Ref. Ares(2024)1585311 - 29/02/2024

DOMINO

Funded by
the European Union

DELIVERABLE

D.4.1 – First data on food metagenomic and genomic diversity of new microbial strains

This project has received funding from the European Union's Horizon Europe
research and innovation programme under grant agreement No 101060218

DOMINO

Data on food metagenomic and genomic diversity of new microbial strains

Deliverable Nr.	D4.1
Due date	29 February 2024
Submission date	29 February 2024
Deliverable type	DATA
Dissemination level	Public
Work package	WP4
Author(s)	Stephane Chaillou, Luca Cocolin, Abelardo Margolles, Kalliopi Rantsiou

Document version	1.0		
Grant agreement	101060218	**Duration**	60 months
Start Date	March 2023	**End date**	February 2028

Contributors

NAME	ORGANISATION
Kalliopi Rantsiou	UNITO
Ilario Ferrocino	UNITO
Luca Cocolin	UNITO
Abelardo Margolles	CSIC
Lorena Ruiz	CSIC
John Kenny	TEAGASC
Fiona Crispie	TEAGASC
Jekaterina Kazantseva	TFTAK
Rain Kuldjärv	TFTAK
Pascal Bonnarme	INRAE
Vincent Herve	INRAE
Christèle Humblot	IRD
Federica Pinto	UNITN
Nicola Segata	UNITN
Vitor Heidrich	UNITN
Tanja Kostic	AIT

Revision history

VERSION	DATE	REVIEWER	MODIFICATIONS
1	22/02/24	Kalliopi Rantsiou	First draft
2	23/02/24	Abelardo Margolles	Modifications
3	24/02/2024	Luca Cocolin	Modifications
4	26/02/2024	Stephane Chaillou	Modifications
5	26/02/2024	Kalliopi Rantsiou	Modifications

Table of content

1. Executive summary

The overall objective of Task 4.1 is to leverage microbial genomic resources from sustainable raw materials and fermented foods. This deliverable is a fundamental step in this direction. The deliverable reports on data that originate from 6 fermented food case studies; namely legumes, table olives, vegetables, cereals, apple pomace and water kefir. During the course of the first 12 months of the DOMINO project, partners (AIT, CSIC, INRAE, IRD, TEAGASC, TFTAK, and UNITO) were engaged in conducting fermentations (under laboratory or industrial conditions) and performing analyses that led to the collection of biological resources, in the form of **microbial isolates,** and genetic material in the form of **total DNA** extracted from the samples. Microbial isolates were identified to species level and characterized phenotypically (relevant characters were investigated based on the fermented food source) and genetically (to verify those isolates that represent individual strains and not clonal populations). A total of 193 strains of bacteria and yeasts were then chosen by partners and subjected to Whole Genome Sequencing (performed by UNITN). In parallel, 188 DNA samples were subjected to shot-gun sequencing (performed by UNITN). This deliverable summarizes the information regarding the sequencing performed. The sequencing data will feed Task 4.2 activities on *in-silico* analysis of genomes/metagenomes to design tailor-made microbial solutions for the fermented food case studies.

2. Introduction

In order to design microbial solutions for healthy and sustainable fermented foods, the DOMINO consortium is harnessing the potential of sequencing and *in silico* analyses to gain unprecedented information and new knowledge on the microbiome of novel and long-established, plant-based fermented foods. The new knowledge/information will drive the decision process for the construction of microbial consortia to be used in fermented foods that respond to the need of augmented, data-supported health benefits and sustainable processes.

3. Approach employed

The approach employed to reach the objectives, outlined above, initiated with extensive sampling of spontaneous fermentation processes performed industrially (table olives and vegetables such as kohlrabi, baby carrot, white cabbage, kimchi) or at lab scale (apple pomace and water kefir). Also, commercial fermented products were included in the analyses (plant-based cheeses, kefir, cereals). Samples were subjected to microbiological analyses that were designed based on prior knowledge regarding the nature of the raw material, the characteristics of the fermentation process and the fermented products. Metadata were collected for all samples according to a pre-defined template. The output of the sampling led to isolation of yeasts and bacteria, putative candidates for the design of microbial consortia to be tested in subsequent steps and in tasks 4.2 to 4.5. Isolates were identified to species level by either rRNA gene encoding sequencing or by MALDI. In addition, some isolates were also characterized genotypically and phenotypically, focusing on characters relevant to the raw material and/or fermentation process with which they may be associated in subsequent tests. For this purpose, partners of WP4 discussed internally and agreed to perform some common phenotypic tests *in vitro* to characterize the isolates. Partners also included case-specific tests if deemed necessary. In addition, partners interrogated strain databases already available. Microorganisms from the microbial collections of the partners involved in WP4 were also used. In this way, additional potential candidates were identified that converged with the new isolates in a collection of DOMINO strains destined for Whole Genome Sequencing (WGS). In total, 193 strains were sequenced; 116 bacteria and 77 yeasts. This deliverable reports the sequencing data of the 193 strains while it is foreseen that additional strains will be sequenced in the following months. WGS was performed by Illumina technology while it is planned that the same strains will undergo long-read sequencing by Nanopore technology and sequencing data will be compared.

Aliquots of 188 samples collected during spontaneous fermentations and/or fermented products, coming from the six different case studies, were also subjected to DNA extraction for subsequent shotgun sequencing. Shotgun sequencing and WGS were performed by UNITN.

Table 1 reports the numbers of strains and metagenomes that were sequenced for each case study.

Table 1 – Summary of numbers of strains and metagenomes sequenced for each case study

Partner	Fermented Food Case Study	Strains subjected to Whole Genome Sequencing (WGS)		Metagenomes
		Bacteria	Yeasts	
INRAE	Legumes	20	-	24
UNITO	Table Olives	-	53	24
TFTAK	Vegetables	40	-	37
IRD	Cereals	10	-	40
CSIC	Apple Pomace	22	20	37
TEAGASC	Water Kefir	24	4	26
TOTAL		116	77	188

4. Output

Following sequencing, the data regarding the output were summarized and are reported in the following tables (2-7; metagenomes, 8-13; strains).

Table 2. Data of sequencing run for metagenomes of samples of legumes (commercial plant-based cheeses from the European market) – partner INRAE

sampleid	n_raw_reads_total	n_raw_reads_R	min_read_len_	median_read_le	max_read_len_	n_raw_reads_R	min_read_len_	median_read_le	max_read_len_	n_preproc_read	min_read_len_p	median_read_le	max_read_len_preproc
DO_INRAE_PBC-01	47217288	23608644	35	151.0	151	23608644	35	151.0	151	46938338	75	150.0	151
DO_INRAE_PBC-02	64087528	32043764	35	151.0	151	32043764	35	151.0	151	63651910	75	150.0	151
DO_INRAE_PBC-03	19375882	9687941	35	151.0	151	9687941	35	151.0	151	18826625	75	150.0	151
DO_INRAE_PBC-04	72223776	36111888	35	151.0	151	36111888	35	151.0	151	70800653	75	150.0	151
DO_INRAE_PBC-05	67775208	33887604	35	151.0	151	33887604	35	151.0	151	66909992	75	150.0	151
DO_INRAE_PBC-06	27482010	13741005	35	151.0	151	13741005	35	151.0	151	27008246	75	150.0	151
DO_INRAE_PBC-07	145364700	72682350	35	151.0	151	72682350	35	151.0	151	144825232	75	150.0	151
DO_INRAE_PBC-08	57888950	28944475	35	151.0	151	28944475	35	151.0	151	56744772	75	150.0	151
DO_INRAE_PBC-09	34069730	17034865	35	151.0	151	17034865	35	151.0	151	33342094	75	150.0	151
DO_INRAE_PBC-10	81635236	40817618	35	151.0	151	40817618	35	151.0	151	81149202	75	150.0	151
DO_INRAE_PBC-11	46086756	23043378	35	151.0	151	23043378	35	151.0	151	45779328	75	150.0	151
DO_INRAE_PBC-12	30941064	15470532	35	151.0	151	15470532	35	151.0	151	30557838	75	150.0	151
DO_INRAE_PBC-13	51246312	25623156	35	151.0	151	25623156	35	151.0	151	50672100	75	150.0	151
DO_INRAE_PBC-14	73696838	36848419	35	151.0	151	36848419	35	151.0	151	72953081	75	150.0	151
DO_INRAE_PBC-15	168324932	84162466	35	151.0	151	84162466	35	151.0	151	164893983	75	150.0	151
DO_INRAE_PBC-16	115901076	57950538	35	151.0	151	57950538	35	151.0	151	114133313	75	150.0	151
DO_INRAE_PBC-17	53308902	26654451	35	151.0	151	26654451	35	151.0	151	53057049	75	150.0	151
DO_INRAE_PBC-18	49593540	24796770	35	151.0	151	24796770	35	151.0	151	49114246	75	150.0	151
DO_INRAE_PBC-19	43184656	21592328	35	151.0	151	21592328	35	151.0	151	42720499	75	150.0	151
DO_INRAE_PBC-20	68571352	34285676	35	151.0	151	34285676	35	151.0	151	66784332	75	150.0	151
DO_INRAE_PBC-21	75614072	37807036	35	151.0	151	37807036	35	151.0	151	74511082	75	150.0	151
DO_INRAE_PBC-22	48636264	24318132	35	151.0	151	24318132	35	151.0	151	47820448	75	150.0	151
DO_INRAE_PBC-23	56640800	28320400	35	151.0	151	28320400	35	151.0	151	55476381	75	150.0	151
DO_INRAE_PBC-24	91079258	45539629	35	151.0	151	45539629	35	151.0	151	89578191	75	150.0	151

Table 3. Data of sequencing run for metagenomes of samples of table olives (during fermentation) – partner UNITO

sampleid	n_raw_reads_t	n_raw_reads_R	min_read_len_	median_read_le	max_read_len_	n_raw_reads_R	min_read_len_	median_read_le	max_read_len_	n_preproc_read	min_read_len_p	median_read_le	max_read_len_preproc
DO_UNITO_B4	1913634	956817	35	151.0	151	956817	35	151.0	151	1890743	75	150.0	151
DO_UNITO_B6	31856958	15928479	35	151.0	151	15928479	35	151.0	151	30784114	75	150.0	151
DO_UNITO_B7	1051186	525593	35	151.0	151	525593	35	151.0	151	1040832	75	150.0	151
DO_UNITO_B8D	1421940	710970	35	151.0	151	710970	35	151.0	151	1401181	75	150.0	151
DO_UNITO_B9D	3501770	1750885	35	151.0	151	1750885	35	151.0	151	3485516	75	150.0	151
DO_UNITO_O1	989792	494896	35	151.0	151	494896	35	151.0	151	968774	75	150.0	151
DO_UNITO_O11	34324	17162	35	151.0	151	17162	35	151.0	151	32958	75	150.0	151
DO_UNITO_O3	858550	429275	35	151.0	151	429275	35	151.0	151	845028	75	150.0	151
DO_UNITO_O6	599568	299784	35	151.0	151	299784	35	151.0	151	592177	75	150.0	151
DO_UNITO_O7	788422	394211	35	151.0	151	394211	35	151.0	151	770453	75	150.0	151
DO_UNITO_O8	630000	315000	35	151.0	151	315000	35	151.0	151	613034	75	150.0	151
DO_UNITO_O9D	3965950	1982975	35	151.0	151	1982975	35	151.0	151	3899139	75	150.0	151
DO_UNITO_S2	4304176	2152088	35	151.0	151	2152088	35	151.0	151	4283892	75	150.0	151
DO_UNITO_S4	5320412	2660206	35	151.0	151	2660206	35	151.0	151	5293142	75	150.0	151
DO_UNITO_S6	2745342	1372671	35	151.0	151	1372671	35	151.0	151	2724479	75	150.0	151
DO_UNITO_S7	10639322	5319661	35	151.0	151	5319661	35	151.0	151	10595371	75	150.0	151
DO_UNITO_S8	75190118	37595059	35	151.0	151	37595059	35	151.0	151	73663224	75	150.0	151
DO_UNITO_S9	15954684	7977342	35	151.0	151	7977342	35	151.0	151	15900442	75	150.0	151
DO_UNITO_V1	624042	312021	35	151.0	151	312021	35	151.0	151	614971	75	150.0	151
DO_UNITO_V2	135098	67549	35	151.0	151	67549	35	151.0	151	129136	75	150.0	151
DO_UNITO_V3	1177028	588514	35	151.0	151	588514	35	151.0	151	1157686	75	150.0	151
DO_UNITO_V7D	493568	246784	35	151.0	151	246784	35	151.0	151	488807	75	150.0	151
DO_UNITO_V8	104264	52132	35	151.0	151	52132	35	151.0	151	101818	75	150.0	151
DO_UNITO_VS6	182374	91187	35	151.0	151	91187	35	151.0	151	173469	75	150.0	151

Table 4. Data of sequencing run for metagenomes of samples of fermented vegetables (baby carrot, sauerkraut, kohlrabi, kimchi)– partner TFTAK

sampleid	n_raw_reads_t	n_raw_reads_R	min_read_len_	median_read_l	max_read_len_	n_raw_reads_R	min_read_len_	median_read_l	max_read_len_	n_preproc_read	min_read_len_	median_read_l	max_read_len_preproc	
DO_TFTAK_BP_1D	92379482	46189741	35	151.0	151	46189741	35	151.0	151	91986771	75	151.0	151	
DO_TFTAK_BP_3D	69758976	34879488	35	151.0	151	34879488	35	151.0	151	69510293	75	151.0	151	
DO_TFTAK_BP_6D	116925544	58462772	35	151.0	151	58462772	35	151.0	151	114886857	75	150.0	151	
DO_TFTAK_HK_11D	88690748	44345374	35	151.0	151	44345374	35	151.0	151	87293801	75	150.0	151	
DO_TFTAK_HK_1D	135519068	67759534	35	151.0	151	67759534	35	151.0	151	134386243	75	150.0	151	
DO_TFTAK_HK_4D	49576352	24788176	35	151.0	151	24788176	35	151.0	151	49098551	75	150.0	151	
DO_TFTAK_HK_8D	98878854	49439427	35	151.0	151	49439427	35	151.0	151	97848243	75	150.0	151	
DO_TFTAK_Hapu_11D	20482746	10241373	35	151.0	151	10241373	35	151.0	151	20412291	75	150.0	151	
DO_TFTAK_Hapu_14D	55778544	27889272	35	151.0	151	27889272	35	151.0	151	55483015	75	150.0	151	
DO_TFTAK_Hapu_4D	43073444	21536722	35	151.0	151	21536722	35	151.0	151	42832343	75	150.0	151	
DO_TFTAK_Hapu_7D	54881966	27440983	35	151.0	151	27440983	35	151.0	151	54665199	75	150.0	151	
DO_TFTAK_HiKa_11D	120769498	60384749	35	151.0	151	60384749	35	151.0	151	120349027	75	150.0	151	
DO_TFTAK_HiKa_14D	108016604	54008302	35	151.0	151	54008302	35	151.0	151	107635189	75	150.0	151	
DO_TFTAK_HiKa_4D	74861056	37430528	35	151.0	151	37430528	35	151.0	151	74596882	75	151.0	151	
DO_TFTAK_HiKa_6D	69288504	34644252	35	151.0	151	34644252	35	151.0	151	69050128	75	150.0	151	
DO_TFTAK_HiKa_8D	51644436	25822218	35	151.0	151	25822218	35	151.0	151	51314796	75	150.0	151	
DO_TFTAK_Kim_11D	23024394	11512197	35	151.0	151	11512197	35	151.0	151	22496766	75	150.0	151	
DO_TFTAK_Kim_1D	62275574	31137787	35	151.0	151	31137787	35	151.0	151	61511066	75	150.0	151	
DO_TFTAK_Kim_4D	5040480	2520240	35	151.0	151	2520240	35	151.0	151	5012345	75	150.0	151	
DO_TFTAK_NK_1D	43465344	21732672	35	151.0	151	21732672	35	151.0	151	43213320	75	150.0	151	
DO_TFTAK_NK_3D	68573626	34286813	35	151.0	151	34286813	35	151.0	151	68170318	75	150.0	151	
DO_TFTAK_NK_6D	101586846	50793423	35	151.0	151	50793423	35	151.0	151	99126895	75	151.0	151	
DO_TFTAK_PkKa_11D	122445520	61222760	35	151.0	151	61222760	35	151.0	151	122018322	75	150.0	151	
DO_TFTAK_PkKa_14D	27203142	13601571	35	151.0	151	13601571	35	151.0	151	27102560	75	150.0	151	
DO_TFTAK_PkKa_4D	127878836	63939418	35	151.0	151	63939418	35	151.0	151	127260498	75	150.0	151	
DO_TFTAK_PkKa_6D	18686806	9343403	35	151.0	151	9343403	35	151.0	151	18626175	75	150.0	151	
DO_TFTAK_PkKa_8D	36099950	18049975	35	151.0	151	18049975	35	151.0	151	35923904	75	150.0	151	
DO_TFTAK_PkPp_11D	129618260	64809130	35	151.0	151	64809130	35	151.0	151	129067560	75	150.0	151	
DO_TFTAK_PkPp_14D	37371566	18685783	35	151.0	151	18685783	35	151.0	151	37193587	75	150.0	151	
DO_TFTAK_PkPp_4D	74202914	37101457	35	151.0	151	37101457	35	151.0	151	73917201	75	151.0	151	
DO_TFTAK_PkPp_6D	55744110	27872055	35	151.0	151	27872055	35	151.0	151	55471374	75	150.0	151	
DO_TFTAK_PkPp_8D	47191576	23595788	35	151.0	151	23595788	35	151.0	151	47034984	75	150.0	151	
DO_TFTAK_Pk_11D	99621060	49810530	35	151.0	151	49810530	35	151.0	151	99297967	75	150.0	151	
DO_TFTAK_Pk_14D	26504600	13252300	35	151.0	151	13252300	35	151.0	151	26376743	75	150.0	151	
DO_TFTAK_Pk_4D	53071660	26535830	35	151.0	151	26535830	35	151.0	151	52877017	75	150.0	151	
DO_TFTAK_Pk_6D	57604244	28802122	35	151.0	151	28802122	35	151.0	151	57360226	75	150.0	151	
DO_TFTAK_Pk_8D	41202130	20601065	35	151.0	151	20601065	35	151.0	151	41013586	75	150.0	151	

Table 5. Data of sequencing run for metagenomes of samples of fermented and non-fermented cereals (Couscous prepared from sorghum, millet or maize) – partner IRD

sampleid	n_raw_reads_t	n_raw_reads_R	min_read_len_	median_read_l	max_read_len_	n_raw_reads_R	min_read_len_	median_read_l	max_read_len_	n_preproc_read	min_read_len_p	median_read_l	max_read_len_preproc	
DO_IRD_Fa_MM1_1	44123706	22061853	35	151.0	151	22061853	35	151.0	151	43477111	75	150.0	151	
DO_IRD_Fa_MM1_2	48411676	24205838	35	151.0	151	24205838	35	151.0	151	47462487	75	150.0	151	
DO_IRD_Fa_MM1_3	33901654	16950827	35	151.0	151	16950827	35	151.0	151	33528429	75	150.0	151	
DO_IRD_Fa_MM1_4	45147126	22573563	35	151.0	151	22573563	35	151.0	151	44487525	75	150.0	151	
DO_IRD_Fa_MM1_5	29860904	14930452	35	151.0	151	14930452	35	151.0	151	29421428	75	150.0	151	
DO_IRD_Fa_MM2_1	62337140	31168570	35	151.0	151	31168570	35	151.0	151	62101491	75	150.0	151	
DO_IRD_Fa_MM2_2	71997224	35998612	35	151.0	151	35998612	35	151.0	151	71751982	75	150.0	151	
DO_IRD_Fa_MM2_3	37965194	18982597	35	151.0	151	18982597	35	151.0	151	37817208	75	150.0	151	
DO_IRD_Fa_MM2_4	58367188	29183594	35	151.0	151	29183594	35	151.0	151	58127117	75	150.0	151	
DO_IRD_Fa_MM2_5	79387440	39693720	35	151.0	151	39693720	35	151.0	151	78971493	75	150.0	151	
DO_IRD_Fa_Ma1_1	34196924	17098462	35	151.0	151	17098462	35	151.0	151	33690894	75	150.0	151	
DO_IRD_Fa_Ma1_2	45410996	22705498	35	151.0	151	22705498	35	151.0	151	44811002	75	150.0	151	
DO_IRD_Fa_Ma1_3	33611098	16805549	35	151.0	151	16805549	35	151.0	151	33123251	75	150.0	151	
DO_IRD_Fa_Ma1_4	72277392	36138696	35	151.0	151	36138696	35	151.0	151	71177826	75	150.0	151	
DO_IRD_Fa_Ma1_5	50038550	25019275	35	151.0	151	25019275	35	151.0	151	49267503	75	150.0	151	
DO_IRD_Fa_Ma2_1	27208038	13604019	35	151.0	151	13604019	35	151.0	151	27026405	75	150.0	151	
DO_IRD_Fa_Ma2_2	26889674	13444837	35	151.0	151	13444837	35	151.0	151	26737312	75	150.0	151	
DO_IRD_Fa_Ma2_3	21027310	10513655	35	151.0	151	10513655	35	151.0	151	20905570	75	150.0	151	
DO_IRD_Fa_Ma2_4	30054074	15027037	35	151.0	151	15027037	35	151.0	151	29916400	75	150.0	151	
DO_IRD_Fa_Ma2_5	21911672	10955836	35	151.0	151	10955836	35	151.0	151	21811319	75	150.0	151	
DO_IRD_Fa_Mi1_1	10767188	5383594	35	151.0	151	5383594	35	151.0	151	10610757	75	150.0	151	
DO_IRD_Fa_Mi1_2	5405842	2702921	35	151.0	151	2702921	35	151.0	151	5317411	75	150.0	151	
DO_IRD_Fa_Mi1_3	5855300	2927650	35	151.0	151	2927650	35	151.0	151	5731996	75	150.0	151	
DO_IRD_Fa_Mi1_4	3470986	1735493	35	151.0	151	1735493	35	151.0	151	3392221	75	150.0	151	
DO_IRD_Fa_Mi1_5	148324	74162	35	150.0	151	74162	35	150.0	151	134001	75	149.0	151	
DO_IRD_Fa_Mi2_1	11385176	5692588	35	151.0	151	5692588	35	151.0	151	11311520	75	150.0	151	
DO_IRD_Fa_Mi2_2	35473712	17736856	35	151.0	151	17736856	35	151.0	151	35329372	75	150.0	151	
DO_IRD_Fa_Mi2_3	31739854	15869927	35	151.0	151	15869927	35	151.0	151	31606605	75	150.0	151	
DO_IRD_Fa_Mi2_4	33752948	16876474	35	151.0	151	16876474	35	151.0	151	33622144	75	150.0	151	
DO_IRD_Fa_Mi2_5	27783344	13891672	35	151.0	151	13891672	35	151.0	151	27620646	75	150.0	151	
DO_IRD_Fa_So1_1	44794110	22397055	35	151.0	151	22397055	35	151.0	151	43688449	75	150.0	151	
DO_IRD_Fa_So1_2	64850600	32425300	35	151.0	151	32425300	35	151.0	151	63740874	75	150.0	151	
DO_IRD_Fa_So1_3	50335800	25167900	35	151.0	151	25167900	35	151.0	151	49453434	75	150.0	151	
DO_IRD_Fa_So1_4	58716238	29358119	35	151.0	151	29358119	35	151.0	151	57699149	75	150.0	151	
DO_IRD_Fa_So1_5	73951232	36975616	35	151.0	151	36975616	35	151.0	151	72426497	75	150.0	151	
DO_IRD_Fa_So2_1	43995436	21997718	35	151.0	151	21997718	35	151.0	151	43789212	75	150.0	151	
DO_IRD_Fa_So2_2	38339432	19169716	35	151.0	151	19169716	35	151.0	151	38199753	75	150.0	151	
DO_IRD_Fa_So2_3	36392952	18196476	35	151.0	151	18196476	35	151.0	151	36248223	75	150.0	151	
DO_IRD_Fa_So2_4	54339126	27169563	35	151.0	151	27169563	35	151.0	151	54105546	75	150.0	151	
DO_IRD_Fa_So2_5	62838	31419	35	124.0	151	31419	35	124.0	151	49299	75	147.0	151	

Table 6. Data of sequencing run for metagenomes of samples of fermented apple pomace -partner CSIC

sampleid	n_raw_reads_t	n_raw_reads_R	min_read_len_	median_read_l	max_read_len_	n_raw_reads_R	min_read_len_	median_read_l	max_read_len_	n_preproc_read	min_read_len_	median_read_l	max_read_len_preproc
DO_CSIC_P9-M1	13916726	6958363	35	151.0	151	6958363	35	151.0	151	13602743	75	150.0	151
DO_CSIC_P9-M10	50912652	25456326	35	151.0	151	25456326	35	151.0	151	50520056	75	150.0	151
DO_CSIC_P9-M11	73180726	36590363	35	151.0	151	36590363	35	151.0	151	72000360	75	150.0	151
DO_CSIC_P9-M12	56174010	28087005	35	151.0	151	28087005	35	151.0	151	54275224	75	150.0	151
DO_CSIC_P9-M13	33918912	16959456	35	151.0	151	16959456	35	151.0	151	32761793	75	150.0	151
DO_CSIC_P9-M14	36070578	18035289	35	151.0	151	18035289	35	151.0	151	34879990	75	150.0	151
DO_CSIC_P9-M15	2322	1161	35	151.0	151	1161	35	151.0	151	2217	77	150.0	151
DO_CSIC_P9-M16	23044584	11522292	35	151.0	151	11522292	35	151.0	151	22467135	75	150.0	151
DO_CSIC_P9-M17	177076476	88538238	35	151.0	151	88538238	35	151.0	151	172935062	75	150.0	151
DO_CSIC_P9-M18	45363042	22681521	35	151.0	151	22681521	35	151.0	151	44656600	75	150.0	151
DO_CSIC_P9-M19	34236286	17118143	35	151.0	151	17118143	35	151.0	151	33926647	75	150.0	151
DO_CSIC_P9-M2	40792020	20396010	35	151.0	151	20396010	35	151.0	151	39927316	75	150.0	151
DO_CSIC_P9-M20	36432554	18216277	35	151.0	151	18216277	35	151.0	151	35514570	75	150.0	151
DO_CSIC_P9-M21	78841740	39420870	35	151.0	151	39420870	35	151.0	151	76208115	75	150.0	151
DO_CSIC_P9-M22	147966488	73983244	35	151.0	151	73983244	35	151.0	151	144658243	75	150.0	151
DO_CSIC_P9-M23	36966190	18483095	35	151.0	151	18483095	35	151.0	151	36386670	75	150.0	151
DO_CSIC_P9-M24	71901484	35950742	35	151.0	151	35950742	35	151.0	151	70079468	75	150.0	151
DO_CSIC_P9-M25	22674046	11337023	35	151.0	151	11337023	35	151.0	151	22104181	75	150.0	151
DO_CSIC_P9-M26	68717538	34358769	35	151.0	151	34358769	35	151.0	151	67202399	75	150.0	151
DO_CSIC_P9-M27	96462236	48231118	35	151.0	151	48231118	35	151.0	151	95536614	75	150.0	151
DO_CSIC_P9-M28	29297650	14648825	35	151.0	151	14648825	35	151.0	151	28741443	75	150.0	151
DO_CSIC_P9-M29	101711344	50855672	35	151.0	151	50855672	35	151.0	151	98987179	75	150.0	151
DO_CSIC_P9-M3	34476832	17238416	35	151.0	151	17238416	35	151.0	151	33994646	75	150.0	151
DO_CSIC_P9-M30	75934188	37967094	35	151.0	151	37967094	35	151.0	151	73484920	75	150.0	151
DO_CSIC_P9-M31	98120160	49060080	35	151.0	151	49060080	35	151.0	151	95784762	75	150.0	151
DO_CSIC_P9-M32	70583892	35291946	35	151.0	151	35291946	35	151.0	151	69103582	75	150.0	151
DO_CSIC_P9-M33	45266262	22633131	35	151.0	151	22633131	35	151.0	151	44420827	75	150.0	151
DO_CSIC_P9-M34	119953958	59976979	35	151.0	151	59976979	35	151.0	151	116752121	75	150.0	151
DO_CSIC_P9-M35	78461594	39230797	35	151.0	151	39230797	35	151.0	151	75978232	75	150.0	151
DO_CSIC_P9-M36	89222088	44611044	35	151.0	151	44611044	35	151.0	151	86723735	75	150.0	151
DO_CSIC_P9-M37	113301666	56650833	35	151.0	151	56650833	35	151.0	151	110417790	75	150.0	151
DO_CSIC_P9-M4	18178220	9089110	35	151.0	151	9089110	35	151.0	151	17794621	75	150.0	151
DO_CSIC_P9-M5	16754456	8377228	35	151.0	151	8377228	35	151.0	151	16508789	75	150.0	151
DO_CSIC_P9-M6	11668042	5834021	35	151.0	151	5834021	35	151.0	151	11342588	75	150.0	151
DO_CSIC_P9-M7	239764102	119882051	35	151.0	151	119882051	35	151.0	151	232962262	75	150.0	151
DO_CSIC_P9-M8	52788342	26394171	35	151.0	151	26394171	35	151.0	151	51924507	75	150.0	151
DO_CSIC_P9-M9	95758902	47879451	35	151.0	151	47879451	35	151.0	151	94329302	75	150.0	151

Table 7. Data of sequencing run for metagenomes of samples of water kefir. Samples used different water kefir grains, each with different food sources (apple pulp, brewers spent grain and fig extract as the traditional control). For those samples where the grains appeared to flourish, the grains were sequenced. In all cases the filtered drink was sequenced. Three samples from King of Kefir were sequenced as commercial control samples. - partner TEAGASC

sampleid	n_raw_reads_total	n_raw_reads_R1	min_read_len_R1	median_read_len_R1	max_read_len_R1	n_raw_reads_R2	min_read_len_R2	median_read_len_R2	max_read_len_R2	n_preproc_reads	min_read_len_preproc	median_read_len_preproc	max_read_len_preproc
DO_TEAGASC_APSp	103568616	51784308	35	151	151	51784308	35	151	151	99201174	75	150	151
DO_TEAGASC_BSGSp	59234190	29617095	35	151	151	29617095	35	151	151	58923543	75	150	151
DO_TEAGASC_KKCMT	29889778	14944889	35	151	151	14944889	35	151	151	28477813	75	150	151
DO_TEAGASC_KKG	47381852	23690926	35	151	151	23690926	35	151	151	44639588	75	150	151
DO_TEAGASC_KKLG	25517128	12758564	35	151	151	12758564	35	151	151	24225304	75	150	151
DO_TEAGASC_WK001_APG	58210830	29105415	35	151	151	29105415	35	151	151	57885071	75	150	151
DO_TEAGASC_WK001_APS	35984734	17992367	35	151	151	17992367	35	151	151	35273117	75	150	151
DO_TEAGASC_WK001_BSGG	69264478	34632239	35	151	151	34632239	35	151	151	68823154	75	150	151
DO_TEAGASC_WK001_BSGSu	63005076	31502538	35	151	151	31502538	35	151	151	62565250	75	150	151
DO_TEAGASC_WK001_FG	108605444	54302722	35	151	151	54302722	35	151	151	107336166	75	150	151
DO_TEAGASC_WK001_FS	40076042	20038021	35	151	151	20038021	35	151	151	39518984	75	150	151
DO_TEAGASC_WK015_APSu	92989152	46494576	35	151	151	46494576	35	151	151	89994207	75	150	151
DO_TEAGASC_WK015_BSGSu	96938514	48469257	35	151	151	48469257	35	151	151	95635344	75	150	151
DO_TEAGASC_WK015_FS	66673856	33336928	35	151	151	33336928	35	151	151	65802757	75	150	151
DO_TEAGASC_WK025_APSu	46174468	23087234	35	151	151	23087234	35	151	151	43727973	75	151	151
DO_TEAGASC_WK025_BSGSu	48519360	24259680	35	151	151	24259680	35	151	151	47828418	75	150	151
DO_TEAGASC_WK025_FS	92990784	46495392	35	151	151	46495392	35	151	151	90818520	75	150	151
DO_TEAGASC_WK054_APSu	112577722	56288861	35	151	151	56288861	35	151	151	111417089	75	150	151
DO_TEAGASC_WK054_BSGSu	204923720	102461860	35	151	151	102461860	35	151	151	203367502	75	150	151
DO_TEAGASC_WK054_FS	112546562	56273281	35	151	151	56273281	35	151	151	111645352	75	150	151
DO_TEAGASC_WK059_APG	69107090	34553545	35	151	151	34553545	35	151	151	68360919	75	150	151
DO_TEAGASC_WK059_APSu	253862804	126931402	35	151	151	126931402	35	151	151	244654339	75	150	151
DO_TEAGASC_WK059_BSGG	42588592	21294296	35	151	151	21294296	35	151	151	42166975	75	150	151
DO_TEAGASC_WK059_BSGSu	48272680	24136340	35	151	151	24136340	35	151	151	47219635	75	150	151
DO_TEAGASC_WK059_FG	65284514	32642257	35	151	151	32642257	35	151	151	64206652	75	150	151
DO_TEAGASC_WK059_FS	43018596	21509298	35	151	151	21509298	35	151	151	41906975	75	150	151

Table 8. Data of sequencing run for strains provided by INRAE

sampleid	type	identification	n_raw_reads_total	n_raw_reads_R1	min_read_len_R1	median_read_len_R1	max_read_len_R1	n_raw_reads_R2	min_read_len_R2	median_read_len_R2	max_read_len_R2	n_preproc_reads	min_read_len_preproc	median_read_len_preproc	max_read_len_preproc
DOg_INRAE_C01	bacteria	*Microbacterium algeriense*	3931494	1965747	35	151.0	151	1965747	35	151.0	151	3901882	75	150.0	151
DOg_INRAE_C02	bacteria	*Bacillus tequilensis*	4532940	2266470	35	151.0	151	2266470	35	151.0	151	4495359	75	150.0	151
DOg_INRAE_C03	bacteria	*Bacillus licheniformis*	4233890	2116945	35	151.0	151	2116945	35	151.0	151	4208297	75	150.0	151
DOg_INRAE_C06	bacteria	*Bacillus tequilensis*	7425816	3712908	35	151.0	151	3712908	35	151.0	151	7384348	75	150.0	151
DOg_INRAE_C08	bacteria	*Weizmannia sp*	3538356	1769178	35	151.0	151	1769178	35	151.0	151	3515442	75	150.0	151
DOg_INRAE_C18	bacteria	*Bacillus safensis*	12664858	6332429	35	151.0	151	6332429	35	151.0	151	12595501	75	150.0	151
DOg_INRAE_C19	bacteria	*Bacillus siamensis*	7590430	3795215	35	151.0	151	3795215	35	151.0	151	7540888	75	150.0	151
DOg_INRAE_CIRM-BIA1909	bacteria	*Bacillus tequilensis*	8272334	4136167	35	151.0	151	4136167	35	151.0	151	8230451	75	150.0	151
DOg_INRAE_CIRM-BIA1910	bacteria	*Priestia megaterium*	7142682	3571341	35	151.0	151	3571341	35	151.0	151	7101922	75	150.0	151
DOg_INRAE_CIRM-BIA2519	bacteria	*Bacillus velezensis*	6441782	3220891	35	151.0	151	3220891	35	151.0	151	6400007	75	150.0	151
DOg_INRAE_D12	bacteria	*Exiguobacterium mexicanum*	4	2	148	149.5	151	2	149	150.0	151	4	148	149.5	151
DOg_INRAE_D4	bacteria	*Exiguobacterium aurantiacun*	11891086	5945543	35	151.0	151	5945543	35	151.0	151	11771485	75	150.0	151
DOg_INRAE_P01	bacteria	*Microbacterium oxydans*	6062288	3031144	35	151.0	151	3031144	35	151.0	151	6029751	75	150.0	151
DOg_INRAE_P03	bacteria	*Brachybacterium aquaticum*	3296084	1648042	35	151.0	151	1648042	35	151.0	151	3268673	75	150.0	151
DOg_INRAE_P09	bacteria	*Lactococcus lactis*	30240204	15120102	35	151.0	151	15120102	35	151.0	151	29995218	75	151.0	151
DOg_INRAE_P10	bacteria	*Leuconostoc mesenteroides*	16849158	8424579	35	151.0	151	8424579	35	151.0	151	16750205	75	151.0	151
DOg_INRAE_P12	bacteria	*Leuconostoc pseudomesente*	9215938	4607969	35	151.0	151	4607969	35	151.0	151	9155794	75	151.0	151
DOg_INRAE_P13	bacteria	*Leuconostoc suionicum*	15505832	7752916	35	151.0	151	7752916	35	151.0	151	15398658	75	150.0	151
DOg_INRAE_P23	bacteria	*Leuconostoc mesenteroides*	12220290	6110145	35	151.0	151	6110145	35	151.0	151	12145745	75	151.0	151
DOg_INRAE_P24	bacteria	*Leuconostoc pseudomesente*	11882256	5941128	35	151.0	151	5941128	35	151.0	151	11786056	75	151.0	151

Table 9. Data of sequencing run for strains provided by UNITO (provisional identification was performed by sequencing of the D1-D2 loop of the 26S rRNA encoding gene and will be confirmed by WGS)

sampleid	type	identification	n_raw_reads_total	n_raw_reads_R1	min_read_len_R1	median_read_len_R1	max_read_len_R1	n_raw_reads_R2	min_read_len_R2	median_read_len_R2	max_read_len_R2	n_preproc_reads	min_read_len_preproc	median_read_len_preproc	max_read_len_preproc
DOg_UNITO_B4-3	yeast	*Candida diddensiae*	11574994	5787497	35	151.0	151	5787497	35	151.0	151	11272367	75	151.0	151
DOg_UNITO_B5-10	yeast	*Candida diddensiae*	79137884	39568942	35	151.0	151	39568942	35	151.0	151	76979853	75	150.0	151
DOg_UNITO_B5-12	yeast	*Candida diddensiae*	13277666	6638833	35	151.0	151	6638833	35	151.0	151	12877449	75	151.0	151
DOg_UNITO_B5-9	yeast	*Candida diddensiae*	8905962	4452981	35	151.0	151	4452981	35	151.0	151	8647881	75	150.0	151
DOg_UNITO_B6-7	yeast	*Candida diddensiae*	24212484	12106242	35	151.0	151	12106242	35	151.0	151	23542857	75	150.0	151
DOg_UNITO_B6-8	yeast	*Candida diddensiae*	5210136	2605068	35	151.0	151	2605068	35	151.0	151	5050234	75	150.0	151
DOg_UNITO_B7-1	yeast	*Candida diddensiae*	21457490	10728745	35	151.0	151	10728745	35	151.0	151	20843064	75	150.0	151
DOg_UNITO_B7-7	yeast	*Candida diddensiae*	11046742	5523371	35	151.0	151	5523371	35	151.0	151	10725003	75	150.0	151
DOg_UNITO_B8-11	yeast	*Candida diddensiae*	7030968	3515484	35	151.0	151	3515484	35	151.0	151	6850853	75	150.0	151
DOg_UNITO_B8-12	yeast	*Candida diddensiae*	27442442	13721221	35	151.0	151	13721221	35	151.0	151	26707854	75	150.0	151
DOg_UNITO_B8-13	yeast	*Candida diddensiae*	9294154	4647077	35	151.0	151	4647077	35	151.0	151	9059320	75	151.0	151
DOg_UNITO_B8-2	yeast	*Candida diddensiae*	28721596	14360798	35	151.0	151	14360798	35	151.0	151	27888048	75	151.0	151
DOg_UNITO_B8-4	yeast	*Candida diddensiae*	23701596	11850798	35	151.0	151	11850798	35	151.0	151	23021528	75	150.0	151
DOg_UNITO_B9-10	yeast	*Candida diddensiae*	24480514	12240257	35	151.0	151	12240257	35	151.0	151	23678028	75	150.0	151
DOg_UNITO_B9-3	yeast	*Pichia manshurica*	4916378	2458189	35	151.0	151	2458189	35	151.0	151	4769101	75	150.0	151
DOg_UNITO_B9-7	yeast	*Candida diddensiae*	12298124	6149062	35	151.0	151	6149062	35	151.0	151	11974271	75	151.0	151
DOg_UNITO_O10-3	yeast	*Wickerhamomyces anomalus*	10452536	5226268	35	151.0	151	5226268	35	151.0	151	10117857	75	151.0	151
DOg_UNITO_O10-6	yeast	*Wickerhamomyces anomalus*	12629348	6314674	35	151.0	151	6314674	35	151.0	151	12166057	75	151.0	151
DOg_UNITO_O10-8	yeast	*Pichia membranifaciens*	25837554	12918777	35	151.0	151	12918777	35	151.0	151	24780999	75	150.0	151
DOg_UNITO_O11-1	yeast	*Wickerhamomyces anomalus*	10558084	5279042	35	151.0	151	5279042	35	151.0	151	10153637	75	151.0	151
DOg_UNITO_O11-8	yeast	*Wickerhamomyces anomalus*	7108642	3554321	35	151.0	151	3554321	35	151.0	151	6852239	75	151.0	151
DOg_UNITO_O2-15_AR	yeast	*Metschnikowia pulcherrima*	16992674	8496337	35	151.0	151	8496337	35	151.0	151	16385855	75	150.0	151
DOg_UNITO_O2-3_AR	yeast	*Metschnikowia pulcherrima*	20762910	10381455	35	151.0	151	10381455	35	151.0	151	20034989	75	150.0	151
DOg_UNITO_O3-6_SC	yeast	*Candida diddensiae*	37267172	18633586	35	151.0	151	18633586	35	151.0	151	36383796	75	151.0	151
DOg_UNITO_O3-7-SC	yeast	*Wickerhamomyces anomalus*	13999796	6999898	35	151.0	151	6999898	35	151.0	151	13482301	75	151.0	151
DOg_UNITO_O4-7_AR	yeast	*Candida diddensiae*	51389374	25694687	35	151.0	151	25694687	35	151.0	151	50101952	75	151.0	151
DOg_UNITO_O4-9_AR_19_11	yeast	*Candida diddensiae*	41248950	20624475	35	151.0	151	20624475	35	151.0	151	40056039	75	150.0	151
DOg_UNITO_O5-15_SC	yeast	*Candida diddensiae*	14665388	7332694	35	151.0	151	7332694	35	151.0	151	14312745	75	151.0	151
DOg_UNITO_O9-1	yeast	*Debaryomyces hansenii*	59720738	29860369	35	151.0	151	29860369	35	151.0	151	57579418	75	151.0	151
DOg_UNITO_O9-12	yeast	*Wickerhamomyces anomalus*	6454324	3227162	35	151.0	151	3227162	35	151.0	151	6191834	75	151.0	151
DOg_UNITO_O9-5	yeast	*Wickerhamomyces anomalus*	27463696	13731848	35	151.0	151	13731848	35	151.0	151	26285662	75	151.0	151
DOg_UNITO_S10-1	yeast	*Wickerhamomyces anomalus*	10406878	5203439	35	151.0	151	5203439	35	151.0	151	10012360	75	151.0	151
DOg_UNITO_S10-10	yeast	*Wickerhamomyces anomalus*	14646238	7323119	35	151.0	151	7323119	35	151.0	151	14012983	75	151.0	151
DOg_UNITO_S10-6	yeast	*Wickerhamomyces anomalus*	11205994	5602997	35	151.0	151	5602997	35	151.0	151	10815747	75	151.0	151
DOg_UNITO_S11-5	yeast	*Wickerhamomyces anomalus*	15801258	7900629	35	151.0	151	7900629	35	151.0	151	15182557	75	151.0	151
DOg_UNITO_S11-6	yeast	*Pichia membranifaciens*	49306788	24653394	35	151.0	151	24653394	35	151.0	151	47383440	75	150.0	151
DOg_UNITO_S11-7	yeast	*Pichia membranifaciens*	9693208	4846604	35	151.0	151	4846604	35	151.0	151	9325437	75	150.0	151
DOg_UNITO_S2-6_AR	yeast	*Candida diddensiae*	90805474	45402737	35	151.0	151	45402737	35	151.0	151	88133117	75	150.0	151
DOg_UNITO_S4-1_SC	yeast	*Wickerhamomyces anomalus*	9583240	4791620	35	151.0	151	4791620	35	151.0	151	8835263	75	151.0	151
DOg_UNITO_S4-5_SC	yeast	*Candida diddensiae*	15785368	7892684	35	151.0	151	7892684	35	151.0	151	15354600	75	150.0	151
DOg_UNITO_S5-14_AR	yeast	*Candida diddensiae*	19373878	9686939	35	151.0	151	9686939	35	151.0	151	18764655	75	151.0	151
DOg_UNITO_S5-8_AR	yeast	*Candida diddensiae*	1507386	753693	35	150.0	151	753693	35	150.0	151	1378135	75	150.0	151
DOg_UNITO_S6-3_AR	yeast	*Wickerhamomyces anomalus*	13667186	6833593	35	151.0	151	6833593	35	151.0	151	13168047	75	151.0	151
DOg_UNITO_S6-9_AR	yeast	*Wickerhamomyces anomalus*	9559764	4779882	35	151.0	151	4779882	35	151.0	151	9152676	75	151.0	151
DOg_UNITO_S7-3	yeast	*Wickerhamomyces anomalus*	12777744	6388872	35	151.0	151	6388872	35	151.0	151	12276919	75	151.0	151
DOg_UNITO_S8-11	yeast	*Wickerhamomyces anomalus*	14	7	44	150.0	151	7	44	150.0	151	12	146	150.15.00	151
DOg_UNITO_S9-2	yeast	*Debaryomyces hansenii*	10298208	5149104	35	151.0	151	5149104	35	151.0	151	9943360	75	151.0	151
DOg_UNITO_S9-7	yeast	*Wickerhamomyces anomalus*	9552818	4776409	35	151.0	151	4776409	35	151.0	151	9207177	75	150.0	151
DOg_UNITO_S9-9	yeast	*Wickerhamomyces anomalus*	10441630	5220815	35	151.0	151	5220815	35	151.0	151	10048058	75	151.0	151
DOg_UNITO_V5-9	yeast	*Candida diddensiae*	30564148	15282074	35	151.0	151	15282074	35	151.0	151	29662887	75	151.0	151
DOg_UNITO_V8-9	yeast	*Pichia membranifaciens*	21963356	10981678	35	151.0	151	10981678	35	151.0	151	20858119	75	150.0	151
DOg_UNITO_V9-14	yeast	*Candida diddensiae*	37157218	18578609	35	151.0	151	18578609	35	151.0	151	36171654	75	150.0	151
DOg_UNITO_V9-5	yeast	*Candida diddensiae*	18951438	9475719	35	151.0	151	9475719	35	151.0	151	18413244	75	150.0	151

Table 10. Data of sequencing run for strains provided by TFTAK (provisional identification will be confirmed by WGS).

sampleid	type	identification	n_raw_reads_total	n_raw_reads_R1	min_read_len_R1	median_read_len_R1	max_read_len_R1	n_raw_reads_R2	min_read_len_R2	median_read_len_R2	max_read_len_R2	n_preproc_reads	min_read_len_preproc	median_read_len_preproc	max_read_len_preproc
DOg_TFTAK_BP_1-1_3	bacteria	*Leuconostoc citreum*	7453034	3726517	35	151.0	151	3726517	35	151.0	151	7388477	75	150.0	151
DOg_TFTAK_BP_1-3_1	bacteria	*Weissella cibaria*	4016824	2008412	35	151.0	151	2008412	35	151.0	151	3997306	75	150.0	151
DOg_TFTAK_BP_4-1_24	bacteria	*Leuconostoc mesenteroides*	7551410	3775705	35	151.0	151	3775705	35	151.0	151	7504009	75	151.0	151
DOg_TFTAK_BP_4-3_14	bacteria	*Leuconostoc mesenteroides*	8854514	4427257	35	151.0	151	4427257	35	151.0	151	8813000	75	151.0	151
DOg_TFTAK_BP_4-4_8	bacteria	*Lacticaseibacillus paracasei*	10503420	5251710	35	151.0	151	5251710	35	151.0	151	10461427	75	150.0	151
DOg_TFTAK_BP_8-2_14	bacteria	*Lactiplantibacillus plantarum*	9332868	4666434	35	151.0	151	4666434	35	151.0	151	9305272	75	150.0	151
DOg_TFTAK_BP_8-2_8	bacteria	*Levilactobacillus brevis*	11240272	5620136	35	151.0	151	5620136	35	151.0	151	11200454	75	150.0	151
DOg_TFTAK_BP_8-4_18	bacteria	*Lactiplantibacillus plantarum*	29400798	14700399	35	151.0	151	14700399	35	151.0	151	29311097	75	150.0	151
DOg_TFTAK_BP_8-4_7	bacteria	*Levilactobacillus brevis*	50440894	25220447	35	151.0	151	25220447	35	151.0	151	50174564	75	150.0	151
DOg_TFTAK_BP_8-4_8	bacteria	*Lacticaseibacillus paracasei*	23761326	11880663	35	151.0	151	11880663	35	151.0	151	23692903	75	150.0	151
DOg_TFTAK_HK_1-1_2	bacteria	*Lactiplantibacillus plantarum*	20800070	10400035	35	151.0	151	10400035	35	151.0	151	20730331	75	150.0	151
DOg_TFTAK_HK_1-1_21	bacteria	*Leuconostoc citreum*	12284918	6142459	35	151.0	151	6142459	35	151.0	151	12240086	75	150.0	151
DOg_TFTAK_HK_1-2_12	bacteria	*Leuconostoc pseudomesenteroides*	19131652	9565826	35	151.0	151	9565826	35	151.0	151	19069631	75	150.0	151
DOg_TFTAK_HK_1-3_14	bacteria	*Lactiplantibacillus paraplantarum*	12457796	6228898	35	151.0	151	6228898	35	151.0	151	12409392	75	150.0	151
DOg_TFTAK_HK_11-1_11	bacteria	*Lactiplantibacillus paraplantarum*	11021254	5510627	35	151.0	151	5510627	35	151.0	151	10989565	75	150.0	151
DOg_TFTAK_HK_11-3_1	bacteria	*Levilactobacillus brevis*	29773658	14886829	35	151.0	151	14886829	35	151.0	151	29550294	75	150.0	151
DOg_TFTAK_HK_11-4_1	bacteria	*Lactiplantibacillus plantarum*	24644072	12322036	35	151.0	151	12322036	35	151.0	151	24574010	75	150.0	151
DOg_TFTAK_HK_4-1_13	bacteria	*Fructilactobacillus fructivorans*	13222402	6611201	35	151.0	151	6611201	35	151.0	151	13177380	75	150.0	151
DOg_TFTAK_HK_4-2_9	bacteria	*Latilactobacillus curvatus*	11092884	5546442	35	151.0	151	5546442	35	151.0	151	11062844	75	150.0	151
DOg_TFTAK_HK_4_3_12	bacteria	*Latilactobacillus curvatus*	10954116	5477058	35	151.0	151	5477058	35	151.0	151	10917643	75	150.0	151
DOg_TFTAK_HK_8-3_11	bacteria	*Lactiplantibacillus plantarum*	49180246	24590123	35	151.0	151	24590123	35	151.0	151	49004283	75	150.0	151
DOg_TFTAK_HK_8-4_18	bacteria	*Lactiplantibacillus plantarum*	9789786	4894893	35	151.0	151	4894893	35	151.0	151	9756960	75	150.0	151
DOg_TFTAK_Kim_4-2_4	bacteria	*Lactiplantibacillus plantarum*	23182170	11591085	35	151.0	151	11591085	35	151.0	151	23112862	75	150.0	151
DOg_TFTAK_Kim_4-3_11	bacteria	*Lactiplantibacillus plantarum*	19683698	9841849	35	151.0	151	9841849	35	151.0	151	19614528	75	150.0	151
DOg_TFTAK_Kim_4-3_8	bacteria	*Lactiplantibacillus plantarum*	32754	16377	35	151.0	151	16377	35	151.0	151	29708	75	150.0	151
DOg_TFTAK_Kim_4-4_4	bacteria	*Levilactobacillus brevis*	11837716	5918858	35	151.0	151	5918858	35	151.0	151	11788245	75	150.0	151
DOg_TFTAK_Kim_4-4_8	bacteria	*Levilactobacillus brevis*	18284240	9142120	35	151.0	151	9142120	35	151.0	151	18237124	75	150.0	151
DOg_TFTAK_Kim_4_3_17	bacteria	*Latilactobacillus curvatus*	9560802	4780401	35	151.0	151	4780401	35	151.0	151	9528760	75	150.0	151
DOg_TFTAK_NK_1-2_1	bacteria	*Leuconostoc pseudomesenteroides*	22439338	11219669	35	151.0	151	11219669	35	151.0	151	22350565	75	150.0	151
DOg_TFTAK_NK_1-3_2	bacteria	*Weissella cibaria*	12868696	6434348	35	151.0	151	6434348	35	151.0	151	12827625	75	150.0	151
DOg_TFTAK_NK_1-3_3	bacteria	*Latilactobacillus sakei*	11010382	5505191	35	151.0	151	5505191	35	151.0	151	10974792	75	150.0	151
DOg_TFTAK_NK_4-1_1	bacteria	*Leuconostoc pseudomesenteroides*	8790666	4395333	35	151.0	151	4395333	35	151.0	151	8762503	75	150.0	151
DOg_TFTAK_NK_4-2_11	bacteria	*Leuconostoc mesenteroides*	31609566	15804783	35	151.0	151	15804783	35	151.0	151	31481795	75	150.0	151
DOg_TFTAK_NK_4-2_14	bacteria	*Leuconostoc citreum*	11591782	5795891	35	151.0	151	5795891	35	151.0	151	11545489	75	150.0	151
DOg_TFTAK_NK_4-3_2	bacteria	*Leuconostoc mesenteroides*	25230816	12615408	35	151.0	151	12615408	35	151.0	151	25130263	75	150.0	151
DOg_TFTAK_NK_8-2_1	bacteria	*Lactiplantibacillus plantarum*	19003570	9501785	35	151.0	151	9501785	35	151.0	151	18935859	75	150.0	151
DOg_TFTAK_NK_8-2_5	bacteria	*Levilactobacillus brevis*	24038868	12019434	35	151.0	151	12019434	35	151.0	151	23970318	75	150.0	151
DOg_TFTAK_NK_8-4_2	bacteria	*Lactiplantibacillus plantarum*	17931346	8965673	35	151.0	151	8965673	35	151.0	151	17871434	75	150.0	151
DOg_TFTAK_NK_8-4_4	bacteria	*Lacticaseibacillus paracasei*	9542922	4771461	35	151.0	151	4771461	35	151.0	151	9511489	75	150.0	151
DOg_TFTAK_NK_8-4_8	bacteria	*Levilactobacillus brevis*	25665854	12832927	35	151.0	151	12832927	35	151.0	151	25585713	75	150.0	151

Table 11. Data of sequencing run for strains provided by IRD (provisional identification will be confirmed by WGS)

sampleid	type	identification	n_raw_reads_total	n_raw_reads_R1	min_read_len_R1	median_read_len_R1	max_read_len_R1	n_raw_reads_R2	min_read_len_R2	median_read_len_R2	max_read_len_R2	n_preproc_reads	min_read_len_preproc	median_read_len_preproc	max_read_len_preproc
DOg_IRD_LP-1105	bacteria	*Lactiplantibacillus plantarum*	6137900	3068950	35	151.0	151	3068950	35	151.0	151	6102630	75	150.0	151
DOg_IRD_LP-1710	bacteria	*Lactiplantibacillus plantarum*	11868356	5934178	35	151.0	151	5934178	35	151.0	151	11818820	75	150.0	151
DOg_IRD_LP-2029	bacteria	*Lactiplantibacillus plantarum*	7300030	3650015	35	151.0	151	3650015	35	151.0	151	7255286	75	150.0	151
DOg_IRD_LP-2047	bacteria	*Lactiplantibacillus plantarum*	899328	449664	35	151.0	151	449664	35	151.0	151	894667	75	150.0	151
DOg_IRD_LP-2653	bacteria	*Lactiplantibacillus plantarum*	3625288	1812644	35	151.0	151	1812644	35	151.0	151	3606183	75	150.0	151
DOg_IRD_LP-2657	bacteria	*Lactiplantibacillus plantarum*	2104458	1052229	35	151.0	151	1052229	35	151.0	151	2094293	75	150.0	151
DOg_IRD_LP-420	bacteria	*Lactiplantibacillus plantarum*	5961516	2980758	35	151.0	151	2980758	35	151.0	151	5936962	75	150.0	151
DOg_IRD_LP-953	bacteria	*Lactiplantibacillus plantarum*	2878598	1439299	35	151.0	151	1439299	35	151.0	151	2866643	75	150.0	151
DOg_IRD_LR-1534	bacteria	*Limosilactobacillus reuteri*	5796848	2898424	35	151.0	151	2898424	35	151.0	151	5765474	75	150.0	151
DOg_IRD_LR-911	bacteria	*Limosilactobacillus reuteri*	7656650	3828325	35	151.0	151	3828325	35	151.0	151	7617851	75	150.0	151

Table 12. Data of sequencing run for strains provided by CSIC (provisional identification of bacteria was performed by 16S rRNA sequencing and provisional identification of yeast was performed by ITS1 and ITS2 sequencing and will be confirmed by WGS)

sampleid	type	identification	n_raw_reads_total	n_raw_reads_R1	min_read_len_R1	median_read_len_R1	max_read_len_R1	n_raw_reads_R2	min_read_len_R2	median_read_len_R2	max_read_len_R2	n_preproc_reads	min_read_len_preproc	median_read_len_preproc	max_read_len_preproc
DOg_CSIC_P9-G12	yeast	*Torulaspora delbruekii*	5159030	2579515	35	151.0	151	2579515	35	151.0	151	4968995	75	150.0	151
DOg_CSIC_P9-G13	yeast	*Torulaspora spp*	6182474	3091237	35	151.0	151	3091237	35	151.0	151	5969295	75	150.0	151
DOg_CSIC_P9-G14	yeast	*Saccharomyces cerevisae*	2131134	1065567	35	151.0	151	1065567	35	151.0	151	1993038	75	151.0	151
DOg_CSIC_P9-G15	yeast	*Saccharomyces cerevisae*	7231008	3615504	35	151.0	151	3615504	35	151.0	151	6867600	75	151.0	151
DOg_CSIC_P9-G17	yeast	*Pichia kluyveri*	7691992	3845996	35	151.0	151	3845996	35	151.0	151	7235918	75	151.0	151
DOg_CSIC_P9-G18	yeast	*Pichia kluyveri*	12795680	6397840	35	151.0	151	6397840	35	151.0	151	11939042	75	151.0	151
DOg_CSIC_P9-G19	yeast	*Pichia kluyveri*	5886922	2943461	35	151.0	151	2943461	35	151.0	151	5540044	75	151.0	151
DOg_CSIC_P9-G2	yeast	*Hanseniaspora uvarum*	6910900	3455450	35	151.0	151	3455450	35	151.0	151	6691043	75	151.0	151
DOg_CSIC_P9-G20	yeast	*Pichia kluyveri*	5833708	2916854	35	151.0	151	2916854	35	151.0	151	5488506	75	151.0	151
DOg_CSIC_P9-G21	yeast	*Pichia kudriavzeuii*	10204146	5102073	35	151.0	151	5102073	35	151.0	151	8914433	75	150.0	151
DOg_CSIC_P9-G22	yeast	*Pichia spp*	12164376	6082188	35	151.0	151	6082188	35	151.0	151	11454671	75	151.0	151
DOg_CSIC_P9-G24	yeast	*Metschnikowia pulcherrima*	9356770	4678385	35	151.0	151	4678385	35	151.0	151	9072394	75	150.0	151
DOg_CSIC_P9-G25	yeast	*Metschnikowia pulcherrima*	9691100	4845550	35	151.0	151	4845550	35	151.0	151	9373783	75	150.0	151
DOg_CSIC_P9-G28	yeast	*Hanseniaspora uvarum*	7084886	3542443	35	151.0	151	3542443	35	151.0	151	6922000	75	150.0	151
DOg_CSIC_P9-G29	yeast	*Hanseniaspora uvarum*	7598886	3799443	35	151.0	151	3799443	35	151.0	151	7353390	75	151.0	151
DOg_CSIC_P9-G30	yeast	*Hanseniaspora uvarum*	4164400	2082200	35	151.0	151	2082200	35	151.0	151	4024807	75	151.0	151
DOg_CSIC_P9-G31	bacteria	*Acetobacter lambici*	5565286	2782643	35	151.0	151	2782643	35	151.0	151	5524875	75	150.0	151
DOg_CSIC_P9-G37	bacteria	*Lactococcus lactis*	4791712	2395856	35	151.0	151	2395856	35	151.0	151	4742963	75	151.0	151
DOg_CSIC_P9-G41	bacteria	*Lactococcus lactis*	16053446	8026723	35	151.0	151	8026723	35	151.0	151	15907096	75	151.0	151
DOg_CSIC_P9-G42	bacteria	*Lactococcus lactis*	46023936	23011968	35	151.0	151	23011968	35	151.0	151	45566120	75	151.0	151
DOg_CSIC_P9-G46	bacteria	*Acetobacter persici*	30488196	15244098	35	151.0	151	15244098	35	151.0	151	30307629	75	150.0	151
DOg_CSIC_P9-G47	bacteria	*Lactiplantibacillus plantarum*	5867794	2933897	35	151.0	151	2933897	35	151.0	151	5835288	75	150.0	151
DOg_CSIC_P9-G48	bacteria	*Lactiplantibacillus plantarum*	10591898	5295949	35	151.0	151	5295949	35	151.0	151	10526509	75	150.0	151
DOg_CSIC_P9-G49	bacteria	*Lactiplantibacillus plantarum*	8257742	4128871	35	151.0	151	4128871	35	151.0	151	8201443	75	150.0	151
DOg_CSIC_P9-G5	yeast	*Hanseniaspora uvarum*	7777660	3888830	35	151.0	151	3888830	35	151.0	151	7507761	75	151.0	151
DOg_CSIC_P9-G51	yeast	*Torulaspora delbruekii*	13365478	6682739	35	151.0	151	6682739	35	151.0	151	13273061	75	150.0	151
DOg_CSIC_P9-G52	yeast	*Pichia membanifaciens*	8106684	4053342	35	151.0	151	4053342	35	151.0	151	7768176	75	150.0	151
DOg_CSIC_P9-G54	bacteria	*Levilactobacillus brevis*	10874318	5437159	35	151.0	151	5437159	35	151.0	151	10812760	75	150.0	151
DOg_CSIC_P9-G55	bacteria	*Lacticaseibacillus rhamnosus*	8621678	4310839	35	151.0	151	4310839	35	151.0	151	8573869	75	150.0	151
DOg_CSIC_P9-G56	bacteria	*Lactiplantibacillus plantarum*	18756442	9378221	35	151.0	151	9378221	35	151.0	151	18635021	75	150.0	151
DOg_CSIC_P9-G57	bacteria	*Lacticaseibacillus paracasei*	27302324	13651162	35	151.0	151	13651162	35	151.0	151	27148309	75	150.0	151
DOg_CSIC_P9-G58	bacteria	*Acetobacter spp*	21536408	10768204	35	151.0	151	10768204	35	151.0	151	21394135	75	150.0	151
DOg_CSIC_P9-G6	yeast	*Hanseniaspora uvarum*	6206764	3103382	35	151.0	151	3103382	35	151.0	151	5977795	75	151.0	151
DOg_CSIC_P9-G61	bacteria	*Acetobacter thailandicus*	23753722	11876861	35	151.0	151	11876861	35	151.0	151	23591792	75	150.0	151
DOg_CSIC_P9-G63	bacteria	*Acetobacter persici*	9190964	4595482	35	151.0	151	4595482	35	151.0	151	9112710	75	150.0	151
DOg_CSIC_P9-G64	bacteria	*Acetobacter fabarum*	5985974	2992987	35	151.0	151	2992987	35	151.0	151	5940310	75	150.0	151
DOg_CSIC_P9-G65	bacteria	*Gluconacetobacter aggeris*	7519894	3759947	35	151.0	151	3759947	35	151.0	151	7468994	75	150.0	151
DOg_CSIC_P9-G67	bacteria	*Gluconobacter cerinus*	4959576	2479788	35	151.0	151	2479788	35	151.0	151	4931845	75	150.0	151
DOg_CSIC_P9-G69	bacteria	*Gluconobacter oxydans*	5359544	2679772	35	151.0	151	2679772	35	151.0	151	5328814	75	150.0	151
DOg_CSIC_P9-G72	bacteria	*Leuconostoc pseudomesenteroides*	13107702	6553851	35	151.0	151	6553851	35	151.0	151	13042738	75	150.0	151
DOg_CSIC_P9-G74	bacteria	*Gluconacetobacter aggeris*	8278500	4139250	35	151.0	151	4139250	35	151.0	151	8219930	75	150.0	151
DOg_CSIC_P9-G75	bacteria	*Acetobacter okinawensis*	31255416	15627708	35	151.0	151	15627708	35	151.0	151	31062645	75	150.0	151

Table 13. Data of sequencing run for strains provided by TEAGASC (provisional identification of bacteria was performed by 16S rRNA sequencing and provisional identification of yeast was performed by ITS1 and ITS2 sequencing was performed by 16S rRNA -bacteria- or ITS -yeasts- Sanger sequencing and will be confirmed by WGS).

sampleid	type	identification	n_raw_reads_total	n_raw_reads_R1	min_read_len_R1	median_read_len_R1	max_read_len_R1	n_raw_reads_R2	min_read_len_R2	median_read_len_R2	max_read_len_R2	n_preproc_reads	min_read_len_preproc	median_read_len_preproc	max_read_len_preproc
DOg_TEAGASC_AO_A3	bacteria	*Acetobacter orientalis*	17134030	8567015	35	151	151	8567015	35	151	151	17048986	75	150	151
DOg_TEAGASC_AO_B2	bacteria	*Acetobacter orientalis*	12859438	6429719	35	151	151	6429719	35	151	151	12763017	75	150	151
DOg_TEAGASC_AS_H1	bacteria	*Acetobacter syzygii*	9968520	4984260	35	151	151	4984260	35	151	151	9909906	75	150	151
DOg_TEAGASC_ASp_A4	bacteria	*Acetobacter spp.*	15135858	7567929	35	151	151	7567929	35	151	151	15056875	75	150	151
DOg_TEAGASC_BA_D2	bacteria	*Bacillus spp.*	17129500	8564750	35	151	151	8564750	35	151	151	17057057	75	151	151
DOg_TEAGASC_B_AC_G2	bacteria	*Bifidobacterium aquikeferi*	12046084	6023042	35	151	151	6023042	35	151	151	11964343	75	151	151
DOg_TEAGASC_C_I_H2	bacteria	*Clostridium spp.*	11645102	5822551	35	151	151	5822551	35	151	151	11586860	75	150	151
DOg_TEAGASC_GC_A2	bacteria	*Gluconobacter cerinus*	16287916	8143958	35	151	151	8143958	35	151	151	16217417	75	150	151
DOg_TEAGASC_GO_C4	bacteria	*Gluconobacter oxydans*	6354848	3177424	35	151	151	3177424	35	151	151	6329613	75	150	151
DOg_TEAGASC_GSp_D1	bacteria	*Gluconobacter spp.*	9821752	4910876	35	151	151	4910876	35	151	151	9774947	75	150	151
DOg_TEAGASC_LG_G1	bacteria	*Liquorilactobacillus ghanensis*	11322430	5661215	35	151	151	5661215	35	151	151	11262015	75	151	151
DOg_TEAGASC_LN_F1	bacteria	*Liquorilactobacillus nagelli*	9482898	4741449	35	151	151	4741449	35	151	151	9434412	75	151	151
DOg_TEAGASC_LN_H3	bacteria	*Liquorilactobacillus nagelli*	12501256	6250628	35	151	151	6250628	35	151	151	12431444	75	151	151
DOg_TEAGASC_LP_B3	bacteria	*Leuconostoc pseudomesenteroide*	9551584	4775792	35	151	151	4775792	35	151	151	9504487	75	151	151
DOg_TEAGASC_LP_C2	bacteria	*Lacticaseibacillus paracasei*	8796276	4398138	35	151	151	4398138	35	151	151	8757595	75	150	151
DOg_TEAGASC_LP_C3	bacteria	*Leuconostoc pseudomesenteroide*	7631304	3815652	35	151	151	3815652	35	151	151	7590458	75	151	151
DOg_TEAGASC_LP_D4	bacteria	*Lacticaseibacillus paracasei*	7497844	3748922	35	151	151	3748922	35	151	151	7450985	75	151	151
DOg_TEAGASC_LP_E3	bacteria	*Leuconostoc pseudomesenteroide*	12782626	6391313	35	151	151	6391313	35	151	151	12718241	75	151	151
DOg_TEAGASC_LP_G3	bacteria	*Lacticaseibacillus paracasei*	10818744	5409372	35	151	151	5409372	35	151	151	10759255	75	150	151
DOg_TEAGASC_LS_B4	bacteria	*Liquorilactobacillus satsumonsis*	15228738	7614369	35	151	151	7614369	35	151	151	15149689	75	150	151
DOg_TEAGASC_LS_D3	bacteria	*Liquorilactobacillus satsumonsis*	16690474	8345237	35	151	151	8345237	35	151	151	16611820	75	150	151
DOg_TEAGASC_LS_E1	bacteria	*Liquorilactobacillus satsumonsis*	11552446	5776223	35	151	151	5776223	35	151	151	11491947	75	150	151
DOg_TEAGASC_LSp_F3	bacteria	*Liquorilactobacillus spp.*	8274174	4137087	35	151	151	4137087	35	151	151	8237476	75	151	151
DOg_TEAGASC_OO_F2	bacteria	*Oenococcus spp.*	10064188	5032094	35	151	151	5032094	35	151	151	10013243	75	150	151
DOg_TEAGASC_SC_F4	yeast	*Saccharomyces cerevisae*	4913120	2456560	35	151	151	2456560	35	151	151	4528671	75	151	151
DOg_TEAGASC_SC_G4	yeast	*Saccharomyces cerevisae*	9349582	4674791	35	151	151	4674791	35	151	151	8905635	75	150	151
DOg_TEAGASC_ZF_E4	yeast	*Zygotorulaspora florentina*	21477446	10738723	35	151	151	10738723	35	151	151	20717877	75	150	151
DOg_TEAGASC_ZF_H4	yeast	*Zygotorulaspora florentina*	7658960	3829480	35	151	151	3829480	35	151	151	7097735	75	151	151

5. Concluding remarks

Considering the output of the metagenome sequencing in terms of raw reads it is concluded that it will be possible to perform the planned investigations regarding the microbiome of the samples and the metabolic models needed to build the synthetic consortia. For a few samples raw reads obtained were low and it may be necessary to perform again the metagenome sequencing and/or DNA extraction and sequencing. Similarly, whole genome sequencing was overall successful. Only for a few strains it will be necessary to perform again the sequencing and/or DNA extraction and sequencing.

- In relation to what was planned, the goal regarding metagenome numbers was reached while for strains a slightly lower number, compared to the plan, was achieved. This is not expected to influence the progress of the activities within the project. First of all, the information collected from the sequenced strains is sufficient to initiate the process of designing the consortia (Task 4. 2). In addition, the number of strains available in the culture collections of partners is much higher than what was planned, and it will be possible to increase the WGS strains in the coming period. This delay is partly attributed to the unavailability of the commercial kit that was supposed to be used by all partners in the various countries.