Cluster analysis for a standardized classification and description of volcanic ash: Case study of the 1983 eruption at Miyakejima, Japan

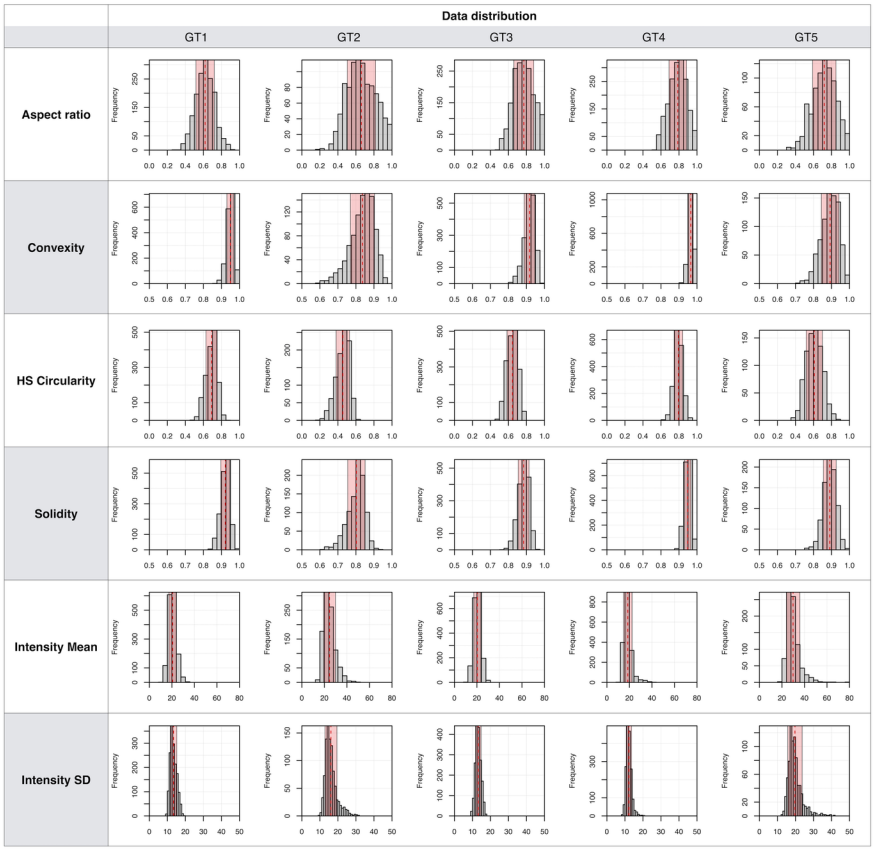
Rina Noguchi[1,1], Nobuo Geshi[2,2], Daigo Shoji[3,3], and Hideitsu Hino[4,4]

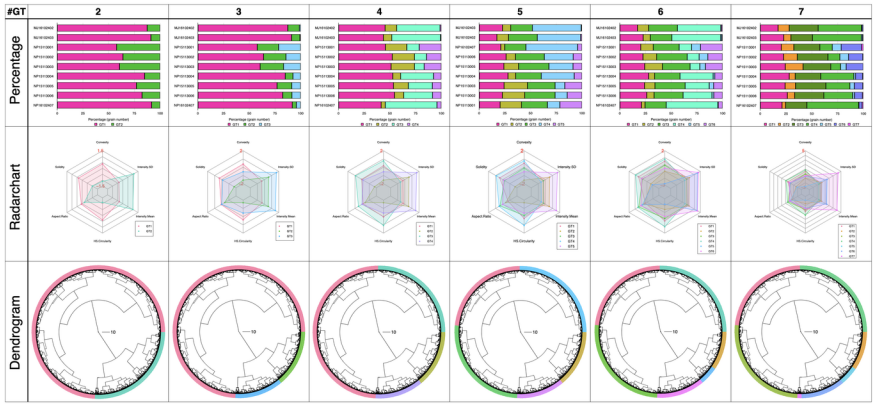
[1]Faculty of Science, Niigata University
[2]Geological Survey of Japan, AIST
[3]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency
[4]The Institute of Statistical Mathematics

January 20, 2023

Abstract

The composition of volcanic ash, which is a source of primary description data in volcanological study, is important information for estimating the eruption styles and sequences. However, its description under a microscope by human operation has difficulties in classification thresholds and time and effort-consumptions. This study demonstrates an accurate and rapid description of volcanic ash samples that consist of thousands of grains. We analyzed nine tephra samples (two magmatic (dry) and seven phreatomagmatic (wet)), which were produced in the 1983 A.D. fissure eruption event at Miyakejima volcano, Japan. Our dataset, which is consists of multivariate shape and transparency parameters, was rapidly obtained using an automated grain analyzer. In this study, we applied a two-step cluster analysis to objectively and quantitatively define grain type and classify samples. To define grain types, we referred to the statistically appropriate number of clusters of whole-ash grains in our samples. For our samples, the appropriate number of clusters for grain type was five. Each grain type is characterized by parameters and has different proportions among our samples. In wet tephra samples, grains that were categorized as transparent and highly irregularly shaped types were relatively abundant. Those grains can be considered as vesicular sideromelane grains, which are often found in products of phreatomagmatic eruptions. Such a standardized description of volcanic ash based on statistically determined grain type will contribute to initial descriptions before subsequent detailed analysis.





Cluster analysis for a standardized classification and description of volcanic ash: Case study of the 1983 eruption at Miyakejima, Japan

R. Noguchi[1], N. Geshi[2], D. Shoji[3], and H. Hino[4]

[1]Faculty of Science, Niigata University, 8050, Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan
[2]Geological Survey of Japan, National Institute of Advanced Science and Technology, AIST Site 7, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8567, Japan
[3]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1, Yoshinodai, Chuo-ku, Sagamihara, Kanagawa 252-5210, Japan
[4]The Institute of Statistical Mathematics, 10-3, Midori-cho, Tachikawa, Tokyo 190-8562, Japan

Key Points:

- Quantitative data treatment of thousands of volcanic ash grains is important for estimating the eruption styles and sequences.
- We perform an accurate and rapid description of volcanic ash samples using an automated grain analyzer and a two-step cluster analysis.
- Such a standardized quantitative and rapid description will contribute to initial descriptions before subsequent detailed analysis.

Corresponding author: Rina Noguchi, r-noguchi@env.sc.niigata-u.ac.jp

Abstract

The composition of volcanic ash, which is a source of primary description data in volcanological study, is important information for estimating the eruption styles and sequences. However, its description under a microscope by human operation has difficulties in classification thresholds and time and effort-consumptions. This study demonstrates an accurate and rapid description of volcanic ash samples that consist of thousands of grains. We analyzed nine tephra samples (two magmatic (dry) and seven phreatomagmatic (wet)), which were produced in the 1983 A.D. fissure eruption event at Miyakejima volcano, Japan. Our dataset, which is consists of multivariate shape and transparency parameters, was rapidly obtained using an automated grain analyzer. In this study, we applied a two-step cluster analysis to objectively and quantitatively define grain type and classify samples. To define grain types, we referred to the statistically appropriate number of clusters of whole-ash grains in our samples. For our samples, the appropriate number of clusters for grain type was five. Each grain type is characterized by parameters and has different proportions among our samples. In wet tephra samples, grains that were categorized as transparent and highly irregularly shaped types were relatively abundant. Those grains can be considered as vesicular sideromelane grains, which are often found in products of phreatomagmatic eruptions. Such a standardized description of volcanic ash based on statistically determined grain type will contribute to initial descriptions before subsequent detailed analysis.

Plain Language Summary

How to treat thousands of volcanic ash grains? The composition of volcanic ash is important information for estimating the eruption styles and sequences, though its data management has difficulties in classification thresholds and time and effort-consumptions. Here, we demonstrate an accurate and rapid description of volcanic ash samples that consist of thousands of grains and were formed by magmatic (dry) and phreatomagmatic (wet) eruptions. We obtained accurate multivariate shape and transparency data using an automated grain analyzer then applied a two-step cluster analysis to objectively and quantitatively classify ash grains/samples based on their parameters. For our samples, grains are classified into five types and each grain type is characterized by parameters and has different proportions among our samples. In wet samples, grains that were categorized as transparent and highly irregularly shaped types were relatively abundant. Those grains can be considered as vesicular sideromelane grains, which are often found in products of phreatomagmatic eruptions. Such a standardized description of volcanic ash based on statistically determined grain type will contribute to initial descriptions before subsequent detailed analysis performed by a human.

1 Introduction

The composition of volcanic ash, which is the source of primary description data in volcanological study, is important information for estimating eruption styles and sequences. For example, the proportion of juvenile material is used to infer the status of ascending magma (e.g., Nakada et al., 1995). Characteristics of texture and shape of juvenile material are used to infer its vesiculation and chilling histories (e.g., Sheridan & Wohletz, 1983; Geshi et al., 2019). In general, the composition of volcanic ash is described by microscopic observation, which has several issues for effective description. One of the main issues is to determine classification thresholds. Even for professionals, it is difficult to determine a threshold for grain classification. Ambiguous and/or featureless shapes of some grains cause difficulty in the class setting and judgment. For these reasons, the threshold highly depends on individual investigators and there are no integrated thresholds for the classification of volcanic ash grains. Furthermore, because it requires large numbers of grains (e.g., Suzuki et al. (2013) classified several hundred or more ash grains

in a size fraction of 250 to 500 µm), the manual (i.e., microscopic observation by human) description of volcanic ash components takes time and effort. Especially for ongoing eruptions, accurate and rapid primary descriptions are important for monitoring and forecasting eruption activity.

The development of grain analyses, statistical techniques, and machine learning methods has enabled automatic grain identification, measurement, and classification (e.g., Liu, Cashman, & Rust, 2015; Leibrandt & Le Pennec, 2015; Shoji et al., 2018). For morphometric analysis of volcanic ash grains, several shape parameters have been proposed (e.g., Dürig et al., 2019, and references therein). In general, those shape parameters are measured using optical projections or scanning electron microscope (SEM) silhouettes. Compared with SEM observation, automated particle analysis systems can measure many ash grains in a short time. For example, using an automated particle analyzer, Leibrandt and Le Pennec (2015) demonstrated the measurement of 3000 ash grains in ∼1.8 hours, while measurement of 500–1000 of ash grains took 3–5 hours using a SEM (Lautze et al., 2012). As many studies have presented, SEM observation is effective for detailed investigation of selected grains aiming to understand magma ascension and chilling processes; however, it is not suitable for primary, rapid, and comprehensive descriptions, such as aiming to understand the fraction of juvenile materials. Using silhouette images obtained by an automated particle analyzer, Shoji et al. (2018) demonstrated the classification of volcanic ash grains with a convolutional neural network, one of the machine learning techniques. Such a technique significantly improves the initial description of volcanic ash, especially from the viewpoint of time and effort consumption; however, dataset collection for training classification machines is a new challenge. Machine learning using a convolutional neural network requires a huge dataset consisting of more than 1000 images and consumes time and effort. The subjective view of the data preparer is unavoidably included in training data. To establish an integrated description procedure using these advanced instruments and techniques, effective and objective preparation and analysis of data are necessary.

In this study, we demonstrate an accurate and rapid description of volcanic ash composition with an automated grain analyzer and cluster analysis using ash samples from an observed eruption. The aim of this study was to develop an automatic assistance system for preparing the primary description before optical microscopic and SEM observations. This study will contribute to summarizing the composition and selecting ash grains before subsequent detailed investigations.

2 Method

2.1 Volcanic ash samples and their geological background

In this study, we analyzed nine tephra samples that were produced in the 1983 A.D. fissure eruption event (hereafter 1983 eruption) of Miyakejima volcano, Japan (Fig. 1). The rock type formed in this event is tholeiitic basalt (Aramaki et al., 1986). During the event, lava flow effusion, fountaining, and Strombolian eruptions ("dry") and phreatomagmatic eruptions ("wet") occurred in the same fissure vent system simultaneously. The initial eruption started on the southwestern flank of Oyama with a NE–SW trend, and then fissure vents propagated toward the northeast and southwest (Fig. 1). The eruption style was initially fountaining and shifted to localized Strombolian (Aramaki et al., 1986). At the southwest edge of the fissure vent, phreatomagmatic eruptions occurred and formed a tuff ring in the Nippana area (e.g., Aramaki et al., 1986). The Nippana tuff ring consists of surge deposits mainly composed of stratified sideromelane (Sumita, 1985). The lower units are characterized by block-sag structure, water-chilled bombs, and water-chilled scoria. The groundmass crystallinity of scoria produced at the flank (i.e., dry eruption) is slightly lower than that of the wet scoria produced at Nippana (Shimano & Nakada, 2006).

Figure 1. Map of Miyakejima Island, Japan. Sampling locations of this study are shown by blue stars.

We used two dry tephra samples and seven wet tephra samples generated in the 1983 eruption (Table 1). The dry tephra samples (MJ16102402 from lower and MJ16102403 from upper units, corresponding to E-1 and E-2, respectively, in the stratigraphic sequence in Endo et al. (1984)), which were generated by fountaining, were collected on the eastern crater rim of the northern Jinan-yama scoria cone (Fig. 1). At this site, samples were collected just above asphalt paving, which was the ground surface before the 1983 eruption (Figs. 2 and S1). The wet tephra samples (NP15113001–06, NP16102407) were collected at the crosscut outcrop of the Nippana tuff ring (Figs. 2 and S2). In the stratigraphic sequence of Endo et al. (1984), NP15113001 corresponds to S2-*l*, NP15113002–06 correspond to S2-*u*, and NP16102407 (the uppermost layer) corresponds to R-3. Because these tephra samples appeared to be unaltered and were collected from their original emplacements, we regarded their original shape and transparency be unmodified by subsequent alteration and transportation in the surface environment.

Table 1. List of samples used in the present study.

Eruption type (sample location)	Sample ID	Number of grains	Note
Magmatic (Jinan-yama)	MJ16102403	348	Above MJ16102402
	MJ16102402	221	Lower layer
Phreatomagmatic (Nippana)	NP16102407	790	Above NP15113006
	NP15113006	782	Above NP15113005
	NP15113005	692	Above NP15113004
	NP15113004	1097	Above NP15113003
	NP15113003	405	Above NP15113002
	NP15113002	697	Above NP15113001
	NP15113001	1820	Lower layer

Figure 2. Stratigraphic columns of the northern Jinan-yama scoria cone and the Nippana tuff ring after Endo et al. (1984)

Most of the grains in the 1983 eruption samples are juvenile glassy fragments. In dry samples, translucent glassy black/dark brown grains are dominant (Fig. 3A). These glassy grains are possibly tachylite, which is formed by rapid cooling of molten basaltic magma (e.g., White & Valentine, 2016). The irregular shape of some of those grains may have been caused by fluidal deformation. In wet samples, glassy light brown grains are dominant (Fig. 3B). Most of those grains are transparent and are considered to be siderome-lane, which dominates in more rapidly cooled magma, as occurs during magma-water interactions (e.g., White & Valentine, 2016). The 1983 eruption tephra contains 3-5 vol% of phenocrysts (Kuritani et al., 2003; Shimano & Nakada, 2006). According to Kuritani et al. (2003), phenocrysts of plagioclase, olivine, augite, titanomagnetite, and rare or-thopyroxene are present (97:1:2 for the modal proportion of plagioclase, olivine, and augite phenocrysts). In the 1983 eruption samples, crystal aggregates consisting of plagioclase, olivine, augite, and titanomagnetite were found (Kuritani et al., 2003). The glassy frag-ments contain microlites (e.g., Shimano & Nakada, 2006).

Figure 3. Microscopic images of example 1983 Miyakejima eruption samples. A: MJ16102402. B: NP15113002. Note that these images were taken under incident lighting.

2.2 Grain measurement and parameter selection

We used an automated grain analyzer to obtain accurate multivariate shape and transparency data rapidly. This study used the Morphologi G3S (Malvern Instrument Ltd.) automated particle analyzer at the Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology (AIST). The sieve size fraction of grains we used was 2ϕ to 3ϕ (125 to 250 µm). We chose this size fraction because it existed in sufficient quantities for statistical analysis. At $\times 5$ magnification, the measurable grain diameter ranged from 6.5 to 420 µm (Malvern, 2013). Therefore, the sample sizes are appropriate for these measurement conditions. In grain measurement, overlapped and combined grains affect the measured shape and transparency. To scatter the volcanic ash grains on the glass plate, we used a sample dispersion unit with an injection pressure of 1.5 bar and an injection time of 20 ms. During the measurement, the illumination was set to diascopic (bottom light), under automatic light calibration (calibration intensity of 80.00 and intensity tolerance of 0.20). The threshold for background separation (0 to 255) was set at 80 to obtain a sharp focus. The measurement lasted approximately 40 min for each sample. After the measurement, we excluded overexposed, unseparated, and cut-off grain images by hand in the Morphologi software. As a result, we obtained grain shape and transparency data for 6,852 grains in total (refer to the Supplementary Data).

The parameter selection step is essential for analyzing multivariate data. Taking the independence of parameters into account, Liu, Cashman, and Rust (2015) proposed adopting four parameters: convexity, solidity, axial ratio, and a form factor (i.e., high-sensitivity (HS) circularity in Morphologi G3S) for grain shape analysis of volcanic ash.

In addition to grain shape, Morphologi G3S can also measure grayscale luminance, which, under the diascopic lighting conditions, indicates grain transparency. Information regarding the grain transparency of volcanic ash is important for identifying the glass and crystal components (e.g., Miwa et al., 2015). In this context, we used four shape parameters (aspect ratio, convexity, HS circularity, and solidity) and two transparency values (intensity mean and standard deviation). Because the axial ratio is not provided by Morphologi G3S, we used the aspect ratio instead. The parameters used in this study—aspect ratio (A_r), convexity (C_v), solidity (S_d), HS circularity (H_c), intensity mean (I_m), and intensity standard deviation (I_{sd})—are derived as follows:

$$A_r = \frac{W}{L},$$

$$C_v = \frac{P_c}{P_g},$$

$$S_d = \frac{A_g}{A_g + A_c},$$

$$H_c = \frac{4 \times \pi \times A_g}{{P_g}^2},$$

$$I_m = \frac{\sum_{i=1}^{N} I_i}{N},$$

$$I_{SD} = \sqrt{\frac{\sum_{i=1}^{N} {I_i}^2 - \frac{\left(\sum_{i=1}^{N} I_i\right)^2}{N}}{N}},$$

where W is the length along the minor axis of the grain, L is the length along the major axis of the grain, P_c is the perimeter of the convex hull, P_g is the perimeter of the grain, A_c is the area of the convex hull, A_g is the area of the grain, I_i is the intensity [0 (opaque) to 255 (transparent)] of pixel i, and N is the total number of pixels in the grain (Malvern, 2013). Therefore, in this study, the low values of these shape parameters indicate more elongated, rough, and/or irregular shapes. These parameters were calculated for each ash grain.

Although each parameter is not directly correlated with the material, some are characteristic in particular materials. For example, fluidally elongated glass fragments and elongated crystals have a low aspect ratio. Concave sides caused by bubble walls affect convexity (small bubbles) and solidity (large bubbles) (e.g., Liu, Cashman, Rust, & Gislason, 2015; Liu, Cashman, & Rust, 2015). Surface tension within a fluid droplet before it cools contributes to decreasing the circularity of magma and lava fragments (e.g., Wohletz, 1983; Fitch & Fagents, 2020). Grains rounded by friction and collisions also show low circularity (e.g., Manga et al., 2011).

2.3 Cluster analysis

We applied a standard cluster analysis technique to objectively and quantitatively classify ash grains based on their shape and transparency parameters. Cluster analysis is a multivariate analysis method for unsupervised classification (e.g., Anderberg, 2014; Tan et al., 2005). In this study, we adopted Ward's hierarchical clustering method (Anderberg, 2014).

We performed a two-step clustering analysis to 1) categorize all ash grains (i.e., 6,852 grains) into a small number of grain types (grain clustering) and 2) represent samples by a feature vector composed of the ash fraction (sample clustering) (Fig. 4). In the first cluster analysis, we categorized all ash grains across the entire sample and then considered each cluster as a statistically determined grain type. After calculating the grain number percentages of each grain type for all samples, considering the proportions of grain types as the feature vector for the sample, we then categorized the samples. The sample clustering was used to consider the appropriate cluster number (i.e., number of grain types) in the grain clustering. Because the range of values differs between shape parameters (0 to 1) and intensity values (0 to 255), ash grain data were standardized (with mean 0 and standard deviation 1) before the analysis. Whole-ash grain data, including images, are presented in the Supplementary Data. To define grain types, we referred to the statistically appropriate number of clusters of whole-ash grains in our samples. There are several methods for determining the number of clusters. For example, the R package NbClust provides 30 different indices for determining the appropriate number of clusters (Charrad et al., 2012). In the NbClust package, the best number of clusters is determined by a majority vote on the optimal numbers of clusters, which are defined for each index based on maximum/minimum differentiation. Some of these indices require a heavy computational burden, particularly considering that our data included 6,852 ash grains. In this case, the package authors recommend the use of only 26 of the 30 indices, which is more computationally efficient and does not depend on visual inspection. Therefore, we used the NbClust package to determine the appropriate cluster number for 6,852 ash grains.

3 Results

3.1 Determination of statistical grain types

The appropriate number of grain clusters in our sample was suggested as five based on the result of the NbClust package. Fig. 5 shows the number of indices with the appropriate cluster number for each and all samples. For most samples, the appropriate cluster number is two or three. MJ16102403 has the maximum appropriate cluster number among all samples, five. For all ash grains in our samples, the appropriate cluster number is three. Fig. S3 shows classification results based on the fraction of grain type in each sample with two to seven clusters. The structure of classification did not change when the number of clusters was four or more. The Euclidean distance became stable when the cluster number was five or more. Given this background, we applied five as the cluster number to determine the grain type for our samples. The results for the other numbers of grain type analyses are shown in Fig. S4.

Fig. 6 shows the characteristics of the shape and transparency parameters for each cluster in the case of five grain types. Here we refer to each cluster as grain types (GTs) 1 to 5. GT1, consisting of 1,586 grains, has the lowest aspect ratio and the highest convexity and solidity. From this feature, GT1 is characterized by grains with an elongated and smooth surface (fewer spiky edges). GT2, consisting of 957 grains, has significantly low shape parameters, especially solidity. The intensity parameters of GT2 are high. These features characterize GT2 as transparent irregular-shaped grains with spiky edges. GT3, consisting of 1,768 grains, is a featureless category except for a slightly high aspect ratio. This indicates that GT3 grains have an equant (blocky) shape. GT4, consisting of

Figure 4. Cluster analysis procedure used in this study.

Figure 5. Rose diagram for appropriate number of grain type clusters calculated by the NbClust package for each sample.

1,728 grains, has the lowest intensity parameters and highest shape parameters, especially HS circularity. These features suggest that GT4 grains are opaque and have an equant rounded shape without spiky edges. GT5, consisting of 813 grains, has low convexity and HS circularity. The intensity parameters of GT5 are significantly high. These features characterize GT5 as very transparent and irregular-shaped grains. The frequency of parameters in each grain type is shown in Fig. S5.

Figure 6. Characteristics of each grain type in cluster analysis. (A) Dendrogram of the cluster analysis. The number after the grain type indicates the number of grains in each type. (B) Parameter characteristics of each grain type. Note that each parameter is shown with a standardized scale (with mean 0 and standard deviation 1).

Fig. 7 shows example silhouette images of statistically classified grains for the five grain types. GT1 grains are relatively opaque and have an elongated angular shape (acute to right angles). Most sides of GT1 grains are flat, but some of them are partially concave. Some GT1 grains are transparent and have a fluidally elongated shape. Some transparent GT1 grains contain bubbles and microlites (possibly plagioclase). GT2 grains are transparent and have significantly irregular and concave outlines mainly derived from bubble walls. Some GT2 grains show a fluidally elongated shape. Cleary, the transparent GT2 grains contain bubbles and microlites. GT3 grains are relatively opaque and have rough outlines. Some sides of GT3 grains are partially concave. Inside the relatively transparent GT3 grains, microlites and bubbles were observed. GT4 grains are relatively opaque and have an equant shape with a smooth outline. Most sides of GT4 grains are convex. Some GT4 grains are microlite-rich and contain bubbles. GT5 grains are transparent and have irregular and angular shapes. Inside GT5 grains, microlites are frequently observed (Fig. S6). Most of the GT5 grains contain bubbles and some of them are as large as penetrating grains.

3.2 Composition based on grain types

Our tephra samples were characterized by a fraction of grain numbers falling into each grain type, which was determined by cluster analysis (Fig. 8). GT2 and GT5 grains are more common in wet samples from lower layers of the Nippana tuff ring (NP15113001, 02, and 03) than in the others. More than 40 % of grains of the dry samples (MJ16102402 and MJ16102403) and wet sample NP16102407 consists of GT4 grains. The dry samples have quite a few numbers of GT5 grains (less than 1.3 %). There are no large differences in fractions of GT1 and GT3 among our samples; the average and the standard deviation are 22.9 $\pm$ 3.2 % and 26.0 $\pm$ 3.4 %, respectively.

Figure 7. Example images of each grain type.

Figure 8. Grain type composition of 1983 eruption tephra.

4 Discussion

4.1 Can we infer eruption details from statistically determined grain types?

When the classification results are combined with microscopic observation, most of the GT2 and GT5 grains could be considered as sideromelane shards. As shown in Figs. 6 and 8, GT2 and GT5 grains have higher I_m values (i.e., transparency) than those of the other grain types, and are common in wet tephra samples. Of course, there is a possibility that they are other transparent materials, such as plagioclase phenocrysts, however, they are minor in our samples, and the existence of microlites inside of those grains is strong evidence that they are glass shards. This interpretation is consistent with the abundance of sideromelane in products of phreatomagmatic eruptions (e.g., White & Valentine, 2016). The other opaque grains of GT1, GT3, and GT4 could correspond to tachylite, other colored minerals, and crystal aggregates. Because the tephra used in this study mostly consisted of juvenile materials, the variety of grains is small. Our procedure will be more effective for more complicated tephra, which contains both juvenile and altered/recycled materials.

The abundance of GT2 grains in lower units of the Nippana tuff ring is considered to be caused by intense vaporization and quenching of magma. GT2 grains are considered as sideromelane, as discussed above, and have vesicular outlines. Vesicular grains have been used as one of the indicators of phreatomagmatic eruptions (e.g., Liu, Cashman, Rust, & Gislason, 2015; Schmith et al., 2017). Liu, Cashman, Rust, and Gislason (2015) suggested that a vesicular grain is generated by pre-fragmentation vesiculation prior to subsequent brittle fragmentation by rapid quenching. An abundance of GT2 grains, as an indicator of such a fragmentation scheme, would imply the effect of external water during an eruption. However, as Liu, Cashman, Rust, and Gislason (2015) suggested, whether a vesicular feature appears depends on the target grain and bubble sizes. They showed that the proportion of bubbly particles (i.e., vesicular grains) in ash samples decreases as the grain size approaches the modal vesicle diameter. If the target grain size and modal vesicle diameter are in the same range, the proportion of vesicular grains could have less sensitivity for scaling with the effect of external water in eruptions. Thus, GT2 grains have the potential to be an indicator of a phreatomagmatic eruption, but the relationship between grain and bubble size should be considered.

Other types of data that can be measured rapidly will increase the volcanological meaning of automatic initial descriptions of grains. In this study, only the shape and transparency data obtained from silhouette images were used. Luminance data, such as RGB composition, will greatly help to identify oxidized/altered grains and colored minerals, as shown in Miwa et al. (2015). Rapid measurement by Raman spectroscopy, which has already been implemented (e.g., Kammrath et al., 2018), also will contribute to ash classification.

Statistically determined grain types in this study are not applicable to other volcanic ash samples because the determination highly depended on the specific dataset. The appropriate number of grain types is not always five and depends on the sample analysis. Perhaps, data profiling and cluster analysis with samples from several eruption styles and geological backgrounds will make such clusterings more generalizable. In the future, after measurement and analysis to establish statistical grain types that cover many eruption styles and volcanology more meaningfully, an automatic classification system can be built. One of the procedures is setting parameter-based thresholds. However, as shown in Fig. S5, because a statistically determined grain type has a wide range of parameters that often overlap with other types, setting a threshold is complicated. To solve this issue, machine learning techniques, such as neural networks, can be applied. Trained classification models can classify volcanic ash accurately and rapidly (Shoji et al., 2018). One of the advantages to applying neural networks is that they can calculate the class probability. Because volcanic ash is a natural product, it contains ambiguous grains that are

difficult to classify. In such a case, the concept of probability in volcanic ash grain classification will be useful. Even if a very accurate classification system could be developed, it would only support the first description. In subsequent detailed analyses, manual microscopic and other observations by experts are required.

4.2 Application to other materials

The initial description scheme for grain materials can be adapted to other fields, such as sedimentology and planetary science. Grain classification is a common theme in geology (e.g., Drolon et al., 2000; MacLeod, 2002). Our procedure could improve the quality of those classifications by applying transparency. In the case of the Martian Moons eXploration mission, which is scheduled to return grain samples from Phobos in the late 2020s, about 1 million grains, which may include not only Phobos material but also Martian samples (Hyodo et al., 2019), will be described and classified without supervised classification. To avoid unexpected pollution, which affects subsequent detailed analyses (e.g., detection of organic matter), the initial description of those returned samples will be performed by observation systems that do not damage the samples. The development of an automatic description and classification method for large numbers of grains based on shape and transparency would play an important role in initial and subsequent analyses to achieve the mission goal.

5 Conclusion

In this study, we demonstrated an accurate and rapid description method for volcanic ash composition with an automated grain analyzer and cluster analysis using ash samples from the 1983 Miyakejima eruption. We used samples produced in magmatic ("dry") and phreatomagmatic ("wet") eruptions that occurred in the same fissure eruption system simultaneously. Using four shape parameters and two transparency parameters, we found five statistically determined grain types following the suggested appropriate cluster number. Each grain type was characterized by parameters that had different proportions among our samples. In wet tephra samples, grains that were categorized as transparent and highly irregular-shaped were relatively abundant. Those grains can be considered as vesicular sideromelane grains, which are often found in the products of phreatomagmatic eruptions. Such a statistically determined grain type could be used as supervised data in a machine learning procedure in further automatic grain classification, though analysis of samples covering many eruption styles is needed. The development and improvement of our procedure will contribute to initial descriptions before subsequent detailed analysis performed by a human.

Acknowledgments

The grain data used in the present study were collected using Morphologi G3S at the Geological Survey of Japan, AIST. Data sets use this study are available online (at `https://doi.org/10.6084/m9.figshare.14676045.v1`). This work was supported by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (grant numbers 17H02063 and 20H01986 for RN, and CREST JPMJCR1761 for HH) and the Joint Usage/Research Center Program of the Earthquake Research Institute, University of Tokyo (grant number 2015-B-04).

References

Anderberg, M. R. (2014). *Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks* (Vol. 19). Academic press.

Aramaki, S., Hayakawa, Y., Fujii, T., Nakamura, K., & Fukuoka, T. (1986). The

October 1983 eruption of Miyakejima volcano. *Journal of Volcanology and Geothermal Research*, *29*(1-4), 203–229.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2012). NbClust Package: finding the relevant number of clusters in a dataset. *J. Stat. Softw.*

Drolon, H., Druaux, F., & Faure, A. (2000). Particles shape analysis and classification using the wavelet transform. *Pattern Recognition Letters*, *21*(6-7), 473–482.

Dürig, T., Bowman, M. H., White, J. D., Murch, A., Mele, D., Verolino, A., & Dellino, P. (2019). PARTIcle Shape ANalyzer PARTISAN–an open source tool for multi-standard two-dimensional particle morphometry analysis. *Annals of Geophysics*, *61*(6 sup), 671.

Endo, K., Miyaji, N., Chiba, T., Sumita, M., & Sakatsume, K. (1984). Tephrastratigraphical study on the 1983 Miyake-jima eruption. *Bulletin of the Volcanological Society Japan*, *2*, 184–207.

Fitch, E. P., & Fagents, S. A. (2020). Characteristics of rootless cone tephra emplaced by high-energy lava–water explosions. *Bulletin of Volcanology*, *82*(8), 1–16.

Geshi, N., Németh, K., Noguchi, R., & Oikawa, T. (2019). Shift from magmatic to phreatomagmatic explosions controlled by the lateral evolution of a feeder dike in the Suoana-Kazahaya eruption, Miyakejima Volcano, Japan. *Earth and Planetary Science Letters*, *511*, 177–189.

Hyodo, R., Kurosawa, K., Genda, H., Usui, T., & Fujita, K. (2019). Transport of impact ejecta from mars to its moons as a means to reveal Martian history. *Scientific reports*, *9*(1), 1–6.

Kammrath, B. W., Koutrakos, A., Leary, P. E., Castillo, J. A., Wolfgang, J., & Huck-Jones, D. (2018). Morphologically directed raman spectroscopic analysis of forensic samples. *Spectroscopy*, *33*(1), 46–53.

Kuritani, T., Yokoyama, T., Kobayashi, K., & Nakamura, E. (2003). Shift and rotation of composition trends by magma mixing: 1983 eruption at Miyake-jima Volcano, Japan. *Journal of Petrology*, *44*(10), 1895–1916.

Lautze, N. C., Taddeucci, J., Andronico, D., Cannata, C., Tornetta, L., Scarlato, P., ... Castro, M. D. L. (2012). SEM-based methods for the analysis of basaltic ash from weak explosive activity at Etna in 2006 and the 2007 eruptive crisis at Stromboli. *Physics and Chemistry of the Earth, Parts A/B/C*, *45*, 113–127.

Leibrandt, S., & Le Pennec, J.-L. (2015). Towards fast and routine analyses of volcanic ash morphometry for eruption surveillance applications. *Journal of Volcanology and Geothermal Research*, *297*, 11–27.

Liu, E. J., Cashman, K. V., & Rust, A. (2015). Optimising shape analysis to quantify volcanic ash morphology. *GeoResJ*, *8*, 14–30.

Liu, E. J., Cashman, K. V., Rust, A., & Gislason, S. (2015). The role of bubbles in generating fine ash during hydromagmatic eruptions. *Geology*, *43*(3), 239–242.

MacLeod, N. (2002). Geometric morphometrics and geological shape-classification systems. *Earth-Science Reviews*, *59*(1-4), 27–47.

Malvern. (2013). Morphologi G3 User Manual [Computer software manual].

Manga, M., Patel, A., & Dufek, J. (2011). Rounding of pumice clasts during transport: field measurements and laboratory studies. *Bulletin of Volcanology*, *73*(3), 321–333.

Miwa, T., Shimano, T., & Nishimura, T. (2015). Characterization of the luminance and shape of ash particles at Sakurajima volcano, Japan, using CCD camera images. *Bulletin of Volcanology*, *77*(1), 5.

Nakada, S., Motomura, Y., & Shimizu, H. (1995). Manner of magma ascent at Unzen Volcano (Japan). *Geophysical Research Letters*, *22*(5), 567–570.

Schmith, J., Höskuldsson, Á., & Holm, P. M. (2017). Grain shape of basaltic ash populations: implications for fragmentation. *Bulletin of Volcanology*, *79*(2),

14.

Sheridan, M. F., & Wohletz, K. H. (1983). Hydrovolcanism: basic considerations and review. *Journal of Volcanology and Geothermal Research*, *17*(1-4), 1–29.

Shimano, T., & Nakada, S. (2006). Vesiculation path of ascending magma in the 1983 and the 2000 eruptions of Miyakejima volcano, Japan. *Bulletin of volcanology*, *68*(6), 549–566.

Shoji, D., Noguchi, R., Otsuki, S., & Hino, H. (2018). Classification of volcanic ash particles using a convolutional neural network and probability. *Scientific reports*, *8*(1), 8111.

Sumita, M. (1985). Ring-shaped cone formed during the 1983 Miyake-jima eruption. *Bull. Volcanol. Soc. Japan*, *30*, 11-32.

Suzuki, Y., Nagai, M., Maeno, F., Yasuda, A., Hokanishi, N., Shimano, T., ... Nakada, S. (2013). Precursory activity and evolution of the 2011 eruption of Shinmoe-dake in Kirishima volcano—insights from ash samples. *Earth, planets and space*, *65*(6), 11.

Tan, P., Steinbach, M., & Kumar, V. (2005). *Chapter 8, cluster analysis: Basic concepts and algorithms in introduction to data mining.* Pearson Press, New York, New, York, USA.

White, J. D., & Valentine, G. A. (2016). Magmatic versus phreatomagmatic fragmentation: absence of evidence is not evidence of absence. *Geosphere*, *12*(5), 1478–1488.

Wohletz, K. H. (1983). Mechanisms of hydrovolcanic pyroclast formation: grain-size, scanning electron microscopy, and experimental studies. *Journal of Volcanology and Geothermal Research*, *17*(1-4), 31–63.

Figure 1.

139.50 °E

N

2000 caldera

Oyama

Ako

Jinan-yama

MJ16102402, 03

Izu Peninsula

Miyakejima

Legend

Sampling locations

1983 fissure vents

1983 lava flows

NP15113001–06,
NP16102407

Nippana

0 500 1,000 m

34.05 °N

Figure 2.

Stratigraphic sequences in
Endo et al., 1984

E-4
E-3
E-2
E-1

MJ16102403
MJ16102402

50 cm

Jinan-yama

R-5
R-4
R-3

NP16102407

S2-*u*

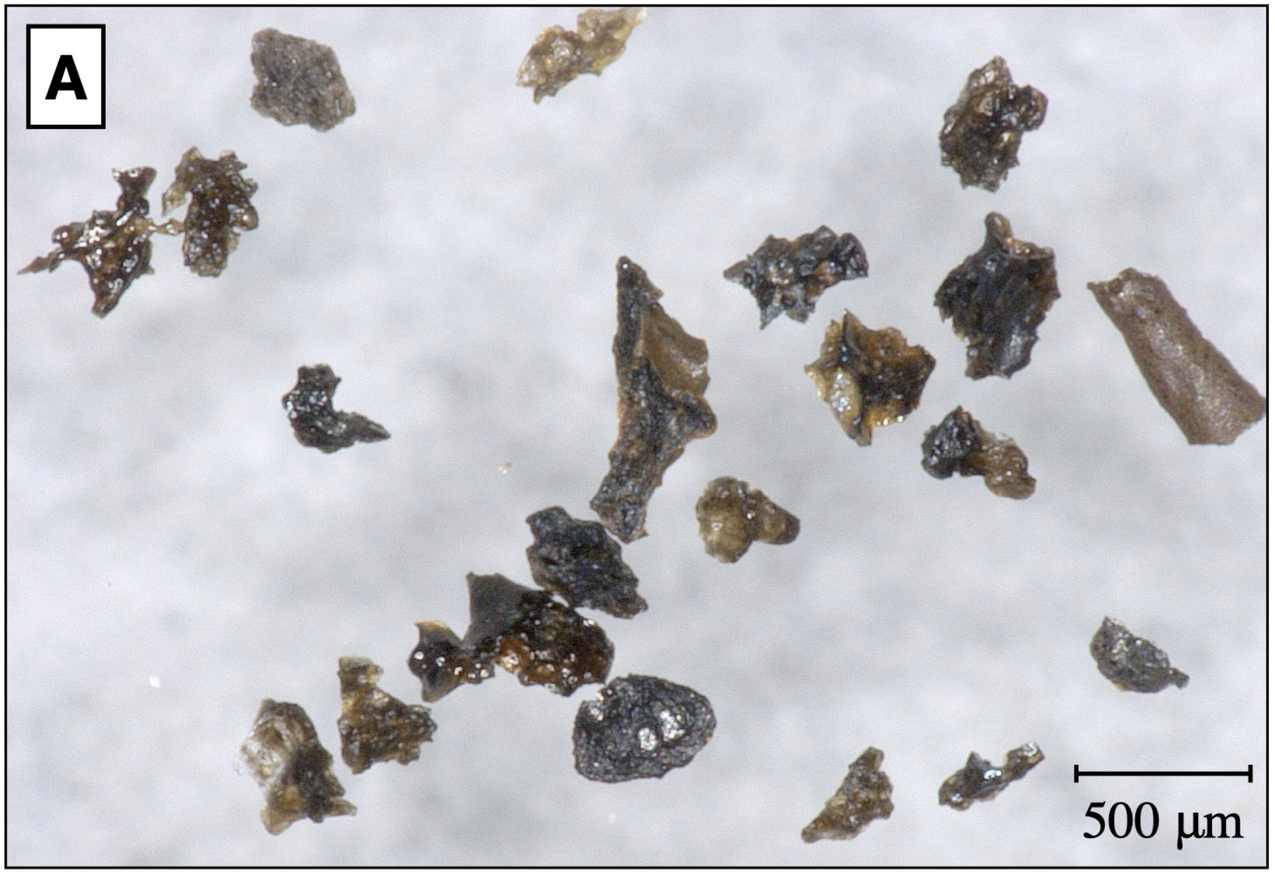
NP15113006
NP15113005
NP15113004
NP15113003
NP15113002

PQ-4
PQ-2
S2-*l*

NP15113001

Nippana

Figure 3.


A
500 µm

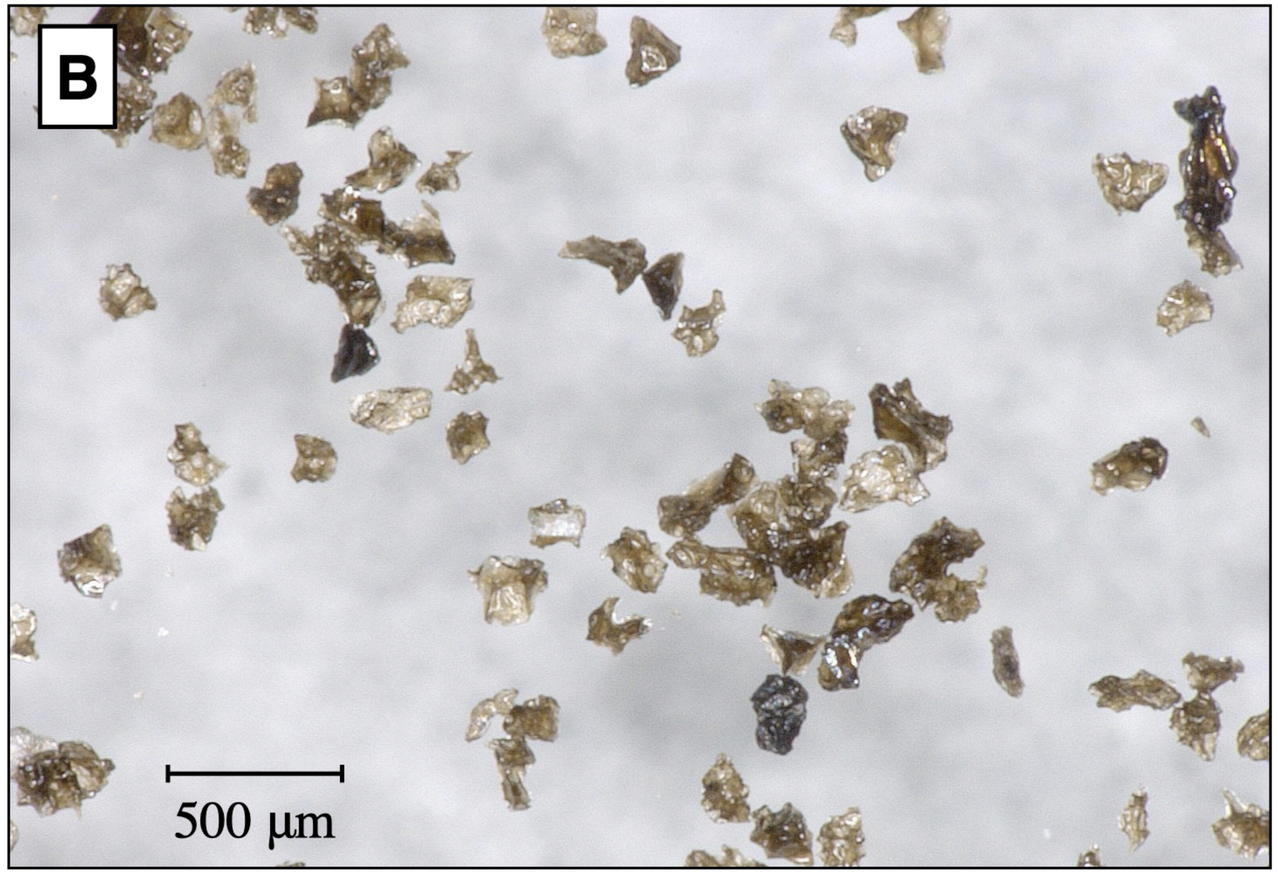

B
500 µm

Figure 4.

Sample 1

Sample 2

Sample N

Cluster analysis 1: Grain-type determination

Grain type 1

Grain type 2

Grain type N

Sample 1

GT 1 : xx %

GT 2: xx %

GT **N** : xx %

Sample 2

GT 1 : xx %

GT 2: xx %

GT **N** : xx %

Sample N

GT 1 : xx %

GT 2: xx %

GT **N** : xx %

Cluster analysis 2: Sample clustering

Sample xx

Sample xx

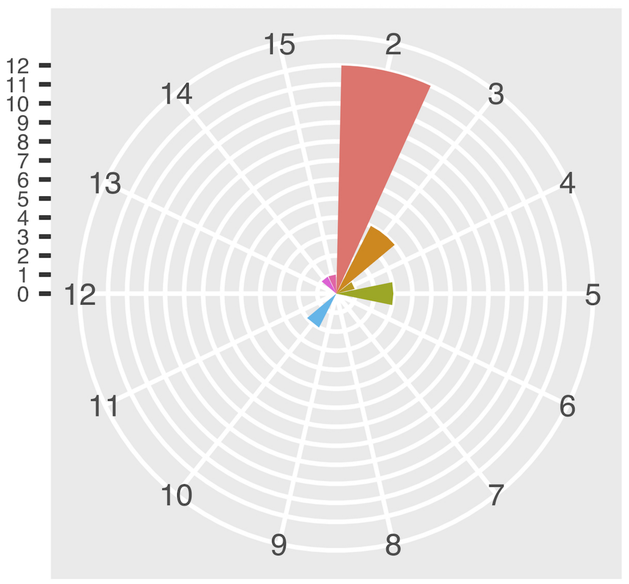
Sample xx

Sample xx

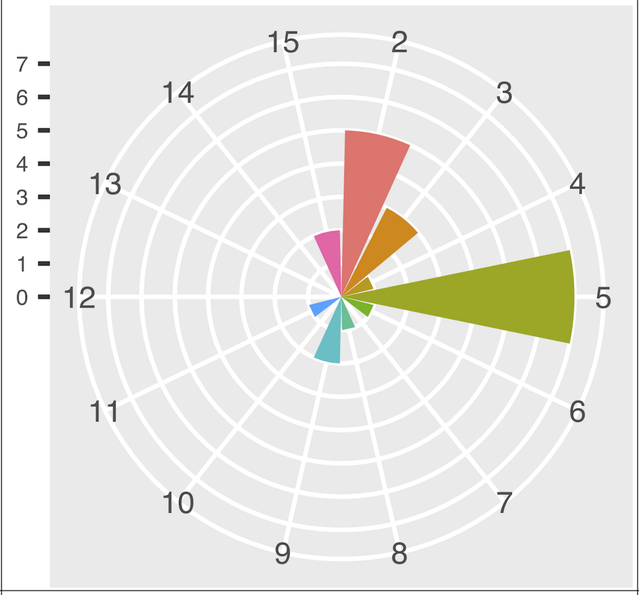
Sample xx

Figure 5.

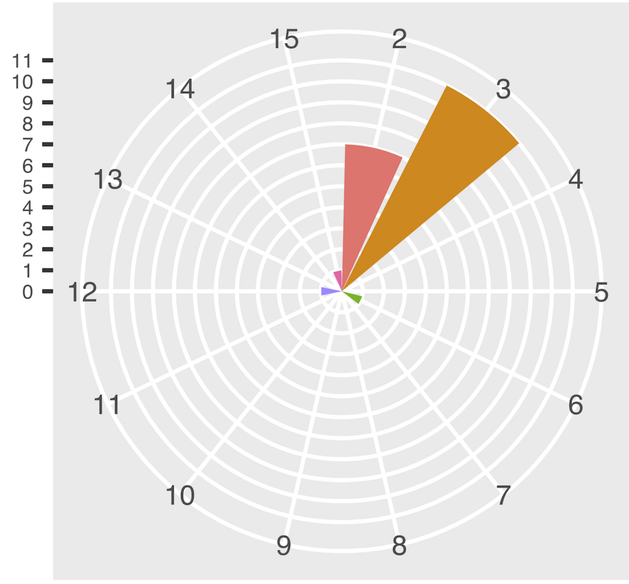
MJ16102402



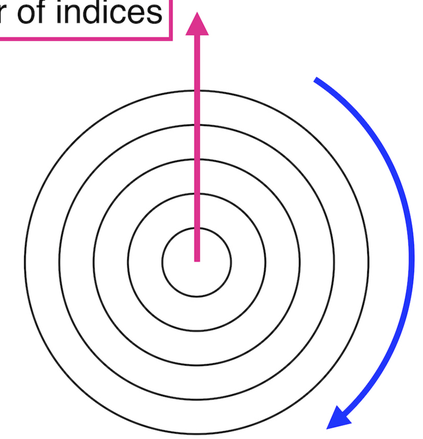
MJ16102403

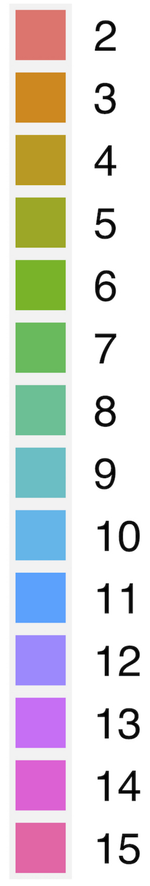


All grains

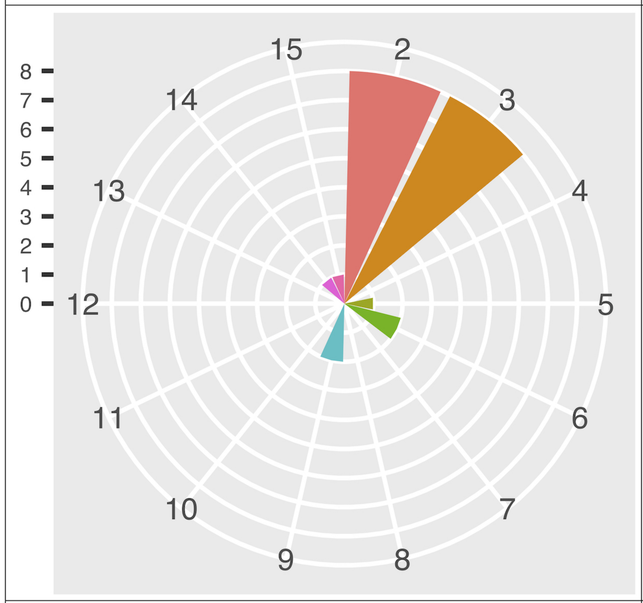


Number of indices

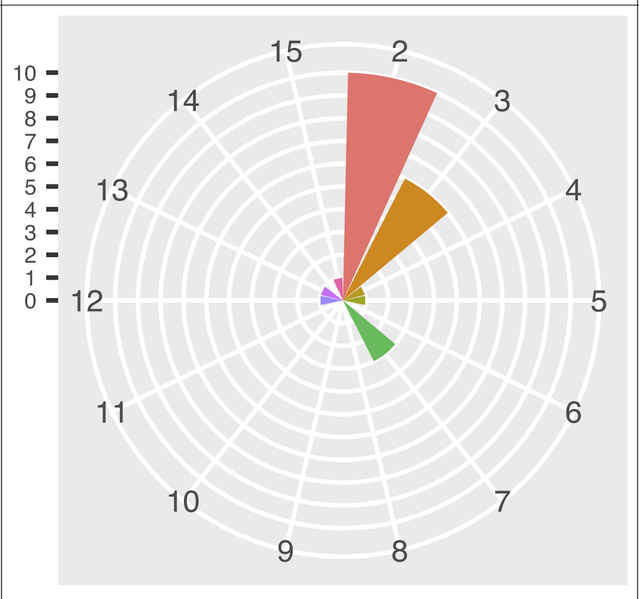




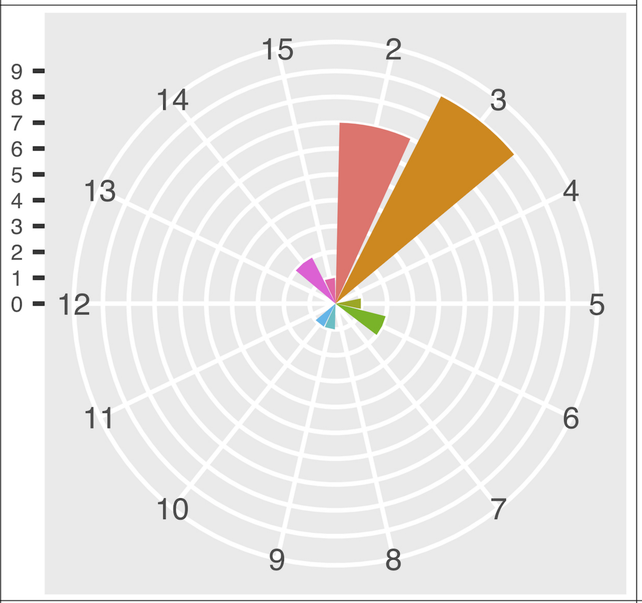
NP15113001



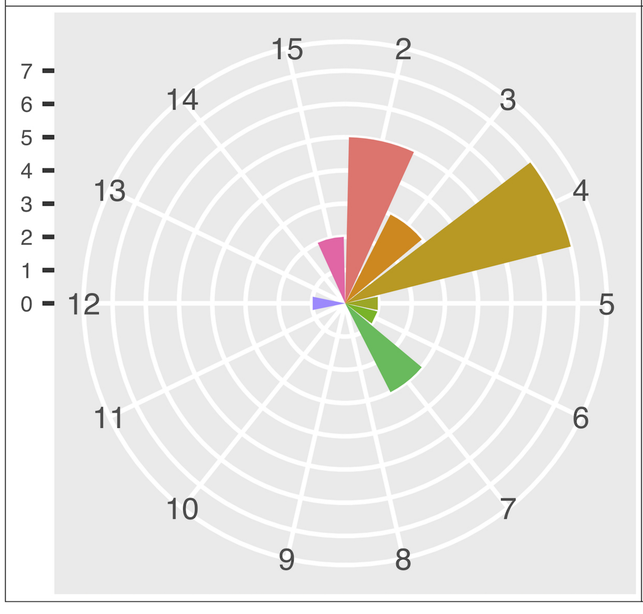
NP15113002



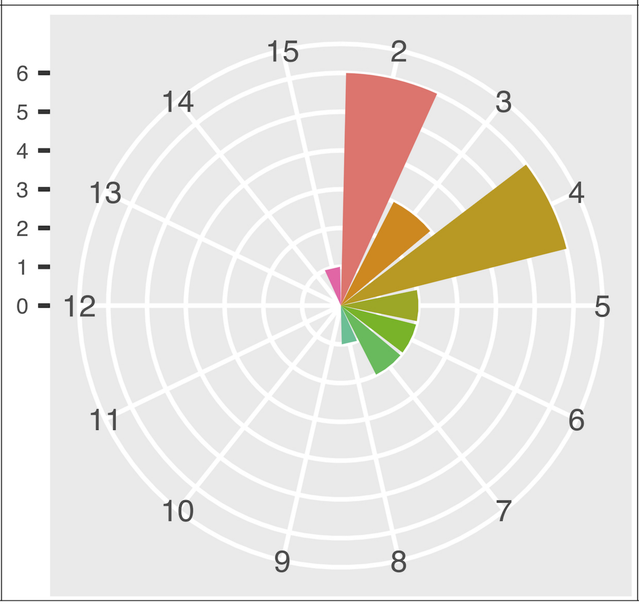
NP15113003



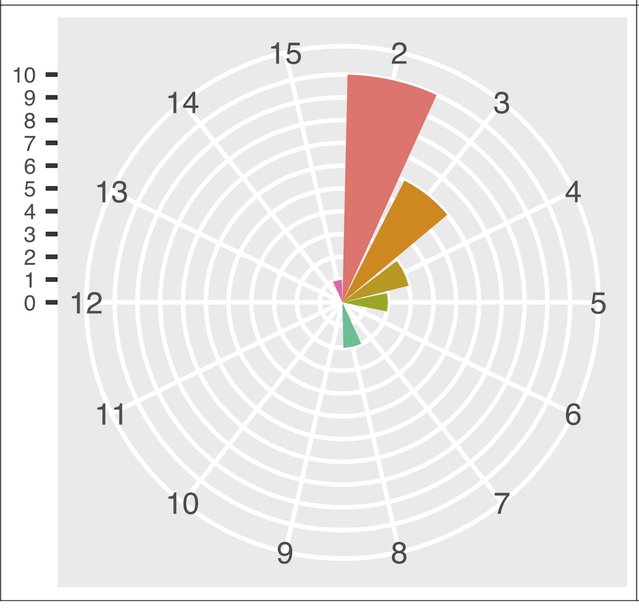
NP15113004



NP15113005



NP15113006



NP16102407

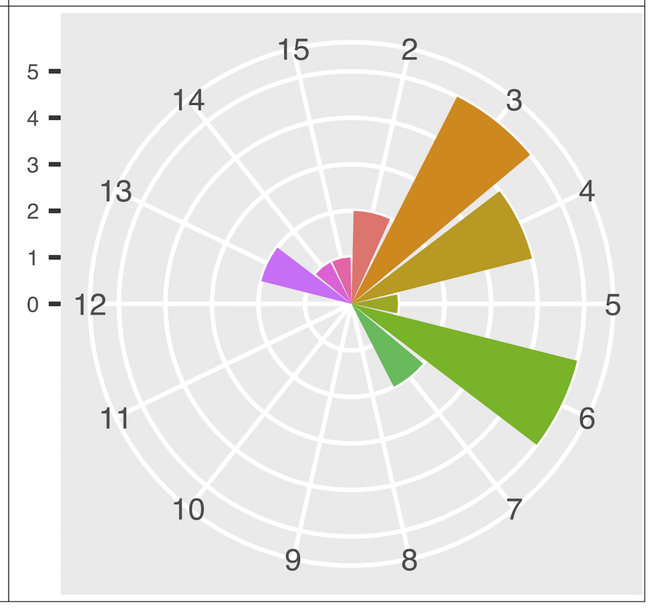


Figure 6.

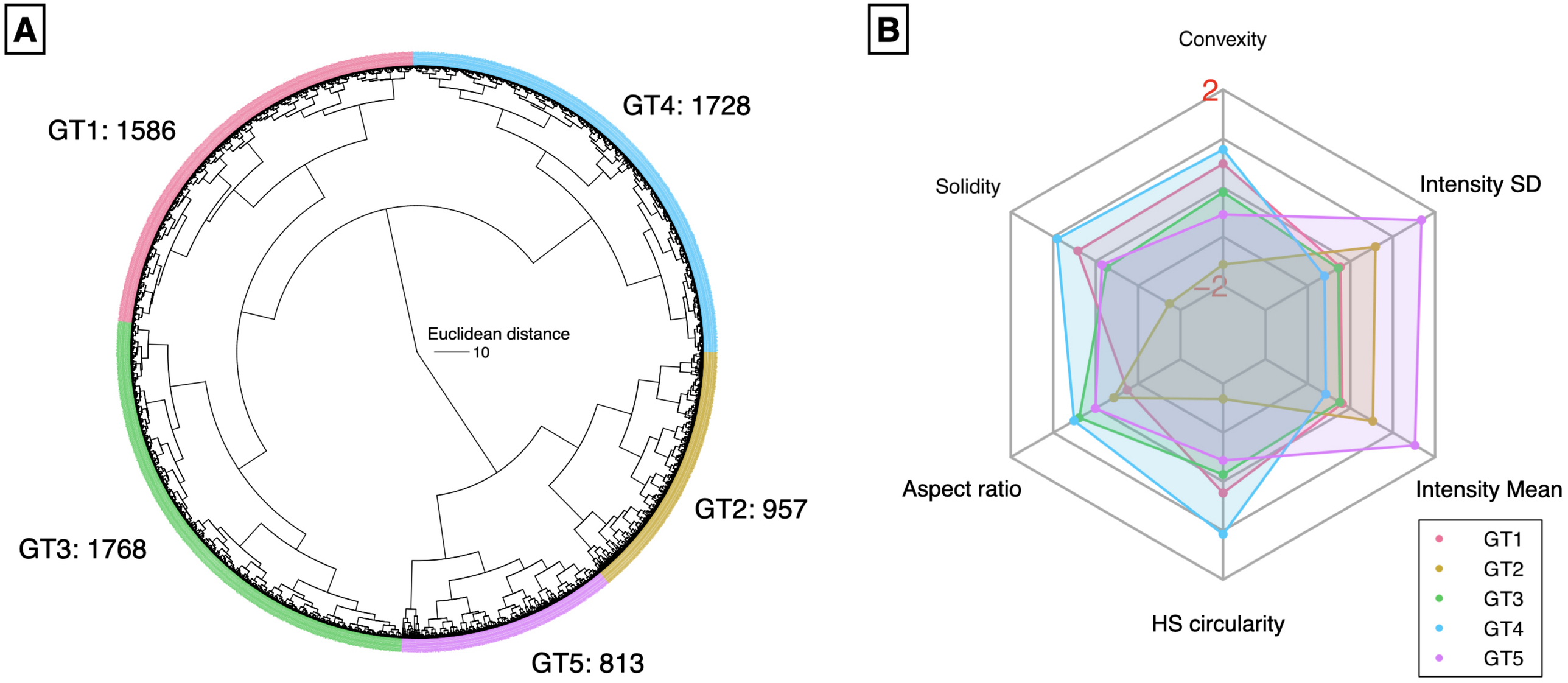

A
GT1: 1586
GT4: 1728
Euclidean distance
10
GT2: 957
GT3: 1768
GT5: 813
B
Convexity
2
−2
Solidity
Intensity SD
Aspect ratio
Intensity Mean
HS circularity
GT1
GT2
GT3
GT4
GT5

Figure 7.

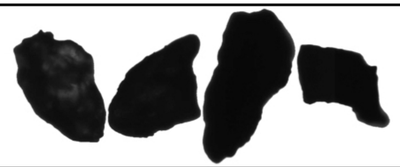
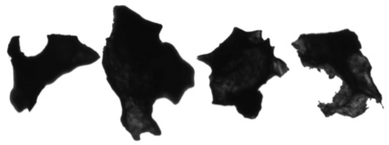
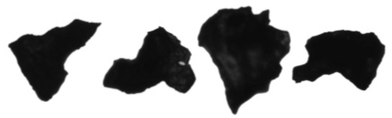
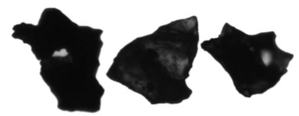
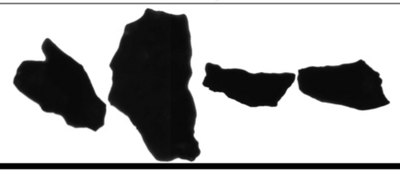
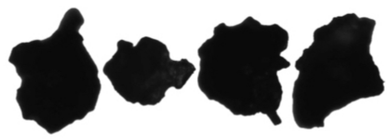
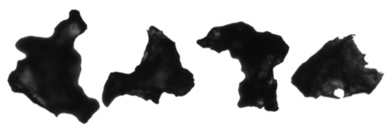
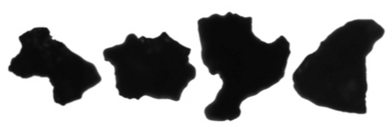
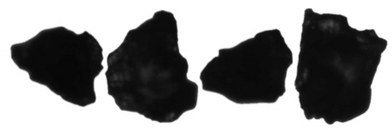
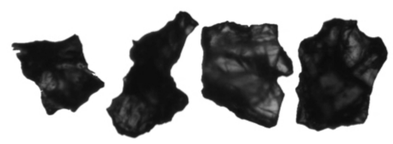
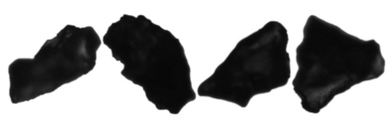
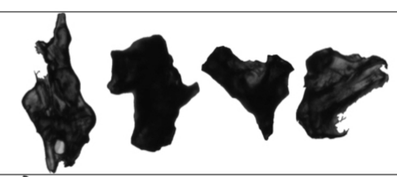
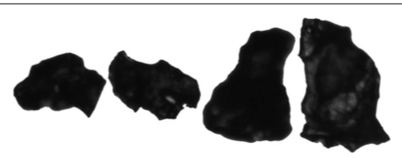
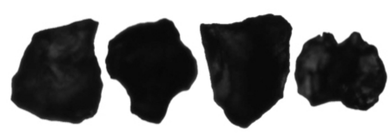
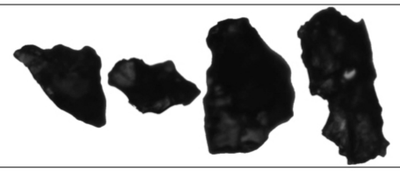
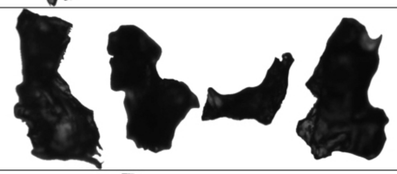
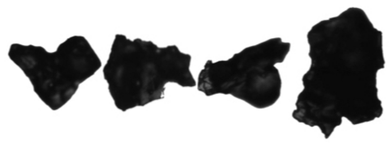
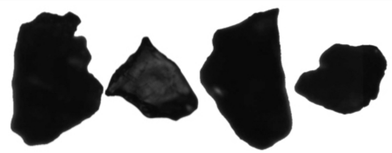
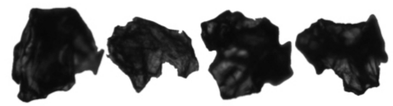
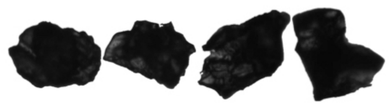
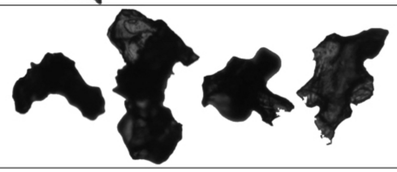
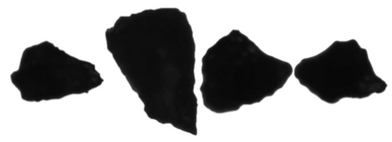
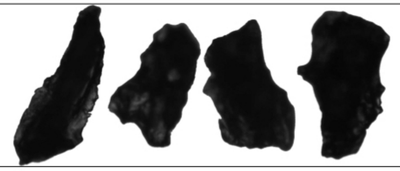
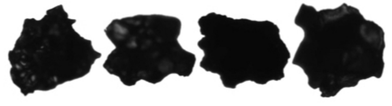
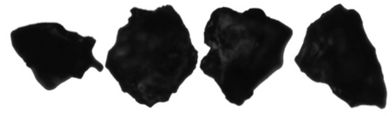
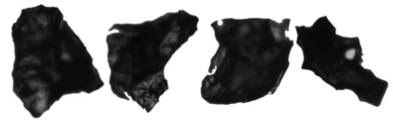
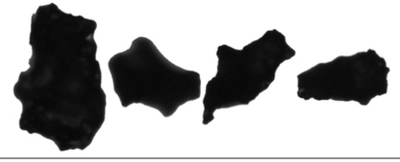
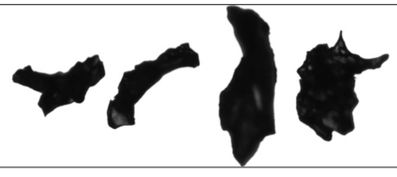
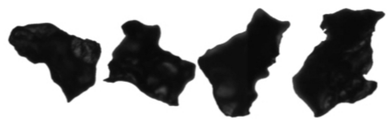
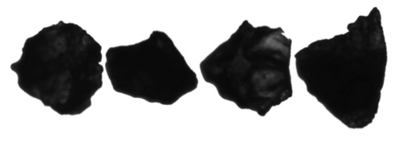
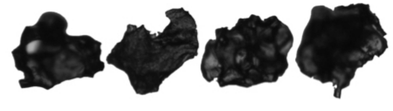
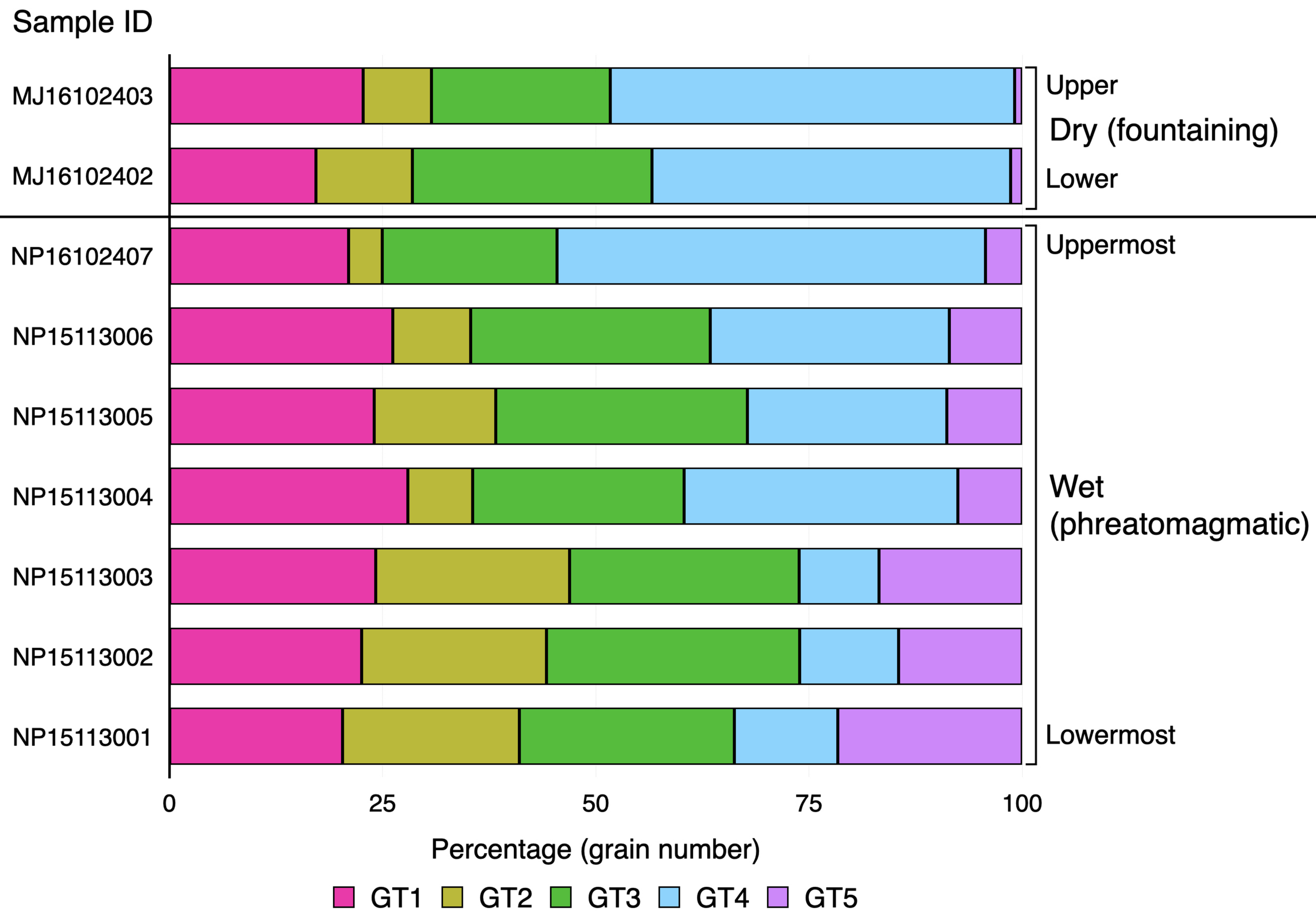
Sample	GT1	GT2	GT3	GT4	GT5
MJ16102403					
MJ16102402					
NP16102407					
NP15113006					
NP15113005					
NP15113004					
NP15113003					
NP15113002					
NP15113001					

Figure 8.


Sample ID
MJ16102403
MJ16102402
NP16102407
NP15113006
NP15113005
NP15113004
NP15113003
NP15113002
NP15113001
Upper
Dry (fountaining)
Lower
Uppermost
Wet
(phreatomagmatic)
Lowermost
0
25
50
75
100
Percentage (grain number)
GT1
GT2
GT3
GT4
GT5

Supporting Information for "Cluster analysis for a standardized classification and description of volcanic ash: Case study of the 1983 eruption at Miyakejima, Japan"

R. Noguchi[1], N. Geshi[2], D. Shoji[3], and H. Hino[4]

[1]Faculty of Science, Niigata University, 8050, Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan

[2]Geological Survey of Japan, National Institute of Advanced Science and Technology, AIST Site 7, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8567, Japan

[3]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1, Yoshinodai, Chuo-ku, Sagamihara, Kanagawa 252-5210, Japan

[4]The Institute of Statistical Mathematics, 10-3, Midori-cho, Tachikawa, Tokyo 190-8562, Japan

Contents of this file

Additional Supporting Information (Files uploaded separately)

Introduction

This supporting information is consists of figures and a dataset.

Supporting figures

Supporting figures are appeared in page X-4 to X-9 of this file.

Supporting data set

Data Set S1. This dataset contain volcanic ash grain data that we obtained and used. The dataset is available online (at `https://doi.org/10.6084/m9.figshare.14676045.v1`). As shown in the main text, grain data was obtained using an automated grain analyzer (Morphologi G3S (Malvern Instrument Ltd.)). The structure of the dataset is

/morphologi_data:

This directory contains grain data for each sample measured by Morphologi G3S (Malvern Panalytical Ltd.).

/results:

This directory contains the results of the cluster analysis.

/results/centroids:

This directory contains calculated centroids by the cluster analysis in each cluster number. The first column in each file shows the grain type. The "result_CA_all" column shows the number of grains fall in each grain type.

/results/labeled_data:

This directory contains grain type-labeled ("hclust.label" column) Morphologi data for each sample and for each grain type number.

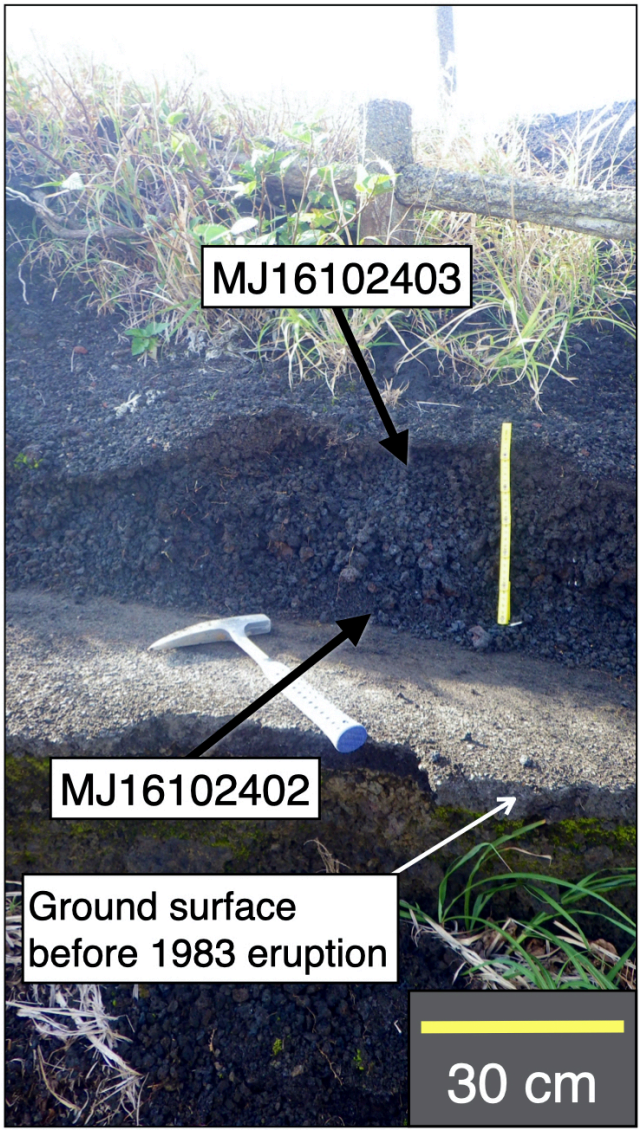


Figure S1. Sampling points on the eastern rim of the northern Jinan-yama scoria cone.

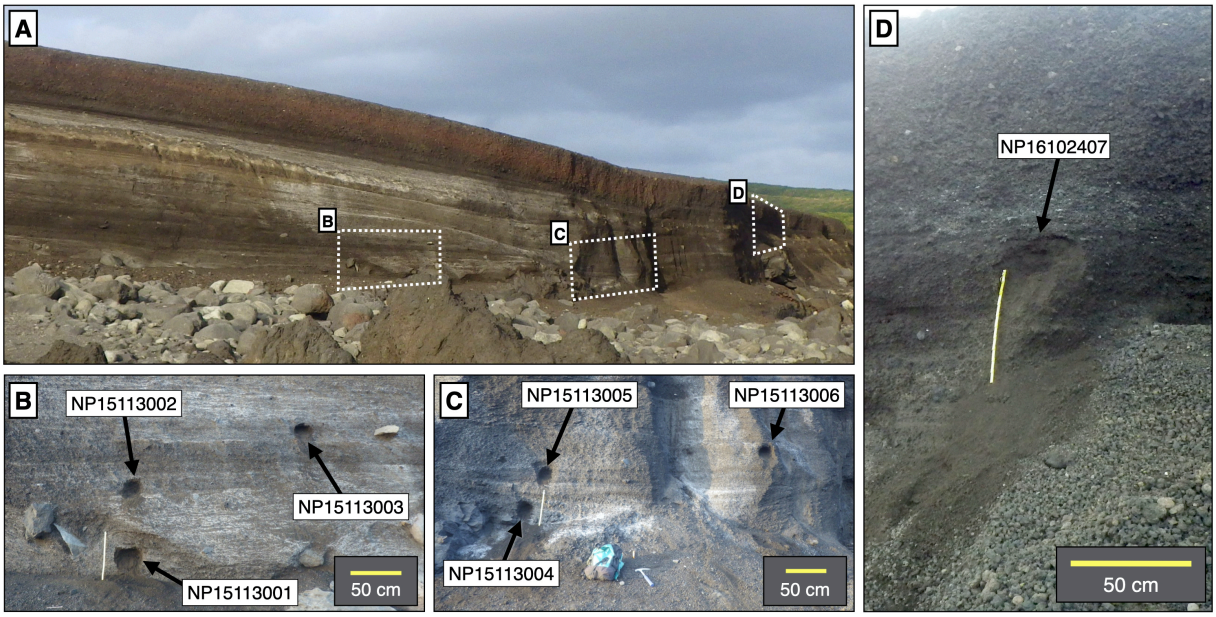


Figure S2. Sampling points at the crosscut outcrop of Nippana tuff ring. A: Overview of the outcrop. B: Lower layers. C: middle layers. D: Upper layers. Note that the uppermost reddish layer is a fallout deposit from a neighboring crater (R crater; e.g., Sumita, 1985).

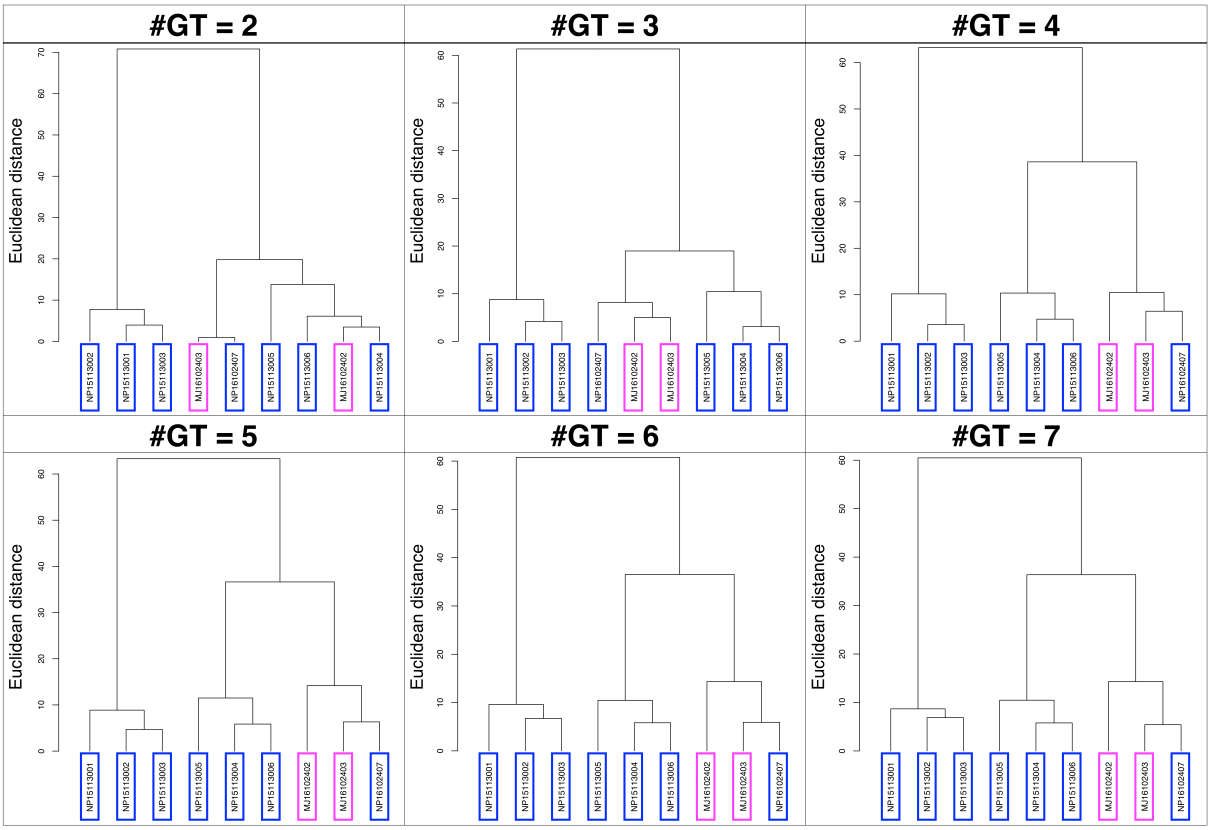


Figure S3. Sample classification dendrograms for grain type numbers 2 through 7. Magenta: dry tephra samples. Blue: wet tephra samples.

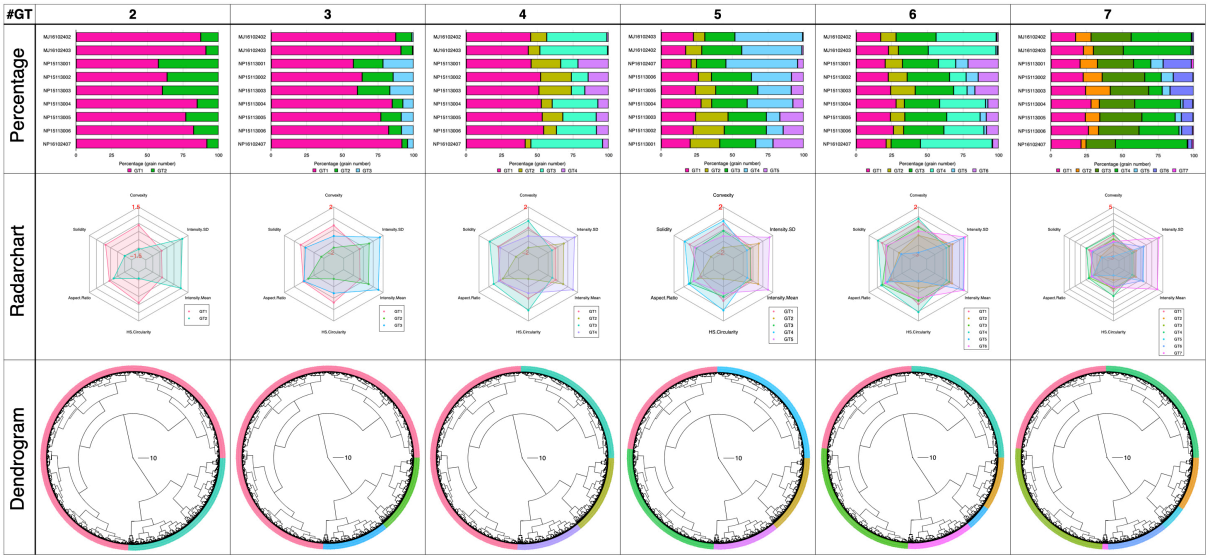


Figure S4. Grain type composition, radargram, and dendrogram for each number of grain types.

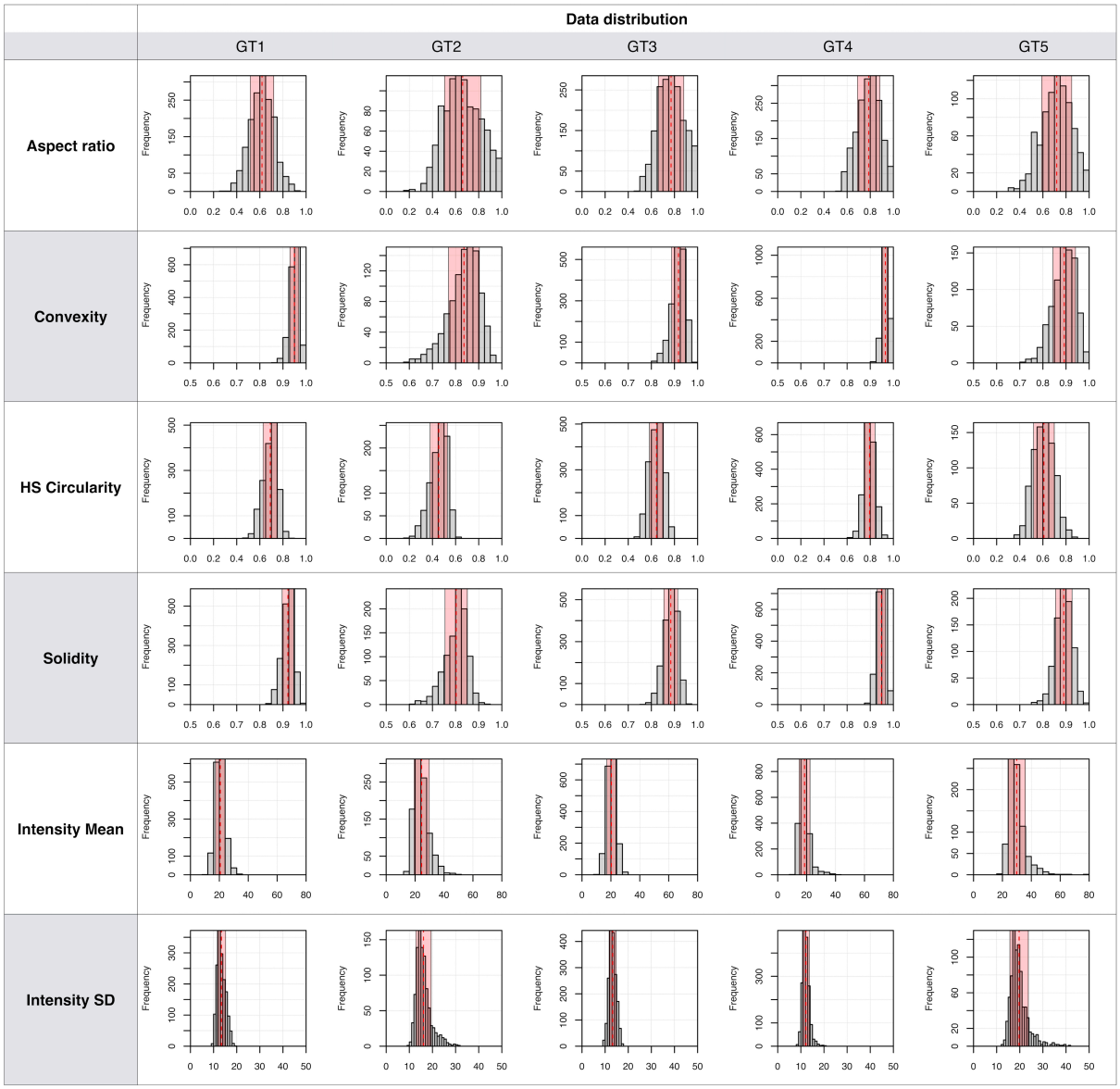


Figure S5. Frequency of parameters in each grain type. The red dashed lines and rectangles indicate the average and standard deviation, respectively.

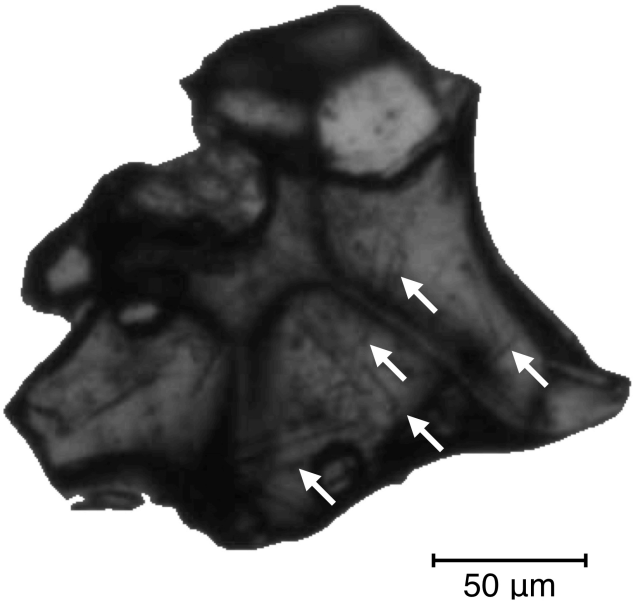


Figure S6. Example of a GT5 grain containing microlites (white arrows).