

Weighted Gene Co-expression Network Analysis of Metal Mixtures in Drinking Water and Their Effects on Zebrafish Behaviour

R. K. K. Dilrukshi^{1,*} , N. Withanage² , J. Nishad³ ,
and P. W. Fernando⁴ 

¹Department of Spatial Sciences, General Sir John Kotelawala Defence University - Southern Campus, Sri Lanka

²Department of Statistics, University of Sri Jayewardenepura, Sri Lanka

³The Nicholas School of the Environment, Duke University,
North Carolina 27708, United States

⁴Department of Mathematics, University of Sri Jayewardenepura, Sri Lanka

*Corresponding author: kdilrukshi@kdu.ac.lk

Received: 19th June, 2025/ Revised: 31st August, 2025/ Published: 30th November, 2025

©IAppstat-SL2025

ABSTRACT

Mixtures of chemical contaminants can pose significant health risks to humans and wildlife, even at levels deemed safe for individuals. Zebrafish (Danio rerio) offer a high-throughput exposure model with the complexity of a vertebrate organism, making them well-suited for evaluating mixture toxicity. However, substantial gaps exist in statistical methods for assessing the associations between chemical mixtures and phenotypic endpoints in toxicity testing. Here, a Weighted Gene Co-expression Network Analysis (WGCNA) was developed to address this challenge, focusing on a larval behavioural assay and leveraging data from 92 well-water samples from Maine and New Hampshire, USA. Our study aims to implement mixture-relevant statistical approaches to elucidate the relationships between chemicals in drinking water and the behavioural responses observed in zebrafish. A WGCNA was employed to uncover gene expression patterns that mediate behavioural effects induced by these chemical mixtures. Individual chemicals within these mixture models exhibit both positive and negative partial effects. For instance, within the turquoise module, Cadmium

(Cd) is positively weighted, while Copper (Cu), Nickel (Ni), and Lead (Pb) are negatively associated with zebrafish behaviour. Similarly, within the grey module, Arsenic (As), Uranium (U) and Chromium (Cr) show positive effects, whereas Selenium (Se) and Antimony (Sb) exhibit negative effects. These findings underscore the importance of evaluating overall mixture exposure effects and highlight the critical need to consider complex interactions within chemical mixtures in environmental toxicology.

Keywords: Chemical mixtures, Behavioral toxicology, Zebrafish (*Danio rerio*), Weighted gene co-expression network analysis

1 Introduction

The contamination of drinking water with chemical pollutants is a growing global concern, with implications for both environmental and human health. Groundwater and private wells are particularly vulnerable to contamination from industrial activities, agricultural practices, and natural geochemical processes (Carpenter et al., 2002; Cole et al., 2013). These water sources often contain mixtures of heavy metals, pesticides, and other pollutants that may interact to produce additive, synergistic, or antagonistic effects. Importantly, such interactions can result in biological responses at concentrations well below current regulatory thresholds, underscoring the need for mixture-focused toxicological research.

Zebrafish (*Danio rerio*) have become a widely used vertebrate model for environmental toxicology. Their rapid development, genetic and physiological similarity to humans, and transparency during early life stages make them particularly suitable for mechanistic studies (Kalueff et al., 2013). Zebrafish behavioral assays, such as locomotor activity and responses to light/dark transitions, are sensitive indicators of neurodevelopmental disruption and provide quantifiable endpoints for evaluating the impacts of environmental exposures (Babich et al., 2021; Dilrukshi et al., 2024). These attributes position zebrafish as an ideal system for investigating the effects of chemical mixtures found in drinking water.

A key challenge in mixture toxicology is identifying the molecular mechanisms linking complex exposure profiles to organismal outcomes. Traditional approaches often assess chemicals individually, which may obscure network-level biological responses. Weighted Gene Co-expression Network Analysis (WGCNA) provides a systems biology framework for uncovering coordinated gene expression patterns associated with environmental exposures (Li et al., 2018; Rager et al., 2021). By constructing networks of co-expressed genes, WGCNA identifies modules that represent shared biological functions and enables the association of these modules with external traits, such as exposure levels or behavioral phenotypes

(Zhang and Horvath , 2005; Langfelder and Horvath , 2008).

In this study, we apply WGCNA to investigate the relationships between chemical mixtures in drinking water and zebrafish larval behavior. Using well-water samples collected from Maine and New Hampshire, USA, we integrate behavioral data with gene co-expression networks to identify modules of genes potentially driving mixture-induced behavioral changes. This approach provides new insights into the molecular pathways underlying environmental mixture toxicity and contributes to advancing risk assessment strategies beyond single-chemical evaluations.

2 Methodology

2.1 Study Design and Data Source

This study reanalyzed data originally generated by Babich et al. (2021), which assessed the developmental and behavioral impacts of drinking water exposures in zebrafish. A total of 92 private well-water samples were collected from households in Maine and New Hampshire, USA. Zebrafish embryos ($n = 24$ per sample) were exposed to each well-water sample as well as to a laboratory control water sample. Behavioral endpoints were measured in larval zebrafish using a standardized locomotor assay, in which movement was tracked over a 25-minute period under alternating light and dark conditions (5 minutes each phase).

Each well-water sample was analyzed for multiple trace metals, including: Arsenic (As), Antimony (Sb), Barium (Ba), Beryllium (Be), Cadmium (Cd), Chromium (Cr), Copper (Cu), Iron (Fe), Lead (Pb), Manganese (Mn), Nickel (Ni), Selenium (Se), Thallium (Tl), and Uranium (U). These data served as the exposure matrix for downstream analyses.

2.2 Weighted Gene Co-expression Network Analysis (WGCNA)

2.2.1 Network Construction

WGCNA was performed to identify modules of co-expressed genes associated with zebrafish behavioral responses. Pairwise Pearson correlations were computed for all genes across samples. These correlations were transformed into an adjacency matrix using a soft-thresholding power (β) to emphasize strong correlations and reduce the influence of weak ones:

$$a_{ij} = |\text{cor}(x_i, x_j)|^\beta$$

where a_{ij} represents the connection strength between genes i and j . The power parameter β was selected to ensure that the network conformed to a scale-free topology, a property typical of biological networks.

2.2.2 Topological Overlap and Module Detection

The adjacency matrix was further transformed into a Topological Overlap Matrix (TOM), which incorporates both direct correlations and shared neighbors between genes. TOM-based dissimilarity was used as input for hierarchical clustering. Modules of co-expressed genes were identified as clusters (branches) of the resulting dendrogram using a dynamic tree-cutting algorithm.

2.2.3 Module Eigengenes and Trait Association

For each module, the module eigengene (ME) was calculated as the first principal component of gene expression within the module, representing its overall expression profile:

$$ME_M = PC_1(X_M)$$

where X_M is the expression matrix of genes in module M . Module eigengenes were then correlated with external traits, including zebrafish behavioral responses, to identify biologically meaningful associations.

2.2.4 Network Statistics

To characterize network structure, intramodular connectivity was computed for each gene, identifying hub genes as those with the highest connectivity within their respective modules. Module preservation statistics were also evaluated to assess the robustness of module detection.

This statistical framework enables WGCNA to capture biologically meaningful patterns of co-expression, providing insights into the molecular mechanisms by which complex environmental mixtures influence zebrafish behavior.

3 Results and Discussion

3.1 Analysis with Outliers

Our initial objective was to categorize chemicals into distinct groups using an unsupervised methodology. This approach aimed to uncover patterns of co-occurrence among 92 drinking water samples. We utilized the WGCNA

technique, which enabled the recognition of chemicals that exhibited parallel increases or decreases in concentration across various exposures. The outcome of this analysis led to the identification of four distinct sets of chemicals (refer to Table 1 and Fig. 1). These sets are labeled as 'modules' and are color-coded as follows, each comprising a specific number of chemicals: Grey module (7 chemicals), Turquoise module (3 chemicals), Blue module (2 chemicals), and Brown module (2 chemicals).

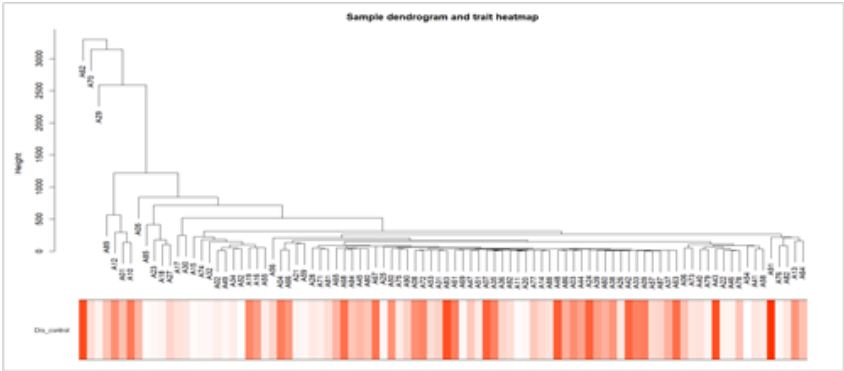
The brown and blue modules represent negative relations with zebrafish behaviour, while the turquoise module represents a moderate negative relation. The brown module contained Cu and Pb. As a mixture, these chemicals exhibited a negative correlation (-0.31) with zebrafish behaviour and a statistically significant p-value of 0.003. This suggests that as concentrations of Cu and Pb increase, zebrafish exhibit increasingly adverse behavioural effects, indicating a potentially toxic impact of these metals when present together. The blue module contained Ni and Cd. This combination of chemicals also exhibited a negative correlation (-0.32) with zebrafish behaviour and a p-value of 0.002, which is statistically significant. This finding implies a compounded toxic effect when these metals coexist, leading to pronounced behavioural disruptions in zebrafish.

The turquoise module included Beryllium, Barium, and Thallium. This group of chemicals exhibited a moderate negative correlation (-0.2) with zebrafish behavior and a p-value of 0.06, which is less statistically significant. Although the impact is less severe compared to the brown and blue modules, it still indicates a detrimental effect on zebrafish behavior when these chemicals are present together. The grey module consisted of Chromium, Arsenic, Selenium, Antimony, Uranium, Manganese, and Iron. This mixture of chemicals collectively had no significant impact (p-value = 0.9) on zebrafish behavior (Fig. 1.C). Despite the presence of these elements, their co-occurrence did not correlate with notable changes in zebrafish behavior, suggesting a lack of synergistic toxicity or possibly balanced interactions that mitigate individual effects.

These identified modules highlight the complexity and interconnectivity of chemical contaminants in drinking water. The patterns of co-occurrence uncovered by the WGCNA provide valuable insights into the potential sources and behaviors of these chemicals in the environment. The negative correlations observed for the brown, blue, and turquoise modules underscore the importance of considering chemical mixtures in toxicological assessments, as combined exposures can have more severe impacts than individual chemicals. Further investigation into the specific properties and interactions of these chemical groups is warranted to understand their collective impact on zebrafish behavior and potentially on human health.

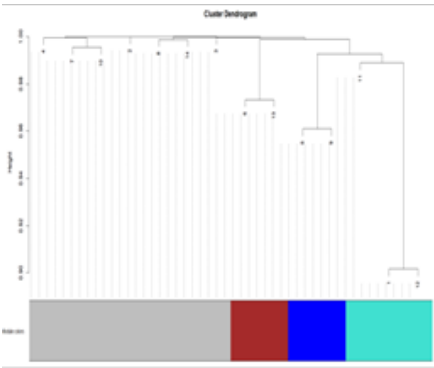
A

Sample dendrogram and heat map



B

Cluster dendrogram



C

Module-distance relationship

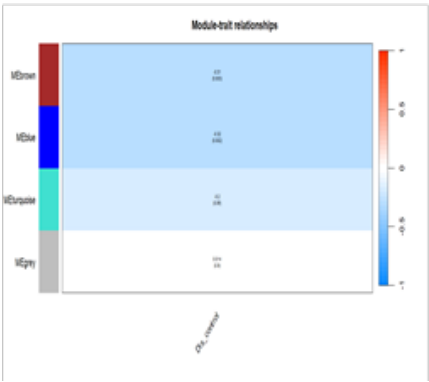


Fig. 1: Visualization of WGCNA analysis, including the presence of outliers. A) Sample dendrogram and heat map. White color represents a low value, and red color represents a high value. B) Cluster dendrogram of 14 chemicals. Numbers represents chemicals: 1-Be, 2-Cr, 3-Mn, 4-Fe, 5-Ni, 6-Cu, 7-As, 8-Se, 9-Cd, 10-Sb, 11-Ba, 12-Tl, 13-Pb, 14-U. C) Module-distance relationship. Correlations between chemical groups and distance traveled by zebrafish. Significant ($p < 0.05$) correlations are shown as colored rectangles, and white rectangle reflects insignificant correlation. Each cell contains the corresponding correlation and p-value; red indicates positive correlation and blue indicates negative correlation.

Table 1: The identification of four different chemical groups using WC-CNA.

Module	Metal component	Color
Module 0	Chromium (Cr)	Gray
	Arsenic (As)	
	Selenium (Se)	
	Antimony (Sb)	
	Uranium (U)	
	Manganese (Mn)	
	Iron (Fe)	
Module 1	Beryllium (Be)	Turquoise
	Barium (Ba)	
	Thallium (Tl)	
Module 2	Nickel (Ni)	Blue
	Cadmium (Cd)	
Module 3	Copper (Cu)	Brown
	Lead (Pb)	

3.2 Analysis without Outliers

The detection and handling of outliers are crucial for ensuring the robustness and reliability of statistical analyses. Outliers can significantly influence results, making robust methods essential in such cases to avoid biased conclusions. In our analysis, we identified outliers using the WGCNA approach. Fig. 2.A illustrates that among the 92 samples, instances of outliers were detected in samples 62, 70, and 29. These outliers are data points that deviate significantly from the majority of the dataset, potentially due to unique characteristics or values.

After removing the identified outliers, leaving us with 89 drinking water samples, we conducted another round of WGCNA analysis. Utilizing the WGCNA technique, we aimed to recognize chemicals exhibiting concurrent increases or decreases in concentration across different exposure scenarios. The results of this analysis led to the identification of four distinct groups of chemicals, referred to as 'modules' (Table 2 and Fig. 2.B). These modules are color-coded as follows, each containing a specific number of chemicals: Grey module (5 chemicals), Turquoise module (4 chemicals), Blue module (3 chemicals), and Brown module (2 chemicals).

The grey module consists of Chromium, Arsenic, Selenium, Antimony, and Uranium. Interestingly, the combined presence of these elements in the Grey module results in weak negative correlations of -0.38, -0.2, and -0.14 with zebrafish behaviour, respectively. This suggests that while in-

Table 2: The identification of four different chemical groups using WC-CNA.

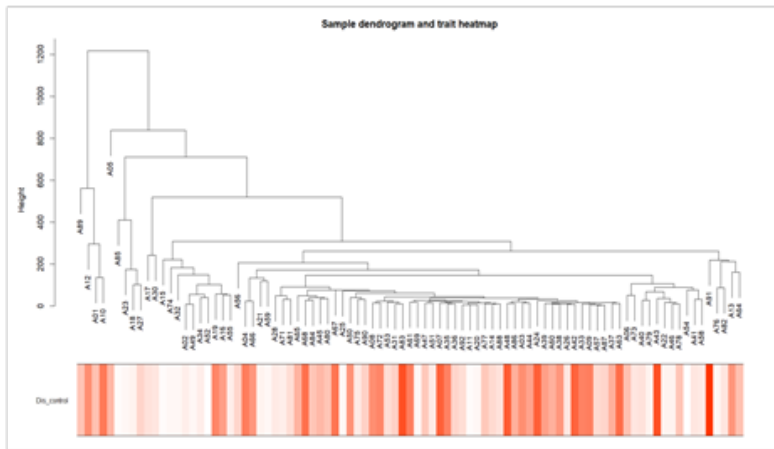
Module	Metal component	Color
Module 0	Chromium (Cr)	Gray
	Arsenic (As)	
	Selenium (Se)	
	Antimony (Sb)	
	Uranium (U)	
Module 1	Nickel (Ni)	Turquoise
	Cadmium (Cd)	
	Copper (Cu)	
	Lead (Pb)	
Module 2	Beryllium (Be)	Blue
	Barium (Ba)	
	Thallium (Tl)	
Module 3	Manganese (Mn)	Brown
	Iron (Fe)	

dividually these chemicals may not strongly affect behaviour, their combined presence shows a tendency towards negative impacts. The Turquoise module includes Nickel, Cadmium, Copper, and Lead. These chemicals collectively exhibit negative correlations with zebrafish behaviour, as observed in the blue module from the previous analysis. This indicates that their presence together continues to show detrimental effects on zebrafish behaviour. The blue module comprises Beryllium, Barium, and Thallium. Similar to the previous analysis, this combination exhibits moderate negative correlations with zebrafish behaviour, indicating their potential adverse effects when co-occurring. The Brown module contains Manganese and Iron. Interestingly, the presence of these elements in the Brown module leads to a weak positive correlation of 0.1 with zebrafish behaviour (Fig. 2.C). This suggests that higher concentrations of Manganese and Iron may have a slight stimulatory effect on zebrafish behaviour under the conditions tested.

These findings underscore the importance of considering both individual chemicals and their interactions in assessing their impacts on zebrafish behavior. The presence of outliers and their removal allowed for a more refined analysis, highlighting distinct chemical modules that collectively influence biological responses. Further exploration into the mechanisms underlying these correlations is essential for understanding how chemical mixtures in drinking water may affect aquatic organisms and potentially human health.

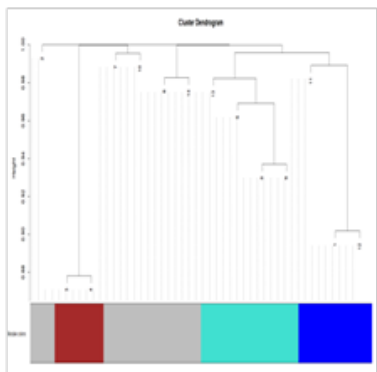
A

Sample dendrogram and heat map



B

Cluster dendrogram



C

Module-distance relationship

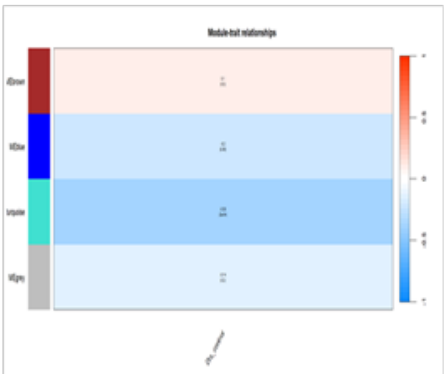


Fig. 2: Visualization of the WGCNA analysis technique without outliers. A) Sample dendrogram and heat map. White color represents a low value, and red color represents a high value. B) Cluster dendrogram of 14 chemicals. Numbers represents chemicals: 1-Be, 2-Cr, 3-Mn, 4-Fe, 5-Ni, 6-Cu, 7-As, 8-Se, 9-Cd, 10-Sb, 11-Ba, 12-Tl, 13-Pb, 14-U. C) Module-distance relationship. Correlations between chemical groups and distance traveled by zebrafish. Significant ($p < 0.05$) correlations are shown as colored rectangles, and white rectangle reflects insignificant correlation. Each cell contains the corresponding correlation and p-value; red indicates positive correlation and blue indicates negative correlation.

4 Conclusion

This study underscores the significant health risks posed by mixtures of chemical contaminants in drinking water, which can adversely affect both human and wildlife health, even when individual chemicals are present at levels considered safe. Utilizing zebrafish (*Danio rerio*) as a model organism, we employed Weighted Gene Co-expression Network Analysis (WGCNA) to identify distinct chemical modules and their associations with behavioral outcomes. Our analysis revealed that certain metal combinations, such as copper (Cu) and lead (Pb), exhibited strong negative correlations with zebrafish behavior, highlighting the compounded toxic effects of these mixtures.

Notably, our findings align with the work of Dilrukshi et al. (2024), who developed a hierarchical modeling framework to evaluate the toxicity of complex metal mixtures in drinking water. Their study identified several "bad-actor" metals, including Pb, Cd, Ni, Cu, and Ba, and demonstrated that exposure to these metals at concentrations deemed safe by regulatory standards can lead to significant behavioral toxicity in zebrafish larvae. This reinforces the notion that current regulatory limits may not adequately account for the combined effects of chemical mixtures, necessitating a reevaluation of safety thresholds.

To enhance the robustness of our findings, we conducted analyses both with and without outliers. The initial WGCNA analysis identified four distinct modules (Grey, Turquoise, Blue, and Brown), reflecting patterns of co-occurring chemicals. However, the presence of outliers introduced variability that affected the strength of correlations among some chemicals. After removing outliers, the overall structure of modules remained similar, but some chemicals were reassigned to different modules. This adjustment improved the clarity and specificity of module definitions, indicating that outliers may have distorted the true co-occurrence patterns.

Associations between the Turquoise module (which included Cd, Cu, Ni, and Pb) and negative zebrafish behavioral responses were observable, yet less pronounced than in the analysis without outliers. The associations strengthened significantly, highlighting that the presence of outliers diluted the strength of the correlations between specific chemicals and behavioral outcomes. This finding underscores the importance of outlier management in toxicological assessments.

In conclusion, our research highlights the critical need to assess the toxicity of chemical mixtures in drinking water, as their combined effects can pose significant health risks. The integration of robust statistical techniques, such as WGCNA, and careful outlier management enhances the reliability

of findings in mixture toxicity studies. Future research should aim to elucidate the mechanisms underlying the interactions between chemical mixtures and biological systems, ultimately informing regulatory policies and strategies to safeguard environmental and public health.

Acknowledgement

This research was supported by a grant from the Research Council of the University of Sri Jayewardenepura, Sri Lanka.

Grant Number: ASP/01/RE/SCI/2022/20.

References

- Babich, E., Craig, A., Muscat, J., Disney, A., Farrell, L., Silka, L., Jayasundara, N. (2021) Defining drinking water metal contaminant mixture risk by coupling zebrafish behavioral analysis with citizen science, *Scientific Reports*, 11(1). Available at: <https://doi.org/10.1038/s41598-021-96244-4>.
- Carpenter, D. O., Arcaro, K., Spink, D. C. (2002). Understanding the human health effects of chemical mixtures. *Environmental Health Perspectives*, 110(Suppl 1), 25–42. 10.1289/ehp.02110s125
- Cole, S. R., Richardson, D. B., Chu, H., Naimi, A. I. (2013) 'Analysis of occupational asbestos exposure and lung cancer mortality using the g formula', *American Journal of Epidemiology*, 177(9), pp. 989–996. Available at: <https://doi.org/10.1093/aje/kws343>.
- Daniel, R. M., DeStavola, B. L., and Cousens, S. N. (2011) 'Gformula: estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula', *Stata Journal*, 11(4), pp. 479–517. Available at: <https://doi.org/10.1177/1536867X1101100401>.
- Dilrukshi, K. R. K., Merutka, I. R., Chernick, M., Rohrbach, S., Babich, R., Withanage, N., Fernando, P. W., Jayasundara, N. (2024). Determining bad actors: A linear mixed effects model approach to elucidate behavioral toxicity of metal mixtures in drinking water. *Ecotoxicology and Environmental Safety*, 287, 117296. <https://doi.org/10.1016/j.ecoenv.2024.117296>
- Kalueff, A. V., Stewart, A. M., & Gerlai, R. (2013). Zebrafish as an emerging model for studying complex brain disorders. *Trends in Pharmacological Sciences*, 35(2), 63–75.
- Keil, A. P., Buckley, J. P., O'Brien, K. M., Ferguson, K. K., Zhao, S., White, A. J. (2020). A quantile-based g-computation

approach to addressing the effects of exposure mixtures. *Environmental Health Perspectives*, 128(4), 047004. Available at: <https://doi.org/10.1289/EHP5838>

Langfelder, P. and Horvath, S. (2008) 'WGCNA: An R package for weighted correlation network analysis', *BMC Bioinformatics*, 9(1). doi:10.1186/1471-2105-9-559.

Li, J., Zhou, D., Qiu, W., Shi, Y., Yang, J.-J., Chen, S., Wang, Q., Pan, H. (2018) 'Application of weighted gene co-expression network analysis for data from paired design', *Scientific Reports*, 8(1). Available at: <https://doi.org/10.1038/s41598-017-18705-z>.

Rager, J. E., Clark, J., Eaves, L. A., Avula, V., Niehoff, N. M., Kim, T. H., Jaspers, I., Gilmour, M. I. (2021) 'Mixtures modeling identifies chemical inducers versus repressors of toxicity associated with wildfire smoke', *Science of The Total Environment*, 775, p. 145759. Available at: <https://doi.org/10.1016/j.scitotenv.2021.145759>.

Zhang, B. and Horvath, S. (2005) 'A general framework for weighted gene co-expression network analysis', *Statistical Applications in Genetics and Molecular Biology*, 4(1). Available at: <https://doi.org/10.2202/1544-6115.1128>.