Statistically Analyzing the Biodiversity and Water Quality of Loudoun County Local Streams

Tanya Singh

Department of Research, Academies of Science

AOS Sophomore Science Research

Dr. Tomlinson

Research Partners: Anjali Potu, Sri Somarapu

February 9, 2022

Abstract

Streams and rivers are an important part of the ecosystem, and organisms can not live properly in streams with poor ecological health. Therefore, the purpose of this experiment is to test the health of our local streams across Loudoun County. To do this, a sample of one of the local streams in Loudoun County was taken. Then, the organisms collected from the stream were classified and recorded in a data table. The same thing was done by 28 groups, who sampled different streams all throughout Loudoun County. Then, the Water Quality Index (WQI), biodiversity (H'), and evenness (J') were calculated for each of these 28 locations and compiled into a data table. From this, the locations were then grouped into 3 categories: streams near Leesburg, streams near Brambleton, and streams near Aldie. Then, three Kruskal-Wallis tests were performed to determine if there was a significant difference between the WQI, H', and J' values throughout all three locations. If the p-value from these tests were below the universal 5% significance level, then we can conclude that there is a significant difference in these values. Though, the p-value for the WQI values was 0.14, which is greater than 0.05; the p-value for the H' values was 0.29, which is also greater than 0.05; and the p-value for the J' values was 0.11, which is also greater than 0.05. Therefore, it was concluded that there is no statistically significant difference between the WQI, H', and J' values throughout all three locations. Consequently, the health and biodiversity of all Loudoun County streams are relatively similar.

Introduction

Biodiversity is an important part of our ecosystem in the modern world. Recently, over pollution has caused climate change, which has affected the world in many ways. One of the main ways climate change has affected the bodies of water is the increase in temperature, which has caused many aquatic organisms to either migrate or slowly die. In other instances, climate change causes droughts, where many aquatic organisms are unable to find enough water and surrounding aquatic life to live. All of these factors may eventually lead to the decrease in biodiversity of our local water systems (*Ecosystem effects of biodiversity loss could rival impacts of climate change and pollution,* 2012).

This loss in biodiversity has many negative impacts, one being on our ecosystem. If a certain species decreases, then this could cause an overpopulation of another species that used to be a prey to the first one. At first, the increase in population of the second species may seem beneficial, though this would decrease the biodiversity and the evenness. Biodiversity measures the variance of species, and if one would overpopulate, this would decrease the population of the others. Additionally, the species would not be evenly distributed. This leads to the ecosystem not functioning well. Therefore, the decrease in biodiversity does not lead to a well functioning ecosystem (Hagan et al., 2021). This is a big concern in our society because if the biodiversity of our local streams decrease, then the food chain would not be properly balanced, eventually hurting humans. This could decrease the lack of food sources, increase diseases, and lose livelihoods. To address this issue it is crucial to investigate the biodiversity of our local streams and observe the change over time.

Previously, research has been done on the relationship of the position in the stream and the biodiversity. This was conducted on Langat River in Malaysia. Four different stations were set up, where there was one upstream, and then each succeeding station was a little downstream. Then, they would collect macroinvertebrates and calculate the WQI at each of the four stations. From this, they were able to conclude that the water quality indices upstream were significantly greater than the water quality indices downstream. (Azrina et al., 2006). Knowing this, the experiment that will be conducted would be slightly different. Firstly, more than one stream will be assessed in Loudoun County, unlike past research where only one stream was assessed. Additionally, the biodiversity (H') and evenness (J') would also be assessed throughout the streams in Loudoun County. Additionally, the data interpretation would differ because previously, a parametric ANOVA test was used due to the large sample size, while this research would assess the data collected using a non-parametric Kruskal-Wallis test to compare the WQI, H', and J' values throughout Loudoun County.

The purpose of this research is to assess the health and biodiversity of our local streams in Loudoun County. Based on previous research, the hypothesis was generated: If the area of Loudoun County is more civilized such as Eastern Loudoun County, then the streams in these locations will have a lower Water Quality Index (WQI), biodiversity (H'), and evenness (J') compared to locations that are less civilized such as Northern and Southern Loudoun County. This is because civilized regions have a greater human population, thus more human waste goes into local streams, causing them to get polluted. Polluted streams indicate a bad water quality because these streams are unlivable for most organisms. This would also result in lower H' and J' values because a lot of organisms aren't able to live in these streams. For this experiment, the location of the stream is the independent variable. The dependent variables for this experiment are the WQI, H', and J' values. The WQI measures how ecologically safe the water is, H' measures how diverse the species in the streams are, and J' represents how evenly distributed the species are amongst the streams. Throughout this experiment, the time the data collection takes place, equipment, and body of water per group remains constant. Additionally, there was no given control group for this experiment. The observations of this experiment were then recorded in a data table, from which valid conclusions were drawn to accept or reject the hypothesis.

Materials and Methods

For this experiment, a river with a slow-moving river with a muddy or sandy bottom was located in Loudoun County. Classification documents to identify macroinvertebrates, 2 aluminum trays, 14-cell ice cube trays, tweezers, plastic spoons, and a D-net were used in the process of this experiment. After locating the stream, the coordinates of that location were taken using Google Maps. Then, a student stood on the edge of the river and slowly lowered the D-net into water. The D-net was then used to slowly scrape the bottom of the river and collect any macroinvertebrates found. The contents from the net were then dumped into the aluminum trays. The aluminum trays must already have water to keep the organisms alive after being transferred over. Then, the ice trays were filled with water and put next to the aluminum tray. Spoons and tweezers were then used to separate similar looking organisms into a separate cell. Then, the number of organisms were counted and then classified using the classification key provided and recorded in Table 1. Finally, the organisms were then released slowly back into the stream. Then, this process was repeated for two more trials, where each trial was conducted 10 feet more upstream compared to the previous one. During this experiment, parental supervision is necessary to make sure no one gets injured due to the slippery rocks located in the streams.

After collecting quantitative measurements for all three trials, the WQI, H’, and J’ were calculated for all 28 groups and recorded in Table 4. Then, ArcGIS was used to group these data points based on 3 different locations in Loudoun County: Leesburg, Brambleton, and Aldie.

After this, Kruskal-Wallis tests were performed to observe if there is a significant difference in the WQI, H', and J' values throughout all three locations. This provided the p-value, and if it was less than 5%, there would be a significant difference in the WQI, H', and J' values. These tests were performed in Table 5 below. Finding the p-value would then help us make valid conclusions regarding the health of Loudoun County streams.

Data

Raw Data of the Macroinvertebrates Collected

				Number of Organisms found			
				Trials			
Phylum	**Class**	**Order**	**Organism**	1	2	3	Total
Platyhelminthes			Flatworm	0	0	0	0
Annelida	Olligochaeta		Aquatic Worm	1	0	1	2
	Hirudinea		Leech	0	0	0	0
Mollusca	Gastropoda		Lunged Snail	0	0	0	0
	Bivalvia		Gilled Snail	0	0	0	0
			Clams/Mussels	43	18	12	73
Arthropoda	Malacostraca	Decapoda	Crayfish	0	0	1	1
		Isopoda	Sowbug	0	0	0	0
		Amphioda	Scud	0	1	0	1
	Insecta	Plecoptera	Stonefly Larvae	1	0	1	2
		Ephemeroptera	Mayfly Larvae	0	0	0	0
		Trichoptera	Caddisfly Larvae	0	0	0	0
			Common Netspinner	5	0	2	7
		Diptera	True Fly Larvae	0	0	0	0
			Midge Larvae	0	0	1	1
			Crane Fly Larvae	0	0	0	0
			Black fly Larvae	0	0	0	0
		Meglaoptera	Dobsonfly Larvae	0	0	0	0
		Colopetra	Riffle Beetle	1	0	0	1
			Water Penny Larvae	0	0	0	0
		Odonata	Dragonfly Larvae	1	0	0	1
			Damselfly Larvae	0	0	0	0

Table 1: The table above represents the number of each of the organisms collected over three trials. These organisms were collected from Broad Run stream at coordinates (38.9774457°N, 77.5044787°W). The complete data table can be found here.

Results

After recording the organisms collected for each Trial in Table 1 above, the Water Quality Index (WQI) of the stream was calculated as shown in Table 2 below. The WQI identifies how well a certain stream is doing ecologically. If the WQI is between 9-12, then the stream has acceptable ecological conditions; if the WQI is 8, then the ecological conditions can not be determined; and if the WQI is 0-7, then the stream has unacceptable ecological conditions. The organisms were distributed into 6 metrics, with a certain number of organisms associated with that metric. Then, the percentage of organisms associated with each metric was calculated to find the metric value. Lastly, the metric value was then used to calculate the WQI. This can be calculated if the metric value follows this key in Table 3. Since Metrics 2 and 3 fit the criteria shown, they receive 2 points each, thus resulting in a WQI of 4.

Metric Distribution to calculate the WQI

Metric	Metric Organism Group	Number of Organism	/	Total Number	x 100	Metric Value	WQI
1	Mayflies	0	/	89	x 100	2.247	4
	Stoneflies	2					
	Caddisflies	0					
	Crane Flies	0					
	Dobsonflies	0					
2	Common Netspinner	7	/	89	x 100	7.865	
3	Lunged Snails	0	/	89	x 100	0	
4	Beetles	1	/	89	x 100	1.124	
	Water Pennies	0					
5	Aquatic Worms	2	/	89	x 100	87.64	
	Flatworms	0					
	Leeches	0					
	Sowbugs	0					
	Scuds	1					
	Dragonflies and Damselflies	1					
	Midge larvae	1					
	Black Fly Larvae	0					
	Clams/Mussels	73					
6	Aquatic Worms	2	/	89	x 100	86.52	
	Flatworms	0					
	Leeches	0					
	Crayfish	1					
	Sowbugs	0					
	Scuds	1					
	Gilled Snails	0					
	Lunged Snails	0					
	Clams/Mussels	73					

Table 2: The above data table represents the classification of the organisms collected into 6 metrics, and then calculating the metric value. These values were then compared to the classification key shown in Table 3 to determine the WQI of 4.

WQI Metric Classification Key

		Points		
Metric Number	Metric Value	2	1	0
1	2.247	> 32.2	16.1 - 32.2	< 16.1
2	7.865	< 19.7	19.7 - 34.5	> 34.5
3	0	< 0.3	0.3 - 1.5	> 1.5
4	1.124	> 6.4	3.2 - 6.4	< 3.2
5	87.64	< 46.7	46.7 - 61.5	> 61.5
6	86.52	< 5.4	5.4 - 20.8	> 20.8
	Total:	4		

Table 3: The table above represents the criteria that must be met to calculate the WQI of a given stream. For the coordinates (38.9774457°N, 77.5044787°W), the WQI of that stream is 4.

In addition to the WQI, the biodiversity of these streams must also be calculated. This can be calculated using the Shannon-Wiener Index (H'), as represented by Equation 1. The biodiversity measures how diverse the species in a certain ecosystem are. Therefore, the greater the biodiversity, the healthier the stream. The organisms recorded in Table 1 were then used to calculate the H' value, which was recorded in Table 4 below. The H' was calculated to be 0.785.

H' Equation

$$H' = \sum_{i=1}^{S} (p_i)(\ln p_i)$$

Equation 1: The equation above calculates the biodiversity of a given ecosystem, where S represents the total number of species and p_i represents the proportion of species i to the population. The H' for coordinates (38.9774457°N, 77.5044787°W) was calculated to be 0.785.

Lastly, the evenness (J') of the species in the streams were calculated using Equation 2 below. The J' value represents how evenly distributed the different species are. Therefore, the closer the value of J’ is to 1, the more evenly distributed the species are. From Table 1, the J’ was then calculated to be 0.357, which was then recorded in Table 4 below.

J’ Equation

$$J' = \frac{H'}{H'_{\max}} = \frac{H'}{\ln(S)}$$

Equation 2: The equation above calculates the evenness of an ecosystem, where S represents the total number of species, H’ (Shannon-Wiener index) represents the biodiversity of the species. The J’ for coordinates (38.9774457°N, 77.5044787°W) was calculated to be 0.357.

The WQI, H’, and J’ were also calculated for other streams throughout Loudoun County. There were a total of 28 different groups who followed the same procedure. The exact coordinates of the stream, WQI, H’, and J’ values are recorded in Table 4 below for each group.

Complied Theoretical Data

Group Number	Latitude	Longitude	WQI	H'	J'
1	38.966472	-77.526111	4	1.566	0.874
2	39.099056	-77.494541	10	0.386	0.278
3	38.96693	-77.52244	6	1.213	0.875
4	38.9784110	-77.5055520	9	1.968	0.767
5	38.967306	-77.519500	5	1.537	0.858
6	38.899661	-77.552223	10	0.941	0.982
7	38.935000	-77.568000	10	0.706	0.3627
8	39.06041	-77.5607	8	1.3778	0.8561
9	38.9657291	-77.5260178	5	1.316	0.818
10	38.9774457	-77.5044787	4	0.785	0.357
11	38.8563241	-77.5352735	8	1.004	0.9141
12	39.180475	-77.623835	8	2.25223	0.85342
13	38.9657351	-77.526048	9	1.573	0.716
14	39.0595527	-77.3335956	9	1.574	0.798
15	38.9767380	-77.5042856	8	1.737	0.969
16	38.9765433	-77.5044368	11	1.399	0.719
17	38.96727	-77.52193	4	1.402	0.783
18	39.099239	-77.4944764	5	1.228	0.533
19	38.966	-77.523	7	1.885	0.906
20	38.89925	-77.5520394	6	1.916	0.872
21	38.99281	-77.49233	5	1.834	0.797
22	39.0994	-77.4947	9	0.794	0.573
23	38.977455	-77.504633	6	0.787	0.716
24	38.978333	-77.505556	6	1.352	0.975
25	39.227988	-77.553025	2	0.8676	0.2807
27	39.0689	-77.449	2	0.95	0.59
28	39.105556	-77.561111	9	0.937	0.523

Table 4: The data table above represents the WQI, H', and J' values for 28 different locations in Loudoun County.

After the data was compiled into Table 4, the data was organized into 3 groups using ArcGIS based on location. The first group was streams located near Dulles and Brambleton, the second group was streams located near Leesburg, and the third group was streams located near Aldie and Chantilly. This can be seen in Figure 1 below. Some groups were left out because their location was further away from the three main locations.

Loudoun County Macroinvertebrate Data Collection Map Groupings

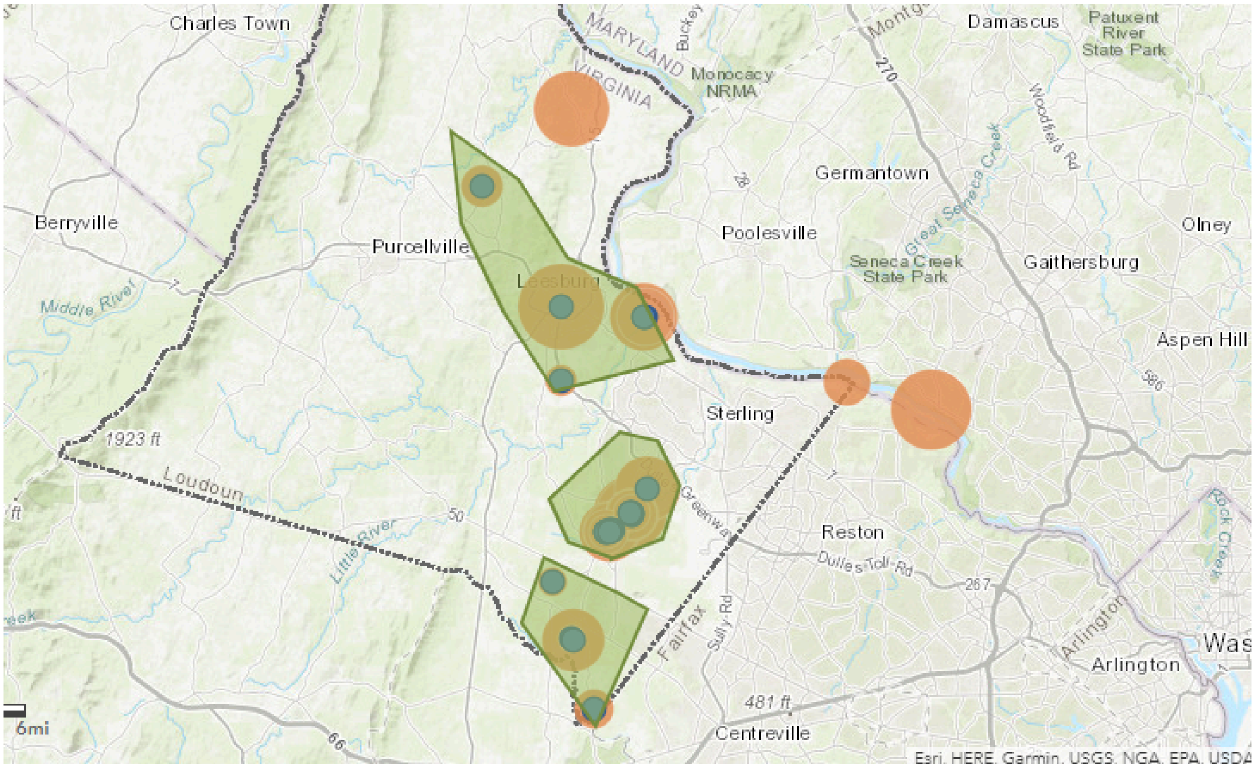

Figure 1: The figure above represents the three groupings of the WQI, H', and J' based on location. The polygons in green represent the area covered in the grouping, and the blue dots represent the exact location of each other data collection points inside the three designated polygons. The orange circles represent an approximate location of all of the data collection sites.

Then, 3 Kruskal-Wallis tests were performed in INSTAT to determine if there is a difference between the WQI, H', and J' values all throughout the three locations classified in Figure 1. These tests calculated the p-value, which was calculated to check if the true median WQI, H', and J' for each of the compared populations were truly similar. If the true medians were significantly different, the p-value would be less than 5%, which is the universal significance level for the Kruskal-Wallis. Though, if the p-value is more than 5%, there would be no statistically significant difference in the sample data in Table 4. These p-values are located in Table 5 below:

Kruskal-Wallis Tests for WQI, H’, and J’

	WQI		
	Dulles/ Rock Ridge/ Brambleton	Leesburg/ Purcellville	Aldie/ Chantilly
	4	10	10
	6	8	8
	9	8	8
	5	5	6
	5	9	
	4	9	
	9		
	8		
	11		
	2		
	7		
	5		
	6		
	6		
Mean	6.214285714	8.166666667	8
Variance	5.7197802	2.9666667	2.6666667
Group size n	14	6	4
H	3.984761905		
df	2		
p	0.1363703473		
a	0.05		
c	5.991		

	H'		
	Dulles/ Rock Ridge/ Brambleton	Leesburg/ Purcellville	Aldie/ Chantilly
	1.566	0.386	0.941
	1.213	1.3778	0.706
	1.968	2.25223	1.004
	1.537	1.228	1.916
	1.316	0.794	
	0.785	0.937	
	1.573		
	1.737		
	1.399		
	1.402		
	1.885		
	1.834		
	0.787		
	1.352		
Mean	1.453857143	1.162505	1.14175
Variance	0.13028829	0.40555011	0.28287225
Group size n	14	6	4
H	2.511428571		
df	2		
p	0.2848722953		
a	0.05		
c	5.991		

	J'		
	Dulles/ Rock Ridge/ Brambleton	Leesburg/ Purcellville	Aldie/ Chantilly
	0.874	0.278	0.982
	0.875	0.8561	0.3627
	0.767	0.85342	0.9141
	0.858	0.533	0.872
	0.818	0.573	
	0.357	0.523	
	0.716		
	0.969		
	0.719		
	0.783		
	0.906		
	0.797		
	0.716		
	0.975		
Mean	0.795	0.6027533333	0.7827
Variance	0.023320769	0.048918884	0.080453647
Group size n	14	6	4
H	4.446428571		
df	2		
p	0.1082605686		
a	0.05		
c	5.991		

Table 5: The table above represents the three Kruskal-Wallis tests performed to determine if there is a difference throughout the three locations from Figure 1.These tests were performed using INSTAT for each of the compared populations to determine the degrees of freedom, confidence intervals, and p-value. The established significance level for these Kruskal-Wallis tests were 0.05, therefore if the p-value is below that, then there is a significant difference in the true mean volumes for the three compared sample populations. Additionally, the mean values for the WQI, H’, and J’ are given for all three locations.

Discussions

Based on Table 5, since the p-value of 0.14 is greater than the 5% significance level, it can be concluded that there is no difference in the median WQI values. If the WQI medians were equal, there would be a 14% chance of getting a H - statistic this high or higher by sampling variability alone. There is insufficient evidence from our sample data to support a difference in WQI median values; therefore, it can be concluded that there is no statistically significant difference in the WQI median values across all 3 locations. Though, when comparing the mean values, it can be seen that streams in Leesburg and Aldie have a WQI average of approximately 8, therefore valid conclusions about the streams ecological conditions can not be made in this area. Though for Dulles and Brambleton, the average WQI is 6.2, which signifies unacceptable ecological conditions.

Additionally, based on Table 5, since the p-value of 0.29 is greater than the 5% significance level, it can be concluded that there is no difference in median H' values. If the H' medians were equal, there would be a 29% chance of getting a H - statistic this high or higher by sampling variability alone. There is insufficient evidence from our sample data to support a difference in H' median values; therefore, it can be concluded that there is no statistically significant difference in the H' median values across all 3 locations. When comparing the means, streams near Dulles and Brambleton had the greatest biodiversity of approximately 1.45, which is significantly greater than the other two locations.

Lastly, based on Table 5, since the p-value of 0.11 is greater than the 5% significance level, it can be concluded that there is no difference in median J' values. If the J' medians were equal, there would be a 11% chance of getting a H - statistic this high or higher by sampling variability alone. There is insufficient evidence from our sample data to support a difference in J' median values; therefore, it can be concluded that there is no statistically significant difference in the J' median values across all 3 locations. Additionally, while comparing the means, the streams near Dulles and Brambleton have the highest J' value, which means that the different organisms are more evenly distributed, as opposed to the streams near Leesburg, which has the lowest J' value of 0.6.

Though, assumptions and errors must be accounted for. To perform a Kruskal-Wallis test, two main assumptions were made: all groups are random and independent, and groups have similar statistical distributions. Additional error includes students not observing and recording all the organisms they caught using the D-net due to their small size. This may have caused the H', J’, and WQI calculations to be inaccurate. Additionally, since students were able to choose a location in Loudoun County to collect organisms from, there is a possibility that there were more trials conducted in a specific area opposed to another. Therefore, not all streams in Loudoun Country were able to be tested. Lastly, each of the groups of students had a different member conduct each trial, which caused inconsistent data collection because some members may have gathered a larger sample size as opposed to others. This may have also caused the biodiversity in any one location for that group to be greater than the rest of the locations. To further improve this experiment, the same group member should be collecting the macroinvertebrates each trial. This would increase consistency in data collection. Additionally, in the future, each of the groups should be assigned a stream so a larger area of Loudoun County could be tested rather than a large number of groups testing the same area.

In conclusion, there is no statistically significant difference in the WQI, H’, and J’ values; therefore, we can reject our original hypothesis because the biodiversity throughout Loudoun County was relatively similar. Additionally, the streams around Leesburg had the greatest average WQI, though the streams around Dulles and Brambleton have the greatest average H’ and J’. This means that the streams near Dulles and Brambleton have a greater biodiversity and these species are more evenly distributed than the streams near Leesburg. Though, the difference was not statistically large enough to draw a conclusion differentiating the biodiversity throughout all three locations. In the future, the conclusions from this experiment may be used to find any ongoing trends about the biodiversity in Loudoun County streams.

References

Ecosystem effects of biodiversity loss could rival impacts of climate change and pollution. (2012). Retrieved February 9, 2022, from https://www.vims.edu/newsandevents/topstories/archives/2012/nature_biodiversity.php

Hagan, J. G., Vanschoenwinkel, B., & Gamfeldt, L. (2021). We should not necessarily expect positive relationships between biodiversity and ecosystem functioning in observational field data. *Ecology Letters*, *24*(12), 2537–2548. https://doi.org/10.1111/ele.13874

Maureaud, A., Hodapp, D., van Denderen, P. D., Hillebrand, H., Gislason, H., Dencker, T. S., Beukhof, E. D., & Lindegren, M. (2019). Biodiversity-ecosystem functioning relationships in fish communities: biomass is related to evenness and the environment, not to species richness. *$2*. https://doi.org/10.1098/rspb.2019.1189

Hong, P., Schmid, B., De Laender, F., Eisenhauer, N., Zhang, X., Chen, H., Craven, D., De Boeck, H. J., Hautier, Y., Petchey, O. L., Reich, P. B., Steudel, B., Striebel, M., Thakur, M. P., & Wang, S. (2022). Biodiversity promotes ecosystem functioning despite environmental change. *Ecology Letters*, *25*(2), 555–569. https://doi.org/10.1111/ele.13936

Azrina, M. Z., Yap, C. K., Rahim Ismail, A., Ismail, A., & Tan, S. G. (2006). Anthropogenic impacts on the distribution and biodiversity of benthic macroinvertebrates and water quality of the Langat River, Peninsular Malaysia. *Ecotoxicology and Environmental Safety*, *64*(3), 337–347. https://doi.org/10.1016/j.ecoenv.2005.04.003