How does surface soil geomorphology and land-use influence the soil microbial ecosystems in south eastern Australia? Insights gained from DNA sequencing of the soil metagenome

Pauline Mele[A,B], Tim Sawbridge[A,B], Helen Hayden[A], Barbera Methe[C], Tim Stockwell[C], Cindi Pfannkoch[C], Matthew Lewis[C], David Tanenbaum[C], Doug Rusch[C] and Karla Heidelberg[D]

[A]Biosciences Research Division, Department of Primary Industries-Victoria, Bundoora, VIC, Australia.
[B]Biosciences Research Centre, Latrobe University, Bundoora.
[C]J. Craig Venter Institute, Rockville, MD, USA.
[D]Department of Biological Sciences-University of Southern California, CA, USA.

Abstract
Soils are a vital resource in Australian agricultural production systems. The sustained health of this resource is dependent on how land-use and management practices impact on the underlying soil microbial community that delivers vital ecosystem goods and services. Collaborative research between Department of Primary Industries Victoria (Melbourne, Australia) and the J.Craig Venter Institute in Rockville (Maryland, USA) research has applied a stepwise DNA-based approached to resolve the influence of contrasting soils (calcarosol and ferrosol) and land-use (managed and remnant vegetation) on the soil bacterial communities. This approach generated small insert clone libraries based on 16S ribosomal RNA (rRNA) sequences for three samples; calcarosol; managed (cropped) and remnant samples and a ferrosol, managed (grazed dairy pasture) sample. A deeper, whole genome shotgun sequencing approach based on Titanium 454 pyro-sequencing technology yielded detailed information on the structural and functional elements of the microbial community in two samples (calcarosol; managed and remnant). All samples yielded unique microbial communities with <1% shared sequences overall. Samples collected from the same regional soils (e.g. highly alkaline calcarosol soil) but with contrasting land-use patterns (e.g. cropped and remnant) were more similar with 13% shared sequences. Samples collected from different regional soils but with similar management shared 4% sequences. The greatest differences in communities were those with contrasting soil, and land-use characteristics with <3% shared sequences. Of the known taxa, Acidobacteria, Cyanobacteria and Planctomycetes were relatively prevalent in the acid soil and Actinobacteria, alpha and delta Proteobacteria prevail in the alkaline, remnant soils. In the alkaline managed soil, Bacteroidetes/chlorobi taxa prevail. We chose the calcarosol samples for additional Sanger and 454 FLX sequencing to evaluate functional genes, with specific emphasis on nutrient cycling and disease suppression pathways. The results provide insight for ecosystem function and management decisions in the context of climate change and resource sustainability.

Key Words
Ecosystem, metagenomics, sequencing, biodiversity, functional genes, 16S rRNA

Introduction

Agricultural production is dependent upon the ecosystem goods and services provided by soil microbial processes (Kibblewhite *et al.,* 2008). An outstanding challenge is to identify the microbial species and functions that underpin these processes so that appropriate and targeted management interventions can be applied. The soil microbial ecosystem presents particular technical challenges because it contains thousands of species of bacteria and fungi per gram that are largely unculturable with only 1-10% of the entire community being identified(Xu, 2006). The emergence of metagenomics, a powerful set of methods that provides a deep assessment of the DNA genetic material contained within microbial communities is offering new opportunities to access the diversity within environmental ecosystems including soils (Handelsman, 2007). Whilst the metagenome of various environmental ecosystems, including soils, have now been partially unravelled (Daniel, 2005; Tringe, 2005; Tyson *et al.,* 2004; Venter *et al.,* 2004), the soil metagenome presents a much more challenging environment due to exceptionally high diversity by most estimates. Because of the technical challenges associated with such a diverse ecosystem, the genetic resource contained within the living component of soils (the microbial community) remains largely untapped (Daniel, 2005).

Methods

Site and sampling description

Sites were selected in Victoria, Australia on two significant soil orders, a calcarosol in north-western Victoria and a ferrosol in the south-east [Figure 1;(Isbell, 2002)]. Three treatments were sampled; one from the ferrosol site under 'managed' grazed dairy pasture (FM) and two from the calcarosol site, on contrasting land-uses; a managed-cropped site (CM) and a remnant native vegetation site (CR). Soils were sampled from within a 50m^2 area from six 1m^2 grids. Five soil cores to depth 5cm were collected per grid and composited.

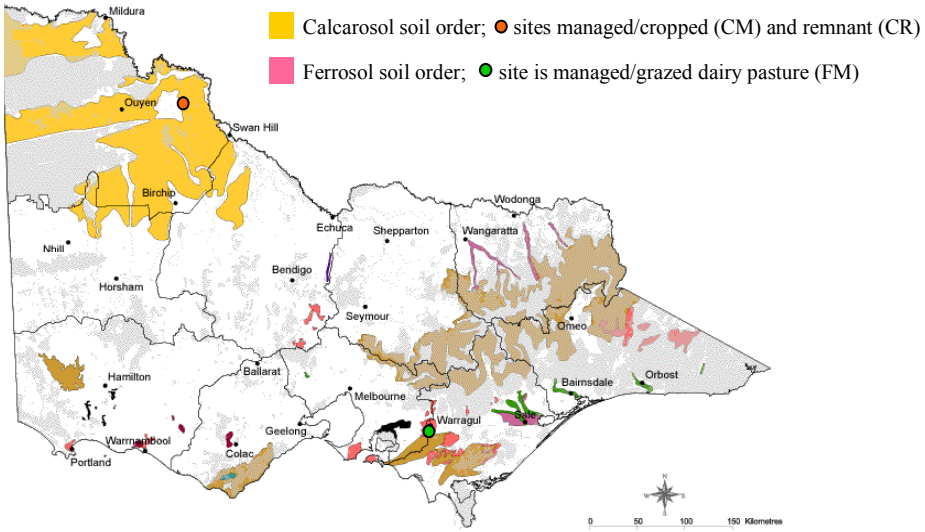

Figure 1. Sample locations in soil geomorphic regions containing the calcarosol and ferrosol soil orders.

The physical and chemical characteristics of each site are described in Table 1. The five cores were then homogenised and frozen immediately for shipping on dry ice to the J.Craig Venter Institute, (MD, USA).

Table 1. Physical and chemical characteristics of soil from each of the three collection sites.

Analyses/soil type & land-use	Calcarosol Managed (cropped) CM	Calcarosol Remnant (native vegetation) CR	Ferrosol Managed (grazed pasture) FM
Particle size			
Clay (<0.002mm)	5.5	16.5	27.5
Coarse Sand (0.2mm - 2.0mm)	47.7	35.4	8.2
Coarse Sand (0.5 - 2.0mm)	40.3	26.8	16
Fine Sand (0.02mm - 0.2mm)	<0.1	5.3	2.2
Silt (0.002mm - 0.02mm)	4.5	7.5	29
pH (H_2O)	6.4	8.1	5.1
pH (Ca)	5.9	7.7	4.5
Electrical conductivity (EC dS/m)	0.07	0.15	0.14
Total soluble salts (%)	0.02	0.05	0.05
Available Nitrogen (mg/kg)			
Ammonium-N mg/kg	0.8	1.3	4
Nitrate-N mg/kg	14	5.4	45
Total N (%)	<0.05	0.14	0.76
Total C (%)	0.62	2.8	9.5
Organic matter (%)	1.1	5.2	18
Oxidisable organic Carbon (mg/kg)			
Organic C	0.73	2.2	9
Organic matter	1.4	4.1	17
Available Phosphorus (mg/kg)	18	5	72
Total cations (mg/kg)			
Calcium	720	19000	2000
Magnesium	920	3800	1100
Phosphorus	160	170	1800
Potassium	2200	5800	2300
Sodium	38	130	150
Sulphur	42	180	920

Soil preparation and DNA extraction
Frozen soils were thawed and the samples from the 1 m^2 grids were combined at each site for the three sites. Soil was then sieved to <3mm, with larger pieces of plant material removed with forceps, on dry ice and 300g subsample was blended at slow speed (Waring blender) in ultra-pure water and then in extraction buffer (10% glycerol, Tween-80). Free DNA was removed using saline phosphate buffer and the ionic strength of the buffer was modified using Chrombach buffer. A Nycodenz gradient was then used to separate cells of specific density which were then isolated and pelleted for storage in buffer. Total nucleic acids were extracted from soils using a MoBio Ultraclean kit according to the manufacturer's instructions.

Library construction and analyses for 16S conserved genes
For library construction, 16S PCR was performed on aliquots of the soil DNA using the prokaryotic 16s primers 27f and 1492r. PCR was performed using pHUSION from Finnzyme. For each sample, three 50 ul reactions were done. Fifteen cycles of PCR were done, the samples were pooled and ethanol precipitated. The PCR products were gel purified. BstXI adaptors were added to the end of the DNA. Excess adaptors were removed by three rounds of gel purification. The PCR product was recovered from the gel and ligated to a vector containing complementary BstXI ends. The resulting ligation was transformed into DH10B competent cells. Clones were sequenced by Sanger sequencing. The 16S rRNA forward and reverse cloned Sanger sequences were concatenated and sequences over 1 kilobases were used as blastn queries against the core set of 16S sequences downloaded from the 'Green Genes' website (http://greengenes.lbl.gov). 16S sequences that produced an HSP of greater than 1kb in this blastn search were trimmed to the segment that produced this match. These trimmed sequences were aligned to the 'core_set_aligned' alignment using the alignment tool in Mothur v 1.4. (www.mothur.org). The resulting alignment was trimmed to: start 400 bp; end 4400bp. These sequences were used to generate Operational Taxonomic Units (OTUs) at 0.03 % divergence and these OTUs were then used as input for phylogenetic analysis of bacterial diversity using the MOTHUR suite of programs (http://schloss.micro.umass.edu/mothur).

Whole shotgun sequencing and analyses using ti454 pyrosequencing
A shotgun library was made for total genomic DNA extracted from CM and CR. DNA was sheared into small fragments, size selected and purified. Standard procedures were followed for ti454 pyrosequencing shotgun library preparation. Adapters were ligated onto purified DNA, which was then prepared by emulsion PCR and titration for addition to the beads. Beads with DNA particles attached were then applied to the glass plates for pyrosequencing. A total of 21.17 million reads of average length of 365 base pairs were generated; 10.6 million for CM and 10.5 million for CR. These sequences were assembled by site and proteins were predicted using Prodigal v1.05. These proteins then used as blastp queries against database of proteins downloaded from completed and partial bacterial genomes from NCBI.

Results
Applying the MOTHUR software (http://schloss.micro.umass.edu/mothur) to the 8,528 OTUs generated from our three treatments (CM, CR and FM) revealed little difference in diversity as determined by the Shannon, Simpson and Chao diversity indices (Table 2).

Table 2. Effect of soil class and land-use on soil microbial diversity using three estimators of diversity, Simpson, Shannon and Chao (http://schloss.micro.umass.edu/mothur).

Diversity indices	CM	CR	FM
Simpson $D_{simpson} = \frac{\sum_{i=1}^{S_{obs}} n_i (n_i - 1)}{N (N-1)}$	0.00038	0.00039	0.00040
Shannon $\hat{H}_{shannon} = \sum_{i=1}^{S_t} \frac{\hat{C}\pi_i \ln\left(\hat{C}\pi_i\right)}{1-\left(1-\hat{C}\pi_i\right)^N}$	6.849	6.7259	6.7168
Chao1 $S_{chao1} = S_{obs} + \frac{n_1 (n_1 - 1)}{2 (n_2 + 1)}$	2846	2322	2986

The microbial community composition however was highly unique between the three sites with < 1% shared sequences overall. Samples collected from the same regional soils (eg highly alkaline soil in very low rainfall site) but with contrasting land-use patterns (eg cropped versus remnant) were more closely related with 13% shared sequences. Samples collected from different regional soils but with agricultural management regimes applied shared 3% sequences. The greatest differences in communities were those with contrasting soil, region and land-use characteristics with <3% shared sequences. Of the known taxa, Acidobacteria, Cyanobacteria and Planctomycetes are relatively prevalent in the acid ferrosol soil (FM) and Actinobacteria, alpha and delta Proteobacteria prevailed in the alkaline calcarosol with remnant vegetation (CR). In the alkaline calcarosol under a cropping management regime (CM), Bacteroidetes/chlorobi taxa prevail.

Whole shotgun sequencing and Ti pyrosequencing of the bacterial metagenomes of two sites, CM and CR revealed the overall dominance of alpha Proteobacteria and Actinobacteria. From alignment with predicted peptides, the dominant taxa in CM compared to CR were the alpha and gamma Proteobacteria. The dominant taxa in CR compared to CM were Actinobacteria, Acidobacteria, Firmicutes, Chloroflexi, delta Proteobacteria and Cyanobacteria. Furthermore in CR, almost 18% of the Actinobacterial peptides were *Rubrobacter* spp, which is highly dominant and contrasts with CM, where Rubrobacter comprised <1% of the Actinobacteria with the strongest representation of Nocardioides at 2%. Alignment with environmental peptides is also likely to reveal significant differences in N, P and C cycles and in disease suppression.

Conclusion

Whilst microbial community diversity measured using several metrics (Simpson's, Shannon's and Chao) were similar at all three sampling locations in south –eastern Australia, the species present at each site were highly unique. Deep sequencing of the metagenome has revealed that the vast majority of species and the functions they perform are not represented by anything presently catalogued in international and publicly accessible sequence databases. Assuming that each taxa provides a vital function, the Ti pyrosequencing approach will resolve the contribution of each grouping of taxa to vital ecosystem services such as N cycling, C fixation and decomposition and disease suppression in these contrasting treatments.

References

Daniel R (2005) The metagenomics of soil. *Nat Rev Micro* **3**, 470-478.

Handelsman JT, Alvarez-Cohen L, Ashburner M, Cann IKO, DeLong E, Doolittle WF, Fraser-Liggett CM, Godzik A, Gordon JI, Riley M, Schmid MB (2007) The New Science of Metagenomics. Revealing the Secrets of Our Microbial Planet. National Academy of Sciences, Washington DC.

Isbell RF (Ed.) (2002) 'The Australian Soil Classification.' (CSIRO Publishing, Melbourne).

Kibblewhite M, Ritz K, Swift M (2008) Soil health in Agricultural Systems. *Phil. Trans.R.Soc.* **362**, 685-701.

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM (2005) Comparative Metagenomics of Microbial Communities *Science* 308.

Tyson GW, Chapman J *et al.* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43.

Venter JC, Remington K *et al.* (2004) Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* **304**, 66-74.

Xu J (2006) Microbial ecology in the age of genomics and metagenomics: concepts, tools and recent advances. *Molecular Ecology* **15**, 1713-1731.