Functional Metagenomic Analysis of the Coffee (*Coffea arabica*) Fermentation

Katerine Vera Pacheco[a], Wilfredo Valdivieso Quintero[a], Andrea Juliana Mantilla-Paredes[a], William Jaimes[b], Jorge Torrado[c], German Zafra[a,d]*

[a]Universidad de Santander UDES, Grupo de Investigación en Ciencias Básicas y Aplicadas para la Sostenibilidad – CIBAS. Campus Lagos del Cacique, Calle 70 No. 55-210, Bucaramanga, Santander, 680003, Colombia.
[b]Penagos Hermanos y Compañía SAS. Bucaramanga, Santander, Colombia.
[c]Telmo J. Díaz y CIA S.A., Hacienda el Roble. Los Santos, Santander, Colombia.
[d]Current address: Universidad Industrial de Santander, Escuela de Microbiología. Bucaramanga, Santander, 680002, Colombia.
gzafra@udes.edu.co

This study focused on studying the influence of the temperature and time on the functional diversity of the microbial populations involved in coffee (*Coffea arabica*) fermentation, using a shotgun metagenomic approach. Fermentations of depulped coffee grains were carried out under controlled and non-controlled temperature conditions for 24 h. Paired-end whole genome sequencing from mucilage samples was performed using an Illumina Hiseq 2x150 platform. Global and specific gene abundance was analyzed using the KEGG orthology (KO). Results showed a predominance of genes involved in carbohydrate and aminoacid metabolism during fermentations. The abundance of genes involved in glycolysis / gluconeogenesis, lactate fermentation and mixed acids were higher during fermentation conducted under non-controlled temperature conditions; however, fermentations carried out at 11 °C induced a significant increase in the abundance of genes involved in the synthesis of aminoacid, lipids and organic acids, as well as protein secretion systems. We concluded that different temperatures and conditions in fermentations produce appreciable changes on the functional potential of both aminoacid and carbohydrate metabolism, especially in the abundance of N-acetyl-lysine deacetylase, pyruvate dehydrogenase and 6-phosphofructokinase genes, which in turn could greatly affect the taste and quality of coffee. This information, together with the results from coffee cupping, provided valuable insights into the role microorganisms involved in coffee fermentation play in obtaining better taste attributes, as well to identify key genes and potential metabolic pathways associated with these special attributes.

1. Introduction

Coffee fermentation is a process carried out by microorganisms, whose main purpose is to remove the remaining mucilage from the coffee pulping process. In Colombia, coffee fermentation is carried out mostly using waterless systems (dry fermentation) in open tanks of various materials. In addition to eliminating the mucilage, the fermentation has a pronounced effect on the final quality of the coffee. This influence may confer special characteristics and added value to coffee, enhancing its flavor and aroma. On the other hand, disastrous losses in quality of the beans may take place if the coffee is over-fermented (Bade-Wegner et al. 1997). Among the different variables influencing this process, the fermentation time, temperature and variations within the microbial populations involved are key to modify the chemical composition of the beans, which ultimately affects the taste and quality of the final product. Although there are reports on the microbial composition of mucilage and the microbial populations involved in the fermentation process (Masoud et al. 2004), most of these studies have used conventional techniques for microbial isolation/count and PCR-based techniques to identify the isolated microorganisms. Conventional microbiological studies applied to this type of fermentations are difficult to carry out, mainly because of the absolute numbers of some taxa are either very high or very low, causing issues with the microbial identification and counts. In addition, the presence of

teleomorph or anamorphic stages make difficult to carry out reliable comparative studies. These type of approaches does not allow to analyze most of non-cultivable populations involved in the fermentation process. Despite there is knowledge about the biochemical conversion of the substrates present in the mucilage and beans, a direct association between the type of microorganisms present and the final organoleptic attributes of the finished coffee has not been stablished (Lee et al. 2015). Thus, since only a small proportion of the microorganisms involved in the fermentation is cultivable, it becomes necessary to use more informative techniques to describe in a greater detail these important aspects of coffee fermentation. Metagenomics, and specially the use of next-generation sequencing (NGS) methods, allow the direct analysis of microorganisms without isolating them, providing also a way to solve the problems mentioned above. Because of the above, this study aimed to provide a better understanding of the microbiome associated with the fermentation of coffee, their functional potential and influence on coffee quality using a metagenomic approach. The study focused on studying the influence of the temperature and time on the functional diversity of microbial population involved in coffee fermentation. This information, together with the results from coffee cupping, would provide valuable insights into the role microorganisms involved in coffee fermentation play in obtaining better taste attributes, as well to identify key genes and potential metabolic pathways associated with these special attributes.

2. Materials and methods

2.1 Coffee fermentation, processing and cupping

The coffee population analyzed in this study corresponded to *Coffea arabica*, var. Caturra. Total sample size consisted of approximately 10 tons of fresh coffee cherries collected from a coffee farm located in Los Santos, Santander, Colombia (06.8628260°, -073.0450050°; 18 °C annual average temperature), in March 2016. Coffee cherries were subjected to a mechanical pulping process in a UDT4 depulper (Penagos Hermanos y CIA, Colombia) and subsequently, two different processes of fermentation (no water added) were carried out during 24 hours: 1) a fermentation under temperature-controlled conditions at 11 °C in a prototype fermenter (2 m height, 1.1 m diameter) provided by Penagos Hermanos y CIA (Colombia), and 2) a fermentation under non-controlled temperature conditions, carried out in a conventional open tiled fermentation tank under the prevailing environmental conditions on the farm (22 to 26 °C). Composite mucilage samples (150 g) were taken in sterile flasks from five different zones and depths of both the open tank and the fermenter, during hours 0, 14, 18 and 24 of fermentation. These mucilage samples were kept at -20 °C until DNA extraction. In addition, 300 g of fermented bean samples from each time point were obtained for coffee production and cupping. These coffee beans underwent the remaining production processing (washing, drying, hulling, sorting, roasting and milling) to obtain a finished coffee product. Subsequently, each brewed coffee was subjected to a cupping process to evaluate flavor notes and overall quality.

2.3 Nucleic acid extraction and sequencing

Metagenomic DNA was extracted from mucilage samples using the PowerSoil® DNA Isolation Kit (MO BIO Laboratories Inc.) according to the manufacturer's instructions. Quantity and quality of DNA was analyzed using a NanoDrop 2000 system (Thermo Scientific). Paired-end whole genome sequencing was performed using the Nextera XT DNA sample preparation kit and the Illumina HiSeq 2500 platform (with the 2x150 paired-end, rapid run mode). This process allows to obtain between 5 and 20 million raw reads from each sample.

2.4 Functional predictive analysis and mapping of metabolic pathways

Raw sequences were checked for sequencing tags and adapters using the FastX-Toolkit. Sequences were then subjected to quality control using the FastQC tool to check data quality. Bowtie (Langmead et al. 2009) was used to remove host (coffee) contaminant sequences based on *Arabidopsis thaliana* (TAIR9) genome. Gene calling, annotations and similarity-based searches of unassembled metagenomic data were carried out using the MG-RAST v4.02 pipeline (Meyer et al. 2008). ORFs and their corresponding protein sequences were identified and annotated using the KEGG orthology (KO), with a maximum e-value of $1e^{-5}$ and a minimum identity cutoff and alignment length of 80 % and 15 bp, respectively. Assigned sequences were retrieved from MG-RAST and analyzed with STAMP v2.1.3 (Parks and Beiko 2010) to calculate the statistical differences between individual samples or groups. To map annotated genes into potential metabolic pathways, metagenomic reads were annotated with KO in MG-RAST and visualized using the KEGG mapper tool (http://www.genome.jp/kegg/tool/map_pathway.html). Statistical analysis of metagenomic profiles were carried out with STAMP v2.1.3, using ANOVA, Fisher´s, Tukey, and Kramer & Welch tests.

3. Results and discussion

After paired-end whole genome sequencing of mucilage samples we obtained the following eight datasets: TNH0: 9,835,879 sequences of 167 bp (on average) for a total of 3.6 Gbp; TNH14: 7,141,248 sequences of 166 bp (on average) for a total of 2.4 Gbp; TNH18: 11,004,665 sequences of 165 bp (on average) for a total of 3.9 Gbp; TNH24: 8,187,076 sequences of 172 bp (on average) for a total of 3.1 Gbp; T11H0: 9,392,753 sequences of 170 bp (on average) for a total of 3.5 Gbp; T11H14: 12,211,152 sequences of 170 bp (on average) for a total of 3.5 Gbp; T11H18: 10,886,246 sequences of 167 bp (on average) for a total of 3.9 Gbp and T11H24: 12,211,152 sequences of 167 bp (on average) for a total of 4.4 Gbp. All these datasets are publicly available in the MG-RAST server under the MG-RAST IDs mgm4716136.3, mgm4716134.3, mgm4716133.3, mgm4716137.3, mgm4716131.3, mgm4716135.3, mgm4716132.3 and mgm4716138.3 respectively, in the static link http://metagenomics.anl.gov/linkin.cgi?project=mgp19659.

3.1 Abundance of genes related to amino acid metabolism

Figure 1 shows the global gene abundance based on KEGG orthology, which showed a predominance of metabolism pathways, especially for carbohydrate, aminoacid and energy metabolism in fermentations conducted under both non-controlled temperature and the 11 °C fermentation (Figure 1).
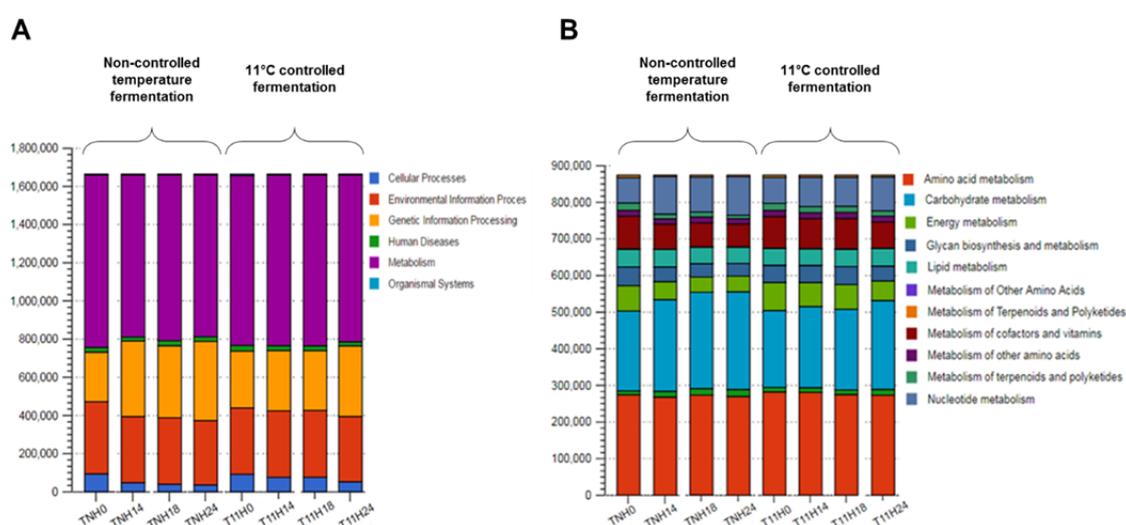

Figure 1. Estimation of gene abundance in coffee mucilage metagenomes using KEGG orthology (KO). A) KO level 1. B) KO level 2 based on the abundance of genes involved in metabolism.

Except for an increase in the abundance of genes related to genetic information processing, no major differences were observed among fermentations. Regarding aminoacid metabolism, the highest proportion of genes detected during both fermentations corresponded to those involved in the alanine, aspartate and glutamate metabolism (ko00250), as well as glycine, serine and threonine metabolism (ko00260), which increased during fermentation (Figure 2). This result agrees with a previous study showing the same aminoacids as predominant in the composition of *Coffea canephora* (robusta) and *Coffea arabica* var. caturra mucilage (Puerta 2011). A higher abundance of genes related to aminoacid metabolism were found in metagenomes from non-controlled temperature fermentation compared to those from the 11 °C fermentation. Even though the most abundant genes corresponded to the metabolism of the before mentioned aminoacids, the most notorious differences between the two fermentations were related to lysine and phenylalanine metabolism genes (Figure 2). Although the abundance of genes involved in the degradation of lysine was similar at the start of both fermentations (4,96 % in non-controlled and 4,79% in 11°C fermentation), it significantly decreased at hours 14 (1 %, $p=0.01$), 18 (0,7 %, $p=0.01$) and 24 (0.3 %, $p=0.008$) during non-controlled temperature conditions; in contrast, during the 11 °C fermentation it remained relatively constant during hours 14 (4.4 %) and 18 (4.5 %) and decreased only by the end of the fermentation (1.6%). On the other hand and similar to lysine degradation genes, a significant decrease in the abundance of genes involved in the metabolism of phenylalanine was observed during the non-controlled temperature fermentation, changing from 13.7 % by hour 0 to 8.2 % and 5.4 % by hour 14 and 18 respectively. Accordingly, this group of

genes was not detected by hour 24, strongly suggesting that microbial metabolism shifted towards other group of genes, such as tyrosine metabolism, by the end of both fermentations (Figure 2).
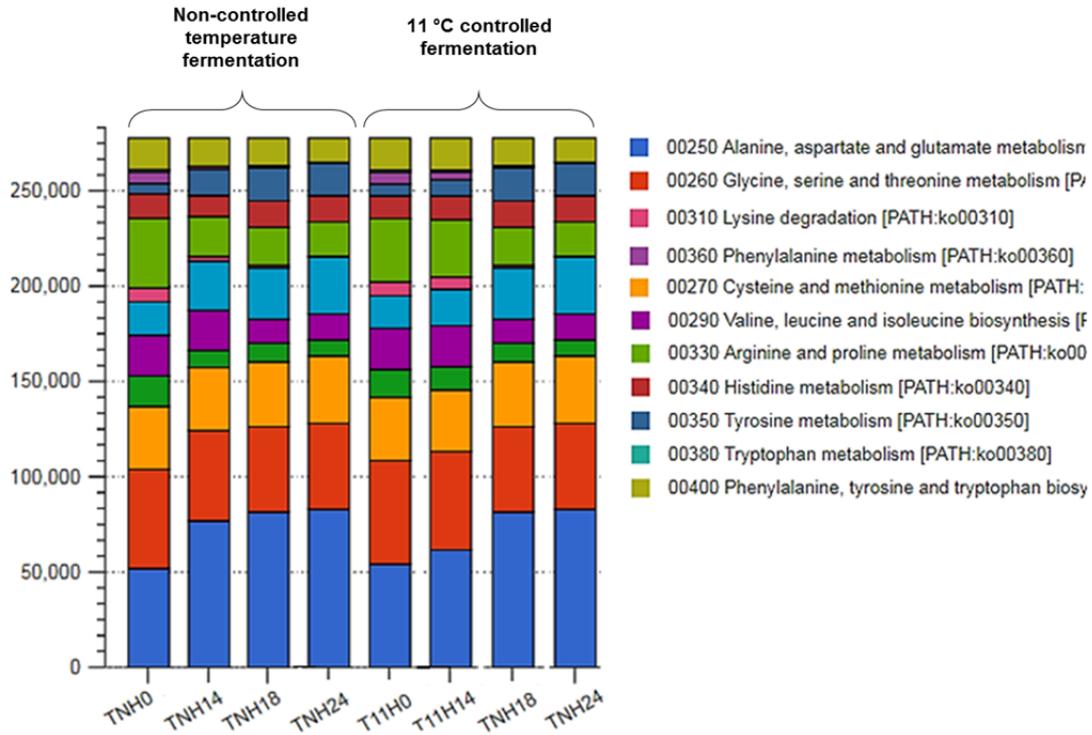

Figure 2. Estimation of the abundance of genes related to aminoacid metabolism in coffee mucilage metagenomes using KEGG orthology (KO). Number denote the pathway entry in the KEGG database.
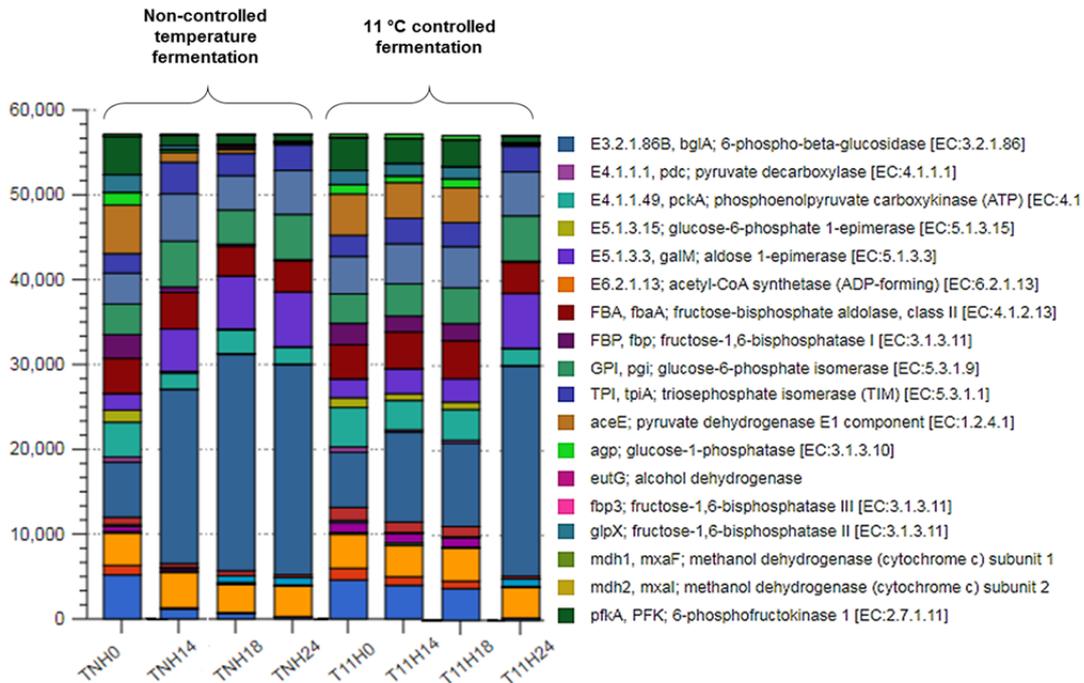

Figure 3. Estimation of the abundance of genes related to carbohydrate metabolism in coffee mucilage metagenomes using KEGG orthology (KO) at level 4 (function level).

3.2 Abundance of genes related to carbohydrate metabolism

Regarding carbohydrate metabolism, the most abundant group of genes were related to glycolysis/gluconeogenesis, pentose phosphate and pyruvate (Entner-Doudoroff) metabolism pathways. This result was expected since these are the three central pathways of the intermediate metabolism of carbohydrates in microorganisms (Varela 2002) which processes carbohydrates and carboxylic acids while also provide metabolic precursors for other metabolic pathways. The abundance of genes involved in several other metabolic pathways significantly increased (i.e. galactose metabolism) or decreased (i.e. inositol phosphate metabolism) during both fermentations. At the function level, we found the gene having the highest abundance was the 6-phosphate-beta-glucosidase [EC 3.2.1.86] (figure 3), an enzyme that acts on β-1,4-glycosidic bonds of oligosaccharides and produces glucose monomers as final product (Casablanca et al. 2011). This enzyme also has transferase and transglucosidase activities, being able to generate products of higher size than the initial oligosaccharides. This is the reason why, along with glucotransferases, glucosidases constitutes the major catalytic machinery for the synthesis and breakdown of glucosidic bonds in microorganisms (Casablanca et al. 2011). The abundance of 6-phosphate-beta-glucosidase gene significantly increased over time in both fermentations, but more noticeably during the fermentation conducted under non-controlled temperature, increasing from 11 % by hour 0 to 36 %, 45 % and 43 % by hours 14, 18 and 24 respectively. During the fermentation carried out at 11 °C this abundance significantly increased only until hour 24 (43 %).

Table 1. Flavor notes found in brewed coffee depending on the temperature and time of coffee fermentation

Fermentation temperature	Fermentation time	Predominant flavor note
Non-controlled temperature conditions (22-26 °C)	Hour 0 (TNH0)	Intense caramel
	Hour 14 (TNH14)	Caramel
	Hour 18 (TNH18)	Coriander grain
	Hour 24 (TNH24)	Sweet chocolate
Controlled temperature conditions (11 °C)	Hour 0 (T11H0)	Bitter chocolate
	Hour 14 (T11H14)	Sweet chocolate
	Hour 18 (T11H18)	Bitter chocolate
	Hour 24 (T11H24)	Caramel

3.3 Abundance of genes during fermentation and relationship to coffee cupping

The complex communities involved in coffee fermentation are responsible of the production of different pectolytic, saccharolytic and lypolityc enzymes which degrade mucilage and generate metabolites, such as acids, alcohols, esters and ketones. These enzymes modifies coffee grains composition, as well as odor, color and pH (Lee et al. 2015). In addition, fermentation temperature, time and water content have a strong effect on the final product. As observed in Table 1, different flavor notes were generated in the final product depending on the type and time of coffee fermentation. In general, the brewed coffees produced from the non-controlled temperature fermentation presented a predominant caramel flavor, whereas chocolate flavors were generated from the controlled temperature fermentation. These flavors are part of the normal range of primary flavors that are obtained in the caturra coffee variety (Stumptown 2017). By analyzing gene abundance variations, their corresponding enzyme activities and the predominant flavor notes in each coffee, we found similarities suggesting a relationship with the results from coffee cupping. For example, the N-acetyl-lysine deacetylase [EC:3.5.1.-] gene was detected only at hour 18 of the non-controlled temperature fermentation, coinciding with the appearance of the coriander grain flavor. The above could indicate this enzyme could be associated with the appearance of this particular flavor. This enzyme is involved in the removal of acetyl groups from lysine residues and is mainly produced by yeasts (Sengupta and Seto 2004). On the other hand, while the proportion of genes involved in the degradation of lysine decreased during 11 °C fermentation, coffee flavor notes also changed from chocolate to caramel notes by the end of the fermentation.

Even though the genes encoding pyruvate dehydrogenase [EC 1.2.4.1] and 6-phosphate-fructokinase [EC 2.7.1.11] were not the most abundant during fermentations, we observed interesting differences when comparing their abundances among the two fermentations. These genes were detected during the first hours of the non-controlled temperature fermentation (hours 0 and 14), but were undetectable by hour 18 or 24. In contrast, both genes were constantly detected during the entire course of the 11 °C controlled temperature fermentation. This could be related to the occurrence of certain flavor notes, particularly chocolate flavor notes.

4. Conclusions

In conclusion, different temperatures and conditions in coffee fermentations produced appreciable changes on the functional profiles, especially those associated with aminoacid and carbohydrate metabolism, which in turn could greatly affect the taste and quality of coffee. Functional diversity was higher in fermentation carried out at higher temperatures, under non-controlled conditions. Several genes, including N-acetyl-lysine deacetylase, pyruvate dehydrogenase and 6-phosphofructokinase appear to be involved in the occurrence of certain specific flavors in coffee, depending on the temperature and time of fermentation. This information, together with the results from coffee cupping, provided valuable insights into the role microorganisms involved in coffee fermentation play in obtaining better taste attributes, as well to identify key genes and potential metabolic pathways associated with these special attributes.

Acknowledgments

References

Bade-Wegner H., Bendig I., Holscher W., Wollmann R., 1997, Volatile compounds associated with the over-fermented flavour defect, In: 17th International Scientific Colloquium on Coffee - Chemistry, Nairobi, 1997. Association for Science and Information on Coffee - ASIC, p 21

Casablanca E., Ríos N., Terrazas Siles E., Álvarez M., 2011, β-glucoside production by thermophilic bacteria indigenous the bolivian altiplano culture, Rev Colomb Biotecnol, 13, 66-72.

Langmead B., Trapnell C., Pop M., Salzberg S.L., 2009, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome biology, 10, R25. DOI:10.1186/gb-2009-10-3-r25

Lee L.W., Cheong M.W., Curran P., Yu B., Liu S.Q., 2015, Coffee fermentation and flavor--An intricate and delicate relationship, Food Chem, 185, 182-191. DOI:10.1016/j.foodchem.2015.03.124

Masoud W., Cesar L.B., Jespersen L., Jakobsen M., 2004, Yeast involved in fermentation of Coffea arabica in East Africa determined by genotyping and by direct denaturating gradient gel electrophoresis, Yeast, 21, 549-556. DOI:10.1002/yea.1124

Meyer F. et al., 2008, The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes, BMC bioinformatics, 9, 386. DOI:10.1186/1471-2105-9-386

Parks D.H., Beiko R.G., 2010, Identifying biologically relevant differences between metagenomic communities, Bioinformatics, 26, 715-721. DOI:10.1093/bioinformatics/btq041

Puerta G.I., 2011, Chemical composition of coffee mucilage according to fermentation ad refrigeration time, Cenicafé, 62, 23-40.

Sengupta N., Seto E., 2004, Regulation of histone deacetylase activities, Journal of cellular biochemistry, 93, 57-67. DOI:10.1002/jcb.20179

Stumptown, 2017, Coffee varieties, Stumptown Coffee Roasters. <https://www.stumptowncoffee.com/varieties/> accessed 2017

Varela G., 2002, Bacterial physiology and metabolism. <http://www.higiene.edu.uy/cefa/Libro2002/Cap%2011.pdf> accessed 2017