In-situ metagenomics: A platform for rapid sequencing and analysis of metagenomes in less than one day

Javier Tamames[1], Diego Jiménez[1], Álvaro Redondo[1], Sandra Martínez[2], Asunción de los Rios[3]

1 Microbiome Analysis Laboratory, Systems Biology Department, CNB-CSIC, 28049 Madrid, Spain
2 Centro de Investigación Mariña, Universidade de Vigo, Departamento de Ecoloxía e Bioloxía Animal, 36310 Vigo, Spain
3 Department of Biogeochemistry and Microbial Ecology, National Museum of Natural Sciences (MNCN-CSIC), 28006 Madrid, Spain

Abstract

We present here a complete system for metagenomic analysis that allows to produce the sequencing and analysis of a medium-size metagenome in less than one day. This unprecedented development was possible due to the conjunction of state-of-the art experimental and computational advances: a portable, autonomous laboratory suitable for DNA extraction and sequencing with nanopore technology; the powerful metagenomic analysis pipeline SqueezeMeta, capable to provide a complete analysis in a few hours and using scarce computational resources; and tools for the automatic inspection of the results via a graphical user interface, that can be coupled to a web server to allow remote visualization of data (SQMtools and ShinySQM). We have tested the feasibility of our approach in the sequencing of the microbiota associated to volcanic rocks in La Palma, Canary Islands. Also, we did a two-day sampling campaign of marine waters in which the results obtained the first day guided the experimental design of the second day.

This system offers the possibility of prospecting any habitat in a quick way, obtaining both taxonomic and functional information about the target microbiome.

Introduction

Metagenomics is a powerful tool for gaining insights on microbial communities, and has become a standard procedure for analysing the structure and functionality of microbiomes.

The popularization of portable sequencers, especially those based on nanopore technologies (Deamer et al., 2016), has facilitated the possibility of having rapid sequencing data which can be very valuable in several instances, for instance in clinical scenarios of disease control or epidemics (Quick et al., 2015, 2016) . Also, the portability of these devices has been explored *in situ*, for instance in oceanographic expeditions or in the Antarctic ice (Gowers et al., 2019; Johnson et al., 2017; Lim et al., 2014), illustrating the capability of producing sequences readily. This allows to envision the capacity of designing dynamic sampling campaigns, where the planning of the whole campaign can be driven by the results being produced. This can be important, for instance, whenever the sampling takes place in remote regions for which is desirable to have prompt data acquisition to prevent suboptimal results. Or in any instance in which following the course of a microbiome in real time is desired, for instance monitoring microbial blooms (Nowinski et al., 2019), assessing the quality of drinking waters (including security and bioterrorism)(Turingan et al., 2013), food processing issues like fermentations (De Filippis et al., 2017). While standard amplification approaches (metabarcoding) can be useful in some of these instances (for instance, for detecting particular organisms in the sample), they are usually limited to taxonomic composition and/or very particular functions, and present issues related to biases in the amplification. When the objective is to obtain a full functional profile of the whole community, or we suspect that our sample contains a great deal of interesting unknown organisms, metagenomics is the most sensible option.

The bottleneck of such design is the complexity of the bioinformatics analysis. In the past, we developed the SqueezeMeta pipeline (Tamames & Puente-Sánchez, 2019) with several objectives in mind: 1) offering a fast and easy-to use platform for performing the complete analysis of metagenomes. Our goal was to include all the usual steps in metagenomic analysis with state-of-the-art tools, but making them attainable to all users, no matter their bionformatic skills. 2) Breaking the dependence on large computers, making it able to run even in laptops. And 3) Providing additional tools for the statistical analysis and sharing of the results.

Since then, we and others have tested the ability of SqueezeMeta to fulfill these requirements in many different instances. These capabilities makes SqueezeMeta an optimal system to analyses real-time metagenomic data in all settings, even in rough conditions and scarcity of computational resources.

Our challenge has been to be able to produce a full metagenomic analysis in less that 24 hours, in a total autonomous way, with no reliance of electric power or internet connectivity. This will make our platform capable to work in any circumstance, in any environment (including the most remote ones) and obtain real-time results that can shared with others on-the-fly. To do so, we devised a platform composed of the different modules:
1) A portable DNA extraction laboratory, small enough to be carried by a single person, to produce required amounts of environmental DNA
2) a MinION DNA sequencer, to sequence that DNA, producing metagenomic DNA sequences.
3) The bioinformatic pipeline SqueezeMeta, running in a small laptop the same one that manages the sequencing), to analyze the metagenomic DNA sequences, and:
4) The stand-alone statistical package SQMTools (Puente-Sánchez et al., 2020) to produce statistical analysis of the data, coupled to our new ShinySQM library (https://github.com/redondrio/SQMxplore) which allows to create interactive web pages and interfaces for openly sharing the results.
These steps are summarized in Figure 1.

For testing the feasibility of in-situ sequencing and the dynamic design of campaigns, we performed two different sampling and sequencing experiments. The first was aimed to set up the protocol, by sequencing the microbiota associated to volcanic rocks in La Palma island. The second was aimed to design a two-days campaign in which the results of the first day can be used to determine the objectives for the second day. We choose sampling oceanic waters in Ria de Vigo (Spain) for that purpose.

Materials and Methods

Portable DNA extraction laboratory

The portable laboratory is composed of the following items:
-MicroSpin centrifuge yielding 12.500 RPM (ThermoFisher, Waltham, MA, USA)
-Table Vortex, lightened by removing the metal base (ThermoFisher, Waltham, MA, USA)
-MicroSpinner (ThermoFisher, Waltham, MA, USA)
-Two MiniBatteries to power up all systems (U'King Shenzhen Zhuo Qiong Technology Co., Ltd., China)
-PowerSoil DNA extraction kit (Weight: Qiagen NV, Venlo, Netherlands)
Optionally, in case of cold conditions, we can heat the devices using:
-3 Hand warmers (up to 60ºC, Shenzhen Ziheng Technology Co., Ltd., China)
-2 portable thermal isolated containers

The extraction protocol implies bead-beating and centrifugations, that can be done efficiently using the portable equipment. Our tests of efficiency indicated that the yield is similar to this obtained with standard lab equipment. Microbial DNA is often scarce in environmental samples. Therefore, it is advisable to process several extraction tubes using the same filtration column, as to gather as much DNA as possible. In our settings, we process 8 tubes per column. In our experience, it is also advisable recommendable to perform a gentle bead-beating, as to maintain as much as possible DNA integrity, which will be very important to obtain good results in the subsequent sequencing.

Also, we have improved results when purifying the extracted DNA using Omega Mag-Bind TotalPure NGS Beads (Omega Bio-Tek, Norcross, GA) . In our experience, the flow cell life span is much better preserved when this cleaning step is used, thus reducing operative costs.

All the devices are powered by a portable battery (222Wh/60000mAh) with autonomy for 12 hours of normal functioning. In case of cold conditions, we isolate the batteries in an insulated lunch bag, heated by hand warmers inside. Cool conditions for storing some reagents are kept by using a thermal insulated container (portable 10 l camping fridge) with cold packs inside.

Laboratory transportation and setting

All devices can be carried in a suitcase or a medium backpack (60 liters). The total weight is around 13 Kg (Figure 5). A light camping tent is used to provide shelter and protection from sun, rain, moisture, or winds. Inside of the tent, a small folding table (1x1 meters) is sited as stable surface, together with a portable chair.

DNA sequencing platform

The sequencing module is composed of the following items:

-Qubit 4 fluorometer (Invitrogen, ThermoFisher, Waltham, MA, USA)
-MinION sequencer (ONT, Oxford, UK)
-MinION flow cell (ONT, Oxford, UK)
-RAPid Sequencing Kit (ONT, Oxford, UK)
-Micro Thermocycler or portable water heater
-Laptop Schenker XMG Fusion 15 (16 Gb RAM, 8 core), with stand-alone MinKnown software (v21.02., ONT, Oxford, UK)

First, the DNA concentration using the Qubit fluorometer was measured. This is needed to adjust correctly the amount of DNA to be introduced in the flow cell. The concentration of DNA obtained from environmental samples is variable,

but can be as low as 40 ng/ml in lava rocks. Then, we calculate the volume of the DNA solution to be added for counting with 400 ng of DNA. As the maximum volume for preparing the library is 10 ml, we estimate that a minimum DNA concentration of 40 ng/ml is needed. Several samples can be multiplexed in the same sequencing run.

The library is prepared using the RAPid kit from ONT, barcoding the diverse samples with different tags. This kit includes a transposase that must be thermally inactivated. This can be done using a mini thermocycler, or simply heating water using a portable water heater and incubating briefly the solution. The sequencing time depends on several factors (DNA concentration, flow cell integrity, etc), until we reach the desired amount of sequence. In cold conditions, MinION device and the laptop are protected by using insulated containers with hand warmers inside.

Bioinformatics platform

The equipment needed for the bioinformatic analysis are the following:

-Same laptop than above (Schenker XMG Fusion 15), running the SqueezeMeta pipeline (https://github.com/jtamames/SqueezeMeta), R, the SQMTools, and Shiny R libraries installed, and ShinySQM. Internet connectivity is not needed for functioning, but of course would be neccesary for sharing the data over the internet, if desired.
-MiniBatteries to power up the laptop (U'King Shenzhen Zhuo Qiong Technology Co., Ltd., China)

SqueezeMeta is a fully automatic software that performs all the steps of the bioinformatic analysis of metagenomic data. The preferred mode of analysis implies assemblying the raw sequences. But when the amount of sequencing is moderate, as in our case, the performance of the assembly decreases and it is advisable to run the analysis directly on the raw reads. Each read is then processed looking for ORFs and performing taxonomic and functional annotation for them, using the sqm_longreads program from the SqueezeMeta

suite. The results are composed by a set of tables compiling all the information found for each read, and statistics on the abundance of taxa and functions. The drawback of using read annotation is that it consumes more resources and time, and therefore compromises our goal of performing the complete pipeline in less than 24 hours. As a consequence, for the marine samples we used the following strategy: Analyze the first three samples using co-assembly (using SPAdes (Bankevich et al., 2012) as assembler), to provide a quick analysis adequate to determine the most interesting spot for additional sampling. The second set of samples was analyzed using careful annotation of reads.

The analysis of the results is facilitated by the SQMtools R package, part of the SqueezeMeta suite. This library imports the tables resulting from the SqueezeMeta run and creates a R object that can be used to perform many different statistical analyses. SQMTools also includes many prefabricated commands to obtain easily the most common types of plots and analyses.

The final step is the visualization and publication of results to make them accessible to the public. For this we use ShinySQM, which is based on Shiny, a R library to build interactive web apps straight from R. ShinySQM takes the results from SqueezeMeta and SQMtools and displays them using a web interface. In this way, the data can be easily explored and shared, for instance by drawing histograms for the taxonomic composition of the sample, or the abundance of different functions. In this way, a remote user is able to access the results for inspection, not needing knowledge of R or programming.

Taxonomic diagrams were plotted with Pavian (Breitwieser & Salzberg, 2020), via the sqm2pavian script of the SqueezeMeta pipeline. KEGG diagrams were done using the SQMtools interface to PathView (Luo & Brouwer, 2013).

Sampling design: volcanic rocks

For the sequencing of microbiota associated to volcanic rocks, lava rock samples of two different ages were taken from lava fields in the south of La Palma island (Canary Island, Spain). Two main volcanic eruptions took place at

the sampling site: San Antonio volcano (1677), and Teneguia volcano (1971). Each of these produced its own lava flows, which are very close and easily identifiable (Figure 2) (28°28'32"N 17°51'04"W).
We took two samples (lava rocks), one in each spot of different age, weighting 1 Kg each. We crumbled down these using a small hammer, to obtain a fine grain powder amenable to use by the PowerSoil extraction kit.

Sampling design: seawater

For the seawater sequencing, we planned an oceanographic cruise in the Ría de Vigo (NW Iberian Peninsula). During the first day (July 12th 2022), surface (2 m) water samples were taken at three different locations: one in the outer part of the Ría importantly influenced by oceanic waters (Cap Home, 42º 14.262´N 8º 52.325´W), one in the middle sector of the embayment in a anthropogenically affected area (Samil Beach, 42º 12.551´N 8º 46.983´W) and another one in the inner part of the Ría with relatively higher influence of riverine discharge (San Simón Bay, 42º 18.707´N 8º 37.926´W) (Figure 3). The three samples were processed and analyzed in order to choose the microbial community with the most interesting metagenomic profile and repeat the sampling at the corresponding site the following day (on July 14th 2022), increasing the sampling resolution in the water column (2 m and 5 m) and the sequencing depth of the analysis.

Vertical profiles down to 22 m (at Cap Home), 6m (at Samil Beach) and 4 and 6 m (at San Simón Bay on 12th and 14th July, respectively) of water column temperature, salinity, Photosynthetic Active Radiation (PAR) and fluorescence were obtained with an SBE 25 CTD probe (Suppl Figure 1). Seawater samples were collected in 5 l acid-cleaned Niskin bottles and filtered through a 200 µm pore size mesh to remove larger zooplankton in order to ensure good replication and facilitate filtration process. Subsequently, 12 l acid-washed polycarbonate bottles were gently filled and kept under dim light conditions until arrival to the laboratory. Microplankton biomass was concentrated by means of

sequential filtration through 3 and 0.2 µm pore-size polycarbonate filters at low vacumm pressure. Particles retained in the 3 µm pore-size filters were discarded and microbial DNA was extracted from the 0.2 µm pore-size polycarbonate filters.

For each sample, 5 liters of water approximately were processed. When processing seawater samples, a first step of microplancton biomass concentration by means of vacuum filtration is needed. Logistics on board did not allow to perform this filtration at sea although this is a procedure often performed during oceanographic cruises. Therefore, instead of doing everything onboard the ship, after collection water samples were taken to the laboratory at Estación de Ciencias Marina de Toralla (ECIMAT, Vigo) for filtration. The rest of the protocol remains unaltered.

Results

Volcanic samples

The goal of this experiment was to assess the differences in community structure in lava rocks of different age, in order to shed light on the patterns of microbial colonization of lavas. Therefore, we were interested in determining the taxonomic profile of both samples.
We were able to reach the objective of completing the full protocol of sampling, DNA extraction, sequencing and analysis in-situ, powering all the equipment with batteries and in absence of data connectivity in less than 24 hours. The amount of DNA obtained from these rocks was very low, and therefore the sequencing had to be extended for several hours, leading to almost complete degradation of the flow cell. We sequenced a total of 286.4 Mb, 191 Mb for San Antonio and 95.4 Mb for Teneguia. Raw reads for these samples were analyzed using the SqueezeMeta pipeline (Suppl Table 1).

The taxonomic profiles obtained by the analysis of the metagenomes can be seen in Figure 4. While the bacterial community structure is rather similar in both samples, marked differences at respect of eukaryotic compounds were found. The composition of Ascomycota assigned to Lecanoromycetes (main class including lichen-forming fungi) differed between both samples. A clear dominance of sequences assigned to *Letharia* genus (Lecanorales) and presence also of *Cladonia* genus (Lecanorales) was observed in Teneguia lava rocks. However in San Antonio samples, sequences assigned to *Letharia* (Lecanorales) and *Lasallia* (Umbilicariales) genera were detected but without the clear dominance of *Letharia* found in Teneguia samples. In addition, sequences assigned to the fungal Chaetotyriales and Leotiomycetes orders were only found in San Antonio samples. On the other hand, sequences assigned to *Trebouxia* (Chlorophyta, Trebouxiales), the most common photobiont of lichen-forming fungi was detected also only in San Antonio samples. At respect of bacterial communities, differences in composition of the phyllum Actionabacteria were also found between both samples. These results reveal that the age of the lava conditionate mainly fungal composition and the establishment of lichen communities. Thus, it is demonstrated that this platform is useful to identify in field differences in microbial composition and focus the sampling.

We also generated functional profiles for both samples, making possible to analyze functional diversity as it is exemplified by the abundance of genes involved in sulfur metabolism (Suppl Figure 2)

Marine samples

The objective of this experiment was to test the feasibility of planning a results-driven cruise, in which an initial sampling of different locations can serve to determine the most interesting spot to be further analyzed in subsequent days. Our goal was to explore the sulphur metabolism in coastal areas. We decided to explore three locations of Ria de Vigo, looking for the one with most interesting or most abundant genes related to sulphur metabolism. We performed two

different samplings. During the first day (12th July) we took microplankton surface samples in three different locations in the Ría de Vigo, sequenced DNA and analyzed sequences in less than 24 hours. Metagenomic information recovered during the first day informed about sulfur metabolism in the three stations and helped to choose the most interesting location to perform a more detailed analysis (increased vertical resolution of sampling and increased sequencing depth) during the second day.

After DNA extraction, we were able to retrieve the following DNA concentrations in the three spots: 8.60 ng DNA · ul^{-1}, 9.75 ng DNA · ul^{-1}in, and 21.8 ng · ul^{-1}, for Cap Home, Samil Beach, and San Simón Bay samplings spots, respectively.

Taking these concentration into account, the tree samples were barcoded and pooled to equimolar amounts of DNA, which was introduced in the MinION flow cell for sequencing. We decided to sequence for just 10 hours, because as this was an exploratory analysis, not much sequencing depth was necessary, and we wanted to maximize flow cell survival.

We obtained 138.640 reads, corresponding to 424 Mb of sequence. Even if we pooled equimolar amount of DNA for the tree samples, the result did not preserve equal quantities for each sample. Indeed, 48% corresponded to San Simón sample, 31% to Cap Home, and 21% to Samil. This can be due to different causes (see discussion), and points out that determining the amount of DNA to be sequenced for each sample is not an easy task.

As we wanted the results to be ready quickly, we decided not to work with individual reads and instead analyze the results of the co-assembly of the three samples. Since the coverage in all samples was low, the proportion of reads that could be assembled was low for all samples (26%, 24% and 23%), yielding just 597 contigs (but long ones: N50=35.5 Kb, longest contig, 179 Kb). To increase the information, we decided to use the option "singletons" in SqueezeMeta, that takes all the unassembled reads and treats them as new contigs. In this way, all reads are represented in the analysis.

This finally provided 65.361 contigs (N50: 6.600 bp), coding for 252.703 ORFs. The analysis took approximately 3 hours to complete in our laptop. Therefore, the total length of the experiment was: Sampling: 5 hours. DNA extraction: 5 hours. Sequencing: 10 hours. Analysis: 3 hours, total 23 hours.

Inspection of the results in SQMtools and ShinySQM quickly determined that San Simón was the most interesting spot for sulphur metabolism, both in terms of abundance and presence of genes related to sulphur. Especially relevant was the presence of Sox genes, being the only sample in which we spotted the presence of SoxA and SoxB genes (Suppl Figure 3). Therefore, we decided that San Simón was the optimal location for the second sampling.

This second sampling was done on July 14th 2022. We took two samples in San Simón sampling point, corresponding to two different depths (2 meters and 5 meters) so sulfur metabolism of microbial communities from this station was characterized in detail.
The concentration of measured DNA was 29.7 ng · ul^{-1} and 22.9 ng · $ul^{-1,}$ for the samples at 2 and 5 meters, respectively. We performed sequencing during x hours in the same flow cell used the previous day. Even if we intended to out equimolar quantities of DNA, the total amount sequenced was 204 Mb and 32 Mb for the two samples, emphasizing our difficulties to achieve equal sequencing depths.
As time was not as demanding this time, we used a more complex approach for the analysis, using the "doublepass" option of SqueezeMeta. This includes an additional step of Blastx homology search in these parts of the sequences where either there was no gene predicted, or the putative ORF did not match anything in the nr database, pointing to a possible prediction mistake. We included also the sample taken the previous day at the same location

A summary of the results can be seen in Suppl Figure 4. We obtained 1148 contigs in the assembly (Longest contig,175796 bp) that contained approximately 30% of the reads. These were supplemented with 67186 singletons (unassembled reads). The final set of contigs contained almost

500.000 ORFs, of which more than 400.000 matched some gene in the nr database.

Different sulfur-related genes were found in the three different locations during the first day of sampling. Overall, the metagenome in San Simon Bay included a relatively higher number of sulfur genes, while Samil Beach metagenome included comparatively less genes than those from the other two sampling sites. Interestingly, the relative abundance of each gene was generally higher in the metagenome collected at San Simon Bay. For example, SoxA and SoxB genes, thiosulfate sulfur transferases (2.8.1.1), TauACB and genes responsible for catabolizing sulfonamides (1.14.117) were relatively more abundant in San Simón Bay, suggesting an important presence of bacteria utilizing thiosulfate and bacteria incorporating taurine at this site. Similarly, dehydrogenation of sulfite (1.8.5.6, 1.8.2.1) and sulfate reduction (2.7.7.4, 2.7.1.25, 3.1.3.7, CysND, CysC, CysH) were also relatively more important at San Simon Bay. Overall, results from the first sampling day suggested that microbial communities from San Simon Bay will be of interest for a second and more intensive (water column depth resolution) sampling the following day.

During the second survey, we found interesting temporal and spatial (vertical) differences in sulfur-related genes in San Simon Bay metagenomes (Figure 5). Most of the sulfur-related genes found were relatively more abundant in surface samples (2 m) than close to bottom (5 m). This result may suggest, for example, that bacteria utilizing thiosulfate and bacteria incorporating taurine at this site are relatively more abundant in surface waters. On the other hand, a tendency to higher relative abundance at surface waters on 14thcompared to 12th July was found fo some of the genes (e.g. SoxB, TauACB). This result suggests temporal changes in the relative importance of specific sulfur metabolisms in San Simón Bay.

Overall, utilization of of in-situ design allowed to make an informed decision on the most interesting site at Ría de Vigo to perform an intensive metagenomic survey on sulfur-related genes.

Discussion

Analysis of metagenomic sequencing results is a work-intensive task involving several steps and different software tools. In addition, this must be followed by careful statistical analysis to determine the desired objectives (e.g. differences in functional or taxonomic content of samples, presence of particular genes or organisms, etc). Therefore, this must usually be performed by bioinformatic experts and often requires the usage of powerful computational resources.

To reduce this burden in resources and expertise, we recently developed several software tools that provide a complete solution for all the bioinformatics involved in metagenomics. The SqueezeMeta software is a complete metagenomic pipeline that automatizes all steps of the analysis. It requires minimal user intervention, and therefore it is amenable to all kind of users no matter their bioinformatics expertise, and is capable to work in limited computational resources, even allowing to analyze metagenomes in a laptop.

The second application is the SQMTools software. This is a R library devoted to facilitate the statistical analysis of the SqueezeMeta results and tables. It automatically loads the different results generated by a SqueezeMeta run (e.g. contigs and gene sequences and annotations, aggregated functional and taxonomic profiles, and/or binning results) and puts them into a single R object that can be explored with a set of simple functions allowing plot and chart drawing, performing multivariate analysis, or connecting to other popular analysis packages in microbial ecology.

Nevertheless, the drawback was that users need to be somehow proficient in R usage to take full advantage of the power of this tool. To overcome this limitation, we have developed a third tool to facilitate the usage for all kind of users. This tool, named ShinySQM, has also the benefit of allowing the sharing of results remotely with other users. ShinySQM is an application written using the R's Shiny library that allows loading the tables created by SQMTools, resulting from a SqueezeMeta metagenomic analysis. This tool leverages the

capacities of Shiny to provide an interactive Graphical User Interface (GUI), offering the possibility of visually inspecting tables, creating and exporting customized plots, and performing multivariate analyses with no need of R programming. Shiny offers dynamic reload of the results, so that any adjustment in the input data is immediately translated to the resulting tables or plots. It also permits to load the results in a web server, allowing remote users to interact with the data, thus facilitating considerably the discussion and dissemination of the results.

The combination of these three tools provide a complete solution for all the bioinformatics involved in metagenomics, and together with the availability of portable sequences, opens the way to be able to analyze metagenomes quickly and directly on the sampling spot. To test this capacity, we have sequenced and analysed metagenomes for soils and marine waters.

We have shown that a portable laboratory fitting a medium backpack can be enough to sequence and analyse a medium-size metagenome. This setting is also autonomous. All devices are powered by batteries, thus not needing connection to a stable power source to work. Internet connectivity is not needed, unless the results wanted to be shared with remote users via the web interface provided by ShinySQM. Even in that case, the amount of data needed to be uploaded is tiny.

The weight of the portable laboratory is around 13 kg, so it can be carried by a single person for some time. Obviously, this weight can be shared between different persons and/or put into some wheeled transporter if the terrain allows it. For instance, when sequencing marine samples, we decided to process the data in the base station to take advantage of the filtering devices that were bulky and heavy. But if needed, we could have also included these pieces of equipment in the portable laboratory, powering these with additional batteries and sharing the weight between different persons in the sampling team. In this scenario, however, we have not tested if the ship movement, affecting the stability of the devices, can be an issue.

In laboratory tests devoted to preparing an Antarctic campaign, we have found that the cold conditions severely affect the performance of the equipments, as observed by others (Gowers et al., 2019; Johnson et al., 2017), but we have been able to provide a solution by the usage of thermal insulated boxes in which we put one or several hand warmers, also battery-powered. This is enough to maintain moderately warm conditions that ensure the proper functioning of the instrumental.

When working with substrates like rocks, where microbial colonization is reduced, often we face a problem with the low concentration of DNA present in the samples. We ameliorate this drawback by processing higher amount of sample. For instance, we start using eight tubes with 200 ng of soil each. All these tubes are later collected in a single column, to concentrate as much as possible. Also, we found that the adjustment of the bead beating procedure to lysate the cells is critical. Vigorous beating can help to break more cells, especially if these are embedded in a solid matrix, but can also lead to extensive DNA fragmentation that will hamper the posterior sequencing. Therefore, we advise to use gentle conditions for this step.

Excessive DNA fragmentation is a problem because, in terms of sequencing performance, it is much better to have fewer long sequences than many shorter ones. We will produce more amount of sequence with less degradation of the flow cell. But also influences the preparation of the library. The same amount of DNA, as measured with the Qubit, will result in many more DNA molecules if these are shorter. This will alter the ratio sequence/adapter in the preparation of the library. It will also make very difficult to equalize the contributions of different samples in multiplexing. As said above, equal DNA concentrations can still produce much different number of DNA molecules if one sample is more fragmented than the other(s), thus producing biases in the abundance of sequenced molecules.

Flow cell survival is a real issue, especially in soil samples prone to have substances that can inactivate or damage the pores. In our first tests, the number of available pores were dropping rapidly as sequencing progressed.

This strongly hindered the reusing of the flow cells, therefore increasing costs very much. In our experience, a cleaning/purifying step using magnetic beads for DNA separation of impurities clearly helps to improve the durability of flow cells, and keep the costs of in-situ sequencing to a reasonable extent. Also, it is advisable to check the flow cells as soon as they are shipped, to check their integrity.

Regarding the bioinformatics analysis, there are basically two approaches when studying a metagenome: perform an assembly or co-assembly, or working with unassembled raw reads. The pros of the co-assembly are that it provides a common reference for all the samples, making it easy the comparison, and also provides longer sequences in the form of contigs that are better for the analysis, since they contain several genes that increase the reliability of taxonomic and functional assignments. On the other hand, the lower the amount of sequence, the less complete and comprehensive the assembly. Using raw reads, thus skipping the assembly, has the advantage of using all information available, not letting back any reads. The main drawback is the more demanding computational costs, since this analysis is carried using Diamond Blastx (Buchfink et al., 2015), implying translation and homology searching of the six frames of each read.

For our goal of producing a full analysis in less than 24 hours using a laptop as computing infrastructure, the analysis of raw reads is unfeasible since it will take a longer time. Therefore, we used co-assembly for analysing the data. The assembly produced rather long contigs but only around 25% of the reads were used. To increase this percentage, we used the singleton mode of SqueezeMeta, which remaps the reads to the assembly and includes the unused reads as new contigs. Then the analysis proceeds as usual, relying on gene prediction steps to determine ORFs. Gene predictors lose accuracy when the sequences are noisy, as it is the case in MINion reads, but since at this stage we just wanted a glimpse of the functional profiles to select the most interesting spot, lower precision is acceptable.

For analyzing the second day samples, we decided to refine the previous strategy by using the “doublepass” option of SqueezeMeta. This includes a step in which the predicted ORFs were evaluated by the results of homology searching. If the ORF has a good hit and the hit is complete (unless in the boundaries of the sequence), the predicted ORF is confirmed and kept. Otherwise, it is discarded and an additional blastx search, not relying in gene prediction, is performed in the parts of the sequences with no confirmed ORFs. This method combines the power of the assembly and gene prediction to save analysis time with blastx searches to increase sensitivity.
Also, base calling runs more efficiently using GPUs, saving valuable time. Therefore it is advisable that the computing device (laptop) is equipped with that feature.

In these two experiments, we have shown conclusively that the goal of obtaining metagenomic information in less than one day, summarized in taxonomic and functional profiles, is feasible. We now plan to take advantage of this capacity to quickly monitor the activity of microbial communities.

References

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V, Sirotkin, A. V, Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Breitwieser, F. P., & Salzberg, S. L. (2020). Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btz715

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

De Filippis, F., Parente, E., & Ercolini, D. (2017). Metagenomics insights into food fermentations. In *Microbial Biotechnology*. https://doi.org/10.1111/1751-7915.12421

Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, *34*, 518–524. https://doi.org/10.1038/nbt.3423

Gowers, G. O. F., Vince, O., Charles, J. H., Klarenberg, I., Ellis, T., & Edwards, A. (2019). Entirely off-grid and solar-powered DNA sequencing of microbial communities during an ice cap traverse expedition. *Genes*. https://doi.org/10.3390/genes10110902

Johnson, S. S., Zaikova, E., Goerlitz, D. S., Bai, Y., & Tighe, S. W. (2017). Real-time DNA sequencing in the antarctic dry valleys using the Oxford nanopore sequencer. *Journal of Biomolecular Techniques*, *28*(1), 2–7. https://doi.org/10.7171/jbt.17-2801-009

Lim, Y. W., Cuevas, D. A., Silva, G. G. Z., Aguinaldo, K., Dinsdale, E. A., Haas, A. F., Hatay, M., Sanchez, S. E., Wegley-Kelly, L., Dutilh, B. E., Harkins, T. T., Lee, C. C., Tom, W., Sandin, S. A., Smith, J. E., Zgliczynski, B., Vermeij, M. J. A., Rohwer, F., & Edwards, R. A. (2014). Sequencing at sea: challenges and experiences in Ion Torrent PGM sequencing during the 2013 Southern Line Islands Research Expedition. *PeerJ*, *2*, e520. https://doi.org/10.7717/peerj.520

Luo, W., & Brouwer, C. (2013). Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btt285

Nowinski, B., Smith, C. B., Thomas, C. M., Esson, K., Marin, R., Preston, C. M., Birch, J. M., Scholin, C. A., Huntemann, M., Clum, A., Foster, B., Foster, B., Roux, S., Palaniappan, K., Varghese, N., Mukherjee, S., Reddy, T. B. K., Daum, C., Copeland, A., … Moran, M. A. (2019). Microbial metagenomes and metatranscriptomes during a coastal

phytoplankton bloom. *Scientific Data*. https://doi.org/10.1038/s41597-019-0132-4

Puente-Sánchez, F., Garciá-Garciá, N., & Tamames, J. (2020). SQMtools: Automated processing and visual analysis of ’omics data with R and anvi’o. *BMC Bioinformatics*. https://doi.org/10.1186/s12859-020-03703-2

Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., Nair, S., Neal, K., Nye, K., Peters, T., De Pinna, E., Robinson, E., Struthers, K., Webber, M., Catto, A., Dallman, T. J., Hawkey, P., & Loman, N. J. (2015). Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. *Genome Biology*, *16*(1), 114. https://doi.org/10.1186/s13059-015-0677-2

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H. J., Becker-Ziaja, B., Boettcher, J. P., Cabeza-Cabrerizo, M., Camino-Sánchez, Á., Carter, L. L., … Carroll, M. W. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, *530*(7589), 228–232. https://doi.org/10.1038/nature16996

Tamames, J., & Puente-Sánchez, F. (2019). SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2018.03349

Turingan, R. S., Thomann, H. U., Zolotova, A., Tan, E., & Selden, R. F. (2013). Rapid Focused Sequencing: A Multiplexed Assay for Simultaneous Detection and Strain Typing of Bacillus anthracis, Francisella tularensis, and Yersinia pestis. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0056093

Breitwieser, F. P., & Salzberg, S. L. (2020). Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btz715

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

De Filippis, F., Parente, E., & Ercolini, D. (2017). Metagenomics insights into food fermentations. In *Microbial Biotechnology*. https://doi.org/10.1111/1751-7915.12421

Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, *34*, 518–524. https://doi.org/10.1038/nbt.3423

Gowers, G. O. F., Vince, O., Charles, J. H., Klarenberg, I., Ellis, T., & Edwards, A. (2019). Entirely off-grid and solar-powered DNA sequencing of microbial communities during an ice cap traverse expedition. *Genes*. https://doi.org/10.3390/genes10110902

Johnson, S. S., Zaikova, E., Goerlitz, D. S., Bai, Y., & Tighe, S. W. (2017). Real-time DNA sequencing in the antarctic dry valleys using the Oxford nanopore sequencer. *Journal of Biomolecular Techniques*, *28*(1), 2–7. https://doi.org/10.7171/jbt.17-2801-009

Lim, Y. W., Cuevas, D. A., Silva, G. G. Z., Aguinaldo, K., Dinsdale, E. A., Haas, A. F., Hatay, M., Sanchez, S. E., Wegley-Kelly, L., Dutilh, B. E., Harkins, T. T., Lee, C. C., Tom, W., Sandin, S. A., Smith, J. E., Zgliczynski, B., Vermeij, M. J. A., Rohwer, F., & Edwards, R. A. (2014). Sequencing at sea: challenges and experiences in Ion Torrent PGM sequencing during the 2013 Southern Line Islands Research Expedition. *PeerJ*, *2*, e520. https://doi.org/10.7717/peerj.520

Luo, W., & Brouwer, C. (2013). Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btt285

Nowinski, B., Smith, C. B., Thomas, C. M., Esson, K., Marin, R., Preston, C. M., Birch, J. M., Scholin, C. A., Huntemann, M., Clum, A., Foster, B., Foster, B., Roux, S., Palaniappan, K., Varghese, N., Mukherjee, S., Reddy, T. B. K., Daum, C., Copeland, A., … Moran, M. A. (2019). Microbial metagenomes and metatranscriptomes during a coastal phytoplankton bloom. *Scientific Data*. https://doi.org/10.1038/s41597-019-0132-4

Puente-Sánchez, F., Garciá-Garciá, N., & Tamames, J. (2020). SQMtools: Automated processing and visual analysis of ’omics data with R and anvi’o. *BMC Bioinformatics*. https://doi.org/10.1186/s12859-020-03703-2

Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., Nair, S., Neal, K., Nye, K., Peters, T., De Pinna, E., Robinson, E., Struthers, K., Webber, M., Catto, A., Dallman, T. J., Hawkey, P., & Loman, N. J. (2015). Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. *Genome Biology*, *16*(1), 114. https://doi.org/10.1186/s13059-015-0677-2

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H. J., Becker-Ziaja, B., Boettcher, J. P.,

Cabeza-Cabrerizo, M., Camino-Sánchez, Á., Carter, L. L., … Carroll, M. W. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, *530*(7589), 228–232. https://doi.org/10.1038/nature16996

Tamames, J., & Puente-Sánchez, F. (2019). SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Frontiers in Microbiology*, *9.* https://doi.org/10.3389/fmicb.2018.03349

Turingan, R. S., Thomann, H. U., Zolotova, A., Tan, E., & Selden, R. F. (2013). Rapid Focused Sequencing: A Multiplexed Assay for Simultaneous Detection and Strain Typing of Bacillus anthracis, Francisella tularensis, and Yersinia pestis. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0056093

Acknowledgments

This research was funded by projects TRAITS (PID2019-110011RB-C31) and ROCKEATERS (PID2019-105469RB-C22).

Figures

Figure 1: Approximate timeline of an in-situ metagenomic experiment. Time points in the left side are estimates, and refer to the starting time of the given step.

Figure 2: Recent eruptions in La Palma island, and location of the sampling in the confluence of Teneguia and San Antonio lava flows (Source: http://www.ign.es/resources/docs/IGNCnig/VLC-Teoria-Volcanologia.pdf)

Figure 3: Sampling locations in Ria de Vigo

Figure 4: Taxonomic profiles of Teneguia and San Antonio metagenomic samples. Plots were done using Pavian and the sqm2pavian script of SqueezeMeta.

Figure 5: Functional profiles for sulfur metabolism from samples in La Palma island. Each box shows the abundance (in TPM) of the enzymes involved in this process. Boxes are divided, with left part showing the abundance in 14 July at 2 meters depth, the middle in 14 July at 5 meters depth, and the right in 12 July at 5 meters depth. Plot was done using the interface to PathView included in the SQMtools library.

Supplementary material

	San Antonio	Teneguía
Total reads	60200	33382
Total bases	191.0 Mb	95.4 Mb
Total ORFs	100387	56317
KEGG functions	48225	26483

Suppl Table 1: Amount of sequencing for volcanic samples

Suppl Figure 1: Depth profiles of temperature (red), salinity (blue), irradiance (pink) and fluorescence (green) in the three sampling locations during the first sampling day (12th July 2012) and in San Simón Bay during the second sampling day (14th July 2022)

Suppl Figure 2: Functional profiles for sulfur metabolism from samples in La Palma island. Each box shows the abundance (in percentage) of the enzymes involved in this process. Boxes are divided, with left half showing the abundance in the Teneguia sample, and the right half in San Antonio sample.

Suppl Figure 3: Functional profiles for sulfur metabolism in Vigo Bay, first sampling

Suppl Figure 4: Data from the SqueezeMeta analysis of the samples in Ria de Vigo

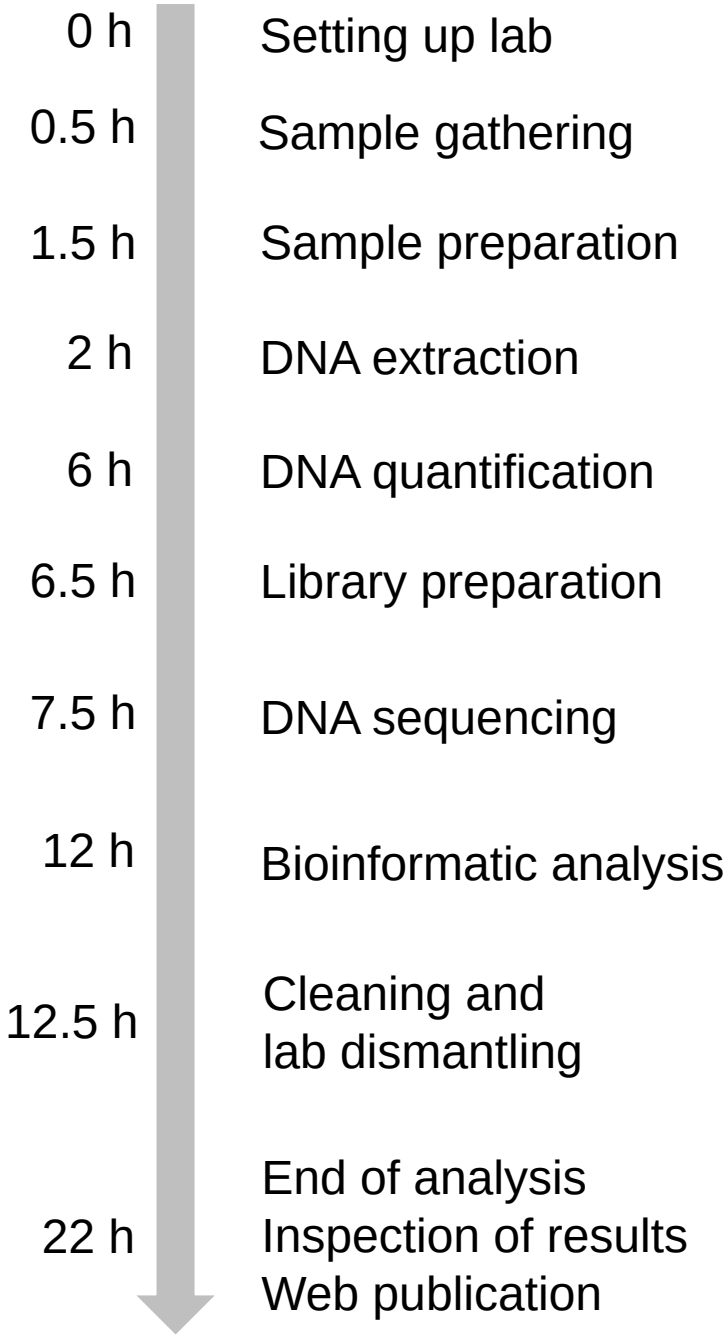

0 h
Setting up lab
0.5 h
Sample gathering
1.5 h
Sample preparation
2 h
DNA extraction
6 h
DNA quantification
6.5 h
Library preparation
7.5 h
DNA sequencing
12 h
Bioinformatic analysis
12.5 h
Cleaning and
lab dismantling
22 h
End of analysis
Inspection of results
Web publication

Figure 1

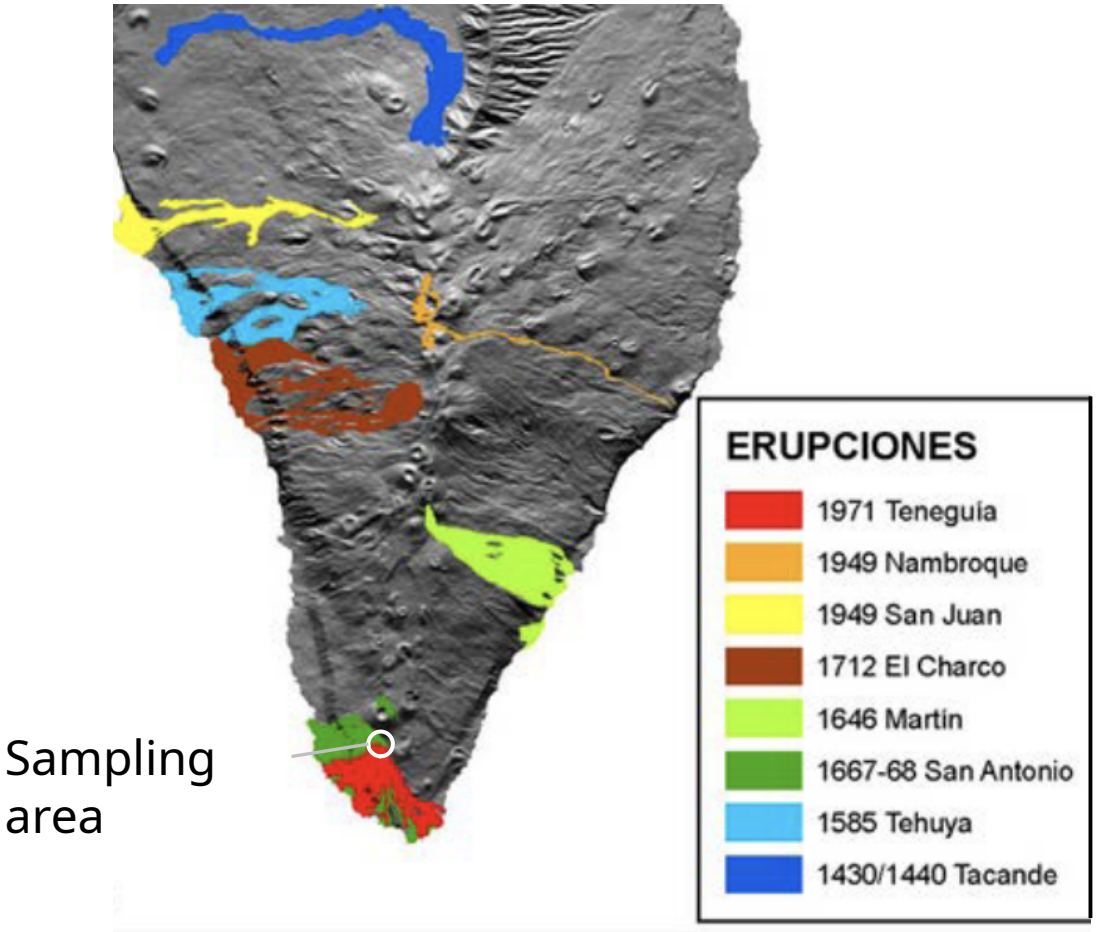

ERUPCIONES
1971 Teneguía
1949 Nambroque
1949 San Juan
1712 El Charco
1646 Martín
1667-68 San Antonio
1585 Tehuya
1430/1440 Tacande
Sampling area

Figure 2

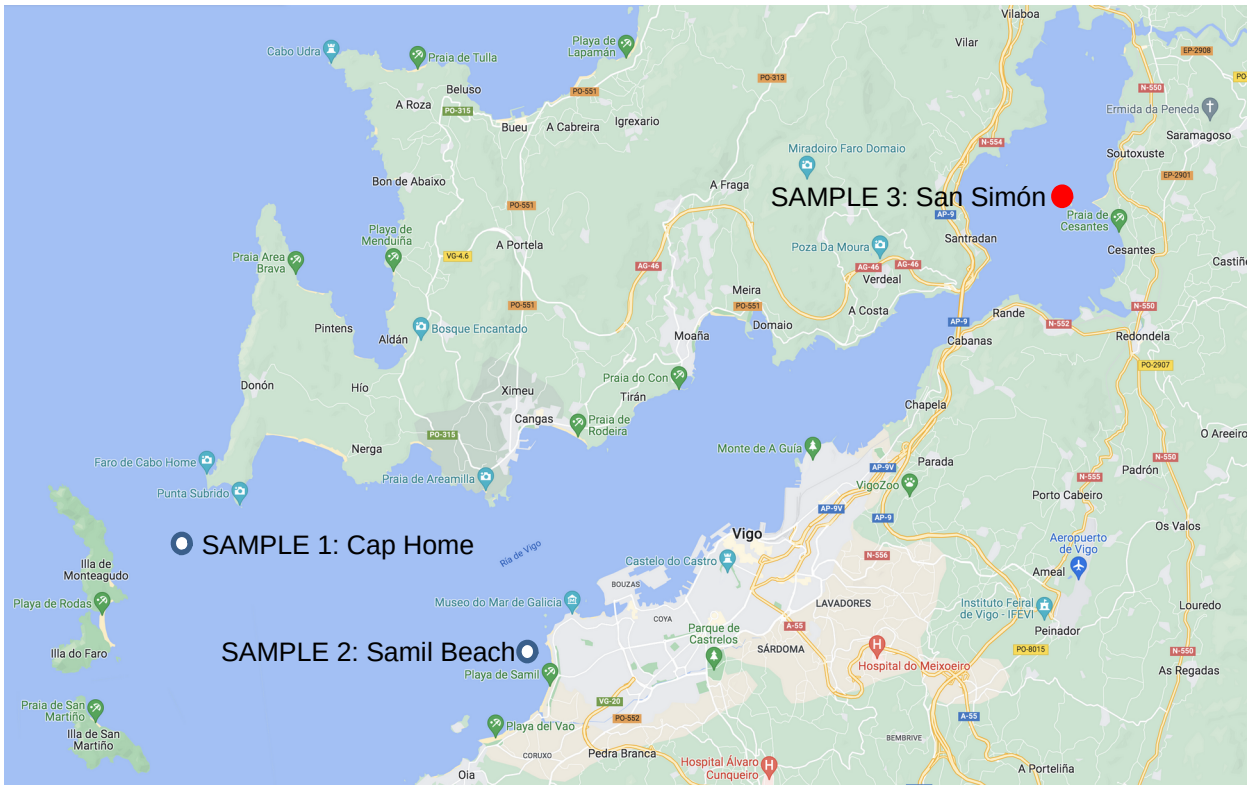
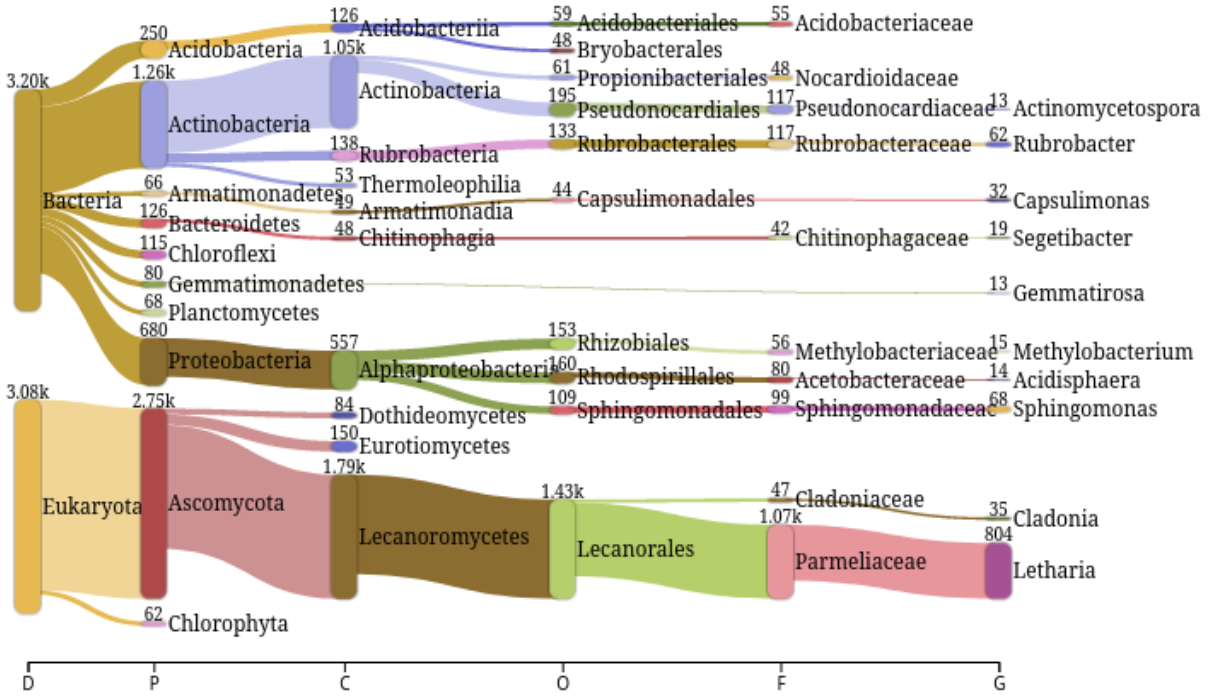
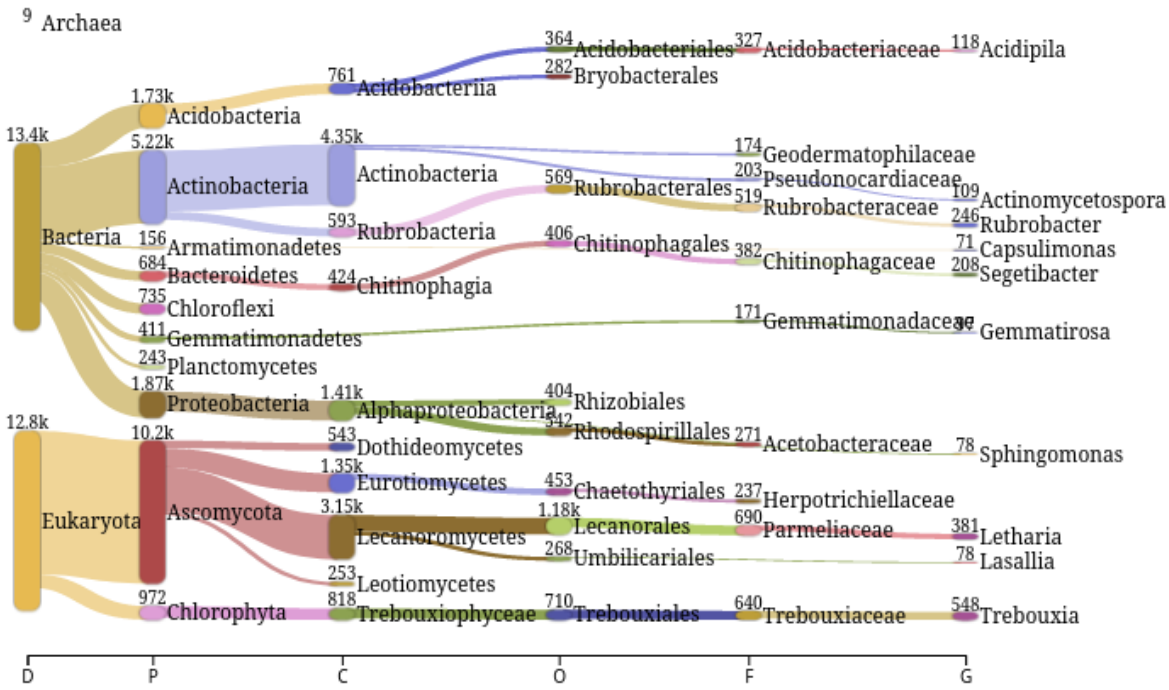

SAMPLE 3: San Simón
SAMPLE 1: Cap Home
SAMPLE 2: Samil Beach
Vilaboa
Vilar
Cabo Udra
Praia de Tulla
Playa de Lapamán
Beluso
A Roza
Bueu
A Cabreira
Igrexario
Ermida da Peneda
Saramagoso
Soutoxuste
Miradoiro Faro Domaio
A Fraga
Bon de Abaixo
Praia de Cesantes
Cesantes
Santradan
Playa de Menduiña
A Portela
Praia Area Brava
Poza Da Moura
Verdeal
Meira
A Costa
Rande
Pintens
Aldán
Bosque Encantado
Moaña
Domaio
Cabanas
Redondela
Donón
Hío
Ximeu
Praia do Con
Tirán
Chapela
Cangas
Praia de Rodeira
O Areeiro
Nerga
Monte de A Guía
Parada
Faro de Cabo Home
Praia de Areamilla
VigoZoo
Padrón
Punta Subrido
Porto Cabeiro
Os Valos
Vigo
Ría de Vigo
Aeropuerto de Vigo
Illa de Monteagudo
Castelo do Castro
Ameal
Playa de Rodas
Museo do Mar de Galicia
LAVADORES
Instituto Feiral de Vigo - IFEVI
Louredo
Parque de Castrelos
Peinador
Illa do Faro
SÁRDOMA
Hospital do Meixoeiro
As Regadas
Playa de Samil
Praia de San Martiño
Playa del Vao
Illa de San Martiño
Pedra Branca
Hospital Álvaro Cunqueiro
Oia
A Porteliña

Figure 3



Teneguía



San Antonio

Figure 4

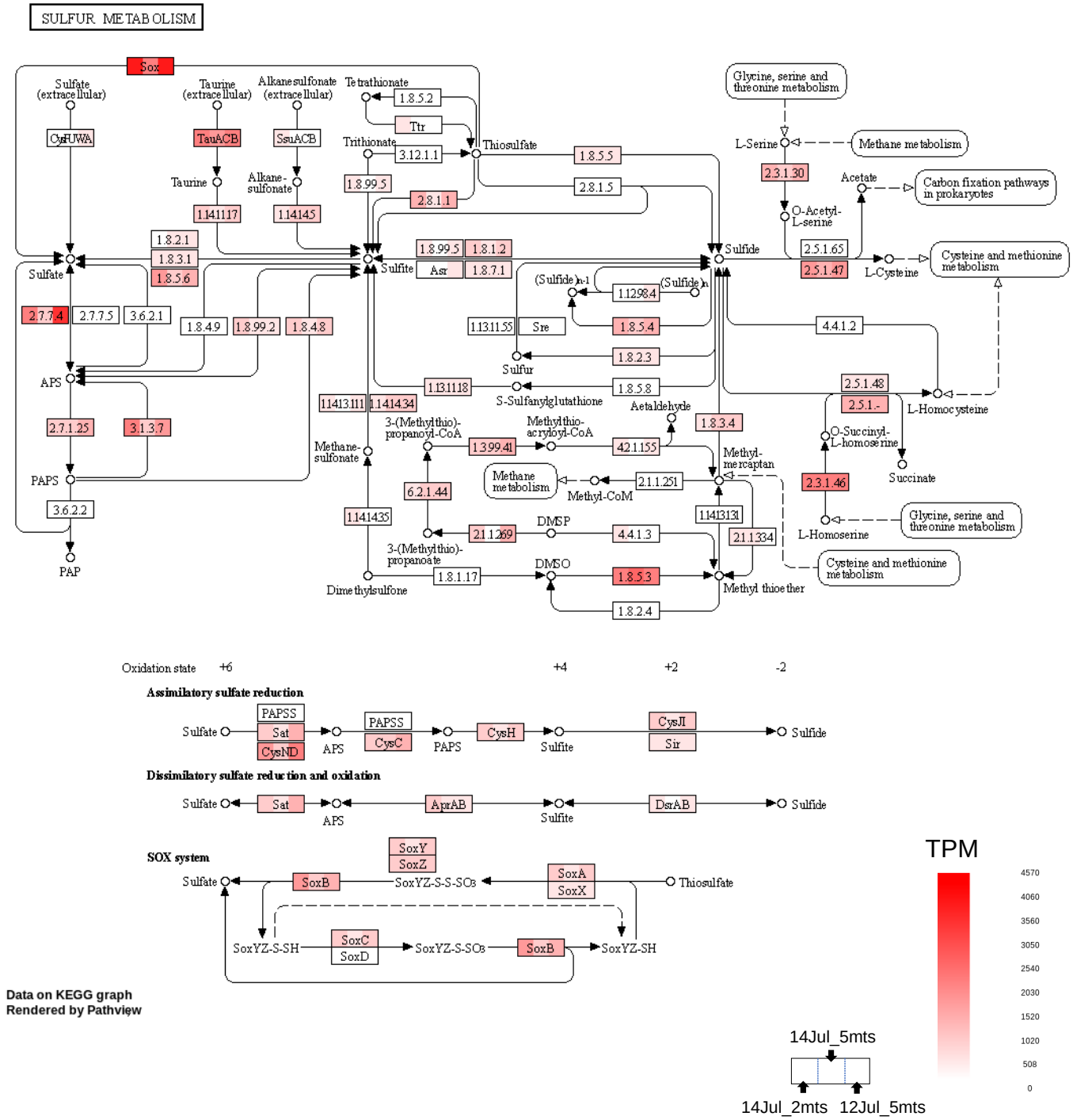

SULFUR METABOLISM
Sox
Sulfate (extracellular)
Taurine (extracellular)
Alkanesulfonate (extracellular)
Tetrathionate
Trithionate
Thiosulfate
Taurine
Alkane-sulfonate
Sulfate
Sulfite
Sulfide
APS
PAPS
PAP
Sulfur
S-Sulfanylglutathione
3-(Methylthio)-propanoyl-CoA
Methylthio-acryloyl-CoA
Methane-sulfonate
3-(Methylthio)-propanoate
Dimethylsulfone
DMSP
DMSO
Methyl-CoM
Methane metabolism
Aetaldehyde
Methyl-mercaptan
Methyl thioether
Glycine, serine and threonine metabolism
L-Serine
Methane metabolism
Acetate
Carbon fixation pathways in prokaryotes
O-Acetyl-L-serine
L-Cysteine
Cysteine and methionine metabolism
L-Homocysteine
O-Succinyl-L-homoserine
Succinate
L-Homoserine
Oxidation state
+6
+4
+2
-2
Assimilatory sulfate reduction
Dissimilatory sulfate reduction and oxidation
SOX system
SoxYZ-S-SO3
SoxYZ-S-SH
TPM
14Jul_5mts
14Jul_2mts
12Jul_5mts
Data on KEGG graph
Rendered by Pathview

Figure 5