



## DATA ARTICLE

# A global dataset of sandstone detrital composition by Gazzi-Dickinson method

Xiaolong Dong<sup>1</sup> | Xiumian Hu<sup>1</sup> | Wen Lai<sup>2</sup> | Weiwei Xue<sup>1</sup> | Shijie Zhang<sup>3</sup> |  
Yiqiu Zhang<sup>1</sup> | Wei An<sup>4</sup> | Haiming Fan<sup>5</sup> | Sijin Chen<sup>5</sup> | Cui Li<sup>5</sup> | Xingyun Wang<sup>5</sup> |  
Yue Wu<sup>5</sup> | Jinlv Chen<sup>6</sup> | Yajun Zhang<sup>6</sup> | Kun Yu<sup>6</sup>

<sup>1</sup>State Key Laboratory of Mineral Deposits Research, School of Earth Sciences and Engineering, Nanjing University, Nanjing, China

<sup>2</sup>Frontiers Science Center for Critical Earth Material Cycling, School of Geography and Ocean Science, Nanjing University, Nanjing, China

<sup>3</sup>School of Tourism, Henan Normal University, Xinxiang, China

<sup>4</sup>China School of Resources and Environmental Engineering, Hefei University of Technology, Hefei, China

<sup>5</sup>Suzhou Deep-time Digital Earth Research Centre, Suzhou, China

<sup>6</sup>Image Sky International Co., Ltd., Suzhou, China

## Correspondence

Xiumian Hu, State Key Laboratory of Mineral Deposits Research, School of Earth Sciences and Engineering, Nanjing University, 210023, Nanjing, China.

Email: [huxm@nju.edu.cn](mailto:huxm@nju.edu.cn)

## Funding information

National Natural Science Foundation of China Project, Grant/Award Number: 42050102; the Second Tibetan Plateau Scientific Expedition and Research Program, Grant/Award Number: 2019QZKK0204

## Abstract

Detrital composition of sandstone is the most important data for siliciclastic studies including sandstone classification, provenance analysis, oil and gas exploration. A large amount of detrital composition data has accumulated over the past decades, however, they are scattered in publications without unified standards. Here we constructed a global dataset of detrital components of sandstones from 646 peer-reviewed publications using Gazzi-Dickinson method. A total of 19,861 samples from Precambrian to Quaternary are involved in this dataset. For each sample, we present details on reference information, geographic information, geological background, depositional age and the original data. It is a high-quality dataset for the information on each sandstone sample from different studies which was standardized. The dataset can be used widely, such as for stratigraphic

**Dataset Identifier:** <https://doi.org/10.1002/gdj3.212>

**Creator:** Xiaolong Dong, Xiumian Hu, Wen Lai, Weiwei Xue, Shijie Zhang, Yiqiu Zhang, Wei An, Haiming Fan, Sijin Chen, Cui Li, Xingyun Wang, Yue Wu, Jinlv Chen, Yajun Zhang, Kun Yu

**Dataset correspondence:** [huxm@nju.edu.cn](mailto:huxm@nju.edu.cn)

**Title:** A global dataset of sandstone detrital composition by Gazzi-Dickinson method

**Publisher:** Geoscience Data Journal

**Publication year:** 2023

**Resource type:** Excel files

**Version:** 1

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Geoscience Data Journal* published by Royal Meteorological Society and John Wiley & Sons Ltd.

comparison, provenance analysis, exploring the general laws of the source-to-sink process and geological engineering.

#### KEYWORDS

detrital composition, Gazzi-Dickinson method, provenance analysis, sandstone, spatiotemporal information

## 1 | INTRODUCTION

The use of framework petrography is a common step in provenance research of sand and sandstone (Augustsson, 2021; Basu, 1976; Chayes, 1956; Garzanti, 2019; Ingersoll et al., 1984; Suttner, 1976). The detrital composition of each sandstone sample was gotten by counting the grains one by one under the microscope, which would take at least 2–3 h for a skilled geologist. However, these data are scattered throughout the publications which contain information on detrital composition all over the world. A large amount of sedimentology-related data has been accumulated, making it possible to collect data on detrital components into a dataset (e.g. Lai et al., 2022). For a long time, the source-to-sink system in sedimentology has attracted much attention (Allen, 2017; Walsh et al., 2016) and the detrital composition is indispensable for provenance analysis, stratigraphic comparison, basin analysis, etc. Therefore, it is urgent to build a global dataset of detrital composition, which would make the scientific research on sedimentary general laws based on big data become possible.

Here, we present and describe a dataset of detrital composition from 646 published scientific papers or theses, and a sandstone detrital dataset of Qinghai-Tibet Plateau (Lai et al., 2022). The geographic extent of the dataset is  $-85$  to  $76^\circ$  and  $-177$  to  $178^\circ$ , including the intracontinental, continental block margins, subduction

zones, orogenic belts, suture zones, etc. Samples are mainly from subaerial exposure, which comprise  $\sim 82\%$  of the total samples, while those from modern marine are rare, and are mainly from the drilled holes including Integrated Ocean Drilling Program (IODP), Deep Sea Drilling Project (DSDP) and Ocean Drilling Program (ODP) (Figure 1).

## 2 | DATA DESCRIPTION AND DEVELOPMENT

Data were compiled from published scientific papers and theses, which were all counted using the “Gazzi-Dickinson” as method in the paper. In the Gazzi-Dickinson method, a suitable matrix node chosen based on the sand grain size, the matrix spacing is typically larger than the majority of the sand grain to ensure that individual particles are not double-counted, then each grain on the node is identified and counted. One of the most important features of the Gazzi-Dickinson method is that for all grains larger than  $62.5\mu\text{m}$  in the rock are counted on an individual mineral, independently of whether they are single grains or bound to other particles. References of the dataset cover a period from 1967 to 2022. In order to realize the effective integration of different types of detritus component data sheets more effectively, a unified standard and information entry format are necessary. The dataset

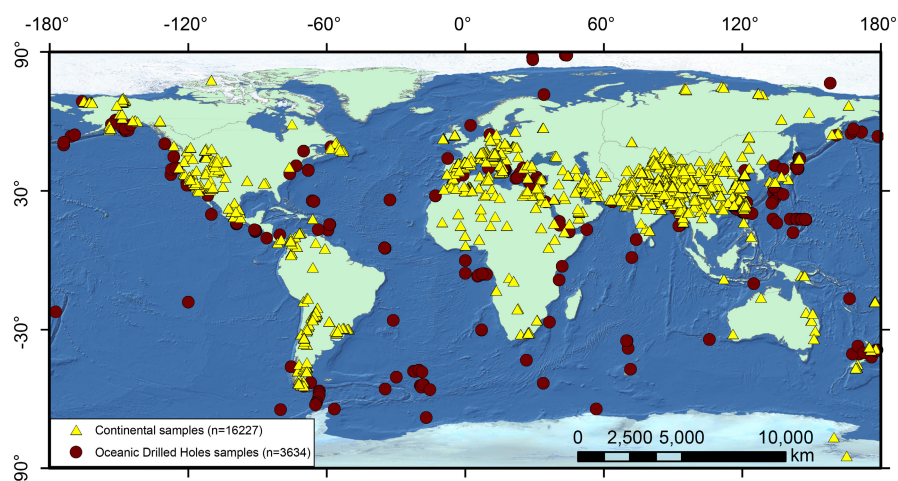


FIGURE 1 Localization of sandstone samples.

is designed as a simple Excel table, including an original metadata sheet containing the original information of the publications, a metadata sheet (100%) which are normalized into 100% for using convenience, a reference sheet and an annotated sheet.

## 2.1 | Standardized header of metadata sheet

This dataset includes five major parts: the reference information, the geographic information, the geological background, the depositional age and the original data. (1) The reference information part contains seven fields including Lead Author, Year, Journal, Volume, Pages, Title and Web Link. (2) The geographic information part contains 6 fields including Country, State/Provenance, Region, Geological Locality, Latitude and Longitude (3) The geological background contains 9 fields including Geological Background, Geological Continent, Geological Unit, Group, Formation, Member, Sedimentary profile, Basin types and Depositional Environment. (4) The depositional age part includes Depositional Age (Era, Period, Epoch, Stage Age), Max-Depositional Age (Ma), Min-Depositional Age (Ma) and Age method. (5) The original data part includes Sample ID, Sample name and the sample data. The sample name is the sample number in the original article while the sample ID is the code that we give to all the samples we collected from the theses. The data in the scientific papers and theses have two kinds—the counting number and proportion. The type and content of detrital fragments as the main part includes 29 columns such as Qm, Qp, Pl, Kf, Lv, Lu, Lm, Lsd, Lsc, Lch, L, HM, etc. The description and relationship among the detrital fragment codes are shown in [Tables 1 and 2](#) and the repository sheet.

## 2.2 | Dataset construction process

The methods are described in four steps depicted in [Figure 2](#) and outlined in detail below:

1. Georeferencing: geosciences papers and theses were searched using “Gazzi-Dickinson” as the keyword, and then check the papers one by one to ensure whether they have relevant point-counting original data.
2. Metadata entry: the literature information is entered into a designed standardized table by professional information entry assistants. The standardization of detrital fragment types and content follow [Lai et al. \(2022\)](#). Identify detritus component data from the selected literature appendix or text table, calibrate, merge or split

different types of detritus components, and enter it into the standard metadata sheet.

3. Geological background entry and metadata checking: other information such as geographical location and geological setting in the body of the text were entered by sedimentology geologists, meanwhile, checking the metadata at the same time.
4. Data cleaning: deduplicate and calibrate data on sedimentary environment, GPS, geological information, etc., and replace other synonymous terms with a uniform professional vocabulary. Check whether the data matches the original paper. If there are other peer-reviewed articles on the same subject that have relevant descriptions of vacant items, find them and enter them into the dataset or if no data are available, do not enter any information.

## 2.3 | Supplementation and calibration of sedimentary environment and spatiotemporal information

This dataset was constructed while complementing and calibrating GPS, epoch and depositional environment as much as possible in order to provide complete temporal and environmental attributes for each rock sample.

GPS information is a digital form of accurate spatial location and a necessary element for future mapping. Although some of the GPS data will be provided in the main text or in the attached table, there are still >30% of rock samples lacking accurate GPS. We read the corresponding GPS value from the geological maps or the lithology columns in the paper or checked on Google Earth. If the GPS values were still not available from the above, we populated the GPS of the region or locality described in the text, though not precise enough.

The sedimentary environment in the geological information section can generally be found in the main text, but ~26% of the samples lack sedimentary environment information. These rock samples without depositional environment information are pending publication of subsequent research and future updates to the dataset.

## 3 | DATA DESCRIPTION

### 3.1 | Dataset location and format

The dataset is available at DDE (<https://repository.deep-time.org/detail/1596796819413078017>). It consists of a metadata Excel table. Each row of the main data table records all the information of each sandstone.

TABLE 1 All the field type of the database.

Part 1, Reference information	Lead_author	Part 4 metadata	Sample ID
	Year		Sample Name
	Journal		Counted Number
	Volume		Proportion (%)
	Pages		Q
	Title		Qm
	WebLink		Qp
Part 2 geographic location	Geological Background	Part 4 metadata	F
	Country		Pl
	State/Provenance		Kf
	Region		L
	Geological Locality		Ls
	Geological Continent		Lp
	Geological Unit		Lcc
	Group		Lcd
	Formation		Lch
	Member		&Ls
	Sedimentary Profile		Lm
	Latitude		Lms
	Longitude		Slate
	Basin types		Phyllite
Depositional Environment	Schist		
Part 3 deposition age	Era	Lmc	
	Period	Lmv	
	Epoch	Lu	
	Stage Age	Lv	
	Depositional Age	Lvf	
	Average Depositional Age	Lvm	
	Minimal Depositional Age	Lvu	
	Age-Method	&Lv	
		Mica	
		HM	
	Others		

### 3.2 | The characteristics of the dataset

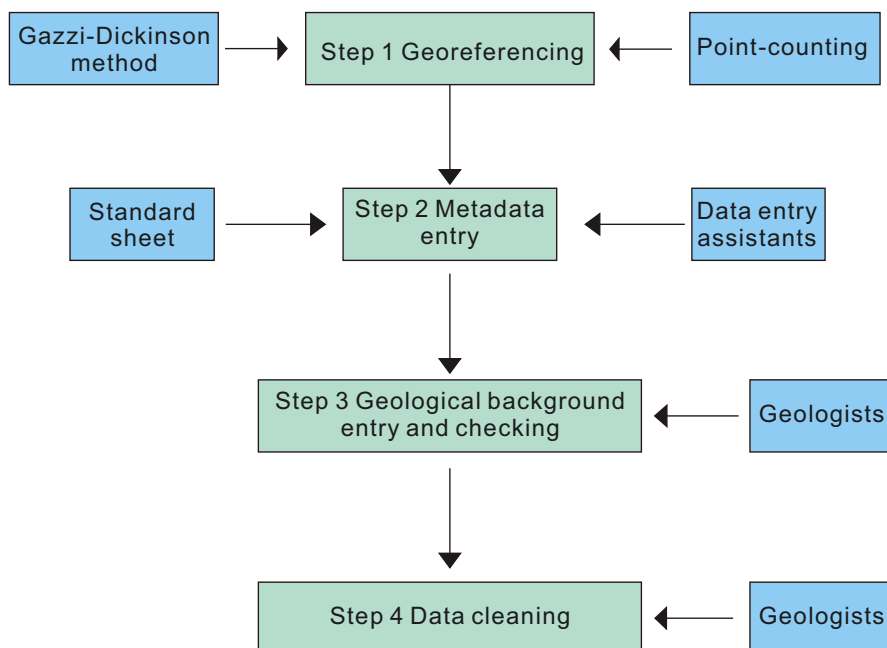
This dataset, collected from 646 peer-reviewed published articles, contains detrital compositional data of 19,861 sandstone samples with depositional ages ranging from Precambrian to Quaternary (Figure 3). A great variations exist in sample volume over different time periods. The Cenozoic and Mesozoic sandstone samples are predominant, containing ~43% and ~37% samples of the dataset, respectively, followed by Palaeozoic sandstones samples (~16%), whereas the Proterozoic samples are the minimum, which is only ~4%. The Cretaceous has the highest number of samples ( $n = 4,119$ ), while the

Paleogene has the second largest ( $n = 3,566$ ), followed by Quaternary ( $n = 2,549$ ) and Neogene ( $n = 2,417$ ). The Jurassic and Triassic samples are 1,696 and 1,499, respectively. The Phanerozoic and Precambrian sandstone samples are all less than 1,000 while the Silurian samples are the minimum, consisting only 114 samples (Figure 3).

These sandstones are mainly formed in 21 types of sedimentary environments, including alluvial, fluvial, glacial, aeolian, delta, submarine, deep marine, etc. Three major sedimentary environments are continental, transition and marine. The continental samples are predominant which account for ~51% of the samples which

**TABLE 2** Fragment codes for framework composition of sandstones (modified after Garzanti et al., 2021; Ingersoll et al., 1984; Lai et al., 2022).

Fragment code	Description	Fragment code	Description
Qm	Monocrystalline Quartz	&Lv	Other volcanic lithic fragments
Qp	Polycrystalline Quartz	Lv	Volcanic lithic fragments (=Lv <sub>f</sub> + Lv <sub>i</sub> + Lv <sub>b</sub> + &Lv)
Q	Total quartz (=Qm + Qp)	Lms	Metamorphic sedimentary lithic fragments
Pl	Plagioclase	Lmv	Metamorphic volcanic lithic fragments
Kf	K-Feldspar	Slate	Slate
F	Feldspars (=Pl + Kf + &F)	Phyllite	Phyllite
Lp	Detrital lithic fragments	Schist	Schist
Lcc	Carbonate lithic fragments	Lmc	Marble
Lcd	Dolomite lithic fragments	Lu	Metamorphic ultrabasic lithic fragments
Lch	Chert	&Lm	Other metamorphic lithic fragments
&Ls	Other sedimentary lithic fragments	Lm	Metamorphic lithic fragments (=Lms + Lmv + Slate + Phyllite + Schist + Lmc + Lu + &Lm)
Ls	Sedimentary lithic fragments (=Lsp + Lsc + Lsd + Lch + &Ls)	L	Lithic fragments (=Ls + Lv + Lm)
Lvf	Felsic volcanic lithic fragments	Mica	muscovite, biotite and other mica
Lvm	Intermediate volcanic lithic fragments	HM	Heavy minerals in total
Lvu	Basic volcanic lithic fragments	Others	Pores, matrix, cements, etc.



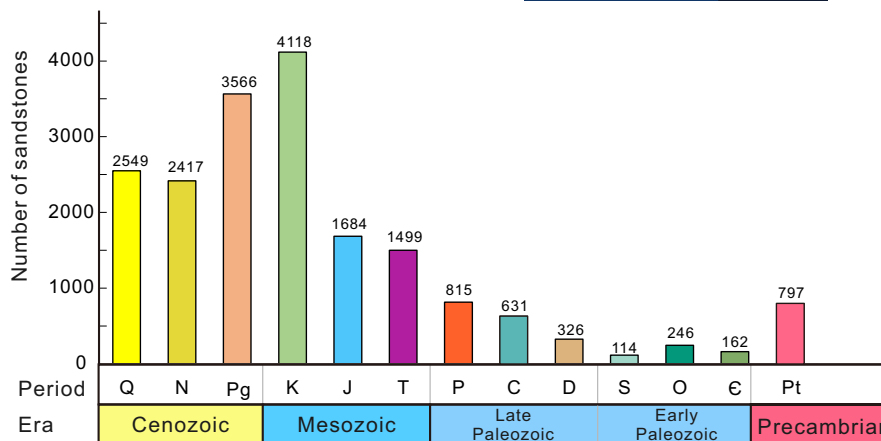
**FIGURE 2** Flowchart of the steps involved in georeferencing, data entry and cleaning to produce a database of detrital composition of sandstones.

contain the sedimentary environment information. The samples in the transition and marine environment consist of ~17% and ~32% of the dataset, respectively. The fluvial samples ( $n = 3,192$ ) are the largest in all kinds of sedimentary environments, followed by shallow marine ( $n = 2,240$ ). Other kind of depositional environments with more than 1,000 samples are alluvial and coast environments. Samples from aeolian, glacial, lacustrine,

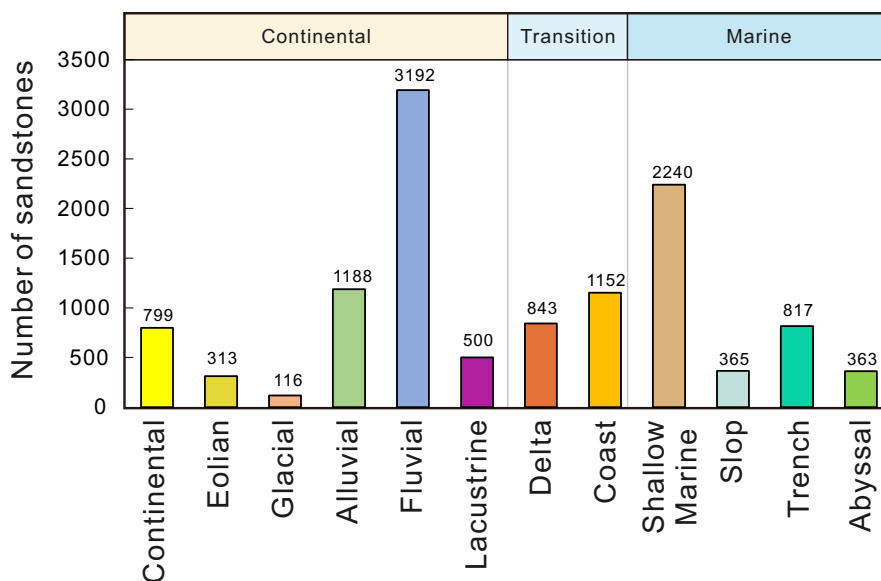
delta, slop, trench and abyssal environment vary from 100 to 800 (Figure 4).

All kinds of sandstones were found in our dataset. The detrital modes of sandstone samples from different periods are shown in Figure 5. The quartzose sandstones are overwhelming, followed by lithic sandstones, while the feldspathic sandstones are the minimum in all periods. As for the lithics, the volcanic lithic fragments are

**FIGURE 3** Relative number of sandstones of different periods. Q, Quaternary; N, Neogene; Pg, Paleogene; K, Cretaceous; J, Jurassic; T, Triassic; P, Permian; C, Carboniferous; D, Devonian; S, Silurian; O, Ordovician; € , Cambrian; Pt, Precambrian.



**FIGURE 4** Types of sedimentary environments and their sandstone quantities included in the dataset.



more common than the metamorphic and sedimentary lithic fragments. The Quaternary samples are dominated by quartzose and lithic sandstones, and the sedimentary lithics are overwhelming. The Neogene, Paleogene and Cretaceous samples are characterized by similar detrital modes, which are dominated by quartzose and feldspathic sandstones with very few lithic sandstones, and the three kinds of lithics are nearly equivalent. The detrital modes of Jurassic to Precambrian samples are similar with quartzose sandstones overwhelming and few feldspathic and lithic sandstones.

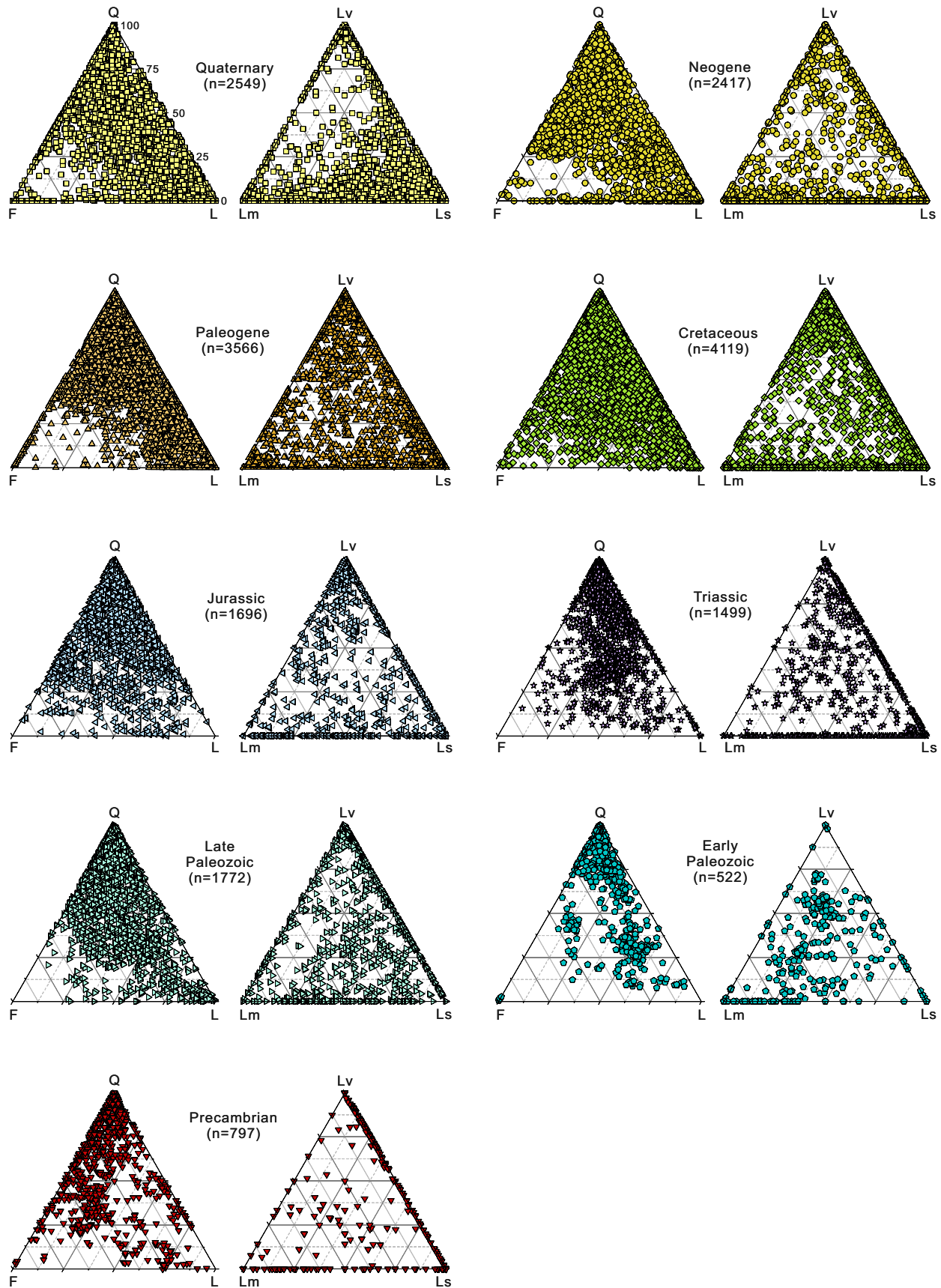
Moreover, the detrital modes of sandstone samples from different sedimentary environments are shown in Figure 6. The alluvial and fluvial environments are dominated by quartzose and lithic sandstones, and the sedimentary lithic fragments are more common in the fluvial system. Quartzose sandstones are overwhelming in the aeolian environment. The detrital modes of shallow marine and deep marine are quite similar, but the volcanic lithic fragments are dominated in the deep marine environment.

The detrital modes vary from different periods and different environments deserve more exploration in the future.

### 3.3 | Limitations and update of dataset

Two limitations should be noted: (1) The dataset is limited by the use of only the Gazzi-Dickinson method and some articles using other methods may be omitted, which may not fully reflect the diversity of detrital composition of the global sandstones; (2) Some important information such as GPS data, depositional age, depositional environment is not fully available or sufficiently accurate in the original article.

Due to these limitations, this dataset will be regularly updated every 1–2 years to supplement newly published or newly discovered literature data under the unified management of DDE. For the stratigraphic units in the existing dataset, if more accurate results such as depositional environment are published in the future, the corresponding part of the information will also be updated or corrected.



**FIGURE 5** The QFL and LmLvLs ternary diagrams for sandstones in different periods. The classification of sandstone from Garzanti (2019). Q, quartzose; F, feldspathic; L, lithic; Lv, volcanic lithic fragments; Lm, metamorphic lithic fragments; Ls, sedimentary lithic fragments.

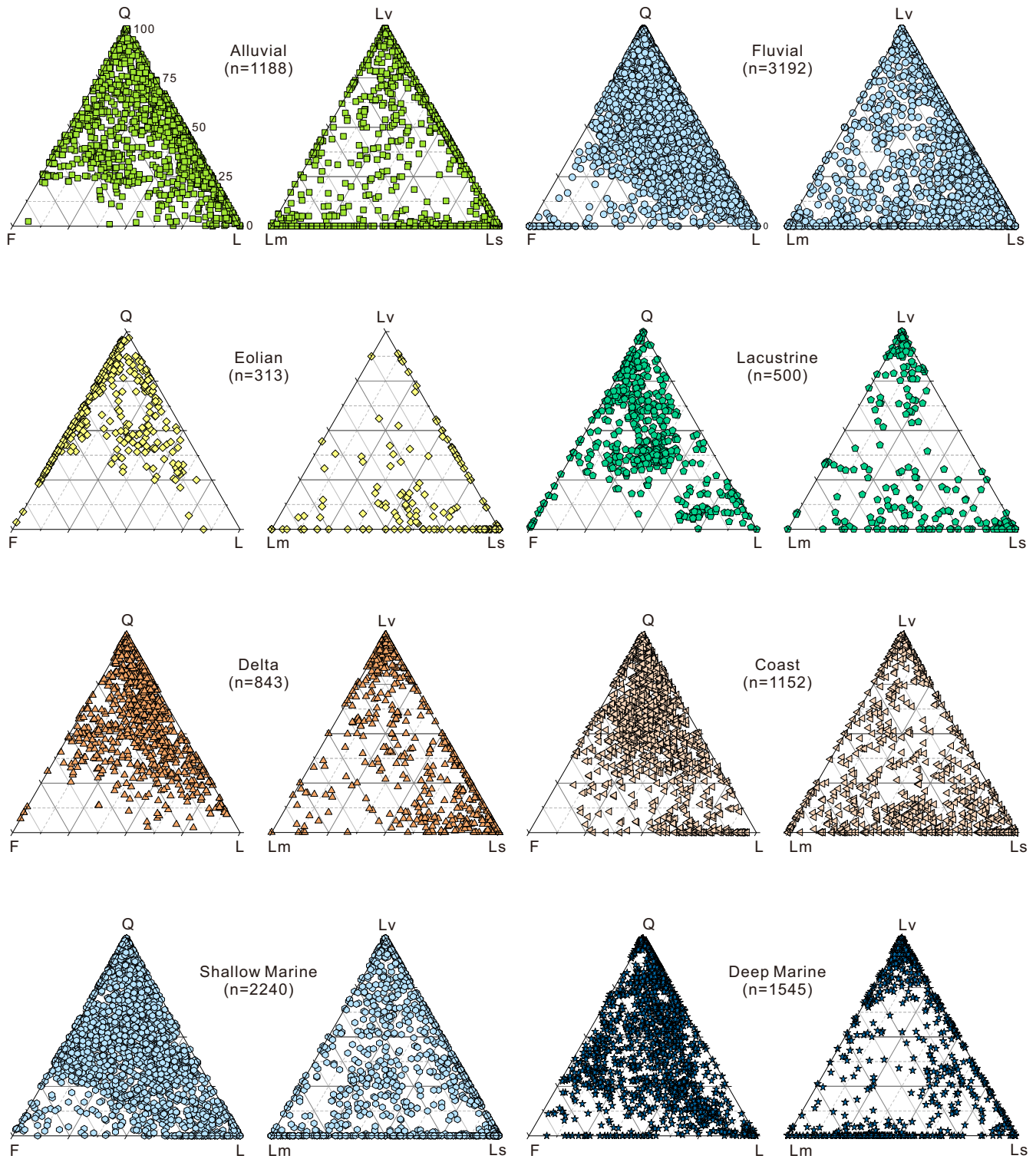


FIGURE 6 The QFL and LmLvLs ternary diagrams for sandstones in different sedimentary environment.

#### 4 | POTENTIAL DATASET USE AND REUSE

The composition of detrital sediments reflects the influence of provenance, tectonics and climate, especially the source area is decisive for the composition of detrital

sediments (Leary et al., 2016; Wang et al., 2015). The most important use of this dataset is for provenance analysis, but many other uses could be explored, such as geological unit comparison, depositional environment and tectonic background inferring, etc. The information such as stratigraphic units or epochs can be easily linked to other



geological studies. The geographic locations or GPS included can be effectively linked to tectonic studies or social applications.

Diverse depositional environments, provenance features and tectonic backgrounds are shown in this dataset. Their output tectonic backgrounds cover intracontinental, continental block margins, subduction zones, orogenic belts, suture zones, etc. The vast majority of background settings. These representative data have important reference and in-depth research value for important hot geoscience issues such as the general laws of source-sink processes and the relationship between sedimentary and tectonic settings. In addition, these sedimentary debris data still have many potential regional geoscience laws and the value of disciplines to be explored.

## AUTHOR CONTRIBUTIONS

**Weiwei Xue:** Data curation (equal); formal analysis (equal). **Yiqiu Zhang:** Data curation (equal). **Wei An:** Data curation (equal). **Sijin Chen:** Data curation (equal). **Cui Li:** Data curation (equal). **Xingyun Wang:** Data curation (equal). **Yue Wu:** Data curation (equal). **Jinlv Chen:** Data curation (equal). **Yajun Zhang:** Data curation (equal). **Kun Yu:** Data curation (equal).

## ACKNOWLEDGEMENTS

The research was funded by the National Natural Science Foundation of China Project (42050102) and the Second Tibetan Plateau Scientific Expedition and Research Program (2019QZKK0204). This is a contribution to the DDE Big Science Program.

## OPEN RESEARCH BADGES



This article has been awarded Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. Data is available at <https://repository.deep-time.org/detail/1596796819413078017>

## ORCID

Xiumian Hu  <https://orcid.org/0000-0002-5401-8682>  
Shijie Zhang  <https://orcid.org/0000-0002-6973-493X>

## REFERENCES

Allen, P.A. (2017) *Sediment routing systems: the fate of sediment from source to sink*. Cambridge: Cambridge University Press.

- Augustsson, C. (2021) Influencing factors on petrography interpretations in provenance research—A case-study review. *Geosciences*, 11(5), 205. Available from: <https://doi.org/10.3390/geosciences11050205>
- Basu, A. (1976) Petrology of Holocene fluvial sand derived from plutonic source rocks: implications to paleoclimatic interpretation[J]. *Journal of Sedimentary Petrology*, 46(3), 694–709.
- Chayes, F. (1956) *Petrographic modal analysis[M]*. New York: John Wiley and Sons, 113.
- Garzanti, E. (2019) Petrographic classification of sand and sandstone. *Earth-Science Reviews*, 192, 545–563. Available from: <https://doi.org/10.1016/j.earscirev.2018.12.014>
- Garzanti, E., He, J., Barbarano, M., Resentini, A., Li, C., Yang, L. et al. (2021) Provenance versus weathering control on sediment composition in tropical monsoonal climate (South China) – 2. Sand petrology and heavy minerals. *Chemical Geology*, 564, 119997. Available from: <https://doi.org/10.1016/j.chemgeo.2020.119997>
- Ingersoll, R.V., Bullard, T.F., Ford, R.L., Grimm, J.P., Pickle, J.D. & Sares, S.W. (1984) The effect of grain size on detrital modes: a test of the Gazzi-Dickinson point-counting method. *Journal of Sedimentary Petrology*, 54, 103–116. Available from: <https://doi.org/10.1306/212f83b9-2b24-11d7-8648000102c1865d>
- Lai, W., Hu, X., Dong, X. & Ma, A. (2022) A dataset of sandstone detrital composition from Qinghai-Tibet plateau. *Geoscience Data Journal*, 1–8. Available from: <https://doi.org/10.1002/gdj3.184>
- Leary, R.J., DeCelles, P.G., Quade, J., Gehrels, G.E. & Waanders, G. (2016) The Liuqu conglomerate, southern Tibet: Early Miocene basin development related to deformation within the great counter thrust system. *Lithosphere*, 8, 427–450. Available from: <https://doi.org/10.1130/l542.1>
- Suttner, L.J. (1976) Sedimentary petrographic provinces: an evaluation. In: Ross, C.A. (Ed.) *Paleogeographic provinces and provinciality*, Vol. 21. United States: SEPM Special Publication, pp. 75–84.
- Walsh, J.P., Wiberg, P.L., Aalto, R., Nittrouer, C.A. & Kuehl, S.A. (2016) Source-to-sink research: economy of the Earth's surface and its strata. *Earth Science Reviews*, 153, 1–6. Available from: <https://doi.org/10.1016/j.earscirev.2015.11.010>
- Wang, J.G., Hu, X.M., BouDagher-Fadel, M., Wu, F.Y. & Sun, G.Y. (2015) Early Eocene sedimentary recycling in the Kailas area, southwestern Tibet: Implications for the initial India–Asia collision. *Sedimentary Geology*, 315, 1–13. Available from: <https://doi.org/10.1016/j.sedgeo.2014.10.009>

**How to cite this article:** Dong, X., Hu, X., Lai, W., Xue, W., Zhang, S., Zhang, Y. et al. (2024) A global dataset of sandstone detrital composition by Gazzi-Dickinson method. *Geoscience Data Journal*, 11, 128–136. Available from: <https://doi.org/10.1002/gdj3.212>