

Barham, M., Kirkland, C.L., and Hollis, J., 2019, Spot the difference: Zircon disparity tracks crustal evolution: *Geology*, <https://doi.org/10.1130/G45840.1>

Supplementary Table DR1. Detrital zircon age data sorted according to geographic region

2019151_Table DR1.xlsx

Supplementary Table DR2 Point-biserial correlation of DZ dissimilarity vs. supercontinent intervals taken from multiple sources. Highlighted timeline C represents the preferred supercontinent periods assessed in this work and shown in figures. Ages of supercontinent episodes derived from the published sources indicated. The average distance to the centroid (average of the distance from each individual point to the average position of all points of a set) and average nearest neighbour (average of the distance from each point to its nearest neighbour in the same set) measures represent common methods of assessing the degree of dispersion of point data in 2d space. In this case the proximity of points within a set refers to a graphical representation of the degree of similarity of detrital zircon age spectra between defined continental regions within any particular age bin. The Pearson correlation coefficient, r , can take a range of values from +1 to -1. A value of 0 indicates that there is no association between the two variables. Negative R-values here indicate a negative correlation such that intervals of supercontinents are correlated with decreased disparity or “difference” in the zircon spectra of continents. P-values indicate the significance of the correlation identified by the R-value. Here, the p-values are very small indicating strong evidence against the null hypothesis (that there is no correlation). Typically p-values ≤ 0.05 are considered significant (95 % confidence), thus our results far exceed this and indicate a >99% confidence in the correlation of global zircon disparity to simplistic supercontinent/no supercontinent states.

Timeline	Pangaea	Gondwana	Rodinia	Columbia/ Nuna	Distance to centroid (CD)		Average nearest neighbor (ANN)		Source
	Ma	Ma	Ma	Ma	R-value	p-value	R-value	p-value	
A	300–250	700–530	1250–1000	1900–1780	-0.266	0.017	-0.272	0.015	[1]
B	300–250	750–530	1300–900	2000–1700	-0.325	0.003	-0.325	0.003	[1]
C	300-200	670-480	1300-900	1975-1450	-0.327	0.003	-0.291	0.009	[2-6]

1. Gardiner, N.J., et al., *Contrasting Granite Metallogeny through the Zircon Record: A Case Study from Myanmar*. Scientific Reports, 2017. **7**(1): p. 748.
2. Schmitt, R.d.S., R.d.A. Frago, and A.S. Collins, *Suturing Gondwana in the Cambrian: The Orogenic Events of the Final Amalgamation*, in *Geology of Southwest Gondwana*, S. Siegesmund, et al., Editors. 2018, Springer International Publishing: Cham. p. 411-432.
3. Pisarevsky, S.A., et al., *Mesoproterozoic paleogeography: Supercontinent and beyond*. Precambrian Research, 2014. **244**: p. 207-225.
4. Meert, J.G. and M. Santosh, *The Columbia supercontinent revisited*. Gondwana Research, 2017. **50**: p. 67-83.
5. Li, Z.X., et al., *Assembly, configuration, and break-up history of Rodinia: A synthesis*. Precambrian Research, 2008. **160**(1–2): p. 179-210.
6. Stampfli, G.M., et al., *The formation of Pangea*. Tectonophysics, 2013. **593**: p. 1-19.

Supplementary Table DR3 Disparity values, data and geographic regions available for each bin interval. Region refers to the defined 11 continental regions defined in Supplementary Table 1. n = the number of age data. Average distance to centroid (average of the distance from each individual point to the average position of all points of a set) and average nearest neighbor distance (average of the distance from each point to its nearest neighbour in the same set) are relative values only and do not have absolute units relevant beyond the multidimensional space dispersion analysis used in this study.

Bin interval (Ma)	Regions with n>30	Average n per region	Data relative to moving av.*	Average distance to centroid (CD)	Average nearest neighbor distance (NN)
0-200	7	1722	<i>H</i>	0.089720329	0.054852
201-400	10	724	<i>L</i>	0.022279942	0.012994
401-600	11	1231	<i>H</i>	0.052892472	0.031339
601-800	11	664	<i>L</i>	0.019646127	0.0076461
801-1000	11	747	<i>L</i>	0.039458813	0.024702
1001-1200	11	1425	<i>H</i>	0.013765103	0.0070748
1201-1400	9	808	<i>L</i>	0.012881978	0.0079176
1401-1600	9	737	<i>L</i>	0.028265015	0.021855
1601-1800	9	1737	<i>H</i>	0.021220682	0.013568
1801-2000	10	1783	<i>H</i>	0.027837731	0.016437
2001-2200	9	782	<i>L</i>	0.053849602	0.025864
2201-2400	9	319	<i>L</i>	0.06346129	0.042061
2401-2600	9	811	-	0.080779517	0.024223
2601-2800	9	1123	<i>H</i>	0.048782938	0.031007
2801-3000	8	429	-	0.013945163	0.010825
3001-3200	8	311	-	0.033166963	0.016423
3201-3400	9	207	-	0.016201431	0.0095767
3401-3600	6	268	-	0.016610162	0.0091579
3601-3800	4	160	-	0.019880269	0.020384
3801-4000	3	58	<i>L</i>	0.063163252	0.073274

**H* = age data for this interval are more abundant than the moving average; *L* = age data are less abundant than the moving average for this temporal interval.

Disparity time-series approach

Data filtering and visualisation

A range of widely available software (Excel, Access, etc.) facilitates the rapid filtering of published, compiled U/Pb age databases of analysed zircon (Voice et al., 2011; Kirkland et al., 2016; Puetz, 2018). Typically, grains must be filtered for the appropriate nature of the zircon analysed (detrital, igneous, metamorphic), as well as the focus of the analysis (timing of crystallization vs. metamorphism), degree of concordance of the U/Pb and Pb/Pb age systems, and ultimately the most appropriate age system to use for a final age to represent the genesis or other geological processes charted by the grain's isotope systematics. These ages are then assessed according to their basinal/geographic/cratonic/tectonic setting, as per requirements. In this work age data were sorted according to their geographic region of sampling and subsequently sub-sampled according to 200 m.y. intervals of crystallization age.

Age data of detrital zircon populations are typically visualised via probability density plots with more common ages of crystallization reflected in corresponding peaks in the age spectrum. Probability density plots stack Gaussian curves for each individual grain analysis (incorporating both age and error) on top of each other to produce a combined population age spectrum. Since younger analyses with smaller errors can overly influence distinctive age peaks in an age spectrum, kernel density estimates produce curves based purely on the density of ages. Cumulative probability plots are a visual representation of the cumulative distributive function, which indicates the proportion of analyses in that population that are younger or older than any given age. Probability and cumulative age plots can be produced through a variety of excel plugins as well as specific scripts for R and java (Sircombe, 2004; Vermeesch, 2009; Gehrels, 2011; Vermeesch, 2012; Vermeesch et al., 2016; Vermeesch, 2018b).

Kolmogorov-Smirnov distance

The Kolmogorov-Smirnov (KS) test assesses the similarity of non-normally distributed age spectra (the K-S test is specifically used to evaluate the validity of the null hypothesis, which holds that there is no significant difference between two populations and that any observed difference relates to insufficient sampling or experimental error). The Kolmogorov-Smirnov distance essentially indicates the degree of similarity of two samples, with larger distance values indicating greater dissimilarity of sample age populations. The KS distance is determined as the maximum vertical separation between cumulative probability curves for any two samples age populations. The KS test can be performed in Excel, SPSS, various packages for R, etc. (Gehrels, 2011; Vermeesch, 2013; Vermeesch et al., 2016; Vermeesch, 2018a). The KS test here was performed on all regionally separated, temporally sub-binned data simultaneously to ensure the relative distances within clusters were maintained during later cluster analyses.

Multi-dimensional scaling

Multidimensional scaling (MDS) offers a means to represent the relative similarity of multiple samples to each other in 2d space. In this work, the KS distance was used to define the similarity of each sample to each other to construct a MDS plot. The distance information is computed as a matrix of pairwise dissimilarities that are rationalised into a mutually consistent plot whereby similar samples plot close to each other and all dissimilar samples plot further apart. Since all data were plotted in the same space, distinct clusters develop whereby age populations for all regions within the same temporal bin plot close to each other, forming 20 clusters (0-200, 201-400.... 3801-4000 Ma). Under closer inspection each of these clusters show differing degrees of dispersion that reflect the similarity of age signatures between regions within that absolute timeframe. Multidimensional scaling can be undertaken in various statistical packages including PAST and packages developed for R (Hammer et al., 2001; Vermeesch, 2013, 2018a). Note that the distances of dissimilarity between samples are not comparable between separate MDS analyses and are only relative to other distances defined within the same MDS space.

Cluster analysis and disparity metrics

Here, the coordinates of the regionally separated, temporally sub-sampled age data in MDS space were extracted from an R-based analysis of the similarity assessment. Extracted coordinates facilitate the next stage of assessment whereby the proximity (similarity/disparity) of temporally- and geographically-binned populations can be absolutely investigated. The nearest neighbour metric defines the average distance between each point and its nearest neighbour, this was calculated using PAST (Hammer et al., 2001). The average distance to a centroid defines the distance from each individual point to a hypothetical average of all the points in the group (centroid), and was calculated simply in Excel using Pythagorean Theorem. Larger distances indicate greater dissimilarity.

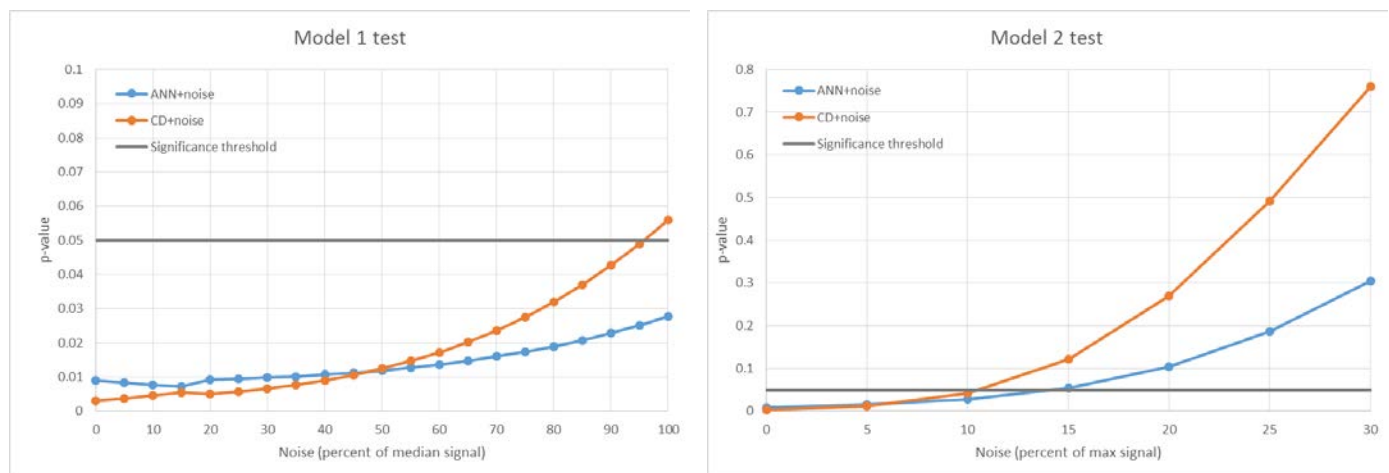
Point Biserial test

The point biserial test assesses the correlation of a continuous variable (in this instance an unbound disparity measure) against a dichotomous variable (one that can only be in two states – in this case either a supercontinent arrangement or dispersed continental arrangement). Point biserial correlation tests can be performed in a number of analytical packages including SPSS. Like standard Pearson correlations the degree and manner of any correlation of the data are expressed in r- and p-values. The correlation coefficient “r-value” measures the strength and sense of a linear relationship between the variables. R-values are always between +1 (indicating that both variables increase in lock-step by the same magnitude) and –1 (indicating that an increase in one variable leads to a perfect decrease in the second variable). P-values indicate the strength/significance of any correlation, with threshold values typically taken at $p \leq 0.05$, indicating >95% likelihood that you can reject the null hypothesis of the correlation being a consequence of chance. In this work while disparity data were binned at 200 m.y. intervals, supercontinent intervals were considered to the nearest 5 Ma of published consensus. As such, for the purposes of the point biserial test, data were resampled at 50 Ma intervals from 25 Ma to 3975 Ma to capture the higher fidelity of the supercontinent cycle.

The sensitivity/robustness of the disparity metric (for testing against the supercontinent cycle) was assessed through the application of two noise models to the data.

- Model 1 was generated as a random positive or negative function of the median value of nearest neighbour or distance to centroid measures across the entire time series.
- Model 2 was generated as a random positive or negative function of the maximum value of nearest neighbour or distance to centroid measures across the entire time series.

Correlation of the disparity time series to the supercontinent cycle was maintained at significance levels of $p < 0.05$ after data modification by (i) up to 95% of the median of CD and NN values obtained, and (ii) up to 10% of the maximum values of disparity obtained for CD and NN metrics (see figures below).



Further Reading

https://docs.google.com/document/d/1MYwm8GcdYFOsfNV62B6PULb_-g2r1AS3vmm4gHMOFvg/preview

- Gehrels, G., 2011, Detrital Zircon U-Pb Geochronology: Current Methods and New Opportunities, Tectonics of Sedimentary Basins, John Wiley & Sons, Ltd, p. 45-62.
- Hammer, Ø., Harper, D. A. T., and Ryan, P. D., 2001, PAST: Paleontological statistics software package for education and data analysis: Palaeontologia Electronica, v. 4, no. 1, p. 1-9.
- Kirkland, C. L., Hollis, J., and Gardiner, N. J., 2016, Greenland U-Pb Geochronology Database. www.greenmin.gl
- Puetz, S. J., 2018, A relational database of global U–Pb ages: Geoscience Frontiers, v. 9, no. 3, p. 877-891.
- Sircombe, K. N., 2004, AgeDisplay: an EXCEL workbook to evaluate and display univariate geochronological data using binned frequency histograms and probability density distributions: Computers & Geosciences, v. 30, no. 1, p. 21-31.
- Vermeesch, P., 2009, RadialPlotter: aJava application for fission track, luminescence and other radial plots: Radiation Measurements, v. 44, no. 4, p. 409-410.
- , 2012, On the visualisation of detrital age distributions: Chemical Geology, v. 312-313, p. 190-194.
- , 2013, Multi-sample comparison of detrital age distributions: Chemical Geology, v. 341, no. Supplement C, p. 140-146.
- , 2018a, Dissimilarity measures in detrital geochronology: Earth-Science Reviews, v. 178, p. 310-321.
- , 2018b, IsoplotR: A free and open toolbox for geochronology: Geoscience Frontiers, v. 9, no. 5, p. 1479-1493.
- Vermeesch, P., Resentini, A., and Garzanti, E., 2016, An R package for statistical provenance analysis: Sedimentary Geology, v. 336, no. Supplement C, p. 14-25.
- Voice, P. J., Kowalewski, M., and Eriksson, K. A., 2011, Quantifying the timing and rate of crustal evolution: global compilation of radiometrically dated detrital zircon grains: The Journal of Geology, v. 119, no. 2, p. 109-126.