

# Supplementary Information

## Taxogenomic assessment and genomic characterisation of *Weissella cibaria* strain 92 able to metabolise oligosaccharides derived from dietary fibres

Anna Månberger<sup>1,\*</sup>, Phebe Verbrugghe<sup>2</sup>, Elísabet Eik Guðmundsdóttir<sup>3</sup>, Sara Santesson<sup>1</sup>, Anne Nilsson<sup>2</sup>, Guðmundur Óli Hreggviðsson<sup>3,4</sup>, Javier A. Linares-Pastén<sup>1,\*</sup> and Eva Nordberg Karlsson<sup>1,\*</sup>

<sup>1</sup> Biotechnology, Department of Chemistry, Lund University, PO Box 124, SE-221 00 Lund, Sweden

<sup>2</sup> Food Technology, Engineering and Nutrition, Lund University, PO Box 124, SE-221 00, Lund, Sweden

<sup>3</sup> Mátis, Vinlandsleid 12, IS-113, Reykjavík, Iceland

<sup>4</sup> Faculty of Life and Environmental Sciences, University of Iceland, Askja, IS-101, Reykjavík, Iceland

\* Corresponding author, e-mail: [anna.manberger@biotek.lu.se](mailto:anna.manberger@biotek.lu.se), [javier.linares\\_pasten@biotek.lu.se](mailto:javier.linares_pasten@biotek.lu.se), [eva.nordberg\\_karlsson@biotek.lu.se](mailto:eva.nordberg_karlsson@biotek.lu.se), +46 222 36 73 (JALP), +46 222 46 26 (ENK)

**Supplementary Table S1.** Sequencing and trimming data of *Weissella* sp. strains. Coverage is calculated based on genome size determined after the assembly, see **Table 1**.

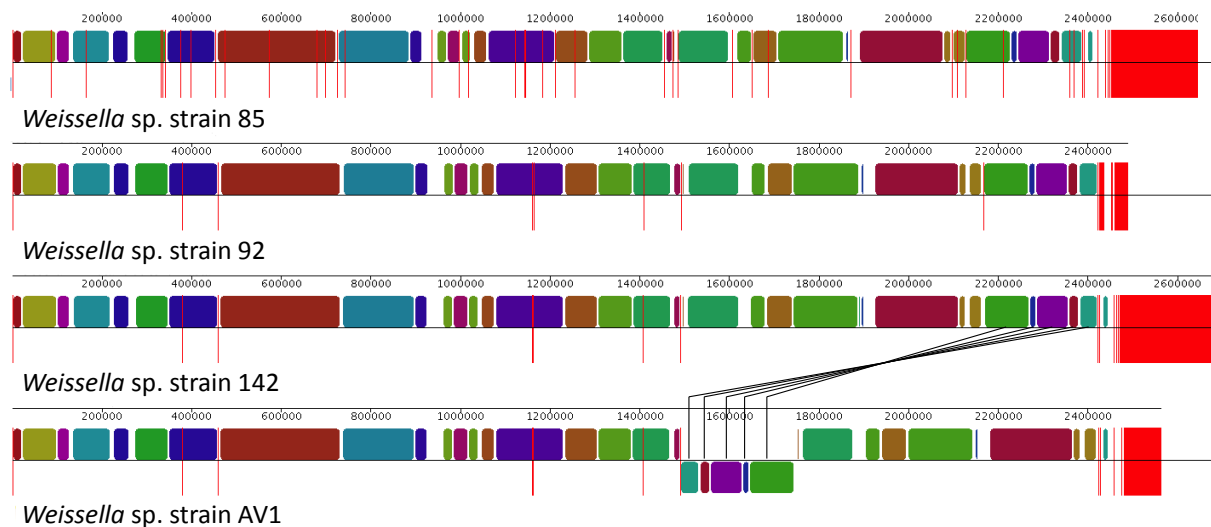
		<b>85</b>	<b>92</b>	<b>142</b>	<b>AV1</b>
<i>Sequencing</i>					
DNA Library	Reads passed filter	2 354 724	995 989	1 522 075	2 188 909
	Total Mbp	1 413	598	913	1 313
	Coverage	534	240	340	513
Mate pair library	Reads passed filter	-	1 226 160	1 429 182	1 612 465
	Total Mbp	-	613	715	806
	Coverage	-	232	287	300
Total coverage		534	472	627	813
<i>Trimming and quality filtering</i>					
DNA Library	Paired reads	1 531 549	846 854	932 023	1 433 933
	Unpaired reads	665 340	123 219	464 090	595 342
	Total Mbp	951	463	594	883
	Coverage	360	186	221	345
Mate pair library	Paired reads	-	637 683	765 720	793 512
	Unpaired reads	-	492 074	531 757	613 272
	Total Mbp	-	398	464	495
	Coverage	-	160	173	193
Total coverage		360	346	394	538

**Supplementary Table S2.** Screening of carbohydrate fermentation in *Weissella cibaria* strain 92. Carbohydrate fermentation was considered present (+) if the specific growth rate,  $\mu$ , was at least  $0.1 \text{ h}^{-1}$  and the pH decreased at least 0.2 units over 48 h of cultivation. Otherwise, no or negligible carbohydrate fermentation (-) was assumed. P - production, U - uptake, n.d. - not detected. Production and uptake are relative to the amount of product that was within the medium at the start of cultivation.

Carbohydrate	Carbohydrate fermentation	Fermentation products				
		Lactate	Ethanol	Acetate	Propionate	Butyrate
<i>1 mL cultivation – microaerophilic</i>						
Glucose	+	P	P	U	n.d.	n.d.
Laminaribiose	+	P	P	P	n.d.	n.d.
Laminaritriose	-					
Laminaritetraose	-					
Arabinose	+	P	n.d.	P	n.d.	n.d.
Arabinobiose (A <sub>2</sub> )	+	P	n.d.	P	n.d.	P
Arabinotriose (A <sub>3</sub> )	-					
Arabinotetraose (A <sub>4</sub> )	-					
Arabinopentose (A <sub>5</sub> )	-					
<i>N</i> -Acetyl-glucosamine	+	P	P	P	n.d.	n.d.
Diacetyl-chitobiose	-					
<i>75 mL cultivation – anaerobic</i>						
Glucose	+	P	P	P	n.d.	n.d.
Arabinose	+	P	n.d.	P	n.d.	P
Xylose	+	P	n.d.	P	n.d.	P
Negative control	-					

**Supplementary Table S3.** Primer sequences used for real-time quantitative reverse transcription-PCR (RT-qPCR). *peg* refers to protein encoding gene from annotation.

<b>Gene_peg</b>	<b>Forward primer</b>	<b>Reverse primer</b>
<i>Reference genes</i>		
<i>gyrA_402</i>	TGACGACCAACAAGGGTGTA	TAAGCGAATAACCCGGACAC
<i>gyrB_403</i>	TTGATGTCACAAAGGCGAAG	ACAATGGACGCATGTAACGA
<i>dnaG_1592</i>	GCGTAATCTGTTTGGGCTGT	CCGCATGAGAAGCAGTGATA
<i>rpoB_730</i>	GTCCATCAATGGAGCAAGGT	TAAACATCATCGCGGATCAA
<i>rpoC_731</i>	TGGACACATCTTGGCACCTA	CCTTCACGACCAGCTTCTTC
<i>rpoD_7593</i>	GATGACACAGCCAAGCAAGA	ATTGCATACCACGACCAACA
<i>ddl_1049</i>	ATGCCATGGACAAGGAAAAG	CTTCCGCCACAATCGTATCT
<i>Target genes</i>		
<i>bgl_197</i>	GATACGGCGTTCAAGGATGT	AAATCGGCATCATAGCGTTC
<i>bglB_302</i>	TTGCGCATACCAAGTATCCA	TTACGCGCTGCATAGACATC
<i>bglB_303+905</i>	TCAATTGCGGATGTGATGTT	ATCCCACGGTGATTTGGATA
<i>bglB_379</i>	CCGTTATCAAAAGCCGTTGT	ATGTGCTCGCGGAAATAATC
<i>bglB_461</i>	CCGTTATGCGGAGACGTTAT	AATACCACCAGCGTTCCAAG
<i>bglB_1258</i>	TTTATGATTGGGGGCATGTT	TGCACGTCTGAGAACCAGTC
<i>bglB_1261</i>	TCCAATGCTGCAAAACTCAG	AATGCGGATGCCAATAACTC
<i>bglB_1679</i>	GCTGACGTTGAGGTTGTGAA	GTAACGCAAGCCTTGTGGAT
<i>bglB_2328</i>	TGTCGACCTGATGGGTTACA	CCCATAACGCTTCGACATTT
<i>eng_661</i>	GCAACGTTAAGCAAGCAACA	CGCAACTTCCTTGGCTTTAG
<i>gh13_853</i>	AATTGGCGCAGTATTTTTGG	CAGTGCTGGATTTTCCAAT
<i>bglB_303</i>	CGCAAACACGGTTGATTATG	GTTGTAGGCGAGCCATTCAT



**Supplementary Fig. S1.** Alignments of the assembled genomes. The border of each contig is marked by a red line. Coloured blocks are regions aligned to regions in the other genomes, marked in the same colour. Blocks under the line indicate that the alignment is made in the reverse direction relative to the first genome (*Weissella* sp. strain 85). For a clearer picture, only major rearrangements of aligned regions (> 10.000 bp) are marked by connecting lines.

	New <i>Weissella</i> sp. strains				<i>Weissella cibaria</i>								<i>Weissella confusa</i>		
	85	AV1	142	FBL5	AB3b	CMU	KACC 11862	CMS3	CMS2	ff3PR	CH2	MG1	LBAE C39-2	DSM 20196	MBF8-1
<i>W. sp.</i> strain 92	100.0	100.0	100.0	99.9	99.9	99.8	99.8	99.8	99.8	99.7	99.7	99.6	88.0	87.4	86.6
<i>W. sp.</i> strain 85	100.0	100.0	99.9	99.9	99.8	99.8	99.8	99.8	99.7	99.8	99.6	88.0	87.3	86.6	
<i>W. sp.</i> strain AV1	100.0	99.9	99.9	99.8	99.8	99.8	99.8	99.8	99.7	99.8	99.6	88.0	87.1	86.7	
<i>W. sp.</i> strain 142	99.9	99.9	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.7	88.0	87.1	86.7	
<i>W. cibaria</i> FBL5	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.7	99.7	99.4	88.7	87.9	88.5	
<i>W. cibaria</i> AB3b	99.7	99.7	99.7	99.7	99.7	99.7	99.7	99.7	99.9	99.8	99.9	87.3	87.3	87.1	
<i>W. cibaria</i> CMU	100.0	100.0	100.0	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	86.4	85.3	85.8	
<i>W. cibaria</i> KACC 11862	100.0	100.0	100.0	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	86.3	85.3	86.0	
<i>W. cibaria</i> CMS3	100.0	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	86.4	85.7	85.8	
<i>W. cibaria</i> CMS2	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	86.4	85.3	86.0	
<i>W. cibaria</i> ff3PR	99.9	99.9	100.0	87.4	87.9	86.1	87.4	87.9	86.1	87.1	87.1	88.6	87.5	87.1	
<i>W. cibaria</i> CH2	99.9	99.9	87.1	87.5	86.8	87.1	87.1	87.1	87.1	87.1	87.1	88.6	87.5	87.1	
<i>W. cibaria</i> MG1	87.1	87.5	86.8	87.1	87.5	86.8	87.1	87.5	86.8	87.1	87.5	87.1	87.5	86.8	
<i>W. confusa</i> LBAE C39-2	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	
<i>W. confusa</i> DSM 20196	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	

	<i>Weissella</i> species												
	sp. DD23	confusa	parames- enteroides	hellenica	bombi	jogaegotgali	ceti	oryzae	kandleri	viridescens	minor	halotolerans	koreensis
<i>W. cibaria</i> CH2	99.8	87.5	72.1	71.8	71.4	69.3	65.3	64.9	63.7	63.7	62.5	62.4	60.6
<i>W. sp.</i> DD23	88.3	74.2	74.1	71.9	68.1	70.5	68.1	58.8	67.1	59.4	65.7	63.8	63.8
<i>W. confusa</i> DSM 20196	72.9	69.2	76.6	71.4	66.3	68.8	65.1	66.8	61.3	66.6	62.7	62.8	62.8
<i>W. paramesenteroides</i> ATCC 33313	81.3	83.6	87.7	64.5	60.2	57.0	60.4	62.2	62.7	62.7	63.1	62.0	62.0
<i>W. hellenica</i> R-53116	92.5	84.7	62.6	62.4	68.8	62.6	60.3	60.3	61.3	61.3	61.3	61.3	61.3
<i>W. bombi</i> R-53094	82.0	65.8	67.5	61.4	67.2	67.6	65.2	67.6	65.2	67.6	65.2	64.3	64.3
<i>W. jogaegotgali</i> FOL01	54.4	59.3	56.5	56.6	65.2	57.7	59.8	59.8	59.8	59.8	59.8	59.8	59.8
<i>W. ceti</i> WS105	64.9	66.8	72.4	64.9	58.4	63.6	63.6	63.6	63.6	63.6	63.6	63.6	63.6
<i>W. oryzae</i> SG25T	64.6	59.3	61.7	59.3	61.7	59.3	61.7	59.3	61.7	59.3	61.7	59.3	61.7
<i>W. kandleri</i> DSM 20593	59.2	61.8	56.8	56.8	56.8	56.8	56.8	56.8	56.8	56.8	56.8	56.8	56.8
<i>W. viridescens</i> DSM 20410	78.6	65.1	56.0	56.0	56.0	56.0	56.0	56.0	56.0	56.0	56.0	56.0	56.0
<i>W. minor</i> DSM 20014	67.8	59.9	59.9	59.9	59.9	59.9	59.9	59.9	59.9	59.9	59.9	59.9	59.9
<i>W. halotolerans</i> DSM 20190	57.3	57.3	57.3	57.3	57.3	57.3	57.3	57.3	57.3	57.3	57.3	57.3	57.3

**Supplementary Fig. S2.** Sequence identity in percentage of bidirectional gene pairs from pairwise comparison of strains within *Weissella*.

