Time series regression relates soil microbial DNA concentration to enzymatic glucose neo-generation

David Ives[1] and Gurudeo Anand Tularam[2]

[1] PhD candidate, Griffith Sciences [ENV, EFRI], Griffith University, Nathan, Qld 4111, Australia
[2] Senior Lecturer, Mathematics and Statistics, Griffith Sciences [ENV, EFRI], Griffith University, Nathan, Qld 4111, Australia
Email: david.ives@griffithuni.edu.au

Abstract: Understanding how microbial DNA from root zone (rhizosphere) soil relates to key metabolic enzymes used by rhizosphere microbes was investigated using time series regression. We aimed to define the influence of these enzymes on DNA concentrations. This objective was achieved using phosphorescent measurement of both enzyme activities and DNA concentrations. Rhizosphere samples from three strains of canola genetically modified (GM) for herbicide resistance (HR) to atrazine, imidazolinone, and glyphosate respectively together with the isoline to the transgenic glyphosate-resistant variety were grown in a greenhouse in pH-neutral Vertisol soil. Analyses were carried out at days 7, 21, 42, and 56 growth representing germination, early growth, maturity, and seed formation-senescence. Enzymes represented key soil microbial metabolic processes and included leucine aminopeptidase (LAP), alkaline phosphatase (PHOS), cellobiose dehydrogenase (CELL), beta-glucosidase (BGL), and aryl sulphatase (SUL). The experiment employed a randomized complete block design with four blocks of eight pots, each pot containing two plants of identical genotype, and each cultivar randomly distributed within each block. Enzyme measurements were taken before substrate addition and again after incubating at 25 degrees Celsius for one hour. DNA concentrations were measured using the Pico Green technique. Mean increases in enzyme activities for each lifespan stage were correlated with DNA concentrations using time series regression. A highly significant relationship ($p = 0.001$) between DNA concentration and beta-glucosidase activity emerged. No other enzyme significantly influenced DNA. The implications of this finding are that during times in the plant lifecycle when microbial-root biochemical interaction are increased (flowering and seed formation), the soil must have adequate SOM available to the microorganisms. It is now known that commensal relationships exist between rhizosphere soil microorganisms and the roots within that soil. The importance of microorganisms in breaking down decaying organic material from plant residues is reflected in contemporary measurement of soil fertility which recognizes that a matrix of physical, chemical, and biological parameters together are required for an adequate assessment. This is because when organically-bound carbon biomass is subjected to transformation, other nutrients like Nitrogen (*N*), Phosphorus *(P)*, and Sulphur (*S)* are released for mineralization by roots. The results reinforce the importance of soil organic matter (SOM) as a component of soil fertility and sustainable soil health.

Keywords: *Rhizosphere, **e**nzymes, soil, DNA, time series regression*

1. INTRODUCTION

1.1. Background information on root-microbial commensalism

Roots respond in an interdependent way with the flora and fauna in the soil. They excrete nutrients into the rhizosphere, sustaining the microbes and the food chain dependent on them. Microbes correspondingly mineralise soil nutrients for root absorption (Janos, 2007). The exudates mediate biological responses from the biota and from roots of other plants through chemical signalling (Bulgarelli, 2012). The effect is to protect the roots from invasion by parasitic organisms (Jain et al., 2012), and to provide competition against adjacent plants, as well as changing plant tissue chemistry and function (Friesen et al., 2011). Recent discoveries by Lundberg et al. (2012) and Bulgarelli et al. (2012) have shown that the rhizosphere contains microorganisms (in particular bacteria) which are important to the plant's survival. The root secretions are controlled by the plant's genes, and the immune system selects those organisms (commensals and mutualists) which benefit the plant by contributing to growth.

1.2. Microbial enzymes

These are important in several ways relating to soil organic matter, soil organic carbon, and the transformation of minerals (mineralization) of organically-bound essential minerals to forms capable of absorption by roots (Feller, 2012). Failure of cellulose- and lignin-based material to decay under the influence of soil microbial cellulase enzymes could therefore have profound effects on soil-plant interaction.

1.3. Experimental aims

Our aim was to discover if significant correlations existed between microbial DNA concentration and the activities of five key microbial enzymes chosen for their unique contributions to plant metabolism. To achieve this aim, we employed panel regression analysis of phosphorescence measurement of enzyme activity and DNA with the objective of developing a dataset suitable for the regression analysis.

2. MATERIALS AND METHODS

2.1. Introduction

Three of the GM cultivars grown were modified for HR to atrazine, imidazolinone, and glyphosate respectively. The genetically engineered variety was the glyphosate-resistant cultivar. The other two HR varieties carried cis-genic modification through selective breeding. An isoline to the glyphosate resistant cultivar was also grown for comparison. A glasshouse system minimized environmental variation to produce consistent growing conditions. Time-series regression was used to correlate the means for all enzyme vectors with DNA concentration. Enzyme activities were measured using the method advocated by Deng et al. (2011) employing phosphorescence. DNA concentrations were measured with PicoGreen method advocated by Petric et al. (2011). SAS was used for statistical analysis.

2.2. Panel regression analysis

The purpose of panel or time series regression is to investigate whether or not sources of variation other than those vectors chosen in the standard regression model are influencing results. An ordinary least squares (OLS) analysis may be biased either because an influential variable has been omitted, or because errors in the covariates are correlated over the time series, or both. In a standard regression, unobserved variable effects are absorbed into the error term, forming a composite error factor. This can result in the phenomenon of endogenicity, and add inefficiency to the statistical analysis. For example, errors generated at time one may be correlated with errors at other times, and thus be biasing the OLS result with serial correlation errors. It is of interest as well that the DNA result may not be independent of any of the enzymes. By correcting for endogenicity, more accurate results may be obtained. The time series vectors associated with panel regression analysis may help to explain the nutrient sources required by the canola plant as it grows and develops, allowing for the above complications. Hence the justification of panel regression analysis as applied to this experimental programme: It is descriptive of the dynamics of plant growth and may produce a causal analysis of plant and soil microbe physiological effects, allowing for the above statistical limitations. In the study, the following pattern of analysis is followed:

- OLS – ordinary least squares analysis: This can be inefficient by failing to use all the information capable of being exploited within the data set, and by not fully exploring error term autocorrelation.

- Fixed effects method analysis (FE). This transforms the data to remove unobserved but constant systemic error. The remaining error is a time-varying error of unexplainable origin also called an "idiosyncratic" error. The "fully demeaned" data can therefore be subjected to OLS analysis.
- Random Effects Methods (RE) are more efficient in estimation and capable of examining within-observation errors. The model assumes that errors across units are uncorrelated, but correlated within each unit; and that variance of the composite error term is the sum of both idiosyncratic error and unobserved errors. Again the data is "demeaned" by removing an estimated demeaning factor between 1 and 0 derived from variance component estimation, in order to choose either the Fixed or Random model, depending on calculation of the Hausman Test which determines if the unobserved errors are exogenous, i.e. uncorrelated.

3. RESULTS

3.1. Standard Panel Regression Model

The model for this Ordinary Least Squares (OLS) regression is given by the following equation

$$DNA_{it} = \beta_0 + LAP_{it}\beta_{LAP} + PHOS_{it}\beta_{PHOS} + BGL_{it}\beta_{BGL} + CELL_{it}\beta_{CELL} + SUL_{it}\beta_{SUL} + v_{it} \quad (1)$$

This model has the following assumptions: Covariates exogenous, errors uncorrelated, errors homoscedastic.

Using the standard regression model where DNA concentration is the Dependent Variable, it is noted that the model is valid (R^2 = 0.9985, adjusted R^2 = 0.9985); yet only one parameter (BGL) is significant in explaining DNA concentration (F = 1346, df = 5, p = 0.05). Those parameters which are not significant statistically have both positive and negative values and large standard deviations.

Table 1. Means, standard deviations, minima, maxima for panel data set

Variable	N	Mean	Standard Deviaton	Minimum	Maximum
DNA	16	24.792	40.427	1.130	93.836
LAP	16	16.662	6.412	5.730	21.300
PHOS	16	47.650	39.857	16.131	117.590
BGL	16	64.969	76.673	16.552	202.118
CELL	16	6.358	7.628	1.470	21.170
SUL	16	1.291	1.331	0.150	3.200

- DNA = double-stranded DNA concentration; LAP = leucine aminopeptidase activity; PHOS = phosphatase activity; BGL = beta-glucosidase activity; CELL = cellobiose dehydrogenase activity; SUL = aryl sulfatase activity

Table 2. Standard panel regression model for dependent variable DNA

Source	DF	Sum Squares	Mean Square	F Value	p
Model	5	24479	4895.865	1346.880	<.0001
Error	10	36.350	3.635		
Corr. Total	15	24516			
R^2	0.998				
Adj R^2	0.998				

The model derived in Table 2 is valid (**R^2** = 0.998) producing a highly significant effect in predicting DNA concentration (p = 0.001).

Table 3. Parameter estimates for standard panel regression model

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-3.999	1.388	-2.880	0.016
LAP	1	-0.410	0.212	-1.930	0.082
PHOS	1	-0.002	0.033	-0.060	0.954
BGL	1	0.633	0.083	7.650	<.0001
CELL	1	-0.974	0.825	-1.180	0.265
SUL	1	0.635	0.951	0.670	0.520

Tables 2 and 3 indicate that the standard regression model gives a valid and highly significant means of assessing DNA through the BGL vector (p = 0.001). LAP is a less significant influence (p = 0.1). Other vectors indicate both positive and negative effects in explaining DNA concentration, although not significant. Here it appears that the enzymatic activity of BGL in delivering glucose to the rhizosphere from soil organic carbon sources has a significant and positive impact on the associated concentration of the DNA in that rhizosphere.

3.2. Fixed effects panel regression model analysis (FE)

By removing an error term C_i by subtracting the within-unit mean from each measurement on that enzyme activity, the time-constant effects are eliminated, expressed as follows:

$$y_{it} - mean\ (y_i) = x'_{it} - mean\ (x_i'\beta) + C_i - mean\ (C_i) \ldots u_{it} - mean\ (u_i),\ t = 1,2,\ldots T \quad (2)$$

A fixed effects transformation has occurred and can be written

$$demeaned\ Y_{it} = demeaned\ x'_{it}\beta + demeaned\ u_{it} \quad (3)$$

and x'_{it} does not contain any error term and applies pooled (OLS) demeaned data to the analysis.

A test for the significance of the No Fixed Effects model was carried out.

H_0 : Demeaning the across unit error term gives a significant effect

H_1: No significant difference to results with this procedure

Dependent Variable: DNA; R^2 = 0.9987

Table 4. Fixed estimates model results for dependent variable DNA

Num DF	Dep DF	F Value	Pr > F
3	7	0.42	0.744

Since the null hypothesis is unable to be rejected, this model cannot provide a suitable method of removing model endogenicity in the data analysis.

Table 5. Panel regression model analysis for FE model

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-4.900	1.940	-2.530	0.0395
LAP	1	-0.337	0.284	-1.180	0.275
PHOS	1	-0.013	0.044	-0.300	0.770
BGL	1	0.639	0.105	6.060	0.0005
CELL	1	-1.058	1.069	-0.990	0.355
SUL	1	0.272	1.299	0.200	0.840

In Table 5 the vector BGL remains significant ($p = 0.001$) in predicting DNA, and its positive vector therefore provides a precise effect for BGL on the dependent variable, since the model is still valid ($R^2 = 0.9987$). Estimated vectors and R^2 differ from the OLS model, thereby indicating greater accuracy. Nevertheless, only one vector is capable of reliably explaining DNA. Because the alternate hypothesis cannot be rejected, the model is of limited value to the analysis. The RE Model therefore follows.

3.3. Random effects panel regression model (RE)

In the RE model, an estimator transforms data by "partially demeaning" each vector using a demeaning factor between 0 and 1, the specific demeaning value being based on an estimation of the variance components. Only a part of the mean is subtracted, instead of subtracting the entire unit-specific mean. This is an attempt to explain the effects of within-cluster correlation and thus provide more efficiency in analysis. Such a process requires the more stringent assumption expressed mathematically as:

$$E(C_i \mid x_{i1}, \ldots, x_{iT}) = E(C_i) = 0 \quad (4)$$

To test whether the RE panel regression model provides a better estimate of DNA concentration than the FE model, the Hausmann test is carried out. If it can be shown that the unobserved errors are exogenous, the FE and RE are equivalent. If not, the RE model is more suitable for analysis because of its efficiency.

Dependent Variable: DNA; $R^2 = 0.9987$

Hausman test for random effects indicated an M value of 0.04, P>m of 1.000, with 5 DF

Table 6. Parameter estimates for the RE model

Variable	DF	Parameter Estimate	Standard Error	t value	Pr > |t
Intercept	1	-4.051	1.478	-2.740	0.0208
LAP	1	-0.370	0.226	-1.640	0.133
PHOS	1	-0.008	0.035	-0.230	0.819
BGL	1	0.635	0.086	7.410	<.0001
CELL	1	-1.007	0.862	-1.170	0.270
SUL	1	0.435	1.025	0.420	0.681

Table 6 indicates that BGL is again significant in predicting DNA concentrations (p <.001). Thus a more accurate estimate of vectors has been derived by eliminating endogenicity and refining nominated vectors.

3.4. Summarizing results

The time series RE model analysis demonstrated a highly significant statistical relationship between BGL activity and DNA concentration. Since BGL is an enzyme associated with mineralization and the release of metabolites from decaying plant residues (sources of lignin and cellulose), it again brings into focus how SOM decay is vital to rhizosphere microbial processes nurturing commensal root-microbe relationships.

4. DISCUSSION AND CONCLUSIONS

4.1. Plant-microbe-enzyme linkages

The panel regression analysis suggests both descriptive and causal explanations for achieved experimental results while minimizing endogenicity. Through similar models, it may be possible to understand how soil organisms influence plant differential growth and development and demand for nutrients. The apparent linkage of beta-gluconase activity with DNA concentration indicates that glucose derived from biochemical breakdown of cellulose, lignin, and other soil organic carbon sources appears to play a vital part in the sustenance of microbes in the soil.

4.2. Implications

Planting without the need to plough, or at least minimizing soil disturbance is a desirable environmental pathway. GM and HR cultivars minimizing pesticide use fit into this scenario well. Long-term no-till strategies which allows no disturbance to SOM decay can release mineralizable plant nutrients. The effects of GM plants on rhizosphere processes has received widespread investigation through ecotoxicological examination, and has so far failed to establish consistently significant differences amongst cis-genic modification (by selective traditional plant breeding), or trans-genic modification (by genetic engineering), or plants with no modification, compared with variation caused by abiotic factors like season, climate, and soil type. Advocates of GM technology prefer to see crop residues remain on or in the soil and to allow residues and roots to decay without the need to plough. Promoting and preserving the organic content of soil is now seen as an essential part of soil stewardship. Therefore, if GM technology is one part of a new way of growing canola and other crops, then gains in SOM and soil C generally will be environmentally beneficial. We predict that GM crops will be part of a mix of new productive technologies capable of supporting a growing world population through *sustainable intensification.*

DECLARATION

Conflicts of interest: none.

ACKNOWLEDGEMENTS

The NSW DPI through Wollongbar Research Institute for access to laboratory analysis and greenhouse. Financial assistance by Griffith University E.F.R.I.

REFERENCES

Bulgarelli, D. (2012). The structure of and assembly cues for the *Arabinopsis* root-inhabiting bacterial microbiota. *Nature*. Volume 487 (August), Issue 7409. doi: 10.1038/nature 11336.

Deng, S., Kang, H. & Freeman, C. (2011). Microplate fluorogenic assay of soil enzymes. In: R.P. Dick (Ed.) *Methods in Soil Enzymology*, SSSA, Madison, WI.

Feller, C., Blanchart, E., & Bernoux, M., et al. (2012). Soil fertility concepts over the past two centuries: the importance attributed to soil organic matter in developed and developing countries. *Arch. Agron. Soil Sci.*, Volume 58, PS3-S21. Doi: /10.1080/03650340.2012.693598

Friesen, M.L., Porter, S.S., & Stark, S.C., et al. (2011). Microbially mediated plant functional traits. *Ann. Rev. Ecol. Evol. System.*, *42*(12), pp. 23-46.

Jain, A., Singh, S., et al. (2012). Microbial consortium-mediated reprogramming of defence network in pea to enhance tolerance against *Sclerotonia sclerotiorum*. *J. Appl. Micro., 112*, pp. 537-550.

Janos, D.P. (2007). Plant responsiveness to Mycorrhizas differs from dependence on Mycorrhizas. *Mycorrhiza, 17*(2), pp. 75-91.

Lundberg, D.S., Lebeis, S.L., & Paredes, S.H., et al. (2012). Defining the core *Arabinopsis thaliana* root microbiome. *Nature*, Volume 488 (08), Issue 7409, (August 2). doi: 10.1038/nature11237

Petric, I., Philippot, L., & Abbate, C.M. (2011). Inter-laboratory evaluation of the ISO standard 11063 "Soil quantification method to directly extract DNA from soil samples." *J. Micro. Meth. 84*(3), pp. 454-460.