Technische Universität München
TUM School of Life Sciences

Characterization and habitat adaptation of *Tetragenococcus halophilus* and *Debaryomyces hansenii* from a lupine seasoning sauce fermentation

Tobias Link

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Michael Rychlik

Prüfende der Dissertation:

1. apl. Prof. Dr. Matthias A. Ehrmann
2. Prof. Dr. J. Phillip Benz

Die Dissertation wurde am 21. 08. 2023 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 18. 12. 2023 angenommen.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

°C	Degree Celsius	**mA**	Milliampere
µg	Microgram	**MALDI-TOF MS**	Matrix-assisted laser desorption/ionization
µL	Microliter	**Mbp**	Megabase pair
µM	Micromolar	**mg**	Milligram
µm	Micrometer	**min**	Minute
ADI	Arginine Deiminase	**mL**	Milliliter
ANI	Average Nucleotide Identity	**mm**	Millimeter
ANIb	Average Nucleotide Identity based on the BLAST algorithm	**mM**	Millimolar
ATP	Adenosine triphosphate	**MRS**	Culture Medium developed by De man Rogosa and Sharpe
BLAST	Basic local alignment search tool	**N_{50}**	Sequence length of the shortest contig at 50 % assembly rate
bp	Base pair	**NaCl**	Sodium Chloride
Cas	CRISPR associated sequence	**ng**	Nanogram
CDS	Coding sequence	**nM**	Nanomolar
CFU	Colony forming units	**OD_{600nm}**	Optical density measured at 600 nm
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats	**ORF**	Open Reading Frame
DDH	DNA-DNA hybridization	**PCR**	Polymerase Chain Reaction
DEGs	Differentially expressed genes	**PGAP**	Prokaryotic Genome Annotation Pipeline from NCBI
DHAP	Dihydroxyacetone phosphate	**pH**	Negative decimal logarithm of hydrogen ion activity
DNA	Deoxyribonucleic acid	**p-value**	Probability value
dNTP	Deoxy nucleoside triphosphate	**qPCR**	Quantitative PCR
dsDNA	Double-stranded DNA	**RAPD**	Random-amplified polymorphic DNA
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen	**RAST**	Rapid annotations using subsystems technology
FDR	False Discovery Rate	**RFO**	Raffinose Family Oligosaccharides
g	G Force	**RM**	Restrictions modification system
g	Gram	**RNA**	Ribonucleic acid
G3P	Glyceraldehyde 3-phosphate	**ROS**	Reactive Oxygen Species
gDNA	Genomic DNA	**rpm**	Rounds per minute
GGDC	Genome-to-Genome Distance Calculator	**RT**	Room temperature
h	Hour	**s**	Second
HDC	Histidine decarboxylase	***sp.***	Species
HPAEC-PAD	High performance anion exchange chromatography with pulsed Amperometric detection	***subsp.***	Subspecies
ITS	Internal transcribed spacer	**TDC**	Tyrosine decarboxylase
kb	Kilobase pair	**TIGR**	The Institute for Genomic Research
KEGG	Kyoto encyclopedia of genes and genomes	**TMW**	Technical microbiology Weihenstephan
L	Liter	**TRIS**	Tris (hydroxymethyl) aminomethane

L_{50}	The smallest number of contigs whose combined length covers 50 % of the total genome size	**UPGMA**	Unweighted Pair Group Method with Arithmetic mean
LAB	Lactic acid bacteria	***var.***	Variant
LMRS	Modified MRS medium with Lupine peptone	**w/v**	Weight per volume
M	Molar	**YGAP**	Yeast Genome Annotation Pipeline

1 Introduction

1.1 Food Fermentations

Fermented foods are made by enzymatic and microbial activity on a specific foodstuff in a desired manner (Marco et al., 2021). Fermentation of foods plays a big role in the human diet. A variety of foods are made by microbial fermentation including cereals, beer, yogurt, cheese, bread, meat in the form of sausages, fermented fish, fermented vegetables such as kimchi, vinegar, coffee, cocoa and soybeans (Gänzle, 2015). Fermentation of food can be traced back 14,000 years, to a time when the hunter-gatherer societies became agricultural societies and preservation of food was needed (Arranz-Otaegui et al., 2018; Hayden et al., 2013). Until today, foods are fermented for preservation as it is a cheap way of increasing the shelf-life. This preservation can be linked to microbial activity as the carbohydrates of the fermented substrate can be broken down into different organic acids such as lactate and acetate in a process called lactic acid fermentation (Gänzle, 2015). The production of lactate lowers the pH of the fermented food. This in turn reduces the growth of other microorganisms. Furthermore, some microorganisms can produce anti-microbial agents such as bacteriocins that kill or inhibit the growth of microorganisms that are associated with food spoilage (Papagianni and Anastasiadou, 2009).

Another reason for fermenting foods is that fermented foods are considered to have health benefits (Marco et al., 2017). Many fermented foods are rich in vitamins as well as minerals and are linked to antioxidant or anti-diabetic activity (Hur et al., 2014; Sivamaruthi et al., 2018). For example, the consumption of ‘Kochujang’ (Korean red pepper paste) can lower the risk of hyperlipidemia and the consumption of ‘Chungkookjang’ (fermented soybeans) is reported to improve the body fat percentage, lean body mass and the waist-to-hip ratio (Byun et al., 2016; Lim et al., 2015). Again, some of these beneficial attributes are linked to microorganisms. For example, some studies revealed that proteins in the fermented foodstuff can be altered into bioactive peptides such as isoleucyl-prolyl-proline (IPP) and valyl-prolyl-proline (VPP) in milk fermentation. These bioactive peptides are recommended for the management of hypertension (Beltrán-Barrientos et al., 2016; Usinger et al., 2009).

INTRODUCTION

1.2 Fermented Asian seasoning sauces and pastes.

The fermentation of fish and soybeans into sauces and pastes has a long tradition, especially in Asia. The following chapter will provide an overview of the fermentation processes of these products with concern to the different countries.

Fish sauces

Fish-based sauces vary from country to country. They are known as ‘Nouc-mam’ and ‘Nouc-mam-gau-ca’ from Cambodia, ‘Yeesui’ originated from Hong Kong, ‘Ketjap-ikan’ from Indonesia, ‘Shottsuru’ from Japan, ‘Aekjeot’ from Korea, ‘Budu’ from Malaysia, ‘Patis’ from the Philippines and ‘Nam-pla’ from Thailand. The sauces vary in the fish species used during the manufacturing, the sodium chloride content and the fermentation time (Beddows, 1998; Fukui et al., 2012; Koo et al., 2016). Although the manufacturing process can vary, the general procedure is similar. The nam-pla fish sauce from Thailand is produced by mixing cleaned fish with salt in the ratio from 2:1 or 3:1 (fish: salt) (w/v). Then, the fish is filled into a tank or tableware and mixed with a brine solution with the correct amount of sodium chloride. After 12- 18 months of fermentation, the supernatant is transferred into a second tableware or tank for ripening. After a ripening period of 2-12 weeks, this solution is called nam-pla. To produce lower-quality nam-pla, the fish residues in the fermentation tank or tableware are filled a second time with brine solution and fermented for up to 4 months (Lopetcharat et al., 2001).

Soy-based sauces and pastes.

Presumably, the most famous Asian seasoning sauces are soy sauces. Similarly to fish sauces, there are many different types such as the Indonesian ‘Kecap’, the Korean ‘Doenjang’ and 'Ganjang’, the Chinese or Japanese soy sauce (Fukushima, 1985, 1981; Roling and van Verseveld, 1996; Röling et al., 1994; Shin and Jeong, 2015). Although all products rely on soybeans and sometimes on the addition of wheat or rice, the production process is very different.

To produce the Korean ‘Ganjang’ (soy sauce) and ‘Doenjang’ (miso), the blocks of soybeans are mixed with salt brine and filled in special tableware called ‘onggi’ for aging. After three to six months the liquid is separated from the solids called ‘Kanjang’ and the solids left in the ‘onggi’ are now called ‘Doenjang’ (Shin and Jeong, 2015).

INTRODUCTION

Japanese soy sauce types are called ‘shoyu’ and can be divided into five types, ‘Koikuchi’, ‘Usukuchi’, ‘Tamari’, ‘Sashikomi’ and ‘Shiro’ (Fukushima, 1985, 1981; Lioe et al., 2010). The most consumed one is koikuchi-shoyu which is produced by mixing roasted and crushed wheat with soaked soybeans at a ratio dependent on the manufacturer. This mixture is then inoculated with an *Aspergillus* and fermented at 25-35 °C for at least 36 h, at this point the mixture is called ‘**Koji**’. Next, the ‘Koji’ is mixed with a brine solution containing up to 20% NaCl (w/v) in a big tank and is now called ‘**Moromi**’. Depending on the type and manufacturer, the aging time of ‘Moromi’ can vary between months to several years (Fukushima, 1985, 1981; Lioe et al., 2010).

The other Japanese shoyu types are manufactured similarly, the major differences are the altered amounts of wheat during the fermentation or in the case of sashikomi, the use of koikuchi shoyu instead of water at the beginning of the ‘Moromi’ fermentation (Fukushima, 1985, 1981).

Traditional Chinese soy sauce and Indonesian ‘Kecap’ production are similar as they use soybeans as the sole ingredient (Fukushima, 1985, 1981). The production process is similar to the Japanese type in a two-step fermentation, with the ‘Koji’ fermentation being done for up to 2 weeks (Lioe et al., 2010).

Fermentation of foodstuff, such as soybeans can be accelerated and controlled by back slopping of finished products into a new ‘Moromi’ fermentation (Fukushima, 1985, 1981). In a similar fashion back slopping of a portion from a sourdough into a new sourdough fermentation is regularly done to preserve and accelerate the sourdough (Decock and Cappelle, 2005). A broad variety of seasoning sauces and pastes based on soybeans with wheat or different types of fish have been developed in Asian cultures over time, based on these simple principles (Lee, 1997; Patra et al., 2016).

Lupine-based seasoning sauce

Recently, as an alternative to soy sauces, some companies started manufacturing a seasoning sauce based on lupine beans. Lupine beans and soybeans have comparable total nutrient content. The main differences are in the distributions of the amino acids and variations in the carbohydrate composition. While the major galactan present in the hemicellulose of soybeans consists of arabinogalactan with a 52:43:5 ratio for galactose:arabinose:galacturonic acid, the major galactan in lupine beans is

β-1,4-galactan with a 88:5:1:1:5 ratio for galactose:arabinose:rhamnose:xylose:galacturonic acid(Sakamoto et al., 2013).In lupine beans, these galactans are reported to be the major storage material in the seed (Kaczmarska et al., 2017). Therefore, the compositional differences in the galactans of soybeans and lupine beans alter the carbohydrates available in the fermentation process. The lupine seasoning sauce has a lower salt content than most soy sauces (13,4 % (w/v)) and is fermented similarly to the Japanese 'Tamari' shoyu type (Figure 1). The microbial dynamics of this fermentation were analyzed at multiple sodium chloride concentrations by Rebekka Lülf. Revealing that the two species *Tetragenococcus (T.) halophilus* and *Debaryomyces (D.) hansenii* are dominating the microbiota in the fermentation (Lülf et al., 2021). Although the fermentation process is similar to the ones for soy sauces, the key microorganisms that contribute to flavor development, as well as the traits needed for dominating the fermentation and the importance of individual species within this fermentation process, are not yet known.

A B

Figure 1: Mainstages in the lupine sauce fermentation process. A) 'lupine koji', toasted, mashed and cracked lupine beans fermented with *A. oryzae*. B) 'lupine moromi', koji is mixed in a 1:1.5 ratio with a brine solution in large vessels.

1.3 Microbiota dynamics of seasoning sauces during fermentation

Due to rising interest in microbiology (in the late 19th and early 20th century) as well as improved technologies, many researchers started to look closer at fermented foodstuff. Recently, the microbial consortia of many Asian fermented foods have been investigated, including several variants of soy sauces, soy pastes and fish sauces (Allwood et al., 2021; Du et al., 2019; Fukui et al., 2012; Jung et al., 2016; Kim et al.,

2009; Röling et al., 1994; Sim et al., 2015; Song et al., 2015b, 2015a; Sulaiman et al., 2014; Tanaka et al., 2012; Yan et al., 2013).

In detail, a change in the dominant species of the traditional Chinese fish sauce from *Shewanella* sp. and *Halanaerobium* sp. to *Pseudomonas* sp., *Psychrobacter* sp., *Carnobacterium* sp. and *Tetragenococcus* sp. was observed during fermentation. Interestingly, *Tetragenococcus* sp. was the most abundant species (Du et al., 2019).
Similar microbiological changes could be detected in Japanese 'Shottsuru'. The microbiota of this fermentation changed from *Staphylococcus* sp. and *Bacillus* sp. to *Bacillus* sp. and *Tetragenococcus* sp. Again, *Tetragenococcus* sp. dominated the microbiota of the fermentation (Fukui et al., 2012).
In Korean 'Ganjang' fermentation the microbiota shifts from a very diverse community to a community consisting of *Tetragenococcus* sp., *Staphylococcus* sp., *Halomonas* sp. and *Marinobacter* sp. Furthermore, *Debaryomyces* sp. was the major yeast in the fermentation (Han et al., 2020).
In Japanese soy sauce and Indonesian 'Kecap' the bacterial community is dominated by *Staphylococcus* sp. and *Weissella* sp. and changed to a community with *Tetragenococcus* sp. as the most dominant species (Röling et al., 1994; Tanaka et al., 2012).

Conclusively, the named fermentation processes and the changes in the microbiota demonstrate that *T. halophilus* is a dominating organism in fish and soy sauce fermentation. This raises the question, of whether *T. halophilus* can also be considered the dominant key fermentative organism in the lupine moromi fermentation. Experiments done by Lülf et al. support this assumption (Lülf et al., 2021). At the start, the microbiota of lupine moromi fermentation was dominated by *Weissella paramesenteroides*. After six months, the community has shifted to *T. halophilus* and *Staphylococcus equorum*. Furthermore, the dominant yeast species was *Debaryomyces hansenii* (Lülf et al., 2021). This experiment demonstrates, that *T. halophilus* also plays a major role in lupine moromi fermentation, the second fermentation stage in the production process of lupine seasoning sauce. Despite, this evidence, it is still unclear which traits are favored and contribute to the dominance of a strain. Furthermore, it is not clear how *T. halophilus* influences the flavor contribution or contributes to food safety in the lupine moromi fermentation.

1.4 Starter cultures in the food industry

To accelerate fermentation processes and ensure a safe and consistent product, starter cultures are being used. Starter cultures have a long history in the food industry. In the beginning, these starter cultures consisted of the finished product that was being backslopped to reinoculate a fresh fermentation or by using special containers that allowed for the survival of the microorganisms that were present in a mature product (Durso and Hutkins, 2003). With the development of techniques for cultivating pure cultures by Pasteur and the identification of lactic acid bacteria by Lister in 1877, the industrial production of pure starter cultures was possible (Lister, 1877; Louis, 1879). Since then, pure starter cultures have been used for a variety of food fermentations including cheese, yogurt, wine, beer, bread, soy sauce and meat (Devanthi et al., 2018; Nardi et al., 2019; Roberts et al., 1992; Roca and Incze, 1990; Van Kerrebroeck et al., 2018; Walker and Stewart, 2016; Xu et al., 2015).

Selecting starter cultures for fermentation is difficult, as many factors and corresponding strain-specific attributes have to be considered and tested in advance (Janßen et al., 2019). The major attributes considered for the selection of starter culture stains are discussed below.

Most importantly, the selected starter strains must be able to dominate and thereby control the fate of the fermentation process. The production of exopolysaccharides may have antimicrobial activity against other species and thereby contribute to the dominance of a starter strain within a fermentation (Abdalla et al., 2021). Furthermore, the production of exopolysaccharides by starter strains is a wanted feature in yogurts or cheese fermentation as it alters the texture during cheese ripening (Badel et al., 2011; De Vuyst et al., 2003; Torino et al., 2015; Zannini et al., 2016).

Another feature contributing to the dominance of a bacterial species in a fermentation process is acidification. Reducing the pH of the fermentation product can help to reduce the growth of other bacterial species (Hammes, 1990; Hu et al., 2022).

Dominance of bacteria can also be provided or aided by using bacteriocin-producing strains. The production of bacteriocins is a strain-specific feature that can help a strain dominate a fermentation and thereby prevent spoilage by other strains of species

(Etchells et al., 1966). Bacteriocins are peptides, proteins or protein complexes with antibiotic properties and usually target species or strains closely related to the producer (Klaenhammer, 1988). Bacteriocin-producing strains have been isolated and characterized for many LABs, including *Lactococcus lactis*, *Lactobacillus acidophilus*, *Lactiplantibacillus plantarum* and *Pediococcus* sp. (Barefoot and Klaenhammer, 1983; Leal-Sánchez et al., 2002; Papagianni and Anastasiadou, 2009; Roberts et al., 1992). Attempts to improve the production of bacteriocins in a mutagenized strain of *L. lactis* strain have also been made (Zhang et al., 2014).

Not only other microorganisms can influence the performance of a desired starter strain in a fermentation but also bacteriophages can have a significant impact on the fermentation fate. Bacteriophages in the lytic cycle can infect a host strain and be replicated in high numbers by this cell. After replication the cells burst, releasing the bacteriophages that then infect more bacterial cells of the same strain. Consequently, this causes a collapse of the bacterial culture and the fermentation is stopped (Garneau and Moineau, 2011; Uchida and Kanbe, 1993). Many bacteriophages have a narrow host specificity, a multi-strain fermentation can help to ensure the quality of the fermentation. Another approach is the generation of a phages insensitive starter strain by controlled exposure to lytic phages and propagating surviving cells (Higuchi et al., 1999). The reason for the insensitivity of a strain can be due to the alteration of the structure recognized by the phages or the degradation of the phage DNA by a CRISPR-Cas system (Lavelle et al., 2021; Marraffini, 2013; Wakinaka et al., 2022).
LABs are used as starter cultures and contribute to the flavor of fermented foods by the production of organic acids, ketones or alcohols.
In salt-reduced soy sauce moromi, the use of *T. halophilus* as a starter culture increases the amount of ethanol, 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF) and maltol (Singracha et al., 2017). The metabolism of cysteine and methionine by *Lacticaseibacillus paracasei* in cheese fermentation creates a sulfuric flavor (Wüthrich et al., 2018). In fermented sourdough, the aroma compounds associated with sourdough bread and the sour aroma are produced by the metabolic action of a multitude of LABs (*Lactiplantibacillus plantarum*, *Furfurilactobacillus rossiae* and *Lacticaseibacillus casei*) and yeasts (Winters et al., 2019).

Besides dominance and flavor contribution, the safety of a starter culture is also an important selection criterion. For example, many species of LAB are associated with the formation of biogenic amines (BAs) in fermented foods. LABs can be associated with the formation of histamine, tyramine, 2-phenylethylamine, cadaverine or putrescine. Low amounts of BAs can be metabolized in the human body by amino oxidases present in the gut (Barbieri et al., 2019). However, high amounts of BAs can cause food poisoning (Nout, 1994). Although many species are associated with the formation of biogenic amines, the ability for the formation is a strain-dependent feature. Therefore, to prevent the accumulation of BAs in food products, starter strains are tested for the formation of BAs and strains that do not produce biogenic amines are selected.

In conclusion, the above-discussed points highlight the need to characterize bacterial strains thoroughly (genomically and physiologically) before selecting them as starter cultures for food fermentation. Especially for the starter culture *T. halophilus*, many of the above-mentioned attributes have not yet been investigated.

1.5 *Tetragenococcus halophilus,* a key organism in seasoning sauce fermentations

In 1919 Orla-Jensen described "*Tetracoccus* No.1" isolated from salted anchovies (Orla-Jensen, 1919). In 1934, Mees further characterized the strain and described the species *Pediococcus (P.) halophilus* (Mees, 1934). In the 1950s the species *Pediococcus (P.) soyae* was described by Sakaguchi and by Iizuka and Yamasato independently. In both publications, *P. soyae* was identified as the major lactic acid bacterium in shoyu moromi (Iizuka and Yamasato, 1959; Sakaguchi, 1958). In more recent history, the names of *P. halophilus* and its synonym *P. soyae* were changed to *Tetragenococcus halophilus* (Collins et al., 1990). In 2012, the species was divided into two subspecies, subsp. *halophilus* with the type strain DSM20339^T and subsp. *flandriensis* with the type strain DSM23766^T based on the ability to grow in sugar-thick juice and to ferment D-lactose, D-raffinose and D-arabinose (Justé et al., 2012). *T. halophilus* is a Gram-positive, facultative anaerobic, cocci-shaped lactic acid bacterium that is most of the time organized in tetrads or pairs (Collins et al., 1990). Besides the fermented foods mentioned above, members of this species could also be isolated from various other sources including cheese rind brine, mountain snow or

contaminated thick juice (Justé et al., 2008b; Rodríguez et al., 2022; Uchida et al., 2014; Unno et al., 2020).

T. halophilus is known for its high tolerance to low aw values caused by high sodium chloride concentration or high amounts of sucrose. The growth is best between pH 7 - 9 in MRS or TSB medium at 30 °C (Justé et al., 2014; Kobayashi et al., 2004; Sakaguchi, 1958). The halotolerance of this organism is achieved by a multitude of mechanisms contributing to different parts of the effects of halo- /osmostress. One mechanism is the accumulation of high amounts of compatible solutes such as glycine betaine, proline, L-carnitine, choline and citrulline inside the cell (Baliarda et al., 2003; Chun et al., 2019; Robert et al., 2000). In bacteria, the pH homeostasis and osmostress are indirectly linked as both affect the intracellular proton concentrations, e.g. addition of 2.5 M NaCl resulted in shrunken cells of *Listeria monocytogenes* and decreased internal pH (Fang et al., 2004). Therefore, another way for *T. halophilus* to cope with both acid and osmostress is to use L-arginine *via* the arginine deiminase (ADI) pathway with a subsequent increase of the intracellular pH (Wakinaka and Watanabe, 2019). Only a few strains have been found so far that are associated with the production of biogenic amines, which also help the organism to increase the pH (Satomi et al., 2008). Regarding flavor contribution, *T. halophilus* is known to produce organic acids such as lactic acid and acetic acid (Harada et al., 2017; Lee et al., 2013). Further on, the species is known for its ability to ferment a diverse spectrum of carbohydrates, which in consequence was used to differentiate strains in a fermentation process (Roling and van Verseveld, 1996; Uchida, 1982; Villar et al., 1985). This raises the question, of whether strains of *T. halophilus* can be differentiated by their source of isolation (lupine moromi or soy sauce) based on their genomic adaptation towards the carbon sources of a specific substrate.

Single strains of *T. halophilus* have been used as a starter culture in ‘nam pla’ Thai fish sauce, fermentation of red alga Nori, Japanese ‘fish nukazuke’, in Chinese ‘Doubanjiang‘ horsebean-chili-paste, in Korean ‘saeu-jeot’ fish fermentation and soy sauce fermentation (Kim et al., 2019; Singracha et al., 2017; Uchida et al., 2014; Udomsil et al., 2011; Wakinaka et al., 2019; Wu et al., 2013). The strain *T. halophilus* MJ4 used as starter culture in ‘saeu-jeot’ could dominate the fermentation and thereby prevent the growth of biogenic amine-producing bacteria (Kim et al., 2019).

Monitoring of *T. halophilus* populations during a fermentation process has been done *via* qPCR. Therefore, a set of primers was designed to amplify the ITS region between the 16S and 23S rDNA sequences based on the genome sequence of NBRC 12172. However, this method does not allow for strain differentiation (Udomsil et al., 2016). Furthermore, Random Amplified Polymorphic DNA PCR (RAPD PCR) has been used to discriminate between strains (Justé et al., 2008a). However, strain differentiation by this method is not perfect as strains could potentially display bands of equal length but different sequences and would not be distinguishable. Furthermore, RAPD PCR is time-consuming and relies on high-quality gDNA. Thus, there is a high need for establishing a new easier way to identify and track strains during the fermentation process.

1.6 The yeast *Debaryomyces hansenii*

Debaryomyces hansenii is a yeast belonging to the family of the *Saccharomycetaceae* and the genus *Debaryomyces*, respectively. It is widely known for its halo- and osmo-tolerance and can tolerate up to 25 % (w/v) NaCl (Sørensen and Jakobsen, 1997; Stevenson et al., 2015). It is also recognized for its ability to ferment various carbon sources, such as D-xylose to xylitol (Dominguez, 1997; Rivas et al., 2009). It is frequently found in cheese brine, on a variety of surface-ripened cheese, in the fermentation of dry sausages or fermented Korean ‘Ganjang’ or ‘Doenjang’ (Osei Abunyewa et al., 2000; Petersen et al., 2002; Song et al., 2021). It can also be isolated from hypersaline lakes or other marine environments (Berrocal et al., 2016; Kumar et al., 2012). Members of this species are associated with lipase and protease activity in the fermentation processes (Jakobsen and Narvhus, 1996). The species presents great potential as a starter culture for grape must in the wine production process (Wyk et al., 2020). The species was previously divided into *D. hansenii* var. *hansenii* and var. *fabryi* based on different electrophoretic mobilities of the glucose-6-phosphate dehydrogenase and the maximum growth temperature (Nakase et al., 1998). However, this segregation was revered in 2008. The new delineation between *D. hansenii*, *D. fabryi* and *D. subglobosus* was done on Phylogenetic analyses from DNA fingerprints, the actin 1 (ACT1) gene and DNA reassociation values (Groenewald et al., 2008). Nonetheless, the species displays a big genetic diversity as chromosomal

polymorphism is common and some strains have been reported to be haploid (Corredor et al., 2003; Van der Walt et al., 1977).

As stated above, recently *D. hansenii* could be isolated from a lupine moromi fermentation (Lülf et al., 2021). As lupine moromi represents a hitherto undescribed origin for this yeast, it is unclear what potential benefit the fermentation process can have from this organism and if it's a spoiler or a wanted member of the microbiota. Furthermore, it is unknown what possible genetic modifications and traits the isolates from lupine moromi have, and how they can be discriminated from other *D. hansenii*. Lastly, it is unknown how many different strains may coexist in the fermentation.

2 Research Hypotheses

Soy sauces are consumed in great quantities around the world. Traditionally brewed soy sauces contain between 16 % - 20 % (w/v) sodium chloride, which is necessary for the preservation and the taste. The microbial consortium during the fermentation process of soybeans consists of different bacterial and yeast species. Especially, the 'Moromi' fermentation enables the growth of yeasts and bacteria including LAB. The key organisms in this stage are *Tetragenococcus halophilus* and *Zygosaccharomyces rouxii*. Recently, a similar process was developed with the use of fermented lupine beans instead of soybeans. The fermentation process of lupine beans has been adapted from the production process of soy sauce, but as the main ingredient and the sodium chloride content have been changed, it is unclear how the microbiota responds to that. In contrast to soybean-based seasoning sauces, the major species in the lupine-based seasoning sauce are *T. halophilus* and *D. hansenii* instead of *T. halophilus* and *Z. rouxii*. It is unclear how this alters the contribution of *T. halophilus* to the fermentation and how *T. halophilus* and *D. hansenii* adapted towards the lupine-based fermentation. Until now, no starter culture for the process based on lupine beans has been defined, to ensure a safe and consistent fermentation product. To tackle these issues, the following working hypotheses were defined in this thesis:

- *T. halophilus* is a key organism in the lupine season sauce fermentation analogous to its role in soy sauce fermentations.
- Replacing soy with lupine requires an adaptation of *T. halophilus* isolates
- *T. halophilus* strains from lupine moromi represent a diverse group and are genomically different compared to strains from other isolation sources.
- The genomic differences of *T. halophilus* strains are reflected by their metabolic properties.
- Lupine-adapted *T. halophilus* strains show increased competitiveness in lupine moromi fermentation.
- The yeasts isolated from lupine moromi are genomically different than the strains of *D. hansenii* from other isolation sources.

3 Material and methods

3.1 Species and strains

In this thesis, bacteria and yeast strains originated from the in-house strain collection of the institution “Technical Microbiology Weihenstephan” (TMW) were previously isolated from lupine moromi fermentations with varying NaCl concentrations by Rebekka lülf (Lülf et al., 2021). This includes 11 strains of *T. halophilus* and one of *D. hansenii*. The strains TMW 2.2260 and TMW 2.2261 were isolated from an experimental buckwheat fermentation, during a research internship of a student from Rebekka lülf. The type strains of *T. halophilus* subsp. *halophilus* DSM 20339^{T}, *T. halophilus* subsp. *flandriensis* DSM 23766^{T} and *T. osmophilus* DSM 23765^{T} as well as one additional *T. halophilus* subsp. *halophilus* DSM 20337 were bought from the “German collection of microorganisms and cell cultures” (DSMZ). *L. sakei* TMW 1.1474 and *L. curvatus* TMW 1.595 were only used one time, as positive control for the formation of biogenic amines. A list of strains, the accession number of the genomic sequence and their respective isolation source is given in Table 1.

MATERIAL AND METHODS

Table 1: Strains and genomes used in this thesis.

Accession number	Organism	Strain	References	Origin
CP027783.1 - 5.1	*T. osmophilus*	DSM 23765^T	(Justé et al., 2012)	Degraded sugar thick juice
CP027768.1 - 9.1	*T. halophilus* subsp. *flandriensis*	DSM 23766^T	(Justé et al., 2012)	Degraded sugar thick juice
GCF_003841405.1	*T. halophilus* subsp. *halophilus*	DSM 20339^T	(Chun et al., 2019)	Salted anchovy
NC_016052.1	*T. halophilus* subsp. *halophilus*	NBRC 12172	NITE 2011	Soy sauce mash
CP012047.1	*T. halophilus*	MJ4	(Kim et al., 2019)	Fish (anchovy) sauce
CP046246.1	*T. halophilus*	YJ1	Yonsei University	Salty fish sauce
CP020017.1	*T. halophilus*	KUD23	(Lee et al., 2018)	Korean soypaste doenjang
GCF_001593045.1	*T. halophilus*	FBL3	(Kim et al., 2017)	Galchijeot, fish sauce
GCF_018327505.1	*T. halophilus*	WJ7	(Shirakawa et al., 2020)	Fish nukazuke with rice bran
GCF_025379915.1	*T. halophilus*	TMW 2.2260	This thesis	Buckwheat moromi
-	*T. halophilus*	TMW 2.2261	This thesis	Buckwheat moromi
GCF_024137165.1	*T. halophilus*	TMW 2.2254	This thesis	Lupine moromi
-	*T. halophilus*	TMW 2.2255	This thesis	Lupine moromi
GCF_024137145.1	*T. halophilus*	TMW 2.2256	This thesis	Lupine moromi
GCF_024137175.1	*T. halophilus*	TMW 2.2257	This thesis	Lupine moromi
-	*T. halophilus*	TMW 2.2258	This thesis	Lupine moromi
-	*T. halophilus*	TMW 2.2259	This thesis	Lupine moromi
-	*T. halophilus*	TMW 2.2262	This thesis	Lupine moromi
GCF_024137125.1	*T. halophilus*	TMW 2.2263	This thesis	Lupine moromi
GCF_024137075.1	*T. halophilus*	TMW 2.2264	This thesis	Lupine moromi
-	*T. halophilus*	TMW 2.2265	This thesis	Lupine moromi
GCF_024137065.1	*T. halophilus*	TMW 2.2266	This thesis	Lupine moromi
GCF_024137205.1	*T. halophilus*	DSM 20337	Sequenced in this thesis	Soy sauce mash
GCF_002897555.1	*T. halophilus* subsp. *halophilus*	11	(Nishimura et al., 2017)	Soy sauce mash
GCF_002897515.1	*T. halophilus* subsp. *halophilus*	D10	(Nishimura et al., 2017)	Soy sauce mash
GCF_002897595.1	*T. halophilus* subsp. *halophilus*	D-86	(Nishimura et al., 2017)	Soy sauce mash
GCF_002897535.1	*T. halophilus* subsp. *halophilus*	NISL 7118	(Nishimura et al., 2017)	Soy sauce mash
GCF_002897435.1	*T. halophilus* subsp. *halophilus*	NISL 7126	(Nishimura et al., 2017)	Soy sauce mash
GCF_013393435.1	*T. halophilus*	KG12		Korean soy sauce
GCF_018327565.1	*T. halophilus*	YG2	(Shirakawa et al., 2020)	Soy sauce mash
GCF_018327545.1	*T. halophilus*	YA5	(Wakinaka and Watanabe, 2019)	Soy sauce mash
GCF_018327525.1	*T. halophilus*	YA163	(Shirakawa et al., 2020)	Soy sauce mash
GCF_016759905.1	*T. halophilus*	8C7	(Unno et al., 2020)	Brie de meaux cheese rind
-	*Latilactobacills sakei*	TMW 1.1474	(Freiding et al., 2011)	Sauerkraut
NZ_CP016470.1 - 1.1	*Latilactobacillus curvatus*	TMW 1.595	(Freiding et al., 2011)	Starter strain
GCA_024256405.1	*Debaryomyces hansenii*	TMW 3.1188	This thesis	Lupine moromi
-	*Debaryomyces hansenii*	281	This thesis	Sausage

-	*Debaryomyces hansenii*	310	This thesis	Sausage
-	*Debaryomyces hansenii*	311	This thesis	Sausage
-	*Debaryomyces hansenii*	325	This thesis	Sausage

3.2 Microbiological methods

3.2.1 Bacterial cultivation

All bacterial strains were grown at 30 °C in 15 mL or 50 mL closed tubes without shaking. All *T. halophilus*, *L. sakei* and *L. curvatus* strains were cultivated in a slightly modified MRS (De Man, Rogosa, Sharpe) media (Table 2) with the pH adjusted to 5.7. For the cultivation of the type strains DSM 23765^T and DSM 23766^T Caso Bouillon (Carl Roth, Karlsruhe, Germany) containing 5 % NaCl (w/v) (Carl Roth, Karlsruhe, Germany) was used.

Table 2: Modified MRS used in this thesis. The pH was set to pH 5.7 if not stated otherwise.

Chemicals	Quantity [g/L]	Company	Order No.
Tryptone/Peptone ex casein	10	Carl Roth GmbH	6681.4
Meat extract	10	Merck Millipore	1039792500
Yeast extract	5	Carl Roth GmbH	2904.4
Tween® 80	1	Gerbu biotechnik	2002
di-Potassium hydrogen phosphate trihydrate	2	Merck Millipore	1050991000
Sodium acetate trihydrate	5	Carl Roth GmbH	6779.1
di-Ammonium hydrogen citrate	2	Carl Roth GmbH	0267.2
Magnesium sulfate heptahydrate	0.2	Sigma-Aldrich	M1880
Manganese (II) sulphate monohydrate	0.05	Carl Roth GmbH	4487.1
Sodium chloride	50	Carl Roth GmbH	3957.2
D-Glucose monohydrate	22	Merck Millipore	108342
Agar-Agar	15	Carl Roth GmbH	1347.4

3.2.2 Cultivation of *T. halophilus* in Lupine MRS (LMRS)

To simulate a more lupine moromi-like environment, a modified version of the MRS media used in this thesis was created. Therefore, the 10 g/L tryptone/peptone ex casein and the 10 g/L meat extract in regular MRS were exchanged for 20 g/L lupin peptone (Solabia, Pantin, France) and the sodium chloride concentration was set to 13.5 % (w/v) NaCl.

3.2.3 Cultivation of *D. hansenii*

D. hansenii TMW 3.1188 was aerobically cultivated in shake flasks on a shaking incubator set to 200 rpm and 25 °C. The media used for the cultivation was a slightly modified malt extract (ME) DSMZ 90 media (Table 3).

Table 3: ME medium for cultivation of *D. hansenii*. The pH was adjusted to 5.6.

Chemicals	Quantity [g/L]	Company	Order No.
Malt extract	30	Carl Roth GmbH	6681.4
peptone ex soya	3	Merck Millipore	1039792500
Sodium chloride	50	Carl Roth GmbH	3957.2
Agar-Agar	15	Carl Roth GmbH	1347.4

3.2.4 Preparation of cryo-stocks from microbial cultures

For preparation of cryo-stocks, 1 mL of cultures from *T. halophilus* or *D. hansenii* was harvested in a 1.5 ml Eppendorf tube (Eppendorf, Hamburg, Germany) and the pellet was resuspended in 1 mL of either MRS or ME media. Next, the resuspended cells were mixed with 800 µL of sterile 80 % (v/v) glycerol (Gerbu Biotechnik GmbH, Heidelberg, Germany) and transferred to a Nunc® cryogenic vial (Sigma Aldrich, St.Louis, MO, USA) and stored at - 80 °C.

3.2.5 Quantification of microbial growth

To quantify the growth of *T. halophilus* in a sample, 100 µL of this sample was diluted with 900 µL full strength Ringer´s solution (Merck, Darmstadt, Germany) with 5 % (w/v) NaCl. This was then further diluted up to twelve times in the same manner. 100 µL of the diluted samples were then streaked out using glass beads (Carl Roth GmbH, Karlsruhe, Germany) on MRS agar plates with 5 % (w/v) NaCl. The agar plates were cultivated for 3 days at 30 °C in anaerobic jars with AnaeroGen™ packs (Fisher Scientific, Waltham, MA, USA).

3.2.6 Carbohydrate utilization of *T. halophilus* strains

The fermentation pattern of *T. halophilus* strains for a wide range of carbohydrates was generated in a small scale 200 µL pouches in the API 50 CHL kit (BioMérieux, Marcy l´Etoile, France). The assay was started by washing a culture of *T. halophilus* with full-strength Ringer´s solution (Merck, Darmstadt, Germany) and then resuspending it in API 50 CHL medium containing 2% (w/v) NaCl to create a solution with an OD_{600nm} of 0.3 (Table 4). Then, 200 µL of this solution was pipetted into every pouch and after 48 h at 30 °C the test strips were evaluated.

Table 4:Contents of the API CHL 50 medium with 2 % NaCl.

Chemicals	Quantity [g/L]	Company	Order No.
Polypeptone™	10	Becton, Dickinson & Company	211910
Yeast extract	5	Carl Roth GmbH	2904.4
Tween® 80	1	Gerbu biotechnik	2002
di-Potassium hydrogen phosphate trihydrate	2	Merck Millipore	1050991000
Sodium acetate trihydrate	5	Carl Roth GmbH	6779.1
di-Ammonium hydrogen citrate	2	Carl Roth GmbH	0267.2
Magnesium sulfate heptahydrate	0.2	Sigma-Aldrich	M1880
Manganese (II) sulphate monohydrate	0.05	Carl Roth GmbH	4487.1
Sodium chloride	20	Carl Roth GmbH	3957.2
Bromcresolpurpur	0,17	Merck Millipore	103025

3.2.7 Formation of biogenic amines in *T. halophilus*

Identification of histamine or tyramine-producing strains was done in a modified MRS medium according to the method previously described (Bover-Cid and Holzapfel, 1999). The medium was further modified as 0.2 mM Pyridoxal-5-phosphate and either with 10 mM L-Tyrosine di-sodium salt •2 H_2O or 10 mM L-Histidine monochloride • H_2O were added (Table 5). Cultivation was done in 1.5 ml Eppendorf tubes for 48 h at 30 °C. The production of histamine or tyramine was confirmed by a color change.

MATERIAL AND METHODS

Latilactobacillus sakei TMW 1.1474 and *L. curvatus* TMW 1.595 were used as positive controls.

Table 5: Medium used for the detection of biogenic amine formation. The pH was set to 5.3.

Chemicals	Quantity [g/L]	Company	Order No.
Tryptone/Peptone ex casein	5	Carl Roth GmbH	6681.4
Meat extract	5	Merck Millipore	1039792500
Yeast extract	5	Carl Roth GmbH	2904.4
Tween® 80	1	Gerbu biotechnik	2002
di-Potassium hydrogen phosphate trihydrate	2	Merck Millipore	1050991000
Sodium acetate trihydrate	5	Carl Roth GmbH	6779.1
di-Ammonium hydrogen citrate	2	Carl Roth GmbH	0267.2
Magnesium sulfate heptahydrate	0.2	Sigma-Aldrich	M1880
Manganese (II) sulphate monohydrate	0.05	Carl Roth GmbH	4487.1
Iron (II) sulfate heptahydrate	0,04	Carl Roth GmbH	3722.1
Sodium chloride	50	Carl Roth GmbH	3957.2
D-Glucose monohydrate	0,5	Merck Millipore	108342
Thiamine hydrochloride	0.01	Sigma Aldrich	T4625
Pyridoxal-5-phosphate	0.05	Sigma Aldrich	P9255
L-Tyrosine di-sodium salt x 2 H_2O	2,6119	Merck Millipore	1.02413.0100
L-Histidine monochloride x H_2O	2,0963	Merck Millipore	1.04350.0100

3.2.8 Salt tolerance of *T. halophilus* strains

Determination of the optimal NaCl concentration of *T. halophilus* strains was done in MRS with 2 % - 18% (w/v) NaCl. Therefore, an overnight culture was washed in full-strength Ringer´s solution (Merck, Darmstadt, Germany) and was then added to fresh media to a final OD_{600nm} of 0.002. After 48 h of static incubation at 30 °C the OD_{600nm} was measured in an SPECTROstar^{nano} plate reader (BMG Labtech, Ortenberg, Germany).

3.2.9 Growth of *T. halophilus* in LMRS

Verification of the growth of *T. halophilus* in the adapted LMRS was done by monitoring the growth and the pH for 40 h at 30 °C. Three different conditions were tested, without supplementation of any carbon source, with 10 mM D-galactose or with 10 mM D-glucose. During the 40 h, samples of the culture were taken and the OD_{600nm} was measured using a Novaspec Plus photometer (Biochrom Ltd., Camboume, UK).

3.2.10 Visual screening for EPS production on MRS agar

To screen for potential EPS production in *T. halophilus* strains, regular MRS medium was supplemented with 5% or 13.5 % NaCl (w/v) and 10 or 60 g/L of a carbon source. The EPS production with four different carbon sources was checked by the addition of D-lactose, D-galactose, D-glucose or sucrose. The agar plates contained 1.5 % (w/v) Agar-Agar. Cultures of *T. halophilus* strains TMW 2.2254, TMW 2.2256 and DSM 20339^{T} were used to test for the formation of slimy colonies after 12 days of incubation at 30 °C under anoxic conditions. The plates were visually evaluated and the results were documented by taking photographs.

3.3 Molecular biology methods

3.3.1 gDNA isolation from *T. halophilus*

Isolation of gDNA of *T. halophilus* strains was done from cultures grown statically in MRS (Table.2) for 48 h at 30 °C using an E.Z.N.A Bacterial DNA-Kit (Omega bio-tek, Norcross, Georgia, USA) according to the manufacturer’s instructions.

3.3.2 gDNA isolation from *D. hansenii*

Genomic DNA of *D. hansenii* was isolated using the E.Z.N.A fungal DNA Mini Kit (Omega bio-tek, Norcross, Georgia, USA) with the following modifications: approx. 150 mg sea sand (Merck KGaA, Darmstadt, Germany), 150 mg Glass beads 0.5 mm diameter (Scientific Industries, Bohemia, NY, USA) and 200 µL of TE buffer (10 mM Tris-HCl (Carl Roth, Karlsruhe, Germany), 1 mM EDTA pH 8 (Carl Roth, Karlsruhe, Germany) were used for resuspending the pellet. Followed by vortexing for 2 min to disrupt and partially lyse the cells. After that, the isolation was carried out following the manufacturer's instructions.

3.3.3 RAPD PCR for the strain identification of *T. halophilus*

Isolates identified as *T. halophilus* by MALDI-TOF MS were screened for unique DNA fingerprints generated by randomly amplified polymorphic DNA (RAPD) PCR using the M13V primer (Ehrmann et al., 2003). 50 ng of gDNA was used as input for the RAPD-PCR. Therefore, gDNA was mixed with a RAPD-PCR mastermix consisting of 2.5 µL 10X PCR-Buffer without $MgCl_2$, 200 µM of DNTPs (final concentration), 5 mM $MgCl_2$, 500 nM (Eurofins Genomics, Ebersberg, Germany) of every primer, 0.25 µL (1.25 U) of taq polymerase (MP Biomedicals, Eschwege, Germany) and 20.25 µL of 0.22 µm filtered H_2O. The PCR was done in a Mastercycler® gradient (Eppendorf, Hamburg, Germany) machine (Table 6).

Table 6: RAPD-PCR protocol for *T. halophilus*. The M13V primer was used as the only primer.

Step	Cycles	Temperature	Time
1	3	94 °C	180 s
2	3	40 °C	300 s
3	3	72 °C	300 s
4	32	94 °C	60 s
5	32	60 °C	120 s
6	32	72 °C	180 s
7	1	72 °C	300 s

3.3.4 Strain-specific PCR using marker genes for strain identification.

The identification of strains was also done *via* the amplification of specific marker genes of each *T. halophilus* strain. These marker genes were found using the BADGE program (Behr et al., 2016). Using these genes, a primer set was developed that allows

the identification of strains by a strain-specific pattern. The primer set can be found in (Table 7). This PCR-Mix consisted of 2.5 µL 10X PCR-Buffer, 200 µM of DNTPs (final concentration), 31.25 nM (Eurofins Genomics, Ebersberg, Germany) of every primer used in the set, 0.25 µL (1.25 U) of taq polymerase (MP Biomedicals, Eschwege, Germany) and 20.25 µL of 0.22 µm filtered H_2O. The PCR was run on a Mastercycler® gradient (Eppendorf, Hamburg, Germany) machine (Table 8).

Table 7: Primer sequences designed for the strain identification using marker genes and their target locus.

Primer name	Strain	Gene	Locus	Sequence
1rbsU 2rbsU	all	D-ribose transporter	IBK48_08720	**F:** GGTTGGGGATTATTTCCAACTATCGCATCTAAATTCG **R:** CTGTTAGGGTAGCTAAGGCAAC
3araA 4araA	DSM20337, DSM20339^T, TMW 2.2256, TMW 2.2263, TMW 2.2266	L-arabinose isomerase	IBK48_08390	**F:** GCACGGGGATCGTGAATATGGTTAT **R:** CTAATCTCAACCGACTCAATACCAAAC
5xylA 6xylA	TMW 2.2266	Xylose isomerase	HXW84_10810	**F:** GTAGTTCGTTAAGGATTCGAGATCTTC **R:** GCTCAATAGGGCGGTATTTGTTAATAG
9BppUphage 10BppUphage	TMW 2.2263	BppU family phage baseplate upper protein	HXW81_02100	**F:** GATCAATGATTACACACGTGCTTC **R:** GCTGATATACTGACAATCTTCCAGAC
11_08505 12_08505	TMW 2.2257	ATP-dependent endonuclease	HXW75_08505	**F:** GAGTGCTTTACGCCACAATTC **R:** GATCTGTCAGTTCATCATGAGTTAGAC
15_11195 16_11195	TMW 2.2254	sigma 54-interacting transcriptional regulator	HV360_11195	**F:** GTCGTACTCAAGCTAGTCGTG **R:** GCGACTTTACGTGCATTGGATA
17_00840 18_00840	TMW 2.2260	ABC transporter ATP-binding protein	HXW78_00840	**F:** CATTTACAAACAGCGATTGCTGAG **R:** GCTGTTGTTGACCTTGTGAAATATTG
19_12850 20_12850	DSM20339^T	MarR family transcriptional regulator	C7K42_12850	**F:** GAGAACTTGCTGATGAATTGGG **R:** TTACATTCTCCTACCAAAGTGCC
21_06690 22_06690	TMW 2.2256	6-phospho-beta-galactosidase	lacG	**F:** GATTTTCTATTAGGTGGTGCAACTG **R:** GATGTAAGATCTCTTCCCTCGTTTAATTC

Table 8: PCR program for the marker gene strain identification PCR.

Step	Cycles	Temperature	Time
1	1	96 °C	30 s
2	24	95 °C	15 s
3	24	52 °C	25 s
4	24	68 °C	160 s
5	1	72 °C	200 s

3.3.5 Primer design and application of strain-specific multiplex PCR

Strain-specific primers were designed with version 1.0 of the program “Rapid identification of PCR primers for unique core sequences” (RUCS) (Thomsen et al., 2017). Every primer pair was checked for potential secondary structures using the NetPrimer tool available on the website from Premier Biosoft (https://www.premierbiosoft.com/netprimer/). Furthermore, only primer pairs were chosen that have a calculated annealing temperature using the NEB tm calculator (https://tmcalculator.neb.com/#!/main#!%2F) of 55 °C using standard Taq polymerase. However, the annealing temperature was increased to 57 °C to increase specificity. The final primer set, targeted gene and loci as well as the sequence can be found in Table 9.

The set consisted of eight strains TMW 2.2254, TMW 2.2256, TMW 2.2257, TMW 2.2260, TMW 2.2263, TMW 2.2264, TMW 2.2266 and DSM 20339^{T}. The strains were picked using sterile toothpicks and smeared into PCR tubes (Starlab International GmbH, Hamburg, Germany). Afterwards, 25 µL of a PCR-Mix was added. This PCR-Mix consisted of 2.5 µL 10X PCR-Buffer, 200 µM of DNTPs (final concentration), 31.25 nM (Eurofins Genomics, Ebersberg, Germany) of every primer used in the set, 0.25 µL (1.25 U) of taq polymerase (MP Biomedicals, Eschwege, Germany) and 21.25 µl of 0.22 µm filtered H_2O. PCR was done on a Mastercycler® gradient (Eppendorf, Hamburg, Germany) machine (Table 10).

Table 9: Multiplex PCR primer sequences and the target genes. The primers were designed using RUCS software for the strain identification. The first column displays the targeted strain. The second and third columns show the targeted gene with the respective loci. The fourth column shows the sequences of the forward and reverse primers.

Strain	Gene	Loci	Sequence
TMW 2.2254	PTS EIIC, alpha/beta hydrolase, Gfo/Idh/MocA family oxidoreductase	HV360_08825-08835	**F:** CGTCCGGGTGTCTTAAAGTCA **R:** AGCTCGCTATTTGGGTGGAAA
TMW 2.2256	Ethanolamine-lyase B	HXW74_05975	**F:** AATCCCGACACAAACTAGCGT **R:** GGACCGTAGCTGTTTCTTGGA
TMW 2.2257	nucleoside hydrolase, multidrug transporter	HXW75_03115-03120	**F:** CCTAAACCGTCTTCACCGTGA **R:** CTCCTCAAGGTCGTCAAAGCT
TMW 2.2260	PTS-IIC, dapA	HXW78_09960-09965	**F:** AAGTGCATGTCCTTCCCCATT **R:** CGCTTGTGTGTTAGCTTTGCT
TMW 2.2263	Phage terminase	HXW81_02015-02020	**F:** GAGCCAACTGTAGACGACACA **R:** CACCACGTCCCATTGTTAGGA
TMW 2.2264	Sulfatase	HXW82_1240-01245	**F:** TGCCAGCTCGTCGAGATATTC **R:** CCAATGTAGGCTCACCACCTT
TMW 2.2266	Transglycosylase	HXW84_06150	**F:** GTCAATGGCTTGGTGTTTCCC **R:** TTCCGTTGGCTGGATTCGTAA
DSM20339^T	Hypothetical protein	C7K42_05055-05065	**F:** CACCCGTTAACCCCATTACCA **R:** AGCAGGTTTGTCGGCATAAGA

Table 10:PCR program for the multiplex PCR for the strain identification of *T. halophilus*.

Step	Cycles	Temperature	Time
1	1	95 °C	300 s
2	24	95 °C	30 s
3	24	57 °C	30 s
4	24	72 °C	120 s
5	1	72 °C	300 s

3.3.6 Strain differentiation of *D. hansenii* isolates using RAPD PCR

D. hansenii isolates from lupine moromi fermentation were differentiated by RAPD PCR using the M13 and M13V primer (Andrade et al., 2006; Ehrmann et al., 2003; Huey and Hall, 1989).

The isolated gDNA was mixed with 24 µL of a PCR-Mix. This PCR-Mix consisted of 2.5 µL 10X PCR-Buffer, 200 µM of dNTPs (final concentration), 500 nM (Eurofins Genomics, Ebersberg, Germany) of every primer, 0.25 µL (1.25 U) of taq polymerase

(MP Biomedicals, Eschwege, Germany) and 20.25 µL of 0.22 µm filtered H_2O. The PCR was done on a Mastercycler® gradient (Eppendorf, Hamburg, Germany) machine. For the RAPD PCR the 10X PCR buffer was exchanged for a 10X buffer without $MgCl_2$ was used and $MgCl_2$ was added to a final concentration of 5 mM (Table 11).

Table 11: RAPD-PCR program for *D. hansenii*. The PCR was done with either the M13 or the M13V primer.

Step	Cycles	Temperature	Time
1	3	94 °C	180 s
2	3	40 °C	300 s
3	3	72 °C	300 s
4	32	94 °C	60 s
5	32	60 °C	120 s
6	32	72 °C	180 s
7	1	72 °C	300 s

3.3.7 Amplification ITS1, 5.8S rRNA and ITS2 regions of *D. hansenii* TMW 3.1188

For the classification of TMW 3.1188 into the species *D. hansenii* the primer V9G and LR5 were used (Hoog and Ende, 1998; Wagner and Fischer, 2001). The PCR was done with the proofreading Q5-Polymerase (New England Biolabs, Ipswich, MA, USA). A 50 µL PCR reaction consisted of 1 µL of gDNA from TMW 3.1188, 10 µL of 5X Q5 Reaction Buffer, 200 µM dNTPs, 0.5 µM of every primer, 0.5 U (0.02 U/µL) of Q5 Polymerase and 32.5 µL of 0.22 µm filtered H_2O. The PCR was done on a Mastercycler® gradient (Eppendorf, Hamburg, Germany) machine. The PCR product included part of the small ribosomal subunit, the ITS1 region, the 5.8S rRNA, the ITS2 region and parts of the large ribosomal subunit. The PCR product was sequenced afterwards from the 5'-3' and from the 3'-5' side using the V9G and LR5 primer (Table 12).

Table 12: PCR-Program for the partial amplification of the small ribosomal subunit of *D. hansenii*, the ITS1 region, the 5.8 s rRNA, the ITS2 region of *D. hansenii* TMW 3.1188 using the primer V9G and LR5.

Step	Cycles	Temperature	Time
1	1	98 °C	30 s
2	24	98 °C	20 s
3	24	62 °C	20 s
4	24	72 °C	80 s
5	1	72 °C	120 s

3.3.8 Agarose gel electrophoresis

All PCR reactions and gDNA samples were analyzed for the correct size or the amount of degradation by agarose gel electrophoresis. Therefore, agarose gels with 1.5 % (w/v) or 0.8 % (w/v) agarose were prepared by mixing 0.5 % (w/v) TBE (45 mM Tris-HCl (Gerbu Biotechnik GmbH, Heidelberg, Germany), 45 mM boric acid, 1 mM Ethylenediaminetetraacetic acid (EDTA) (Gerbu Biotechnik GmbH, Heidelberg, Germany) with the correct amount of agarose (Biozym Scientific GmbH, Hessisch Oldendorf, Germany). Afterwards, the mixture was repeatedly heated until the solution was clear. Then, the solution was cooled down to 45 °C and then poured into gel chambers with combs (Peqlab Biotechnologie GmbH, Erlangen, Germany). After polymerization and cooling of the gels, the PCR reactions or the gDNA samples were mixed with loading dye (Thermo Fisher Scientific, Waltham, MA, USA) and filled into the pockets of the gel. To analyze the size of the fragments, either a 100 bp ladder or a 1 kb Plus DNA ladder was used (Thermo Fisher Scientific, Waltham, MA, USA). To separate the samples a current between 200 mA and 400 mA was applied to gel chambers by an electrophoresis Power Pack P25 power supply (Biometra GmbH, Göttingen, Germany). The runtime dependent on the gel size and the expected fragment size. After separation, the gels were stained with dimidium bromide (Carl Roth GmbH, Karlsruhe, Germany) and then illuminated with and UVT-28M transilluminator (Herolab, Wiesloch, Germany).

3.4 Analytical methods

3.4.1 MALDI-TOF MS profiling for the verification of microorganisms

Identification and verification of bacterial and yeast colonies were done initially and routinely with MALDI-TOF-MS biotyper (Bruker Daltonics, Billerica, MA, USA). Therefore, the organic solvent solution was made by mixing HPLC grade ddH2O with acetonitrile (ACN) and trifluoroacetic acid (TFA) in the ratios 19:20:1. Next, the matrix solution was prepared by solubilizing a-cyano-4-hydroxycinnamic acid with a concentration of 10 mg/ml in the organic solvent solution. Using sterile toothpicks, single colonies were smeared onto a stainless-steel target (Bruker Daltonics, Billerica, MA, USA) and overlaid with 1 µL 70% (w/v) formic acid (Sigma Aldrich St. Louis, MO, USA). After airdrying, 1 µL of the matrix solution is pipetted onto each dried spot. After another drying period, the target is measured in the MALDI-TOF-MS biotyper.

3.4.2 Qualitative carbohydrate measurement with HPAEC-PAD

A Dionex ICSW 3000 HPAEC-PAD (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a CarboPacPA1 preparative IC column (4 × 250 mm) (Thermo Fisher Scientific, Waltham, MA, USA) secured by a CarboPacPA1 standard bore guard column (4 × 50 mm) (Thermo Fisher Scientific, Waltham, MA, USA) was used for the qualitative detection of carbohydrates. For the analysis, a 27 min isocratic method with 10 mM sodium hydroxide in deionized water at a flow rate of 1 ml /min was used. The column temperature was set to 30°C.

3.5 Genomic analyses

3.5.1 Genome *de novo* assembly and analyses for *T. halophilus*

The genomes of all *T. halophilus* TMW strains in this thesis were sequenced by Eurofins Genomics (Konstanz, Germany) using an Illumina NovaSeq 6000 system using the sequence mode S2 PE150 XP.

The assembly was done with the version 0.4.8 of the unicycler tool at the galaxy website (https://usegalaxy.eu/). The settings were set to default, besides the exclusion of contigs shorter than 1 kb. The assembled genomes were annotated with three databases the rapid annotation using subsystems technology (RAST) (Aziz et al., 2008), NCBI PGAP (Li et al., 2021) and the “The institute for Genomic Research” (TIGR)(Ouyang et al., 2007). Genomic relatedness was determined using the average nucleotide indices (ANIb) values and the Genome-to-Genome distance calculator ver. 2.1 (Meier-Kolthoff et al., 2013). ANIb values were calculated using the JSpeciesWS version 3.7.8 (Richter et al., 2016). To determine strain-specific ORFs the “Blast Diagnostic Gene Finder” (BADGE) was used (Behr et al., 2016). The predicted proteins were further checked with pBlast (https://blast.ncbi.nlm.nih.gov/Blast.cgi), smartBlast (https://blast.ncbi.nlm.nih.gov/smartblast/) and CDS search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) from NCBI (Lu et al., 2020). Prophage regions were predicted using the **PHA**ge **S**earch **T**ool **E**nhanced **R**elease (PHASTER) (Arndt et al., 2016). Identification of CRISPR regions and CAS proteins was done using Crisprdb (https://crisprcas.i2bc.paris-saclay.fr) (Couvin et al., 2018). BAGEL4 was used to predict potential bacteriocin clusters (de Jong et al., 2006). Perseus version 1.6.14.0 (Tyanova et al., 2016) was used for the hierarchical clustering of cluster/genes with the distance calculation set to Euclidean and the linkage set to average. One thousand iterations were set as the maximum.

Assignment of the “Clusters of Orthologous Genes” (COGs) categories to the pan and core genome of *T. halophilus* set was done by generating a BADGE analysis output with all settings set to standard based on nucleotide sequences (Galperin et al., 2015; Tatusov et al., 1997). From this output, the pan-, core- and accessory-genome could be filtered by formatting the output as an Excel table and filtering in the “percent_occurence” column for only “100” for the core or all numbers but “100” and “empty” for the accessory genome. A second table was generated using the eggnog-mapper website (http://eggnog-mapper.embl.de/) with eggnog-mapper V 2.1.9, with all settings kept at standard (Cantalapiedra et al., 2021). For the eggnog-mapper software, a concatenated fasta file consisting of the nucleotide sequences from all CDS from all *T. halophilus* genomes was used as an input and the “CDS” data was selected. Finally, both output tables were combined using the “matching rows by name” function in Perseus v 1.6.14.0 and selecting the columns in both output tables containing the locus tags from the strains (Tyanova et al., 2016).

3.5.2 *De novo* assembly and analyses for *D. hansenii* TMW 3.1188

D. hansenii was cultivated as described above and high molecular gDNA was isolated from an overnight culture by lysis of the cells using R-zymolyase (Zymo Research, Freiburg, Germany) followed by a phenol-chloroform extraction and digestion of proteins with proteinase K.

The isolated gDNA was then sent for whole genome sequencing using the PacBio Single Molecule, Real-Time (SMRT) technology at the Functional Genomics Center Zurich (Zurich, Switzerland). The SMRTbell library was produced using the SMRTbell Express Template Prep Kit 2.0 (Pacific Bioscience, Menlo Park, California, USA). Quality control of the input DNA and the resulting library was done using a Qubit Fluorometer dsDNA High Sensitivity assay (Life Technologies, Carlsbad, California, USA) and a Femto Pulse Device (Agilent, Santa Clara, California, USA). The library was size selected to enrich fragments >17.5 Kb using a Blue Pippin Device (Sage Science, Beverly, Massachusetts, USA). Next, a Pacific Biosciences Sequel instrument was programmed to sequence libraries in Continuous Long Reads (CLR) mode using a Sequel Sequencing Kit 3.0 (Pacific Biosciences, Menlo Park, California, USA). Lastly, data quality was checked using the Pacific Biosciences SMRT Link software. *De novo* genome assembly was done with HGAP4 in SMRTLink version

10.0.0.108728 with the expected genome size set to 12 Mbp and all other parameters set to standard.

FASTANI version 1.3 (Jain et al., 2018) was used to calculate ANIb values to other *D. hansenii* strains and genomediff from GenomeTools V. 1.6.2 (Gremme et al., 2013) was used to calculate kr distance (Haubold et al., 2009) between *D. hansenii* strains. D-genies online tool (Cabanettes and Klopp, 2018) was used to map contigs of TMW 3.1188 to the type strain CBS767^{T}.

Gene prediction and annotation were done with YGAP (Proux-Wéra et al., 2012) and KEGG (Kanehisa et al., 2022).

3.5.3 Phylogenetic analyses

Sequence alignments for the *D. hansenii ACT1* genes and the alignments of the concatenated sequences of *fusA*, *gyrA*, *gyrB*, *lepA*, *pyrG*, *recA* and *rpoD* of *T. halophilus* as well as the alignments of the 16S rRNA genes of *T. halophilus* were done in mega7 with Clustalω algorithm (Kumar et al., 2016). For both species, the dendrograms were constructed with the neighbor-joining algorithm (Saitou and Nei, 1987) with 1000 bootstrap replicates (Felsenstein, 1985). Additionally, for *T. halophilus* the dendrograms were also constructed using the Maximum Composite Likelihood method (Tamura et al., 2004; Tamura and Nei, 1993) followed by bootstrapping with 1000 replicates (Felsenstein, 1985).

3.5.4 Clustering of RAPD DNA fingerprints from *T. halophilus*

RAPD fingerprints of *T. halophilus* isolates were processed and clustered using Bionumerics v 7.6.2 (Applied Maths, Sint-Martens-Latem, Belgium). To normalize the RAPD fingerprints and align the DNA bands, the DNA ladder was used as a reference. The UPGMA (unweighted pair group method with arithmetic mean) algorithm with a 4 % dice similarity coefficient and 1 % tolerance change was used to calculate similarities and distances for the hierarchical cluster analysis.

3.6 Transcriptome analyses

3.6.1 Cultivation and sample preparation

TMW 2.2254 and TMW 2.2256 were grown in LMRS with 13.5 % NaCl (w/v) and 2 %(w/v) D-galactose for 48 h at 30 °C, to obtain a preculture. 2% (w/v) of these cultures were then used to inoculate 100 mL of LMRS with 13.5 % NaCl (w/v) and grown for another 24 h under the same conditions. On the day of the experiment, 100 mL of fresh LMRS medium was then inoculated with 1x10^8 CFU/mL of *T. halophilus* cells and incubated statically for 10 h at 25 °C in a 125 ml Erlenmeyer flask sealed with a cotton plug. After 1 and 10 h, 40 ml of the cells were harvested by centrifugation at 10.000 x *g* for 10 min at 4 °C. The pelleted cells were then handled on ice and mixed with 2 mL RNA-later solution (Thermo Fisher Scientific, Waltham, MA, USA), after 5 min the cells were flash-frozen in liquid nitrogen.

To determine the growth, 100 µL of diluted culture was plated onto LMRS plates containing 13.5 % (w/v) NaCl using glass beads (Carl Roth GmbH, Karlsruhe, Germany). The agar plates were cultivated for 3 days at 30 °C in anaerobic jars with AnaeroGen™ packs (Fisher Scientific, Waltham, MA, USA). Plates with 20 to 200 colonies were considered for the calculation of the cell count.

RNA isolation was done with the RNAeasy mini kit (Qiagen, Hilden, Germany) with several steps added to ensure complete lysis. In the beginning, frozen cell pellets were thawed and resuspended in the liquid surrounding them. Then, the liquid containing the cells was transferred to a 1.5 mL Eppendorf tube (Eppendorf, Hamburg, Germany) and pellet again by centrifugation at 16.000 x *g* for 5 min, the RNA-later solution was then discarded without disruption of the cell pellet. TE buffer pH 8.5 supplemented with 50 mg/mL lysozyme (24,000 kU/mL, SERVA Electrophoresis GmbH, Heidelberg, Germany) was added to lyse the cells for 25 min at RT. Next, RLT with 2 M DTT and acid-washed glass beads (212 - 300 µm) (Sigma Aldrich, St. Louis, MO, USA) were added. The tubes were then transferred into a homogenizer (MP Fastprep-24, Fisher Scientific, Waltham, MA, USA) set to a shaking frequency of 6.5 m/s for 25 s, this step was repeated three times. Afterwards, the glass beads were collected by centrifugation at 10.000 rpm for 10 min at RT, and the supernatant containing the lysate was transferred to a fresh 2 ml tube and mixed with 500 µL 96% (v/v) ethanol. Then, all steps were carried out following the manufacturer`s instructions.

3.6.2 RNA sequencing and bioinformatic analyses

RNA integrity, library preparation and sequencing were done by Eurofins Genomics (Eurofins Genomics, Konstanz, Germany). Sequencing was done on an Illumina HiSeq 2500 machine. Prior to the shipment to Eurofins, the RNA content was determined using a Nanodrop spectrophotometer (Nanodrop 1000 3.6.0, PeQLab Biotechnologie GmbH, Erlangen, Germany).

After Sequencing the raw reads were mapped to genomes using BWA-MEM (Li, 2013). After the mapping of the reads the raw read counts were counted using the tool featurecounts (Liao et al., 2014), only counting the overlapping features with unique mapping positions and a minimum mapping quality score of 10. Paired-end counts were counted only once and only if they mapped onto the same contig. Reads were assigned based on the highest number of matching bases, in case they mapped multiple times. Normalization was done using the Trimmed Mean of M-values (TMM) in the edgeR package (version 3.16.5) (Robinson et al., 2010). Genes with a fold-change of ≥ 2 or ≤ - 2 and p-value < 0.05 and a false discovery rate (FDR) <0.01 were defined as differentially expressed genes (DEGs).

3.7 Lupine moromi pilot scale fermentation

Lupine moromi pilot-scale fermentations were conducted to analyze the growth performance of *T. halophilus* strains in a near real-world environment at a lab scale. Therefore, lupine koji was prepared by the Purvegan company (Ramsen, Germany) by toasting lupine seeds that were then soaked in water, cracked, mashed and subsequently fermented at 28 - 35 °C for two days by an *Aspergillus oryzae* in an industrial fermentation tank in the factory. Prior to the use, the koji was heated to 80°C for 10 min in an oven and then packaged and sealed using a vacuum packing machine. This was done to stop the koji fermentation and store it for a longer time.

To ensure equal distribution of the selected *T. halophilus* strains across all biological replicates, sterile tap water containing 13.5 % (w/v) NaCl was inoculated with cells from every strain to create an inoculation solution. Finally, to create a lab-scale pilot fermentation 20 g of lupine koji was mixed with inoculation solution to a volume of 50 mL in a 50 mL falcon tube.

4 Results

4.1 Screening and identification of *T. halophilus* strains from lupine moromi

4.1.1 Multiple strains could be identified among the isolates from lupine moromi.

To analyze the strain diversity among 68 *T. halophilus* isolates RAPD PCR with the primer M13V was done. This set included 58 isolates from lupine moromi, seven from an experimental buckwheat fermentation and three from the DSMZ. The resulting fingerprints were compared to reference strains from the DSMZ strain collection, including the type strain from *T. osmophilus* DSM 23765^T, the type strain from *T. halophilus* subsp. *flandriensis* DSM 23766^T and DSM 20337 isolated from soy sauce mash. For comparison and clustering of the fingerprints Bionumerics v.7.6.2 was used (Figure 2).

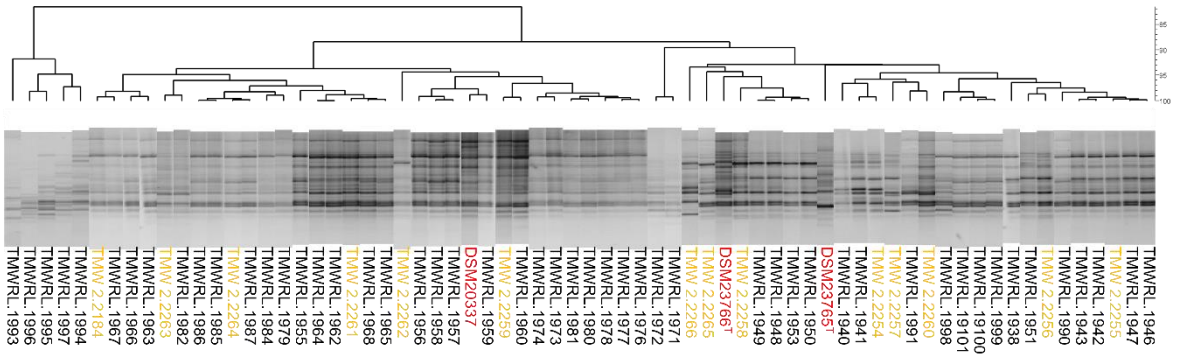

Figure 2: Resulting fingerprints of the RAPD PCR of *T. halophilus* strains. Clustering and phylogenetic analysis of the DNA fingerprints was done using Bionumerics v7.6.2. The clustering was carried out with the UPGMA algorithm and with the Pearson correlation chosen as the similarity coefficient with 8% optimization and 1 % curve smoothing. Strains marked red were taken from the DSMZ strain collection. Yellowed-colored strains were added to the TMW in-house strain collection. Black-colored strains were not added to the in-house strain collection and were not further analyzed.

As a result of the comparison, 14 isolates were chosen to be added to the in-house strain collection. These 14 strains were chosen as the RAPD PCR yielded enough bands for specific identification of the strains. No isolates from the cluster with TMWRL. 1967, 1971, 1972, 1979, 1984, 1993, 1995, 1996 and 1997 were not considered for the final strain set as the RAPD PCR only yielded a few clear bands. The strains TMW 2.2254 – TMW 2.2266 were chosen for whole genome sequencing using Illumina. Additionally, the strain DSM 20337 was also sequenced as no sequence for this strain was available in the NCBI database. After *de novo* genome assembly with

unicycler, the genomes were compared using the L50 and N50 values of the assemblies as well as the ANI values to each other. This enabled the identification of identical strains, as TMW 2.2255, TMW 2.2259, TMW 2.2261 and TMW 2.2264 had ANI values of between 99.99 to 100 to each other (Table 22). A similar high identity was observed for the set consisting of TMW 2.2256, TMW 2.2258, TMW 2.2262 and TMW 2.2265. This set had an ANI value of 99.93 to 100 to each other. As a result of this analysis, the final set of strains isolated from lupine moromi consisted of TMW 2.2254, TMW 2.2256, TMW 2.2257, TMW 2.2263, TMW 2.2264 and TMW 2.2266. The strain TMW 2.2260 was also used in this thesis, although it was isolated from an experimental buckwheat fermentation supplied with lupine protein (Figure 3).

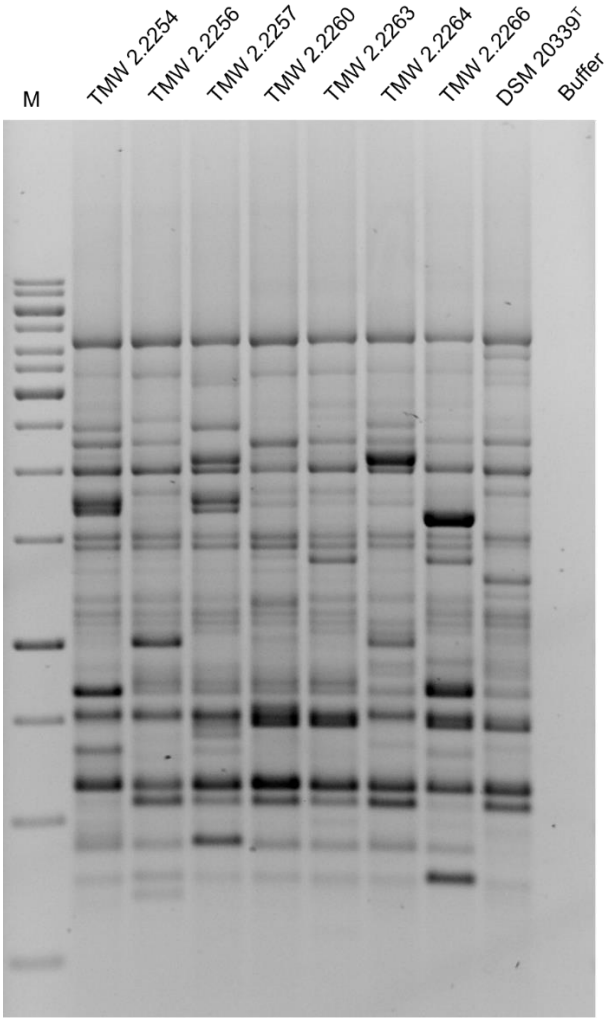

Figure 3: RAPD DNA fingerprints from the final set of *T. halophilus* strains. The RAPD PCR was done with the M13V primer. Marker = Generuler 1 Kb DNA ladder (Thermo Scientific, Waltham, USA); Buffer = PCR master mix without any DNA. Figure modified from Link et al., 2023.

The resulting DNA fingerprints from the RAPD PCR are unique within the final set. This set is used for further characterization and analyses.

4.2 Physiological characterization of *T. halophilus* strains from lupine moromi

To characterize the selected set in more detail and potentially detect characteristics only present in strains from lupine moromi, the strains were analyzed for their ability to form histamine or tyramine, the optimal concentration of sodium chloride for growth and the ability to ferment various carbohydrate sources.

Parts of the results have been published in Link et al., 2021.

4.2.1 Biogenic amine formation of *Tetragenococcus* strains including new isolates from lupine moromi.

Biogenic amines are unwanted compounds in a fermentation, the newly isolated strains from lupine moromi were screened for the formation of tyramine or histamine (Figure 4). The type strains from *T. osmophilus* DSM 23765^T and from *T. halophilus* subsp. *flandriensis* DSM 23766^T as well as the strain DSM 20337 were also screened for the formation as no data on the formation of biogenic amines was available. The strains *L. sakei* TMW 1.1474 and *L. curvatus* TMW 1.595 were used as positive controls for the formation of histamine or tyramine.
After visual evaluation, none of the strains but the positive controls produced biogenic amines which resulted in a color change of the medium.

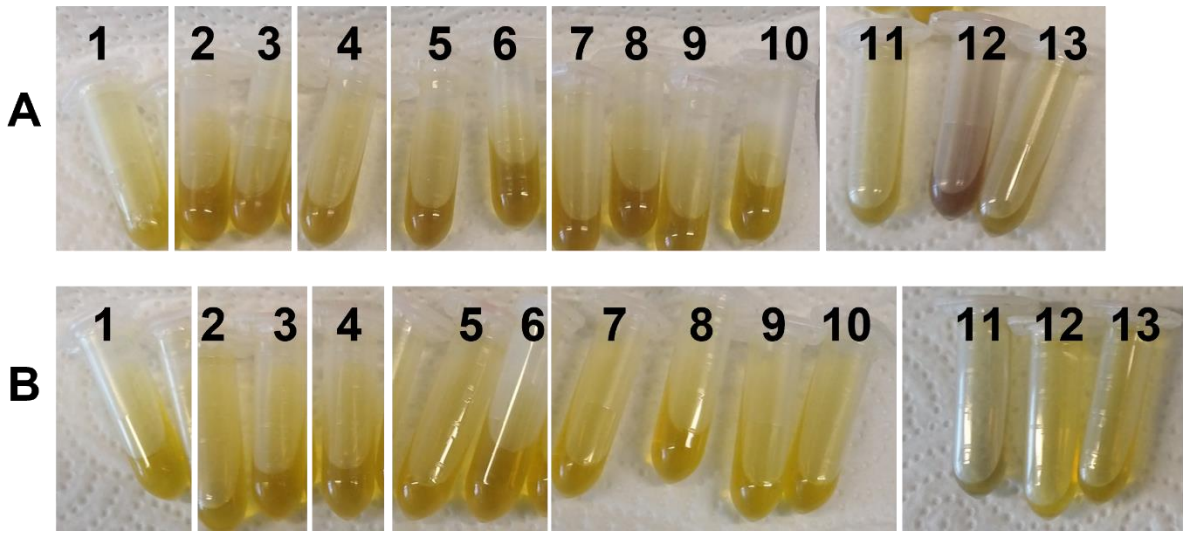

Figure 4: Visual screening of the biogenic amine formation in *T. halophilus* strains from lupine moromi. Detection of biogenic amine-producing strains using indicator medium. Photos were taken after 48 h of incubation at 30 °C. A) Medium supplemented with 10 mM L-tyrosine. B) Medium supplemented with 10 mM L-histidine. Numbering: 1 = TMW 2.2254; 2 = TMW 2.2256; 3 = TMW 2.2257; 4 = TMW 2.2260; 5 = TMW 2.2263; 6 = TMW 2.2264; 7 = TMW 2.2266; 8 = DSM 23765^T; 9 = DSM 23766^T; 10 = DSM 20337. Positive controls: 11 = TMW 1.1474 (HDC +); 12 = TMW 1.595 (TDC +). Negative control 13 = uninoculated medium.

4.2.2 Fermentable carbohydrate sources

The analytical profile index test (API-ID) was used to characterize and potentially detect an origin-specific fermentation pattern among the strains isolated from lupine moromi and the strain from a buckwheat fermentation. The type strains of *T. osmophilus*, *T. halophilus* subsp. *flandriensis* and *T. halophilus* subsp. *halophilus* and the strain DSM 20337 were used as references (Table 13). After 48 h at 30 °C the stripes were visually evaluated.

Table 13: Determination of the fermentable carbohydrate sources of *T. halophilus*. The API-ID 50 CHL stripes were used. The standard API-CHL 50 medium was supplemented with 2% NaCl (w/v). All tested strains were able to produce acid from D-glucose, D-fructose, D-mannose, N-acetyl-Glucosamine, arbutin, esculin/iron citrate, salicin, D-cellobiose, D-maltose, sucrose, D-trehalose, D-gentobiose, D-turanose. None of the strains could produce acid from erythritol, L-xylose, D-adonitol, methyl-beta-D-xylopyranosid, L-sorbose, L-rhamnose, dulcitol, inositol, inulin, adonitol, glycogen, xylitol, D-lyxose, D-fucose, L-fucose, L-arabitol, potassium gluconate, potassium-2-ketogluconate, potassium-5-ketogluconate; +, positive reaction; w, weak reaction; -, negative reaction. The evaluation was done after 48 h at 30 °C. Figure modified from Link et al., 2021.

				T. halophilus						
	T. osmophilus **DSM 23765^T**	**subsp. *flandriensis* DSM 23766^T**	**subsp. *halophilus* DSM 20339^T**	**DSM 20337**	**TMW 2.2260**	**TMW 2.2254**	**TMW 2.2256**	**TMW 2.2257**	**TMW 2.2263**	**TMW 2.2264**	**TMW 2.2266**
Glycerol	-	-	-	-	**+**	-	**+**	**+**	-	-	-
D-Arabinose	-	**+**	-	-	-	-	-	-	-	-	-
L-Arabinose	-	-	**+**	**+**	-	-	**+**	-	**+**	-	**+**
D-Ribose	-	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**
D-Xylose	-	-	-	-	-	-	-	-	-	-	**+**
D-Galactose	-	**w**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**
D-Mannitol	**+**	-	-	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**
D-Sorbitol	-	**+**	-	**+**	-	-	**+**	**+**	**+**	-	-
Methyl-alpha-D-Mannopyranoside	**+**	-	-	-	-	-	-	**+**	-	-	-
Methyl-alpha-D-Glucopyranoside	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	-	-	-
Amygdalin	-	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**
D-Lactose	-	-	-	-	**+**	-	**+**	-	**w**	-	-
D-Melibiose	-	**+**	-	**+**	-	**+**	-	-	-	**+**	-
D-Melecitose	-	-	**+**	-	-	-	-	**+**	-	-	-
D-Raffinose	-	**+**	-	-	-	-	-	-	-	-	-
D-Tagatose	-	-	**+**	**+**	**+**	**+**	**+**	**+**	**+**	-	**+**
D-Arabitol	**+**	-	**+**	-	**+**	**+**	**+**	**+**	**+**	**+**	-

RESULTS

All strains tested were able to ferment D-glucose, D-fructose, D-mannose, N-acetyl-glucosamine, arbutin, esculin/iron-citrate, salicin, D-cellobiose, D-maltose, sucrose, D-trehalose, D-gentobiose and D-turanose. No acids were produced by any strain tested from erythritol, L-xylose, D-adonitol, methyl-beta-D-xylopyranoside, L-sorbose, L-rhamnose, dulcitol, inositol, inulin, adonitol, glycogen, xylitol, D-lyxose, D-fucose, L-fucose, L-arabitol, potassium gluconate, potassium-2-ketogluconate or potassium-5-ketogluconate (Table 13).

Besides that, the type strain of *T. osmophilus* of fermented D-mannitol, methyl-alpha-D-mannopyranoside, methyl-alpha-D-glucopyranoside and D-arabitol. The type strain of *T. halophilus* subsp. *flandriensis* fermented D-arabinose, D-ribose, D-sorbitol, amygdalin, methyl-alpha-D-glucopyranoside, D-melibiose, D-raffinose and a weak reaction with galactose. The type strain of *T. halophilus* subsp. *halophilus* fermented L-arabinose, D-ribose, D-galactose, amygdalin, methyl-alpha-D-glucopyranoside, D-melecitose, D-tagatose and D-arabitol. The strain DSM 20337 from soy sauce mashes produced acid from L-arabinose, D-ribose, D-galactose, D-mannitol, D-sorbitol amygdalin, methyl-alpha-D-glucopyranoside, D-melibiose, D-tagatose. The strain TMW 2.2260 from buckwheat moromi fermented glycerol, D-ribose, D-galactose, D-mannitol, amygdalin, methyl-alpha-D-glucopyranoside, D-lactose, D-tagatose and D-arabitol.

Notably among all strains from lupine moromi, no fermentation pattern was duplicated and did not match any pattern from the type strains. Among the lupine moromi strains all were able to use D-mannitol, amygdalin, D-ribose and D-galactose. Furthermore, the strains from lupine moromi either ferment D-lactose or D-melibiose or neither but never both. Only the strains TMW 2.2256 and TMW 2.2257 were able to ferment glycerol to acid. L-arabinose was also only fermented by the strains TMW 2.2256, TMW 2.2263 and TMW 2.2266. D-sorbitol is only fermented by TMW 2.2256, TMW 2.2257 and TMW 2.2263. Only TMW 2.2254, TMW 2.2256 and TMW 2.2257 ferment Methyl-alpha-D-glucopyranoside. The strain TMW 2.2266 was the only strain to ferment D-xylose to acid. TMW 2.2266 was also the only strain from lupine moromi or buckwheat moromi to not ferment D-arabitol. TMW 2.2264 was the only from strain lupine moromi or buckwheat moromi that did not ferment D-tagatose. TMW 2.2257 was the only strain isolated from lupine moromi to ferment D-melecitose and methyl-alpha-D-mannopyranoside.

4.2.3 Growth at different sodium chloride concentrations

Sodium chloride is essential or improves growth for some species of *Tetragenococcus* and the optimal concentration can vary between species (Justé et al., 2014). To test for the variance among the newly isolated strains from lupine moromi, the strain from buckwheat moromi, DSM 20337 and the type strain DSM 20339^T the growth was analyzed after 48 h at 30 °C (Figure 5).

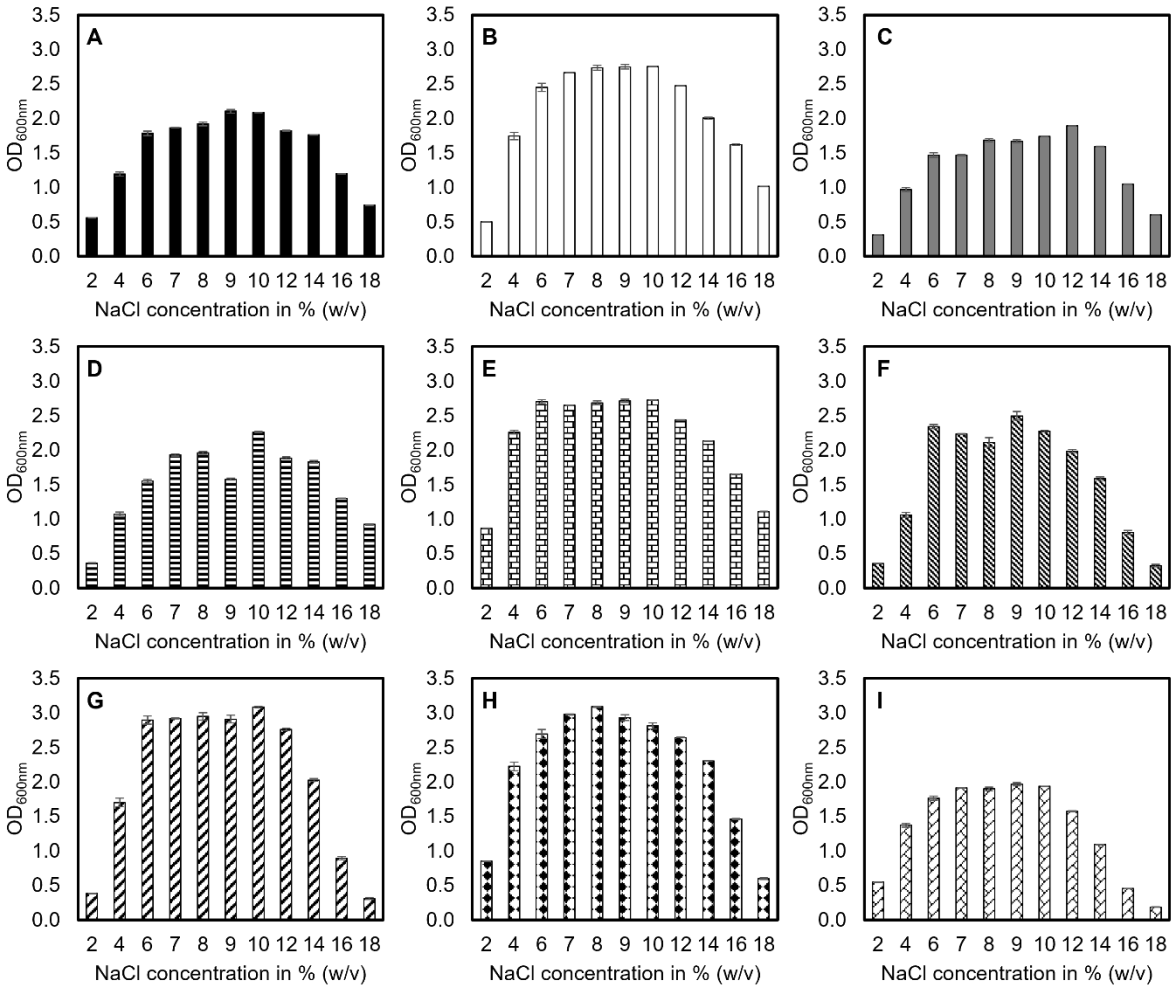

Figure 5: Salt tolerance of *T. halophilus* at 30 °C in MRS. The bars represent the mean values of the OD_{600nm} after 48 h from three independent biological replicates. The order of the strains from the most left to right: A = TMW 2.2254 ( ); B = TMW 2.2256 ( ); C = TMW 2.2257 ( ); D = TMW 2.2263 ( ); E = TMW 2.2264 ( ); F = TMW 2.2266 ( ); G = TMW 2.2260 ( ); H = DSM 20337 ( ); I = DSM 20339^T ( ).

All tested strains had their optimum between 8 % to 10 % NaCl (w/v). The strain DSM 20337 had the lowest NaCl optimum with 8 % NaCl but reached the highest OD_{600nm} with 3.09 at 8% NaCl (w/v). The optimal NaCl concentration for TMW 2.2254, TMW 2.2266 and type strain DSM 20339^T was at 9 % NaCl (w/v). The strains TMW 2.2256, TMW 2.2260, TMW 2.2263 and TMW 2.2264 had the optimum at 10%

NaCl (w/v). The optimal NaCl concentration for TMW 2.2257 was at 12 % NaCl (w/v). Notably, the strain set tested could be divided into two groups of strains that reach low cell densities and ones that reach higher cell densities. The set of strains with higher cell densities included TMW 2.2256, TMW 2.2260, TMW 2.2264 and DSM 20337. The set with low cell densities included TMW 2.2254, TMW 2.2257, TMW 2.2263, TMW 2.2266 and DSM 20339^T (Figure 5).

4.3 Comparative genomic analysis of the species *T. halophilus*

To explore the genomic diversity that is present among the strains from lupine, the genomes of the strains TMW 2.2254, TMW 2.2256, TMW 2.2257, TMW 2.2263, TMW 2.2264 and TMW 2.2266 were annotated using the NCBI PGAP, RAST, TIGR and eggnog-mapper. Then the genomes were compared to the reference strains and a set of strains from the NCBI database on a phylogenetic basis, for the possibility of encoded prophages, encoding of CRISPR/Cas sequences and encoding of carbohydrate clusters.

Parts of these results have been published in Link et al., 2021.

4.3.1 Pan/core and accessory genome of *T. halophilus*

The genome of a species or set of strains can be grouped into three categories, the pan-genome, the accessory genome and the core genome. The core genome includes all sequences that are present in all sequences. The accessory genome includes sequences that are absent in at least one strain but present in all other strains but also sequences that are only present in one strain and absent in all other strains. The pan-genome is the sum of the core genome and the accessory genome.
To visualize the pan and core genome of the species *T. halophilus* the program CMG biotools v2.2 (Vesth et al., 2013) was used (Figure 6). The pan – and core-genome plot was generated using the CDS found by the prodigalrunner program as implemented in CMG biotools, as this was a requirement for the correct function of the pan- and core-plot program from CMG biotools.

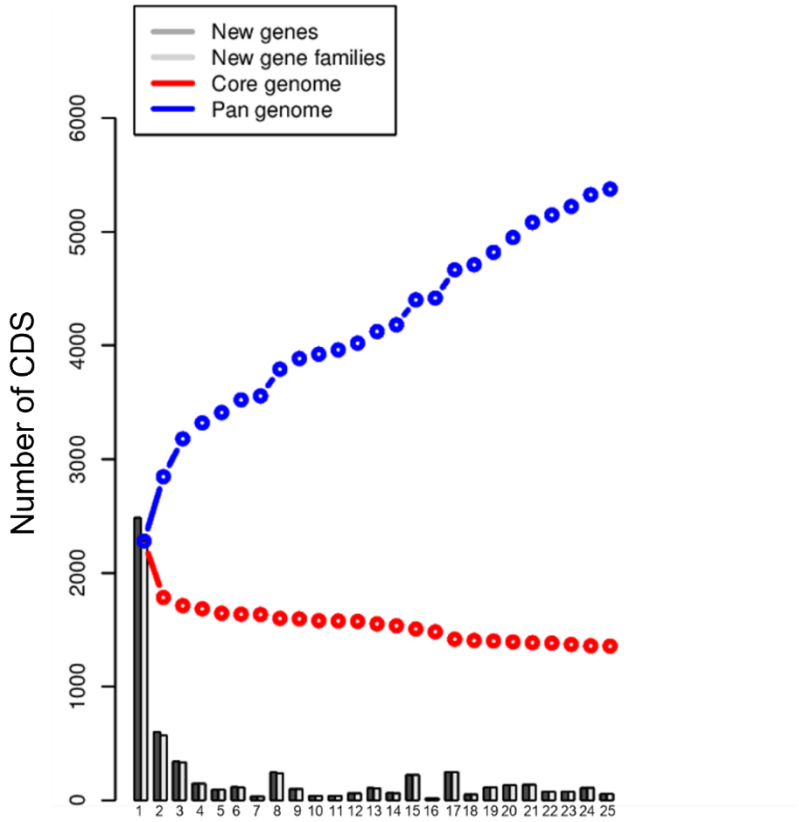

Figure 6: Plotted pan- (blue) and core-(red) genome of *T. halophilus*. 1: DSM 20339^T, 2: DSM 23766^T, 3: NBRC 12172, 4: NISL 7126, 5: 11, 6: D10, 7: D-86, 8: KG12, 9: YA163, 10: YA5, 11: YG2, 12: WJ7, 13: DSM 20337, 14: NISL 7118, 15: KUD23, 16: MJ4, 17: FBL3, 18: YJ1, 19: TMW 2.2254, 20: TMW 2.2256, 21: TMW 2.2257, 22: TMW 2.2263, 23: TMW 2.2264, 24: TMW 2.2266, 25: 8C7. Figure modified from Link et al., 2021.

The generated pan/core -plot shows that the species *T. halophilus* as analyzed with this set displays great diversity as the core genome only represents ≈ 25% of the pan genome. The pan genome with 5375 CDS and the core genome was predicted with 1356 CDS.

The annotation using eggNOG-mapper v 2.1.9 allowed the grouping of CDS from the core and accessory genome into clusters of orthologous genes (COG) categories. For this analysis, the same set of strains was used, but the CDSs as predicted by the NCBI PGAP were used as the basis for the core/accessory genome and were filtered using the BADGE program (Figure 7).

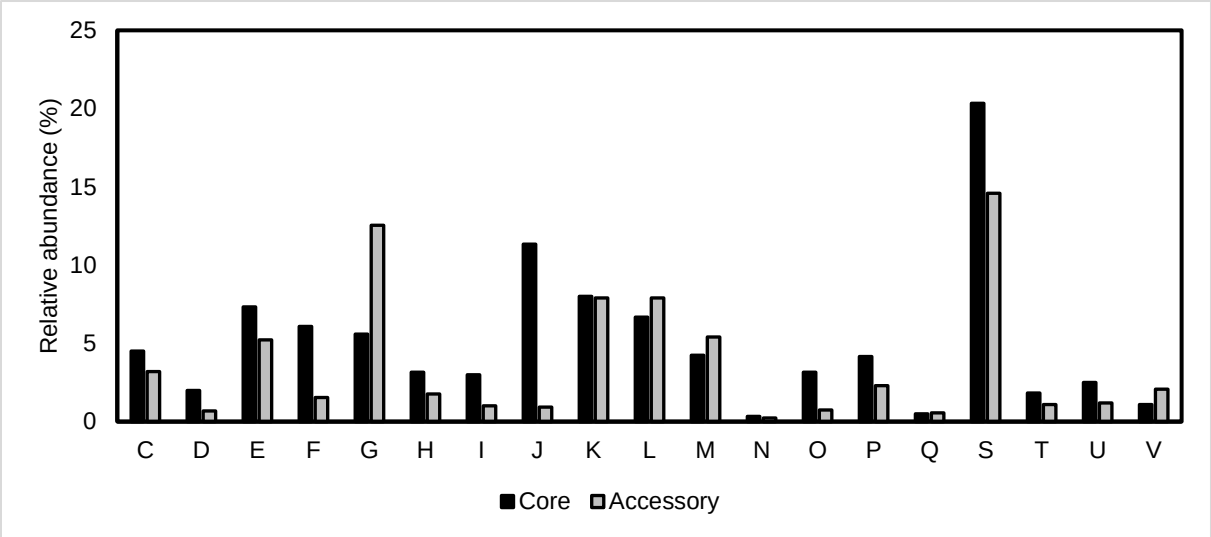

Figure 7: Relative abundance of the genes assigned to the different COG categories from the core or accessory genome from *T. halophilus* using eggnog-mapper v 2.1.9. The values are expressed as the relative abundances of a given category within the core or the accessory genome. C = energy production and conversion; D = cell cycle control, cell division, chromosome partitioning; E = amino acid transport and metabolism; F = nucleotide transport and metabolism; G = carbohydrate transport and metabolism; H = coenzyme transport and metabolism; I = lipid transport and metabolism; J = translation, ribosomal structure and biogenesis; K = transcription; L = replication, recombination and repair; M = cell wall/membrane/envelope biogenesis; N = cell motility; O = posttranslational modification, protein turnover, chaperones; P = inorganic ion transport and metabolism; Q = secondary metabolites biosynthesis, transport and catabolism; S = function unknown; T = signal transduction mechanisms; U = intracellular trafficking, secretion and vesicular transport; V = defense mechanisms. 1626 CDSs could not be assigned to a category and are not shown in this graph.

The core genome consisted of 1200 CDS and the accessory genome consisted of 5418 CDS. Therefore, the pan-genome was predicted with 6618 CDS. The eggnog-mapper pipeline assigned the CDS to a total of 19 different categories (Figure 7). The values are displayed as the relative abundance of a category within the core or the accessory genome.

A total of 4.17 % of the CDS from the core genome and 29.09 % of the accessory were not assigned to a category. 4.5 % of the core genome and 3.19 % of the accessory genome were assigned to category C, energy production and conversion. To category D, cell cycle control, cell division, chromosome partitioning, 2 % of the core genome and 0.68 % of the accessory genome were assigned. Category E, amino acid transport and metabolism, was assigned to 7.33 % of the core genome and 5.22 % of the accessory genome. Category F, nucleotide transport and metabolism, was assigned to 6.08 % of the core genome and 1.55 % of the accessory genome. Category G, carbohydrate transport and metabolism was assigned to 5.58 % of the core genome and 12.55 % of the accessory genome. Category H, coenzyme transport and metabolism, was assigned to 3.17 % of the core genome and 1.77 % of the accessory genome. Category I, lipid transport and metabolism was assigned to 3 % of the core

genome and 1.02 % of the accessory genome. Category J, translation, ribosomal structure and biogenesis, was assigned to 11.33 % of the core genome and 0.92 % of the accessory genome. Category K, transcription, was assigned 8 % of the core genome and 7.9 % of the accessory genome. Category L, replication, recombination and repair, was assigned to 6.67 % of the core genome and 7.90 % of the accessory genome. Category M, cell wall/membrane/envelope biogenesis, was assigned to 4.25 % of the core genome and 5.41 % of the accessory genome. Category N, cell motility, was assigned to 0.33 % of the core genome and 0.24 % of the accessory genome. Category O, posttranslational modification, protein turnover, chaperones, was assigned to 3.17 % of the CDS from the core genome and 0.74 % of the accessory genome. Category P, inorganic ion transport and metabolism, was assigned to 4.17 % of the core genome and 2.31 % of the accessory genome. Category Q, secondary metabolites biosynthesis, transport and catabolism, was assigned to 0.5 % of the core genome and 0.55 % of the accessory genome. The biggest category S, function unknown, was assigned to 20.33 % of the core genome and 14.60 % of the accessory genome. Category T, signal transduction mechanisms, was assigned to 1.83 % of the core genome and 1.09 % of the accessory genome. Category U, intracellular trafficking, secretion and vesicular transport, was assigned to 2.5 % of the core genome and 1.2 % of the accessory genome. Category V, defense mechanisms, was assigned to 1.08 % of the core genome and 2.07 % of the accessory genome.

4.3.2 The species *T. halophilus* can be separated into distinctive three lineages

Phylogenetic analyses of the 16S rDNA of the strains from lupine moromi assigned all to the species *T. halophilus* (Figure 24). To determine the subspecies affiliation of the strains from lupine moromi, the *in-silico* DDH values were calculated. The genomic relationship of the strains from lupine moromi to other strains in the set was analyzed phylogenetically using concatenated housekeeping genes. To generate the concatenated housekeeping genes sequences, the nucleotide sequences of the housekeeping genes *fusA*, *gyrA*, *gyrB*, *lepA*, *pyrG*, *recA* and *rpoD* were concatenated in this order. The genomic relationship was further characterized using ANIb values.

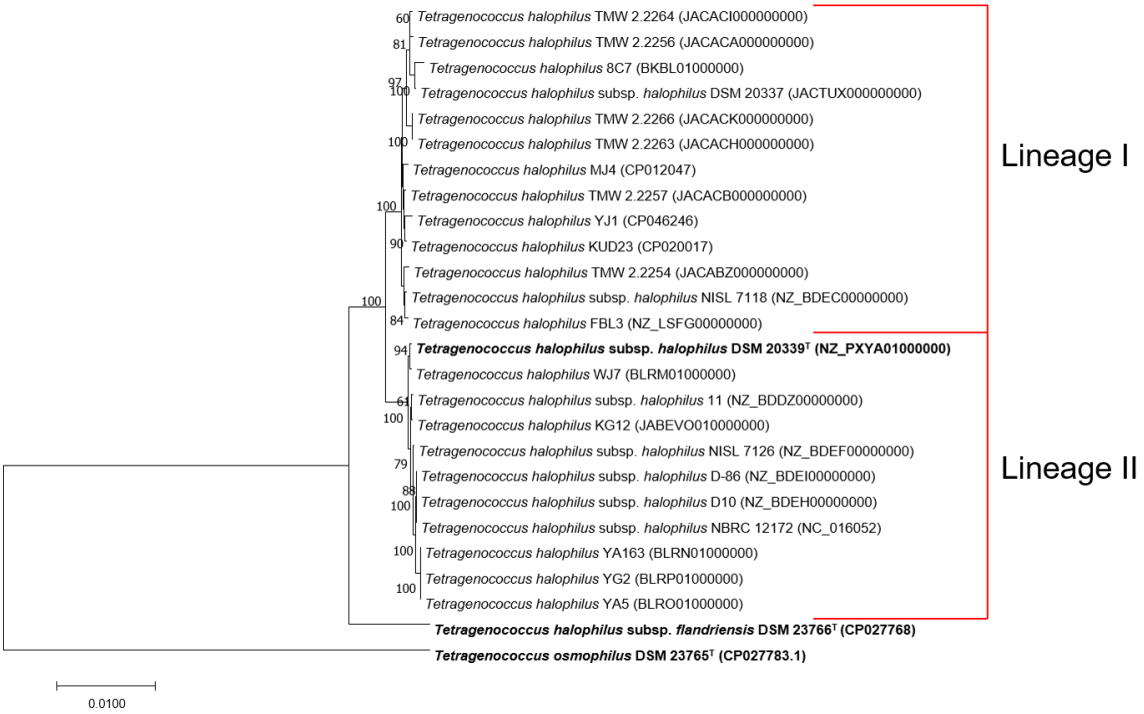

Figure 8: Phylogenetic tree of *T. halophilus* based on concatenated nucleotide sequences of the housekeeping genes (*fusA*, *gyrA*, *gyrB*, *lepA, pyrG*, *recA*, *rpoD*) using the Neighbor-Joining method (Saitou and Nei, 1987). The sum of the branch length is 0.10333351. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. *T. osmophilus* DSM 23765^T was used as an outgroup. Figure modified from Link et al., 2021.

A phylogenetic analysis using concatenated housekeeping genes (Figure 8) shows that the species *T. halophilus* is separated into three lineages. The lineage I includes all strains isolated from lupine moromi but also isolated strains from fermented fish, soy sauces or cheese brine. The lineage II includes the type strain DSM 20339^T, several other strains from soy sauce fermentation and the strain WJ7 from fermented fish. The lineage III only consists of the strain DSM 23766^T. The calculated ANIb values were calculated for the entire set (Table 22). Sorting the ANIb values into the new lineages allowed the analysis of the diversity within the species as well as the comparison to the other lineages. The median intra-lineage ANIb value of lineage I was 97.4. The median intra-lineage ANIb value for lineage II was 98.4. The median of the inter-lineage ANIb values between lineage I and lineage II was 96.8. The median of the inter-lineage ANIb values between lineage I and III is 96.3. The median of the inter-lineage ANIb values between lineage II and III is 95.8. As lineage III only consisted of one strain the intra lineage median ANIb value is 100 (Figure 9). The calculated *in-*

silico DDH values from the lupine moromi strains to the type strains DSM 20339^T and DSM 23766^T were <79 %, which is above the species delineation of 70% (Table 14).

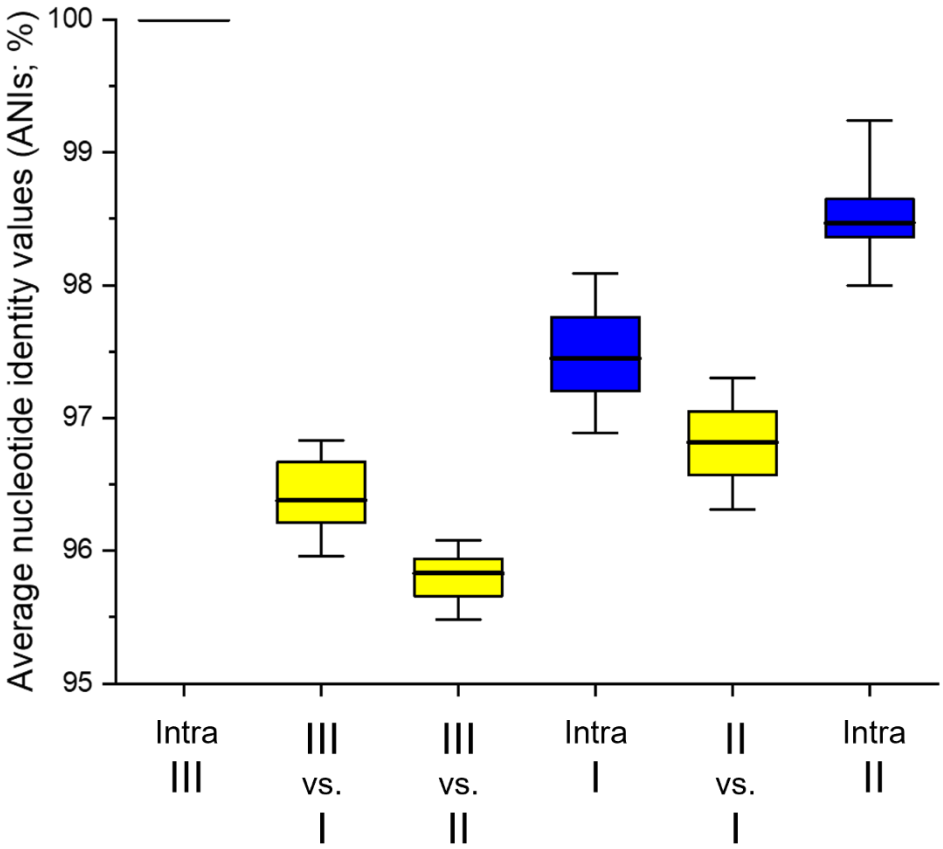

Figure 9: ANIb values between the *T. halophilus* lineages. ANIb values were calculated using JSpecies WS and displayed in percent. Intra III = Intra lineage values of the lineage III; III vs. I = Inter lineage values of the strains from lineage III to lineage I; III vs. II = Inter lineage values of the strains from lineage III to lineage II; Intra I = Intra lineage values of lineage I; II vs. I = Inter lineage values of lineage II to lineage I; Intra II = Intra lineage values of lineage II. Figure modified from Link et al., 2021.

Table 14: DDH values for the *T. halophilus* strains from lupine moromi to the type strains. The DDH values from the second formula are displayed. These values were calculated using the GGDC 2.1 from the DSMZ website (https://ggdc.dsmz.de/). The table is ordered by descending DDH values to DSM 20339^T.

	T. halophilus* subsp. *halophilus	***T. halophilus* subsp. *flandriensis***
	DSM 20339^T	**DSM 23766^T**
Strains	**DDH in %**	**DDH in %**
TMW 2.2256	77.1 (74.1-79.8)	76.4 (73.4-79.1)
TMW 2.2264	77.0 (74-79.7)	73.7 (70.7-76.5)
TMW 2.2263	76.7 (73.7-79.4)	74.0 (71-76.8)
TMW 2.2254	76.3 (73.3-79)	75.5 (72.5-78.3)
TMW 2.2266	73.1 (70.1-75.9)	74.5 (71.5-77.3)

4.3.3 Clustering upon functional carbohydrate gene clusters

The second most frequent CDSs in the accessory genome were assigned to the COG category G (the category for carbohydrate metabolism and transport) Therefore, the carbohydrate metabolism was analyzed in more detail. The BADGE program was used to compare the nucleotide sequences of all strains. Only ORFs of which the function was clear and had a sound annotation between the NCBI PGAP, RAST and TIGR pipeline were used. The resulting differences were then filled into a binary matrix and hierarchical clustered using Perseus (Figure 10).

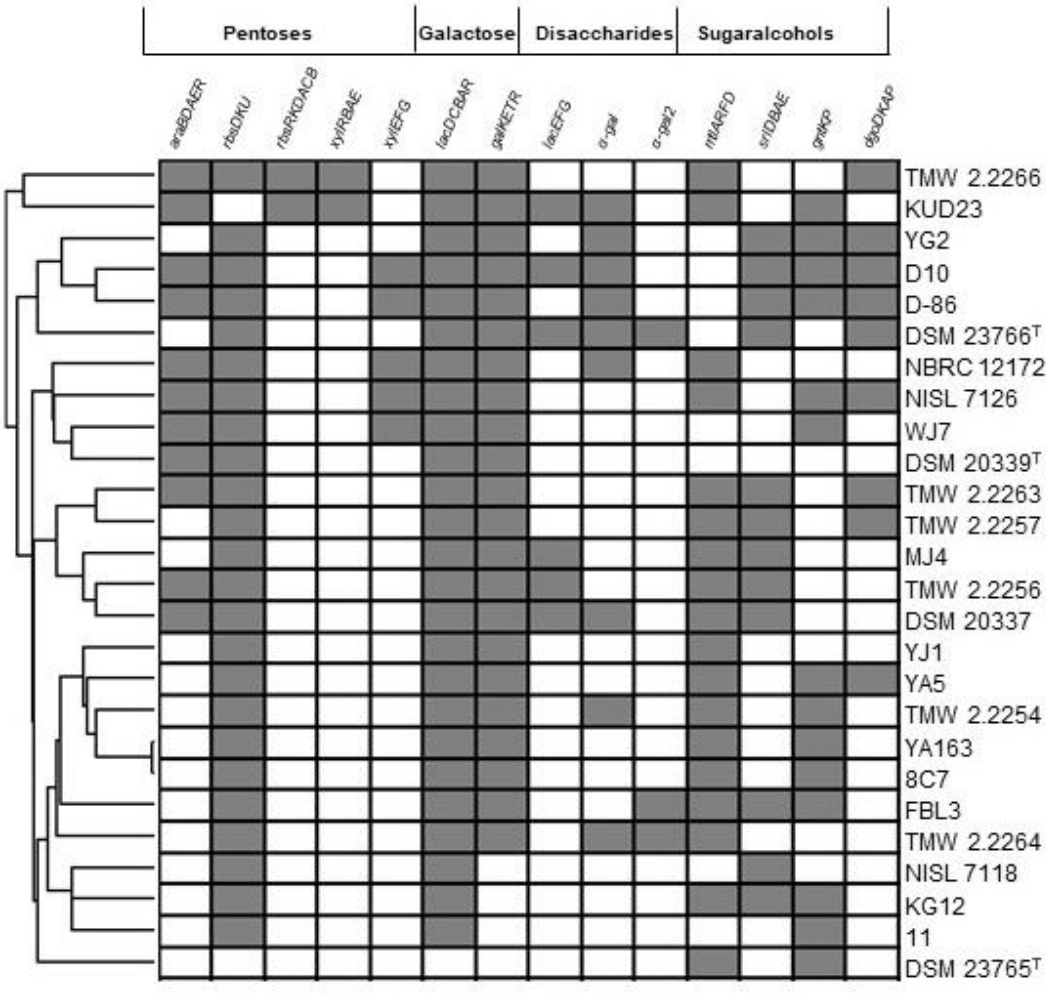

Figure 10: Clustering of the functional carbohydrate cluster in *T. halophilus*. The clustering was done using the hierarchical clustering function with Euclidean distance clustering in Perseus v1.6.14.0. Grey box= functional and complete cluster. White box= incomplete/unfunctional or missing cluster. The strain *T. osmophilus* DSM 23765^T was used as an outgroup. *araBDAER* = L-arabinose metabolizing cluster; *rbsDKU* = D-ribose metabolizing cluster with PTS-transporter; *rbsRKDACB* = D-ribose metabolizing cluster with ABC-transporter; *xylRBAE* = D-xylose metabolizing cluster; *xylEFG* = putative pentoses ABC-transporter; *lacDCBAR* = cluster encoding for the tagatose-6P pathway; *galKETR* = leloir pathway; *lacEFG* = D-lactose metabolizing cluster; *α-gal* = alpha galactosidase; *α-gal2* = non-identical second alpha galactosidase; *mtlARFD* = D-mannitol metabolizing cluster; *srlDBAE* = D-sorbitol metabolizing cluster;

***gntKP* = gluconate metabolizing cluster; *dgoDKAP* = putative D-galactonate metabolizing cluster. Figure modified Link et al., 2021.**

The clustering of the functional carbohydrate cluster reveals that the strains differ in the presence or absence of carbohydrate metabolizing operons across all origins and lineages. The plot also shows that the metabolic pathways for D-ribose and D-galactose are conserved among the members of the species *T. halophilus*. The clustering reveals that the *lacEFG* operon and the alpha-galactosidase are traits that are only present in a few strains. The *araBDAER* operon for the metabolism of l-arabinose is more commonly present in strains originating from soy-based fermentations. The operons *rbsRKDACB* and *xylRBAE* are only present in the strains KUD23 and TMW 2.2266. The putative pentose transporter *xylEFG* is only present in a group of strains from soy sauce fermentations. The distribution of operons for the metabolism of sugar alcohols such as mannitol and sorbitol does not follow a pattern across all isolation sources.

4.3.4 Distribution of key amino acid pathways and enzymes in *T. halophilus*

Most *T. halophilu*s strains in the set were isolated from protein-rich fermentations. The distribution of enzymes and pathways associated with the degradation of amino acids and the production of biogenic amines was analyzed with BADGE (Table 15).

Table 15: Distribution of key amino acid pathways and enzymes in *T. halophilus*. The arginine deiminase pathway (ADI), the agmatine deiminase pathway (AgDI), the aspartate decarboxylase with adjacent aspartate transporter (*aspDT*) and the pyruvoyl-dependent histidine decarboxylase (BAG14318.1) (*hdcA*) in *T. halophilus*. Strains are ordered by the isolation sources. + = complete and predicted as functional cluster; - = incomplete or predicted as not functional cluster.

Isolation source	Strain	ADI	AgDI	*AspDT*	*hdcA*
Lupine moromi	TMW 2.2263	-	-	-	-
Lupine moromi	TMW 2.2254	-	-	+	-
Lupine moromi	TMW 2.2266	-	+	-	-
Lupine moromi	TMW 2.2256	+	-	-	-
Lupine moromi	TMW 2.2257	+	-	-	-
Lupine moromi	TMW 2.2264	+	-	+	-
Degraded sugar beet juice	DSM 23766^T	-	-	-	-
Brie de meaux cheese rind	8C7	-	-	-	-
Fish nukazuke	WJ7	-	+	-	-
Anchovy fish sauce	MJ4	+	-	-	-
Fish sauce	YJ1	+	-	-	-
Fish sauce	FBL3	+	-	-	-
Salted anchovy	DSM 20339^T	+	-	-	-
Soy sauce mash	NISL 7118	-	-	-	-
Soy sauce mash	11	-	-	-	+
Soy sauce mash	YA5	-	-	+	-
Soy sauce mash	D10	-	+	+	-

Soy sauce mash	D-86	-	+	+	-
Soy sauce mash	DSM 20337	+	-	-	-
Soy sauce mash	YG2	+	-	+	-
Soy sauce mash	YA163	+	-	+	-
Soy sauce mash	NISL 7126	+	-	+	-
Soy sauce mash	NBRC 12172	+	-	+	-
Korean Soy sauce	KG12	+	-	+	-
Korean soypaste	KUD23	+	-	+	-

The comparative analysis showed that the ADI pathway is predicted in 14 strains as active and complete. The two proteins *AspD* and *AspT* are predicted in 11 strains as active and complete. The pathway for the conversion of agmatine to putrescine *AgDI* is only present and predicted as active in four strains. Only in one strain a histidine decarboxylase was found. None of the pathways or enzymes are present in the four strains TMW 2.2263, DSM 23766^T, NISL 7118 and 8C7.

4.3.5 CRISPR/CAS systems, prophages and restriction-modification systems in *T. halophilus* strains

Bacteriophages represent a big problem in some fermentations and have been reportedly isolated from soy sauce fermentations (Garneau and Moineau, 2011; Uchida and Kanbe, 1993). The species *T. halophilus* was screened for the presence of prophage sequences with subsequent prediction of intact prophages using PHASTER.

Defense mechanisms involve CRISPR/Cas systems and restriction-modification systems (Allison and Klaenhammer, 1998; Horvath and Barrangou, 2010). The presence or absence and the diversity of these regions were analyzed in the set of strains using CRISPRdb and BADGE (Table 16).

Table 16: CRISPR/CAS systems and predicted intact prophages detected in *T. halophilus*. The CRISPR/CAS sequences were identified using the CRISPRdb (https://crisprcas.i2bc.paris-saclay.fr) (Couvin *et al*., 2018). The evidence level indicates how many marker proteins of a specific CAS type were found. Strains are ordered by isolation source, CAS type detected and evidence level. The intact prophages were predicted using PHASTER (https://phaster.ca)(Arndt *et al*., 2016). The presence of all components of an RM system is indicated by an "+"; Versions with duplicated *HsdS* and transposase insertion are indicated by an "+*"; Absence of rm system are indicated by "-"; Figure modified from Link et al., 2021.

Isolation source	Strain	CAS Type	Evidence level	Predicted intact prophages	TYPE I RM	TYPE II RM
Lupine moromi	TMW 2.2257	IIC + IIIAD	4	3	-	-
Lupine moromi	TMW 2.2264	IC	4	1	+*	-
Lupine moromi	TMW 2.2254	IE	4	1	-	-
Lupine moromi	TMW 2.2256	I	1	1	-	-
Lupine moromi	TMW 2.2266	I	1	2	+*	-
Lupine moromi	TMW 2.2263	I	0	2	+*	-
Fish sauce	YJ1	IE + IIIAD	4	0	+*	-
Anchovy fish sauce	MJ4	IC + IIIA	4	0	+*	-
Fish nukazuke	WJ7	IIIC	4	0	+*	-
Fish sauce	FBL3	I + IIIC	0	1	-	-
Salted anchovy	DSM 20339^T	X	0	0	-	-
Degraded sugar beet juice	DSM 23766^T	X	0	2	-	+
Brie de meaux cheese rind	8C7	X	0	0	+*	-
Soy sauce mash	11	IC + IIIC	4	0	+*	-
Soy sauce mash	YG2	IE + IC	4	0	+*	-
Soy sauce mash	YA5	IE + IC	4	0	-	-
Soy sauce mash	YA163	IC	4	1	-	-
Korean Soy sauce	KG12	IE	4	2	-	-
Soy sauce mash	NISL 7118	I + IIIC	0	0	+*	-
Soy sauce mash	DSM 20337	I + IIIC	0	0	-	+
Soy sauce mash	D10	I + IIIC	0	0	+*	-
Soy sauce mash	D-86	I + IIIC	0	1	+*	-
Soy sauce mash	NISL 7126	I + IIIC	0	2	+*	+
Soy sauce mash	NBRC 12172	X	0	1	-	-
Korean soypaste	KUD23	IIIA	4	0	-	-

The analysis with PHASTER reveals that all strains from lupine moromi have at least one prophage sequence predicted as intact in their genome (Table 16). The strains TMW 2.2263 and TMW 2.2266 have two and TMW 2.2257 encodes for three sequences. The strains YA163, D-86 and NBRC 12172 are strains from soy sauce mash with one predicted intact prophage sequence. The strains NISL 7126 and KG12

are also from soy sauce mash and encode for two prophage sequences predicted as intact. FBL3 is the only strain from fish sauce to encode for a prophage sequence predicted as intact. The type strain DSM 23766^T does encode for two prophage sequences that are predicted as intact.

The analysis of the same strains for the presence of CRISPR regions and CAS proteins reveals that in 12 strains nine different CAS type combinations were identified with an evidence level of four. A manual blast search revealed only strains with an evidence level of four had CAS proteins. Resulting from that, three different CAS type combinations were present in the strains from lupine moromi, with TMW 2.2254 encoding for CAS type IE, TMW 2.2257 encoding for the CAS types IIC and III AD and TMW 2.2264 encoding for CAS type IC. The strains from fish sauces encoded for a total of three different CAS type combinations with the types IE and III AD in the strain YJ1, the types IC and IIIA in the strain MJ4 and the IIIC type in the strain WJ7. The strains from soy sauce encode for a total of five different combinations among six strains. The strain 11 encodes for IC and IIIC types, YG2 and YA5 for the types IE and IC, YA163 for IC type, KG12 for the IE type and the strain KUD23 for the IIIA type.

Marker genes of the restriction-modification (RM) systems type I and type II were found in some strains of *T. halophilus*. No sequences corresponding to RM systems were found in ten strains. Twelve strains encoded for the *hsdMSR* type I system. Two strains encoded for a type II system, consisting of a AlwI type endonuclease and a DAM methylase. One strain encoded for *hsdMSR* and the same type II system. Notably all *hsdMSR* system have an integrase inserted between two annotated copies of *hsdS*. Therefore, it is unclear if this system is active or not.

4.4 Mucoid phenotype of *T. halophilus* TMW 2.2256

During the experiments, it was noted that the strain TMW 2.2256 forms mucoid colonies when cultivated on agar supplied with D-galactose or D-lactose (Figure 11 A, B). A ropy phenotype was observable when the colonies were pulled with an inoculation loop (Figure 11, C). The formation was compared to two other strains and examined visually.

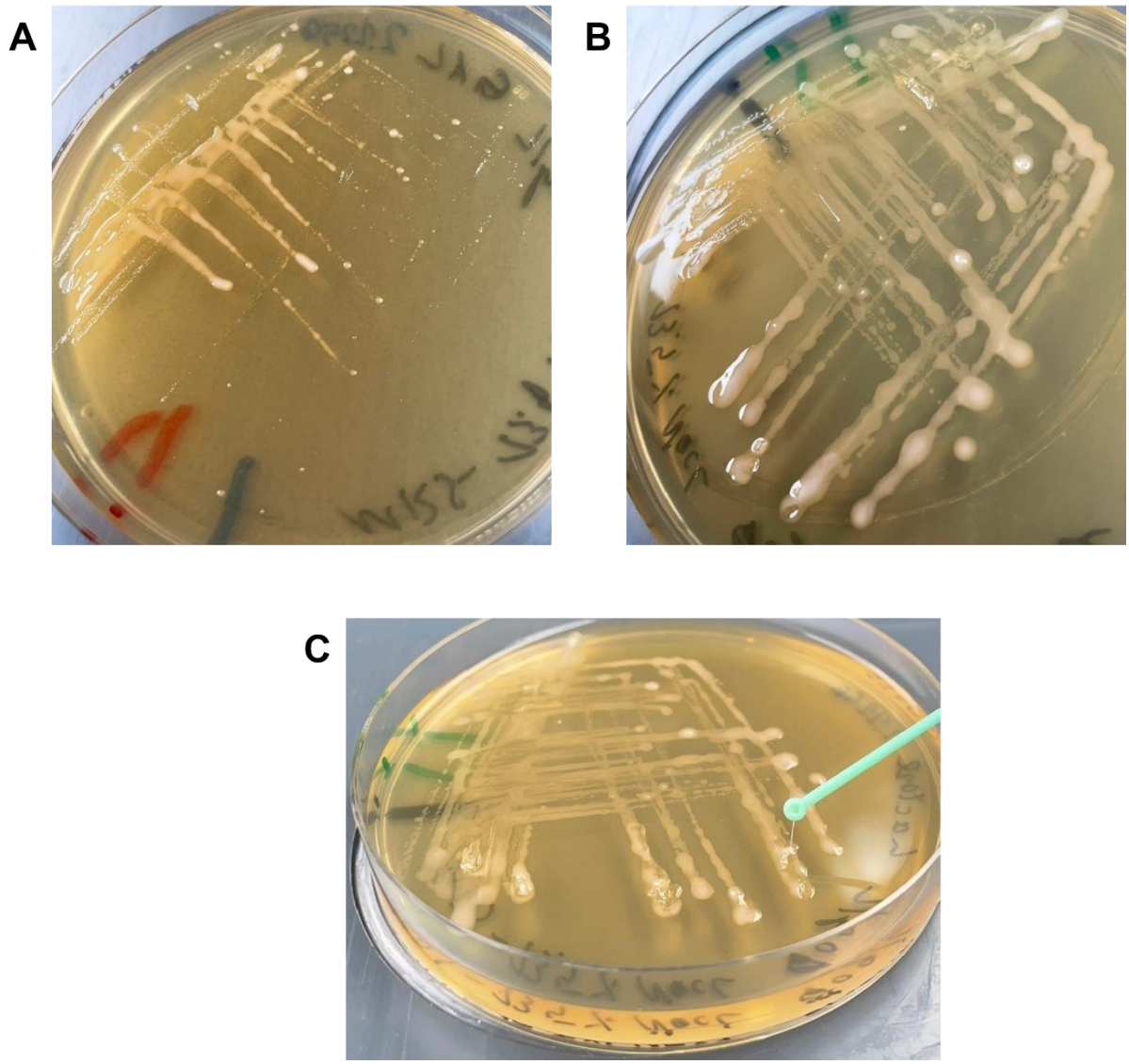

Figure 11: Mucoid phenotype screening of TMW 2.2256. MRS agar supplied with 13.5% (w/v) NaCl and D-galactose or D-lactose were used for the screening. A) Colony morphology of TMW 2.2256 on MRS agar with 60 g/L D-galactose. B) Colony morphology of TMW 2.2256 on MRS agar with 10 g/L D-lactose. C) Ropy phenotype of TMW 2.2256 when grown on MRS agar with D-lactose with 10 g/L.

The formation of mucoid colonies of TMW 2.2256 was compared with other representative strains from the species, TMW 2.2254 and DSM 20339^T (Table 17). The appearance of mucoid colonies was the strongest in TMW 2.2256 and the lowest in

DSM 20339^T. The formation was increased when the sodium chloride concentration was increased to 13.5 % NaCl (w/v).

Table 17: Qualitative screening of mucoid colony formation on different agars. The plates were incubated for 12 days at 30 °C before examination.

Carbon source	NaCl (w/v)	TMW 2.2256	TMW 2.2254	DSM 20339^T
D-glucose 60 g/L	5 %	-	-	-
D-glucose 60 g/L	13.5 %	++	-/+	-
D-galactose 60 g/L	5 %	-	-	-
D-galactose 60 g/L	13.5 %	+	-	-
sucrose 60 g/L	5 %	-	-	-
sucrose 60 g/L	13.5 %	-/+	-	-
D-lactose 10 g/L	5 %	-	-	-
D-lactose 10 g/L	13.5 %	+++	++	-/+

4.5 Transcriptomic profiling of *T. halophilus* TMW 2.2254 and TMW 2.2256 in LMRS reveals strain-dependent differences.

To determine traits that are important for the growth in lupine moromi a lupine-adapted model medium, LMRS was designed. The protein contents were exchanged for a lupine peptone which still had some residual carbohydrates in the powder as declared by the manufacturer. To identify these residual carbohydrates unused lupine peptone was qualitatively measured on a HPAEC-PAD system. Next, the growth of two representative strains isolated from lupine moromi was monitored in LMRS with two added carbon sources, D-glucose or D-galactose. Then, the metabolism was analyzed in more detail using transcriptomics.

Parts of the results were published (Link and Ehrmann, 2023a).

4.5.1 Growth of *T. halophilus* in LMRS

To modify the MRS medium to a more lupine-like medium, the protein compounds were exchanged for a peptone based on the proteins from lupine beans. The monosaccharides in the unused lupine peptone were detected *via* HPAEC-PAD (Figure 25). D-glucose and D-galactose were identifiable. Two adjacent peaks were not clearly identifiable. Relative quantification to each other showed that peak two was the most abundant, then D-glucose, then peak one and then D-galactose.
To verify that the growth of *T. halophilus* was possible in this new medium, the growth was monitored over 40 h at 30 °C in LMRS supplemented with 13.5 % NaCl (w/v) and no added carbon source, addition of 10 mM D-glucose or 10 mM galactose (Figure 12, A). The pH was also monitored for 40 h (Figure 12, B).

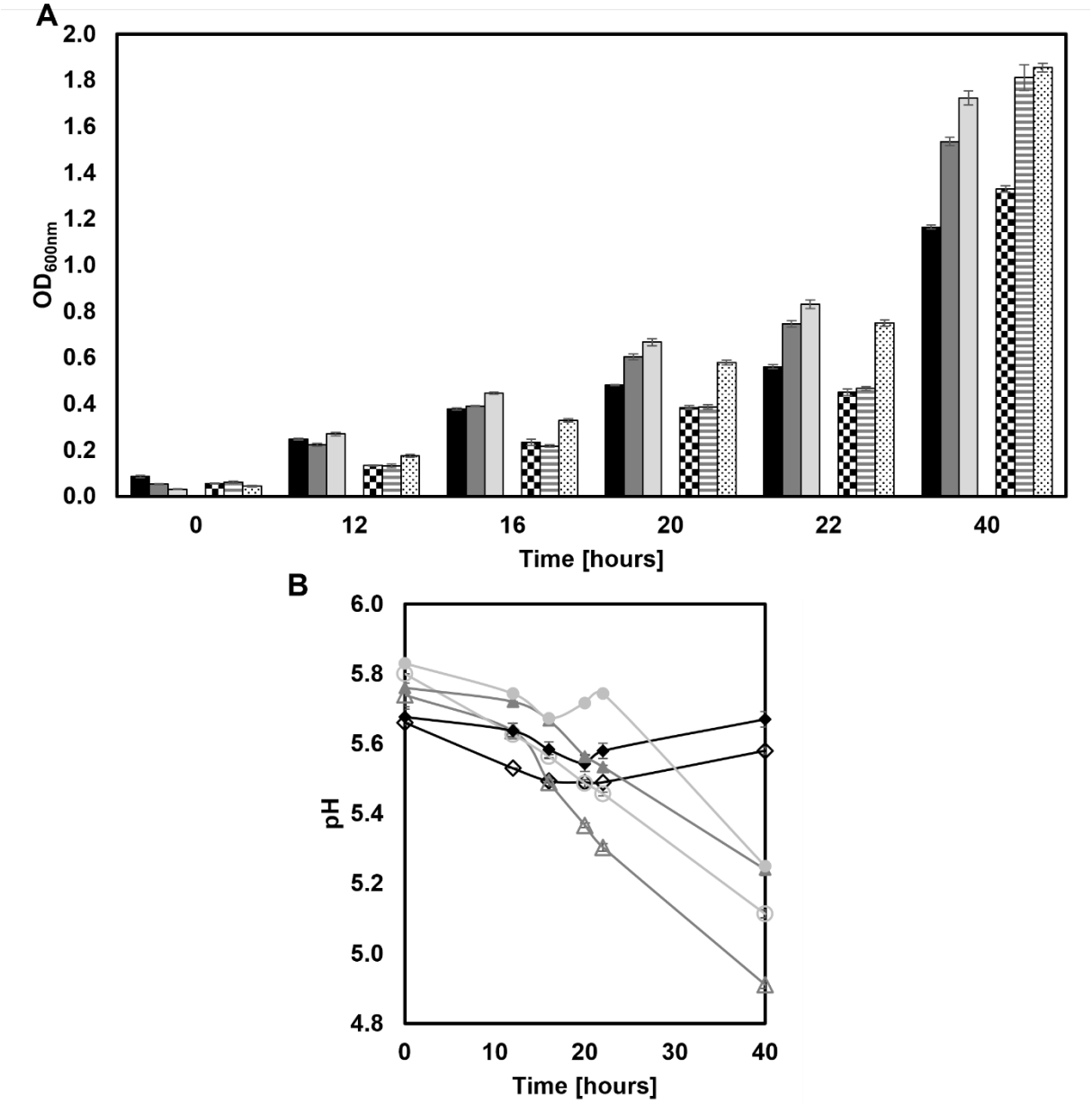

Figure 12: Growth behavior of *T. halophilus* strains TMW 2.2254 and TMW 2.2256 in LMRS. The medium was supplemented with either D-glucose, D-galactose or no added carbon sources. A= OD_{600nm} of TMW 2.2254 (black= w/o added carbon sources; dark gray= 10 mM D-glucose; light gray= 10 mM D-galactose); OD_{600nm} of TMW 2.2256 (checked= w/o added carbon sources; grey lines= 10 mM D-glucose; black dots= 10 mM D-galactose). B= pH values of the culture broth of TMW 2.2254 (open symbols) and TMW 2.2256 (filled symbols) (black= w/o added carbon sources; gray= 10 mM D-glucose; light gray= 10 mM D-galactose). The standard deviation is based on 3 biological replicates. Figure modified from Link and Ehrmann, 2023a.

Both strains grew under all conditions (Figure 12, A). After 40 h, TMW 2.2254 reached an OD_{600nm} of 1.16 without the addition of any extra carbon source, an OD_{600nm} of 1.72 with the addition of D-galactose and an OD_{600nm} of 1.53 with the addition of D-glucose. TMW 2.2256 reached an OD_{600nm} of 1.33 without the addition of any extra carbon source, an OD_{600nm} of 1.85 with the addition of D-galactose and an OD_{600nm} of 1.81 with the addition of D-glucose. The pH values decreased under all conditions but rose

for both strains almost back to the starting pH in the cultures grown without the addition of any carbon sources. The decrease was the biggest for both strains when D-glucose was added, with TMW 2.2254 reaching a pH of 4.91 and TMW 2.2256 reaching a pH of 5.24 after 40 h. The final pH when grown with D-galactose was 5.1 for TMW 2.2254 and 5.25 for TMW 2.2256.

4.5.2 Transcriptomic profile of TMW 2.2254 and TMW 2.2256 reveal differences in the metabolism when grown in LMRS

The strains TMW 2.2254 and TMW 2.2256 were grown in LMRS medium for 10 h at 25 °C and sampled after 1 h and 10 h. After RNA sequencing the reads were mapped against the respective genomes. The mapping yielded 22.06 million to 43.54 million unique reads for TMW 2.2254 and 24.65 million to 48.13 million unique reads for TMW 2.2256. The reads covered 98.3 % to 99.4 % of the genome from TMW 2.2254 and 98.7 % to 99.7 % of the genome of TMW 2.2256. Then, features that were different in the reactions, DEGs, were filtered by applying a cutoff of Log2 ≥ 2 or ≤ - 2, a p-value < 0.05 and an FDR of <0.01 (Table 18).

Table 18: Selected DEGs from the transcriptomic profile of TMW 2.2254 and TMW 2.2256 in LMRS. Substrates and putative substrates of respective genes are listed in column one. NCBI PGAP annotation of the CDSs is shown in column two. Columns three and four contain the respective locus tag in the strain, missing of the CDSs is represented by an "X". Column five contains an EC number if available. Columns six and seven show the Log2 fold change of the respective CDSs, missing CDSs are represented by "X". CDSs that are present in the strain but do not count as DEGs are indicated as small italics numbers in white cells. FS indicates a frameshift that results in a premature stop codon. Cutoffs: Log2 foldchange of ≥ 2 or ≤ - 2, p-value ≤ 0.05 and FDR ≤ 0.01. Figure modified from Link and Ehrmann, 2023a.

Substrates	NCBI annotation	Locus tag TMW2.2254	Locus tag TMW2.2256	EC	TMW2.2254	TMW2.2256
Tagatose (putatively)	tagatose-bisphosphate aldolase	HV360_03980	HXW74_10490	4.1.2.40	5.95	5.88
	PTS sugar transporter subunit IIA	HV360_03985	HXW74_10495		5.55	6.13
	PTS transporter subunit EIIC	HV360_03990	HXW74_10500		4.55	5.80
	1-phosphofructokinase	HV360_03995	HXW74_10505	2.7.1.101	4.98	5.69
Galactose-6P	PTS sugar transporter subunit IIA	HV360_07360	HXW74_08040		2.04	*0.56*
	galactose-6-phosphate isomerase subunit LacA	HV360_07365	HXW74_08045	5.3.1.26	2.79	3.67
	galactose-6-phosphate isomerase subunit LacB	HV360_07370	HXW74_08050	5.3.1.26	3.62	3.97
	tagatose-6-phosphate kinase lacC	HV360_07375	HXW74_08055	2.7.1.144	4.02	5.29
	tagatose-bisphosphate aldolase LacD	HV360_07380	HXW74_08060	4.1.2.40	4.28	6.29
Galactose-6P	tagatose-bisphosphate aldolase LacD2	HV360_11235	HXW74_11230	4.1.2.40	3.22	5.59
	tagatose-6-phosphate kinase LacC2	HV360_11240	HXW74_11235	2.7.1.144	3.47	5.30
	HAD family phosphatase pgbmB	HV360_11245	HXW74_11240		3.56	4.25
	galactose-6-phosphate isomerase subunit LacB2	HV360_11250	HXW74_11245	5.3.1.26	2.98	3.57
	galactose-6-phosphate isomerase subunit LacA2	HV360_11255	HXW74_11250	5.3.1.26	2.67	3.31
	PTS sugar transporter subunit IIA	HV360_11260	HXW74_11255		2.57	2.39
	PTS fructose transporter subunit IIB	HV360_11265	HXW74_11260		2.70	2.79
	PTS glucitol transporter subunit IIA	HV360_11270	HXW74_11265		2.80	3.40
Galactose	LacI family DNA-binding transcriptional regulator	HV360_04875	HXW74_07385		3.42	3.14
	UDP-glucose--hexose-1-phosphate uridylyltransferase galT	HV360_04880	HXW74_07390	2.7.7.12	3.73	2.97
	UDP-glucose 4-epimerase GalE	HV360_04885	HXW74_07395	5.1.3.2	3.49	2.67
	galactokinase galK	HV360_04890	HXW74_07400	2.7.1.6	3.46	2.35
Melibiose/Raffinose	sugar O-acetyltransferase	HV360_04870	X		2.47	X
	alpha-galactosidase	HV360_04860	X	3.2.1.22	2.12	X
Lactose	PTS lactose/cellobiose transporter subunit	X	HXW74_08065	2.7.1.207	X	8.09
	6-phospho-beta-galactosidase	X	HXW74_08070	3.2.1.85	X	6.81
	PTS transporter subunit EIIC	X	HXW74_08075	2.7.1.207	X	6.30
Cellobiose	6-phospho-beta-glucosidase (2)	HV360_09000	HXW74_10725	3.2.1.86	*-0.11*	4.74
	PTS cellobiose transporter subunit IIB	HV360_08995	HXW74_10730	2.7.1.205	NV	6.25
	PTS cellobiose transporter subunit IIA	HV360_08990	HXW74_10735	2.7.1.205	NV	6.13
	hypothetical protein	HV360_08985	HXW74_10740		NV	6.60
	PTS cellobiose transporter subunit IIC	HV360_08980	HXW74_10745	2.7.1.205	FS	6.25
Trehalose	trehalose operon repressor	HV360_06435	HXW74_03230		3.06	3.59
	beta-phosphoglucomutase	HV360_06440	HXW74_03225	5.4.2.6	3.05	3.57
	glycoside hydrolase family 65 protein	HV360_06445	HXW74_03220	2.4.1.216	2.81	2.58
	PTS system trehalose-specific EIIBC component	HV360_06450	HXW74_03215	2.7.1.201	2.02	2.34
Mannose	mannose-6-phosphate isomerase, class I	HV360_09365	HXW74_06890	5.3.1.8	3.18	*1.06*
Glucosamine (putatively)	SIS domain-containing protein	HV360_06680	X		2.33	X
	SIS domain-containing protein	HV360_06685	X		2.48	X
	PTS system mannose/fructose/sorbose family transporter subunit IID	HV360_06690	X		2.51	X
	PTS sugar transporter subunit IIC	HV360_06695	X		2.71	X
	PTS sugar transporter subunit IIB	HV360_06700	X		2.90	X
	PTS mannose transporter subunit IIA	HV360_06705	X		2.61	X

Table 18 continued.

	sigma 54-interacting transcriptional regulator	HV360_11195	X		3.00	X
	PTS sugar transporter subunit IIA	HV360_11200	X		3.98	X
	PTS sugar transporter subunit IIB	HV360_11205	X		3.52	X
	PTS sugar transporter subunit IIC	HV360_11210	X		2.70	X
	PTS system mannose/fructose/sorbose family transporter subunit IID	HV360_11215	X		2.92	X
	SIS domain-containing protein	HV360_11220	X		2.99	X
Glucosamine (putatively)	chromate transporter	HV360_11225	X		3.46	X
	chromate transporter	HV360_11230	X		2.81	X
	SDR family oxidoreductase	X	HXW74_09615	1.1.1.140	X	2.24
Sorbitol	HTH domain-containing protein	X	HXW74_09620		X	2.07
	glycerol dehydrogenase	HV360_08255	HXW74_06190	1.1.1.6	-2.37	*1.78*
	PTS-dependent dihydroxyacetone kinase	HV360_08250	HXW74_06195	2.7.1.121	*-0.70*	2.19
	dihydroxyacetone kinase subunit DhaK	HV360_08245	HXW74_06200	2.7.1.121	*-0.89*	2.04
Glycerol	dihydroxyacetone kinase subunit L	HV360_08240	HXW74_06205	2.7.1.121	*-0.84*	2.31
	glycerol kinase GlpK	HV360_09185	HXW74_09205	2.7.1.30	*0.54*	2.21
	type 1 glycerol-3-phosphate oxidase	HV360_09180	HXW74_09210	1.1.3.21	*0.69*	2.77
Glycerol	aquaporin family protein	HV360_09175	HXW74_09215		*0.63*	2.52
Mannitol	PTS mannitol transporter subunit IICBA	HV360_05275	HXW74_04665	2.7.1.197	*-1.59*	-3.65
Salicin/arbutin (putatively)	6-phospho-beta-glucosidase (1)	HV360_08285	HXW74_06160	3.2.1.86	-4.25	-3.98
	PTS glucose transporter subunit IIA	HV360_08280	HXW74_06165		-4.31	-4.27
	PTS sugar transporter subunit IIA	HV360_00385	HXW74_00510	2.7.1.202	-4.22	-4.89
	1-phosphofructokinase	HV360_00390	HXW74_00505	2.7.1.56	-4.00	-4.90
Fructose	DeoR/GlpR transcriptional regulator	HV360_00395	HXW74_00500		-4.20	-5.07
Glucosamine	glutamine--fructose-6-phosphate transaminase (isomerizing)	HV360_00345	HXW74_00550	2.6.1.16	*-1.01*	-2.96
	PTS fructose transporter subunit IIA	HV360_09995	X		4.95	X
	PTS sugar transporter subunit IIB	HV360_10000	X		4.94	X
	PTS sugar transporter subunit IIC	HV360_10005	X		5.27	X
	PTS system mannose/fructose/sorbose family transporter subunit IID	HV360_10010	X		4.96	X
Fructolysin	SIS domain-containing protein	HV360_10015	X		5.43	X
Sucrose (putatively	endonuclease/exonuclease/phosphatase family protein	HV360_03845	HXW74_06245		*1.75*	2.09
	2-hydroxycarboxylate transporter family protein	HV360_05475	HXW74_02240		2.18	2.48
	[citrate (pro-3S)-lyase] ligase citC	HV360_05490	HXW74_02255	6.2.1.22	2.19	2.93
	citrate lyase acyl carrier protein citD	HV360_05495	HXW74_02260		2.27	3.36
	citrate (pro-3S)-lyase subunit beta CitE	HV360_05500	HXW74_02265	4.1.3.34	2.10	3.52
	citrate lyase subunit alpha citF	HV360_05505	HXW74_02270	2.8.3.10	2.06	3.50
Citrate	citrate lyase holo-[acyl-carrier protein] synthase CitX	HV360_05510	HXW74_02275	2.7.7.61	2.23	3.72
	sodium ion-translocating decarboxylase subunit beta	HV360_10380	HXW74_02215	7.2.4.2	*1.87*	3.15
	biotin/lipoyl-binding protein	HV360_10385	HXW74_02220	7.2.4.2	*1.68*	3.10
	OadG family protein	HV360_10390	HXW74_02225	7.2.4.2	2.02	3.30
Oxalacetate	oxaloacetate decarboxylase subunit alpha	HV360_10395	HXW74_02230	7.2.4.2	*1.91*	2.62

Table 18 continued.

	pyruvate, phosphate dikinase	HV360_08015	HXW74_08830		0.73	2.50
	dihydrolipoyllysine-residue acetyltransferase	HV360_00115	HXW74_00795	2.3.1.12	0.84	2.31
	alpha-ketoacid dehydrogenase subunit beta	HV360_00110	HXW74_00800	1.2.4.1	1.02	2.52
	pyruvate dehydrogenase (acetyl-transferring) E1 component subunit alpha	HV360_00105	HXW74_00805	1.2.4.1	1.00	2.65
	alanine dehydrogenase	HV360_07735	HXW74_07135	1.4.1.1	1.18	2.15
	L-lactate dehydrogenase	HV360_07515	HXW74_08205	1.1.1.27	0.40	2.49
Pyruvate	aldose 1-epimerase family protein	HV350_07250	HXW74_07925		-1.89	-2.11
	pyruvate:ferredoxin (flavodoxin) oxidoreductase	HV360_01935	HXW74_01520	1.2.7.1	-0.31	2.26
	SUF system NifU family Fe-S cluster assembly protein	HV360_01615	HXW74_03680		0.88	2.19
	lipoate--protein ligase family protein	HV360_04570	HXW74_10860		-0.01	4.26
Co-factors	transporter substrate-binding domain-containing protein	HV360_04575	HXW74_10865		0.70	3.72
	arginine deiminase	X	HXW74_08090	3.5.3.6	X	5.73
	ornithine carbamoyltransferase	X	HXW74_08095	2.1.3.3	X	5.87
	carbamate kinase	X	HXW74_08100	2.7.2.2	X	5.65
	Crp/Fnr family transcriptional regulator	X	HXW74_08105		X	5.39
L-arginine	YfcC family protein	X	HXW74_08110		X	4.02
Polar amino Acids	amino acid ABC transporter permease	X	HXW74_07015		X	2.13
	ABC transporter permease subunit	HV360_11155	HXW74_05020		-3.45	-3.20
Branch-chain amino acids	amino acid ABC transporter ATP-binding protein	HV360_11160	HXW74_05025		-3.11	-2.95
Aromatic amino ac	3-phosphoshikimate 1-carboxyvinyltransferase aroA	HV360_00675	HXW74_00200	2.5.1.19	-0.61	-2.10
	DNA-directed RNA polymerase subunit beta'	HV360_06645	HXW74_10050		-2.29	-1.21
	anaerobic ribonucleoside-triphosphate reductase activating protein, nrdG	HV360_06095	HXW74_01150	1.97.1.4	0.24	2.30
	bifunctional pyr operon transcriptional regulator/uracil phosphoribosyltransferase PyrR	HV360_00770	HXW74_00105	2.4.2.9	-0.36	-2.17
	uracil permease uraA	HV360_00765	HXW74_00110		-0.78	-2.66
	aspartate carbamoyltransferase catalytic subunit, pyrB	HV360_00760	HXW74_00115	2.1.3.2	-1.06	-2.85
	dihydroorotase, pyrC	HV360_00755	HXW74_00120	3.5.2.3	-0.90	-2.72
	carbamoyl phosphate synthase small	HV360_00750	HXW74_00125	6.3.5.5	-0.85	-2.43
Nucleotides	carbamoyl phosphate synthase large subunit, carB	HV360_00745	HXW74_00130	6.3.5.5	-0.91	-2.19
	ferrous iron transport protein B	HV360_04080	HXW74_01545		-2.14	-0.59
Ions	zinc ABC transporter substrate-binding protein	HV360_07445	HXW74_10635		-2.03	-1.44
Compatible solutes	betaine-aldehyde dehydrogenase	HV360_08380	HXW74_06065	1.2.1.8	-3.99	-0.78
	iron-containing alcohol dehydrogenase	HV360_08375	HXW74_06070	1.1.1.1	-4.32	-1.14

The mapping of the biological function of the DEGs revealed, that many are associated with carbohydrate metabolism, transport of carbohydrates, central carbon metabolism or amino acid metabolism (Table 18).

Galactose metabolizing pathways such as the two non-identical versions of the tagatose-6-P pathway were increased in TMW 2.2254 and TMW°2.2256. In TMW°2.2254 version 1 showed a log_2-fold increase of 2 to 4.2, while version 2 showed a log_2-fold increase of 2.5 to 3.5. In TMW 2.2256 the respective CDS showed a log2-fold increase of 3.6 to 6.2 for version 1 while version 2 showed a log_2-fold increase of 2.3 to 5.5. The difference between these two versions in either strain is the addition of an additional HAD phosphatase CDS. Comparison of the amino acid sequence of each CDS in these operons revealed a 70 to 90% identity across all CDSs besides the *lacC* CDS. *lacC* and *lacC2* have an identity of 66 to 67 % across the entire amino acid sequence (Figure 13).

The Leloir pathway, another galactose metabolizing pathway consisting of the genes *galT*, *galE* and *galK,* showed a log_2-fold increase in TMW 2.2254 by 3.4 to 3.7. In TMW 2.2256 this pathway showed a log_2-fold increase of 2.3 to 3.1.

An operon associated with putative metabolism of tagatose or fructose showed a log_2-fold increase of 4.5 to 5.9 and 5.6 to 6.1 in TMW 2.2254 and TMW 2.2256, respectively. The fructose-utilizing operon *fruAKR* showed a log_2-fold decrease of 4 to 4.2 and 4.8 to 5 in TMW 2.2254 and TMW 2.2256. In TMW 2.2254 a mannose-6-phosphate isomerase showed a log_2-fold increase of 3.1.

Operons and CDSs associated with the metabolism of disaccharides were also changed. In TMW 2.2256 a phosphatase putatively involved in sucrose metabolism showed a log_2-fold increase of 2. The operon associated with the metabolism of cellobiose showed a log_2-fold increase of 2.9 to 3.5-fold in TMW 2.2256. The alpha-galactosidase, associated with the metabolism of melibiose or raffinose showed a log_2-fold increase of 2.11 in TMW 2.2254. The lactose metabolizing operon was only encoded by the strain TMW 2.2256, this operon showed a log_2-fold increase of 6.3 to 8.1-fold. The operon responsible for the transport and the metabolism of trehalose showed a log_2-fold increase of 2 to 3-fold in TMW 2.2254 and 2.3 to 3.5-fold in TMW 2.2256. Both strains encode for a 6-phospho-beta-glucosidase with adjacent PTS-IIA of which the function is unclear. However, these CDSs showed a log_2-fold decrease of 4.2 and 4.3-fold in TMW 2.2254 and 3.9 and 4.2-fold in TMW 2.2256.

Changes in the abundance of CDSs responsible for the metabolism or transport of sugar alcohols were also found. A mannitol transporter spanning the PTC-IICBA domains showed a log_2-fold decrease of 3.6 in TMW 2.2256. The operon encoding for the phosphorylating pathway for utilization of glycerol *glpKOF* showed a log_2-fold increase of 2.2 to 2.7-fold in TMW 2.2256. The operon encoding for the dehydrogenation pathway for the utilization of glycerol *gldA-dhaMKL* was only partly changed in the strains. In TMW 2.2256 the *dhaMKL* subunits showed a log_2-fold increase of 2 to 2.3-fold and in TMW 2.2254 only the glycerol dehydrogenase *gldA* was decreased by 2.3-fold. Two CDSs putatively associated with the metabolism of sorbitol showed a log_2-fold increase of 2 to 2.2-fold in TMW 2.2256.

The strain TMW 2.2254 has three operons increased with putative functions. The first operon (HV360_06680-HV360_06705) encodes for two SIS domain proteins and showed a log_2-fold increase of 2.3 and 2.4-fold. An adjacent PTS-IIABCD system was increased by 2.5-2.9-fold.

The second operon (HV360_11195-HV360_11230) encodes for a sigma 54-interacting transcriptional regulator that showed a log_2-fold increase of 3, a PTS system that showed a 2.9 to 3.9-fold increase, an SIS domain protein that showed a 2.9-fold increase and two adjacent chromate transporter that showed a 3.4 to 2.8-fold increase. The third operon consisted of an SIS domain-containing protein that showed a log_2-fold increase of 5.4 and an adjacent PTS system that showed a 4.9 to 5.2-fold increase. Fewer changes were found in amino acid metabolizing pathways. The ADI pathway of TMW 2.2256 showed a log_2-fold increase of 4 to 5.8 while at the same time, the CDS from *aroA* showed a log_2-fold decrease of 2. Furthermore, an alanine dehydrogenase was increased in TMW 2.2256 by a log_2-fold value of 2.1. CDSs responsible for the transport of amino acid were also changed, as an ABC amino acid permease subunit was increased in TMW 2.2256 by a log_2-fold value of 2.1. Another permease subunit and ATP-binding protein CDS were decreased by a log_2-fold value of 3.1 to 3.4-fold in TMW 2.2254 and 2.9 to 3.2-fold in TMW 2.2256.

The citrate lyase with an adjacent transporter was increased in both strains. TMW 2.2254 increased the operon with the transporter by a log_2-fold value of 2.06 to 2.2. TMW 2.2256 increased the operon with the transporter by a log_2-fold value of 2.4 to 3.7. The genomically closely located oxalacetate decarboxylase was increased by a log_2-fold value of 2.6 to 3.3 in TMW 2.2256. TMW 2.2254 only increased the *oadG* subunit by a log_2-fold value of 2.

Changes in the CDSs associated with the central carbon metabolism were also detected. In the strain TMW 2.2256 a pyruvate: ferredoxin oxidoreductase was increased by a log_2-fold value of 2.2, the pyruvate dehydrogenase was increased by 2.3 to 2.6-fold, the pyruvate phosphate dikinase was increased by 2.5-fold and a lactate dehydrogenase was increased by 2.4-fold. Furthermore, a lipoate-protein ligase and adjacent transporter were increased by a log_2-fold value of 4.2 and 3.7 in TMW 2.2256. A cluster assembly protein of the SUF system (*NifU*) was increased by a log_2-fold value of 2.1 and an aldose 1-epimerase was decreased by 2.1-fold in TMW 2.2256.

The operon responsible for the pyrimidine synthesis, consisting of *pyrR*, *uraA*, *carA* and *carB,* was decreased by a log_2-fold value of 2.1 to 2.6 in TMW 2.2256. At the same time, the anaerobic ribonucleotide-triphosphate reductase activating protein (*nrdG*) was increased by a log_2-fold value of 2.2 in TMW 2.2256. In TMW 2.2254 only the DNA-directed RNA polymerase beta was decreased by a log_2-fold value of 2.2.

RESULTS

The CDSs of proteins associated with osmo-stress and osmo-balance such as the betaine-aldehyde dehydrogenase or the choline dehydrogenase were decreased in TMW 2.2254 by a log_2-fold value of 3.9 to 4.3. Two CDSs associated with ions were decreased in TMW 2.2254, namely the iron transporter *fetB* and a zinc substrate binding unit. *FetB* was decreased by a log_2-fold value of 2 and the zinc substrate binding unit was decreased by 2.1-fold.

A total of 64 DEGs could not be assigned to a metabolic pathway (Table 26).

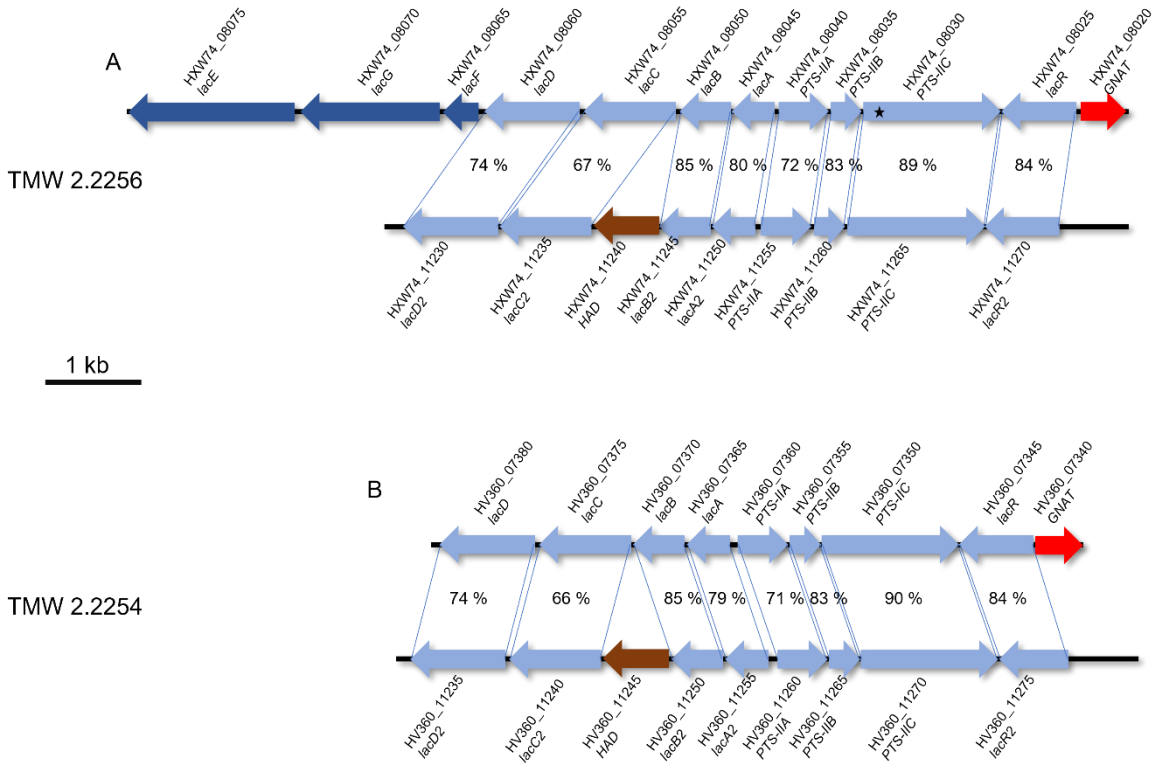

Figure 13: Genetic organization of the CDSs for the tagatose-6P-Pathways in *T. halophilus* TMW 2.2254 and TMW 2.2256. A= Organization of the tagatose-6P-Pathway *lacCDBAR* and *lacCDBAR2* in TMW 2.2256 with the amino acid identity shown in percent over the entire sequence of the respective CDS; * = indicates a frameshift with a resulting premature stop codon. B= Organization of the tagatose-6P-Pathway *lacCDBAR* and *lacCDBAR2* in TMW 2.2254 with the amino acid identity shown in percent over the entire sequence. lacAB/*lacAB2*= galactose-6-phosphate isomerase subunit A or B; lacC/*lacC2*= tagatose 6-phosphate kinase; lacD/*lacD2*= tagatose 1,6-diphosphate aldolase*; lacG = 6-phospho-β-galactosidase LacEF* = lactose specific PTS system*; PTS-IIABC*/*PTS-IIABC2*= Galactose specific PTS systems; *lacR*/*lacR2* = putative regulator; *HAD*= HAD family phosphatase; GNAT= GNAT acetyltransferase. The scale bar represents 1 Kb. The sequence similarities are displayed in percent. Alignment was performed using the NCBI blastP program on the NCBI website (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch). Figure modified from Link and Ehrmann, 2023a.

4.6 Monitoring the strain-dependent growth dynamics of *T. halophilus* strains in a lupine moromi pilot fermentation

To determine the most competitive *T. halophilus* starter strain, a model lupine moromi fermentation was inoculated with a set of selected *T. halophilus* strains. Then, the fermentations were monitored for three weeks in terms of their pH and the bacterial cells. The bacterial growth was monitored by counting of cells grown on MRS plates. The growth dynamics of selected *T. halophilus* strains were monitored by identification by colony-PCR using the strain-specific multiplex PCR system.

Parts of these results have been published (Link and Ehrmann, 2023b).

4.6.1 Development of multiplex PCR-based methods to discriminate between *T. halophilus* strains

The only way to discriminate the strains used in this thesis was RAPD PCR, a method that is time-consuming and unpredictable. The outcome relies on the random amplification of unknown fragments and it is not applicable to a higher throughput. To tackle these problems a new way of identifying the strains has been developed.
The first attempt was based on the identification of a strain-specific pattern based on the presence or absence of marker ORFs that were identified by the BADGE program. These marker ORFs were the L-arabinose isomerase *araA*, the xylose isomerase *xylA*, a *BppU* phage baseplate protein, an ATP-endonuclease, a sigma 54 regulator, a 6-phospho-beta-galactosidase, a MarR family regulator, an ABC transporter and the ribose transporter *rbsU*. The primers only amplified parts of the ORFs. The fragment of the ribose transporter *rbsU* served as a PCR control as all the strains possess this ORF. The pattern for the strains was chosen as follows: DSM 20339^T= 498 bp, 745 bp and 1041 bp; TMW 2.2254 = 745 bp and 2401 bp; TMW 2.2256 = 745 bp, 1041 bp and 1385 bp; TMW 2.2257 = 745 bp and 1828 bp; TMW 2.2260 = 745 bp and 1315 bp; TMW 2.2263 = 745 bp, 1041 bp and 1842 bp; TMW 2.2264 = 745 bp; TMW 2.2266 = 745 bp, 1041 bp and 1442 bp (Figure 14).

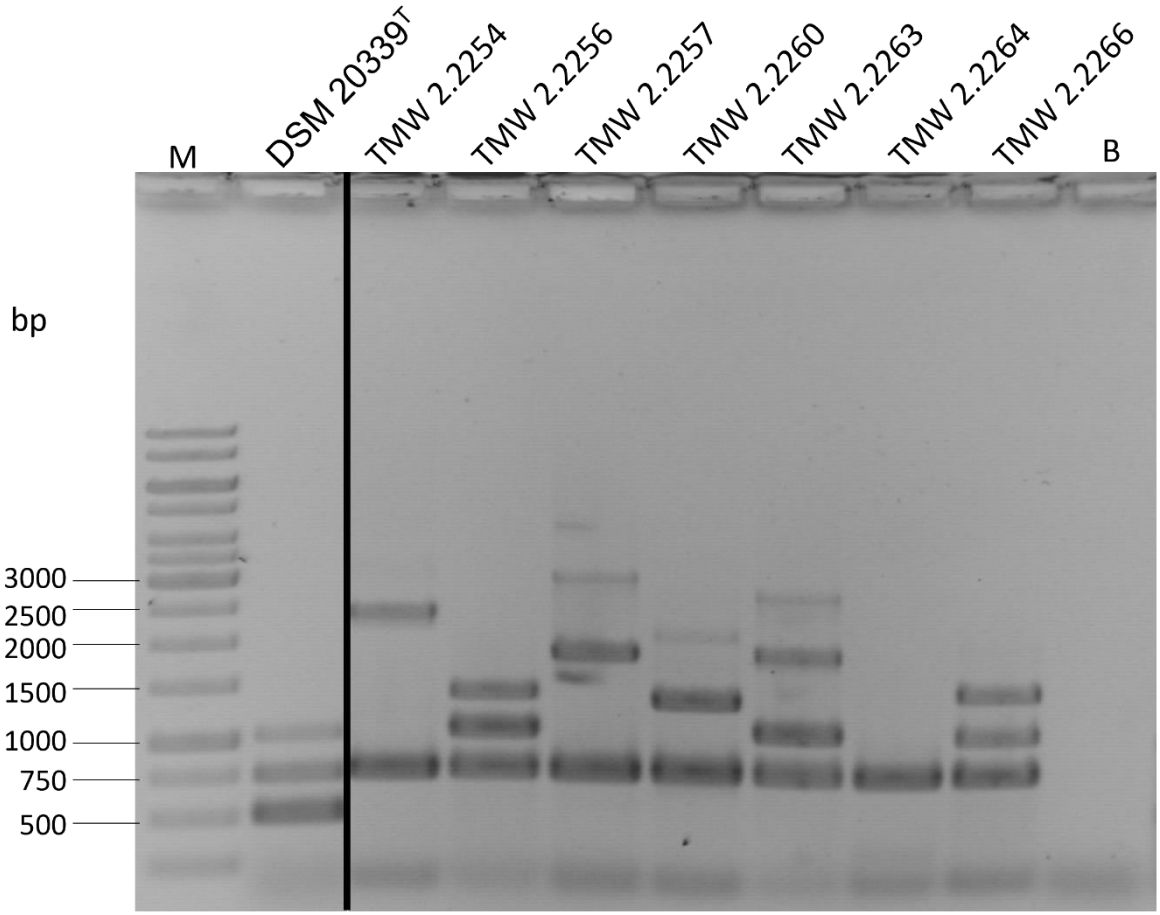

Figure 14: Fragments generated by the multiplex PCR primer set using marker ORFs to generate a strain-specific pattern. The expected bands for the strains DSM 20339^T= 498 bp, 745 bp and 1041 bp; TMW 2.2254 = 745 bp and 2401 bp; TMW 2.2256 = 745 bp, 1041 bp and 1385 bp; TMW 2.2257 = 745 bp and 1828 bp; TMW 2.2260 = 745 bp and 1315 bp; TMW 2.2263 = 745 bp, 1041 bp and 1842 bp; TMW 2.2264 = 745 bp; TMW 2.2266 = 745 bp, 1041 bp and 1442 bp. M = Generuler 1 Kb DNA ladder (Thermo Scientific, Waltham, USA); Buffer = PCR master mix without any DNA.

With the multiplex PCR system discrimination between the strains was possible. All strains showed the expected band pattern. The reaction also yielded several unspecific products in the lanes of TMW 2.2257, TMW 2.2260 and TMW 2.2263 (Figure 14).

The second approach for the identification was a primer set designed by using the "Rapid identification of PCR primers for unique core sequences "(RUCS). All genomes from set TMW 2.2254, TMW 2.2256, TMW 2.2257, TMW 2.2260, TMW 2.2263, TMW 2.2264, TMW 2.2266 and DSM 20339^T were used as input for the primer design. The program then compared the genomes and filtered regions specific to the genomes and designed primers for these regions with different lengths. The primer sets from the output were further checked for their melting temperature using the Taq polymerase and for the probability of creating secondary structures. The final primer sets were selected in a way that every strain had one specific PCR band with a unique length. In TMW 2.2254 the primers targeted a PTS-EIIC subunit and an oxidoreductase, with a

resulting fragment size of 2552 bp. The primers for TMW 2.2256 amplified parts of the ethanolamine ammonia-lyase *EutB*, with a fragment size of 596 bp. The primer for TMW 2.2257 amplified parts of a nucleoside hydrolase and a multidrug transporter, with a predicted fragment size of 822 bp. In TMW 2.2260 the strain-specific primers targeted the 4-hydroxy-tetrahydrodipicolinate synthase *dapA* and an adjacent PTS-IIC subunit, generating an 1119 bp long fragment. In TMW 2.2263 the primers targeted a phage region. The small terminase and an adjacent large terminase subunit are partially amplified generating a 333 bp long fragment. In TMW 2.2264 the selected primers amplified parts of a sulfatase and an adjacent SPASM domain protein generating a fragment with a length of 1583 bp. Parts of a transglycosylase SLT domain-containing protein inside a phage region were amplified for identification of TMW 2.2266 by generating a fragment with a 1961 bp size. For the identification of the type strain DSM 20339^T a fragment with a length of 2882 bp consisting partly of a PTS-IIC transporter, a hypothetical protein and a methionine ABC transporter was chosen (Figure 15).

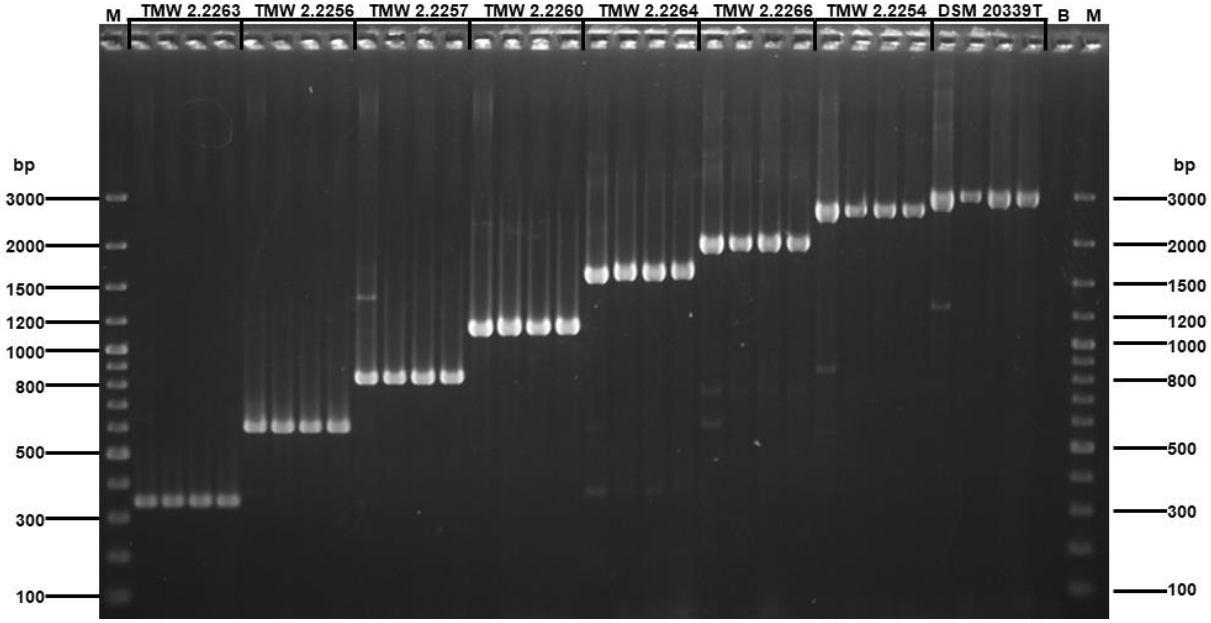

Figure 15: Amplified DNA fragments of the multiplex PCR primer set using single ORFs for the strain-specific identification. Lane one for every strain used gDNA as input. Lanes two, three and four used representative colonies from the respective strains. The expected bands for the strains: TMW 2.2263 = 333 bp, TMW 2.2256 = 596 bp, TMW 2.2257 = 822 bp, TMW 2.2260 = 1119 bp, TMW 2.2264 = 1583 bp, TMW 2.2266 = 1961 bp, TMW 2.2254 = 2552 bp, DSM 20339^T = 2882 bp. M = Generuler 100 bp (Thermofisher Scientific, Waltham, MA, USA). B = PCR mastermix control. Figure modified from Link and Ehrmann, 2023b.

Four PCR reactions per strain were done, the first lane for every strain had gDNA as input. Lanes two to four for all strains were single representative colonies from the respective strains. All PCR reactions yielded the expected major bands. The bands

using gDNA as input were slightly stronger than the ones using colonies as input. In the first lanes of TMW 2.2257, TMW 2.2264, TMW 2.2266, TMW 2.2254 and DSM 20339^T faint unspecific bands were visible.

4.6.2 Lupine moromi pilot fermentation

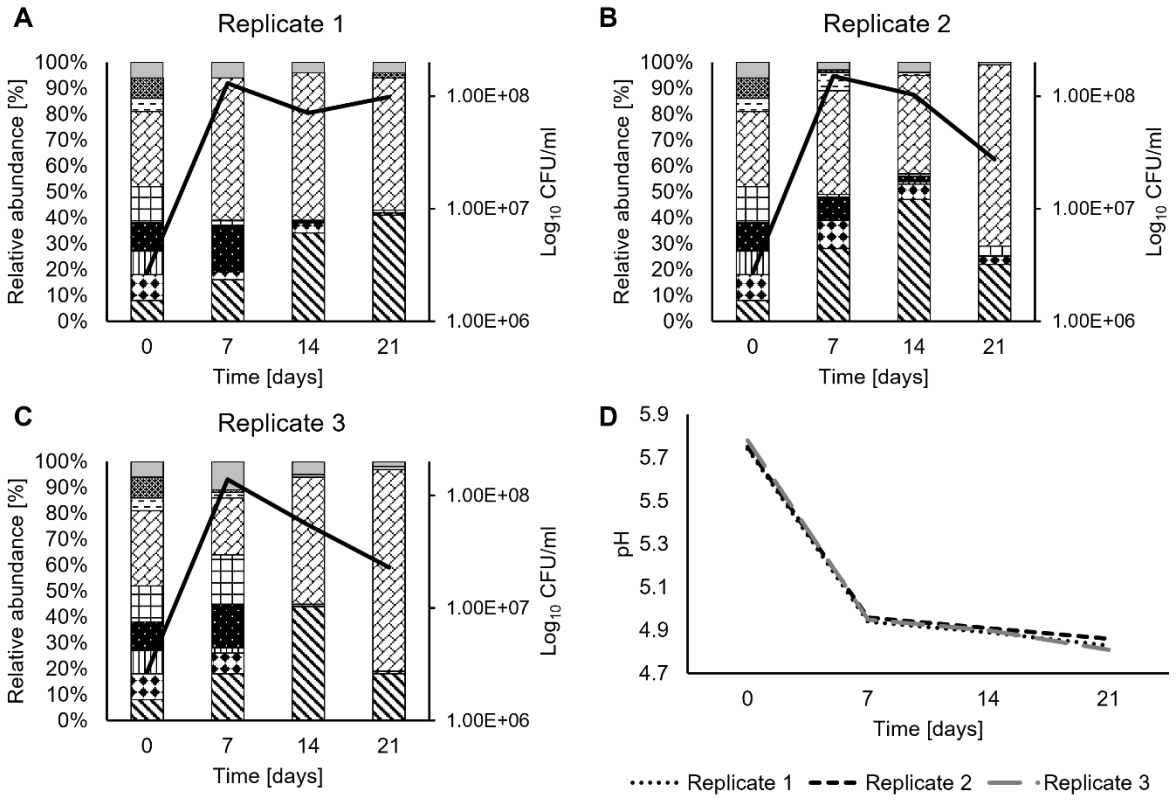

Figure 16: Monitoring of the competitive lupine moromi fermentation. The strain composition and the pH development were monitored in a lupine moromi pilot fermentation. A= Strain composition and cell counts of the first pilot fermentation replicate; B= Strain composition and cell counts of the second pilot fermentation replicate; C= Strain composition and cell counts of the third pilot fermentation replicate; D= pH values of the three replicates over the course of 21 days. Legend in A, B and C starting from the bottom of every column: ( )= TMW 2.2254; ( )= TMW 2.2256; ( )= TMW 2.2257; ( )= TMW 2.2260; ( )= TMW 2.2263; ( )= TMW 2.2264; ( ) = TMW 2.2266; ( ) = DSM 20339^T; gray= not clearly identifiable, meant that none or multiple bands occurred and this colony was not assigned to a specific strain. The black lines in A, B and C represent the respective cell count in CFU/ml of the replicate at every sampling point. Figure modified from Link and Ehrmann, 2023b.

As all fermentations started with identical inoculate, all the replicates started with the same cell count (2.68 x 10^6 CFU/ml) and strain composition (Figure 16). After seven days all replicates exceeded cell counts of 10^8 CFU/ml. After 14 days the cell counts decreased in all replicates to a value between 5.45 x 10^7 to 1.03 x 10^8 CFU/ml. After 21 days the cell counts for all replicates were in the range of 2.3 x 10^7 to 9.85 x 10^7 CFU/ml. The pH dropped below 5 within the first seven days and further decreased to a value between 4.8 and 4.86 after 21 days.

RESULTS

The strain TMW 2.2254 grew to cell counts between 2.4 and 4.84 x 10^7 CFU/ml across all replicates over the course of 14 days. A decrease to cell counts between 4.14 and 6.05 x 10^6 CFU/ml was detected in two replicates, while an increase to 4.04 x 10^7 CFU/ml was detected in one replicate. The strain TMW 2.2256 grew up to a cell count of 1.67 x 10^7 CFU/ml after seven days across all replicates. Then, the cell counts steadily decreased to a cell count between 2.30 and 9.85 x 10^5 CFU/ml after 21 days. The strain TMW 2.2257 was only detected twice after the initial inoculation, after seven days in replicate 3 with a cell count of 2.77 x 10^7 CFU/ml and after 14 days in replicate 2 with a cell count of 1.03 x 10^6 CFU/ml. TMW 2.2260 reached cell counts between 1.36 and 2.35 x 10^7 CFU/ml after seven days, followed by a steady decrease in the next weeks across all replicates. After 14 days the strain was only detected in two replicates with cell counts between 7.1 x 10^5 and 2.06 x 10^6 CFU/ml. After 21 days the strain was not detected in any replicates. After seven days TMW 2.2263 was detectable with cell counts within the range of 1.5 to 2.6 10^7 CFU/ml across all replicates. After 14 days the strain was only detected in two replicates with cell counts between 5.45 x 10^5 and 1.03 x 10^6 CFU/ml. After 21 days the strain was detected in two replicates with cell counts between 9.85 x 10^5 and 1.1 x 10^6 CFU/ml. The strain TMW 2.2264 was detected in all replicates across all sampling time points. The strain reached cell counts between 3.05 and 7.15 x 10^7 CfU/ml after seven days across all replicates. Followed by a steady decrease to 1.79 and 1.93 x 10^7 CFU/ml in two replicates, while increasing to 5.02 x 10^7 CFU/ml in one replicate. The strain TMW 2.2266 was only detected in two replicates after seven days with cell counts ranging from 2.77 10^6 to 1.06 x 10^7 CFU/ml. Then cell counts decreased to 5.45 x10^5 and 1.03 x 10^6 CFU/ml after 14 days. After 21 days the strain was only detected in one replicate with a cell count of 2.3 x10^5 CFU/ml. The type strain DSM 20339^T was only detected in two replicates after seven days with cell counts ranging from 1.39 to 1.52 x 10^6 CFU/ml. After 21 days the strain was only detected with 1.97 x 10^6 CFU/ml in one replicate. PCR reactions yielding more than one band or none were counted and classified as “not clearly identifiable” (Table 19).

Table 19: Calculated cell counts of every strain within the lupine moromi model fermentation in all replicates. The cell counts for every strain were calculated by multiplying the total CfU/ml with the percentual distribution of every strain within a replicate. To calculate the percentual distribution, one hundred colonies were evaluated *via* colony PCR as described in the materials and methods section. n.D = Not detected; not clearly identifiable = reactions that yield none or more than one band were not assigned to one specific strain.

	Replicate 1				Replicate 2				Replicate 3			
Sample day	0 Days	7 Days	14 Days	21 Days	0 Days	7 Days	14 Days	21 Days	0 Days	7 Days	14 Days	21 Days
TMW 2.2254	2.14E+05	2.08E+07	2.41E+07	4.04E+07	2.14E+05	4.24E+07	4.84E+07	6.05E+06	2.14E+05	2.49E+07	2.40E+07	4.14E+06
TMW 2.2256	2.68E+05	3.90E+06	2.84E+06	9.85E+05	2.68E+05	1.67E+07	6.18E+06	8.25E+05	2.68E+05	1.11E+07	n.D	2.30E+05
TMW 2.2257	2.41E+05	n.D	n.D	n.D	2.41E+05	n.D	1.03E+06	n.D	2.41E+05	2.77E+06	n.D	n.D
TMW 2.2260	2.95E+05	2.34E+07	7.10E+05	n.D	2.95E+05	1.36E+07	2.06E+06	n.D	2.95E+05	2.35E+07	n.D	n.D
TMW 2.2263	3.75E+05	2.60E+06	n.D	9.85E+05	3.75E+05	1.52E+06	1.03E+06	1.10E+06	3.75E+05	2.63E+07	5.45E+05	n.D
TMW 2.2264	7.77E+05	7.15E+07	4.05E+07	5.02E+07	7.77E+05	6.06E+07	3.91E+07	1.93E+07	7.77E+05	3.05E+07	2.67E+07	1.79E+07
TMW 2.2266	1.34E+05	n.D	n.D	n.D	1.34E+05	1.06E+07	1.03E+06	n.D	1.34E+05	2.77E+06	5.45E+05	2.30E+05
DSM 20339^{T}	2.14E+05	n.D	n.D	1.97E+06	2.14E+05	1.52E+06	n.D	n.D	2.14E+05	1.39E+06	n.D	n.D
Not clearly identifi	1.61E+05	7.80E+06	2.84E+06	3.94E+06	1.61E+05	4.55E+06	4.12E+06	2.75E+05	1.61E+05	1.52E+07	2.73E+06	4.60E+05
total CfU/ml	2.68E+06	1.30E+08	7.10E+07	9.85E+07	2.68E+06	1.52E+08	1.03E+08	2.75E+07	2.68E+06	1.39E+08	5.45E+07	2.30E+07

4.7 Characterization of the yeast *D. hansenii* TMW 3.1188

The strain diversity from twenty-five isolates identified as *D. hansenii* by MALDI-TOF MS was investigated with RAPD-PCR. Then, a unique strain was selected and the consensus sequence spanning the ITS1, 5.8s rDNA and ITS2 region was submitted to GenBank. For further characterization, the whole genome of this strain was sequenced.

Parts of this was published in Link and Lülf et al., 2022.

4.7.1 Screening of isolates using RAPD PCR

The RAPD-PCR-based screening method was used to analyze the strain diversity among *D. hansenii* isolates from Lupine moromi. Two different primers were used, a M13 primer and another M13V primer. The M13 primer has been previously used to differentiate between strains of *D. hansenii* (Andrade et al., 2006; Huey and Hall, 1989). The M13V primer was composed of a different sequence and therefore might result in different DNA-fingerprints (Ehrmann et al., 2003). *D. hansenii* 281, 310, 311 and 325 were isolated from dry sausage fermentation and served as positive controls for the differentiation on a strain level. Genomic DNA from *S. cerevisiae* served as PCR-positive control (Figure 17 and Figure 18).

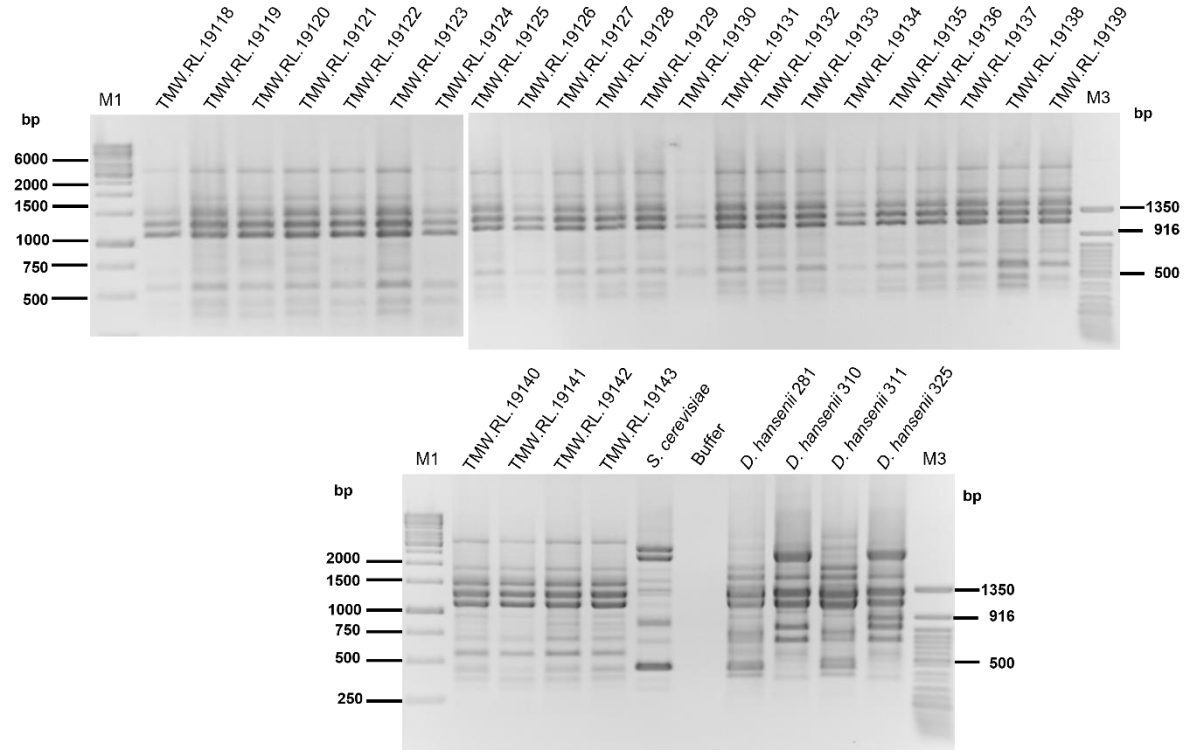

Figure 17: DNA-Fingerprints for *D. hansenii* isolates obtained with M13V primer (Ehrmann et al., 2003). M1 = Generuler 1Kb DNA ladder. M3 = Quick-Load® Purple 50 bp DNA ladder (New England biolabs). TMW.RL 19118 – 19143 D. hansenii isolates from lupine moromi. gDNA from S. cerevisiae was used as PCR-positive control. Buffer = PCR mastermix control.

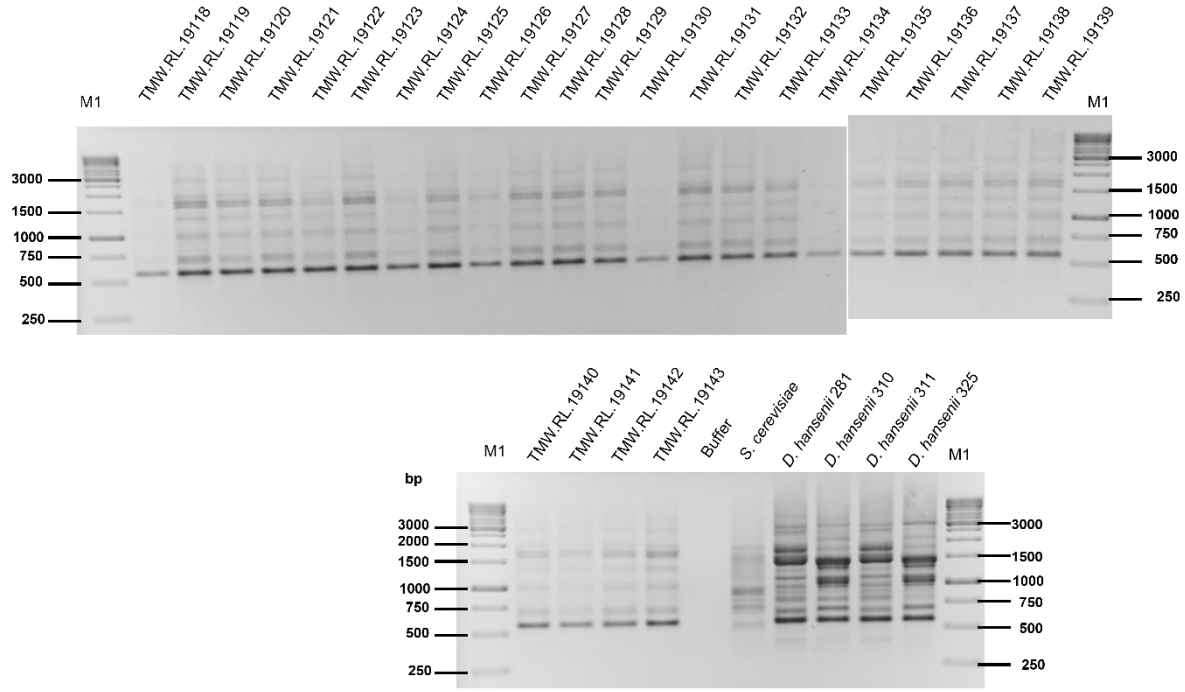

Figure 18: DNA-Fingerprints for *D. hansenii* isolates obtained with the primer M13 (Huey and Hall, 1989). M1 = Generuler 1Kb DNA ladder. M2 = Generuler 100 bp Plus DNA Ladder. TMW.RL 19118 – 19143 *D. hansenii* isolates from lupine moromi. gDNA from *S. cerevisiae* was used as PCR-positive control. Buffer = PCR mastermix control.

The fingerprints from the RAPD-PCR allowed differentiation between the species *D. hansenii* and *S. cerevisiae* with both primers. Furthermore, all isolates from lupine moromi could be differentiated from the isolates from dry sausages (281,310,311,325) (Figure 17 and Figure 18).

4.7.2 Genomic analyses, annotation and comparison to other *D. hansenii* strains

The DNA-Fingerprints generated using RAPD-PCR were identical for all isolates from lupine moromi, it was concluded that the isolates are just one strain isolated multiple times. Then, the isolate TMWRL.19143 was renamed to TMW 3.1188 and the entire genome was sequenced. After genome sequencing was done using the Pacbio Sequel technology. the assembled genome was quality-checked and afterwards annotated using YGAP. The genome was then compared to other genomes of the same species using the kr distance and ANIb values. Phylogenetic analyses were conducted using the consensus sequence of the ITS1, 5.8S rRNA and ITS2 region as well as the *ACT1* gene.

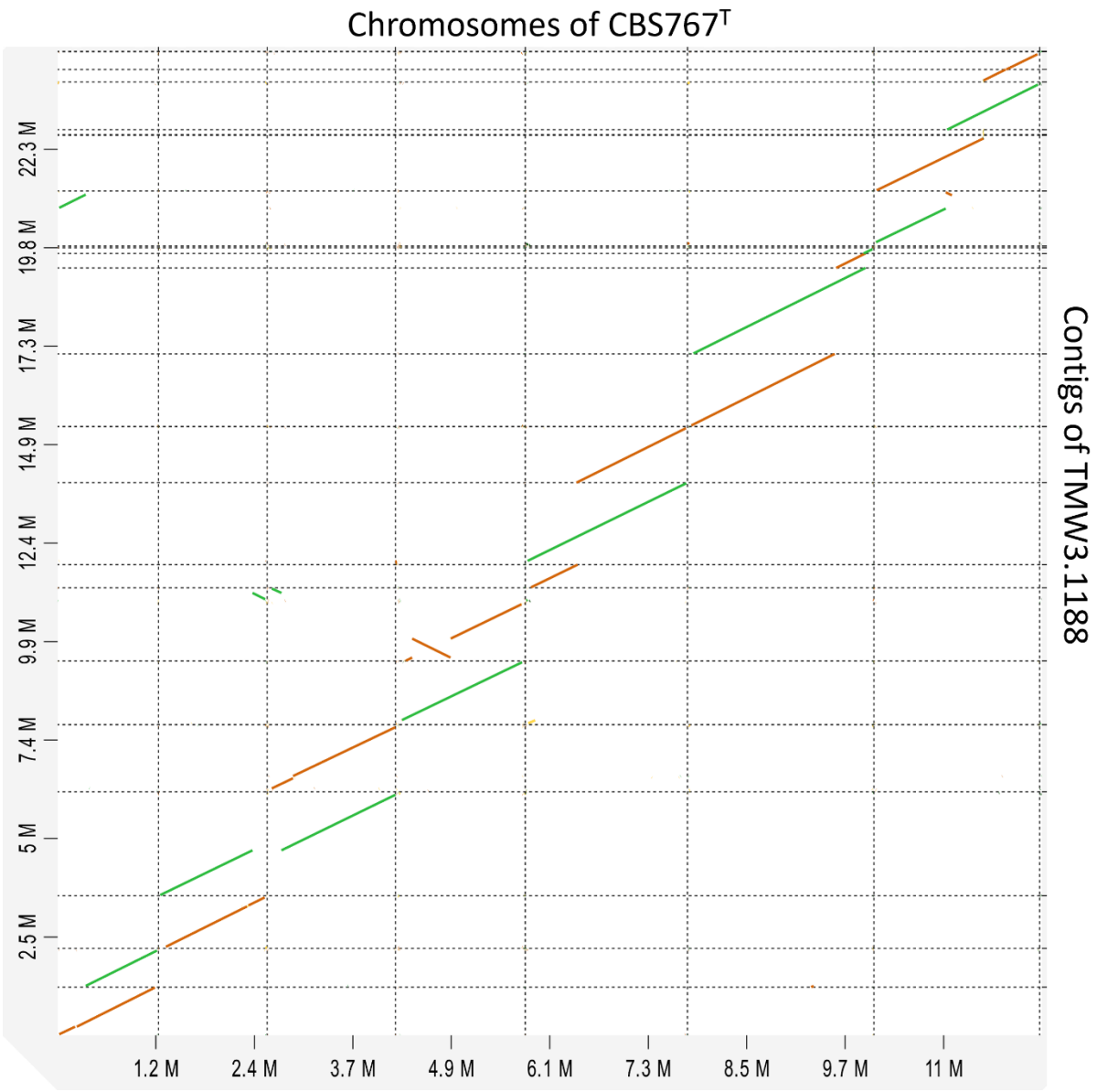

Figure 19: Visual alignment of the contigs from TMW 3.1188 to the chromosome from CBS767^T. The graph was created using D-genies (Cabanettes and Klopp, 2018). Green bars indicate an alignment with high identity. Orange to red bars indicate an alignment with low identity.

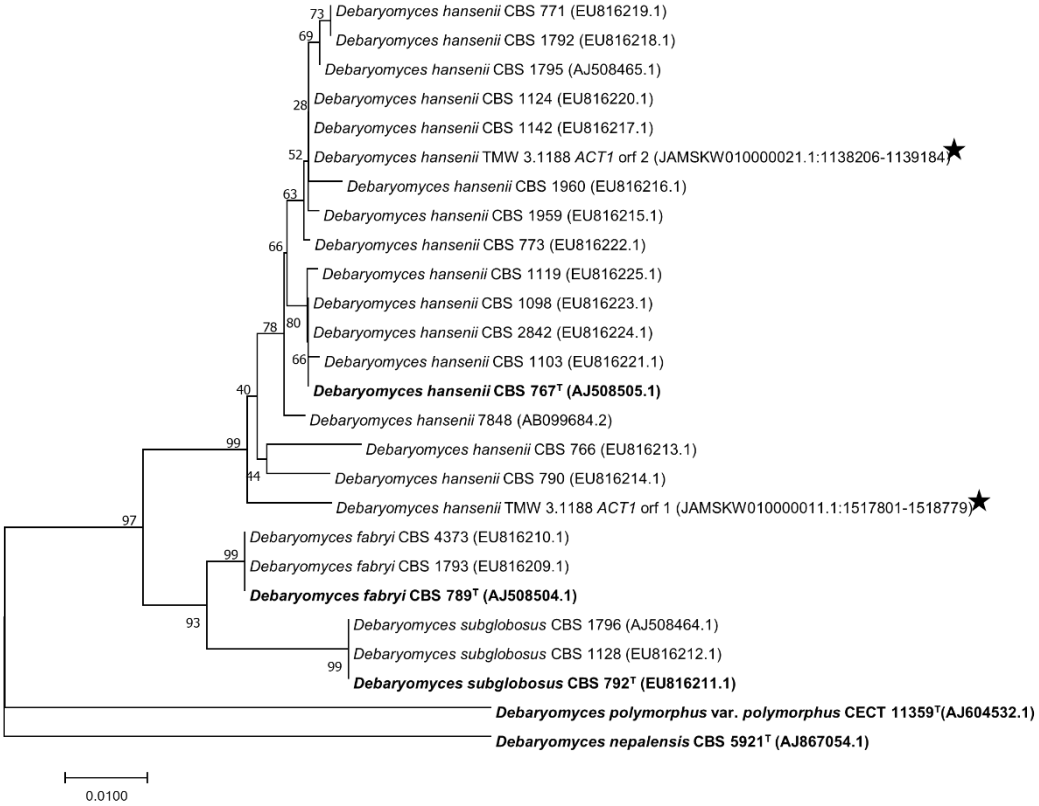

Figure 20:Phylogenetic analysis using the *ACT1* gene of *D. hansenii*. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 0.23095609 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated. There was a total of 752 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). Type strains were indicated by bold letters and "T". Black stars indicate the position of the sequences from TMW 3.1188. The accession numbers or positions in the chromosome are written in brackets. Figure modified from Link and Lülf et al., 2022.

Table 20: kr-distance and ANI values from TMW 3.1188 to selected other strains from the species *D. hansenii* and *D. fabryi*.

	Kr distance TMW 3.1188	ANI values TMW 3.1188
D hansenii MTCC234	**0.129**	**84.50**
D fabryi CBS789^T	**0.102**	**85.33**
D hansenii J6	**0.119**	**84.51**
D hansenii CBS767^T	**0.056**	**95.72**

RESULTS

The data quality was checked using the Pacific Biosciences SMRT Link software. 10,724,419,855 bases were sequenced throughout a 10 h sequencing movie. The mean subread length was 14,003 bases and the longest subread had 14,126 bases. *De novo* genome assembly was done with HGAP4 in SMRTLink version 10.0.0.108728 with the expected genome size set to 12 Mbp and all other parameters set to standard. The assembly yielded 26 contigs with an N_{50} value of 1,604,673 and a total sequence length of 24,773,645 base pairs. The GC content across the entire genome was 36.23%. The assembled genome was deposited at NCBI (GCA_024256405.1). The consensus sequence spanning the ITS1, 5.8S rRNA and ITS2 region was submitted to Genbank (OP179623).

The YGAP annotation detected 11090 CDS and 413 tRNAs. The comparison with the type strain of the species *D. hansenii* CBS 767^T showed that at least two contigs of *D. hansenii* TMW 3.1188 map to every chromosome of CBS767^T (Figure 19). The kr distance from TMW 3.1188 was 0.129 to the *D. hansenii* MTCC234, 0.119 to J6, 0.102 to the type strain of the species *D. fabryi* CBS769T and 0.056 to the type strain of the species *D. hansenii* CBS767^T. The strain TMW 3.1188 had the highest ANI value to the strain CBS767^T. To the strains MTC234 and J6 the ANI values were only 84.5 and 84.51. To the type strain of the species *D. fabryi* CBS789^{T,} the strain TMW 3.1188 had an ANI value of 85.33 (Table 20). The fact that the genome size of TMW 3.1188 was almost twice the expected size and that each contig of the reference genome had 2 contigs of the TMW 3.1188 genome aligned to it, led to the hypothesis that TMW 3.1188 is a diploid strain.

The phylogenetic analysis using both *ACT1* genes from the genome of TMW 3.1188 confirmed the affiliation of the strain with the species *D. hansenii* (Figure 20).

5 Discussion

This study investigates the genetic diversity of *T. halophilus* strains isolated from lupine fermentation in comparison to strains from other isolation sources such as soybean or fish fermentations. It further characterizes the species *T. halophilus* towards their genetic and metabolic adaptation to lupine moromi as a fermentation substrate. In detail, pathways and genes were identified, enabling their adaptation to the new substrate and linkages to their metabolic properties were established. A competitive lupine moromi fermentation using *T. halophilus* strains isolated from fish, buckwheat moromi or lupine moromi was performed, revealing that *T. halophilus* strains isolated from lupine moromi dominate the lupine moromi fermentation. Specific enzymes and pathways could be linked to the competitiveness of those strains. Lastly, the diversity of yeasts isolated from lupine moromi was investigated and the genome of a diploid *D. hansenii* strain was compared to other strains of the species.

From this work and the initial hypotheses, the following theses can be derived:

- Six strains of *T. halophilus* could be isolated from several lupine moromi fermentations, revealing the intra-species diversity within these fermentations.
- The species *T. halophilus* can be separated into three distinctive intra-species lineages with *T. halophilus* strains from lupine moromi all belonging to the same lineage.
- Genomic comparison of *T. halophilus* strains from lupine moromi with strains from other isolation sources reveals no predictable adaption to the lupine moromi environment but rather reveals the general adaptation of *T. halophilus* towards plant-based carbohydrate sources.
- Transcriptomic analysis of strains grown in LMRS with galactose reveals that *T. halophilus* has multiple pathways of utilizing galactose simultaneously.
- The growth behavior of *T. halophilus* strains in lupine moromi can be linked to strain-specific traits, such as the presence of alpha galactosidases and the duplication of the enzyme encoding for the tagatose pathway.
- *T. halophilus* strains from lupine moromi show a mucoid phenotype that is increased in the presence of galactose and high sodium chloride.
- The *D. hansenii* strain TMW 3.1188 isolated from lupine moromi adapted to the new environment by diploidy.

DISCUSSION

5.1 *T. halophilus* is an essential bacterial species in lupine moromi fermentations processes and strains isolated from lupine moromi cluster within one lineage.

At the beginning of this study, a total of 58 isolates identified as *T. halophilus* from different lupine moromi fermentations and multiple time points were analyzed for their genomic and physiological diversity. The lupine moromi fermentations consisted of a brine solution with sodium chloride concentrations between 10 - 20 % (w/v) that was mixed with lupine 'Koji' at a ratio of 1:1.5. *T. halophilus* plays an important role in the fermentation of soy sauces, it was up to the question if this species plays a similar role in the lupine moromi (Devanthi and Gkatzionis, 2019). RAPD-PCR followed by whole genome sequencing of the strains displaying a unique fingerprint led to the identification of 11 strains isolated from lupine moromi. After genomic comparison based on assembly statistics, ANIb value and annotation, six unique strains from lupine moromi were identified. Further comparison with other members from the species *T. halophilus* allowed the grouping into two lineages (Figure 8). Lineage I consisted of strains from lupine moromi, cheese brine, fish and soy sauce. Lineage II mainly consisted of strains from soy sauce with two exceptions from fish sauces. Using BADGE analysis six ORFs could be identified that were unique for lineage I and ten ORFs that were unique for lineage II (Table 23 and Table 24). Interestingly the clone analysis, based on ANIb values, revealed that TMW 2.2256 and TMW 2.2264 were isolated multiple times and were also isolated from the same fermentation at the same time point (15 % NaCl after two weeks) (Table 21). This fact proves the hypothesis that multiple strains can be present at the same time in the lupine moromi fermentation broth. A similar strain diversity can be found in soy sauce fermentations (Roling and van Verseveld, 1996; Uchida, 1982). Previous studies have shown that *T. halophilus* is capable of fermenting a vast range of carbohydrate sources in a strain-dependent manner (Roling and van Verseveld, 1996). A similarly variable carbohydrate metabolism allows the species *L. curvatus* to adapt to a plant-derived environment (Terán et al., 2018). This ability to adapt to a new environment *via* the carbohydrate metabolism might be the reason why *T. halophilus* is present in high numbers in the lupine moromi fermentation regardless of the added carbon sources (Lülf et al., 2023). Although, carbon sources play an important role in the adaptation to an environment, factors such as the sodium chloride levels can also influence the microbial consortium within a fermentation (Lülf et al., 2021). A comparison of the optimal sodium chloride

concentration between nine strains from 4 different isolation sources revealed that this feature can vary in a strain-dependent manner (Figure 5). The biggest difference could be seen in the final OD_{600nm} after 48 h grown at the optimal NaCl concentration. The strains TMW 2.2256, TMW 2.2260, TMW 2.2264 and DSM 20337 reached the highest final OD_{600nm} values. While the type strain DSM 20339^T, TMW 2.2257 and TMW 2.2266 reached the lowest values (Figure 5). Reaching high cell density in a short amount of time is a key feature of a starter strain, as it prevents spoilage by other organisms.

To cope with the rising levels of acids and the associated lowering of the pH, some strains of *T. halophilus* and its closely related species *T. muriaticus* have been reported to produce biogenic amines (Satomi et al., 2008). This is not true for the strains isolated from lupine moromi. None of the strains produce biogenic amines from L-tyrosine or L-histidine and are therefore not a health concern (Figure 4).

Taken together, the strain diversity of *T. halophilus* present in the lupine moromi fermentation may be influenced and shaped by the available carbohydrate sources and sodium chloride concentration. Nevertheless, it can be stated that *T. halophilus* is a ubiquitous member of the lupine moromi microbiota and thereby contributes majorly to the fate of the fermentation.

5.1.1 Genomic comparison of *T. halophilus* strains reveals no unique correlation of specific genes/pathways to the source of isolation.

Genomic analysis of the *T. halophilus* pan and core genome using the COGs categories as assigned by the eggnog-mapper revealed that most CDSs with a known function were assigned to the category “carbohydrate transport and metabolism” (Figure 7). Notably, the relative abundance of CDS assigned to the category G (carbohydrate transport and metabolism) was 2.2 times higher in the accessory genome than in the core genome. This result is in line with the knowledge that *T. halophilus* has a diverse carbohydrate metabolism (Roling and van Verseveld, 1996; Uchida, 1982). In contrast to that, the CDS that contribute to the sodium chloride tolerance of *T. halophilus* such as the ones associated with the import of choline, proline, glycine betaine or ATP-dependent export of Na+ ions *via* the V-type ATPases (categories C, E, F and U) are more abundant in the core genome (Chun et al., 2019; He et al., 2017b; Heo et al., 2019; Robert et al., 2000). These results indicate that the sodium chloride tolerance is a species-wide conserved feature with little variance, but

that the adaptation to a niche occurs *via* the carbohydrate metabolism. To investigate this variance in the genomes of 24 *T. halophilus* strains from various isolation sources, the three annotation pipelines NCBI PGAP, RAST and TIGR were used. The presence or absence of the same cluster of genes or single genes was visualized by a matrix generated with Perseus (Figure 10), *T. osmophilus* was used as an outgroup. Among all members of the *T. halophilus* species from this set the tagatose-6P pathway is conserved, while the Leloir pathway is not as conserved in the species. These results indicate a clear adaptation towards a plant-derived environment of the species *T. halophilus*. Furthermore, the higher abundance of the arabinose degrading operon among the isolates from soybean-based fermentations is in line with the fact that soybeans contain more arabinose than e.g., lupine beans (Sakamoto et al., 2013). The abundance of operons for the utilization of D-mannitol, D-sorbitol, gluconate and galactonate is the highest among the strains from lupine moromi (Figure 10). This is potentially due to the fact, that the main yeast in lupine moromi is *D. hansenii* which is known to produce a variety of sugar alcohols (Dominguez, 1997; Loman et al., 2018). Although lupine beans are *Leguminosae* and therefore can have considerable amounts of raffinose family oligosaccharides (RFOs), the abundance of enzymes such as the alpha-galactosidases associated with the degradation of RFOs cannot be linked to an environment (Cerning-Béroard and Filiatre-Verel, 1980). This is potentially due to the fact, that the moromi fermentations are a multi-species habitat and that therefore RFOs can also be utilized by other species within it (Lülf et al., 2021).

5.22 % of the accessory genome was assigned to the category E (amino acid transport and metabolism) including pathways and enzymes that are important for the survival and energy generation under acidic or saline conditions such as the ADI pathway or by *AspDT*. The ADI pathway is known to be highly upregulated in *T. halophilus* under acid and saline stress, contributing to the survival and energy generation (Lin et al., 2017; Liu et al., 2015). Only 50% of the strains from lupine moromi encode for a functional ADI pathway, the reason for that might be the different pH optima or that other mechanisms towards these acidic conditions are sufficient. The two-gene operon *AspDT* is present in two strains from lupine moromi and in multiple strains from soybean fermentations (Table 15). The L-aspartate-4-decarboxylase (AspD) and especially its adjacent aspartate: alanine antiporter AspT are well-studied in *T. halophilus* (Abe et al., 2002; Higushi et al., 1998). AspD catalyzes the decarboxylation reaction of aspartic acid to alanine and CO_2. AspT then exports alanine and imports

aspartic acid, which together generates a PMF(Abe et al., 1996). AspT (TC:2.A.81.1.1) has a higher affinity towards l-aspartate than towards l-alanine and an optimum pH of 5 (Sasahara et al., 2011). Together with the high abundance of l-aspartate in lupine beans and the fact that pH 5 is in the growth range for *T. halophilus*, the presence of these two genes could be beneficial for a strain (Jezierny et al., 2011).

The abundance of CDS assigned to the category V (defense mechanisms) was 1.9 times higher in the accessory genome than in the core genome. A major reason for this is the variance in the CRISPR/Cas regions or the absence/presence of restriction-modification systems in the genomes (Table 16). Within 12 strains including 3 from lupine moromi, a total of 9 different combinations of CAS type CDS are present. Recently, a comparative study of 18 *T. halophilus* strains determined 9 potential insertion points of CRISPR arrays using genomic alignment and the genome of NBRC 12172 as a reference (Matsutani et al., 2021). The fact, that not all members of *T. halophilus* species contain at least one type of CRISPR/Cas array increases the probability that this genomic feature is strain-dependently acquired through mobile genetic elements that harbor CRISPR/Cas sequences (Pinilla-Redondo et al., 2022). Despite the presence or absence of CRISPR/Cas sequences in a genome many of the strains in the analyzed set contain predicted prophages (Table 16). The most predicted and annotated as intact prophage sequences are found among the strains lupine moromi. Hence, it can be said that the occurrence of phages is higher in lupine moromi than in soy moromi or sauces derived from fish or that the CRISPR/Cas arrays were acquired after the insertion of the prophage into the genome.

The comparative genomic analyses of the entire set showed that the diversity of the carbohydrate metabolism enables *T. halophilus* to adapt to new environments. Further, it reveals that features important for growth and survival under saline conditions (ADI, *AgDI* and *AspDT*) can be strain dependent. Furthermore, it can be hypothesized that CRISPR/Cas systems could help *T. halophilus* when grown in lupine moromi, as the occurrence of phages and thereby prophage sequences seem to be higher in lupine-moromi compared to soy moromi or fish-derived sauces.

5.2 The species *T. halophilus* is genomically adapted towards a plant-derived environment.

Regarding the hypothesis of the metabolic adaption of *T. halophilus* towards a lupine or plant-derived medium, the transcriptomic profile of two *T. halophilus* strains was investigated in a lupine moromi model medium. This medium was composed of 20 g/L lupine peptone added to the MRS medium. The pH of the medium was adjusted to 5.7, which is the starting pH in a lupine moromi fermentation.

As a pretest for the transcriptomic experiment, the growth of both strains of *T. halophilus* in this medium (LMRS, pH 5.7) was monitored when either D-galactose, D-glucose or no additional carbon source was added (Figure 12). Interestingly, both strains displayed the highest growth with D-galactose compared to D-glucose or no additional carbon source (Figure 12). Furthermore, both strains exhibited good growth without the addition of any additional carbon source. This shows that the residual carbon sources in the peptone can be utilized by both strains. Qualitative analysis with HPAEC-PAD showed that D-glucose, D-galactose and two unidentified carbon sources were present in non-fermented lupine peptone (Figure 25). The pH during the fermentation without the addition of D-galactose and D-glucose decreases within the first 22 h, followed by an increase after 22 h (Figure 12, B). Both strains reached the lowest pH values with D-glucose. However, regardless of the addition of D-glucose or D-galactose, TMW 2.2254 reaches lower pH values compared to TMW 2.2256. This difference can be linked to the alkalization potential of the ADI pathway in TMW 2.2256 as this is inactive in TMW 2.2254 and the ADI pathway is known to be highly upregulated under high NaCl conditions in *T. halophilus* (Liu et al., 2015).

Based on the results of this pretest, the fermentation experiment for the transcriptomic profiling was done in LMRS supplemented with galactose at a pH of 5.7. To resemble the conditions in the moromi fermentation, *T. halophilus* cultures were cultivated in Erlenmeyer flasks under microaerophile non-shaking conditions at 25 °C for 10 h. Precultures were also prepared in the same medium. This experimental setup should allow a snapshot of the gene expression in an environment simulating the lupine moromi fermentation.

DISCUSSION

Several differentially expressed genes (DEGs) could be identified by comparing the transcriptomic profile at 1 h and 10 h of fermentation. The DEGs correspond to the carbohydrate metabolism followed by the central carbon metabolism, amino acid metabolism, pyrimidine synthesis, transport of ions and synthesis of compatible solutes (Table 18). The operon for the utilization of trehalose was increased in both strains after 10 h, despite no addition of trehalose to the medium. Since trehalose is a compound with osmo-protective functions and accumulates in *T. halophilus* under salt stress, these results indicate accumulation of trehalose in *T. halophilus* when grown in LMRS instead of trehalose degradation (He et al., 2017a; Roberts, 2005). This is further supported by the fact that the glycoside hydrolase family 65 protein, which showed enhanced expression in the analyzed *T. halophilus* strains, putatively encodes for a trehalose 6-phosphate phosphorylase *trePP.* The latter enzyme was described in *L. lactis* to catalyze the formation of trehalose-6-phosphate rather than its phosphorolysis (Andersson et al., 2001). Nevertheless, it must be said, that this enzyme produces trehalose 6-phosphate and it is not clear to this point whether trehalose 6-phosphate provides a similar osmo-protection compared to trehalose. Notably, the transcripts of the betaine-aldehyde dehydrogenase and an iron-containing alcohol dehydrogenase were decreased in TMW 2.2254 after 10 h of fermentation. As these two enzymes can be linked to the synthesis of betaine from choline and the accumulation of glycine betaine is favored under saline conditions, these results indicate that the availability of choline in the medium is low (Chun et al., 2019). Another important way to respond to osmotic stress is the ADI pathway. This pathway was highly increased in TMW 2.2256 LMRS after 10 h of fermentation. In comparison, transcripts of this pathway were not found in TMW 2.2254 as this pathway is inactive due to a deletion in the genome. In TMW 2.2256 this pathway potentially provides the cell with NH_3, which increases the intracellular pH (He et al., 2017a). An intermediate of this pathway, citrulline, accumulates in *T. halophilus* cells under salt stress and is linked to reduce ROS in plants (Akashi et al., 2001; Chun et al., 2019). A similar effect might be present in *T. halophilus.* However, this hypothesis lacks evidence and still needs further investigation in *T. halophilus*.

The synthesis of aromatic amino acid was predictively decreased in TMW 2.2256, due to the 3-phosphoshikimate 1-carboxyvinyltransferase (*aroA*) gene being decreased by 2-fold after 10 h of fermentation. *aroA* is a bottleneck enzyme in the shikimate pathway, catalyzing the step from shikimate 3-P to 5-O-(1-Carboxyvinyl)-3-phosphoshikimate (a

precursor of Chorismate). According to the manufacturer, the used lupine peptone is rich in free tyrosine and phenylalanine. Therefore, it can be hypothesized that synthesis of aromatic amino acids was not needed at the time of sampling.

The transcripts of the Leloir pathway (*galKETR*) and the tagatose-6-P pathway (*lacDCBAR*) were increased in both strains after 10 h of fermentation. Interestingly both strains predictively utilize galactose *via* both pathways simultaneously. Additionally, both strains encode for a non-identical version of the tagatose pathway (*lacDCBAR2*) with 67 % to 90 % amino acid sequence similarity (Figure 13). The *lacDCBAR2* operon encodes also for an additional HAD phosphatase of which the exact function is unknown. In *L. lactis*, an unidentified phosphatase enables the use of galactose-6-phosphate *via* the Leloir pathway by dephosphorylation and thereby connects the Leloir pathway and the tagatose-6P pathway on a metabolic level (Solopova et al., 2018). Although a similar connection could be present in *T. halophilus*, further experiments are needed. The genetic organization and the identity of the amino acid sequence indicate, that the *lacDCBAR2* has been duplicated or has been acquired through genetic mobile elements. *lacDCBAR2* is only present in some strains of the species *T. halophilus*, including TMW 2.2254, TMW 2.2256 and TMW 2.2264. The strain NBRC 12172 (DSM20338) has *lacDCBAR2* duplicated and thus has three copies of the tagatose-6P pathway (Unno et al., 2020). The fact that not all strains of the species encode the *lacDCBAR2* operon and that it is active in the presence of galactose strengthens the hypothesis, that this feature represents an adaptation towards a galactose-rich environment and a competitive advantage against other bacteria. Although a duplication of a gen does not necessarily double the transcript amount, the gene dosage in the cell can be influenced (Kondrashov, 2012). Another study supporting this assumption was done with *S. cerevisiae.* Growth of this species under glucose-limiting conditions lead to a significant increase in the mRNA level because of multiple tandem duplications of two genes into three chimeric transporters. The evolved strain could outperform the parental strain in the glucose uptake assays (Brown et al., 1998). In conclusion, the identified *lacDCBAR2* operon could help *T. halophilus* grow faster in a galactose environment by increasing the gene dosage and by connecting the Leloir to the tagatose-6P pathway.

DISCUSSION

StringTie analysis of the transcriptomic raw data revealed, that the increased alpha-galactosidase and sugar O-acetyltransferase adjacent to the Leloir pathway in TMW 2.2254 were not transcribed as a polycistronic mRNA together with the genes of the Leloir pathway, indicating the activation of a different regulator that is activated in LMRS. In contrast to that, the increased transcripts of the 6-phospho-beta-galactosidase (*lacG*) and adjacent transporters (*lacEF*) in TMW 2.2256 can be linked towards the activation through D-galactose. StringTie analysis of the raw data showed, that the tagatose-6P operon is transcribed together with *lacG* and *lacEF* as a polycistronic mRNA. Thus, it can be hypothesized that the transcription of the alpha-galactosidase is regulated differently and potentially by a different activator, which is present in LMRS, compared with the activation of the *lacG* and the adjacent transporters *lacEF*.

The *glpKOF* pathway of *T. halophilus* showed a stronger increase in TMW 2.2256 than in TMW 2.2254. The *glpKOF* pathway is responsible for the glycolysis-independent synthesis of glycerophospholipid sources such as DHAP and G3P (Doi, 2019). This pathway is induced in *E. faecalis* only in the presence of oxygen (Doi, 2015). Due to the fact, that the cultivation was done under aerobic conditions a similar activation can be assumed. The synthesis of other glycerophospholipids may be required as saline stress can induce changes in the membrane compositions in *T. halophilus* (He et al., 2017a).

Furthermore, despite the absence of glycerol in the LMRS medium, the operons encoding glycerol dissimilation, the dehydrogenation pathway using the glycerol dehydrogenase (*gldA*) and the *dhaMKL* operon, were increased in TMW 2.2256 after 10 hours of fermentation. This pathway is reported to be essential for the dissimilation of glycerol in LAB (Doi, 2019). In *E. faecalis* disruption of *gldA* and the absence of heme lead to the incapability of *E. faecalis* to utilize glycerol (Bizzini et al., 2010). A comparable increase of transcript corresponding to this operon could not be identified for TMW 2.2254. This might be due to the disruption and therefore inactivation of *dhaL* in TMW 2.2254. The incapability of glycerol utilization of TMW 2.2254 was also observed in the API 50 CHL test (Table 13). The utilization of glycerol *via* the dehydrogenation pathway in TMW 2.2256 might lead to an alteration of the

NADH/NAD+ balance inside the cell, due to the increased flow of NADH per mole pyruvate from this pathway.

This assumption was further supported by an observed increase (at least two-fold) of transcripts corresponding to the pyruvate dehydrogenase complex and the lactate dehydrogenase for TMW 2.2256. Indeed, no increase could be detected for TMW 2.2254 (Table 18). Additionally, part of the generated pyruvate might also be converted to phosphoenol-pyruvate to provide enough PEP for the PEP-dependent DHA kinase *dhaKL*. This assumption was supported by an increase of transcripts for the pyruvate phosphate dikinase in TMW 2.2256 after 10 h of fermentation.

Taken together, several operons and genes needed for the fermentation of mono- and di-saccharides were highly upregulated in *T. halophilus* when grown in LMRS. Transcripts of the tagatose-6P pathway and the Leloir pathway were increased in *T. halophilus* in the presence of galactose, simultaneously. This might predictively increase the uptake and utilization of galactose. Together with the strain dependent utilization of sugar alcohols such as glycerol, an increased flow of DHAP might explain the increase of the pyruvate metabolizing enzymes. The ADI pathway majorly contributes to the alkalization potential and pH homeostasis in LMRS. Thus, it can be stated that *T. halophilus* is well adapted towards an environment rich in galactose.

5.2.1 The competitiveness of *T. halophilus* strains in a lupine moromi fermentation can be linked to strain-specific features.

To investigate the competitiveness of *T. halophilus* strains in a simulated moromi fermentation, a small-scale moromi fermentation was inoculated with a set of *T. halophilus* strains. The fermentation was carried out in triplicate batches. The *T. halophilus* strain set consisted of the six strains from lupine moromi, the type strain DSM 20339^{T} and the strain TMW 2.2260 from buckwheat moromi. The strains were distinguishable by a multiplex PCR primer set based on marker ORFS identified by the BADGE program (Figure 14). This primer set enabled a unique identification of all strains in a shorter time compared to the RAPD-PCR procedure used for the initial identification (Figure 3). However, to further improve the identification, a new multiplex PCR primer set was designed using the software “Rapid identification of PCR primers for unique core sequences” (RUCS) (Thomsen et al., 2017). With this set every strain exhibited a single specific fragment length after PCR (Figure 15). This improved primer

set was then used to monitor the strain composition in the lupine moromi during the fermentation process (Figure 16).

The strain compositional changes were comparable across all fermentations (replicates). TMW 2.2264 and TMW 2.2254 dominated the microbiota after 14 days. Both strains reached cell counts above 10^7 per ml, among the highest across the strains (Table 19). The CFU/ml for the other six strains decreased over time while the decrease of the cell counts for the strains TMW 2.2264 and TMW 2.2254 was not as steep. It can be hypothesized that the rapid decrease of the pH (5.7 to 4.9 within the first week) led to acid stress for all *T. halophilus* strains. Thus, only strains with a higher pH tolerance were able to survive or even grow, resulting in the observed cell counts per ml. One of the major pathways to cope with acid or saline stress in *T. halophilus* is the ADI pathway (Liu et al., 2015). This pathway was also highly increased in TMW 2.2256 in the LMRS lupine moromi simulation medium, further supporting this assumption (Table 18).

Since lupine beans are rich in RFOs, the presence of an alpha-galactosidase (α-gal) might also be beneficial (Cerning-Béroard and Filiatre-Verel, 1980). This hypothesis is further supported by the fact, that out of all strains, TMW 2.2254 and TMW 2.2264 are the only ones that encoded for at least one copy of an α-gal. Additionally, this gene also exhibited an enhanced expression during the transcriptomic experiment (Table 18). TMW 2.2264 also encodes for a second α-gal (α-gal2) (HXW82_03205) that can only be found in DSM23766^T (originating from degraded sugar beet) but in none of the other analyzed strains. This α-gal2 only has 43% amino acid identity over the entire sequence compared to the original α-gal gene. Thus, it is unlikely that α-gal2 originated from a duplication event. However, to clarify the relevance of α-gal2, more biochemical assays are needed e.g., to determine the substrate spectrum or to compare the activity on RFOs with the α-gal gene which is more commonly found in *T. halophilus*.

Furthermore, both dominating strains TMW 2.2254 and TMW 2.2264 harbor the *lacDCBAR2* operon that was reported as active in the LMRS transcriptomic experiment (Table 18). However, strains missing this operon also grew or survived within 21 days of the competitive fermentation. Thus, the impact of this operon might be not as high compared to the ability to cope with the increased acid stress.

In conclusion, a multiplex PCR system was developed, that allowed monitoring of the strain compositional changes for *T. halophilus* in a small-scale lupine moromi fermentation. Furthermore, observed growth and survival of specific strains could be explained by resistance to acid stress which is mainly due to the presence or absence of the ADI pathway. Adaptation to specific carbon sources present in lupine beans also contributes to the competitiveness of the strains. This was highlighted by the presence and activity of at least one alpha-galactosidase in the genome of dominant strains.

5.3 The mucoid phenotype of TMW 2.2256

During the experiments performed in this thesis, mucoid colonies were detected belonging to *T. halophilus* TMW 2.2256 which was grown on MRS-agar supplied with D-galactose or D-lactose (Figure 11, A-B). At the start of this thesis, such behavior was not described for *T. halophilus*, consequently, experiments to investigate this mucoid phenotype were conducted. The colonies of this strain exhibited the strongest visual phenotype when cultivated in a regular MRS medium supplied with 13.5% NaCl (w/v) and 10 g/L D-lactose. Two other analyzed representative strains of *T. halophilus* also exhibited an increase of this phenotype when grown on the same agar plates. However, the phenotype of the strain TMW 2.2256 was the strongest. These results raise the question of whether *T. halophilus* is capable of producing exopolysaccharides (EPS). Furthermore, it was unclear which pathway may be associated with the synthesis of these EPS and how it is organized in the genome.

EPS-forming lactic acid bacteria are widely known and are often categorized into Mucoid or ropy phenotypes. The culture media and carbon sources required for these phenotypes are strain dependent. However, in most cases when a disaccharide such as D-lactose is used, a heteropolysaccharide (HePS) is produced (Prete et al., 2021). Due to the fact, that TMW 2.2256 exhibited the strongest phenotype in the presence of D-lactose, it is likely that a HePS is formed. In general, exopolysaccharides can be divided into homopolysaccharides (HoPS) and heteropolysaccharides (HePS). The major difference between both polysaccharides is, that HoPS consists of repeating units of the same sugar monomer whereas HePS can be made up of multiple different sugar monomers. Another difference is that HoPS are generally synthesized extracellularly whereas HePS are synthesized intracellularly e.g. *via* the wzx/wzy-dependent pathway (De Vuyst and Degeest, 1999; Monsan et al., 2001; Zeidan et al.,

2017). Recently, two publications reported the formation of HePS by *T. halophilus* (strain CGMCC 3792 and SNTH-8). The strain SNTH-8 produced two HePS, THPS-1 and THPS-2, both consisting of arabinose, xylose, fucose, galactose, glucose and glucuronic acid but in different molar ratios. Similarly, two fractions of EPS, EPS-1 and EPS-2, were obtained from CGMCC 3792. EPS-1 consisted of galactose, mannose and glucuronic acid and EPS-2 consisted of glucose and mannose with traces of rhamnose and galactose. Both EPS-1 and EPS-2 exhibited antioxidant activities and potentially protect from cryo- or lyophilization damage (Yang et al., 2022; Zhang et al., 2022).

However, despite the isolation and the characterization of these exopolysaccharides little is known about the genetic background of the responsible pathways and enzymes in *T. halophilus*. Analysis of the genome of TMW 2.2256 revealed a cluster of 18 Kb in size, encoding for seventeen CDSs (Figure 21).

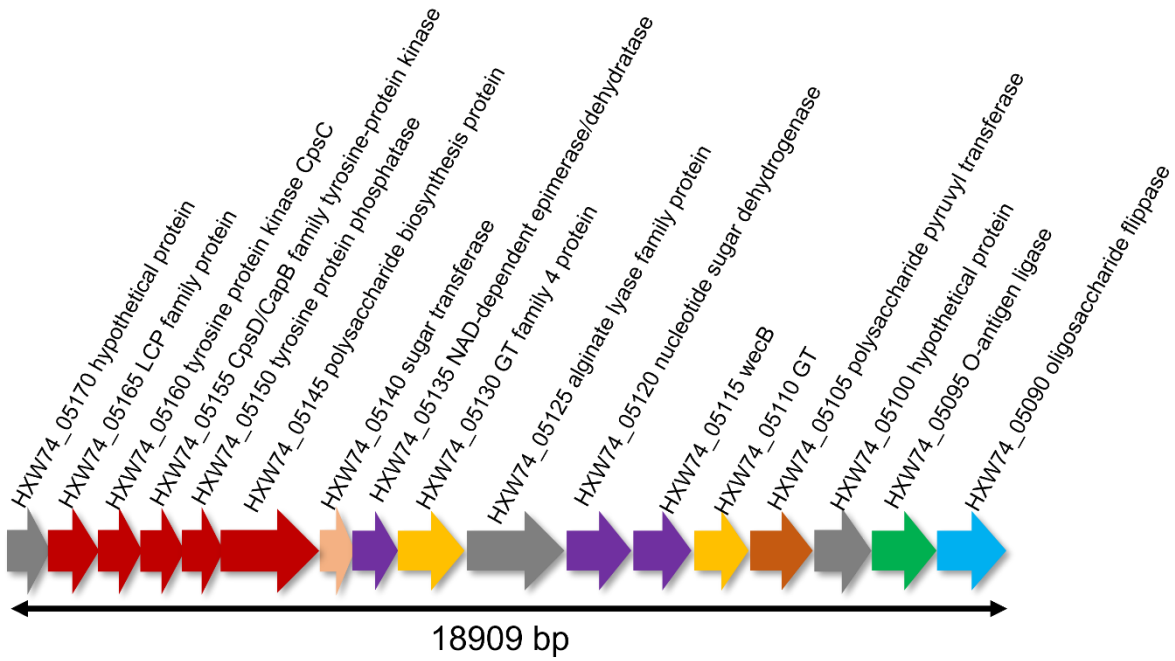

Figure 21: Organization of the predicted wzx/wzy-dependent pathway of *T. halophilus* TMW 2.2256

The first CDS, HXW74_05170 is a hypothetical protein with an unknown function. HXW74_05165 encodes for an LCP (lyTR-Cps2A-psr) protein. LCP proteins are associated with the transfer of polysaccharides or teichoic acids to peptidoglycan (Kawai et al., 2011). The CDS HXW74_05160, HXW74_05155 and HXW74_05150 encode for homologs to the *epsCDB* (wzd, wze and wzh) with an identity between 85

% and 98 % to the respective amino acid sequence of the recently described *capACDB* operon of *T. halophilus* YA5 (Wakinaka et al., 2023). The sequences of TMW 2.2256 have an identity of 32 % and 43% to the respective proteins in *S. pneumoniae* TIGR4 and *Lactiplantibacillus plantarum* TMW 1.1478. In TMW 1.1478 *epsCDB* (wzd, wze and wzh) are part of the phosphoregulatory system of the wzx/wzy-dependent pathway. All of these results confirm that the amino acid sequences of HXW74_05165 to HXW74_05150 belong to the wzx/wzy-dependent pathway of *T. halophilus*. The phosphoregulatory parts are important for the synthesis of polysaccharides and determine the chain length of the polysaccharides (Zeidan et al., 2017). The two CDSs HXW74_05095 and HXW74_05090 are located at the end of the cluster in TMW 2.2256 and are annotated as O-antigen ligase and oligosaccharide flippase. The conserved domains of HXW74_05090 reveal that this is a putative wzx flippase responsible for the export of the polysaccharides.

A more detailed analysis was done for HXW74_05095, which predictively encodes for a wzy polymerase. The topology of the O-antigen ligase (HXW74_05095) revealed 13 transmembrane domains and a similar arrangement to the wzy polymerases from *S. thermophilus*, *L. plantarum* and *S. pneumoniae* (Figure 22, A-D).

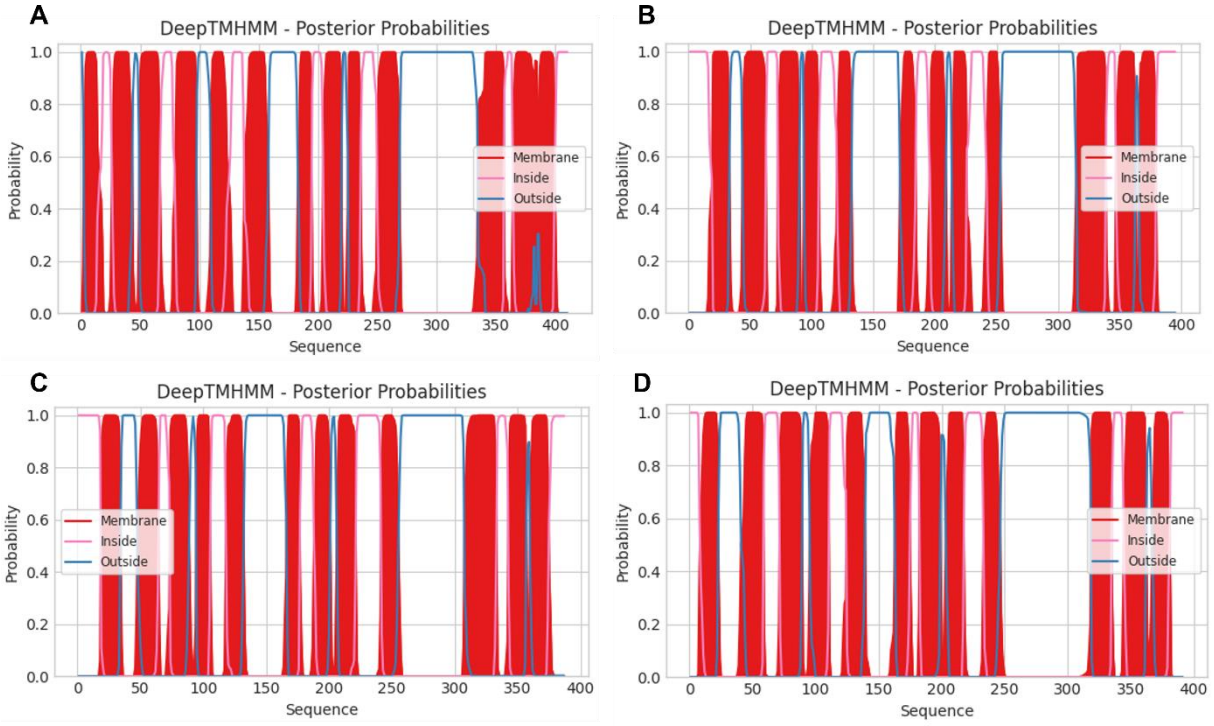

Figure 22: Comparison of the predicted transmembrane domains from *T. halophilus* TMW 2.2256 with other known wzy polymerases. The prediction was done using the web version of DeepTMHMM (version 1.0.20) (Hallgren et al., 2022)**. A) Predicted transmembrane domain of the putative wzy polymerase of TMW 2.2256 (locus HXW74_05095). B) Predicted transmembrane domain of the polysaccharide polymerase of the wzy pathway of *S. thermophilus* S-3 (NC_008532) (locus STER_RS05265). C) Predicted transmembrane domain of the polysaccharide polymerase cps1I of the wzy pathway of *L. plantarum* WCFS1 S-3 (NC_004567) (locus**

LP_RS05055). D) Predicted transmembrane domain of the polysaccharide polymerase of the wzy pathway of *S. pneumoniae* serotype 11a (CR931653) (locus SPC11A_0013).

The similarity of the predicted structure (Figure 22) together with the fact that 9-14 Tm domains are typical for wzy polymerases, strengthens the hypothesis that HXW74_05095 is a wzy polymerase homolog (Jolly and Stingele, 2001).

The coding sequences HXW74_05140 to HXW74_05105 predictively encode for different glycosyltransferases, dehydrogenases or epimerases. The predictions of the specificity of these enzymes using bioinformatic tools can be vague and therefore more experimental data are needed to further determine the function of these CDS.

Taken together, the 18 kb operon from TMW 2.2256 most likely encodes for the wzx/wzy-dependent pathway in *T. halophilus*. This pathway is a highly variable region within the species *T. halophilus*, with low GC content and many different glycosyltransferases and dehydrogenases. Recently, the wzx/wzy-dependent pathway has been implicated in the production of CPSs in *T. halophilus*. The variability of this region was linked to the defense against bacteriophages, as it was shown that *T. halophilus* CPS plays a role in the susceptibility to bacteriophages (Wakinaka et al., 2023). Although it has been shown that the wzx/wzy-dependent pathway of *T. halophilus* can produce a CPS, it remains unclear whether it could also be responsible for the formation of an EPS such as that produced by *T. halophilus* strains CGMCC 3792 and SNTH-8.

5.4 The diploidy of the *D. hansenii* strain TMW 3.1188 represents an adaptation to lupine moromi.

D. hansenii was highly abundant in the sampled lupine moromi's and a total of 25 yeast isolates were identified by MALDI-TOF MS (Lülf et al., 2021). Yeasts can majorly influence the quality of a fermentation product by the formation of a variety of aroma-contributing compounds (Song et al., 2015b). The 25 *D. hansenii* isolates originated from six different moromi fermentations with sodium chloride concentrations ranging from 10 – 20 % (w/v) (Table 21). To determine the biodiversity among the strains, gDNA was isolated and used for genotyping. RAPD-PCR with two different M13 primers was performed to differentiate isolates based on their unique fingerprint. This M13 primer was previously used to differentiate *D. hansenii* strains in other studies (Andrade et al., 2006; Huey and Hall, 1989). The resulting fingerprint of this study allowed the differentiation on species and strain level. The resulting fingerprint of the RAPD-PCR for the *D. hansenii* isolates from lupine moromi allowed the differentiation from isolates from fermented sausages and lupine moromi (Figure 17 and Figure 18). Interestingly, the fingerprint was identical for all isolates from lupine moromi. These results indicate that all isolates from lupine moromi represent the same strain. Thus, it can be said that the lupine moromi fermentations are dominated by one yeast strain of *D. hansenii,* regardless of the sodium chloride concentration.

To gain further insight into the genomic nature of this dominant *D. hansenii* strain (TMW 3.1188), the genome was fully sequenced using Pacbio. The data from partial sequencing of the 5.8S RNA, ITS2 and parts of the LSU sequencing confirmed the affiliation with the species *D. hansenii*. Phylogenetic analysis using kr distance and ANIb values against a set of completely sequenced strains reveals that the strain is closely related to the type strain CBS767^T (Table 20). Furthermore, the assembled genome of TMW 3.1188 was compared to the genome of the type strain CBS767^T. This comparison revealed that the genome of TMW 3.1188 was almost twice as big. Furthermore, at least two contigs of TMW 3.1188 mapped to every chromosome of CBS767^T, one with a higher identity and one with a lower identity (Figure 19). Annotation of this genome was done using the YGAP pipeline. The annotation yielded almost double the amount of CDS, compared with the CDS found in CBS767^T, further strengthening the hypothesis of genome duplication.

To answer the arising question, of whether the duplication of the genome of TMW3.1188 originated from another closely related *Debaryomyces* species or from *D.*

hansenii, a bioinformatic approach to differentiate the diverse species *D. hansenii* on strain level was chosen. Typically, the ITS1/5.8SRNA/ITS2 region is used for the phylogenetic analysis of yeast strains. However, in the case of the species *D. hansenii,* it was known that this region does exhibit not enough diversity to differentiate between strains consistently. Other studies described the *ACT1* gene as a good alternative for strain differentiation (Groenewald et al., 2008). Both versions of the *ACT1* gene in the genome of TMW 3.1188 were used for a phylogenetic analysis. Based on this *ACT1* gene locus, it can be hypothesized that the duplication of the genome of TMW3.1188 originated from the species *D. hansenii*, rather than another species (Figure 20). This finding suggests an intraspecies hybrid strain. Similar hybrids were reported for *Zygosaccharomyces rouxii* (James et al., 2005; Sato et al., 2017). In *D. hansenii* the tolerance to high salinity is associated with the intracellular accumulation of glycerol, mutants deficient in glycerol synthesis have reduced tolerance to salinity (Andre et al., 1988). Since *D. hansenii* TMW 3.1188 is a hybrid strain potentially encoding for two intact glycerol production pathways, we hypothesize that this strain has an increased tolerance to high salinity. Similarly, the glycerol production and the tolerance to sodium chloride were increased in the natural hybrid *Z. rouxii* ATCC 42981 compared to the wild-type strain ATCC2623 (Pribylova et al., 2007). Even though yeast hybrid strains have been known for a long time and early studies of *D. hansenii* suggested a chromosomal polymorphism, only recently has the genome of a *D. hansenii* hybrid isolated from ‘Gangjang’ been completely sequenced (Corredor et al., 2003; Jeong et al., 2022). Taken together, this study provides strong evidence for the presence of an intraspecies hybrid strain of *D. hansenii*.

Natural yeast hybrids occur in many man-made environments such as lager beer or wine and such hybridization events can contribute to the adaptation to these artificial environments (Gabaldón, 2020). Thus, this hybridization event might be an adaptation of *D. hansenii* to the artificial environment which is the lupine-based moromi. However, the consequences of this hybridization event are not known yet.

In conclusion, the isolated and sequenced strain TMW 3.1188 represents a strain that dominates the lupine moromi fermentation, thus contributing to the fermentation characteristics. The dominance during lupine fermentation might be explained by the genome duplication of the strain, providing metabolic advantages compared to other *D. hansenii* strains.

6 Summary

Fermentation of food has a long tradition in the history of mankind. Over the centuries, many fermented foods have become icons of a particular region and part of the culture of the people who produced them. A well-recognized product for the Asian region is the traditionally brewed soy sauce. Recently, a similar bean-based seasoning sauce was launched in the form of lupine sauce. Although the general concepts and steps are almost identical to those from the soy sauce brewing process, this lupine sauce is fundamentally different. Initial studies of the microbial consortium of this lupine moromi revealed that it was dominated by *T. halophilus* and *D. hansenii*.

In this thesis, the genomic and phylogenetic diversity of *T. halophilus* isolated from lupine moromi was investigated and compared with strains from soy sauce, fish sauce and cheese rind. Therefore, 58 isolates of *T. halophilus* from different time points and lupine moromi were analyzed by comparing RAPD fingerprints. This analysis revealed the presence of multiple strains within a fermentation. Whole-genome sequencing followed by comparative analysis of the strains exhibiting unique RAPD clusters led to the identification of six unique strains isolated from lupine moromi. Comparison of the genomes of these six strains with the genomes of strains from other isolation sources revealed that the species *T. halophilus* can be grouped into three lineages. The first lineage included strains from many different isolation sources including lupine moromi. The second lineage consisted of strains originating from soy or fish sauces. The third lineage consisted only of the type strain of the subspecies *T. halophilus* subsp. *flandriensis*. Further genomic analysis of the lineages did not reveal genes that could be linked to an isolation source. However, it can be noted that the abundance of the arabinose degrading operons was higher in the strains from soy sauce, compared to the strains from any other isolation source. Further analysis showed that the occurrence of prophage sequences in the genomes of strains from lupine moromi is higher compared to the occurrence in strains from other environments. This led to the assumption that more phages might be present in the lupine moromi compared to other habitats. Analysis of the occurrence and the types of CRISPR/Cas systems revealed that only some strains encode for a CRISPR/Cas system and that a wide range of different types occurs in *T. halophilus*. Notably all strains from lupine moromi either do not encode for a CRISPR/Cas system or a different type. The strains from fish sauce

have a similar diversity. The strains based on soy products (soy sauce and soy paste) have a combined number of four different CRISPR/Cas types. Concluding, genetic diversity within *T. halophilus* strains isolated from lupine moromi could be demonstrated. However, no specific genes could be linked to the source of isolation.

In this thesis, the analysis of the transcriptomic profile of *T. halophilus* TMW 2.2254 and TMW 2.2256 grown in LMRS was performed. This experiment was done to identify beneficial genes for growth in lupine moromi and link metabolic properties with the regulation of genes. Therefore, samples were taken after 1 h and 10 h of fermentation. Comparative transcriptomic analysis revealed an increase in the expression of operons and genes associated with the degradation of galactose after 10 h. This could be traced back to the high galactose content of the fermentation media. Additionally, simultaneous usage of the tagatose-6P and the Leloir pathway in *T. halophilus* could be predicted from the transcriptomic data. This simultaneous action enables a high flux of galactose into the cell. This explains the enhanced growth of *T. halophilus* with galactose compared to glucose. Genomic analysis also revealed that a second version of the tagatose-6P operon is present in most strains of the species. This second version is also highly upregulated in LMRS after 10 h, further contributing to the influx of galactose. Strain-dependent utilization of sugar alcohols and the resulting increased flow of DHAP from the high influx of galactose might also explain the observed increase of pyruvate metabolizing transcripts after 10 h. Furthermore, an enhanced expression of the genes of the ADI pathway could be observed in TMW 2.2256, which is consistent with the observed increase in the pH during fermentation. Taking together, several key pathways and enzymes could be identified that are beneficial for the growth of *T. halophilus* in lupine moromi.

Another part of this thesis aimed to analyze the competitiveness of *T. halophilus* strains in a lupine-based habitat as well as to identify specific genes contributing to the competitiveness. Therefore, different strains of *T. halophilus* (isolated from fish, buckwheat moromi and lupine moromi) were used to inoculate a small-scale lupine moromi. To identify each strain, a multiplex PCR primer set was developed. After 14 days, the strains TMW 2.2254 (from lupine moromi) and TMW 2.2264 (from lupine moromi) became the dominant strains in all fermentations. Interestingly, both strains do encode for an alpha-galactosidase (whereas the other strains do not). In the case

of TMW 2.2264, even two non-identical versions of this enzyme can be found. Legumes such as lupin beans are rich in RFOs which can be degraded by alpha galactosidases. Thus, TMW 2.2254 and TMW 2.2264 seem to have an advantage over the other strains of the tested set. This supports the hypothesis that specific strains of *T. halophilus* can use the RFOs in lupin beans as a carbon source and thereby reduce the amount of RFOs in the product. In conclusion, these results (together with the fact that in LMRS the galactose degrading operons are highly upregulated) support the hypothesis that *T. halophilus* strains from lupine moromi encode for adapted traits and have increased competitiveness in lupine moromi fermentation.

In this thesis, a mucoid phenotype of *T. halophilus* could be identified during regular cultivation experiments. Colonies of this strain (TMW 2.2256) displayed a mucoid phenotype, which was (up to the date of discovery) not known for this species. In order to explain this slime formation, a genomic analysis for operons and putative genes associated with the formation of the EPS/CPS was conducted. Genomic analysis revealed the presence of a functional and complete wzx/wzy-dependent pathway of *T. halophilus*. The gene organization in *T. halophilus* TMW 2.2256 was comparable to other EPS/CPS operons in EPS-producing LABs. However, it remains unclear whether EPS or CPS is produced by *T. halophilus*. Furthermore, other *T. halophilus* strains with a weaker phenotype also encode for this pathway. Thus, no clear linkage between the wzx/wzy-dependent pathway and the observed slimy phenotype could be done.

Previous studies have shown that the lupine moromi fermentation is dominated by *D. hansenii*. Yeasts belonging to the species *D. hansenii* could be isolated multiple times from lupine moromi. Genotyping experiments of these isolates using RAPD fingerprints were done in this thesis. This revealed that only a single strain of *D. hansenii* was present in all fermentations. Thus, as a major key player in the fermentation process, this strain TMW 3.1188 was completely sequenced by Pacbio in this thesis. The sequencing data showed, that the genome of *D. hansenii* TMW 3.1188 is almost twice the size compared to the genome of the type strain CBS 767^{T}. Further conducted phylogenetic analysis using the *ACT1* gene revealed that TMW 3.1188 is a hybrid strain from two *D. hansenii* strains. This proves the hypothesis, that the yeast strain TMW 3.1188 dominating the lupine moromi is genomically distinct from other strains of the species.

7 Zusammenfassung

Fermentierte Lebensmittel sind seit Jahrhunderten Teil der menschlichen Nahrung. Einige Lebensmittel werden aufgrund dieser Tatsache eng mit der Kultur der Ursprungsländer verknüpft. Ein gutes Beispiel dafür ist die traditional gebraute Sojasoße. Basierend auf dem traditionellen Herstellungsverfahren wird seit kurzem eine auf Lupinenbohnen basierende Gewürzsoße hergestellt. Obwohl die generelle Herstellung und die Produktionsschritte sehr ähnlich sind, ist die Fermentation aufgrund des unterschiedlichen Ausgangsmaterials klar unterscheidbar. Erste Untersuchungen der auf Lupinenbohnen basierenden Moromifermentationen wiesen darauf hin, dass *T. halophilus* und *D. hansenii* in großer Anzahl vorhanden sind.

In dieser Arbeit wurden 58 *T. halophilus* Isolate aus verschiedenen Lupinenmoromi Ansätzen mittels RAPD-Fingerprinting gescreent. Es zeigte sich, dass in den Fermentationen mehrere *T. halophilus* Stämme präsent sind. Eine Sequenzierung des kompletten Genoms von den Isolaten mit einem einzigartigen RAPD-Muster führte zu der Identifikation von sechs unterschiedlichen *T. halophilus* Stämmen. Vergleichende Analysen dieser sechs *T. halophilus* Stämme mit den Genomsequenzen von andere *T. halophilus* Stämmen aus anderen Habitaten, ermöglichte die Gruppierung der Spezies *T. halophilus* in drei Linien. Die erste und größte Line beinhaltet Stämme aus Lupinenmoromi, Fischsaucen, Sojasoßen, Buchweizen moromi und Käserinde. Die zweite Linie beinhaltet viele Stämme, welche aus Sojasoßen oder Fischsoßen Fermentationen isoliert wurden. Die dritte Linie besteht nur aus dem Typ Stamm von *T. halophilus* subsp. *flandriensis*. Die in dieser Arbeit durchgeführten vergleichenden Analysen der genomischen Ausstattung des Testsets ermöglichten keine eindeutige Identifikation von einzelnen Genen, welche direkt auf die Herkunft der Stämme schließen lässt. Dennoch ist die Abundanz des L-Arabinose abbauenden Operons in Stämmen, welche aus Sojasoßen Fermentationen isoliert wurden, höher im Vergleich mit der Abundanz dieses Operons in den Stämmen aus allen anderen Isolationsorten.

Weitere genomische Analysen mittels einer Phagen- und Viren-Datenbank zeigten, dass in der Genomsequenz von Stämmen aus Lupinenmoromi öfter Bakteriophagen Sequenzen vorhanden sind. Basierend auf der Annahme, dass in Lupinenmoromi Bakteriophagen ein Problem darstellen könnten, wurden alle verfügbaren Genome von *T. halophilus* auf die Anwesenheit und den Typus des CRISPR/Cas Systems

untersucht. Innerhalb der Spezies *T. halophilus* konnten nicht nur ein- sondern mehrere CRISPR/Cas Typen gefunden werden, welche vermutlich auch miteinander kombinierbar sind. Interessanterweise konnten für die *T. halophilus* Stämme aus Lupinenmoromi drei verschiedene CRISPR/Cas Typen identifiziert werden. Drei Stämme aus Lupinenmoromi besaßen kein CRISPR/Cas System. Für die *T. halophilus* Isolate aus fermentierten Fischsoßen konnten ebenfalls drei verschiedene CRISPR/Cas Typen gefunden werden. Für die Isolate aus Sojasoßen Fermentationen (inklusive Doenjang Isolat) konnten in Summe vier verschiedene CRISPR/Cas Typen identifiziert werden. Schlussendlich lässt sich kein Zusammenhang zwischen dem Vorhandensein eines funktionalen CRISPR/Cas Systems, Isolationsortes und der Menge an Bakteriophagen im Genom ziehen.

Um ein besseres Verständnis der Lebensweise von *T. halophilus* in dem Habitat Lupinenmoromi zu erlangen, wurde eine Analyse des Stoffwechsels von zwei Vertretern der sechs Lupinen Stämme durchgeführt. Hierfür wurde das Wachstum der Stämme TMW 2.2254 und TMW 2.2256 in dem Lupinenmodell medium „LMRS“ untersucht. Ein deutlich verbessertes Wachstum zeigte sich für beide Stämme, wenn Galaktose anstelle von Glukose als extra Kohlenstoffquelle in das Medium hinzugegeben wurde.

Aufgrund dieser Tatsache und der Tatsache, dass Lupinenbohnen reich an Galaktose sind, wurde die Stoffwechselanalyse beider Stämme in einem mit Galaktose angereichtem LMRS durchgeführt. Beide Stämme wurden nach 1 Stunde und 10 Stunden Kultivierung im LMRS beprobt und ihr Transkriptom analysiert. Nach 10 Stunden Kultivierung konnten für beide Stämme vermehrt Transkripte von Genen für den Abbau und die Verwertung von Galaktose identifiziert werden. Dabei war auffällig, dass beide Stämme den Leloir-Weg und den Tagatose-6-P Weg simultan benutzen. Eine zweite Kopie des Tagatose-6-P-Operons konnte ebenfalls in beiden Stämmen identifiziert werden. Die simultane Aufnahme und Verstoffwechselung von Galaktose durch mehrere Stoffwechselwege erklärt das erhöhte Wachstum mit Galaktose (im Vergleich mit Glukose) der beiden *T. halophilus* Stämme. Die Transkriptom-Analyse zeigte ebenfalls eine erhöhte Anzahl an Transkripten welche für Pyruvat metabolisierenden Gene kodieren. Dies konnte mit der unterschiedlichen Nutzung von Zuckeralkoholen und dem daraus resultierenden erhöhtem Vorhandensein von DHAP begründet werden. Der erhöhte Galaktose-Abbau über den Tagatose-6-P Weg verstärkte diesen Effekt noch.

Ebenso konnten Gene, welche für den Arginin Deiminase Abbauweg kodieren, in den Proben von TMW 2.2256 nach 10 Stunden Fermentation in großer Zahl identifiziert werden. Der Arginine Deiminase Abbauweg ist nicht im Genom von TMW 2.2254 kodiert. Da dieser Stoffwechselweg zu basischen Metaboliten (Harnstoff, NH_3) führt, lässt sich somit auch der beobachtete Anstieg des pH-Wertes bei der Fermentation von TMW 2.2256 erklären. Dementsprechend wurde kein Anstieg des pH-Wertes für die Fermentation von TMW 2.2254 beobachtet.

Ein weiterer Teil dieser Arbeit befasst sich mit der Durchsetzung von *T. halophilus* Stämmen unter möglichst realitätsnahen Bedingungen. Hierfür wurden Lupinenmoromi Fermentationen in kleinem Maßstab mit einer Mischung von *T. halophilus* Stämmen aus verschiedensten Isolationsorten (Fisch Fermentation, Buchweizenmoromi und Lupinenmoromi) beimpft und fermentiert. Um die Identifikation der Stämme zu gewährleisten, wurde ein Multiplex-fähiges PCR-Primer Set designt, bei dem jeder Stamm klar mit einer Bande identifiziert werden konnte. Nach 14 Tagen dominierten die Stämme TMW 2.2254 (Lupinenmoromi) und TMW 2.2264 (Lupinenmoromi) die Fermentation. Interessanterweise sind beide Stämme die einzigen des getesteten Sets, welche für mindestens eine Kopie (im Fall von TMW 2.2264 sogar zwei Kopien) einer Alpha-Galaktosidase kodieren. Somit scheint die Verstoffwechselung dieser Zucker in dem RFO reichen Substrat Lupinenbohne einen Vorteil für die Durchsetzungsfähigkeit darzustellen. Die in dieser Arbeit gewonnenen Ergebnisse zeigen, dass sich *T. halophilus* an eine Umgebung pflanzlichen Ursprungs adaptieren kann und das Vorhandensein bestimmter Gene, wie z. B. die Alpha-Galaktosidase einen Vorteil für das Wachstum in Lupinenmoromi darstellen.

Eine weitere interessante Entdeckung dieser Arbeit war die Identifizierung eines mukoiden Phänotyps von *T. halophilus* TMW 2.2256. Kolonien des *T. halophilus* Stammes zeigten eine auffällige Schleimbildung. Dies war bis zum Zeitpunkt der Entdeckung kein beschriebenes Merkmal dieser Spezies. Um diesen Phänotyp zu erklären, wurde das Genom von TMW 2.2256 auf das Vorhandensein von Genen, welche mit EPS- oder CPS-Bildung in Verbindung stehen, durchsucht. Die Analysen deuteten auf den wzx/wzy-abhängigen Polysaccharid Synthese Weg. Dieser ist in

ZUSAMMENFASSUNG

TMW 2.2256 in einem 18 Kb großen Operon organisiert und kodiert alle nötigen Komponenten, um ein Polysaccharid mit diesem Weg zu synthetisieren. Die Organisation dieses Operons in *T. halophilus* ist ähnlich zu der Organisation in anderen schleimbildenden LABs. Allerdings ist unklar ob in *T. halophilus* ein EPS oder ein CPS gebildet wird. Weiterhin besitzen auch andere *T. halophilus* Stämme mit weniger Schleimbildung den wzx/wzy-abhängigen Polysaccharid Synthese Weg, wodurch keine eindeutige Zuordnung des Synthesewegs zu dem beobachteten Phänotyp möglich ist.

Vorangegangene Arbeiten zu dieser Dissertation zeigten, dass *D. hansenii* in allen Lupinenmoromi Fermentationen in hoher Anzahl vorkommt. Es konnten viele Isolate gesammelt werden. Im Zuge dieser Dissertation wurden die Isolate dann weiter analysiert. Mittels RAPD-PCR wurde dann versucht Stammunterschiede zwischen den Isolaten aus Lupinenmoromi sichtbar zu machen. Die Isolate von Lupinenmoromi konnten klar von den Isolaten aus fermentierten Würsten unterschieden werden. Allerdings zeigt alle Isolate aus Lupinenmoromi den gleichen DNA-Fingerabdruck. Dieses Ergebnis bedeutete, dass alle Isolate aus Lupinenmoromi auf ein und denselben Stamm zurückzuführen sind, TMW 3.1188. Als Keyplayer der Lupinenmoromi Fermentation wurde daher dieser Stamm mittels Pacbio vollständig sequenziert. Dabei zeigten sich Auffälligkeiten in der Größe des Genoms. Das Genom war doppelt so groß wie das des Referenzstammes CBS767^{T}. Phylogenetische Analysen des *ACT1* ORFs von nah verwandten *D. hansenii* Stämmen zeigten, dass es sich bei TMW 3.1188 um eine Hybridhefe handelt. Somit konnte eindeutig nachgewiesen werden, dass sich die aus Lupinenmoromi Fermentation isolierte Hefe TMW 3.1188 genomisch von anderen Stämmen derselben Spezies (mit anderen Isolationsort) unterscheidet.

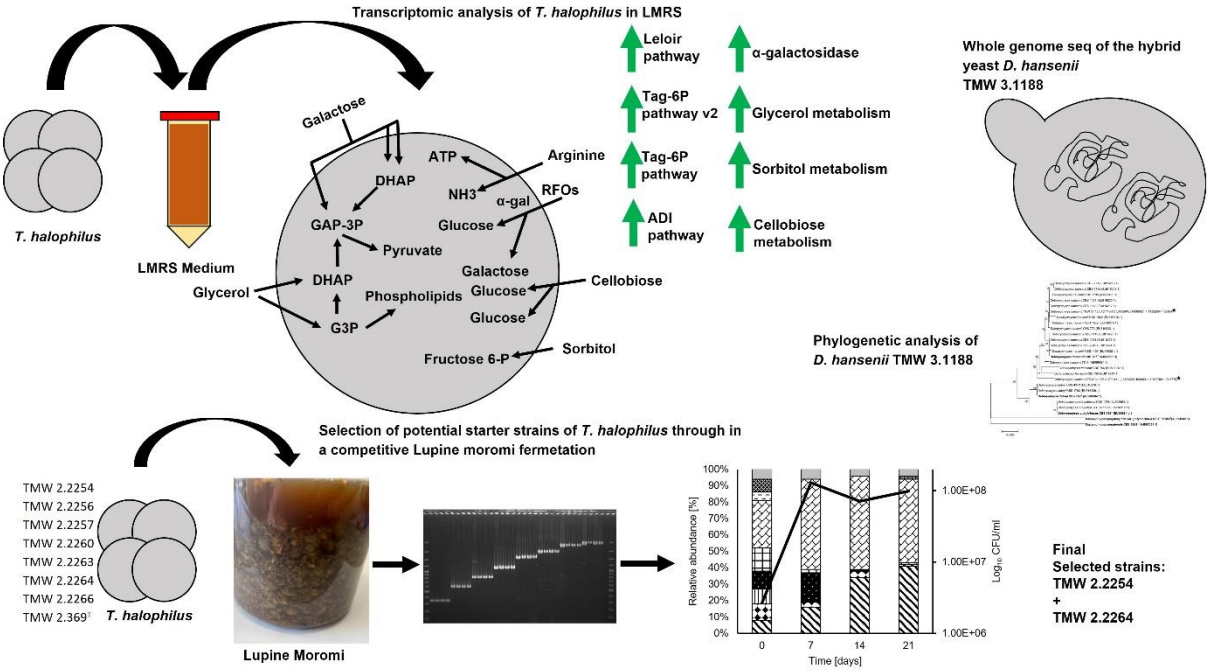

Figure 23: Graphical abstract of this dissertation with the major findings.

8 References

Abdalla, A.K., Ayyash, M.M., Olaimat, A.N., Osaili, T.M., Al-Nabulsi, A.A., Shah, N.P., Holley, R., 2021. Exopolysaccharides as Antimicrobial Agents: Mechanism and Spectrum of Activity. Front. Microbiol. 12. https://doi.org/10.3389/fmicb.2021.664395

Abe, K., Hayashi, H., Malone, P.C., 1996. Exchange of aspartate and alanine: Mechanism for development of a proton-motive force in bacteria. J. Biol. Chem. 271, 3079–3084. https://doi.org/10.1074/jbc.271.6.3079

Abe, K., Ohnishi, F., Yagi, K., Nakajima, T., Higuchi, T., Sano, M., Machida, M., Sarker, R.I., Maloney, P.C., 2002. Plasmid-encoded *asp* operon confers a proton motive metabolic cycle catalyzed by an aspartate-alanine exchange reaction. J. Bacteriol. 184, 2906–2913. https://doi.org/10.1128/JB.184.11.2906-2913.2002

Akashi, K., Miyake, C., Yokota, A., 2001. Citrulline, a novel compatible solute in drought-tolerant wild watermelon leaves, is an efficient hydroxyl radical scavenger. FEBS Lett. 508, 438–442. https://doi.org/10.1016/S0014-5793(01)03123-4

Allison, G.E., Klaenhammer, T.R., 1998. Phage resistance mechanisms in lactic acid bacteria. Int. Dairy J. 8, 207–226. https://doi.org/10.1016/S0958-6946(98)00043-0

Allwood, J.G., Wakeling, L.T., Bean, D.C., 2021. Fermentation and the microbial community of Japanese koji and miso: A review. J. Food Sci. 86, 2194–2207. https://doi.org/10.1111/1750-3841.15773

Andersson, U., Levander, F., Rådström, P., 2001. Trehalose-6-phosphate Phosphorylase Is Part of a Novel Metabolic Pathway for Trehalose Utilization in *Lactococcus lactis*. J. Biol. Chem. 276, 42707–42713. https://doi.org/10.1074/jbc.M108279200

Andrade, M.J., Rodríguez, M., Sánchez, B., Aranda, E., Córdoba, J.J., 2006. DNA typing methods for differentiation of yeasts related to dry-cured meat products. Int. J. Food Microbiol. 107, 48–58. https://doi.org/10.1016/j.ijfoodmicro.2005.08.011

Andre, L., Nilsson, A., Adler, L., 1988. The Role of Glycerol in Osmotolerance of the Yeast Debaryomyces hansenii. Microbiology 134, 669–677. https://doi.org/10.1099/00221287-134-3-669

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44, W16–W21. https://doi.org/10.1093/nar/gkw387

Arranz-Otaegui, A., Carretero, L.G., Ramsey, M.N., Fuller, D.Q., Richter, T., 2018. Archaeobotanical evidence reveals the origins of bread 14,400 years ago in northeastern Jordan. Proc. Natl. Acad. Sci. U. S. A. 115, 7925–7930. https://doi.org/10.1073/pnas.1801071115

Aziz, R.K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics 9, 1–15. https://doi.org/10.1186/1471-2164-9-75

Badel, S., Bernardi, T., Michaud, P., 2011. New perspectives for Lactobacilli exopolysaccharides. Biotechnol. Adv. 29, 54–66. https://doi.org/10.1016/j.biotechadv.2010.08.011

Baliarda, A., Robert, H., Jebbar, M., Blanco, C., Deschamps, A., Le Marrec, C., 2003. Potential osmoprotectants for the lactic acid bacteria *Pediococcus pentosaceus* and

Tetragenococcus halophila. Int. J. Food Microbiol. 84, 13–20. https://doi.org/10.1016/S0168-1605(02)00388-4

Barbieri, F., Montanari, C., Gardini, F., Tabanelli, G., 2019. Biogenic amine production by lactic acid bacteria: A review. Foods 8, 1–27. https://doi.org/10.3390/foods8010017

Barefoot, S.F., Klaenhammer, T.R., 1983. Detection and activity of lactacin B, a bacteriocin produced by *Lactobacillus acidophilus*. Appl. Environ. Microbiol. 45, 1808–1815. https://doi.org/10.1128/aem.45.6.1808-1815.1983

Beddows, C.G., 1998. Fermented fish and fish products, in: Microbiology of Fermented Foods. Springer US, Boston, MA, pp. 416–440. https://doi.org/10.1007/978-1-4613-0309-1_13

Behr, J., Geissler, A.J., Schmid, J., Zehe, A., Vogel, R.F., 2016. The identification of novel diagnostic marker genes for the detection of beer spoiling *Pediococcus damnosus* strains using the BlAst diagnostic gene findEr. PLoS One 11, 1–23. https://doi.org/10.1371/journal.pone.0152747

Beltrán-Barrientos, L.M., Hernández-Mendoza, A., Torres-Llanez, M.J., González-Córdova, A.F., Vallejo-Córdoba, B., 2016. Invited review: Fermented milk as antihypertensive functional food. J. Dairy Sci. 99, 4099–4110. https://doi.org/10.3168/jds.2015-10054

Berrocal, C.A., Rivera-Vicens, R.E., Nadathur, G.S., 2016. Draft Genome Sequence of the Heavy-Metal-Tolerant Marine Yeast *Debaryomyces hansenii* J6. Genome Announc. 4, 6–7. https://doi.org/10.1128/genomeA.00983-16

Bizzini, A., Zhao, C., Budin-Verneuil, A., Sauvageot, N., Giard, J.-C., Auffray, Y., Hartke, A., 2010. Glycerol Is Metabolized in a Complex and Strain-Dependent Manner in *Enterococcus faecalis*. J. Bacteriol. 192, 779–785. https://doi.org/10.1128/JB.00959-09

Bover-Cid, S., Holzapfel, W.H., 1999. Improved screening procedure for biogenic amine production by lactic acid bacteria. Int. J. Food Microbiol. 53, 33–41. https://doi.org/10.1016/s0168-1605(99)00152-x

Brown, C.J., Todd, K.M., Rosenzweig, R.F., 1998. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol. Biol. Evol. 15, 931–942. https://doi.org/10.1093/oxfordjournals.molbev.a026009

Byun, M.-S., Yu, O.-K., Cha, Y.-S., Park, T.-S., 2016. Korean traditional *Chungkookjang* improves body composition, lipid profiles and atherogenic indices in overweight/obese subjects: a double-blind, randomized, crossover, placebo-controlled clinical trial. Eur. J. Clin. Nutr. 70, 1116–1122. https://doi.org/10.1038/ejcn.2016.77

Cabanettes, F., Klopp, C., 2018. D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. PeerJ 2018. https://doi.org/10.7717/peerj.4958

Cantalapiedra, C.P., Herṇandez-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol. Biol. Evol. 38, 5825–5829. https://doi.org/10.1093/molbev/msab293

Cerning-Béroard, J., Filiatre-Verel, A., 1980. Characterization and distribution of soluble and insoluble carbohydrates in lupin seeds. Z. Lebensm. Unters. Forsch. 171, 281–5. https://doi.org/10.1007/BF01042478

Chun, B.H., Han, D.M., Kim, K.H., Jeong, S.E., Park, D., Jeon, C.O., 2019. Genomic and metabolic features of *Tetragenococcus halophilus* as revealed by pan-genome and transcriptome analyses. Food Microbiol. 83, 36–47. https://doi.org/10.1016/j.fm.2019.04.009

Collins, M.D., Williams, A.M., Wallbanks, S., 1990. The phylogeny of *Aerococcus* and

Pediococcus as determined by 16S rRNA sequence analysis: description of *Tetragenococcus* gen. nov. FEMS Microbiol. Lett. 70, 255–262. https://doi.org/10.1111/j.1574-6968.1990.tb14006.x

Corredor, M., Davila, A.-M., Casarégola, S., Gaillardin, C., 2003. Chromosomal polymorphism in the yeast species *Debaryomyces hansenii*. Antonie Van Leeuwenhoek 84, 81–8. https://doi.org/10.1023/a:1025432721866

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E.P.C., Vergnaud, G., Gautheret, D., Pourcel, C., 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 46, W246–W251. https://doi.org/10.1093/nar/gky425

de Jong, A., van Hijum, S.A.F.T., Bijlsma, J.J.E., Kok, J., Kuipers, O.P., 2006. BAGEL: A web-based bacteriocin genome mining tool. Nucleic Acids Res. 34, 273–279. https://doi.org/10.1093/nar/gkl237

De Vuyst, L., Degeest, B., 1999. Heteropolysaccharides from lactic acid bacteria. FEMS Microbiol. Rev. 23, 153–177. https://doi.org/10.1016/S0168-6445(98)00042-4

De Vuyst, L., Zamfir, M., Mozzi, F., Adriany, T., Marshall, V., Degeest, B., Vaningelgem, F., 2003. Exopolysaccharide-producing *Streptococcus thermophilus* strains as functional starter cultures in the production of fermented milks. Int. Dairy J. 13, 707–717. https://doi.org/10.1016/S0958-6946(03)00105-5

Decock, P., Cappelle, S., 2005. Bread technology and sourdough technology. Trends Food Sci. Technol. 16, 113–120. https://doi.org/10.1016/j.tifs.2004.04.012

Devanthi, P.V.P., Gkatzionis, K., 2019. Soy sauce fermentation: Microorganisms, aroma formation, and process modification. Food Res. Int. 120, 364–374. https://doi.org/10.1016/j.foodres.2019.03.010

Devanthi, P.V.P., Linforth, R., Onyeaka, H., Gkatzionis, K., 2018. Effects of co-inoculation and sequential inoculation of *Tetragenococcus halophilus* and *Zygosaccharomyces rouxii* on soy sauce fermentation. Food Chem. 240, 1–8. https://doi.org/10.1016/j.foodchem.2017.07.094

Doi, Y., 2019. Glycerol metabolism and its regulation in lactic acid bacteria. Appl. Microbiol. Biotechnol. 103, 5079–5093. https://doi.org/10.1007/s00253-019-09830-y

Doi, Y., 2015. L-Lactate Production from Biodiesel-Derived Crude Glycerol by Metabolically Engineered *Enterococcus faecalis*: Cytotoxic Evaluation of Biodiesel Waste and Development of a Glycerol-Inducible Gene Expression System. Appl. Environ. Microbiol. 81, 2082–2089. https://doi.org/10.1128/AEM.03418-14

Dominguez, J.M., 1997. Production of xylitol from D-xylose by *Debaryomyces hansenii*. Appl. Biochem. Biotechnol. - Part A Enzym. Eng. Biotechnol. 63–65, 117–127. https://doi.org/10.1007/BF02920418

Du, F., Zhang, X., Gu, H., Song, J., Gao, X., 2019. Dynamic changes in the bacterial community during the fermentation of traditional Chinese fish sauce (TCFS) and their correlation with TCFS quality. Microorganisms 7. https://doi.org/10.3390/microorganisms7090371

Durso, L., Hutkins, R., 2003. Starter Cultures, in: Encyclopedia of Food Sciences and Nutrition. Elsevier, pp. 5583–5593. https://doi.org/10.1016/B0-12-227055-X/01146-9

Ehrmann, M.A., Müller, M.R.A., Vogel, R.F., 2003. Molecular analysis of sourdough reveals *Lactobacillus mindensis* sp. nov. Int. J. Syst. Evol. Microbiol. 53, 7–13. https://doi.org/10.1099/ijs.0.02202-0

REFERENCES

Etchells, J.L., Borg, A.F., Kittel, I.D., Bell, T.A., Fleming, H.P., 1966. Pure Culture Fermentation of Green Olives. Appl. Microbiol. 14, 1027–1041. https://doi.org/10.1128/am.14.6.1027-1041.1966

Fang, W., Siegumfeldt, H., Budde, B.B., Jakobsen, M., 2004. Osmotic stress leads to decreased intracellular pH of *Listeria monocytogenes* as determined by fluorescence ratio-imaging microscopy. Appl. Environ. Microbiol. 70, 3176–3179. https://doi.org/10.1128/AEM.70.5.3176-3179.2004

Felsenstein, J., 1985. Confidence Limits on Phylogenies: an Approach Using the Bootstrap. Evolution (N. Y). 39, 783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Freiding, S., Gutsche, K.A., Ehrmann, M.A., Vogel, R.F., 2011. Genetic screening of *Lactobacillus sakei* and *Lactobacillus curvatus* strains for their peptidolytic system and amino acid metabolism, and comparison of their volatilomes in a model system. Syst. Appl. Microbiol. 34, 311–320. https://doi.org/10.1016/j.syapm.2010.12.006

Fukui, Y., Yoshida, M., Shozen, K. ichi, Funatsu, Y., Takano, T., Oikawa, H., Yano, Y., Satomi, M., 2012. Bacterial communities in fish sauce mash using culture-dependent and -independent methods, in: Journal of General and Applied Microbiology. pp. 273–281. https://doi.org/10.2323/jgam.58.273

Fukushima, D., 1985. Fermented vegetable protein and related foods of Japan and China. Food Rev. Int. 1, 149–209. https://doi.org/10.1080/87559128509540768

Fukushima, D., 1981. Soy proteins for foods centering around soy sauce and tofu. J. Am. Oil Chem. Soc. 58, 346–354. https://doi.org/10.1007/BF02582376

Gabaldón, T., 2020. Hybridization and the origin of new yeast lineages. FEMS Yeast Res. 20, 1–8. https://doi.org/10.1093/femsyr/foaa040

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E. V., 2015. Expanded Microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43, D261–D269. https://doi.org/10.1093/nar/gku1223

Gänzle, M.G., 2015. Lactic metabolism revisited: metabolism of lactic acid bacteria in food fermentations and food spoilage. Curr. Opin. Food Sci. 2, 106–117. https://doi.org/10.1016/j.cofs.2015.03.001

Garneau, J.E., Moineau, S., 2011. Bacteriophages of lactic acid bacteria and their impact on milk fermentations. Microb. Cell Fact. 10, S20. https://doi.org/10.1186/1475-2859-10-S1-S20

Gremme, G., Steinbiss, S., Kurtz, S., 2013. Genome tools: A comprehensive software library for efficient processing of structured genome annotations. IEEE/ACM Trans. Comput. Biol. Bioinforma. 10, 645–656. https://doi.org/10.1109/TCBB.2013.68

Groenewald, M., Daniel, H.-M., Robert, V., Poot, G.A., Smith, M.T., 2008. Polyphasic re-examination of *Debaryomyces hansenii* strains and reinstatement of *D. hansenii*, *D. fabryi* and *D. subglobosus*. Persoonia 21, 17–27. https://doi.org/10.3767/003158508X336576

Hallgren, J., Tsirigos, K.D., Damgaard Pedersen, M., Juan, J., Armenteros, A., Marcatili, P., Nielsen, H., Krogh, A., Winther, O., 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv 2022.04.08.487609. https://doi.org/10.1101/2022.04.08.487609

Hammes, W.P., 1990. Bacterial starter cultures in food production. Food Biotechnol. 4, 383–397. https://doi.org/10.1080/08905439009549750

Han, D.M., Chun, B.H., Feng, T., Kim, H.M., Jeon, C.O., 2020. Dynamics of microbial communities and metabolites in ganjang, a traditional Korean fermented soy sauce,

during fermentation. Food Microbiol. 92, 103591. https://doi.org/10.1016/j.fm.2020.103591

Harada, R., Yuzuki, M., Ito, K., Shiga, K., Bamba, T., Fukusaki, E., 2017. Influence of yeast and lactic acid bacterium on the constituent profile of soy sauce during fermentation. J. Biosci. Bioeng. 123, 203–208. https://doi.org/10.1016/j.jbiosc.2016.08.010

Haubold, B., Pfaffelhuber, P., Domazet-Lošo, M., Wiehe, T., 2009. Estimating mutation distances from unaligned genomes. J. Comput. Biol. 16, 1487–1500. https://doi.org/10.1089/cmb.2009.0106

Hayden, B., Canuel, N., Shanse, J., 2013. What Was Brewing in the Natufian? An Archaeological Assessment of Brewing Technology in the Epipaleolithic, Journal of Archaeological Method and Theory. https://doi.org/10.1007/s10816-011-9127-y

He, G., Deng, J., Wu, C., Huang, J., 2017a. A partial proteome reference map of *Tetragenococcus halophilus* and comparative proteomic and physiological analysis under salt stress. RSC Adv. 7, 12753–12763. https://doi.org/10.1039/C6RA22521G

He, G., Wu, C., Huang, J., Zhou, R., 2017b. Effect of exogenous proline on metabolic response of *Tetragenococcus halophilus* under salt stress. J. Microbiol. Biotechnol. 27, 1681–1691. https://doi.org/10.4014/jmb.1702.02060

Heo, S., Lee, J., Lee, J.-H., Jeong, D.-W., 2019. Genomic Insight into the Salt Tolerance of *Enterococcus faecium*, *Enterococcus faecalis* and *Tetragenococcus halophilus*. J. Microbiol. Biotechnol. 29, 1591–1602. https://doi.org/10.4014/jmb.1908.08015

Higuchi, T., Uchida, K., Abe, K., 1999. Preparation of Phage-insensitive Strains of *Tetragenococcus halophila* and Its Application for Soy Sauce Fermentation. Biosci. Biotechnol. Biochem. https://doi.org/10.1271/bbb.63.415

Higushi, T., Uchida, K., Abe, K., 1998. Aspartate Decarboxylation Encoded on the Plasmid in the Soy Sauce Lactic Acid Bacterium, *Tetragenococcus halophila* D10. Biosci. Biotechnol. Biochem. 62, 1601–1603. https://doi.org/10.1271/bbb.62.1601

Hoog, G.S., Ende, A.H.G.G., 1998. Molecular diagnostics of clinical strains of filamentous Basidiomycetes. Mycoses 41, 183–189. https://doi.org/10.1111/j.1439-0507.1998.tb00321.x

Horvath, P., Barrangou, R., 2010. CRISPR/Cas, the Immune System of Bacteria and Archaea. Science (80-.). 327, 167–170. https://doi.org/10.1126/science.1179555

Hu, Y., Zhang, L., Wen, R., Chen, Q., Kong, B., 2022. Role of lactic acid bacteria in flavor development in traditional Chinese fermented foods: A review. Crit. Rev. Food Sci. Nutr. 62, 2741–2755. https://doi.org/10.1080/10408398.2020.1858269

Huey, B., Hall, J., 1989. Hypervariable DNA fingerprinting in *Escherichia coli*: minisatellite probe from bacteriophage M13. J. Bacteriol. 171, 2528–2532. https://doi.org/10.1128/jb.171.5.2528-2532.1989

Hur, S.J., Lee, S.Y., Kim, Y.-C., Choi, I., Kim, G.-B., 2014. Effect of fermentation on the antioxidant activity in plant-based foods. Food Chem. 160, 346–356. https://doi.org/10.1016/j.foodchem.2014.03.112

Iizuka, H., Yamasato, K., 1959. *Pediococcus soyae* nov. sp. Main Lactic Acid Bacteria in “Shoyu Moromi “∗. J. Gen. Appl. Microbiol. 5, 58–73. https://doi.org/10.2323/jgam.5.58

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9, 5114. https://doi.org/10.1038/s41467-018-07641-9

Jakobsen, M., Narvhus, J., 1996. Yeasts and their possible beneficial and negative effects on

the quality of dairy products. Int. Dairy J. 6, 755–768. https://doi.org/10.1016/0958-6946(95)00071-2

James, S.A., Bond, C.J., Stratford, M., Roberts, I.N., 2005. Molecular evidence for the existence of natural hybrids in the genus *Zygosaccharomyces*. FEMS Yeast Res. 5, 747–55. https://doi.org/10.1016/j.femsyr.2005.02.004

Janßen, D., Ehrmann, M.A., Vogel, R.F., 2019. Monitoring of assertive *Lactobacillus sakei* and *Lactobacillus curvatus* strains using an industrial ring trial experiment. J. Appl. Microbiol. 126, 545–554. https://doi.org/10.1111/jam.14144

Jeong, D.M., Yoo, S.J., Jeon, M.-S., Chun, B.H., Han, D.M., Jeon, C.O., Eyun, S.-I., Seo, Y.-J., Kang, H.A., 2022. Genomic features, aroma profiles, and probiotic potential of the *Debaryomyces hansenii* species complex strains isolated from Korean soybean fermented food. Food Microbiol. 105, 104011. https://doi.org/10.1016/j.fm.2022.104011

Jezierny, D., Mosenthin, R., Sauer, N., Roth, S., Piepho, H.-P., Rademacher, M., Eklund, M., 2011. Chemical composition and standardised ileal digestibilities of crude protein and amino acids in grain legumes for growing pigs. Livest. Sci. 138, 229–243. https://doi.org/10.1016/j.livsci.2010.12.024

Jolly, L., Stingele, F., 2001. Molecular organization and functionality of exopolysaccharide gene clusters in lactic acid bacteria. Int. Dairy J. 11, 733–745. https://doi.org/10.1016/S0958-6946(01)00117-0

Jung, J.Y., Lee, H.J., Chun, B.H., Jeon, C.O., 2016. Effects of temperature on bacterial communities and metabolites during fermentation of Myeolchi-Aekjeot, a traditional Korean fermented anchovy sauce. PLoS One 11, 1–20. https://doi.org/10.1371/journal.pone.0151351

Justé, A., Lievens, B., Frans, I., Marsh, T.L., Klingeberg, M., Michiels, C.W., Willems, K.A., 2008a. Genetic and physiological diversity of *Tetragenococcus halophilus* strains isolated from sugar- and salt-rich environments. Microbiology 154, 2600–2610. https://doi.org/10.1099/mic.0.2008/018168-0

Justé, A., Lievens, B., Klingeberg, M., Michiels, C.W., Marsh, T.L., Willems, K.A., 2008b. Predominance of *Tetragenococcus halophilus* as the cause of sugar thick juice degradation. Food Microbiol. 25, 413–421. https://doi.org/10.1016/j.fm.2007.10.012

Justé, A., Lievens, B., Rediers, H., Willems, K.A., 2014. The genus *Tetragenococcus*. Lact. Acid Bact. Biodivers. Taxon. 9781444333, 213–227. https://doi.org/10.1002/9781118655252.ch16

Justé, A., Van Trappen, S., Verreth, C., Cleenwerck, I., De Vos, P., Lievens, B., Willems, K.A., 2012. Characterization of *Tetragenococcus* strains from sugar thick juice reveals a novel species, *Tetragenococcus osmophilus* sp. nov., and divides *Tetragenococcus halophilus* into two subspecies, *T. halophilus* subsp. *halophilus*. Int. J. Syst. Evol. Microbiol. 62, 129–137. https://doi.org/10.1099/ijs.0.029157-0

Kaczmarska, K.T., Chandra-Hioe, M. V., Zabaras, D., Frank, D., Arcot, J., 2017. Effect of Germination and Fermentation on Carbohydrate Composition of Australian Sweet Lupin and Soybean Seeds and Flours. J. Agric. Food Chem. 65, 10064–10073. https://doi.org/10.1021/acs.jafc.7b02986

Kanehisa, M., Sato, Y., Kawashima, M., 2022. KEGG mapping tools for uncovering hidden features in biological data. Protein Sci. 31, 47–53. https://doi.org/10.1002/pro.4172

Kawai, Y., Marles-Wright, J., Cleverley, R.M., Emmins, R., Ishikawa, S., Kuwano, M., Heinz, N., Bui, N.K., Hoyland, C.N., Ogasawara, N., Lewis, R.J., Vollmer, W., Daniel, R.A., Errington, J., 2011. A widespread family of bacterial cell wall assembly proteins. EMBO

J. 30, 4931–4941. https://doi.org/10.1038/emboj.2011.358

Kim, E., Kim, J., Yang, S., Suh, S., Kim, H.-J., Kim, C.-G., Choo, D.-W., Kim, H.-Y., 2017. Draft Genome Sequence of *Tetragenococcus halophilus* Strain FBL3, a Probiotic Bacterium Isolated from Galchijeot, a Salted Fermented Food, in the Republic of Korea. Genome Announc. 5, 17–18. https://doi.org/10.1128/genomeA.00304-17

Kim, K.H., Lee, S.H., Chun, B.H., Jeong, S.E., Jeon, C.O., 2019. *Tetragenococcus halophilus* MJ4 as a starter culture for repressing biogenic amine (cadaverine) formation during saeu-jeot (salted shrimp) fermentation. Food Microbiol. 82, 465–473. https://doi.org/10.1016/j.fm.2019.02.017

Kim, T.W., Lee, J.H., Kim, S.E., Park, M.H., Chang, H.C., Kim, H.Y., 2009. Analysis of microbial communities in doenjang, a Korean fermented soybean paste, using nested PCR-denaturing gradient gel electrophoresis. Int. J. Food Microbiol. 131, 265–271. https://doi.org/10.1016/j.ijfoodmicro.2009.03.001

Klaenhammer, T.R., 1988. Bacteriocins of lactic acid bacteria. Biochimie 70, 337–349. https://doi.org/10.1016/0300-9084(88)90206-4

Kobayashi, T., Kajiwara, M., Wahyuni, M., Hamada-Sato, N., Imada, C., Watanabe, E., 2004. Effect of culture conditions on lactic acid production of *Tetragenococcus* species. J. Appl. Microbiol. 96, 1215–1221. https://doi.org/10.1111/j.1365-2672.2004.02267.x

Kondrashov, F.A., 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc. R. Soc. B Biol. Sci. 279, 5048–5057. https://doi.org/10.1098/rspb.2012.1108

Koo, O.K., Lee, S.J., Chung, K.R., Jang, D.J., Yang, H.J., Kwon, D.Y., 2016. Korean traditional fermented fish products: jeotgal. J. Ethn. Foods 3, 107–116. https://doi.org/10.1016/j.jef.2016.06.004

Kumar, S., Randhawa, A., Ganesan, K., Raghava, G.P.S., Mondal, A.K., 2012. Draft genome sequence of salt-tolerant yeast *Debaryomyces hansenii* var. *hansenii* MTCC 234. Eukaryot. Cell 11, 961–962. https://doi.org/10.1128/EC.00137-12

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054

Lavelle, K., Sinderen, D. van, Mahony, J., 2021. Cell wall polysaccharides of Gram positive ovococcoid bacteria and their role as bacteriophage receptors. Comput. Struct. Biotechnol. J. 19, 4018–4031. https://doi.org/10.1016/j.csbj.2021.07.011

Leal-Sánchez, M. V., Jiménez-Díaz, R., Maldonado-Barragán, A., Garrido-Fernández, A., Ruiz-Barba, J.L., 2002. Optimization of bacteriocin production by batch fermentation of *Lactobacillus plantarum* LPCO10. Appl. Environ. Microbiol. 68, 4465–4471. https://doi.org/10.1128/AEM.68.9.4465-4471.2002

Lee, C.H., 1997. Lactic acid fermented foods and their benefits in Asia. Food Control 8, 259–269. https://doi.org/10.1016/s0956-7135(97)00015-7

Lee, J.H., Heo, S., Jeong, K., Lee, B., Jeong, D.W., 2018. Genomic insights into the non-histamine production and proteolytic and lipolytic activities of *Tetragenococcus halophilus* KUD23. FEMS Microbiol. Lett. 365, 1–4. https://doi.org/10.1093/femsle/fnx252

Lee, K.E., Lee, S.M., Choi, Y.H., Hurh, B.S., Kim, Y.-S., 2013. Comparative volatile profiles in soy sauce according to inoculated microorganisms. Biosci. Biotechnol. Biochem. 77, 2192–200. https://doi.org/10.1271/bbb.130362

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM

00, 1–3.

Li, W., O'Neill, K.R., Haft, D.H., Dicuccio, M., Chetvernin, V., Badretdin, A., Coulouris, G., Chitsaz, F., Derbyshire, M.K., Durkin, A.S., Gonzales, N.R., Gwadz, M., Lanczycki, C.J., Song, J.S., Thanki, N., Wang, J., Yamashita, R.A., Yang, M., Zheng, C., Marchler-Bauer, A., Thibaud-Nissen, F., 2021. RefSeq: Expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. Nucleic Acids Res. 49, D1020–D1028. https://doi.org/10.1093/nar/gkaa1105

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–30. https://doi.org/10.1093/bioinformatics/btt656

Lim, J.-H., Jung, E.-S., Choi, E.-K., Jeong, D.-Y., Jo, S.-W., Jin, J.-H., Lee, J.-M., Park, B.-H., Chae, S.-W., 2015. Supplementation with *Aspergillus oryzae*-fermented kochujang lowers serum cholesterol in subjects with hyperlipidemia. Clin. Nutr. 34, 383–387. https://doi.org/10.1016/j.clnu.2014.05.013

Lin, J., Liang, H., Yan, J., Luo, L., 2017. The molecular mechanism and post-transcriptional regulation characteristic of *Tetragenococcus halophilus* acclimation to osmotic stress revealed by quantitative proteomics. J. Proteomics 168, 1–14. https://doi.org/10.1016/j.jprot.2017.08.014

Link, T., Ehrmann, M.A., 2023a. Transcriptomic profiling reveals differences in the adaptation of two *Tetragenococcus halophilus* strains to a lupine moromi model medium. BMC Microbiol. 23, 14. https://doi.org/10.1186/s12866-023-02760-w

Link, T., Ehrmann, M.A., 2023b. Monitoring the growth dynamics of *Tetragenococcus halophilus* strains in lupine moromi fermentation using a multiplex-PCR system. BMC Res. Notes 16, 115. https://doi.org/10.1186/s13104-023-06406-y

Link, T., Lülf, R.H., Parr, M., Hilgarth, M., Ehrmann, M.A., 2022. Genome Sequence of the Diploid Yeast *Debaryomyces hansenii* TMW 3.1188. Microbiol. Resour. Announc. 11, 5–7. https://doi.org/10.1128/mra.00649-22

Link, T., Vogel, R.F., Ehrmann, M.A., 2021. The diversity among the species *Tetragenococcus halophilus* including new isolates from a lupine seed fermentation. BMC Microbiol. 21, 320. https://doi.org/10.1186/s12866-021-02381-1

Lioe, H.N., Selamat, J., Yasuda, M., 2010. Soy sauce and its umami taste: A link from the past to current situation. J. Food Sci. 75. https://doi.org/10.1111/j.1750-3841.2010.01529.x

Lister, J., 1877. Introductory Address Delivered in the Medical Department of King's College. BMJ 2, 465–469. https://doi.org/10.1136/bmj.2.875.465

Liu, L., Si, L., Meng, X., Luo, L., 2015. Comparative transcriptomic analysis reveals novel genes and regulatory mechanisms of *Tetragenococcus halophilus* in response to salt stress. J. Ind. Microbiol. Biotechnol. 42, 601–616. https://doi.org/10.1007/s10295-014-1579-0

Loman, A.A., Islam, S.M.M., Ju, L.K., 2018. Production of arabitol from enzymatic hydrolysate of soybean flour by Debaryomyces hansenii fermentation. Appl. Microbiol. Biotechnol. 102, 641–653. https://doi.org/10.1007/s00253-017-8626-5

Lopetcharat, K., Choi, Y.J., Park, J.W., Daeschel, M.A., 2001. Fish Sauce Products and Manufacturing: A REVIEW. Food Rev. Int. 17, 65–88. https://doi.org/10.1081/FRI-100000515

Louis, P., 1879. Studies on Fermentation. The Diseases of Beer, their Causes, and the means of Preventing them. A Translation, made with the Author's Sanction, of "Etudes sur la biere" with Notes, Index, and Original Illustration., Robb Macmilan and Co.,.

REFERENCES

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A., 2020. CDD/SPARCLE: The conserved domain database in 2020. Nucleic Acids Res. 48, D265–D268. https://doi.org/10.1093/nar/gkz991

Lülf, R.H., Selg-Mann, K., Hoffmann, T., Zheng, T., Schirmer, M., Ehrmann, M.A., 2023. Carbohydrate Sources Influence the Microbiota and Flavour Profile of a Lupine-Based Moromi Fermentation. Foods 12, 197. https://doi.org/10.3390/foods12010197

Lülf, R.H., Vogel, R.F., Ehrmann, M.A., 2021. Microbiota dynamics and volatile compounds in lupine based Moromi fermented at different salt concentrations. Int. J. Food Microbiol. 354, 109316. https://doi.org/10.1016/j.ijfoodmicro.2021.109316

Marco, M.L., Heeney, D., Binda, S., Cifelli, C.J., Cotter, P.D., Foligné, B., Gänzle, M., Kort, R., Pasin, G., Pihlanto, A., Smid, E.J., Hutkins, R., 2017. Health benefits of fermented foods: microbiota and beyond. Curr. Opin. Biotechnol. 44, 94–102. https://doi.org/10.1016/j.copbio.2016.11.010

Marco, M.L., Sanders, M.E., Gänzle, M., Arrieta, M.C., Cotter, P.D., De Vuyst, L., Hill, C., Holzapfel, W., Lebeer, S., Merenstein, D., Reid, G., Wolfe, B.E., Hutkins, R., 2021. The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on fermented foods. Nat. Rev. Gastroenterol. Hepatol. 18, 196–208. https://doi.org/10.1038/s41575-020-00390-5

Marraffini, L.A., 2013. CRISPR-Cas Immunity against Phages: Its Effects on the Evolution and Survival of Bacterial Pathogens. PLoS Pathog. 9, 1–4. https://doi.org/10.1371/journal.ppat.1003765

Matsutani, M., Wakinaka, T., Watanabe, J., Tokuoka, M., Ohnishi, A., 2021. Comparative Genomics of Closely Related *Tetragenococcus halophilus* Strains Elucidate the Diversity and Microevolution of CRISPR Elements. Front. Microbiol. 12, 1–10. https://doi.org/10.3389/fmicb.2021.687985

Mees, R.H., 1934. Onderzoekingen over de Biersarcina. Thesis, Tech. Univ.

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.P., Göker, M., 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14. https://doi.org/10.1186/1471-2105-14-60

Monsan, P., Bozonnet, S., Albenne, C., Joucla, G., Willemot, R.M., Remaud-Siméon, M., 2001. Homopolysaccharides from lactic acid bacteria. Int. Dairy J. 11, 675–685. https://doi.org/10.1016/S0958-6946(01)00113-3

Nakase, T., Suzuki, M., Phaff, H.J., Kurtzman, C.P., 1998. *Debaryomyces* Lodder & Kreger-van Rij Nom. Cons, in: The Yeasts. Elsevier, pp. 157–173. https://doi.org/10.1016/B978-044481312-1/50030-7

Nardi, T., Panero, L., Petrozziello, M., Guaita, M., Tsolakis, C., Cassino, C., Vagnoli, P., Bosso, A., 2019. Managing wine quality using *Torulaspora delbrueckii* and *Oenococcus oeni* starters in mixed fermentations of a red Barbera wine. Eur. Food Res. Technol. 245, 293–307. https://doi.org/10.1007/s00217-018-3161-x

Nishimura, I., Shiwa, Y., Sato, A., Oguma, T., Yoshikawa, H., Koyama, Y., 2017. Comparative genomics of *Tetragenococcus halophilus*. J. Gen. Appl. Microbiol. 63, 369–372. https://doi.org/10.2323/jgam.2017.02.003

Nout, M.J.R., 1994. Fermented foods and food safety. Food Res. Int. 27, 291–298. https://doi.org/10.1016/0963-9969(94)90097-3

Orla-Jensen, S., 1919. The Lactic Acid Bacteria. Biologiske skrifter.

REFERENCES

Osei Abunyewa, A.A., Laing, E., Hugo, A., Viljoen, B.C., 2000. The population change of yeasts in commercial salami. Food Microbiol. 17, 429–438. https://doi.org/10.1006/fmic.1999.0333

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., Buell, R.C., 2007. The TIGR Rice Genome Annotation Resource: Improvements and new features. Nucleic Acids Res. 35, 8–11. https://doi.org/10.1093/nar/gkl976

Papagianni, M., Anastasiadou, S., 2009. Pediocins: The bacteriocins of Pediococci. Sources, production, properties and applications. Microb. Cell Fact. 8, 1–16. https://doi.org/10.1186/1475-2859-8-3

Patra, J.K., Das, G., Paramithiotis, S., Shin, H.S., 2016. Kimchi and other widely consumed traditional fermented foods of Korea: A review. Front. Microbiol. 7, 1–15. https://doi.org/10.3389/fmicb.2016.01493

Petersen, K.M., Westall, S., Jespersen, L., 2002. Microbial succession of *Debaryomyces hansenii* strains during the production of Danish surfaced-ripened cheeses. J. Dairy Sci. 85, 478–486. https://doi.org/10.3168/jds.S0022-0302(02)74098-8

Pinilla-Redondo, R., Russel, J., Mayo-Muñoz, D., Shah, S.A., Garrett, R.A., Nesme, J., Madsen, J.S., Fineran, P.C., Sørensen, S.J., 2022. CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids. Nucleic Acids Res. 50, 4315–4328. https://doi.org/10.1093/nar/gkab859

Prete, R., Alam, M.K., Perpetuini, G., Perla, C., Pittia, P., Corsetti, A., 2021. Lactic acid bacteria exopolysaccharides producers: A sustainable tool for functional foods. Foods 10, 1–27. https://doi.org/10.3390/foods10071653

Pribylova, L., de Montigny, J., Sychrova, H., 2007. Osmoresistant yeast *Zygosaccharomyces rouxii*: the two most studied wild-type strains (ATCC 2623 and ATCC 42981) differ in osmotolerance and glycerol metabolism. Yeast 24, 171–180. https://doi.org/10.1002/yea.1470

Proux-Wéra, E., Armisén, D., Byrne, K.P., Wolfe, K.H., 2012. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. BMC Bioinformatics 13. https://doi.org/10.1186/1471-2105-13-237

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., Peplies, J., 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32, 929–931. https://doi.org/10.1093/bioinformatics/btv681

Rivas, B., Torre, P., Domínguez, J.M., Converti, A., 2009. Maintenance and growth requirements in the metabolism of *Debaryomyces hansenii* performing xylose-to-xylitol bioconversion in corncob hemicellulose hydrolyzate. Biotechnol. Bioeng. 102, 1062–1073. https://doi.org/10.1002/bit.22155

Robert, H., Le Marrec, C., Blanco, C., Jebbar, M., 2000. Glycine betaine, carnitine, and choline enhance salinity tolerance and prevent the accumulation of sodium to a level inhibiting growth of *Tetragenococcus halophila*. Appl. Environ. Microbiol. 66, 509–517. https://doi.org/10.1128/AEM.66.2.509-517.2000

Roberts, M.F., 2005. Organic compatible solutes of halotolerant and halophilic microorganisms. Saline Systems 1, 5. https://doi.org/10.1186/1746-1448-1-5

Roberts, R.F., Zottola, E.A., Mckay, L.L., 1992. Use of a Nisin-Producing Starter Culture Suitable for Cheddar Cheese Manufacture. J. Dairy Sci. 75, 2353–2363. https://doi.org/10.3168/jds.S0022-0302(92)77995-8

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for

differential expression analysis of digital gene expression data. Bioinformatics 26, 139–40. https://doi.org/10.1093/bioinformatics/btp616

Roca, M., Incze, K., 1990. Fermented sausages. Food Rev. Int. 6, 91–118. https://doi.org/10.1080/87559129009540862

Rodríguez, J., González-Guerra, A., Vázquez, L., Fernández-López, R., Flórez, A.B., de la Cruz, F., Mayo, B., 2022. Isolation and phenotypic and genomic characterization of *Tetragenococcus* spp. from two Spanish traditional blue-veined cheeses made of raw milk. Int. J. Food Microbiol. 371, 109670. https://doi.org/10.1016/j.ijfoodmicro.2022.109670

Roling, W., van Verseveld, H.W., 1996. Characterization of *Tetragenococcus halophila* Populations in Indonesian Soy Mash (Kecap) Fermentation. Appl. Environ. Microbiol. 62, 1203–1207. https://doi.org/10.1128/aem.62.4.1203-1207.1996

Röling, W.F.M., Timotius, K.H., Budhi Prasetyo, A., Stouthamer, A.H., Van Verseveld, H.W., 1994. Changes in microflora and biochemical composition during the Baceman stage of traditional indonesian Kecap (soy sauce) production. J. Ferment. Bioeng. 77, 62–70. https://doi.org/10.1016/0922-338X(94)90210-0

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Sakaguchi, K., 1958. Studies on the activities of bacteria in soy sauce brewing: Part III. Taxonomic studies on *Pediococcus soyae* nov. sp., The soy sauce lactic acid bacteria. J. Agric. Chem. Soc. Japan 22, 354–362. https://doi.org/10.1080/03758397.1958.10857506

Sakamoto, T., Nishimura, Y., Makino, Y., Sunagawa, Y., Harada, N., 2013. Biochemical characterization of a GH53 endo-β-1,4-galactanase and a GH35 exo-β-1,4-galactanase from *Penicillium chrysogenum*. Appl. Microbiol. Biotechnol. 97, 2895–2906. https://doi.org/10.1007/s00253-012-4154-5

Sasahara, A., Nanatani, K., Enomoto, M., Kuwahara, S., Abe, K., 2011. Substrate specificity of the aspartate:Alanine antiporter (AspT) of *Tetragenococcus halophilus* in reconstituted liposomes. J. Biol. Chem. 286, 29044–29052. https://doi.org/10.1074/jbc.M111.260224

Sato, A., Matsushima, K., Oshima, K., Hattori, M., Koyama, Y., 2017. Draft Genome Sequencing of the Highly Halotolerant and Allopolyploid Yeast *Zygosaccharomyces rouxii* NBRC 1876. Genome Announc. 5, 1–2. https://doi.org/10.1128/genomeA.01610-16

Satomi, M., Furushita, M., Oikawa, H., Yoshikawa-Takahashi, M., Yano, Y., 2008. Analysis of a 30 kbp plasmid encoding histidine decarboxylase gene in *Tetragenococcus halophilus* isolated from fish sauce. Int. J. Food Microbiol. 126, 202–209. https://doi.org/10.1016/j.ijfoodmicro.2008.05.025

Shin, D., Jeong, D., 2015. Korean traditional fermented soybean products: Jang. J. Ethn. Foods 2, 2–7. https://doi.org/10.1016/j.jef.2015.02.002

Shirakawa, D., Wakinaka, T., Watanabe, J., 2020. Identification of the putative *N*-acetylglucosaminidase CseA associated with daughter cell separation in Tetragenococcus halophilus. Biosci. Biotechnol. Biochem. 84, 1724–1735. https://doi.org/10.1080/09168451.2020.1764329

Sim, K.Y., Chye, F.Y., Anton, A., 2015. Chemical composition and microbial dynamics of *budu* fermentation, a traditional Malaysian fish sauce. Acta Aliment. 44, 185–194. https://doi.org/10.1556/AAlim.2014.0003

Singracha, P., Niamsiri, N., Visessanguan, W., Lertsiri, S., Assavanig, A., 2017. Application of lactic acid bacteria and yeasts as starter cultures for reduced-salt soy sauce (moromi)

fermentation. LWT 78, 181–188. https://doi.org/10.1016/j.lwt.2016.12.019

Sivamaruthi, B., Kesika, P., Prasanth, M., Chaiyasut, C., 2018. A Mini Review on Antidiabetic Properties of Fermented Foods. Nutrients 10, 1973. https://doi.org/10.3390/nu10121973

Solopova, A., Bachmann, H., Teusink, B., Kok, J., Kuipers, O.P., 2018. Further Elucidation of Galactose Utilization in *Lactococcus lactis* MG1363. Front. Microbiol. 9, 1–9. https://doi.org/10.3389/fmicb.2018.01803

Song, D.H., Chun, B.H., Lee, S., Son, S.Y., Reddy, C.K., Mun, H.I., Jeon, C.O., Lee, C.H., 2021. Comprehensive metabolite profiling and microbial communities of *Doenjang* (Fermented soy paste) and *Ganjang* (fermented soy sauce): A comparative study. Foods 10. https://doi.org/10.3390/foods10030641

Song, Y.R., Jeong, D.Y., Baik, S.H., 2015a. Monitoring of Yeast Communities and Volatile Flavor Changes During Traditional Korean Soy Sauce Fermentation. J. Food Sci. 80, 2005–2014. https://doi.org/10.1111/1750-3841.12995

Song, Y.R., Jeong, D.Y., Baik, S.H., 2015b. Effects of indigenous yeasts on physicochemical and microbial properties of Korean soy sauce prepared by low-salt fermentation. Food Microbiol. 51, 171–178. https://doi.org/10.1016/j.fm.2015.06.001

Sørensen, B.B., Jakobsen, M., 1997. The combined effects of temperature, pH and NaCl on growth of *Debaryomyces hansenii* analyzed by flow cytometry and predictive microbiology. Int. J. Food Microbiol. 34, 209–220. https://doi.org/10.1016/S0168-1605(96)01192-0

Stevenson, A., Cray, J.A., Williams, J.P., Santos, R., Sahay, R., Neuenkirchen, N., McClure, C.D., Grant, I.R., Houghton, J.D., Quinn, J.P., Timson, D.J., Patil, S. V., Singhal, R.S., Antón, J., Dijksterhuis, J., Hocking, A.D., Lievens, B., Rangel, D.E.N., Voytek, M.A., Gunde-Cimerman, N., Oren, A., Timmis, K.N., McGenity, T.J., Hallsworth, J.E., 2015. Is there a common water-activity limit for the three domains of life. ISME J. 9, 1333–1351. https://doi.org/10.1038/ismej.2014.219

Sulaiman, J., Gan, H.M., Yin, W.F., Chan, K.G., 2014. Microbial succession and the functional potential during the fermentation of Chinese soy sauce brine. Front. Microbiol. 5, 1–9. https://doi.org/10.3389/fmicb.2014.00556

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. U. S. A. 101, 11030–11035. https://doi.org/10.1073/pnas.0404206101

Tanaka, Y., Watanabe, J., Mogi, Y., 2012. Monitoring of the microbial communities involved in the soy sauce manufacturing process by PCR-denaturing gradient gel electrophoresis. Food Microbiol. 31, 100–106. https://doi.org/10.1016/j.fm.2012.02.005

Tatusov, R.L., Koonin, E. V., Lipman, D.J., 1997. A genomic perspective on protein families. Science (80-.). 278, 631–637. https://doi.org/10.1126/science.278.5338.631

Terán, L.C., Coeuret, G., Raya, R., Zagorec, M., Champomier-Vergès, M.C., Chaillou, S., 2018. Phylogenomic analysis of Lactobacillus curvatus reveals two lineages distinguished by genes for fermenting plant-derived carbohydrates. Genome Biol. Evol. 10, 1516–1525. https://doi.org/10.1093/gbe/evy106

Thomsen, M.C.F., Hasman, H., Westh, H., Kaya, H., Lund, O., 2017. RUCS: Rapid identification of PCR primers for unique core sequences. Bioinformatics 33, 3917–3921. https://doi.org/10.1093/bioinformatics/btx526

REFERENCES

Torino, M.I., de Valdez, G.F., Mozzi, F., 2015. Biopolymers from lactic acid bacteria. Novel applications in foods and beverages. Front. Microbiol. 6, 1–16. https://doi.org/10.3389/fmicb.2015.00834

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., Cox, J., 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods 13, 731–740. https://doi.org/10.1038/nmeth.3901

Uchida, K., 1982. Multiplicity in soy pediococci carbohydrate fermentation and its application for analysis of their flora. J. Gen. Appl. Microbiol. 28, 215–223. https://doi.org/10.2323/jgam.28.215

Uchida, K., Kanbe, C., 1993. Occurrence of bacteriophages lytic for *Pediococcus halophilus*, a halophilic lactic-acid bacterium, in soy sauce fermentation. J. Gen. Appl. Microbiol. 39, 429–437. https://doi.org/10.2323/jgam.39.429

Uchida, M., Miyoshi, T., Yoshida, G., Niwa, K., Mori, M., Wakabayashi, H., 2014. Isolation and characterization of halophilic lactic acid bacteria acting as a starter culture for sauce fermentation of the red alga *Nori* (*Porphyra yezoensis*). J. Appl. Microbiol. 116, 1506–1520. https://doi.org/10.1111/jam.12466

Udomsil, N., Chen, S., Rodtong, S., Yongsawatdigul, J., 2016. Quantification of viable bacterial starter cultures of *Virgibacillus* sp. and *Tetragenococcus halophilus* in fish sauce fermentation by real-time quantitative PCR. Food Microbiol. 57, 54–62. https://doi.org/10.1016/j.fm.2016.01.004

Udomsil, N., Rodtong, S., Choi, Y.J., Hua, Y., Yongsawatdigul, J., 2011. Use of *Tetragenococcus halophilus* as a Starter Culture for Flavor Improvement in Fish Sauce Fermentation. J. Agric. Food Chem. 59, 8401–8408. https://doi.org/10.1021/jf201953v

Unno, R., Matsutani, M., Suzuki, T., Kodama, K., Matsushita, H., Yamasato, K., Koizumi, Y., Ishikawa, M., 2020. Lactic acid bacterial diversity in Brie cheese focusing on salt concentration and pH of isolation medium and characterisation of halophilic and alkaliphilic lactic acid bacterial isolates. Int. Dairy J. 109, 104757. https://doi.org/10.1016/j.idairyj.2020.104757

Usinger, L., Ibsen, H., Jensen, L.T., 2009. Does fermented milk possess antihypertensive effect in humans? J. Hypertens. 27, 1115–1120. https://doi.org/10.1097/HJH.0b013e3283292716

Van der Walt, J.P., Taylor, M.B., Liebenberg, N.V.D.W., 1977. Ploidy, ascus formation and recombination in *Torulaspora* (*Debaryomyces*) *hansenii*. Antonie Van Leeuwenhoek 43, 205–218. https://doi.org/10.1007/BF00395675

Van Kerrebroeck, S., Comasio, A., Harth, H., De Vuyst, L., 2018. Impact of starter culture, ingredients, and flour type on sourdough bread volatiles as monitored by selected ion flow tube-mass spectrometry. Food Res. Int. 106, 254–262. https://doi.org/10.1016/j.foodres.2017.12.068

Vesth, T., Lagesen, K., Acar, Ö., Ussery, D., 2013. CMG-Biotools, a Free Workbench for Basic Comparative Microbial Genomics. PLoS One 8, e60120. https://doi.org/10.1371/journal.pone.0060120

Villar, M., de Ruiz Holgado, A.P., Sanchez, J.J., Trucco, R.E., Oliver, G., 1985. Isolation and characterization of *Pediococcus halophilus* from salted anchovies (*Engraulis anchoita*). Appl. Environ. Microbiol. 49, 664–666. https://doi.org/10.1128/aem.49.3.664-666.1985

Wagner, T., Fischer, M., 2001. Natural groups and a revised system for the European poroid *Hymenochaetales* (*Basidiomycota*) supported by nLSU rDNA sequence data. Mycol. Res. 105, 773–782. https://doi.org/10.1017/S0953756201004257

REFERENCES

Wakinaka, T., Iwata, S., Takeishi, Y., Watanabe, J., Mogi, Y., Tsukioka, Y., Shibata, Y., 2019. Isolation of halophilic lactic acid bacteria possessing aspartate decarboxylase and application to fish sauce fermentation starter. Int. J. Food Microbiol. 292, 137–143. https://doi.org/10.1016/j.ijfoodmicro.2018.12.013

Wakinaka, T., Matsutani, M., Watanabe, J., Mogi, Y., Tokuoka, M., Ohnishi, A., 2023. Identification of Capsular Polysaccharide Synthesis Loci Determining Bacteriophage Susceptibility in *Tetragenococcus halophilus*. Microbiol. Spectr. https://doi.org/10.1128/spectrum.00385-23

Wakinaka, T., Matsutani, M., Watanabe, J., Mogi, Y., Tokuoka, M., Ohnishi, A., 2022. Ribitol-Containing Wall Teichoic Acid of *Tetragenococcus halophilus* Is Targeted by Bacteriophage phiWJ7 as a Binding Receptor. Microbiol. Spectr. https://doi.org/10.1128/spectrum.00336-22

Wakinaka, T., Watanabe, J., 2019. Transposition of IS*4* Family Insertion Sequences IS*Teha3*, IS*Teha4*, and IS*Teha5* into the arc Operon Disrupts Arginine Deiminase System in Tetragenococcus halophilus. Appl. Environ. Microbiol. 85, 1–12. https://doi.org/10.1128/AEM.00208-19

Walker, G.M., Stewart, G.G., 2016. *Saccharomyces cerevisiae* in the production of fermented beverages. Beverages 2, 1–12. https://doi.org/10.3390/beverages2040030

Winters, M., Panayotides, D., Bayrak, M., Rémont, G., Viejo, C.G., Liu, D., Le, B., Liu, Y., Luo, J., Zhang, P., Howell, K., 2019. Defined co-cultures of yeast and bacteria modify the aroma, crumb and sensory properties of bread. J. Appl. Microbiol. 127, 778–793. https://doi.org/10.1111/jam.14349

Wu, C., Liu, C., He, G., Huang, J., Zhou, R., 2013. Characterization of a multiple-stress tolerance *Tetragenococcus halophilus* and application as starter culture in Chinese Horsebean-Chili-Paste manufacture for quality improvement. Food Sci. Technol. Res. 19, 855–864. https://doi.org/10.3136/fstr.19.855

Wüthrich, D., Wenzel, C., Bavan, T., Bruggmann, R., Berthoud, H., Irmler, S., 2018. Transcriptional regulation of cysteine and methionine metabolism in *Lactobacillus paracasei* FAM18149. Front. Microbiol. 9, 1–11. https://doi.org/10.3389/fmicb.2018.01261

Wyk, N. Van, Pretorius, I.S., Wallbrunn, C. Von, 2020. Assessing the oenological potential of *Nakazawaea ishiwadae*, *Candida railenensis* and *Debaryomyces hansenii* strains in mixed-culture grape must fermentation with *Saccharomyces cerevisiae*. Fermentation 6. https://doi.org/10.3390/FERMENTATION6020049

Xu, Z., Li, S., Gong, G., Liu, Z., Wu, Z., Ma, C., 2015. Influence of different acidifying strains of *Lactobacillus delbrueckii* subsp. *bulgaricus* on the quality of yoghurt. Food Sci. Technol. Res. 21, 263–269. https://doi.org/10.3136/fstr.21.263

Yan, Y.Z., Qian, Y.L., Ji, F. Di, Chen, J.Y., Han, B.Z., 2013. Microbial composition during Chinese soy sauce koji-making based on culture dependent and independent methods. Food Microbiol. 34, 189–195. https://doi.org/10.1016/j.fm.2012.12.009

Yang, X., Wu, J., An, F., Xu, J., Bat-Ochir, M., Wei, L., Li, M., Bilige, M., Wu, R., 2022. Structure characterization, antioxidant and emulsifying capacities of exopolysaccharide derived from *Tetragenococcus halophilus* SNTH-8. Int. J. Biol. Macromol. 208, 288–298. https://doi.org/10.1016/j.ijbiomac.2022.02.186

Zannini, E., Waters, D.M., Coffey, A., Arendt, E.K., 2016. Production, properties, and industrial food application of lactic acid bacteria-derived exopolysaccharides. Appl. Microbiol. Biotechnol. 100, 1121–1135. https://doi.org/10.1007/s00253-015-7172-2

REFERENCES

Zeidan, A.A., Poulsen, V.K., Janzen, T., Buldo, P., Derkx, P.M.F., Øregaard, G., Neves, A.R., 2017. Polysaccharide production by lactic acid bacteria: from genes to industrial applications. FEMS Microbiol. Rev. 41, S168–S200. https://doi.org/10.1093/femsre/fux017

Zhang, M., Zeng, S., Hao, L., Yao, S., Wang, D., Yang, H., Wu, C., 2022. Structural characterization and bioactivity of novel exopolysaccharides produced by *Tetragenococcus halophilus*. Food Res. Int. 155, 111083. https://doi.org/10.1016/j.foodres.2022.111083

Zhang, Y.F., Liu, S.Y., Du, Y.H., Feng, W.J., Liu, J.H., Qiao, J.J., 2014. Genome shuffling of *Lactococcus lactis* subspecies *lactis* YF11 for improving nisin Z production and comparative analysis. J. Dairy Sci. 97, 2528–2541. https://doi.org/10.3168/jds.2013-7238

9 Appendix

Table 21: *T. halophilus* and *D. hansenii* isolates used for the RAPD screening. The first column displays the in-house strain collection number (TMW) and the second column shows the corresponding temporary strain collection number (TMWRL.). The isolation source and time point can be found in third column.

TMW Nr.	TMWRL. Nr	Isolation source and time point
	TMW RL.1938	Moromi wild 10 % NaCl t8w
TMW 2.2254	TMW RL.1939	Moromi wild 10 % NaCl t8w
	TMW RL.1940	Moromi wild 10 % NaCl t8w
	TMW RL.1941	Moromi wild 10 % NaCl t8w
	TMW RL.1942	Moromi wild 10 % NaCl t12w
	TMW RL.1943	Moromi wild 10 % NaCl t12w
TMW 2.2255	TMW RL.1944	Moromi wild 10 % NaCl t12w
TMW 2.2256	TMW RL.1945	Moromi inoculated 10 % NaCl t2w
	TMW RL.1946	Moromi inoculated 10 % NaCl t2w
	TMW RL.1947	Moromi inoculated 10 % NaCl t2w
	TMW RL.1948	Moromi wild 13,5 % NaCl t12w
	TMW RL.1949	Moromi wild 13,5 % NaCl t12w
	TMW RL.1950	Moromi wild 13,5 % NaCl t12w
	TMW RL.1951	Moromi wild 13,5 % NaCl t12w
TMW 2.2257	TMW RL.1952	Moromi wild 13,5 % NaCl t12w
	TMW RL.1953	Moromi wild 13,5 % NaCl t12w
TMW 2.2258	TMW RL.1954	Moromi wild 13,5 % NaCl t12w
	TMW RL.1955	Moromi inoculated 13,5 % NaCl t2w
	TMW RL.1956	Moromi inoculated 13,5 % NaCl t2w
	TMW RL.1957	Moromi inoculated 13,5 % NaCl t2w
	TMW RL.1958	Moromi inoculated 13,5 % NaCl t2w
	TMW RL.1959	Moromi inoculated 13,5 % NaCl t2w
	TMW RL.1960	Moromi inoculated 13,5 % NaCl t2w
TMW 2.2259	TMW RL.1961	Moromi inoculated 13,5 % NaCl t2w
	TMW RL.1962	Moromi inoculated 13,5 % NaCl t2w
	TMW RL.1963	Moromi inoculated 20 % NaCl t2d
	TMW RL.1964	Buckwheat moromi
	TMW RL.1965	Buckwheat moromi
	TMW RL.1966	Buckwheat moromi
	TMW RL.1967	Buckwheat moromi
	TMW RL.1968	Buckwheat moromi
TMW 2.2260	TMW RL.1969	Buckwheat moromi
TMW 2.2261	TMW RL.1970	Buckwheat moromi
	TMW RL.1971	Moromi inoculated 10 % NaCl t8w
	TMW RL.1972	Moromi inoculated 10 % NaCl t8w
	TMW RL.1973	Moromi wild 13,5 % NaCl t8w
	TMW RL.1974	Moromi wild 13,5 % NaCl t8w
TMW 2.2262	TMW RL.1975	Moromi wild 13,5 % NaCl t8w
	TMW RL.1976	Moromi wild 13,5 % NaCl t8w
	TMW RL.1977	Moromi inoculated 13,5 % NaCl t8w
	TMW RL.1978	Moromi inoculated 13,5 % NaCl t8w
	TMW RL.1979	Moromi inoculated 13,5 % NaCl t8w
	TMW RL.1980	Moromi inoculated 13,5 % NaCl t8w
	TMW RL.1981	Moromi wild 15 % NaCl t12w
	TMW RL.1982	Moromi wild 15 % NaCl t12w
TMW 2.2263	TMW RL.1983	Moromi wild 15 % NaCl t12w
	TMW RL.1984	Moromi wild 15 % NaCl t12w
	TMW RL.1985	Moromi inoculated 15 % NaCl t2w
	TMW RL.1986	Moromi inoculated 15 % NaCl t2w
	TMW RL.1987	Moromi inoculated 15 % NaCl t2w
TMW 2.2264	TMW RL.1988	Moromi inoculated 15 % NaCl t2w
TMW 2.2265	TMW RL.1989	Moromi inoculated 15 % NaCl t2w
	TMW RL.1990	Moromi inoculated 15 % NaCl t2w
	TMW RL.1991	Moromi inoculated 15 % NaCl t12w
TMW 2.2266	TMW RL.1992	Moromi inoculated 15 % NaCl t12w

	TMW RL.1993	Moromi inoculated 15 % NaCl t12w
	TMW RL.1994	Moromi inoculated 15 % NaCl t12w
	TMW RL.1995	Moromi inoculated 15 % NaCl t8w
	TMW RL.1996	Moromi inoculated 15 % NaCl t8w
	TMW RL.1997	Moromi inoculated 15 % NaCl t8w
	TMW RL.1998	Moromi inoculated 20 % NaCl t12w
	TMW RL.1999	Moromi inoculated 20 % NaCl t12w
	TMW RL.19100	Moromi inoculated 20 % NaCl t12w
	TMW RL.19101	Moromi wild 20 % NaCl t12w
	Debaryomyces hansenii isolates	
	TMW RL.19118	Moromi inoculated 15 % NaCl t2w
	TMW RL.19119	Moromi inoculated 15 % NaCl t2w
	TMW RL.19120	Moromi inoculated 15 % NaCl t2w
	TMW RL.19121	Moromi inoculated 15 % NaCl t2w
	TMW RL.19122	Moromi inoculated 13,5 % NaCl t8w
	TMW RL.19123	Moromi inoculated 13,5 % NaCl t8w
	TMW RL.19124	Moromi inoculated 13,5 % NaCl t8w
	TMW RL.19125	Moromi inoculated 13,5 % NaCl t12w
	TMW RL.19126	Moromi inoculated 13,5 % NaCl t12w
	TMW RL.19127	Moromi inoculated 13,5 % NaCl t12w
	TMW RL.19128	Moromi inoculated 13,5 % NaCl t12w
	TMW RL.19129	Moromi wild 20 % NaCl t12w
	TMW RL.19130	Moromi wild 20 % NaCl t12w
	TMW RL.19131	Moromi wild 20 % NaCl t12w
	TMW RL.19132	Moromi wild 20 % NaCl t12w
	TMW RL.19133	Moromi wild 13,5 % NaCl t2w
	TMW RL.19134	Moromi wild 13,5 % NaCl t2w
	TMW RL.19135	Moromi wild 13,5 % NaCl t2w
	TMW RL.19136	Moromi wild 13,5 % NaCl t2w
	TMW RL.19137	Moromi inoculated 10 % NaCl t2w
	TMW RL.19138	Moromi inoculated 10 % NaCl t2w
	TMW RL.19139	Moromi inoculated 10 % NaCl t2w
	TMW RL.19140	Moromi inoculated 10 % NaCl t2w
	TMW RL.19141	Moromi wild 10 % NaCl t7d
	TMW RL.19142	Moromi wild 10 % NaCl t12w
TMW 3.1188	TMW RL.19143	Moromi wild 10 % NaCl t12w

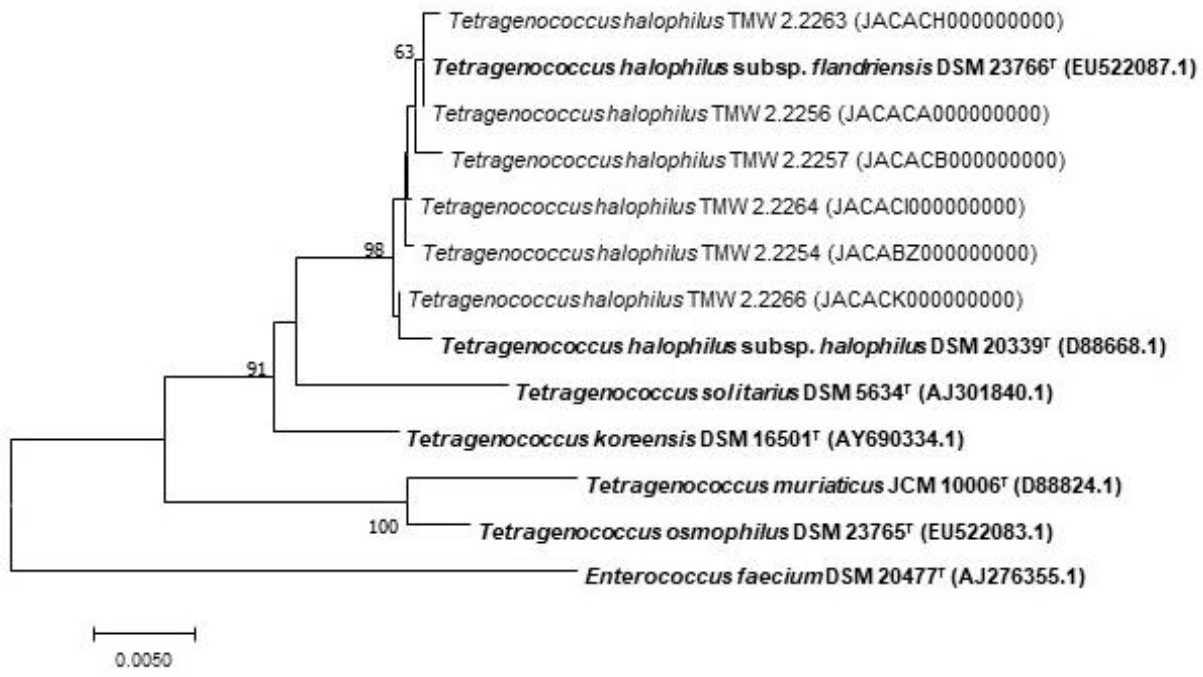

Figure 24: Phylogenetic tree of *T. halophilus* 16S rRNA gene sequences. The type strain of every species was included. The tree was constructed using the Neighbor-Joining method (Saitou and Nei, 1987). The accession number of the 16S rRNA of the type strains is given in brackets. For the newly isolated strains, the genome accession number is present in brackets. The optimal tree with the sum of branch length = 0.09391473 is shown. Bootstrap values (1000 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 1214 positions in the final dataset. The type strain of *Enterococcus faecium* DSM 20477^T served as an outgroup. Adapted from Link et al., 2021.

APPENDIX

Table 22: Calculated ANIb values for every *T. halophilus* strain sequenced in this thesis.

	TMW 2.2258	TMW 2.2259	TMW 2.2262	TMW 2.2261	TMW 2.2256	TMW 2.2263	TMW 2.2260	TMW 2.2265	TMW 2.2257	TMW 2.2255	TMW 2.2266	TMW 2.2254	TMW 2.2264	DSM20337
TMW 2.2258	*													
TMW 2.2259	97.68	*												
TMW 2.2262	100	97.51	*											
TMW 2.2261	97.71	99.99	97.69	*										
TMW 2.2256	99.98	97.49	99.98	97.49	*									
TMW 2.2263	97.83	97.22	97.83	97.22	97.81	*								
TMW 2.2260	98.27	97.81	98.27	97.81	98.27	97.85	*							
TMW 2.2265	99.98	97.43	99.98	97.43	99.94	97.57	98.01	*						
TMW 2.2257	97.89	97.58	97.89	97.58	97.9	97.7	98.03	97.87	*					
TMW 2.2255	97.63	100	97.63	100	97.67	97.44	97.83	97.67	97.71	*				
TMW 2.2266	97.53	97.11	97.52	97.11	97.45	98.77	97.83	97.5	97.38	97.09	*			
TMW 2.2254	97.88	97.84	97.87	97.84	97.87	97.57	97.7	97.88	97.89	97.82	97.47	*		
TMW 2.2264	97.65	100	97.66	99.99	97.7	97.41	97.83	97.69	97.75	99.99	97.38	97.98	*	
DSM20337	97.95	98.16	97.95	98.16	97.95	97.89	98.01	97.95	98.22	98.15	97.6	98.37	98.15	*

Table 23: BADGE analysis output filtered for ORFs specific lineage I.

Annotation	Reference ORF	8C7	DSM 20337	MJ4	YJ1	KUD23	NISL 7118	FBL3	TMW 2.2254	TMW 2.2256	TMW 2.2257	TMW 2.2263	TMW 2.2264	TMW 2.2266
fumarate reductase flavoprotein subunit	Tetragenococcus_halophilus_8C7_TH6N_06470	+	-	+	+	+	+	+	+	+	+	+	-	+
hypothetical protein	Tetragenococcus_halophilus_8C7_TH6N_09370	+	-	+	+	-	+	+	+	+	+	+	+	+
cellobiose-specific PTS system IIC component	Tetragenococcus_halophilus_8C7_TH6N_16770	+	+	+	+	+	+	-	-	+	+	+	+	+
periplasmic binding protein _ LacI transcriptional regulator	Tetragenococcus_halophilus_8C7_TH6N_16780	+	+	+	+	+	+	-	-	+	+	+	+	+
alpha-L-fucosidase	Tetragenococcus_halophilus_8C7_TH6N_16790	+	+	+	+	+	+	-	-	+	+	+	+	+
betaine_carnitine_choline ABC transporter substrate-binding protein	Tetragenococcus_halophilus_8C7_TH6N_21790	+	-	+	+	-	+	+	+	+	+	+	+	+

Table 24: BADGE analysis output filtered for ORFs specific lineage II.

Annotation	Reference ORF	KG12	11	WJ7	DSM 20339^T	NISL 7126	D-86	D10	NBRC 12172	YA163	YG2	YA5
site-specific integrase	Tetragenococcus_halophilus_KG12_HLV32_10265	+	+	+	+	+	+	+	+	+	+	+
group-specific protein	Tetragenococcus_halophilus_KG12_HLV32_10270	+	+	+	+	+	+	+	+	+	+	+
hypothetical protein	Tetragenococcus_halophilus_KG12_HLV32_10275	+	+	+	+	+	+	+	+	+	+	+
hypothetical protein	Tetragenococcus_halophilus_KG12_HLV32_10280	+	+	+	+	+	+	+	+	+	+	+
Hsp33 family molecular chaperone HslO	Tetragenococcus_halophilus_KG12_HLV32_10290	+	+	+	+	+	+	+	+	+	+	+
MerR family transcriptional regulator	Tetragenococcus_halophilus_KG12_HLV32_10295	+	+	+	+	+	+	+	+	+	+	+
CopG family transcriptional regulator	Tetragenococcus_halophilus_KG12_HLV32_10300	+	+	+	+	+	+	+	+	+	+	+
type II toxin-antitoxin system death-on-curing family toxin	Tetragenococcus_halophilus_KG12_HLV32_10320	+	+	+	+	+	+	+	+	+	+	+
helix-turn-helix transcriptional regulator	Tetragenococcus_halophilus_KG12_HLV32_10325	+	+	+	+	+	+	+	+	+	+	+
alpha_beta hydrolase	Tetragenococcus_halophilus_KG12_HLV32_12445	+	+	+	+	+	+	+	+	+	+	+

Table 25: BADGE analysis output filtered for ORFs specific to some strains lineage II.

Annotation	Reference ORF	KG12	11	WJ7	DSM 20339^T	NISL 7126	D-86	D10	NBRC 12172	YA163	YG2	YA5
sugar ABC transporter sugar-binding protein	Tetragenococcus_halophilus_NBRC_12172_TEH_04330	-	-	+	-	+	+	+	+	-	-	-
sugar ABC transporter ATP-binding protein	Tetragenococcus_halophilus_NBRC_12172_TEH_04340	-	-	+	-	+	+	+	+	-	-	-
sugar ABC transporter permease protein	Tetragenococcus_halophilus_NBRC_12172_TEH_04350	-	-	+	-	+	+	+	+	-	-	-

APPENDIX

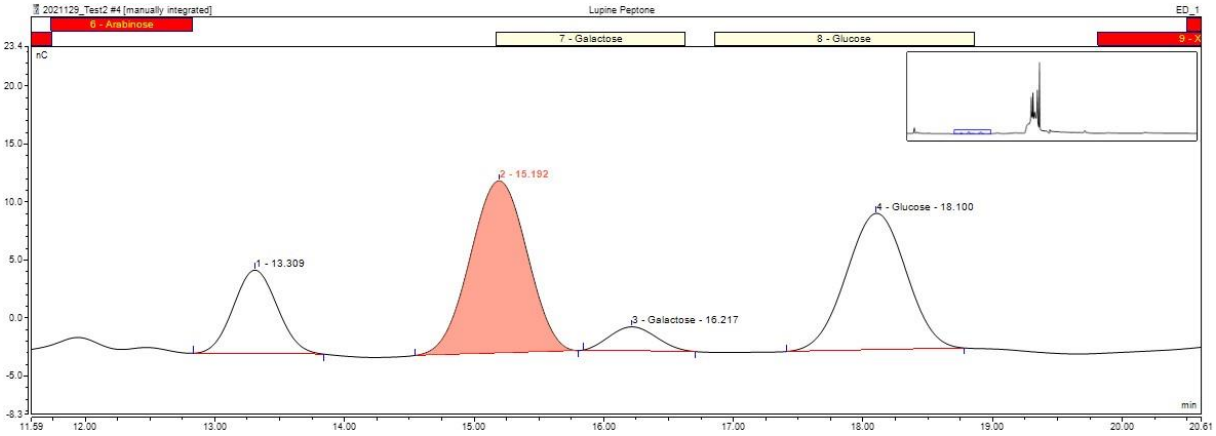

Figure 25: Qualitative measurement of LMRS using HPAEC-PAD. LMRS separated and analyzed using an HPAEC-PAD.

Table 26: DEGs from TMW 2.2254 and TMW 2.2256 not assigned to a specific substrate or pathway. Column one shows the annotation of the CDS according to the NCBI PGAP. Columns two and three the locus tag in the respective strains, missing of a CDS is represented by an "X". Columns four and five show the fold change DEGs within the given strains missing CDS are represented by an "X". Cutoffs: Log2 Foldchange of ≥ 2 or ≤ - 2, p-value ≤ 0.05 and FDR ≤ 0.01.

NCBI annotation	Locus tag TMW 2.2254	Locus tag TMW 2.2256	TMW 2.2254	TMW 2.2256
IS4/IS5 family transposase	X	HXW74_02125	X	3.09
metal ABC transporter substrate-binding protein	X	HXW74_07330	X	2.72
IS4/IS5 family transposase	X	HXW74_07980	X	2.63
IS4 family transposase	X	HXW74_08130	X	2.46
methanol dehydrogenase	HV360_05425	HXW74_11805		2.34
50S ribosomal protein L32	HV360_05255	HXW74_04645		2.22
hypothetical protein	HV360_03820	HXW74_06270		2.19
PTS glucitol transporter subunit IIA	HV360_07350	HXW74_08030		2.18
acyclic terpene utilization AtuA family protein	HV360_06800	HXW74_05390		2.10
DUF3042 family protein	HV360_04895	HXW74_07405		2.09
ThuA domain-containing protein	HV360_03730	HXW74_06360		2.05
hypothetical protein	HV360_10740	HXW74_10680		-2.05
ECF transporter S component	HV360_07690	HXW74_07100		-2.07
hypothetical protein	X	HXW74_11650	X	-2.15
hypothetical protein	HV360_05405	HXW74_11760		-2.30
acyltransferase family protein	HV360_02430	HXW74_06460		-2.40
LysM peptidoglycan-binding domain-containing protein	X	HXW74_09465	X	-2.54
hypothetical protein	HV360_06520	HXW74_03145		-2.58
GNAT family N-acetyltransferase	HV360_07340	HXW74_08020	2.73	
ASCH domain-containing protein	HV360_07130	HXW74_08625	2.70	
macro domain-containing protein	HV360_09385	HXW74_07310	2.68	
DEAD/DEAH box helicase	HV360_03000	HXW74_03970	2.59	
iron-containing alcohol dehydrogenase family protein	HV360_07140	HXW74_08635	2.50	
hypothetical protein	HV360_00615	HXW74_00265	2.43	
YbaK/EbsC family protein	HV360_00620	HXW74_00270	2.35	

DUF4387 domain-containing protein	HV360_06795	HXW74_05385	2.24		3.21
chromate transporter	HV360_11785	X	2.15	X	
cupin domain-containing protein	HV360_03695	HXW74_06395	2.14		
hypothetical protein	HV360_00380	HXW74_00515	2.01		
hypothetical protein	HV360_05440	HXW74_11895	-2.01		
beta-ketoacyl-ACP synthase II	HV360_04655	HXW74_05845	-2.02		
hypothetical protein	HV360_04145	HXW74_01610	-2.06		
endonuclease MutS2	HV360_04175	HXW74_01640	-2.07		
addiction module antitoxin	HV360_02190	HXW74_10190	-2.09		
hypothetical protein	HV360_02635	X	-2.10	X	
hypothetical protein	HV360_07005	HXW74_10965	-2.11		
tRNA-dihydrouridine synthase	HV360_08225	HXW74_06220	-2.13		
DEAD/DEAH box helicase	HV360_09220	HXW74_07585	-2.14		
helix-turn-helix transcriptional regulator	HV360_04525	HXW74_01990	-2.16		
group II intron reverse transcriptase/maturase	HV360_11775	X	-2.23	X	
ATP-binding cassette domain-containing protein	HV360_11345	HXW74_07325	-2.32		-2.74
PTS sugar transporter subunit IIC	HV360_11285	X	-2.34	X	
LysM peptidoglycan-binding domain-containing protein	HV360_04930	HXW74_07440	-2.38		
hypothetical protein	HV360_09245	HXW74_11650+07560	-2.38		-2.15
type I toxin-antitoxin system Fst family toxin	HV360_08335	HXW74_06110	-2.42		-2.42
ATP-binding protein	HV360_08855	X	-2.80	X	
TetR family transcriptional regulator	HV360_01245	HXW74_04960	-2.88		
YitT family protein	HV360_10145	HXW74_11105	-2.94		
FtsX-like permease family protein	HV360_02825	HXW74_10820	-3.09		
ABC transporter ATP-binding protein	HV360_01250	HXW74_04955	-3.13		
transglycosylase family protein	HV360_08140	X	-3.27	X	
ABC transporter ATP-binding protein	HV360_02830	HXW74_10815	-3.28		
YaiI/YqxD family protein	HV360_09235	HXW74_07570	-3.39		
D-alanyl-D-alanine carboxypeptidase family protein	HV360_09240	HXW74_07565	-3.42		
ABC-F family ATP-binding cassette domain-containing protein	HV360_09370	HXW74_07325	-3.42		-2.74
ABC transporter ATP-binding protein	HV360_01255	HXW74_04950	-3.59		-2.10
transglycosylase family protein	HV360_08110	HXW74_09435	-4.42		-3.82

10 Acknowledgements

The work in this thesis was part of the project “Entwicklung regionaler Bio-Würzsaucen auf Lupinenbasis als salzreduzierte, glutenfreie Alternative zu Sojaprodukten” funded by the German Ministry of Food and Agriculture (BMEL) under the projectname “Flavorloop” and the project number 28-1-A4.001-17.

First and foremost, I would like to thank Prof. Dr. Matthias A. Ehrmann for supervising my thesis and giving me the opportunity to work on this project. In particular, I would like to mention the many hours of discussions about scientific theses, experimental ideas and microbiology in general.

I want to thank Prof. Dr. J. Philipp Benz for being my second examiner and for giving me the opportunity to do some experiments on his HPAEC.

I also want to thank Prof. Dr. Rudi F. Vogel for mentoring my work and acquiring the funding for this project, as well as for his insightful advice on microbiology.

Furthermore, I want to thank all the people from the entire chair but especially from my office. Whether it was throwing footballs or Frisbees while discussing ideas or even drinking way too much coffee, it’s been fun with you all.

Last but not least, I want to thank my girlfriend for her support and sometimes guidance if my ideas were a bit too extravagant. In you I have found a person who is as enthusiastic about science as I am, I love you.

Finally, I want to thank my family and my friends for continuously supporting my journey.

11 Publications, presentations, posters

Link, T., Vogel, R.F., Ehrmann, M.A., 2021. The diversity among the species *Tetragenococcus halophilus* including new isolates from a lupine seed fermentation. BMC Microbiol. 21, 320.
https://doi.org/10.1186/s12866-021-02381-1

Link, T., Lülf, R.H., Parr, M., Hilgarth, M., Ehrmann, M.A., 2022. Genome Sequence of the Diploid yeast *Debaryomyces hansenii* TMW 3.1188. Microbiol. Resour. Announc. 11, 5-7.
https://doi.org/10.1128/mra.00649-22

Link, T., Ehrmann, M.A., 2023. The diversity among the species *Tetragenococcus halophilus* including new isolates from a lupine seed fermentation. BMC Microbiol. 23, 14.
https://doi.org/10.1186/s12866-023-02760-w

Link, T., Ehrmann, M.A., 2023. Monitoring the growth dynamics of *Tetragenococcus halophilus* strains in lupine moromi fermentation using a multiplex-PCR system. BMC Res. Notes 16, 115.

https://doi.org/10.1186/s13104-023-06406-y

Poster presentation at the Lactic Acid Bacteria Conference 2021

Link, T., apl., Prof. Dr. Matthias A. Ehrmann, Prof. Dr. Rudi F. Vogel, 2021. Genomic analysis and metabolic traits of *Tetragenococcus halophilus* isolates from lupin moromi.