

## **(Meta)genomic approaches to gain insight into the cocoa bean fermentation ecosystem and its associated candidate functional starter cultures**

(Meta)genomics is a scientific field that enables determination of the genetic material of microorganisms, which has led to new research opportunities in microbiology. The present study focused on the (meta)genomics of the cocoa bean fermentation process as a case study.

The fermentation of raw cocoa beans is an important step in the production of chocolate, during which the mucilaginous pulp surrounding the cocoa beans is removed, important colour and flavour precursors in the beans are formed, and the seed embryo is killed. Insights into the fermentation ecosystem composition and the functional roles of the microorganisms involved is of great interest, as this will provide knowledge to better select functional starter culture strains to perform controlled cocoa bean fermentation processes.

First, a combination of several similarity-based and composition-based taxonomic profiling tools was applied on metagenomic sequence data of the microbial community DNA of a single cocoa bean fermentation process sample from Brazil. Overall, this approach revealed a wider bacterial and fungal diversity than previously reported and identified *Hanseniaspora uvarum*, *Hanseniaspora opuntiae*, *Saccharomyces cerevisiae*, *Lactobacillus fermentum*, and *Acetobacter pasteurianus* as the prevailing microbial species. Secondly, a functional analysis of the metagenomic sequence data of this single sample was carried out to gain insight into the metabolic capacities of the different microorganisms involved. By reconstructing bacterial meta-pathways, their association with the microbial members of the cocoa bean fermentation ecosystem was revealed.

Whole-genome shotgun sequencing and in-depth functional analysis of the lactic acid bacteria strains *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80 revealed functional roles of these strains during the cocoa bean fermentation process. To investigate the mechanisms that enable *Acetobacter pasteurianus* 386B to dominate the cocoa bean fermentation process, a (comparative) genome sequence analysis using whole-genome mate-pair sequencing was applied. This revealed important characteristics for this strain, including the presence of strain-specific genes such as a gene coding for an endopolygalacturonase and mechanisms involved in tolerance towards various stress conditions. The genome sequences of *Acetobacter ghanensis* LMG 23848<sup>T</sup> and *Acetobacter senegalensis* 108B, species that are often found in the beginning of cocoa bean fermentation processes, were determined and analysed using whole-genome mate-pair sequencing. This approach revealed unique functional properties that are of interest for the cocoa bean fermentation process.

Overall, this work extended the knowledge of the cocoa bean fermentation ecosystem by using a metagenomic sequencing approach, in particular towards the microorganisms involved and the underlying mechanisms of their metabolic features. In addition, genome-based analysis of five (candidate) functional starter culture strains of potential application in the cocoa bean fermentation process was performed and proved to be a powerful tool to facilitate the selection of appropriate starter culture strains for controlled cocoa bean fermentation processes.

## **(Meta)genomische aanpak om inzicht te verwerven in het cacaofoonfermentatieproces en geassocieerde kandidaat functionele starterculturen**

(Meta)genomica is een wetenschapsveld dat de bepaling van het genetisch materiaal van micro-organismen mogelijk maakt, hetgeen heeft geleid tot een nieuwe microbiologische onderzoeks aanpak. Onderhavige studie was gericht op de (meta)genomica van het cacaofoonfermentatieproces als casus.

De fermentatie van rauwe cacao bonen is een belangrijke stap in de productie van chocolade, waarbij de viskeuze pulp rond de cacao bonen verwijderd wordt, belangrijke kleuren- en smaakprecursoren in de bonen gevormd worden en het zaadembryo gedood wordt. Inzichten in de samenstelling van het fermentatie-ecosysteem en de functionele rol van de betrokken micro-organismen is van groot belang, aangezien dit kennis verschaft om functionele startercultuur stammen beter te selecteren teneinde gecontroleerde cacaofoonfermentatieprocessen uit te voeren.

Vooreerst werd een combinatie van verscheidene compositie- en similariteitsgebaseerde taxonomische profileringsmethoden toegepast op de metagenoomsequentie data van DNA van een staal van de volledige microbiële gemeenschap van een Braziliaans cacaofoon-fermentatieproces. Algemeen onthulde deze aanpak een bredere diversiteit aan bacteriën en schimmels dan voorheen gerapporteerd werd en werden *Hanseniaspora uvarum*, *Hanseniaspora opuntiae*, *Saccharomyces cerevisiae*, *Lactobacillus fermentum* en *Acetobacter pasteurianus* geïdentificeerd als de meest voorkomende microbiële soorten. Ten tweede werd een functionele analyse van de metagenoomsequentie data van dit ene staal uitgevoerd om inzicht te verkrijgen in de metabole capaciteiten van de verschillende betrokken micro-organismen. Door het reconstrueren van bacteriële metabole meta-paden werd hun associatie met de microbiële leden van het cacaofoonfermentatie-ecosysteem onthuld.

*Shotgun*-sequentiebepaling en diepgaande functionele analyse van het volledige genoom van de melkzuurbacteriestammen *Lactobacillus fermentum* 222 en *Lactobacillus plantarum* 80 onthulden functionele rollen van deze stammen tijdens het cacaofoonfermentatieproces. Om de mechanismen te onderzoeken die *Acetobacter pasteurianus* 386B in staat stellen om het cacaofoonfermentatieproces te domineren, werd een (vergelijkende) genoomsequentieanalyse toegepast aan de hand van *mate-pair* sequentie bepaling. Dit onthulde belangrijke eigenschappen van deze stam, inclusief de aanwezigheid van stamspecifieke genen zoals een gen coderend voor een endopolygalacturonase en mechanismen betrokken bij de tolerantie voor verschillende stresscondities. De genoomsequenties van *Acetobacter ghanensis* LMG 23848<sup>T</sup> en *Acetobacter senegalensis* 108B, soorten die vaak worden aangetroffen in het begin van cacaofoonfermentatieprocessen, werden bepaald en geanalyseerd aan de hand van *mate-pair* sequentie bepaling. Deze aanpak onthulde unieke functionele eigenschappen die belangrijk zijn voor het cacaofoonfermentatieproces.

Dit werk breidde in het algemeen de kennis van het cacaofoonfermentatie-ecosysteem uit, op basis van een metagenoomsequentiebepalingsaanpak, in het bijzonder omtrent de aanwezige micro-organismen en de onderliggende mechanismen van hun metabole eigenschappen. Bovendien werd een genoomgebaseerde analyse van vijf (kandidaat) functionele startercultuur stammen toepasbaar in het cacaofoonfermentatieproces uitgevoerd, hetgeen een krachtige methode bleek te zijn om de selectie van geschikte startercultuur stammen voor gecontroleerde cacaofoonfermentatieprocessen te vergemakkelijken.