

further differentiated, by re-melting, fractional crystallisation and/or mixing to produce the wide range of igneous rock compositions that are exposed at surface (Hawkesworth and Kemp, 2006). The compositional symmetry between the evolved continental crust and the relatively depleted mantle is well documented and helps to define our unique place in the solar system (Allègre, 1997; Taylor and McLennan, 2009).

Despite isotopic evidence that suggests the Archean depleted mantle source (at least by 3.8 Ga) was likely similar to today (Vervoort and Blichert-Toft, 1999), the mechanisms driving the production and composition of the continental crust in the geologic past are less understood (Bédard *et al.*, 2013). Andesitic rocks are central to this debate because they represent the signature rock type at many young arcs and their occurrence in the Precambrian rock record is a key piece of evidence for tracing modern tectonic processes back into the geologic past. Whether these ancient intermediate rock compositions represent primary mantle melts, mixing between differentiated basaltic and rhyolitic melts and/or are crystal fractionation products of basaltic magmas, however, remains controversial (Reubi and Blundy, 2009; Lee and Bachmann, 2014).

Herein the shifting compositional relationships between the three main igneous rock types (*i.e.* basalt, andesite, and rhyolite) map the major geochemical changes that have shaped the continental crust through Earth's history. The focus is on estimating the geochemical centre, or central tendency, for each element and rock type through time (Aitchison, 1986; Egozcue *et al.*, 2003). This approach is thus distinct from, but complementary to, most previous studies on smaller, well-characterised sample sets that tend to emphasise geochemically anomalous sample sub-sets and/or focus on more specific rock sub-types (*e.g.* calc-alkaline versus tholeiitic basalts).

## Compositional Data Analysis

Geochemical data, like other forms of compositional data, sum to constant (*i.e.* closed data) and convey only relative information (Aitchison, 1986). Log-ratios have emerged as one of the preferred approaches for treating such compositional datasets (Egozcue *et al.*, 2003), however, these alternative data treatments can be difficult to interpret in the conventional way, which has limited the application of log-ratio analysis to petrologic problems. Herein log-ratios are based on the order of element compatibility during mantle melting to facilitate their petrologic interpretation. Elements were also grouped to further simplify log-ratio analysis and to summarise temporal trends. This particular form of log-ratio analysis is thus comparable to the usual approach of linking petrogenetic processes to patterns on multi-element diagrams, or so-called spider plots. Slope changes on these classic diagrams, which can be further reduced to ratios (or log-ratios in this case) between elements of contrasting compatibility (incompatible versus compatible elements), reflect igneous differentiation processes such as the degree of partial melting and/or fractional crystallisation (Pearce, 2008). However, in

## Compositional symmetry between Earth's crustal building blocks

C.J.M. Lawley<sup>1\*</sup>



doi: 10.7185/geochemlet.1612

### Abstract

Arc magmatism drives the production of modern continental crust. However, the mode of crustal differentiation in the geologic past, particularly in the Archean, remains controversial. Herein I adopt a compositional approach to interrogate a global, igneous geochemical database (EarthChem Library) and document the evolving compositional history of basalt, andesite, and rhyolite, which represent the three main crustal building blocks. Basalt and andesite yield synced geochemical trends and progressively incompatible element-rich and compatible element-poor compositions that are consistent with extensive partial mantle-melting during the Archean. Post-Archean basaltic to andesitic rocks also tend to be more alkaline in character, which coupled with their high field-strength- and large-ion lithophile-element signature, points to the increased influence of Phanerozoic-style intra-plate magmatism on the global, rock record. Coeval rhyolitic rocks are depleted in these same elements, suggesting that post-Archean felsic magmas track the evolving compositions of their basaltic to andesitic source, which are, in turn, controlled by the partial melting trend. These complementary, or symmetric, geochemical trends between rock types shifted during the Proterozoic and heralded the onset of modern compositional relationships between crustal building blocks.

Received 09 December 2015 | Accepted 22 February 2016 | Published 11 March 2016

### Introduction

Igneous rocks are the main building blocks of the continental crust and represent a long-lived and accessible record of Earth's dynamic geochemical landscape. New continental crust, generated today at convergent and intra-plate settings, is predominately basaltic and thus contrasts with the andesitic composition of the bulk continental crust (Rudnick, 1995; Taylor and McLennan, 1995). This apparent disconnect points to a two-stage, igneous-driven process whereby relatively primitive basaltic magmas are extracted from the mantle and are

1. Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario, K1A 0E8, Canada

\* Corresponding author (email: christopher.lawley@canada.ca)



comparison to interpretations based on raw element concentrations, log-ratios reflect the actual relationships between element groups rather than the spurious effects related to the closed nature of compositional data (Egozcue *et al.*, 2003).

## Data Processing

The full methodological approach is described in the Supplementary Information. Data processing and figures can also be reproduced using the free, open source R software environment (R Core Team, 2015) and code available from the corresponding author. Geochemical data and age data were compiled from the EarthChem library (Fig. S-1). This global geochemical database contains a relatively large number of missing values for some elements. Most previous studies exclude these missing values and consider only the so-called complete cases; however, the consequences of this default option are rarely, if ever, discussed. Log-ratio analysis requires complete datasets and thus, in this case, missing values for elements with minor to moderate degrees of missingness (typically 10–30 % missing) were replaced using column-wise multivariate imputation by chained equations (van Buuren and Groothuis-Oudshoorn, 2011). The impact of the selected imputation method is evaluated in the Supplementary Methods (Figs. S-2 and S-3).

Precambrian samples ( $\geq 541$  Ma;  $n = 66,739$ ) were then discretised into basaltic, andesitic and rhyolitic compositions based on the Total Alkali Silica (Le Bas *et al.*, 1986) and Pearce (Pearce, 1996) classification diagrams (Figs. S-4 and S-5). This classification approach uses mobile- and least-mobile elements to reject metasomatised samples, a particularly important step for Precambrian metamorphosed igneous rocks. The remaining igneous samples ( $n = 18,246$ ) were then split into age groups based on a K-means clustering approach. Weighted bootstrap re-sampling of the data (weighted inversely to age) was used to minimise the effect of unevenly distributed crystallisation ages (Fig. S-6), which are inherent to the punctuated, igneous rock record (Keller and Schoene, 2012). Log-ratios were then calculated from the re-sampled dataset and are presented in Figures 1 and 2.

## Results and Discussion

Isometric log-ratios (ilr-1–8) represent the preferred approach for compositional datasets (Egozcue *et al.*, 2003) and, as defined herein, divide the classic spider diagram into a series of non-overlapping segments (Fig. 1). The first log-ratio (ilr-1) represents the overall proportion of incompatible versus compatible elements (Fig. 1a), whereas the remaining ilrs split the mildly incompatible to compatible (ilr-2–4) and incompatible (ilr-5–8) element portions of the spider diagram into successively smaller segments (Fig. 1a). These smaller segments reflect key petrogenetic indicators such as large ion lithophile element (LILE)- and high

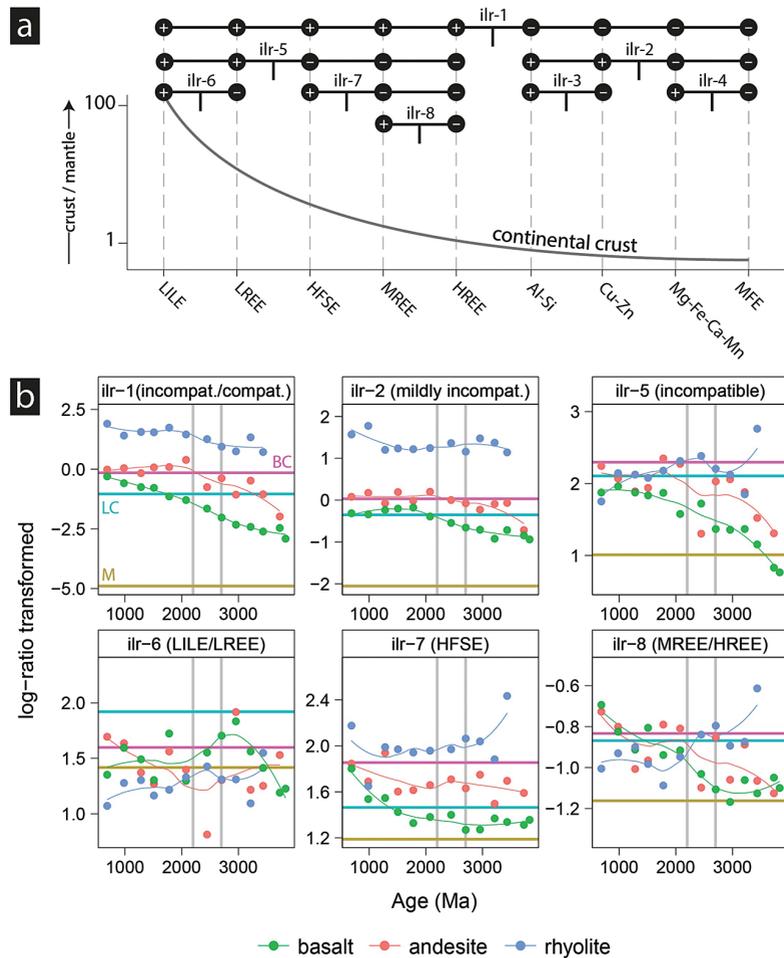
field strength element (HFSE)-enrichment (ilr-6 and ilr-7, respectively), and can also be constructed to reproduce commonly used ratios, such as Sm/Yb, which monitor the influence of key mineral phases (*e.g.*, garnet) during melting (ilr-8; Pearce, 2008).

Basalt and andesite yield negative sloping log-ratios (ilr-1; Fig. 1b), suggesting that these rock types have become increasingly incompatible-element-rich and compatible-element-poor since the Archean. Temporal trends between basaltic and rhyolitic rock types are broadly parallel and yield steadily increasing incompatible versus compatible element log-ratios (ilr-1–8) despite grouping disparate basaltic and andesitic rock sub-types (Fig. 1). On a conventional spider diagram, systematic incompatible element enrichment would correspond to steepening multi-element profiles through time (Fig. 1a). This well-documented pattern is consistent with extensive partial mantle-melting in the geologic past and was possibly driven by hotter, Archean mantle temperatures (Herzberg *et al.*, 2010; Keller and Schoene, 2012).

These element trends can alternatively be presented as shifting mineralogic controls and physical rock properties, such as rock density (Fig. S-9). Modal mineralogy, as inferred from CIPW norms calculated using raw, major-element concentrations (Fig. S-9), also suggest that larger mantle-melt fractions during the Archean may have locally exhausted clinopyroxene and fused more refractory mantle minerals such as orthopyroxene and olivine. These evolving rock properties and estimated modal mineral abundances are thus consistent with decreasing mantle-melt fractions demonstrated by the systematic increase in incompatible versus compatible element log-ratios (Fig. 1), but are based on raw, major-element concentrations.

However, geochemical trends between rock types are only approximately parallel (Figs. 1 and 2). For some element groups (ilr-1–2), basalt and andesite yield diverging, but synced log-ratio patterns during the Archean, suggesting that andesitic rock compositions are more similar to basalt today than they were in the geologic past (Figs. 1 and 2; *n.b.*, except perhaps for the Mesoarchean). Modern andesitic rocks are often interpreted as hybrid melts, produced during mixing of approximately equal proportions of basalt and rhyolite, presumably in upper crustal reservoirs (Rudnick, 1995). These mixing models are based, in part, on the bi-modal SiO<sub>2</sub> distribution of melt inclusions and rock textures that record disequilibrium conditions between crystals and melt (Reubi and Blundy, 2009). Hybridised melt models can reproduce many of the key geochemical features of some Archean andesitic rock suites (Bédard *et al.*, 2013), however, binary mixing between basaltic and rhyolitic rocks are inconsistent with the curved, crystal fraction trends for the multiple time slices presented in Figure 3.

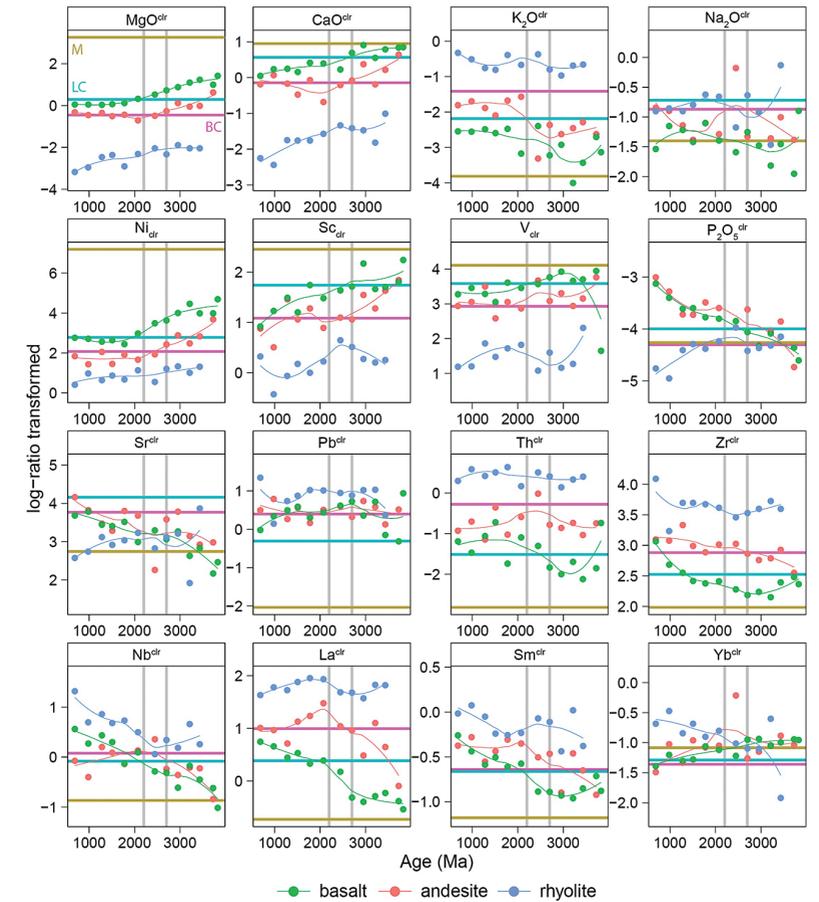




**Figure 1** (a) Schematic spider diagram plotting the composition of the continental crust, normalised to primitive mantle, in relation to isometric-log-ratios (ilir). Numerators and denominators for each ilir-balance are denoted by plus and minus signs, respectively. Each ilir represents a ratio between elements with contrasting compatibility during mantle melting. (b) Geochemical evolution of basalt, andesite, and rhyolite rocks based on ilir. Aggregated log-ratio means and loess-curves are colour coded by rock type. Vertical, grey lines represent the Archean-Palaeoproterozoic boundary (2.7–2.2 Ga; Condie and O’Neill, 2010). Basaltic and andesitic rocks have become incompatible element-rich through time resulting in a negatively sloping ilir-1. Rhyolitic rocks have also become incompatible element-rich (ilir-1), but to a lesser extent and, for some log-ratios (ilir-5–8), yield contrasting trends (positive slopes) that may track the evolving composition of their basaltic to andesitic source. Horizontal lines represent mantle (M; Palme and O’Neill, 2014) and crustal compositions (LW = lower crust; BC = bulk continental crust; Rudnick and Gao, 2003) and are shown for reference. Abbreviations: LILE =

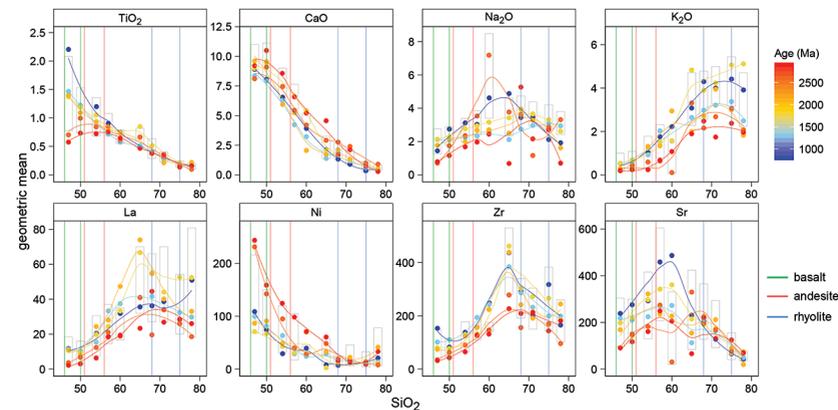


large ion lithophile (K, Na, Ba, Sr, Rb, Pb, Cs), LREE = light rare earth element (La, Ce, Pr, Nd); HFSE = high field strength elements (Ti, P, Zr, Hf, Nb, Ta, U, Th); MREE = middle rare earth element (Sm, Eu, Gd); HREE = heavy rare earth element (Tb, Dy, Ho, Er, Yb, Lu, Y); MFE = mantle forming elements (Cr, Ni, Co, Ga, Sc, V).



**Figure 2** Geochemical evolution of basalt, andesite and rhyolite on an element-by-element basis following a centred log-ratio (clr) approach (Aitchison, 1986). Vertical, grey lines represent the Archean-Palaeoproterozoic boundary (2.7–2.2 Ga; Condie and O’Neill, 2010). Horizontal lines represent mantle (M; Palme and O’Neill, 2014) and crustal compositions (LW = lower crust; BC = bulk continental crust; Rudnick and Gao, 2003) and are shown for reference. This log-ratio transformation preserves a one-to-one relationship with the original variables, but adding or removing variables changes the absolute value of each clr (see Supplementary Information). Isometric log-ratios (Fig. 1) are not restricted by this constraint and represent the preferred approach for compositional data analysis (Egozcue *et al.*, 2003).





**Figure 3** Harker diagrams based on the geometric mean of raw  $\text{SiO}_2$  concentrations. Individual curves are colour coded to age and represent globally averaged, liquid lines of descent for each time slice. Vertical lines represent the inter-quartile range for each rock type. Vertical grey boxes represent the inter-quartile range for each element at any given  $\text{SiO}_2$  bucket. Overall, ancient crystal fractionation for compatible elements are displaced to lower values at any given  $\text{SiO}_2$  bucket, consistent with the partial mantle-melt trend based on log-ratios.

Inflexions on these diagrams likely reflect some crystallising mineral phase akin to classic liquid lines of descent (Fig. 3). However, in this case, crystal fractional curves represent globally averaged melting conditions rather than the evolution of a single, co-genetic magma suite (Keller and Schoene, 2012; Keller *et al.*, 2015). Compatible elements yield concave-down patterns, which for younger time slices are displaced to lower values at any given  $\text{SiO}_2$  concentration (Fig. 3). These displaced curves support the partial melting trends described above (Keller and Schoene, 2012), but critically, are preserved despite the large range of rock compositions at any given  $\text{SiO}_2$  interval.

Archean rocks with intermediate compositions are relatively rare and, where present, occur as thin layers intercalated with tholeiitic basaltic flows (Furnes *et al.*, 2013). These apparent stratigraphic relationships, along with the synched geochemical evolution and curved fractionation trends, tend to suggest that ancient, intermediate rocks represent the fractionated melt products of coeval basaltic magmas. Mixing, or assimilation, with older rocks as is often demonstrated by relatively more sensitive contamination proxies, such as radiogenic isotopes and/or inherited zircon (Allègre and Othman, 1980; Hildreth and Moorbath, 1988), must have also occurred, but did not significantly disrupt the coupled, geochemical evolution of andesite to its parental basalt at any given age interval.

Archean rhyolitic rock compositions, such as the tonalite-trondhjemite-granodiorite (TTG) suite, are also geochemically distinct and, along with lesser basaltic rocks that define narrow, coeval greenstone belts, represent the dominant rock types prior to the Palaeoproterozoic (Furnes *et al.*, 2013). The fractionated

heavy rare earth element (HREE) TTG signature has been the focus of numerous previous studies and likely reflects a garnet ( $\pm$  hornblende) residue during partial melting of hydrous basaltic rocks that were unlike those at modern, mid-ocean ridges (Bédard, 2006; Martin *et al.*, 2014). This unusual trace element behaviour is evident in the current compilation by the anomalous compositions of Meso- to Neoproterozoic rhyolite (ilr-5, -7, -8; Fig. 1). The data underscore the distinct melting conditions responsible for the TTG suite and their obvious impact on bulk, rhyolitic rock chemistry during the Archean.

Post-Archean rocks show a major, composition shift between large-ion lithophile element (LILE)-, high field strength element (HFSE)- and light rare earth element (LREE)-enrichment (Fig. 1; ilr-6–8) in basaltic-andesitic and a corresponding depletion of these elements in similarly-aged rhyolitic magmas. The patterns are clearly defined on an element-by-element basis using log-ratios (Fig. 2), multivariate projections (Figs. S-7 and S-8), and with raw element concentrations (Fig. 3). Post-Archean rocks also tend to be more alkaline in character (Fig. S-5), which coupled with their HFSE- and LILE-rich signature, points to the increasing importance and/or preservation of Phanerozoic-style intra-plate volcanism across the Archean-Proterozoic boundary (2.7–2.2 Ga; Condie and O'Neill, 2010). The extent to which this signal is influenced by the fragmented, Precambrian rock record, however, is difficult to assess and requires further study (Hawkesworth and Kemp, 2006). Nevertheless, this trace element signature (Fig. 1; ilr-6–7) forms the basis for most tectonic discrimination diagrams and is distinct from modern, basaltic arc-magmas, which along with rocks comprising many Archean greenstone belts, are generally depleted in HFSE (Pearce, 2008). The corresponding HFSE-, LILE- and LREE-poor composition of post-Archean rhyolite seems to suggest that these more differentiated rocks track the evolving composition of their basaltic to andesitic source. As a result, the diverging geochemical patterns of rhyolitic versus basaltic-andesite rocks during the Proterozoic may also reflect, at least in part, the decreasing partial melting trend described above and represent a key geochemical feature of the global, igneous rock record. The Archean-Proterozoic boundary thus represents an important period of Earth's history that coincided with these modern compositional relationships between crustal building blocks (Figs. 1 and 2).

In summary, statistical analysis of global, igneous rock chemistry document the shared geochemical history of basalt, andesite and rhyolite that is synched with decreasing, partial-melting through time (Fig. 1). Diverging, element trends emphasise atypical petrogenetic processes that operated during specific periods of Earth's history such as: (1) HFSE-rich (ilr-7; Fig. 1) and HREE-poor Archean rhyolitic magmas (*i.e.* TTG suite; ilr-8; Fig. 1); (2) fractionated Archean andesitic compositions (Figs. 1, 2 and 3); and (3) a major shift towards LREE-, LILE- and HFSE-rich basalt-andesite and corresponding depletion of these elements within coeval rhyolite during the Proterozoic (Figs. 1, 2 and 3). Log-ratios (Egozcue *et al.*, 2003), informed by element behaviour during mantle melting (Fig. 1a), represent a rigorous framework for investigating closed, geochemical data. Future studies



will benefit from the addition of high-quality geochemical and geochronologic datasets to the EarthChem Library, particularly for intermediate rock compositions during poorly-represented, but significant periods of Earth's history such as the Archean-Proterozoic and Neoproterozoic-Cambrian boundaries.

## Acknowledgements

This work was completed under the auspices of the Targeted Geoscience Initiative program and represents ESS contribution (20150469). The hydrogeology group at the GSC is thanked for access to their computing facility. Bruce Kjarsgaard, Eric Grunsky and Vera Pawlowsky-Glahn are thanked for suggestions and discussions that improved an earlier version of the manuscript. John Percival, Peter Cawood and one anonymous reviewer are thanked for thoughtful reviews that improved the clarity of the manuscript.

Editor: Bruce Watson

## Additional Information

**Supplementary Information** accompanies this letter at [www.geochemicalperspectivesletters.org/article1612](http://www.geochemicalperspectivesletters.org/article1612)

**Reprints and permission information** is available online at <http://www.geochemicalperspectivesletters.org/copyright-and-permissions>

**Cite this letter as:** Lawley, C.J.M. (2016) Compositional symmetry between Earth's crustal building blocks. *Geochem. Persp. Let.* 2, 117-126.

## References

- AITCHISON, J. (1986) The statistical analysis of compositional data. Chapman & Hall, Ltd. London, UK, 416 pp.
- ALLÈGRE, C. (1997) Limitation on the mass exchange between the upper and lower mantle: the evolving convection regime of the Earth. *Earth and Planetary Science Letters* 150, 1–6.
- ALLÈGRE, C., OTHMAN, D. (1980) Nd-Sr isotopic relationship in granitoid rocks and continental crust development: A chemical approach to orogenesis. *Nature* 286, 335–342.
- BÉDARD, J.H. (2006) A catalytic delamination-driven model for coupled genesis of Archean crust and sub-continental lithospheric mantle. *Geochimica et Cosmochimica Acta* 70, 1188–1214, doi: 10.1016/j.gca.2005.11.008.
- BÉDARD, J.H., HARRIS, L.B., THURSTON, P.C. (2013) The hunting of the snArc. *Precambrian Research* 229, 20–48.
- CONDIE, K.C., O'NEILL, C. (2010) The archean-proterozoic boundary: 500 my of tectonic transition in earth history. *American Journal of Science* 310, 775–790, doi: 10.2475/09.2010.01.

- EGOZCUE, J., PAWLOWSKY GLAHN, V., MATEU-FIGUERAS, G., BARCELÓ VIDAL, C. (2003) Isometric logratio for compositional data analysis. *Mathematical Geosciences* 35, 279–300, doi: 10.1023/A:1023818214614.
- FURNES, H., DILEK, Y., DE WIT, M. (2013) Precambrian greenstone sequences represent different ophiolite types. *Gondwana Research* 27, 649–685, doi: 10.1016/j.gr.2013.06.004.
- HAWKESWORTH, C., KEMP, A. (2006) Evolution of the continental crust. *Nature* 443, 811–817, doi: 10.1038/nature05191.
- HERZBERG, C., CONDIE, K., KORENAGA, J. (2010) Thermal history of the Earth and its petrological expression. *Earth and Planetary Science Letters* 292, 79–88, doi: 10.1016/j.epsl.2010.01.022.
- HILDRETH, W., MOORBATH, S. (1988) Crustal contributions to arc magmatism in the Andes of Central Chile. *Contributions to Mineralogy and Petrology* 98, 455–489, doi: 10.1007/BF00372365.
- KELLER, C.B., SCHOENE, B. (2012) Statistical geochemistry reveals disruption in secular lithospheric evolution about 2.5 Gyr ago. *Nature* 485, 490–493, doi: 10.1038/nature11024.
- KELLER, C.B., SCHOENE, B., BARBONI, M., SAMPERTON, K.M., HUSSON, J.M. (2015) Volcanic-plutonic parity and the differentiation of the continental crust. *Nature* 523, 301, doi: 10.1038/nature14584.
- LE BAS, M., LE MAITRE, R., STECKEISEN, A., ZANETTIN, B. (1986) The construction of the total alkali-silica chemical classification of volcanic rocks. *Journal of Petroleum Science and Engineering* 27, 745–750.
- LEE, C.-T.A., BACHMANN, O. (2014) How important is the role of crystal fractionation in making intermediate magmas? Insights from Zr and P systematics. *Earth and Planetary Science Letters* 393, 266–274, doi: 10.1016/j.epsl.2014.02.044.
- MARTIN, H., MOYEN, J.-F., GUITREAU, M., BLICHERT-TOFT, J., LE PENNEC, J.-L. (2014) Why Archean TTG cannot be generated by MORB melting in subduction zones. *Lithos* 198–199, 1–13, doi: 10.1016/j.lithos.2014.02.017.
- PALME, H., O'NEILL, H.S.C. (2014) Cosmochemical Estimates of Mantle Composition. In: Holland, H., Turekian, K. (Eds.) *Treatise on Geochemistry*. Elsevier Ltd., 1–38, doi: 10.1016/B978-0-08-095975-7.00201-1.
- PEARCE, J. (1996) A users guide to basalt discrimination diagrams. In: Pearce, J., Wyman, D. (Eds.) *Geological Association of Canada, Short Course Notes*. Geological Association of Canada, 79–113.
- PEARCE, J.A. (2008) Geochemical fingerprinting of oceanic basalts with applications to ophiolite classification and the search for Archean oceanic crust. *Lithos* 100, 14–48, doi: 10.1016/j.lithos.2007.06.016.
- R CORE TEAM (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- REUBI, O., BLUNDY, J. (2009) A dearth of intermediate melts at subduction zone volcanoes and the petrogenesis of arc andesites. *Nature* 461, 1269–1273, doi: 10.1038/nature08510.
- RUDNICK, R. (1995) Making continental crust. *Nature* 378, 571–577.
- RUDNICK, R.L., GAO, S. (2003) Composition of the continental crust. In: Holland, H., Turekian, K. (Eds.) *Treatise on Geochemistry*, Second Edition. Elsevier, 1–64.
- TAYLOR, S., MCLENNAN, S. (2009) Planetary Crusts. Cambridge University Press, Cambridge, 404 pp.
- TAYLOR, S., MCLENNAN, S. (1995) The geochemical evolution of the continental crust. *Reviews of Geophysics and Space Physics* 33, 241–265.
- VAN BUUREN, S., GROOTHUIS-ODUSHOORN, K. (2011) mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 45, 1–67.
- VERVOORT, J.D., BLICHERT-TOFT, J. (1999) Evolution of the depleted mantle: Hf isotope evidence from juvenile rocks through time. *Geochimica et Cosmochimica Acta* 63, 533–556, doi: 10.1016/S0016-7037(98)00274-9.



## Compositional symmetry between Earth's crustal building blocks

C.J.M. Lawley<sup>2\*</sup>

### Supplementary Information

The Supplementary Information includes:

- Supplementary Methods
- Supplementary Information References

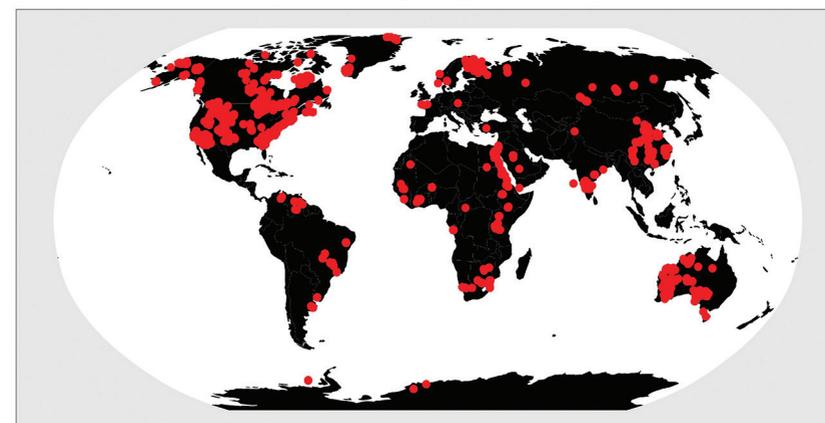
### Supplementary Methods

**Data compilation and missing values.** All data processing and interpretation was completed in R (R Core Team, 2015). Code and processed datasets are available from the corresponding author. Igneous rock geochemistry and ages were downloaded and compiled from the EarthChem Library ([www.earthchem.org](http://www.earthchem.org)). Selected Precambrian ( $\geq 541$  Ma) samples ( $n = 66,739$ ) contained a relatively large proportion of missing values, particularly for some elements. Log-ratio transformations, the preferred approach for compositional data (discussed below) and, many multivariate statistical methods, require complete datasets (Palarea-Albaladejo and Martín-Fernández, 2015). In this case, missing values are likely due to a combination of element concentrations that were below the analytical detection limit, elements that were not collected as part of that particular analytical suite and/or potentially a variety of other unknown reasons. Unfortunately, differentiating between these different types of missing values was impractical, if not impossible, for such a large and diverse dataset and thus all missing values were treated similarly.

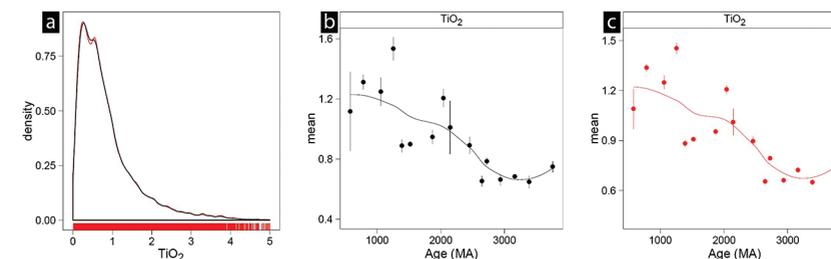
Samples (rows) and elements (columns) with greater than 50 % missing values were removed prior to imputing the remaining missing values using column-wise multivariate imputation by chained equations (van Buuren and Groothuis-Oudshoorn, 2011). Missing values comprised between 10 and 30 % for most elements. Other imputation approaches are available (Palarea-Albaladejo

and Martín-Fernández, 2015) and the selected method, particular for elements with a relatively large proportion of missing values, is pertinent to the secular geochemical trends that are the focus of the current study. Ideally, imputed concentrations are realistic and reflect a value that could have been observed had it not been missing. The plausibility of imputed values was checked qualitatively, and on a univariate basis, by comparing the distribution of five imputation iterations with the kernel density distribution of complete-cases (Figs. S-2 and S-3). For most elements, the imputed distributions compare well with the actually observed distribution (Fig. S-2a), suggesting that imputed values are geologically reasonable. Whether these missing values shifted the centre of the data, which is the focus herein, was also tested by comparing boxplot distributions before and after imputation for all elements (Fig. 3).

EarthChem Library Sample Distribution



**Figure S-1** Geographic distribution of Precambrian EarthChem Library samples included within the present study.

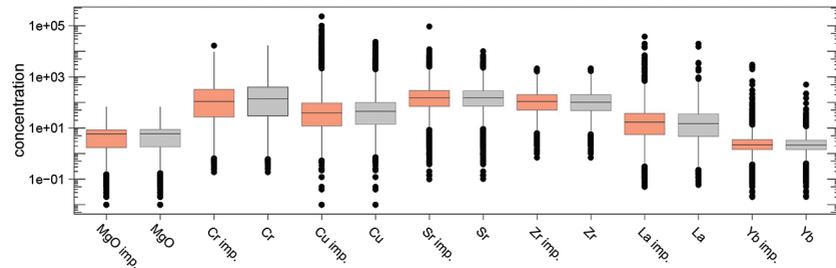


**Figure S-2** (a) Complete cases (black) versus imputed (red) kernel density distributions for  $\text{TiO}_2$ . (b-c) Overall, the imputed values are realistic and do not affect the secular trends that are the focus herein. Error-bars are reported as  $2 \times$  standard error of the mean.

2. Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario, K1A 0E8, Canada

\* Corresponding author (email: [christopher.lawley@canada.ca](mailto:christopher.lawley@canada.ca))

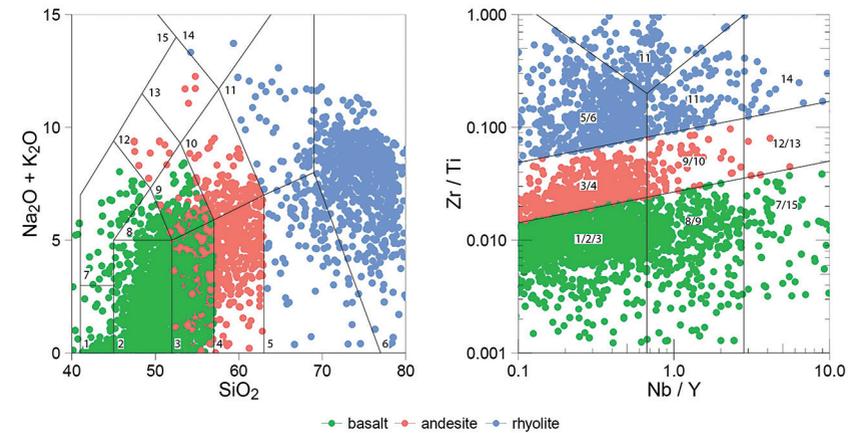




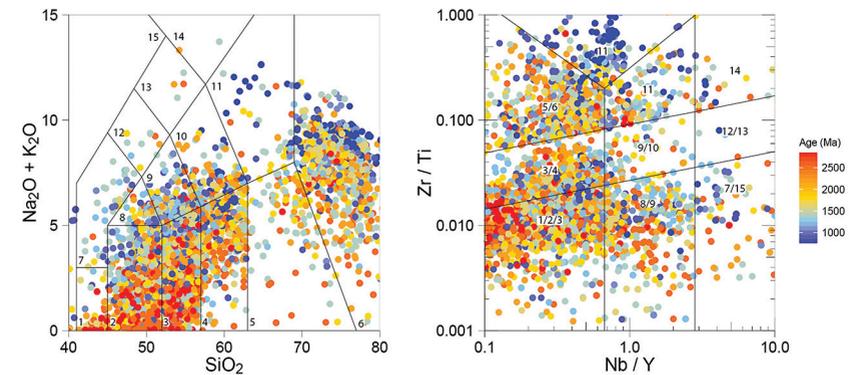
**Figure S-3** Boxplot distributions for complete-cases (black) versus the imputed dataset (dots represent outliers). Imputed values do not significantly affect the central tendency for elements that possess between 10 and 50 % missing values (left to right).

The imputed dataset was then discretised into basaltic, andesitic and rhyolitic compositions based on the classic TAS (Le Bas *et al.*, 1986) and Pearce (Pearce, 1996) classification diagrams. Samples that were not correctly classified in both diagrams were excluded from the following statistical analysis. This combined classification approach, which is based on mobile- and least-mobile elements, reduces the impact of metasomatised samples and identifies unrealistic rock compositions (Figs. S-4 and S-5). Metamorphosed and hydrothermally altered rocks represent a significant proportion of the dataset ( $n = 8,839$ ) and suggest that identifying least-metasomatised samples ( $n = 18,246$ ) is particularly important for the Precambrian igneous rock record. Normative mineralogy was calculated for this least-altered sample-set using major element concentrations on a volatile-free basis (Guzman, 2015).

**Weighted bootstrap re-sampling.** The rock record is unevenly distributed in space and time due to the combined effects of igneous processes and preservation bias (Cawood *et al.*, 2013). A significant proportion of the EarthChem database also contains poorly-defined ages that contribute to the sample clustering at specific ages (*e.g.*, 3.1 Ga; Fig. S-6a). Herein the effect of this uneven distribution on secular geochemical trends was evaluated with weighted bootstrap re-sampling (Keller and Schoene, 2012). Samples were weighted inversely proportional to age using kernel-based density estimations (Duong, 2007). Subsets of these weighted samples ( $n = 100$ ) were then discretised into age groups using K-means clustering (Bivand, 2015). This process was repeated 10,000 times prior to log-transforming the data and calculating means and standard errors for each variable by rock- and age-group. In general, the re-sampled data resulted in more uniform, posterior age distributions due to the reduced influence of peak crystallisation ages (Fig. S-6). Although qualitative, the re-sampled data appear to produce a close-fit with the geochemical trends defined by the original data, suggesting that the uneven distribution of crystallisation ages or the selected age-groupings do not impact the overall conclusions discussed herein (Fig. S-6). Nevertheless, re-sampled data were used for all further statistical processing to limit the influence of those samples with poorly-defined ages.

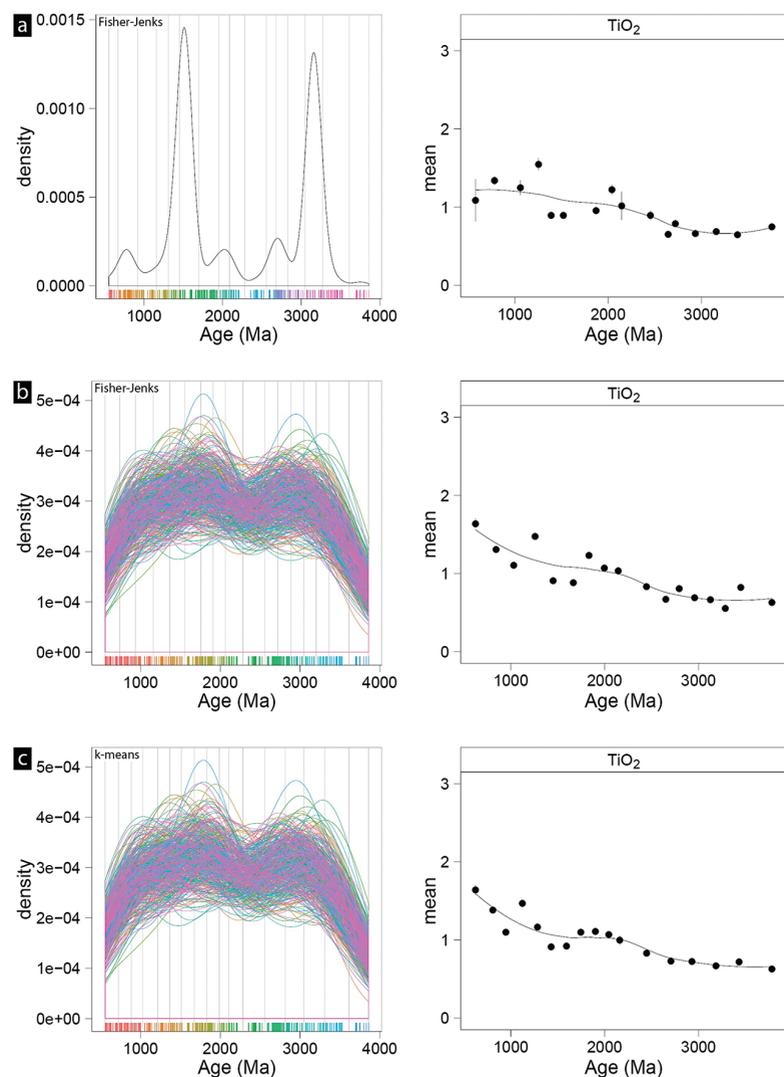


**Figure S-4** TAS (left) and Pearce (right) rock classification diagrams colour coded to rock type. Only the first 100 (out of 10,000) re-sampled datasets are shown for clarity. Rock names: 1 = picro-basalt; 2 = basalt; 3 = basaltic andesite; 4 = andesite; 5 = dacite; 6 = rhyolite; 7 = basanite; 8 = trachy-basalt; 9 = basaltic trachy-andesite; 10 = trachyte; 12 = phonotephrite; 13 = tephriphronolite; 15 = foidite.



**Figure S-5** TAS (left) and Pearce (right) rock classification diagrams colour coded to age. Only the first 100 (out of 10,000) re-sampled datasets are shown for clarity. The tendency for modern rock types to be alkali element-rich is clear even within this simple data projection. Rock names: 1 = picro-basalt; 2 = basalt; 3 = basaltic andesite; 4 = andesite; 5 = dacite; 6 = rhyolite; 7 = basanite; 8 = trachy-basalt; 9 = basaltic trachy-andesite; 10 = trachyte; 12 = phonotephrite; 13 = tephriphronolite; 15 = foidite.





**Figure S-6** (a) Original data distribution and secular TiO<sub>2</sub> trend fitted with loess. Data clustering is due to the uneven distribution of the igneous rock record and false age-peaks related to poorly-defined ages within the EarthChem library. The impact of this punctuated igneous rock record was minimised by weighted bootstrap re-sampling (weighted inversely to age density; see text for discussion). Vertical lines are age buckets defined by the Fisher-Jenks method. Rug plots are colour coded to age bucket. (b–c) Kernel density distributions for the first 1,000 re-sampled iterations (out of 10,000; density curves are colour coded to iteration number) and secular TiO<sub>2</sub> trends fitted with loess. Vertical lines are age buckets defined by the

Fisher-Jenks and K-means methods, respectively. Rug plots colour coded to age bucket. Overall, re-sampled data for even this relatively small-number of iterations closely approximates the trend defined by the original dataset.

**Log-ratio transformations.** Element concentrations, like other compositional data, sum to a constant and have unique statistical properties (Aitchison, 1986). Log-ratio approaches (*i.e.* logarithms of ratios of compositional parts) provide a framework for handling compositional data and work by transforming a D-part composition, known as the simplex ( $x = [x_1, x_2, \dots, x_D]$ ), into multi-dimensional real-space (Aitchison, 1986). Transformed data are then analysed in the conventional way.

Herein two log-ratio transformations were employed. The centred log-ratio (clr) approach transforms a D-part composition into a D-real-vector (Aitchison, 1986). Each clr can be considered as a compositional coordinate and is calculated by dividing each compositional vector  $x$  by the geometric mean  $g(x)$  and then taking the logarithms:

$$\text{clr}(x) = \log \left( \frac{x_1}{g(x)}, \frac{x_2}{g(x)}, \dots, \frac{x_D}{g(x)} \right) \quad \text{Eq. S-1}$$

Where  $g(x) = \left( \prod_j^D x_j \right)^{1/D}$

Calculated clr coordinates possess a one-to-one relationship with the original variables (Fig. 2), but are sub-compositionally incoherent (*i.e.* adding or removing variables changes the calculated clrs). Isometric log-ratio (ilr) transformations provide an alternate representation of compositional data that generates D-1 coordinates and an orthogonal system of axes in real space that facilitate further statistical analysis (Egozcue *et al.*, 2003). Orthonormal ilr-coordinates can be defined by a sequential binary partition (SBP) using the concept of balances (Egozcue *et al.*, 2003; Van Der Boogaart and Tolosana-Delgado, 2006; Pawlowsky-Glahn and Egozcue, 2011). First, a compositional vector is divided into two, non-overlapping groups of parts. Each group is further divided until all groups contain a single part (Fig. 1). This process results in a series of partitions that correspond to D-1 ilr-coordinates, which for the  $j^{\text{th}}$  balance is defined as:

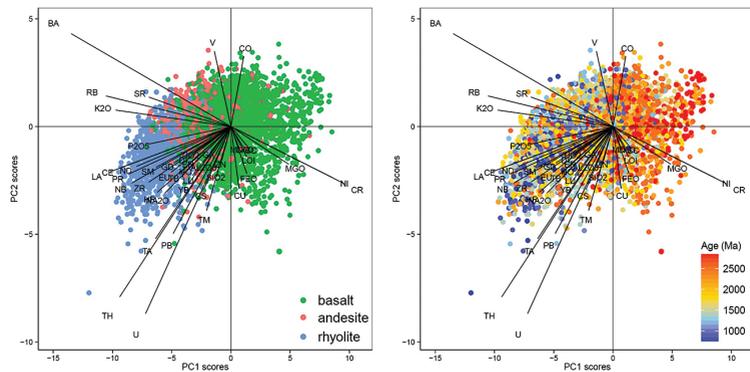
$$\text{ilr}(j) = \sqrt{\frac{n_j^+ n_j^-}{n_j^+ + n_j^-}} \ln \frac{g(c_j^+)}{g(c_j^-)} \quad \text{Eq. S-2}$$

Where  $n_j^+$  and  $n_j^-$  are the number of parts in the numerator and denominator, respectively, in the  $j^{\text{th}}$  row of the sequential binary partition;  $g(c_j^+)$  and  $g(c_j^-)$  are the geometric mean of the components in the numerator and denominator, respectively.

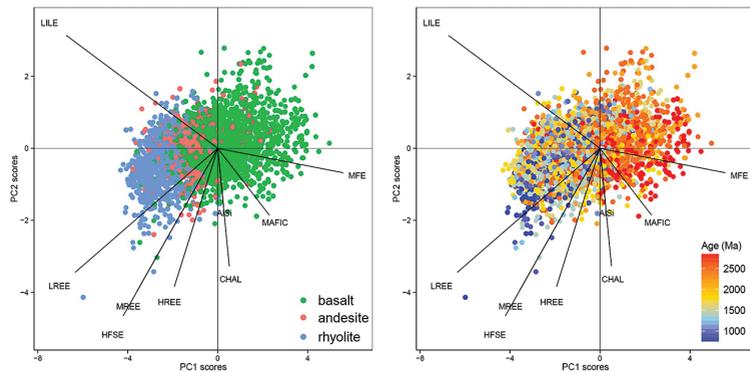
Balances were simplified by splitting elements into groups prior to log-ratio transformations. Element groupings correspond to known geochemical behaviour and were guided, in this specific case, by clusters of elements following



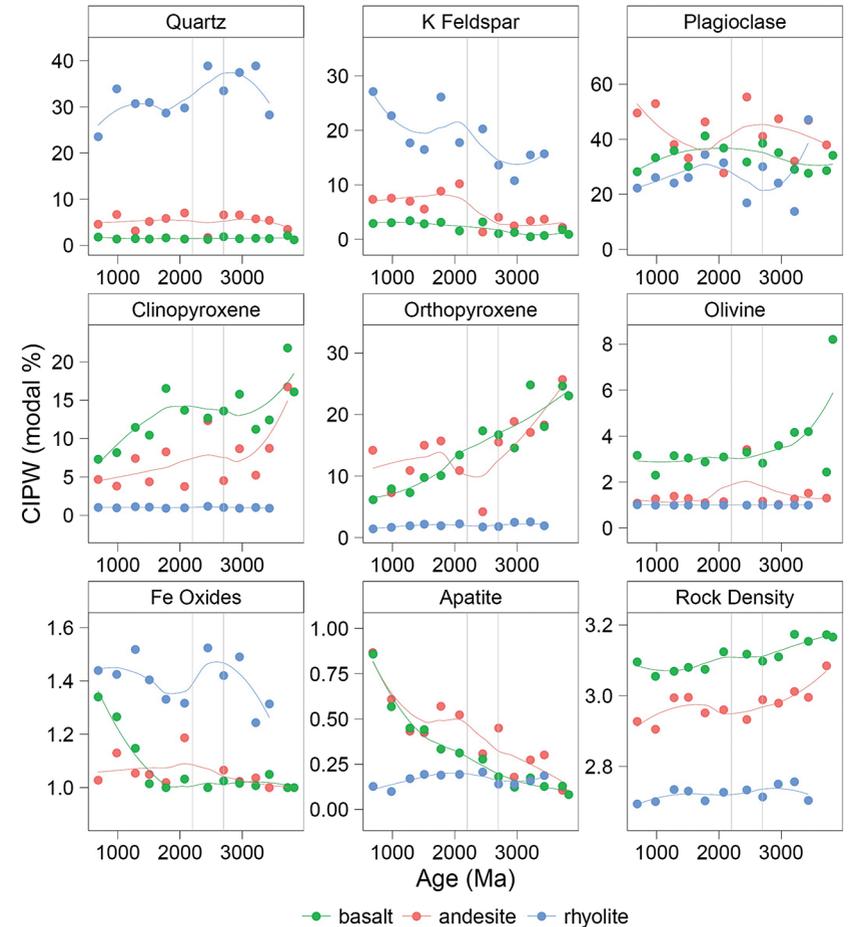
robust Principal Component Analysis (PCA; Figs. S-7 and S-8; Templ *et al.*, 2011). Each element group is defined as: (1) large ion lithophile elements (LILE) = K, Na, Ba, Sr, Rb, Pb, Cs; (2) light rare earth elements (LREE) = La, Ce, Pr, Nd; (3) high field strength elements (HFSE) = Ti, P, Zr, Hf, Nb, Ta, U, Th; (3) middle rare earth element (MREE) = Sm, Eu, Gd; (4) heavy rare earth elements (HREE) = Tb, Dy, Ho, Er, Yb, Lu, Y; (5) Al-Si; (6) chalcophile elements = Cu, Zn; (7) Mg, Fe, Ca, Mn; and (8) mantle forming elements (MFE) = Cr, Ni, Co, Ga, Sc, V.



**Figure S-7** Robust principal component analysis (Templ *et al.*, 2011) colour coded to rock type (left) and age (right) using the original variables. Only the first 1,000 (out of 10,000) re-sampled datasets are shown for clarity.



**Figure S-8** Robust principal component analysis (Templ *et al.*, 2011) colour coded to rock type (left) and age (right) using element groups. The similarities between this projection and principal component analysis using the original elements (Fig. S-7) suggests that reducing the number of variables does not impact the results. Only the first 1,000 (out of 10,000) re-sampled datasets are shown for clarity.



**Figure S-9** Normative mineralogy (modal %) based on CIPW calculations using raw major element concentrations in R (Guzman, 2015). Vertical, grey lines represent the Archean-Palaeoproterozoic boundary (2.7–2.2 Ga; Condie and O'Neill, 2010). Extensive partial melting during the Archean exhausted clinopyroxene and fused more refractory minerals such as orthopyroxene and olivine.

Element groups were then ordered according to their compatibility during mantle melting (Fig. 1). The order is equivalent to classic, multi-element diagrams, or spider plots (Hofmann, 1988; Pearce, 2008). As a result, each ilr can be interpreted in the conventional petrological sense. Simply put, a positive ilr suggests that the geometric mean of the numerator is greater than the geometric mean of the denominator. For ilrs that represent ratios of incompatible and compatible



elements through time (Fig. 1b), an increasing *ilr* suggests that these rocks have become more incompatible element rich through time. The geologic significance of each *ilr* is displayed graphically in Figure 1a.

### Supplementary Information References

- AITCHISON, J. (1986) The statistical analysis of compositional data. Chapman & Hall, Ltd. London, UK, 416 pp.
- BIVAND, R. (2015) classInt: Choose Univariate Class Intervals. <https://cran.r-project.org/web/packages/classInt/index.html>.
- CAWOOD, P., HAWKESWORTH, C., DHUIME, B. (2013) The continental record and the generation of continental crust. *Geological Society of America Bulletin* 125, 14–32.
- CONDIE, K.C., O'NEILL, C. (2010) The archean-proterozoic boundary: 500 my of tectonic transition in earth history. *American Journal of Science* 310, 775–790, doi: 10.2475/09.2010.01.
- DUONG, T. (2007) ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software* 21, 1–16.
- EGOZCUE, J., PAWLOWSKY GLAHN, V., MATEU-FIGUERAS, G., BARCELÓ VIDAL, C. (2003) Isometric logratio for compositional data analysis. *Mathematical Geosciences* 35, 279–300, doi: 10.1023/A:1023818214614.
- GUZMAN, R. (2015) NORRRM: Geochemical Toolkit for R. <https://cran.r-project.org/web/packages/NORRRM/index.html>.
- HOFMANN, A. (1988) Chemical differentiation of the Earth: The relationship between mantle, continental crust, and oceanic crust. *Earth and Planetary Science Letters* 90, 297–314, doi: 10.1016/0012-821X(88)90132-X.
- KELLER, C.B., SCHOENE, B. (2012) Statistical geochemistry reveals disruption in secular lithospheric evolution about 2.5 Gyr ago. *Nature* 485, 490–493, doi: 10.1038/nature11024.
- LE BAS, M., LE MAITRE, R., STECKEISEN, A., ZANETTIN, B. (1986) The construction of the total alkali-silica chemical classification of volcanic rocks. *Journal of Petroleum Science and Engineering* 27, 745–750.
- PALAREA-ALBALADEJO, J., MARTÍN-FERNÁNDEZ, J.A. (2015) zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and Intelligent Laboratory Systems* 143, 85–96, doi: 10.1016/j.chemolab.2015.02.019.
- PAWLOWSKY-GLAHN, V., EGOZCUE, J. (2011) Exploring Compositional Data with the CoDa-Dendrogram. *Austrian Journal of Statistics* 40, 103–113.
- PEARCE, J. (1996) A users guide to basalt discrimination diagrams. In: Pearce, J., Wyman, D. (Eds.) *Geological Association of Canada, Short Course Notes*. Geological Association of Canada, 79–113.
- PEARCE, J.A. (2008) Geochemical fingerprinting of oceanic basalts with applications to ophiolite classification and the search for Archean oceanic crust. *Lithos* 100, 14–48, doi: 10.1016/j.lithos.2007.06.016.
- R CORE TEAM (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- TEMPL, M., HRON, K., FILZMOSER, P. (2011) robCompositions: an R-package for robust statistical analysis of compositional data. In: Pawlowsky-Glahn, V., Buccianti, A. (Eds.) *Compositional Data Analysis. Theory and Applications*. John Wiley & Sons, Chichester (UK), 341–355.
- VAN BUUREN, S., GROOTHUIS-OUDSHOORN, K. (2011) mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 45, 1–67.
- VAN DER BOOGAART, K.G., TOLOSANA-DELGADO, R. (2006) Compositional data analysis with “R” and the package “compositions.” Geological Society London, *Special Publication* 264, 119–127, doi: 10.1144/GSL.SP.2006.264.01.09.

