UNIVERSIDAD AUTÓNOMA DE GUERRERO

Facultad de Ciencias Químico Biológicas

Facultad de Ciencias de la Tierra

MAESTRÍA EN BIOCIENCIAS

“Análisis de la diversidad bacteriana y funcional de la tuba con un enfoque bioinformático”

T E S I S

QUE PARA OBTENER EL GRADO DE

MAESTRO EN BIOCIENCIAS

PRESENTA:

I.B.Q. Fernando Astudillo Melgar

DIRECTOR: **Dr. Gerardo Huerta Beristain**

CODIRECTOR: **Dr. José Utrilla Carreri**

Chilpancingo de los Bravo, Gro., junio 2018

UNIVERSIDAD AUTÓNOMA DE GUERRERO
FACULTAD DE CIENCIAS QUÍMICO BIOLÓGICAS
FACULTAD DE CIENCIAS DE LA TIERRA
Maestría en Biociencias
Av. Lázaro Cárdenas S/N. Ciudad Universitaria. Chilpancingo, Gro. C.P. 39090. Tel 01 747 47 1 93 10 ext: 3601

APROBACIÓN DE TESIS

En la ciudad de Chilpancingo, Guerrero, siendo los 11 días del mes de diciembre de dos mil diecisiete, se reunieron los miembros del Comité Tutoral designado por la Academia de Posgrado de la Maestría en Biociencias, para examinar la tesis titulada "Análisis funcional de la diversidad bacteriana de la tuba con un enfoque bioinformático", presentada por el alumno Fernando Astudillo Melgar, para obtener el Grado de Maestría en Biociencias. Después del análisis correspondiente, los miembros del comité manifiestan su aprobación de la tesis, autorizan la impresión final de la misma y aceptan que, cuando se satisfagan los requisitos señalados en el Reglamento General de Estudios de Posgrado e Investigación Vigente, se proceda a la presentación del examen de grado.

El Comité Tutoral

Dr. Gerardo Huerta Beristain
Dirección de tesis

Dra. Ma. Elena Moreno Godínez

Dra. Jeiry Toribio Jiménez

Dr. Miguel Ángel Mendoza Catalán

Dr. José Utrilla Carreri

UNIVERSIDAD AUTÓNOMA DE GUERRERO
UAGro
Coordinación del Posgrado de la FCQB
Coordinación 2014-2018

Vo. Bo

Dra. Isela Parra Rojas
Coordinadora del Posgrado de la Facultad de Ciencias Químico Biológicas

UNIVERSIDAD AUTÓNOMA DE GUERRERO
UAGro
Unidad Académica de Ciencias Químico Biológicas
DIRECCIÓN 2014 - 2018

Vo. Bo

Dra. Amalia Vences Velázquez
Directora de la Facultad de Ciencias Químico Biológicas

Este trabajo fue realizado en el Laboratorio de Investigación en Biotecnología de la Facultad de Ciencias Químico Biológicas de la Universidad Autónoma de Guerrero. Se contó con la colaboración del Laboratorio de Biología de Sistemas y Biología Sintética del Centro de Ciencias Genómicas de la Universidad Nacional Autónoma de México y con el Departamento de Microbiología Molecular del Instituto de Biotecnología de la Universidad Nacional Autónoma de México

Bajo la dirección de:
Dr. Gerardo Huerta Beristaín

La codirección de:
Dr. José Utrilla Carreri

La asesoría externa de:
Dr. Adrián Ochoa Leyva
Biol. Filiberto Sánchez López
M.B. José Raunel Tinoco Valencia
Dr. Carlos Ortuño Pineda

Al comité tutoral:
Dra. Ma. Elena Moreno Godínez
Dra. Jeiry Toribio Jiménez
Dr. Miguel Ángel Mendoza Catalán

Durante los estudios en la Maestría en Biociencias, el C. Fernando Astudillo Melgar, recibió beca CONACYT con No. registro: 597135. Y apoyo del proyecto FOMIX-CONACyT-Edo. De Guerrero 249671.
También se agradece el apoyo proporcionado del proyecto DGAPA-PAPIIT IA201518

Agradecimientos.

Al Dr. Gerardo. No tengo palabras para agradecer todo lo que ha hecho por mi, es una excelente persona, confió en mi sin ni siquiera conocerme y eso es algo que no cualquiera hace. Hizo mi estadía en la maestría tan agradable, aprendí tanto de usted y no nada más como profesionista, sino también como persona. Gracias por las enseñanzas, las clases, los momentos de risa y el bully. Espero que siga teniendo mucho éxito Dr. Lo admiro mucho.

A mis asesores externos. Al Dr. José. Gracias por la oportunidad de trabajar en su laboratorio por brindarme las herramientas para el desarrollo del trabajo y por las observaciones y consejos durante el desarrollo del proyecto. Al Dr. Adrián. Por todo su apoyo, por el material y orientación brindada, por permitirme trabajar en su laboratorio y por su asesoría.

A mi sínodo. Al Dr Migue, Dra. Ma Elena y Dra. Jeiry. Por su apoyo, revisiones y consejos que tuvieron en bien darme durante el desarrollo del proyecto. Además gracias por facilitarme equipos y material cuando los necesité.

Biol. Filiberto y al M.B. Raunel. Por su apoyo y asesoría en la realización de experimentos.

A los profesores de seminario. Al Dr. Carlos. Muchas gracias por todo su apoyo, paciencia, consejos y tanto más, es una gran persona, me ayudo tanto en muchos aspectos, es un gran profesionista, lo admiro mucho. A la Dra. Paty. Gracias por todas las veces que me ayudo, siempre que necesite un favor usted tuvo en bien ayudarme, sus consejos, sus pláticas, la admiro mucho. Gracias a ambos son unos excelentes profesores de seminario y amigos.

A la Maestra Daysi. Gracias por brindarme la oportunidad de convivir con su familia, de abrirme las puertas de su casa, las pláticas, los momentos de risa, los consejos, los juegos y por muchas cosas más muchas gracias!. A Canek y Nikte son unos chicos muy geniales, gracias por todos los juegos y aventuras que tuvimos, estoy seguro que serán unos excelentes profesionistas y personas, ya que tienen muy buenos ejemplos en casa, los aprecio mucho.

A mis amigos de generación. No pongo compañeros porque para mí, todos son mis amigos, gracias por todo! Fue una aventura muy bonita, llena de momentos buenos, algunos malos, mucho aprendizaje, convivios, risas, estrés, bueno mucho de todo. Ana, Carlos, Dave, Mary, Augusto, Karen y Yonas, gracias por la oportunidad de haberlos conocido, de brindarme su amistad, espero siempre sigamos en contacto.

A todos los GHB team (Banduqui). Itzel, Muchas gracias por todo, eres una gran amiga y te quiero mucho, compartimos tanto estrés y alegrías, aventuras en donde estuvimos solos trabajando mientras toda la escuela en fiesta, bueno tanto, eres una gran amiga. Sofi, Romy, Angy, Ana, Rous, Alhee, son unas personas muy lindas, gracias por todo su apoyo y cariño, los momentos de locura y risa que me brindaron, la confianza, y la enseñanza de que con esfuerzo se pueden lograr muchas cosas, las quiero mucho. Marino, Leyva gracias por la ayuda y los momentos de risa. Ada, gracias por todas las buenas pláticas, los favores, las aventuras, el bully, pero sobre todo por tu amistad. Itzel, Aydee y Brian, mis "niños" gracias por confiar y seguir este proyecto tan bonito, gracias por permitirme conocerlos y en especial por ser muy buenos amigos. A todos los miembros "nuevos" del equipo, Gema, Ale, Liss, Alberto, Venecia, Genesis y Perla, por la convivencia compartida.

A mi amiga Amparo, por todos los buenos ratos vividos, los consejos, las pláticas, los momentos de estrés, bueno por muchas cosas, eres una gran amiga te quiero mucho.

Al Dr. Javier. Por su apoyo, las pláticas, sus consejos y por el equipo que me facilito durante toda la maestría.

A todos los profesores que conforman esta gran maestría que es Biociencias. Por todo el apoyo y enseñanzas obtenidas durante el transcurso del posgrado. A la Biol. Lupita por el apoyo en el área administrativa.

Atte:

Fernando Astudillo Melgar

Dedicatoria.

El presente trabajo va dedicado principalmente a dos grandes personas que se han sacrificado y luchado por brindarme una buena educación y valores, y a los que les debo todo. A mi padre, al cual le agradezco todos sus consejos, amor, cuidado, apoyo, orientación, los cuales me han servido a lo largo de mi vida. A mi madre, que siempre ha estado cuidándome con sus consejos apoyo y amor incondicional. Son unos excelentes seres humanos, siempre han sido y serán mi motor para salir adelante, un ejemplo a seguir como profesionistas y como personas; los admiro y amo con todo el corazón.

A mis hermanos, mis sobrinos, mi cuñada. Gracias por todo el apoyo brindado, los consejos, los ánimos, los encargos que hacía, las vueltas que a veces los hice pasar, los abrazos que me llenan de energía para seguir adelante, los momentos que se preocupaban junto conmigo y por muchas cosas más gracias. Con este escrito quiero compartirles uno de mis sueños que se realiza porque es un gran logro que también les pertenece ya que me han brindado todo cuanto he necesitado, en especial el cariño y apoyo durante el transcurso de mi vida.

A mis tíos, mi abuelita y mi prima. Que durante mis estancias en Morelos siempre me estuvieron cuidando y al pendiente de mi. Sus apapachos, comidas, paseos, ánimos, consejos, en fin... toda la energía que a uno le hace falta cuando está lejos del hogar base, y digo base porque sin duda ustedes me hicieron sentir en un hogar. Gracias.

Fernando Astudillo Melgar.

ANÁLISIS DE LA DIVERSIDAD BACTERIANA Y FUNCIONAL DE LA TUBA CON UN ENFOQUE BIOINFORMÁTICO.

Table of contents

Bacterial diversity, population dynamics and functional analysis of commercial and laboratory fermented palm wine (Tuba).

Fernando Astudillo-Melgar[1,3], Adrián Ochoa-Leyva[2], José Utrilla[3]* Gerardo Huerta-Beristain[1]*

1.- Laboratorio de Investigación en Biotecnología, Universidad Autónoma de Guerrero, Chilpancingo, México.

2.- Departamento de Microbiología molecular, Instituto de Biotecnología- Universidad Nacional Autónoma de México, Cuernavaca, México.

3.- Programa de Biología de Sistemas y Biología Sintética, Centro de Ciencias Genómicas - Universidad Nacional Autónoma de México, Cuernavaca, México.

*Co-corresponding authors

e-mail: utrilla@ccg.unam.mx, hbgerardo@gmail.com

Abstract.

Palm wine is obtained by fermentation of palm tree sap recollected from palm spathe cut. In the pacific coast of Mexico, palm wine is called Tuba and consumed as a traditional fermented beverage. Tuba has different empirical applications as an auxiliary in gastrointestinal diseases, vitamins and minerals sources. In the present study, a next generation 16S sequencing approach was employed to analyze bacterial diversity and population dynamics during the fermentation process of Tuba, both in laboratory controlled conditions and commercial samples from local vendors. Taxonomic identification showed that *Fructobacillus* was the main genus in all the samples, following for *Leuconostoc*, *Gluconacetobacter*, *Sphingomonas* and *Vibrio*, respectively. Alpha diversity analysis demonstrated variability between all the samples. Beta diversity grouped the bacterial population in according to the collection origin of the sample. Metabolic functional inference showed that the members of the bacterial communities may present the vitamin, antibiotic, amino acid and antioxidant biosynthesis genes. Additionally, we further investigated the correlation between the predominant genera and metabolic characteristics of this beverage. This study will provide the basis for the identification of functional characteristics and the isolation of native strains that may serve as probiotics and that allow standardization of process of Tuba production.

Keywords: Tuba, fermented beverage, bacterial diversity, functionality, massive sequencing.

Introduction.

A wide variety of fermented food products such as yogurt, alcoholic beverages, bread and sauces are produced worldwide. During the production process of these fermented foods different microorganisms contribute to the organoleptic and biochemical characteristics (Tang et al. 2017). Recent studies in fermented food have shown that microbial ecology aspects such as diversity, their spatial distribution and ecological interaction, have a strong influence on metabolic production and chemical composition (Escalante et al. 2015). Bacterial consortia interactions in fermented foods promote process of polymer degradation and production of metabolites of interest such as alcohol, aromatics, acetate, lactate among others that contribute to functional and organoleptic properties (Tamang et al. 2016).

Palm wine is a traditional beverage made using the sap collected from palm trees. It is consumed in different parts of the world, in Africa it is known as "legmi", in South India as "kallu", while in Borneo it has the names of "bahar" and "goribon" (Velázquez-Monreal et al., 2011). The differences among these beverages are the production process, the coconut tree species and the plant part where the sap is collected (Santiago-Urbina & Ruíz-Terán 2014). In Mexico, several traditional fermented beverages are produced such as pulque (Escalante et al. 2016), pozol (Díaz-Ruíz et al. 2003) and Tuba (De la Fuente-Salcido et al. 2015). Tuba was brought to Mexico by Philippine influence during the Spanish colonial period. This beverage is produced in the southern pacific coast of Mexico (Guerrero, Colima, Michoacan states). It is obtained from the sap of the inflorescences of *Cocos nucifera* L and it is consumed as a traditional beverage empirically used as an aid in gastrointestinal problems and as a rehydration drink (Velázquez-Monreal et al. 2011; de la Fuente-Salcido et al. 2015).

The importance of bacteria in fermented foods has promoted the application of different strategies to analyze the bacterial diversity and role during elaboration of fermented products. The use of massive sequencing technologies together with recent bioinformatics methods, such as QIIME for diversity analysis (Navas-Molina et al. 2015; Caporaso et al. 2011) and PICRUST for functional inference (Langille et al. 2013), have increased the taxonomic and functional information of uncultured bacterial communities in different ecosystems (Filippis et al. 2017). However, those methods have been used mainly in projects such as Human Microbiome and Earth Microbiome (Creer et al. 2016). Nevertheless, in the food area, the applications of them are limited. Some studies in traditional Asian liquors and sauces have established a correlation between microbial diversity and organoleptic properties, increasing the information about bacterial communities in Asian products such as Yucha (Tang et al. 2017; Zhang et al. 2016) and some Mexican traditional beverage as Pulque (Escalante et al. 2008).

Here, we study the fermentation profile, population dynamics and bacterial diversity of Tuba produced in the Guerrero coast of Mexico. We sampled sap that was fermented under controlled conditions and sampled commercial Tuba. Using 16S amplicon sequencing and metabolic characteristics, we were able to analyze the diversity and infer functionality of bacterial communities present in all tuba samples. This work provides a basis for the further functional characterization of Tuba in its production process, probiotic potential and other functions as antibiotic and antioxidant biosynthesis.

Material and Methods.

Sample collection.

Sap samples were collected from three visibly healthy palm trees in a rural area in Acapulco, Guerrero, Mexico. Commercial Tuba samples were obtained from four different artesian producers in diamante zone from Acapulco, Guerrero Mexico (Figure 1). The climatological conditions of the samples collection site at the sampling day are described in table 1. Samples were transported in sanitized coolers to the laboratory for fermentation and analysis. The sap from palm trees was tagged with the following code a “P” followed by the number of the palm tree and “T” which means the fermentation time (i.e. P1T0). Commercial samples were tagged using the letter L followed by a consecutive number that symbolize the number of the establishment where each sample was obtained.

Fermentation in laboratory controlled conditions.

Each sap sample (100 mL as working volume) was fermented in four 250 mL Erlenmeyer flasks corresponding to 0, 12, 24 and 35 hours of fermentation. They were incubated at 30°C and 100 rpm of shaking speed in an orbital incubator. Samples were centrifuged (4000 rpm x 15 min) and the pellets were used for DNA extraction, while the supernatants were stored at -20°C for further analysis.

Metabolic composition characterization.

Sucrose, glucose, fructose, water-soluble proteins, acetic acid, ethanol and pH.

Sugars, organic acids and ethanol from laboratory fermented and commercial samples were quantified using two HPLC methods following column manufacturer conditions. Glucose, fructose, sucrose and xylose were quantified using an Aminex HPX-87P (Biorad) column with an IR detector. Acetate and ethanol concentrations were measured using Aminex HPX-87H (Biorad) column and a UV 210 nm detector. Water-soluble proteins were measured by Bradford method modified by Fernández & Galván, 2015. The pH was measured using a potentiometer with 1 mL of the sample.

16S amplicon library preparation and sequencing.

The DNA extraction from all the samples was performed using the ZR Soil Microbe DNA MiniPrep™ kit (Zymo Research) according to the manufacturer protocol. The DNA was quantified using Qubit Fluorometric Quantitation (Thermo Fisher Scientific). 12.5 ng of total DNA was used for PCR of amplicons of the V3-V4 regions of the 16S rRNA ribosomal gene (Table 2) as described by the Illumina Protocol. All the PCR products were purified (AMPure XP beads - Illumina products) and quantified (Qubit). Finally, all the libraries were sequenced by Illumina MiSeq.

Bioinformatics and Statistical analysis

The sequences were analyzed using QIIME version -1.9.1software (Caporaso et al. 2011) in Python 2.7. The total sequences were clustered using UCLUST into OTUs tables (operational taxonomic units) using the Greengene database (GG 13_8_otus) as reference with a range of 97% of similarity and using the closed system with the command pick_closed_reference_otus.py. Taxonomy summaries including relative abundance data were generated using summarize_taxa.py, plot_taxa_summary.py and plot_taxa_through_plots.py commands. In all the cases, we used the data filtering option of 0.01% in abundances because it is reported that filtering data base decreases the estimation error (Kuczynski et al. 2012; Navas-Molina et al. 2015).

Alpha diversity was evaluated using the function of alpha_rarefaction.py from QIIME, that calculate alpha diversity on each sample in an OTUs table, using a variety of alpha diversity metrics as Shannon-Wiener index, Simpson index, Otus_observed and Chao1 value. Each metrics result were analyzed by ANOVA applying the Tukey-Kramer test (0.95 confidence interval) to estimate significance difference between the samples. Beta diversity was calculated by beta_diversity_through_plots.py, a workflow script for computing beta diversity distance matrices (UniFrac unweighted method) and generating Principal coordinates analysis (PCoA) plots from QIIME.

The normalized OTUs table (0.01% abundance filter) was used to estimate functional features present in the samples, using PICRUSt version 1.1.0 (Langille et al. 2013) and the Greengenes databases 16S_13_5 and KO_13_5. The OTUs table was normalized to obtain the metagenomic functional predictions at different hierarchical KEGG levels using normalize_metagenomes.py, predict_metagenomes.py and categorize_by_function.py scripts of the same software.

For the statistical studies of the functions, we used STAMP (Statistical analysis of taxonomic and functional profiles version 2.1.3), through ANOVA analysis applying the Tukey-Kramer test (0.95 confidence interval) to evaluated gene abundance of each function. R statistical program (version 3.3.3) was used to make plots using "ggplot2" and "dplyr" libraries.

Results.

Sample Composition

To determine the microenvironmental conditions that affect the microbial communities and metabolic characteristics of the Tuba samples, we evaluated the sugars (sucrose, glucose and fructose), water-soluble proteins, ethanol and acetic acid concentrations as well as the pH value (Supplementary Table 1S). Tuba P1 was the sample with the highest concentration in glucose and fructose with 61.4 and 47.3 g/L respectively at 12 hours, 4.7% (v/v) in ethanol and 6.0 g/L in acetate at 35 hours (Figure 2A). Tuba P2 was the sample with lowest concentration of monosaccharides at the beginning of the fermentation and high sucrose concentration (121.7 g/L), however, at the last fermentation time the ethanol and acetate concentrations were low with 3.5 g/L and 0.6% (v/v) respectively (Figure 2B). Tuba P3 showed the highest concentration of ethanol (5% v/v) at the end of the fermentation, nevertheless, the glucose and fructose concentration were 39.8 and 29.1 g/L respectively at 12 hours (Figure 2C). The pH values in Tuba P1, P2 and P3 decreased from 3.7 to 2.8 during the fermentation process. The water-soluble protein concentration of the Tuba samples showed low values from 0.006 to 0.01 g/L. In the case of the commercial samples, all of them presented different composition values, nevertheless they had an average values of 40.5 g/L of sucrose, 40.0 g/L of glucose, 42.53 g/L of fructose, 1.6 g/L of acetic acid, 0.1% (v/v) of ethanol and a pH of 4 (Figure 2D).

Taxonomic classification.

A total of 302,398 sequences were obtained from the Tuba amplicon libraries, with an average of 75,594 sequences per Tuba fermented under controlled conditions (distributed as follows, for the Tuba P1 74,860 reads were obtained; for the Tuba P2, 75,623; for the Tuba P3 76,298) and the four commercial samples had an average of 75,617 sequences. A total of 123 OTUs were detected in all Tuba samples. However, filtering data base with 0.01%

relative abundance filter, the OTUs were reduced to 28 as the more abundance. The taxonomic identification was elaborated using the last filter mentioned, which demonstrates the 10 most representative genera of the 16 Tuba samples (Figure 3). The genera that predominate in all the samples were *Fructobacillus, Leuconostoc, Gluconacetobacter, Sphingomonas*, *Vibrio* and some genera of the Enterobacteriaceae family. Additionally, analyzing the Enterobacteriaceae populations with the lower abundance we found genera as *Erwinia, Klebsiella, Serratia* and *Cronobacter* (Supplementary Figure 1S). The population dynamic had a similar trend in Tuba fermented in controlled conditions but with different percentage in the abundances; we observed an increase of lactic acid bacteria (LABs) until 24 h, acetic acid bacteria (AABs) and some proteobacteria as *Sphingomonas* through the fermentation time and a decrease of *Vibrio* genus (Figure 3).

Diversity analysis.

Alpha diversity tests were performed using the OUTs table obtained with the 0.01 % filter and grouped according to the origin of the sample. We observed a similar behavior in all the four analysis, that means, no matter what base-priority was in the analysis as richness (observed_otus), dominance (Simpson), equity (Shannon index) or singletones (Chao1 value) it did not affect diversity results (Supplementary Figure 2S). Tuba P1 was the most diverse with the highest values in the four diversity index, then Tuba P3 and commercial Tuba samples had similar index values, and finally Tuba P2 was the least diverse with the lowest values. After of ANOVA statistical analysis, we found that in Chao1 and Observed_otus tests Tuba P2 was the only showing significant difference. Nevertheless, in Shannon and Simpson index the four groups showed significant difference among each other (Table 3).

Beta diversity with Unweighted UniFrac distance was determined using the 0.01% filter. We did not observe groupings by fermentation time (Figure 4A) however, a grouping was observed by origin of the samples (Figure 4B). In the graphic of origin of the sample we also observe a grouping by quadrant of the all the Tuba samples, however Tuba P2 showed the greatest dispersion in the data, which indicated a big difference between the fermentation times in Tuba P2. Similar effect is observed in Tuba P1 where two fermentation times (0 h and 35 h) show similar beta diversity values compared to commercial samples and Tuba P3. Otherwise, the samples, which were in the same quadrant as Tuba P3 and the commercial Tuba, were considered strongly related (Figure 4).

Functional inference

After diversity distribution analysis we sought to understand the functionality of the bacterial community in Tuba fermentation, therefore we used PICRUSt algorithm to predict the metagenomic profiles of the samples. Initially, we obtained functional characteristics of the 3 KEGG levels (Level 1: general cellular functions, Level 2: Specific functions i.e. different metabolism, and Level 3: Specific pathway associate with specific function) (http://www. genome.ad.jp/kegg/). We limited our analysis to the level 3 and we discarded elementary cellular functions such as replication, translation, and functions associated with human diseases (cancer) or poorly characterized functions, to analyze specific genes related with functions of biotechnological relevance. Considering the 328 registered functions on KEGG, we found the 19 most abundant functions associated with carbohydrates metabolic process, vitamins, amino acid, antibiotics and antioxidant molecules biosynthesis (Figure 5), this suggested that the production of those compounds may be taking place during Tuba fermentation. After an ANOVA test, we found functions without significant difference as the carotenoid biosynthesis (Figure 6A), this means that no matter what is the sample origin, this function may have present at the same gene abundance in the four groups. Otherwise, there were functions with significant difference, such as peptidases biosynthesis that had more gene abundance in Tuba P2 samples (Figure 6B). Each sample had more abundance in genes associated with a specific function, for example, antioxidant, antibiotic compounds, and folate biosynthesis in Tuba P3, lipopolysaccharide biosynthesis and Lysine genes in Tuba P1, finally the 4 commercial Tuba samples may have bacteria with genes associated mainly with folate biosynthesis and peptidases. Our study allowed to analyze if some of bacterial genera found in Tuba may had gene associated with enzymes of carotenoid biosynthesis, we observed that *Sphingomonas and Gluconacetobacter* had more abundance percentage in the enzyme *15-cis-phytoene synthase* (Figure 7).

Discussion.

In the present study, we carried out for the first time the identification of bacterial diversity, the fermentation dynamics in terms of bacterial populations and metabolic changes during Tuba fermentations comparing between laboratory controlled conditions and commercial samples. This comparison was realized by a combination of metabolic analysis and 16S amplicon sequencing during the Tuba fermentation, as well as to infer functions of biotechnological interest that the Tuba may present during the fermentative process.

We found that the average of total sugar concentration in sap of the palm trees in the Tuba samples was 130 g/L. Where it contained 77.06% of sucrose, 12.81% of glucose and 10.15% of fructose, without presence of xylose. In a study with sap of *Phoenix dactylifera* was reported that it contained 95.27 % of sucrose, 2.51% glucose and 1.61% of fructose with a neutral pH of 7-7.4 (Santiago-Urbina & Ruíz-Terán 2014). These results suggested that the sap composition is dependent of the palm type. The concentration of sucrose from the sap samples at the start of the laboratory controlled fermentation was high, from 85 g/L to 121 g/L (Supplementary Table 1S). Then, after 12 hours of fermentation for Tuba P1 and P3, and after 24 hours for Tuba P2 the concentration of sucrose was reduced, increasing the concentration of glucose and fructose, presumably by invertase-mediated hydrolysis. The different behavior of Tuba P2 sample may be related with low sucrose hydrolysis, delaying the fermentation process. This may be caused by a lower yeast abundance than other samples, the yeast abundance was not measured in this study, but we did not identify any ethanologenic bacteria, then we may attribute all the ethanol production to yeasts present in the samples. Also, the high concentration of sucrose may also cause retro-inhibition of the invertase enzyme, and the pH values may reduce its catalytic activity (Hsieh et al. 2006; Goosen et al. 2007). Hence, we can only propose that the hydrolysis of sucrose is associated with the presence of *Fructobacillus* and *Leuconostoc* genera, because those microorganism present the genes that codes for the invertase β-fructofuranosidase (Supplementary Figure 3S), a further study is needed to show yeast abundance in Tuba samples and its implication in sucrose hydrolysis.

In the laboratory controlled fermentations we observed near complete sucrose hydrolysis and lower ethanol production compared to other fermented beverages such as pulque. Pulque shown an absence of sucrose hydrolysis, high ethanol concentration and a high abundance of ethanolic bacteria such as *Zymomonas mobilis* (Escalante et al. 2008), that genus was not found in our work. Thus, these results suggest that the composition of the bacterial community in Tuba play an important role in the hydrolysis of sucrose at the start of the fermentation. These characteristics are related with the bacterial diversity, because several bacterial genera present in the sap has different metabolism and regulation types that in consequence may inhibit or delay the fermentative process (Tamang et al., 2016).

We found the 10 more abundant bacterial genera that belong to three main groups, lactic acid bacteria (LABs), acetic acid bacteria (AABs) and proteobacteria (Figure 3). It has been reported that LABs are the main antibiotic and folate producers in fermented products (De la Fuente-Salcido et al. 2015; Rossi et al. 2011) both functions have an important impact on human health. Moreover, some LABs reported in here such as *Fructobacillus* and *Leuconostoc* genera are similar phylogenetic and metabolically, nevertheless, *Fructobacillus* is unable to produce ethanol, redirecting the carbon flow to the production of lactate, (Endo et al. 2015). Other genera found was *Lactococcus* that produce more lactate than ethanol (Makarova et al. 2006).

The acetate production is related with the abundance of AABs such as *Gluconacetobacter* and *Acetobacter* genera that was found in all samples. Interestingly, sample P1 showed the higher abundances of *Acetobacter*, which contributed with the acetate and ethanol production in comparison with the Tuba P2 and P3. Nevertheless, we observed a smaller abundance of the AABs in the commercial samples; contributing with a lower acetate and ethanol concentrations with respect to the laboratory fermented samples. This result is in agreement with other studies, where the authors propose that the growth of the AABs of the *Gluconacetobacter* and *Acetobacter* genera is dependent on the presence of acetate and ethanol in the environment (Lisdiyanti et al. 2003). Other researches have stablished these genera as the main acetate producers in products from fruit fermentation (Dellaglio et al. 2017).

Both *Vibrio* genera and Enterobacteriaceae family (both proteobacterias) were detected in all the analyzed Tuba samples. *Vibrio* have been reported as a "natural" pollutant of fermented products (Lee et al. 2015). The abundances of these bacterial groups was reduced through the Tuba fermentation process, this abundance in commercial Tuba samples was similar with the abundance to the initial fermentation points. Finally, we observed a relation between the increase of the abundance of LABs and the decrease of Enterobacteriaceae family. It has been shown that the secretion of peptidases by LAB and AAB limits the cell growth of pathogen such as *Vibrio* (De la Fuente-Salcido et al. 2015;

Lee et al. 2015). Hence, in this case the limitation of the growth of some proteobacteria in the Tuba, may be caused by compounds produced by the bacterial community (such as the peptidases).

The alpha diversity tests, showed that in Chao1 and Observed_otus the Tuba P2 had significant difference but with Shannon and Simpsons index all Tuba samples (P1, P2, P3 and commercial) showed significant difference. That difference was due to the focus of each test, Observed_otus and Chao1 had low values for Tuba P2 that means low number of bacterial genus and high dominance. Although, Shannon and Simpson index analyzed the abundance and equity of the population, which means that the four Tuba groups have the same 10 genera but in different abundance (Figure 3). The low values of Tuba P2 in alpha tests may have related to high concentrations of sucrose and low acetate and ethanol. In a study of the bacterial diversity in pulque was established that the diversity is strongly correlated with ethanolic fermentation conditions and aguamiel and pulque composition (Escalante et al. 2008). The microbial beta diversity data showed no significant differences between the samples of each palm. Hence, the 10 most abundant genera of the 16 analyzed samples were associated with the origin of the samples. In some studies, the biotic and abiotic conditions (seasonality, plant physiology, age, soil conditions, and other abiotic variables such as water irrigation and other environmental factors) affected the bacterial diversity at different times of the fermentation (Staley et al. 2014; Fonseca-García et al. 2016; Coleman-Derr et al. 2016). Hence, we propose that the sugar concentration and the pH of the Tuba, has an effect on the bacterial diversity of this beverage, contributing to define the metabolic composition and the dominant bacterial genera. Additionally, the sap samples were collected after it was harvested by the producer and we took all the precautions to conserve the initial bacterial community and took it to the laboratory for fermentation (all handling was done in aseptic conditions). Therefore the observed differences in bacterial diversity in the samples is a combination of the palm related abiotic variables and the harvesting procedure itself.

Palm wine is consumed in many places in the world, the Tuba type that is the subject of this study has its own characteristics. It is produced near coconut palm production sites in the Mexican south pacific coast and is consumed as refreshing, hydration drink and empirically used as traditional aid for gastrointestinal discomfort, here we are showing its low alcohol content, however it can reach higher concentrations if fermented for longer time (Velázquez-Monreal et al. 2011). In this study, the functional analysis of the Tuba P1, P2, P3 and commercial samples using PICRUSt showed 18 functions of biotechnological interest (Figure 6), some of them showed significant differences as folate biosynthesis, antibiotic production and peptidases. Other functions were present on all Tuba samples without a significant difference among them such as terpene and carotenoid biosynthesis. These functions have been described in other fermented beverages such as pulque, where they proposed it as a functional product because it has mainly prebiotic and probiotic action with antimicrobial activity and production of nutrients (Escalante et al. 2016). In other study *Cocos nucifera* L. (Palmacea) water (CW), variety Chandrasankara, was tested for its ability to scavenge free radicals, and they found a good antioxidant activity percentage

(Mantena et al. 2003). Beverages made from plants, seeds or fruits have high contents of phenolic compounds that have the capacity to stabilize reactive oxygen and nitrogen species (Richelle et al. 2001), especially red, pink and white color fermented beverages (Martins de Sá et al. 2014). In addition, as we found in this work (Figure 7), microbial communities may be able to produce antioxidant compounds, there are evidence that described LABs genera as antioxidant compound producers, mainly glutathione, folate and butyrate (Wang et al. 2017). Other studies reported bacteria that produce antioxidant compound but it was not been identified yet (Tabbene et al. 2010), or it is a pigment produced in specific conditions by the bacteria (Radhakrishnan et al. 2016).

In this work we reported for the first time the bacterial diversity and potential functional analysis through the fermentation process of the Tuba. With the knowledge of microbiota diversity and metabolic functional inference, the Tuba production can be controlled adjusting the presence and abundance of beneficial genera that contributes with the functional characteristics of the Tuba. It also contributed to the stablishing of microbiological basis of its empirical uses. Additionally, the bacterial isolation from these samples may provide us with new species with probiotic potential.

References.

Caporaso, J.G. et al., 2011. QIIME allows analysis of high-throughput community sequencing data. *Nature America*, 7(5), pp.335–336.

Coleman-Derr, D. et al., 2016. Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species. *New Phytologist*, 209, pp.798–811.

Creer, S. et al., 2016. The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, 7, pp.1008–1018.

De la Fuente-Salcido, N.M. et al., 2015. Isolation and characterization of bacteriocinogenic lactic bacteria from M-Tuba and Tepache, two traditional fermented beverages in México. *Food Science & Nutrition*, 3(5), pp.434–442. Available at: http://doi.wiley.com/10.1002/fsn3.236.

Dellaglio, F. et al., 2017. Description of Gluconacetobacter swingsii sp. nov. and Gluconacetobacter rhaeticus sp. nov., isolated from Italian apple fruit. *International Journal of Systematic and Evolutionary Microbiology*, 55(2005), pp.2365–2370.

Díaz-Ruíz, G. et al., 2003. Microbial and Physiological Characterization of Weakly Amylolytic but Fast-Growing Lactic Acid Bacteria : a Functional Role in Supporting Microbial Diversity in Pozol , a Mexican Fermented Maize Beverage. *Applied and Enviromental Microbiology*, 69(8), pp.4367–4374.

Endo, A. et al., 2015. Comparative genomics of Fructobacillus spp . and Leuconostoc spp . reveals niche- specific evolution of Fructobacillus spp . *BMC Genomics*, 16(1117). Available at: http://dx.doi.org/10.1186/s12864-015-2339-x.

Escalante, A. et al., 2008. Analysis of bacterial community during the fermentation of pulque, a traditional Mexican alcoholic beverage, using a polyphasic approach. *International Journal of Food Microbiology*, 124(2), pp.126–134.

Escalante, A. et al., 2016. Pulque, a Traditional Mexican Alcoholic Fermented Beverage : Historical , Microbiological , and Technical Aspects. *frontiers in Microbiology*, 7, pp.1–18.

Escalante, A.E. et al., 2015. Ecological perspectives on synthetic biology : insights from microbial population biology. *frontiers in Microbiology*, 6, pp.1–10.

Fernández & Galván, 2015. Métodos para la cuantificación de proteínas. Aviable at: http://sgpwe.izt.uam.mx/files/users/uami/acym/27_metodos_para_la_cuantificacion_de_proteinas.pdf.

Filippis, F. De et al., 2017. A comparison of bioinformatic approaches for 16S rRNA gene pro fi ling of food bacterial microbiota. *International Journal of Food Microbiology*, 265, pp.9–17. Available at: http://dx.doi.org/10.1016/j.ijfoodmicro.2017.10.028.

Fonseca-García, C. et al., 2016. The Cacti Microbiome : Interplay between Habitat-Filtering and. *frontiers in Microbiology*, 7, pp.1–16.

Goosen, C. et al., 2007. Molecular and Biochemical Characterization of a Novel Intracellular Invertase from Aspergillus niger with Transfructosylating Activity. , 6(4), pp.674–681.

Hsieh, C.-W. et al., 2006. Molecular Cloning and Functional Identification of Invertase Isozymes from Green Bamboo Bambusa oldhamii. *Agricultural and Food Chemestry*, 54, pp.3101–3107.

Kuczynski, J. et al., 2012. Using QIIME to analyze 16s rRNA gene sequences from microbial communities. *Current Protocols in Microbiology*, pp.1–28.

Langille, M.G.I. et al., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, 31(9), pp.814–821. Available at: http://dx.doi.org/10.1038/nbt.2676.

Lee, S.H., Jung, J.Y. & Jeon, C.O., 2015. Bacterial community dynamics and metabolite changes in myeolchi-aekjeot , a Korean traditional fermented fi sh sauce, during fermentation. *International Journal of Food Microbiology*, 203, pp.15–22. Available at: http://dx.doi.org/10.1016/j.ijfoodmicro.2015.02.031.

Lisdiyanti, P. et al., 2003. Diversity of Acetic Acid Bacteria in Indonesia, Thailand, and the Philippines. *Microbiology and Culture Collections*, 19(2), pp.91–99.

Makarova, K. et al., 2006. Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences*, 103(42), pp.15611–15616. Available at: www.pnas.org/cgi/doi/10.1073/pnas.0607117103.

Mantena, S.K. et al., 2003. In vitro evaluation of antioxidant properties of Cocos nucifera. *Nahrung /Food*, 47(2), pp.126–131.

Martins de Sá, L.Z.C. et al., 2014. Antioxidant potential and vasodilatory activity of fermented beverages of jabuticaba berry (Myrciaria jaboticaba). *Journal of Functional Foods*, 8, pp.169–179. Available at: http://dx.doi.org/10.1016/j.jff.2014.03.009.

Navas-Molina, J.A. et al., 2015. Advancing our understanding of the human microbiome using QIIME. *Methods Enzymol*, 531(2013), pp.371–444.

Radhakrishnan, M. et al., 2016. In vitro antioxidant activity and antimicrobial activity against biofilm forming bacteria by the pigment from Desert soil Streptomyces sp D25. , 6(6), pp.148–150.

Richelle, M., Tavazzi, I. & Offord, E., 2001. Comparison of the Antioxidant Activity of Commonly Consumed Polyphenolic Beverages (Coffee, Cocoa, and Tea) Prepared per Cup Serving. *Journal of Agricultural and Food Chemistry*, 49(7), pp.3438–3442.

Rossi, M., Amaretti, A. & Raimondi, S., 2011. Folate Production by Probiotic Bacteria. *nutrients*, 3, pp.118–134.

Santiago-Urbina, J.A. & Ruíz-Terán, F., 2014. Microbiology and biochemistry of traditional palm wine produced around the world. *International Food Research Journal*, 21(4), pp.1261–1269.

Staley, C. et al., 2014. Core functional traits of bacterial communities in the Upper Mississippi River show limited variation in response to land cover. *frontiers in Microbiology*, 5(414), pp.1–11.

Tabbene, O. et al., 2010. A new antibacterial and antioxidant S07-2 compound produced by Bacillus subtilis B38. *FEMS Microbiology Letters*, 303, pp.176–182.

Tamang, J.P., Watanabe, K. & Holzapfel, W.H., 2016. Review : Diversity of Microorganisms in Global Fermented Foods and Beverages. *frontiers in Microbiology*, 7, pp.1–28.

Tang, J. et al., 2017. Analysis of the Bacterial Communities in Two Liquors of Soy Sauce Aroma as Revealed by High-Throughput Sequencing of the 16S rRNA V4 Hypervariable Region. *BioMed Research International*, 2017, pp.1–9.

Velázquez-Monreal, J. et al., 2011. Tuba: una bebida fermentada tradicional de Colima. Cienciacierta 25:11–14.

Wang, Y. et al., 2017. Antioxidant Properties of Probiotic Bacteria. *nutrients*, 9(521), pp.1–15.

Zhang, J. et al., 2016. Metagenomic approach reveals microbial diversity and predictive microbial metabolic pathways in Yucha , a traditional Li fermented food. *Scientific Reports*, 6, pp.1–9.

Figures.

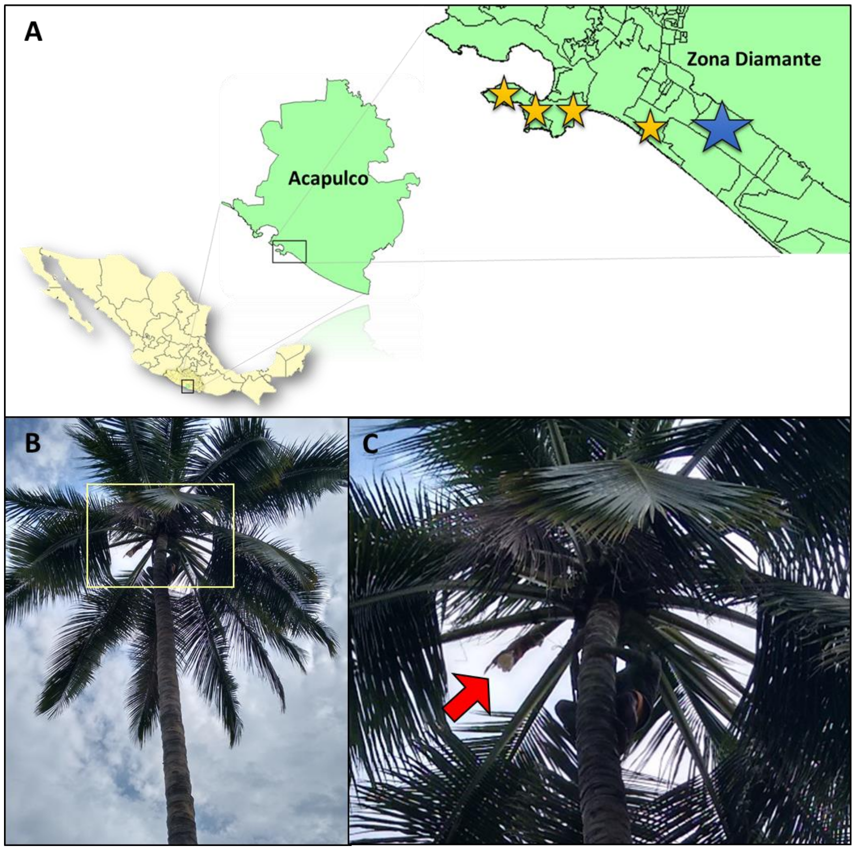

Figure 1. Sample collection. A) Sampling sites. The yellow stars represent the location of the four commercial establishments (commercial samples) and the blue star show the area where sap samples for the laboratory controlled fermentation were obtained. B) *Cocos nucifera* L (palm tree). Yellow square signaling sap collection zone. C) Sap collection zone. Red arrow indicate the palm structure (inflorescence) where the sap is collected.

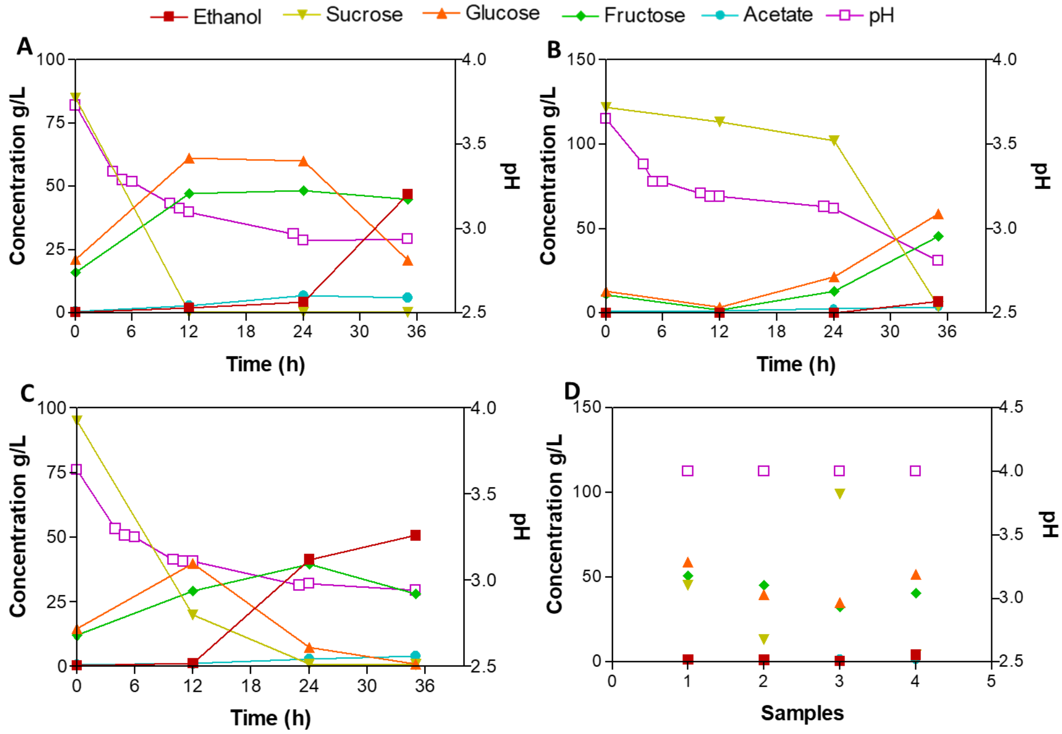

Figure 2. Metabolic composition of the laboratory fermented Tuba and commercial Tuba. A) Tuba P1, B) Tuba P2, C) Tuba P3 and D) Commercial Tuba samples. Each number correspond to one sample. Right axis represented pH value.

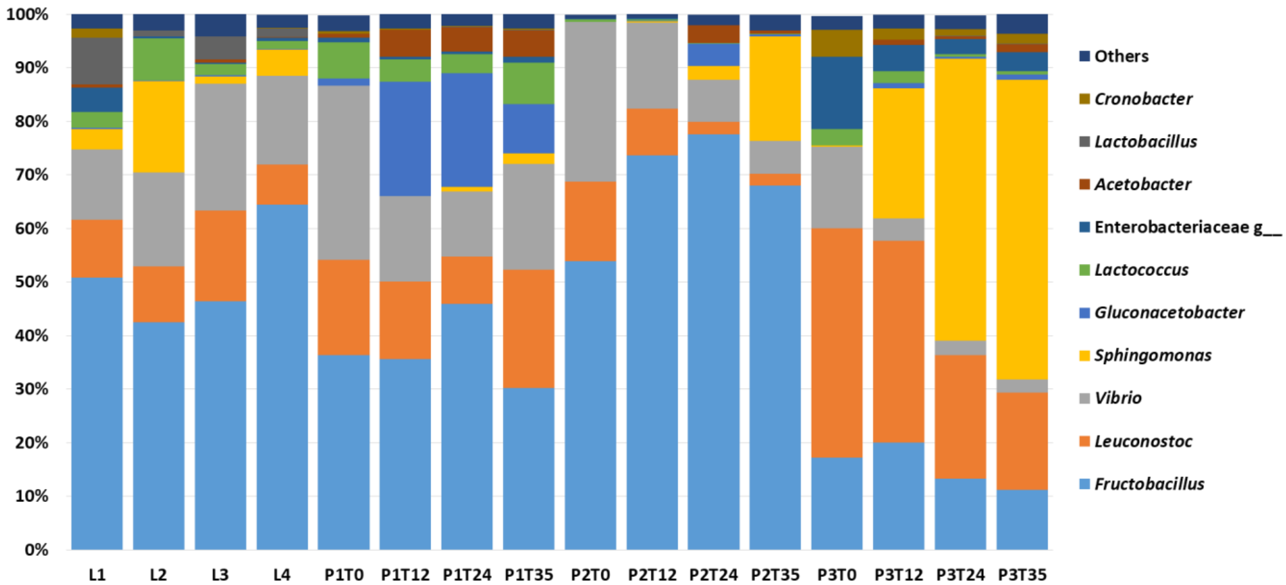

Figure 3. Taxonomic identification. The graph represented the top ten of genera using 0.01% abundance filter OTUs table.

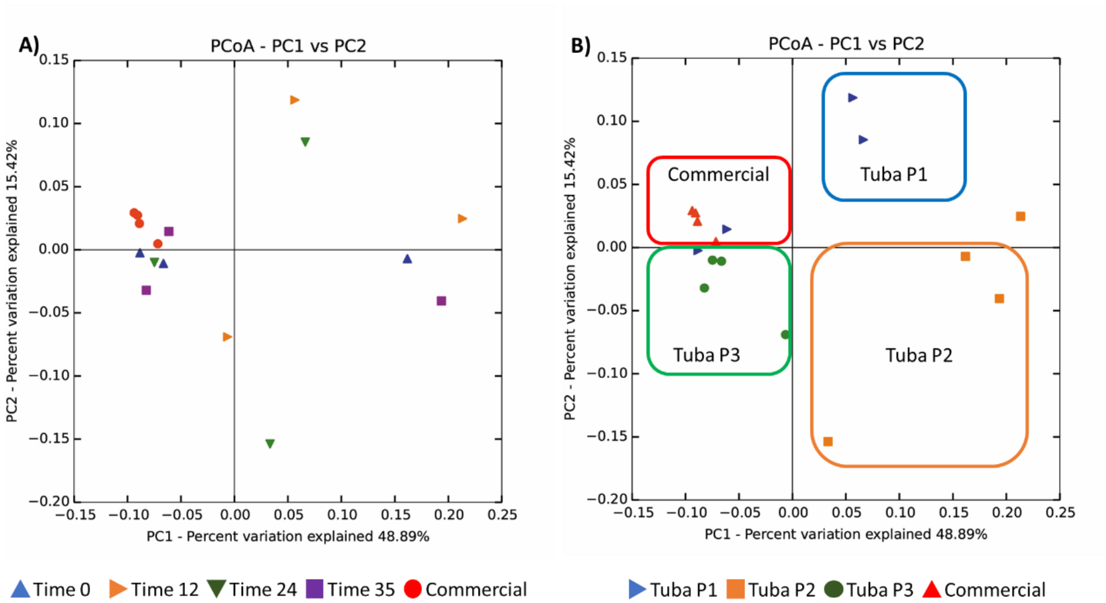

Figure 4. Beta diversity. A) Associate with respect to the fermentation time. B) Associate with respect to the origin of the sample. Analysis performed by the Unifrac unweighted technique with 0.01% abundance filter and plotted with the Principal Coordinates Analysis (PCoA). The color boxes show a grouping data.

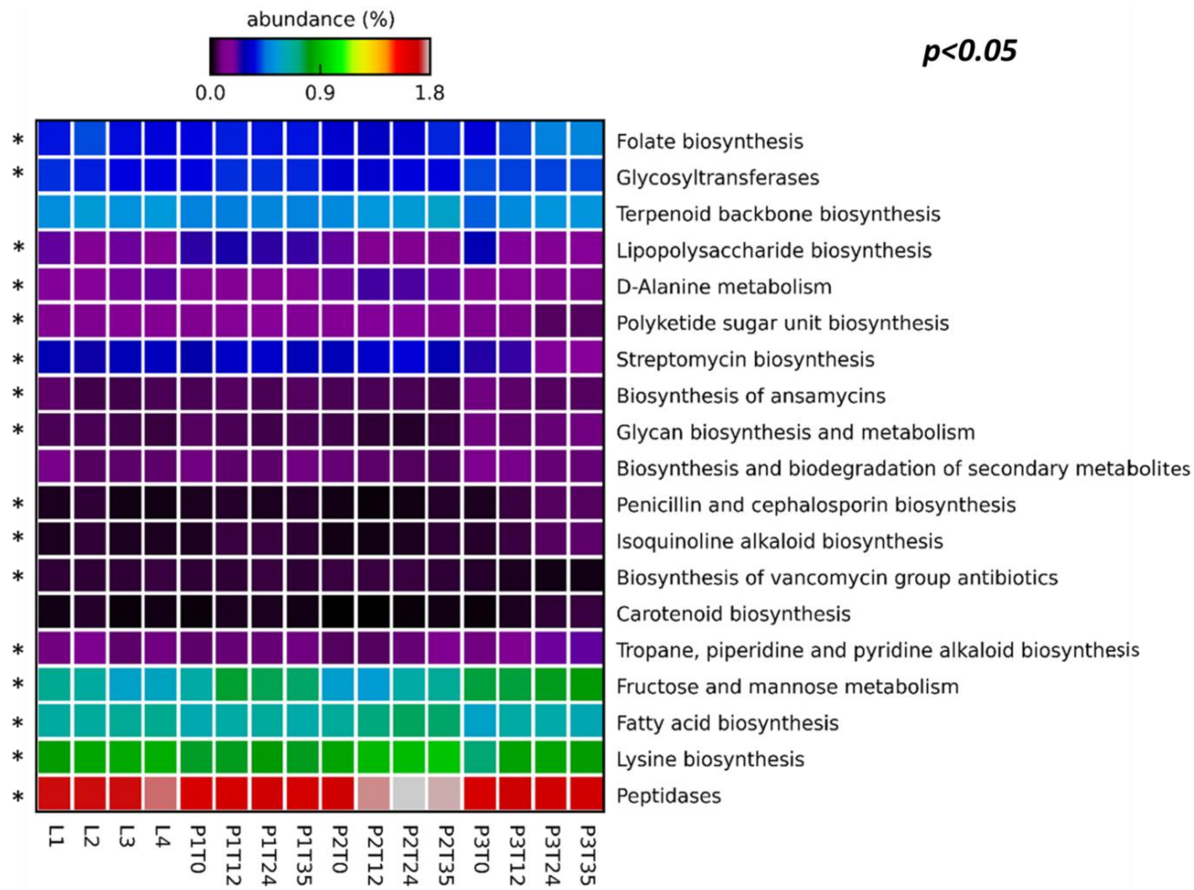

Figure 5. Abundance of sequences associate with functions. An ANOVA was performed with Tukey-Kramer (0.95), the percentage of genes associated with functions, discarding elementary cellular functions. Asterisk show functions with significant difference (*p<0.05*).

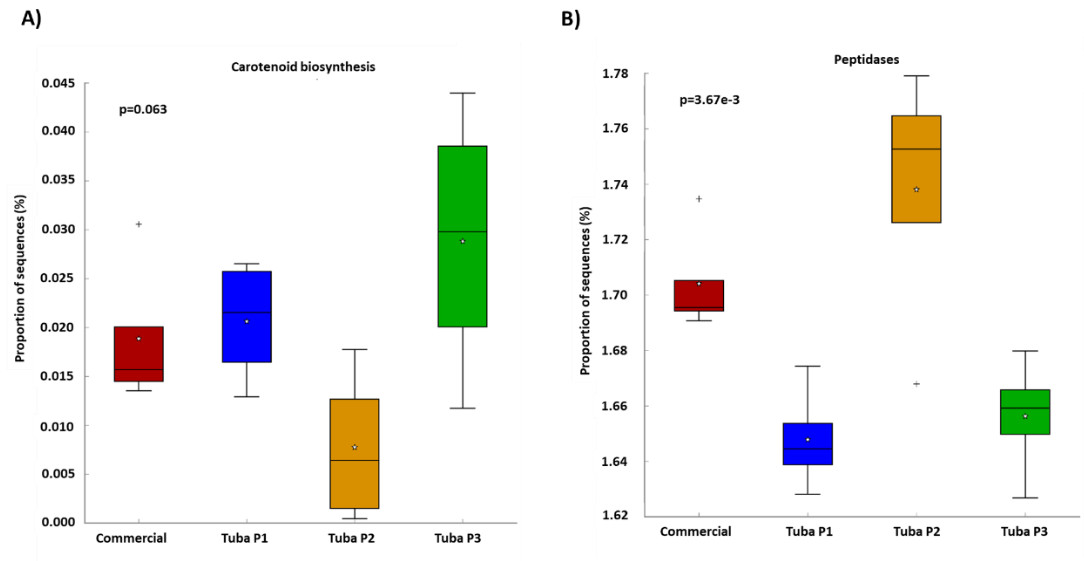

Figure 6. Box plot of two functions of interest. With A) not significant difference and B) significant difference. An ANOVA was performed with Tukey-Kramer (0.95) and plotted with STAMP.

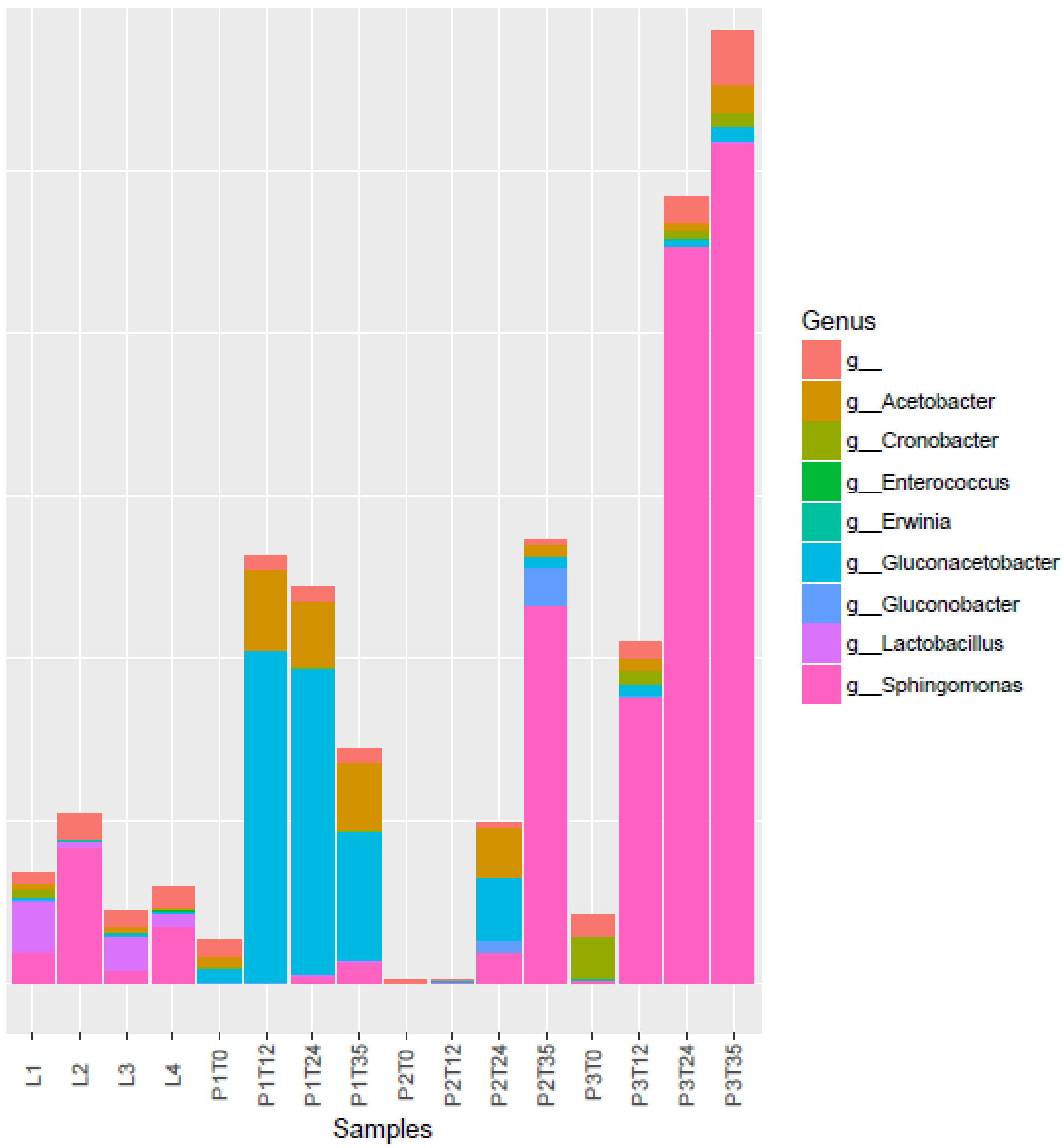

Figure 7. Main bacteria with 15-cis-phytoene synthase gene (K02291 KEGG code). Analysis performed with the function "metagenome_contributions.py" obtained by PICRUSt analysis and plotted with R studio.

Tables.

Table 1. Climatological conditions of the study sites.

PARAMETER	PALM SAP COLLECT	COMMERCIAL COLLECT
Date	14/07/2016	16/08/2016
Coordinates	North 16 ° 46'54.53 '' West 99 ° 47'02.73 ''	
Altitude	12	
Temperature	Max. 32°C y Min. 24°C	Max. 30°C y Min. 24°C
Humidity	87%	89%
Pressure	0.996 atm	0.996 atm
Weather	Light rain	Light rain

Data provided by Comisión Nacional del Agua (CONAGUA).

Table 2. PCR primers targeting 16S rRNA V3-V4 region of bacteria

Amplicon size:	550 bp
Forward	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3'
Reverse	5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3'.

Table 3. Alpha diversity. Asterisk show significant difference with *p<0.05*.

Alpha diversity metrics/Sample	Chao1 index	Observed_otus	Shannon index	Simpson index
Commercial Tuba	49.240 ± 3.758	42.600 ± 5.609	2.619 ± 0.006*	0.692 ± 0.002*
Tuba P1	49.715 ± 5.307	42.508 ± 6.621	2.876 ± 0.010*	0.781 ± 0.001*
Tuba P2	35.863 ± 4.654*	26.588 ± 5.235*	1.595 ± 0.011*	0.491 ± 0.002*
Tuba P3	49.559 ± 5.305	41.229 ± 7.006	2.468 ± 0.014*	0.704 ± 0.001*

Supplementary material.

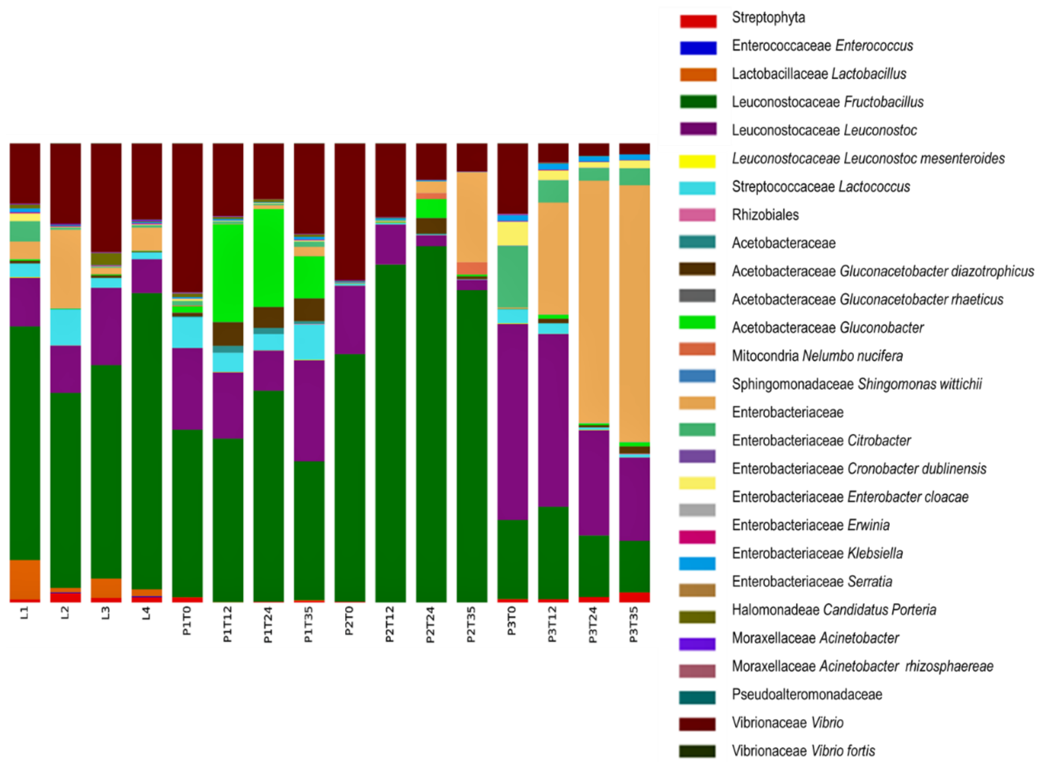

Figure 1S. Taxonomic identification. Most abundant OTU's using the 0.01% abundance filter OTUs table.

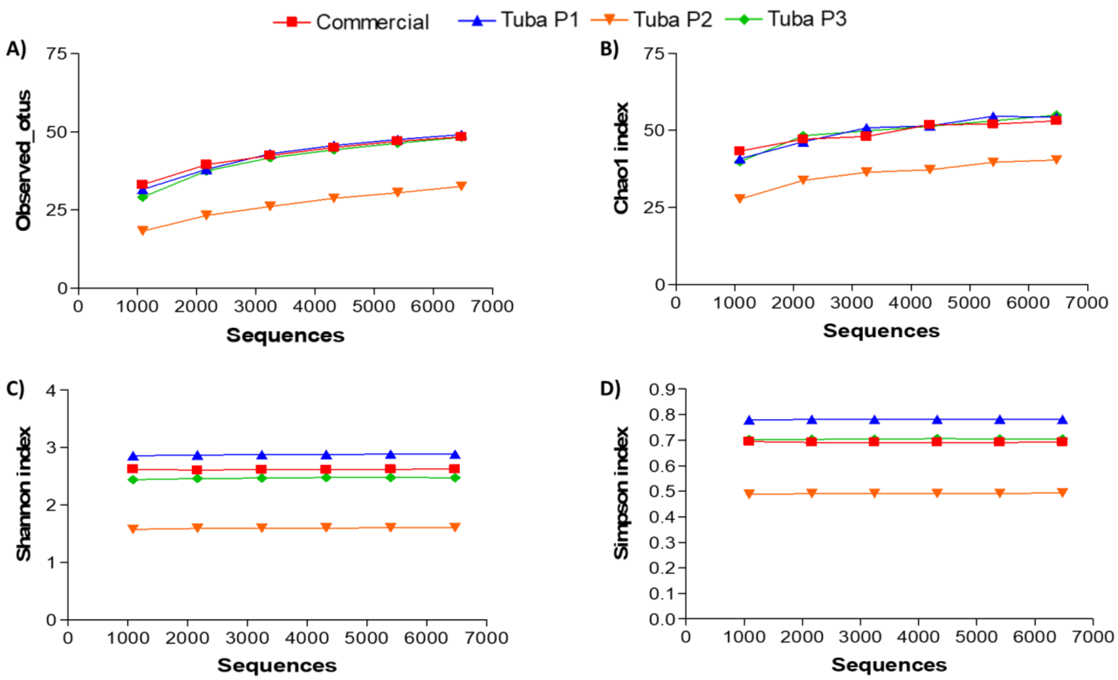

Figure 2S. Alpha diversity rarefaction plots with 0.01%. A) Observed_otus, B) Chao1, C) Shannon and D) Simpson. Each population is represented for a specific color in all the graphics**.**

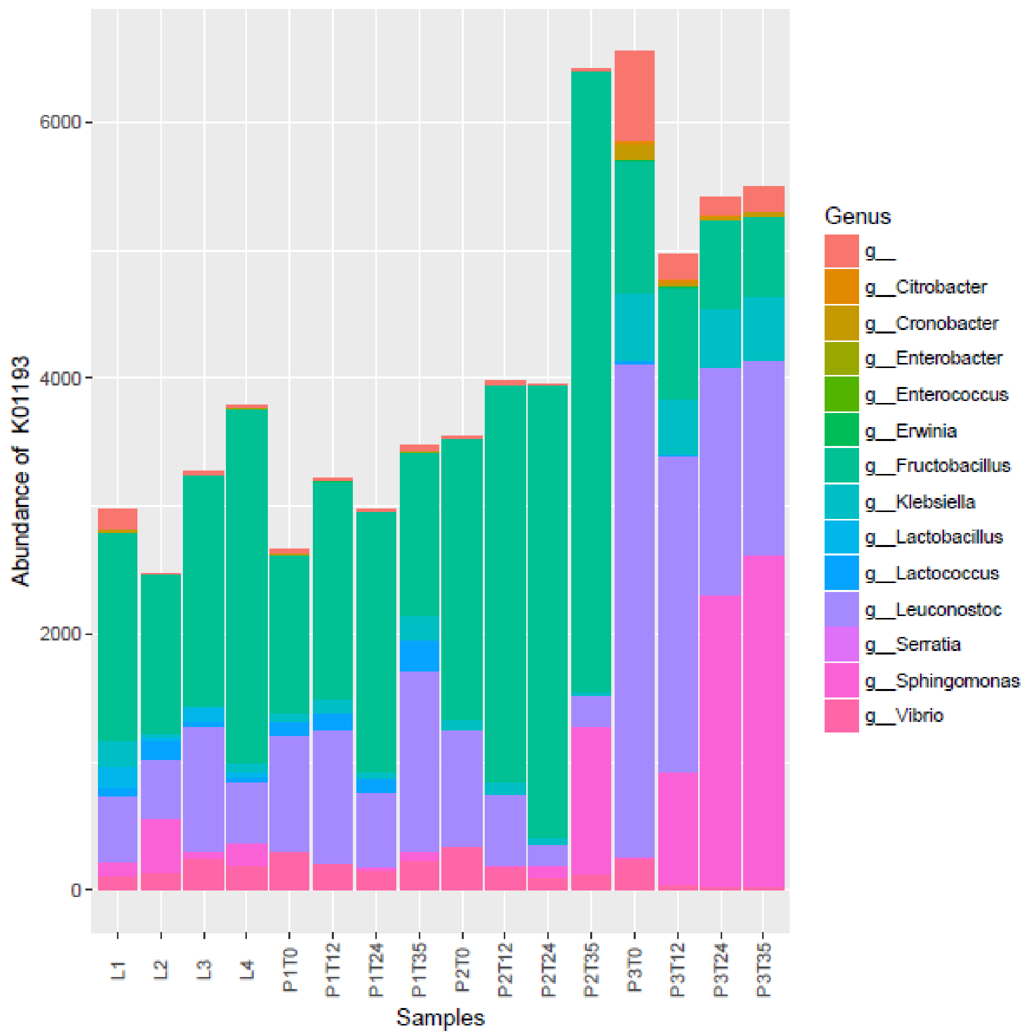

Figure 3S. Abundance of invertase gene (K01193). Analysis performed with the function "metagenome_contributions.py" obtained by PICRUSt analysis and plotted with R studio.

Table 1S. Chemical composition of the Tuba.

	Tuba P1	Tuba P2	Tuba P3	Commercial L1	Commercial L2	Commercial L3	Commercial L4
Time (h)				**Proteins (g/L)**			
0	0.0068	0.0115	0.0097	0.0087	0.0173	0.0166	0.0054
12	0.0094	0.0038	0.0109				
24	0.0057	0.0084	0.0046				
35	0.0082	0.0147	0.0042				
Time (h)				**Sucrose (g/L)**			
0	85.1478	121.7608	95.2033	45.3103	13.4554	99.2911	4.3160
12	0.3932	113.2938	19.8372				
24	0.3748	102.1780	0.8837				
35	0.2629	4.2062	0.7476				
Time (h)				**Glucose (g/L)**			
0	21.1467	13.1798	14.6654	59.2407	39.8149	35.0787	51.8465
12	61.4149	3.4526	39.8145				
24	60.1442	21.6494	7.2646				
35	21.0363	59.0502	0.8876				
Time (h)				**Fructose (g/L)**			
0	16.0169	11.0782	11.9225	50.8742	45.5284	32.7653	40.9536
12	47.3585	1.9186	29.1776				
24	48.5620	13.0289	39.7342				
35	44.8364	45.6336	28.1220				
Time (h)				**Acetate (g/L)**			
0	0.6867	0.8935	0.6452	1.5209	1.6449	1.5271	1.9845
12	3.0672	1.4066	1.1047				
24	6.9141	2.7640	2.8354				
35	6.0741	3.5594	3.9577				
Time (h)				**Ethanol (g/L)**			
0	0.4116	0.1669	0.2528	1.5176	1.4198	0.7016	4.4201
12	2.2396	0.0000	1.1756				
24	4.4278	0.0000	41.3655				
35	47.1141	6.8062	50.6929				
Time (h)				**pH**			
0	3.73	3.65	3.64	4.0	4.0	4.0	4.0
4	3.34	3.38	3.3				
5	3.29	3.28	3.26				
6	3.28	3.28	3.25				
10	3.15	3.21	3.12				
11	3.12	3.19	3.11				
12	3.1	3.19	3.11				
23	2.97	3.13	2.97				
24	2.93	3.12	2.98				
35	2.94	2.81	2.94				