of variance (PERMANOVA) using the “adonis” function of the “vegan” package. To correlate the relative abundance of genes implicated in the nitrogen and phosphorous cycles to yields, Spearman's ρ (rho) correlation analyses were performed using the "cor" function of the “stats” package. Associations between genes and a particular genotype were tested using the function “multipatt” from the “indicspecies” package (Cáceres & Legendre, 2009) by the analysis of ecological preferences based on correlation indices with 999 permutation, using confidence levels of 0.001. One-way ANOVA analyses were performed with the function “aov” to detect differences between genotypes regarding the relative abundance of functional genes. For ANOVA, Permanova and correlation analyses, P-values were corrected using the Bonferroni method when appropriate.

Results

Wheat yields and other agronomic parameters

The average grain yields under low fertilization were significantly different between the genotypes tested (F = 12.95, P = 0.00066) (Table 1). Results showed that the durum genotypes Strongfield, CDC Verona and Pelissier had significantly higher yields, by up to 20%, than most bread wheat genotypes, with the notable exception of the Red Fife genotype. This latter genotype, the most ancient one, ranked third for yield, well above all other bread wheat genotypes. The other bread wheat genotypes, CDC Teal, Lillian, CDC Kernen, AC Barrie and CDC Stanley had similar, intermediate yields. Finally, the Marquis genotype had the lowest yield, 28% less than the Strongfield and CDC Verona genotypes (Table 1). As expected, grain yields were nearly significantly lower under the low fertilization conditions used here as compared to reference values (paired t-test, t= -1.96, P= 0.08076). Unexpectedly, the only exception was the Red Fife genotype which was 22% more productive under lower fertilization than under optimal.

Other agronomic parameters such as plant height (HT), days to maturity (MAT), and quality test such as test weight (TWT) and thousand kernel weight (TKW) were also measured (Supplementary Table 2). All the parameters were significantly different between the genotypes ($P<0.001$) except for TWT. The most ancient cultivars Red fife, Pelissier and Marquis were taller, highlighting the selection for shorter plants less prone to lodging in more recent genotypes. Regarding the values of days to maturity (MAT), the oldest genotypes Red Fife and Pelissier showed longer times to maturity, with the newer genotypes showing medium to early maturity and Strongfield having the earliest maturity. Selection for early maturity is important for the regions where the growing season is short and late-maturing genotypes are often badly damaged due to frost (McCallum & DePauw, 2008). Thousand kernel weight values were higher for Pelissier, followed by Verona, Strongfield and Red Fife. Generally, the newer bread wheat genotypes showed lower values for TKW. Comparison with reference values under optimal fertilization showed significant differences for TWT (P=0.002), but not for HT, MAT, and TKW (Supplementary Table 2).

Differences in the metagenomic community composition between genotypes

We assembled 5,480,054 contigs containing a total of 5,996,993 genes that were classified in 288 genomic bins. Based on the taxonomic affiliation of the genes, the gene relative abundance table was separated into two subsets, one containing 1,515,515 bacterial genes and the other 25,362 fungal genes, regardless of whether the gene was functionally annotated or not. The similarity between the 30 samples based on these subsets was visualized using principal coordinate analysis (PCoA) ordinations based on Bray-Curtis dissimilarity (Figure 1). There was no clear grouping of genotypes, both for the bacterial and the fungal datasets, with a large variability between samples from different experimental blocks. Permanova confirmed this visual interpretation, with a lack of significant differences in the bacterial and fungal related gene relative abundance patterns between the genotypes, and a significant block effect. *Proteobacteria*, *Actinobacteria* and *Bacteroidetes* were the dominant bacterial phyla whereas the phylum *Ascomycota* dominated among the fungal genes.

Differences in the relative abundance of specific functional genes between genotypes

We searched the annotation of genes for pathways related to macronutrient (N, P, and K) and to plant-microbe interactions. We kept all genes, even if some had an annotation pointing toward two or more different pathways. In total, 217,883 genes (3.63 % of all genes) were assigned to nitrogen cycling related pathways (KEGG orthology reference pathways ko00910, ko00250, ko00680, ko00630, ko00380, ko02020, ko00071, ko00627, ko00360, ko00340, ko00260, ko00020, ko00010, ko00380, ko00330, ko00270): most of these pathways were related to two component systems (sensor-response regulation) and amino acid metabolism, such as glutamate, glutamine, tryptophan, histidine, cysteine, alanine and GABA. For phosphorus, 4,363 genes were assigned to the P starvation pathway ko02020 (K07636, K07768) corresponding also to two component regulation systems *PhoR-PhoB* and *SenX3-RegX3*, and 12,518 genes were assigned to P assimilation pathways (ko00627, ko00030) coding for alkaline/acid phosphatase and quinoprotein glucose dehydrogenase, respectively. For potassium, 490 genes were assigned to the K^+ transport pathway ko02020 (K01546, K01547, K01548, K01545) corresponding to two component systems of ATPases, namely, *kdpA*, *kdpB*, *kdpC* and *kdpF*. For pathways related to plant-microbe interactions, 25 genes were designated for 3-Indol Acetic Acid (IAA) pathways, 12 of them assigned to tryptophan metabolism, ko00380 (K00466), related to hypothetical or several gene affiliations, and the remaining 13 were assigned to the pathway K12940, specific to the gene *abgA* which codes for the utilization of aminobenzoyl-glutamate. Finally, 161 genes were related to the ACC deaminase pathway, ko00640 (K01505), coding for the enzyme ACC deaminase.

The genes mentioned above were grouped into nine categories, for which significance of genotype effect was tested using two complementary approaches. First, the effect of wheat genotypes on the sum of the relative abundance for all genes within a category was tested using ANOVA. Although there were some variations between genotypes (Figure 2), these variations were not significant in ANOVA tests. Secondly, the effect of genotypes on the structure of the gene table containing all the genes in a category was tested using Permanova based on Bray-Curtis dissimilarity. This analysis also revealed no significant differences between the genotypes.

Similarly, when looking at the genes individually and testing the effect of genotypes on their relative abundance using ANOVA, very few genes showed significant patterns. For the genes related to K, the only gene significantly different among genotypes was a K+ transporting ATPase (*kdp* sub-unit C) for which the taxonomy was unresolved. For N, only one gene, a nitrogenase component 1 type oxidoreductase (*nifK*) affiliated with the *Rhodocyclaceae*, was identified as varying significantly between genotypes. All the other genes from the categories mentioned above were not significantly different between the genotypes.

Differences in the relative abundance of all genes between genotypes

In addition to the selected functional genes involved in nutrient cycling or plant-microbe interactions, we also used the entire gene dataset to 1) identify genes that were significantly affected by the wheat genotypes using ANOVA, and 2) identify genes strongly associated with one or the other wheat genotype using indicator “species” analysis. ANOVA revealed 57 genes being highly significantly affected by genotypes (at a Bonferroni corrected α_B=8.33E-09) (Supplementary Table 3). For more than half of these genes, function and taxonomic affiliations were unknown. To identify associations between single genotype and specific genes, indicator “species” analysis was carried out on the subsets of genes affiliated to bacteria and fungi. There were more significant indicator genes for bacteria than for fungi, and significance was generally much higher. For instance, 100 bacterial genes were significant indicators for one or the other genotype at a $P< 0.001$, whereas only two fungal genes were significant at this level. The genotypes AC Barrie, Strongfield and CDC Teal had the highest number of bacterial gene indicators (Table 2). The indicator genes belonged to a wide variety of bacteria, some of which, such as *Sinorhizobium* (CDC Teal), *Azospirillum* (CDC Teal), *Mesorhizobium* (Strongfield), *Nitrosomonas* (Strongfield), *Frankia* (Red Fife), *Natronococcus* (Red Fife), and *Nitrospira* (AC Barrie) are involved in the nitrogen cycle (Table 2). In terms of function, many indicator genes could not be identified through our homology search in all major databases (Table 2). The indicator genes that could be identified were related to various functions, some of which could be related to nutrient cycling or life in the rhizosphere, such as the sulfur cycle (Marquis), amino acid metabolism (many genotypes), resistance to antibiotics (Pelissier and CDC Teal), synthesis of osmoprotectants (Strongfield and CDC Teal) and vitamin biosynthesis (CDC Kernen and AC Barrie). However, the 57 genes identified by ANOVA and the 102 genes identified by indicator species analysis made up an insignificant portion of the entire gene dataset (5,996,993 genes).

Correlation between yields and functional genes

Using Bonferroni correction for multiple testing, we could not find any significant correlations between the relative abundance of functional genes involved in nutrient cycling and plant-microbe interactions and the wheat yields.

Discussion

Many authors have hypothesized that breeding of wheat genotypes under high nutrient inputs would result in a lowered capacity to recruit and maintain association with rhizosphere microorganisms involved in the cycling of nutrients (Siciliano *et al.*, 1998, Germida & Siciliano, 2001, Paterson *et al.*, 2007). Since most breeding programs for grain crops have not specifically targeted belowground traits, the agroecological context in which and for which genotypes are bred is thought to determine the selection pressure on these traits (Schmidt *et al.*, 2016). In the rhizosphere, positive species interactions are more likely to emerge and be maintained in nutrient-poor environments as nutrient enrichment has the potential to reduce the nutrient limitations that make mutualists beneficial (Verbruggen & Toby Kiers, 2010, Emmett *et al.*, 2018). Fertilizer inputs can make microbial interactions costly and even parasitic for crops under field conditions, as it was shown in the case of mycorrhizal fungi (Kiers *et al.*, 2002, Ryan *et al.*, 2005, Kiers & Denison, 2008). When exposed to high nutrient levels, plants may severely decrease or cease the resource allocation to their roots, thus to their microbial partners, resulting in microbial community shifts as competition for limited carbon resources increases (Verbruggen & Kiers, 2010). This would ultimately result in assembly patterns mainly driven by soil type and climatic conditions (Kiers & Denison, 2008). As such, the continuous selection for yields under variable environmental conditions and under optimal fertilization regimes imposes general patterns of root activity and rhizosphere C flows that do not allow rhizosphere microbial communities to diverge. It would therefore be expected that modern high-yield genotypes would show similar microbial associations, different from the ancestral genotypes, which would mainly be driven by environmental factors, and that they would be much less productive in soils with lower nutrient concentrations.

Under low fertilizer inputs, our results did show a significant decrease in yields for almost all genotypes as compared to reference values, except for the most ancient bread wheat genotype, Red Fife. However, the trends in the yield losses under low nutrient conditions were not reflected in the rhizosphere metagenome. Indeed, we did not find any significant differences between the genotypes in the general gene community structure, with some significant shifts for a minority of genes. In our study, the metagenome of the rhizosphere was strongly affected by variation in the soil across the field, as confirmed by the strong block effect observed for many parameters. Our results are in sharp contrast with recent publications that highlighted that different closely related wheat genotypes have contrasting microbial communities and functional gene abundance in their rhizosphere. A recent field experiment showed that two modern genotypes of wheat harbored significantly different abundance of functional genes related to the N-cycle (based on qPCR), across fertilization treatments that ranged from 0 to 120 NH_4NO_3 kg/ha (Yergeau *et al.*, 2020). These differences were mirrored in the contrasting grain yields and quality between the two genotypes (Yergeau *et al.*, 2020). However, the differences between the genotypes were often dwarfed by the variation between the two fields sampled (Yergeau *et al.*, 2020). Similarly, a pot experiment has shown that four modern wheat genotypes harbored significantly different microbial communities (based on 16S rRNA gene and ITS region amplicon sequencing), but that this effect was stronger inside the roots and leaves than in the rhizosphere, and, in all cases, a distant second to the effect of soil water stress history (Azarbad *et al.*, 2020). Furthermore, in the same pot experiment, it was also shown that microbial-driven processes and bacterial and fungal abundance in the rhizosphere of wheat were significantly influenced by genotype (Azarbad *et al.*, 2018). Taken together, these studies indicated that different plant genotypes do harbor significantly different microbial communities, which results in different activities and process rates, but that this effect is often rather subtle and varies with soil type and plant compartment.

In contrast to our results, differences between genotypes were previously reported for other plants. For instance, in a field study using willows, the genotype was shown to significantly influence the rhizosphere microbial communities, but only for fungi and when willows were growing under high contaminant stress (Bell *et al.*, 2014). Similarly, in the same field study, the willow genotypes that were the least tolerant to soil contaminant stress also showed the largest shifts in the metatranscriptome of their rhizosphere following contamination (Yergeau *et al.*, 2018). These results are difficult to compare to our study, as stress appeared to strengthen the genotype effect on soil microbial communities and the nutrient limitation imposed here was probably not as stressful as the presence of contaminants. Some studies have shown significant differences between the microbial communities associated with wild and domesticated genotypes of beans (Pérez-Jaramillo *et al.*, 2017, Pérez-Jaramillo *et al.*, 2019), barley (Bulgarelli *et al.*, 2015), sunflower (Leff *et al.*, 2017) and sugar beets (Zachow *et al.*, 2014). Other studies have shown that the geographical origin of the genotype had an influence on the capacity to associate with particular soil fungi (Bell *et al.*, 2014). However, here, our genotypic gradient was relatively short, spanning only approximately 100 years of wheat breeding in Canada for two major lines, and we did not include wild or foreign representatives in the comparison. Alternatively, the use of shotgun metagenomics could explain part of the discrepancies between our study and previous work. Indeed, the taxonomic shifts observed in previous studies using amplicon sequencing will not necessarily result in functional shifts when using shotgun metagenomics, as previously shown for plant associated microbial communities (Louca *et al.*, 2016). This is probably caused by the high functional redundancy among soil microorganisms. Soil shotgun metagenomics reads are also widely dominated by bacteria, and only a minor part of the reads is normally associated with fungi, which reduces the resolution. As fungi are often more sensitive to a genotype effect (Bell *et al.*, 2014, Azarbad *et al.*, 2020, Yergeau *et al.*, 2020), this limitation of shotgun metagenomics could partly explain the lack of significance of the genotype effect.

Nonetheless, some genes were identified as particularly linked to certain genotypes using indicator “species” analysis. Although some of these genes were annotated as encoding for proteins potentially important in nutrient cycling and in plant-microbe interaction, no clear trend emerged regarding a difference between ancient vs. modern genotypes. The genotypes with the most indicator genes were modern genotypes (Strongfield, AC Barrie and CDC Teal), suggesting that modern genotypes might associate with a more unique set of microorganisms and associated functional genes than ancestral genotypes. In all cases, the small and subtle differences found between genotypes were too small to cause significant shifts in the rhizosphere metagenome when taken in its entirety. However, it is an open question as to whether these small changes would result in significant effects on wheat nutrient acquisition.

In summary, using shotgun metagenomics we have found only a handful of significant differences between modern and ancestral wheat genotypes grown in the same field under limiting nutrient conditions. We cannot therefore reject the null hypothesis that 100+ years of Canadian wheat breeding has not changed the rhizosphere microbial functional potential. This is the first time that this hypothesis was tested using a shotgun metagenomic approach, and it was in sharp contrast to previous culture-based and amplicon sequencing studies. In view of the enormous soil microbial diversity and its functional redundancy, it is not unlikely that taxonomic shifts could occur without concomitant shifts in the overall functions.

Acknowledgments

This work was supported by the Wheat Flagship Program of the National Research Council Canada and by a Natural Sciences and Engineering Research Council Discovery grant to EY (RGPIN-2014-05274). We would like to thank Sylvie Sanschagrin and Brenda Haug for technical help in the laboratory and during sampling. We acknowledge Compute Canada for access to the Guillemin (McGill University) and Graham (University of Waterloo) systems.

References

Aira M, Gómez-Brandón M, Lazcano C, Bååth E & Domínguez J (2010) Plant genotype strongly modifies the structure and growth of maize rhizosphere microbial communities. *Soil Biol Biochem* **42**: 2276-2281.

Azarbad H, Constant P, Giard-Laliberté C, Bainard LD & Yergeau E (2018) Water stress history and wheat genotype modulate rhizosphere microbial response to drought. *Soil Biol Biochem* **126**: 228-236.

Azarbad H, Tremblay J, Giard-Laliberté C, Bainard L & Yergeau E (2020) Four decades of soil water stress history together with host genotype constrain the response of the wheat microbiome to soil moisture. *FEMS Microbiol Ecol*.

Bakker M, Manter D, Sheflin A, Weir T & Vivanco J (2012) Harnessing the rhizosphere microbiome through plant breeding and agricultural management. *Plant Soil* **360**: 1-13.

Bell TH, El-Din Hassan S, Lauron-Moreau A, Al-Otaibi F, Hijri M, Yergeau E & St-Arnaud M (2014) Linkage between bacterial and fungal rhizosphere communities in hydrocarbon-contaminated soils is related to plant phylogeny. *ISME J* **8**: 331-343.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J & Sayers EW (2009) GenBank. *Nucleic Acids Res* **37**: D26-D31.

Boisvert S, Raymond F, Godzaridis É, Laviolette F & Corbeil J (2012) Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol* **13**: R122.

Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*.

Bulgarelli D, Garrido-Oter R, Münch PC, Weiman A, Dröge J, Pan Y, McHardy AC & Schulze-Lefert P (2015) Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell host & microbe* **17**: 392-403.

Bulgarelli D, Rott M, Schlaeppi K, *et al.* (2012) Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* **488**: 91-95.

Cáceres MD & Legendre P (2009) Associations between species and groups of sites: indices and statistical inference. *Ecology* **90**: 3566-3574.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K & Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD & Costello EK (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335-336.

Clarke JM, McCaig TN, DePauw RM, Knox RE, Clarke FR, Fernandez MR & Ames NP (2005) Strongfield durum wheat. *Canadian Journal of Plant Science* **85**: 651-654.

Cornell University SoIPS (2016) 2012-15 OREI Organic Spring Wheat Trial. *Cornell Cals, College of Agriculture and Life Science* **Plant Breeding and Genetic Section**.

Cuthbert RD, DePauw RM, Knox RE, Singh AK, McCaig TN, McCallum B & Fetch T (2017) AAC Brandon hard red spring wheat. *Canadian Journal of Plant Science* **97**: 393-401.

DePauw RM, Townley-Smith TF, Humphreys G, Knox RE, Clarke FR & Clarke JM (2005) Lillian hard red spring wheat. *Canadian Journal of Plant Science* **85**: 397-401.

Donn S, Kirkegaard JA, Perera G, Richardson AE & Watt M (2015) Evolution of bacterial communities in the wheat crop rhizosphere. *Environ Microbiol* **17**: 610-621.

Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**: e1002195.

Emmett BD, Buckley DH, Smith ME & Drinkwater LE (2018) Eighty years of maize breeding alters plant nitrogen acquisition but not rhizosphere bacterial community composition. *Plant Soil* **431**: 53-69.

Felipe de Mendiburu (2016) Agricolae: Statistical Procedures for Agricultural Research. p.^pp.

Finn RD, Bateman A, Clements J*, et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res* **42**: D222-D230.

Germida J & Siciliano S (2001) Taxonomic diversity of bacteria associated with the roots of modern, recent and ancient wheat cultivars. *Biol Fertil Soils* **33**: 410-415.

Hughes GR & Hucl P (1993) CDC Teal hard red spring wheat. *Canadian Journal of Plant Science* **73**: 193-197.

Huntemann M, Ivanova NN, Mavromatis K, *et al.* (2016) The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). *Standards in Genomic Sciences* **11**: 17.

Kiers ET & Denison RF (2008) Sanctions, Cooperation, and the Stability of Plant-Rhizosphere Mutualisms. *Annu Rev Ecol Evol Syst* **39**: 215-236.

Kiers ET, West SA & Denison RF (2002) Mediating mutualisms: farm management practices and evolutionary changes in symbiont co-operation. *J Appl Ecol* **39**: 745-754.

Leff JW, Lynch RC, Kane NC & Fierer N (2017) Plant domestication and the assembly of bacterial and fungal communities associated with strains of the common sunflower, Helianthus annuus. *New Phytol* **214**: 412-423.

Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Li H & Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**: 589-595.

Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, Farjalla VF & Doebeli M (2016) High taxonomic variability despite stable functional structure across microbial communities. *Nature Ecology & Evolution* **1**: 0015.

Lundberg DS, Lebeis SL, Paredes SH, *et al.* (2012) Defining the core Arabidopsis thaliana root microbiome. *Nature* **488**: 86-90.

Mahoney AK, Yin C & Hulbert SH (2017) Community Structure, Species Variation, and Potential Functions of Rhizosphere-Associated Bacteria of Different Winter Wheat (Triticum aestivum) Cultivars. *Frontiers in Plant Science* **8**.

McCallum BD & DePauw RM (2008) A review of wheat cultivars grown in the Canadian prairies. *Canadian Journal of Plant Science* **88**: 649-677.

Mendes LW, Raaijmakers JM, de Hollander M, Mendes R & Tsai SM (2017) Influence of resistance breeding in common bean on rhizosphere microbiome composition and function. *The Isme Journal* **12**: 212.

Nelson A. G., Quideau S., Frick B., Niziol D., Clapperton J. & D.. S (2011) Spring wheat genotypes differentially alter soil microbial communities and wheat breadmaking quality in organic and conventional systems. *Canadian Journal of Plant Science* **91**: 485-495.

Nohno T, Noji S, Taniguchi S & Saito T (1989) The narX and narL genes encoding the nitrate-sensing regulators of Escherichia coli are homologous to a family of prokaryotic two-component regulatory genes. *Nucleic Acids Res* **17**: 2947-2957.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R, Simpson GL, Solymos P, Stevens MHH & Wagner H (2013) Package 'vegan'. *Community ecology package, version* **2**.

Okubara PA, Kornoely JP & Landa BB (2004) Rhizosphere colonization of hexaploid wheat by Pseudomonas fluorescens strains Q8r1-96 and Q2-87 is cultivar-variable and associated with changes in gross root morphology. *Biol Control* **30**: 392-403.

Paterson E, Gebbing T, Abel C, Sim A & Telfer G (2007) Rhizodeposition shapes rhizosphere microbial community structure in organic soil. *New Phytol* **173**: 600-610.

Pawlowski K, Klosse U & de Bruijn FJ (1991) Characterization of a novel Azorhizobium caulinodans ORS571 two-component regulatory system, NtrY/NtrX, involved in nitrogen fixation and metabolism. *Molecular and General Genetics MGG* **231**: 124-138.

Pérez-Jaramillo JE, Mendes R & Raaijmakers JM (2016) Impact of plant domestication on rhizosphere microbiome assembly and functions. *Plant Mol Biol* **90**: 635-644.

Pérez-Jaramillo JE, de Hollander M, Ramírez CA, Mendes R, Raaijmakers JM & Carrión VJ (2019) Deciphering rhizosphere microbiome assembly of wild and modern common bean (Phaseolus vulgaris) in native and agricultural soils from Colombia. *Microbiome* **7**: 114.

Pérez-Jaramillo JE, Carrión VJ, Bosse M, Ferrão LFV, de Hollander M, Garcia AAF, Ramírez CA, Mendes R & Raaijmakers JM (2017) Linking rhizosphere microbiome composition of wild and domesticated Phaseolus vulgaris to genotypic and root phenotypic traits. *The ISME Journal* **11**: 2244-2257.

Pozniak CJ, Fox SL & Knott DR (2009) CDC Verona durum wheat. *Canadian Journal of Plant Science* **89**: 321-324.

Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

R Development Core Team (2010) R: A language and environment for statistical compuring. R Foundation for statistical computing, Viena, Austria.

Robinson MD, McCarthy DJ & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Ryan MH, Van Herwaarden AF, Angus JF & Kirkegaard JA (2005) Reduced growth of autumn-sown wheat in a low-P soil is associated with high colonisation by arbuscular mycorrhizal fungi. *Plant Soil* **270**: 275-286.

Saskatchewan Crop Insurance Corporation (2018) Wheat - Hard Red Spring Provincial Average Yields by Variety.

Sayers EW, Barrett T, Benson DA, *et al.* (2011) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **39**: D38-D51.

Schmidt JE, Bowles TM & Gaudin ACM (2016) Using Ancient Traits to Convert Soil Health into Crop Yield: Impact of Selection on Maize Root and Rhizosphere Function. *Frontiers in Plant Science* **7**.

Schulz AA, Collett HJ & Reid SJ (2001) Nitrogen and carbon regulation of glutamine synthetase and glutamate synthase in Corynebacterium glutamicum ATCC 13032. *FEMS Microbiol Lett* **205**: 361-367.

Siciliano SD, Theoret CM, de Freitas JR, Hucl PJ & Germida JJ (1998) Differences in the microbial communities associated with the roots of different cultivars of canola and wheat. *Can J Microbiol* **44**: 844-851.

Slinkard AE, Knott, Douglas R. (1995) Harvest of Gold : The History of Field Crop Breeding in Canada. *Saskatoon, Saskatchewan: University Extension Press, University of Saskatchewan* 367.

Tang S & Borodovsky M (2015) Ab Initio Gene Identification in Metagenomic Sequences. *Encyclopedia of Metagenomics: Genes, Genomes and Metagenomes: Basics, Methods, Databases and Tools*,(Nelson KE, ed.) p.^pp. 13-19. Springer US, Boston, MA.

Tremblay J, Yergeau E, Fortin N, Cobanli S, Elias M, King TL, Lee K & Greer CW (2017) Chemical dispersants enhance the activity of oil- and gas condensate-degrading marine bacteria. *The ISME Journal* **11**: 2793-2808.

Valente J, Gerin F, Le Gouis J, Moënne-Loccoz Y & Prigent–Combaret C (2020) Ancient wheat varieties have a higher ability to interact with plant growth-promoting rhizobacteria. *Plant Cell Environ* **43**: 246-260.

Venter JC, Remington K, Heidelberg JF, *et al.* (2004) Environmental shotgun sequencing of the Sargasso Sea. *Science* **304**.

Verbruggen E & Toby Kiers E (2010) Evolutionary ecology of mycorrhizal functional diversity in agricultural systems. *Evolutionary Applications* **3**: 547-560.

Yergeau E, Tremblay J, Joly S, Labrecque M, Maynard C, Pitre FE, St-Arnaud M & Greer CW (2018) Soil contamination alters the willow root and rhizosphere metatranscriptome and the root–rhizosphere interactome. *The ISME Journal* **12**: 869-884.

Yergeau É, Quiza L & Tremblay J (2020) Microbial indicators are better predictors of wheat yield and quality than N fertilization. *FEMS Microbiol Ecol* **96**.

Zachow C, Müller H, Tilcher R & Berg G (2014) Differences between the rhizosphere microbiome of Beta vulgaris ssp. maritima-ancestor of all beet crops-and modern sugar beets. *Frontiers in microbiology* **5**: 415-415.

Figure legends

Figure 1. Microbial community composition. Principal coordinate analysis based on Bray-Curtis dissimilarity of gene relative abundances for A) genes affiliated with fungi, and B) genes affiliated with bacteria. Colors represent the genotypes, and blocks are identified by shapes.

Figure 2. Summed relative abundances of the functional genes related to A) phytohormones and B) N and P cycling. Error bars represent the standard error.

Table 1. Yields of the ten wheat genotypes under low fertilization regime, compared to reference values under optimal fertilization levels.

Genotype (Release year)	Yield (kg ha^{-1})		Performance change %	Source data
	Low fertilization (2013) Average ±std. dev.	Optimal fertilization		
Strongfield (2004)	2460 ± 208 a	4030	-39	(Clarke *et al.*, 2005)
CDC Verona (2008)	2456 ± 280 a	3840	-36	(Pozniak *et al.*, 2009)
Red Fife (1845)	2310 ± 193 ab	1877	+22	(Cornell University, 2016)
Pelissier (1929)	2289 ± 322 ab	N/A	-	
CDC Teal (1991)	1990 ± 160 bc	3004	-34	(Hughes & Hucl, 1993, Saskatchewan Crop Insurance Corporation, 2018)
Lillian (2003)	1984 ± 300 bc	3594	-45	(DePauw *et al.*, 2005)
CDC Kernen (2009)	1969 ± 379 bc	4731	-58	(Cuthbert *et al.*, 2017)
AC Barrie (1994)	1931 ± 354 bc	3509	-45	(DePauw *et al.*, 2005)
CDC Stanley (2009)	1907 ± 223 bc	3684	-48	(Slinkard, 1995, Saskatchewan Crop Insurance Corporation, 2018)
Marquis (1911)	1773 ± 242 c	2289	-23	(Cornell University, 2016)

Different letters indicate significant differences at $\alpha<0.05$ according Tukey Honestly Significant Difference post-hoc test.

Table 2. Functional and taxonomic affiliation of significant (P<0.001) bacterial genes identified by indicator species analysis.

gene_id	KEGG ID	Function	Stat	Species
AC Barrie (33)				
Gene ID 3999372	NULL	NULL	0.830	Unknown
Gene ID 2035590	K00936	NULL	0.803	*Desulfurispirillum indicum*
Gene ID 3251592	K02988	Ribosome, bacteria==Ribosome, archaea	0.798	*Mycobacterium sp. JDM601*
Gene ID 5786201	NULL	NULL	0.762	*Bordetella bronchiseptica*
Gene ID 2251928	NULL	NULL	0.750	uncultured bacterium
Gene ID 284465	K02004	Putative ABC transport system	0.741	uncultured bacterium contig00351
Gene ID 3843697	K11690	NULL	0.733	uncultured marine bacterium HF10_25F10
Gene ID 5704810	K03695	NULL	0.729	*Burkholderia sp. YI23*
Gene ID 3992123	K03924	NULL	0.725	*Candidatus Nitrospira defluvii*
Gene ID 234626	K01434	NULL	0.707	*Burkholderia mallei*
Gene ID 2561106	NULL	NULL	0.703	*Pleurocapsa minor*
Gene ID 3660176	K00937	NULL	0.703	*Bartonella clarridgeiae*
Gene ID 4802338	K03518	NULL	0.693	*Thermanaerovibrio acidaminovorans*
Gene ID 2129591	NULL	NULL	0.693	*Collimonas sp. MPS11E8*
Gene ID 4504727	K00540	NULL	0.685	uncultured bacterium contig00184
Gene ID 5227558	K09930	NULL	0.681	Unknown
Gene ID 842716	K08884	NULL	0.678	Unknown
Gene ID 1426909	K02014	NULL	0.675	*Gemmatimonas aurantiaca*
Gene ID 1881505	K10947	NULL	0.674	uncultured bacterium contig00449
Gene ID 5836336	K08884	NULL	0.669	Unknown
Gene ID 679771	K03296	NULL	0.657	uncultured bacterium contig00128
Gene ID 3242231	NULL	NULL	0.650	*Streptomyces tendae*
Gene ID 2368097	K15923	NULL	0.647	Unknown

Gene ID 2624689	K11690	NULL	0.646	*Leptothrix cholodnii*
Gene ID 2320612	NULL	NULL	0.633	Unknown
Gene ID 2623136	K07714	AtoS-AtoC (cPHB biosynthesis) two-component regulatory system	0.629	*Acidiphilium multivorum*
Gene ID 4919969	K02291	beta-Carotene biosynthesis, GGAP => beta-carotene	0.627	uncultured bacterium
Gene ID 5836775	K02470	NULL	0.615	*Amycolatopsis pretoriensis*
Gene ID 2723931	K02014	NULL	0.604	*Corallococcus coralloides*
Gene ID 5511188	NULL	NULL	0.597	uncultured bacterium contig00095
Gene ID 4538019	K01950	NAD biosynthesis, aspartate => NAD	0.588	*Methylibium petroleiphilum*
Gene ID 3734199	K01999	Branched-chain amino acid transport system	0.551	*Syntrophobacter fumaroxidans*
Gene ID 3670026	K09810	Lipoprotein-releasing system	0.541	*Agrobacterium sp.* H13-3
CDC Kernen (11)				
Gene ID 2653112	K02004	Putative ABC transport system	0.772	Unknown
Gene ID 1630857	K15670	NULL	0.773	*Sorangium cellulosum*
Gene ID 722943	NULL	NULL	0.774	*[Cellvibrio] gilvus*
Gene ID 4937480	NULL	NULL	0.775	*Sphingobium japonicum*
Gene ID 3067392	K03529	NULL	0.776	*Catenulispora acidiphila*
Gene ID 1129186	K08884	NULL	0.777	*Rhodobacter sphaeroides*
Gene ID 3709603	K13641	NULL	0.778	Unknown
Gene ID 942139	K00928	Lysine biosynthesis, succinyl-DAP pathway, aspartate => lysine==Methionine biosynthesis.	0.779	*Ramlibacter tataouinensis*
Gene ID 5103385	K00344	NULL	0.780	*Serratia marcescens*
Gene ID 536020	K02529	NULL	0.781	*Saccharomonospora viridis*
Gene ID 191875	K00798	Cobalamin biosynthesis, cobinamide => cobalamin	0.782	*Methylobacterium extorquens*
CDC Stanley (3)				

Gene ID 5906598	K03724	NULL	0.739	*Burkholderia sp. KJ006*
Gene ID 4279595	K02051	NitT/TauT family transport system	0.585	*Pseudomonas fluorescens*
Gene ID 3926067	K00024	Citrate cycle (TCA cycle, Krebs cycle)==Dicarboxylate-hydroxybutyrate cycle	0.581	*Rhodothermus marinus*
CDC Teal (18)				
Gene ID 4227965	K04761	NULL	0.846	*Burkholderia sp. YI23*
Gene ID 1299618	K10943	FlrB-FlrC (polar flagellar synthesis) two-component regulatory system	0.815	*Pseudovibrio sp. FO-BEG1*
Gene ID 1529777	K07690	EvgS-EvgA (acid and drug tolerance) two-component regulatory system	0.776	*Edwardsiella tarda*
Gene ID 1273055	K03501	NULL	0.756	Unknown
Gene ID 5425221	NULL	NULL	0.735	Unknown
Gene ID 2884903	NULL	NULL	0.715	*Corynebacterium jeikeium*
Gene ID 5512026	K02952	Ribosome, bacteria==Ribosome, archaea	0.708	*Brevundimonas subvibrioides*
Gene ID 5560261	K13950	NULL	0.701	*Sinorhizobium medicae*
Gene ID 975665	K00099	C5 isoprenoid biosynthesis, non-mevalonate pathway	0.698	*Sodalis glossinidius*
Gene ID 5082960	K01784	Nucleotide sugar biosynthesis, eukaryotes. Leloir pathway, galactose => alpha-D-glucose-1P	0.666	*Desulfotomaculum kuznetsovii*
Gene ID 4607043	K00145	Ornithine biosynthesis, glutamate => ornithine	0.664	uncultured Acidobacteria bacterium cosmid p2H8
Gene ID 1603508	K03272	ADP-L-glycero-D-manno-heptose biosynthesis	0.655	*Streptomyces verticillus*
Gene ID 5283033	K01090	NULL	0.606	*Salinispora arenicola*
Gene ID 399606	K08884	NULL	0.606	*Microbacterium testaceum*
Gene ID 948818	K00333	NADH:quinone oxidoreductase, prokaryotes	0.566	*Salinispora arenicola*

Gene ID 4341080	K02779	PTS system, glucose-specific II component	0.564	*Azospirillum lipoferum*
Gene ID 1190455	K03296	NULL	0.519	Candidatus *Solibacter usitatus*
CDC Verona (1)				
Gene ID 2974755	K01697	Methionine degradation==Cysteine biosynthesis, homocysteine + serine => cysteine	0.643	*Variovorax paradoxus*
Lilian (1)				
Gene ID 4087572	K15975	NULL	0.626	*Symbiobacterium thermophilum*
Marquis (9)				
Gene ID 3898574	K02004	Putative ABC transport system	0.863	Unknown
Gene ID 2708209	K02048	Sulfate transport system	0.751	Unknown
Gene ID 915210	NULL	NULL	0.713	Unknown
Gene ID 177898	K01999	Branched-chain amino acid transport system	0.669	*Rhodoferax ferrireducens*
Gene ID 4402551	K07807	NULL	0.650	*Deinococcus radiodurans*
Gene ID 5990271	K01797	NULL	0.615	*Ralstonia syzygii*
Gene ID 1881382	K02479	NULL	0.599	Unknown
Gene ID 3559744	K13924	CheA-CheYBV (chemotaxis) two-component regulatory system	0.570	Unknown
Gene ID 5662970	K06951	NULL	0.542	*Niastella koreensis*
Pelissier (6)				
Gene ID 3975373	K01262	NULL	0.777	*Streptomyces lavendulae*
Gene ID 1835174	K13893	Microcin C transport system	0.734	*Rhodopseudomonas palustris*
Gene ID 5746516	K10001	Glutamate/aspartate transport system	0.721	Unknown
Gene ID 4987075	NULL	NULL	0.696	*Gramella forsetii*
Gene ID 5442032	K07147	NULL	0.611	*Pelagibacterium halotolerans*

Gene ID 831402	K01768	NULL	0.477	*Pseudonocardia dioxanivorans*
Red Fife (3)				
Gene ID 4626467	K03092	NULL	0.664	*Pirellula staleyi*
Gene ID 1920270	K04091	NULL	0.611	*Frankia sp. EAN1pec*
Gene ID 2700809	K01848	Hydroxypropionate-hydroxybutylate cycle==3-Hydroxypropionate bi-cycle	0.510	*Natronococcus occultus*
Strongfield (25)				
Gene ID 1463806	K13924	CheA-CheYBV (chemotaxis) two-component regulatory system	0.75	Unknown
Gene ID 3789738	K09458	Fatty acid biosynthesis, elongation==Pimeloyl-ACP biosynthesis, BioC-BioH pathway, malonyl-ACP => pimeloyl-ACP	0.73	*Geobacillus thermoleovorans*
Gene ID 3240233	K07397	NULL	0.72	*Sphingobium chlorophenolicum*
Gene ID 3160721	K02968	Ribosome, bacteria	0.71	Unknown
Gene ID 4432120	K00294	NULL	0.71	*Burkholderia sp. YI23*
Gene ID 5561172	K01772	Heme biosynthesis, glutamate => protoheme/siroheme	0.71	*Anaeromyxobacter dehalogenans*
Gene ID 1241594	K00641	NULL	0.70	*Variovorax paradoxus*
Gene ID 5470347	K01998	Branched-chain amino acid transport system	0.70	*Variovorax paradoxus*
Gene ID 106891	K03391	NULL	0.70	*Burkholderia sp. YI23*
Gene ID 1983603	K01142	NULL	0.68	*Nitrosomonas europaea*
Gene ID 760709	K00924	NULL	0.67	*Weeksella virosa*
Gene ID 439292	K07485	NULL	0.66	uncultured bacterium contig00155
Gene ID 5821176	K02480	NULL	0.66	uncultured Acidobacteria bacterium
Gene ID 907933	K01283	NULL	0.66	*Anaeromyxobacter sp*. Fw109-5
Gene ID 4848490	K07093	NULL	0.66	uncultured bacterium
Gene ID 3875585	K02004	Putative ABC transport system	0.64	Unknown

Gene ID 2688963	K17879	NULL	0.62	*Propionibacterium freudenreichii*
Gene ID 2781511	K02835	NULL	0.62	*Gemmatimonas aurantiaca*
Gene ID 5230296	K00939	Adenine ribonucleotide biosynthesis, IMP => ADP,ATP	0.62	*Burkholderia ambifaria*
Gene ID 2987209	K00975	Trehalose biosynthesis, D-glucose 1P => trehalose	0.60	*Thioalkalivibrio sulfidiphilus*
Gene ID 4224518	K07712	GlnL-GlnG (nitrogen regulation) two-component regulatory system	0.59	*Pantoea vagans*
Gene ID 804818	K02863	Ribosome, bacteria==Ribosome, archaea	0.58	*Slackia heliotrinireducens*
Gene ID 912294	K01322	NULL	0.57	*Shewanella loihica*
Gene ID 1369058	K01684	D-galactonate degradation, De Ley-Doudoroff pathway, D-galactonate => glycerate-3P	0.57	*Mesorhizobium opportunistum*
Gene ID 1547139	K02004	Putative ABC transport system	0.55	Unknown

Figure 1

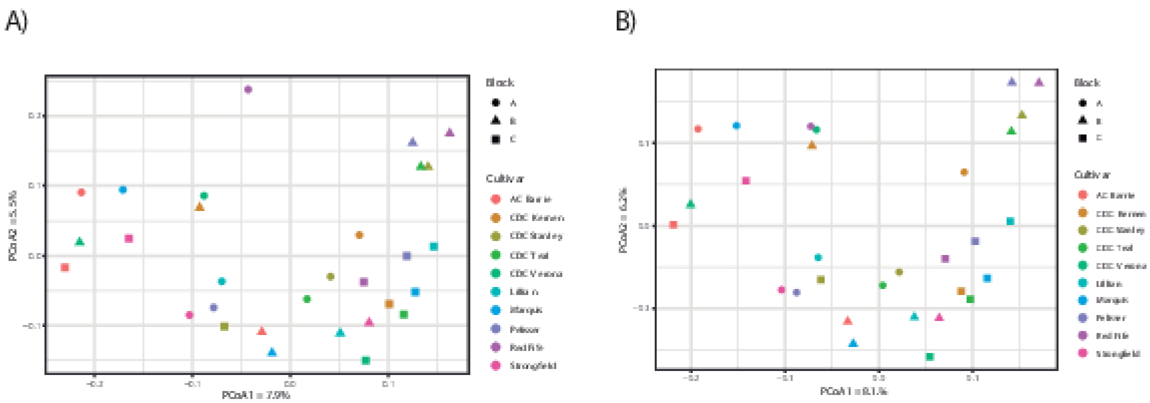
A)
B)
Block
Cultivar
AC Barrie
CDC Stanley
Marquis
Red Fife
Strongfield
PCoA1 = 7.9%
PCoA2 = 5.5%
PCoA1 = 8.1%
PCoA2 = 6.2%

Figure 2