Title: Following Rapoport's Rule: The geographic range and genome size of bacterial taxa decline at warmer latitudes.

Running title: Stream bacterial communities follow Rapoport's Rule

[1]Gavin Lear, [1]Kelvin Lau, [1,+]Anne-Marie Perchec, [2]Hannah L. Buckley, [2]Bradley S. Case, [3]Martin Neale, [4]Noah Fierer, [4]Jonathan W. Leff, [1]Kim M. Handley, [1]Gillian Lewis

[1]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand.
[2]School of Science, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand
[3]Martin Jenkins, Auckland 1010, New Zealand.
[4]Department of Ecology and Evolutionary Biology and Cooperative Institute for Research in Environmental Biology, University of Colorado at Boulder, Boulder, CO, USA, 80309-0216

[+]current address: Ministry for Primary Industries, PO Box 2526, Wellington 6140, New Zealand.

Correspondence: Dr. Gavin Lear, School of Biological Sciences, The University of Auckland, 3a Symonds Street, Auckland 1010, New Zealand.
E-mail: g.lear@auckland.ac.nz.
Tel: +64-9-3739799

Conflict of Interest: The authors declare no conflict of interest.

SIGNIFICANCE STATEMENT

Bacterial communities conform to Rapoport's Rule, exhibiting reduced range sizes at warmer latitudes. The average latitudinal range of freshwater bacteria declined by 28 km for every 100 km north travelled across New Zealand. Using genome size as a proxy for metabolic versatility, we show that reductions in range sizes at warmer latitudes are likely correlated with reduced tolerance to variation in environmental conditions. These findings highlight the universality of Rapoport's Rule across multiple domains of life and the likely importance of genome size in determining microbial distributions.

ABSTRACT

We sought to test whether stream bacterial communities conform to Rapoport's Rule, a pattern commonly observed for plants and animals whereby taxa exhibit decreased latitudinal range sizes closer to the equator. Using a DNA sequencing approach, we explored the biogeography of biofilm bacterial communities in 204 streams across a ~1,000km latitudinal gradient. The range sizes of bacterial taxa were strongly correlated with latitude, decreasing closer to the equator, which coincided with a greater than five-fold increase in bacterial taxonomic richness. The relative richness and range size of bacteria were associated with spatially-correlated variation in temperature and rainfall. These patterns were observed despite enormous variability in catchment environmental characteristics. Similar results were obtained when restricting the same analyses to native forest catchments, thereby controlling for spatial biases in land use. We analysed genomic data from ~500 taxa detected in this study, for which data were available and found that bacterial communities at cooler latitudes also tended to possess greater potential metabolic potential. Collectively, these data provide the first evidence of latitudinal variation in the range size distributions of freshwater bacteria, a trend which may be determined, in part, by a trade-off between bacterial genome size and local variation in climatic conditions.

INTRODUCTION

Bacteria comprise a major component of the biomass and diversity in rivers and streams where they play essential roles as primary producers, decomposers and pathogens[1]. The relative abundance and distribution of bacteria vary considerably across global freshwater systems and yet the factors controlling their biogeography remain poorly understood. Microbial communities are impacted by a wide variety of environmental factors, responding strongly to drivers including temperature[2] and pH[3]. In addition to these 'niche-based' patterns, distance-based gradients in bacterial community composition are frequently reported[4], suggesting a possible role for dispersal limitation in shaping microbial community structure. With both niche processes and dispersal constraints predicted to drive the assembly of environmental microbial communities, there is growing interest in determining whether prokaryotes conform to the same ecological patterns commonly observed for multicellular eukaryotes[5], patterns that have served as the foundation for much of the research in macroecology and biogeography.

The increased richness of plant and animal species from the poles towards the equator is one of the strongest and most consistent large-scale ecological patterns observable on earth. The existence of these gradients has been documented for a wide range of multicellular organisms[6,7], and more recently in unicellular organisms[4,8]. Latitudinal gradients in species richness have been attributed to factors including greater primary production potential[9] and higher speciation rates[10] closer to the equator. 'Rapoport's Rule'[11] provides an alternative, but not contradictory explanation for latitudinal gradients in species richness, based on the principal of changing environmental tolerance with latitude. It is hypothesised that organisms living further from the equator are exposed to, and thus have more tolerance to, a wider range of environmental conditions. At warmer latitudes, the reduced seasonal variability in environmental conditions is predicted to favour organisms that have developed more specialised habitat requirements. This may permit the coexistence of a greater number of species while also restricting the latitudinal range of these more specialised taxa. Rapoport's Rule of decreasing species ranges closer to the equator has been observed for a wide variety of organisms[12,13]. However, the generality of Rapoport's Rule remains equivocal, receiving mixed support and suggestions that this trend could be restricted to certain systems, taxa

and geographic areas[14,15]. Rapoport's Rule has never, to our knowledge, been directly tested in populations of free-living [i.e., non-symbiotic] non-marine bacterial communities. Exploration of latitudinal gradients in bacterial community attributes provides substantial opportunities to assess if Rapoport's 'Rule' is also observable for microbial community data.

Using stream biofilm bacterial community data collected across a latitudinal range of ~1,000 km, we asked three key questions. First, do bacterial communities adhere to Rapoport's Rule of decreasing latitudinal range closer to the equator? The answer to this question has important implications for understanding if Rapoport's Rule may be universal, or at least observable among communities spanning multiple domains of life, including prokaryotes. Second, how important are natural, versus anthropogenic factors, in explaining latitudinal gradients in bacterial geographic ranges and richness? Evidence demonstrating that non-anthropogenic factors, such as climate, are correlated with latitudinal patterns in taxon ranges and richness would suggest the existence of a pervasive Rapoport effect on bacterial communities. Finally, is there a relationship between latitude and average bacterial genome size? Confirmation of a positive relationship would indicate that bacterial range size may indeed be impacted by their tolerance to environmental change, since organisms with larger genomes could be expected to exhibit greater versatility to changing environmental conditions[16-19].

MATERIALS AND METHODS

Sample Collection and Molecular Analysis

Stream biofilm samples were removed from a total of 204 New Zealand stream sites and processed as described in Lear *et al*[4]. We chose to collect stream epilithic biofilm since these bacterial communities are easy to sample and are relatively sessile. This means that biofilm community responses are more likely to reflect changes in local environmental conditions, as compared to bacteria within the water column which are typically transient in nature, originating further upstream. At each site, biofilm biomass was scraped from the upper surface of five rocks (within a 10-m reach) during the Austral summer season (February-March 2010). The method of Miller et al[20] was used to extract

DNA from 0.25-g of each pelleted biofilm sample. To characterise the diversity and composition of biofilm bacterial communities at each site, the V4/V5 region of the 16S rRNA gene was amplified using the Archaea- and Bacteria-specific primer set 515f/806r[21]. Primers were modified to include Illumina flowcell adapter sequences. Reverse primers, specific to each sample contained 12-bp error-correcting barcodes to allow for sample multiplexing[21]. PCR amplification was completed in triplicate for each sample using the following thermocycling routine: (1) 94°C for 3 min, (2) 35 cycles of 94°C for 45 s, 50°C for 60 s, and 72°C for 90 s, and (3) 72°C for 10 min. Products were then pooled and purified using SequelPrep Normalization kits (Invitrogen Ltd., NZ) and run on an Illumina MiSeq instrument using 2 x 150 bp chemistry.

Bioinformatics Analyses

Following an approach similar to Ramirez et al.[22], paired end sequences were demultiplexed, merged, quality filtered and clustered into operational taxonomic units (OTUs) using the UPARSE pipeline[23], based on a sequence identity threshold equal to or greater than 97%. Prokaryote OTUs were classified to corresponding taxonomy using the RDP classifier[24] trained on the Greengenes 13_8 database[25]. After quality checking and the removal of chloroplast and mitochondrial DNA, all samples were rarefied to 5,000 randomly selected reads per sample and three samples per site to standardize sequence depth and sampling effort across sites. We deposited all amplicon sequence data associated with this manuscript in the NCBI Sequence Read Archive under accession number SRP328535.

We adapted the method of Barberán et al.[18], using the USEARCH algorithm[26] to determine which OTUs were matches to those bacterial taxa for which whole genome data are available. To achieve this, we matched representative 16S rRNA gene sequences from the present study against the 5,424 complete bacterial genomes in the GenBank database (02-Jun-15); sequences were considered to match if they shared ≥97% identity. For 522 of the original 34,000 OTUs for which full genome data were available, we downloaded genome size information from the Integrated Microbial Genomes (IMG[27]) database.

Spatial Analysis of Upstream Catchment Data

We coupled our bacterial dataset with detailed upstream catchment information collated for each sampling location using GIS overlay procedures available within ARCGIS 10.3, as described in Lear et al[4]. This approach was used to generate (1) land cover, (2) climate, and (3) soil taxonomic and chemical information for the catchment area upstream of each sampling location. Relative proportions of the different soil taxonomic groups were subjected to a data reduction procedure using principal coordinates analysis (PCoA); the individual sample data scores along the first three PCoA axes, which explained 41.2% of the variance in this dataset, were extracted and provided the univariate data used in subsequent analyses. A list of all explanatory variables included in this study is provided in Table S1. We used the NZTM2000 geodetic coordinate system for our analyses. With this system, the standard units of latitude, termed northings, are measured in metres and values increase closer to the equator.

Quantitative Methods

Our analyses used either bacterial (i) richness (the average number of different OTUs in each sample), (ii) mean latitudinal range (the average latitudinal range of all OTUs at each site), (iii) composition (the 'bacterial abundance matrix' was the number of reads for each OTU at each site) or (iv) mean compositional dissimilarity (calculated as the average Bray-Curtis distance for each sample compared to all of the samples using the bacterial abundance matrix). To test for statistically significant variance among multivariate bacterial community data obtained from different regions of New Zealand, permutational multivariate analyses of variance (PERMANOVA) were completed by comparing Bray-Curtis distances among sample profiles using an 'add-on' package for PRIMER, PERMANOVA+[28]. We used the 'glm' function in R (version 3.2.0)[29] to fit quasi-Poisson regression models of the relationship between average taxon richness and either the absolute latitude or elevation of the sample site. Scatter plots showing the relationship between sample site distance (or altitude) and community similarity were plotted and the 'heatscatter' function in the R package 'LSD' used to show the extent to which data points overlapped. To determine the mean latitudinal range size of the

bacterial taxa identified in each sample, we first determined the latitudinal range of each OTU identified in this study (i.e., the latitudinal distance, in km, between the northernmost and southernmost location that each OTU was found). We then calculated the average latitudinal range of the 5,000 bacteria (or DNA sequence reads) identified in each sample. To test for Rapoport's Rule, we correlated the mean latitudinal range size of taxa at each site with the latitude that each community was sampled from. To assess potential biases on range size calculations that may be caused by the geographic distribution of our sample sites, we used a randomised resampling approach and recalculate the mean latitudinal range size and regression statistics; we repeated this procedure 199 times. Specifically, for each permutation of the data we randomly shuffled sample data among all 204 sites and recalculated the strength of correlation between the latitude of each sample site and the mean latitudinal range size of taxa at each site (now linked to randomly allocated sample OTU data).

Variance partitioning procedures outlined in Borcard et al.[30] were used to indicate how much total variation in the bacterial community data can be explained by each of four groups of explanatory variables: 'Space', 'Climate', Soil' and 'Catchment', as well as the component of shared variation, e.g., spatially structured variation in environmental variables. Spatial variables were derived using the Moran's eigenvector maps (MEMs) procedure[30] implemented using the 'pcnm' function in the R package 'vegan', which uses a principal coordinates analysis to represent different scales of spatial variation for the given set of sample locations[30]. The variance partitioning procedure computes R^2 canonical values analogous to the adjusted R^2 values produced in multiple regression. The multivariate sample data were also related to explanatory matrices of spatial and environmental data using distance-based redundancy analysis (db-RDA) and a forward selection procedure with the 'capscale' and 'ordistep' functions in the 'vegan' package in R. These data were used to calculate multivariate regression trees using the 'ctree' package, and also to determine the threshold values of key soil, catchment or climate related attributes most closely associated with each nodal data cluster in the tree.

RESULTS

Spatial trends in the bacterial community data

Significant latitudinal trends were detected in mean bacterial range size, taxon richness and composition across New Zealand. Overall, the similarity in bacterial community composition between any pair of samples decreased with increasing geographic distance (Fig. S1). Significant differences in bacterial community composition were detected among all regions studied (pairwise PERMANOVA comparing data from all regions, all $P < 0.001$) and communities in the regions separated by the greatest geographic distance, AK and CB were the least similar. Although located approximately 600 km apart on different islands, the regions CB and HR which each contain large tracts of relatively flat agricultural land, shared the greatest average similarity in bacterial community composition (Fig. S2). This resulted in an obvious deviation from an otherwise largely monotonic distance decay relationship (Fig S1; Fig. 1).

We confirmed that stream samples located further north (closer to the equator) supported a greater average richness of bacterial taxa (Fig. 1). The average OTU richness for each sample in the northernmost region (AK) was $1,400 \pm 41$ ($n = 60$), compared to only 345 ± 29 ($n = 39$) for the southernmost region (CB). The maximum number of OTUs detected in any one sample was 1,988. Overall, taxon richness increased by 17% to 23% per 100 km north travelled (estimated from the 95% confidence intervals of a quasi-Poisson regression model). The number of taxa that could only be found in the northernmost third of samples was six times greater than the number of taxa that could only be detected in the southernmost third (Fig S3).

We present clear evidence of a latitudinal gradient in mean bacterial taxonomic range (Fig. 2a), supporting the existence of Rapoport's Rule for freshwater bacterial communities. The mean latitudinal range of the OTUs in each sample was strongly correlated with the latitudinal location of sample collection (Spearman's rank correlation coefficient, rho = -0.83, $P < 0.001$), declining closer to the equator. Overall, the average latitudinal range of the OTUs identified in each sample declined by 28 km for every 100 km north travelled. The mean correlation detected between site latitude and mean latitudinal range after sample data were randomly shuffled between sites was far weaker (rho = -0.12; Figure S4). This indicates that an uneven sample site distribution (i.e., with sample sites located in more northern regions of New Zealand being in closer proximity

to each other than those in more southerly regions) is sufficient to drive a spurious Rapoport effect. Nevertheless, the influence of this effect on the computed latitudinal ranges is weak compared to the trends observed for the underlying bacterial community data.

Additionally, we found evidence of a weak, but significant, relationship between the average genome size of each bacterial community and site latitude (Spearman's rho = -0.25, P < 0.001). Bacteria within sample sites located south of northing 5400000 had genomes that were, on average, over 0.6 Mbp larger than those located north of northing 5750000 (Fig. 2b).

Stream sampling locations spanned a range of elevations from sea level (0 m) to 756 m. Overall, taxon richness declined by an average of 10 OTUs for every ten metre increase in elevation above sea level (Fig. 3) across this range. Sample elevation was poorly correlated to latitude (PPMC r^2 = 0.08, P < 0.004) indicating that the significant impacts of elevation and latitude occur independent of each other. A quasi-regression model provided strong evidence elevation is a reliable predictor of the mean elevational range of OTUs (P < 0.001); range size increased with elevation. Despite this decline in taxonomic richness and increase in range with elevation, no clear decay in mean bacterial compositional similarity could be attributed to variation in site elevation, possibly caused by the relatively low number of high elevation samples.

A multitude of anthropogenic factors have been shown to cause spatial variation in microbial community composition and diversity, particularly in stream bacterial communities[31]. Since the human population density of New Zealand is skewed towards the north of the country, a greater abundance of urban land uses as well as an increased diversity of rural land use types in the upper North Island could impact latitudinal trends in bacterial community attributes. For this reason, we also chose to repeat our analysis to include only data from catchments dominated by native forest land uses, and with few recorded urban areas. Latitudinal declines in species range were in fact stronger (i.e., the slope of the trend line steeper) for data obtained from native forest catchments (Fig. 4) than as observed for the wider data set (Fig. 1) or for catchments dominated by urban and rural land uses. Significant gradients in taxon richness were similarly detected from the native forest catchment data.

The relative role of spatial and environmental factors in determining latitudinal gradients in bacterial taxon range, richness and composition.

As Rapoport's Rule of decreasing taxonomic range at higher latitudes is hypothesised to be driven by latitudinal gradients in environmental, and particularly climate variables, we quantified the relative importance of natural environmental and climatic factors versus catchment land use factors in driving the observed community patterns. Variance partitioning using Moran's eigenvector maps (MEMs) confirmed that catchment land use factors alone were not significant predictors of the variability in latitudinal range size. However, spatially structured variation in catchment land use accounted for 34% of the observed variation in bacterial range size distributions (Fig. S5). Spatially structured variation in climate and soil factors explained a further 49% of the observed variance in bacterial range sizes. Similar patterns were observed for bacterial community richness, where catchment land use factors alone accounted for just 1% of the variability in bacterial community composition, but spatially structured variation in catchment land use attributes accounted for 36% of observed variation. The explanatory variables measured explained far more of the variability in the richness and latitudinal range size of bacterial OTUs than their compositional similarity (the percentage of unexplained variance was 16%, 19% and 70%, for mean range size, richness and compositional similarity, respectively; Figs S5 and S6).

According to the multivariate regression tree (Fig. S7), the greatest mean latitudinal range size was observed at sites with the highest seasonal variation in temperature. Within sites exposed to higher seasonal variation (>17°C difference between mean summer and winter atmospheric temperature), latitudinal range was also significantly correlated with precipitation variation. The main determinant of bacterial taxon richness was also seasonal temperature variation; greater richness was detected in sites with less temperature fluctuation and in catchments containing a lower diversity of land uses (Fig. S8). Finally, variation in precipitation was significantly and positively correlated to the relative richness of OTUs in sites also exposed to greater variation in temperature.

DISCUSSION

Since the term was first coined by Stevens et al[11], Rapoport's Rule has motivated ecologists to explore latitudinal gradients in the geographic ranges of animals and plants[32]. Many studies have observed trends consistent with this effect[33,34], but there also appear to be important exceptions to this rule[35] which has stoked healthy debate. Long since Gaston et al[14] questioned the need for an 'epitaph' for Rapoport's Rule, the existence of latitudinal trends in microbial species ranges remains poorly studied. A key reason for this has been the methodological limitations associated with trying to characterize the enormous diversity of complex natural microbial communities. However, advances in DNA sequencing provide new opportunities to explore latitudinal gradients in microbial diversity by more comprehensively characterizing the abundances and distributions of diverse environmental microbial populations. Here, in support of recent studies by Amend *et al.*[36] and Sul *et al.*[37], we confirm the existence of latitudinal gradients in bacterial range in Southern Hemisphere samples. Due to the geographic constraints of our study, we remain unable to confirm if minimal range size is reached in equatorial regions. Nevertheless, our findings, generated from over 40,000 pairwise comparisons of the sample data contribute towards a growing body of evidence that, similar to communities of macroorganisms, the richness and average range size of aquatic microbial communities is strongly related to latitude[8,38]. We believe this to be the first time such a relationship has been shown among free-living non-marine bacterial communities. Further, we provide evidence of a significant latitudinal gradient in bacterial genome size, implying a characteristic reduction in microbial metabolic versatility among communities exposed to less varied climate conditions. Collectively, these findings provide valuable insights into the potential mechanisms regulating the distribution and diversity of bacterial communities across the globe.

Gradients in environmental factors, including temperature and precipitation, have been proposed as the driving forces behind Rapoport's Rule[11]. Greater seasonal variability at cooler latitudes is proposed to select for greater climatic tolerances and therefore wider latitudinal range sizes. The greater latitudinal range of bacterial OTUs located in sites with increased variation in annual precipitation and temperature suggests that bacterial range size distributions may indeed be driven by latitudinal gradients in

environmental tolerance. Since climate is all that consistently varies among sites separated by latitude or elevation, extending Rapoport's Rule to elevation was a simple way of exploring how species richness and range size may vary across other climate gradients. Our observation of strong latitudinal *and* elevational gradients in taxon richness and range size firmly implicates climate as a primary determinant of microbial range size distributions.

Numerous alternative mechanisms have been proposed to drive declines in average range size from high to low latitudes and elevations. In particular, geographic range sizes are suggested to be truncated by barriers to dispersal that are unrelated to species tolerances[39]. Geographic features such as mountain ranges and the open ocean can decrease the potential range of species that would occur had there been no barrier to dispersal. These barriers to dispersal have been suggested by some to exert a stronger effect on the latitudinal extent of species than climatic factors[40]. Additionally, dispersal barriers are suggested as possible reasons as to why latitudinal range size distributions are reported more frequently in the northern hemisphere, as decreased land area coincides with reductions in species ranges closer to the equator[41]. Since dispersal constraints are likely to be less important for non-symbiotic bacterial taxa[42] as compared to higher plants and animals, evidence of latitudinal gradients in range size distributions for bacteria indicates that strong barriers for dispersal are not essential to observe Rapoport's Rule. Importantly, we observed no truncation in taxon range related to the location of Cook Straight, which divides the North and South Islands of New Zealand and is over 22 km wide at its narrowest point. Together, these findings contribute to a growing body of evidence that bacterial distributions are little impacted by geographic barriers and that barriers to dispersal are not essential to observe Rapoport's Rule.

Decreased range sizes at lower latitude have also been suggested to be caused by differential rates of extinction, possibly correlated with the range and duration of glacial maxima.[43]. This idea was disputed by Sax et al.[32] who observed patterns in the geographic range of endemic, as well as exotic macroorganisms; the latter could only reasonably be limited by recent conditions in their local environment and not previous climate landscapes to which they were not exposed. The findings of the present study support the conclusions of Sax et al[32] since rapid rates of speciation and dispersal in

bacterial communities make it highly unlikely that modern day bacterial distributions are dominated by past glaciation events.

A further alternative explanation for Rapoport's Rule is that the greater geographic range exhibited by species living at cooler latitudes is a consequence of reduced competition, which may be intrinsically linked to lower species richness at these latitudes. Again, this is deemed unlikely to be the sole, or dominant driver of Rapoport's Rule[14], not least because many species present at higher latitudes are also present at lower latitudes, as was observed in the present study (Fig S3). Finally, since gradients in resource availability are known to impact species distributions, they too have been proposed to determine latitudinal gradients in species richness and range[44]. While data on C and N resources in stream water columns are limited, substantial data on concentrations of C and N in the catchment soil reveal no clear latitudinal gradient in these resources across the New Zealand landscape[45]. Additionally, we found no evidence to indicate that concentrations of C or N in upstream catchment soils are significant determinants of bacterial range or richness. It is clear that no one single factor exists to explain all of the observed spatial variation in bacterial latitudinal range size distributions, with multiple predictors (and their interactions) likely to operate at different spatial and temporal scales. Additionally, it is important to note that factors including latitudinal gradients in sampling intensity are capable of generate spurious Rapoport patterns. Nonetheless, with variation in average range size distributions correlating most strongly with variation in average temperature, precipitation and solar radiation, perhaps the most parsimonious explanation is that the observed latitudinal gradients in bacterial range size are dominated by latitudinal gradients in climate variation.

Confirmation of significant latitudinal gradients in environmental tolerance can be seen as an important step for understanding the likely mechanisms of Rapoport's Rule. Despite the presumed importance of environmental tolerance for describing species range size distributions, latitudinal gradients in environmental tolerance have largely been inferred but not tested. Capitalising on recent increases in the availability of microbial genome data, we confirmed for the first time, the existence of a latitudinal gradient in microbial genome size, which we use here as a proxy for metabolic versatility. Genome reduction is a characteristic trait of symbiotic organisms[46] which dwell in the relatively

constant environment provided by the host. Also predicted and observed in free-living prokaryotes, larger genomes are associated with increased metabolic versatility, allowing bacteria to produce a greater range of enzymes such that they may tolerate and exploit a wider range of environmental conditions[16,17]. Positive correlations have previously been reported between genome size and the number of sites any bacteria may occupy. A study of ~600 soil samples collected across Central Park in New York[18] indicated a tendency for organisms inhabiting a greater range of environments to have larger and potentially more versatile genomes than those organisms with restricted distributions. The present study confirms that such occupancy patterns also occur across much larger spatial scales (i.e., >1000 km) and may cause, or be driven by, latitudinal gradients in average genome size. This latitudinal trend in presumed metabolic versatility is particularly important since gradients in environmental tolerance form a necessary tenet of Rapoport's Rule. Additionally, whereas predicted range sizes are inevitably impacted, to some degree, by sample size artefacts, average genome size data are independently calculated for each sample. Such data are far less likely to generate spurious Rapoport patterns[15] related to sampling intensity and distribution. Thus, we demonstrate the utility of microbial genome data for interpreting ecological patterns and their drivers, even from large sample data and across large spatial scales.

CONCLUSION

The range size distribution of freshwater biofilm bacteria conform to Rapoport's Rule. Lawton[47] redefined Rapoport's 'Rule' as a 'pattern', highlighting that the increased range size of organisms at cooler latitudes may not be universal across all samples and taxonomic groups. Nevertheless, it would appear that Rapoport's Rule is not only regularly detected for macroorganisms, but that this same pattern can be demonstrated for freshwater bacterial communities. As suggested by Amend *et al.*[36], the similarities in Rapoport's Rule for communities of macro- and microorganisms provides no proof that the same underlying processes shape their patterns in range and richness. However, given the propensity of microbial taxa for long-range dispersal and migration, it is intuitive to suggest that geographic boundaries for dispersal are not required to observe Rapoport's Rule. While there are likely to be multiple drivers of latitudinal gradients in species' range sizes, we provide the first evidence of a linkage between latitudinal range size

distribution and genome size, used here as a proxy for metabolic versatility. Comparing latitudinal gradients of genome size across a wider range of taxonomic groups may provide a fruitful avenue for future research into the underlying drivers of Rapoport's Rule.

ACKNOWLEDGEMENTS.

REFERENCES

1   Lear, G. *et al.* Biofilms in freshwater. In: Microbial Biofilms, Current Research and Applications (Eds. Lear G, Lewis GD). 129-151 (Caister Academic Press, Norfolk, UK, 2012).

2   Yannarell, A. C. & Triplett, E. W. Within- and between-lake variability in the composition of bacterioplankton communities: investigations using multiple spatial scales. *Appl Environ Microbiol* 70, 214-223 (2004).

3   Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* 75, 5111-5120 (2009).

4   **Lear, G.** *et al.* The biogeography of stream bacteria. *Global Ecol Biogeogr* 22, 544-554 (2013).

5   Bryant, J. A. *et al.* Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *P Natl Acad Sci USA* 105, 11505-11511 (2008).

6   Roy, K., Jablonski, D., Valentine, J. W. & Rosenberg, G. Marine latitudinal diversity gradients: Tests of causal hypotheses. *P Natl Acad Sci USA* 95, 3699-3702 (1998).

7   Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *P Natl Acad Sci USA* 104, 5925-5930 (2007).

8   Fuhrman, J. A. *et al.* A latitudinal diversity gradient in planktonic marine bacteria. *P Natl Acad Sci USA* 105, 7774-7778 (2008).

9   Hawkins, B. A., Porter, E. E. & Felizola Diniz-Filho, J. A. Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology* 84, 1608-1623 (2003).

10   Mittelbach, G. G. *et al.* Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol Lett* 10, 315-331 (2007).

11   Stevens, G. C. The latitudinal gradient in geographical range: how so many species coexist in the tropics. *Amer Nat* 133, 240-256 (1989).

12   Whitton, F. J. S., Purvis, A., Orme, C. D. L. & Olalla-Tarraga, M. A. Understanding global patterns in amphibian geographic range size: does Rapoport rule? *Global Ecol Biogeogr* 21, 179-190 (2012).

13   Arita, H. T., Rodriguez, P. & Vazquez-Dominguez, E. Continental and regional ranges of North American mammals: Rapoport's rule in real and null worlds. *J Biogeog* 32, 961-971 (2005).

14   Gaston, K. J., Blackburn, T. M. & Spicer, J. I. Rapoport's rule: time for an epitaph? *Trends Ecol Evol* 13, 70-74 (1998).

15   Colwell, R. K. & Hurtt, G. C. Nonbiological gradients in species richness and a spurious Rapoport effect. *The American Naturalist* 144, 570-595 (1994).

16   Bentkowski, P., Van Oosterhunt, C. & Mock, T. A model of genome size evolution for prokaryotes in stable and fluctuating environments. *Genome Biology and Evolution* 7, 2344-2351 (2015).

17   Ranea, J. A., Grant, A., Thornton, J. M. & Orengo, C. A. Microeconomic principles explain an optimum genome size in bacteria. *Trends in Genetics* 21, 21-25 (2005).

18   Barberan, A. *et al.* Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecol Lett* 17, 794-802 (2014).

19   Guieyesse, B. & Wuertz, S. Matebolically versatile large-genome prokaryotes. *Current Opinion in Biotechnology* 23, 467-473 (2012).

20   Miller, D. N., Bryant, J. E., Madsen, E. L. & Ghiorse, W. C. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl Environ Microbiol* 65, 4715-4724 (1999).

21   Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6, 1621-1624 (2012).

22   Ramirez, K. S. *et al.* Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proc R Soc B* 281, 20141988 (2014).

23   Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10, 996-998 (2013).

24   Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73, 5261-5267 (2007).

25   McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary anaylses of Bacteria and Archaea. *ISME J* 6, 610-618 (2012).

26   Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460-2461 (2010).

27   Markowitz, V. M. *et al.* IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* 40, D115-D122 (2012).

28   Anderson, M. J., Gorley, R. N. & Clarke, K. R. *PERMANOVA+ for PRIMER: guide to software and statistical methods. PRIMER-E Ltd., Plymouth, United Kingdom.* (2008).

29   Oksanen, J. *et al.* Vegan: Community Ecology Package version 2.0-10. http://CRAN.R-project.org/package=vegan. (2013).

30   Borcard, D., Gillert, F. & Legendre, P. *Numerical Ecology with R.* (Springer, 2011).

31   Lau, K. E. M. *et al.* A novel bacterial community index to assess stream ecological health. *Freshwat Biol* 60, 1988-2002 (2015).

32   Sax, D. F. Latitudinal gradients and geographic ranges of exotic species: implications for biogeography. *J Biogeog* 28, 139-150 (2001).

33   Letcher, A. J. & Harvey, P. H. Variation in geographical range size among mammals of the Palearctic. *Amer Nat* 144, 30-42 (1994).

34   Kaufman, D. M. Diversity of New-World mammals - Universality of the latitudinal gradients of species and bauplans. *J Mammal* 76, 322-334 (1995).

35   Tomasovych, A., Jablonski, D., Berke, S. K., Krug, A. Z. & Valentine, J. W. Nonlinear thermal gradients shape broad-scale patterns in geographic range size and can reverse Rapoport's rule. *Global Ecol Biogeogr* 24, 157-167 (2015).

36   Amend, A. S. *et al.* Macroecological patterns of marine bacteria on a global scale. *J Biogeog* 40, 800-811 (2012).

37   Sul, W. J., Oliver, T. A., Ducklow, H. W., Amaral-Zettler, L. A. & Sogin, M. L. Marine bacteria exhibit a bipolar distribution. *P Natl Acad Sci USA* 110, 2342-2347 (2013).

38   Ladau, J. *et al.* Global marine bacterial diversity peaks at high latitudes in winter. *ISME J* 7, 1669-1677 (2013).

39   Sizling, A. L., Storch, D. & Keil, P. Rapoport's rule, species tolerances, and the latitudinal diversity gradient: geometric considerations. *Ecology* 90, 3575-3586 (2009).

40   Roy, K., Jablonski, D. & Valentine, J. W. Eastern Pacific molluscan provinces and latitudinal diversity gradient: No evidence for "Rapoport's rule". *P Natl Acad Sci USA* 91, 8871-8874 (1994).

41   Orme, C. D. L. *et al.* Global patterns of geographic range size in birds. *PLos Biology* 4, 1276-1283 (2006).

42   Fenchel, T. & Finlay, B. J. The ubiquity of small species: patterns of local and global diversity. *Bioscience* 54, 777-784 (2004).

43   Jansson, R. Global patterns in endemism explained by past climatic change. *Proceedings: Royal Society* 270, 583-590 (2003).

44   Bonn, A., Storch, D. & Gaston, K. J. Structure of the species-energy relationship. *Proc R Soc B* 271, 1685-1691 (2004).

45   Scott, N. A. *et al.* Monitoring land-use change effects on soil carbon in New Zealand: quantifying baseline soil carbon stocks. *Environ Pollut* 116, S167-S188 (2002).

46   Gottlieb, Y., Lalzar, I. & Klasson, L. Distinctive genome reduction rates revealed by genomic analysis of two *Coxiella*-like endosymbionts in ticks. *Trends in Genetics* 7, 1779-1796 (2015).

47   Lawton, J. H. Are there general laws in ecology? *Oikos* 84, 177-192 (1999).

FIGURES

Figure 1. Relationship of bacterial community composition and average OTU richness with latitude. (a) Map showing the OTU taxon richness at each sampling location. Data are OTU richness obtained from 5,000 DNA sequences per sample (● 150 - 300; ● 300 – 600; ● 600 – 1000; ● 1000 -1500; ● 1500-2000). A total of 204 streams were sampled within seven administrative districts, each identified using two-letter codes (AK, Auckland Region; WK, Waikato Region; HR, Horizons [Manawatu-Wanganui] Region; HB, Hawkes Bay Region; WG, Greater Wellington Region; TS, Tasman District; CB, Canterbury Region) across the North and South Islands of New Zealand; (b) Relationship between dissimilarity of bacterial OTU data (in multivariate space, using a Bray-Curtis measure) and latitudinal distance between samples (in km). Colours on the scatter plot indicate the extent to which points on the plot overlap on a gradient from light grey (no overlap) through to yellow (substantial overlap). Also plotted on the figure is a solid black trendline showing differences in mean Bray Curtis dissimilarity, calculated every 20 km. (c) Relationship between bacterial taxon richness (i.e., number of distinct OTUs obtained from 5,000 DNA sequences per sample) and latitude (measured as distance north of the southernmost sampling location (in km) for samples collected within various New Zealand regions (● AK; ● CB; ● HB; ● HR; ● TS; ● WG; ● WK). Regression line shows the fit of a quasi-Poisson regression model. Error bars are 1 x SE.

Figure 2. Relationship between sample site latitude (in degrees Northings) and mean latitudinal range of bacterial OTUs, or genome size (in Mbp), for samples collected within various New Zealand regions (● AK; ● CB; ● HB; ● HR; ● TS; ● WG; ● WK). (a) Mean latitudinal range of bacterial OTUs; line of best fit is $y = -0.33x + 2350$, $R^2 = 0.70$, $P < 0.001$. (b) Average genome size for samples collected at different latitudes; boxplots show median and 1.5 x interquartile range. Error bars are 1 x SE.

Figure 3. Relationship of bacterial community composition, richness and range with site elevation. (a) Relationship between dissimilarity of bacterial OTU data (in multivariate space, using a Bray-Curtis measure) and elevational distance between samples (in m). Colours on the scatter plot indicate the extent to which points on the plot overlap on a gradient from light grey (no overlap) through to yellow (substantial overlap). Also plotted on the figure as a solid black line showing differences in mean Bray Curtis dissimilarity, calculated every 20 m. (b) Relationship between average bacterial taxon richness (number of distinct OTUs obtained from 5,000 DNA sequences per sample) and altitude above sea level. Regression line shows the fit of a quasi-Poisson regression model (c) The mean altitudinal range of bacterial OTUs detected at each sampling location. Line of best fit is $y = 0.37x + 466$, $R^2 = 0.32$, $P < 0.001$. Colours refer to each of the seven regions of New Zealand included in this study (● AK; ● CB; ● HB; ● HR; ● TS; ● WG; ● WK). Error bars are 1 x SE.

Figure 4. Spatial patterns in bacterial community (OTU) attributes for samples collected in catchments dominated by native forest land uses. (a) Relationship between community dissimilarity (Bray-Curtis distance between sample data) and latitudinal distance between samples (in km). Colours on the scatter plot indicate the extent to which points on the plot

overlap on a gradient from light grey (no overlap) through to yellow (substantial overlap). Also plotted on the figure is a solid black trend line showing differences in mean Bray Curtis dissimilarity, calculated every 50 km. (b) Relationship between average bacterial taxon richness (number of distinct OTUs obtained from 5,000 DNA sequences per sample) and latitude (measured as distance north of southernmost sample location (in km). Regression line shows the fit of quasi-Poisson regression model. (c) Relationship between average bacterial taxon richness (number of distinct OTUs) and altitude above sea level; Regression line shows the fit of a quasi-Poisson regression model. (d) Mean latitudinal range of taxa present. Colours indicate samples located within various New Zealand regions (● AK; ● CB; ● HB; ● HR; ● TS; ● WG; ● WK). Line of best fit is $y = -0.33x + 2437$, $R^2 = 0.63$, $P < 0.001$. Error bars are 1 x SE.
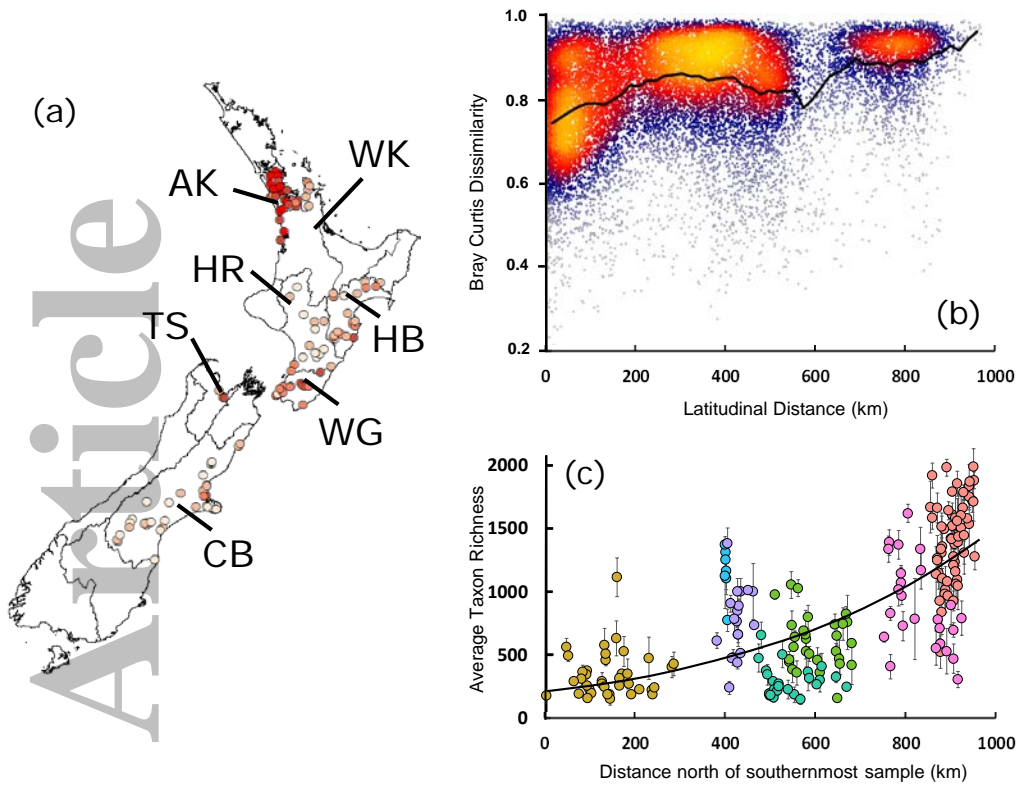
(a)

WK

AK

HR

HB
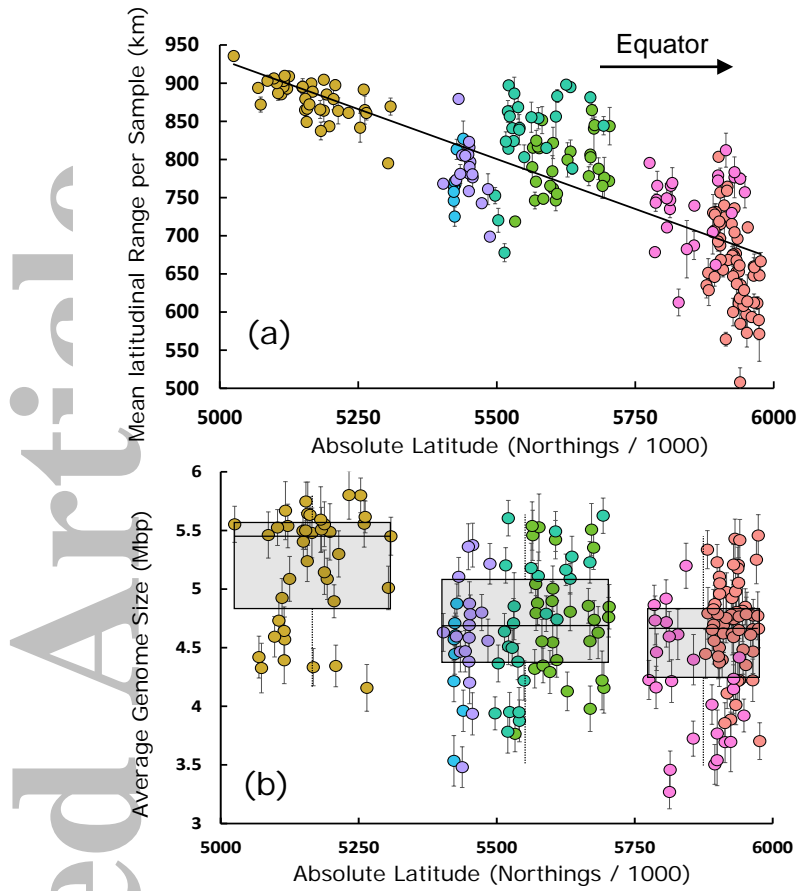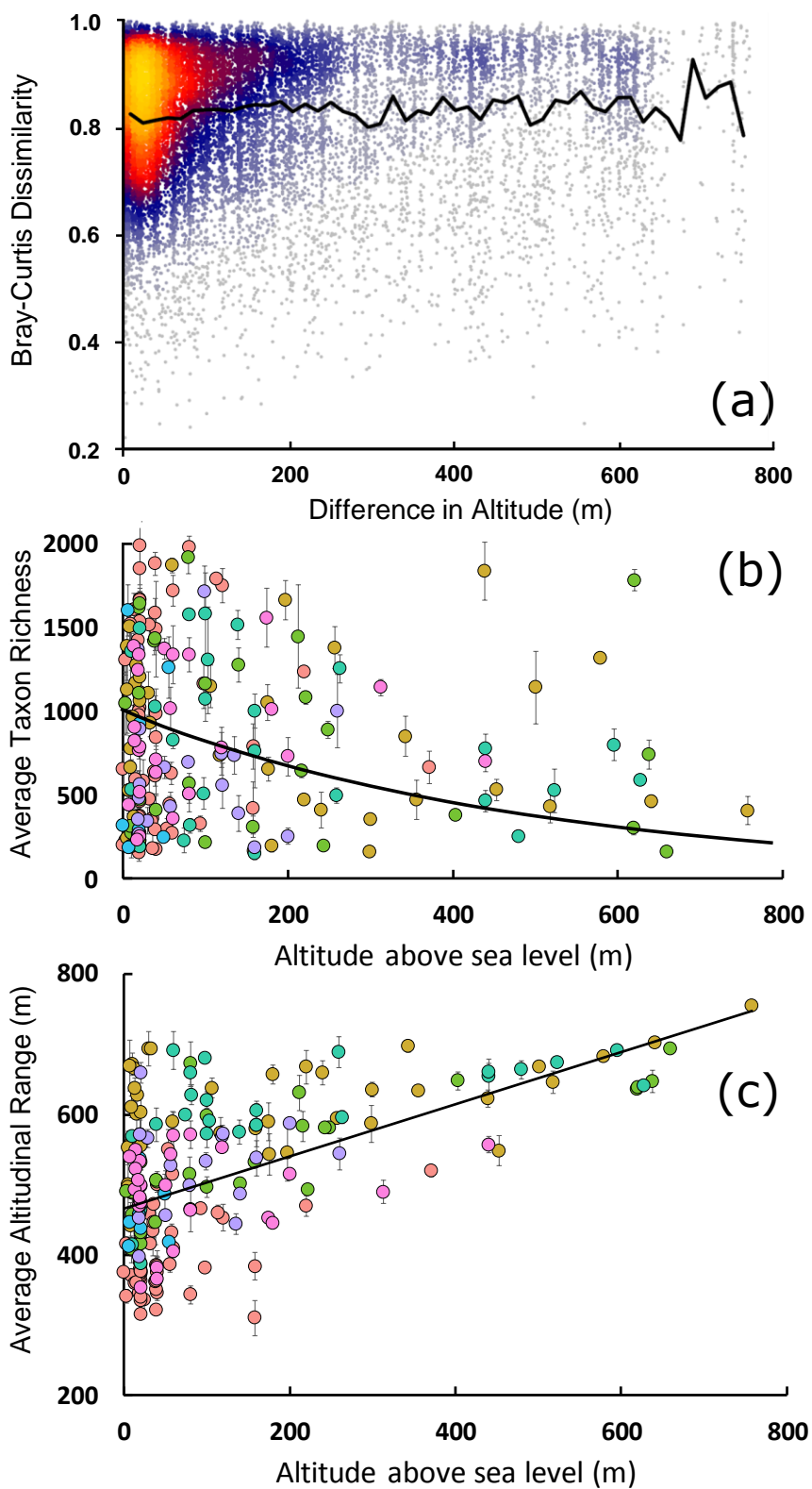
TS

WG

CB

(b)

(c)

Figure 1 – Lear et al

Figure 2 – Lear et al

Figure 3 – Lear et al

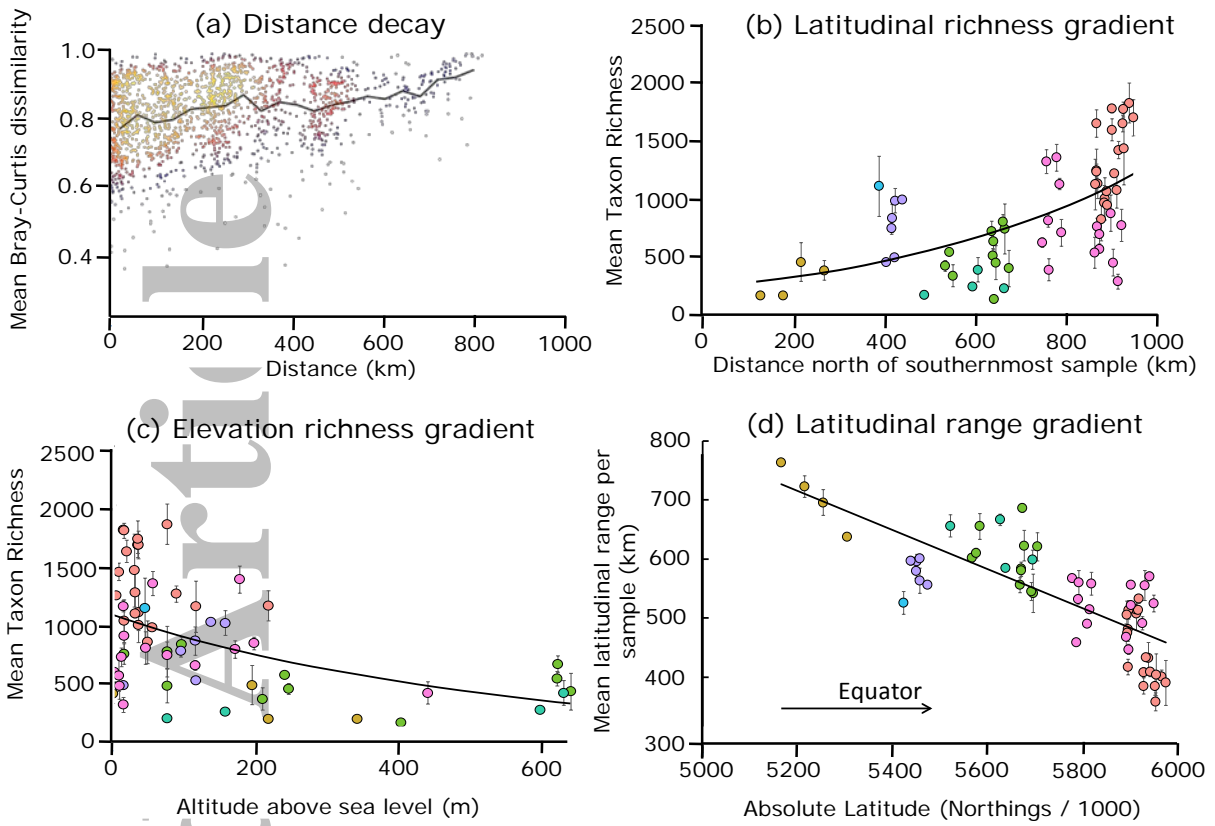Figure 4 – Lear et al