How the environment shapes evolution?

Andrey Gizdov, Martin Ricardo Del Rio Grageda, Madlene Haddad

United Kingdom, Mexico, Israel

Jagoda Jablonska & Prof. Dan Tawfik

The Department of Biomolecular Sciences

Abstract

Agricultural development is a key factor for providing resources to our growing population, so maximizing crop yields has been a target of fertilizer use for many decades. The soils which such fertilizers are applied to, are home to large and diverse microorganismal populations. Evolution shapes organisms in response to their lifestyles and environmental niches they occupy. Rapid agricultural development over the past decades and introduction of man-made chemical compounds to soils induced the adaptation of soil-residing microorganisms to live in the presence of these substances, to degrade them and utilize the degradation products for energy and vital resources production. To achieve it, organisms evolved specialized enzymes that are able to boost such chemical reactions. In this study, we aim at proving that chemical changes introduced to soil drive an adaptation which is reflected by the broadened repertoire of special classes of oxygen utilizing enzymes.

We apply state-of-art bioinformatics methods, such as metagenomic data analysis, protein homology detection, combined with a self-developed pipeline.

Introduction

Oxygen is a very important molecule that has enabled life as we know it. The introduction of oxygen to the atmosphere drove immense changes to the living organisms. The Great Oxygenation Event (ca. 2.5 billion years ago) marks the beginning of oxygen being present on Earth. It had a great impact on life, as it gave rise to many enzymes, metabolic pathways and, as a consequence, the life as we know today. It is anticipated that the introduction and accumulation of oxygen in the atmosphere drove the emergence of around 650 new metabolites and 700 new enzymatic reactions. As a result, many organisms switched from anaerobic (oxygen-independent) to aerobic (oxygen-dependent lifestyle), making it the most abundant phenotype.

The aerobic phenotype is pronounced in enzymes (molecular machines) that can catalyze redox reactions using an oxygen molecule as electron acceptor. Compared to anaerobic analogues of

such reactions, reactions with oxygen are much more thermodynamically and kinetically efficient - oxygen acts as a chemical boost.

It is known that organisms to live and proliferate need basic chemical elements C (carbon)-H (hydrogen)-N (nitrogen)-O (oxygen)-P (phosphorus)-S (sulfur) which are the building blocks of life. In order to recruit these elements, organisms need to break down more complex molecules (eg. glucose) with the help of enzymes. Some organisms, soil-inhabiting microorganisms in particular, live surrounded by man made chemical compounds (eg. fertilizers, xenobiotics, pollutants) that can be degraded only using the power of atmospheric oxygen and further used as a source of basic chemical building blocks.

The human population is in constant increase and so is the global food demand.To counter this issue, soils are cultivated with extractive crops which deplete the nutrient reserves, leading to a negative balance of nutrients and a removal of the organic matter from the soil. This removal of organic matter and nutrient rich layer of soil profile causes nutrient depletion, leading to loss of soil fertility. The structure and water holding capacity depends on the availability of nutrients such as nitrogen (N), phosphorus (P), potassium (K) and iron (Fe) to support plant growth. To overcome this nutrient deficiency, farmers have started to use more and more chemical fertilizers. By the end of 2016 the global requirement of chemical fertilizers was expected to reach 194 million tons per year. Apart from fertilizers, ecological catastrophes, such as oil spills, lead to the introduction of certain compounds to the surrounding environment.

Throughout the past decades, besides affecting the development of agriculture, fertilizers unintentionally became an inseparable part of the environment of soil-inhabiting bacteria. Microorganisms evolved multiple strategies to extract basic building blocks of life from these human-introduced compounds. A few examples of commonly used fertilizers include Urea (CH_4N_2O) and Ammonium Nitrate (NH_4NO_3). Does after a long enough exposure to fertilizer evolution come into play and induce the development of appropriate fertilizer-degrading enzymes? This study is directed towards answering this question.

If evolution played a role, in order to boost the break-down process, the bacterial genome will be especially rich in oxygen-utilizing enzymes (an enzyme which uses the oxygen as a reactant), as this is the main catalyst known to be involved in breaking the stable bonds which the fertilizer compounds possess.

Therefore, if the hypothesis is true, the expectation is to find a greater number of oxygen utilizing enzymes in the fertilized soil samples compared to natural untreated soil samples.

Materials and Methods

Input data

As an input data for our predictions we used 235 samples of metagenomes collected from different collection sites all over the world (https://www.ebi.ac.uk/metagenomics/). The nature of collection sites differ and spans from highly developed agricultural fields, contaminated soils to natural unmaintained environments, such as caves and forests.

This data contains information about the geographical area from which the samples come from, and it also contains the DNA sequences of microorganisms that live in the soil, and estimated number of bacterial species in a given sample.

The sequences are encoded in fasta format which is the most common format for representing nucleotide or protein sequences, using single-letter codes. A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line (defline) is distinguished from the sequence data by a greater-than (">") symbol at the beginning. An example sequence in FASTA format is:

```
1   >M04033:2:000000000-AKUBP:1:2112:3113:14621-1:N:0:15_1_253_-
2   CCAATTCATATAAGCTGGACAGGTGGAACAGAGAAAGTATGGGTAATATTGGATAACTTAAATTTTGGTCAATC
```

Oxygen-utilizing enzymes search with Oxyphen

Oxyphen is a software written in Python that uses BLAST program in order to search for oxygen-utilizing enzymes in an input genome (or metagenome). Since a normal computer would take much time to process all 235 samples - about 14 days according to the fact that it processed about 5 samples in 7 hours - the program was run on WEXAC, the Weizmann Institute cluster for extensive calculations, a large-scale supercomputing resource designed to perform large jobs, parallel processing, visualization and scientific applications.

Moreover, in order to analyze data returned by Oxyphen, we wrote a program in Python to gather information from Oxyphen output files. Our program analyzes the following Oxyphen output files:

1) The first type of file analyzed was the input file where we could find all the sequences found in a given sample in FASTA format, from which we extracted the total number of sequences.
2) The second type of file analyzed was results for individual metagenomes (one sample):

```
hit subject id  len eval    cov EC
M04033:3:000000000-AKUBR:1:2108:7350:12279-1:N:0:15_1_136_+ P98000  51.351000000000006  1.29e-09    84      1.9.3.1
M04033:3:000000000-AKUBR:1:2108:26914:12293-1:N:0:15_136_245_-  Q6UIL3  43.333  0.01    86      1.14.99.47
```

Which contains important information about the oxygen-utilizing enzymes that exist in that given sample and various measures of quality. The total number of enzymes in the file was extracted to be used as part of the calculations.

3) The third type of file analyzed was the summary table with results for all of the samples analyzed:

```
ERR1559800_MERGED_FASTQ_CDS_unannotated.txt 265 1.9.3.1,1.14.99.47,1.14.12.25,1.14.17.4,1.14.13.39,
```

4) Finally, the fourth file was a table that contains the predicted organisms identified in a given sample.

```
#OTU ID ERR1527876  taxonomy
244992  1.0 k__Bacteria; p__Actinobacteria; c__Acidimicrobiia; o__Acidimicrobiales; f__C111; g__; s__
810416  1.0 k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhodospirillales; f__Rhodospirillaceae; g__; s__
```

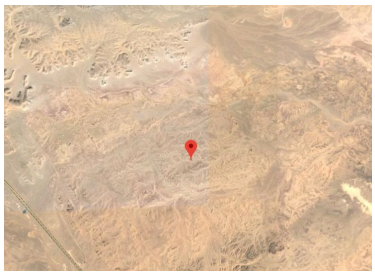
Method for assessing O_2-enzymes enrichment per metagenome

Finally, all the information (number of Oxygen-utilizing enzymes, total number of sequences, number of enzymes and number of organisms) were taken to perform different normalizations:

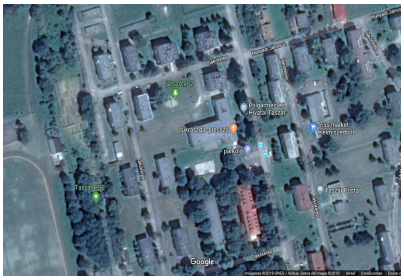
$$a = \frac{Oxygen{-}utilizing\ enzymes}{Total\ number\ of\ different\ Oxygen{-}utilizing\ enzymes} \quad b = \frac{Oxygen{-}utilizing\ enzymes}{Number\ of\ sequences} \quad c = \frac{Total\ number\ of\ enzymes}{number\ of\ sequences\ in\ the\ file}$$

$$d = \frac{Oxygen{-}utilizing\ enzymes}{Number\ of\ organisms}$$

In order to determine the level of agricultural development of the soil where the sample was taken from, we developed and combined two methods. First, the samples were analyzed based on the satellite image of the terrain, provided by Google Maps:



1



2

3

And a different ranking was assigned to them based on the following criteria:

1- Most unlikely to not be intervened by humans.

2- Urban landscapes

3- Field (crops)

Second method was text mining using the publications associated with given samples, where authors provide information about the type of location.

Having these two datasets, we managed to obtain information for every sample whether it comes from agricultural, natural or contaminated soil and used it to analyze data.

Results

In this study we wanted to investigate whether organisms in samples coming from human-impacted environments (contaminated / agricultural) encode more oxygen-utilizing enzymes than organisms found in wild environments that have not been modified.

First, we applied a simple input data quality control to make sure our data is complete. We checked the mean length of the sequences and rejected samples with short reads:

```
C:\Users\admin\Desktop>python pruebaWIS.py
Number of lines:
194777.0
Mean length
195.53
```

When we combined all data listed in a "***Oxygen-utilizing enzymes search***" subsection of Materials and Methods , we obtained table with results (Table 1) that we further analyzed:

ERR1559800	0.363013698630137	0.001360530247411142	0.006196830221227352	1.8150684931506849
ERR1559801	0.3410958904109589	0.0013199745547073792	0.006440839694656489	2.3714285714285714
ERR1559802	0.3219178082191781	0.0014213996249924394	0.006568680820177826	2.7976190476190474

Table 1: Sample id, the numbers a,b,c,d which are explained in the method

The samples used come from 25 different places, for example, the following sample was collected from an oil-contaminated military base in Hungary

Fig 1: Location of soil sample which was collected from an oil-contaminated military base from in Hungary indicated with a blue dot.

After analyzing all the different normalization methods (as described in "***Method for assessing O_2-enzymes enrichment per metagenome"*** subsection***)***, we plotted our number of

oxygen-utilizing enzymes per sample on a world map and designated the type of sample (contaminated, agricultural,natural) with a corresponding color and the number of oxygen-utilizing enzymes per organism represented by the size of the bubble.

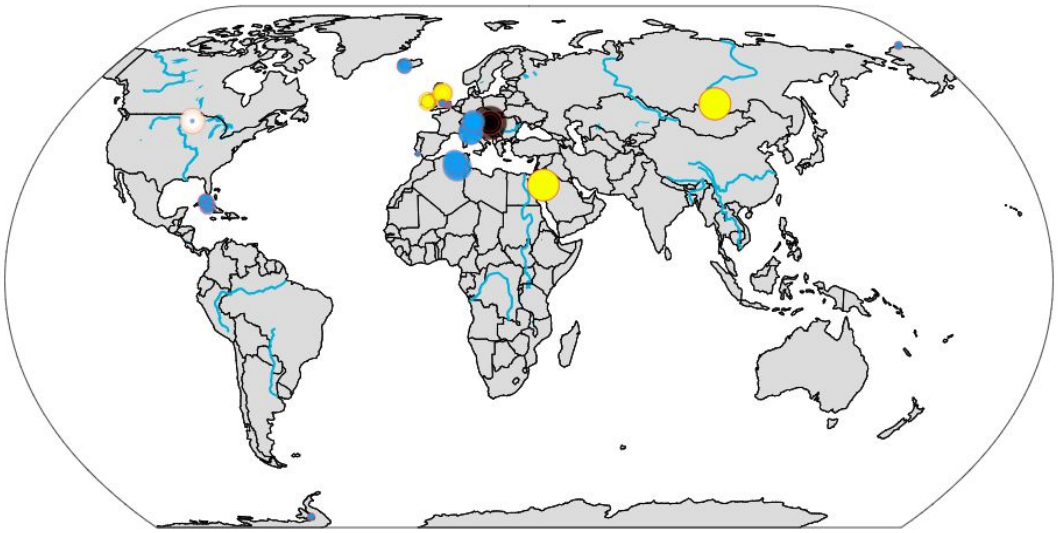


Fig 2: Detection of the number of oxygen utilizing per organism enzymes from different samples and types of soil using a map. The size of the bubble represents the number of oxygen utilizing enzymes per organism, the color of the bubble represents the type of the soil (yellow=agricultural,black=contaminated,blue=natural)

We also analyzed the distribution of the number of oxygen utilizing enzymes in samples collected from aforementioned environments (Fig 3). As can be clearly noticed, a big number of oxygen-utilizing enzymes per organism are more common in the agricultural.

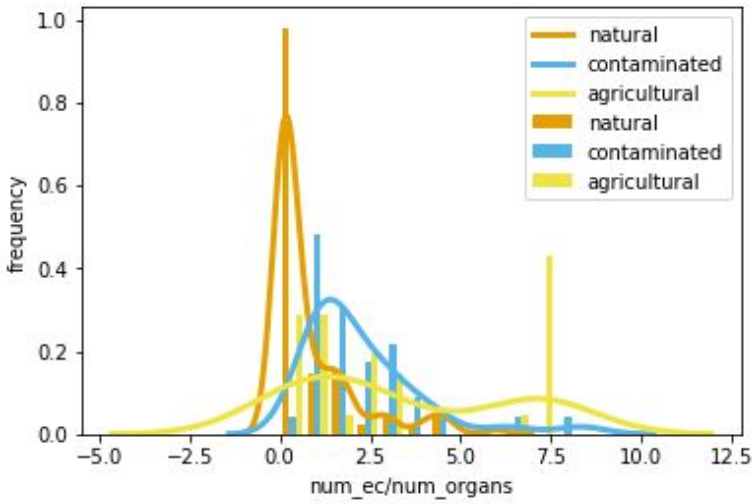


Fig 3: Detection of the number of oxygen utilizing enzymes per organism in natural,contaminated and agricultural soil .The x axis represents the number of oxygen utilizing enzymes per organism while the y axis represents the frequency.

We extracted the most common enzymes from all three environments and included them in Table 2.

EC	NAME	REACTION
		agricultrued
1.10.3.2	Laccase	a benzenediol + a benzenediol + a benzenediol + a benzenediol + oxygen
1.14.19.30	Acyl-lipid (8-3)-desaturase.	
1.11.1.21	Catalase peroxidase	hydrogen peroxide + hydrogen peroxide → H_2O + H_2O + oxygen
4.1.1.39	Ribulose-bisphosphate carboxylase	3-phospho-D-glycerate + 3-phospho-D-glycerate ← D-ribulose-1,5-bisphosphate

Table 2: The names of the EC classes which are the most common oxygen utilizing enzymes in the agricultural soil, and the reactions they are associated with.

As well, the difference in the enzyme types contained in different sample groups:

```
Natural - Agricultural
set()
Natural - contaminated
{'', '1.13.12.22', '1.14.13.130', '1.13.12.13'}
Agricultural - contaminated
{'', '1.13.12.22', '1.14.13.130', '1.13.12.13'}
```

Table 3: This table displays the differences in the types of enzymes (represented by EC class numbers **[5]**) between each pair of sample categories (natural, contaminated, agricultural) - as can be noted, there is no difference in enzyme types between the natural and agricultural soil, while there is a difference between the natural and contaminated soil and between the agricultural and contaminated soil

Discussion

The hypothesis was confirmed even though the results were not very differentiating or clear. Nevertheless, the correlation between oxygen utilizing enzymes and soils either agricultural or contaminated was noted (Fig 2,Fig 3). In that regard, it was also identified that the enzymes in

natural soils versus the ones in agricultural ones were the same, meaning that the diversity there doesn't change (Table 3).Moreover as we see in Table 2 the most common types of Oxygen-utilizing enzymes in the agricultural soil are involved with reactions that have substrates with ring bonds that are characteristic of aromatic compounds present in fertilizers.
In the future, we might be able to analyze more samples using our developed pipeline. Moreover, the analysis of every differentiating enzyme and what role plays in the environment could add more specificity to the study.

References

[1] List of commonly used agricultural fertilizers
https://www.hunker.com/12405093/list-of-common-agricultural-fertilizers
[2] Jabłońska J, Tawfik DS (2019) Free Radical Biology & Medicine, The number and type of O2-utilizing enzymes indicates aerobic vs. anaerobic phenotype.
[3]Muhammad Imtiaz Rashida,Liyakat Hamid Mujawar,Tanvir Shahzad,Talal Almeelbia,b, Iqbal M.I. Ismail, Mohammad Oves(2016)Microbiological Research,Bacteria and fungi can contribute to nutrients bioavailability and aggregate formation in degraded soils
[4] European Bioinformatics Institute (2019) MGNFY website: //www.ebi.ac.uk/metagenomics/
[5] Enzyme nomenclature database (2019), website: https://enzyme.expasy.org/

Acknowledgement

We thank our mentor Jagoda Jablonska and Professor Dan Tawfik, the head of the laboratory as well as all the lab members. We also thank the ISSI program and everyone funding it.