



OPEN

Metagenomic analysis reveals rapid development of soil biota on fresh volcanic ash

Hokyung Song¹, Dorsaf Kerfahi², Koichi Takahashi³, Sophie L. Nixon¹, Binu M. Tripathi⁴, Hyoki Kim⁵, Ryunosuke Tateno⁶✉ & Jonathan Adams⁷✉

Little is known of the earliest stages of soil biota development of volcanic ash, and how rapidly it can proceed. We investigated the potential for soil biota development during the first 3 years, using outdoor mesocosms of sterile, freshly fallen volcanic ash from the Sakurajima volcano, Japan. Mesocosms were positioned in a range of climates across Japan and compared over 3 years, against the developed soils of surrounding natural ecosystems. DNA was extracted from mesocosms and community composition assessed using 16S rRNA gene sequences. Metagenome sequences were obtained using shotgun metagenome sequencing. While at 12 months there was insufficient DNA for sequencing, by 24 months and 36 months, the ash-soil metagenomes already showed a similar diversity of functional genes to the developed soils, with a similar range of functions. In a surprising contrast with our hypotheses, we found that the developing ash-soil community already showed a similar gene function diversity, phylum diversity and overall relative abundances of kingdoms of life when compared to developed forest soils. The ash mesocosms also did not show any increased relative abundance of genes associated with autotrophy (*rbc*, *coxL*), nor increased relative abundance of genes that are associated with acquisition of nutrients from abiotic sources (*nifH*). Although gene identities and taxonomic affinities in the developing ash-soils are to some extent distinct from the natural vegetation soils, it is surprising that so many of the key components of a soil community develop already by the 24-month stage. In this system, however, rapid development may be facilitated by the relatively moderate pH of the Sakurajima ash, proximity of our mesocosms to propagule sources, and the rapid establishment of a productive bryophyte and lichen layer on the surface. Ash from other volcanoes richer in acids or more distant from propagule sources could show a different pattern and slower soil biota development.

Volcanic ash is one of the world's main sources of soils, particularly in the more active regions of the Pacific 'ring of fire', such as Japan^{1,2}. Unlike its namesake wood ash, volcanic 'ash' is a powder of silicate minerals, and when it first falls it is sterile or nearly so, and essentially free of soluble nutrients³. The ash also has little water retaining capacity and erodes easily. Chemical and biological processes are required to release nutrients, and contribute organic matter to sustain food webs³. It is clear that on a timescale of decades, volcanic ash can become a fertile soil with a diverse biota—and indeed the soils of much of Japan are developed over volcanic ash^{1,2}. The eventual result of most successions on volcanic ash in Japan and elsewhere are organic rich soils under either forest or (in colder environments) scrub and tundra².

Studies on developing volcanic ash soils have so far concentrated on chemical characterisation of the ash soils⁴⁻⁷ or description of bacterial communities using only 16S rRNA gene sequencing⁸⁻¹¹. There have been few studies of the functional aspects of the biota of developing ash soils. In one exception, a metagenome study by Fujimura et al.³ compared the functional gene profile of 3.5-, 6.6-, and 9.5-year old ash from the Miyakejima volcano in Japan and noted that the *rbcL* gene for CO₂ fixation and *nifH* gene for nitrogen fixation were more abundant in the developing ash-soils. These genes have been suggested to be important for energy, carbon and

¹Department of Earth and Environmental Sciences, The University of Manchester, Manchester, UK. ²School of Natural Sciences, Department of Biological Sciences, Keimyung University, Taegu 42601, Republic of Korea. ³Department of Biology, Faculty of Science, Shinshu University, Matsumoto 390-8621, Japan. ⁴Korea Polar Research Institute, Incheon, Republic of Korea. ⁵Celeomics Inc., 19F, Bldg. A, BYC High City, 131, Gasandigital 1-ro, Gwumcheon-gu, Seoul 153-718, Korea. ⁶Field Science Education and Research Center, Kyoto University, Kyoto 606-8502, Japan. ⁷School of Geography and Ocean Sciences, Nanjing University, Nanjing 210023, Jiangsu, China. ✉email: rtateno@kais.kyoto-u.ac.jp; foundinkualalumpur@yahoo.com

nutrient acquisition of microbes in ash soils that lack organic matter^{12–16}. However, the study of Fujimura et al.³ did not consider the even earlier stages—the first 2 years—of biota development. The ash system they studied was also one that was relatively acidic at inception, being rich in sulphuric acid, while many other ash deposits are not so acidified³.

The relative paucity of detailed studies on ecosystem development on volcanic ash may be partly explained by the fact that volcanic explosion events are sporadic and unpredictable, and the resulting ash field sites are often off-limits for sampling for several years afterwards due to safety concerns. While observations of macroscopic life in the earliest stages of volcanic ash succession have long been recorded^{17–19}, there have been few or no studies of soil biota during the first 2 or 3 years of soil biota development in volcanic ash. Most studies of soil biota have begun at five or more years after the ash initially fell^{20,21}, partly because most areas of ash deposition are off limits for sampling for several years until volcanic activity has subsided. Here we set out to circumvent some of the practical difficulties of observing soil development processes on freshly fallen volcanic ash by setting up outdoor mesocosms positioned in various environments across Japan, using the same source of freshly fallen volcanic ash from the Sakurajima volcano, and by using a metagenome approach. Our intention was to understand how rapidly a soil biota appears in the ash, how it may differ from more developed soils, and whether climate and location have a large effect on the development of biota.

The major questions behind this study were:

1) *How rapidly can the major functional components of a soil ecosystem appear in volcanic ash?* We hypothesized that due to the relatively extreme conditions in the ash—such as lack of available mineral nutrients and organic matter—and also due to limitations on arrival and establishment of biota due to slow dispersal, there would be: (1) a lower diversity of broad taxonomic categories of organisms, such as phyla of fungi, archaea, Protista and Metazoa than are normally found in a developed soil. (2) A lower diversity of categories of functional genes of soil biota than are normally found in a developed soil, indicating less functional complexity in the early ash-soil system.

2) *Which distinct functional aspects characterize the early development of ash-soil biota?* We hypothesized that the distinct chemical environment of the developing ash would result in increased relative abundance of a number of gene categories. For example: (1) Increased relative abundance of genes associated with autotrophy (e.g. *rbc* gene and *coxL* gene) due to lack of organic matter^{12–16}, and abundance of photosynthetic and chemoautotrophic groups of microbes^{8,21–23}. Since the usual heterotrophic routes are assumed not to be available initially in the newly developing ash ecosystem, and oxidizable minerals are abundant, microbial chemosynthesis may be expected to be relatively more important than in a developed soil. (2) Increased relative abundance of genes that are associated with acquisition of nutrients from abiotic sources rather than decomposition of organic matter (e.g. *nif* genes). (3) Increased relative abundance of stress response genes and dormancy related genes, due to more frequent drying of the ash-soil which will presumably be well drained and low in organic matter. (4) Decreased relative abundance of interaction related genes (CRISPR genes and genes classified as regulation and cell signalling and virulence genes in the SEED subsystem database²⁴).

Results

Soil physicochemical conditions. Supplementary Fig. S1 shows total carbon and total nitrogen concentrations and pH of the ash soil and forest soils at each site. Total carbon and total nitrogen concentrations were significantly lower in the ash soils, irrespective of location and age. pH was also significantly lower in the ash soils in Sakurajima and Takakuma, but no such pattern was found in other sites. At Norikura Low, the third year ash soil had higher pH.

Bacterial community composition and diversity analysed based on the 16S rRNA gene amplification sequencing. Two way crossed ANOSIM test results showed significant differences in bacterial community composition between 2nd year ash pots, 3rd year ash plots and natural forest soils (Fig. 1, Supplementary Table S1). Also, the community composition was significantly different between sites (Fig. 1, Supplementary Table S1). The Bray–Curtis distances of the 2nd-year pot samples and the forest samples were not significantly different from the Bray–Curtis distances of the 3rd-year pot samples and the forest samples at Norikura and Sakurajima. However, the Bray–Curtis distance between the 2nd-year ash soils and forests was significantly higher than the Bray–Curtis distance between the 3rd-year ash soils and forests at Kamigamo (Supplementary Fig. S2).

For both 24 and 36 month samples, we found significant differences between ash-soils and natural forest soils in the relative abundance of major bacterial phyla (Supplementary Fig. S3), continuing the trend established for the 24 month samples by Kerfahi et al.¹⁰. The relative abundance of Acidobacteria was significantly lower ($p < 0.05$) in the ash soils compared to forest at all sites except for Sakurajima, where the 3rd-year ash soil showed no significant difference in the relative abundance of Acidobacteria compared to forest soil ($p = 0.459$). The relative abundance of Patescibacteria was also overall lower in the ash soils compared to forest (no significant difference was found at Kamigamo, but the trend was similar). In contrast, the relative abundance of Chloroflexi in the ash soils was significantly higher compared to the forest soils at all sites.

Among the major bacterial families (Supplementary Fig. S4), we found significantly higher relative abundance of Ktedonobacteraceae in the ash soils compared to forest soils at all sites. The relative abundance of Acetobacteraceae was also higher in the ash soils in all sites except for Sakurajima, where there was no significant difference in the relative abundance of Acetobacteraceae found between the 3rd-year ash soils in comparison with forest ($p = 0.349$). In contrast, the relative abundance of Xanthobacteraceae was significantly higher in the forest soils at Norikura and Sakurajima but no significant difference found at Kamigamo. The relative abundance of Pedosphaeraceae was always higher in the forest soils regardless of location and age.

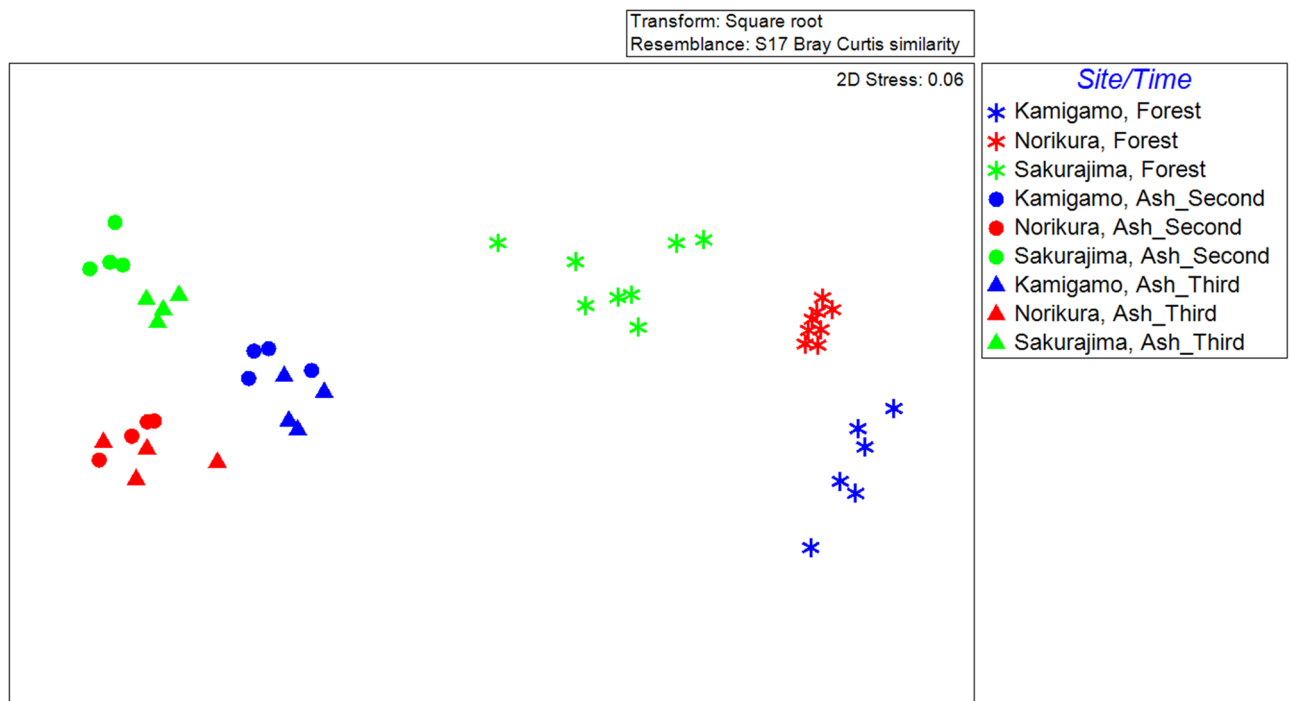


Figure 1. nMDS plot drawn based on Bray–Curtis dissimilarity of bacterial community between samples, based on amplicon sequencing which new 36 month results added. Second = 24 month, third = 36 month.

The total richness of bacterial OTUs and Shannon diversity of the forest soils were significantly higher than in the ash soils at Norikura and Sakurajima, for both 24 months and for 36 months (Fig. 2). There was no significant difference found at Kamigamo, but the trend was similar.

Comparison of bacterial community composition and diversity between 16S rRNA and metagenome sequencing. Family level bacterial community composition based on the metagenome sequencing data are shown in Supplementary Fig. S5. The major families identified by metagenome sequencing were similar to those identified by 16S rRNA amplicon sequencing (Supplementary Fig. S4). However, the relative abundance of minor families in the metagenomes was higher than identified by 16S rRNA amplicon sequencing. Also, the proportion of unclassified/unidentified sequences was lower in the metagenome sequencing data. Family diversity accumulation curve shows higher diversity of families calculated based on metagenome sequencing data (Supplementary Fig. S6).

Phylogenetic assembly and turnover of bacterial community. SES.MNTD values calculated based on 16S rRNA data in all samples were below zero suggesting that co-occurring OTUs were phylogenetically more related than expected by chance (Supplementary Fig. S7). SES.MNTD values of the natural forest soils were significantly lower than the ash soils suggesting stronger phylogenetic clustering in the forest soils at Norikura and at Sakurajima (Supplementary Fig. S7). However, no significant difference was found at Kamigamo.

Archaeal and eukaryotic phyla. In total, five archaeal phyla were found in the metagenomes in this study (Crenarchaeota, Euryarchaeota, Korarchaeota, Nanoarchaeota, Thaumarchaeota). The number of archaeal phyla annotated based on metagenome sequence data was quite consistent, varying only from four to five phyla amongst the various samples. The number of eukaryotic phyla annotated based on metagenome sequence was significantly different between ash soils and forest soils in Kamigamo and Takakuma (Supplementary Fig. S8). In both sites, forest soil had a lower total number of Eukaryotic phyla. The phyla that were present in the ash-soil samples but not in any forest samples were all minor taxonomic groups which had relative abundance of 0–0.032%. The overall non-bacterial phylum composition was similar between the ash and the forest samples (Supplementary Fig. S9).

Bacterial functional gene composition analyzed based on metagenome sequencing. As most of the sequences annotated based on the metagenome sequencing were bacterial (on average, 97% of reads were bacterial)—and in order to make comparison with the findings based on the 16S rRNA gene amplicon sequencing—we have only focused on bacterial functional genes in this and the following section.

Two way crossed ANOSIM test results showed a significant difference between the 2nd-year ash soils, 3rd-year ash soils, and forest soils in terms of bacterial functional gene composition (Fig. 3, Supplementary Table S2). Also, the community composition was significantly different between sites (Fig. 3, Supplementary Table S2). The

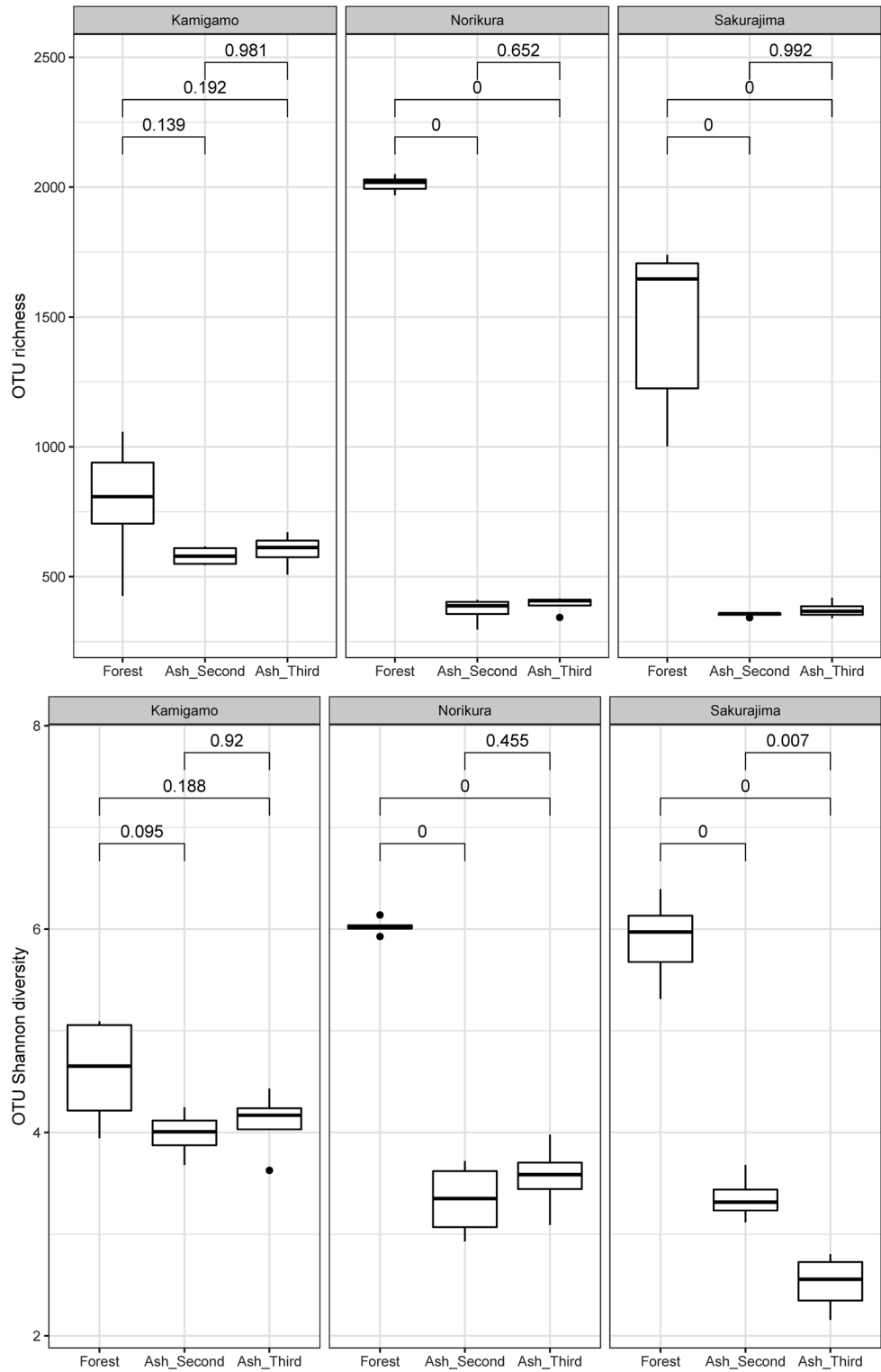


Figure 2. Box plots showing OTU richness (upper) and OTU Shannon diversity based on 16S rRNA amplicon sequencing.

Bray–Curtis distances of the 2nd-year pot samples and the forest samples were not significantly different from the Bray–Curtis distances of the 3rd-year pot samples and the forest samples at all sites (Fig. S10). Figure 4 shows the heatmap of standardized averaged relative abundance of Level 1 functional gene categories.

In terms of specific gene categories, we found the relative abundance of *rbcL* genes showed no consistent pattern (Fig. 5a). The relative abundance of these genes was highest in the forest at the lower elevation of Norikura compared to the ash mesocosms, but the opposite pattern was shown in Takakuma. In contrast to our

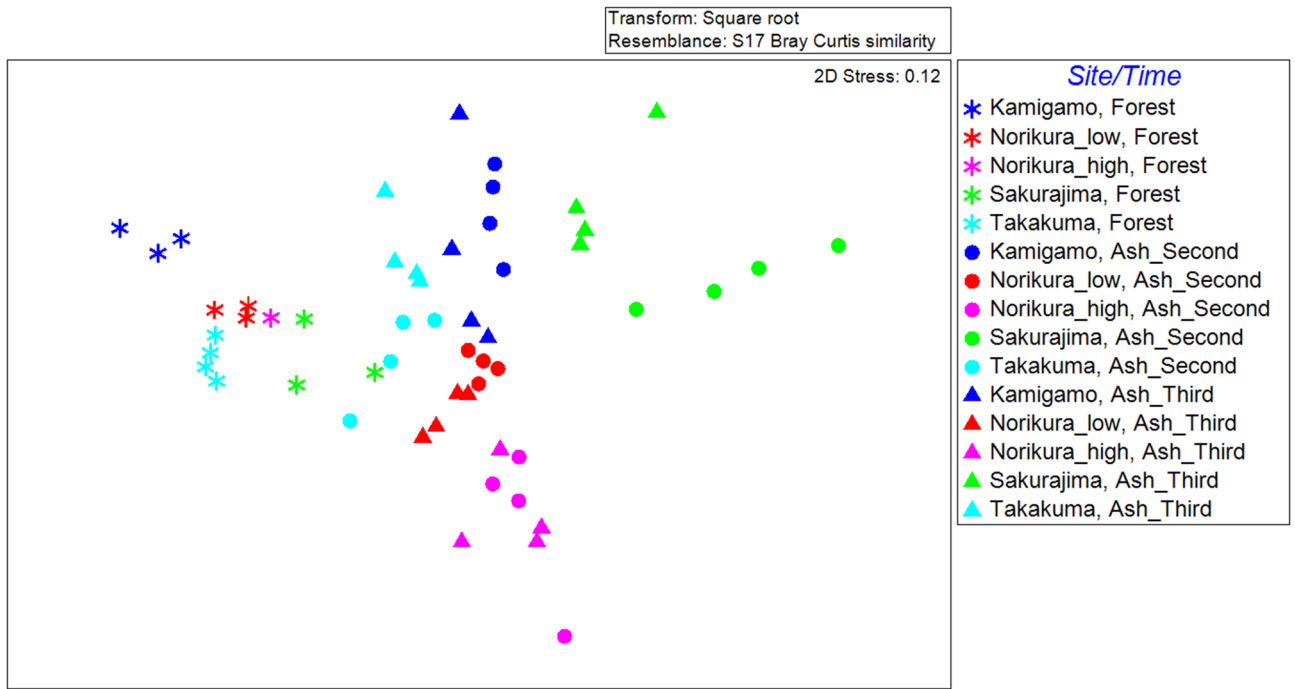


Figure 3. nMDS plot drawn based on Bray–Curtis dissimilarity of Subsystem level4 functional gene composition between samples.

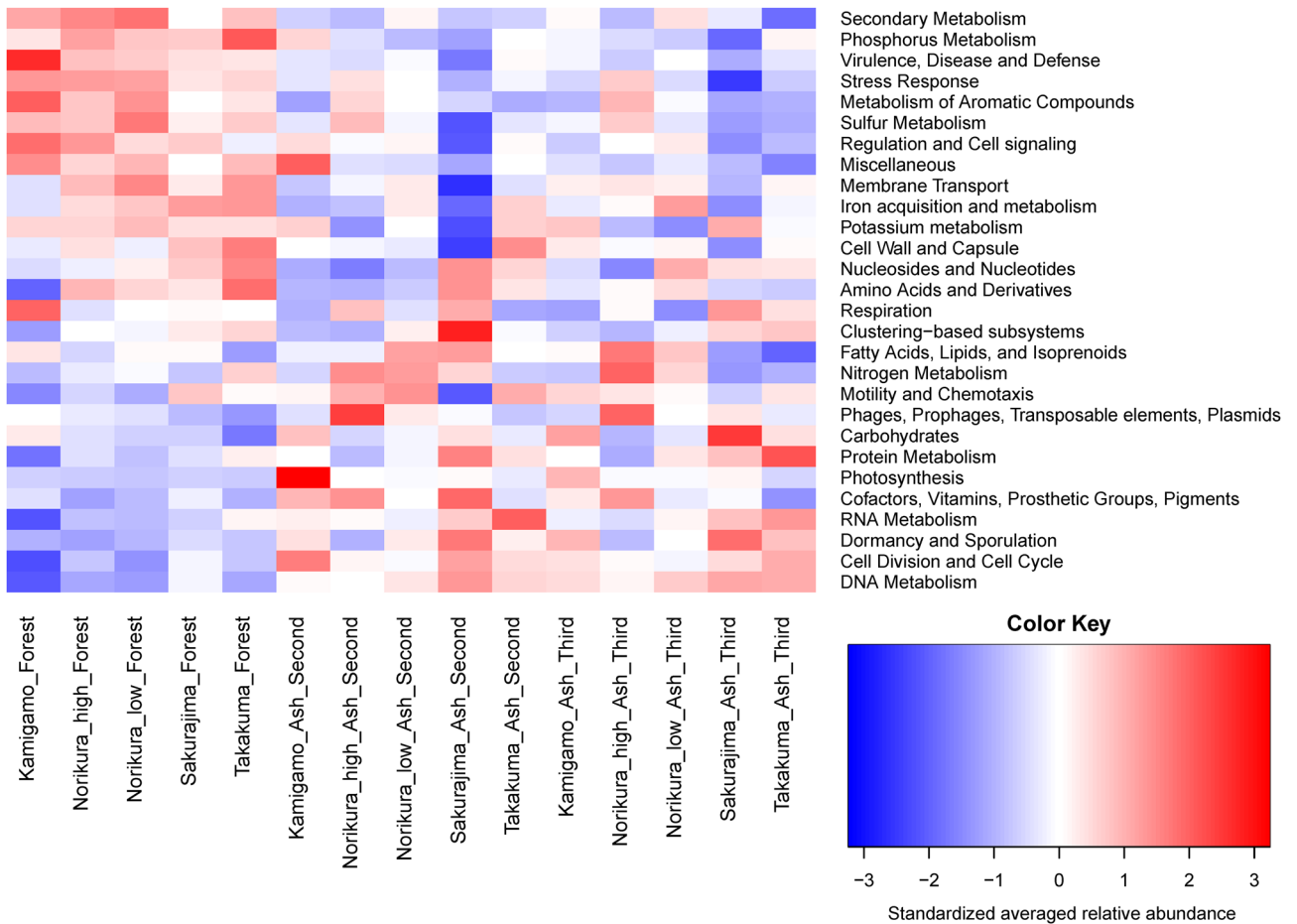


Figure 4. Heatmap of standardized averaged relative abundance of Level 1 functional gene categories.

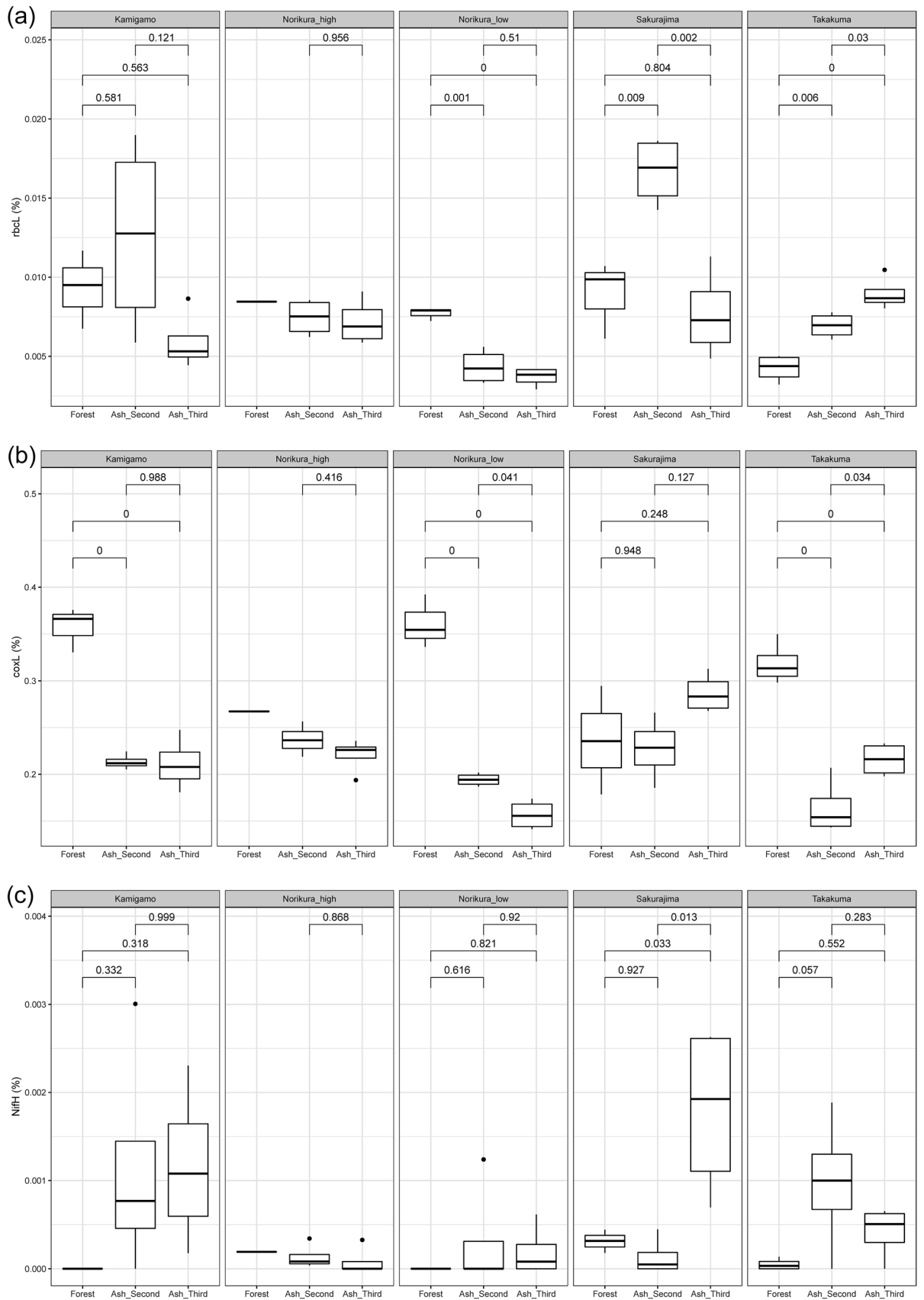


Figure 5. (a) Relative abundance of *rbcL* (ribulose biphosphate carboxylase large chain) gene in the ash soils and forest soils. (b) Relative abundance of *coxL* (Carbon monoxide dehydrogenase large chain) gene in the ash soil and forest soil. (c) Relative abundance of *nifH* gene in the ash soil and forest soil.

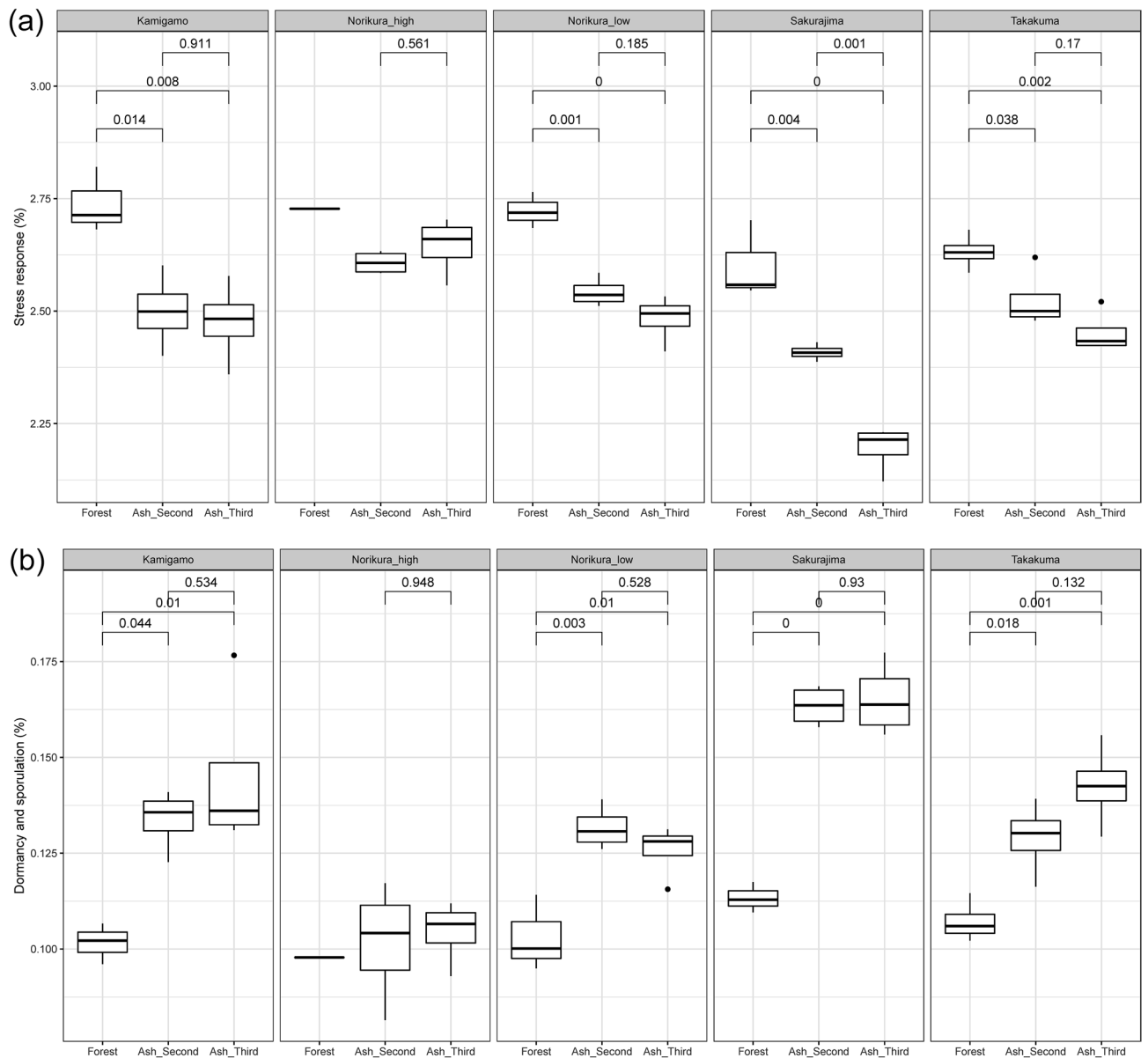


Figure 6. (a) Relative abundance of stress response gene category in the ash soil and forest soil. (b) Relative abundance of dormancy and sporulation gene category in the ash soil and forest soil.

expectation, the relative abundance of *coxL* genes was higher in the forest soils (Fig. 5b) in most of the sites except for Sakurajima. The relative abundance of *nifH* genes was very low overall, and showed no significant difference between forest and ash soils (Fig. 5c).

The relative abundance of stress response genes was in general, higher in the forest soils (Fig. 6a). Dormancy and sporulation related genes were relatively more abundant in the ash soils (Fig. 6b). Gene categories related to cell–cell interactions were relatively more abundant in the forests in most of the sites (Fig. 7a,b). However, the relative abundance of CRISPR genes was lower overall in the forest soils (Fig. 7c).

Bacterial functional gene diversity. Figure 8 shows that there were no consistent trends found in the functional gene diversity between ash soils and forest soils. Functional gene diversity was higher in forest soils than in ash soils at Sakurajima. However, the opposite pattern was found at Kamigamo. There was no significant difference in the functional gene diversity between ash soils and forest soils in Norkura and Takakuma.

Discussion

How rapidly can the major functional components of a soil ecosystem appear in volcanic ash? The overall conclusion from comparing the functional composition of the developing ash-soil with natural forest soils from the same climates, is that the major biological aspects of the system have developed sur-

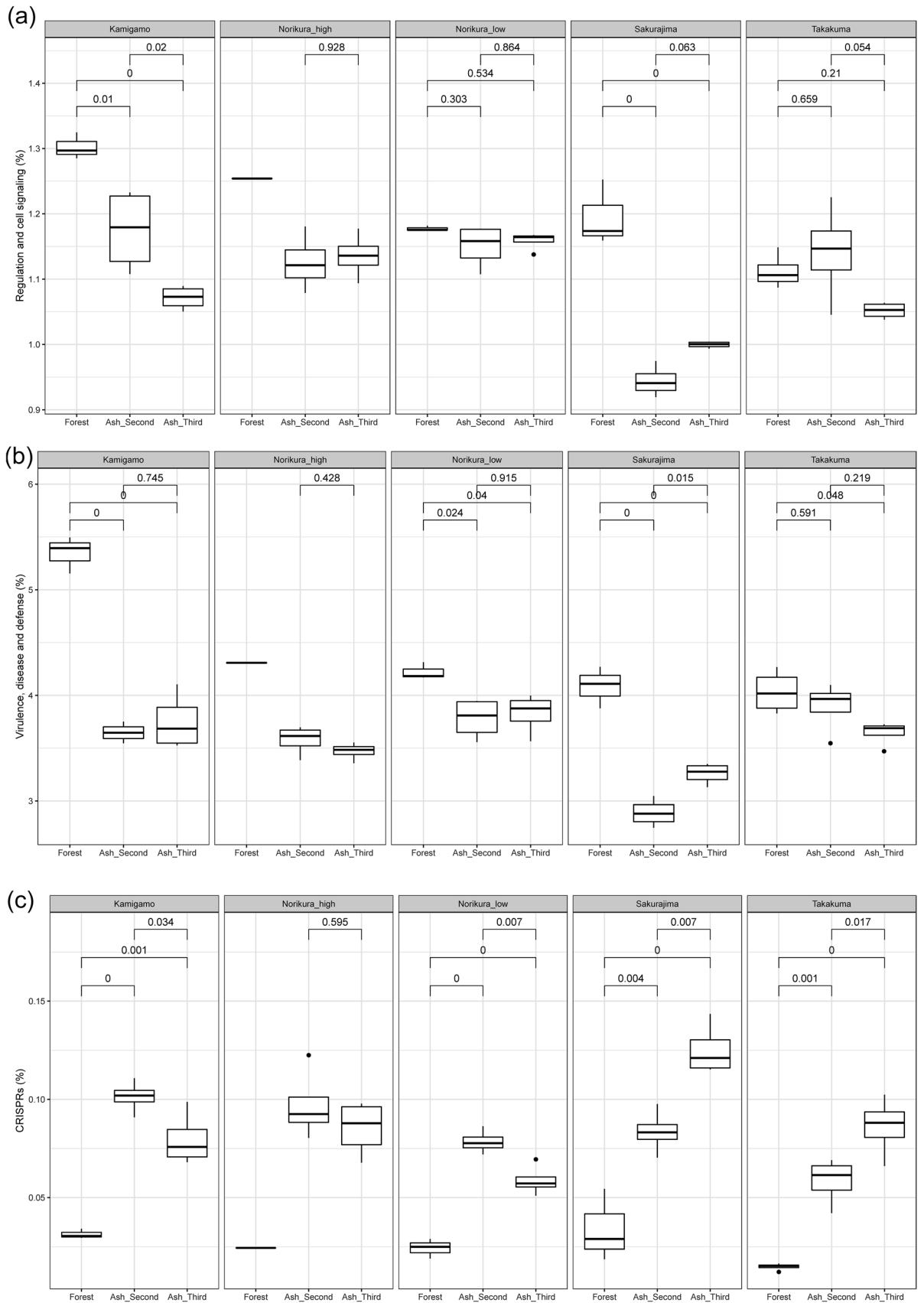


Figure 7. (a) Relative abundance of regulation and cell signalling gene category in the ash soil and forest soil. (b) Relative abundance of virulence, disease, and defense gene category in the ash soil and forest soil. (c) Relative abundance of CRISPRs in the ash soil and forest soil.

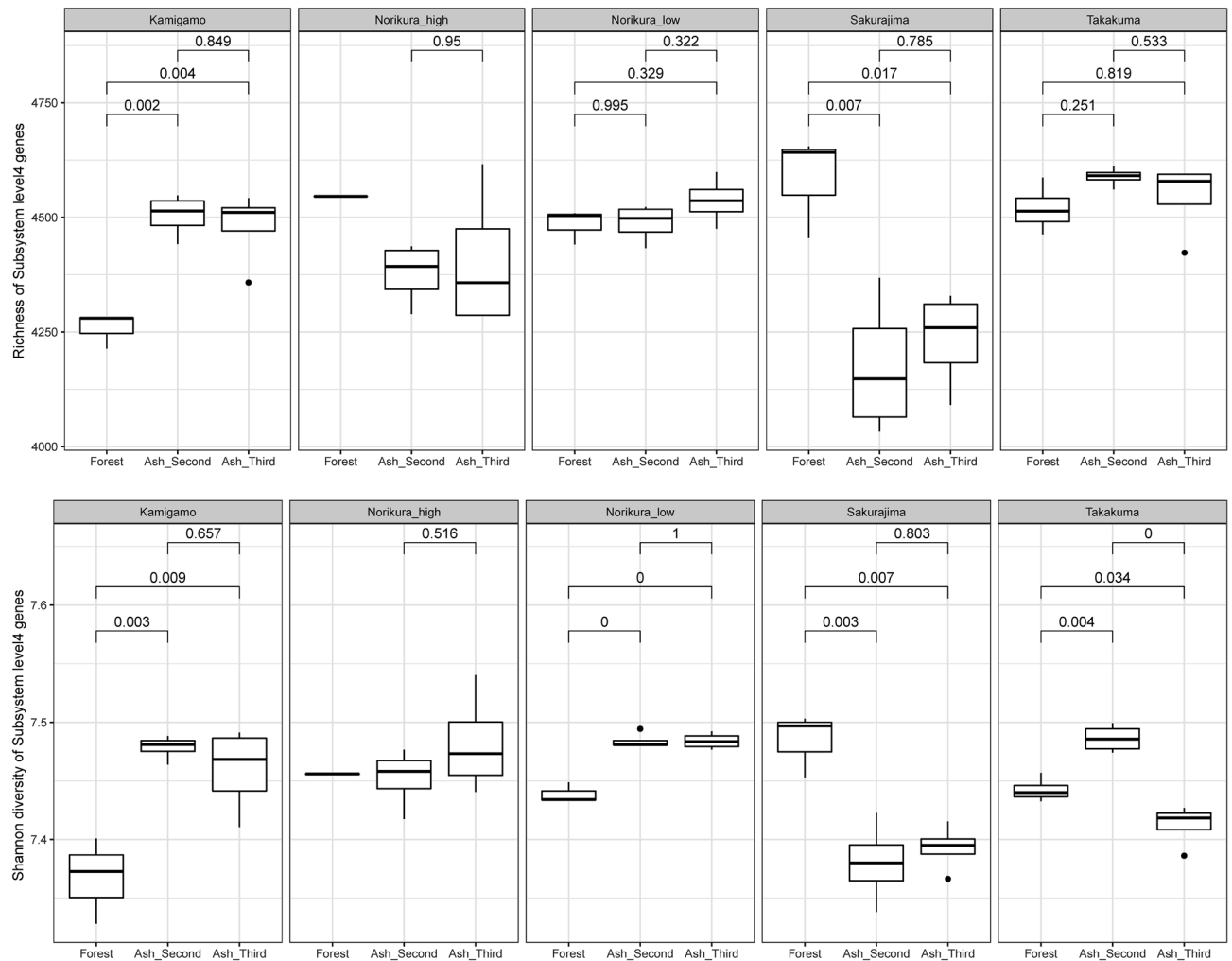


Figure 8. Box plots showing richness of Subsystem level 4 functional genes (upper) and Shannon diversity of Subsystem level 4 functional genes (lower).

prisingly quickly. Already by the 24-month and 36-month stage, the ash-soil had a similar diversity of bacterial functional genes to a forest soil (Fig. 8).

This contrasts with the findings of Fujimura et al.³, who found that ash deposited by the Miyakejima volcano still showed a very distinct and low diversity community compared to that expected of normal forest soils. This difference could be attributable to differences in the initial chemical composition of the ash: the pH value of the Miyakejima ash studied by Fujimura et al.³ was much lower than the ash we sampled, due to the persistence of acids from the initial eruption, or acidic rainfall from later outgassing. Our Sakurajima ash samples had only a slightly acidic pH (5.2) at the time of harvesting, which was similar to the average value (5.14) of 251 ash samples from Sakurajima collected from 1955 to 2001²⁵, and they remained in the 5–6 pH range at both 24 months and 36 months. Even the ash soil mesocosm samples that were situated on the lower slopes of the Sakurajima volcano, and the nearby natural forest soils at Sakurajima, remained in the pH 5–6 range and had a similar gene composition to other sites, suggesting that the area around the Sakurajima volcano is not subject to frequent highly acidic rain events. The difference between the Sakurajima ash and the Miyakejima ash hints at the importance of details of the chemical environment of the ash for determining its development into a functioning soil ecosystem.

Another factor which may explain the difference between the paths of biota development of the ash from the two volcanoes is the relative availability of propagules. Our ash soils were surrounded by developed ecosystems which could provide a rain of propagules of microorganisms and bryophytes in windblown dust or in rain splash. By contrast the Miyakejima ash field was an area of extensive ecosystem destruction. If bryophytes and lichens had not been able to establish due to unavailability of propagules, an important source of photosynthetically-fixed energy would have been unavailable to the Miyakejima system. This would have impeded the development of a decomposer food chain, of mineral weathering by organic acids and chelates, and of buffering of soil pH and water content and nutrient storage on organic matter and clay surfaces.

Was there a lower taxonomic diversity than is normally found in a developed soil? At the broadest level, from the metagenomes, the taxonomic composition of the biota of the developing ash-soil resembled that of the developed forest soils—with archaea, fungi, protists and metazoans occurring at similar relative

abundances in the ash-soils and forest soils. In this sense, the ash-soil had developed remarkably quickly in terms of the basic taxonomic framework of a functioning soil ecosystem. Bryophytes and lichens were able to establish^{10,26} with their propagules small enough to enter through the gauze covering, and presumably played an important role in providing carbon and extra niches to the developing ecosystem—even though they are unlikely to be as supportive of a diverse soil biota as the extensive root systems of vascular plants.

There was however—continuing a pattern seen in the 24-month results reported by Kerfahi et al.¹⁰—a lower OTU and Shannon diversity of bacteria based on the 16S rRNA amplicon sequencing. The continuing lower OTU-level 16S taxonomic diversity of the ash soils might have had multiple causes.

Firstly, limitations on dispersal and colonization of the ash soil systems might have kept OTU diversity lower. It is assumed that the biota present were mostly derived from the surrounding forest ecosystems—as windblown or rain splashed material. Dispersal limitation of soil biota is likely to play some role in all volcanic ash fields, although the existence of nearby areas of surviving ecosystem is also common^{20,27}. Most volcanic explosions or ash fields result in no more than a few square kilometres of landscape being completely devastated, and isolated pockets of vegetation that survive are common^{27,28}—in this respect the proximity of our mesocosms to natural vegetation may be fairly realistic. In natural ash deposits, upwards colonization by soil biota from buried ‘legacy soils’ is also possible²⁰. In our experiment, upwards colonization from soil below the pots was not possible, and the plastic gauze covering may have prevented access by insects, birds or seeds that could have brought incidental soil biota. It is thus difficult to know for sure whether the overall role of dispersal limitation in our mesocosm systems is stronger or weaker than would generally be the case in natural ash fields. Since natural ash fields themselves are also very heterogenous in terms of area, depth and degree of devastation of natural vegetation, it is especially hard to generalise.

Secondly, the relatively extreme chemical and physical conditions of the volcanic ash itself are also likely play a major role in limiting the bacterial taxonomic diversity of the ash-soil biota in the mesocosms. Although the pH of the ash in our mesocosm systems was not extreme, it may be expected to be droughty due to low organic matter content, and low in available nutrients and organic matter that could sustain soil food chains. Even by 36 months, the organic carbon of the ash-soil was still orders of magnitude lower than in the surrounding forest soils (Supplementary Fig. S1)^{10,26}. Thus, fewer niche types may be viable in the environment of the developing ash.

Nevertheless, although lower level taxonomic diversity of bacteria is clearly lower than in the established forest soils, the metagenome results from the ash soil mesocosms show that most higher level taxonomic groups of bacteria, archaea and eukaryotes are already present by the 24-month and 36-month stage, emphasizing that in some respects the establishment of soil biota has been rapid. While DNA of dead microorganisms blown as dust could have provided the impression of populations being present, it seems unlikely that dead material would be able to disperse in the same relative proportions as in a forest soil, to give a metagenome that resembled a developed soil in all its major groups, in roughly the same relative abundances (Please refer to Fig. 1 in Kerfahi et al.¹⁰).

Was there a lower diversity of categories of functional genes of soil biota than are normally found in a developed soil, indicating less functional complexity in the early ash-soil system?

Surprisingly, the richness and diversity of functional gene categories at Level 4 of the SEED Subsystem was not significantly different between the ash mesocosms and the forest soils. Around 97% of the metagenome reads in the ash mesocosms and the forest soil were bacterial, so the much lower OTU diversity of bacteria in the mesocosms would be expected to give a lower functional diversity of genes. The contrast between patterns of taxonomic and functional diversity in the ash mesocosms emphasises the redundancy of gene functions in soil organisms—such that losing a high proportion of taxonomic diversity apparently has no effect on functional diversity. If the level of functional gene diversity is high, this implies that in a general sense the ecosystem can potentially perform a wide variety of functions. Greater functional diversity is also considered to result in greater resilience of the ecosystem²⁹.

We had also hypothesized that the distinct chemical environment of the developing ash would result in increased relative abundance of a number of specific gene categories:

Is there an increased relative abundance of genes (or of prokaryotic genera) associated with autotrophy (e.g. *Rubisco gene*, *coxL gene*) in the ash-soils?

We searched for potential taxa and genes which might indicate autotrophic carbon fixation through chemosynthetic oxidation processes. We found a higher relative abundance of Ktedonobacteraceae (Chloroflexi) in the ash soils compared to forest soils, which is a pattern that was also found in previous ash studies^{21,30}. Ktedonobacteraceae is a novel taxonomic group that has only been recently added to the phylogenetic tree of Chloroflexi^{30,31}. Their genome has been reported to be large and they are known to have a wide potential for different metabolic mechanisms³². Ktedonobacteraceae are mostly found in extreme environments, including volcanic ash and hydrothermal vents³⁰.

We had hypothesized that the *coxL* (carbon monoxide dehydrogenase large chain) gene—which is implicated in autotrophy—would be more abundant in the ash soil samples. Likewise, the *rbcl* (ribulose biphosphate carboxylase large chain) gene is involved in carbon fixation and we hypothesized its greater abundance in the ash-soil mesocosms. However, in fact both *rbcl* and *coxL* had higher relative abundances in the forest soils.

Although *coxL* has been mostly linked with chemotrophic organisms living in an extreme environment, due to its potential role for assisting heterotrophic growth in the environments lacking organic matter^{21,33}, it is also one of the more abundant genes in natural forest soils³⁴. Noting that natural forest is one of the largest global sinks of atmospheric CO₂^{34,35}, the high relative abundance of *coxL* gene might be due to high abundance of CO in natural forests. Also, the *coxL* gene might have the potential to participate in other metabolic pathways. The presence of *coxL* genes in many different types of organisms (e.g. plant symbionts, animal pathogens, etc.) suggests other roles of CO-oxidation by the *coxL* gene³³.

The higher abundance of *rbcL* in the forest soils in our study might be related to photosynthetic C fixation by cyanobacteria in both the forest soil and ash soils. *rbcL* gene also have been found in many natural forest sites. By contrast, Fujimura et al.³ found *rbcL* to be more abundant in developing volcanic ash soils, which might be attributable to differences in the chemical composition of ash or due to details of the mesocosm system used in our study.

Is there increased relative abundance of genes that are associated with acquisition of nutrients from abiotic sources rather than decomposition of organic matter (e.g. nitrogen fixation genes)? Contrary to expectations we found no differences in the relative abundance of nitrogen fixation gene (*nifH*) between the ash soils and forest soils. It is possible that this reflects nitrogen being relatively more limiting in the forest soils, where there is an abundance of organic carbon. It is also possible that in the forest soils, the abundance of labile carbon in the rhizosphere of N-limited plants encourages N fixing bacteria through root secretions³⁶.

Is there increased relative abundance of stress response genes and dormancy related genes? As hypothesized, we found that dormancy related genes were more abundant in the ash mesocosms than in the forest soils. This would seem to be adaptive in terms of survival of a well-drained ash-soil, low in organic matter, which is apt to dry out. This would depend, however, on the detailed microclimate. Field measurements suggested that in sunny conditions the microclimate in the static air in the trays in which our pots were held was generally around 0.5–1 °C higher than in the open air a meter away, although this difference disappeared in cloudy conditions and at night¹⁰. It is possible that our mesocosms (exposed in open sunlight, but covered by a gauze) would have either retained or lost moisture more easily than the soils, and that this might have affected the frequency of drying. We had also anticipated that stress response genes would be more common in the ash mesocosms. However, in this case the stress response category was less common than in the forest soil, surprisingly suggesting that at a cellular level stresses are in fact less common in the ash-soil.

Is there decreased relative abundance of cell–cell interaction related genes? As anticipated, genes related to cell–cell interactions such as regulation and cell signalling genes and virulence, disease, and defense genes (Fig. 7a,b) were relatively less abundant in the ash mesocosms. This seems to emphasize that complex interactions are less common in the developing ash system than in a fully developed soil with its more abundant and diverse carbon sources. However, bacterial genes found as part of CRISPRs were relatively less abundant in the forests, implying lower intensity of defense mechanisms of bacteria against viruses in the forests compared to ash soils. The relatively low OTU diversity of the ash soils may perhaps allow greater spread and mortality of viruses through populations, necessitating greater investment in defences.

Since Odum³⁷, it has been noted that developing ecosystems tend to be based on tighter and more effective interactions as succession continues. For example, Morriën et al.³⁸ found that during secondary succession of old field ecosystems, co-occurrences of taxa become more predictable and carbon cycling in the decomposer chain more efficient. Empirically our findings seem to agree with the same paradigm, as we found cell–cell interaction genes found to be more abundant in the developed forest soils. It would be very interesting to directly measure decomposer carbon cycling efficiency in ash-soil mesocosm systems as they develop.

Although we focused on 16S rRNA amplicon sequence data for comparing OTU level diversity between samples as taxonomic assignment based on MG-RAST pipeline is very limited^{39,40}, we had briefly compared the family level bacterial community composition assigned based on metagenome sequence data and based on 16S rRNA amplicon sequence data. It is still debatable if whole genome sequencing, which is free from primer bias, is superior in this respect to 16S rRNA amplicon sequencing⁴⁰. Our data supports the idea that at the bacterial family level, metagenome sequencing covered a larger variety of taxonomic groups than 16S rRNA sequencing.

Conclusions

Overall, the biota of the 2- and 3-year old ash-soil mesocosms resembled that of forest soils more closely than we had expected, even though the total carbon and the total nitrogen concentration (and microbial biomass¹⁰) was much lower than in a developed forest soil. While bacterial diversity and OTU composition in our mesocosms was very distinct from the forest soils, a wide range of higher level taxonomic groups of prokaryotes and eukaryotes were present. The gene function profile of our ash-soils broadly resembled the B layer of a 'normal' temperate forest soil. We had expected to find evidence for a much greater role of chemosynthetic carbon fixation in the ash-soil, but the gene composition and taxonomic composition did not suggest this. However, it is possible that amongst the abundant unclassified prokaryote OTUs in our samples, such activity is actually important.

Bacterial functional gene diversity in the ash-soil mesocosms was also surprisingly similar to that of the forest soils, suggesting that relatively low bacterial taxonomic diversity has little impact on the functional potential of the soil due to redundancy of genes.

Perhaps most importantly in determining the path of development of this system, bryophytes, lichens and cyanobacteria had already begun to establish on the surface of the ash soils by the 24-month stage and were further developed by the 36-month stage^{10,26}, and this may be seen as providing the starting point for a rapid positive feedback process in ecosystem development. Also, the relatively moderate pH of the Sakurajima ash may have contributed to the rapid development of the ash system.

It would be interesting to follow the trajectory of developing ash-soil mesocosms over further years to understand how long they take to more closely resemble a developed soil ecosystem. However, it is important to bear in mind that our mesocosm systems are in certain ways limited in simulating natural volcanic ash soil biota succession. The proximity of a developed ecosystem surrounding the mesocosms may mean there was a ready

source of dispersal of propagules of microbes and small soil animals, and the spores of bryophytes, which once established are likely a major source of photosynthetic carbon. Nevertheless, studies of volcanic landscape succession have often shown that bryophyte and lichen spores are capable of long distance dispersal on the wind^{17–19}. Unlike a ‘real’ volcanic ash field, our mesocosms have also been relatively disadvantaged in terms of arrivals, being cut off by the mesh covering from wind dispersed seeds, insects, and from bird droppings which could bring vascular plants, nutrients, and soil microbes. Additionally, our pots could not readily be colonized from below, whereas buried legacy soils may be important in many natural volcanic ash fields²⁰. In ecology, there is an almost inevitable trade-off between control of the experimental conditions, and the complex and confusing realism of a natural ecosystem. Both approaches are important and complimentary, and there is a need for more field based studies—ideally beginning as soon as possible after the eruption itself—to understand the development of volcanic ash into soil ecosystems in a more natural context.

Materials and methods

Source of volcanic ash. The volcanic ash used in this study was derived from the Sakurajima volcano, on the southern island Kyushu in Japan (31° 34′ 50″ N 130° 39′ 29″ E). Sakurajima is an andesitic stratovolcano, which in recent decades has been in a continuous state of low level eruption showering ash on its surrounding slopes and for several kilometres beyond⁴¹. Being andesitic, the ash from Sakurajima is rich in silica and low in calcium and magnesium. Sakurajima ash does not usually contain any large amounts of acids such as sulphuric acid, hydrochloric acid and hydrofluoric acid, although acidic rainfall events around the volcano have sometimes occurred⁴². At time of our sampling, the freshly deposited volcanic ash had a pH of about 5.2, no more acidic than many soils in southern Japan.

Setting up microcosms, soil sampling, and DNA extraction. After several days of eruption, freshly fallen ash was collected on plastic sheets near the Sakurajima volcano, and sterilized at 200 °C in dry heat in an oven for an hour. We set up several groups of pot mesocosms of this ash at various locations and climates around Japan, in clearings in natural woodland or scrub vegetation (methods described in detail in Kerfahi et al.¹⁰). Briefly, these locations were (a) the lower slopes of the Sakurajima volcano, is a sparsely wooded area of young pines (*Pinus* sp.) with large openings of bare ash and herbaceous vegetation (mainly *Fallopia* sp.) (b) Takakuma, mixed evergreen broadleaf hill forest some 10 km from Sakurajima, (c) Kamigamo, mixed evergreen broadleaf forest some 600 km away near Kyoto, (d) two sites on Mt Norikura, 800 km northeast of Sakurajima. One site was at 1450 m asl. (above sea level) (referred to as “Norikura_Low” in the manuscript) surrounded by cool climate deciduous broadleaves and conifers (mainly *Larix* sp.), the second site was at 2800 m asl. (referred to as “Norikura_High” in the manuscript), surrounded by open dwarf pine (*Pinus pumila*) scrub with many bare areas of loose gravel and scree.

Pots of the sterilized ash were placed in replicate plastic trays (drained by holes, several pots per tray for repeat sampling over several years), the trays placed 5 m apart outside and each tray covered by 1 mm hole size plastic gauze, as described in Kerfahi et al.¹⁰. Replicates of the pots were sampled from their respective trays at 24 months (2 years) and 36 months (3 years) after being set up. The top 2 cm of the ash in each pot were used for DNA extraction and analysis. To compare functional gene composition of ash soils with natural vegetated soils, we collected one to four cores (same width as the pots), each 5 m apart from one another of the top 2 cm of soil from the adjacent natural ‘forest’ (in some cases more resembling scrub) vegetation, <200 m from each experimental site. Soil DNA was extracted using the Power Soil DNA extraction kit (MO BIO Laboratories, Carlsbad, CA, USA) and kept at –20 °C for further processing.

Year 1 samples (12 months) did not yield enough DNA for amplicon or shotgun sequencing. Bacterial community data based on the 16S rRNA sequencing for Year 2 (24 months) are already published as Kerfahi et al.¹⁰. For a more complete perspective, we added another year’s set of samples for 16S community data, as well as shotgun sequencing the 24 month and 36 month samples for metagenomes. To compare bacterial communities with natural forest soils, we obtained publicly available sequence data from other studies on the same site^{10,43}. Details of the samples can be found in Supplementary Table S3 for 16S rRNA data and Supplementary Table S4 for metagenome data.

Soil physicochemical analysis. To determine total carbon and total nitrogen concentration of samples, an NC analyzer (NC-900; Shimadzu, Kyoto, Japan) was used. pH was measured using pH meters (D-51, Horiba, Kyoto, Japan) after extraction from 10 g of dry ash with 25 mL of ion exchange water^{10,26}.

PCR and sequencing. To amplify 16S rRNA gene from soils, universal bacterial primers 341F (5′-CCT AGGGGNGGCWGCAG-3′) and 805R (5′-GACTACHVGGGTATCTAATCC-3′) were used. Polymerase chain reaction were performed under the following conditions: initial denaturation (95 °C, 3 min) followed by 25 cycles of denaturation (95 °C, 30 s), annealing (55 °C, 30 s), and elongation (72 °C, 30 s), and final extension (72 °C, 5 min). After purifying PCR products using AMPure XP beads (Beckman Coulter, Inc., Brea, CA, USA), index PCR was performed under following conditions: initial denaturation (95 °C, 3 min) followed by 8 cycles of denaturation (95 °C, 30 s), annealing (55 °C, 30 s), and elongation (72 °C, 30 s), and final extension (72 °C, 5 min). PCR products were purified and pooled together and sequenced under Illumina MiSeq platform (Illumina, Inc.). Whole genome shotgun sequencing was carried out by Celemics (Celemics, Inc., Seoul, Korea), using an Illumina NextSeq sequencers (Illumina, Inc.).

16S rRNA amplicon sequence processing. Paired end sequences were joined together using pandaseq v.2.8.⁴⁴. Sequences were further processed using Mothur v. 1.42.3.⁴⁵ following the MiSeq SOP (<https://www.illumina.com>).

[mothur.org/wiki/MiSeq_SOP](https://www.mothur.org/wiki/MiSeq_SOP)). Sequences were aligned and classified based on Silva database v.132.⁴⁶ Since natural forest sample DNAs from Kerfahi et al.¹⁰ have been amplified using different primer sets from our study and Cho et al.⁴³ (Supplementary Table S3), we used only the commonly aligned nucleotide sequences of each read. 16S rRNA sequence reads from this study were deposited in MG-RAST under the project ID of X. Operational taxonomic units (OTUs) were assigned based on 97% similarity. Singleton sequences were removed. The OTU table was subsampled into 10,020 reads per samples for diversity analysis.

Phylogenetic analysis. Maximum-likelihood tree was generated using FastTree v.2.1.3⁴⁷ based on representative OTU sequences. For evaluation of phylogenetic community assembly processes, standardized effect size measure of the mean nearest taxon distances (SES.MNTD)^{48,49} with 999 randomized null modes were calculated using “ses.mntd” in R package “picante”⁵⁰.

Whole genome shotgun sequence processing. Metagenome sequences were uploaded to MG-RAST under project IDs of mgp92027, mgp87971, and mgp85837 for further processing. Sequence reads with > 5 low quality base pairs (< 15 phred score) were removed. For taxonomic annotation of sequence reads, the RefSeq database⁵¹ was used and for functional gene annotation, the SEED database²⁴ was used. Sequences were annotated using default settings of MG-RAST (maximum e-value cutoff of 10⁻⁵, minimum % identity cutoff of 60%, and minimum alignment length cutoff of 15 bp). The Subsystem Function level (the finest level) data was subsampled to 160,910 reads per sample for diversity analysis.

Statistical analysis. To visualize ecological distance between each sample set, OTU table and level 4 functional gene table were square-root transformed and Bray–Curtis dissimilarities between samples were calculated. nMDS plots were drawn and statistical significance of the effect of sites and time were tested using two way crossed ANOSIM test in PRIMER software v.6 (<http://www.primers-e.com>) with 999 permutations⁵². To compare the Bray–Curtis dissimilarity between the 2nd-year ash soils and forest soils in comparison with the Bray–Curtis dissimilarity between the 3rd-year ash soils and forest soils, we performed t test. P values were adjusted using the Benjamini and Hochberg’s method⁵³. To compare soil physicochemical conditions, relative abundance of bacterial taxa, relative abundance of functional genes, and bacterial taxonomic and functional gene diversity between the ash soil and natural forest, ANOVA test was performed. Where the results are significant, Tukey’s HSD test was performed as a posthoc test. For soil physicochemical factors, conditions for ANOVA tests were in some cases not formally met, but given the very clear distinct differences in C and N (ash soils consistently have only 1/10 to 1/100 as much C and N as the forest soils) we present the results here to provide further evidence for the patterns shown.

Data availability

The datasets generated during the current study are available in the metagenomics RAST (rapid annotation using subsystems technology) server under the project numbers mgp92027, mgp87971, and mgp85837. For further information on MG-RAST sample accession numbers, refer to Supplementary Table S3 and S4. Authors can confirm that all relevant data are included in the article and/or its supplementary information files.

Received: 11 June 2020; Accepted: 29 October 2020

Published online: 08 December 2020

References

- Ooi, N. Palynological study of the peat sediments and around the latest glacial maximum at Hikone, the east shore of Lake Biwa, Japan. *J. Phytogeogr. Taxon.* **37**, 37–42 (1989).
- Oor, N., Minaki, M. & Noshiro, S. Vegetation changes around the last glacial maximum and effects of the Aira-Tn ash, at the Itai-Teragatani site, central Japan. *Ecol. Res.* **5**, 81–91. <https://doi.org/10.1007/bf02348465> (1990).
- Fujimura, R. *et al.* Unique pioneer microbial communities exposed to volcanic sulfur dioxide. *Sci. Rep.* **6**, 19687–19687. <https://doi.org/10.1038/srep19687> (2016).
- Cui, J., Hirota, M., Kamiyo, T., Yoshitake, S. & Katoh, K. Soil net nitrogen mineralization at different ecosystem development stages after the year 2000 eruption on Miyakejima island. *J. Ecosys. Ecograph* **8**, 2 (2018).
- Mukai, M., Aiba, S.-I. & Kitayama, K. Soil-nutrient availability and the nutrient-use efficiencies of forests along an altitudinal gradient on Yakushima Island, Japan. *Ecol. Res.* **31**, 719–730. <https://doi.org/10.1007/s11284-016-1381-8> (2016).
- Stewart, C. *et al.* Assessment of leachable elements in volcanic ashfall: A review and evaluation of a standardized protocol for ash hazard characterization. *J. Volcanol. Geoth. Res.* **392**, 106756. <https://doi.org/10.1016/j.jvolgeores.2019.106756> (2020).
- Ugawa, S., Inagaki, Y., Karibu, F. & Tateno, R. Effects of soil compaction by a forestry machine and slash dispersal on soil N mineralization in Cryptomeria japonica plantations under high precipitation. *New Forest.* **51**, 887–907. <https://doi.org/10.1007/s11056-019-09768-z> (2020).
- Hernández, M., Dumont, M. G., Calabi, M., Basualto, D. & Conrad, R. Ammonia oxidizers are pioneer microorganisms in the colonization of new acidic volcanic soils from South of Chile. *Environ. Microbiol. Rep.* **6**, 70–79. <https://doi.org/10.1111/1758-2229.12109> (2014).
- Kelly, L. C., Cockell, C. S., Thorsteinsson, T., Marteinson, V. & Stevenson, J. Pioneer microbial communities of the Fimmvörðuháls lava flow, Eyjafjallajökull, Iceland. *Microbial. Ecol.* **68**, 504–518. <https://doi.org/10.1007/s00248-014-0432-3> (2014).
- Kerfahi, D. *et al.* Development of soil bacterial communities in volcanic ash microcosms in a range of climates. *Microb. Ecol.* **73**, 775–790. <https://doi.org/10.1007/s00248-016-0873-y> (2017).
- Lathifah, A. N. *et al.* Comparative characterization of bacterial communities in moss-covered and unvegetated volcanic deposits of Mount Merapi, Indonesia. *Microbes Environ.* **34**, 268–277. <https://doi.org/10.1264/jsme2.ME19041> (2019).
- Fujimura, R. *et al.* Analysis of early bacterial communities on volcanic deposits on the island of Miyake (Miyake-jima), Japan: A 6-year study at a fixed site. *Microbes Environ.* **27**, 19–29. <https://doi.org/10.1264/jsme2.ME11207> (2009).

13. Nanba, K., King, G. M. & Dunfield, K. Analysis of facultative lithotroph distribution and diversity on volcanic deposits by use of the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase. *Appl. Environ. Microbiol.* **70**, 2245–2253. <https://doi.org/10.1128/aem.70.4.2245-2253.2004> (2004).
14. Crews, T. E., Kurina, L. M. & Vitousek, P. M. Organic matter and nitrogen accumulation and nitrogen fixation during early ecosystem development in Hawaii. *Biogeochemistry* **52**, 259–279. <https://doi.org/10.1023/A:1006441726650> (2001).
15. Gomez-Alvarez, V., King, G. M. & Nüsslein, K. Comparative bacterial diversity in recent Hawaiian volcanic deposits of different ages. *FEMS Microbiol. Ecol.* **60**, 60–73. <https://doi.org/10.1111/j.1574-6941.2006.00253.x> (2007).
16. King, G. M. Contributions of atmospheric CO and hydrogen uptake to microbial dynamics on recent Hawaiian volcanic deposits. *Appl. Environ. Microbiol.* **69**, 4067–4075. <https://doi.org/10.1128/aem.69.7.4067-4075.2003> (2003).
17. Dale, V. H., Swanson, F. J. & Crisafulli, C. M. *Ecological Responses to the 1980 Eruption of Mount St. Helens* (Springer, Berlin, 2005).
18. Bishop, J. G. Early primary succession on Mount St. Helens: Impact of insect herbivores on colonizing lupines. *Ecology* **83**, 191–202. [https://doi.org/10.1890/0012-9658\(2002\)083\[0191:Epsoms\]2.0.Co;2](https://doi.org/10.1890/0012-9658(2002)083[0191:Epsoms]2.0.Co;2) (2002).
19. Tsuyuzaki, S. & Hase, A. Plant community dynamics on the volcano Mount Koma, northern Japan, after the 1996 eruption. *Folia Geobot.* **40**, 319–330. <https://doi.org/10.1007/BF02804282> (2005).
20. Zeglin, L. H., Wang, B., Waythomas, C., Rainey, F. & Talbot, S. L. Organic matter quantity and source affects microbial community structure and function following volcanic eruption on Kasatochi Island, Alaska. *Environ. Microbiol.* **18**, 146–158. <https://doi.org/10.1111/1462-2920.12924> (2016).
21. Hernández, M., Calabi, M., Conrad, R. & Dumont, M. G. Analysis of the microbial communities in soils of different ages following volcanic eruptions. *Pedosphere* **30**, 126–134. [https://doi.org/10.1016/S1002-0160\(19\)60823-4](https://doi.org/10.1016/S1002-0160(19)60823-4) (2020).
22. Sato, Y. *et al.* *Cupriavidus pinatubonensis* sp. nov. and *Cupriavidus laharis* sp. nov., novel hydrogen-oxidizing, facultatively chemolithotrophic bacteria isolated from volcanic mudflow deposits from Mt. Pinatubo in the Philippines. *Int. J. Syst. Evol. Microbiol.* **56**, 973–978. <https://doi.org/10.1099/ijs.0.63922-0> (2006).
23. Weber, C. F. & King, G. M. Volcanic soils as sources of novel CO-oxidizing paraburkholderia and burkholderia: *Paraburkholderia hiiakae* sp. nov., *Paraburkholderia metrosideri* sp. nov., *Paraburkholderia paradisi* sp. nov., *Paraburkholderia peleeae* sp. nov., and *Burkholderia alpina* sp. nov. a member of the *Burkholderia cepacia* complex. *Front. Microbiol.* **8**, 207. <https://doi.org/10.3389/fmicb.2017.00207> (2017).
24. Overbeek, R. *et al.* The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* **42**, D206–D214. <https://doi.org/10.1093/nar/gkt1226> (2014).
25. Sakamoto, H. Changes of water soluble constituents of volcanic ashes and volcanic activity at the Sakurajima Volcano, Kagoshima, Japan. *Nat. Kagoshima* **39**, 177–189 (2013).
26. Tateno, R. *et al.* Temperature effects on the first three years of soil ecosystem development on volcanic ash. *CATENA* **172**, 1–10. <https://doi.org/10.1016/j.catena.2018.08.009> (2019).
27. Moral, R. D., Wood, D. M. & Titus, J. H. In *Ecological Responses to the 1980 Eruption of Mount St. Helens* (eds Dale, V. H. *et al.*) 93–109 (Springer, New York, 2005).
28. Fuller, R. N. & del Moral, R. The role of refugia and dispersal in primary succession on Mount St. Helens, Washington. *J. Veg. Sci.* **14**, 637–644 (2003).
29. Fonseca, C. R. & Ganade, G. Species functional redundancy, random extinctions and the stability of ecosystems. *J. Ecol.* **89**, 118–125. <https://doi.org/10.1046/j.1365-2745.2001.00528.x> (2001).
30. Yabe, S., Sakai, Y., Abe, K. & Yokota, A. Diversity of Ktedonobacteria with actinomycetes-like morphology in terrestrial environments. *Microbes Environ.* **32**, 61–70. <https://doi.org/10.1264/jsme2.ME16144> (2017).
31. Cavaletti, L. *et al.* New lineage of filamentous, spore-forming, gram-positive bacteria from soil. *Appl. Environ. Microbiol.* **72**, 4360–4369. <https://doi.org/10.1128/aem.00132-06> (2006).
32. Zheng, Y. *et al.* Genome features and secondary metabolites biosynthetic potential of the class Ktedonobacteria. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2019.00893> (2019).
33. King, G. M. & Weber, C. F. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* **5**, 107–118. <https://doi.org/10.1038/nrmicro1595> (2007).
34. Lalonde, I. & Constant, P. Identification of unknown carboxydovore bacteria dominant in deciduous forest soil via succession of bacterial communities, *coxL* genotypes, and carbon monoxide oxidation activity in soil microcosms. *Appl. Environ. Microbiol.* **82**, 1324–1333. <https://doi.org/10.1128/aem.03595-15> (2016).
35. Badr, O. & Probert, S. D. Sinks and environmental impacts for atmospheric carbon monoxide. *Appl. Energy* **50**, 339–372. [https://doi.org/10.1016/0306-2619\(95\)98803-A](https://doi.org/10.1016/0306-2619(95)98803-A) (1995).
36. Fitzpatrick, C. R. *et al.* Assembly and ecological function of the root microbiome across angiosperm plant species. *Proc. Natl. Acad. Sci.* **115**, E1157–E1165. <https://doi.org/10.1073/pnas.1717617115> (2018).
37. Odum, E. P. The strategy of ecosystem development. *Science* **164**, 262–270. <https://doi.org/10.1126/science.164.3877.262> (1969).
38. Morriën, E. *et al.* Soil networks become more connected and take up more carbon as nature restoration progresses. *Nat. Commun.* **8**, 14349. <https://doi.org/10.1038/ncomms14349> (2017).
39. Randle-Boggis, R. J., Helgason, T., Sapp, M. & Ashton, P. D. Evaluating techniques for metagenome annotation using simulated sequence data. *FEMS Microbiol. Ecol.* <https://doi.org/10.1093/femsec/fiw095> (2016).
40. Brumfield, K. D., Huq, A., Colwell, R. R., Olds, J. L. & Leddy, M. B. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *PLoS One* **15**, e0228899. <https://doi.org/10.1371/journal.pone.0228899> (2020).
41. Kazahaya, R., Shinohara, H., Mori, T., Iguchi, M. & Yokoo, A. Pre-eruptive inflation caused by gas accumulation: Insight from detailed gas flux variation at Sakurajima volcano, Japan. *Geophys. Res. Lett.* **43**, 11219–11225. <https://doi.org/10.1002/2016gl070727> (2016).
42. Kawaratani, R. K. & Fujita, S.-I. Wet deposition of volcanic gases and ash in the vicinity of Mount Sakurajima. *Atmos. Environ. Part A. Gener. Top.* **24**, 1487–1492. [https://doi.org/10.1016/0960-1686\(90\)90057-T](https://doi.org/10.1016/0960-1686(90)90057-T) (1990).
43. Cho, H. *et al.* Soil pH rather than elevation determines bacterial phylogenetic community assembly on Mt. Norikura. *FEMS Microbiol. Ecol.* <https://doi.org/10.1093/femsec/fiy216> (2018).
44. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: Paired-end assembler for illumina sequences. *BMC Bioinform.* **13**, 31. <https://doi.org/10.1186/1471-2105-13-31> (2012).
45. Schloss, P. D. *et al.* Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541. <https://doi.org/10.1128/aem.01541-09> (2009).
46. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596. <https://doi.org/10.1093/nar/gks1219> (2012).
47. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490–e9490. <https://doi.org/10.1371/journal.pone.0009490> (2010).
48. Webb, C. O., Ackerly, D. D., McPeck, M. A. & Donoghue, M. J. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33**, 475–505. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150448> (2002).
49. Tripathi, B. M. *et al.* Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *ISME J.* **12**, 1072–1083. <https://doi.org/10.1038/s41396-018-0082-4> (2018).

50. Kembel, S. W. *et al.* Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464. <https://doi.org/10.1093/bioinformatics/btq166> (2010).
51. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**, D61–D65. <https://doi.org/10.1093/nar/gkl842> (2006).
52. Clarke, K. & Gorley, R. *PRIMER: Getting Started with v6* (PRIMER-E Ltd, Plymouth, 2005).
53. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300 (1995).

Author contributions

H.S.: coordination, methodology, metagenome interpretation, writing—original draft, visualization. D.K.: field sampling, DNA extraction. K.T.: field sampling. Soil surveys. S.L.N.: sequence analysis. B.M.T.: sequence analysis. H.K.: metagenome sequencing. R.T.: conceptualization, field sampling, soil surveys. J.A.: conceptualization, study monitoring, writing—original draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-78413-z>.

Correspondence and requests for materials should be addressed to R.T. or J.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020