TESIS DOCTORAL

# DIVERSIDAD METAGENÓMICA DE BACTERIAS EXTREMÓFILAS EN VOLCANES ANDINOS CAYAMBE Y SUMACO DEL ECUADOR

PhD THESIS

# EXTREMOPHILE BACTERIAL METAGENOMIC DIVERSITY IN ANDEAN VOLCANOES CAYAMBE AND SUMACO OF ECUADOR

AUTORA

MAGDALENA DE LOS ANGELES DIAZ ALTAMIRANO

DIRECTORES

FELIX J. SANGARI

CARLOS ALFONSO MOLINA

UNIVERSIDAD DE CANTABRIA

Escuela de Doctorado de la Universidad de Cantabria

Santander 2023

UNIVERSIDAD DE CANTABRIA

ESCUELA DE DOCTORADO DE LA UNIVERSIDAD DE CANTABRIA


# TESIS DOCTORAL

# DIVERSIDAD METAGENÓMICA DE BACTERIAS EXTREMÓFILAS EN VOLCANES ANDINOS CAYAMBE Y SUMACO DEL ECUADOR

# PhD THESIS

# EXTREMOPHILE BACTERIAL METAGENOMIC DIVERSITY IN ANDEAN VOLCANOES CAYAMBE AND SUMACO OF ECUADOR

Realizada por: Magdalena de los Ángeles Díaz Altamirano.

Dirigida por: Félix J. Sangari García

Codirigida por: C. Alfonso Molina.

Santander 2023

UNIVERSIDAD DE CANTABRIA

ESCUELA DE DOCTORADO DE LA UNIVERSIDAD DE CANTABRIA

# TESIS DOCTORAL

## DIVERSIDAD METAGENÓMICA DE BACTERIAS EXTREMÓFILAS EN VOLCANES ANDINOS CAYAMBE Y SUMACO DEL ECUADOR

# PhD THESIS

## EXTREMOPHILE BACTERIAL METAGENOMIC DIVERSITY IN ANDEAN VOLCANOES CAYAMBE AND SUMACO OF ECUADOR


Realizada por:

**Magdalena de los Ángeles Díaz Altamirano**

Dirigida por:

**Félix J. Sangari García**

Codirigida por:

**C. Alfonso Molina**


Febrero 2023

*” Una vez en la cima siempre regresarás a ver el camino que costó, irás por la tierra con la vista hacia el cielo, pensando que lo mejor fue la subida y que siempre desearás volver a hacerlo".*

*Magdalena*

*Dedicado a*:

*A Daniela Valentina y Andrea Isabella*
*Por su tierno y gran amor*
*Por ser la inspiración y motivación para seguir*
*Por ser parte de mí en cada paso, desde el más pequeño hasta el más grande*

*A Carlos*
*Por amarme y formar parte del proyecto de vida*
*Por permitir siempre, que aprenda más a su lado*
*Por todo su esfuerzo para que pudiese conseguir esta meta*

*A mis padres y hermanos*
*Por su infinito e incomparable amor*
*Por su entera confianza y apoyo incondicional para lo que sea.*

***Crónica de ascenso al volcán:***

*"Los vi y estaban ahí, imponentes y atemorizantes con una fuerza de atracción que me incitaba a subir, el majestuoso Cayambe, blanco el más blanco que he podido ver, y el majestuoso Sumaco místico y peligroso a la vez, olía a selva pura, frondosa vegetación, sonaba a la nada y al todo, al mismo tiempo."*

*Muy grande fue la ilusión de subir y coronarlos pensaba en que era el único momento, que debía decidirlo y lo hice, al iniciar el ascenso al Cayambe, un gran APU en forma de toro apareció, era tan negro e imponente que contrastaba con el radiante blanco del glaciar, me reto a pedir permiso para subir, sus grandes ojos no dejaban de mirarme fijamente e impedían que pase, sutilmente baje la ventana del cuatro por cuatro en el que subíamos y le dije, por favor me deja pasar y, lo autorizó. Gracias al gran creador de la naturaleza y sus obras maravillosas de las que solo formamos una mínima parte, por su grandeza y su bondad. Seguí mi ruta hacia la cumbre, el agotamiento empieza a aumentar, la falta de oxígeno te hace sentir la fragilidad de la vida y el inmenso motor que te mueve a pensar que puedes morir, te hace reaccionar y seguir, cada paso a 5200 msnm, cargada tan solo una mochila vacía, es uno de los mayores esfuerzos físicos que he hecho y representaba el gran primer paso que se debía dar. Pero gran paso tuve que dar para subir al Sumaco, cuando en medio de la densa selva donde solo el inicio, es un camino y luego ya no ves por donde ir, solo seguir era la meta, cuantas veces me caí de cara a un charco lleno de lodo, me hundía en él y me levantaba constantemente, sin fuerza, todo en mí, era puro barro y agua lodo, pensaba en la lección y el concepto de vida que me dejaba saber el ascenso al Sumaco, no podía más, quería quedarme ahí donde estaba, no avanzaba más, tenía muchos rasguños de las ramas y tierra en mis ojos, no lograba ver nada, todos mis músculos se paralizaban, solo el llanto de no poder más, me limpió los ojos, pero de repente llegué al mayor descanso y me quedé dormida. Un raro y fuerte sonido de la frondosidad de la selva, las hojas y los árboles me despertó, estaba completamente sola, me acerqué a la ventana del refugio y lo vi, un gigante oso que usaba anteojos, solo paseaba por allí, pero algo lo atraía, fue la fuerza que emitía lo que me sedujo y me invito a pensar, que nada me pasaría si él estaba allí, gracias por aparecer en mi vida y darme tu amor, tu cariño y por cuidar de las dos pequeñas ozesnas que te seguían, solo buscaban comida, me atreví a mirarlas y sentí el incomparable amor de una madre por sus crías, por conseguirles comida y protección y por la forma en que les enseñaba a ganársela. Gracias a esas ozesnas por seguirme ciegamente, con ese puro amor tan natural e inocente que solamente, te sigue a dónde vas sin cuestionar nada. El ruido de la selva Amazónica y sus paisajes me abrumaba, un nuevo día, allí, luego de haber sentido el cambio de temperatura de calor, al frío páramo andino, donde el agua de la gran laguna era helada, y me calaba, repentinamente escuché dos grandes lobos que aullaban a lo lejos, aunque no los veía, los sentía muy cerca de mí, como siempre, me sentía agradecida por su cuidado y su guía, cerca de ellos, una figura se dibujaba en el agua, formaba círculos grandes y pequeños daba vueltas y desaparecía, era un gran cocodrilo que se acercaba y se alejaba, su poderío y gran desconfianza lo mantenía alerta, el gran amor por su cría lo vulneraba, pero estuvo siempre ahí, me recordaba la importancia de la desconfianza y a la vez de la lealtad. Un pequeño conejito sorpresa apareció y me siguió toda la travesía, gracias a Uds., por ese infinito amor y por su incondicionalidad para acompañarme, por seguir ahí siempre y por lograr que me mueva.*
*En el Cayambe seguía subiendo y seguía decayendo, aún faltaba mucho,aún me veía en el medio de la nada, siempre me preguntaba la definición de la nada, y es eso, estar muy cerca del cielo donde sólo ves blanco a cada lado, blanco en el suelo y blanco sobre ti, esa sensación de ver la nada a tu alrededor, como si dejaras de existir por momentos, de pronto el poderoso vuelo de un Cóndor y su pasar tan cerca de ti, te recuerda que aun estás vivo y puedes ser fácil presa de*

*alguien o algo más grande, que te puede dañar pero que a la vez te reta a conocerlo, la imponencia de sus alas extendidas me hizo sentir esa gran curiosidad por volar, gracias a tu vuelo Aucuntur mallcu por motivarme a buscar nuevos horizontes y ver cosas más grandes que siempre me las contabas y las danzabas.*

*Una veloz y muy colorida Lizarda estaba posada en la esquina de madera del refugio, era demasiado rápida para mis ojos, me recordaba la astucia y la velocidad con la que hay que actuar frente a situaciones de peligro y frente a los que gustan de hacer daño. Gracias a ti por todas tus advertencias y cuidados. Un pato y su familia a la vez paseaban en el charco, me indicaron lo importante de ser unidos, la protección y el cuidado que debemos lograr a pesar de las adversidades, gracias por propiciar una lección de familia unida y dejar saber que eso te mantiene a flote.*

*Observaba fijamente como una colonia de hormigas trazaba una hilera ordenada que subía con mucha carga y me dejaba ver la importancia del trabajo organizado y en equipo, un hormiga cargaba mucho peso sobre sí, bajar el Sumaco era casi tan difícil como subirlo, pero se denotaba su esfuerzo por dejar una huella sobre la tierra, por ser engreído y soberano a la hora de trabajar y decidir, humilde y sencillo a la hora de compartir, gracias por todo el apoyo y por marcarme la hilera a seguir desde una perspectiva de crítica, son tus consejos y fuertes acusaciones, lo que permitió ver mis constantes errores, hizo que analizara las cosas desde otra perspectiva y deje la comodidad para levantarme y trabajar por los demás, porque aprendí que esforzarse por impulsar el desarrollo de otros, te llena de mayor satisfacción y te hace sentir más grande.*

*¡¡¡Mira Maggy!!!, ya falta poco, ya estás aquí, es el paraíso, gritaba la cabra andina no te dejes, faltaba ya tan solo tres o cuatro pasos para llegar, habíamos atravesado por tantos obstáculos y al final me esperaba la meta, no daba ni un paso más, esa sensación donde ya tu cuerpo no responde, a pesar de que tu mente lo desea, esa fragilidad de perder tu control físico, esa misma angustia me hizo llegar a la cima. Gracias a toda mi familia y amigos que me acompañaron a escalar, que me hicieron ver que lo más importante fue la subida y todo lo que vivimos en ella.*

***¡¡Gracias de todo corazón!!***

*A Alfonso por motivarme a tener curiosidad en el camino de la investigación, por aceptar el reto de ser mi director, por enseñarme la exigencia y la perseverancia que se debe tener en un doctorado y por dejarme conocer tus sabios consejos y sugerencias. Gracias a Félix por tu apoyo incondicional como tutor y como persona, por permitir que comparta con todo tu grupo en el IBBTEC, porque a pesar de todos los malos ratos, estuviste siempre con una cara feliz para mí.*

*Gracias mil al gran Pablo Jarrín por tu poderío amigo y por generarme valor, por tu bondad y por tu criterio a la hora de tomar decisiones, al gran Daniel Quiroz por ser la fuerza y la confianza plasmadas en la juventud, las que permitieron que haga clic en una tecla, sin temor a echarlo a perder todo. Gracias Dany por tu ayuda y por apoyarme, por tu infinita paciencia. Gracias al gran amigo Sevillano Pablo Castillejo por tu alegría y disfrute en las salidas de campo. Muchas gracias a Spiros Agathos por ser tan especial, por acogerme y hacer posible mi estancia en Bélgica, por su don, de ser humano, por todo su apoyo y por compartir su gran experiencia.*

*Gracias a Juanma, el más temerario de todos los investigadores que conozco, por su tenacidad y por hacerme llorar al principio y al final, por enseñarme que no todo es sencillo que requiere de esfuerzo y mucho sacrificio, por dejarme conocer la cara más fea del miedo. A Candela, Yelina, Matxalen, Santi, Mapi por sus sabias palabras y su compañerismo en el programa. A Júlia, por*

*acompañarme al final del camino por su cuidado y aprecio. A la Universidad de Cantabria y todo el IBBTEC por acogerme y dejarme sentir en un Centro de muy alto nivel de Investigación.*

*Gigas de gracias a mis grandes amigos bioinformáticos de Valencia, Vicente por dejarme saber que aún existen personas como tú, que el desinterés y la amabilidad en este mundo aún se pueden dar, y a Wladi por ser un gran maestro y por colaborarme en todo momento, pero sin Samuel y sin Pablo Monfort no hubiese logrado plasmar los datos, a Carlos por su gracia y su forma de llegar a las personas, gracias a todos Uds. por sus veloces clases y porque pude disfrutar del Máster de Bioinformática, conocer a quienes fueron mis compañeros de clase Ma. José, Germán, Julia, Irene, Verónica, Ernesto y más, porque en su momento supieron apoyarme. Gracias a todos Uds. conseguí entender el mundo de la Bioinformática, y a todo el I2SysBio y su gente, Marina y Ester y Chari, por acogerme en sus conversaciones y darme confianza, gracias, amigos valencianos por recibirme en su seno y permitir que forme parte de su diario vivir en el instituto.*

*A mis grandes amigas españolas Raquel, Ana Poveda, Ana Glez García, Lola y María que aparecieron en mi vida y que Dios las puso en mi camino como los ángeles de la guarda, porque en ausencia de mi madre, cada una de ellas tuvieron ese rol en mi vida, fuera de mi país y de mi entorno familiar, porque me acogieron en sus casas, me cuidaron y me compartieron su vida en España, me hicieron sentir en casa y que el amor no me faltara. Porque María a pesar de haberla encontrado casi al final se convirtió en mi Ada Madrina, todo lo que yo pedía, ella lo hacía posible, mil gracias, mi María por todo tu amor en Galizano y en Santander. Porque Lola me auxilio en Valencia y Pto Sagunto, me enseñaste la importancia del amor a tu pareja y la dicha de vivir feliz aun cuando todo puede estar en contra, gracias por tus grandes lecciones de vida Lolita querida. Porque mi gran amiga y compañera Ana Poveda supo confiar en mí y supo darme su amistad y su familia, sin fronteras, en Valencia y Pto Sagunto, gracias a tu madre, a Carlitos y Chloe que supieron cautivarme y hacerme feliz con su inocencia cuando más sola me sentí. Y, a Ana Glez en Santander que supo motivarme y alentarme siempre a seguir subiendo sin desmayar, siempre supo decirme "fenomenal y porque tú te lo mereces" cuando más oportuno era. Porque mi gran amiga Raquel y toda su linda familia, supiste ser la persona que te hace pensar que, en la vida todo da vueltas, y si haces el bien, lo recibirás cuando menos lo imagines, gracias, mi bella amiga por ti, por tu madre Pepita, que siempre me abrieron su casa en Madrid para conocerlo y saber que las tengo allí, como parte de mi familia.*

*Muchas gracias a mis viejas amigas de siempre Pili, Gabita, Marce, Susy, Silvita y Azu por apoyarme, por ser las más grandes amigas por perdurar y auspiciar mis logros, por sus frases de aliento y sus lindos consejos, por su compañía lejos de casa a través de su corazón.*

*Mil gracias a mis queridos y muy estimados ayudantes y colaboradores, Dany Cuyo por tu desinterés y apoyo incondicional, por tu amistad, a Christian Cruz, Paolita y Elsy, a Sara, Cristina, Jefferson, Christofer, Jorge Echeverría, Steven, Álvaro, Emerson, Irvin, Mónica, Julissa, Carlita por ayudarme, por motivarme, por asistirme y auspiciar que pueda lograr esta meta desde su perspectiva de estudiantes y por su gran esfuerzo y joven amistad.*

*A la Universidad Central del Ecuador por su apoyo y manutención, por su poderío y exigencia académica que me ha hecho llegar hasta aquí, al Instituto de investigación en salud Pública y Zoonosis CIZ a Maritza, Erika Rivadeneira, Omar y Eli Yugsi que en su momento supieron enseñarme sus métodos y soporte, Sandrita, Frank, Eli , Richard, Lenin, Jimenita, Javier, Gus y todos los compañeros de investigación que me enseñaron a trabajar con minuciosidad y a sentir*

*que no siempre se tiene todo lo necesario, pero se lo puede resolver. A la Facultad de Ingeniería Química y sus autoridades, que han facilitado y apoyado a través de su gestión el culminar de este reto, Gracias Humberto, Diego, Morita, José, Edward, Paty, Gus, Jorge J, Merceditas y todos los compañeros y colaboradores de lucha diaria en la facultad, Gracias a Lucy Baldeón del INBIOMED por su apoyo, y a todos mis estudiantes, por ser el motor que me incita a la superación.*

*A CENTROCESAL, por el importante soporte técnico a la hora de los análisis, y el tratamiento del metadato, centro calificado y eficiente en sus resultados del cual aprendí mucho porque me permitió formar parte de su equipo.*

*A la Universidad Internacional SEK, por ser la promovedora de todo este proceso, a Katy Coral, y la familia de su hermano por su atención y cariño, gracias al mijo Jorge Oviedo, Anita Lucía Rodríguez por lo compartido en Santander con Uds. Conocí por primera vez esta hermosa ciudad que Katty acertadamente nos recomendó. Alonsito Moreta y Janeth García. Grandes compañeros y amigos de aquellos momentos de mucha alegría en la SEK.*

*Gracias a mi amado Ecuador país Mega maravilloso y su verdad, por hacerme sentir que el único camino a seguir para sacarlo adelante es el estudio. Un especial agradecimiento a la comunidad de Pacto Sumaco, por hacer posible mi asenso y toma de muestras, gracias, José Andi y Germánico por apoyar la subida, por guiarme y darme la mano para levantarme del charco y por ayudarme a cargar la mochila llena de muestras en el Sumaco. Al majestuoso Cayambe por su luz a pesar de que oscurecía, a Ramiro de Ecuadorean Alpine Institute por su experiencia en escalar y por su experticia para hacer la escuela de andinismo para subir a la cumbre, junto a él todos los amigos que decidieron acompañarme en los ascensos. Pablo Castillejo, Alfonso, Pablo Jarrín, Dany Quiroz, Christian Rochina, Erika Rivadeneira, Esteban, Isa y muchos más que estuvieron allí.*

*Gracias también debo dar a Alex y Unai de Pto Sagunto, Uds. supieron colaborarme y llenar un espacio en medio de tanta soledad. Muchas gracias a Henry Quiroz por su paciente ayuda y soporte. Infinitas Gracias a todas las personas, si acaso olvido a alguien, que me acompañaron en esta subida y que hicieron posible este camino, gracias al amor, a la amistad y al compañerismo, al trabajo en equipo, a la naturaleza y sus pequeños hijos Thomas, Teodoro, Pucho, Tadeo, Lena, Narco, Piro y Rita por increíble que parezca, ellos marcaron mi fuerza, el amor a la vida y a la naturaleza durante todas mis estancias.*

*Finalmente agradezco a toda España y su gente, en especial a Santander, Valencia y a este Doctorado por hacerme saber, qué es lo importante, lo prioritario y lo verdadero, que la verdad es cuestión de sobrevivencia, que nada es prioritario, pues a todo hay que dedicarle tiempo y que estar un paso más adelante de todos, es la clave., que al final de todo, sólo te tienes a ti mismo, sin embargo siempre se dará el momento, el lugar y la persona adecuados para ser feliz, no debemos olvidar de soñar, de hacer deporte, de leer , de la buena música, y de hacer amigos, esas son, las únicas herramientas necesarias para viajar y crecer, me enseñaron que suponer algo es un gran error, pero actuar en función de eso es aún peor.*

# INDEX

# INDEX

# OVERVIEW

## 1.1 Background

For over two decades, global warming has been considered a significant threat to vulnerable ecosystems, such as glaciers and ice-capped volcanoes, which are affected by sharp changes in temperature (Oerlemans, 1994). Global melting and glacier retreat is one primary effect of global warming (Raper & Braithwaite, 2006; Y. Shi & Liu, 2000; Sorg et al., 2015). The retreat of tropical Andean glaciers is considered a climate change indicator, mainly as glaciers are sensitive to climate perturbations (Rabatel et al., 2013a, 2018). Over the past forty years, a consistent retreat has been evident at various Andean glaciers (Małecki et al., 2018). It is, therefore important to understand glacier ecosystems in the Andes before their possible disappearance (Stibal et al., 2020).

Amazonian soils are a vast source of novel bacteria (Bruce et al., 2010). Despite widespread and incessant deforestation in favor of agricultural lands, there are yet relatively few studies in the Amazon region that have quantified microbial diversity and linked it to ecological roles and services, especially when compared to other groups of living organisms such as plants or animals (Ranjan et al., 2015). Soil health and structure, the flux of nutrients, and symbiotic relationships with plants depend on the microbial community in the ecosystems (Fonseca et al., 2018). Soil biogeochemistry is intimately connected to bacterial communities, and modifications to the original forest cover inexorably alter both the available nutrients and the bacterial communities present in the affected area (Merloti et al., 2019). Forest soils in the Amazon region are being primarily affected by anthropogenic activities and disturbances that are associated with an expanding agricultural frontier, which also affects bacterial diversity and the presence of functional microbial genes related to soil biogeochemistry (Melo et al., 2021) The transformation of forests into pastures or agricultural lands in the Amazon region increases pH, bacterial richness, and alpha diversity; it also alters community structure with yet unknown consequences for the functions and services of ecosystems (De Carvalho et al., 2016). Increased soil methane emissions and reduced carbon sinks are expected as Amazon forests are replaced by pastures (Kroeger et al., 2021) or warm-up

Along altitudinal gradients to highland environments, they are accelerating bacterial carbon metabolism (Looby and Martin, 2020). Consequences of deforestation are wide-ranging and with unknown or poorly understood effects on human health, as land conversion in the Amazon could also promote an increased frequency of antibiotic resistance genes in soil bacteria (Nascimento et al., 2021).

One of the objectives of environmental microbiology is to decipher and understand the complex terrestrial and aquatic ecosystems. In particular, the microbial assessment of extreme ecosystems. Highly adapted organisms inhabit the most adverse landscapes in this type of ecosystem. The adaptation to unfavorable habitats is highly dynamic and complex and often more than one evolutive fore is involved in this process. For example, in glaciers, temperatures below 0 ºC, high UV radiation, and source scarcity are challenging factors that microbes in these habitats must encounter successfully (Shivaji & Prakash, 2010). Different pathways have been discovered to overcome these evolutive forces. Phycrophylic microbes can reduce energy consumption and express high catalytic activity enzymes that can be applied in the industry (Kuddus & Ramteke, 2009). On the other side, we have microbes that are adapted to survive to high temperatures, thermophilic microorganisms. These microorganisms have gained much more attention due to their enzymes being structurally stable at high temperatures, at which different reactions are improved (Van den Burg, 2003). This introduction will cover the most important aspect of the two-study location, Cayambe glacier and the Sumaco volcano, and the technology employed to accomplish the objectives.

### 1.1.1 Cayambe Glacier

The Cayambe glacier is a massive active explosive volcanic center with a base extension of 24 × 18 Km. It rises to an altitude of 5790 meters above sea level, it is covered by a vast ice cap of nearly 22 km2, with an approximate volume of 0.7 km3, and with a thickness that reaches 100 m at places (Cours, Dernières, Et, & (Monzier et al., 1996)Samaniego, 1996; Guillier & Chatelain, 2006) (Figure. 1.1a). The Cayambe ice cap is constant above 4800 m.a.s.l. It descends to ~4600 meters above the sea level (m.a.s.l) on its western flank and ~4200 m.a.s.l on its eastern flank (Bax & Francesconi, 2019)(Detienne et al., 2017; Samaniego, Monzier, Robin, & Hall, 1998) (Figure 1.1a and 1.2c). The glacier retreat of the Cayambe has been estimated at 25.58% from 1979 to 2009 (Gallegos Castro et al., 2018) (Brito Chasiluisa, Serrano Giné, & Galárraga Sánchez, 2018).

The Cayambe glacier is unique in its geographical location, which is essentially at zero latitudes (0.03º N;77.988º W) with one natural lake called “Laguna Verde” (Figure 1.1d). This location adds Cayambe to the tropical glacier category. Tropical glaciers are particularly interesting due to their high sensibility to environmental variations, especially temperature increment (Rabatel et al., 2013b). Several studies point to faster glacier retreat for tropical glaciers compared with non-tropical glaciers. During the last 4000 years, the Cayambe glacier has experienced 21 volcanic eruptions, the most recent occurring in 1785–1786 (Samaniego et al., 1998). The glacier of the Cayambe also serves as a water source for communities downstream, including large cities such as Quito.

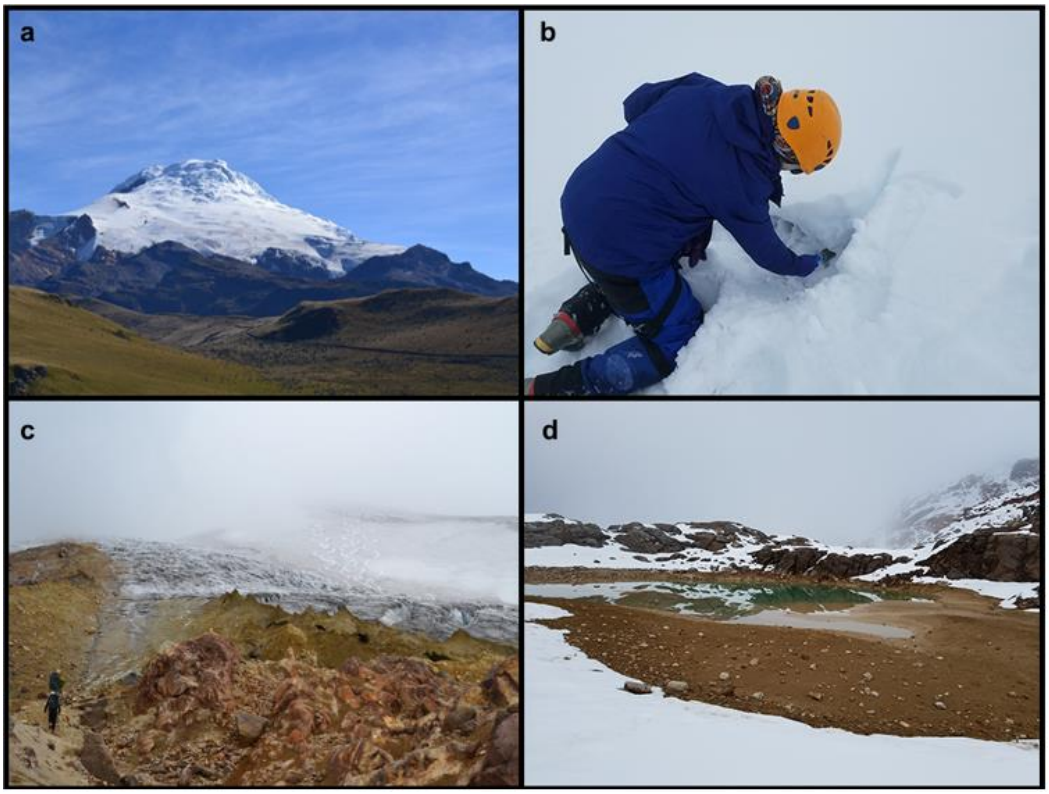



**Figure 1.1.** The Cayambe glacier. (a) A view of the western face of the glacier. (b) The sampling process. (c) Researchers on their way to the glacier. (d) A view of the lake called “Laguna Verde” (2015-2019).

Microbial communities should be perceived not only as the presence and interactions of microscopic living organisms but also as the biological matrix which plays a vital role in shaping ecosystems and communities of multicellular organisms (J. F. Stolz, 2017). Microbial communities at mountain glaciers are often first colonizers and critical players in soil formation, which enable subsequent processes of plant colonization and growth, the transformation of compounds, rock weathering, and nutrient enrichment of downstream ecosystems ,(Ragot et al., 2013) yet, it is unknown, particularly for the Andes, which are the consequences of rapid glacier melting, due to climate change, on the microbial communities and their ecological function (Ciccazzo et al., 2016).

Substantial amounts of biodiversity for multicellular organisms are well known in the tropical Andes (Bax and Francesconi, 2019); however, there are still few studies on microbial diversity in the region, particularly in glaciers and high-altitude mountain environments (Ciccazzo et al., 2016; Hotaling et al., 2017; Nayfach et al., 2020). Most of the studies of microbial communities at mountain glaciers come from the European Alps or the USA; this makes the need for information from the neotropical Andes indispensable for a broader vision of climate change effects and ecological processes on a global scale (Ciccazzo et al., 2016). A thorough assessment of microbial diversity in the Andes is crucial to establish the potential for further prospection into the use of psychrophilic microorganisms and derived bioproducts in microbial metabolism (Borda-Molina et al., 2017). Environmental services, as the result of bacterial metabolism, are also an important reason why we need to understand bacterial communities in these fragile and rapidly changing environments (Margesin et al., 2009).

Is necessary to explore the structure of bacterial communities in extreme ecosystems in the soil, water, and ice of the Cayambe glacier and provide insights into the relationships between the observed system and the associated physicochemical environmental variables. Glaciers run the risk of disappearing and with them, their evolved microbiomes (Staley, 1997). Recording the most remarkable aspects of these endangered psychrophilic microbial communities is essential to understand the potential losses of biodiversity and how this may further impact the environment (Peter & Sommaruga, 2016a).

### 1.1.2 Sumaco Volcano

Pristine forests in the Amazon region provide opportunities to study untouched microbiotas and how environmental parameters, such as those along an altitudinal gradient, can cause shifts in microbial diversity (de Carvalho et al., 2016). Information on the diversity profile of healthy soils from pristine forests in the Amazon region provides a component of value to ecosystems and a source of comparison for degraded environments (de Carvalho et al., 2016; Merloti et al., 2019). Thus, the study of microbiome profiles in pristine environments and changes in response to environmental gradients can provide information to establish downstream applications, such as bioremediation, estimates of the effects of climate change on bacterial diversity and soil nutrient enrichment, or

other associated services from bacterial biodiversity (de Carvalho et al., 2016; Looby and Martin, 2020).

The Sumaco volcano is placed in the Amazon region, and jointly with El Reventador, Pan de Azúcar are the three volcanoes placed in the rear-arc, which is also considered the southernmost extent of rear-arc volcanism in Ecuador (Garrison et al., 2018; Hoffer et al., 2008) (Figure 1.2a). The Sumaco volcano, and its biome were declared a National Park in 1994 by the Ecuadorian Environment Minister. It has an extent of 205,751 ha. According to the international union, this distinction aims to protect natural biodiversity in a natural state as possible to conserve nature (Figure 1.2b). Moreover, the Ecuadorian policies allow tourism activities in this area (Simkin & Siebert, 1994), a source of ecosystem perturbation. Since the last possible Sumaco eruption took place in 1933 (Simkin & Siebert, 1994), now it hosts a rich plant (Lozano, Cabrera, Peyre, Cleef, & Toulkeridis, 2020) and animal (Hodge & Arbogast, 2016) diversity (Figure 1.2c). The unique landscape characteristics that exhibit the Sumaco biosphere to dabble into a field trip to the submission have contributed to making access to this biome a real and complex challenge (Bühler, Robert, & J, n.d.). Although a few studies have been conducted to document the plant and carnivore Sumaco diversity, its microbiome remains completely unknown. Therefore, a microbiome survey in the lesser unperturbed ecosystem is urgent to discover the Sumaco microcosm (Figure 1.2d).


**Figure 1.2.** The Sumaco Volcano. (a) Panoramic view of the Sumaco volcano. (b) The sampling process. (c) Researchers on the way to leave (d) A view of the lake top in the volcanic crater. (2016-2019)

### 1.1.3 Sequencing

To improve the understanding of the complex microbial interactions and ecosystem services, many studies have been conducted worldwide. Nowadays, a total of 43,224 results can be retrieved if we search the phrase “Environmental soil microbiology” in the famous PubMed repository (Díaz et al., 2021). It does not only provide an overview of how much attention has gained in the last two decades. The first reports that aims to elucidate the soil microcosm are back to 1945 (Figure 1.3). However, in 1977 Sanger revealed to the world what was the predominant method to reveal the hidden DNA sequence (Sanger et al., 1977). It took a couple of years to start creating machinery that were capable to implement this protocol. Since then, we can appreciate an exponential increment in research related to this topic. Moreover, the sanger sequencing initiative also present some drawbacks. For example, it is capable to sequence short pieces of DNA (300-1000 bp), with a proper quality. The first nucleotides sequenced has low quality because that is where the primer binds. Additionally, the sequence quality start to degrades after 700 to 900 bases (Schuster, 2008). It was not until 2005 where the first sign of what could be a sequencing game changer appeared, the 454 pyrosequencing technologies, by Roche (Margulies et al., 2005). This

technology promises to reduce time and cost by a highly parallel sequencing system by using the sequencing by synthesis approach. The heart of this protocol lays in the luciferase enzyme. It generates a light signal from free pyrophospate when a nucleotide is incorporated, and then, the signal is captured by a CCD sensor. Although the first put in scene of this technology was the most promising technology at that moment, since 2016 Roche discontinued this technology. Two main factors for this decision can be the high cost compared with other technologies and the sequencing error (1000 to 100000 nucleotides), which is even ten times higher than sanger sequencing (Liu et al., 2012; Quinlan, Stewart, Strömberg, & Marth, 2008).

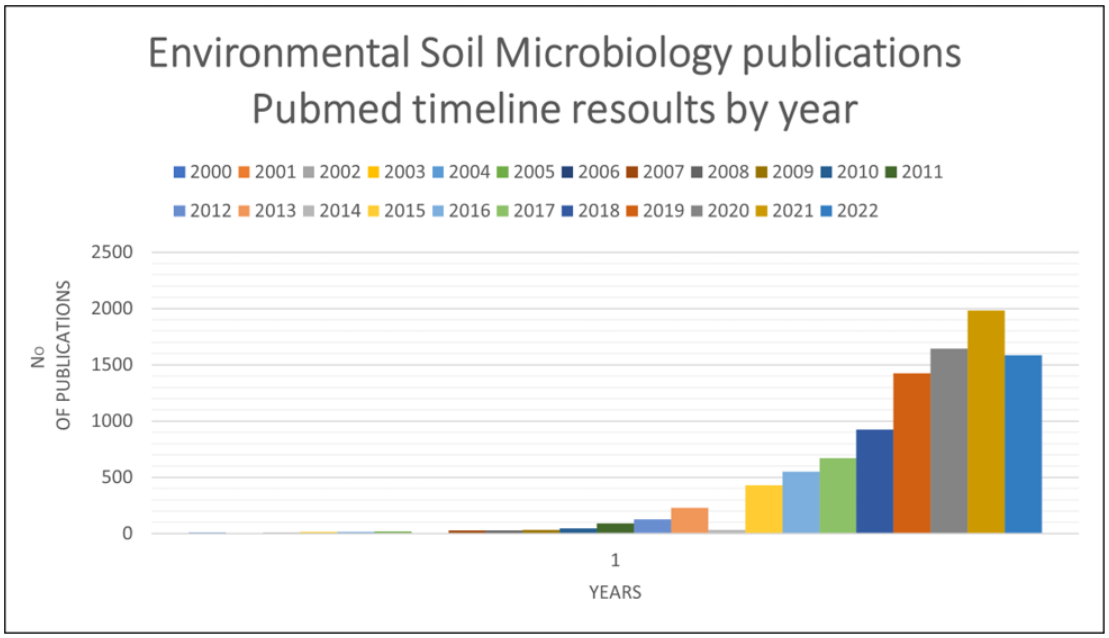



**Figure 1.3.** Number of publications found by PubMed with the "Environmental soil microbiology" key. The search was conducted in October 2022

The advent of the era of Omics techniques, culture-free microbiology research has become possible, thanks to the development of high-throughput technologies and the introduction of Metagenomics, defined as the analysis of a set of genomes extracted from a natural environment, which has allowed understanding the biology of microorganisms focusing on their genes and their individual influence on the functions they collectively contribute as a microbial community. (Pavlovic et al.,2020)

For the characterization of non-culturable microbiota, there are two fundamental principles in genomic research: 1. amplicon sequencing of marker genes such as 16S rRNA for bacteria and archaea and 2. whole genome shotgun sequencing. Metagenomic studies of the 16S rRNA gene are used as "Molecular Clock" as phylogenetic markers, i.e., establishing a lineal association to the phylogenetic distance, very useful to identify microorganisms in a sample and evaluate their taxonomic and phylogenetic identity. The 16S rRNA gene contains conserved and variable dispersed regions, early evolutionary events can be observed in the conserved regions, while in the variable or less conserved regions there may be evidence of recent changes. (Pavlovic et al.,2020)

Two years later after the apparition of 454 Roche technologies, in 2007, Illumina presented the short new read sequencing method that has been widely employed. This protocol is based on the “bridge amplification” technique, since oligonucleotide sequences are fixed on a solid support and are complementary to adapters that are previously ligated at each end of the DNA molecules. Then, several cycles of amplifications are applied to create clusters of each oligonucleotide fragment. Each solid support can allocate millions of simultaneous cluster reactions. During the amplification reaction, one of the four bases are incorporated. Since each of the four nucleotides are fluorescent labeled, the signal is acquired with a direct imaging detector which increase the detected speed (Slatko, Gardner, & Ausubel, 2018). Although Illumina provides several sequencer options, two of them are the most employed for environmental microbiology, MiSeq and HiSeq. The main difference is the sequence capacity. While MiSeq can generate up to 25 million read clusters, HiSeq can produce 5 billion read clusters. This characteristic also delimits them applications. For example, MiSeq is intended for targeted amplicon sequencing, meanwhile HiSeq is intended to shotgun metagenomics (Reuter, Spacek, & Snyder, 2015).

In environmental microbiology, both approaches, targeted amplicon and shotgun sequencing are aimed to elucidate the structure of microcosm community. However, the targeted amplicon approach is focused to sequence high conserved small subunit of the bacterial genome that can help to distinguish between organism, while the shotgun approach is more comprehensive sequencing the whole microbial DNA and then using this information to distinguish between species (Steven, Gallegos-Graves, Starkenburg, Chain, & Kuske, 2012). Although the targeted amplicon sequencing is limited to provided

taxonomic information, shotgun sequencing can also show microbes functions such as metabolism and microbe bioactivity (Cobo-Díaz et al., 2017). When the microcosm of a certain ecosystem is unknown, one of the first analysis to conduct is to explore the taxonomic profile, and the best cost-efficient way to do this, is by targeted amplicon sequencing.

## 1.2 Objectives

### 1.2.1 Hypothesis

Based on the arguments exposed by Ciccazzo et al. (2016), who proposed that distance from the retreating glacier can be used as a proxy to soil age (i.e., a chronosequence), was hypothesized that elevation would be a significantly correlated component to differences in the composition of the observed communities. Was also hypothesized that these differences will be related to significant correlations in the concentration of nutrients and other physicochemical properties that are relevant for bacterial life. This study provides insights into the composition of bacterial communities of two Andean volcanoes.

By considering that there is no definitive consensus on how bacterial and microbial communities respond along elevational gradients, this survey on the Cayambe and Sumaco volcanoes adds additional sources of evidence on the role of soil chemistry on richness, diversity, and community structure along an elevational gradient. Amidst so many varying results in previous studies, that took a conservative stance and followed the most recent synthesis by Looby and Martin (2020).

Was hypothesized that bacterial communities would respond most strongly to soil pH, regardless of varying elevation along the sampled altitudinal gradient (H1). A null hypothesis is also possible to establish for no effect of the measured physicochemical parameters on the structure of bacterial communities (H0). Thus, was tested the most recent proposal on how microbial diversity responds to montane gradients, placing a central role in soils as potentially independent of other abiotic factors along an elevational gradient (Looby & Martin, 2020).

## 1.2.2 Objectives

### 1.2.2.1 General Objective

To survey the bacterial biodiversity of the Andean Volcanos Cayambe and Sumaco in the search to catalog and identify the bacterial communities of these ecosystems.

### 1.2.2.2 Specific Objectives

- To analyze the bacterial diversity and community structure of the Cayambe and Sumaco Volcanoes through a three-year survey using a universal molecular marker.
- To relate the Cayambe and Sumaco volcanoes' bacterial communities with their physical-chemical parameters.
- To catalog the microbiome of the Cayambe and Sumaco Volcanoes.

## 1.2.3 General Methods

### 1.2.3.1 Sample collection and environmental analysis of Cayambe Glacier

The sampling location, the ice-capped Cayambe volcano (0,03º N; 77,98º W), is located in the equator line in a country whit the same name, Ecuador. In general, three sampling collection steps were conducted from 2015 to 2019 in an altitudinal gradient ranging from 4721 to 5810 meters above sea level (m.a.s.l). The first procedure occurred between November 28$^{th}$, 2015, and January 28$^{th}$, 2016; the second sampling effort was between February 22$^{nd}$, 2018, and October 7$^{th,}$ 2018, and the third one between May 2$^{nd}$ and 12$^{th}$, 2019. We reached the summit in each sampling collection procedure, about 5800 m.a.s.l, by climbing the rough frozen landscape. Once in the submit, we collected different sample types that can be grouped into soil and water samples (Figure 3.1). The main criteria for this grouping were based on the AND extraction method. For example, ice and snow samples at room temperature are water in the liquid state. Moreover, for downstream comparison, these subcategories, ice, snow, and lake samples, are conserved and presented as "S*ubtype".*

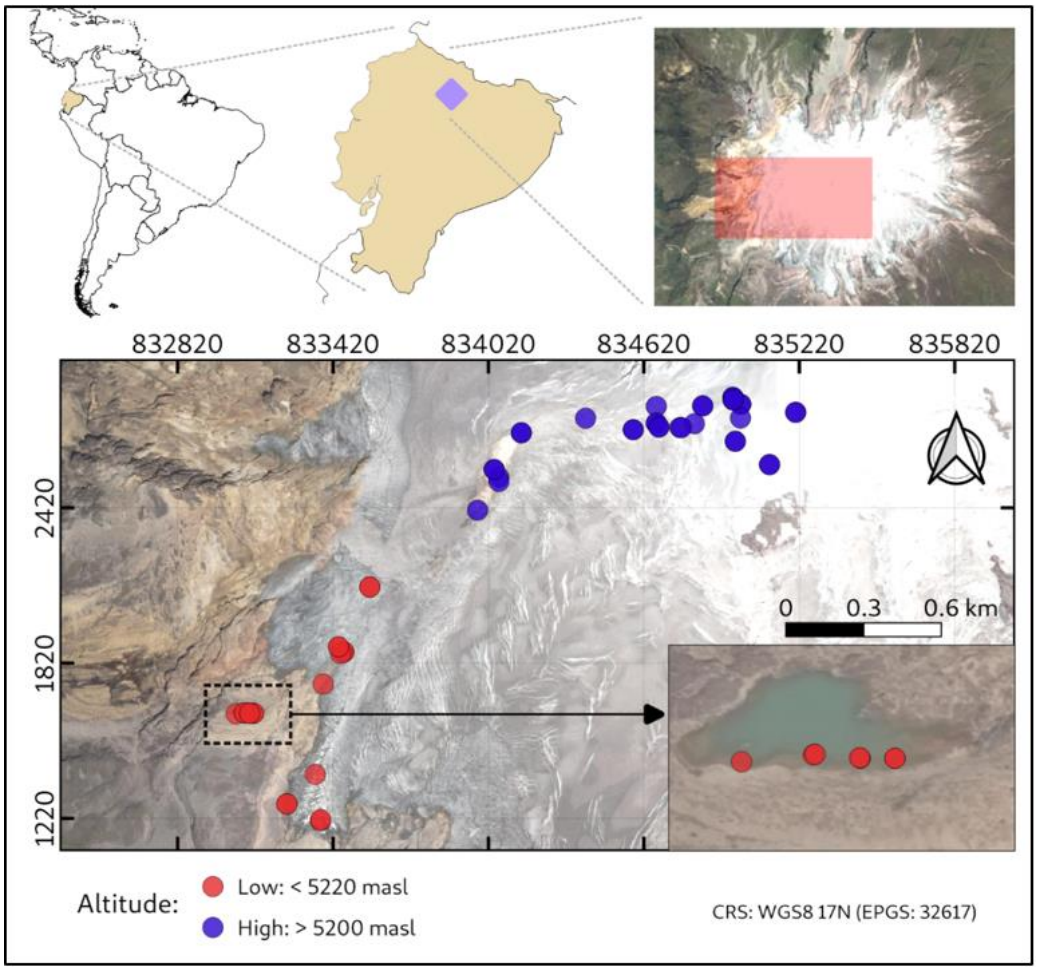



**Figure 1.4.** Location of the Cayambe Volcanic Complex and map of the collected samples along the glacier ascension route (red and blue dots). Samples were categorized into high-altitude and low altitude. (2015-2019)

To contrast samples, we define two main categories based on the landscape: low (<5220 m.a.s.l) and high (>5200 m.a.s.l) altitudes samples. It was based on a visible gap in the geographic layout of the sampling route. Therefore, we expect to find different microcosm behavior and composition over time. For soil and water samples, 1 Kg and 1 L were collected for each type, respectively. Each sample was used for the isolation of total genomic target DNA and the determination of physical-chemical properties. A total of 51 samples were collected over the studied time ranges. Samples were labeled as “CAY##” and sequentially numbered as it is presented in the following Table 1.1.

#### 1.2.3.2 Sample collection and environmental analysis of SUMACO

The sampling location, the Sumaco volcano (0.54º S, 77.63º N) is a continental volcano located around the 0 latitude in Ecuador. Three sampling collection approaches were collected between 2016, to 2019. during a three-day survey towards the summit of the Sumaco volcano and spanned an elevation gradient between 1584 to 3809 m.a.s.l.; The first sample recovery was conducted on 22–24 July 2016, the second effort was on 3rd March 2018, and the third one was on 19th September 2019. According to the tropic's seasons, the first and third sampling approach was conducted in summer, while the second sampling effort was carried out in winter. In each sampling collection procedure, we reached the volcano summit, which is about 3800 m.a.s.l; by climbing the rough path and the last transect, we walked through the dense vegetation. Once at the summit, we collected soil samples through the altitudinal gradient and water samples from the two lakes (Figure 1.4 and Table 1.2). Thus, covering a range of vegetation transitions that included tropical, subtropical, and paramo ecosystems (Lozano et al. 2020). The climbing route extended over 20 kilometers from the town of "Pacto Sumaco" to the summit. After reaching the summit on the second day, we descended during the third and collected nine soil samples (1 kilogram each).

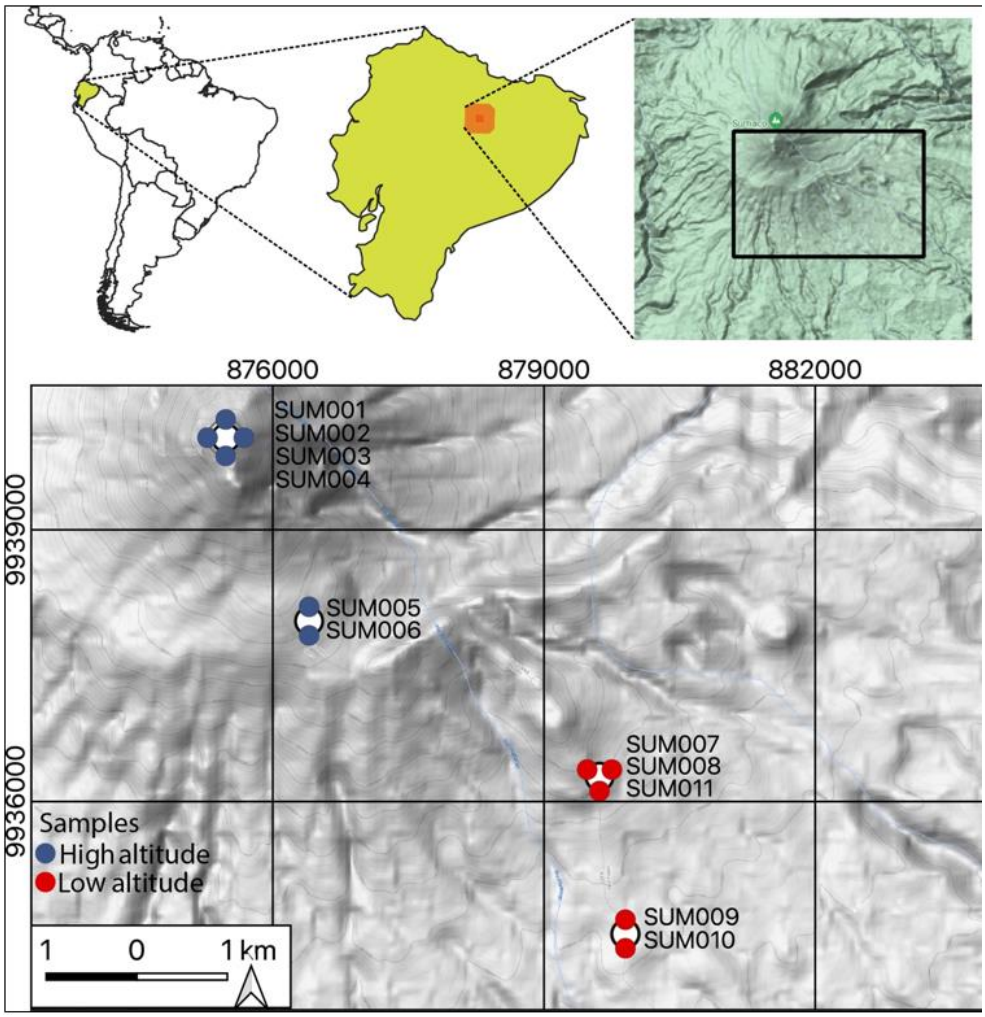



**Figure 2.5.** Sampling locations along an elevational gradient in the Sumaco Volcano. Locations that were close together are represented as points in a circle. Samples have been classified into two altitudinal categories, above (blue) and below 2900 m.a.s.l. (red).

A shovel or ice axe was used to dig into the sampling point approximately 10 to 25 centimeters below the surface, removing roots and rocks. Soil samples were taken in duplicate with a shovel previously washed and disinfected with 70% alcohol. Each duplicate was used for sequencing and physical-chemical characterization, respectively. Each soil sample was a unique observation along the altitudinal gradient, such that physicochemical measurements characterized each one. Samples were labeled as “SUM##” and sequentially numbered as presented in Table 1.2.

**Table 1.1.**

Volcanoes Samples Collection M. Díaz. (Cayambe)

| No | Sampling Date | Sec ID | Altitude (m.a.s.l) | UTM Coordinates W | UTM Coordinates N/S | Rep. | Year Collection | Detail | Type |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 28/11/2015 | CAY001 | 4944 | 17N-083346 | 1863 | a | 1 | Glacial ground | soil |
| 2 | 28/11/2015 | CAY002 | 4944 | 17N-083346 | 1863 | a | 1 | Snow | water |
| 3 | 28/11/2015 | CAY003 | 4944 | 17N-083346 | 1863 | a | 1 | Reventador ash ground | soil |
| 4 | 28/11/2015 | CAY004 | 4944 | 17N-083346 | 1877 | a | 1 | Water in the glacier | water |
| 5 | 28/11/2015 | CAY005 | 4783 | 17N-083315 | 1629 | a | 1 | Water-Lagoon | water |
| 6 | 28/11/2015 | CAY006 | 4783 | 17N-083315 | 1629 | a | 1 | Sediment-Lagoon | soil |
| 7 | 28/11/2015 | CAY007 | 4799 | 17N-083305 | 1616 | a | 1 | Water-Lagoon | water |
| 8 | 28/01/2016 | CAY008 | 5583 | 18N-0166858 | 652 | a | 1 | Snow-ice | water |
| 9 | 28/01/2016 | CAY009 | 5569 | 18N-0166802 | 2730 | a | 1 | Snow-ice | water |
| 10 | 28/01/2016 | CAY010 | 5533 | 18N-0166711 | 2812 | a | 1 | Snow-ice | water |
| 11 | 28/01/2016 | CAY011 | 5434 | 18N-0166438 | 2766 | a | 1 | Ice-Snow (Reventador ash). | water |
| 12 | 28/01/2016 | CAY012 | 5375 | 18N-0166190 | 2710 | a | 1 | Soil (Ice Flower). | soil |
| 13 | 28/01/2016 | CAY013 | 5375 | 18N-0166079 | 2710 | a | 1 | Ice - ice (High Jarrin Peaks) | water |
| 14 | 28/01/2016 | CAY014 | 5306 | 18N-0166103 | 2538 | a | 1 | Dry soil (Jarrin Peaks Summit). | soil |
| 15 | 28/01/2016 | CAY015 | 5293 | 18N-0166103 | 2522 | a | 1 | Water (Jarrin Peaks Below-well) | water |
| 16 | 28/01/2016 | C1 | 4944 | 17N-083346 | 1863 | b | 1 | Glacial ground | soil |
| 17 | 28/01/2016 | C2 | 4944 | 17N-083346 | 1863 | b | 1 | Snow | water |
| 18 | 28/01/2016 | C3 | 4944 | 17N-083346 | 1863 | b | 1 | Reventador ash ground | soil |
| 19 | 28/01/2016 | C4 | 4944 | 17N-083346 | 1877 | b | 1 | Water in te glacier | water |
| 20 | 28/01/2016 | C5 | 4783 | 17N-083315 | 1629 | b | 1 | Water-Lagoon | water |
| 21 | 28/01/2016 | C6 | 4783 | 17N-083315 | 1629 | b | 1 | Sediment-Lagoon | soil |

[a] data obtained from Original Sample

[b] data obtained from Replicate Sample

**Table 1.1.**

Volcanoes Samples Collection M. Díaz. (Cayambe)

| No | Sampling Date | Sec ID | Altitude (m.a.s.l) | UTM Coordinates W | UTM Coordinates N/S | Rep. | Year Collection | Detail | Type |
|---|---|---|---|---|---|---|---|---|---|
| 22 | 28/01/2016 | C7 | 4799 | 17N-083305 | 1616 | b | 1 | Water-Lagoon | water |
| 23 | 28/01/2016 | C8 | 5569 | 18N-0166802 | 2730 | b | 1 | Ice-Snow | water |
| 24 | 28/01/2016 | C9 | 5434 | 18N-0166438 | 2766 | b | 1 | Ice- Snow (Reventador Ash). | water |
| 25 | 28/01/2016 | C10 | 5375 | 18N-0166190 | 2710 | b | 1 | Soil (Ice Flower). | soil |
| 26 | 28/01/2016 | C11 | 5375 | 18N-0166079 | 2710 | b | 1 | Ice - ice (High Jarrin Peaks) | water |
| 27 | 28/01/2016 | C12 | 5306 | 18N-0166103 | 2538 | b | 1 | Dry soil (Jarrin Peaks summit). | soil |
| 28 | 28/01/2016 | C13 | 5293 | 18N-0166103 | 2522 | b | 1 | Water (Jarrin Peaks below-well) | water |
| 29 | 28/01/2016 | C14 | 5237 | 17N-0833978 | 2411 | a | 1 | Ground (Start of Jarrin Peaks). | soil |
| 30 | 12/01/2018 | C15 | 4799 | 17N-083305 | 1616 | a | 2 | Lagoon water | water |
| 31 | 12/01/2018 | C16 | 4799 | 17N-083305 | 1616 | b | 2 | Lagoon water | water |
| 32 | 12/01/2018 | C17 | 5010 | 2114 | 833562 | a | 2 | Cayambe Medium Snow | water |
| 33 | 12/01/2018 | C18 | 5010 | 2114 | 833562 | b | 2 | Soil-Ice | soil |
| 34 | 12/01/2018 | C19 | 5024 | 528 | 17N0833561 | a | 2 | Snow | water |
| 35 | 12/01/2018 | C20 | 5024 | 528 | 17N0833561 | a | 2 | Ground-Ice | water |
| 36 | 12/01/2018 | C21 | 4940 | 1858 | 17N0833450 | a | 2 | Snow | water |
| 37 | 12/01/2018 | C22 | 4940 | 1858 | 17N0833450 | a | 2 | Ground | soil |
| 38 | 12/01/2018 | C23 | 4942 | 1883 | 17N0833441 | a | 2 | Snow | water |
| 39 | 12/01/2018 | C24 | 4942 | 1883 | 17N0833441 | a | 2 | Ground-Ice | water |
| 40 | 12/01/2018 | C25 | 4895 | 1739 | 17N0833382 | a | 2 | Snow | water |
| 41 | 12/01/2018 | C26 | 4895 | 1739 | 17N0833382 | a | 2 | Ground | soil |
| 42 | 12/01/2018 | C27 | 4801 | 1638 | 17N0833120 | a | 2 | Lagoon water | water |

[a] data obtained from Original Sample

[b] data obtained from Replicate Sample

**Table 1.1.**

Volcanoes Samples Collection M. Díaz. (Cayambe)

| No | Sampling Date | Sec ID | Altitude (m.a.s.l) | UTM Coordinates W | UTM Coordinates N/S | Rep. | Year Collection | Detail | Type |
|---|---|---|---|---|---|---|---|---|---|
| 43 | 12/01/2018 | C28 | 4801 | 1638 | 17N0833120 | a | 2 | Lagoon water | water |
| 44 | 12/01/2018 | C29 | 4801 | 1638 | 17N0833120 | a | 2 | Lagoon water | water |
| 45 | 12/01/2018 | C30 | 4801 | 1638 | 17N0833120 | a | 2 | Lagoon water | water |
| 46 | 07/10/2018 | C31 | 5768 | 167037.3 E | 2819.6 N | a | 2 | Cayambe snow summit | water |
| 47 | 07/10/2018 | C32 | 5690 | 167006.5 E | 2846.6 N | a | 2 | Snow/ Ice | water |
| 48 | 07/10/2018 | C33 | 5649 | 166918.5 E | 2821.5 N | a | 2 | Snow/Ice | water |
| 49 | 07/10/2018 | C34 | 5662 | 166890.2 E | 2813.9 N | a | 2 | Ice | water |
| 50 | 07/10/2018 | C35 | 5608 | 166857.4 E | 2802.1 N | a | 2 | Snow | water |
| 51 | 07/10/2018 | C36 | 5569 | 166784.2 E | 2757.3 N | a | 2 | Snow | water |
| 52 | 07/10/2018 | C37 | 5508 | 166622.3 E | 2721.5 N | a | 2 | Snow | water |
| 53 | 07/10/2018 | C38 | 5332 | 166084.6 E | 2567.8 N | a | 2 | Ground | soil |
| 54 | 07/10/2018 | C31 | 5768 | 167037.3 E | 2819.6 N | b | 2 | Cayambe snow summit | water |
| 55 | 07/10/2018 | C32 | 5690 | 167006.5 E | 2846.6 N | b | 2 | Snow/ Ice | water |
| 56 | 07/10/2018 | C33 | 5649 | 166918.5 E | 2821.5 N | b | 2 | Snow/Ice | water |
| 57 | 07/10/2018 | C34 | 5662 | 166890.2 E | 2813.9 N | b | 2 | Ice | water |
| 58 | 07/10/2018 | C35 | 5608 | 166857.4 E | 2802.1 N | b | 2 | Snow | water |
| 59 | 07/10/2018 | C36 | 5569 | 166784.2 E | 2757.3 N | b | 2 | Snow | water |
| 60 | 07/10/2018 | C37 | 5508 | 166622.3 E | 2721.5 N | b | 2 | Snow | water |
| 61 | 07/10/2018 | C38 | 5332 | 166084.6 E | 2567.8 N | b | 2 | Ground | soil |
| 62 | 07/10/2018 | C39 | 5010 | 2114 | 833562 | b | 2 | Cayambe snow summit | water |
| 63 | 07/10/2018 | C40 | 5010 | 2114 | 833562 | b | 2 | Ground-Ice | water |
| 64 | 07/10/2018 | C41 | 5024 | 528 | 17N0833561 | b | 2 | Snow | water |
| 65 | 07/10/2018 | C42 | 5024 | 528 | 17N0833561 | b | 2 | Ground-Ice | water |

[a] data obtained from Original Sample

[b] data obtained from Replicate Sample

**Table 1.1.**

Volcanoes Samples Collection M. Díaz. (Cayambe)

| No | Sampling Date | Sec ID | Altitude (m.a.s.l) | UTM Coordinates W | UTM Coordinates N/S | Rep. | Year Collection | Detail | Type |
|---|---|---|---|---|---|---|---|---|---|
| 66 | 07/10/2018 | C43 | 4940 | 1858 | 17N0833450 | b | 2 | Snow | water |
| 67 | 07/10/2018 | C44 | 4940 | 1858 | 17N0833450 | b | 2 | Ground | soil |
| 68 | 07/10/2018 | C45 | 4942 | 1883 | 17N0833441 | b | 2 | Snow | water |
| 69 | 07/10/2018 | C46 | 4942 | 1883 | 17N0833441 | b | 2 | Ground-Ice | water |
| 70 | 07/10/2018 | C47 | 4895 | 1739 | 17N0833382 | b | 2 | Snow | water |
| 71 | 07/10/2018 | C48 | 4895 | 1739 | 17N0833382 | b | 2 | Ground | soil |
| 72 | 07/10/2018 | C49 | 4801 | 1638 | 17N0833120 | b | 2 | Lagoon water | water |
| 73 | 07/10/2018 | C50 | 4801 | 1638 | 17N0833120 | b | 2 | Lagoon water | water |
| 74 | 07/10/2018 | C51 | 4801 | 1638 | 17N0833120 | b | 2 | Lagoon water | water |
| 75 | 07/10/2018 | C52 | 4801 | 1638 | 17N0833120 | b | 2 | Lagoon water | water |
| 76 | 02/05/2019 | C61 | 5072 | 1276 | 17N0833242 | a | 3 | Snow- Ice | water |
| 77 | 02/05/2019 | C62 | 5072 | 1276 | 17N0833242 | b | 3 | Snow- Ice | water |
| 76 | 02/05/2019 | C61 | 5072 | 1276 | 17N0833242 | a | 3 | Snow- Ice | water |
| 77 | 02/05/2019 | C62 | 5072 | 1276 | 17N0833242 | b | 3 | Snow- Ice | water |
| 78 | 02/05/2019 | C63 | 4721 | 1390 | 17N0833352 | a | 3 | Snow- Ice | water |
| 79 | 02/05/2019 | C64 | 4721 | 1390 | 17N0833352 | b | 3 | Snow- Ice | water |
| 80 | 12/05/2019 | C65 | 5810 | 2589 | 18N0167249 | a | 3 | Snow-Ice Summit Highest Point | water |
| 81 | 12/05/2019 | C66 | 5810 | 2589 | 18N0167249 | b | 3 | Snow- Ice | water |
| 82 | 12/05/2019 | C67 | 5754 | 2587 | 18N0167148 | a | 3 | Snow- Ice | water |
| 83 | 12/05/2019 | C68 | 5754 | 2587 | 18N0167148 | b | 3 | Snow- Ice | water |
| 84 | 12/05/2019 | C69 | 5686 | 2677 | 18N0167016 | a | 3 | Snow- Ice | water |
| 85 | 12/05/2019 | C70 | 5686 | 2677 | 18N0167016 | b | 3 | Snow- Ice | water |
| 86 | 12/05/2019 | C71 | 5534 | 2747 | 18N0166711 | a | 3 | Snow- Ice | water |
| 87 | 12/05/2019 | C72 | 5534 | 2747 | 18N0166711 | b | 3 | Snow- Ice | water |

[a] data obtained from Original Sample

[b] data obtained from Replicate Sample

**Table 1.1.**

Volcanoes Samples Collection M. Díaz. (Cayambe)

| No | Sampling Date | Sec ID | Altitude (m.a.s.l) | UTM Coordinates W | UTM Coordinates N/S | Rep. | Year Collection | Detail | Type |
|---|---|---|---|---|---|---|---|---|---|
| 88 | 12/05/2019 | C73 | 5438 | 2772 | 18N0166428 | a | 3 | Snow- Ice | water |
| 89 | 12/05/2019 | C74 | 5438 | 2772 | 18N0166428 | b | 3 | Snow- Ice | water |

[a] data obtained from Original Sample

[b] data obtained from Replicate Sample

**Table 1.2.**

Volcanoes Samples Collection M. Díaz. (Sumaco)

| No | Sampling Date | Sec ID | altitude (m.a.s.l) | UTM Coordinates W | UTM Coordinates N/S | | Year Collection | Detail | Type |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 22/07/2016 | SUM001 | 3771 | 18S 207540.26 | 9940095.12 | a | 1 | Laguna Summit crater water | water |
| 2 | 22/07/2016 | SUM002 | 3783 | 18S 207534.00 | 9940093 | a | 1 | Crater ground | soil |
| 3 | 22/07/2016 | SUM003 | 3808.3 | 18S 207538.54 | 9940022.11 | a | 1 | Center crater ground | soil |
| 4 | 22/07/2016 | SUM004 | 3808.2 | 18S 207537.31 | 9940022.12 | a | 1 | Crater Rim ground | soil |
| 5 | 22/07/2016 | SUM005 | 2998.4 | 18S 208430.60 | 9938201.02 | a | 1 | Ground near path | soil |
| 6 | 22/07/2016 | SUM006 | 2990 | 18S 208491.86 | 9937869.78 | a | 1 | Clay ground near trail lagoon | soil |
| 7 | 22/07/2016 | SUM007 | 2452 | 18S 211221.93 | 9937005.55 | a | 1 | Wawa Sumaco lagoon water | water |
| 8 | 22/07/2016 | SUM008 | 2261.3 | 18S 211763.44 | 9935930.67 | a | 1 | Stop 4 ground | soil |
| 9 | 22/07/2016 | SUM009 | 1953 | 18S 211960.31 | 9934588.01 | a | 1 | Stop 3 ground | soil |
| 10 | 22/07/2016 | SUM010 | 1942 | 18S 211963.31 | 9934587 | a | 1 | Stop 2 ground | soil |
| 11 | 22/07/2016 | SUM011 | 1584 | 18S 212036.28 | 9936029.25 | a | 1 | Stop 1 ground | soil |
| 12 | 22/07/2016 | S1 | 3771 | 18S 207540.26 | 9940095.12 | b | 1 | Laguna Summit crater water | water |
| 13 | 22/07/2016 | S2 | 3783 | 18S 207534.00 | 9940093 | b | 1 | Crater ground | soil |
| 14 | 22/07/2016 | S3 | 3808.3 | 18S 207538.54 | 9940022.11 | b | 1 | Center crater ground | soil |
| 15 | 22/07/2016 | S4 | 3808.2 | 18S 207537.31 | 9940022.12 | b | 1 | Crater Rim ground | soil |
| 16 | 22/07/2016 | S5 | 2998.4 | 18S 208430.60 | 9938201.02 | b | 1 | Ground near path | soil |
| 17 | 22/07/2016 | S6 | 2990 | 18S 208491.86 | 9937869.78 | b | 1 | Clay ground near trail lagoon | soil |
| 18 | 22/07/2016 | S7 | 2452 | 18S 211221.93 | 9937005.55 | b | 1 | Wawa Sumaco lagoon water | water |
| 19 | 22/07/2016 | S8 | 2261.3 | 18S 211763.44 | 9935930.67 | b | 1 | Stop 4 ground | soil |
| 20 | 22/07/2016 | S9 | 1953 | 18S 211960.31 | 9934588.01 | b | 1 | Stop 3 ground | soil |
| 21 | 22/07/2016 | S10 | 1942 | 18S 211963.31 | 9934587 | b | 1 | Stop 2 ground | soil |
| 22 | 22/07/2016 | S11 | 1584 | 18S 212036.28 | 9936029.25 | b | 1 | Stop 1 ground | soil |
| 23 | 08/03/2018 | S12 | 3771 | 18M 207556 | 9940112 | a | 2 | Laguna Summit crater water | water |
| 24 | 08/03/2018 | S13 | 3783 | 18M 207544 | 9940102 | a | 2 | Crater ground | soil |
| 25 | 08/03/2018 | S14 | 3808.3 | 18M 207468 | 9940110 | a | 2 | Center crater ground | soil |

[a] data obtained from Original Sample

[b] data obtained from Replicate Sample

**Table 1.2.**
Volcanoes Samples Collection M. Díaz. (Sumaco)

| No | Sampling Date | Sec ID | altitude (m.a.s.l) | UTM Coordinates W | UTM Coordinates N/S | | Year Collection | Detail | Type |
|---|---|---|---|---|---|---|---|---|---|
| 26 | 08/03/2018 | S15 | 3808.2 | 18M 207444 | 9940068 | a | 2 | Crater Rim ground | soil |
| 27 | 08/03/2018 | S16 | 2452 | 18M 211221 | 9937005 | a | 2 | Wawa Sumaco lagoon water | water |
| 28 | 08/03/2018 | S17 | 2261.3 | 18M 207439 | 9936068 | a | 2 | Stop 4 ground | soil |
| 29 | 08/03/2018 | S18 | 1953 | 18M 212222 | 9934932 | a | 2 | Stop 3 ground | soil |
| 30 | 08/03/2018 | S19 | 1942 | 18S 211552 | 9933068 | a | 2 | Stop 2 ground | soil |
| 31 | 08/03/2018 | S20 | 1584 | 18S 211545 | 9932116 | a | 2 | Stop 1 ground | soil |
| 32 | 08/03/2018 | S21 | 3771 | 18M 207556 | 9940112 | b | 2 | Laguna Summit crater water | water |
| 33 | 08/03/2018 | S22 | 3783 | 18M 207544 | 9940102 | b | 2 | Crater ground | soil |
| 34 | 08/03/2018 | S23 | 3808.3 | 18M 207468 | 9940110 | b | 2 | Center crater ground | soil |
| 35 | 08/03/2018 | S24 | 3808.2 | 18M 207444 | 9940068 | b | 2 | Crater Rim ground | soil |
| 36 | 08/03/2018 | S25 | 2998.4 | 18M 211221 | 9937005 | b | 2 | Wawa Sumaco lagoon water | water |
| 37 | 08/03/2018 | S26 | 2990 | 18M 207439 | 9936068 | b | 2 | Stop 4 ground | soil |
| 38 | 08/03/2018 | S27 | 2452 | 18M 212222 | 9934932 | b | 2 | Stop 3 ground | soil |
| 39 | 08/03/2018 | S28 | 2261.3 | 18S 211552 | 9933068 | b | 2 | Stop 2 ground | soil |
| 40 | 08/03/2018 | S29 | 1953 | 18S 211545 | 9932116 | b | 2 | Stop 1 ground | soil |
| 41 | 23/09/2019 | S21 | 3783 | W 077°37'38.5" | S 00°32'29.0" | a | 3 | Crater dark brown solid ground | soil |
| 42 | 23/09/2019 | S22 | 3808.3 | W 077°37'42.0" | S 00°32'27.0" | a | 3 | Solid dark brown ground center crater | soil |
| 43 | 23/09/2019 | S23 | 3808.2 | W 077°37'39.3" | S 00°32'31.6" | a | 3 | Dark brown solid ground with vegetables residues Crater Rim | soil |
| 44 | 23/09/2019 | S24 | 2800 | W 077°36'54.2" | S 00°33'50.3" | a | 3 | Dark brown solid ground Pavayacu shelter | soil |
| 45 | 23/09/2019 | S25 | 2261.3 | W 077°35'37.0" | S 00°34'10.2" | a | 3 | Solid dark brown ground Stop 4 | soil |
| 46 | 23/09/2019 | S26 | 1953 | W 077°35'22.9" | S 00°34'43.7" | a | 3 | Solid dark brown ground Stop 3 | soil |

[a] data obtained from Original Sample
[b] data obtained from Replicate Sample

**Table 1.2.**

Volcanoes Samples Collection M. Díaz. (Sumaco)

| No | Sampling Date | Sec ID | altitude (m.a.s.l) | UTM Coordinates W | UTM Coordinates N/S | | Year Collection | Detail | Type |
|---|---|---|---|---|---|---|---|---|---|
| 47 | 23/09/2019 | S27 | 1942 | W 077°35'14.4" | S 00°35'29.5" | a | 3 | Solid dark brown ground Stop 2 | soil |
| 48 | 23/09/2019 | S28 | 1700 | W 077°35'24.6" | S 00°36'52.9" | a | 3 | Solid dark brown ground Stop 1 | soil |
| 49 | 23/09/2019 | S29 | 1584 | W 077°35'28.4" | S 00°37'13.9" | a | 3 | Solid dark brown ground Stop 1 park boundary | soil |
| 50 | 23/09/2019 | S30 | 2452 | W 077°36'52.8" | S 00°33'51.8" | a | 3 | Wawa Sumaco lagoon water | water |
| 51 | 23/09/2019 | S31 | 3771 | W 077°37'38.6" | S 00°32'28.7" | a | 3 | Crater Volcano lagoon water | water |
| 52 | 23/09/2019 | S32 | 3783 | W 077°37'38.5" | S 00°32'29.0" | b | 3 | Crater dark brown solid ground | soil |
| 53 | 23/09/2019 | S33 | 3808.3 | W 077°37'42.0" | S 00°32'27.0" | b | 3 | Solid dark brown ground center crater | soil |
| 54 | 23/09/2019 | S34 | 3808.2 | W 077°37'39.3" | S 00°32'31.6" | b | 3 | Dark brown solid ground with vegetables residues Crater Rim | soil |
| 55 | 23/09/2019 | S35 | 2800 | W 077°36'54.2" | S 00°33'50.3" | b | 3 | Dark brown solid ground Pavayacu shelter | soil |
| 56 | 23/09/2019 | S36 | 2261.3 | W 077°35'37.0" | S 00°34'10.2" | b | 3 | Solid dark brown ground Stop 4 | soil |
| 57 | 23/09/2019 | S37 | 1953 | W 077°35'22.9" | S 00°34'43.7" | b | 3 | Solid dark brown ground Stop 3 | soil |
| 58 | 23/09/2019 | S38 | 1942 | W 077°35'14.4" | S 00°35'29.5" | b | 3 | Solid dark brown ground Stop 2 | soil |
| 59 | 23/09/2019 | S39 | 1700 | W 077°35'24.6" | S 00°36'52.9" | b | 3 | Solid dark brown ground Stop 1 | soil |
| 60 | 23/09/2019 | S40 | 1584 | W 077°35'28.4" | S 00°37'13.9" | b | 3 | Solid dark brown ground Stop 1 park boundary | soil |
| 61 | 23/09/2019 | S41 | 2452 | W 077°36'52.8" | S 00°33'51.8" | b | 3 | Wawa Sumaco lagoon water | water |
| 62 | 23/09/2019 | S42 | 3771 | W 077°37'38.6" | S 00°32'28.7" | b | 3 | Crater Volcano lagoon water | water |

[a] data obtained from Original Sample

[b] data obtained from Replicate Sample

### 1.2.3.3 Physico-Chemical properties

Each soil sample was a unique observation along the altitudinal gradient, such that physicochemical measurements characterized each one. Samples were labeled as “SUM##” and “CAY##” sequentially numbered. The physicochemical properties of each sample were measured at the Center for Integral Analytical Solutions (CENTROCESAL Cía.Ltda., Ecuador. Accreditation No SAE LEN 12-001) and consisted of the following parameters: electrical conductivity (EC), organic matter content (Org), humidity (Hum), cation exchange capacity (CEC), phosphorus (P), nitrogen (N), calcium (Ca2+), magnesium (Mg2+), manganese (Mn2+), sulfur (S), potassium (K+), iron (Fe), and sodium (Na+). These parameters were obtained according to the procedures described in (Baird et al., 2017). pH was evaluated in the laboratory with a pH meter (Mettler-Toledo SevenGO, Millipore, Columbus, OH, USA). Sample coordinates, elevation, and values for the physicochemical parameters are included in Supplementary Anexo 1.

### 1.2.3.4 DNA extraction, 16S rRNA gene library preparation, and sequencing

Total soil genomic DNA was isolated with the PowerSoil DNA Isolation kit (Cat. No. 112888-50, MoBio Laboratories, Inc.). Total water-based samples of DNA were isolated with the PowerWater DNA Isolation kit (Cat. No. 14900-50 NF MoBio Laboratories, Inc.). DNA integrity and concentration have been evaluated with the Thermo Scientific™ Nanodrop™ 200,0 and concentrations have ranged from 0.8 to 60 nm/uL. The total extracted genomic DNA according to the supplier's instructions, is currently stored at -80°C in the Ecuadorian Microbiome Project (EcuMP) collection at the Institute of Research on Zoonoses (CIZ) of the Central University of Ecuador. The development of amplicon libraries is carried out by the protocol 16S Metagenomic Sequencing Library Preparation (Illumina), “*Preparing 16S ribosomal RNA gene Amplicons for the Illumina MiSeq System”* This protocol describes a method for preparing samples for sequencing the variable V3 and V4 regions of the 16S rRNA gene. The protocol combined with a benchtop sequencing system, on-board primary analysis, and secondary analysis provides a comprehensive workflow for 16S rRNA amplicon sequencing, including the primer pair sequences for the V3 andV4 region that create a single

amplicon of approximately ~460 bp. The protocol also includes overhang adapter sequences that must be appended to the primer pair sequences for compatibility with Illumina index and sequencing adapters. A 500 bp partial region, including the V3 and V4 hypervariable regions of the 16S rRNA genes was amplified using the default primers for V3-V4, based on previous work (Klindworth et al., 2013): Total soil genomic DNA was isolated with the PowerSoil DNA Isolation kit (Cat. No. 112888-50, MoBio Laboratories, Inc.). Total water-based samples of DNA were isolated with the PowerWater DNA Isolation kit (Cat. No. 14900-50 NF MoBio Laboratories, Inc.). DNA integrity and concentration have been evaluated with the Thermo Scientific™ Nanodrop™ 200,0 and concentrations have ranged from 0.8 to 60 nm/uL. The total extracted genomic DNA according to the supplier's instructions, is currently stored at -80°C in the Ecuadorian Microbiome Project (EcuMP) collection at the Institute of Research on Zoonoses (CIZ) of the Central University of Ecuador. The development of amplicon libraries is carried out by the protocol 16S Metagenomic Sequencing Library Preparation (Illumina), "*Preparing 16S ribosomal RNA gene Amplicons for the Illumina MiSeq System"* This protocol describes a method for preparing samples for sequencing the variable V3 and V4 regions of the 16S rRNA gene. The protocol combined with a benchtop sequencing system, on-board primary analysis, and secondary analysis provides a comprehensive workflow for 16S rRNA amplicon sequencing, including the primer pair sequences for the V3 V4 region that create a single amplicon of approximately ~460 bp. The protocol also includes overhang adapter sequences that must be appended to the primer pair sequences for compatibility with Illumina index and sequencing adapters. A 500 bp partial region, including the V3 and V4 hypervariable regions of the 16S rRNA genes was amplified using the default primers for V3-V4, based on previous work (Klindworth et al., 2013):

**Bakt_341F: CCTACGGGNGGCWGCAG**

**Bakt_805R: GATCHVGGGTATCTAATCC**

Prepare library‑The protocol describes the steps to amplify the V3 and V4 region and using a limited cycle PCR, add Illumina sequencing adapters and dual-index barcodes to the amplicon target. Using the full complement of Nextera XT indices, up to 96 libraries can be pooled together for sequencing

#### 1.2.3.5 Sequence processing and taxonomic analysis

Bacterial phylogenetic clades were estimated with Mothur v.1.43.0 (Schloss et al., 2009a) and followed the MiSeq Standard Operational Procedure (Kozich et al., 2013). Forward and reverse reads were assembled into contigs followed by filtering and processing the resulting sequences. In brief, sequences with a minimum overlap of 20 bp, which included a length between 580 bp and 348 bp were retained. Sequences with ambiguities or homopolymers longer than 14 bp were removed from the analysis. The filtered sequences were deduplicated and aligned against the V3-V4 region of the SILVA v132 reference small subunit rRNA gene alignment database (Quast et al., 2013). Those sequences that did not span the complete alignment were filtered out by optimizing the start and end positions using a 95% criterion. The resulting alignments were further processed by eliminating columns that contained gaps or dot characters, and the sequences were deduplicated for a second time. Denoising was performed by preclustering sequences with less than one difference per 100 bp, and chimeras were removed using Mothur's implementation of the VSEARCH algorithm (Rognes et al., 2016). Sequences were classified by the Wang method (Q. Wang et al., 2007), using the naïve Bayesian classifier against the SILVA v132 reference taxonomy database and a 70% bootstrap threshold. Sequences belonging to Chloroplast, Mitochondria, and Eukarya were removed. The opticlust algorithm (Westcott & Schloss, 2017) was applied to the resulting sequences for estimating clusters and OTUs at 99% identity. Consensus classifications and representative sequences were determined for each OTU based on the most abundant sequence within each group. After processing the data in Mothur, OTUs were further filtered. We only retained OTUs that belonged to the Bacteria kingdom. Moreover, OTUs not classified at the phylum level and singletons were filtered out with the assistance of the phyloseq package (McMurdie & Holmes, 2013) in R (R Core Team., 2020). Bacterial community structure was explored at various taxonomic levels. All the commands used in the Mothur pipeline for sequence processing are available in the file that is included as supplemental material ("Mothur_v1.43_V3V4_DEF.batch") at annexes.

# CHAPTER 1: CAYAMBE VOLCANO

# 2.1 Cayambe Volcano in Ecuador

## 2.1.1. Background

Glaciers cover about 10% of the earth's surface and ice layers, the largest reservoirs for freshwater (Meier et al., 2007). Despite the ecosystem services that glaciers provide, they are also inhabited by microorganisms, which plays a vital role in these ecosystems by colonizing these extreme environments (Ciccazzo et al., 2016). Although the first factor that comes to mind when we think about glaciers is temperatures below zero, less notorious factors may threaten life at the microscopic level. For example, the UV radiation at high altitudes (i.e., glaciers, mountains) is higher than at low altitudes (at sea level) (Blumthaler et al., 1997; Pfeifer et al., 2006). Since UV radiation is a high-energy source, it can damage microbial macromolecules and induce mutations in the genetic code. The microcosm response to this factor seems to be the use of almost all resources to repair molecular damages (Loveland-Curtze et al., 2009; Miteva et al., 2004). Moreover, in terms of nutrient content, glaciers are a nutrient scarcity niche with the nutrient input of organic dust brought by wind (Duarte et al., 2006; Okin et al., 2004).

Furthermore, we cannot consider temperatures in this type of frozen landscape. For example, the average Arctic temperatures range between -15 and 0 ºC. Moreover, due to the UV radiation, temperatures beneath rock can reach 17-20 ºC for many hours a day (Mevs et al., 2000). In contrast, tropical glaciers, such as Antisana, reported temperatures above 0 ºC with a monthly average of 5 ºC (FARRONA et al., 2016). Microorganism has adapted their molecular machinery and cellular structures to the cold environment. For example, rigidity increments of the cell membrane, the upregulation of cold-shock and cold-active proteins, and other proteins and enzymes essential for the cell cycle were reported (Salwan et al., 2010; Shivaji & Prakash, 2010).

Additionally, we need to add a new factor to the Anthropocene, climate change, to the aforementioned challenging factors. Glacier melting and retreat are one of the main effects of global warming (Raper & Braithwaite, 2006; Shi & Liu, 2000; Sorg et al., 2015). Glaciers in the tropics are considered climate

change biomarkers for their high sensitivity to temperate changes (A. Rabatel et al., 2013; Antoine Rabatel et al., 2018). Therefore, they are more susceptible to a faster retreat, documented over the past 40 years (Małecki et al., 2018). The ice-capped Cayambe volcano has a unique position at zero latitude (0.03º N; 77.988º W), contributing to the glacier retreat of 25.28% from 1979 to 2009 (Gallegos Castro et al., 2018).

Andean glaciers are a fascinating ecosystem where biodiversity has been surveyed at the multicellular level (Bax & Francesconi, 2019; Jacobsen et al., 2010). Moreover, more studies about tropical glaciers' microbial diversity and ecology must be conducted. Therefore, assessing bacterial diversity is of utmost importance to establish the potential for future bioprospection into the psychrophilic word, which can derive from microbial bioproducts (Kuddus & Ramteke, 2009; Margesin et al., 2009). The contribution of microcosm metabolism and ecology, such as primary succession and environmental services, are important reasons to understand the microcosm architecture in these sensitive and rapidly changing ecosystems (Ciccazzo et al., 2016).

Before the invention of high-throughput sequencing technology, the microbiome was surveyed through culture-dependent methods (Ohtonen et al., 1999). It was a minimal approach since microorganisms that were unable to be cultured were left apart. Next-generation sequencing technologies open new microbial assessment methods, especially since they overcome this issue (Zhou et al., 2010). Although the most potent next-generation technology enables the whole metagenome sequence, amplicon metagenomics is a valuable and vital method for the taxonomic profiling of novel environments (Hillmann et al., 2018; Shokralla et al., 2012). Therefore, amplicon metagenomics has been the most accepted method for microbiome surveys in glaciers (Brown, Milner et al., 2007; Wilhelm et al., 2013).

The present chapter’s objective is to explore the temporal microbiome architecture of bacterial communities at the extreme Cayambe glacier ecosystem and provide insights into the relationship between the observed diversity and the associated physicochemical factors. The assessed chronosequence will reveal the microbial dynamics in a glacier that is condemned to disappear and its evolved microbiome with it. Despite the ecosystemic services that Cayambe is currently offering, it provides water to big cities like Quito. Furthermore, is also

dives deeper in ecology theory and proposes the first tropic glacier metacommunity.

## 2.2 Diversity Analysis

### 2.2.1 Alpha diversity analysis

As first glance at the Cayambe volcano's microbial composition over the sampling years, the relative abundance of the ten most abundant families is presented (Figure 2.1). The Effective Shannon diversity and Simpsons index were calculated for all samples. The diversity indexes were compared across sample sites and the two altitude categories with a one-sided Wilcoxon signed-rank test. Rarefaction curves with steps of 600 samples for soil and water were estimated using the ranacapa package's back-end functions.

Then richness and evenness were estimated by rarefaction. For each sample, a subsampling was applied using the minimum library size. For each subsample the InvSimpson and richness observed were calculated. This procedure was repeated 100 times to obtain the mean values. Finally, the alpha diversity was compared according to the year and the altitude of the sample.

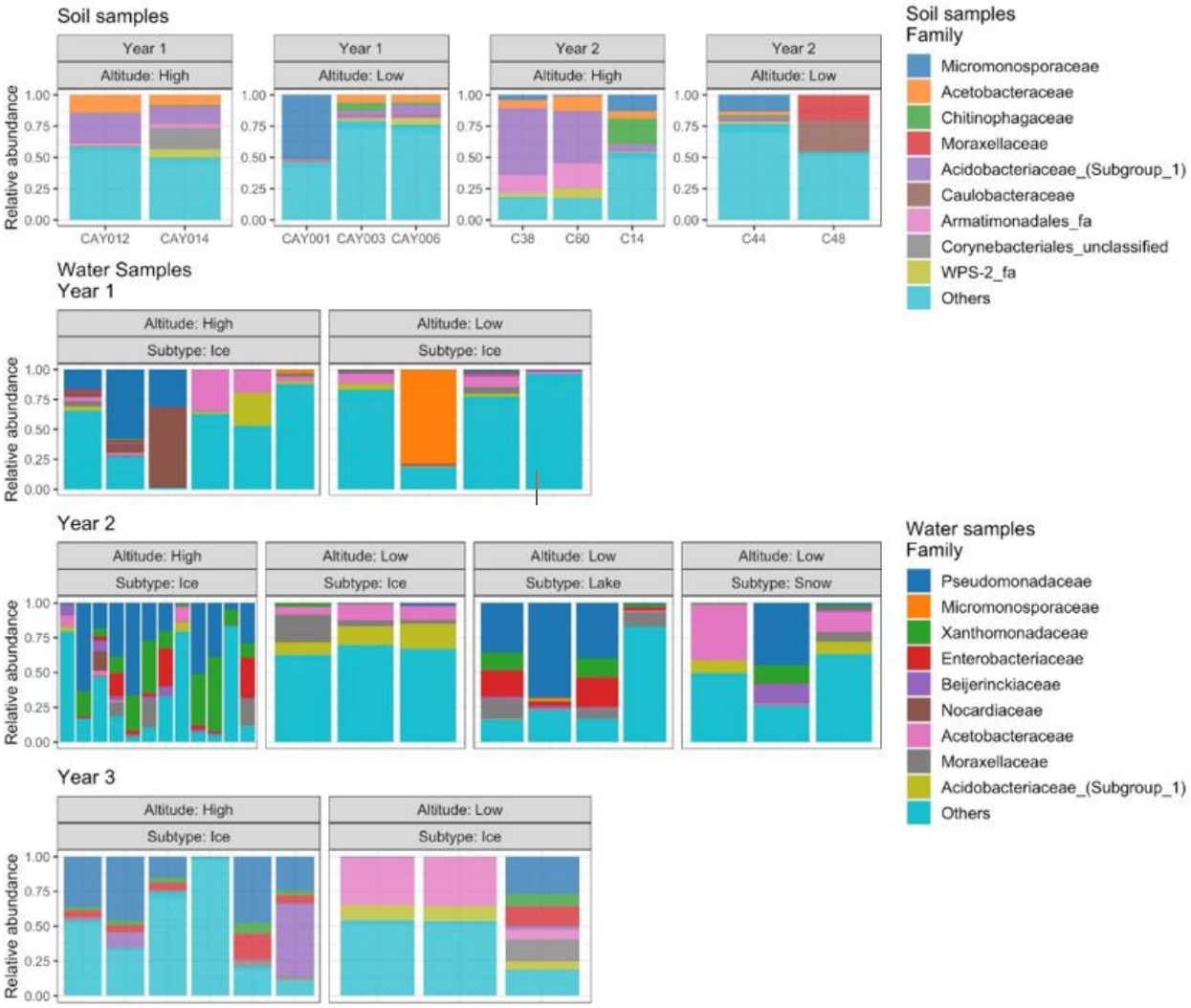



**Figure 2.1. Bacterial Community structure of Cayambe Volcano at the Family level of the through the three years of survey.** The compositional plot of the relative abundance of the 10 most abundant families is grouped for soil (a) and water-based samples (b) (ice-snow-lake). The sampling year and the sample subtype categories are displayed as plot titles.

### 2.2.2 Beta diversity analysis

Before any beta diversity analysis, the abundance table was processed as follows. To eliminate OTU's with a low detection rate, was employed two criteria: (1) each OTU has to be detected at least five times and (2) in more than half of the samples. The filtered phyloseq object was transformed into a Deseq2 object by the phyloseq_to_deseq2() routine (McMurdie & Holmes, 2013). This function requires the experimental design as input; in our case, the altitude (high/low) and time were provided. Then, the abundance table was normalized

with the proposed median of ratios Deseq2 procedure. Here, counts are divided by sample-specific size factors determined by the median ratio of OTUs counts relative to geometric mean per OTU. Finally, since the samples were processed in batches (sampling years), a batch correction was performed. Was employ the suggested routine in the limma package for this task, destined to remove unwanted batch effects in genomics data.

Beta diversity was assessed through a non-multidimensional scaling (NMDS) with fitted environmental (physico-chemical) variables. The algorithm for fitting environmental variables to the NMDS space finds the direction in which environmental vectors change most rapidly. The correlation with the sample scores is the largest; the associated statistical significance in this context is for a null hypothesis the correlation equals zero (Oksanen et al., 2018). The NMDS was based on Bray-Curtis distances obtained from the corrected matrix of abundances for families across samples. We used a radar plot to show the distribution of scaled physical-chemical variables for water and soil samples and grouped by two altitude categories (<5220 m.a.s.l and >5200 m.a.s.l).

### 2.2.3 Differential genus analysis

In order to discover differences in the presence of genera between low and high-altitude samples, a differential genus detection analysis based on a negative binomial distribution was performed with the DESeq2 package (Love et al., 2014). This analysis returned the computed log2 fold change and corresponding p-values. The latter was corrected by the Benjamini-Hochberg method (Benjamini et al., 1995), as a threshold to minimize the false discovery ratio. Genera that was accurate classified at the genus level were projected into a volcano plot, with $-\log_e(p)$ against the log2 fold change. Since the fold change was obtained by low altitude/high altitude abundance ratios, those genera with a positive fold change will express larger abundance at low-altitudes, and those with a negative fold change will express larger abundance at high altitudes.

# 2.3 Results

## 2.3.1 Alpha diversity

A total of 51 samples were recovered from the Cayambe ice-capped Volcanic landscape, ranging from 4,721 to 5,810 m.a.s.l. A total of 2 894,191 high-quality reads were obtained from all samples. On average, soil samples account for 48,668 ± 13,956 reads, while water/ice samples have 58,719 ± 21,402 reads. These totals result after the process of quality control, demultiplexing, denoising, the removal of chimeras, and any sequence from other taxa (e.g., chloroplast, eukarya, archaea, and mitochondrial sequences). The available sequences were classified into 727 OTU.

43 phyla are included in the taxonomic classification of Cayambe microcosms. Was found *Proteobacteria*, *Actinobacteria,* and *Acidobacteria*, in descending order, among the most abundant phyla*.* Although soil and water samples share the order, they differ in relative abundance. For example, while the relative abundance for the phylum mentioned above are *Proteobacteria* (32%)*, Actinobacterial* (22%)*,* and *Acidobacteria* (19%) in water samples, the relative abundance are *Proteobacteria* (65%)*, Actinobacteria* (11%)*,* and *Acidobacteria* (4%) in soil samples (Figure 2.1a).

Soil and water samples share 40% of the 10 most abundant families, being those *Acetobacteraceae, Acidobacteriaceae, Moraxellaceae,* and *Micromonosporaceae.* In soil communities the *Micromonosporaceae* family is superabundant in the 2015 to 2019 years at low-altitudes, but this distribution is not seen as high altitude. Moreover, the *Acidobacteria* is present along the 2015 to 2019 years and in both altitudes (Figure 2.1b). This family's presence is also seen along with the majority of water comunities; it could suggest a cosmopolitan Cayambe family.

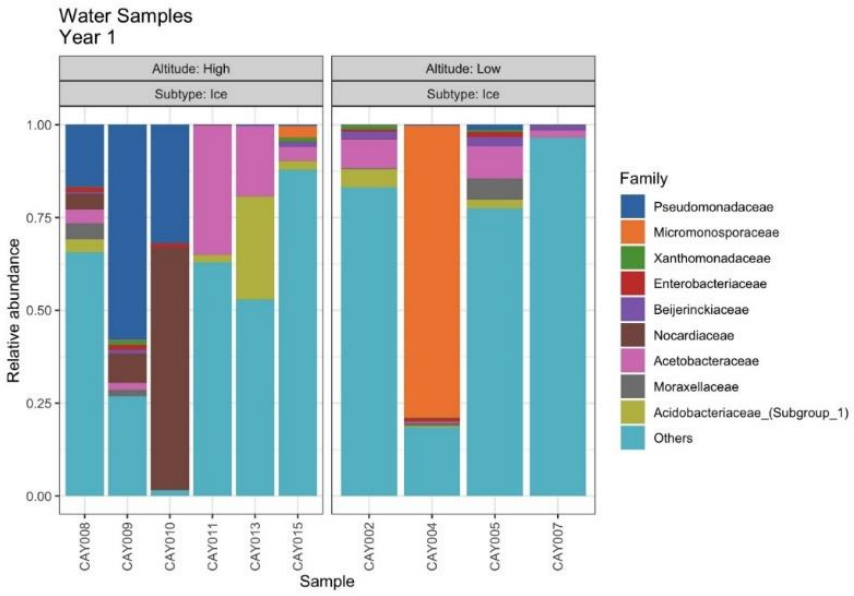



**Figure 2.1a. Bacterial Community structure of Cayambe Volcano at the Family level of the through the three years of survey.** The compositional plot is grouped in the relative abundance of the 10 most abundant families for water-based samples (ice-snow-lake). The sampling year 1 and the sample subtype categories are displayed as plot titles.

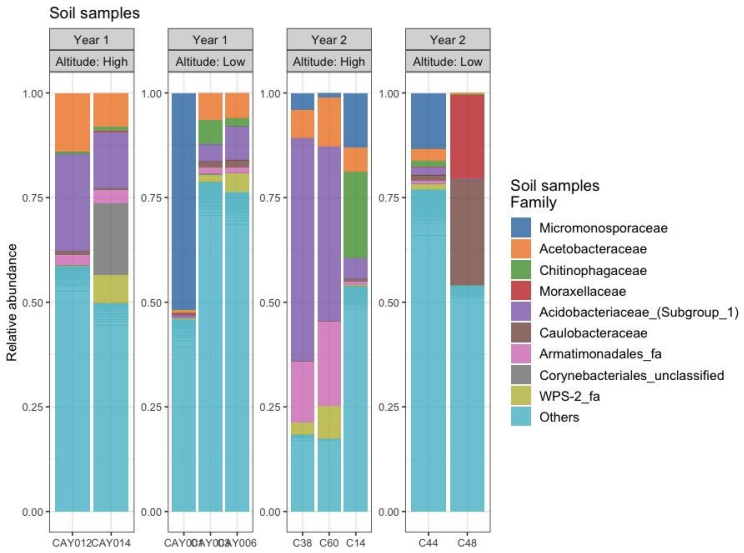



**Figure 2.1b. Bacterial Community structure of Cayambe Volcano at the Family level of the through the three years of survey.** The compositional plot is grouped in the relative abundance of the 10 most abundant families for soil. The sampling year 1,2 and the sample subtype categories are displayed as plot titles.

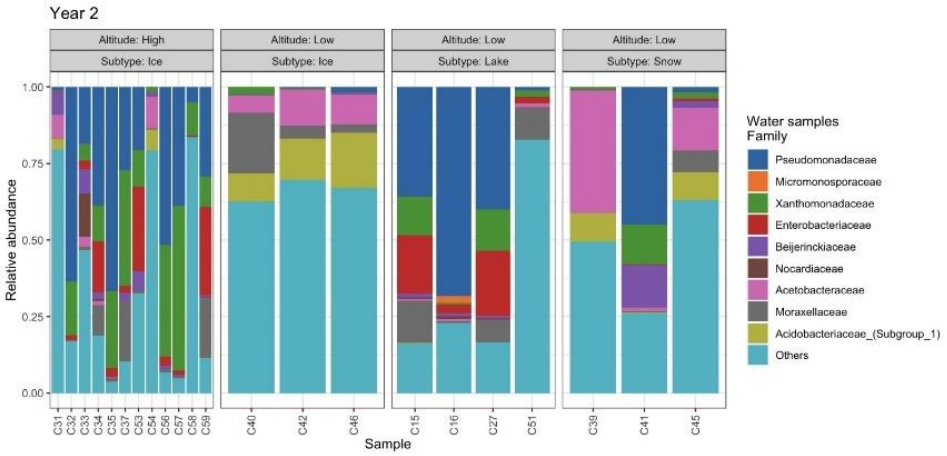



**Figure 2.1c. Bacterial Community structure of Cayambe Volcano at the Family level of the through the three years of survey.** The compositional plot is grouped in the relative abundance of the 10 most abundant families for water-based samples (ice-snow-lake). The sampling year 2 and the sample subtype categories are displayed as plot titles.

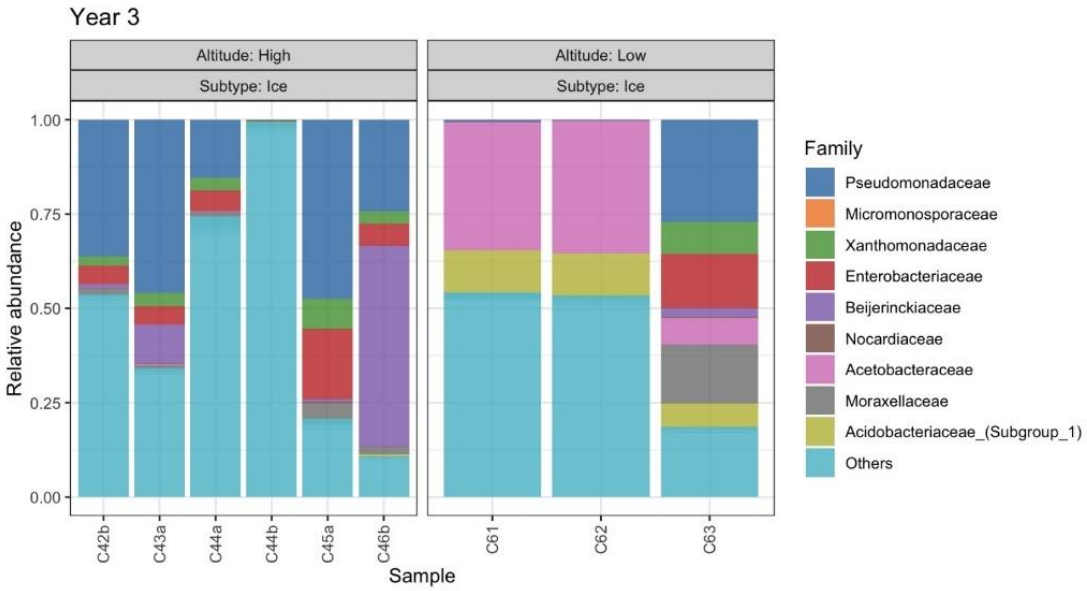



**Figure 2.1d. Bacterial Community structure of Cayambe Volcano at the Family level of the through the three years of survey.** The compositional plot is grouped in the relative abundance of the 10 most abundant families for water-based samples (ice-snow-lake). The sampling year and the sample subtype categories are displayed as plot titles.

Distribution of the alpha-diversity values of soil communities depicts a clear pattern that higher diversity is found at low-altitudes. Even though, the comparison of the Effective Shannon and Simpson index based only on altitudes (high vs. low) did not result in a statistically significant difference ($p < 0.05$) (Figure 2.1a and 2.1b). Moreover, the resulting p-values are around the significance threshold Effective Shannon ($p = 0.11$) and Simpson ($p = 0.15$) what support the diversity trend based on an altitudinal gradient.

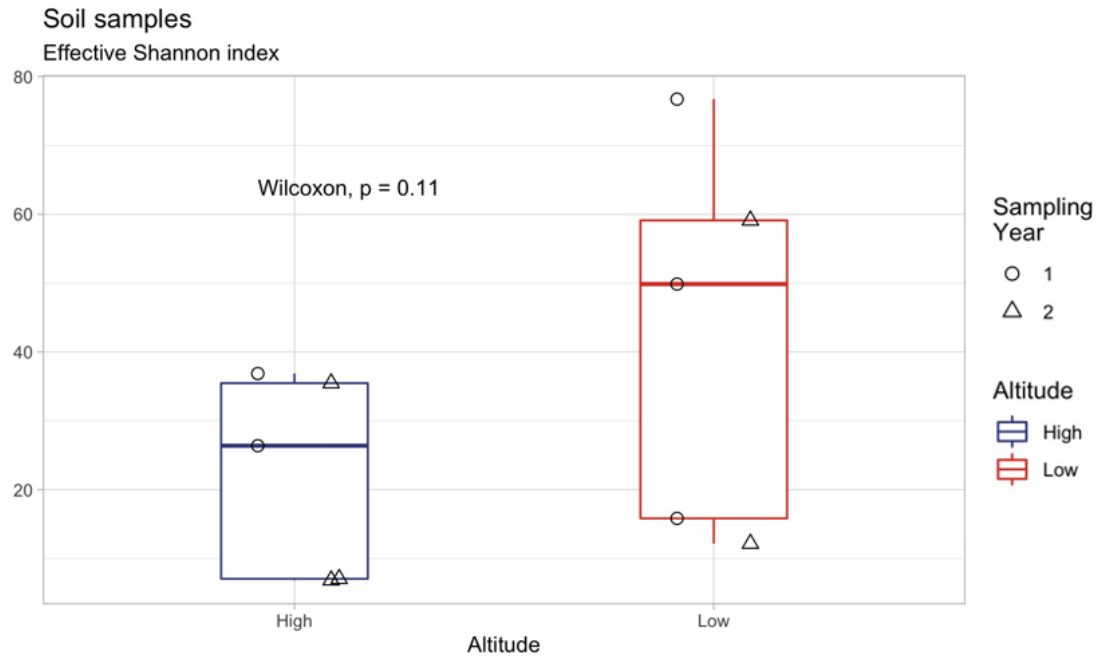



**Figure 2.2a. Cayambe Volcano, Alpha-diversity analysis of the through all years of survey.** Effective Shannon index were compared in term of the altitudinal gradient (High Vs Low) for soil-based samples.

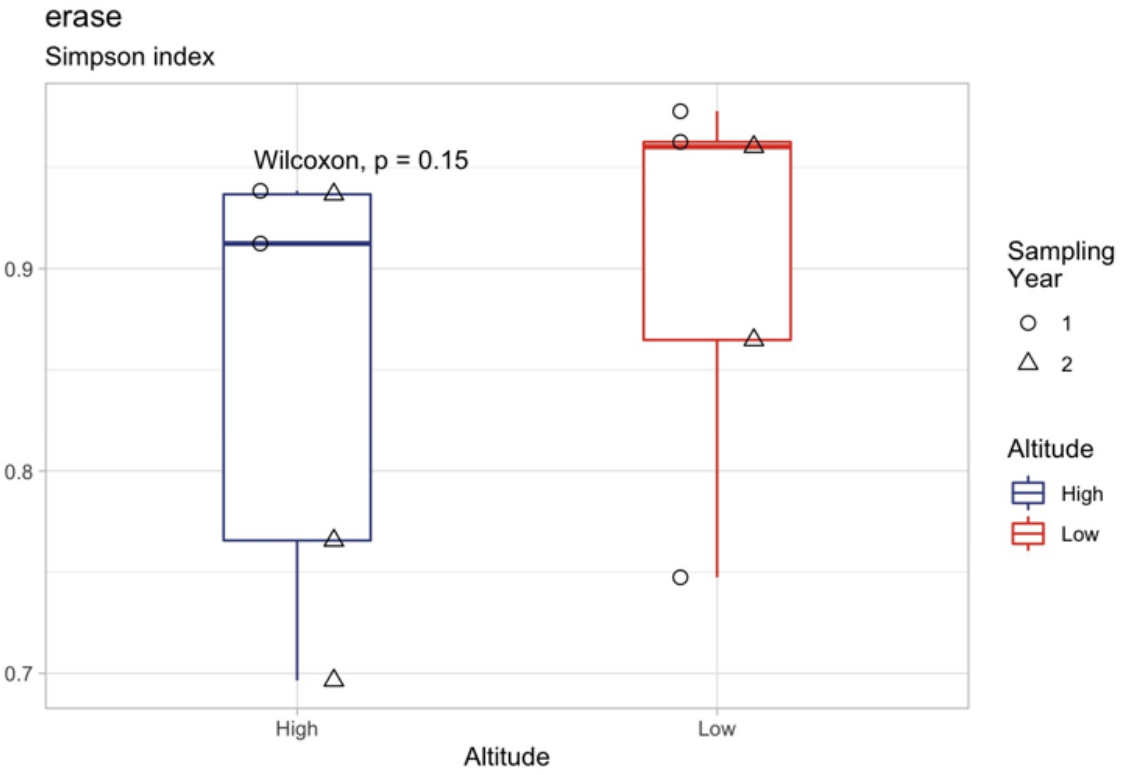



**Figure 2.3b. Cayambe Volcano, Alpha-diversity analysis of the through all years of survey.** Simpson index were compared in term of the altitudinal gradient (High Vs Low) for soil-based samples.

Furthermore, when the time variable is added to the comparison, the difference between high and low-altitudes is more visible (Figure 2.2c).

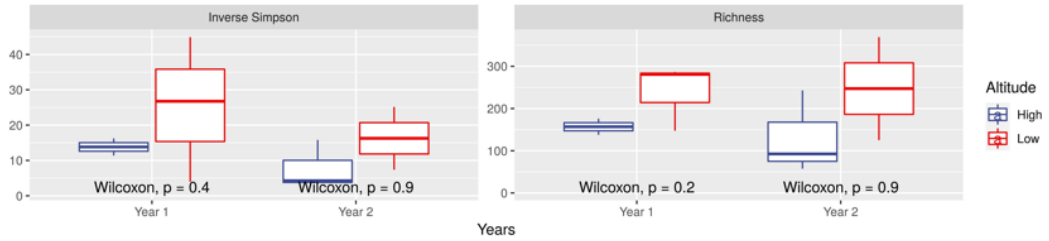



**Figure 2.2c. Cayambe Volcano, Alpha-diversity analysis of the through all years of survey.** Comparison between High and low altitude samples according to the glacier chronosequence for soil-based samples.

The same trend can be seen in water samples. For example, comparing low and high-altitude samples, low altitude samples slightly tend to have a greater alpha-diversity, which also leads to a not statistical significance with either Effective Shannon ($p$ = 0.12) and Simpson indexes ($p$ = 0.13) (Figure 2.2d and 2.2e).

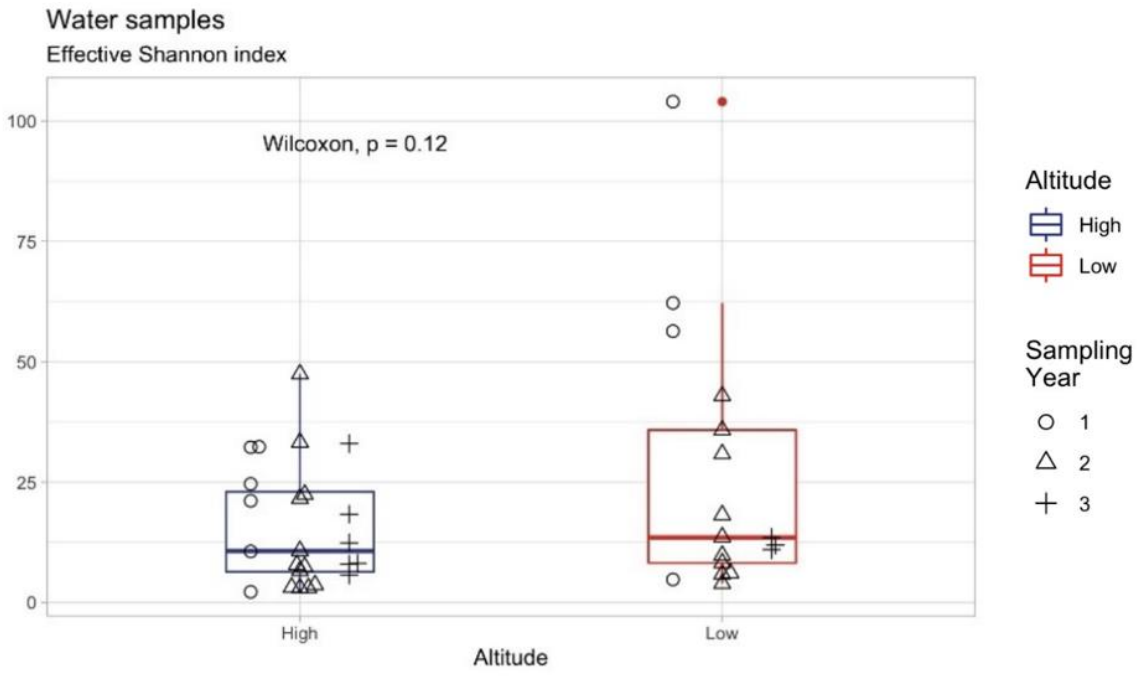



**Figure 2.2d. Cayambe Volcano, Alpha-diversity analysis of the through all years of survey.** Effective Shannon index were compared in term of the altitudinal gradient (High Vs Low) for water-based samples.

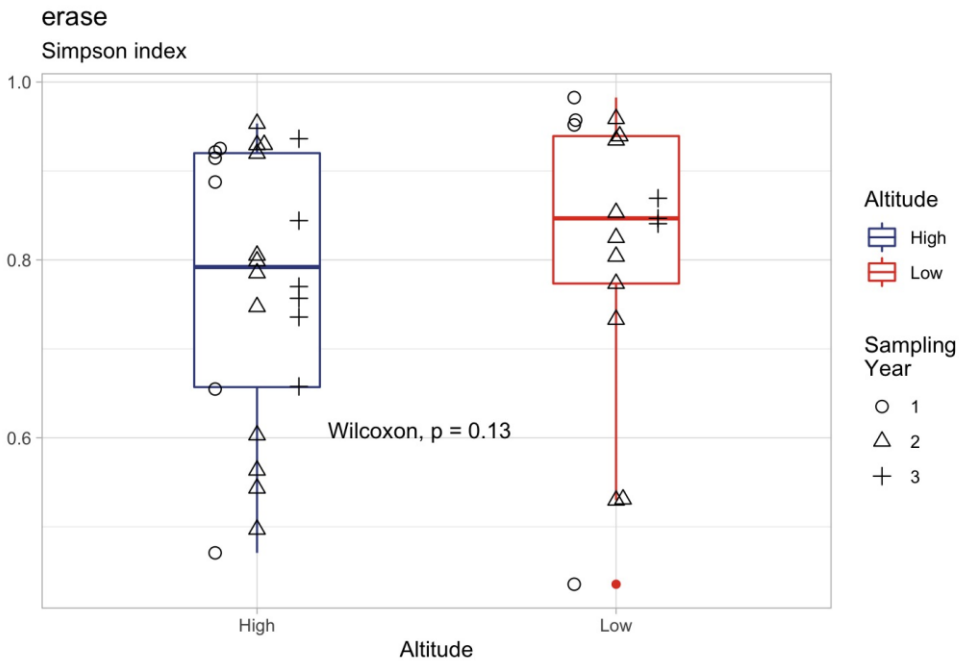



**Figure 2.2e. Cayambe Volcano, Alpha-diversity analysis of the through all years of survey.** Simpson index were compared in term of the altitudinal gradient (High Vs Low) for water-based samples.

In contrast to the alpha-diversity trend over time for soil samples, the difference between high and low-altitudes tends to decrease over time. In other words, the difference decreases over time, making them more similar. Therefore, the lowest p-value obtained for this comparison was for year 1 (p = 0.13) (Figure 2.3). Furthermore, the calculated richness presents an additional behavior; it tends to decrease over time, suggesting a diversity loss.

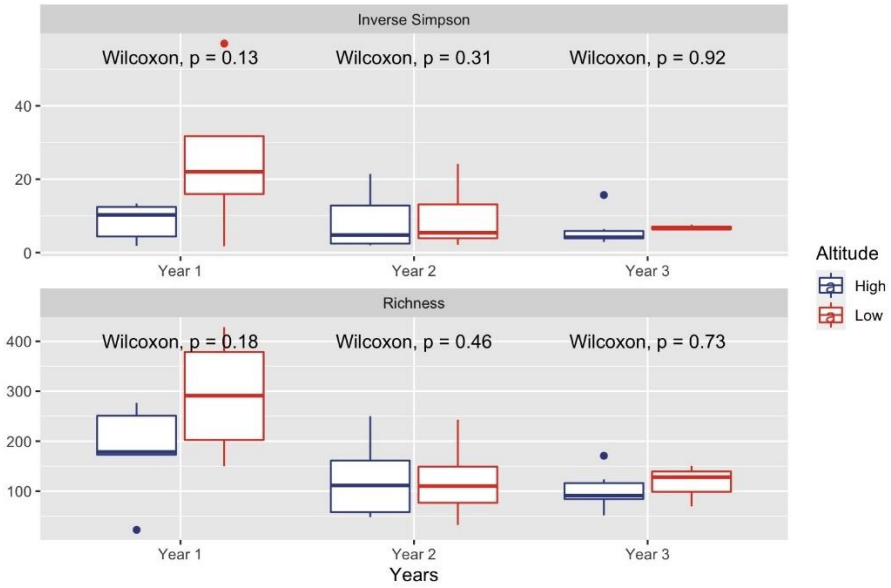



**Figure 2.3. Cayambe Volcano, Alpha-diversity analysis of the through all years of survey.** Comparison between High and low altitude samples according to the glacier chronosequence for water-based samples.

The rarefaction curves for all samples show that they reached the asymptote, with at least 60% of reads, which indicates a sufficient sequencing depth. Surprisingly, with a more significant number of water samples, its curves also exhibit decreasing maximum richness genus over time, being the last year, the third one, with the lowest richness ever sampled (Figure 2.4 and 2.5) For all these analysis, alpha-diversity indexes were subjected to one-sided Wilcoxon test. Rarefaction curves for soil and water-based samples.

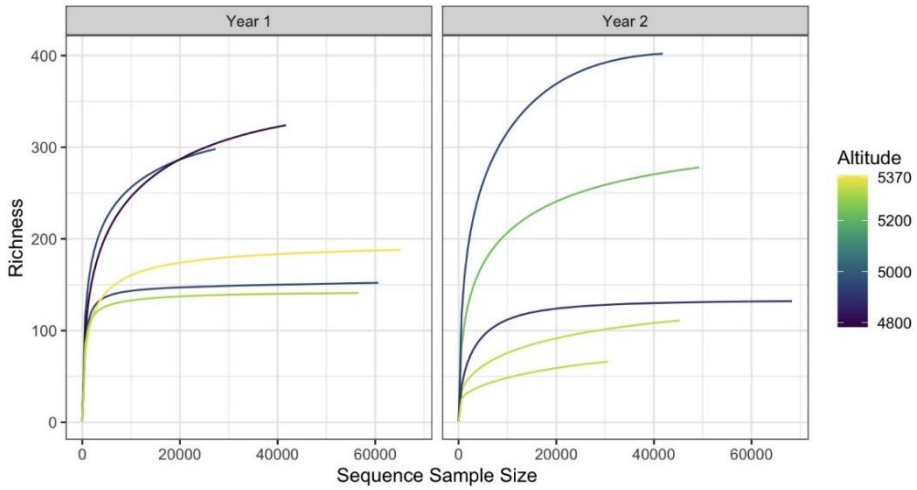



**Figure 2.4. Cayambe Volcano, rarefaction curves of the through all years of survey.** The curves show the asymptote with al least 60% of reads, based only in all soil-samples.

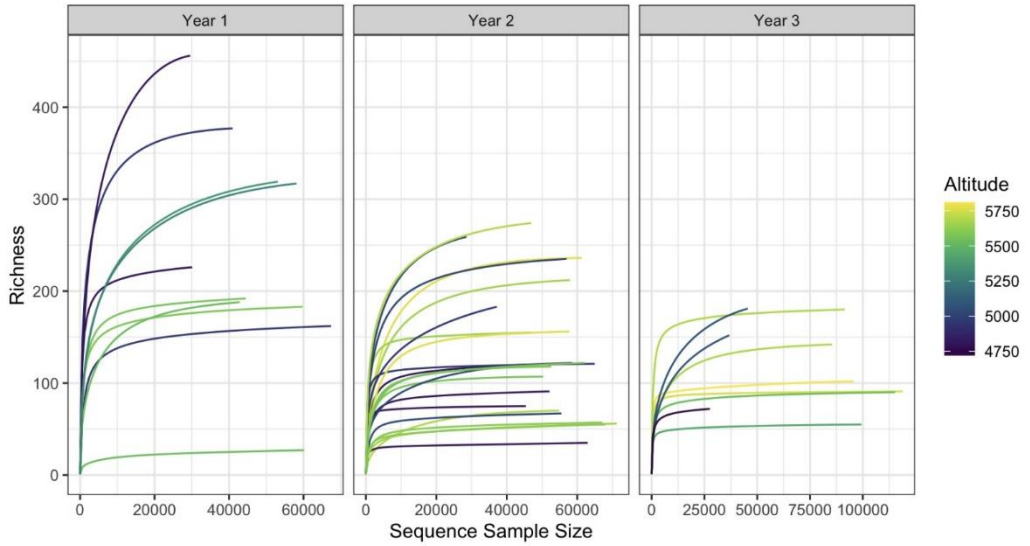



**Figure 2.5. Cayambe Volcano, rarefaction curves of the through all years of survey.** The curves show the asymptote with at least 60% of reads, based only in all water-based samples.

### 2.3.2 Beta diversity

On the other hand, the beta diversity analysis comprises two main approaches, the NMDS, and hierarchical clustering analysis. The NMDS analysis represents the distances among samples regarding the differences in abundance for genera, soil, and water communities.

The NMDS analysis for soil samples can be separated into two well-differentiated clusters. High-altitude samples populate the right NMDS dimension with a narrow dispersion, while low-altitudes are broader in the left NMDS space (Figure 2.6a). Moreover, only one high-altitude sample (C14) can have a different microbiome composition since it is located outside the high-altitude sample cluster. It can be firstly approached with its most abundant families (*Burkholderiaceae, Caulobacteraceae, Moraxellaceae*). Two of these three families are the most abundant in this sample, and *Burkholderiaceae* is the more abundant in this sample than in others. Three out of the fourteen environmental variables measured resulted significant being Mn ($p$ = 0.041), $N$ (p = 0.013), Humidity ($p$ = 0.013), (Figure 2.6b).

On the other hand, since the NMDS of water samples was performed with all sample subtypes as a unified dataset, the resulting NMDS did not show a clear

cluster either for the altitudinal gradient. Therefore, we decided to split the resulting NMDS coordinates into three different layouts for a cleaner view per sample subtype. In other words, we took the same NMDS sample coordinates, and we display them in three different cartesian planes for ice (Figure 2.6c), snow (Figure 2.6d), and lake (Figure 2.6e).

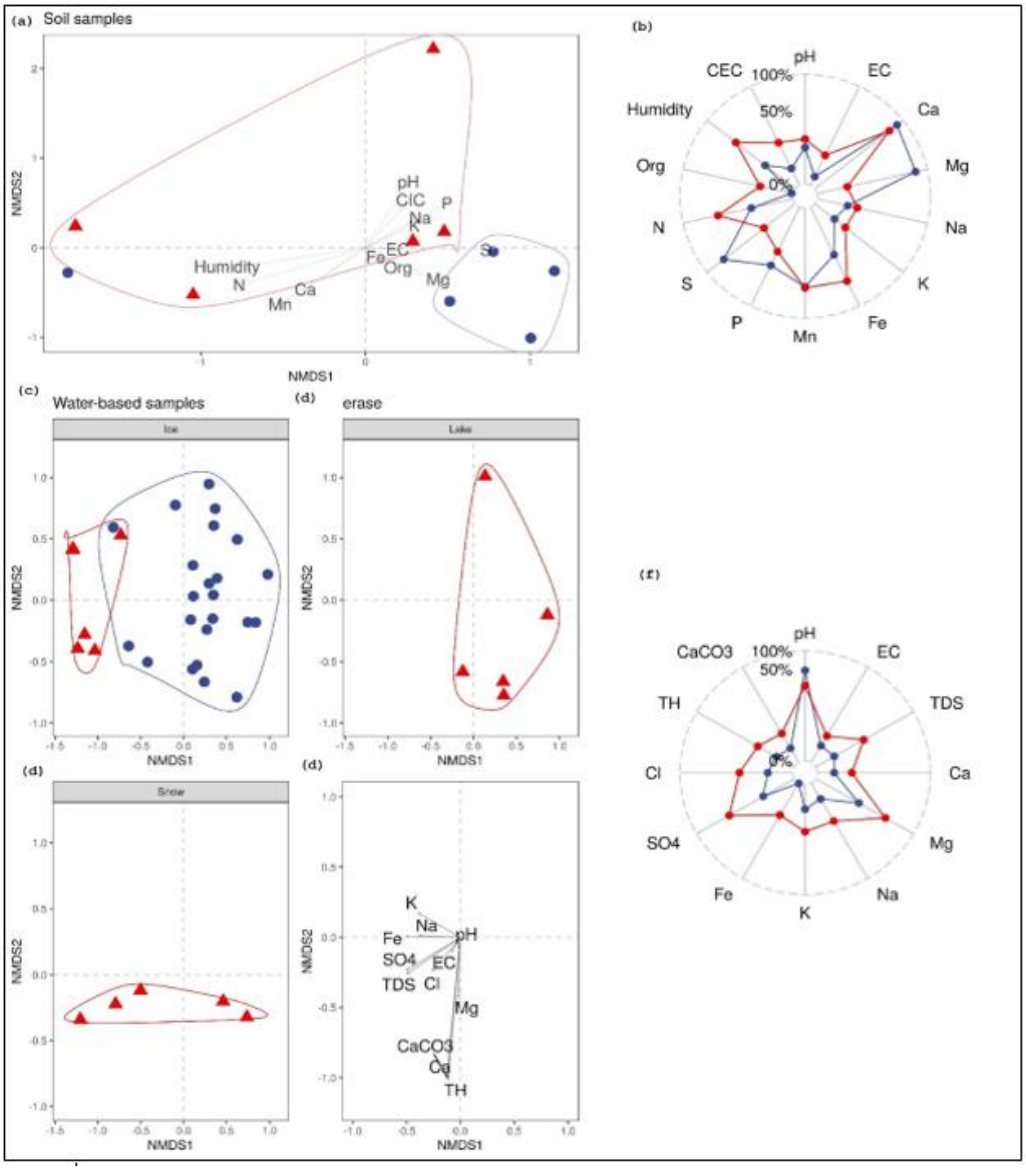



**Figure 2.6. Beta-diversity analysis**. (a) NMDS of the soil samples (stress = 0.17), environmental variables were fitted to the NMDS space. (b) Radar plot of the assessed environmental soil samples. (c, d, e, and f) NMDS of the water-based samples. The ordination analysis was conducted with the three sample subtypes (ice, snow, lake), and, as a visualization technique, the scores are displayed in three different cartesian planes for ice (c), lake (d), and snow (e). The fitted environmental variables are displayed in an additional plane (f). (g) Radar plot of the measured environmental variables for water-based samples.

Additionally, an extra NMDS plot were added to present the fitted environmental variables onto the NMDS space (Figure 3.4f). After this data visualization trick, each subtype's altitudinal gradient becomes clear. In contrast with the three different environmental variables found significant for soil matrices, in water matrices, three variables intrinsically connected were found significant (Ca, TH, and $CaCO_3$). The Ca content was measured directly as an element, and TH is a variable related to a compound with calcium in it (calcium carbonate). Finally, the $CaCO_3$ content was calculated in function of TH; therefore, it is not surprising, that these three variables resulted significantly.

### 2.3.3 Differential genus analysis

The differential genus analysis can be grouped in two different comparison categories for each, soil, and water-based samples separately. The first one is related to the altitudinal gradient (high vs. low), and the second is about the temporal variable (sampling years). In the comparison for the altitudinal gradient, the fold change is based on the high/low relationship. Therefore, a genus with negative fold change values means the specific genus is found at greater abundance at low-altitudes, and genus with positive fold change values means the specific genus is found at greater abundance at high altitudes.

Soil samples subjected to this analysis, resulted in two genera with a p-value < 0.05, *Nitrobacter* (p = 0.003) and *Granulicella* (p = 0.048) (Figure 2.7).

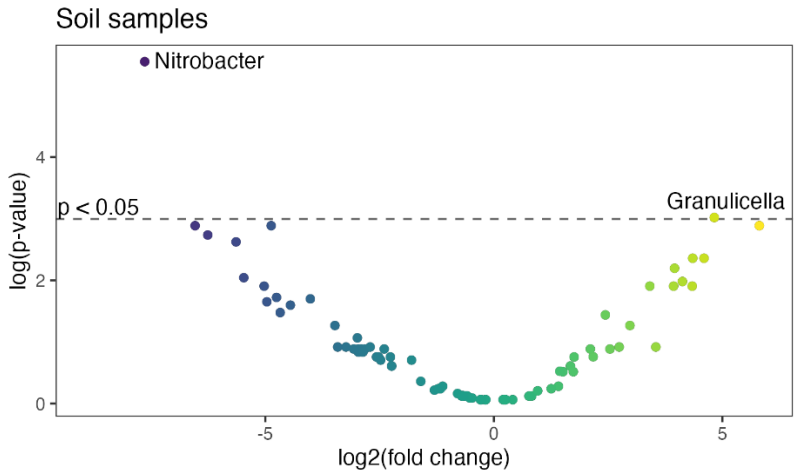



**Figure 2.7. Cayambe Volcano plot.** Shows two genera with a p-value < 0.05. based only in all soil-based samples communities.

While *Nitrobacter* has a negative FC value, *Granulicella* has a positive FC value. In other words, *Nitrobacter* can be found at the greater abundance in low altitude samples, meanwhile, *Granulicella* can be found at the greater abundance in high altitude samples. On the other hand, water samples subjected to this analysis resulted in five significant genera, moreover, three of them were classified until the genus taxonomic level. In this list we have *Pseudomonas* (p = 0.005), *Stenotrophomonas* (p = $0.6x10^{-3}$), *Bacillus* (p = $0.3x10^{-3}$) (Figure 2.8).

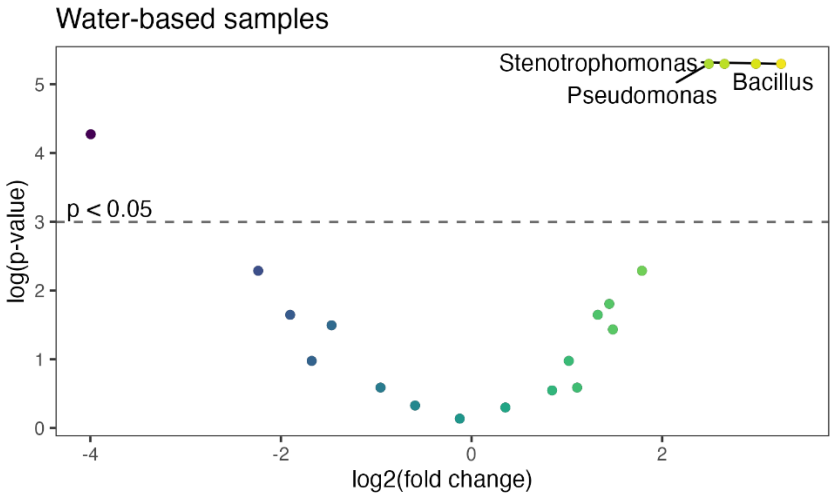



**Figure 2.8. Cayambe Volcano plot.** Shows significant genera with a p-value >= 0.05. based only in all water-based samples communities

## 2.4 Discussion

It is well known that global warming is one of the most threatening consequences of climate change. Between its downstream environmental and ecosystems havoc, glacier retreats are one of them. Although glaciers are located worldwide, glaciers located in the tropics, especially those around the equatorial line, are more susceptible to global warming (Vuille et al., 2015). For example, in 2007, Chacaltaya glacier located in Bolivia, disappeared (Beeman & Ruiz, 2018). Therefore, the haste to understand the glacier biome before its imminent disappearance (Vuille et al., 2008). Our study on the ice-capped andean volcano microcosm encompassed three main variables: glacier biomes (i.e. water and soil), altitudinal gradient (i.e. high and low), and temporal sampling (year 1, 2, and 3).

### 2.4.1 Alpha diversity

Briefly, our findings suggest a clear difference in soil samples for high and low altitude for each year, exhibiting a greater diversity at low altitudes. On the other hand, this pattern is only seen in water sample in the first year, and the difference between high and low altitudes decreases over the time, making them more similar. Furthermore, other behavior can also be established in water samples, the average richness tends to decrease over time, no matter what altitude the sample is. Since several studies has been carried out in other glaciers around the world, at a given point of the time, we can state that a similar pattern can be found in other microcosm such us in Tianshan Mountains (central Asia) (Ren et al., 2017) and also in macrofauna in arctic and alpine glacier-fed streams (Jacobsen & Dangles, 2012). Another measure related to the altitudinal gradient is the distance

To the summit, in this case, the closer the sample is to the summit, the less biodiverse it is. Since glaciers are landscape with a low nutrient concentration (Jacobsen et al., 2014), the contribution of snow melts and ice melt from high altitudes increase the nutrient concentration at low-altitudes (Cooper, 2010) with a major implication in the microbiome (Brown, Hannah, et al., 2007).

Furthermore, the primary succession of glacier retreated areas depict a clear biodiversity pattern. It started with the lowest diversity indexes, but then, after macro and micro colonization, diversity increases until reach a maximum value. This behavior was documented since 70's in the chronosequence at glacier Bay (Alaska) (Reiners et al., 1971). This approach was replicated on glacier stream invertebrate communities in southest Alaska from 1978 to 2004, with a similar increasing richness over time (Brown & Milner, 2012). Additionally, a 150-year chronosequence in a glacier foreland of the high Artic, focused on the bacterial diversity, showed an increasing richness over time (Schütte et al., 2010a). In this study we found a different behavior for the richness, a decreasing one. We notice that the sampling range for this study is five years, and it is not comparable with the previous mentioned studies. With this narrow time window, we are assessing the biome at probably one of the finest scales of the Cayambe chronosequence. Therefore, we can expect, with a continuous and future sampling efforts, to have a long-term picture of the glacier ecosystem.

Moreover, the decreasing diversity pattern over time for water-based samples might be explained for environmental factors and sample characteristics. In the two first sampling time points, we can still notice differences between high and low altitudes for the aforementioned reasons. Meanwhile, the decreasing average richness over the time might be explained for the continuous temperature elevations in high altitude Andean mountains (Vuille et al., 2008, 2015). With the temperature increment, microbiome, previously adapted for colder temperature, will no longer resist this change. Moreover, the temperature increment is not the only change that the microbiome has to deal with, UV radiation has to be added to the formula. Since the temperature is dependent on the UV radiation, the microcosm is greater affected by it. The UV radiation can damage macromolecules, which lead to the cell dead. Therefore, it is hypothesized that in this nutrient scarcity environment, microbes use the entire resource input in repairing macromolecular damages (Miteva et al., 2004; Schütte et al., 2010a). Finally, the similarity between high and low altitude samples can be explained by its matrix characteristics, ice matrix. The ice matrix has been shown to harvest microbial communities (Miteva et al., 2004; Schütte et al., 2010a). Thanks to solutes' presence, the freezing point of glacier melt waters is lower than pure waters. It creates a mixed environment inside ice cores, where micron-sized liquid water veins can be found (Lutz et al., 2015). Therefore, as a result of the ice formation process, solutes are left out from crystals units, which creates a favorable environment for microbes (Price, 2007).

Additionally, the ice sampled in Cayambe, was taken after removing a superficial layer of snow. The sampling location results in a more constant temperature than the highly variable temperatures at the immediate snow surface. These matrix features create a more amenable environment with less threatened variables that cause similarities between high and low altitude diversities.

### 2.4.2 Beta diversity

In terms of beta diversity, soil samples seem to have a more structured and fixed community, specifically at high altitudes. This argument is supported by the lower dispersion of the high-altitude samples in the NMDS space, while

low-altitudes are spread in a higher NMDS space (Figure 4a). Only one exception must be taken into concentration for high altitude samples, placed in the most negative coordinate of the first axis. The answer to this displacement of the high-altitude samples cloud is merely its community composition, which can be seen in the soil samples heatmap. The *Micromonospora* genus is a superabundant genus in this sample, making a single clade in the HC analysis for this sample. The genus *Micromonospora* has been previously reported in glaciers in China (Christner et al., 2002; Mosley-Thompson et al., 2000), in Svalbard (Amato et al., 2007), and in Himachal Pradesh (Raja & Prabakaran, 2011). The peculiar high abundance in this sample might be due to the presence of a *Micromonospora* antibiotic specie producer (Boumehira et al., 2016).

The environmental variable that resulted in a significant difference for high and low soil samples is N. It was expected since nitrogen constitutes a building block for biomolecules, and it is only available through the nitrogen fixation process. In contrast with agricultural soils where Proteobacteria almost entirely carry out the nitrogen fixation process, *Firmicutes* were found a vital actor of this process in polar soil environments (Deslippe & Egger, 2006). According to our results, the sightly difference of 1% in the relative abundance of *Firmicutes* between high (3%) altitude and low altitude (4%) samples could lead to the difference in nitrogen content. Humidity is among the significant environmental factors. It can influence the microorganism activities, even though a specific variable-genus relationship could be made. Finally, the Mn content was found significant for the altitudinal gradient. However, any bacterial species was strongly related to Mn content in glaciers; fungal species were described to dissolve glacier granite power by exudating citrate, malate, and oxalate increases the Mn concentration (Brunner et al., 2011). It increases the attention to include other microorganisms rather than bacteria to better understand the glacier microcosm, which can be done in future works.

On the other hand, only the ice sample subtype can be contrasted between high and low altitude samples. Due to lower temperatures at high altitudes, there are more samples at this altitude. Moreover, the low-altitude ice samples were recovered in the third sampling year, and as aforementioned

explained, they were very similar in terms of alpha-diversity. Therefore, it is not surprising that both, high and low altitude clouds are together. Moreover, the lake microbiome is the particular interest for future work. High altitude lakes could act as harbors of airborne bacteria. For example, the lake's air-water interface in a remote area in the Pyrenees was like airborne bacteria than those in deeper water layers (Hillmann et al., 2018). It constitutes a new input for the glacier microcosm. Moreover, this input source is material for further research.

Additionally, it has been reported that the lake bacterial community is shaping by seasons, where the winter growth snowpack and summer melting phase plays a vital role (Llorens-Marès et al., 2012). Future teams can develop a study to answer what is the seasonal bacterial variation of the Cayambe lake. Finally, three intimate environmental variables were found significant for water samples. Moreover, we could not specify a relationship between these variables and the bacterial community architecture.

### 2.4.3 Differential genus analysis

*Nitrobacter* and *Granulicella* were found significantly different over time in soil samples for the altitudinal gradient. *Nitrobacter* has an average relative abundance in low altitude samples of 0.6%, while in high altitude samples are found at 0.01%. This genus was reported previously in glacier soils for playing an essential role in the nitrogen cycle by using energy from the oxidization of nitrate to fix $CO_2$ via the Calvin cycle (Latha et al., 2009), and has been proposed as responsible for carbon fixation as a chemoautotrophic bacterium (Werner & Newton, 2005). In contrast with the *Firmicutes* genus found at a greater abundance in Cayambe, *Nitrobacter* has a well specialized nitrate-oxidizing molecular machinery (S. Starkenburg et al., 2011; S. R. Starkenburg et al., 2006) that might have a greater contribution in the nitrogen cycle. *Granulicella* has been described to be present in various points of glaciers chronosequences. For example, four new species of this genus were reported from Arctic (Männistö et al., 2012), in tourism-related contaminated alpine glaciers (Cappa et al., 2014), and the glacier forefield chronosequence (Franzetti et al., 2020; Jiang et al., 2018), what leads us to suggest that *Granulicella* is a cosmopolitan glacier

bacterium for the long-term glacier chronosequence, from ice-capped glaciers to forefileds.

*Pseudomonas, Stenotrophomonas,* and *Bacillus* were found significantly different in water samples based on the altitudinal gradient. Pseudomonas genus plays a variety of roles in the glacier ecosystem. For example, *Pseudomonas fluorescencs* is an ice-nucleating bacteria (Obata et al., 1998). The higher abundance of *Pseudomonas* at high altitudes (28%) than at low-altitudes (14%) may be related to this bacteria's ice-nucleation capability. Additionally, *Pseudomonas* has been reported as a psychrophilic P solubilizer (Pandey et al., 2002, 2006). Therefore, the higher presence of the *Pseudomonas* at high altitudes could also explain the greater P concentration. Antibiotic resistance was also proved on *Pseudomonas* isolates from glaciers in India, where the susceptibility test showed that at least 25% of the isolates (6 isolates) had antibiotic resistance for the four-antibiotic employed (Ampicillin, Streptomycin, Vancomycin, and Tetracycline) (Sherpa et al., 2019). Even though *Pseudomonas* showed antibiotic resistance, isolates from glaciers in Pakistan presented biotechnology potential due to the production of thermolabile protease and alkaliphilic lipase, both oriented to the detergent industry.

*Stenotrophomonas* has been found in different glacier ecosystems, such as in China (Mosley-Thompson et al., 2000) and India (Josiah et al., 2018; Salwan et al., 2010). This genus has also caught the industry's attention, especially the detergent industry, since a cold-active, alkaline, and detergent-stable protease was isolated from the Gangotri glacier (Kuddus & Ramteke, 2009; Mohammed & Pramod, 2011). According to our results, *Stenotrophomonas* is found at a greater abundance at high altitudes (10%) than at low-altitudes (3%), being high altitude the best option to for bioprospection. Bacillus genus has been widely reported to be found in glacier ecosystems, for example, in glacier sediment from Svalbard Islands (Stathopoulou et al., 2019), Glacier Cryoconite of Nepali Himalaya and Greenland, Arctic (Singh et al., 2020), North Sikkim, India (Sherpa et al., 2018, 2020), and Loahugou glacier, China (Ali et al., 2020). However, its ecological role has not been discovered yet.

### 2.4.4 Cayambe microbiome metacommunity

While traditional ecology emphasizes niches as discrete environments, the metacommunity theory assumes that niches are not isolates and might be considered continuous until a certain extent. We can define the spatial dynamics in niche and community interactions, which accounts for mass effect, rescue effect, colonization, and dispersal effect between the main interaction factors (Hanski, 1991). The ecology of glacier fit as a permanent habitat with indistinct boundaries, since glacier soil, snow, ice, and glacier lakes are intimately in contact, facilitating colonization and dispersal effects (Leibold et al., 2004). The glacier metacommunity is built up with the contribution of different glacier niches such as cryoconite holes (Edwards et al., 2011), glacial ice (Margesin et al., 2009), groundwater (Griebler & Lueders, 2009; Margesin et al., 2009), adjacent soil (Schütte et al., 2010b), and lakes (Hervas & Casamayor, 2009; Llorens-Marès et al., 2012).

The glacier metacommunity was firstly reported in 2013 for the Alpine glaciers (Wilhelm et al., 2013). According to our data, the Cayambe glacier's metacommunity that was found in all sampled years, at both altitudes and at any sample type, is composed of 31 genera (Figure 5a). The predominat phylum in the metacomunity is *Proteobacteria* with 51%, *Bacteroidetes* and *Firmicutes* with 12% each, *Acidobacteria* and *Actinobacteria* with 9% each, and *Planctomycetes* with 3%. The first five phyla that we report are in line with the metacommunity previously reported for Alpine glaciers (Wilhelm et al., 2013), even when the sample types (ice, biofilm, and streamwater) in that study differ from our sample types. Additionally, except for *Planctomycetes,* the rest of the phylum are reported to be the most abundant in different glacier landscapes (Lutz et al., 2015; Malešević et al., 2019; Simon et al., 2009; Srinivas et al., 2011; Xiang et al., 2009). Therefore, it was expected that the metacommunity belongs to this phylum.

Additionally, we explore beyond the entire Cayambe glacier metacommunity, and we build a specific metacommunity for water-based samples, which were split into three different glacial matrixes: ice, snow, and lake. In this case, a total of twenty-three genera were found to be part of the metacommunity. In compass with the entire glacier metacommunity, *Proteobacteria* is the predominant phyla with 47%. The ternary plot for these

three glacier niches showed that most bacterial genera are inclined to be found more often in the snow and ice niches. This can be explained due to the intimate contact that snow and ice niches have through the dispersal effect, while the lake microbiome is built up with snow and ice biome contribution through the colonization effect (Leibold et al., 2004).

### 2.4.5 Bioinformatic sources

Different factors such as the method of choice for sequencing, the type of Illumina sequencer, the sample preparation technique, the clear use of primers and indexers, the PCR itself which apparently does not involve an observable error effect, constitute errors of next generation sequencing (NGS) currently employed to obtain a large number of sequences.(Pfeiffer et al., 2018) However such sequencing errors hinder the recognition of biologically real existing nucleotides against all 16S rRNA sequencing artifacts. In order to avoid introducing variables coming from these sequencing techniques, what is usually done is to group sequences into Operational Taxonomic Units (OTU's) whose identity threshold is usually 97%, although risking taxonomic resolution. However, this is now beginning to be resolved thanks to the proliferation of different bioinformatics programs whose main workflow is rather to "de-noise" the sequences, and thus aim to avoid these errors and improve taxonomic identity. Therefore, many algorithms exist and differ from each other, by the way they correct these errors introduced in the sequences (Nearing et al., 2018). In this work we used Mothur v.1.42.0 open-source software used by the Schloss lab to process their 16S rRNA gene sequences that are generated using Illumina's MiSeq platform using paired end reads. The approach we take is to use index reads to multiplex many samples in a single run, clustering the sequences into OTUs'' and, eliminating reads derived from predicted errors of neighboring sequences, further employs a sample-by-sample approach that reduces both memory requirements and computational demand (Schloss et al., 2009). Instead DADA2 an open source software described to model and correct errors in Illumina sequenced amplicons is trained throughout the sequencing process by generating a parametric error model and then applies that model to correct and collapse sequence errors into so-called amplicon sequence variants

(ASVs) This approach is advantageous as it builds unique error models for each sequencing run.(Callahan et al., 2016) In the joint data analysis studies that DADA2 performed it consistently attracted more ASVs than two other denoising pipelines employed in those studies and that again the open reference OTU clustering called the most overall. But there is something important to note and that is that they reported in the soil dataset studied both DADA2, and the OTU clustering pipeline were able to cluster more low abundance organisms (Pfeiffer et al., 2018). Finally, Mothur was employed in this work because at the time 2018-2019 it was the best performing algorithm for 16S rRNA amplicon sequence analysis and its implementation for the analysis of these samples from water and volcanic soils was under development. If we submit to a comparison between the data obtained from the analysis with Mothur Vs DADA2, considering only the clustering factor in OTU for the one case and ASV for the other respectively, we can observe that the results are approximate.

## 2.5 Conclusions

The Cayambe volcano showed a noticeable inverse relationship between altitude and diversity, within a range of 1089 m between 5810 and 4721 m.a.s.l. In contrast with other glacier forefield chronosequence that directly increments bacterial community richness due to the long-time scale (decades), the Cayambe chronosequence showed the opposite behavior, primarily for the narrow five-year range. We expect to have a similar chronosequence when the primary succession takes place in this ecosystem. We also proposed a tropical glacier's first metacommunity considering different niches in Cayambe, such as soil, ice, snow, and lake subenvironments. The metacommunity is mainly assembly by *Proteobacteria, Bacteroidetes, Firmicutes, Acidobacteria*, *Actinobacteria, Planctomycetes.* It also converges with the dominant phyla previously reported for glaciers. We propose that the mechanisms that creates this highly structured community are the colonization and dispersal effects.

Except for Ca, $CaCO_3$, and TH in water samples, we found low correlations between the geochemical variables and the observed bacterial community structure. Moreover, N, Mn, and humidity were found to have an essential role as community shaping variables. For example, Nitrobacter, a

nitrite-oxidizing genus, was found at high altitudes, where N concentration is higher. *Pseudomonas* is one of the most exciting genera found in the glacier biome since it has been reported with various ecosystem roles. For example, P-solubilizer bacteria as well as ice-nucleating bacteria. However, a robust antibiotic resistance was also reported for *Pseudomonas* isolates from glaciers.

Furthermore, the most promising genus, according to our analysis, is *Stenotrophomonas*. This genus was reported to be capable of producing cold-active enzymes with applications in the detergent industry. According to our data, the best place to start the Stenotrophomonas bioprospection is at high altitudes. Furthermore, although *Granulicella* has been described as a possible biomarker of primary forefield succession, we found *Granulicella* to be part of the Cayambe biome, suggesting a cosmopolitan presence throughout the entire glacier chronosequence.

Although this work is the first in the equatorial tropic glaciers, various questions remain unanswered. The main question is related to the bacterial community architecture for the upcoming chronosequence and can only be revealed by future microbial surveys. Since we employed the amplicon metagenomic methodology, we were able only to perform taxonomic profiling. The next step is to employ more advanced techniques to discover functional diversity. In consequence with the finding of the high Stenotrophomonas abundance, bacterial culture and isolation should be carried out to determine the potential expression of bioproducts of interest before the glacier retreat.

# CHAPTER 2: SUMACO VOLCANO

# 3.1 Sumaco Volcano

## 3.1.1 Introduction

Altitude is one of the major driving variables to shape biodiversity as well as edaphic properties. High altitude environments are generally characterized by low temperature, nutrient scarcity, fluctuating precipitation, and more UV radiation exposure (Blumthaler et al., 1997a; Morán-Tejeda et al., 2013). These cold landscapes have been successfully colonized by microorganisms that are well adapted to the permanent cold temperatures. Between its main adaptation there are reduced enzyme activity, decreased membrane fluidity, and decreased transcription rates (D'Amico et al., 2006).

The extinct Sumaco volcano is in the Napo province, 50 km east of the Andean range. In the first botanical expedition in 1979, this volcano's maximum altitude was defined to 3840 meters above sea level (m.a.s.l) (Løjtnant & Molau, 1983). Sumaco volcano is a peculiar geological formation that, although located in the rear-arc with other two volcanos (El Reventador and Pan de azúcar), share neither the geochemical characteristics nor the mineralogy (Garrison et al., 2018a). Due to the lack of accessibility to Sumaco volcano and extreme weather, Sumaco is one of the least hotspots which has not been broadly explored (Bühler et al., n.d.; Colony & Sinclair, 1928). Ecology research conducted in the Sumaco extinct volcano primarily focuses on microorganisms (Hodge & Arbogast, 2016a; Lozano et al., 2020a; Torres et al., 2018a). However, no research has been conducted to elucidate the Sumaco microcosm.

Soil harbors tremendous biodiversity and represents a fraction of the planet's living biomass with 103-104kg/ha (Fierer et al., 2007). Soil bacteria also displays a complex and intricate assemblage of communities (Shen et al., 2015). In addition, microbes play a vital role in nutrient cycling (Hopkins & Dungait, 2010; Jiao et al., 2019). On the other hand, microbes can also be seen as living factories that produce complex

Machinery that can be used to solve world problems. For example, the *Geobacter* genus is a promising bacterium with multiple purposes such as electricity production (Bond & Lovley, 2003a) and electrobiosyntesis (Gregory et al.,

2004a). Furthermore, when discussing high-altitude landscapes, we cannot only simplify the soil niche but also account for water-based niches (Kopáček et al., 2000a; Ordoñez et al., 2009a; Peter & Sommaruga, 2016b). High-altitude lakes are a clear example of this type of niches. In addition, microbes found in lakes can also have an industrial application. This is the case of *Novosphingobium* sp. THN1, which was isolated in Taihu lake, showed to be capable to degradate mycrosystin-LR, a tumor-promoting compound (Jiang et al., 2011a).

Before the invention of high throughput sequencing technology, the microbiome was surveyed through culture-dependent methods (Demergasso et al., 2004). It was a minimal approach since microorganisms that were unable to be cultured, were left apart. Nowadays, next-generation sequencing technologies open new microbial assessment methods, especially since they overcome this issue (Zhou et al., 2010). Although the most potent next-generation technology enables the whole metagenome sequence, amplicon metagenomics is a valuable and vital method for taxonomic profiling of novel environments (Hillmann et al., 2018). Therefore, amplicon metagenomics has been the most accepted method for microbiome survey in high-altitude environments (Kumar et al., 2019a; D. Liu et al., 2016a).

The present work aims to explore the temporal microbiome architecture of bacterial communities at the extinct Sumaco volcano ecosystem and provide insights into the relationship between the observed diversity and the associated geochemical factors. The assessed chronosequence will reveal the microbial dynamics on a time scale. Despite the ecosystemic services that Sumaco volcano is currently offering, we provide the hotspot for microbes bioprospection with biotechnology potential. Furthermore, we also give deeper into ecology theory and propose the first tropic volcano metacommunity.

## 3.2 Diversity Analysis

### 3.2.1 Alpha diversity analysis

The microbial composition of the Sumaco volcano over the years of sampling, presents the relative abundance of the ten most abundant families (Figure 4.2). The Effective Shannon diversity and Simpsons index were calculated for all samples at the genus level. The diversity indexes were compared across sample sites and the two categories of altitude with a one-sided Wilcoxon signed-rank test. Rarefaction curves with steps of 600 samples for soil and water were estimated with the back-end functions of the ranacapa package.

Then richness and evenness were estimated by rarefaction. For each sample, a subsampling was applied using the minimum library size. For each subsample the InvSimpson and richness observed were calculated. This procedure was repeated 100 times to obtain the mean values. Finally, the alpha diversity was compared according to the year and the altitude of the sample.

### 3.2.2 Beta diversity analysis

Before any beta diversity analysis, the abundance table was processed as follows. To eliminate genera with a low detection rate, we employed two criteria: (1) each genus has to be detected at least five times and (2) in more than half of the samples. The filtered was transformed to a Deseq2 object by the phyloseq_to_deseq2() routine (McMurdie & Holmes, 2014). This function requires the experimental design as input, in our case, the altitude (high/low) and time were provided. Then, the abundance table was normalized with the proposed median of ratiosDeseq2 procedure. Here, counts are divided by sample-specific size factors determined by median ratio of OTUs counts relative to the geometric mean per OTU. Finally, since our samples were processed in batches (sampling years), a batch correction was performed. We employ the suggested routine in the limma package for this task, destined to remove unwanted batch effects in genomics data.

Beta diversity was assessed through a non-multidimensional scaling (NMDS) with fitted environmental (physical chemical) variables. The algorithm

for fitting environmental variables to the NMDS space finds the direction in which environmental vectors change most rapidly. The correlation with the sample scores is the largest; the associated statistical significance in this context is for a null hypothesis the correlation equals zero (Anders & Huber, 2010). The NMDS was based on Bray-Curtis distances obtained from the corrected matrix of abundances for families across samples. We used a radar plot to show the distribution of scaled physical-chemical variables for water and soil samples and grouped them by two altitude categories.

### 3.2.3 Differential general analysis

In order to discover differences in the presence of genera between low and high-altitude samples, a differential OTU detection analysis based on a negative binomial distribution was performed with the DESeq2 package [34]. This analysis returned the computed log2 fold change and corresponding p-values. The latter was corrected by the Benjamini-Hochberg method (Benjamini et al., 1995), as a threshold to minimize the false discovery ratio. OTUs classified at the genus level were projected into a volcano plot, with -loge(p) against the log2 fold change. Since the fold change was obtained by low altitude/high altitude abundance ratios, those genera with a positive fold change will express larger abundance at low-altitudes, and those with a negative fold change will express larger abundance at high altitudes.

## 3.3 Results

A total of 49 samples were recovered from the Sumaco volcano with a range from 1584 to 3809 m.a.s.l. A total of 2058223 high-quality reads were obtained for the entire sample set. On average, soil samples account for 1568191 ± reads, while lake samples have 490032 ± 21917. These totals result after the process of quality control, demultiplexing, denoising, the removal of chimeras, and any sequence from other taxa (e.g., chloroplast, eukarya, archaea and mitochondrial sequences). The available sequences were classified into 1330 genera.

### 3.3.1 Alpha diversity

The taxonomic classification of the Sumaco microcosm is included in 44 phyla. In soil and lake samples, the Proteobacteriaphylum is the most abundant with 32% and 72%, respectively. In the case of soil samples, the most abundant phyla are: *Proteobacteria* (32%), *Acidobacteria* (30%), *Chloroflexi* (8%), and *Bacteroidetes* (6%). In contrast, the most abundant phyla in lake samples are: *Proteobacteria* (72%), *Bacteroidetes* (12%), *Verrumicrobia* (3%), and *Patescibacteria* (3%).

The most abundant families for soil and lake samples are presented in the Figure 3.1 and 3.2, respectively.

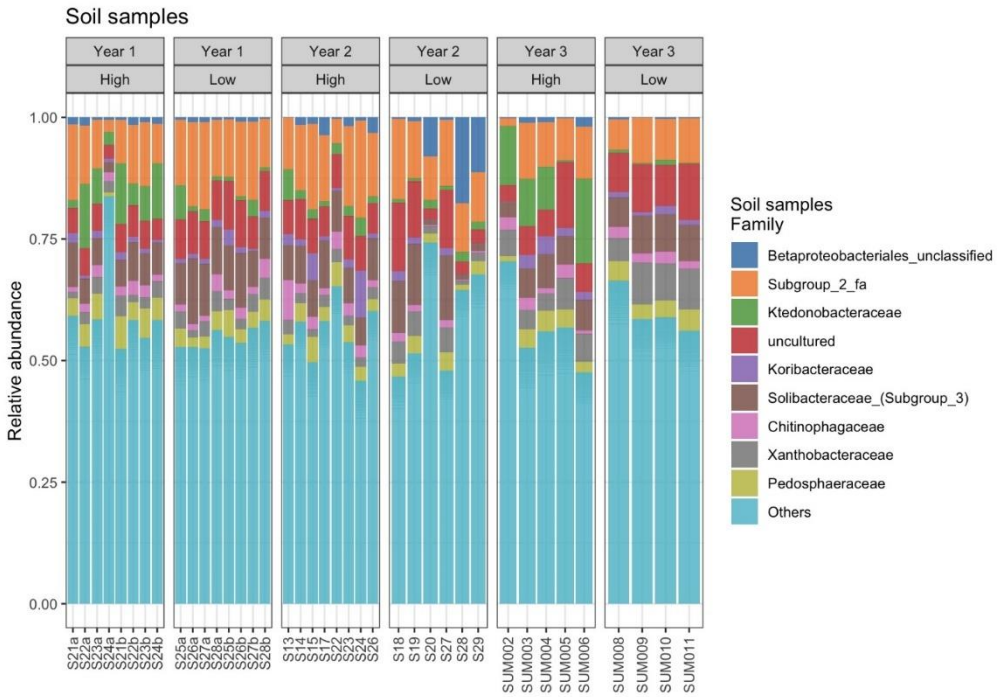



**Figure 3.1. Bacterial Community structure of Sumaco Volcano at the Family level of the through the three years of survey.** The compositional plot is grouped in the relative abundance of the 10 most abundant families for soil. The sampling year 1,2 and the sample subtype categories are displayed as plot titles.

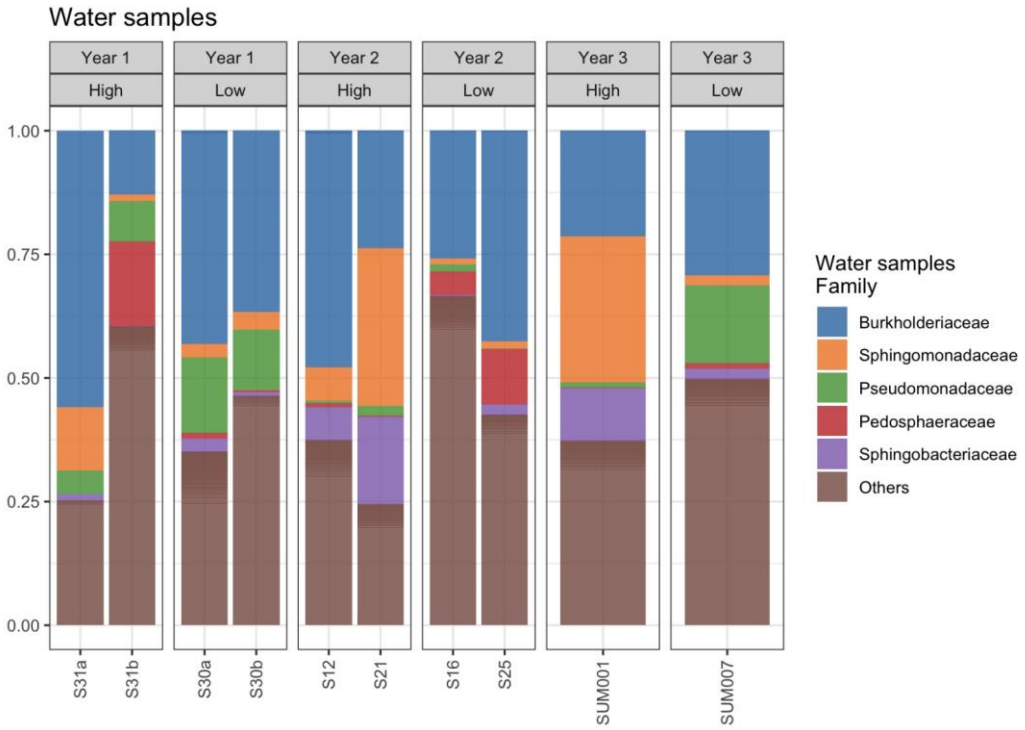



**Figure 3.2. Bacterial Community structure of Sumaco Volcano at the Family level of the through the three years of survey.** The compositional plot is grouped in the relative abundance of the most abundant families for water. The sampling year 1,2,3 and the sample subtype categories are displayed as plot titles.

The distribution of the alpha-diversity values of soil communities, when compared only based on the altitudinal gradient, did not result in a significant difference, based on the Effective Shannon and Simpson index for soil communities (p-value = 0.85) (Figure 3.3).

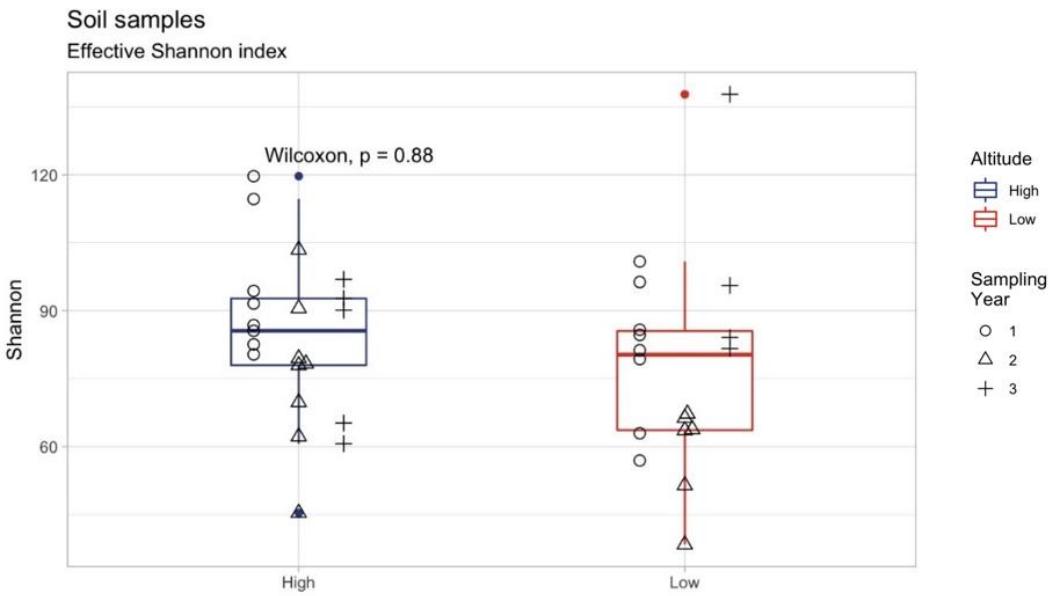



**Figure 3.3. Sumaco Volcano, Alpha-diversity analysis of the through all years of survey.** Effective Shannon index were compared in term of the altitudinal gradient (High Vs Low) for soil-based communities.

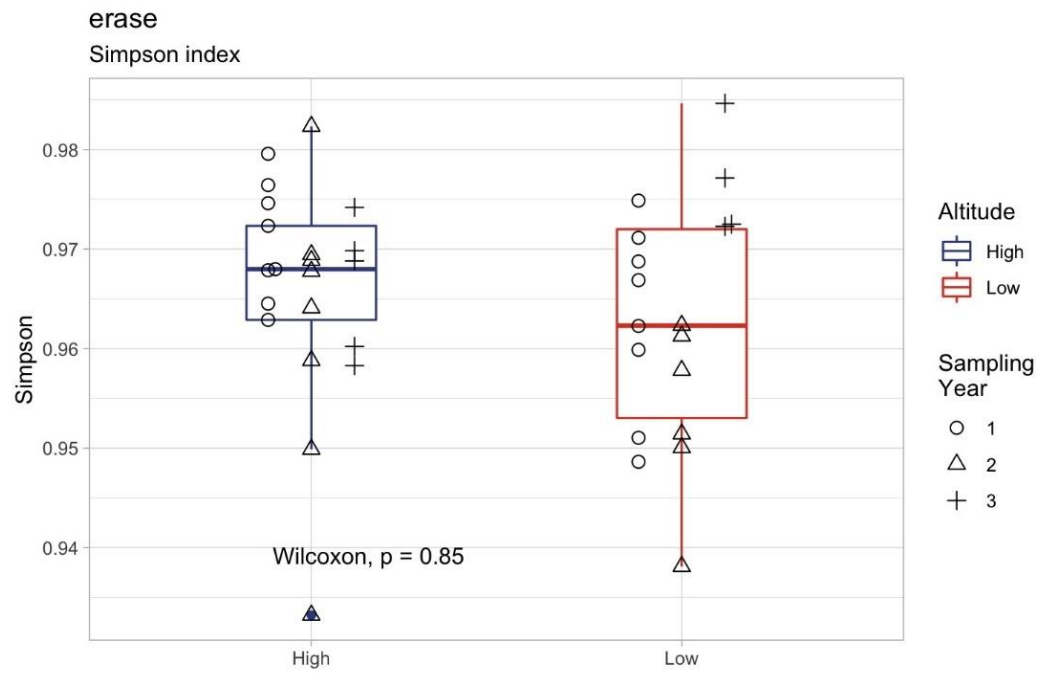



**Figure 3.4. Sumaco Volcano, Alpha-diversity analysis of the through all years of survey.** Simpson indexes were compared in term of the altitudinal gradient (High Vs Low) for soil-based communities

Although a lower p-value was obtained for lake samples (p-value = 0.42), it is still not significant (Figure 3.5).

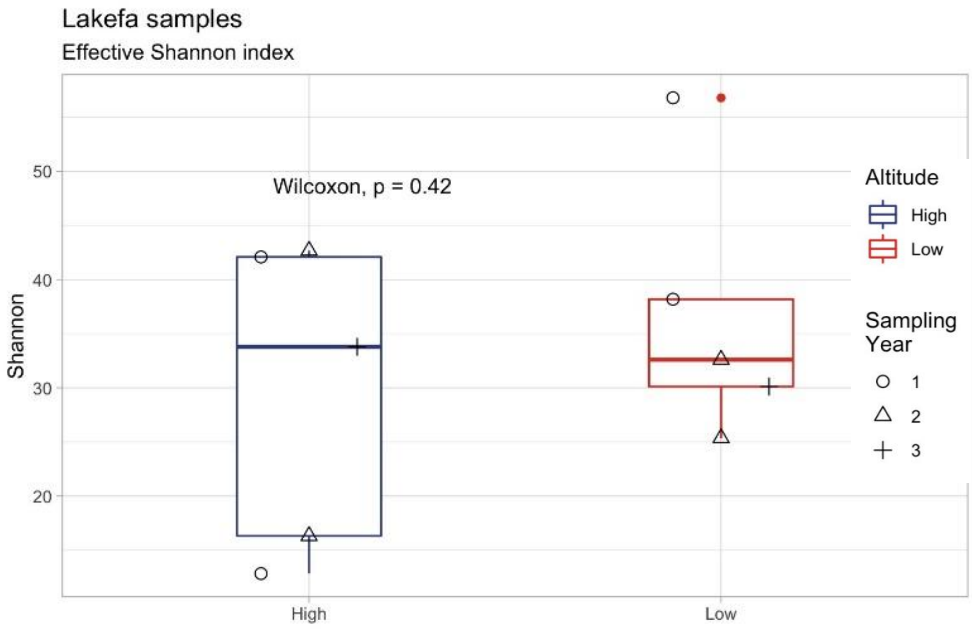



**Figure 3.5. Sumaco Volcano, Alpha-diversity analysis of the through all years of the survey.** Simpson index were compared in term of the altitudinal gradient (High Vs. Low) for lake-based communities

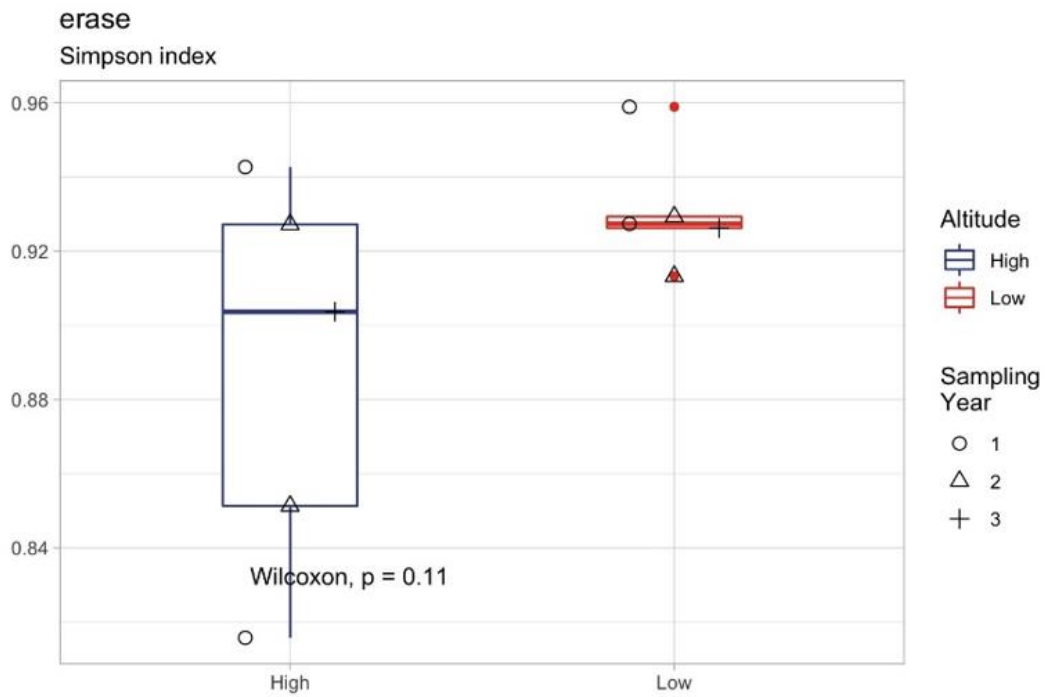



**Figure 3.6. Sumaco Volcano, Alpha-diversity analysis of the through all years of the survey.** Simpson indexes were compared in terms of the altitudinal gradient (High Vs Low) for lake-based communities.

However, we need no consider that we are comparing two different lakes, each at high and low altitudes. On the other hand, when samples are compared according to the altitudinal gradient and the sampling year, exciting patterns are found. For both alpha-diversity indexes, inverse Simpson and richness indexes, a U shape can be seen for the chronosequence, being more evident in the inverse Simpson index (Figure 3.7).

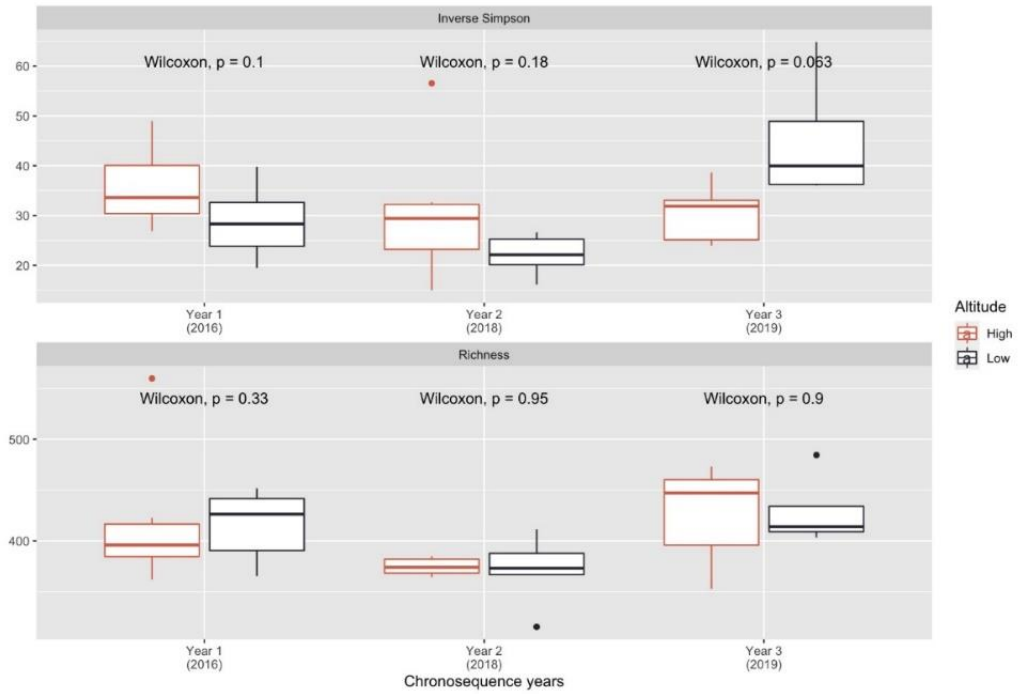



**Figure 3.7. Sumaco Volcano, Alpha-diversity analysis of the through all years of the survey.** Inverse Simpson indexes were compared in terms of the altitudinal gradient (High vs. Low) for soil-based communities

In addition, the comparison of the third year for the inverse Simpson index was the lowest p-value obtained (p-value = 0.063). In contrast, the highest p-value found in these comparisons was for the richness in the second year (p-value = 0.95), indicating almost the same sample richness distribution except for one outlier at a low altitude. Comparing the chronosequence and the altitude for lakes with the statistical test was impossible due to the number of samples recovered (Figure 3.8). Moreover, it seems to have a vague pattern in the compass with any proposed variable.

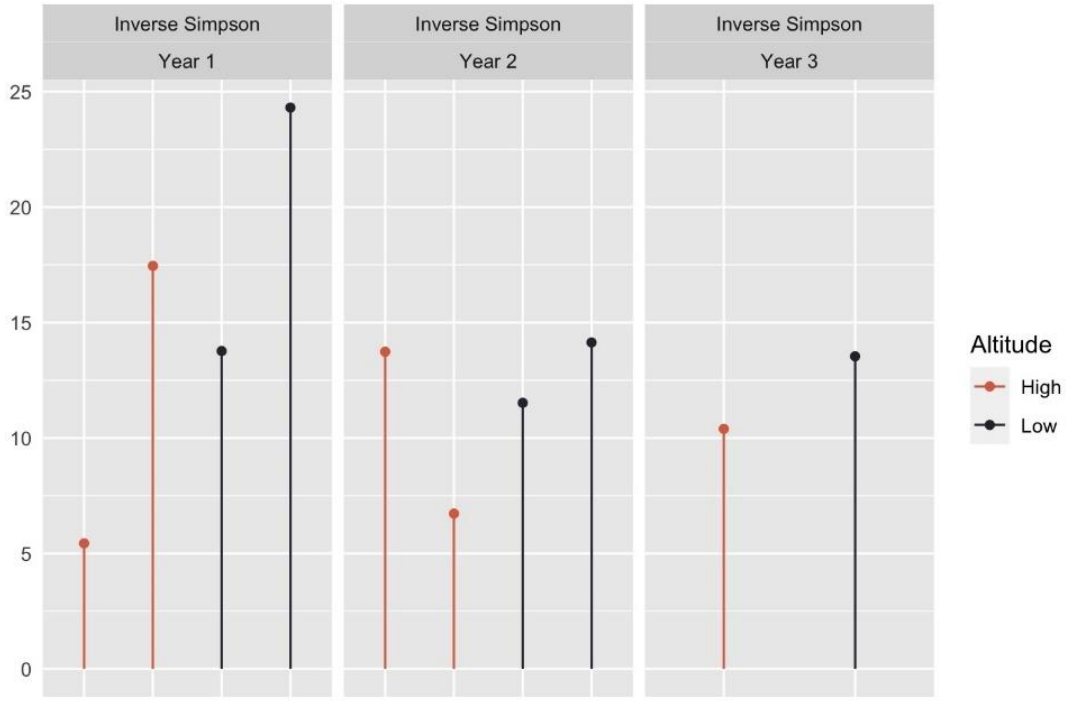



**Figure 3.8. Sumaco Volcano, Alpha-diversity analysis of the through all years of survey.** Inverse Simpson indexes were compared in terms of the altitudinal gradient (High Vs Low) for laked-based communities.

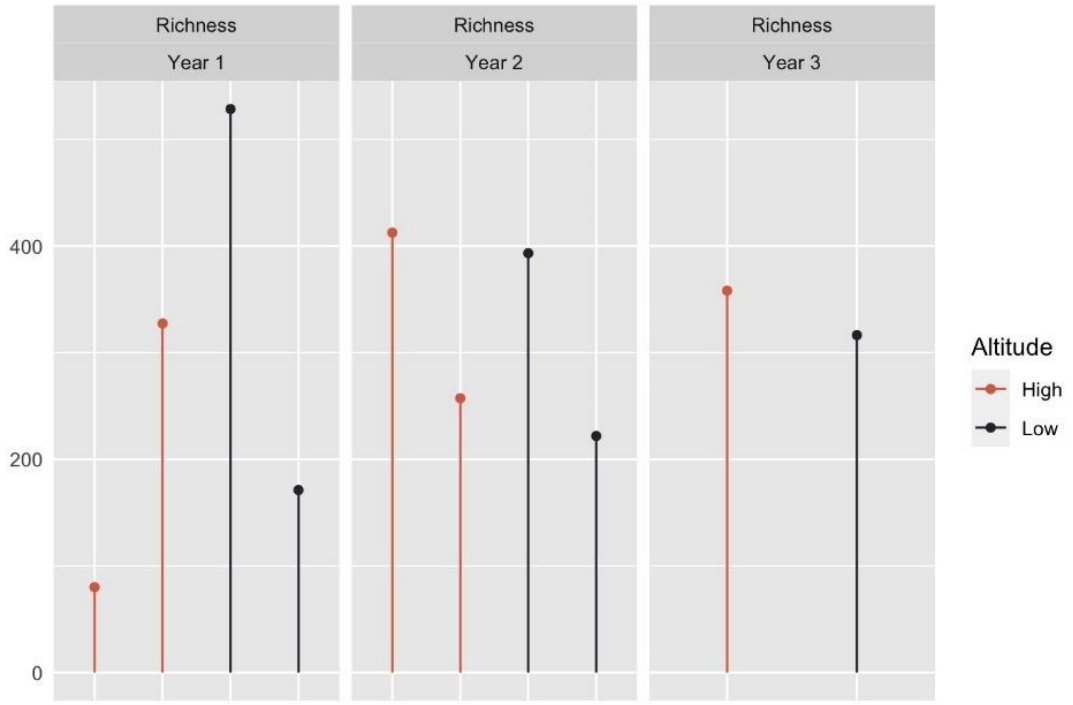



**Figure 3.9. Sumaco Volcano, Alpha-diversity analysis of the through all years of the survey.** Richness indexes were compared in term of the altitudinal gradient (High Vs Low) for laked-based communities

The rarefaction curves for all communities show that they reached the asymptote with at least 85% of reads, which indicates a sufficient sequencing depth. Surprisingly, no trend establishes a relationship between the maximum richness and the altitude (Figure 3.10 and 3.11).

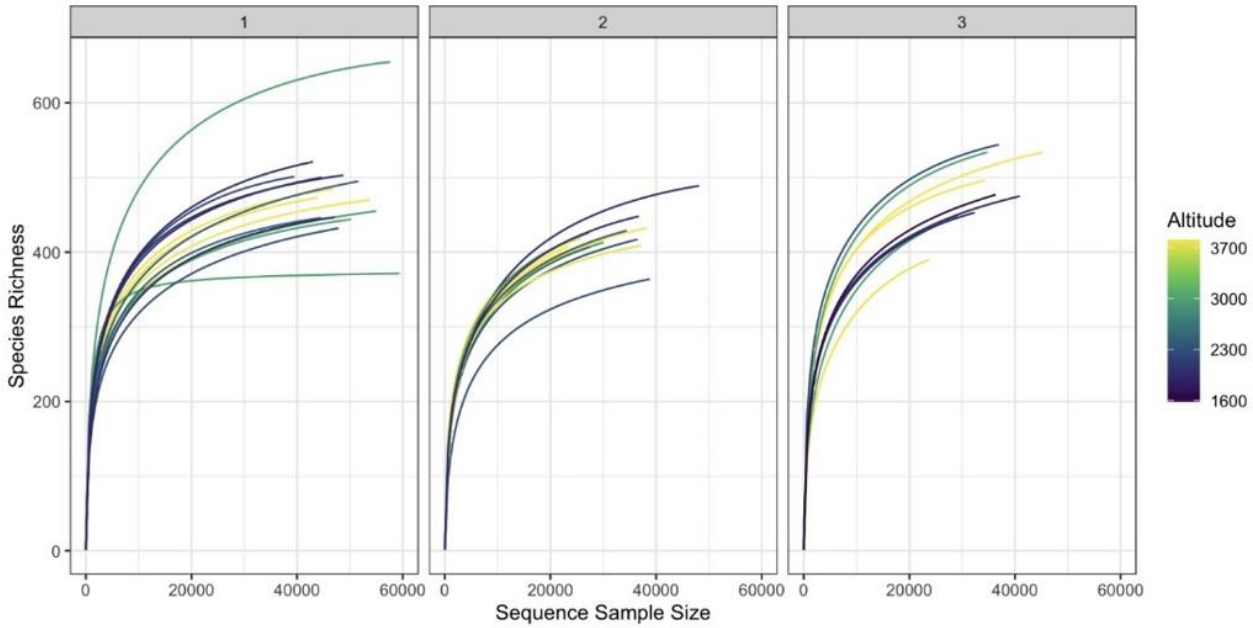



**Figure 3.10. Sumaco Volcano, rarefaction curves of the through all years of sthe urvey.** The curves show the asymptote with al least 60% of reads based only in all soil communities.

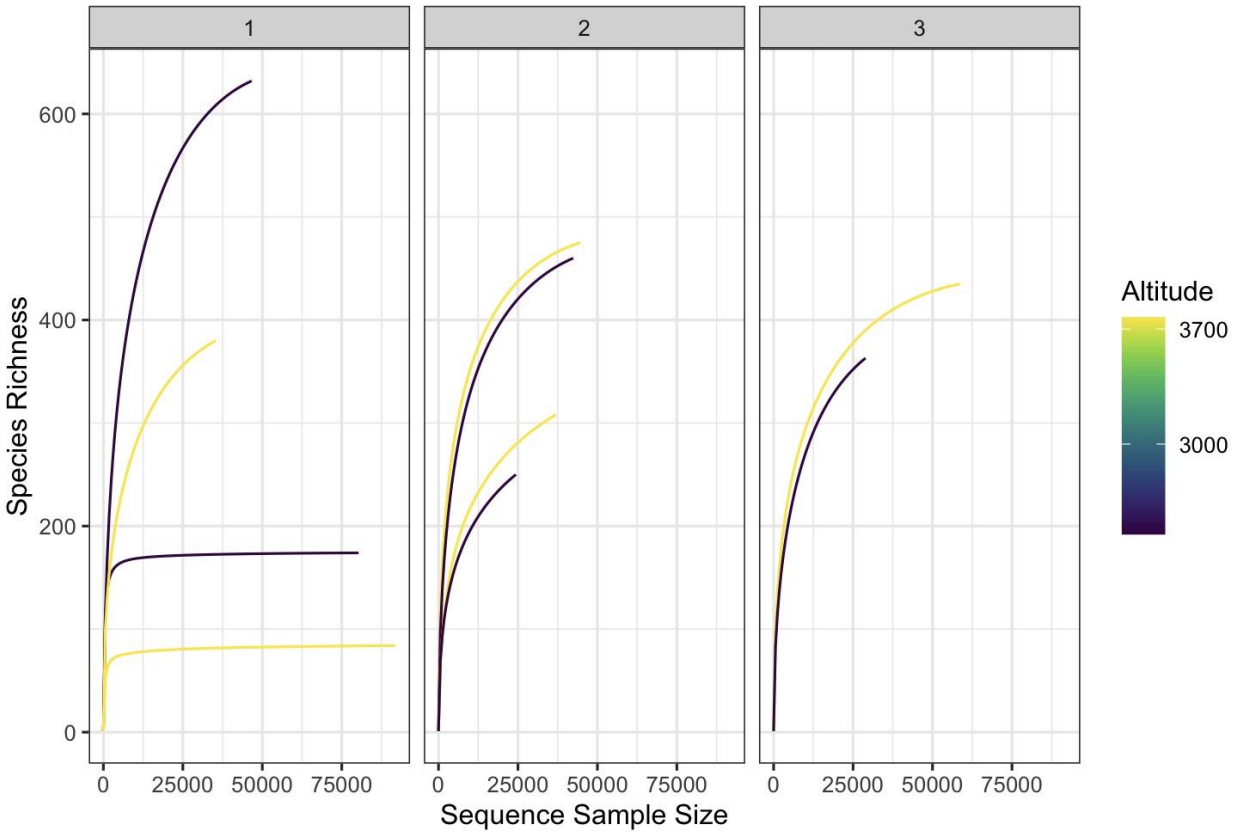



**Figure 3.11. Sumaco Volcano, rarefaction curves of the through all years of the survey.** The curves show the asymptote with at least 40% of reads, based only on all lakes-communities

### 3.3.2 Beta diversity

In the NMDS analysis for soil samples, a clear cloud point between high and low-altitude samples can be observed (Figure 3.12a). The first NMDS dimensions play a central role in differentiating the altitudinal gradient. Meanwhile, the high-altitude samples are located on the negative side of the first dimension, and high-altitude samples tend to be on the positive side of the first dimension. The dispersion of the high-and low-altitude samples can also be noticed; for example, high-altitude sample has a wider dispersion than low-altitude samples. The trend mentioned above is seen for soil samples and can also be applied for the environmental variables. We can find half of the environmental variables in the high-altitude sample cloud, while the other half tend to be located in the low-altitude sample cloud. However, any of these variables were found significantly different between high- and low-altitude samples (Figure 3.12e).

On the other hand, the lake samples constrained analysis showed an elongated dispersion for both lakes. The CCA analysis could capture the 9% and the 7% of the microcosm architecture variance (Figure 3.12b). Meanwhile, the high-altitude lake tends to be placed in the negative side of the first CCA dimension, the low-altitude lake tends to disperse towards the positive side of the first CCA dimension. A cloud point overlapping can be distinguished between high-and low-altitude lakes. In contrast with the constrained analysis of soil samples where any environmental factor was found significantly different for altitudes, the CCA of lakes samples, environmental variables got fewer p-values. For example, pH (p-value = 0.062) and Na (p-value = 0.060). Although they did not meet the criteria of significance (p-value < 0.05), these variables are close to meet these criteria. A peculiar behavior of the two Sumaco lakes' environmental variables was observed, all variables have greater values for the low-altitude lake (Figure 3.12f).

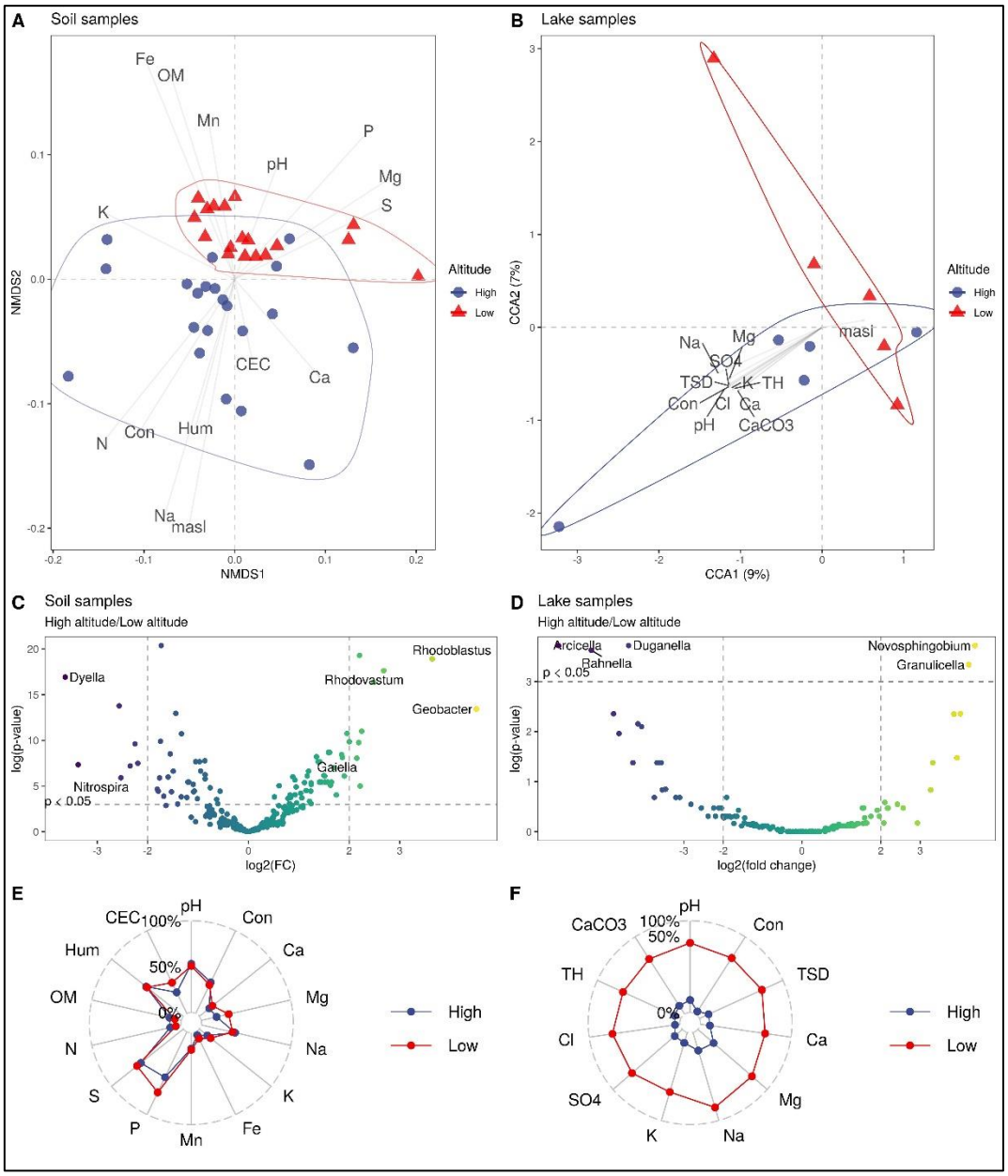



**Figure 3.12.** Beta diversity analysis. Ordination analysis for the microcosm community, an NMDS analysis for soil samples (a) and a CCC analysis for lake (b) samples. Volcano plot for the differential genus analysis for soil (c) and lake (d) samples. Radar plot for the environmental variables for soil (e) and lake (f) samples.

### 3.3.3 Diferential general analysis

0The differential genus analysis can be grouped in two different comparison categories for each, soil and water-based samples separately. The first one is related to the altitudinal gradient (high vs low), and the second is about the temporal variable (sampling years). The comparison for the altitudinal gradient, the fold change is based on the high/low relationship. Therefore, a genus with negative fold change values means the specific genus is found at greater abundance at low-altitudes and genus with positive fold change values means the specific genus is found at greater abundance at high altitudes.

Soil samples subjected to the differential genus analysis resulted in several taxons that meet the criteria to have an adjusted p-value < 0.05 and a -2 < log2(FC) < 2. However, only five taxons were classified properly until the genus level (Figure 4.3.2.c). *Dyella* (p-value = 4.47x10-8), *Nitrospira* (p-value = 2.71x10-3), *Rhodovastum* (p-value = 2.22x10-8), Gaiella (p-value = 3.14x10-4), *Geobacter* (p-value = 1.47x10-6), *Rhodoblastus* (6.08x10-9). In the case of the differential analysis for lake samples, five genera were found significant between the two Sumaco lakes (Figure 4.3.2.d); *Arcicella* (p-value = 0.02), *Duganella* (p-value = 0.02), *Rahnella* (p-value = 0.02), *Novosphingobium* (p-value = 0.02), *Granulicella* (p-value = 0.03).

## 3.4 Discussion

Our study on a tropic volcano encompassed a gradient of elevation and two primary substrates (soil and two lakes). In this study, two types of alpha-diversity behavior were found. The first is related to the driving variable, the elevation gradient and the second trend related to the year's chronosequence year. In addition, the microbiome of the two Sumaco lakes seems to have a random pattern in the chronosequence. A linear relationship between the altitude and diversity has been documented, where higher diversity is found at low-altitudes (Kumar et al., 2019b; X. Liu et al., 2017). In the Sumaco volcano case, this pattern is not well differentiated with a slightly greater alpha-diversity at high altitudes. In a more in-depth look, the only chronosequence year that follows this trend is the third year, which is also the year which almost achieves significant differences. Although the diversity as mentioned earlier behavior pattern is widely

documented in most of the multicellular organisms (Lomolino, 2001), even in the same study area (Hodge & Arbogast, 2016b; Lozano et al., 2020b), this is not always the case of the microbiome (Fierer et al., 2011). In the Wolong Nature Reserve case, in China, the diversity trends fit a hollow pattern with the elevation (D. Liu et al., 2016b).

### 3.4.1 Alpha diversity

Furthermore, when the general trend through the chronosequence is seen, a U shape can be depicted. The second chronosequence year has the lowest values for the Simpson's index which creates a hollow pattern. This change can be explained by the sampling methodology, precisely the date. The first and second sampling approaches were conducted in summer, while the second sampling effort was carried out in winter. This strategy has a significant impact on the resulting lowest abundance in the second chronosequence year. The seasonal variations play a vital role in shaping the microbial community, especially for environmental conditions such as temperature, UV radiation, and precipitation (Dubey et al., 2019; Jansson & Hofmockel, 2018). The seasonal variation has been described in other latitudes where samples clusters according to the sampling date (Delmont et al., 2012). The response of microbes to seasons can be related to the substrate and nutrient availability. For example, in the soil microcosm of beech forest, winter communities had a higher capacity for degradation of complex C substrates than summer communities (Koranda et al., 2013).

### 3.4.2 Beta diversity

In terms of beta-diversity, the difference microcosm architecture between high and low altitude samples is better elucidated. The low-altitude samples are placed in the NMDS space with a lower dispersion than high-altitude samples. We believe that low altitude samples have a steadier microcosm architecture than high-altitude samples. In other words, high-altitude samples are more variable comparing to low-altitude samples. We believe that two main factors could lead to this outcome, the type of vegetation and UV radiation. The vegetation in the Sumaco volcano encompasses an altitudinal gradient, where at

low-altitudes, dense vegetation can be found, while at high altitudes a greater number of shrubs can be found (Lozano et al., 2020b). The dense vegetation provides shelter from the UV radiation, as well it maintains a more steady temperature at the soil strata. This temperate controlling strategy is not available at high-altitudes. Furthermore, one of the most bacterial life-threatening factors is UV radiation. It can damage bacterial macromolecules, such as proteins and DNA, leading to cell death (Miteva et al., 2004). It has also seen a positive relationship between altitude and UV radiation (Blumthaler et al., 1997b), which creates a more selection pressure for microbes at high-altitudes, which can explain the wider sample dispersion for high-altitude samples.

On the other hand, the environmental factor's positioning in the NMDS space shows a well-differentiated two groups. The direction of one group's environmental factor points to the positive side of the first NMDS dimension, while the second group points to the negative side. It is in clear agreement with the high-and low-altitude sample clustering. The majority of the elements, such as Mn, Mg, P, S, and Fe are found at greater concentrations at low-altitudes. This finding is in line with the pedogenesis process, since weathering from high to low-altitudes has been documented (Cole, 1995; Vitousek et al., 2003). On the other hand, not all environmental variables follow the expected trend according to the literature. For example, OM is expected to be at a greater concentration at low-altitudes (Garcia-Pausas et al., 2007a), however, our data shows the opposite. Three main factors seem to be dominant to determine the soil OM content: altitude (Du et al., 2014), the vegetation (Gutiérrez-Girón et al., 2015; Kammer et al., 2009), and topography (Garcia-Pausas et al., 2007b). Since the altitude and vegetation are two well-differentiated variables in our study, we hypothesize that the greater OM at high-altitude samples is uncovered in the Sumaco topography. Since the Sumaco lava analysis showed to be unique for Sumaco and does not share common characteristics with the other Ecuadorian rear-arc volcanos (Garrison et al., 2018b).

In the case of the two Sumaco lakes' beta diversity, one lake at low altitude and one lake at high altitude, they have a clear pattern with a minimum overlapping area. We suggest that the difference between these two lakes is primarily driven by vegetation. For example, it has been proven that lake

chemical characteristics are related to its catchment vegetation (Kopáček et al., 2000b). For the two Sumaco lakes, the low-altitude lake, its catchment vegetation is principally vascular plants and in the high-altitude lake is principally shrubs (Lozano et al., 2020b). On the other hand, in contrast with the soil samples' environmental vector directions, environmental factors seem to be randomly located. The only factor that is not directed to the third quadrant in the altitude is expected since the elevation gradient is one of our governing variables. However, a clear and homogeneous trend is appreciated for the environmental factors, all of them are greater or are in a greater concentration in the lake at low altitude. The nutrient transport process plays a major role in inducing a higher nutrient concentration, since runoff can transport nutrients from high altitude to lower altitudes (Guzman et al., 2019; Leonard et al., 1979).

### 3.4.3 Differential genera analysis

The differentia genera analysis was conducted to determine if the abundance of a particular genus is a difference based on the elevational gradient across the chronosequence. In order to present our results in with a more standard approach, we employed the -log(p-value) and log2(fold change) metrics to create a volcano plot. Since the fold change is based on the high-altitude/low-altitude ratio, positive fold changes indicate higher abundance at high altitudes. Using a threshold of p-value < 0.05, a total of six genera were found significantly different in soil samples, while five genera were found significantly different in lake samples.

In soil samples, *Dyella* and *Nitrospira* have negative fold change values while *Gaiella, Geobacter, Rhodovastum,* and *Rhodoblastushave* positive fold change values. The genus *Dyellawas* first reported in 2005 in isolates taken from a garden in Tokyo (Xie & Yokota, 2005). The first report included this genus can grow in a pH range from 6.8 to 10, and in temperatures from 10 ºC to 37 ºC. A later work proved this genus could live even in more acidic soils, with a minimum pH of 4 (Weon et al., 2009). This minimum pH value is in line with the reported values at both altitudes in Sumaco soil, ranging from 3.9 to 6.9. In addition, several reports of this genus at different soil types, such as sunflower rhizosphere (Anandham et al., 2011), greenhouse soil (Kim et al., 2006), and cliff soil (Lee & Lee,

2009), suggest a cosmopolitan soil presence. *Nitrospira* is a well-known genus related to the nitrogen cycle, specifically for its nitrate-oxidizing metabolism (Attard et al., 2010; Le Roux et al., 2016). The general mechanism of is based on the use of urea as substrate. *Nitrospira* genome encodes for urease enzymes that cleave urea to ammonia and $CO_2$ (Koch et al., 2015). Research conducted in the Southeastern Amazon region showed that soils related to agriculture show a higher abundance of *Nitrospira* genus and a greater Shannon diversity index than native forest (Mendes et al., 2015). Therefore, if the Sumaco land-use change to agriculture, an increment of the abundance of this genus can be expected. The presence of crop systems and local indigenous settlements has been documented in the buffer area of the Sumaco volcano, which can contribute to the diversification of land use (Jadán et al., 2012; Torres et al., 2018b). However, the lack of accessibility and the current state of the Sumaco volcano as a national park could make the shift more challenging.(Díaz et al., 2022)

The *Gaiella* genus was reported in tropical (Mun & Wong Vui Ling, 2020; C. Yang & Vui Ling, 2017) and uncultivated (Wolinska et al., 2019) soils. *Rhodovastum* is an acidophilic bacterium capable of producing purple pigments that have been isolated from paddy (Okamura et al., 2009a; Salama et al., 2020), glacier (Shivaji et al., 2011), and bacterial communities related to Poa annua turf (Beirn et al., 2017). However, the ecology of these two genera in the soil macrocosm has not been elucidated yet. *Geobacter* was also found significant at different altitudes. This genus is principally attractive to be able to reduce Fe(III) (Caccavo et al., 1994). This functionality has been applied with different purposes (M. M. Shi et al., 2019), for example, hydrocarbon-contaminated soil remediation (H. Wang et al., 2019), electricity production (Bond & Lovley, 2003b), electrobiosynthesis (Gregory et al., 2004b). Our result is the baseline for future studies to discover *Geobacter* species with a wide spectrum of possible uses, where low altitude soils will be the first exploration site. *Rodoblastus* was is previously known as *Rhodopseudomonas* (Imhoff, 2001). Members of this genus were reported to be able to produce purple and orange pigments and be able to grow in acid soils at pH of 5 (Okamura et al., 2009b). Since the closer pH value to 5 across the chronosequence is found at low altitude samples, it is completely expected a greater abundance at this altitude.

In the case of Sumaco lakes, *Arcicella, Rahnellaand Duganellagenera* have negative fold change values, while *Novosphingobiumand Granulicella* has positive fold change values. Members of the Arcicella genus has been found in stream water *Arcicella* (Sheu et al., 2010), freshwater biofilms (Nikitin et al., 2004), glacier lakes (Y. Liu et al., 2011), and in turbid lakes after glacier retreats (Peter & Sommaruga, 2016c). The *Rahnella* genus has been described to be present in lakes in different latitudes such as African Rift Valley (Duckworth et al., 1996), United States (Derlet et al., 2004), and Andean region (Ordoñez et al., 2009b). The *Duganella* genus has also been described in the Andean region [18], in the control group in Dongting lake in China (Y. Yang et al., 2012), in the Gurudongmar lake in India (Yadav et al., 2017), and in the Fryxell lake in Antartica (Brambilla et al., 2001). The *Granulicella* genus was found in an Andean high-altitude lake (Aszalós et al., 2016), cosely related to *Granulicella artica* (Männistö et al., 2012). Although the genera mentioned above has been widely described in high-altitude lakes, their ecology role has neither elucidated yet. On the other hand, the *Novosphingovium* genus is reported to be capable to degradate hazardous compounds (A. Stolz, 2009). This genus can be found in different environments such as soil, sea, wood, sediments and lakes with its degradative metabolism is found on plasmids (J. Wang et al., 2018). The most promising *Novosphigobium* lake strain was isolated from thelake Taihu and is capable of degradate 91.2% of mycrocystin-LC (a tumor promoter compound) in the first 12 h (Jiang et al., 2011b).

### 3.4.4 Bioinformatic sources

Different factors such as the method of choice for sequencing, the type of Illumina sequencer, the sample preparation technique, the clear use of primers and indexers, the PCR itself which apparently does not involve an observable error effect, constitute errors of next generation sequencing (NGS) currently employed to obtain a large number of sequences.(Pfeiffer et al., 2018). However, such sequencing errors hinder the recognition of biologically real existing nucleotides against all 16S rRNA sequencing artifacts. To avoid introducing variables coming from these sequencing techniques, what is usually done is to group sequences into Operational Taxonomic Units (OTU's) whose identity

threshold is usually 97%, although risking taxonomic resolution. However, this is now beginning to be resolved thanks to the proliferation of different bioinformatics programs whose main workflow is rather to "de-noise" the sequences, and thus aim to avoid these errors and improve taxonomic identity. Therefore, many algorithms exist and differ from each other, by the way they correct these errors introduced in the sequences (Nearing et al., 2018). In this work we used Mothur v.1.42.0 open-source software used by the Schloss lab to process their 16S rRNA gene sequences that are generated using Illumina's MiSeq platform using paired end reads.

The approach we take is to use index reads to multiplex a large number of samples in a single run, clustering the sequences into OTUs" and, eliminating reads derived from predicted errors of neighboring sequences, further employs a sample-by-sample approach that reduces both memory requirements and computational demand(Schloss et al., 2009b). Instead DADA2 an open source software described to model and correct errors in Illumina sequenced amplicons is trained throughout the sequencing process by generating a parametric error model and then applies that model to correct and collapse sequence errors into so-called amplicon sequence variants (ASVs) This approach is advantageous as it builds unique error models for each sequencing run (Callahan et al., 2016). In the joint data analysis studies that DADA2 performed it consistently attracted more ASVs than two other denoising pipelines employed in those studies and that again the open reference OTU clustering called the most overall. But there is something important to note and that is that they reported in the soil dataset studied both DADA2, and the OTU clustering pipeline were able to cluster more low abundance organisms (Pfeiffer et al., 2018). Finally, Mothur was employed in this work because at the time 2018-2019 it was the best performing algorithm for 16S rRNA amplicon sequence analysis and its implementation for the analysis of these samples from water and volcanic soils was under development. If we submit to a comparison between the data obtained from the analysis with Mothur Vs DADA2, considering only the clustering factor in OTU for the one case and ASV for the other respectively, we can observe that the results are approximate.

## 3.5 Conclusions

Within a range of The Sumaco volcano microcosm did not show a relationship between the elevational gradient and the alpha-diversity, Sine one sample recovery approach was conducted in winter. It allowed us to discover that in this season the alpha-diversity reaches the lowest diversity indexes. On the other hand, when the microcosm architecture is introduced in the analysis, a clear separation between high-and low-altitude samples can be achieved. In other words, the beta-diversity analysis showed a different microbial composition between the two altitudes. Additionally, low-altitude samples exhibit a more homogeneous microbial community than high-altitude samples. We also provide the first proposal of the metacommunity of a tropical volcano considering two different niches (soil and lake). This metacommunity is primarly dominated by *Proteobacteria, Verrumicrobia, Acidobacteria,* and *Actinobacteria.*

However, due to the sample size relationship between soil and lakes samples, we have to recognize the suggested metacommunity is highly influenced by the most abundant phyla in the soil niche. Furthermore, the proposed metacommunity is in line with previously reported phyla in high-altitude ecosystems.

Although any of the evaluated environmental variables resulted in significant differences between altitudes, some interesting patterns may influence the Sumaco microcosm. For example, soil OM content did not follow the expected behavior to be in greater concentrations at low-altitudes. We suggest the Sumaco topography plays a vital role in creating this behavior. Nowadays, there is a lack of geological research in the Sumaco volcano, which obscures the Sumaco geological contribution to its microbiome. Also, we found all environmental factors to be in greater concentration in the lake at low-altitude, and it can be explained by the nutrient and element from high-altitude to low-altitudes by multiple factors, but mainly by runoff.

We found several genera that are always different across the chronosequence at both altitudes. Although almost half of these genus' ecology functions are not elucidated yet, some are of great interest. For example, *Nitrospira* can use urea as energy to transform urea to ammonia and CO2. The

*Geobacter* and *Novosphigobium* are the most attractive genera found due to its possible application in electricity production and hazardous compounds degradation, respectively.

Although our work is the first in the equatorial tropic glaciers, various questions remain unanswered. The main question is related to the bacterial community architecture for the upcoming chronosequence and can only be revealed by future microbial surveys. Since we employed the amplicon metagenomic methodology, we were able only to perform taxonomic profiling. The next step is to employ more advanced techniques to discover functional diversity. In consequence with the finding of the high *Geobacter* and *Novosphigobium* abundance, bacterial culture and isolation should be carried out to determine the potential expression of bioproducts of interest.

## 3.6 References

Anandham, R., Kwon, S. W., Indira Gandhi, P., Kim, S. J., Weon, H. Y., Kim, Y. S., Sa, T. M., Kim, Y. K., & Jee, H. J. (2011). Dyella thiooxydans sp. nov., a facultatively chemolithotrophic, thiosulfate-oxidizing bacterium isolated from rhizosphere soil of sunflower (Helianthus annuus L.). *International Journal of Systematic and Evolutionary Microbiology*, *61*(2), 392–398. https://doi.org/10.1099/ijs.0.022012-0

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

Aszalós, J. M., Krett, G., Anda, D., Márialigeti, K., Nagy, B., & Borsodi, A. K. (2016). Diversity of extremophilic bacteria in the sediment of high-altitude lakes located in the mountain desert of Ojos del Salado volcano, Dry-Andes. *Extremophiles*, *20*(5), 603–620. https://doi.org/10.1007/s00792-016-0849-3

Attard, E., Poly, F., Commeaux, C., Laurent, F., Terada, A., Smets, B. F., Recous, S., & Roux, X. Le. (2010). Shifts between Nitrospira- and Nitrobacter-like nitrite oxidizers underlie the response of soil potential nitrite oxidation to changes in tillage practices. *Environmental Microbiology*, *12*(2), 315–326. https://doi.org/10.1111/j.1462-2920.2009.02070.x

Baird, R., Eaton, A., & Rice, E. (2017). *Standard methods for the examination of water and wastewater* (23rd Ed). American Public Health Association.

Bax, V., & Francesconi, W. (2019). Conservation gaps and priorities in the Tropical Andes biodiversity hotspot: Implications for the expansion of protected areas. *Journal of Environmental Management*, *232*(June 2018), 387–396. https://doi.org/10.1016/j.jenvman.2018.11.086

Beirn, L. A., Hempfling, J. W., Schmid, C. J., Murphy, J. A., Clarke, B. B., & Crouch, J. A. (2017). Differences among soil-inhabiting microbial communities in Poa annua turf throughout the growing season. *Crop Science*, *57*(august), S-262-S-273. https://doi.org/10.2135/cropsci2016.06.0463

Benjamini, Y., Hochberg, Y., & Benjamini, Yoav, H. Y. (1995). Controlling the False Discovery Rate: A practical and Powerfull Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.2307/2346101

Blumthaler, M., Ambach, W., & Ellinger, R. (1997). Increase in solar UV radiation with altitude. *Journal of Photochemistry and Photobiology B: Biology*, *39*(2), 130–134. https://doi.org/10.1016/S1011-1344(96)00018-8

Bond, D., & Lovley, D. (2003). Electricity Production by Geobacter sulfurreducens Attached to

Electrodes. *Applied and Environmental Microbiology*, *69*(3), 1548–1555. https://doi.org/10.1128/AEM.69.3.1548

Borda-Molina, D., Montaña, J. S., Zambrano, M. M., & Baena, S. (2017). Mining lipolytic enzymes in community DNA from high Andean soils using a targeted approach. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *110*(8), 1035–1051. https://doi.org/10.1007/s10482-017-0877-8

Brambilla, E., Hippe, H., Hagelstein, A., Tindall, B. J., Stackebrandt, E., Fryxell, L., & Valleys, M. D. (2001). 16S rDNA diversity of cultured and uncultured prokaryotes of a mat sample. *International Journal*, 23–33.

Bühler, R., Robert, D., & J, V. A. (n.d.). *Climbing Volcano Sumaco : The Beautiful Beast*. *January 2003*. https://doi.org/10.13140/RG.2.2.24297.03685

Caccavo, F., Lonergan, D. J., Lovley, D. R., Davis, M., Stolz, J. F., & Mcinerneyl, M. J. (1994). Oxidizing dissimilatory metal-reducing microorganism. *Applied and Environmental Microbiology*, *60*(10), 3752–3759.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Ciccazzo, S., Esposito, A., Borruso, L., & Brusetti, L. (2016). Microbial communities and primary succession in high altitude mountain environments. *Annals of Microbiology*, *66*(1), 43–60. https://doi.org/10.1007/s13213-015-1130-1

Cole, D. W. (1995). Soil nutrient supply in natural and managed forests. *Plant and Soil*, *168–169*(1), 43–53. https://doi.org/10.1007/BF00029312

Colony, R. J., & Sinclair, J. H. (1928). The lavas of the Volcano Sumaco, eastern Ecuador, South America. In *American Journal of Science* (Vols. s5-16, Issue 94, pp. 299–312). https://doi.org/10.2475/ajs.s5-16.94.299

D'Amico, S., Collins, T., Marx, J. C., Feller, G., & Gerday, C. (2006). Psychrophilic microorganisms: Challenges for life. *EMBO Reports*, *7*(4), 385–389. https://doi.org/10.1038/sj.embor.7400662

De Carvalho, T. S., Da Conceição Jesus, E., Barlow, J., Gardner, T. A., Soares, I. C., Tiedje, J. M., & De Souza Moreira, F. M. (2016). Land use intensifcation in the humid tropics increased both alpha and beta diversity of soil bacteria. *Ecology*, *97*(10), 2760–2771. https://doi.org/10.1002/ecy.1513

Delmont, T. O., Prestat, E., Keegan, K. P., Faubladier, M., Robe, P., Clark, I. M., Pelletier, E.,

Hirsch, P. R., Meyer, F., Gilbert, J. A., Le Paslier, D., Simonet, P., & Vogel, T. M. (2012). Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME Journal*, *6*(9), 1677–1687. https://doi.org/10.1038/ismej.2011.197

Demergasso, C., Casamayor, E. O., Chong, G., Galleguillos, P., Escudero, L., & Pedrós-Alió, C. (2004). Distribution of prokaryotic genetic diversity in athalassohaline lakes of the Atacama Desert, Northern Chile. *FEMS Microbiology Ecology*, *48*(1), 57–69. https://doi.org/10.1016/j.femsec.2003.12.013

Derlet, R. W., Carlson, J. R., & Noponen, M. N. (2004). Coliform and pathologic bacteria in Sierra Nevada national forest wilderness area lakes and streams. *Wilderness and Environmental Medicine*, *15*(4), 245–249. https://doi.org/10.1580/1080-6032(2004)015[0245:CAPBIS]2.0.CO;2

Díaz, M., Jarrín-V, P., Simarro, R., Castillejo, P., Tenea, G. N., & Molina, C. A. (2021). The Ecuadorian Microbiome Project: a plea to strengthen microbial genomic research. *Neotropical Biodiversity*, *7*(1), 223–237. https://doi.org/10.1080/23766808.2021.1938900

Díaz, M., Quiroz-Moreno, C., Jarrín-V, P., Piquer-Esteban, S., Monfort-Lanzas, P., Rivadeneira, E., Castillejo, P., Arnau, V., Díaz, W., Sangari, F. J., & Molina, C. A. (2022). Soil Bacterial Community Along an Altitudinal Gradient in the Sumaco, a Stratovolcano in the Amazon Region . In *Frontiers in Forests and Global Change* (Vol. 5). https://www.frontiersin.org/article/10.3389/ffgc.2022.738568

Du, B., Kang, H., Pumpanen, J., Zhu, P., Yin, S., Zou, Q., Wang, Z., Kong, F., & Liu, C. (2014). Soil organic carbon stock and chemical composition along an altitude gradient in the Lushan Mountain, subtropical China. *Ecological Research*, *29*(3), 433–439. https://doi.org/10.1007/s11284-014-1135-4

Dubey, A., Malla, M. A., Khan, F., Chowdhary, K., Yadav, S., Kumar, A., Sharma, S., Khare, P. K., & Khan, M. L. (2019). Soil microbiome: a key player for conservation of soil health under changing climate. *Biodiversity and Conservation*, *28*(8–9), 2405–2429. https://doi.org/10.1007/s10531-019-01760-5

Duckworth, A. W., Grant, W. D., Jones, B. E., & Van Steenbergen, R. (1996). Phylogenetic diversity of soda lake alkaliphiles. *FEMS Microbiology Ecology*, *19*(3), 181–191. https://doi.org/10.1016/0168-6496(96)00003-7

Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R. A., Felts, B., Rayhawk, S., Knight, R., Rohwer, F., & Jackson, R. B. (2007). Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied and Environmental Microbiology*, *73*(21), 7059–7066.

https://doi.org/10.1128/AEM.00358-07

Fierer, N., Mccain, C. M., Meir, P., Zimmermann, M., Joshua, M., Silman, M. R., Knight, R., Ecology, S., April, N., Fierer, N., Mccain, C. M., Meir, P., Zimmermann, M., Rapp, J. M., Silman, M. R., & Knight, R. (2011). *Microbes do not follow the elevational diversity patterns of plants and animals*. *92*(4), 797–804.

Fonseca, J. P., Hoffmann, L., Cabral, B. C. A., Dias, V. H. G., Miranda, M. R., de Azevedo Martins, A. C., Boschiero, C., Bastos, W. R., & Silva, R. (2018). Contrasting the microbiomes from forest rhizosphere and deeper bulk soil from an Amazon rainforest reserve. *Gene*, *642*(December 2016), 389–397. https://doi.org/10.1016/j.gene.2017.11.039

Gallegos Castro, E., Brito Chasiluisa, C., Serrano Giné, D., & Galárraga Sánchez, R. (2018). Análisis de la variación temporal y espacial de la cobertura glaciar del Nevado Cayambe, Ecuador, mediante fotografías aéreas e imágenes Landsat. *GeoFocus Revista Internacional de Ciencia y Tecnología de La Información Geográfica*, *22*, 97–113. https://doi.org/10.21138/gf.577

Garcia-Pausas, J., Casals, P., Camarero, L., Huguet, C., Sebastià, M. T., Thompson, R., & Romanyà, J. (2007). Soil organic carbon storage in mountain grasslands of the Pyrenees: Effects of climate and topography. *Biogeochemistry*, *82*(3), 279–289. https://doi.org/10.1007/s10533-007-9071-9

Garrison, J. M., Sims, K. W. W., Yogodzinski, G. M., Escobar, R. D., Scott, S., Mothes, P., Hall, M. L., & Ramon, P. (2018). Shallow-level differentiation of phonolitic lavas from Sumaco Volcano, Ecuador. *Contributions to Mineralogy and Petrology*, *173*(1), 0. https://doi.org/10.1007/s00410-017-1431-4

Gregory, K. B., Bond, D. R., & Lovley, D. R. (2004). Graphite electrodes as electron donors for anaerobic respiration. *Environmental Microbiology*, *6*(6), 596–604. https://doi.org/10.1111/j.1462-2920.2004.00593.x

Gutiérrez-Girón, A., Díaz-Pinés, E., Rubio, A., & Gavilán, R. G. (2015). Both altitude and vegetation affect temperature sensitivity of soil organic matter decomposition in Mediterranean high mountain soils. *Geoderma*, *237–238*, 1–8. https://doi.org/10.1016/j.geoderma.2014.08.005

Guzman, C. D., Hoyos-Villada, F., Da Silva, M., Zimale, F. A., Chirinda, N., Botero, C., Morales Vargas, A., Rivera, B., Moreno, P., & Steenhuis, T. S. (2019). Variability of soil surface characteristics in a mountainous watershed in Valle del Cauca, Colombia: Implications for runoff, erosion, and conservation. *Journal of Hydrology*, *576*(June), 273–286. https://doi.org/10.1016/j.jhydrol.2019.06.002

Hillmann, B., Al-ghalith, G. A., Shields-cutler, R. R., Zhu, Q., Gohl, D. M., Beckman, K. B., Knight, R., & Knights, D. (2018). Evaluating the Information Content of Shallow Shotgun Metagenomics. *MSystems*, *3*(6), 1–12.

Hodge, A. M. C., & Arbogast, B. S. (2016). Carnivore diversity at a montane rainforest site in Ecuador's Gran Sumaco Biosphere Reserve. *Oryx*, *50*(3), 474–479. https://doi.org/10.1017/S0030605315000101

Hopkins, D., & Dungait, J. (2010). Soil Microbiology and Nutrient Cycling. In *Soil Microbiology and Sustainable Crop Production* (pp. 1–436). https://doi.org/10.1007/978-90-481-9479-7

Imhoff, J. F. (2001). Transfer of Rhodopseudomonas acidophila to the new genus Rhodoblastus as Rhodoblastus acidophilus gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, *51*(5), 1863–1866. https://doi.org/10.1099/00207713-51-5-1863

Jadán, O., Torres, B., & Günter, S. (2012). Influencia del uso de la tierra sobre almacenamiento de carbono en sistemas productivos y bosque primario en Napo, Reserva de Biosfera Sumaco, Ecuador. *Revista Amazónica Ciencia y Tecnología*, *1*(3), 173–184.

Jansson, J. K., & Hofmockel, K. S. (2018). The soil microbiome — from metagenomics to metaphenomics. *Current Opinion in Microbiology*, *43*, 162–168. https://doi.org/10.1016/j.mib.2018.01.013

Jiang, Y., Shao, J., Wu, X., Xu, Y., & Li, R. (2011). Active and silent members in the mlr gene cluster of a microcystin-degrading bacterium isolated from Lake Taihu, China. *FEMS Microbiology Letters*, *322*(2), 108–114. https://doi.org/10.1111/j.1574-6968.2011.02337.x

Jiao, S., Xu, Y., Zhang, J., Hao, X., & Lu, Y. (2019). Core Microbiota in Agricultural Soils and Their Potential Associations with Nutrient Cycling. *MSystems*, *4*(2), 1–16. https://doi.org/10.1128/msystems.00313-18

Kammer, A., Hagedorn, F., Shevchenko, I., Leifeld, J., Guggenberger, G., Goryacheva, T., Rigling, A., & Moiseev, P. (2009). Treeline shifts in the Ural mountains affect soil organic matter dynamics. *Global Change Biology*, *15*(6), 1570–1583. https://doi.org/10.1111/j.1365-2486.2009.01856.x

Kim, B. Y., Weon, H. Y., Lee, K. H., Seok, S. J., Kwon, S. W., Go, S. J., & Stackebrandt, E. (2006). Dyella yeojuensis sp. nov., isolated from greenhouse soil in Korea. *International Journal of Systematic and Evolutionary Microbiology*, *56*(9), 2079–2082. https://doi.org/10.1099/ijs.0.64175-0

Koch, H., Lücker, S., Albertsen, M., Kitzinger, K., Herbold, C., Spieck, E., Nielsen, P. H., Wagner, M., & Daims, H. (2015). Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus Nitrospira. *Proceedings of the National Academy of Sciences of the United*

*States of America*, *112*(36), 11371–11376. https://doi.org/10.1073/pnas.1506533112

Kopáček, J., Stuchlík, E., Straškrabová, V., & Pšenáková, P. (2000). Factors governing nutrient status of mountain lakes in the Tatra Mountains. *Freshwater Biology*, *43*(3), 369–383. https://doi.org/10.1046/j.1365-2427.2000.00569.x

Koranda, M., Kaiser, C., Fuchslueger, L., Kitzler, B., Sessitsch, A., Zechmeister-Boltenstern, S., & Richter, A. (2013). Seasonal variation in functional properties of microbial communities in beech forest soil. *Soil Biology and Biochemistry*, *60*, 95–104. https://doi.org/10.1016/j.soilbio.2013.01.025

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Applied and Environmental Microbiology*, *79*(17), 5112–5120. https://doi.org/10.1128/AEM.01043-13

Kroeger, M. E., Meredith, L. K., Meyer, K. M., Webster, K. D., Barbosa, P., Fonseca, L., Siu, D. S., Tsai, M., Haren, J. Van, Saleska, S., & Bohannan, B. J. M. (2021). Rainforest-to-pasture conversion stimulates soil methanogenesis across the Brazilian Amazon. *The ISME Journal*, 658–672. https://doi.org/10.1038/s41396-020-00804-x

Kuddus, M., & Ramteke, P. W. (2009). Cold-active extracellular alkaline protease from an alkaliphilic Stenotrophomonas maltophilia: production of enzyme and its industrial applications. *Canadian Journal of Microbiology*, *55*(11), 1294–1301. https://doi.org/10.1139/W09-089

Kumar, S., Suyal, D. C., Yadav, A., Shouche, Y., & Goel, R. (2019). Microbial diversity and soil physiochemical characteristic of higher altitude. *PLoS ONE*, *14*(3), 1–15. https://doi.org/10.1371/journal.pone.0213844

Le Roux, X., Bouskill, N. J., Niboyet, A., Barthes, L., Dijkstra, P., Field, C. B., Hungate, B. A., Lerondelle, C., Pommier, T., Tang, J., Terada, A., Tourna, M., & Poly, F. (2016). Predicting the responses of soil nitrite-oxidizers to multi-factorial global change: A trait-based approach. *Frontiers in Microbiology*, *7*(MAY), 1–13. https://doi.org/10.3389/fmicb.2016.00628

Lee, D. W., & Lee, S. D. (2009). Dyella marensis sp. nov., isolated from cliff soil. *International Journal of Systematic and Evolutionary Microbiology*, *59*(6), 1397–1400. https://doi.org/10.1099/ijs.0.000968-0

Leonard, R. L., Kaplan, L. A., Elder, J. F., Coats, R. N., & Goldman, C. R. (1979). Nutrient transport in surface runoff from a subalpine watershed, Lake Tahoe Basin, California. *Ecological Monographs*, *49*(3), 281–310. https://doi.org/10.2307/1942486

Liu, D., Wu, X., Shi, S., Liu, H., & Liu, G. (2016). A hollow bacterial diversity pattern with elevation in Wolong Nature Reserve, Western Sichuan Plateau. *Journal of Soils and Sediments*, *16*(10), 2365–2374. https://doi.org/10.1007/s11368-016-1422-5

Liu, X., Cong, J., Lu, H., Xue, Y., Wang, X., Li, D., & Zhang, Y. (2017). Community structure and elevational distribution pattern of soil Actinobacteria in alpine grasslands. *Acta Ecologica Sinica*, *37*(4), 213–218. https://doi.org/10.1016/j.chnaes.2017.02.010

Liu, Y., Yao, T., Jiao, N., Tian, L., Hu, A., Yu, W., & Li, S. (2011). Microbial diversity in the snow, a moraine lake and a stream in Himalayan glacier. *Extremophiles*, *15*(3), 411–421. https://doi.org/10.1007/s00792-011-0372-5

Løjtnant, B., & Molau, U. (1983). Analysis of a virgin páramo plant community on Volcán Sumaco, Ecuador. *Nordic Journal of Botany*, *2*(6), 567–574. https://doi.org/10.1111/j.1756-1051.1983.tb01051.x

Lomolino, M. V. (2001). Elevation gradients of species-density: Historical and prospective views. *Global Ecology and Biogeography*, *10*(1), 3–13. https://doi.org/10.1046/j.1466-822x.2001.00229.x

Looby, C. I., & Martin, P. H. (2020). Diversity and function of soil microbes on montane gradients: the state of knowledge in a changing world. *FEMS Microbiology Ecology*, *96*(9). https://doi.org/10.1093/femsec/fiaa122

Lozano, P., Cabrera, O., Peyre, G., Cleef, A., & Toulkeridis, T. (2020). Plant diversity and composition changes along an altitudinal gradient in the isolated volcano sumaco in the ecuadorian amazon. *Diversity*, *12*(6). https://doi.org/10.3390/D12060229

Małecki, J., Lovell, H., Ewertowski, W., Górski, Ł., Kurczaba, T., Latos, B., Miara, M., Piniarska, D., Płocieniczak, J., Sowada, T., Spiralski, M., Warczachowska, A., & Rabatel, A. (2018). The glacial landsystem of a tropical glacier: Charquini Sur, Bolivian Andes. *Earth Surface Processes and Landforms*, *43*(12), 2584–2602. https://doi.org/10.1002/esp.4417

Männistö, M. K., Rawat, S., Starovoytov, V., & Häggblom, M. M. (2012). Granulicella arctica sp. nov., granulicella mallensis sp. nov., granulicella tundricola sp. nov. and granulicella sapmiensis sp. nov., novel acidobacteria from tundra soil. *International Journal of Systematic and Evolutionary Microbiology*, *62*(9), 2097–2106. https://doi.org/10.1099/ijs.0.031864-0

Margesin, R., Shinner, F., Marx, J.-C., & Gerday. (2009). *Psychrophiles from biodiversity to biotechnology*. Berlin: Springer Science & Business Media.

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

McMurdie, P. J., & Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Computational Biology*, *10*(4). https://doi.org/10.1371/journal.pcbi.1003531

Melo, V. F., Barros, L. S., Silva, M. C. S., Veloso, T. G. R., Senwo, Z. N., Matos, K. S., & Nunes, T. K. O. (2021). Soil bacterial diversities and response to deforestation, land use and burning in North Amazon, Brazil. *Applied Soil Ecology*, *158*(December 2019), 103775. https://doi.org/10.1016/j.apsoil.2020.103775

Mendes, L. W., de Lima Brossi, M. J., Kuramae, E. E., & Tsai, S. M. (2015). Land-use system shapes soil bacterial communities in Southeastern Amazon region. *Applied Soil Ecology*, *95*, 151–160. https://doi.org/10.1016/j.apsoil.2015.06.005

Merloti, L. F., Mendes, L. W., Pedrinho, A., de Souza, L. F., Ferrari, B. M., & Tsai, S. M. (2019). Forest-to-agriculture conversion in Amazon drives soil microbial communities and N-cycle. *Soil Biology and Biochemistry*, *137*(February), 107567. https://doi.org/10.1016/j.soilbio.2019.107567

Miteva, V. I., Sheridan, P. P., & Brenchley, J. E. (2004). Phylogenetic and Physiological Diversity of Microorganisms Isolated from a Deep Greenland Glacier Ice Core. *Applied and Environmental Microbiology*, *70*(1), 202–213. https://doi.org/10.1128/AEM.70.1.202-213.2004

Monzier, M., Samaniego, P., & Robin, C. (1996). Le volcan Cayambe (Equateur) : son activité au cours des 5000 dernières années et les menaces qui en résultent. In *Bull. Inst. fr. études andines* (Vol. 25, Issue 3).

Morán-Tejeda, E., López-Moreno, J. I., & Beniston, M. (2013). The changing roles of temperature and precipitation on snowpack variability in Switzerland as a function of altitude. *Geophysical Research Letters*, *40*(10), 2131–2136. https://doi.org/10.1002/grl.50463

Mun, C. L., & Wong Vui Ling, C. M. (2020). Effects of elevated temperature on the tropical soil bacterial diversity. *Sains Malaysiana*, *49*(10), 2335–2344. https://doi.org/10.17576/jsm-2020-4910-01

Nascimento, L., Pedrinho, A., Tereza, A., Vasconcelos, R. De, Mui, S., & William, L. (2021). Amazon deforestation enriches antibiotic resistance genes. *Soil Biology and Biochemistry*, *153*(December 2020), 108110. https://doi.org/10.1016/j.soilbio.2020.108110

Nearing, J. T., Douglas, G. M., Comeau, A. M., & Langille, M. G. I. (2018). Denoising the Denoisers: An independent evaluation of microbiome sequence error- correction approaches. *PeerJ*, *2018*(8). https://doi.org/10.7717/peerj.5364

Nikitin, D. I., Strömpl, C., Oranskaya, M. S., & Abraham, W. R. (2004). Phylogeny of the ring-forming bacterium Arcicella aquatica gen. nov., sp. nov. (ex Nikitin et al. 1994), from a freshwater neuston biofilm. *International Journal of Systematic and Evolutionary Microbiology*, *54*(3), 681–684. https://doi.org/10.1099/ijs.0.02896-0

Oerlemans, J. (1994). Quantifying global warming from the retreat of glaciers. *Science*, *264*(5156), 243–245. https://doi.org/10.1126/science.264.5156.243

Okamura, K., Hisada, T., Kanbe, T., & Hiraishi, A. (2009). Rhodovastum atsumiense gen. nov., sp. nov., a phototrophic alphaproteobacterium isolated from paddy soil. *Journal of General and Applied Microbiology*, *55*(1), 43–50. https://doi.org/10.2323/jgam.55.43

Ordoñez, O. F., Flores, M. R., Dib, J. R., Paz, A., & Farías, M. E. (2009). Extremophile culture collection from Andean lakes: Extreme pristine environments that host a wide diversity of microorganisms with tolerance to UV radiation. *Microbial Ecology*, *58*(3), 461–473. https://doi.org/10.1007/s00248-009-9527-7

Peter, H., & Sommaruga, R. (2016). Shifts in diversity and function of lake bacterial communities upon glacier retreat. *ISME Journal*, *10*(7), 1545–1554. https://doi.org/10.1038/ismej.2015.245

Pfeiffer, F., Gröber, C., Blank, M., Händler, K., Beyer, M., Schultze, J. L., & Mayer, G. (2018). *Systematic evaluation of error rates and causes in short samples in next- generation sequencing*. *July*, 1–14. https://doi.org/10.1038/s41598-018-29325-6

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team. (2020). *R: a language and environment for statistical computing. R Foundation for Statistical Computing*.

Rabatel, A., Francou, B., Soruco, A., Gomez, J., Cáceres, B., Ceballos, J. L., Basantes, R., Vuille, M., Sicart, J. E., Huggel, C., Scheel, M., Lejeune, Y., Arnaud, Y., Collet, M., Condom, T., Consoli, G., Favier, V., Jomelli, V., Galarraga, R., … Wagnon, P. (2013a). Current state of glaciers in the tropical Andes: a multi-century perspective on glacier evolution and climate change. *Cryosphere*, *7*(1), 81–102. https://doi.org/10.5194/tc-7-81-2013

Rabatel, A., Francou, B., Soruco, A., Gomez, J., Cáceres, B., Ceballos, J. L., Basantes, R., Vuille, M., Sicart, J. E., Huggel, C., Scheel, M., Lejeune, Y., Arnaud, Y., Collet, M., Condom, T., Consoli, G., Favier, V., Jomelli, V., Galarraga, R., … Wagnon, P. (2013b). Current state of glaciers in the tropical Andes: A multi-century perspective on glacier evolution and climate

change. *Cryosphere*, *7*(1), 81–102. https://doi.org/10.5194/tc-7-81-2013

Rabatel, Antoine, Ceballos, J. L., Micheletti, N., Jordan, E., Braitmeier, M., González, J., Mölg, N., Ménégoz, M., Huggel, C., & Zemp, M. (2018). Toward an imminent extinction of Colombian glaciers? *Geografiska Annaler, Series A: Physical Geography*, *100*(1), 75–95. https://doi.org/10.1080/04353676.2017.1383015

Ragot, S., Zeyer, J., Zehnder, L., Reusser, E., Brandl, H., & Lazzaro, A. (2013). Bacterial community structures of an alpine apatite deposit. *Geoderma*, *202–203*, 30–37. https://doi.org/10.1016/j.geoderma.2013.03.006

Ranjan, K., Paula, F. S., Mueller, R. C., Jesus, E. da C., Cenciani, K., Bohannan, B. J. M., Nüsslein, K., & Rodrigues, J. L. M. (2015). Forest-to-pasture conversion increases the diversity of the phylum Verrucomicrobia in Amazon rainforest soils. *Frontiers in Microbiology*, *6*(JUL), 1–9. https://doi.org/10.3389/fmicb.2015.00779

Raper, S. C. B., & Braithwaite, R. J. (2006). Low sea level rise projections from mountain glaciers and icecaps under global warming. *Nature*, *439*(7074), 311–313. https://doi.org/10.1038/nature04448

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *2016*(10), 1–22. https://doi.org/10.7717/peerj.2584

Salama, D. M., Meyer, T., & Kyndt, J. (2020). Genome Sequence of the Acidophilic Nonsulfur Purple Photosynthetic Alphaproteobacterium Rhodovastum atsumiense, a Divergent Member of the Acetobacteraceae Family. *Microbiology Resource Announcements*, *January*, 16–18.

Samaniego, P., Monzier, M., Robin, C., & Hall, M. L. (1998). Late Holocene eruptive activity at Nevado Cayambe volcano, Ecuador. *Bulletin of Volcanology*, *59*(7), 451–459. https://doi.org/10.1007/s004450050203

Schloss, P. D. (2020). Reintroducing mothur: 10 years later. *Applied and Environmental Microbiology*, *86*(2), 1–13. https://doi.org/10.1128/AEM .02343-19

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09

Shen, C., Ni, Y., Liang, W., Wang, J., & Chu, H. (2015). Distinct soil bacterial communities along

a small-scale elevational gradient in alpine tundra. *Frontiers in Microbiology*, *6*(JUN), 1–12. https://doi.org/10.3389/fmicb.2015.00582

Sheu, S. Y., Yang, C. S., Chen, M. H., Arun, A. B., Young, C. C., & Chen, W. M. (2010). Arcicella aurantiaca sp. nov., isolated from stream water. *International Journal of Systematic and Evolutionary Microbiology*, *60*(12), 2979–2983. https://doi.org/10.1099/ijs.0.021527-0

Shi, M. M., Jiang, Y. G., & Shi, L. (2019). Electromicrobiology and biotechnological applications of the exoelectrogens Geobacter and Shewanella spp. *Science China Technological Sciences*, *62*(10), 1670–1678. https://doi.org/10.1007/s11431-019-9509-8

Shi, Y., & Liu, S. (2000). Estimation on the response of glaciers in China to the global warming in the 21st century. *Chinese Science Bulletin*, *45*(7), 668–672. https://doi.org/10.1007/BF02886048

Shivaji, S., Pratibha, M. S., Sailaja, B., Hara Kishore, K., Singh, A. K., Begum, Z., Anarasi, U., Prabagaran, S. R., Reddy, G. S. N., & Srinivas, T. N. R. (2011). Bacterial diversity of soil in the vicinity of Pindari glacier, Himalayan mountain ranges, India, using culturable bacteria and soil 16S rRNA gene clones. *Extremophiles*, *15*(1), 1–22. https://doi.org/10.1007/s00792-010-0333-4

Sorg, A., Kääb, A., Roesch, A., Bigler, C., & Stoffel, M. (2015). Contrasting responses of Central Asian rock glaciers to global warming. *Scientific Reports*, *5*, 1–6. https://doi.org/10.1038/srep08228

Stibal, M., Bradley, J. A., Edwards, A., Hotaling, S., Zawierucha, K., Rosvold, J., Lutz, S., Cameron, K. A., Mikucki, J. A., Kohler, T. J., Šabacká, M., & Anesio, A. M. (2020). Glacial ecosystems are essential to understanding biodiversity responses to glacier retreat. *Nature Ecology & Evolution*, *4*(5), 686–687. https://doi.org/10.1038/s41559-020-1163-0

Stolz, A. (2009). Molecular characteristics of xenobiotic-degrading sphingomonads. *Applied Microbiology and Biotechnology*, *81*(5), 793–811. https://doi.org/10.1007/s00253-008-1752-3

Stolz, J. F. (2017). *Gaia and her microbiome*. *October 2016*, 1–13. https://doi.org/10.1093/femsec/fiw247

Torres, B., Vasco, C., Günter, S., & Knoke, T. (2018). Determinants of agricultural diversification in a hotspot area: Evidence from colonist and indigenous communities in the Sumaco Biosphere Reserve, Ecuadorian Amazon. *Sustainability (Switzerland)*, *10*(5). https://doi.org/10.3390/su10051432

Van den Burg, B. (2003). Extremophiles as a source for novel enzymes. *Current Opinion in Microbiology*, *6*(3), 213–218. https://doi.org/10.1016/S1369-5274(03)00060-2

Vitousek, P., Chadwick, O., Matson, P., Allison, S., Derry, L., Kettley, L., Luers, A., Mecking, E., Monastra, V., & Porder, S. (2003). Erosion and the Rejuvenation of Weathering-derived Nutrient Supply in an Old Tropical Landscape. *Ecosystems*, *6*(8), 762–772. https://doi.org/10.1007/s10021-003-0199-8

Wang, H., Lu, L., Mao, D., Huang, Z., Cui, Y., Jin, S., Zuo, Y., & Ren, Z. J. (2019). Dominance of electroactive microbiomes in bioelectrochemical remediation of hydrocarbon-contaminated soils with different textures. *Chemosphere*, *235*, 776–784. https://doi.org/10.1016/j.chemosphere.2019.06.229

Wang, J., Wang, C., Li, J., Bai, P., Li, Q., Shen, M., Li, R., Li, T., & Zhao, J. (2018). Comparative Genomics of Degradative Novosphingobium Strains With Special Reference to Microcystin-Degrading Novosphingobium sp. THN1. *Frontiers in Microbiology*, *9*(September), 1–17. https://doi.org/10.3389/fmicb.2018.02238

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Weon, H. Y., Anandham, R., Kim, B. Y., Hong, S. B., Jeon, Y. A., & Kwon, S. W. (2009). Dyella soli sp. nov. and Dyella terrae sp. nov., isolated from soil. *International Journal of Systematic and Evolutionary Microbiology*, *59*(7), 1685–1690. https://doi.org/10.1099/ijs.0.004838-0

Westcott, S. L., & Schloss, P. D. (2017). OptiClust, an improved method for assigning amplicon-based sequence data to Operational Taxonomic Units. *MSphere*, *2*(2), e00073-17. https://doi.org/10.1128/mSphereDirect.00073-17

Wolinska, A., Górniak, D., Zielenkiewicz, U., Kuzniar, A., Izak, D., Banach, A., & Błaszczyk, M. (2019). Actinobacteria structure in autogenic, hydrogenic and lithogenic cultivated and non-cultivated soils: A culture-independent approach. *Agronomy*, *9*(10). https://doi.org/10.3390/agronomy9100598

Xie, C. H., & Yokota, A. (2005). Dyella japonica gen. nov., sp. nov., a γ-proteobacterium isolated from soil. *International Journal of Systematic and Evolutionary Microbiology*, *55*(2), 753–756. https://doi.org/10.1099/ijs.0.63377-0

Yadav, A. N., Verma, P., Sachan, S. G., & Saxena, A. K. (2017). Biodiversity and Biotechnological Applications of Psychrotrophic Microbes Isolated from Indian Himalayan Regions. *EC Microbiology ECO*, *1*, 48–54.

Yang, C., & Vui Ling, W. (2017). Effects of simulated warming on bacterial diversity and abundance in tropical soils from East Malaysia using open top chambers. *Canadian Journal of Microbiology*.

Yang, Y., Wang, X., Shi, J., & Li, J. (2012). The influence of the discharging sewage on microbial diversity in sludge from Dongting Lake. *World Journal of Microbiology and Biotechnology*, *28*(2), 421–430. https://doi.org/10.1007/s11274-011-0831-2

Zhou, X. G., Ren, L. F., Li, Y. T., Zhang, M., Yu, Y. De, & Yu, J. (2010). The next-generation sequencing technology: A technology review and future perspective. *Science China Life Sciences*, *53*(1), 44–57. https://doi.org/10.1007/s11427-010-0023-6

# Resumen en Castellano

# 4..1 Introducción

Uno de los objetivos de la microbiología ambiental es descifrar y comprender el complejo ecosistema terrestre y acuático. En particular, la evaluación microbiana de ecosistemas extremos. En este tipo de ecosistemas, los organismos altamente adaptados habitan los paisajes más adversos. La adaptación a hábitats adversos es muy dinámica y compleja, y a menudo más de un frente evolutivo está involucrado en este proceso. Por ejemplo, en los glaciares, las temperaturas por debajo de 0 ºC, la alta radiación ultravioleta y la escasez de fuentes son factores desafiantes que los microbios en este tipo de hábitats deben enfrentar con éxito (Shivaji & Prakash, 2010). Se han descubierto diferentes rutas para superar estas fuerzas evolutivas. Los microbios fitofílicos son capaces de reducir el consumo de energía y de expresar enzimas de alta actividad catalítica que pueden aplicarse en la industria (Kuddus & Ramteke, 2009). En el otro lado, se tiene microbios que están adaptados para sobrevivir a altas temperaturas, microorganismos termófilos. Estos microorganismos han ganado mucha más atención debido a que las enzimas son estructuralmente estables a altas temperaturas, a las cuales se mejoran diferentes reacciones (Van den Burg, 2003). Esta introducción cubrirá el aspecto más importante de las dos ubicaciones de estudio, el glaciar Cayambe y el volcán extinto Sumaco, así como la tecnología empleada para lograr los objetivos.

## 4.1.1 Cayambe

El glaciar Cayambe es un centro volcánico explosivo activo masivo con una extensión de base de 24 × 18 Km. Se eleva a una altitud de 5790 m, está cubierto por un vasto casquete de hielo de casi 22 $km^2$, con un volumen aproximado de 0,7 $km^3$, y con un espesor que alcanza los 100 m en algunos lugares (Cours, Dernières, Et, & Samaniego, 1996; Guillier y Chatelain, 2006) (Figura 1.1a). El casquete glaciar de Cayambe se mantiene constante por encima de los 4800 msnm y desciende a ~ 4600 metros sobre el nivel del mar (m.s.n.m) en su flanco occidental y ~ 4200 m.s.n.m en su flanco oriental (Bax & Francesconi, 2019; Detienne et al., 2017; Samaniego, Monzier, Robin y Hall, 1998) (Figura 1.1ay 1.2c). El retroceso de los glaciares del Cayambe se ha

estimado en un 25,58% entre 1979 y 2009 (Gallegos Castro, Brito Chasiluisa, Serrano Giné y Galárraga Sánchez, 2018).

El glaciar Cayambe es único en su ubicación geográfica, ya que se encuentra esencialmente en latitud cero (0.03º N; 77.988º W) con un lago natural llamado “Laguna verde” (Figura 1.1d). Esta ubicación agrega a Cayambe a la categoría de glaciares tropicales. Los glaciares tropicales son de especial interés por su alta sensibilidad a las variaciones ambientales, especialmente al incremento de temperatura (Rabatel et al., 2013). Varios estudios apuntan a un retroceso más rápido de los glaciares tropicales en comparación con los glaciares no tropicales. Durante los últimos 4000 años, el glaciar Cayambe ha experimentado 21 erupciones volcánicas, la más reciente ocurrió en 1785-1786 (Samaniego et al., 1998). El glaciar también sirve como fuente de agua para las comunidades río abajo, incluidas las grandes ciudades como Quito.

### 4.1.2 Sumaco

El volcán Sumaco está ubicado en la región amazónica, y junto con El Reventador, Pan de Azúcar son los tres volcanes ubicados en el arco posterior, que también se considera como el vulcanismo de arco posterior más al sur de Ecuador (Garrison et al. 2018; Hoffer et al., 2008) (Figura 1.2). El volcán Sumaco y su bioma fueron declarados Parque Nacional en 1994 por el Ministro de Medio Ambiente de Ecuador. Tiene una extensión de 205,751 ha. Según la unión internacional, esta distinción tiene como objetivo proteger la biodiversidad natural en el estado natural posible para la conservación de la naturaleza. Además, las políticas ecuatorianas permiten la realización de actividades turísticas en esta zona (Simkin & Siebert, 1994), fuente de perturbación del ecosistema. Dado que la última posible erupción de Sumaco tuvo lugar en 1933 (Simkin & Siebert, 1994), ahora alberga una rica diversidad vegetal (Lozano, Cabrera, Peyre, Cleef, & Toulkeridis, 2020) y animal (Hodge & Arbogast, 2016). Las características únicas del paisaje que exhibe la biosfera de Sumaco han contribuido a hacer del acceso a este bioma un desafío real y complejo (Bühler, Robert, & J, n.d.). Aunque se han realizado algunos estudios para documentar la diversidad de plantas y carnívoros Sumaco, su microbioma sigue siendo completamente desconocido. Por lo tanto, es urgente realizar un estudio del

microbioma en el ecosistema menos perturbado para descubrir el microcosmos de Sumaco.

## 4.2 Objetivos

### 4.2.1 Objetivo general

Revelar la biodiversidad bacteriana de los volcanes andinos Sumaco y Cayambe del Ecuador en la búsqueda de catalogar e identificar las comunidades bacterianas de estos ecosistemas.

### 4.2.2 Objetivos específicos

- Analizar la diversidad y estructura de la comunidad bacteriana de los volcanes Cayambe y Sumaco a través de un estudio de tres años con un marcador molecular universal.
- Relacionar las comunidades bacterianas de los volcanes Cayambe y Sumaco con sus parámetros fisicoquímicos.
- Ensamblar una librería metagenómica a partir del microbioma del suelo de los volcanes Cayambe y Sumaco.

## 4.3 Discusión y Conclusiones

### 4.3.1 Cayambe

El complejo volcánico Cayambe mostró una notable relación inversa entre altitud y diversidad, dentro de un rango de 1089 m entre 5810 y 4721 m.s.n.m. En contraste con otra cronosecuencia de primer plano de glaciares que incrementa directamente la riqueza de la comunidad bacteriana debido a la escala de tiempo prolongada (décadas), la cronosecuencia de Cayambe mostró el comportamiento opuesto, principalmente para el rango estrecho de cinco años. Esperamos tener una cronosecuencia similar cuando tenga lugar la sucesión primaria en este ecosistema. También propusimos la primera metacomunidad de un glaciar tropical considerando diferentes nichos en Cayambe, como suelos, hielo, nieve y subambientes lacustres. La metacomunidad está principalmente ensamblada por *Proteobacteria, Bacteroidetes, Firmicutes, Acidobacteria, Actinobacteria, Planctomycetes.*

También converge con el filo dominante previamente informado para los glaciares. Proponemos que los mecanismos que crea esta comunidad altamente estructurada son los efectos de colonización y dispersión.

A excepción de Ca, $CaCO_3$ y TH en muestras de agua, encontramos bajas correlaciones entre las variables geoquímicas y la estructura de la comunidad bacteriana observada. Además, se descubrió que el N, el Mn y la humedad tienen un papel esencial como variables de configuración de la comunidad. Por ejemplo, *Nitrobacter*, un género oxidante de nitritos, se encontró en altitudes elevadas, donde la concentración de N es mayor. *Pseudomonas* es uno de los géneros más interesantes que se encuentran en el bioma de los glaciares, ya que se ha informado con varios roles en el ecosistema. Por ejemplo, bacterias solubilizadoras de P y bacterias nucleantes de hielo. Sin embargo, también se informó una fuerte resistencia a los antibióticos para los aislados de *Pseudomonas* de los glaciares.

Además, el género más prometedor, según nuestro análisis, es *Stenotrophomonas*. Se informó que este género es capaz de producir enzimas activas en frío con aplicaciones en la industria de los detergentes. Según nuestros datos, el mejor lugar para iniciar la bioprospección de *Stenotrophomonas* es a gran altura. Además, aunque *Granulicella* ha sido descrita como un posible biomarcador de la sucesión del primer campo, encontramos que *Granulicella* es parte del bioma de Cayambe, lo que sugiere una presencia cosmopolita en toda la cronosecuencia del glaciar.

Aunque nuestro trabajo es el primero en los glaciares del trópico ecuatorial, varias preguntas siguen sin respuesta. La pregunta principal está relacionada con la arquitectura de la comunidad bacteriana para la próxima cronosecuencia y solo puede ser revelada por estudios microbianos futuros. Dado que empleamos la metodología metagenómica de amplicón, solo pudimos realizar perfiles taxonómicos. El siguiente paso es emplear técnicas más avanzadas para descubrir la diversidad funcional. En consecuencia, con el hallazgo de una alta abundancia de *Stenotrophomonas*, se debe realizar un cultivo y aislamiento bacteriano para determinar la expresión potencial de los bioproductos de interés antes del retroceso del glaciar.

### 4.3.2 Sumaco

El microcosmos del volcán Sumaco no mostró una relación entre el gradiente de elevación y la diversidad alfa, dentro de un intervalo. Dado que en invierno se realizó un muestreo, nos permitió descubrir que en esta temporada la diversidad alfa alcanza los índices de diversidad más bajos. Por otro lado, cuando se introduce la arquitectura del microcosmos en el análisis, se puede lograr una clara separación entre las muestras de alta y baja altitud. En otras palabras, el análisis de diversidad beta mostró una composición microbiana diferente entre las dos altitudes. Además, las muestras de baja altitud exhiben una comunidad microbiana más homogénea que las muestras de gran altitud. También se brinda la primera propuesta de la metacomunidad de un volcán tropical considerando dos nichos diferentes (suelo y lago). Esta metacomunidad está dominada principalmente por *Proteobacteria, Verrumicrobia, Acidobacteria y Actinobacteria.*

Sin embargo, debido a la relación del tamaño de la muestra entre las muestras de suelo y de lagos, tenemos que reconocer que la metacomunidad sugerida está altamente influenciada por los filos más abundantes en el nicho del suelo. Además, la metacomunidad propuesta está en línea con el filo previamente reportado en ecosistemas de gran altitud.

Aunque cualquiera de las variables ambientales evaluadas resultó en diferencias significativas entre altitudes, algunos patrones interesantes pueden influir en el microcosmos de Sumaco. Por ejemplo, el contenido de MO del suelo no siguió el comportamiento esperado de estar en mayores concentraciones a bajas altitudes. Se sugiere que la topografía de Sumaco juega un papel vital en la creación de este comportamiento. Hoy en día, existe una falta de investigación geológica en el volcán Sumaco, lo que oscurece la contribución geológica de Sumaco a su microbioma. Además, se encontró que todos los factores ambientales están en mayor concentración en el lago a baja altitud, y puede explicarse por el nutriente y el elemento de gran altitud a baja altitud por múltiples factores, pero principalmente por la escorrentía.

Se encuentra varios géneros que siempre son diferentes a través de la cronosecuencia en ambas altitudes. Aunque casi la mitad de las funciones

ecológicas de estos géneros aún no se han aclarado, algunas son de gran interés. Por ejemplo, *Nitrospira* puede usar urea como energía para transformar urea en amoníaco y $CO_2$. *Geobacter y Novosphigobiu* son los géneros más atractivos encontrados debido a su posible aplicación en la producción de electricidad y degradación de compuestos peligrosos, respectivamente.

Aunque este trabajo es el primero en los volcanes del trópico ecuatorial, varias preguntas siguen sin respuesta. La pregunta principal está relacionada con la arquitectura de la comunidad bacteriana para la próxima cronosecuencia y solo puede ser revelada por estudios microbianos futuros. Dado que se empleó la metodología metagenómica de amplicón, solo pudimos realizar perfiles taxonómicos. El siguiente paso es emplear técnicas más avanzadas para descubrir la diversidad funcional. En consecuencia, con el hallazgo de alta abundancia *de Geobacter y Novosphigobiuma,* se debe realizar cultivo y aislamiento bacteriano para determinar la expresión potencial de bioproductos de interés.

Este documento contiene información parcial, que ya ha sido publicada en dos artículos descritos en los anexos.

1. Frontiers | Soil Bacterial Community Along an Altitudinal Gradient in the Sumaco, a Stratovolcano in the Amazon Region (frontiersin.org) https://doi.org/10.3389/ffgc.2022.738568

2. The Ecuadorian Microbiome project: a plea to strengthen microbial genomic research. https://doi.org/10.1080/23766808.2021.1938900

# Annex

frontiers in Forests and Global Change

ORIGINAL RESEARCH
published: 24 March 2022
doi: 10.3389/ffgc.2022.738568

# Soil Bacterial Community Along an Altitudinal Gradient in the Sumaco, a Stratovolcano in the Amazon Region

*Magdalena Díaz[1,2,3,4], Cristian Quiroz-Moreno[5], Pablo Jarrín-V[6], Samuel Piquer-Esteban[4], Pablo Monfort-Lanzas[4], Erika Rivadeneira[3], Pablo Castillejo[7], Vicente Arnau[4], Wladimiro Díaz[4], Félix J. Sangari[1,8] and C. Alfonso Molina[3,9]**

[1] Departamento de Biología Molecular, Universidad de Cantabria, Santander, Spain, [2] Facultad de Ingeniería Química, Universidad Central del Ecuador, Quito, Ecuador, [3] Instituto de Investigación en Zoonosis (CIZ), Universidad Central del Ecuador, Quito, Ecuador, [4] Institute of Integrative Systems Biology (I2Sysbio), University of Valencia and Consejo Superior de Investigaciones Científicas (CSIC), Valencia, Spain, [5] Department of Horticulture and Crop Science, The Ohio State University, Columbus, OH, United States, [6] Instituto Nacional de Biodiversidad, Quito, Ecuador, [7] Grupo de Investigación en Biodiversidad, Medio Ambiente y Salud (BIOMAS), Universidad de Las Américas, Quito, Ecuador, [8] Instituto de Biomedicina y Biotecnología de Cantabria, CSIC – Universidad de Cantabria, Santander, Spain, [9] Facultad de Medicina Veterinaria y Zootecnia, Universidad Central del Ecuador, Quito, Ecuador

OPEN ACCESS

**Edited by:**
*Matthias Ros, Unidad Oaxaca, Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional, Instituto Politécnico Nacional, Mexico*

**Reviewed by:**
*Alan Feest, University of Bristol, United Kingdom*
*Benito Mendoza, National University of Chimborazo, Ecuador*

***Correspondence:**
*C. Alfonso Molina*
*camolina@uce.edu.ec*

**Specialty section:**
*This article was submitted to Tropical Forests, a section of the journal Frontiers in Forests and Global*

Our study is a pioneering exploration of the microbiome in the soil of the Sumaco stratovolcano and an assessment of the effects of an elevational gradient and related physicochemical soil parameters on richness and community structure. The Sumaco, as an isolated Amazonian stratovolcano, may be among one of the least studied ecosystems in Ecuador and perhaps the Amazon region. Universal patterns remain unresolved or available information inconclusive to establish a supported consensus on general governing processes by which elevation and its associated environmental gradients may determine the microbial richness and community structure. We tested a recent proposal on how microbial diversity responds to montane gradients, placing a central role in soils as potentially independent of altitude along an elevational gradient. Correlations and effects among soil physicochemical parameters and altitude were contrasted against richness and community structure through quantitative ecology. The

NEOTROPICAL BIODIVERSITY
2021, VOL. 7, NO. 1, 223–237
https://doi.org/10.1080/23766808.2021.1938900

Taylor & Francis
Taylor & Francis Group

REVIEW

OPEN ACCESS Check for updates

# The Ecuadorian Microbiome Project: a plea to strengthen microbial genomic research

Magdalena Díaz [a,b], Pablo Jarrín-V [c], Raquel Simarro [d], Pablo Castillejo [e,f], Gabriela N. Tenea [g] and C. Alfonso Molina [a,h]

[a]Institute of Research on Zoonoses (CIZ), Central University of Ecuador, Quito, Ecuador; [b]Chemistry Engineering Faculty, Central University of Ecuador, Quito, Ecuador; [c]Health and Environment Research Group, Universidad Regional Amazónica Ikiam, Tena, Ecuador; [d]Department of Biology, Geology, Physics and Inorganic Chemistry,ESCET, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain; [e]Faculty of Environmental Sciences, SEK International University, Quito, Ecuador; [f]Applied Sciences and Engineering Faculty, Universidad De Las Américas, Quito, Ecuador; [g]Biofood and Nutraceutics Research and Development Group, Faculty of Engineering in Agricultural and Environmental Sciences, Technical University of the North, Ibarra, Ecuador; [h]Faculty of Veterinary Medicine and Zootechnics, Central University of Ecuador, Quito, Ecuador

**ABSTRACT**

A technological revolution provides nations with access to unprecedented quantities of molecular information, and this is particularly evident in the vast and yet poorly understood realm of the microbiome. Traditionally, many developing countries in Asia, Africa, and South America remain marginal participants in the global flow of biodata, which will eventually affect their productivity and economies. Here, we present the Ecuadorian Microbiome Project (EcuMP) as an integrative initiative to close the research gap in the microbiome for Ecuador. We discuss the relevance that the study of the microbiome has for our understanding of diversity and new forms of production and biocapital. We also evaluate the state of research in metagenomics and the microbiome for South America, with emphasis on Ecuador as a small but biodiverse country. In the strict sense of access, understanding, and technological innovation based on molecular data, we propose the definition of bioliteracy. As indirect estimates of bioliteracy, we measured the number of indexed publications, BioProjects, monthly global internet traffic to GenBank, and patent applications in Espacenet. South America has a notable unevenness in scientific productivity related to the microbiome and metagenomics. Brazil leads productivity, with most of the measured parameters remaining one order of magnitude higher than other countries in the region. Participation of South American countries in the global flow of genomic information dwarfs when compared to the US. To reduce the effects of technological dependency and the associated lack of economic productivity, Ecuador should address the technological gap in the study of the microbiome. Our assessment reveals the urgency to translate the study of microbiomes into a source of technological prowess and the basis for local biocapitals.

**ARTICLE HISTORY**
Received 28 March 2019
Accepted 2 June 2021

**KEYWORDS**
Microbial diversity; microbiome; metagenomics; biocapital; bioliteracy

# Data tables

**TABLA 1.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN CAYAMBE (CENTROCESAL, 2016).

| IDENTIFICACIÓN DE LA MUESTRA | ALTITUD | pH | Conductividad eléctrica (µs/cm) | Sólidos Totales Disueltos (mg/L) | Calcio mg/Kg | Magnesio mg/Kg | Sodio mg/Kg | Potasio mg/Kg |
|---|---|---|---|---|---|---|---|---|
| CAY002 | Low | 7,085 | 17,3 | 8 | 1,61 | 0,21 | 0,81 | 0,31 |
| CAY004 | Low | 7,129 | 2,2 | 1 | 0,14 | 0 | 0,17 | 0 |
| CAY005 | Low | 4,917 | 9,8 | 4 | 0,44 | 0,23 | 0,42 | 0,11 |
| CAY007 | Low | 4,815 | 14,7 | 6 | 1,03 | 0,12 | 1,38 | 0,15 |
| CAY008 | High | 5,654 | 2,3 | 1 | 0,19 | 0 | 0,21 | 0 |
| CAY009 | High | 5,773 | 8,1 | 4 | 0,58 | 0,16 | 0,43 | 0,1 |
| CAY010 | High | 6,061 | 5,3 | 2 | 0,29 | 0,1 | 0,24 | 0,12 |
| CAY011 | High | 6,638 | 1,9 | 1 | 0,14 | 0 | 0,12 | 0,08 |
| CAY013 | High | 5,651 | 73 | 6 | 0,71 | 0,09 | 1,28 | 0,13 |
| CAY015 | High | 6,132 | 13,5 | 6 | 0,68 | 0,09 | 1,25 | 0,08 |

**TABLA 1.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN CAYAMBE (CENTROCESAL, 2016).

| IDENTIFICACIÓN DE LA MUESTRA | Hierro mg/Kg | Manganeso mg/Kg | Fosfato mg/L | Sulfatos mg/L | Nitrógeno mg/L | Cloruros mg/L | Dureza Total expresada como CaCO3 mg/L | Ca como CaCO3 mg/L |
|---|---|---|---|---|---|---|---|---|
| CAY002 | 0,07 | 0 | 0 | 1,84 | 0 | 2,83 | 4,88 | 4,02 |
| CAY004 | 0 | 0 | 0 | 0 | 0 | 0,47 | 0,35 | 0,35 |
| CAY005 | 0 | 0 | 0 | 0,54 | 0 | 1,57 | 2,05 | 1,1 |
| CAY007 | 0 | 0 | 0 | 0,35 | 0 | 3,41 | 3,07 | 2,57 |
| CAY008 | 0 | 0 | 0 | 0 | 0 | 0,64 | 0,47 | 0,47 |
| CAY009 | 0 | 0 | 0 | 0,69 | 0 | 1,58 | 2,11 | 1,45 |
| CAY010 | 0 | 0 | 0 | 0,33 | 0 | 0,89 | 1,14 | 0,72 |
| CAY011 | 0 | 0 | 0 | 0,21 | 0 | 0,34 | 0,35 | 0,35 |
| CAY013 | 0 | 0 | 0 | 0,37 | 0 | 2,84 | 2,14 | 1,77 |
| CAY015 | 0 | 0 | 0 | 0,49 | 0 | 2,89 | 2,07 | 1,7 |

**TABLA 1.1.** RESULTADOS DE ANÁLISIS FISICO-QUÍMICOS DE MUESTRAS DE SUELO DEL VOLCÁN CAYAMBE (CENTROCESAL, 2016).

| IDENTIFICACIÓN DE LA MUESTRA | ALTITUD | pH | Conductividad eléctrica (µs/cm) | Magnesio mg/Kg | Sodio mg/Kg | Potasio mg/Kg | Hierro mg/Kg | Manganeso mg/Kg | Magnesio mg/Kg |
|---|---|---|---|---|---|---|---|---|---|
| CAY001 | Low | 4,071 | 47,3 | 4,96596796 | 20,1739734 | 51,6753583 | 334,9484 | 4,81741285 | 4,8174 |
| CAY003 | Low | 4,335 | 35,1 | 5,47041214 | 16,5174433 | 56,1676748 | 520,1705 | 4,2616859 | 4,2617 |
| CAY006 | Low | 5,627 | 23,8 | 4,83129707 | 20,5370129 | 56,5142876 | 187,6938 | 4,95876182 | 4,9588 |
| CAY012 | High | 4,753 | 12,5 | 5,01541083 | 25,4943391 | 58,1754386 | 190,2108 | 5,15474262 | 5,1547 |
| CAY014 | High | 3,784 | 97 | 6,14564624 | 20,941937 | 50,8889621 | 50,3623 | 4,89308132 | 5,48931 |

**TABLA 1.1.** RESULTADOS DE ANÁLISIS FISICO-QUÍMICOS DE MUESTRAS DE SUELO DEL (CENTROCESAL, 2016).

| IDENTIFICACIÓN DE LA MUESTRA | Fósforo mg/Kg | Azufre mg/Kg | Nitrógeno %P/P | Materia Orgánica %P/P | Humedad %P/P | Capacidad de intercambio cationico meq/100 g |
|---|---|---|---|---|---|---|
| CAY001 | 3,89 | 12,29846021 | 10,45 | 0,14 | 47,5 | 0,65 |
| CAY003 | 3,36 | 5,491262837 | 8,24 | 21 | 31,8 | 0,66 |
| CAY006 | 3,35 | 3,931115892 | 6,87 | 0,1 | 68,7 | 0,73 |
| CAY012 | 3 | 2,910281836 | 9,25 | 0,12 | 56,3 | 0,69 |
| CAY014 | 3,26 | 37,97415069 | 3,17 | 0,08 | 18,3 | 0,65 |

**TABLA 2.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN CAYAMBE (CENTROCESAL, 2017).

| IDENTIFICACIÓN DE LA MUESTRA | pH | Temperatura | Conductividad eléctrica (µs/cm) | Sólidos Totales Disueltos (mg/L) | Calcio mg/L | Magnesio mg/L | Sodio mg/L | Potasio mg/L |
|---|---|---|---|---|---|---|---|---|
| CAY 017 | 6.762 | 23.0 | 11.9 | 6.0 | 0,204 | 0,083 | 1,215 | 0,987 |
| CAY018 | 5.864 | 23.3 | 3.3 | 1.0 | <0.003 | 0,098 | 0,219 | 0,177 |
| CAY019 | 4.676 | 23.3 | 12.3 | 6.0 | <0.003 | 0,063 | 2,105 | 0,468 |
| CAY020 | 5.286 | 23.4 | 6.5 | 3.0 | <0.003 | 0,118 | 0,681 | 0,359 |
| CAY021 | 5.562 | 23.3 | 2.3 | 1.0 | <0.003 | 0,056 | 0,209 | 0,135 |
| CAY023 | 5.702 | 23.5 | 3.1 | 1.0 | <0.003 | 0,074 | 0,187 | 0,105 |
| CAY024 | 5.935 | 23.5 | 2.0 | 1.0 | <0.003 | 0,102 | 0,212 | 0,143 |
| CAY025 | 5.460 | 23.6 | 4.9 | 2.0 | <0.003 | 0,070 | 0,977 | 0,122 |
| CAY027 | 4.387 | 23.7 | 35.9 | 17.0 | <0.003 | 0,119 | 8,285 | 2,104 |
| CAY028 | 5.712 | 24.4 | 21.6 | 10.0 | <0.003 | 0,106 | 3,271 | 1,305 |
| CAY029 | 4.310 | 23.6 | 38.8 | 18.0 | <0.003 | 0,151 | 7,295 | 2,005 |
| CAY030 | 4.226 | 23.7 | 42.9 | 20.0 | <0.003 | 0,161 | 8,066 | 1,594 |
| CAY031 | 5.553 | 23.4 | 3.3 | 1.0 | <0.003 | 0,060 | 0,187 | 0,318 |
| CAY 032 | 6.236 | 23.7 | 1.4 | 1.0 | <0.003 | 0,043 | 0,158 | 0,235 |
| CAY033 | 5.77 | 23.8 | 2.4 | 1.0 | <0.003 | 0,039 | 0,164 | 0,214 |
| CAY034 | 4.950 | 26.6 | 16.4 | 8.0 | 0,164 | 0,143 | 2,627 | 1,032 |
| CAY035 | 6.095 | 23.8 | 1.9 | 1.0 | <0.003 | 0,081 | 0,287 | 0,087 |
| CAY036 | 5.911 | 23.7 | 1.5 | 1.0 | <0.003 | 0,064 | 0,307 | 0,105 |
| CAY037 | 5.822 | 23.7 | 2.0 | 1.0 | <0.003 | 0,040 | 0,214 | 0,121 |
| CAY039 | 4.678 | 23.6 | 31.2 | 15.0 | <0.003 | 0,104 | 5,272 | 1,896 |

**TABLA 2.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN CAYAMBE (CENTROCESAL, 2017).

| IDENTIFICACIÓN DE LA MUESTRA | Hierro mg/L | Manganeso mg/L | Fosfato mg/L | Sulfatos mg/L | Cloruros mg/L | Nitrógeno mg/L | Dureza Total expresada como CaCO3 mg/L | Ca como CaCO3 mg/L |
|---|---|---|---|---|---|---|---|---|
| CAY 017 | <0.004 | <0.008 | < 0.05 | 1,127 | 2,173 | <0.50 | 0,850 | 0,509 |
| CAY018 | <0.004 | <0.008 | < 0.05 | 0,214 | 0,248 | <0.50 | 0,412 | <0.075 |
| CAY019 | <0.004 | <0.008 | < 0.05 | 1,534 | 1,983 | <0.50 | 0,260 | <0.075 |
| CAY020 | <0.004 | <0.008 | < 0.05 | 0,481 | 0,817 | <0.50 | 0,486 | <0.075 |
| CAY021 | <0.004 | <0.008 | < 0.05 | 0,217 | 0,262 | <0.50 | 0,231 | <0.075 |
| CAY023 | <0.004 | <0.008 | < 0.05 | 0,183 | 0,271 | <0.50 | 0,303 | <0.075 |
| CAY024 | <0.004 | <0.008 | < 0.05 | 0,241 | 0,263 | <0.50 | 0,419 | <0.075 |
| CAY025 | <0.004 | <0.008 | < 0.05 | 0,374 | 0,684 | <0.50 | 0,289 | <0.075 |
| CAY027 | 0,110 | <0.008 | < 0.05 | 2,176 | 7,673 | <0.50 | 0,491 | <0.075 |
| CAY028 | <0.004 | <0.008 | < 0.05 | 1,373 | 4,067 | <0.50 | 0,438 | <0.075 |
| CAY029 | 0,150 | <0.008 | < 0.05 | 2,008 | 6,271 | <0.50 | 0,621 | <0.075 |
| CAY030 | 0,009 | <0.008 | < 0.05 | 3,629 | 6,194 | <0.50 | 0,664 | <0.075 |
| CAY031 | <0.004 | <0.008 | < 0.05 | 0,108 | 0,387 | <0.50 | 0,245 | <0.075 |
| CAY 032 | <0.004 | <0.008 | < 0.05 | 0,097 | 0,293 | <0.50 | 0,178 | <0.075 |
| CAY033 | <0.004 | <0.008 | < 0.05 | 0,115 | 0,386 | <0.50 | 0,159 | <0.075 |
| CAY034 | <0.004 | <0.008 | < 0.05 | 1,934 | 2,088 | <0.50 | 0,996 | 0,409 |
| CAY035 | <0.004 | <0.008 | < 0.05 | 0,131 | 0,294 | <0.50 | 0,332 | <0.075 |
| CAY036 | <0.004 | <0.008 | < 0.05 | 0,099 | 0,314 | <0.50 | 0,265 | <0.075 |
| CAY037 | <0.004 | <0.008 | < 0.05 | 0,157 | 0,266 | <0.50 | 0,164 | <0.075 |
| CAY039 | 0,126 | <0.008 | < 0.05 | 2,319 | 4,167 | <0.50 | 0,428 | <0.075 |

**TABLA 2.1.** RESULTADOS DE ANÁLISIS FISICO-QUÍMICOS DE MUESTRAS DE SUELO DEL VOLCÁN CAYAMBE (CENTROCESAL, 2017).

| IDENTIFICACIÓN DE LA MUESTRA | pH | Temperatura (°C) | Conductividad eléctrica (µs/cm) | Calcio mg/Kg | Magnesio mg/Kg | Sodio mg/Kg | Potasio mg/Kg | Hierro mg/Kg |
|---|---|---|---|---|---|---|---|---|
| CAY022 | 4.571 | 24.6 | 36.2 | 43,43 | 3,57 | 74,30 | 84,21 | 191,32 |
| CAY026 | 3.156 | 26 | 696 | 2,53 | 5,02 | 195,75 | 201,32 | 335,50 |
| CAY038 | 5.026 | 24.1 | 59.8 | 52,67 | 9,44 | 84,33 | 97,32 | 319,71 |

**TABLA 2.1.** RESULTADOS DE ANÁLISIS FISICO-QUÍMICOS DE MUESTRAS DE SUELO DEL VOLCÁN CAYAMBE (CENTROCESAL, 2017).

| IDENTIFICACIÓN DE LA MUESTRA | Manganeso mg/Kg | Fósforo mg/Kg | Azufre mg/Kg | Nitrógeno %P/P | Materia Orgánica %P/P | Humedad %P/P | Capacidad de intercambio cationico meq/100 g |
|---|---|---|---|---|---|---|---|
| CAY022 | 0,32 | 21,32 | 21,01 | 0,94 | 0,842 | 18.43 | 1,70 |
| CAY026 | 0,27 | 31,04 | 18,47 | 1,81 | 1,6741 | 18.39 | 2,50 |
| CAY038 | 0,77 | 27,41 | 28,14 | 0,65 | 0,4287 | 4.40 | 1,10 |

**TABLA 3.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN CAYAMBE (CENTROCESAL, 2019).

| IDENTIFICACIÓN DE LA MUESTRA | MUESTRA | DESCRIPCION DE LA MUESTRA | pH | Temperatura | Conductividad eléctrica (µs/cm) | Sólidos Totales Disueltos (mg/L) | Calcio mg/L |
|---|---|---|---|---|---|---|---|
| CAY040A | Water | líquido incoloro | 5.16 | 22,1 | 7.6 | 3.0 | 0.078 |
| CAY040B | Water | líquido incoloro | 5.10 | 22,4 | 7.7 | 3.1 | 0.079 |
| CAY041A | Water | líquido incoloro | 5.57 | 22,2 | 7.0 | 3.1 | 0.013 |
| CAY041B | Water | líquido incoloro | 5.01 | 22,2 | 7.0 | 3.3 | 0.015 |
| CAY 42A | Water | líquido incoloro | 7,051 | 22,5 | 7,7 | 4,0 | 0,087 |
| CAY 42B | Water | líquido incoloro | 4,731 | 22,2 | 14,7 | 7,0 | 0,127 |
| CAY 43A | Water | líquido incoloro | 5,314 | 22,5 | 9,1 | 4,0 | 0,073 |
| CAY 43B | Water | líquido incoloro | 5,426 | 22,1 | 4,9 | 2,0 | 0,003 |
| CAY 44A | Water | líquido incoloro | 5,527 | 21,6 | 6,1 | 3,0 | 0,058 |
| CAY 44B | Water | líquido incoloro | 5,884 | 22,1 | 2,4 | 1,0 | 0,003 |
| CAY 45A | Water | líquido incoloro | 5,727 | 21,1 | 2,1 | 1,0 | 0,003 |
| CAY 45B | Water | líquido incoloro | 5,752 | 22,0 | 1,6 | 0,7 | 0,003 |
| CAY 46A | Water | líquido incoloro | 5,282 | 22,8 | 12 | 6,0 | 0,102 |
| CAY 46B | Water | líquido incoloro | 5,324 | 21,9 | 3,5 | 2,0 | 0,003 |

**TABLA 3.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN CAYAMBE (CENTROCESAL, 2019).

| IDENTIFICACIÓN DE LA MUESTRA | Magnesio mg/L | Sodio mg/L | Potasio mg/L | Hierro mg/L | Manganeso mg/L | Fosfato mg/L | Sulfatos mg/L | Cloruros mg/L | Nitrógeno mg/L | Dureza Total expresada como CaCO3 mg/L | Ca como CaCO3 mg/L |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CAY040A | 0.099 | 1,9 | 0,8 | 0.04 | 0.08 | 0.05 | 0.117 | 1,5 | 0.500 | 0.648 | 0.075 |
| CAY040B | 0.067 | 2,0 | 0,8 | 0.04 | 0.08 | 0.05 | 0.112 | 1,0 | 0.500 | 0.786 | 0.075 |
| CAY041A | 0.071 | 1,0 | 0,1 | 0.04 | 0.08 | 0.05 | 0.162 | 1,1 | 0.500 | 0.200 | 0.075 |
| CAY041B | 0.070 | 0,0 | 0,2 | 0.04 | 0.08 | 0.05 | 0.166 | 1,0 | 0.500 | 0.076 | 0.075 |
| CAY 42A | 0,095 | 1,875 | 0,785 | 0,004 | 0,008 | 0,05 | 0,357 | 1,597 | 0,500 | 0,608 | 0,218 |
| CAY 42B | 0,120 | 2,129 | 0,944 | 0,004 | 0,008 | 0,05 | 0,163 | 5,834 | 0,500 | 0,813 | 0,318 |
| CAY 43A | 0,086 | 1,926 | 0,427 | 0,004 | 0,008 | 0,05 | 0,309 | 2,817 | 0,500 | 0,535 | 0,182 |
| CAY 43B | 0,065 | 1,021 | 0,128 | 0,004 | 0,008 | 0,05 | 0,118 | 1,706 | 0,500 | 0,267 | 0,075 |
| CAY 44A | 0,072 | 0,947 | 0,254 | 0,004 | 0,008 | 0,05 | 0,127 | 1,134 | 0,500 | 0,439 | 0,144 |
| CAY 44B | 0,068 | 0,325 | 0,104 | 0,004 | 0,008 | 0,05 | 0,117 | 0,368 | 0,500 | 0,280 | 0,075 |
| CAY 45A | 0,059 | 0,376 | 0,124 | 0,004 | 0,008 | 0,05 | 0,119 | 0,401 | 0,500 | 0,241 | 0,075 |
| CAY 45B | 0,049 | 0,206 | 0,087 | 0,004 | 0,008 | 0,05 | 0,088 | 0,167 | 0,500 | 0,200 | 0,075 |
| CAY 46A | 0,128 | 2,420 | 0,371 | 0,004 | 0,008 | 0,05 | 1,673 | 1,417 | 0,500 | 0,779 | 0,254 |
| CAY 46B | 0,128 | 0,782 | 0,086 | 0,004 | 0,008 | 0,05 | 0,118 | 0,841 | 0,500 | 0,528 | 0,075 |

**TABLA 4.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN SUMACO (CENTROCESAL, 2016).

| IDENTIFICACIÓN DE LA MUESTRA | pH | Conductividad eléctrica (µs/cm) | Sólidos Totales Disueltos (mg/L) | Calcio mg/L | Magnesio mg/L | Sodio mg/L | Potasio mg/L | Hierro mg/L |
|---|---|---|---|---|---|---|---|---|
| SUM001 | 6.554 | 2.9 | 1 | 0.12 | 0 | 0.27 | 0.09 | < 0,05 |
| SUM007 | 6.211 | 4.9 | 2 | 0.31 | 0 | 0.32 | 0.11 | < 0,05 |

**TABLA 4.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN SUMACO (CENTROCESAL, 2016).

| IDENTIFICACIÓN DE LA MUESTRA | Manganeso mg/L | Fosfato mg/L | Sulfatos mg/L | Cloruros mg/L | Nitrógeno mg/L | Dureza Total expresada como CaCO3 mg/L | Ca como CaCO3 mg/L |
|---|---|---|---|---|---|---|---|
| SUM001 | < 0,05 | < 0,05 | 0.14 | 0.53 | < 0,5 | 0.32 | 0.32 |
| SUM007 | < 0,05 | < 0,05 | 0.19 | 1.08 | < 0,5 | 0.77 | 0.77 |

**TABLA 4.1.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE SUELO DEL VOLCÁN SUMACO (CENTROCESAL, 2016).

| IDENTIFICACIÓN DE LA MUESTRA | pH | Conductividad eléctrica (µs/cm) | Calcio mg/Kg | Magnesio mg/Kg | Sodio mg/Kg | Potasio mg/Kg | Hierro mg/Kg |
|---|---|---|---|---|---|---|---|
| SUM002 | 4.474 | 66.7 | 95.3 | 8.97 | 29.9 | 64.44 | 418.16 |
| SUM003 | 4.612 | 27.6 | 93.39 | 7.68 | 41.85 | 82.55 | 936.51 |
| SUM004 | 4.704 | 17 | 104.37 | 6.15 | 26.64 | 69.35 | 810 |
| SUM005 | 3.809 | 85.4 | 210.84 | 46.43 | 27.72 | 66.82 | 119.2 |
| SUM006 | 4.69 | 18 | 92.93 | 6.55 | 29.28 | 70 | 664.54 |
| SUM008 | 4.019 | 20 | 136.21 | 25.7 | 65.26 | 422.02 | 4222.22 |
| SUM009 | 3.827 | 67.2 | 371.79 | 55.23 | 23.52 | 66.03 | 154.16 |
| SUM010 | 4.13 | 20 | 201.34 | 10.52 | 58.15 | 98.07 | 494.86 |
| SUM011 | 4.161 | 34.1 | 225.51 | 41.71 | 49.9 | 95.31 | 449.28 |

**TABLA 4.1.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE SUELO DEL VOLCÁN SUMACO (CENTROCESAL, 2016).

| IDENTIFICACIÓN DE LA MUESTRA | Manganeso mg/Kg | Fósforo mg/Kg | Azufre mg/Kg | Nitrógeno %P/P | Materia Orgánica %P/P | Humedad %P/P | Capacidad de intercambio cationico meq/100 g |
|---|---|---|---|---|---|---|---|
| SUM002 | 91.323 | 2.89 | 8.34 | 2.18 | 73.81 | 26.7 | 0.85 |
| SUM003 | 338.771 | 3.24 | 3.43 | 87.64 | 3.45 | 29.5 | 0.92 |
| SUM004 | 5.519 | 4.35 | 3.4 | 164.57 | 4.68 | 29.8 | 0.87 |
| SUM005 | 507.503 | 4.05 | 4.4 | 0.97 | 45.93 | 35.3 | 1.71 |
| SUM006 | 114.251 | 4.23 | 3.18 | 154.76 | 4.06 | 40.8 | 0.82 |
| SUM008 | 65.95 | 5.1 | 4.44 | 0.68 | 38.57 | 30.2 | 1.02 |
| SUM009 | 126.285 | 4.48 | 5.12 | 0.49 | 29.28 | 25.8 | 2.58 |
| SUM010 | 100.988 | 5.47 | 5.37 | 76.91 | 2.58 | 32.7 | 1.05 |
| SUM011 | 187.392 | 6.42 | 6.23 | 80.48 | 3.05 | 38.9 | 1.92 |

**TABLA 5.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN SUMACO (CENTROCESAL, 2017).

| IDENTIFICACIÓN DE LA MUESTRA | pH | Temperatura °C | Conductividad eléctrica (µs/cm) | Sólidos Totales Disueltos (mg/L) | Calcio mg/L | Magnesio mg/L | Sodio mg/L | Potasio mg/L |
|---|---|---|---|---|---|---|---|---|
| SUM012 | 6.103 | 23.2 | 3.5 | 2.0 | <0.003 | 0.069 | 0.8654 | 0.1577 |
| SUM016 | 6.208 | 23.3 | 4.0 | 2.0 | <0.003 | 0.0783 | 0.9127 | 0.0982 |
| SUM012 | 6.103 | 23.2 | 3.5 | 2.0 | <0.003 | 0.069 | 0.8654 | 0.1577 |
| SUM016 | 6.208 | 23.3 | 4.0 | 2.0 | <0.003 | 0.0783 | 0.9127 | 0.0982 |

**TABLA 5.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN SUMACO (CENTROCESAL, 2017).

| IDENTIFICACIÓN DE LA MUESTRA | pH | Temperatura °C | Conductividad eléctrica (µs/cm) | Sólidos Totales Disueltos (mg/L) | Calcio mg/L | Magnesio mg/L | Sodio mg/L | Potasio mg/L |
|---|---|---|---|---|---|---|---|---|
| SUM012 | 6.103 | 23.2 | 3.5 | 2.0 | <0.003 | 0.069 | 0.8654 | 0.1577 |
| SUM016 | 6.208 | 23.3 | 4.0 | 2.0 | <0.003 | 0.0783 | 0.9127 | 0.0982 |
| SUM012 | 6.103 | 23.2 | 3.5 | 2.0 | <0.003 | 0.069 | 0.8654 | 0.1577 |
| SUM016 | 6.208 | 23.3 | 4.0 | 2.0 | <0.003 | 0.0783 | 0.9127 | 0.0982 |

**TABLA 5.1.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE SUELO DEL VOLCÁN SUMACO (CENTROCESAL, 2017).

| IDENTIFICACIÓN DE LA MUESTRA | pH | Temperatura (°C) | Conductividad eléctrica (µs/cm) | Calcio mg/Kg | Magnesio mg/Kg | Sodio mg/Kg | Potasio mg/Kg | Hierro mg/Kg |
|---|---|---|---|---|---|---|---|---|
| SUM013 | 3938 | 24.2 | 122.8 | 1.274.172 | 29.176 | 2.412.514 | 1.373.518 | 275.786 |
| SUM014 | 5334 | 24.1 | 92.7 | 49.109 | 69.272 | 2.053.274 | 806.541 | 670.257 |
| SUM015 | 5574 | 24.2 | 83.0 | 58.716 | 73.258 | 1.972.586 | 920.394 | 1.472.005 |
| SUM017 | 5575 | 24.3 | 83.1 | 68.716 | 83.258 | 1.982.586 | 930.394 | 1.482.005 |
| SUM018 | 5576 | 24.4 | 83.2 | 78.716 | 93.258 | 1.992.586 | 940.394 | 1.492.005 |
| SUM019 | 5577 | 24.5 | 83.3 | 88.716 | 103.258 | 2.002.586 | 950.394 | 1.502.005 |
| SUM020 | 5578 | 24.6 | 83.4 | 98.716 | 113.258 | 2.012.586 | 960.394 | 1.512.005 |
| SUM013 | 3938 | 24.2 | 122.8 | 1.274.172 | 29.176 | 2.412.514 | 1.373.518 | 275.786 |
| SUM014 | 5334 | 24.1 | 92.7 | 49.109 | 69.272 | 2.053.274 | 806.541 | 670.257 |
| SUM015 | 5574 | 24.2 | 83.0 | 58.716 | 73.258 | 1.972.586 | 920.394 | 1.472.005 |
| SUM017 | 5575 | 24.3 | 83.1 | 68.716 | 83.258 | 1.982.586 | 930.394 | 1.482.005 |
| SUM018 | 5576 | 24.4 | 83.2 | 78.716 | 93.258 | 1.992.586 | 940.394 | 1.492.005 |
| SUM019 | 5577 | 24.5 | 83.3 | 88.716 | 103.258 | 2.002.586 | 950.394 | 1.502.005 |
| SUM020 | 5578 | 24.6 | 83.4 | 98.716 | 113.258 | 2.012.586 | 960.394 | 1.512.005 |

**TABLA 5.1.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE SUELO DEL VOLCÁN SUMACO (CENTROCESAL, 2017).

| IDENTIFICACIÓN DE LA MUESTRA | Manganeso mg/Kg | Fósforo mg/Kg | Azufre mg/Kg | Nitrógeno %P/P | Materia Orgánica %P/P | Humedad %P/P | Capacidad de intercambio cationico meq/100 g |
|---|---|---|---|---|---|---|---|
| SUM013 | 100.575 | 14.39 | 24.51 | 1.93 | 18.427 | 82.07 | 2.9 |
| SUM014 | 315.951 | 15.67 | 26.74 | 1.87 | 17.634 | 77.57 | 2.7 |
| SUM015 | 168.342 | 31.24 | 29.41 | 1.73 | 16.867 | 79.83 | 2.5 |
| SUM017 | 178.342 | 32.24 | 30.41 | 2.73 | 26.867 | 79.84 | 3.5 |
| SUM018 | 188.342 | 33.24 | 31.41 | 3.73 | 36.867 | 79.85 | 4.5 |
| SUM019 | 198.342 | 34.24 | 32.41 | 4.73 | 46.867 | 79.86 | 5.5 |
| SUM020 | 208.342 | 35.24 | 33.41 | 5.73 | 56.867 | 79.87 | 6.5 |
| SUM013 | 100.575 | 14.39 | 24.51 | 1.93 | 18.427 | 82.07 | 2.9 |
| SUM014 | 315.951 | 15.67 | 26.74 | 1.87 | 17.634 | 77.57 | 2.7 |
| SUM015 | 168.342 | 31.24 | 29.41 | 1.73 | 16.867 | 79.83 | 2.5 |
| SUM017 | 178.342 | 32.24 | 30.41 | 2.73 | 26.867 | 79.84 | 3.5 |
| SUM018 | 188.342 | 33.24 | 31.41 | 3.73 | 36.867 | 79.85 | 4.5 |
| SUM019 | 198.342 | 34.24 | 32.41 | 4.73 | 46.867 | 79.86 | 5.5 |
| SUM020 | 208.342 | 35.24 | 33.41 | 5.73 | 56.867 | 79.87 | 6.5 |

**TABLA 6.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN SUMACO (CENTROCESAL, 2019).

| IDENTIFICACIÓN DE LA MUESTRA | pH | Temperatura °C | Conductividad eléctrica (µs/cm) | Sólidos Totales Disueltos (mg/L) | Calcio mg/L | Magnesio mg/L | Sodio mg/L | Potasio mg/L |
|---|---|---|---|---|---|---|---|---|
| SUMA 030 PAVA YACU | 7,255 | 22,5 | 26,0 | 13,1 | 2,154 | 0,457 | 2,678 | 1,324 |
| SUM 031 cráter volcán | 6,114 | 22,5 | 3,5 | 2 | 0,284 | 0,087 | 0,416 | 0,241 |

**TABLA 6.** RESULTADOS DE ANÁLISIS FISCO QUÍMICOS DE MUESTRAS DE AGUA DEL VOLCÁN SUMACO (CENTROCESAL, 2019).

| IDENTIFICACIÓN DE LA MUESTRA | Hierro mg/L | Manganeso mg/L | Fosfato mg/L | Sulfatos mg/L | Cloruros mg/L | Nitrógeno mg/L | Dureza Total expresada como CaCO3 mg/L | Ca como CaCO3 mg/L |
|---|---|---|---|---|---|---|---|---|
| SUMA 030 PAVA YACU | 0,114 | <0.008 | < 0.05 | 2,231 | 3,684 | <0.50 | 7,260 | 6,688 |
| SUM 031 cráter volcán | <0.004 | <0.008 | < 0.05 | 0,351 | 0,672 | <0.50 | 1,070 | 1,039 |

**TABLA 6.1.** RESULTADOS DE ANÁLISIS FISICO-QUÍMICOS DE MUESTRAS DE SUELO DEL VOLCÁN SUMACO (CENTROCESAL, 2019).

| IDENTIFICACIÓN DE LA MUESTRA | pH | Temperatura (°C) | Conductividad eléctrica (µs/cm) | Calcio mg/Kg | Magnesio mg/Kg | Sodio mg/Kg | Potasio mg/Kg | Hierro mg/Kg |
|---|---|---|---|---|---|---|---|---|
| SUM 021 | 6,571 | 22,4 | 25,8 | 27,68 | 4,21 | 41,24 | 18,79 | 47,62 |
| SUM 022 | 6,971 | 22,4 | 31,4 | 40,32 | 12,97 | 57,14 | 28,64 | 51,78 |
| SUM 023 | 6,514 | 22,3 | 24,7 | 38,51 | 14,02 | 49,67 | 21,84 | 48,17 |
| SUM 024 | 5,987 | 22,5 | 26,1 | 32,94 | 15,31 | 45,38 | 22,17 | 41,27 |
| SUM 025 | 5,864 | 22,4 | 35,3 | 42,17 | 17,64 | 47,37 | 24,91 | 47,68 |
| SUM 026 | 6,128 | 22,1 | 41,3 | 48,31 | 20,34 | 51,03 | 31,04 | 52,41 |
| SUM 027 | 5,694 | 22,2 | 45,3 | 37,25 | 18,67 | 54,18 | 35,64 | 48,61 |
| SUM 028 | 6,259 | 22,3 | 39,4 | 29,64 | 24,53 | 47,14 | 27,16 | 53,41 |
| SUM 029 | 7,012 | 22,4 | 38,6 | 34,12 | 27,01 | 40,46 | 30,24 | 43,14 |

**TABLA 6.1.** RESULTADOS DE ANÁLISIS FISICO-QUÍMICOS DE MUESTRAS DE SUELO DEL VOLCÁN SUMACO (CENTROCESAL, 2019).

| IDENTIFICACIÓN DE LA MUESTRA | Manganeso mg/Kg | Fósforo mg/Kg | Azufre mg/Kg | Nitrógeno %P/P | Materia Orgánica %P/P | Humedad %P/P | Capacidad de intercambio cationico meq/100 g |
|---|---|---|---|---|---|---|---|
| SUM 021 | 3,27 | 25,71 | 40,37 | 0,87 | 1,847 | 48,97 | 1,90 |
| SUM 022 | 3,62 | 31,62 | 42,97 | 1,98 | 17,62 | 39,74 | 3,60 |
| SUM 023 | 4,01 | 27,14 | 38,71 | 1,86 | 18,41 | 40,15 | 2,90 |
| SUM 024 | 3,87 | 31,24 | 37,62 | 0,94 | 5,67 | 48,37 | 1,30 |
| SUM 025 | 4,31 | 35,17 | 34,81 | 0,86 | 4,97 | 44,38 | 1,20 |
| SUM 026 | 5,37 | 27,41 | 39,20 | 0,67 | 2,51 | 47,68 | 2,30 |
| SUM 027 | 4,68 | 35,48 | 43,12 | 0,74 | 2,14 | 45,37 | 2,00 |
| SUM 028 | 5,64 | 29,34 | 35,69 | 0,81 | 3,15 | 48,32 | 1,20 |
| SUM 029 | 4,37 | 31,21 | 41,13 | 0,73 | 2,08 | 46,93 | 1,70 |

# Model method of análisis

## Análisis estadísticos para las secuencias de Cayambe: año 3

## Contents



## Importación de librerías

En primer lugar, se van a cargar los diferentes datos generados con Mothur. El dataset recoge las muestras de los tres años de Cayambe. Las muestras corresponden a las región V3-V4 del 16S, ha sido clasificado con Silva empleando un cutoff del 99%.

Se ha generado el shared, el constaxonomy y el tree file.

Librerias necesarias: * La libreia principal para el manejo de los datos de 16S es phyloseq. Phyloseq permite unificar los datos de la taxonomia, la OTU table, y el árbol filogenético. Phyloseq emplea la libreria de vegan para el tratamiento de los datos.

- Se emplea la libreia metagMisc, microbiome y microbiomeutilities como complemento a phyloseq. Las tres añaden un gran numero de funciones y utilidades a phyloseq.

## Importación de datos

Tal como se ha explicado en otras ocasiones para cargar los datos de mothur, Qiime y Dada2 a phyloseq se puede hacer mediante los archivos de formar idependiente o mediante el archivo biom, que se trata de un formato standarizado. En esta caso hemos cargado el objeto a traves del .shared, el cons.taxonomy y el tree.file obtenido por Mothur.

- El *shared* contiene la OTU table
- El *taxonomy* contiene la filogenia de cada OTU hasta el nivel de género

```
#Ruta al shared
sharedfile = 'data/cay_1_2_3.shared'

#Ruta al constaxonomy
taxfile = "data/cay_1_2_3.taxonomy"

# Import mothur data
mothur_data = import_mothur(mothur_shared_file = sharedfile,
                            mothur_constaxonomy_file = taxfile)

#Adición de las taxonomias
colnames(tax_table(mothur_data)) = c("Kingdom", "Phylum", "Class", "Order", "Family", "Genus")
meta_tmp <- data.frame(Sample = sample_names(mothur_data))
rownames(meta_tmp) <- meta_tmp$Sample
sample_data(mothur_data) <- sample_data(meta_tmp)
```

Ahora podemos unir la metadata al objeto phyloseq.

```
meta_y1 <- read_xlsx("metadata/metadata_shared_all.xlsx", sheet = 1)
meta_y2 <- read_xlsx("metadata/metadata_shared_all.xlsx", sheet = 2)
meta_y3 <- read_xlsx("metadata/metadata_shared_all.xlsx", sheet = 3)
meta_shared_all <- bind_rows(meta_y1, meta_y2, meta_y3) %>% as.data.frame()
meta_shared_all <- meta_shared_all %>%
  mutate(Altitude = ifelse(Altura > 5200, "High", "Low")) %>%
  filter( !(SeqID %in% c("C42a", "C43b", "C45b", "C46a")) )

meta_shared_all <- meta_shared_all %>%
  mutate(Years = ifelse(Tanda == 1, "Year 1",
                        ifelse(Tanda == 2, "Year 2", "Year 3")))

rownames(meta_shared_all) <- meta_shared_all$SeqID
```

```
mothur_data <- subset_samples(mothur_data,
                              Sample %in% meta_shared_all$SeqID)
sample_data(mothur_data) <- sample_data(meta_shared_all)
```

Encontramos que el objeto phyloseq crudo, es decir, que no se ha aplicado ningun filtro, consta de 311 459 OTUs. En la siguiente seccion se empleara varios filtros para disminuir el ruido que pudieron haber causado factores como errores de secuenciacion y OTU picking, entre otros.

### Importacion Metadata Agua

```
meta_water_y1 <- readxl::read_xlsx("metadata/fisicoquimico_year1.xlsx",
                                   sheet = "Water")
meta_water_y2 <- readxl::read_xlsx("metadata/fisicoquimico_year2.xlsx",
                                   sheet = "Water")
meta_water_y3 <- readxl::read_xlsx("metadata/fisicoquimico_year3.xlsx",
                                   sheet = "Water")
meta_water <- bind_rows(meta_water_y1, meta_water_y2, meta_water_y3) %>%
  dplyr::select(-Mn, -PO4, -N)
```

### Importacion Metadata Suelo

```
meta_soil_y1 <- readxl::read_xlsx("metadata/fisicoquimico_year1.xlsx",
                                   sheet = "Soil")
meta_soil_y2 <- readxl::read_xlsx("metadata/fisicoquimico_year2.xlsx",
                                   sheet = "Soil")
meta_soil <- bind_rows(meta_soil_y1, meta_soil_y2)
```

## Preprocesamiento

### Remocion de singletons

Si miramos el contenido de esta tabla veremos que hay muchas OTUs que presentan una única secuencia y que solo se encuentran en una única muestra (singletons).Existe un sesgo muy fuerte de la aparición de secuencias erróneas reportadas por las tecnologías de secuenciación NGS de amplicones para secuencias con abundancias bajas, secuencias singleton y OTUs singleton en particular.

```
mothur_data <- prune_taxa(taxa_sums(mothur_data) > 1, mothur_data)
mothur_data
```

Despues de aplicar este filtro, notamos que hay un drastico decrecimiento del numero de OTUs, en especial por que la mayoria de ellos constaba solo de un conteo, llevandonos a un numero de 36 199 OTUs.

### Filtrado de Unknowns

Ahora vamos a ver si hay algún taxón desconocido(Unknowns). Es posible que haya secuencias que no han podido ser clasificadas por nuestra base de datos. Por lo general estas OTUs se deberían eliminar. Podríamos comprobar la secuencia consenso y hacer un BLAST(de hecho, por eso no las eliminamos en mothur, aunque podríamos haberlo hecho) ya que podría tratarse de alguna secuencia nueva que no se encontrara en nuestra base de datos. Pero en la mayoría de los casos no serán nada.

```
table(tax_table(mothur_data)[,1])
```

Como podemos ver, tenemos 9 OTUs que a nivel de reino estan clasificados como desconocidos, por lo que se procede a removerlos

```
mothur_data = subset_taxa(mothur_data , Kingdom != "unknown")
```

### Filtrado de High Unclassifieds

También podemos encontrarnos OTUs con taxones no clasificados a niveles taxonómicos superiores como Filo. Vamos a comprobarlo:

```
table(tax_table(mothur_data)[,2])
```

Vemos 493 OTUs de Bacterias sin clasificación taxonómica("Bacteria_unclassified"). Vamos a eliminarlos porque no nos van a aportar mucho.

Y eliminamos las OTUs Sin clasificar a nivel de Filo:

```
mothur_data <-  subset_taxa(mothur_data,
                            Phylum != "Bacteria_unclassified")
mothur_data
```

Despues de eliminar aquellos OTUs que no estan clasificados a nivel de filo, se procede a observar que teneos un total de 35697 OTUs de alta calidad, con los que podemos trabajar en los analisis subsecuentes.

## Agrupar los OTUs por genero

El paquete phyloseq incluye las funciones de agrupación, tip_glom y tax_glom, para fusionar todas las OTU en un experimento que son similares más allá de un umbral filogenético o taxonómico, respectivamente. Por lo tanto para Agrupar OTUs por genero emplearemos la función tax_glom. Este método combina especies que tienen la misma taxonomía en un cierto rango taxonómico. Su enfoque es análogo a tip_glom, pero utiliza datos categóricos en lugar de un árbol. En principio, otros datos categóricos conocidos para todos los taxones también podrían usarse en lugar de la taxonomía. Además, los rangos a la derecha del rango elegido para la aglomeración se reemplazarán con NA, ya que no deberían tener sentido después de la aglomeración.

```
mothur_data_agglomerated = tax_glom(mothur_data, "Genus")
```

Despues de realizar esta aglomeracion, el numero final de Generos a los que trabajaremos seran 1 375.

```
mothur_data_agglomerated
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 1375 taxa and 51 samples ]
## sample_data() Sample Data:       [ 51 samples by 9 sample variables ]
## tax_table()   Taxonomy Table:    [ 1375 taxa by 6 taxonomic ranks ]
```

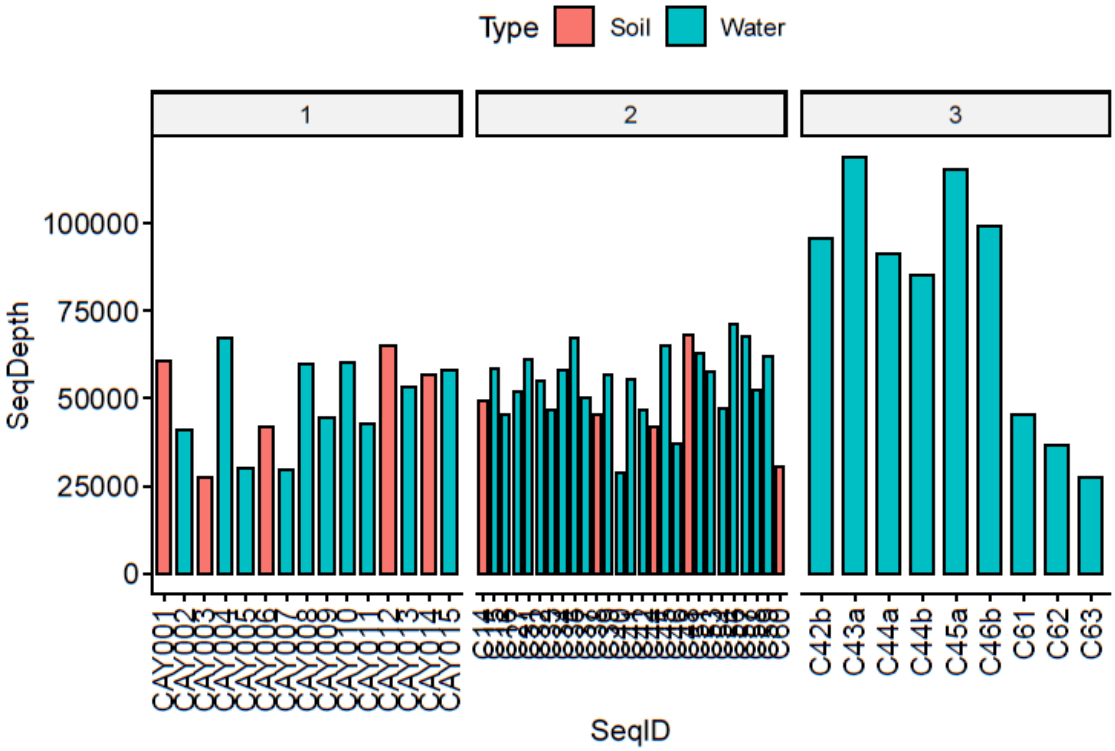
## Obtener el número de reads.

El siguiente paso es comprobar la profundidad de secuenciación de la muestra. Es decir el número de Reads presente en cada una de las muestras. Puede pasar que haya grandes diferencias entre el número de reads de cada muestra. Si esta diferencia es muy grande puede haber sido causada por errores de secuenciación y deberíamos eliminarla.

Para ello vamos a representar el número de reads de cada muestra. La librería ggpubr, permite generar representaciones de ggplot de forma más sencilla.

Como cada OTU esta representada por el número de reads que coinciden en su secuencia, se obtiene la profundida de secuencia sumando la OTU table para cada muestra.

```
#Obtenemos la profundida de secuenciación de cada muestras
SeqDepth = colSums(otu_table(mothur_data_agglomerated))

#Añadimos el SeqDepth a la metadata
sample_data(mothur_data_agglomerated)$SeqDepth = SeqDepth

ggpubr::ggbarplot(meta(mothur_data_agglomerated), "SeqID", "SeqDepth",
                  fill = "Type") +  rotate_x_text() +
  facet_wrap("Tanda", scales = "free_x")
```




Estadistica el numero de reads

```
meta(mothur_data_agglomerated) %>% group_by(Type) %>%
  summarise(mean = mean(SeqDepth), sd = sd(SeqDepth), Total = sum(SeqDepth) )
```

```
## # A tibble: 2 x 4
##   Type    mean     sd   Total
##   <chr>  <dbl>  <dbl>   <dbl>
## 1 Soil  48668  13956.  486680
## 2 Water 58720. 21402. 2407511
```

## Soil Samples

Primero, seleccionaremos solo las muestras que correspondan a la matriz de suelo. Dado que los singletones no pudieron haber sido eliminados eficientemente, procedemos una vez mas a eliminarlos. La diferencia con el procedimiento anterior es que aquellos singletones que se encuentran en el las muestras de suelo, dejan de ser singletones una vez estos OTUs sean evaluados con las muestras de agua, por lo que dejarian de ser singletones.

```
mothur_suelo <- subset_samples(mothur_data_agglomerated, Type == "Soil")
mothur_suelo <- prune_taxa(taxa_sums(mothur_suelo) > 1, mothur_suelo)
mothur_suelo
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 727 taxa and 10 samples ]
## sample_data() Sample Data:       [ 10 samples by 10 sample variables ]
## tax_table()   Taxonomy Table:    [ 727 taxa by 6 taxonomic ranks ]
```

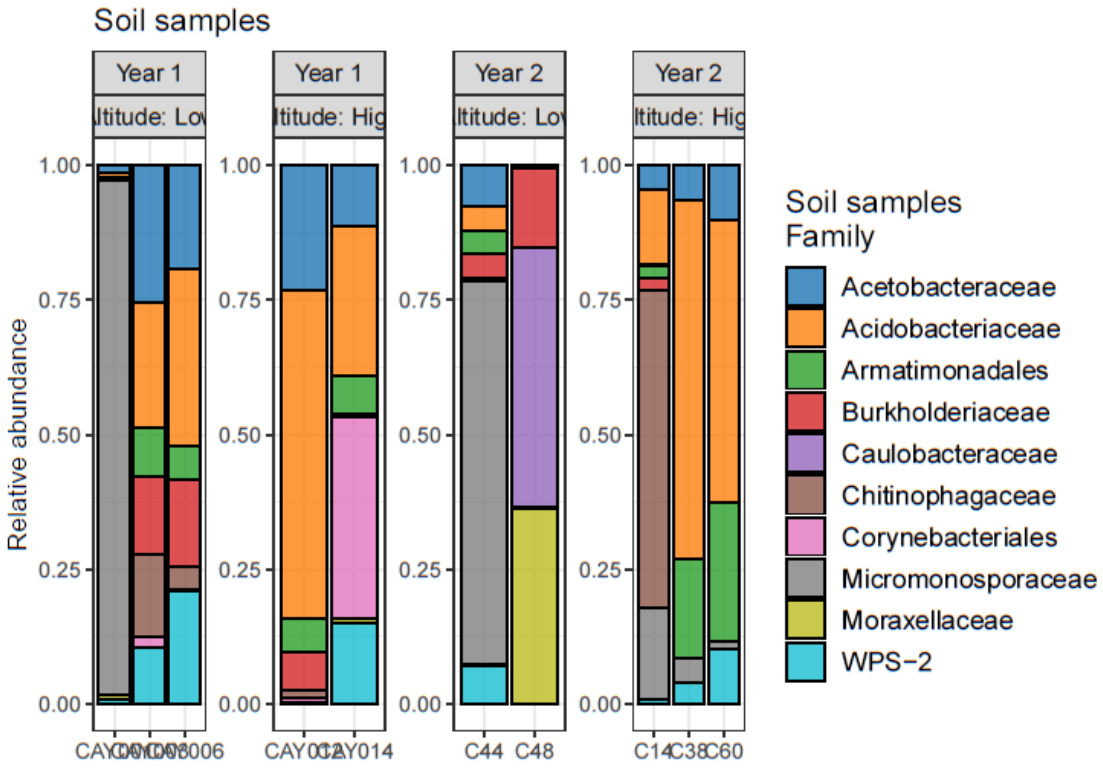
Como se puede observar, del inicio que teniamos 1375, despues de eliminar singletones especificos para suelo, tenemos un total de 727 OTUs.

## Alpha Diversity

### Composición de las muestras

Para tener una imagen de las familias más abundantes de cada muestra, se procederá a realizar una figura de barras con la abundancia relativa de las muestras.

```
plt_sl_rel <- relative_abun(mothur_suelo, n_top = 10) +
  ggtitle("Soil samples") + xlab("") +
  facet_wrap(Tanda~Altitude, scales = "free", nrow = 1,
             labeller = labeller(Tanda = c(`1` = "Year 1", `2` = "Year 2"),
                                 Altitude = c("Low" = "Altitude: Low",
                                              "High" = "Altitude: High") ) ) +
  theme_bw()+ labs(fill = "Soil samples\nFamily") +
   theme(axis.title.x=element_blank(),
         #axis.text.x=element_blank(),
         #axis.ticks.x=element_blank(),
         strip.text = element_text(size = 10),
         legend.title=element_text(size=13),
    legend.text=element_text(size=11))
plt_sl_rel
```

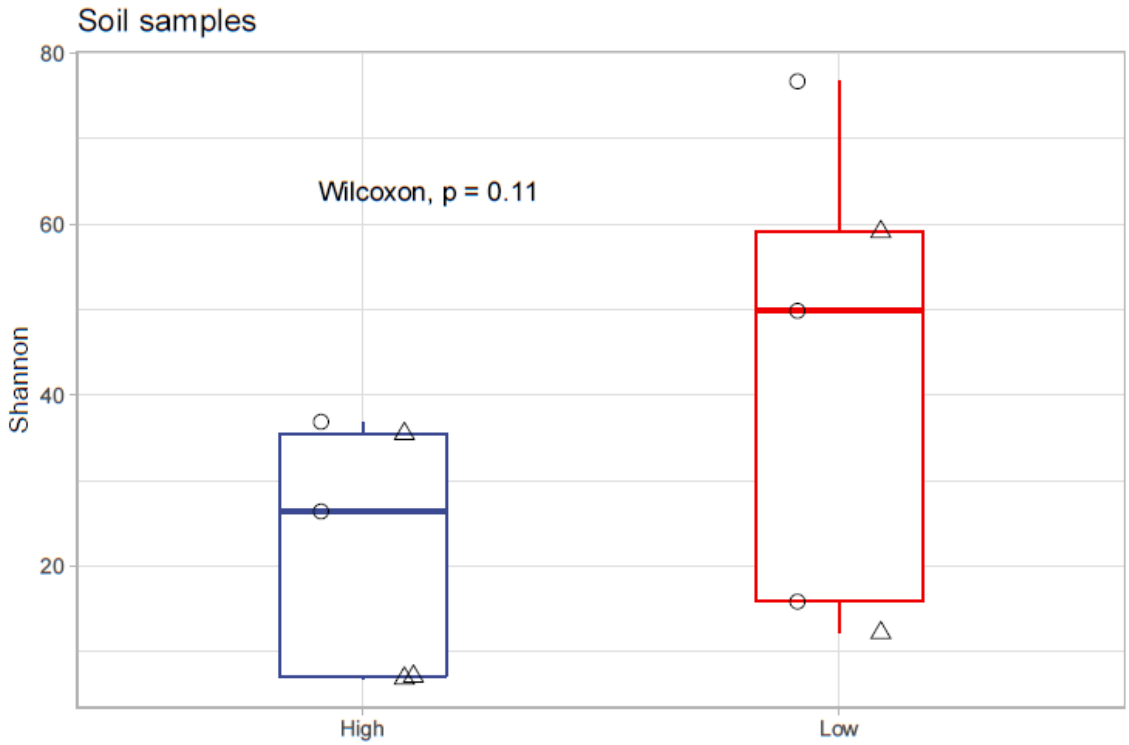


Medidas de alpha diversity

```
alpha_boxplot <- function(mothur, index, y_position = 1) {
  diversity_data <- estimate_richness(mothur,
                                      measures = index ) %>%
    mutate(SeqID = rownames(.)) %>% merge(microbiome::meta(mothur))

  diversity_data <- switch (index,
                            Shannon = mutate(diversity_data,
                                             Shannon = exp(Shannon)),
                            Simpson = diversity_data)
  diversity_data <- diversity_data %>%
    arrange(Altitude, Tanda) %>%
    mutate(SeqID = factor(SeqID, levels = SeqID)) %>%
    mutate(Tanda = factor(Tanda))

  ggplot(diversity_data, aes_string(as.name("Altitude"), index)) +
  geom_boxplot(width = 0.35, aes(color = Altitude)) +
  ggpubr::stat_compare_means(label.y.npc = y_position,
                             method.args = list(alternative = "greater"))+
  geom_beeswarm(size = 2.5, aes(shape = Tanda), dodge.width = 0.35)+
  theme_light() + labs(shape = "Sampling\nYear") +
  scale_shape_manual(values= seq(1, length(unique(diversity_data$Tanda) ) ) ) +
  scale_color_aaas()
}
```

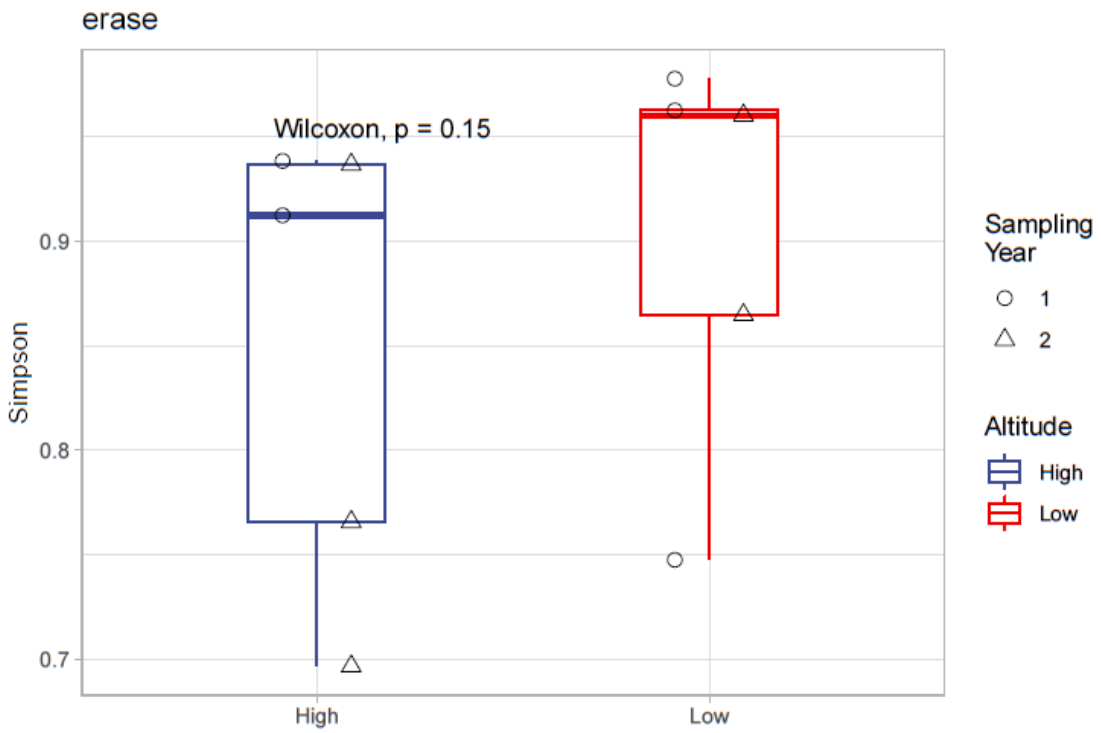
Shannon index

```
soil_alpha_shannon <- alpha_boxplot(mothur_suelo, "Shannon", y_position = 0.8) +
  guides(color = F, shape = F) + xlab("") + labs(title ="Soil samples")
soil_alpha_shannon
```




Simpson index

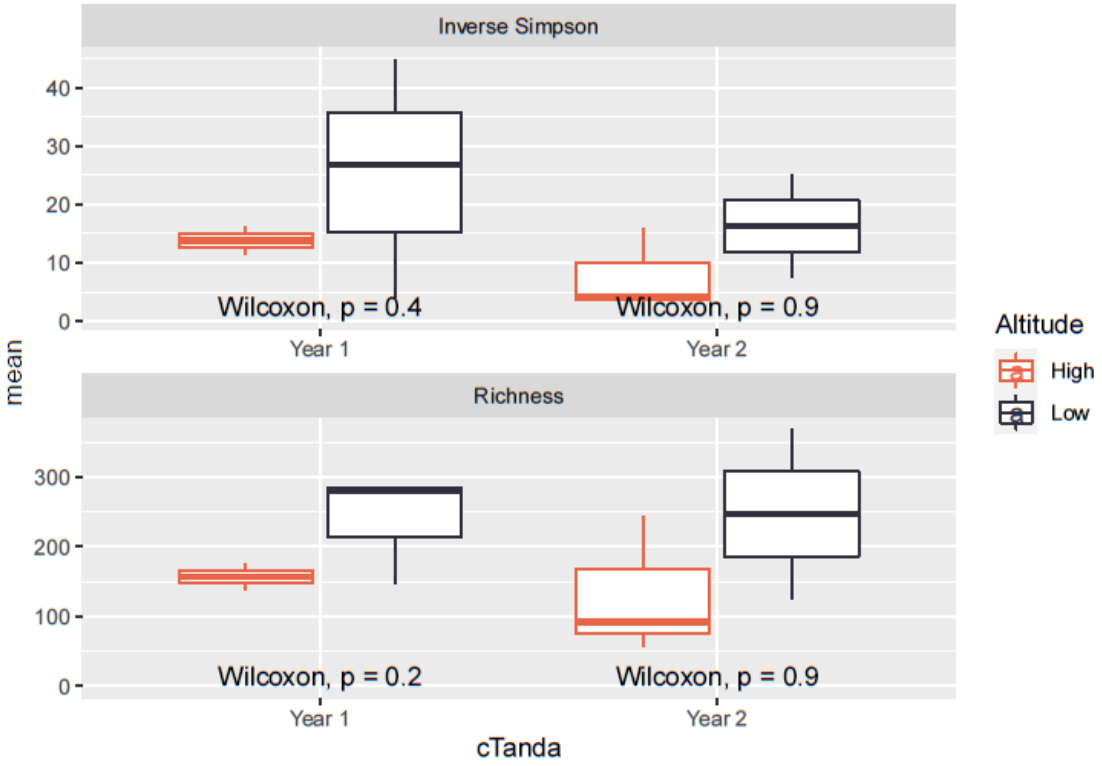
```
soil_alpha_Simpson <- alpha_boxplot(mothur_suelo, "Simpson", y_position = 0.9) +
  xlab("") +  labs(title = "erase")
soil_alpha_Simpson
```




Alpha diversity in time function

```r
min_lib <- min(sample_sums(mothur_suelo))

# Initialize matrices to store richness and evenness estimates
nsamp = nsamples(mothur_suelo)
trials = 100

richness <- matrix(nrow = nsamp, ncol = trials)
row.names(richness) <- sample_names(mothur_suelo)

evenness <- matrix(nrow = nsamp, ncol = trials)
row.names(evenness) <- sample_names(mothur_suelo)

# It is always important to set a seed when you subsample so your result is replicable
set.seed(3)

for (i in 1:100) {
  # Subsample
  r <- rarefy_even_depth(mothur_suelo, sample.size = min_lib, verbose = FALSE, replace = TRUE)

  # Calculate richness
  rich <- as.numeric(as.matrix(estimate_richness(r, measures = "Observed")))
  richness[ ,i] <- rich

  # Calculate evenness
  even <- as.numeric(as.matrix(estimate_richness(r, measures = "InvSimpson")))
  evenness[ ,i] <- even
}
```

```r
# Create a new dataframe to hold the means and standard deviations of richness estimates
SeqID <- row.names(richness)
mean <- apply(richness, 1, mean)
sd <- apply(richness, 1, sd)
measure <- rep("Richness", nsamp)
rich_stats <- data.frame(SeqID, mean, sd, measure)

# Create a new dataframe to hold the means and standard deviations of evenness estimates
SeqID <- row.names(evenness)
mean <- apply(evenness, 1, mean)
sd <- apply(evenness, 1, sd)
measure <- rep("Inverse Simpson", nsamp)

even_stats <- data.frame(SeqID, mean, sd, measure)

#Now we will combine our estimates for richness and evenness into one dataframe

alpha <- rbind(rich_stats, even_stats)

#Let's add the sample metadata into this dataframe using the merge() command

s <- data.frame(sample_data(mothur_suelo))

alphadiv <- merge(alpha, s, by = "SeqID")

#Lastly, we will reorder some factors in this dataset before plotting them

#alphadiv <- order_dates(alphadiv)

alphadiv <- alphadiv %>%
  mutate(cTanda = ifelse(Tanda == 1, "Year 1",
                         ifelse(Tanda == 2, "Year 2", "Year 3")))
#Finally, we will plot the two alpha diversity measures in a timeseries using a facet

alphadiv$cTanda <- factor(alphadiv$cTanda,
                          levels = c("Year 1", "Year 2", "Year 3"))

ggplot(alphadiv, aes(x = cTanda, y = mean, color = Altitude)) +
  geom_boxplot() +
  facet_wrap(~measure, ncol = 1, scales = "free") +
  scale_color_manual(values = c("#E96446", "#302F3D", "#87CEFA")) +
  ggpubr::stat_compare_means(method = "wilcox",label.y = 0.9,
                             method.args = list(alternative = "less"))
```

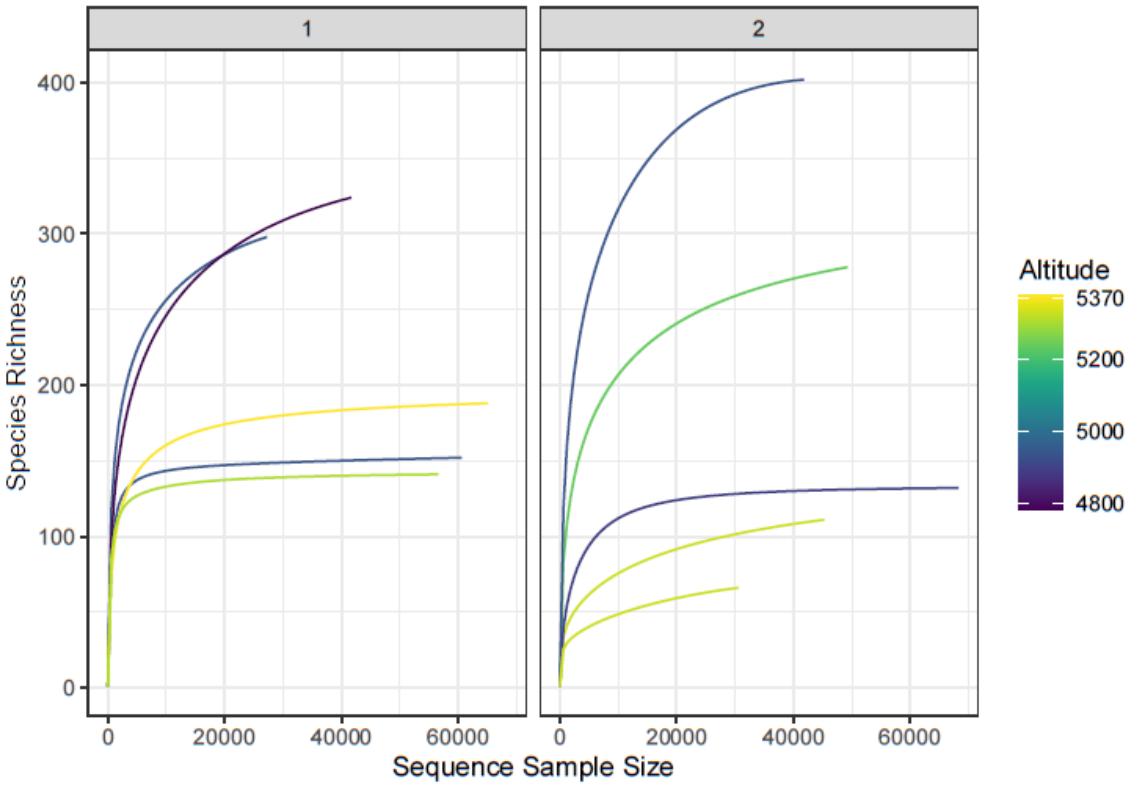



Rarefraction curves

```
rarefraction_soil <- ranacapa::ggrare(mothur_suelo, step = 600,
                                      color = 'Altura', se = F, plot = F) +
    scale_color_viridis(breaks = c(4800, 5000, 5200, 5370))  +
  labs(color = "Altitude") + theme_bw() + facet_wrap("Tanda")
```

```
## rarefying sample C14
## rarefying sample C38
## rarefying sample C44
## rarefying sample C48
## rarefying sample C60
## rarefying sample CAY001
## rarefying sample CAY003
## rarefying sample CAY006
## rarefying sample CAY012
## rarefying sample CAY014
```

```
rarefraction_soil
```




## Water Samples

```
mothur_water <- subset_samples(mothur_data_agglomerated, Type == "Water")
mothur_water <- prune_taxa(taxa_sums(mothur_water) > 1, mothur_water)
mothur_water
```

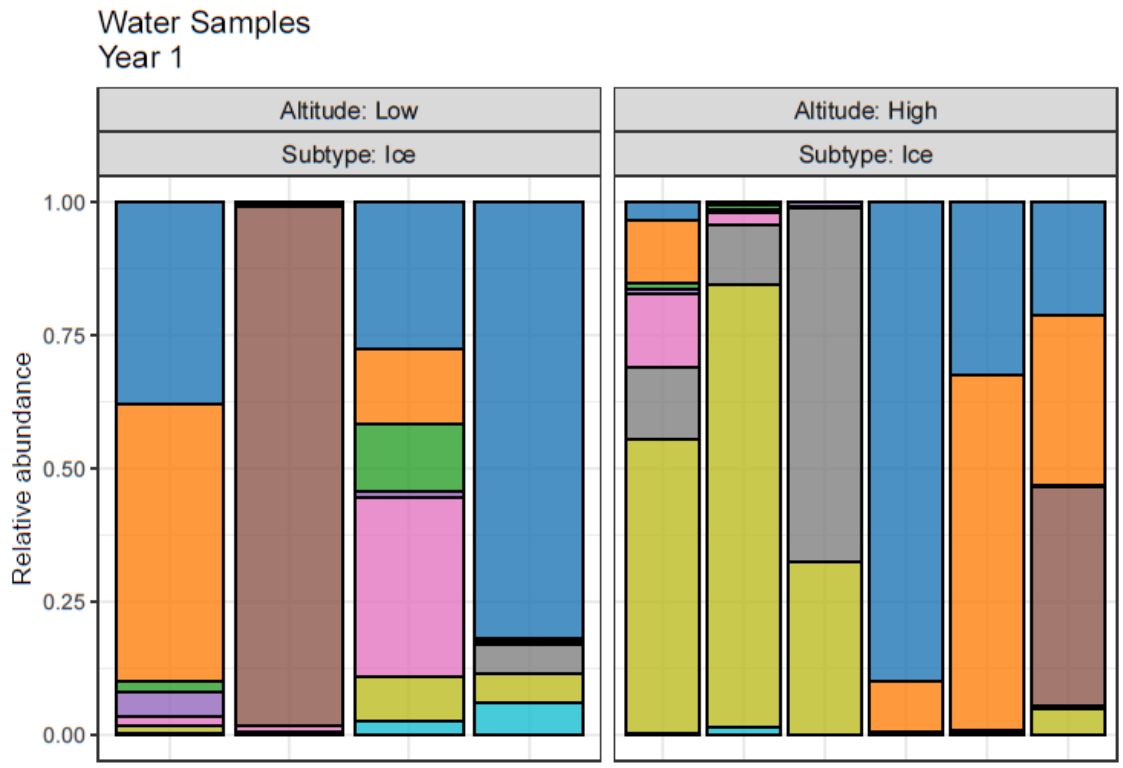
```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 1286 taxa and 41 samples ]
## sample_data() Sample Data:       [ 41 samples by 10 sample variables ]
## tax_table()   Taxonomy Table:    [ 1286 taxa by 6 taxonomic ranks ]
```

```
plt_wt_rel_data <- relative_abun(mothur_water, n_top = 10)
```

### Año 1

```
plt_rel_year1 <-  plt_wt_rel_data$data %>%
  filter(Tanda == "1") %>% mutate(Subtype = "Ice") %>%
  ggplot( aes(x = Sample, y = Abundance, fill = Family)) +
     geom_bar(stat = "identity", position = "fill",
              color = "black", width = 0.9)  +
  facet_wrap(Altitude~Subtype, scales = "free_x",
             labeller =  labeller(Altitude = c("Low" = "Altitude: Low",
                                               "High" = "Altitude: High"),
                                  Subtype = c("Ice" = "Subtype: Ice"))) +
  labs(y = "Relative abundance") +
  scale_fill_d3(alpha = 0.8) +
  theme_bw() +
```

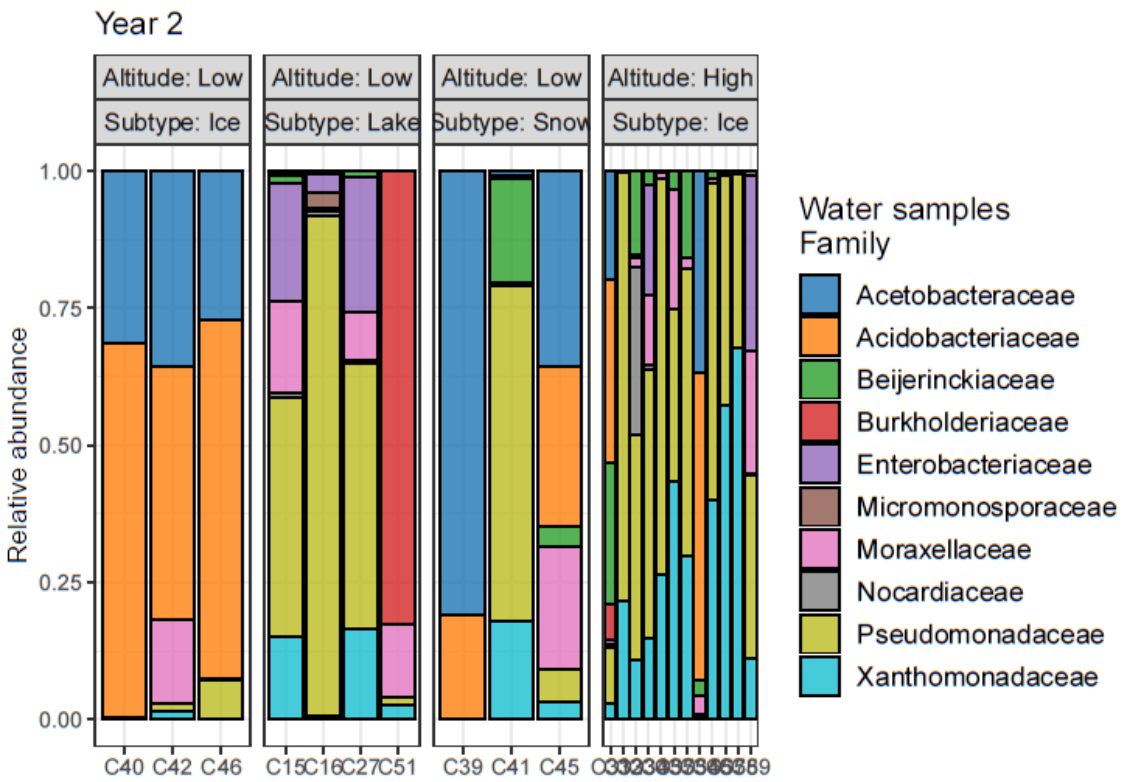
```
   theme(axis.title.x=element_blank(),
         axis.text.x=element_blank(),
         axis.ticks.x=element_blank(),
         strip.text = element_text(size = 10)) +
  labs(title = "Water Samples\nYear 1") +
  guides(fill = F)
plt_rel_year1
```




**Año 2**

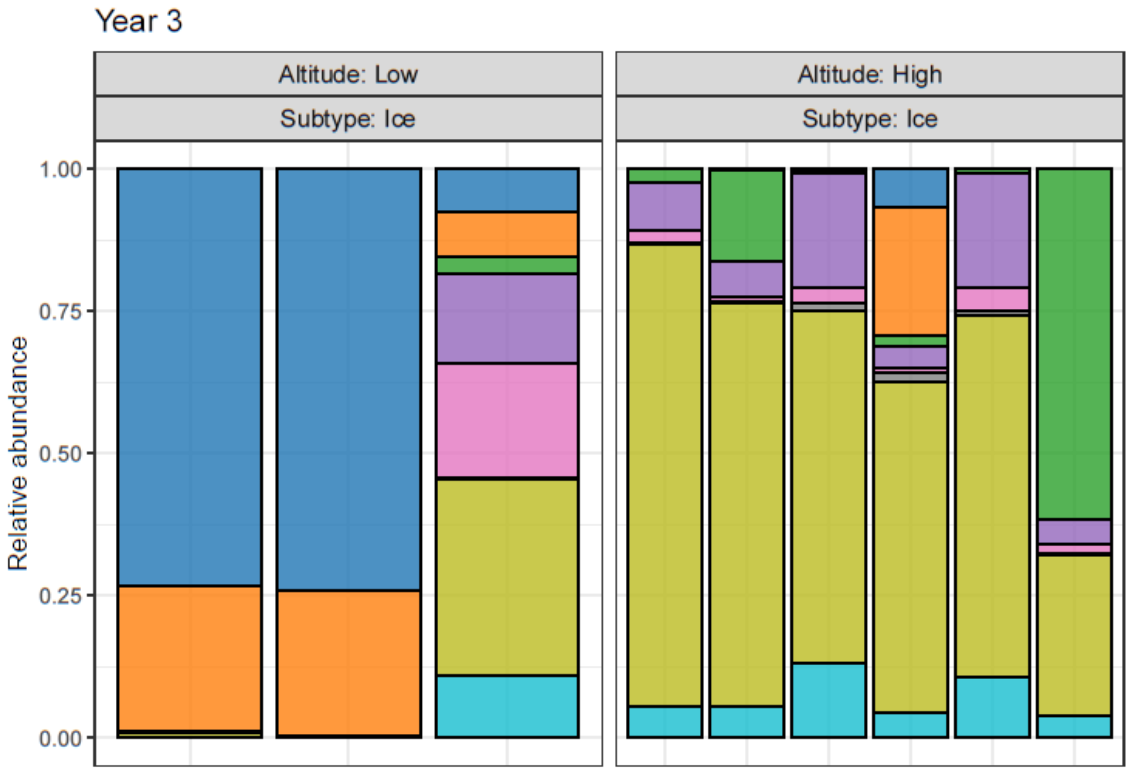
```
plt_rel_year2 <-  plt_wt_rel_data$data %>%
  filter(Tanda == "2") %>%
  ggplot( aes(x = Sample, y = Abundance, fill = Family)) +
     geom_bar(stat = "identity", position = "fill",
              color = "black", width = 0.9)  +
  facet_wrap(Altitude~Subtype, scales = "free_x",nrow = 1,
             labeller =  labeller(Altitude = c("Low" = "Altitude: Low",
                                               "High" = "Altitude: High"),
                                  Subtype = c("Ice" = "Subtype: Ice",
                                              "Snow" = "Subtype: Snow",
                                              "Lake" = "Subtype: Lake") )) +
  labs(y = "Relative abundance", fill = "Water samples\nFamily") +
  scale_fill_d3(alpha = 0.8) +
  theme_bw()+
   theme(axis.title.x=element_blank(),
         #axis.text.x=element_blank(),
         #axis.ticks.x=element_blank(),
```

```
        strip.text = element_text(size = 10),
        legend.title=element_text(size=13),
    legend.text=element_text(size=11)) +
  labs(title = "Year 2")
plt_rel_year2
```




**Año 3**

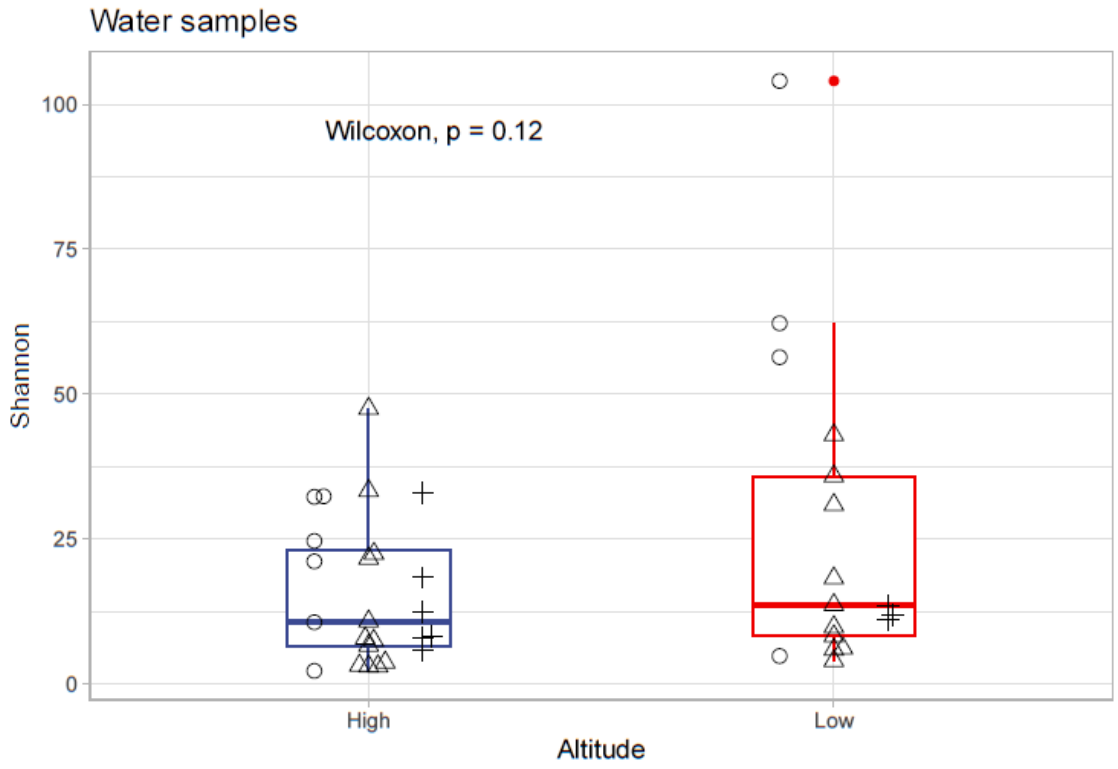
```
plt_rel_year3 <-  plt_wt_rel_data$data %>%
  filter(Tanda == "3") %>% mutate(Subtype = "Ice") %>%
  ggplot( aes(x = Sample, y = Abundance, fill = Family)) +
     geom_bar(stat = "identity", position = "fill",
              color = "black", width = 0.9)  +
  facet_wrap(Altitude~Subtype, scales = "free_x",
             labeller =  labeller(Altitude = c("Low" = "Altitude: Low",
                                               "High" = "Altitude: High"),
                                  Subtype = c("Ice" = "Subtype: Ice") )) +
  labs(y = "Relative abundance") +
  scale_fill_d3(alpha = 0.8) +
  theme_bw() +
   theme(axis.title.x=element_blank(),
         axis.text.x=element_blank(),
         axis.ticks.x=element_blank(),
         strip.text = element_text(size = 10)) +
  labs(title = "Year 3") +
  guides(fill = F)
plt_rel_year3
```




### Medidas de alpha diversity

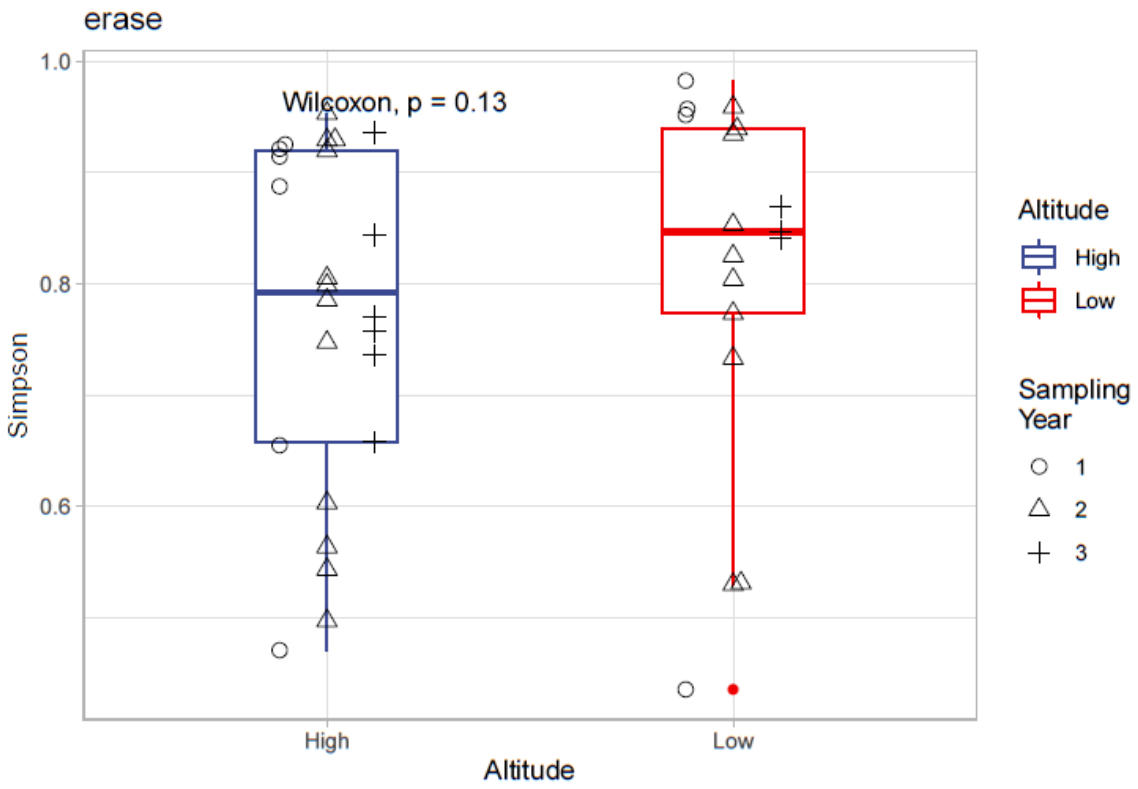
#### Shannon index

```
water_alpha_shannon <- alpha_boxplot(mothur_water, "Shannon", y_position = 0.9) +
  guides(color = F, shape = F) + labs(title = "Water samples")
water_alpha_shannon
```



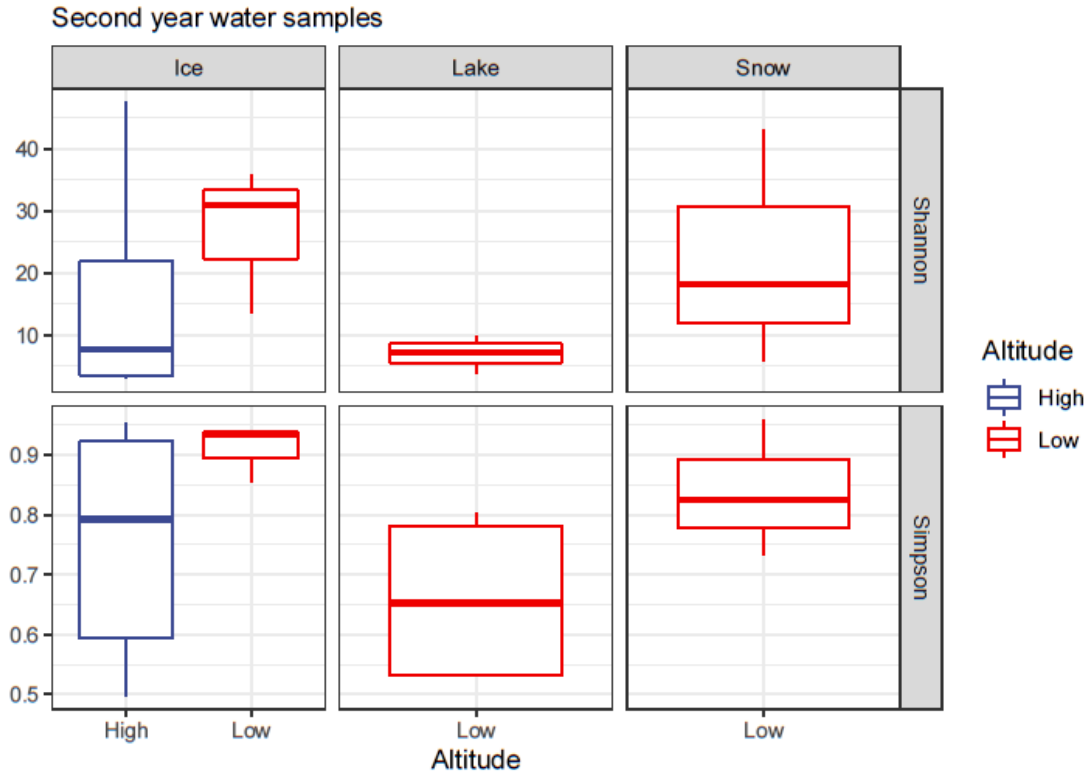

Simpson index

```
water_alpha_Simpson <- alpha_boxplot(mothur_water, "Simpson", y_position = 0.95) +
  labs(title = "erase")
water_alpha_Simpson
```




Separacion Año 2

```
year2_alpha <- estimate_richness(mothur_water,
                                 measures = c("Shannon", "Simpson") ) %>%
  mutate(Shannon = exp(Shannon), SeqID = rownames(.) ) %>%
  left_join(microbiome::meta(mothur_water) ) %>% filter(Tanda == "2") %>%
  gather(key = "Index", value = "Value", Shannon:Simpson)

year2_alpha <- ggplot(year2_alpha, aes(Altitude, Value)) +
  geom_boxplot(aes(color = Altitude)) +
  facet_grid(Index ~ Subtype, scales = "free") +
  labs(y = "") +
  theme_bw() +
  scale_color_aaas() + labs(subtitle = "Second year water samples")
year2_alpha
```



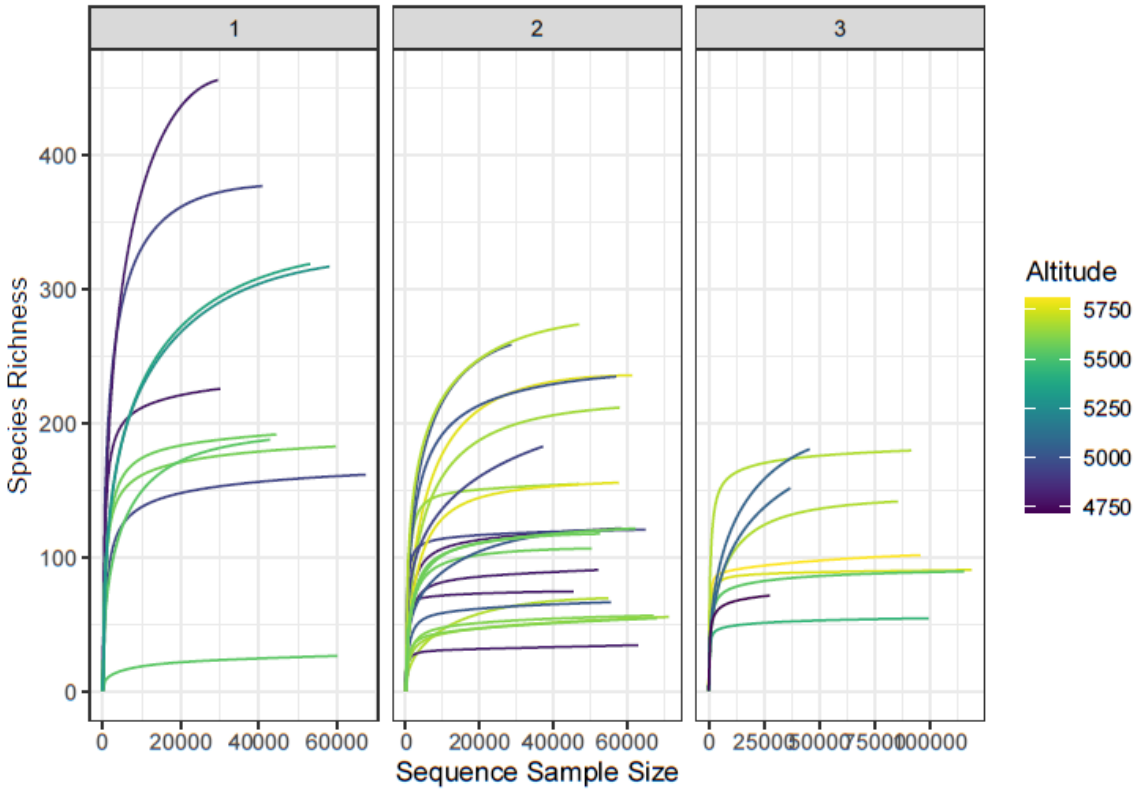

Curvas de rarefraccion

```
rarefraction_water <- ranacapa::ggrare(mothur_water, step = 600,
                                       color = 'Altura', se = F, plot = F) +
    scale_color_viridis()+ facet_wrap("Tanda", scales = "free_x") +
  labs(color = "Altitude") +
  theme_bw()
```

```
## rarefying sample C15
## rarefying sample C16
## rarefying sample C27
## rarefying sample C31
## rarefying sample C32
## rarefying sample C33
## rarefying sample C34
## rarefying sample C35
## rarefying sample C37
## rarefying sample C39
## rarefying sample C40
## rarefying sample C41
## rarefying sample C42
## rarefying sample C42b
## rarefying sample C43a
## rarefying sample C44a
## rarefying sample C44b
## rarefying sample C45
## rarefying sample C45a
```

```
## rarefying sample C46
## rarefying sample C46b
## rarefying sample C51
## rarefying sample C53
## rarefying sample C54
## rarefying sample C56
## rarefying sample C57
## rarefying sample C58
## rarefying sample C59
## rarefying sample C61
## rarefying sample C62
## rarefying sample C63
## rarefying sample CAY002
## rarefying sample CAY004
## rarefying sample CAY005
## rarefying sample CAY007
## rarefying sample CAY008
## rarefying sample CAY009
## rarefying sample CAY010
## rarefying sample CAY011
## rarefying sample CAY013
## rarefying sample CAY015
```

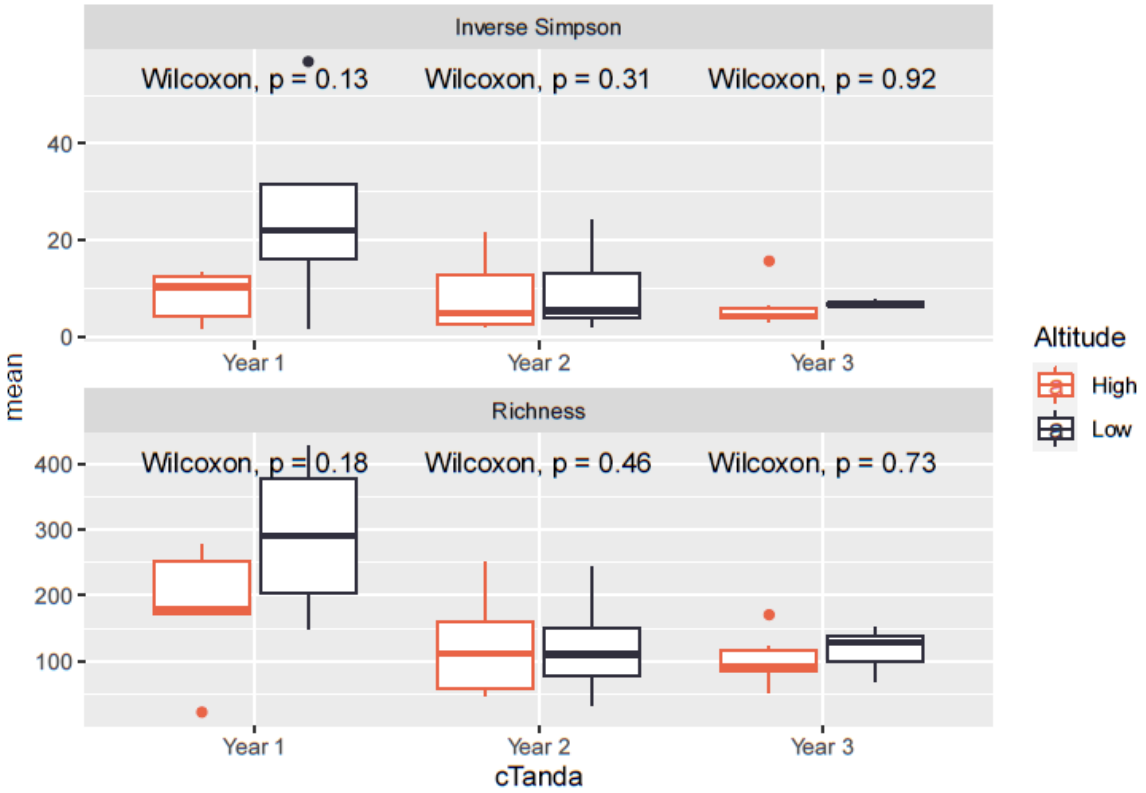
```
rarefraction_water
```




Alpha diversity in time function

```
min_lib <- min(sample_sums(mothur_water))

# Initialize matrices to store richness and evenness estimates
nsamp = nsamples(mothur_water)
trials = 100

richness <- matrix(nrow = nsamp, ncol = trials)
row.names(richness) <- sample_names(mothur_water)

evenness <- matrix(nrow = nsamp, ncol = trials)
row.names(evenness) <- sample_names(mothur_water)

# It is always important to set a seed when you subsample so your result is replicable
set.seed(3)

for (i in 1:100) {
  # Subsample
  r <- rarefy_even_depth(mothur_water, sample.size = min_lib, verbose = FALSE, replace = TRUE)

  # Calculate richness
  rich <- as.numeric(as.matrix(estimate_richness(r, measures = "Observed")))
  richness[ ,i] <- rich

  # Calculate evenness
  even <- as.numeric(as.matrix(estimate_richness(r, measures = "InvSimpson")))
  evenness[ ,i] <- even
}

# Create a new dataframe to hold the means and standard deviations of richness estimates
SeqID <- row.names(richness)
mean <- apply(richness, 1, mean)
sd <- apply(richness, 1, sd)
measure <- rep("Richness", nsamp)
rich_stats <- data.frame(SeqID, mean, sd, measure)

# Create a new dataframe to hold the means and standard deviations of evenness estimates
SeqID <- row.names(evenness)
mean <- apply(evenness, 1, mean)
sd <- apply(evenness, 1, sd)
measure <- rep("Inverse Simpson", nsamp)

even_stats <- data.frame(SeqID, mean, sd, measure)

#Now we will combine our estimates for richness and evenness into one dataframe

alpha <- rbind(rich_stats, even_stats)

#Let's add the sample metadata into this dataframe using the merge() command

s <- data.frame(sample_data(mothur_water))

alphadiv <- merge(alpha, s, by = "SeqID")
```

```r
#Lastly, we will reorder some factors in this dataset before plotting them

#alphadiv <- order_dates(alphadiv)

alphadiv <- alphadiv %>%
  mutate(cTanda = ifelse(Tanda == 1, "Year 1",
                         ifelse(Tanda == 2, "Year 2", "Year 3")))
#Finally, we will plot the two alpha diversity measures in a timeseries using a facet

alphadiv$cTanda <- factor(alphadiv$cTanda,
                          levels = c("Year 1", "Year 2", "Year 3"))

ggplot(alphadiv, aes(x = cTanda, y = mean, color = Altitude)) +
  geom_boxplot() +
  facet_wrap(~measure, ncol = 1, scales = "free") +
  scale_color_manual(values = c("#E96446", "#302F3D", "#87CEFA")) +
  stat_compare_means(label.y.npc = 0.9, method = "wilcox",
                     method.args = list(alternative = "less"))
```




### Medidas de Beta Diversidad

Funcion para normalizar con Deseq2

```r
normalDeseq2 = function(phyloseq){
  require(DESeq2)

  diagdds = phyloseq_to_deseq2(phyloseq, ~ Altitude + Tanda)
  gm_mean = function(x, na.rm=TRUE){
```

```
    exp(sum(log(x[x > 0]), na.rm=na.rm) / length(x))
  }
  geoMeans = apply(counts(diagdds), 1, gm_mean)

  #Estimamos la varianza
  diagdds = estimateSizeFactors(diagdds, geoMeans = geoMeans)
  diagdds = estimateDispersions(diagdds)

  #Extraemos la matriz normalizada
  diagvst = getVarianceStabilizedData(diagdds)
  dim(diagvst)

  diagvst[diagvst<0]<-0

  otu_table(phyloseq) <- otu_table(diagvst, taxa_are_rows = TRUE)

  return(phyloseq)

}
```

### Beta Diversidad

#### Suelo *Remoción de OTUs pobremente representados*

Se ha elegido que OTUs que no son detectados más de cinco veces en más de la mitad de las muestras sean removidos. Conservando de esta manera 28 y 58 OTUs para muestras de agua y suelo, respectivamente.

```
##  High samples
# Create the criteria to filter
criteria_5counts_wt = genefilter_sample(mothur_suelo,
                                        filterfun_sample(function(x) x > 5),
                                        A=0.5*nsamples(mothur_suelo))
# Filter samples
mothur_suelo_filt = prune_taxa(criteria_5counts_wt, mothur_suelo)
mothur_suelo
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 727 taxa and 10 samples ]
## sample_data() Sample Data:       [ 10 samples by 10 sample variables ]
## tax_table()   Taxonomy Table:    [ 727 taxa by 6 taxonomic ranks ]
```

```
mothur_suelo_filt
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 91 taxa and 10 samples ]
## sample_data() Sample Data:       [ 10 samples by 10 sample variables ]
## tax_table()   Taxonomy Table:    [ 91 taxa by 6 taxonomic ranks ]
```

*Normalizacion por Deseq2*

```
mothur_suelo_filt_desq = normalDeseq2(mothur_suelo_filt)
```

```
## converting counts to integer mode
```

```
## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors
```

```
##   the design formula contains one or more numeric variables with integer values,
```

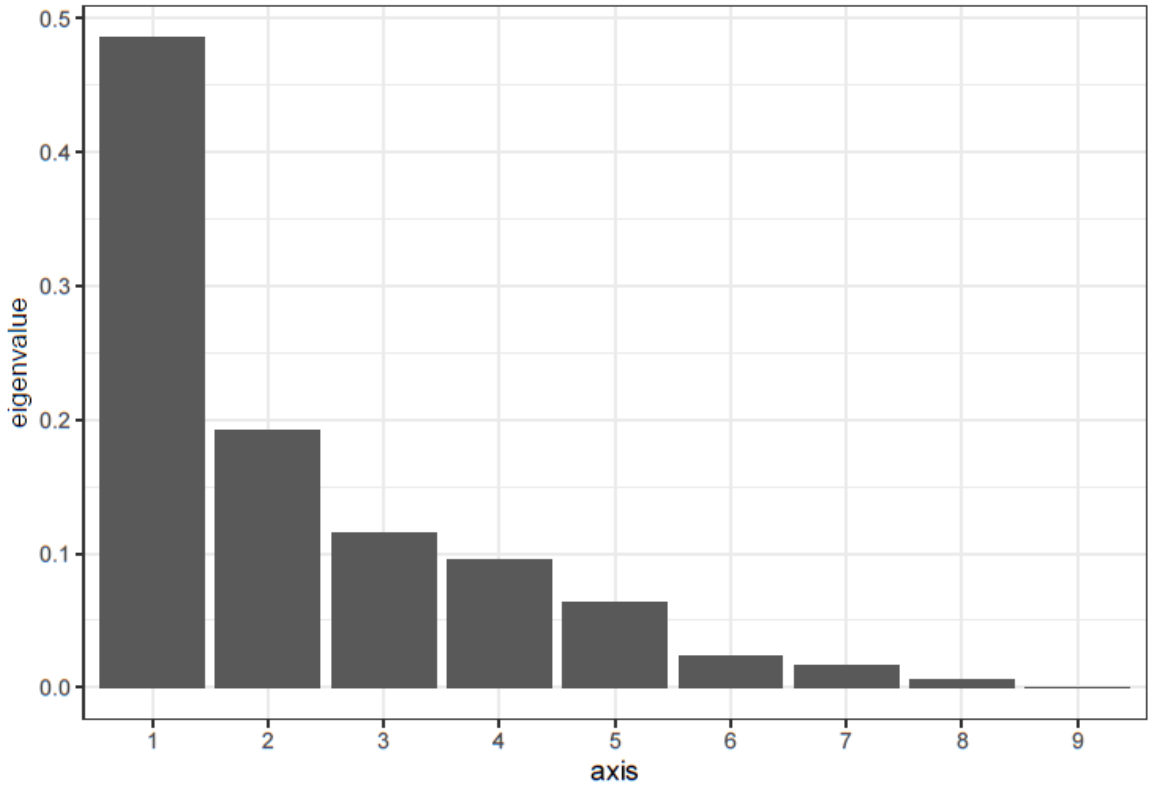
```
##   specifying a model with increasing fold change for higher values.
##   did you mean for this to be a factor? if so, first convert
##   this variable to a factor using the factor() function
```

```
## gene-wise dispersion estimates
```

```
## mean-dispersion relationship
```

```
## final dispersion estimates
```

```
bx.ord_pcoa_bray <- ordinate(mothur_suelo_filt_desq, "PCoA", "bray")

#Scree plot
plot_scree(bx.ord_pcoa_bray) + theme_bw()
```



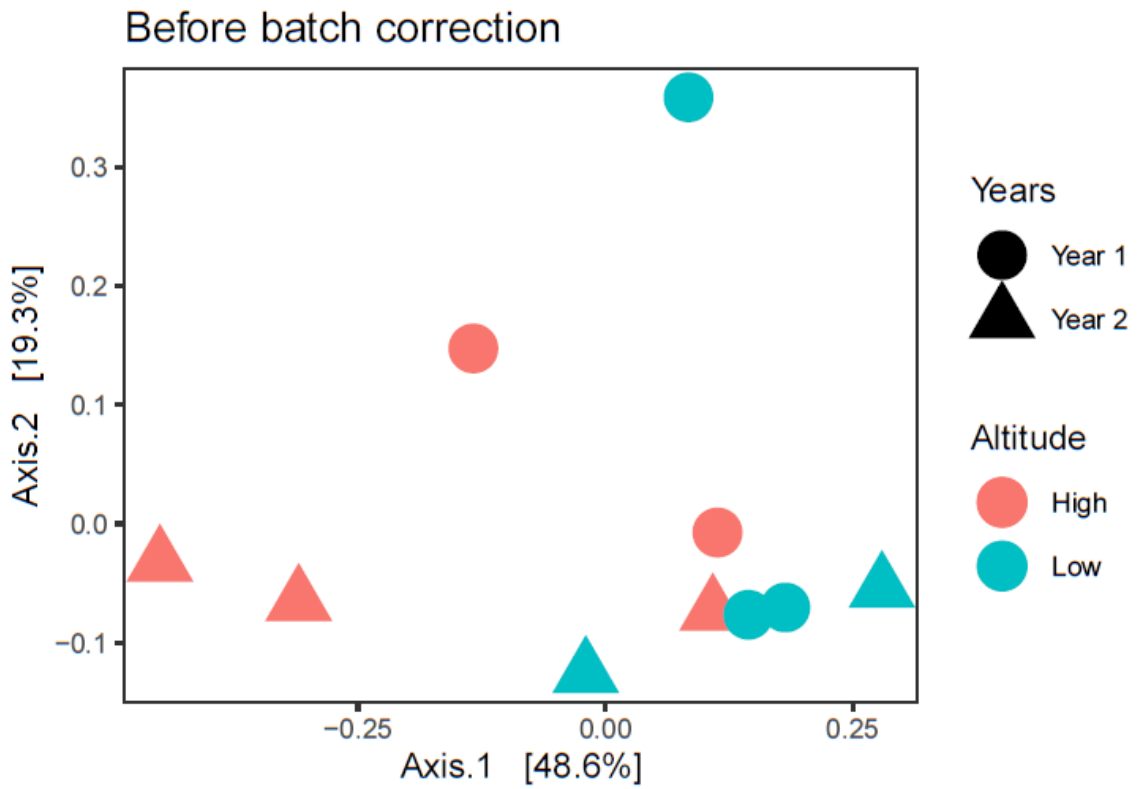

En la representación se observa que los Ejes 1 y 2 explican el 70% de la muestra.

```
beta.ps1 <- plot_ordination(mothur_suelo_filt_desq,
                            shape="Years",
                            color= "Altitude",
                            bx.ord_pcoa_bray,
                            label = "X")  +
  theme(plot.title = element_text(hjust = 0, size = 12))
```

```
## Warning in plot_ordination(mothur_suelo_filt_desq, shape = "Years", color =
## "Altitude", : Label variable was not found in the available data you provided.No
## label mapped.
```

```
beta.ps1 <- beta.ps1 + theme_bw(base_size = 14) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
```

```
# Now we can join the biopsy and stool from same subject
beta.ps1 + geom_point(data=beta.ps1$data, size=9) +
  ggtitle("Before batch correction")
```




*Permanova*

```
set.seed(1)

# Calculate bray curtis distance matrix
bray <- phyloseq::distance(mothur_suelo_filt_desq, method = "bray")

# make a data frame from the sample_data
sampledf <- data.frame(sample_data(mothur_suelo_filt_desq))

# Adonis test
adonis(bray ~ Altitude * Tanda * Altura , data = sampledf)
```

```
##
## Call:
## adonis(formula = bray ~ Altitude * Tanda * Altura, data = sampledf)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##                      Df SumsOfSqs  MeanSqs F.Model      R2 Pr(>F)
## Altitude              1   0.21537 0.215366  3.0030 0.21740  0.058 .
```

```
## Tanda                  1   0.12262 0.122619  1.7098 0.12378  0.204
## Altura                 1   0.03746 0.037457  0.5223 0.03781  0.784
## Altitude:Tanda         1   0.10099 0.100993  1.4082 0.10195  0.253
## Altitude:Altura        1   0.07890 0.078899  1.1002 0.07965  0.409
## Tanda:Altura           1   0.20980 0.209798  2.9254 0.21178  0.063 .
## Altitude:Tanda:Altura  1   0.08207 0.082069  1.1444 0.08285  0.377
## Residuals              2   0.14343 0.071716         0.14479
## Total                  9   0.99063                  1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Con el permanova observamos que obtenemos como variables significativas la Altitude y la combinatoria de cTanda y Altura

```
# Homogeneity of dispersion test
beta <- betadisper(bray, sampledf$Altitude)
permutest(beta)
```

```
##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 999
##
## Response: Distances
##           Df   Sum Sq   Mean Sq      F N.Perm Pr(>F)
## Groups     1 0.009994 0.0099938 0.8036    999  0.391
## Residuals  8 0.099492 0.0124365
```

Esta salida nos dice que nuestra prueba de Adonis es significativa, por lo que podemos rechazar la hipótesis nula de que nuestros tres sitios tienen el mismo centroide.

### Batch Correction

```
otu_soil <- otu_table(mothur_suelo_filt_desq)  + 0.0001
otu_soil <- t(otu_soil)
soil_otu_mat <-  logratio.transfo(otu_soil, logratio = 'CLR')
class(soil_otu_mat) <- "matrix"
```

```
## Warning in class(soil_otu_mat) <- "matrix": Setting class(x) to "matrix" sets
## attribute to NULL; result will no longer be an S4 object
```

```
metadata_only_sl <- microbiome::meta(mothur_suelo_filt) %>% as.data.frame()
```

```
# Correct batch effect
correct_batch <-  function(abundance, batch){
  res.lm <- lm(abundance ~ batch)
  summary.res <- summary(res.lm)
  p <- summary.res$coefficients[2,4]
}

otu_soil_batch <- apply(X = soil_otu_mat,MARGIN =  2,
                        FUN =  correct_batch, batch = metadata_only_sl$Tanda)
soil_batch_adjp <- p.adjust(otu_soil_batch, method = 'fdr')
```

```
# Detech batch effect
soil_mod <- model.matrix( ~  metadata_only_sl$Tanda) # Full model
soil_mod0 <- model.matrix( ~ 1, data =  factor(metadata_only_sl$Tanda) ) # null model
soil_sva_n <- num.sv(dat = t(soil_otu_mat), mod = soil_mod)
```

```r
# Apply SVA
soil_sva <- sva(dat = t(soil_otu_mat), mod = soil_mod,
                mod0 = soil_mod0, n.sv = soil_sva_n)
```

```
## Number of significant surrogate variables is:  2
## Iteration (out of 5 ):1  2  3  4  5
```

```r
# Remove batch effect with limma
soil_limma <- t(removeBatchEffect(t(soil_otu_mat),
                                  batch = metadata_only_sl$Tanda,
                                  design = soil_mod0))
```

```r
soil_limma <- compositions::clrInv(soil_limma)
class(soil_limma) <- "matrix"
soil_limma_nmds <- metaMDS(soil_limma)
```

```
## Run 0 stress 0.0833171
## Run 1 stress 0.05873726
## ... New best solution
## ... Procrustes: rmse 0.1433251  max resid 0.2478996
## Run 2 stress 0.05873707
## ... New best solution
## ... Procrustes: rmse 9.558203e-05  max resid 0.0001631405
## ... Similar to previous best
## Run 3 stress 0.1851773
## Run 4 stress 0.05924776
## Run 5 stress 0.05873712
## ... Procrustes: rmse 0.0001286106  max resid 0.0002114498
## ... Similar to previous best
## Run 6 stress 0.05873715
## ... Procrustes: rmse 0.0001340529  max resid 0.0002233263
## ... Similar to previous best
## Run 7 stress 0.0587374
## ... Procrustes: rmse 0.0002859761  max resid 0.0004732644
## ... Similar to previous best
## Run 8 stress 0.2063824
## Run 9 stress 0.05924793
## Run 10 stress 0.2367489
## Run 11 stress 0.2134393
## Run 12 stress 0.0587375
## ... Procrustes: rmse 0.0003315577  max resid 0.0005482939
## ... Similar to previous best
## Run 13 stress 0.05924785
## Run 14 stress 0.2148004
## Run 15 stress 0.05924782
## Run 16 stress 0.08331557
## Run 17 stress 0.2313119
## Run 18 stress 0.05873707
## ... Procrustes: rmse 2.398459e-05  max resid 3.817129e-05
## ... Similar to previous best
## Run 19 stress 0.08331562
## Run 20 stress 0.05873707
## ... New best solution
## ... Procrustes: rmse 1.977934e-05  max resid 3.144095e-05
## ... Similar to previous best
```

```
## *** Solution reached
```

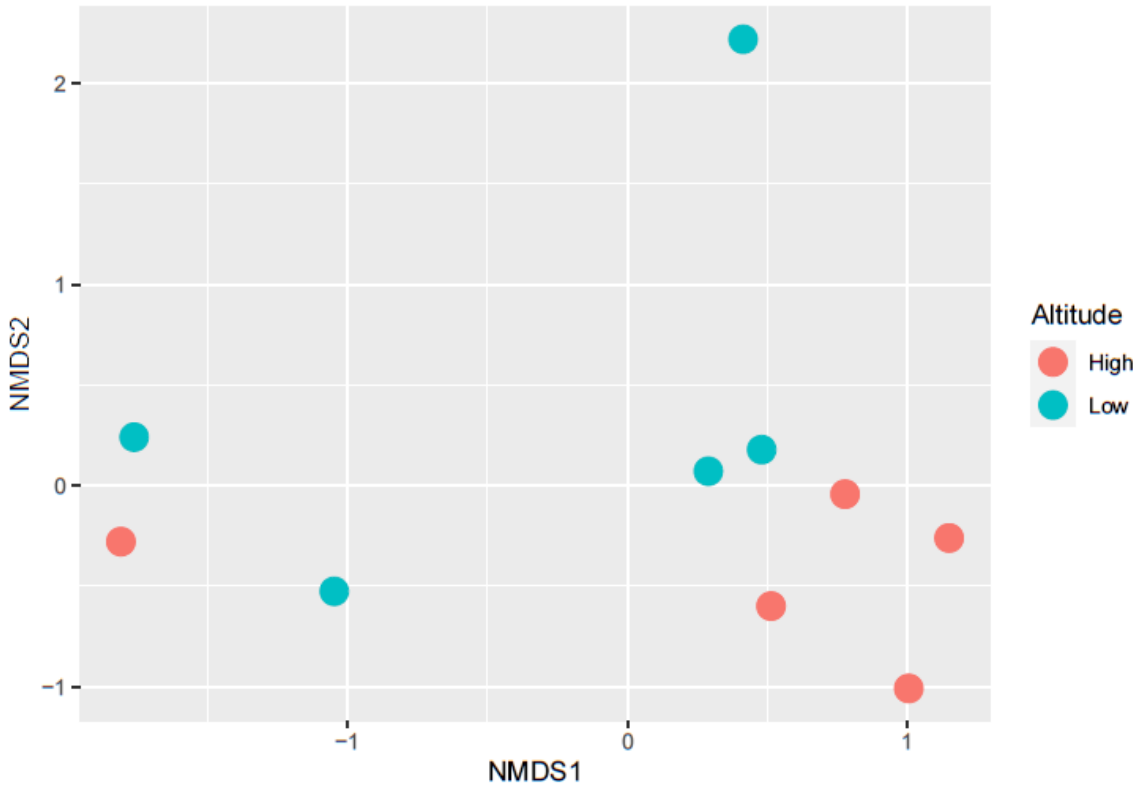
```
meta_soil2 <- meta_soil %>% dplyr::select(-Sample, -PO4) %>% as.data.frame
rownames(meta_soil2) <- meta_soil2$SeqID
meta_soil2 <- dplyr::select(meta_soil2, -SeqID)
soil_nmds_env <- vegan::envfit(soil_limma_nmds, meta_soil2,
                               permutations = 1e3, na.rm = T)
soil_nmds_env
```

```
##
## ***VECTORS
##
##              NMDS1    NMDS2     r2  Pr(>r)
## pH         0.34076  0.94015 0.5825 0.07393 .
## EC         0.99616 -0.08760 0.0379 0.84715
## Ca        -0.61764 -0.78646 0.3521 0.27473
## Mg         0.76583 -0.64304 0.3327 0.29670
## Na         0.70789  0.70632 0.2219 0.44156
## K          0.77408  0.63308 0.1465 0.60639
## Fe         0.50210 -0.86481 0.0145 0.94905
## Mn        -0.63542 -0.77216 0.6671 0.04196 *
## P          0.70540  0.70881 0.4909 0.13487
## S          0.99951 -0.03139 0.5341 0.09690 .
## N         -0.87929 -0.47628 0.7637 0.01399 *
## Org        0.63994 -0.76843 0.0957 0.70729
## Humidity -0.96439 -0.26447 0.7632 0.01399 *
## CIC        0.45841  0.88874 0.3600 0.26973
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Permutation: free
## Number of permutations: 1000
##
## 1 observation deleted due to missingness
```

```
soil_limma_scores <- scores(soil_limma_nmds)
soil_scores_samples <- soil_limma_scores %>% as.data.frame %>%
  dplyr::mutate(Sample = rownames(.))

metadata_only_sl2 <- data.frame(metadata_only_sl) %>%
  dplyr::select(-Sample) %>%
  dplyr::rename(Sample = SeqID) %>%
  dplyr::mutate(Tanda = factor(Tanda))

soil_scores_samples <- left_join(soil_scores_samples, metadata_only_sl2)
```

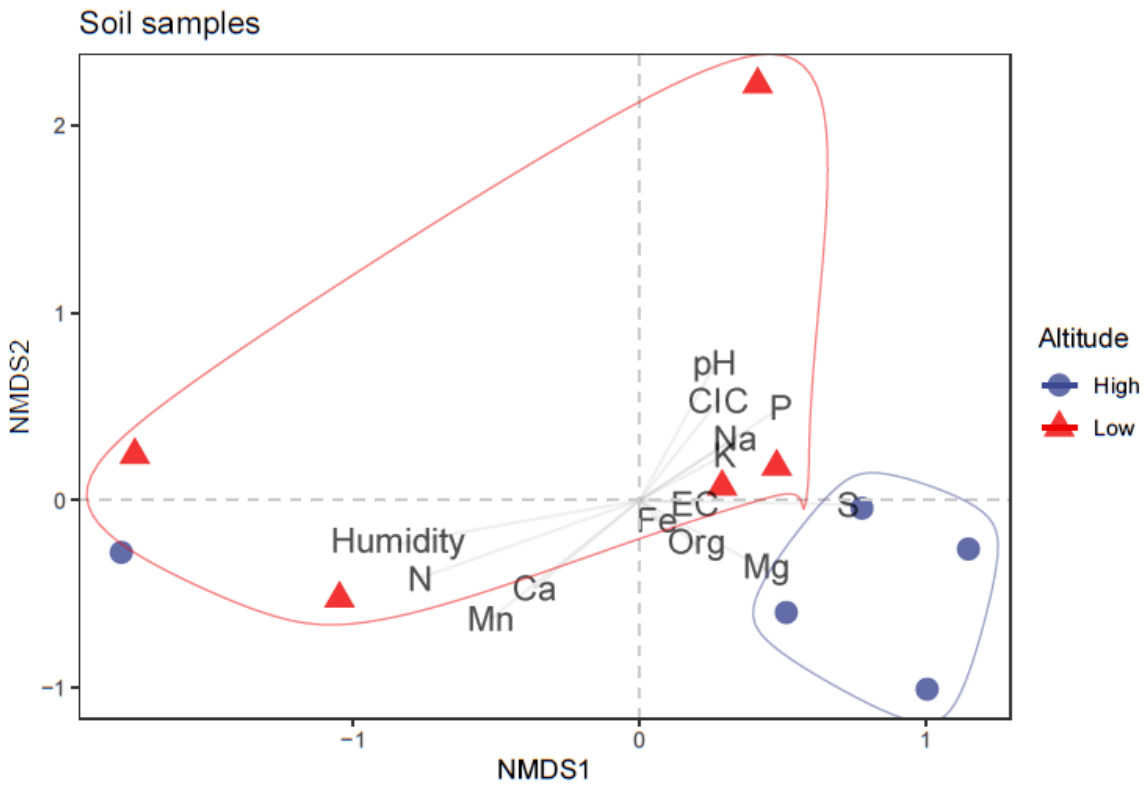
```
## Joining, by = "Sample"
```

```
ggplot(soil_scores_samples, aes(NMDS1, NMDS2, color = Altitude)) +
  geom_point(size = 5)
```




```
soil_env_coord <- scores(soil_nmds_env, "vectors") %>% as.data.frame %>%
  mutate(Variables = rownames(.))
```

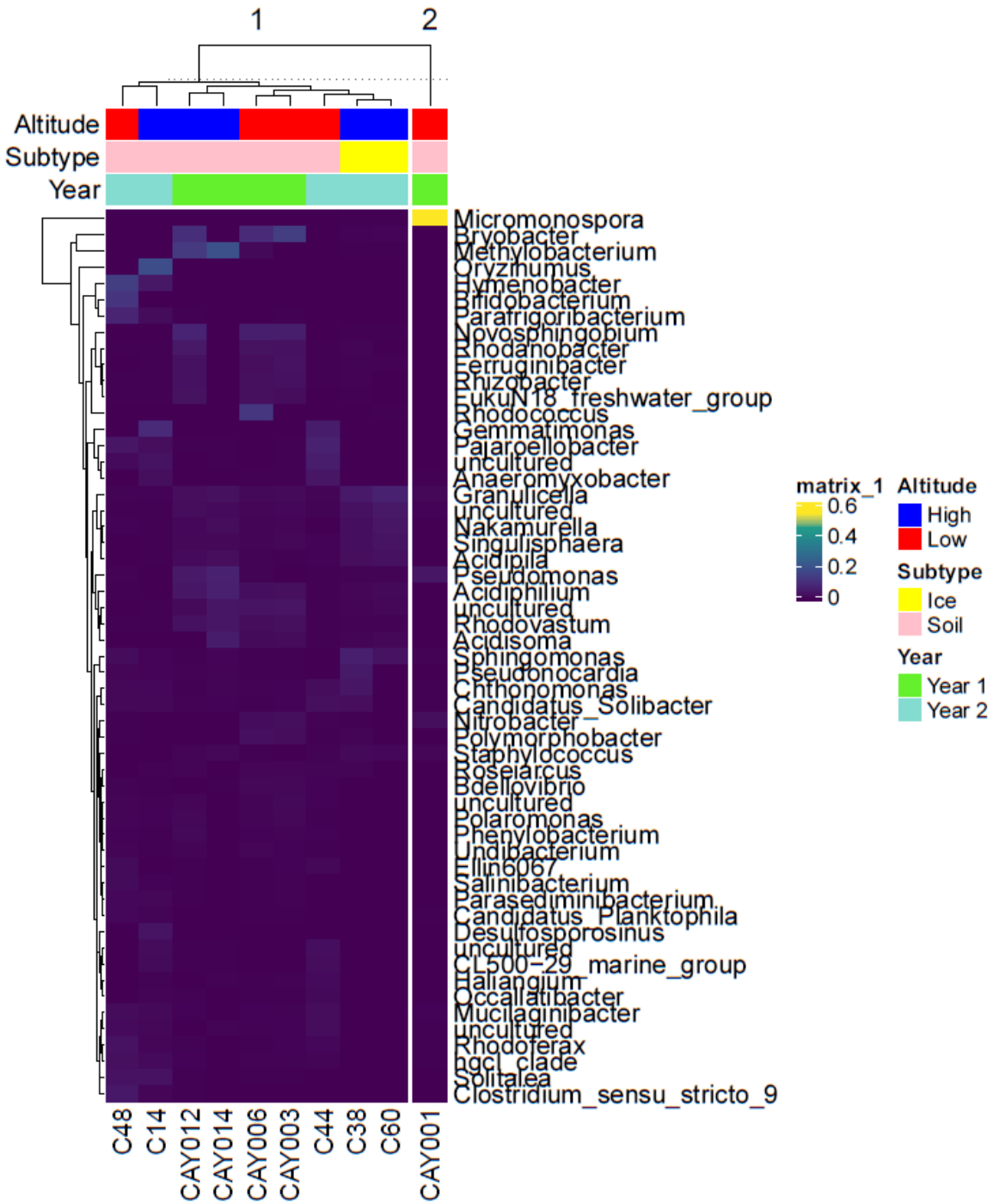
```
plt_sl_nmds <- ggplot(soil_scores_samples, aes(NMDS1, NMDS2)) +
  geom_point(aes(shape = Altitude, color = Altitude),
             size = 4, alpha = 0.8)  +
  scale_color_aaas() + theme_test() +
  labs(x = "NMDS1", y = "NMDS2") +
   #ggrepel::geom_text_repel(aes(label =Sample), size = 4) +
  geom_segment(data = soil_env_coord , aes(x =0, y =0,
                                           xend= NMDS1,
                                           yend = NMDS2),
               arrow = arrow(length = unit(0.05, 'cm')), alpha = 0.1,
               color = 'gray40', lineend = 'square') +
    geom_text(data = soil_env_coord, aes(NMDS1, NMDS2, label = Variables),
              alpha = 0.7, size = 5, parse = T, color = "black") +
  geom_hline(yintercept = 0, color = "gray80", linetype = 2) +
  geom_vline(xintercept = 0,  color = "gray80", linetype = 2) +
   ggtitle("Soil samples") +
  geom_encircle(data = filter(soil_scores_samples, Sample != "C14"),
                aes(NMDS1, NMDS2, color = Altitude), alpha = 0.5, inherit.aes = F)
  #theme(legend.position=c(0.9,0.8))
plt_sl_nmds
```




#### Heatmap

```
soil_dt_heatmap <- t(soil_limma) %>% data.frame() %>% as.matrix()
rownames(soil_dt_heatmap) <- data.frame(tax_table(mothur_suelo_filt))[, "Genus"]

soil_dt_heatmap <-  soil_dt_heatmap[-grep("_ge", rownames(soil_dt_heatmap)), ]
soil_dt_heatmap <-  soil_dt_heatmap[-grep("_unclassified",
                                          rownames(soil_dt_heatmap)), ]

ha_soil <- HeatmapAnnotation(Altitude = sample_data(mothur_suelo)$Altitude,
                             Subtype = sample_data(mothur_suelo)$Subtype,
                             Year = sample_data(mothur_suelo)$Years,
                             col = list(Altitude = c(High = "blue", Low = "red"),
                                        Subtype = c(Ice = "yellow", Soil = "pink")),
                             annotation_name_side = "left")
```

```
Heatmap(log(soil_dt_heatmap + 1), col = viridis(10)[-c(7, 8, 9)],
        top_annotation = ha_soil, column_km = 2)
```




Agua

***Remoción de OTUs pobremente representados***

Se ha elegido que OTUs que no son detectados más de cinco veces en más de la mitad de las muestras sean removidos. Conservando de esta manera 28 y 58 OTUs para muestras de agua y suelo, respectivamente.

```
##  High samples
# Create the criteria to filter
criteria_5counts_wt = genefilter_sample(mothur_water,
                                        filterfun_sample(function(x) x > 5),
                                        A=0.5*nsamples(mothur_water))
# Filter samples
mothur_water_filt = prune_taxa(criteria_5counts_wt, mothur_water)
mothur_water
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 1286 taxa and 41 samples ]
## sample_data() Sample Data:       [ 41 samples by 10 sample variables ]
## tax_table()   Taxonomy Table:    [ 1286 taxa by 6 taxonomic ranks ]
```

```
mothur_water_filt
```

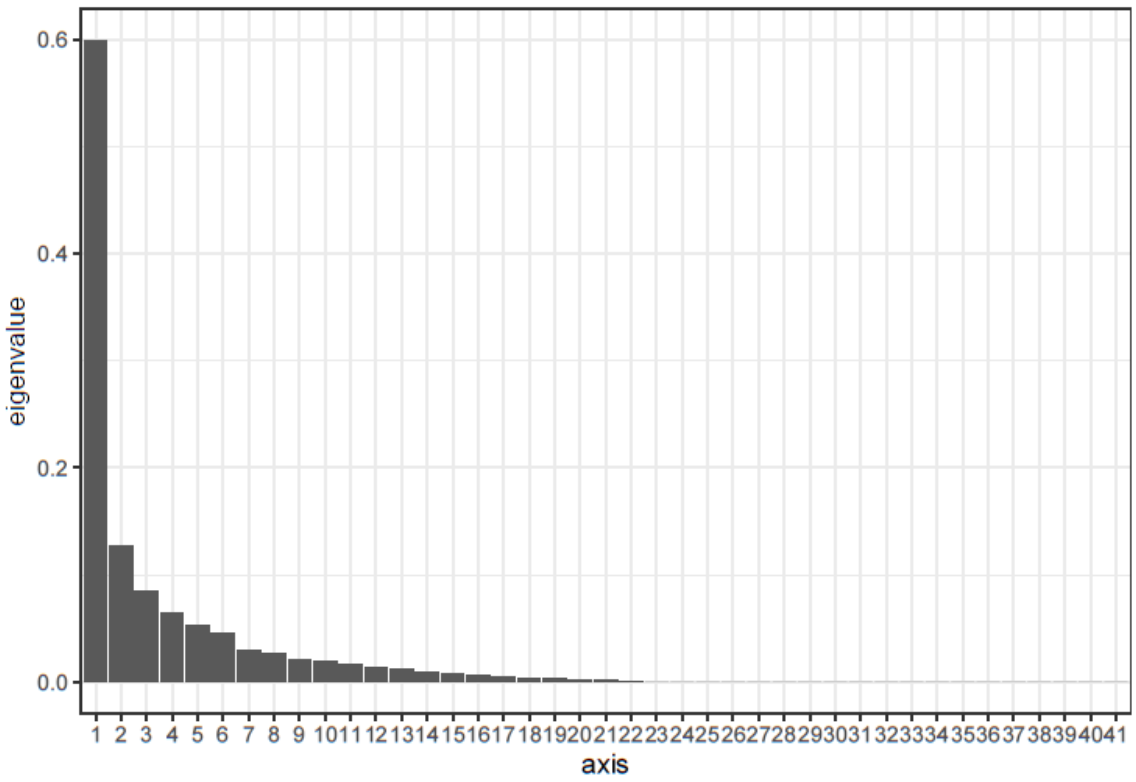
```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 24 taxa and 41 samples ]
## sample_data() Sample Data:       [ 41 samples by 10 sample variables ]
## tax_table()   Taxonomy Table:    [ 24 taxa by 6 taxonomic ranks ]
```

*Normalizacion por Deseq2*

```
mothur_water_filt_desq = normalDeseq2(mothur_water_filt)
```

```
## converting counts to integer mode

## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors

##   the design formula contains one or more numeric variables with integer values,
##   specifying a model with increasing fold change for higher values.
##   did you mean for this to be a factor? if so, first convert
##   this variable to a factor using the factor() function

## gene-wise dispersion estimates

## mean-dispersion relationship

## -- note: fitType='parametric', but the dispersion trend was not well captured by the
##    function: y = a/x + b, and a local regression fit was automatically substituted.
##    specify fitType='local' or 'mean' to avoid this message next time.

## final dispersion estimates
```

```
bx.ord_pcoa_bray <- ordinate(mothur_water_filt_desq, "PCoA", "bray")

#Scree plot
plot_scree(bx.ord_pcoa_bray) + theme_bw()
```

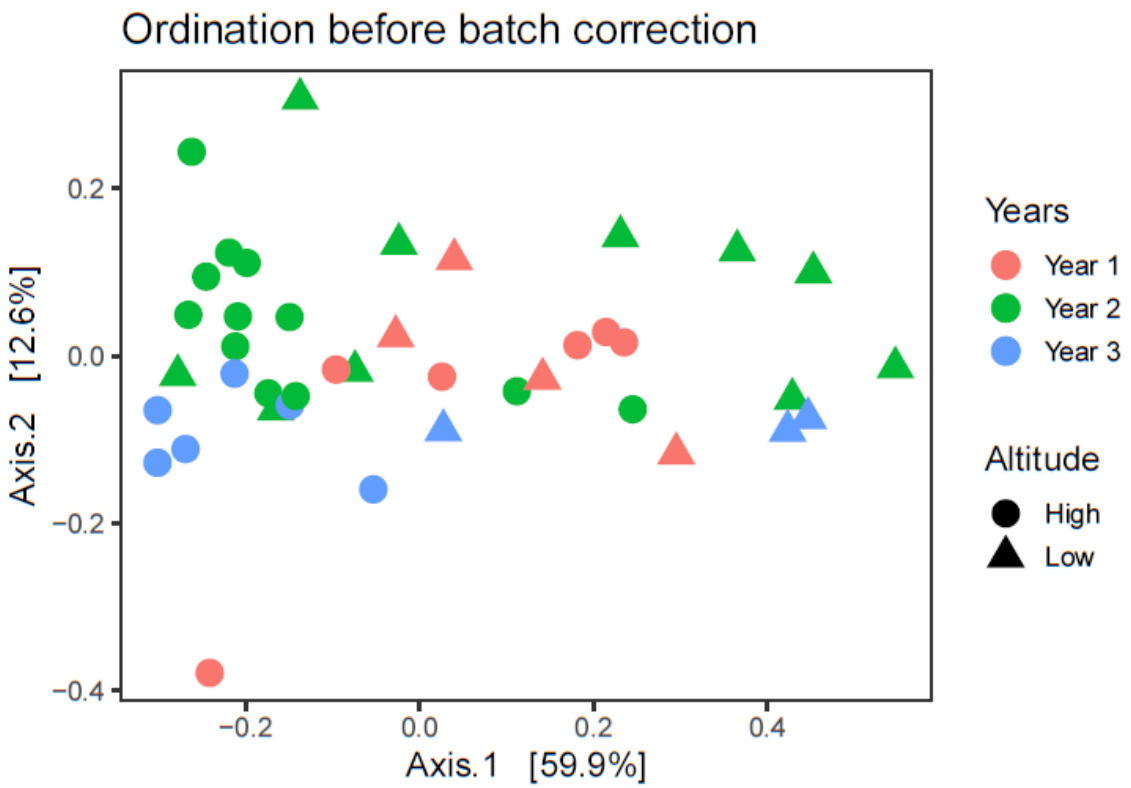



En la representación se observa que los Ejes 1 y 2 explican el 70% de la muestra.

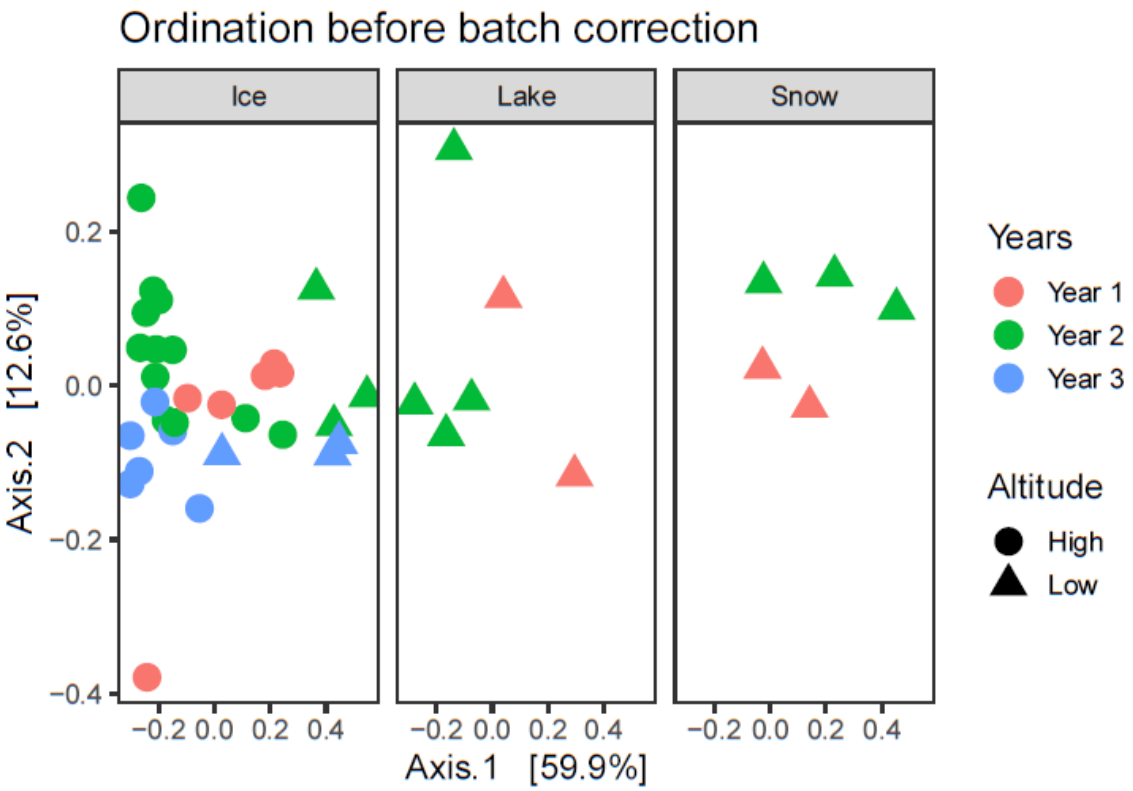
```
beta.ps1 <- plot_ordination(mothur_water_filt_desq,
                            color="Years", shape = "Altitude",
                            bx.ord_pcoa_bray,
                            label = "X")  +
  theme(plot.title = element_text(hjust = 0, size = 12))
```

```
## Warning in plot_ordination(mothur_water_filt_desq, color = "Years", shape =
## "Altitude", : Label variable was not found in the available data you provided.No
## label mapped.
```

```
beta.ps1 <- beta.ps1 + theme_bw(base_size = 14) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())

# Now we can join the biopsy and stool from same subject
beta.ps1 + geom_point(size = 5) + ggtitle("Ordination before batch correction")
```




```
beta.ps1 + geom_point(size = 5) + ggtitle("Ordination before batch correction") +
  facet_wrap("Subtype")
```




***Permanova***

```
set.seed(1)

# Calculate bray curtis distance matrix
bray <- phyloseq::distance(mothur_water_filt_desq, method = "bray")

# make a data frame from the sample_data
sampledf <- data.frame(sample_data(mothur_water_filt_desq))

# Adonis test
adonis(bray ~ Altitude * Years * Altura , data = sampledf)
```

```
##
## Call:
## adonis(formula = bray ~ Altitude * Years * Altura, data = sampledf)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##                      Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## Altitude              1    0.8342 0.83423 13.3386 0.19655  0.001 ***
## Years                 2    0.3527 0.17635  2.8197 0.08310  0.014 *
## Altura                1    0.1821 0.18206  2.9110 0.04289  0.048 *
## Altitude:Years        2    0.3751 0.18757  2.9991 0.08839  0.016 *
## Altitude:Altura       1    0.2477 0.24768  3.9602 0.05836  0.022 *
```

```
## Years:Altura           2    0.3692 0.18461  2.9518 0.08699  0.020 *
## Altitude:Years:Altura  2    0.0696 0.03479  0.5563 0.01640  0.785
## Residuals             29    1.8137 0.06254         0.42733
## Total                 40    4.2444                 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Con el permanova observamos que obtenemos como variables significativas la Altitude y la combinatoria de cTanda y Altura

```
# Homogeneity of dispersion test
beta <- betadisper(bray, sampledf$Altitude)
permutest(beta)
```

```
##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 999
##
## Response: Distances
##           Df  Sum Sq  Mean Sq      F N.Perm Pr(>F)
## Groups     1 0.02907 0.029072 2.6185    999  0.131
## Residuals 39 0.43300 0.011103
```

Esta salida nos dice que nuestra prueba de Adonis es significativa, por lo que podemos rechazar la hipótesis nula de que nuestros tres sitios tienen el mismo centroide.

```
otu_water <- otu_table(mothur_water_filt_desq)  + 0.0001
otu_water <- t(otu_water)
water_otu_mat <-  logratio.transfo(otu_water, logratio = 'CLR')
class(water_otu_mat) <- "matrix"
```

```
metadata_only_wt <- sample_data(mothur_water_filt) %>% as.data.frame()
```

```
# Correct batch effect
correct_batch <-  function(abundance, batch){
  res.lm <- lm(abundance ~ batch)
  summary.res <- summary(res.lm)
  p <- summary.res$coefficients[2,4]
}

otu_water_batch <- apply(X = water_otu_mat,MARGIN =  2,
                         FUN =  correct_batch, batch = metadata_only_wt$Tanda)
watere_batch_adjp <- p.adjust(otu_water_batch, method = 'fdr')
```

```
# Detech batch effect
water_mod <- model.matrix( ~  metadata_only_wt$Tanda) # Full model
water_mod0 <- model.matrix( ~ 1, data =  factor(metadata_only_wt$Tanda) ) # null model
water_sva_n <- num.sv(dat = t(water_otu_mat), mod = water_mod)
```

```
# Apply SVA
water_sva <- sva(dat = t(water_otu_mat), mod = water_mod,
```

```
                    mod0 = water_mod0, n.sv = water_sva_n)
```

Batch correction

```
## Number of significant surrogate variables is:  1
## Iteration (out of 5 ):1  2  3  4  5
```

```
# Remove batch effect with limma
water_limma <- t(removeBatchEffect(t(water_otu_mat),
                                   batch = metadata_only_wt$Tanda,
                                   design = water_mod0))
```

```
water_limma <- compositions::clrInv(water_limma)
class(water_limma) <- "matrix"
water_limma_nmds <- metaMDS(water_limma)
```

```
## Run 0 stress 0.1728408
## Run 1 stress 0.2475632
## Run 2 stress 0.1728408
## ... Procrustes: rmse 1.70546e-05  max resid 5.437182e-05
## ... Similar to previous best
## Run 3 stress 0.2077591
## Run 4 stress 0.2078657
## Run 5 stress 0.1728408
## ... New best solution
## ... Procrustes: rmse 6.583659e-06  max resid 1.9553e-05
## ... Similar to previous best
## Run 6 stress 0.1728408
## ... Procrustes: rmse 2.663696e-05  max resid 0.000107618
## ... Similar to previous best
## Run 7 stress 0.1728408
## ... Procrustes: rmse 1.199094e-05  max resid 3.439449e-05
## ... Similar to previous best
## Run 8 stress 0.1963932
## Run 9 stress 0.2260845
## Run 10 stress 0.1728408
## ... Procrustes: rmse 9.847853e-06  max resid 3.290168e-05
## ... Similar to previous best
## Run 11 stress 0.1728408
## ... Procrustes: rmse 1.85453e-05  max resid 3.551051e-05
## ... Similar to previous best
## Run 12 stress 0.1728408
## ... Procrustes: rmse 2.554301e-05  max resid 7.923704e-05
## ... Similar to previous best
## Run 13 stress 0.1728408
## ... Procrustes: rmse 8.006013e-05  max resid 0.0002609242
## ... Similar to previous best
## Run 14 stress 0.2338903
## Run 15 stress 0.21297
## Run 16 stress 0.2129696
## Run 17 stress 0.1728408
## ... Procrustes: rmse 4.433413e-05  max resid 0.000175521
## ... Similar to previous best
## Run 18 stress 0.2047542
## Run 19 stress 0.1728408
```

```
## ... Procrustes: rmse 2.764162e-06  max resid 6.840913e-06
## ... Similar to previous best
## Run 20 stress 0.1728408
## ... Procrustes: rmse 1.590611e-05  max resid 5.520041e-05
## ... Similar to previous best
## *** Solution reached
```

Envifit

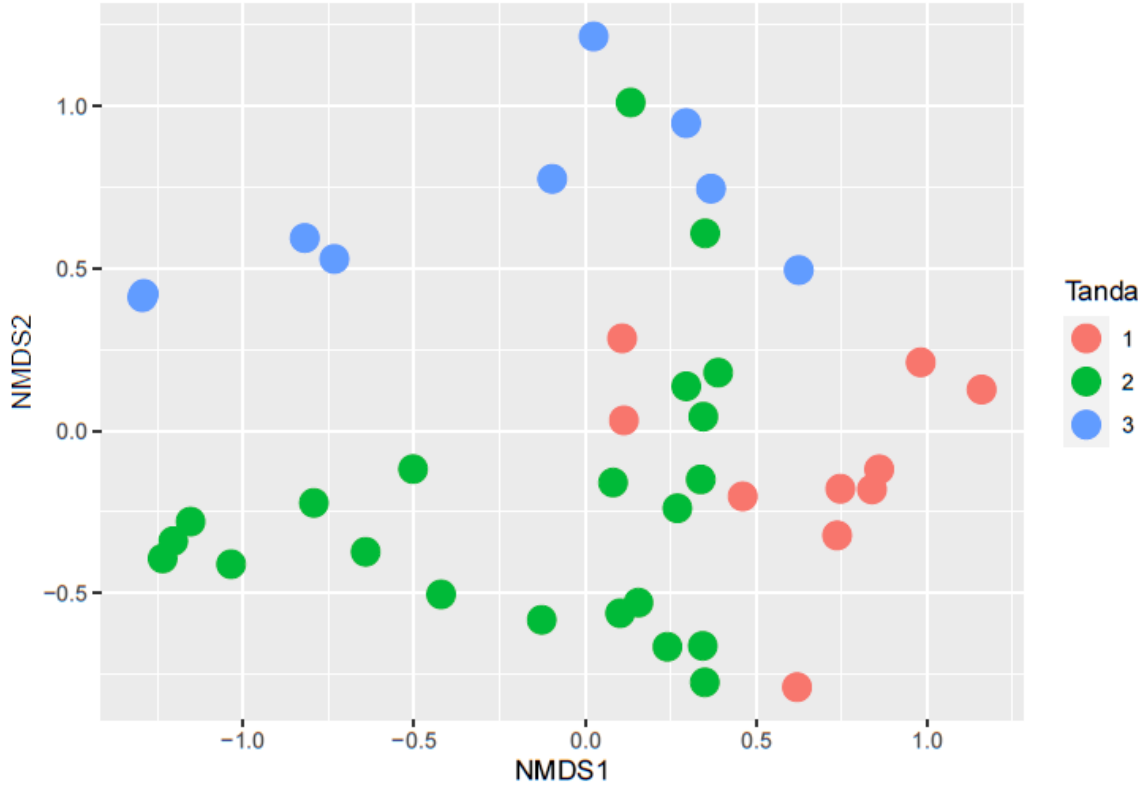
```
meta_water2 <- meta_water %>% dplyr::select(-Sample) %>% as.data.frame
rownames(meta_water2) <- meta_water2$SeqID
meta_water2 <- dplyr::select(meta_water2,-SeqID)
water_nmds_env <- vegan::envfit(water_limma_nmds, meta_water2,
                                permutations = 1e3,na.rm = T)
water_nmds_env
```

```
##
## ***VECTORS
##
##           NMDS1    NMDS2     r2  Pr(>r)
## pH    -0.21081 -0.97753 0.0129 0.81319
## EC    -0.56919 -0.82221 0.0209 0.75025
## TDS   -0.85297 -0.52196 0.0831 0.24076
## Ca    -0.10575 -0.99439 0.1671 0.03696 *
## Mg    -0.03715 -0.99931 0.0709 0.29870
## Na    -0.99942  0.03398 0.0545 0.44256
## K     -0.88978  0.45638 0.0625 0.37063
## Fe    -0.99990  0.01448 0.0717 0.32667
## SO4   -0.87921 -0.47644 0.0810 0.25375
## Cl    -0.68395 -0.72953 0.0542 0.43457
## TH    -0.09536 -0.99544 0.1696 0.03996 *
## CaCO3 -0.10598 -0.99437 0.1671 0.03696 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Permutation: free
## Number of permutations: 1000
##
## 5 observations deleted due to missingness
```

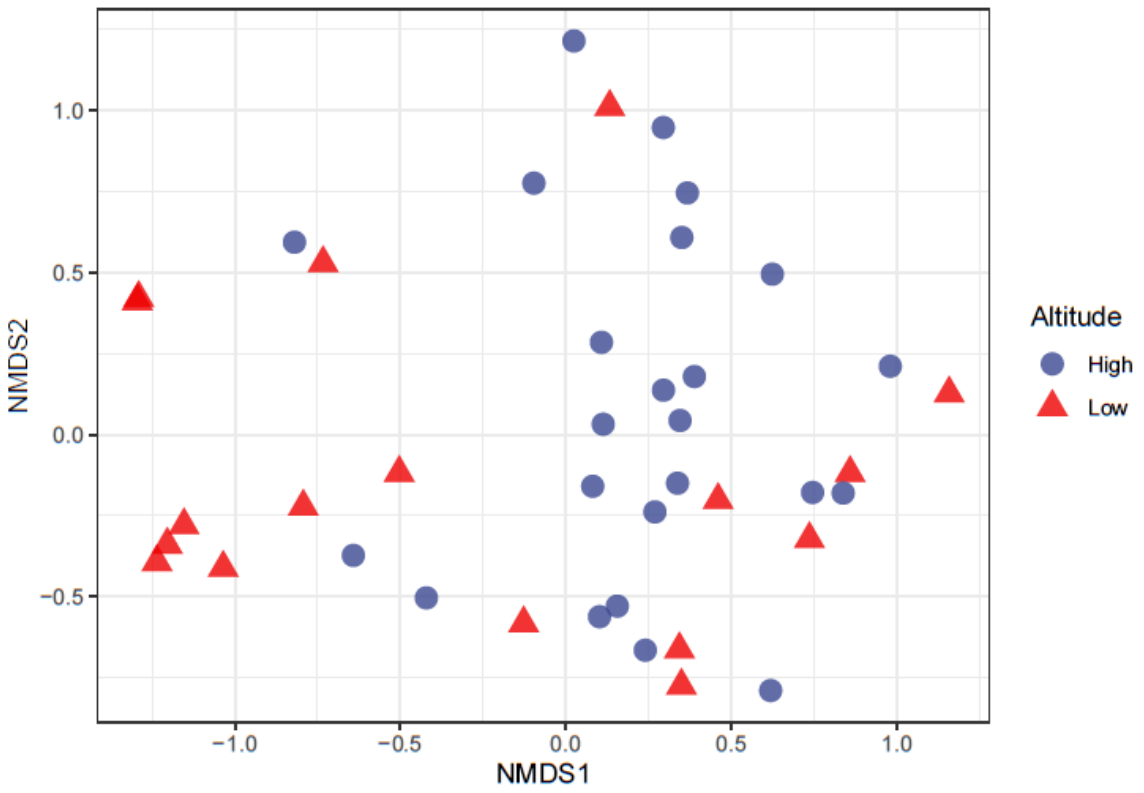
```
water_limma_scores <- scores(water_limma_nmds)
water_scores_samples <- water_limma_scores %>% as.data.frame %>%
  dplyr::mutate(Sample = rownames(.))

metadata_only_wt2 <- data.frame(metadata_only_wt) %>%
  dplyr::select(-Sample) %>%
  dplyr::rename(Sample = SeqID) %>%
  dplyr::mutate(Tanda = factor(Tanda))

water_scores_samples <- left_join(water_scores_samples, metadata_only_wt2)
```

```
## Joining, by = "Sample"
```

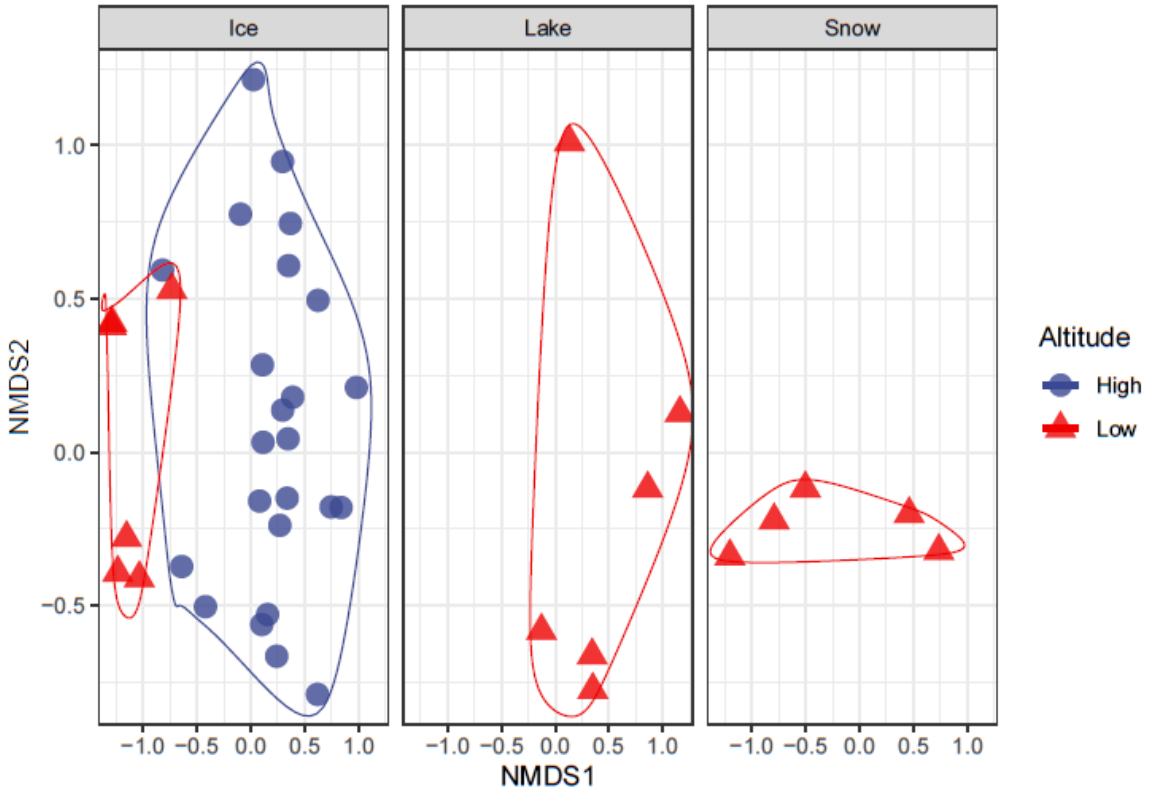
```
ggplot(water_scores_samples, aes(NMDS1, NMDS2, color = Tanda)) +
  geom_point(size = 5)
```




```
base_water_nmds <- ggplot(water_scores_samples,
       aes(NMDS1, NMDS2, color = Altitude, shape = Altitude)) +
  geom_point(size = 4, alpha = 0.8)  +  theme_bw() +
  scale_color_aaas()
base_water_nmds
```

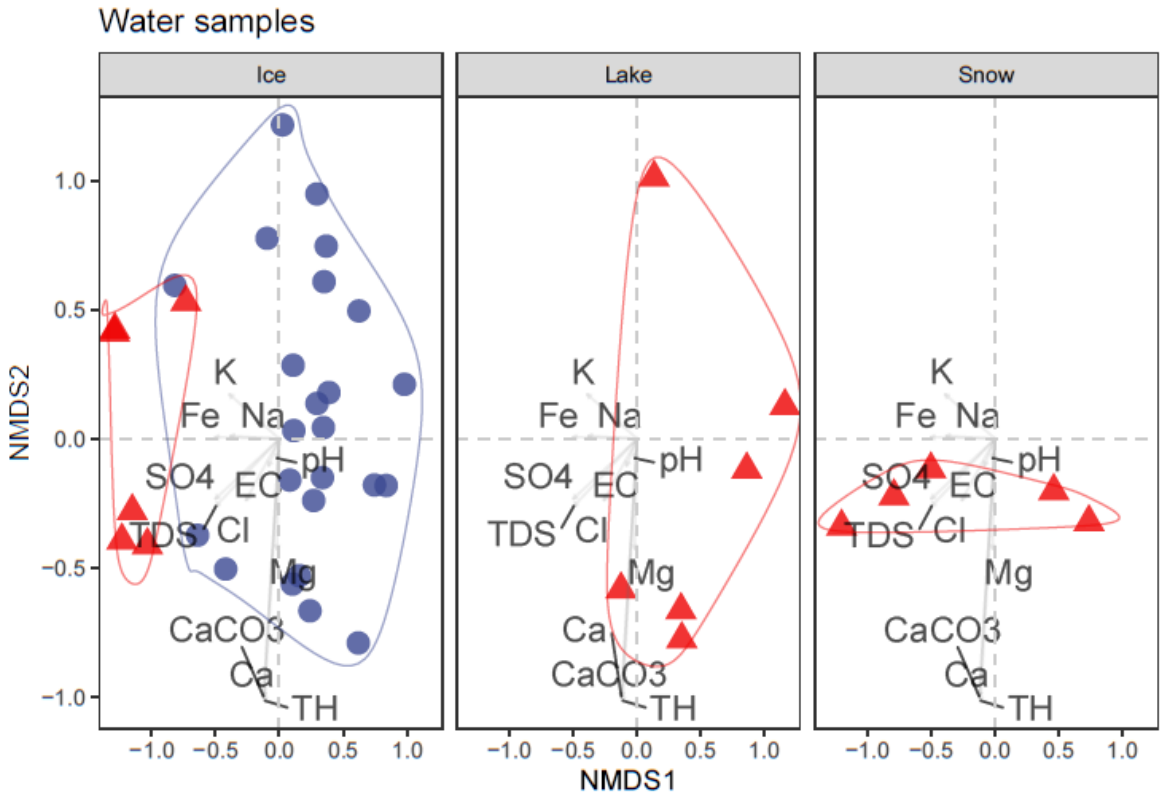



```
split_water_nmds <- base_water_nmds + facet_wrap("Subtype") +
  geom_encircle()
split_water_nmds
```




```
water_env_coord <- scores(water_nmds_env, "vectors") %>% as.data.frame %>%
  mutate(Variables = rownames(.)) %>%
  mutate(NMDS1 = NMDS1 * sqrt(water_nmds_env$vectors$r) * 7,
         NMDS2 = NMDS2 * sqrt(water_nmds_env$vectors$r) * 6)
```

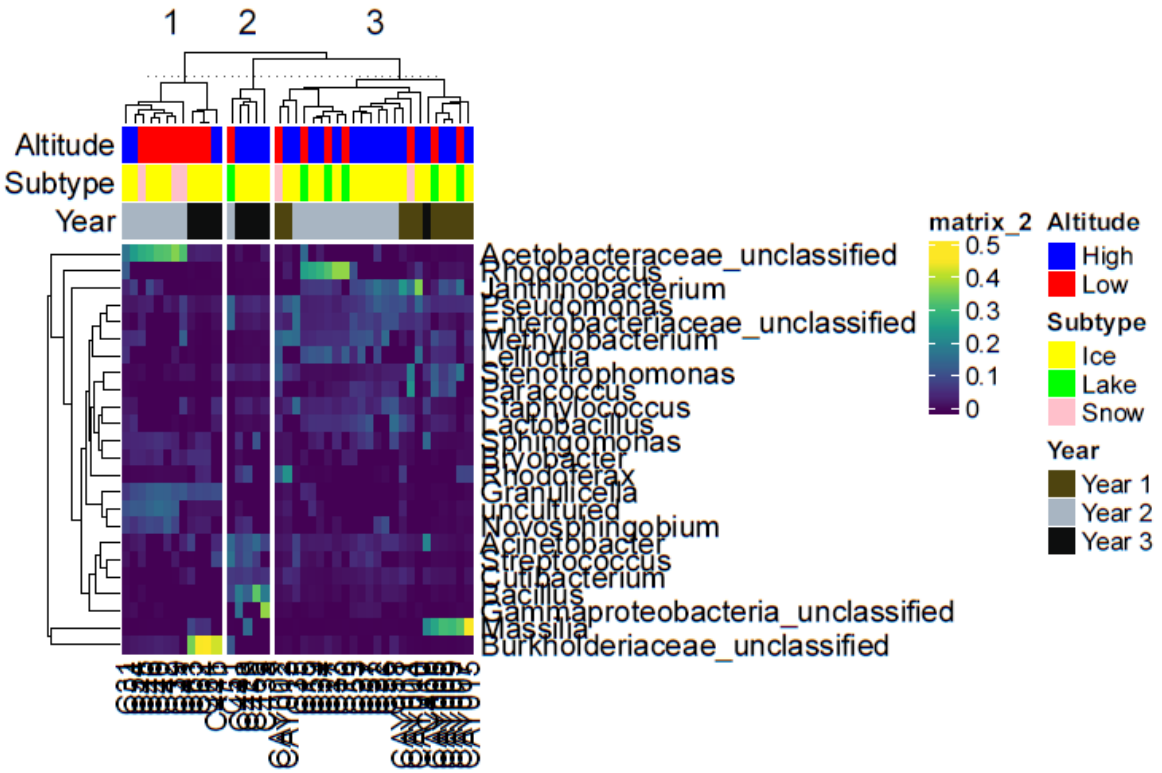
```
plt_wt_nmds <- ggplot(water_scores_samples, aes(NMDS1, NMDS2)) +
  geom_point(aes(shape = Altitude, color = Altitude),
             size = 4, alpha = 0.8)  +
  scale_color_aaas() + theme_test() +
  labs(x = "NMDS1", y = "NMDS2") +
   #ggrepel::geom_text_repel(aes(label =Sample), size = 4) +
  geom_segment(data = water_env_coord , aes(x =0, y =0,
                                            xend= NMDS1,
                                            yend = NMDS2),
               arrow = arrow(length = unit(0.05, 'cm')), alpha = 0.1,
               color = 'gray40', lineend = 'square') +
    ggrepel::geom_text_repel(data = water_env_coord,
                             aes(NMDS1, NMDS2, label = Variables),
            alpha = 0.7, size = 5, parse = T, color = "black") +
  geom_hline(yintercept = 0, color = "gray80", linetype = 2) +
  geom_vline(xintercept = 0,  color = "gray80", linetype = 2) +
   ggtitle("Water samples") + guides(color = F, shape = F) +
  facet_wrap("Subtype") + geom_encircle(aes(color = Altitude),alpha = 0.5)
  #theme(legend.position=c(0.9,0.8))
plt_wt_nmds
```




#### Heatmap

```
water_dt_heatmap <- t(water_limma) %>% data.frame() %>% as.matrix()
rownames(water_dt_heatmap) <- data.frame(tax_table(mothur_water_filt_desq))[, "Genus"]

ha_water <- HeatmapAnnotation(Altitude = sample_data(mothur_water)$Altitude,
                              Subtype = sample_data(mothur_water)$Subtype,
                              Year = sample_data(mothur_water)$Years,
                              col = list(Altitude = c(High = "blue", Low = "red"),
                                         Subtype = c(Lake = "green",
                                                     Ice = "yellow", Snow = "pink")),
                              annotation_name_side = "left")
```

```
Heatmap(log(water_dt_heatmap + 1), col = viridis(10),
        top_annotation = ha_water, column_km = 3)
```




Soil

### Otus Diferenciales

Deseq2 posee dos tipos de hipótesis para trajar con los datos. En CAYAMBE añó 1 trabajamos con el test de Wald que busca cambios en log2 para obtener los grupos diferenciales. En este caso trabajaremos con el test likelihood Ratio Test (LRT).

Este test, trabaja con dos modelos para los recuentos, un modelo completo y un modelo reducido. En nustreo caso el modelo reducido contemplara únicamente la variable Altitud, y el modelo completo añadira la variable Tanda. La prueba determina si la mayor probabilidad de que los datos utilicen los términos adicionales en el modelo completo es mayor de lo esperado si esos términos adicionales son realmente cero.

Por lo tanto, la LRT es útil para probar varios términos a la vez,o todas las interacciones entre dos variables. El LRT para los datos de recuento es conceptualmente similar a un cálculo de análisis de varianza (ANOVA) en regresión lineal.

```
require(DESeq2)
diagdds = phyloseq_to_deseq2(mothur_water_filt,
                             design =  ~ Tanda * Altitude)
```

```
## converting counts to integer mode
```

```
## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors
```

```
##   the design formula contains one or more numeric variables with integer values,
##   specifying a model with increasing fold change for higher values.
##   did you mean for this to be a factor? if so, first convert
##   this variable to a factor using the factor() function
```

```
gm_mean = function(x, na.rm=TRUE){
    exp(sum(log(x[x > 0]), na.rm=na.rm) / length(x))
}
geoMeans = apply(counts(diagdds), 1, gm_mean)

  #Estimamos la varianza
diagdds = estimateSizeFactors(diagdds, geoMeans = geoMeans)
diagdds = estimateDispersions(diagdds)
```

```
## gene-wise dispersion estimates
## mean-dispersion relationship
## -- note: fitType='parametric', but the dispersion trend was not well captured by the
##    function: y = a/x + b, and a local regression fit was automatically substituted.
##    specify fitType='local' or 'mean' to avoid this message next time.
## final dispersion estimates
```

```
dds = DESeq(diagdds, test = "LRT", reduced = ~ Tanda + Altitude)
```

```
## using pre-existing size factors
## estimating dispersions
## found already estimated dispersions, replacing these
## gene-wise dispersion estimates
## mean-dispersion relationship
## -- note: fitType='parametric', but the dispersion trend was not well captured by the
##    function: y = a/x + b, and a local regression fit was automatically substituted.
##    specify fitType='local' or 'mean' to avoid this message next time.
## final dispersion estimates
## fitting model and testing
## -- replacing outliers and refitting for 7 genes
## -- DESeq argument 'minReplicatesForReplace' = 7
## -- original counts are preserved in counts(dds)
## estimating dispersions
## fitting model and testing
```

```
res <- results(dds)

res
```

```
## log2 fold change (MLE): Tanda.AltitudeLow
## LRT p-value: '~ Tanda * Altitude' vs '~ Tanda + Altitude'
## DataFrame with 24 rows and 6 columns
##             baseMean log2FoldChange     lfcSE         stat     pvalue      padj
##            <numeric>      <numeric> <numeric>    <numeric>  <numeric> <numeric>
## Otu000002 15543.354      2.8240289   1.13344  1.355804581   0.244266  0.407110
## Otu000005  5868.367     -0.1247999   1.43390 -0.000109514   1.000000  1.000000
## Otu000007  1002.088      0.0644719   1.96997  0.000517641   0.981848  1.000000
## Otu000008   935.718     -2.5282968   1.40336  1.646318461   0.199461  0.362656
## Otu000013  1723.835      3.2380247   1.95659  2.614324132         NA        NA
## ...              ...            ...       ...          ...        ...       ...
```

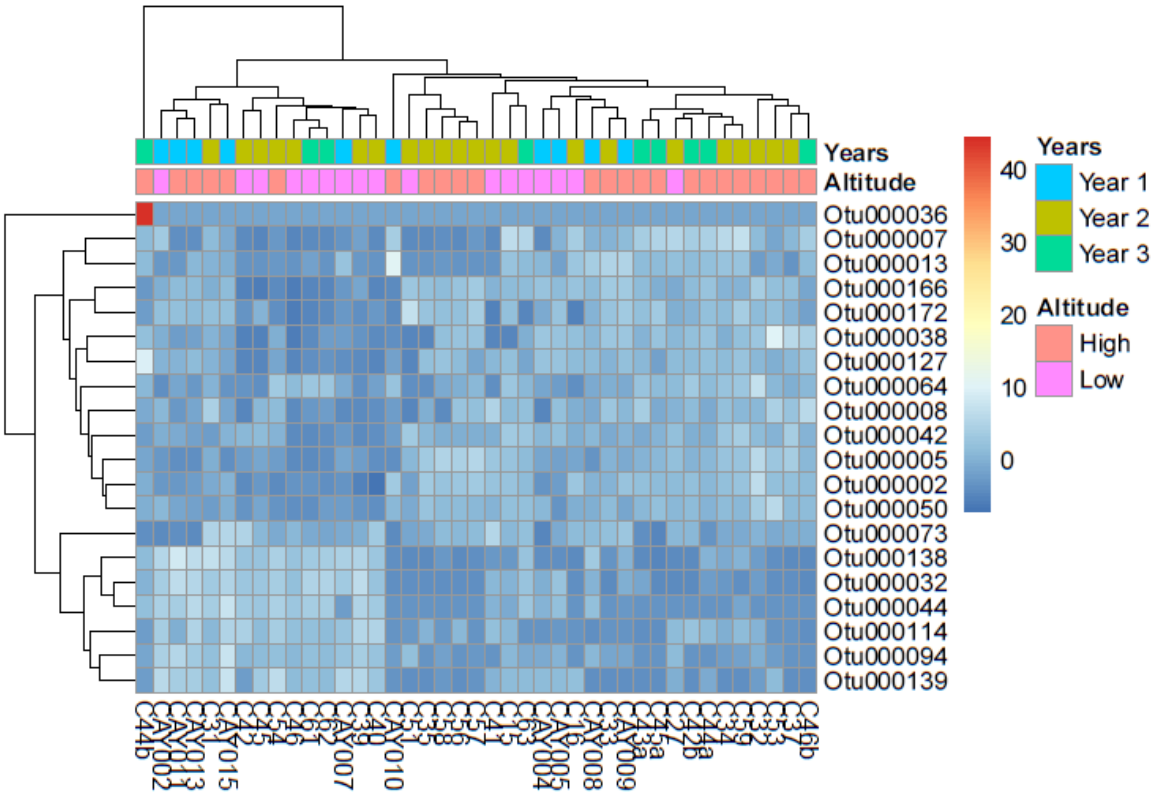
```
## Otu000172   43.0398     -3.0942150   1.10580   4.83786165 0.02784140 0.1392070
## Otu000226  336.0974      0.0758237   1.53937   0.00196467 0.96464571 1.0000000
## Otu000294  272.8573      0.8710711   1.87171   0.18398425 0.66797136 0.8349642
## Otu000546   61.8319     -2.6507853   1.35282   2.71042535 0.09969441 0.2848412
## Otu001256   50.8879     -3.6597830   1.30784  10.35525198 0.00129107 0.0129107
```

### Heatmap

```
# make the DeSeqDataSet
library(pheatmap)
# use the log transform on the data set
rld <- rlog(dds, blind=F)
```

```
## rlog() may take a few minutes with 30 or more samples,
## vst() is a much faster transformation
```

```
topVarianceGenes <- head(order(rowVars(assay(rld)), decreasing=T),20)
matrix <- assay(rld)[1:20,]
matrix <- matrix - rowMeans(matrix)

# select the 'contrast' you want
annotation_data <- as.data.frame(colData(rld)[c("Altitude","Years")])
pheatmap(matrix, annotation_col=annotation_data)
```




Podemos ajustar los resultados en funcion del contrats, mediante la funcion res(), a partir del parametro design podemos ir obteniendo loas OTUS presentes en A no en B.

Tendriamos que comentar como obtener estas diferencias en la reunión.

ref dese12 https://www.bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html

# Radar plots

## Soil

```
soil_radar_data <- meta(mothur_suelo) %>% merge(meta_soil) %>%
   filter(SeqID != "C14") %>%
  dplyr::select(-Sample, -Replica, -Altura, -Type, -Tanda, -Subtype,
                -Years, -SeqDepth, -PO4, -SeqID) %>%
  mutate_each(funs(rescale), -Altitude) %>% group_by(Altitude) %>%
  summarise_all(~mean(.)) %>% rename(CEC = CIC)
```

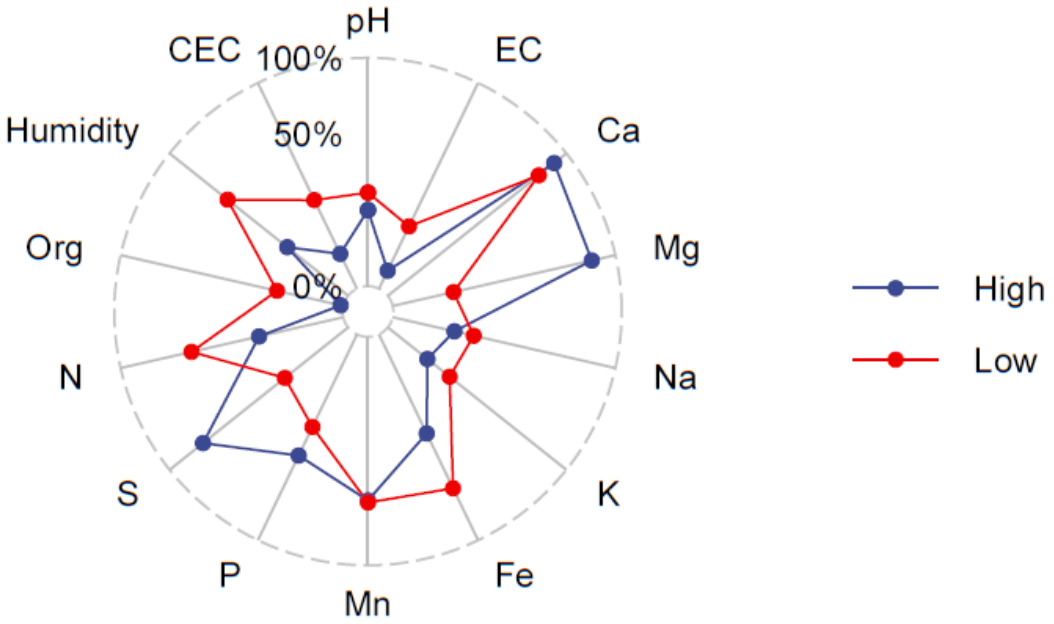
```
## Warning: `funs()` is deprecated as of dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```
## Warning: `mutate_each_()` is deprecated as of dplyr 0.7.0.
## Please use `across()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```
soil_radar <- ggradar(soil_radar_data, group.line.width = 0.5, grid.max = 0.75,
          group.point.size = 2.5, background.circle.colour = "white",
          gridline.mid.colour = "white", legend.position = "right") +
  scale_color_aaas()
```

```
## Scale for 'colour' is already present. Adding another scale for 'colour',
## which will replace the existing scale.
```

```
soil_radar
```




### Water Samples

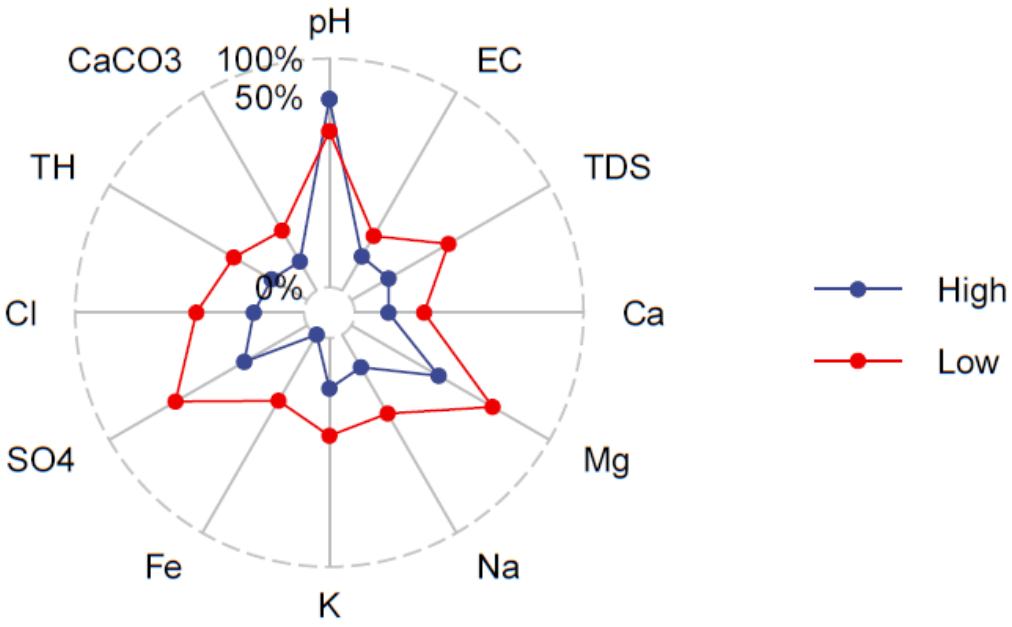
```
water_radar_data <- meta(mothur_water) %>% merge(meta_water) %>%
  filter(SeqID != c("C61", "C62", "C63", "C15", "C16")) %>%
  dplyr::select(-Sample, -Replica, -Altura, -Type, -Tanda, -Subtype,
                -Years, -SeqDepth, -SeqID) %>%
  mutate_each(funs(rescale), -Altitude) %>% group_by(Altitude) %>%
  summarise_all(~mean(.))# %>% rename(CEC = CIC)
```

```
## Warning in SeqID != c("C61", "C62", "C63", "C15", "C16"): longer object length
## is not a multiple of shorter object length
```

```
soil_radar <- ggradar(water_radar_data, group.line.width = 0.5, grid.max = 0.6,
          group.point.size = 2.5, background.circle.colour = "white",
          gridline.mid.colour = "white", legend.position = "right") +
  scale_color_aaas()
```

```
## Scale for 'colour' is already present. Adding another scale for 'colour',
## which will replace the existing scale.
```

```
soil_radar
```


pH
EC
TDS
Ca
Mg
Na
K
Fe
SO4
Cl
TH
CaCO3
100%
50%
0%
High
Low

# List of figures and Tables

# List of figures

## Overview



## Chapter 1

**Figure 2.2c**. Cayambe Volcano, Alpha-diversity analysis of the through all years of survey. Comparison between High and low altitude samples according to the glacier chronosequence for soil-based samples.

**Figure 2.2d.** Cayambe Volcano, Alpha-diversity analysis of the through all years of survey. Effective Shannon index were compared in term of the altitudinal gradient (High Vs Low) for water-based samples.

**Figure 2.2e.** Cayambe Volcano, Alpha-diversity analysis of the through all years of survey. Simpson index were compared in term of the altitudinal gradient (High Vs Low) for water-based samples.

**Figure 2.3.** Cayambe Volcano, Alpha-diversity analysis of the through all years of survey. Comparison between High and low altitude samples according to the glacier chronosequence for water-based samples.

**Figure 2.4.** Cayambe Volcano, rarefaction curves of the through all years of survey. The curves show the asymptote with al least 60% of reads, based only in all soil-samples.

**Figure 2.5.** Cayambe Volcano, rarefaction curves of the through all years of survey. The curves show the asymptote with at least 60% of reads, based only in all water-based samples.

**Figure 2.6.** Beta-diversity analysis. (a) NMDS of the soil samples (stress = 0.17), environmental variables were fitted to the NMDS space. (b) Radar plot of the assessed environmental soil samples. (c, d, e, and f) NMDS of the water-based samples. The ordination analysis was conducted with the three sample subtypes (ice, snow, lake), and, as a visualization technique, the scores are displayed in three different cartesian planes for ice (c), lake (d), and snow (e). The fitted environmental variables are displayed in an additional plane (f). (g) Radar plot of the measured environmental variables for water-based samples.

**Figure 2.7**. Cayambe Volcano plot. Shows two genera with a p-value < 0.05. based only in all soil-based samples communities.

**Figure 2.8.** Cayambe Volcano plot. Shows significant genera with a p-value >= 0.05. based only in all water-based samples communities

## Chapter 2

**Figure 3.1.** Bacterial Community structure of Sumaco Volcano at the Family level of the through the three years of survey. The compositional plot is grouped in the relative abundance of the 10 most abundant families for soil. The sampling year 1,2 and the sample subtype categories are displayed as plot titles.

**Figure 3.2.** Bacterial Community structure of Sumaco Volcano at the Family level of the through the three years of survey. The compositional plot is grouped in the relative abundance of the most abundant families for water. The sampling year 1,2,3 and the sample subtype categories are displayed as plot titles.

**Figure 3.3.** Sumaco Volcano, Alpha-diversity analysis of the through all years of survey. Effective Shannon index were compared in term of the altitudinal gradient (High Vs Low) for soil-based communities.

**Figure 3.4.** Sumaco Volcano, Alpha-diversity analysis of the through all years of survey. Simpson indexes were compared in term of the altitudinal gradient (High Vs Low) for soil-based communities

**Figure 3.5.** Sumaco Volcano, Alpha-diversity analysis of the through all years of the survey. Simpson index were compared in term of the altitudinal gradient (High Vs. Low) for lake-based communities

**Figure 3.6.** Sumaco Volcano, Alpha-diversity analysis of the through all years of the survey. Simpson indexes were compared in terms of the altitudinal gradient (High Vs Low) for lake-based communities.

**Figure 3.7.** Sumaco Volcano, Alpha-diversity analysis of the through all years of the survey. Inverse Simpson indexes were compared in terms of the altitudinal gradient (High vs. Low) for soil-based communities

**Figure 3.8.** Sumaco Volcano, Alpha-diversity analysis of the through all years of survey. Inverse Simpson indexes were compared in terms of the altitudinal gradient (High Vs Low) for laked-based communities.

**Figure 3.9.** Sumaco Volcano, Alpha-diversity analysis of the through all years of the survey. Richness indexes were compared in term of the altitudinal gradient (High Vs Low) for laked-based communities

**Figure 3.10.** Sumaco Volcano, rarefaction curves of the through all years of sthe urvey. The curves show the asymptote with al least 60% of reads based only in all soil communities.

**Figure 3.11.** Sumaco Volcano, rarefaction curves of the through all years of the survey. The curves show the asymptote with at least 40% of reads, based only on all lakes-communities

**Figure 3.12.** Beta diversity analysis. Ordination analysis for the microcosm community, an NMDS analysis for soil samples (a) and a CCC analysis for lake (b) samples. Volcano plot for the differential genus analysis for soil (c) and lake (d) samples. Radar plot for the environmental variables for soil (e) and lake (f) samples.

# List of tables

## Overview



## ANEXX

Esta Investigación aporta a la Microbiología ambiental en descifrar y comprender el complejo sistema terrestre y acuático, particularmente, la evaluación microbiana de ecosistemas extremos. En este tipo de ambientes, los organismos altamente adaptados habitan los paisajes más adversos. Dicha adaptación es muy dinámica y compleja, y a menudo más de un frente evolutivo está involucrado en este proceso. Por ejemplo, en los glaciares, las temperaturas por debajo de 0ºC, la alta radiación ultravioleta y la escasez de fuentes energéticas, son factores desafiantes que los microorganismos en este tipo de hábitats deben enfrentar con éxito. Se han descubierto diferentes rutas para superar estas fuerzas evolutivas. Muchos de este tipo de microorganismos son capaces de reducir el consumo de energía y de expresar enzimas de alta actividad catalítica que pueden aplicarse en la industria. Este estudio cubrió tanto la tecnología empleada para lograr los objetivos, así como el aspecto más importante acerca de la estructura e identificación proximal de las comunidades bacterianas presentes en las dos ubicaciones de los Andes ecuatorianos, el glaciar Cayambe y el volcán Sumaco, no reportadas hasta la actualidad.

This research contributes to environmental microbiology in deciphering and understanding the complex terrestrial and aquatic system, particularly, the microbial evaluation of extreme ecosystems. In such environments, highly adapted organisms inhabit the most adverse landscapes. Such adaptation is highly dynamic and complex, and often more than one evolutionary process is involved. For example, in glaciers, temperatures below 0°C, high ultraviolet radiation and scarcity of energetic sources are challenging factors that microorganisms in this type of habitat must successfully cope with. Many of these types of microorganisms can reduce energy consumption and express enzymes with high catalytic activity that can be applied to industry. This proximal study covers both the technology employed to achieve the objectives, as well as the most important aspect about the structure and identification of the bacterial communities present in the two locations in the Ecuadorian Andes, the Cayambe glacier and the Sumaco volcano, not reported until now.

UC UNIVERSIDAD DE CANTABRIA Escuela de Doctorado