Bull Volcanol (2007) 70:203–220
DOI 10.1007/s00445-007-0132-7

RESEARCH ARTICLE

Cladistic analysis applied to the classification of volcanoes

D. W. E. Hone · S. H. Mahony · R. S. J. Sparks · K. T. Martin

Received: 25 January 2006 / Accepted: 14 February 2007 / Published online: 24 April 2007

Abstract Cladistics is a systematic method of classification that groups entities on the basis of sharing similar characteristics in the most parsimonious manner. Here cladistics is applied to the classification of volcanoes using a dataset of 59 Quaternary volcanoes and 129 volcanic edifices of the Tohoku region, Northeast Japan. Volcano and edifice characteristics recorded in the database include attributes of volcano size, chemical composition, dominant eruptive products, volcano morphology, dominant landforms, volcano age and eruptive history. Without characteristics related to time the volcanic edifices divide into two groups, with characters related to volcano size, dominant composition and edifice morphology being the most diagnostic. Analysis including time based characteristics yields four groups with a good correlation between these groups and the two groups from the analysis without time for 108 out of 129 volcanic edifices. Thus when characters are slightly changed the volcanoes still form similar groupings. Analysis of the volcanoes both with and without time yields three groups based on compositional, eruptive products and morphological characters. Spatial clusters of volcanic centres have been recognised in the Tohoku region by Tamura et al. (*Earth Planet Sci Lett* 197:105–106, 2002). The groups identified by cladistic analysis are distributed unevenly between the clusters, indicating a tendency for individual clusters to form similar kinds of volcanoes with distinctive but coherent styles of volcanism. Uneven distribution of volcano types between clusters can be explained by variations in dominant magma compositions through time, which are reflected in eruption products and volcanic landforms. Cladistic analysis can be a useful tool for elucidating dynamic igneous processes that could be applied to other regions and globally. Our exploratory study indicates that cladistics has promise as a method for classifying volcanoes and potentially elucidating dynamic and evolutionary volcanic processes. Cladistics may also have utility in hazards assessment where spatial distributions and robust definitions of a volcano are important, as in locating sensitive facilities such as nuclear reactors and repositories.

Editorial responsibility: MA Clynne

D. W. E. Hone · S. H. Mahony · R. S. J. Sparks
Department of Earth Sciences, University of Bristol,
Wills Memorial Building, Queens Road,
Bristol, BS8 1RJ, UK

K. T. Martin
Department of Geology, University of South Florida,
4202 E. Fowler Ave,
Tampa, FL 33620, USA

D. W. E. Hone (✉)
Bayerische Staatssammlung für Paläontologie und Geologie,
Richard-Wagner-Straße 10,
80333 München, Germany
e-mail: d.hone@lrz.uni-muenchen.de

Keywords Volcano classification · Cladistics · Japan · Volcano clusters

Introduction

Classification schemes for volcanoes have largely been based on either morphological characteristics or style of eruption. Factors such as size, tectonic setting and dominant chemical composition of the volcanic products have also been included, usually informally and non-systematically or by relating such factors qualitatively to landform or eruptive style through volcanic processes. Classification can be described as the systematic arrangement of the representatives of a group, such as volcanoes, based on natural relationships (Bullard 1962). Classification schemes allow one word or a few words to be used for a complex object or concept. The different groups should provide universally consistent definitions and may have significance in terms of volcano behaviour and evolution. Classifications are more likely to have lasting value if based on observations rather than theories of volcanic processes that might not endure.

Classifications of volcanoes in common usage have a long history and are mostly based around terms related either to landforms or to characteristic styles of eruption. In a survey of relatively modern textbooks (e.g. Cotton 1944; Bullard 1962; Rittmann 1962; Macdonald 1972; Williams and McBirney 1979; Francis and Oppenheimer 2004; Schmincke 2004) we found that terms, such as caldera, stratocone, dome, shield volcano and cinder cone, are introduced and described without reference to etymology. Many terms originated in the 18th and 19th centuries; for example in seminal studies by George Poulett Scrope and Charles Lyle of volcanic landforms of the Auvergne region (Scrope 1872). Rittmann (1962) provides an example of a classification based on simple landforms, such as central volcanoes, fissure volcanoes, lava volcanoes and mixed volcanoes. His schemes include relationships of volcanic landforms to physical and chemical properties of the magmas. In contrast, classification into monogenetic and polygenetic volcanoes has a ‘process’ definition. In general there is a lack of standardised nomenclature and different textbooks and papers have used different terms for the same features (Simkin and Siebert 1994). The most recent attempt to classify volcanoes on the basis of morphology is in the Smithsonian “Volcanoes of the World” (Simkin and Siebert 1994). They recognised seven main types of volcano and provide a comprehensive and systematic account of attributes of volcanoes that might be used in classification.

Volcanic eruptions are classified after particular volcanoes, where the behaviour was first witnessed and described, e.g. Hawaiian, Strombolian, Vulcanian, and Peleean eruptions. Such classifications can be traced back to the work of Stoppanni, Mercalli, Lacroix and Sapper in the early twentieth century (Bullard 1962). Plinian eruptions are an exception, being named after Pliny the Elder, who died in the AD 79 eruption of Vesuvius. Although these terms have descriptive value, this classification scheme is not appropriate to describe polygenetic volcanoes that typically display several different types of activity even in a single eruption. Eruption style on its own is thus not a very good way of classifying volcanoes as entities.

As many volcanoes are investigated in detail, complex evolutionary histories are revealed; some volcanoes have intermediate characteristics between end member types, and others evolve from one type to another. An end member view is that there is a continuum between all volcano types so that definitions are arbitrary and result in spurious divisions. However, the pervasive use of volcano types in contemporary volcanological literature indicates that classification has value and further implies that volcanic systems may indeed fall into distinctive groups, which reflect the underlying dynamics of magmatism. Interest in classification is being rejuvenated by investigations of the spatial and temporal distributions of volcanoes. This interest arises partly because such distributions are likely to be related to fundamental processes of magma generation and geodynamical controls (e.g. Marsh 1981; Bremond d’Ars et al. 1995; Condit and Connor 1996; Martin et al. 2004). There is also an imperative to make sure that critical structures, such as underground nuclear waste storage facilities, are located in places well away from future volcanic activity; hence the interest in understanding the statistical distributions of volcanic vents and their underlying controls. Studies of volcano distribution require systematic and robust definition of what constitutes a volcano in developing a database for analysis. Different definitions may result in different outcomes for statistical modelling. If volcanoes in fact divide naturally into several groups, or alternatively form a continuum, then these attributes need to be clarified.

This paper explores the application of cladistics as a way to classify volcanoes. Cladistics, also known as parsimony analysis, is a technique widely applied in biology to classify organisms and identify evolutionary trends. However, in principle the method can be applied to any type of information where it suspected that there are underlying relationships and groupings. For example cladistics has also been applied to linguistics to identify relationships and groupings in languages (Kitching et al. 1998). Here we apply cladistic methods to a dataset of Quaternary volcanoes from the Tohoku region of NE Japan.

Cladistics

Cladisitic (or systematic) methods were first developed by Willi Hennig in 1950, but were not applied to modern biology (the first scientific application of cladistics) until translation of Hennig’s work from German into English in 1966. The objective was to elucidate Darwin’s ideas about evolution (Darwin 1859; Hennig 1966) by providing a systematic method of arranging species into an inclusive hierarchical structure (e.g. species in a genus and genera in a family etc.). The term ‘cladistics’ derives from the word ‘clade’ which is a general term for any group of organisms at any level which share an evolutionary history (e.g. a clade of species which form a genus). Kitching et al. (1998) provide a comprehensive review. While cladistics was designed to identify evolutionary relationships, its fundamental process is to group together those individuals that have the most in common over a given set of data. This is done in a manner such that these groups are based on the overall similarity of the individual units (e.g. volcanoes), and not just one or a few of the more obvious characteristics. It can therefore be applied to any system where the observer wishes to determine which units are the most similar. There is no intrinsic requirement for the relationships between the identified clades to be evolutionary in character (Brower 2000) and the method has also been applied, for example, to linguistics where relationships between languages are more complex than simple divergence from an ancestral language. In the case of volcanoes, the method might identify groupings of volcanoes that have had similar histories and dynamic characteristics. Cladistics requires as many good quality, nonrelated characters as possible; this then diminishes human bias and removes homoplasy (i.e. characters which do not reflect the processes that link the clades). The survey design is therefore crucial to produce reliable results. The general principle is that the more characteristics that can be defined for analysis, the more robust the clades are likely to be.

There are parallels between cladistics, cluster analysis and multivariate statistics (Davis 2002). In cluster analysis a feature or features, such as distance from one volcano to the next, are used to recognise groups such as spatial clusters (e.g. Connor and Hill 1995). Multivariate statistics can recognise groups of parameters that are correlated and therefore define groups. Cluster analysis can be used to build hierarchical arrangements of objects that are based on measures of similarity or distance derived from a set of characteristics. Dendrograms are constructed from the similarity measures and so there are parallels with cladistics. Cluster analysis is, however, most suited to characters that are continuous variables that can be measured. For example, correlation coefficients between variables are commonly used as the distance measure in constructing dendrograms (Davis 2002). Many of the characters of a volcano database, however, are not measurements of continuous variables. Some characters are discrete categories (e.g. landforms) or presented as course bins based on qualitative attributes (e.g. measures of rock type and abundance). Cladistics has the advantage that it can deal with the discretised data of qualitative descriptive characteristics that are typical of volcano databases. In the case of volcanoes, the method might identify groupings of volcanoes that have had similar histories and dynamic characteristics.

Cladistic and systematic methods are based around the 'Principle of Parsimony' (Kitching et al. 1998). This assumes that the simplest arrangement by which individuals can be related (a variation of Occam's razor) should be chosen. The simplest arrangement is not necessarily the truth, but it is the only philosophically valid method available (Kitching et al. 1998). In contrast, the traditional 'phenetic' methods in biology relied on forming groups based on selected observations and may not be parsimonious. Cladistics is preferable to phenetic methods, which rely on a small number of prominent observations or overall form. In contrast, cladistic analysis uses a large dataset of characteristics with no assumption being made about their relative significance. To use a classic biological example, elephants, rhinoceroses and hippopotami were phenetically grouped into the Pachyderms as they were all large, grazing African mammals with thick skins. However, by looking at the detailed structure of the skeletons and their development and comparing it to other mammals, it became obvious, by cladistic analysis (Fig. 1), that the rhinoceroses are related to horses, the hippopotami to antelope and deer, and the elephants, to the rodent-like hyraxes and the manatees.

Cladistics is performed by assembling data on the chosen system and then converting these data into a useable format. Modern systematic computer programmes use only discrete data. Thus a continuous variable must be broken

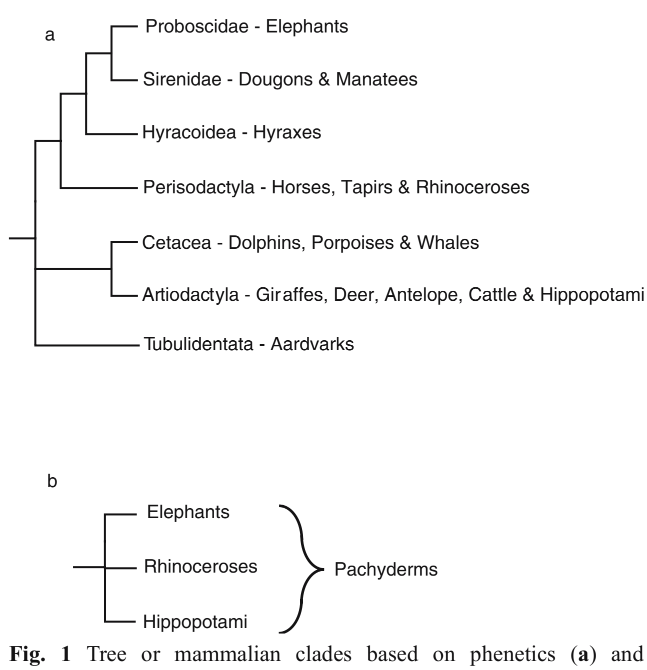

Fig. 1 Tree or mammalian clades based on phenetics (**a**) and cladistics (**b**). Effectively those pairs of units at the tips of the tree are the most similar to each other and successive units or sets of units (clades) are the next most similar. Where there is insufficient data to separate pairs, a polytomy is left of three or more units. Not only does cladistics provide a more accurate representation of the true relationships of these clades, but also provides an increased degree of resolution by removing polytomies for bifurcations. For simplicity, not all clades are shown

down into discrete bins each of which is assigned an integer: these are frequently binary. Computer algorithms are then used to assemble a tree or cladogram (Fig. 2), which identifies the most similar individuals and then successively adds the next most similar individuals until the tree is complete. Here an individual entity (e.g. a volcano, an edifice or a species) is ascribed the term 'unit'. The tree is assembled so that those units that share the most characteristics in common are successively paired together to form 'branches' (which can be of any number of units). Branches are then grouped together that share many characteristics. Groups (clades) are identified where the analysis reveals large differences—either a fundamental shift in the nature of the units (e.g. reptiles and amphibians are separated by a whole suite of characters—their eggs, skin, skeletons etc.), or where there is an obvious split between two branches (e.g. on Fig. 2, the difference between groups 1 and 2). Groups are hierarchical: i.e. the units (Fig. 2) G-M and N-P could be considered separate smaller groups within group 2.

A polytomy is formed where it is not possible to separate two or more units, branches or groups. Thus the relationships between the groups in the polytomy are left unresolved; i.e. it is not possible to establish which units are more or less similar to one another (on Fig. 2, nodes I and II are polytomies). The identification of boundaries between groups is the most subjective aspect of the procedure. The assembly of trees can be done in many different ways and branches are swapped to find the most parsimonious tree or trees. Commonly, many 'most parsimonious trees' are found and these can be combined into consensus trees to show the most significant patterns present.

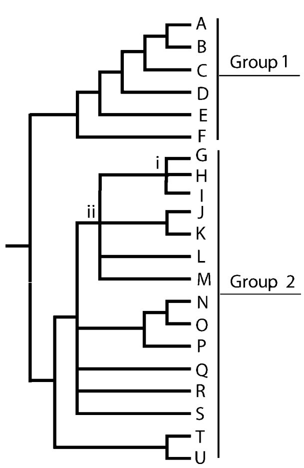

Fig. 2 Example of a cladistic tree, showing bifurcating branches, nodes, polytomies, and groups. Group 1 shows a 'Hennigian comb' of successive taxa (*A–F*) suggesting a good definition of the associations between the units to produce this branching pattern. Group 2 (*G–U*) are characterised by their large number of polytomies (e.g. *i* and *ii*—exact relationships cannot be determined so there are multiple branches from individual nodes). Whilst the inter-relationships are poorly defined in 2, the group as a whole is well separated from 1 as there is clearly a split between the groups

Dataset of Japanese volcanoes

The volcanoes studied are from the Quaternary Period of the Tohoku region of Japan. The islands of Honshu and southwestern Hokkaido in NE Japan lie on the Tohoku volcanic arc, where the Pacific plate converges with Japan (Fig. 3). Data on the volcanoes were taken from the Catalogue of Quaternary Volcanoes in Japan (Committee for Catalog of Quaternary Volcanoes in Japan 1999) produced by the Volcanological Society of Japan. The Catalogue lists volcanoes and volcanic edifices as different categories. We generated a derivative database for the cladistic analysis and made some changes and removed slight inconsistencies as follows. We recognise that the database is partly presented in Japanese and the authors of this paper may have overlooked some of the details.

A volcano comprises one or more edifices. The Committee consisted of over 50 scientists who were volcanological experts in particular areas and groups of volcanoes in Japan. The categorisation of volcanic features into volcanoes and edifices belonging to a volcano was based on interpretation; in particular a volcano was defined as a single coherent magmatic system (Sintaro Hayashi and Koji Umeda, personal communication 2006) with edifices being part of that system. In Japan, magmatic systems are commonly interpreted as the consequence of the rise of a mantle diapir. This concept was used by the Committee to identify volcanoes and associated edifices. Local experts on particular volcanoes or volcanic groups made their interpretations independently (Koji Umeda, personal communication 2006), a process which potentially introduces subjectivity and inconsistency. Each volcano or volcanic edifice is depicted as a dot on the map that accompanies the database. The location of the dot is fixed at a representative location. For 52 of the 59 volcanoes, the location is at the edifice with the highest altitude. In one case the location is at a central vent. One volcano is not located on any of its edifice locations, nor is it an average of its edifice location. In five cases there are several edifices all at the same location. For a few volcanoes the edifice chosen for the location of the volcano is either the oldest, or the youngest or the tallest. A few designations are inconsistent or enigmatic. For example Nasu volcano consists of 6 edifices aligned north to south. The only historically active volcano is Chausudake volcano, but this is designated as an edifice, while the location of the volcano is given as the Sandon–Yaridake edifice, which has no record of historic activity. These differences and inconsistencies have been recognised

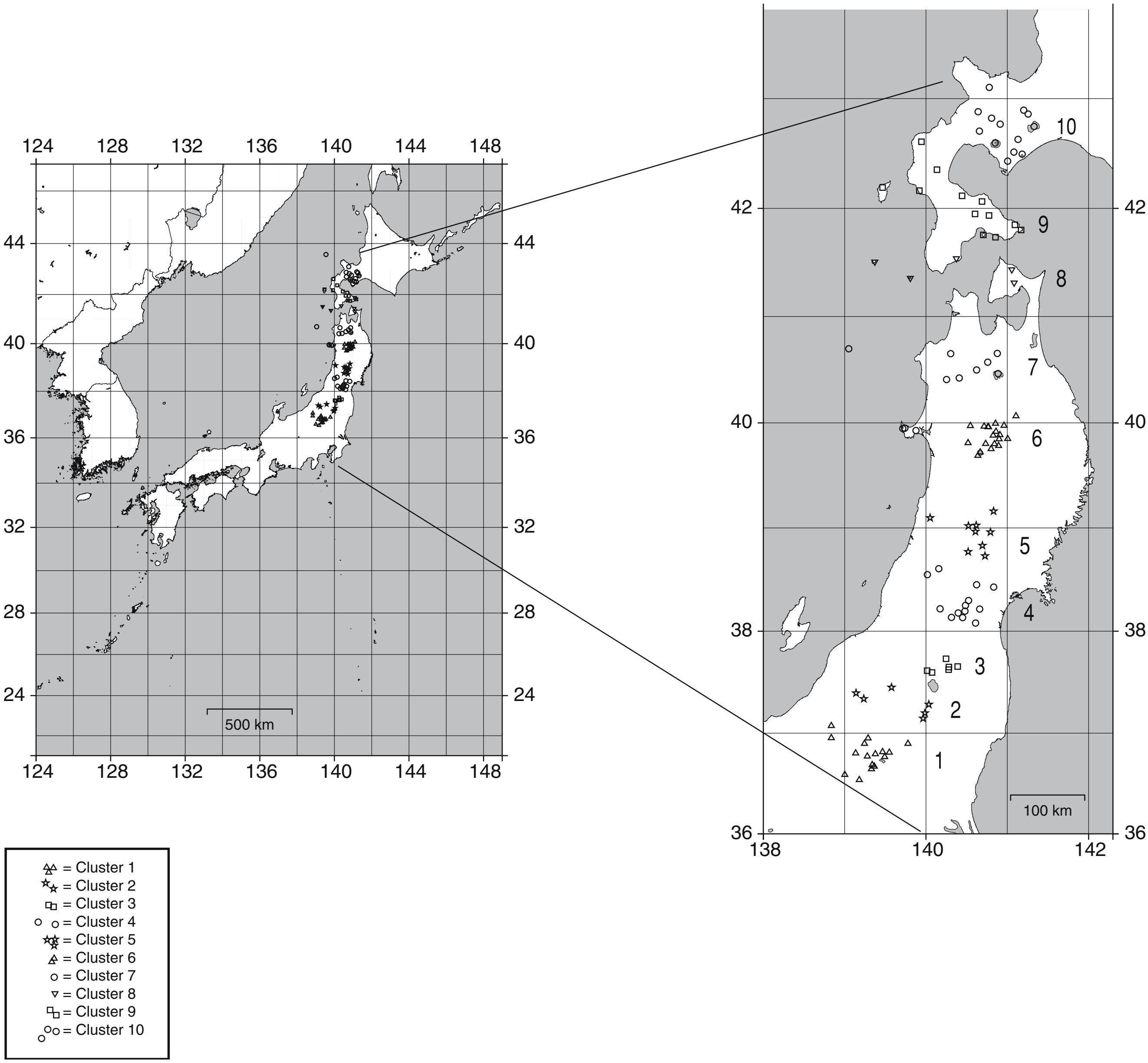

Fig. 3 Map of Japan and outlying islands showing the locations of the volcanoes used in this study. The 10 volcanic 'clusters' identified by Tamura et al. (2002) are separated by tone and shape and are numbered *1–10*

only in 7 of the 59 cases. So the database is largely systematic and consistent.

Volcanic edifices are divided into several different "types" of volcanic landforms or features. Several edifice types are identified in the database: steep and gently sided stratovolcanoes, lava cones, lava flows, lava domes, maar deposits, collapse caldera, pyroclastic flow plateaus, volcanic necks, and alluvial fans consisting of debris and mud flow deposits. In contrast, the volcanoes in the Catalogue are not subdivided into different types; they are simply classed as volcanoes. Here we define an edifice as 'a place where magma reached the surface'. However, not all the edifice types in the database are places were magma has reached the Earth's surface (e.g. debris avalanches). Edifices were removed from our derivative database that did not fulfill this definition. In the case of lava flows, the source vent represents the location and is included in our analyses. Censoring only affected a few data, namely one volcano (which comprised only one edifice) and eight edifices from different volcanoes.

Age data are used to define characters of youthfulness, longevity and maturity of a volcano and also whether particular groupings of volcanoes, identified by cladistic methods, can be linked to specific time periods or evolutionary processes. Tamura et al. (2002) divided the volcanoes of the Tohoku region into ten 'clusters' (Fig. 3) and interpreted these clusters in terms of hot fingers of the mantle rising in the mantle wedge with preferred regions of melt generation. We were interested to find out whether spatial and temporal relationships of volcano groupings would emerge from cladistic analysis.

The following additional definitions were used in forming a derivative database from the Catalogue. The volcanoes and volcanic edifices were categorised as monogenetic (VM) or polygenetic (VP). Thirty-four of the edifices are monogenetic, 66 are polygenetic and for 29 there is insufficient information to determine either way. Forty-four volcanoes as whole entities are polygenetic, seven are monogenetic and there is insufficient information about eight of them to determine either way. Volcanic edifices were divided into: 80 edifices that have a record of historic eruptions but no Holocene record (A1); 14 edifices with a record of Holocene activity but no historic eruptions (A2) and 30 edifices that have evidence of Holocene and historic activity (A3). Five edifices were not classified in this way due to lack of information. This aspect of the database clearly reflects variation in the detail of investigation and might change as more information becomes available.

The derivative dataset comprises a variety of volcanoes and volcanic edifices with large variations in size, age, style of activity, magma composition and eruption type, which constitute the characteristics used in the cladistic analysis. Here we use both the volcanoes and the volcanic edifices as alternative basic units for the cladistic analysis. We analysed a subset of data from 59 volcanoes and 129 associated volcanic edifices from clusters 4 to 8 (Fig. 3). This subset was chosen to define a relatively small region appropriate for an exploratory study focused on a well studied dataset.

Methods

Data on the volcanic edifices were divided into 34 characters (Table 1) and 24 characters for volcanoes (Table 2) based on the morphological features, size, style of activity, age and rock composition. Characters should be independent of each other to avoid certain features being weighted more strongly than others (through pseudo-replication of data). If characters are weighted differently, this might lead to biased groupings. For example, if there were three characters all based on the oldest age of volcanism, age of the volcanoes would dominate the results, so the groups would tend to be split according to age. Thus, as far as is practical, the characters were chosen to be independent. However, it is hard to demonstrate rigorously that characters are truly independent since apparently unrelated parameters might be linked through some unknown common physical mechanism. All that can be done is avoid parameters that are essentially synonymous.

Each character is given an integer number. Each character has a number of character states (subdivided into discrete bins) assigned to it which describe the variability of that character. The data from each volcano or volcanic edifice were then assigned to the relevant character state. In the methodology a character for an individual unit can only be assigned to one state. The nature of each character and character state was decided by looking at the available data and are illustrated here for the case of volcanic edifices. Age data were used to define the following characteristics: the oldest recorded volcanic edifice age (character 1), age of the youngest activity (character 14), what kind of activity the edifice or volcano displays (i.e. Holocene only, historical activity only or a combination of both) (character 34), and the oldest age of an edifice (character 25). Age data were also divided into discrete time windows (characters 26–31). Character 32 is whether the unit is identified as monogenetic or polygenetic. The chemical composition, characters related to morphology (type of landform), and edifice type are represented by characters 2–11 and 16–24 respectively. Some of the characters listed in the database as ‘material produced’ are similar to some of the ‘edifice type’ fields. However, both types of data can be validly used as characters without data replication because the characters are different for these two categories of data. Character 33 is the distance of the edifice from the volcano. The size of the volcano or edifice is given by volume (character 12), edifice volume as a percent of the volcano volume (character 15) and volcano area (character 13).

Each character is coded which splits a characteristic into a small number of discrete bins, each of which is given a code. Assigning values to the individual character states can be subjective. The volume and area characters were split using a logarithmic scale to give order of magnitude measures of volcano or edifice size. The percentage of a certain compositional group of lavas, is split, so those that contain 0–20% are ranked ‘0’, 20–50% ranked ‘1’ etc. to break continuous variables into simple integers.

The assembled data-matrix was run using software called PAUP* (Phylogenetic Analysis Using Parsimony-ver 4.0 Swofford 1998). The PAUP* code is widely used by the biological community for systematic analysis. For example, the classification of life into the ‘tree of life’. PAUP*’s basic tree assembly method is ‘TBR’ (Tree Branching and Reconnection) and works by creating random trees and mapping the characters to give a score for each tree (i.e. the number of character changes, or steps, required to fit the data to that tree). A tree is one way of arranging the taxa into a classified order or several groups, with the most similar taxa positioned most closely together.

Table 1 List of characters and the coding definition used to form the raw data for the analysis of edifices

Characteristics and coding definitions
1. *Volcano age (oldest date). States: 0=0–0.25 Ma; 1=0.25–0.5 Ma; 2=0.5–1 Ma; 3=1–1.5 Ma; 4=1.5–2.0 Ma; 5=>2.0 Ma. Comments: This is the oldest recorded age.
2. *Compositional type—basalt.* States: 0=None; 1=Very small; 2=Small; 3=Medium; 4=Large; 5=Substantial.
3. *Compositional type—basaltic andesite.* States: 0=None; 1=Very small; 2=Small; 3=Medium; 4=Large; 5=Substantial.
4. *Compositional type—andesite.* States: 0=None; 1=Very small; 2=Small; 3=Medium; 4=Large; 5=Substantial.
5. *Compositional type—dacite.* States: 0=None; 1=Very small; 2=Small; 3=Medium; 4=Large; 5=Substantial.
6. *Compositional type—rhyolite.* States: 0=None; 1=Very small; 2=Small; 3=Medium; 4=Large; 5=Substantial.
7. *Material produced—lava flows.* States: 0=absent, 1=present, ? = unknown, no data recorded.
8. *Material produced—pyroclastic rocks.* States: 0=absent, 1=present, ? = unknown, no data recorded.
9. *Material produced—lava domes.* States: 0=absent, 1=present, ? = unknown, no data recorded.
10. *Material produced—scoria.* States: 0=absent, 1=present, ? = unknown, no data recorded.
11. *Material produced—welded spatter.* States: 0=absent, 1=present, ? = unknown, no data recorded.
12. *Volcano volume.* States: 0=<0.1 km^3; 1=0.1–1 km^3; 2=1–<10 km^3; 3=10–<100 km^3; 4=>100 km^3.
13. *Area of volcano.* States: 0=0–1 km^2; 1=1–10 km^2; 2=10–100 km^2; 3=100–1,000 km^2.
14. *Volcano age (youngest date). States: 0=0–0.25 Ma; 1=0.25–0.5 Ma; 2=0.5–1 Ma; 3=1–1.5 Ma; 4=1.5–2.0 Ma; 5=>2.0 Ma. Comments: This is youngest recorded age.
15. *Volume of edifice/volume of volcano (%).* States: 0=0–20%; 1=20–40%; 2=40–60%; 3=60–80%; 4=80–100%. Comments: This is the volume of the edifice as a fraction of the volume of the entire volcano.
16. *Stratovolcano with gentle lava slope.* States: 0=feature recorded as absent; 1=feature recorded as present; ? = feature not recorded as present or absent.
17. *Stratovolcano with steep slope and conical form.* States: 0=feature recorded as absent; 1=feature recorded as present; ? = feature not recorded as present or absent.
18. *Collapse caldera.* States: 0=feature recorded as absent; 1=feature recorded as present; ? = feature not recorded as present or absent.
19. *Pyroclastic cone.* States: 0=feature recorded as absent; 1=feature recorded as present; ? = feature not recorded as present or absent.
20. *Lava cone with gentle slope and conical form (mostly made of lava flows.* States: 0=feature recorded as absent; 1=feature recorded as present; ? = feature not recorded as present or absent.
21. *Lava dome.* States: 0=feature recorded as absent; 1=feature recorded as present; ? = feature not recorded as present or absent.
22. *Debris avalanche.* States: 0=feature recorded as absent; 1=feature recorded as present; ? = feature not recorded as present or absent.
23. *Pyroclastic flow plateau* States: 0=feature recorded as absent; 1=feature recorded as present; ? = feature not recorded as present or absent.
24. *Lava flow, lava plateau.* States: 0=feature recorded as absent; 1=feature recorded as present; ? = feature not recorded as present or absent.
25. *Edifice age (oldest). States: 0=0–0.25 Ma; 1=0.25–0.5 Ma; 2=0.5–1 Ma; 3=1–1.5 Ma; 4=1.5–2.0 Ma; 5=>2.0 Ma.
26. *Number of periods of volcanic activity within the time period 0–0.25 Ma. States: 0=0 ages; 1=1 age; 2=2 ages; 3=3, 4, 5 or 6 ages; 4=7 or more ages.
27. *Number of periods of volcanic activity within the time period 0.25–0.5 Ma. States: 0=0 ages; 1=1 age; 2=2 ages; 3=3, 4, 5 or 6 ages; 4=7 or more ages.
28. *Number of periods of volcanic activity within the time period 0.5–1.0 Ma. States: 0=0 ages; 1=1 age; 2=2 ages; 3=3, 4, 5 or 6 ages; 4=7 or more ages.
29. *Number of periods of volcanic activity within the time period 1.0–1.5 Ma. States: 0=0 ages; 1=1 age; 2=2 ages; 3=3, 4, 5 or 6 ages; 4=7 or more ages.
30. *Number of periods of volcanic activity within the time period 1.5–2.0 Ma. States: 0=0 ages; 1=1 age; 2=2 ages; 3=3, 4, 5 or 6 ages; 4=7 or more ages.
31. *Number of periods of volcanic activity within the time period >2.0 Ma. States: 0=0 ages; 1=1 age; 2=2 ages; 3=3, 4, 5 or 6 ages; 4=7 or more ages.
Additional characters used in analyses 3 & 4
32. *Monogenetic vs. polygenetic. States: 0=Monogenetic, 1=Polygenetic.
33. *Distance of edifice from volcano*: 0=0–0.1 km, 1=0.1–1.0 km, 2=1–10 km, 3=>10 km Comments: Rounding down where necessary i.e. if a value has 0.1 it goes in the 0–0.1 category.
34. *Activity states: 0=Record of historic eruptions but no Holocene eruptions, 1=Record of Holocene but no historic eruptions, 2=Holocene and historic eruptions.

*Denotes removed from dataset for analysis without time characters

For each tree, the code cuts off two branches (subsets within the specified tree) and swaps them and scores again. The tree with the lowest score is accepted (as it is more parsimonious) and the other discarded. The algorithm continues until it ceases finding major reductions in the score, i.e. it only finds less parsimonious trees. The code

Table 2 List of characters and the coding definition used to form the raw data for the analysis of volcanoes

Characteristics and coding definitions
Volcano age—oldest. States: 0=0–0.25 Ma; 1=0.25–0.5 Ma; 2=0.5–1 Ma; 3=1–1.5 Ma; 4=1.5–2.0 Ma; 5=>2.0 Ma. Comments: This is the oldest recorded age.
Volcano age— youngest. States: 0=0–0.25 Ma; 1=0.25–0.5 Ma; 2=0.5–1 Ma; 3=1–1.5 Ma; 4=1.5–2.0 Ma; 5=>2.0 Ma. Comments: This is youngest recorded age.
Volcano volume. States: 0=<0.1 km^3; 1=0.1–1 km^3; 2=1–<10 km^3; 3=10–<100 km^3; 4=>100 km^3.
Volcano area. States: 0=0–1 km^2; 1=1–10 km^2; 2=10–100 km^2; 3=100–1,000 km^2.
Monogenetic/polygenetic. States: 0=Monogenetic, 1=Polygenetic.
Activity. States: 0=Record of historic eruptions but no Holocene, 1=Record of Holocene but no historic eruptions, 2=Holocene and historic eruptions.
Edifice type: stratocone with gentle lava slope. States: 0=absent, 1=present.
Edifice type: stratovolcano with steep slope and conical form States: 0=absent, 1=present.
Edifice type: collapse caldera. States: 0=absent, 1=present.
Edifice type: pyroclastic cone. States: 0=absent, 1=present.
Edifice type: lava cone with gentle slope and conical form. States: 0=absent, 1=present.
Edifice type: lava dome. States: 0=absent, 1=present.
Edifice type: debris avalanche. States: 0=absent, 1=present.
Edifice type: pyroclastic flow plateau. States: 0=absent, 1=present.
Edifice type: lava flow, lava plateau. States: 0=absent, 1=present.
Material produced—lava flows. States: 0=absent, 1=present, ? = unknown, no data recorded.
Material produced—pyroclastic rocks. States: 0=absent, 1=present, ? = unknown, no data recorded.
Material produced—lava domes. States: 0=absent, 1=present, ? = unknown, no data recorded.
Material produced—scoria. States: 0=absent, 1=present, ? = unknown, no data recorded.
Material produced—welded spatter. States: 0=absent, 1=present, ? = unknown, no data recorded.
Compositional type—basalt. States: 0=absent; 1=present
Compositional type—basaltic andesite. States: 0=absent; 1=present
Compositional type—andesite. States: 0=absent; 1=present
Compositional type—dacite. States: 0=absent; 1=present
Compositional type—rhyolite. States: 0=absent; 1=present

then swaps individual units within this tree to find slightly lower scores. Only those trees with the smallest identified score are returned and are defined as the Most Parsimonious Trees (MPTs).

We used a heuristic search method and Wagner optimised ordered characters. The Wagner method of ordering characters requires that character state changes have to ‘pass through’ intermediate steps, i.e. to get from 1 to 3, the character must use 2. Therefore, a change from 1 to 3 counts as two steps, not one. The number of steps is important as it is used as a measure of similarity, and so affects the position of the taxon within the generated tree. For example a character state of 2 is more similar to 3 than a character of 1 is to 3, so the taxa from the first example will be more closely associated in the generated tree than in the latter example. However, these characters are freely reversible, 3 can return to 2, 1 or 0 with the requisite number of steps. This is pertinent as the multistate (i.e. not binary) characters can be processed in this way; a volcano that has reached 1,000 km^3 in volume must have previously been 500 km^3 in volume for example. There are alternative methods (see Kitching et al. 1998) that do not have the restrictions that Wagner optimization places on the data. However, for most characteristics of volcanoes the nature of the data suggested that it was the most appropriate.

A heuristic search was used because of the large size of the dataset. Due to the exponential growth of possible trees as more units are added, currently computers can only deal with datasets of 16 or less units. Above this number algorithms have to search randomly among the possible trees available and break and reconnect branches to identify the most parsimonious solution. The number of permutations are vast: for 16 units there are 2.1×10^{14}, at just 63, the total possible trees is greater than 10^{100}, and here the database has 129 units, giving a possible of 6.54×10^{252} permutations. The search procedure is not guaranteed to find the most parsimonious solution, as it cannot identify all possible trees, but does have a high probability of success. If the same solutions are repeatedly recovered by running the heuristic search multiple times then confidence is gained that the most parsimonious solutions available for the dataset have been found. Parsimony acts to minimise the amount of changes required to fit the data to each proposed tree. Thus ‘perfect’ data would fit only one tree ‘perfectly’. Each time a character state has to change (e.g. from a ‘1’ to a ‘0’) on a branch, a single step is added. We are therefore searching for the tree or trees with the fewest steps and where the data best fits the tree, this being the most parsimonious resolution.

Due to memory and run-time constraints we collected 10,000 MPTs (Most Parsimonious Trees), in each run. In principle many more MPTs are likely to exist. The impact of this potential loss can, however, by reduced by repeating the analysis. If each search finds the same length MPTs, and produces a similarly shaped resultant tree then we can have confidence in the results.

Each MPT is typically very similar and so methods are applied to represent them by a consensus tree. Consensus trees show an 'average' tree from all the 10,000 trees generated, which gives the most likely outcome if the data were perfect. The degree of similarity can be assessed by inspecting the percentage of times that particular branches occur in all 10,000 alternative MPTs. A well-resolved solution will have a high proportion of branches at 100%. The MPTs are combined using two different methods into two individual consensus trees to provide single solutions to the high number of retained trees. In the strict–consensus method only those branches and clades (groups of branches/taxa) are shown that are not contradicted on any of the other 10,000 trees; any subtle differences between any of the 10,000 trees will result in a polytomy for those affected branches. In the Majority Rule Consensus (MRC) method, the groups that appear in the consensus tree are those are present in more than 50% of the retained MPTs. This is merely a different manner of showing the results, which allows a greater degree of resolution to be observed in the tree. This also gives a measure of the confidence of the resolution as the percentage retention of branches are given, i.e. in how many of the 10,000 trees each branch appears. Generally, values over 80% are accepted within biological cladistics, and any group with a retention of 100% will appear in the strict consensus tree. Analyses were performed on the data for volcanic edifices with (Analysis 1) and without (Analysis 2) time characteristics, and for volcanoes with (Analysis 3) and without (Analysis 4) time. Time-based characteristics are identified In Tables 1 and 2 by an asterisk (*).

An additional analysis included a line of data representing the Soufrière Hills Volcano, Montserrat to see if cladistic analysis would place Montserrat in an appropriate group. This study was used to explore whether a global volcano classification system could be envisaged.

Results

For all analyses a total of 10,000 MPTs trees were recovered. This is to be expected, given the relatively low numbers of characters available and the high numbers of units. The number of steps were as follows: Analysis 1 (811 steps), Analysis 2 (441 steps), Analysis 3 (212 steps) and Analysis 4 (134 steps). Analyses 2 and 4 have much fewer steps relative to 1 and 3 respectively as they have fewer characters.

The strict–consensus trees for each analysis were only moderately resolved. Neither the strict nor semi-strict trees are shown as no conclusions could be drawn from them due to the lack of overall resolution. This is a consequence of the large number of MPTs and relatively few characters. A significant number of units lie in different positions on the retained trees, causing branches to collapse. Here, where a unit could lie in one of two positions, it is 'dropped' down

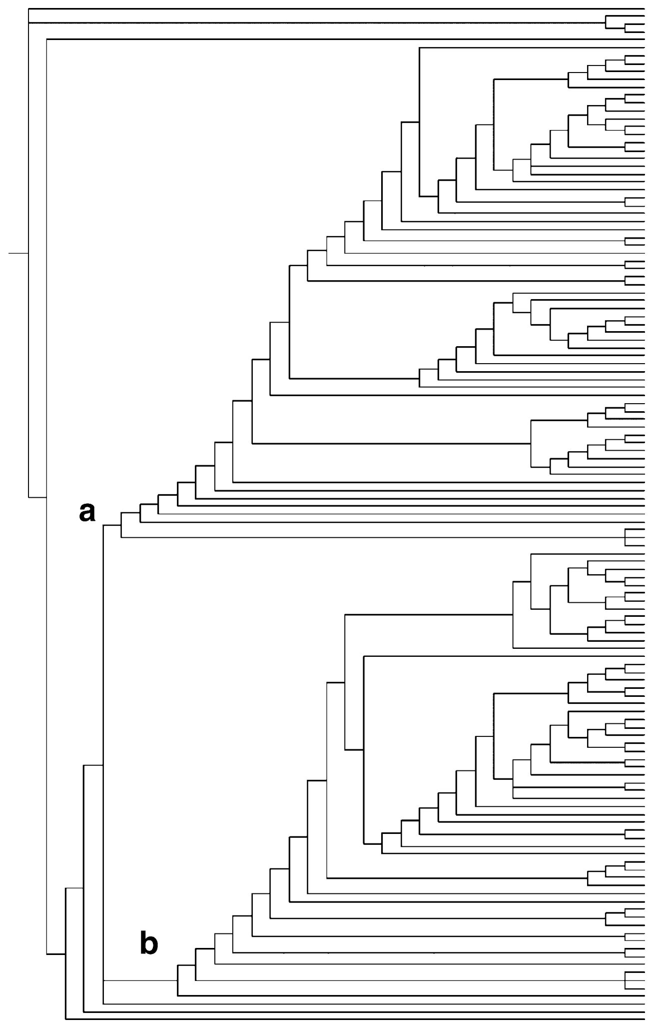

Fig. 4 Majority Rule Consensus (MRC) tree of 10,000 MPTs for the analysis 2. The tree is 'unrooted' and so the branches at the bottom of the tree should not be treated as the base, instead the tree is better visualised as being able to branch in any direction from any node, this is however the easiest graphical representation. The two groups *A* and *B* are labelled at their most basal node (thus every branch above them is in the group). The remainder of the branches are indeterminate units. This structure is then compressed to form Fig. 6(*2*). * *IU* Indeterminate Units

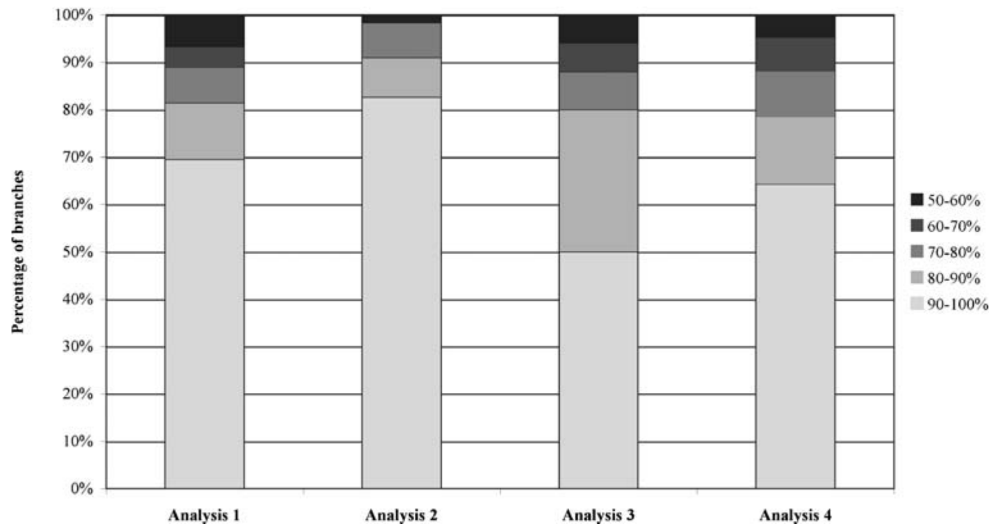

Fig. 5 Bar chart showing levels of retention for the MRC trees (e.g. Fig. 4). In each case they are very high numbers of branches at 100% (i.e. they appeared in all the Most Parsimonious Cladograms). On this evidence, analysis 2 is especially strong, and analysis 3 rather weak

the tree to leave a polytomy at the highest possible position which is in common with both of its two possible positions. However, as this is a strict consensus tree, only a few units need to be unstable in order to collapse large sections of the tree, since one unit, which has a high variability could appear in many places in different trees.

The MRC trees are almost completely resolved (Fig. 4 shows the MRC tree for Analysis 2 as an example), with few polytomies. Generally there are very high percentage numbers on those branches below 100% (Fig. 5). Datasets with large numbers of units and low numbers of characters, as is the case with the volcano data, are commonly characterised by large and frequent polytomies. Both the MRC and strict consensus do not have this problem, suggesting that the analysis is robust. Furthermore, unlike a biological investigation, we are less interested in the relationship between each individual unit and are primarily concerned with establishing major groups, as a basis for classification. Using a biological analogy, we want to separate the mammals from the birds, and not just determine if rats are more similar to gerbils than mice. Thus, the robustness and support of major groups is emphasised here. High levels of conflict (e.g. large numbers of polytomies) are not of great concern provided they are *within* major clades. We would also expect this as there are insufficient characters to resolve all the individual nodes.

Bootstrap support tests were applied to the data to examine the strength of the tree. The bootstrap randomly resamples characters over many replicates (here 1,000) and reconstructs a consensus tree. Datasets with conflicting characters will produce poorly resolved trees as the random resampling will provide trees based on both extremes of the character coding and the consensus of the two will collapse to a large polytomy. This is the case with this analysis. However, the result is less likely to be due to high levels of character conflict, and more likely to be associated with the limited number of characters from which the bootstrap can resample (just 34). We now describe the main features of each group for the cladistic analysis of the data with and without time characteristics, for edifices and volcanoes.

The MRC tree for each analysis (1–4) has been reduced to show just the major groups and their relationships (Fig. 6). These groups are based on major splits in the tree, and are used in the analysis below. A few units were indeterminate (i.e. they did not fit with any major group) and are represented on the trees as such (Fig. 4).

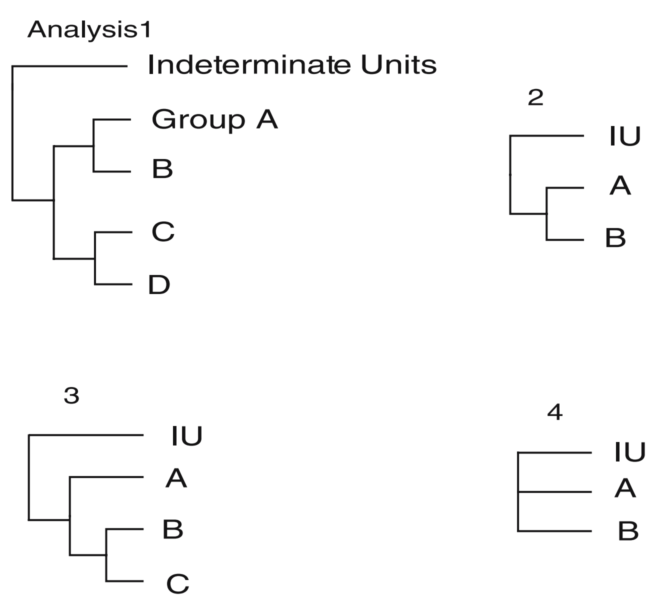

Fig. 6 Reduced trees for analyses *1–4*. As shown in Fig. 4 large trees are divided into their principal groups and then reduced. Major groups are discussed below and their relationships can be examined from this figure, thus for Analysis 3 we would expect the units in *B* and *C* to be more similar to each other than those units in *A*

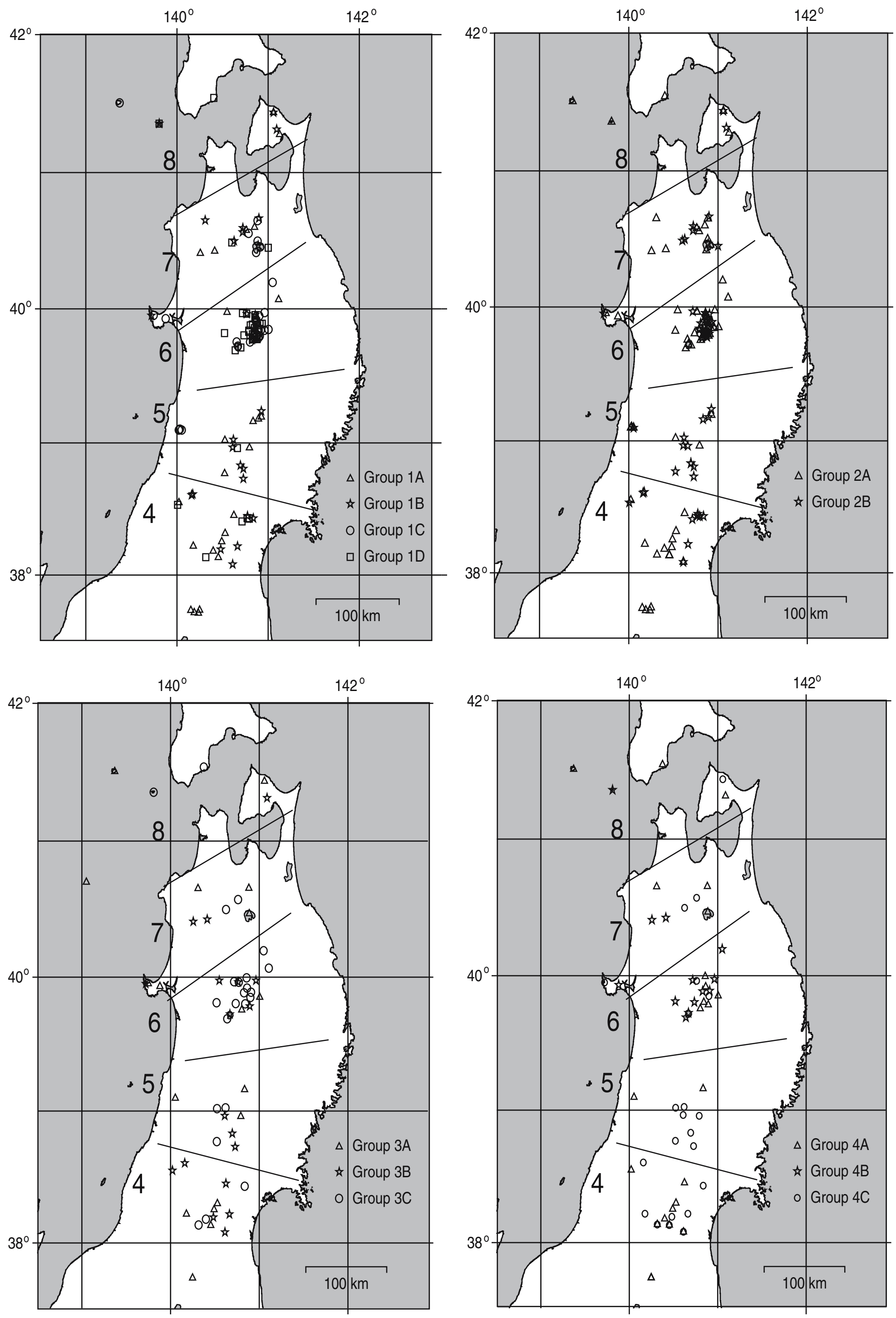

Fig. 7 a–d show the distributions of the groups over the Tohoku region

Analysis

Maps of where the cladistic groups are distributed, in relation to the clusters (Tamura et al. 2002) are illustrated in Fig. 7 a–d. Figure 8 compares the relative composition of each cladistic grouping. Key features of the cladistic groups are given in Table 3 and are now described in more detail below.

Analysis 1 (Edifices, with time)

The analysis yields four groups. Table 4 gives the number of edifices in each group and how they divide between the volcano clusters identified by Tamura et al. (2002). The analysis shows that the different groups are unevenly distributed between the clusters (Fig. 7a). For example cluster 4 largely consists of volcanic edifices from groups A

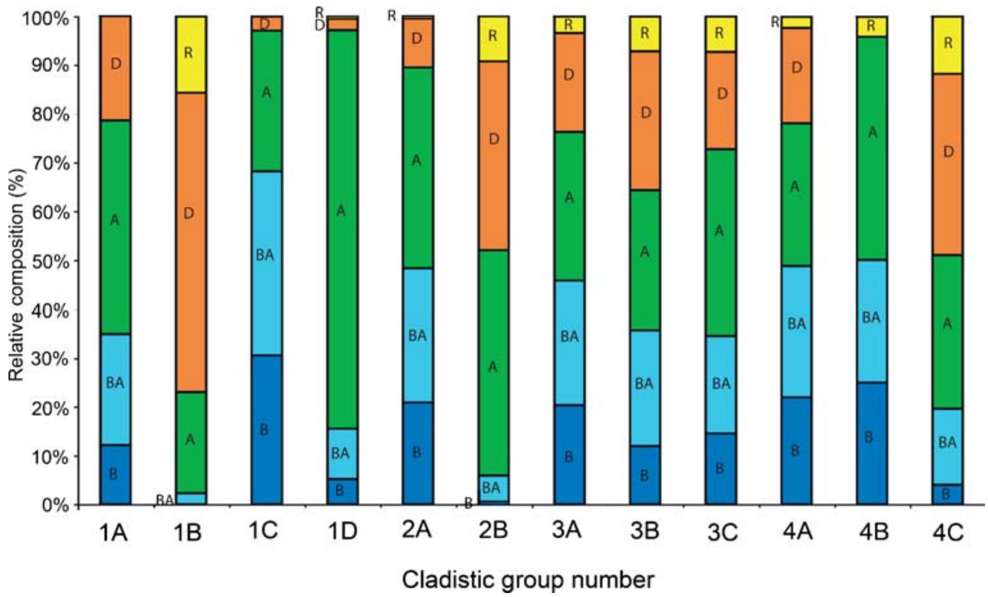

Fig. 8 Comparison of the relative composition percentages for each cladistic grouping

and B, and group D edifices largely occur in cluster 6. These results suggest that edifices with similar characteristics are clustered in space and time. We now describe each group.

Group 1A (26 units) contains 15 steep sided polygenetic stratovolcanoes, and 11 gently sided stratocones. Seventeen of the edifices have dates in the 0.5–1 Ma range, with the most recent volcanism occurring over the range 0.25–1 Ma. Eleven out of the 26 volcanoes show Holocene activity. Andesite is the most abundant composition, also with substantial

Table 3 Table to describe the overall characteristics of the cladistic groupings

Cladistics group	Age	Composition	Features	Stratocone type	Volume	Number in group
1A	0–1.5 Ma	Mixed	Lava flows, Pyroclastic rocks, lava cones	Steep slopes	Varied	26
1B	Holocene with varied ‘oldest’ ages	Predominantly dacite	Pyroclastic flow plateaus, lava domes, calderas	Steep and gentle slopes	0.1–10 km^3	35
1C	0–0.5 Ma	Basalt to andesite	Dominantly lava flows and some pyroclastic rocks	Steep and gentle slopes	>100 km^3	28
1D	Varied	Andesite	Pyroclastic rocks and lava domes	Gentle slopes	Varied	40
2A	Analysis does not involve ‘age’ factors	Basalt to andesite	Calderas, lava domes, lava flows, pyroclastic rocks	Gentle and steep slopes	Varied	65
2B	Analysis does not involve ‘age’ factors	Dacite to rhyolite	Pyroclastic flow plateaus and lava domes	Predominantly gentle slopes	Varied	64
3A	Varied	Mixed	Lava domes and pyroclastic rocks	Steep slopes	10–100 km^3	18
3B	0.5–1 Ma	Dacite to andesite	Calderas, lava domes, pyroclastic flow plateaus	Steep slopes	10–100 km^3 mostly near 10 km^3	18
3C	>2 Ma	Basaltic andesite to dacite	Lava cones, lava domes, pyroclastic flow plateaus	Gentle slopes	1–10 km^3	14
4A	Analysis does not involve ‘age’ factors	Mixed	Lava domes, lava flows, pyroclastic rocks	Steep slopes	10–100 km^3	24
4B	Analysis does not involve ‘age’ factors	Basalt to andesite	Lava cones	Gentle slopes	1–10 km^3	13
4C	Analysis does not involve ‘age’ factors	Andesite to rhyolite	Pyroclastic flow plateau, lava flows, lava domes, calderas	Steep and gentle slopes	10–100 km^3	22

Table 4 Results from edifices with time, analysis 1

Cluster	Group				
	A	B	C	D	Total edifices per cluster
4	12	8	0	5	25
5	6	10	6	1	23
6	2	4	13	30	49
7	4	11	8	2	25
8	2	2	1	2	7
Total edifices per group	26	35	28	40	129

Results from cladistic analyses on volcanoes and edifices with and without time

amounts of dacite and basaltic andesite (Fig. 8). The main products are lava flows, pyroclastic flows, and lava domes. 18 of the 26 edifices have volumes in the range 10–100 km^3. The percentage of the total volcano volume made up by edifice volume is high, with 14 edifices containing 80–100% of the total volcano volume.

Group 1B (35 units) contains 13 monogenetic edifices and 18 that are polygenetic, there was insufficient data regarding four of the edifices to tell if they were monogenetic or polygenetic. There is a large range of ages with 4 being active as far back as 2 Ma, but nine edifices with oldest ages of only 0.25 Ma. Youngest volcano age is commonly (22 edifices) 0–0.25 Ma, with some (13 edifices) having youngest ages of 1–1.5 Ma, suggesting two age groups within this group. Thirteen edifices show Holocene activity and the most active period is from 0–0.25 Ma. The edifices have evolved compositions with abundant andesite, dacite and rhyolite (Fig. 8). Basalt is absent and only three edifices have small amounts of basaltic andesite (Fig. 8). The edifice products are variable, with combinations including 24 having pyroclastic material as a major product, 12 having lava flows and 7 with lava domes, calderas and stratocones. Nineteen of the 35 units have volumes in the range of 10–100 km^3, of the remainder 7 have volumes between 1–10 km^3 and 6 have volumes >100 km^3. Distance of edifices from the volcanoes varies from 0.1–10 km, but the most common (17 edifices) is only 0–0.1 km.

Group 1C (28 units) edifices are a mixture of monogenetic and polygenetic gently and steeply sided stratovolcanoes. The oldest date of volcanism for 15 of these edifices is 0.5–1 Ma. The most recent volcanism in this group was generally (18 units) in the last 0.25 Ma. Compositionally, group 1C are andesite, basaltic andesite and basalt (Fig. 8). Twenty-four produced lava flows and 17 edifices have some pyroclastic deposits. These volcanic edifices have a large volume, 10 of the 28 have >100 km^{3}volume and 11 others have volumes 10–100 km^3. Distance of the edifices from the volcano is typically 1–10 km.

Group 1D (40 units), are mostly gently sided stratovolcanoes with some lava cones, 30 of which are from cluster 6 (Fig. 7a). Seventeen of the 40 edifices have oldest ages of 0.5–1 Ma and also have their youngest ages in this range. Sixteen of the others have oldest ages of greater than 2 Ma with youngest ages in the range 1.5–2 Ma. These characteristics suggest that this group represents a 'type' of volcanic event that has a relatively short lifetime. Compositionally 36 out of the 40 edifices are predominantly andesitic (Fig. 8). Major edifice products include pyroclastic deposits and lava domes. Twenty-three out of the 40 edifices have volumes (1–10 km^3) and the edifices make up 0–20% of the volume of the associated volcano.

Analysis 2 (Edifices, without time characteristics)

Analysis 2 divides up the edifices into two almost equally sized groups (Table 5 and Fig. 7b). The two groupings are more evenly divided between the clusters, but clusters 4 and 8 have significantly more group A than B edifices, whereas cluster 5 has significantly more group B than A edifices. In this analysis the strict consensus tree is much less well resolved, an artefact of the reduced number of characters.

Group 2A (65 units) edifices are made up of stratovolcanoes (steep and gently sided), calderas and lava domes. This group has no rhyolite and is mostly composed of basalt to andesite (Fig. 8). The edifices dominantly produce lava flows, pyroclastic flow deposits and debris avalanches. The volume of the edifice as a proportion of the total volcano volume is high for this group. Group 2A and 2B edifices show large ranges of volumes. Group 2A has 3 edifices with volumes in the 0–0.1 km^{3}range, 4

Table 5 Results from edifices without time, analysis 2

Cluster	Group		Total edifices per cluster
	A	B	
4	15	10	25
5	8	15	23
6	24	25	49
7	13	12	25
8	5	2	7
Total edifices per group	65	64	129

within 0.1–1 km^3, 16 between 1–10 km^3, 30 between 10–100 km^3, and 11>100 km^3.

Group 2B (64 units) edifices are a mixture of gently (22) and steeply (6) sloping stratovolcanoes producing pyroclastic flow plateaus (18), lava domes (18) and lava flows. Compositionally this group is dacitic with some andesite and rhyolite (Fig. 8). Volumes are typically 10–100 km^3. The volume of material produced by the edifices in group 2B is less varied than that produced by the edifices in group 2A. In group 2B, the volumes range with 3 between 0.1–1 km^3, 23 between 1–10 km^3, 29 between 10–100 km^3, with only a few (9) edifices producing >100 km^3.

Analysis 3 (Volcanoes, with time)

Analysis 3 yields three groups (Table 6). Uneven distribution between the clusters is again evident, in particular over half the volcanoes in Group C occur in cluster 6 (Fig. 7c).

Group 3A volcanoes (18 units) are steep sided stratovolcanoes and lava domes. The oldest ages of volcanism vary from 0–2 Ma, but 16 units are polygenetic and have been active in the past 0.25 Ma with historic and Holocene activity. Composition is variable, all 18 of the volcanoes have andesite products, 15 of those also range from basaltic andesite to basalt with some dacite (Fig. 8). Only two volcanoes have any rhyolite products. Products are lava flows and pyroclastic deposits. Eight of the 18 volcanoes have volumes that are 10–100 km^3.

Group 3B volcanoes (18 units) are stratovolcanoes, the majority are steep sided (8), but some gently sided (5). These volcanoes are mainly polygenetic with 13 volcanoes having historical activity only. Oldest ages vary (from 0–1.5 Ma) but average out at 0.5–1 Ma. Youngest ages are from 0–0.25 Ma. Dacite, andesite and basaltic andesite are the dominant compositions (Fig. 8). Four of these volcanoes have calderas, others have lava domes and pyroclastic flow plateaus. Seven of the 18 volcanoes have volumes in the range of 1–10 km^3, 8 of the 18 have volumes 10–100 km^3.

Group 3C volcanoes (14 units) are gently sloping stratovolcanoes. These volcanoes generally have oldest ages of 2.0 Ma, with youngest ages of 1 Ma and 18 out of 23 are polygenetic. Compositionally 21 of the 23 have andesite with large proportions of dacite and basaltic andesite (Fig. 8). Seven of these stratovolcanoes have pyroclastic flow plateaus, six with lava domes, and four with lava cones. Volumes of 1–10 km^3 are typical for 9 of the 14 volcanoes in this group, with another 8 of the 14 having volumes between 10–100 km^3.

Analysis 4 (Volcanoes, without time)

Analysis 4 divides volcanoes into three groups, (Table 6). There is uneven distribution of the groups between the clusters. In particular group B volcanoes only occur in clusters 6, 7 and 8 (Fig. 7d).

Group 4A volcanoes (24 units) mostly consist of steep sided stratovolcanoes with lava domes. There

Table 6 Results from cladistic analyses on volcanoes and edifices with and without time

Cluster	Group			Total volcanoes per cluster
	A	B	C	
Results from volcanoes with time, analysis 3				
4	5	6	3	14
5	3	3	3	9
6	2	5	12	19
7	6	3	3	12
8	2	1	2	5
Total volcanoes per group	18	18	23	59
Results from volcanoes without time, analysis 4				
4	9	0	5	14
5	2	0	7	9
6	7	7	4	18
7	3	5	5	13
8	3	1	1	5
Total volcanoes per group	24	13	22	59

is a broad range of compositions from basalt through to dacite (Fig. 8). Dominant products are lava domes, lava flows, and 20 out of the 24 volcanoes also produced pyroclastic rocks. Typically Group 4A volcanoes typically have volumes of 10–100 km^3 (14 of the 24 volcanoes), but there are several (7/24) with volumes 1–10 km^3.

Group 4B volcanoes (13 units) are all small gently sided stratovolcanoes and 2 of these have lava cones. They are dominantly andesitic with some basaltic andesite and basalt (Fig. 8). Seven out of the 13 have produced lava flows as a major product and 5 have pyroclastic flow deposits. The majority (7/13) of the volcanoes range from 1–10 km^3.

Group 4C volcanoes (22 units) contains five gently sloping stratocones. Group 4C volcanoes are dominated by andesite, dacite, and rhyolite compositions (Fig. 8). Products of these volcanoes include 19 with pyroclastic flow deposits (12 with pyroclastic flow plateaus), 13 with lava flows, 5 with lava domes and 7 formed calderas. A large proportion of the volcanoes (9 out of 22) have volumes in the 10–100 km^3 range, 6 have volumes of 1–10 km^3.

Comparison between analyses; with and without time characters

Comparison of Analysis 1 and Analysis 2 (Table 7) shows good correspondence between edifices of groups 1A and 1C and those of group 2A, while those in groups 1B and 1D form group 2B. This suggests that the analysis incorporating time is good quality, as the group signal is not lost between the two analyses. Comparison between analyses 3 and 4 (Table 8), is less systematic. Groups 4A and 3A correspond very well, and group 4B is made up of groups 3B and 3C. However, group 4C comprises several from each of groups 3A, B and C.

Analysis of the Japanese dataset including data from the Soufrière Hills Volcano (Montserrat), found that the Soufrière Hills Volcano did indeed classify into volcano groups 3A and 4C, groups which contain volcanoes with similar volumes, areas and activity. The polygenetic Soufrière Hills Volcano is characterised by andesitic lava domes producing pyroclastic flows. In common with members of this group, the Soufrière Hills Volcano has a volume of 10–100 km^3 and has been active both in the Holocene and historically.

The various cladistic groupings are used to characterise the volcano clusters (Tamura et al. 2002). Cluster 4 shows a broad change in composition with time; from predominantly basalt, andesite and dacite between 1–1.5 Ma to andesite to rhyolite at 0–0.25 Ma. There is also a change from the older steep sided stratovolcanoes with large (10–100 km^3) volumes, to younger volcanoes having lower volumes (1–10 km^3), calderas and pyroclastic flow plateaus. Cluster 5 volcanoes older than 1 Ma produce dacite products, forming gently sided stratovolcanoes. Volcanoes younger than 1 Ma produce rhyolitic pyroclastic flow plateaus, calderas, with edifices that are closer to the volcano and with larger volumes, such as Chokai volcano located on the western edge of cluster 5. Cluster 6 shows a pronounced compositional evolution with time. First we note that the oldest volcanic products (>1 Ma) are voluminous (approx. 10 km^3) rhyolitic ignimbrites (Suto 1987) but these units were not included in the cladistic analysis because no source structure could be identified. Andesite is the dominant product of the edifices and volcanoes. As the volcanoes get younger there are minor compositional changes from more basalt-rich to more dacite-rich. The older volcanoes in this cluster developed

Table 7 Comparison between analysis 1 (edifices with time characters) and analysis 2 (edifices without time characters

Analysis 1	Analysis 2		Total edifices per group
	2A	2B	
1A	23	3	26
1B	4	31	35
1C	26	2	28
1D	12	28	40
Total edifices per group	65	64	129

Comparisons for volcanoes and for edifices showing how many volcanoes or edifices remain in a similar group when time related characters are used in the analysis versus when they are not used.

Table 8 Comparison between analysis 3 (volcanoes with time characters) and analysis 4 (volcanoes without time characters)

Analysis 3	Analysis 4			Total volcanoes per group
	4A	4B	4C	
3A	12	1	5	18
3B	6	4	8	18
3C	6	8	9	23
Total volcanoes per group	24	13	22	59

Comparisons for volcanoes and for edifices showing how many volcanoes or edifices remain in a similar group when time related characters are used in the analysis versus when they are not used.

calderas, whereas there are three young predominantly basaltic and andesitic stratocones located at the edges of this volcano cluster (Iwate, Akita-Komagatake and Akita Yake-yama). These may be located at the cluster edges as the magma finds an easier pathway to the surface there. The edifices attached to the volcanoes >1 Ma have large values for the volume of the edifice as a proportion of the volume of the volcano, whereas the more recent volcanoes have lower values. Cluster 7 has older volcanoes (>0.25 Ma), which have mainly basalt and basaltic andesite compositions, whereas the younger volcanoes (0–0.25 Ma) have basaltic andesite and andesite compositions. The volcanoes >0.25 Ma are gently sided stratovolcanoes, which change to steep sided stratovolcanoes for the volcanoes <0.25 Ma. A significant feature of the last 0.2 Ma of cluster 7 is the formation of the 10 km×10 km Lake Towada. Post caldera volcanics initiated 10,000 years ago, with a large eruption and second caldera forming phase occurring 8,600 years ago. After the 915 eruption 2 lava domes formed. This agrees with the cladistic analyses that suggest the younger volcanoes have lava domes, pyroclastic flow plateaus, lava cones and calderas. Cluster 8 does not show any marked evolution.

Discussion

This analysis is based on data with much fewer characters and many more units in comparison to biological systems. Biologists typically work with tens or hundreds of characters, and molecular biologists often work with thousands. As a general rule, there should usually be 2–3 times as many characters as units, but for the volcanic edifice dataset there are 34 characters for 129 units. However, many of the volcano characters are multistate and therefore provide more information than a simple binary character. Databases in principle could be constructed for volcanoes with many more characters defined. However, here our choice of characteristics was restricted by the existing database.

Early cladistic analyses were commonly based on low numbers of generalised characters. The characters have rapidly swelled in number and complexity with increasing levels of study and use. In most biological studies the aim is to find the relationship between every single unit. Here, however, our aim is to identify major groups of units, and we have not attempted to obtain high-level resolution. Large-scale trees can be assembled and used to provide a framework to classify volcanoes systematically. The nature of cladistics means that the most important and decisive characters can be determined and a systematic descriptive framework can be established around them, with the subjectivity of phenetic methods removed.

The resultant trees for Analysis 1 were well resolved, a result consistent with the low numbers of characters. Four well-defined groups were identified based on major branching patterns on the tree. Within biology, species and genera would normally have been identified before a large-scale analysis and therefore would be able to readily identify major relationships. Here, we make no *a priori* assumptions and there are no previous studies for comparison. Cladistics has proved to be an enduring force in biology and is now the basis for classification in biology (species, genera, families etc.). Systematic methods may have similar influences and explanatory power in the study of volcanoes.

The results of cladistic analysis can depend on what characteristics are chosen. Here we have compared analysis of the same data with and without characteristics related to the age and duration of volcanism. Although the groupings are different, the majority of the volcanic edifices or volcanoes shared common groupings. The differences indicate that time characteristics can conflict with other characteristics.

The cladistic groupings are unevenly distributed between the volcano clusters identified by Tamura et al. (2002). This observation is consistent with the concept that each volcanic system has a certain coherence, generating similar kinds of volcanoes repeatedly in space and time. The cladistic analysis also shows that the clusters evolve with the character of the volcanism changing with time. Evolution of composition from more mafic to more silicic volcanism is evident in clusters 4, 5 and 7. The dominant magma composition affects eruptive styles, landforms and products. The cladistic analysis appears to pick out these relationships because the time characteristics are linked to the compositional controls. Some of the factors that might account for these differences may include magma supply rate, longevity, crustal structure and magma rheology. The classification results invite enquiry into the underlying dynamical causes of variability.

A future objective might be to apply cladistic analysis to regional and global datasets to look for patterns of volcanism in space and time. The analysis containing data from the Soufrière Hills Volcano, Montserrat fit into an appropriate grouping suggests how this method could be used develop a global volcano classification. For example, a comparison of cladistic groups from different types of island arc and volcanoes from different tectonic settings might reveal new relationships that have significance for understanding magmatic processes. This would be a daunting task on a global level, but smaller scale studies focusing on selected regions may well prove productive. As in biology some relationships might already be well-established or obvious without the need for cladistic analysis, but unexpected relationships might also emerge.

Characters are expected to improve over time through the scientific principles of Francis Bacon which state that more research and known facts should bring us closer to the truth. New characters and datasets in biological studies tend to be the least accurate in the long-term, but characters and data are modified and adjusted by subsequent authors and workers as errors and ambiguities are removed and the results improve in both resolution and repeatability. Therefore the data and results presented here form a starting point from which systematic classification methods in volcanology can develop. The approach is currently limited by the type and quality of information in volcano databases. Although the Japanese database is one of the most comprehensive, it is still largely qualitative and characteristics are limited. There is much opportunity for improving the systematic description of volcanoes and making them more quantitative.

There remains an issue with computing power. Currently, datasets with more than 16–17 taxa are too large to complete a full search of trees (i.e. one guaranteed to find all the MPTs). Here we are analysing over 125, far too many to perform a complete search. However, there are well-established methods to deal with this problem and yield well-supported results. New algorithms based in a new software package (TNT, Goloboff and Pol 2002) can deal with very large datasets and which have very high numbers of character variables, features ideal for analysing volcano based data.

Conclusions

Cladistic analysis offers a systematic approach to classification of volcanoes that reduces bias and subjectivity, and takes account of all the information provided in descriptive databases of volcanoes. Our exploratory study of the database of Japanese volcanoes in NE Japan demonstrates its utility in identifying groupings of volcanoes with similar characteristics. Analyses with and without time characteristics generate groups which mostly correspond. Volcano clusters have been previously identified in NE Japan (Tamura et al. 2002), and cladistic groupings are unevenly distributed between the clusters. This result indicates that each cluster has some distinctive but coherent styles of volcanism and, for groupings with time characteristics, distinctive evolution as magmatic systems. In turn the results suggest that cladistic analysis can be a useful tool for elucidating dynamic igneous processes that could be applied to other regions and globally. The utility of the method can be expected to increase if more detailed and quantitative volcano databases are generated. More rigorous and systematic classification of volcanoes as entities is also required for statistical studies of spatial and temporal distributions of volcanoes with hazards applications.

Acknowledgements We much appreciate the support and encouragement of the Nuclear Waste Management of Japan (NUMO). We acknowledge helpful discussion with Neil Chapman, Chuck Connor, Yoshi Tamura and Hirofumi Kondo. Help on comparative statistical methods and the application of cladistics to this problem, was kindly provided by Phil Donoghue, Mike Benton, Oyvind Hammer, Dave Unwin and Simon Braddy. RSJS acknowledges a Royal Society-Wolfson merit Award.

References

Bremond d'Ars J, Jaupart C, Sparks RSJ (1995) The distribution of volcanoes in active margins. J Geophys Res 100:20421–20432
Brower AVZ (2000) Evolution is not a necessary assumption of cladistics. Cladistics 16:143–154
Bullard FM (1962) Volcanoes: in history, in theory and in eruption. University of Texas Press, Austin
Committee for Catalog of Quaternary Volcanoes in Japan (eds) (1999) Catalog of Quaternary volcanoes in Japan. The Volcanological Society of Japan
Condit CD, Connor CB (1996) Recurrence rates of volcanism in basaltic volcanic fields: an example from the Springerville volcanic field, Arizona. Geol Soc Amer Bull 108:1225–1241
Connor CB, Hill BE (1995) Three non homogenous Poisson models for the probability of basaltic volcanism: application to the Yucca Mountain region, Nevada. J Geophys Res 100:10107–10125
Cotton CA (1944) Volcanoes as landscape forms. Victoria University College, Whitcombe and Tombi Limited, Wellington, NZ
Darwin C (1859) On the origin of species by means of natural selection. Murray, London
Davis JC (2002) Statistics and data analysis in geology. Wiley, New York, pp 530
Francis P, Oppenheimer C (2004) Volcanoes, 2nd edn. Oxford University Press, Oxford
Goloboff PA, Pol D (2002) Semi-strict supertrees. Cladistics 18:514–525
Hennig W (1966) Phylogenetic systematics. University of Illinois Press, Urbana
Kitching IJ, Forey PL, Humphries CJ, Williams DM (1998) Cladistics: theory and practice of parsimony analysis. Oxford University Press, Oxford
Macdonald GA (1972) Volcanoes. Prentice Hall, Inc., Englewood Cliffs
Marsh BD (1981) Physics of magmatic processes (book review). Am Sci 69:970
Martin AJ, Umeda K, Connor CB, Weller JN, Zhao D, Takahashi M (2004) Modelling long-term hazards through Bayesian inference: an example from the Tohoku volcanic arc, Japan. J Geophys Res 109:B10208, DOI 10.1029/2004JB003201
Rittmann A (1962) Volcanoes and their activity. Interscience, New York
Schmincke HU (2004) Volcanism. Springer, Berlin Heidelberg New York
Scrope GP (1872) Volcanoes the character of their phenomena, their share in the structure and composition of the surface of the Globe, and their relation to its internal forces with a descriptive

catalogue of all known volcanoes and volcanic formations, 2nd edn. Longmans, Green, Reader and Dyer, London
Simkin T, Siebert L (1994) Volcanoes of the World, 2nd edition: Geoscience Press, Tucson
Suto S (1987) Geological structure of the central part of the Sengan geothermal area, northeast Japan. Rep Geol Surv Japan 266:43–76
Swofford DL (1998) PAUP* Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer, Massachusetts
Tamura Y, Tatsumi Y, Zhao D, Kido Y, Shukuno H (2002) Hot fingers in the mantle wedge: new insights into magma genesis in subduction zones. Earth Planet Sci Lett 197:105–106
Williams H, McBirney AR (1979) Volcanology. Freeman, Cooper and Company, San Franscisco