

# Complete Genome Sequencing and Comparative Genomic Analysis of the Thermotolerant Acetic Acid Bacterium, *Acetobacter pasteurianus* SKU1108, Provide a New Insight into Thermotolerance

MINENOSUKE MATSUTANI<sup>1\*</sup>, HIDEKI HIRAKAWA<sup>2</sup>, ERI HIRAOKA<sup>1</sup>, GUNJANA THEERAGOOL<sup>3</sup>, TOSHIHARU YAKUSHI<sup>1,4</sup>, and KAZUNOBU MATSUSHITA<sup>1,4</sup>

<sup>1</sup>Department of Biological Chemistry, Faculty of Agriculture, Yamaguchi University, Yamaguchi 753–8515, Japan; <sup>2</sup>Kazusa DNA Research Institute, Kazusa-kamatari, Kisarazu, Chiba 292–0818, Japan; <sup>3</sup>Department of Microbiology, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand; and <sup>4</sup>Research Center for Thermotolerant Microbial Resources, Yamaguchi University, Yamaguchi 753–8515, Japan

(Received February 9, 2016—Accepted August 3, 2016—Published online September 24, 2016)

*Acetobacter pasteurianus* SKU1108 is a typical thermotolerant acetic acid bacterium. In this study, the complete genome sequence of the SKU1108 strain was elucidated, and information on genomic modifications due to the thermal adaptation of SKU1108 was updated. In order to obtain a clearer understanding of the genetic background responsible for thermotolerance, the SKU1108 genome was compared with those of two closely related complete genome strains, thermotolerant *A. pasteurianus* 386B and mesophilic *A. pasteurianus* NBRC 3283. All 24 “thermotolerant genes” required for growth at higher temperatures in the thermotolerant *Acetobacter tropicalis* SKU1100 strain were conserved in all three strains. However, these thermotolerant genes accumulated amino acid mutations. Some biased mutations, particularly those that occurred in xanthine dehydrogenase XdhA, may be related to thermotolerance. By aligning whole genome sequences, we identified ten SKU1108 strain-specific regions, three of which were conserved in the genomes of the two thermotolerant *A. pasteurianus* strains. One of the regions contained a unique paralog of the thermotolerant gene *xdhA*, which may also be responsible for conferring thermotolerance. Thus, comparative genomics of complete genome sequences may provide novel insights into the phenotypes of these thermotolerant strains.

**Key words:** acetic acid bacterium, thermotolerant strain, complete genome sequence

Strictly aerobic acetic acid bacteria (AAB) are classified as a sub-group of the *Acetobacteraceae* family in the class *Alphaproteobacteria*. AAB strongly oxidize various sugars, alcohols, and sugar alcohols. Vinegar is industrially produced by oxidative fermentation using AAB, particularly those in *Acetobacter* and *Komagataeibacter* (37). In ethanol-containing medium, the bacterial strains of these two genera exhibit three growth phases; an ethanol oxidation phase, acetic acid resistance phase, and acetate over oxidation phase (14, 19). In the ethanol oxidation phase, cells oxidize ethanol to acetic acid via acetaldehyde with membrane-bound alcohol and aldehyde dehydrogenases linked to the respiratory chain (20). In the acetic acid resistance phase, cells resist autoproducted acetic acid using several mechanisms without assimilating it. In the overoxidation phase, cells assimilate acetate by oxidizing it to CO<sub>2</sub> via the TCA cycle (14, 19).

The *Acetobacter pasteurianus* SKU1108, isolated from fruits in Thailand, has been shown to efficiently perform acetic acid fermentation at higher temperatures than the required temperature range for other *A. pasteurianus* strains, such as NBRC 3283 and IFO 3191 (=NBRC 3191) (21, 28). We previously reported the complete genome sequence of the mesophilic strain, *A. pasteurianus* NBRC 3283 (2). Furthermore, the complete genome sequence of the thermotolerant *A. pasteurianus* 386B has recently been published (13). Therefore,

in order to identify the genes responsible for the specific thermotolerance phenotype, we elucidated the complete genome sequence of the *A. pasteurianus* SKU1108. Comparisons of this sequence with those of other thermotolerant and mesophilic strains enabled us to identify the genomic regions conserved in thermotolerant bacteria.

## Materials and Methods

### Genomic DNA and library preparation

Genomic DNA from SKU1108 for genome sequencing was prepared as previously reported (23). Genomic DNA (~20 µg) was purified with the AMPure Xp Kit (Beckman Coulter, Beverly, MA, USA). DNA was sheared using a Covaris g-TUBE (Covaris, Woburn, MA, USA) following the manufacturer’s recommendations for 20-kb fragments. A sequencing library was prepared using the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences of California, Menlo Park, CA, USA). Libraries for 2 SMRT cells (~10 µg) were constructed using two different methods. Small library fragments were removed using BluePippin from one SMRT cell (Sage Science, Beverly, MA, USA) with a 10-kb cut-off. In the other cell, small library fragments were not removed.

### DNA sequencing and assembly

Genome sequencing was performed by Takara Bio using PacBio RSII single-molecule real-time (SMRT) sequencing technology (Pacific Biosciences of California, Menlo Park, CA, USA). The sequence data obtained from 2 SMRT cells were used for the subsequent sequence assembly. Sequencing reads were assembled using Hierarchical Genome Assembly Process 3 (HGAP3) in PacBio SMRT portal version 2.3.0. Three large contigs were assembled

\* Corresponding author. E-mail: mine@yamaguchi-u.ac.jp;  
Tel: +81-83-933-5857; Fax: +81-83-933-5820.

with a mean coverage of 387-fold. The assembly was corrected with the Quiver consensus algorithm to obtain a high-accuracy genome assembly (7). The overlap sequences of each contig end were manually edited. Previously reported Illumina sequence reads of the SKU1108 strain were mapped onto these three contigs using Bowtie 2 (18, 22). Unmapped reads were collected and a *de novo* assembly was performed using SPAdes 3.0.0 (3). The resulting assembly revealed two additional small plasmids: plasmids 3 and 4.

#### Genome annotation

The gene detection and genome annotation of the chromosome and four plasmid sequences were performed using the auto-annotation package Prokka (29). Protein-coding sequences (CDS) of the complete genome sequence were predicted using Prodigal 2.62 (11, 12). ARAGORN 1.26 and Barrnap 0.4 were used to predict tRNA and rRNA regions, respectively (16). The functional assignments of the predicted CDSs were based on a BLASTP homology search against the previously reported *Acetobacter pasteurianus* genome and the NCBI non-redundant (NR) database (1). The positions of the start codon were manually checked and edited using the commercial genome sequence editing software, *in silico* Molecular Cloning v5.3.75 (In Silico Biology, Kanagawa, Japan). All signal peptide genes encoded by the SKU1108 genome were predicted by SignalP 4.1 (26). Clustered-regularly interspaced short palindromic repeats (CRISPRs) in the genome sequence were revealed by MinCED 0.16 (4).

#### Whole genome alignment and genome map

The genomic sequences of *A. pasteurianus* 386B (HF677570–HF677572) and *A. pasteurianus* IFO 3283-01 (=NBRC 3283) (AP011121–AP011127) were downloaded from the NCBI FTP website at ftp.ncbi.nlm.nih.gov. The chromosome sequences of the two *A. pasteurianus* strains, 386B and NBRC 3283, were independently aligned against that of the strain SKU1108 using NUCmer (15). Unaligned regions were manually assigned as specific regions 1 to 10. The genes located at these specific regions were identified and listed. We generated graphic illustrations of genome alignments using CGView (33). Graphical alignments were generated using the progressiveMauve program with default parameters (8).

#### Identification of genes associated with thermotolerance

The previously reported mutated sites in the genomes of the TI and TH-3 strains associated with thermotolerance were re-confirmed using a previously reported method (22). The 24 thermotolerant genes conserved in *A. tropicalis* NBRC 101654 (=SKU1100) and their homologous sequences were also searched for in the three *A. pasteurianus* complete genomes by BLASTP with an E-value cut-off of  $10^{-10}$  and sequence overlap (query and subject)  $\geq 70\%$  (1, 30).

#### Phylogenetic tree construction

A BLASTP search against all proteins encoded by 13 *Acetobacter* and 1 *Gluconacetobacter diazotrophicus* Pal 5 genomes was performed using the amino acid sequences of AarC and AarC1 in *A. pasteurianus* SKU1108 as a query. The resulting hits were aligned using MUSCLE v.3.8.31 at the amino acid sequence level (9, 10). Poorly aligned regions were removed using GBLOCKS version

0.91b (6, 34). A phylogenetic tree was constructed using the PROTGAMMAWAG model in RAxML 8.0.14 with 1,000 bootstrap replicates and visualized with the MEGA 6.1 package (31, 32, 35).

#### Sequence data deposition

The *A. pasteurianus* SKU1108 (=NBRC 101655) genome sequence was deposited in DDBJ/EMBL/GenBank under the accession numbers AP014881 to AP014885. The versions described here are the first versions. The BioProject ID is PRJDA65545.

## Results and Discussion

#### General genome features

The genomic DNA of *A. pasteurianus* SKU1108 consisted of a 2,902,389-bp circular chromosome and four plasmids: plasmid 1 (187,193 bp), plasmid 2 (6,331 bp), plasmid 3 (2,799 bp), and plasmid 4 (2,278 bp), with a GC content of 52.75% (Table 1). In total, 2,662 and 214 CDSs were identified in the chromosome and plasmids, respectively. Putative functions were assigned to 2,002 genes. Signal sequences were searched against all CDSs, and detected in 252 cases (data not shown). In total, 56 tRNAs, 1 transfer-messenger RNA (tmRNA), and 5 sets of ribosomal RNA operons (*rrn*) were predicted in the chromosome sequence (Table 1).

The genomes of *A. pasteurianus* are known to contain numerous genes encoding membrane-bound oxidoreductases. In the SKU1108 genome, we revealed genes that code for the membrane-bound PQQ-dependent alcohol dehydrogenase (ADH), *adhAB* operon (locus tag: APT\_00084–APT\_00083) and its subunit III, *adhS* (APT\_00654), and other PQQ enzymes, such as the membrane-bound glucose dehydrogenase (APT\_00581), PQQ-dependent dehydrogenase 1 (APT\_02219), and PQQ-dependent dehydrogenase 4 (APT\_01465). In addition, genes encoding two membrane-bound aldehyde dehydrogenase operons (APT\_00973–APT\_00975 and APT\_00426–APT\_00428) were identified. These genes were also conserved in complete genomes of the strains 386B and NBRC 3283. The respiratory chains of *Acetobacter* species are known to play crucial roles in energy metabolism (24, 27). Therefore, the gene repertoires of respiratory chains were investigated. The cytochrome *ba<sub>3</sub>* ubiquinol oxidase *cyaBACD* operon (APT\_00087–APT\_00090), cytochrome *bd* ubiquinol oxidase *cydBACD* operon (APT\_01543–APT\_01546), cyanide-insensitive ubiquinol oxidase (CIO) *cioBA* operon (APT\_02213–APT\_02214), type I NADH-quinone oxidoreductase, *nuoA-nuoN* operon (APT\_01737–APT\_01725) and *nuoM* (APT\_00690), two type II NADH dehydrogenase *ndh* (APT\_00547 and APT\_02111), heme A synthase *ctaA* (APT\_00075), heme O

**Table 1.** Summary of the complete genome sequence of *Acetobacter pasteurianus* SKU1108

Genome information	Nucleotide sequence length (bp)	GC content (%)	Number of ORFs	Number of tRNAs	Number of tmRNAs	Number of rRNAs	GenBank accession number
chromosome	2,902,389	52.79	2,662	56	1	15	AP014881
plasmid 1	187,193	52.02	201	—	—	—	AP014882
plasmid 2	6,331	52.88	8	—	—	—	AP014883
plasmid 3	2,799	58.20	3	—	—	—	AP014884
plasmid 4	2,278	55.71	2	—	—	—	AP014885
Total	3,135,082	52.75	2,876	56	1	15	—

(5 sets of the rRNA operon)

synthase *ctaB* (APT\_01774), and cytochrome *c* oxidase subunit *ctaD* (APT\_01775) were conserved in the complete genomes of the strains SKU1108, 386B, and NBRC 3283. In contrast, the cytochrome *b* subunit *petB* (APT\_01924) of the ubiquinol-cytochrome *c* reductase (*bc<sub>1</sub>* complex) operon (APT\_01924-APT\_01926) was disrupted with a stop codon insertion around the start codon site in SKU1108, suggesting that this *bc<sub>1</sub>* complex is not functional in SKU1108. However, a paralog sequence of this gene was also found in the SKU1108 genome (APT\_00754). Therefore, a combination of these remaining subunits may construct an active *bc<sub>1</sub>* complex.

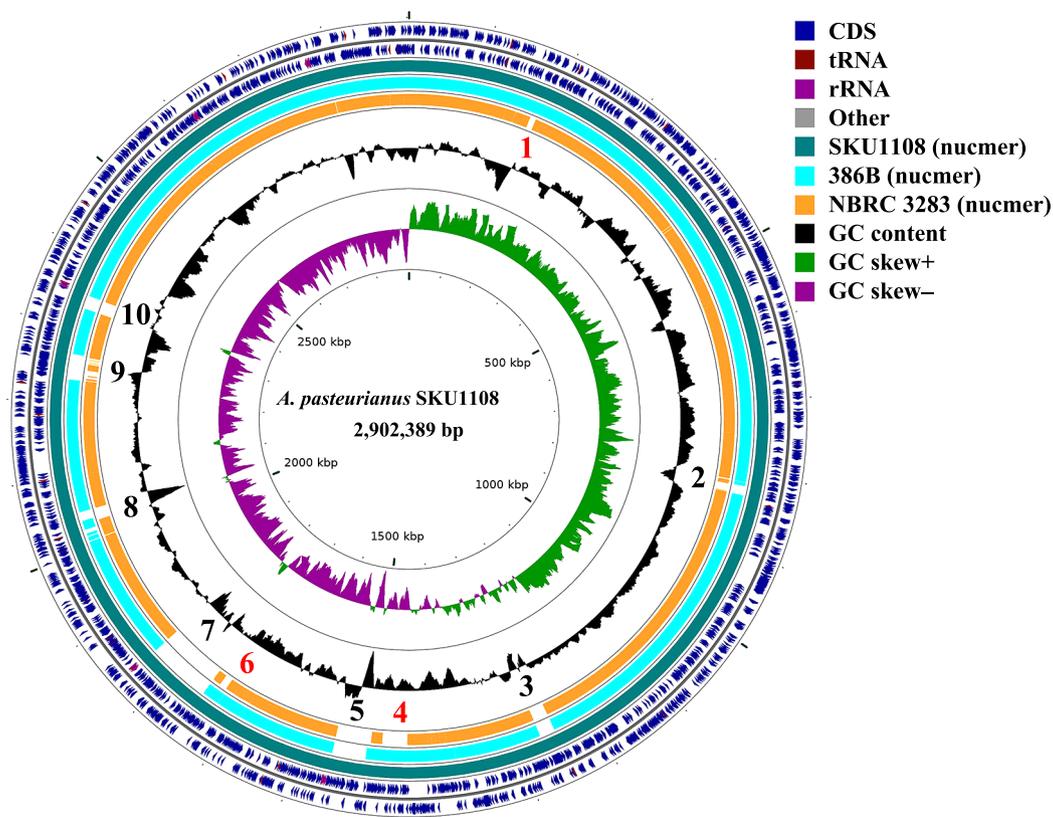
#### Whole genome level comparison in complete genome sequences of three closely related strains

Illegheems and co-workers elucidated the complete genome sequence of the thermotolerant *A. pasteurianus* 386B (13). They isolated this strain from cocoa bean heap fermentation with heating up to approximately 42–43°C (5). They also showed that the strain 386B rapidly produced acetic acid at a high concentration (17). We previously reported that SKU1108 is more thermotolerant than NBRC 3283 and performs acetic acid fermentation even at 38.5°C (22). Therefore, SKU1108 and 386B were defined as thermotolerant acetic acid fermentation strains. On the other hand, NBRC 3283 was defined as a mesophilic strain, which has the ability to perform fermentation at temperatures up to 37°C, but not at 38.5°C. Thus, in order to elucidate the mechanism underlying thermotolerance, a comparative genome analysis was performed using the complete genome sequences of one meso-

philic and two thermotolerant strains. Whole genome level synteny was compared in these three complete genomes using Mauve aligner (Fig. S1). A genome synteny analysis revealed that the synteny of these three genomes was highly conserved.

In order to define the uniquely conserved genomic regions in the SKU1108 genome sequence, three complete genome sequences were mapped and aligned against the SKU1108 chromosome sequence using NUCmer, which is a part of the MUMmer genome alignment package. A circular map of the SKU1108 chromosome sequence with alignments is shown in Fig. 1. Ten strain-specific regions were identified (Table S1). Genes located in these regions are also shown in Table S2. Of these, regions 1, 4, and 6 were only conserved between the thermotolerant strains SKU1108 and 386B. Therefore, genes from these regions may be responsible for the thermotolerance phenotype of these strains and may be classified as thermotolerance-conferring conserved regions. Of the ten specific regions, regions 1, 2, 3, and 9 encode prophage-related proteins such as phage integrase and phage terminase. Therefore, these regions may have been acquired by prophage-insertion.

We also compared the repertoire of plasmid sequences among the three studied strains. SKU1108 had one large 187-kb plasmid 1 and three small plasmids (Table 1). On the other hand, 386B and NBRC 3283 had one (194 kb) and two (191 and 182 kb) large plasmids, respectively (2, 13). In order to clarify differences between the large plasmids, we aligned their sequences against the SKU1108 plasmid 1 sequence using NUCmer. Large plasmid 1 of SKU1108 and that of 386B, Apa386Bp1 (HF677571) were highly conserved (Fig.



**Fig. 1.** Comparison of complete genome sequences of three *Acetobacter pasteurianus* strains. Closely related genome sequences were aligned and mapped onto the complete genome sequence (AP014881) of the *A. pasteurianus* SKU1108 using NUCmer. The illustration was constructed using CG-view. Red and black colored numbers mean thermotolerant strains and SKU1108-specific regions, respectively.

S2). Whole plasmid sequence comparisons also revealed that the synteny of these two plasmids was highly conserved (data not shown). In contrast, the two large plasmids of the strain NBRC 3283 were distinct from that of SKU1108 and not conserved. The results of the whole genome-based phylogenetic analysis suggested that SKU1108 and 386B are more closely related to each other than to NBRC 3283 (Fig. S3). Therefore, since the large plasmids were conserved in the two thermotolerant strains, they may have existed in the ancestor strain, *i.e.* before the divergence of SKU1108 and 386B into two separate thermotolerant strains.

#### Genes associated with thermotolerance

In a previous study, we performed the thermal adaptation of SKU1108 by its repeated cultivation under acetic acid fermentation conditions. Consequently, two individually adapted strains, TI and TH-3, were obtained, and their mutational sites were identified using the draft SKU1108 genome assembly as a reference (22). In order to identify additional mutational sites, the previously reported Illumina sequence reads of two adapted strains were mapped onto the complete genomic DNA sequence of SKU1108 that also included the DNAs of four plasmids. No additional mutational sites were detected, indicating that the mapping analysis against the draft genome assembly constructed from Illumina reads was sufficient for a mutation site analysis of the adapted strains. All the mutational sites were detected on the chromosome sequence. Therefore, we have now changed the locus tag number based on the complete genome sequence (Table S3).

In order to identify the genes whose expression is required for growth at higher temperatures, we previously mutated the *A. tropicalis* SKU1100, a thermotolerant AAB that grows even at 42°C, by transposon mutagenesis, thereby inducing random insertions in the *Tn10* transposon using a conjugation method (30). We revealed that 24 “thermotolerant” genes (see Table 2 or Table S4) were associated with growth at higher temperatures. Therefore, in the present study, we clarified whether these 24 genes and their corresponding nucleotide and amino acid mutations were present in complete genomes. A phylogenetic analysis performed for each gene demonstrated that the thermotolerant genes from SKU1100 and three complete *A. pasteurianus* genomes were clustered in the same clade (data not shown). Therefore, together with the result that all thermotolerant genes were encoded in the conserved regions of all three complete genomes, we concluded that all known thermotolerant genes are conserved in the three genomes, even in the mesophilic strain NBRC 3283 (Table S4). The number of nucleotide/amino acid mutations in the 24 thermotolerant genes and additional 6 genes composed of operons with four genes (DNA methyltransferase [APT\_00916], MinC [APT\_00954], RpoE [APT\_01242], and XdhA [APT\_01267]) of the thermotolerant genes, *rhnB* (APT\_00195), *minE* (APT\_00952), *minD* (APT\_00953), *rpoE* (APT\_01243), *xdhC* (APT\_01265), and *xdhB* (APT\_01266) (30 genes in total) are shown in Table 2. The nucleotide sequences of the nine genes were identical in the three strains. There was some variation in nucleotide mutations in 21 other genes, 16 of which mainly

**Table 2.** Number of nucleotide and amino acid mutations of 24 thermotolerant genes among three analyzed strains

Gene	Product	Locus_tag	Nucleotide mutation			Amino acid mutation		
			SKU1108	NBRC 3283	386B	SKU1108	NBRC 3283	386B
—	hypothetical protein	APT_00053	0	0	0	0	0	0
<i>hpnB</i>	glycosyl transferase	APT_00180	0	0	0	0	0	0
<i>poxA</i>	lysyl-tRNA synthetase	APT_00246	13	0	0	2 (1)	0	0
<i>lepA</i>	GTP-binding protein LepA	APT_00584	9	0	0	1 (0)	0	0
—	hypothetical protein	APT_00590	4	0	0	0	0	0
<i>mmpA</i>	putative metalloprotease MmpA	APT_00603	8	0	0	1 (1)	0	0
<i>hsp20</i>	heat shock protein Hsp20	APT_00623	0	0	0	0	0	0
—	hypothetical protein	APT_00856	5	0	0	0	0	0
<i>gluE</i>	glutamyl-ammonia-ligase adenylyltransferase	APT_00857	0	0	0	0	0	0
<i>rhnB</i>	ribonuclease HII	APT_00915	0	0	0	0	0	0
<i>yhdJ</i>	DNA methyltransferase	APT_00916	0	1	0	0	0	0
<i>minE</i>	cell division topological specificity factor MinE	APT_00952	3	0	0	2 (0)	0	0
<i>minD</i>	cell division inhibitor MinD	APT_00953	9	0	0	0	0	0
<i>minC</i>	cell division inhibitor MinC	APT_00954	4	0	0	1 (0)	0	0
—	hypothetical protein	APT_00961	16	0	0	0	0	0
<i>nhaK2</i>	Na <sup>+</sup> /H <sup>+</sup> antiporter	APT_00984	69	0	0	1 (0)	0	0
<i>smc</i>	chromosome segregation protein SMC	APT_01022	44	0	0	9 (0)	0	0
<i>ttg2C</i>	toluene ABC transporter periplasmic protein	APT_01083	0	0	0	0	0	0
<i>serA</i>	D-3-phosphoglycerate dehydrogenase	APT_01136	6	0	0	0	0	0
<i>rpoE</i>	RNA polymerase sigma-E factor (sigma-24) protein 2	APT_01242	0	0	0	0	0	0
<i>rpoE</i>	RNA polymerase sigma-E factor (sigma-24) protein 1	APT_01243	0	0	0	0	0	0
<i>xdhC</i>	xanthine dehydrogenase accessory protein XdhC	APT_01265	7	1	0	1 (1)	1 (0)	0
<i>xdhB</i>	xanthine dehydrogenase molybdopterin binding subunit XdhB	APT_01266	48	0	0	9 (2)	0	0
<i>xdhA</i>	xanthine dehydrogenase XdhA	APT_01267	28	0	0	11 (3)	0	0
—	asparagine synthetase	APT_01828	1	1	30	1 (1)	1 (0)	5 (0)
—	flavodoxin/nitric oxide synthase	APT_01927	5	5	4	2 (0)	1 (0)	1 (0)
<i>amiA</i>	N-acetylmuramoyl-L-alanine amidase	APT_02041	11	8	9	2 (0)	0	1 (0)
<i>cysG</i>	siroheme synthase	APT_02252	16	5	8	3 (1)	0	2 (0)
—	hypothetical protein	APT_02411	0	0	0	0	0	0
<i>htrA</i>	endopeptidase DegP/Do	APT_02556	21	0	0	3 (0)	0	0

Numbers in parentheses indicate residues substituted with dissimilar amino acids. An additional six genes: APT\_00915, APT\_00952, APT\_00953, APT\_01243, APT\_01265, and APT\_01266, which are clustered with 24 thermotolerant genes, are also included in the mutational analysis.

**Table 3.** List of paralogous sequences of 24 thermotolerant genes conserved in three *A. pasteurianus* strains

Gene SKU1108	Blast query	Blast hits	Gene (hits)	Product (hits)	E-value	Identity (%)
<i>lepA</i>	APT_00584	APT_00570	<i>typA</i>	GTP-binding protein TypA	5.00E-32	26
		APA01_06200	<i>bipA</i>	GTP-binding protein TypA	5.00E-32	26
		APA386B_2113	<i>typA</i>	GTP-binding protein TypA/BipA	5.00E-32	26
<i>ttg2C</i>	APT_01083	APA386B_1720	<i>ttg2C</i>	toluene transporter substrate-binding periplasmic protein Ttg2C	4.00E-14	35
		APT_00211	<i>ttg2C</i>	toluene ABC transporter periplasmic protein	4.00E-14	35
		APA01_02300	<i>ttg2C</i>	toluene ABC transporter periplasmic protein	4.00E-14	35
<i>xdhA</i>	APT_01267	APA386B_251	<i>xdhA</i>	xanthine dehydrogenase small subunit	4.00E-53	32
		APT_01390	<i>xdhA</i>	xanthine dehydrogenase XdhA	4.00E-53	32
<i>htrA</i>	APT_02556	APA386B_1296	<i>degP</i>	endopeptidase DegP/Do	1.00E-94	42
		APT_02481	<i>degP</i>	endopeptidase DegP/Do	1.00E-94	42
		APA01_24860	<i>degP</i>	endopeptidase DegP/Do	2.00E-94	42

accumulated nucleotide mutations in NBRC 3283 (Table 2 and Fig. S3). In contrast, the gene encoding asparagine synthetase (APT\_01828) accumulated nucleotide mutations in SKU1108 only, and the gene for DNA methyltransferase in 386B only. In the remaining three genes, flavodoxin/nitric oxide synthase (APT\_01927), *amiA* (APT\_02041), and *cysG* (APT\_02252), nucleotide mutations accumulated in all three strains. Non-synonymous mutations in these thermotolerant genes were also investigated and amino acid mutations were detected in 15 (Table 2). Of these, seven genes had amino acid mutations (10 mutations in total) that conferred substitutions by dissimilar amino acids. Amino acid mutations with dissimilar residues only occurred in NBRC 3283, suggesting that they are responsible for the less thermotolerant phenotype of NBRC 3283. Their mutation sites were APT\_00246 (A115D), APT\_00603 (S95R), APT\_01265 (G106E), APT\_01266 (T394R and E413G), APT\_01267 (R161C, C162R, and T220I), APT\_01828 (R471S), and APT\_02252 (E185A). In particular, genes encoding the *xdhABC* operon (APT\_01265–APT\_01267) accumulated 6 amino acid mutations. Of these, the T220 residue (T206 in *Rhodobacter capsulatus*) of *xdhA* (APT\_01267), shown to be in the active site in the crystal structure of *xdhAB* from *R. capsulatus*, is particularly important because it mediates hydrogen bonding with the co-factor FAD (36). Since mutated T220I cannot form hydrogen bonds with FAD, this mutation may affect the xanthine dehydrogenase activity of NBRC 3283.

In order to identify other thermotolerant genes only conserved in the two thermotolerant strains, we performed a BLASTP search against all proteins encoded in the three complete genomes using 24 genes of SKU1108 as the query. BLASTP hits (Table 3) showed that four thermotolerant genes, *lepA*, *ttg2C*, *xdhA*, and *degP*, had paralogs in the chromosome. However, a paralog of *xdhA* (APT\_01390) encoding xanthine dehydrogenase XdhA (APT\_01267) was only conserved in the two thermotolerant strains. This is a part of the *xdhAB* operon (APT\_01390–APT\_01389) that is encoded in specific region 4 found in the genomes of both thermotolerant strains (Fig. 1 and Table S2). Xanthine dehydrogenase is known to regulate heat shock and stress responses (30). This *xdhAB* paralog may partly contribute to the thermotolerance of SKU1108 and 386B. Thus, amino acid mutations accumulated in the thermotolerant genes and their paralogs encoded in the conserved thermotolerance-conferring region may be responsible for the thermotolerance

of the two thermotolerant strains. As described above, two thermotolerant strains have the same large plasmid 1 (Fig. S2). Since these plasmids are only conserved in two thermotolerant strains, we speculated that the genes responsible for thermotolerance are encoded in these two large plasmids. Therefore, we also searched the paralog sequences encoded in these plasmid sequences by a BLASTP search. In addition, a BLASTP search was performed against all other plasmid sequences of the three strains. However, we were unable to identify any paralog sequences corresponding to the 24 thermotolerant genes in these plasmids.

Incidentally, related to the acetic acid resistance or acetic acid assimilation ability of *Acetobacter* species, we found an *aarC* paralog (*aarCI*) sequence in strain-specific region 7. *Acetobacter* species have a specialized citric acid cycle in which the *aarC* gene encoding acetate CoA-transferase plays a crucial role by regulating acetate assimilation (25). A phylogenetic analysis of the AarC protein, conserved in published *Acetobacter* genome sequences, produced the phylogenetic tree that was divided into two large clades: genes belonging to the *aarC* clade are conserved in all *Acetobacter* species, whereas those of the *aarCI* clade are only conserved in genomes of the three *A. pasteurianus* strains studied including SKU1108 (Fig. S4). Thus, it will be interesting to clarify the role of the *aarCI* gene in acetate assimilation.

## Conclusion

In the present study, we elucidated the complete genome sequence of the thermotolerant AAB, *A. pasteurianus* SKU1108, and compared it with those of other thermotolerant and mesophilic *A. pasteurianus* strains. By using a comparative genomic analysis of closely related strains, we revealed several candidate genes that underpin the thermotolerance phenotype. Further investigations of these closely related strains may provide novel insights into genetic causes of their specific phenotypes.

## Acknowledgements

This work was supported financially by the Advanced Low Carbon Technology Research and Development Program (ALCA) of the Japan Science and Technology Agency (JST). Part of this work was performed through a collaboration within the Core to Core Program, which was supported by the Japan Society for the Promotion of Science (JSPS) and the National Research Council of

Thailand (NRCT). This work was also financially supported by JSPS KAKENHI Grant Number 26830126 to MM.

## References

- Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Azuma, Y., A. Hosoyama, M. Matsutani, *et al.* 2009. Whole-genome analyses reveal genetic instability of *Acetobacter pasteurianus*. *Nucleic Acids Res.* 37:5768–5783.
- Bankevich, A., S. Nurk, D. Antipov, *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455–477.
- Bland, C., T.L. Ramsey, F. Sabree, M. Lowe, K. Brown, N.C. Kyrpides, and P. Hugenholtz. 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209.
- Camu, N., T. De Winter, K. Verbrugge, I. Cleenwerck, P. Vandamme, J.S. Takrama, M. Vancanneyt, and L. De Vuyst. 2007. Dynamics and biodiversity of populations of lactic acid bacteria and acetic acid bacteria involved in spontaneous heap fermentation of cocoa beans in Ghana. *Appl. Environ. Microbiol.* 73:1809–1824.
- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17:540–552.
- Chin, C.S., D.H. Alexander, P. Marks, *et al.* 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10:563–569.
- Darling, A.E., B. Mau, and N.T. Perna. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147.
- Edgar, R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
- Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Hyatt, D., G.L. Chen, P.F. Locascio, M.L. Land, F.W. Larimer, and L.J. Hauser. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
- Hyatt, D., P.F. LoCasico, L.J. Hauser, and E.C. Uberbacher. 2012. Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* 28:2223–2230.
- Illegheems, K., L. De Vuyst, and S. Weckx. 2013. Complete genome sequence and comparative analysis of *Acetobacter pasteurianus* 386B, a strain well-adapted to the cocoa bean fermentation ecosystem. *BMC Genomics* 14:526.
- Kanchanarach, W., G. Theeragool, T. Inoue, T. Yakushi, O. Adachi, and K. Matsushita. 2010. Acetic acid fermentation of *Acetobacter pasteurianus*: relationship between acetic acid resistance and pellicle polysaccharide formation. *Biosci. Biotechnol. Biochem.* 74:1591–1597.
- Kurtz, S., A. Phillippy, A.L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S.L. Salzberg. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12.
- Laslett, D., and B. Canback. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32:11–16.
- Lefeber, T., M. Janssens, N. Camu, and L. De Vuyst. 2010. Kinetic analysis of strains of lactic acid bacteria and acetic acid bacteria in cocoa pulp simulation media toward development of a starter culture for cocoa bean fermentation. *Appl. Environ. Microbiol.* 76:7708–7716.
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Matsushita, K., T. Inoue, G. Theeragool, J. Trcek, H. Toyama, and O. Adachi. 2005. Acetic acid production in acetic acid bacteria leading to their ‘death’ and survival, p. 169–181. *In* M. Yamada (ed.), *Survival and Death in Bacteria*. Research Signpost, Kerala.
- Matsushita, K., H. Toyama, and O. Adachi. 1994. Respiratory chain and bioenergetics of acetic acid bacteria, p. 247–301. *In* A.H. Rose, and D.W. Tempest (ed.), *Advances in Microbial Physiology*. vol. 36, Academic Press, London.
- Matsutani, M., H. Hirakawa, N. Saichana, W. Soemphol, T. Yakushi, and K. Matsushita. 2012. Genome-wide phylogenetic analysis of differences in thermotolerance among closely related *Acetobacter pasteurianus* strains. *Microbiology* 158:229–239.
- Matsutani, M., M. Nishikura, N. Saichana, T. Hatano, U. Masud-Tippayarak, G. Theeragool, T. Yakushi, and K. Matsushita. 2013. Adaptive mutation of *Acetobacter pasteurianus* SKU1108 enhances acetic acid fermentation ability at high temperature. *J. Biotechnol.* 165:109–119.
- Matsutani, M., H. Suzuki, T. Yakushi, and K. Matsushita. 2014. Draft genome sequence of *Gluconobacter thailandicus* NBRC 3257. *Stand. Genomic Sci.* 9:614–623.
- Miura, H., T. Mogi, Y. Ano, C.T. Migita, M. Matsutani, T. Yakushi, K. Kita, and K. Matsushita. 2013. Cyanide-insensitive quinol oxidase (CIO) from *Gluconobacter oxydans* is a unique terminal oxidase subfamily of cytochrome *bd*. *J. Biochem.* 153:535–545.
- Mullins, E.A., J.A. Francois, and T.J. Kappock. 2008. A specialized citric acid cycle requiring succinyl-coenzyme A (CoA): acetate CoA-transferase (AarC) confers acetic acid resistance on the acidophile *Acetobacter aceti*. *J. Bacteriol.* 190:4933–4940.
- Petersen, T.N., S. Brunak, G. von Heijne, and H. Nielsen. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8:785–786.
- Richhardt, J., B. Luchterhand, S. Bringer, J. Buchs, and M. Bott. 2013. Evidence for a key role of cytochrome *bo3* oxidase in respiratory energy metabolism of *Gluconobacter oxydans*. *J. Bacteriol.* 195:4210–4220.
- Saeki, A., G. Theeragool, K. Matsushita, H. Toyama, H. Lotong, and O. Adachi. 1997. Development of thermotolerant acetic acid bacteria useful for vinegar fermentation at higher temperatures. *Biosci. Biotechnol. Biochem.* 61:138–145.
- Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069.
- Soemphol, W., A. Deeraksa, M. Matsutani, T. Yakushi, H. Toyama, O. Adachi, M. Yamada, and K. Matsushita. 2011. Global analysis of the genes involved in the thermotolerance mechanism of thermotolerant *Acetobacter tropicalis* SKU1100. *Biosci. Biotechnol. Biochem.* 75:1921–1928.
- Stamatakis, A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stamatakis, A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Stothard, P., and D.S. Wishart. 2005. Circular genome visualization and exploration using CGView. *Bioinformatics* 21:537–539.
- Talavera, G., and J. Castresana. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56:564–577.
- Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30:2725–2729.
- Truglio, J.J., K. Theis, S. Leimkühler, R. Rappa, K.V. Rajagopalan, and C. Kisker. 2002. Crystal structures of the active and alloxanthine-inhibited forms of xanthine dehydrogenase from *Rhodobacter capsulatus*. *Structure* 10:115–125.
- Yamada, Y., P. Yukphan, H.T. Lan Vu, Y. Muramatsu, D. Ochaikul, S. Tanasupawat, and Y. Nakagawa. 2012. Description of *Komagataeibacter* gen. nov., with proposals of new combinations (*Acetobacteraceae*). *J. Gen. Appl. Microbiol.* 58:397–404.