Sains Malaysiana 54(3)(2025): 769-783
http://doi.org/10.17576/jsm-2025-5403-13

Identification and *In-Silico* Analysis of a Cellulose Synthase-Like Gene on Eddoe Taro

(Pengenalpastian dan Analisis *In-Silico* Gen Selulosa Seperti Sintase pada Eddoe Taro)

DELVI MARETTA[1,*], IS HELIANTI[2], DEVIT PURWOKO[3], EDI SANTOSA[4] & SOBIR[4]

[1]*Research Center for Horticulture, National Research and Innovation Agency, Republic of Indonesia (BRIN), Cibinong Science Center, Jl. Raya Bogor KM 46 Cibinong, West Java 16915, Indonesia*
[2]*Research Center for Genetic Engineering, National Research and Innovation Agency, Republic of Indonesia (BRIN), Cibinong Science Center, Jl. Raya Bogor KM 46 Cibinong, West Java 16915, Indonesia*
[3]*Research Center for Applied Botany, National Research and Innovation Agency, Republic of Indonesia (BRIN), Cibinong Science Center, Jl. Raya Bogor KM 46 Cibinong, West Java 16915, Indonesia*
[4]*Department Agronomy and Horticulture, Faculty of Agriculture, Bogor Agricultural University, Jl Meranti Kampus IPB Darmaga, Bogor, West Java 16680, Indonesia*

Received: 14 May 2024/Accepted: 16 December 2024

ABSTRACT

Glucomannan is a non-starch carbohydrate predominantly found in tubers, serving as a significant resource for food and health industries. Despite the widely recognized glucomannan content in taro tubers, there is limited understanding regarding the gene level. Therefore, this research aimed to identify putative Cellulose Synthase-like A (*CslA*) gene sequences associated with glucomannan biosynthesis in eddoe taro plant. Genome isolation was carried out on six genotypes of eddoe taro, each showing different glucomannan content. A pair of primers designed from the mannan synthase encoding gene sequences obtained from the NCBI. Subsequently, sequences of the PCR product were analyzed for identification and *in-silico* analysis. The result of *in-silico* RFLP analysis showed that six genotypes had polymorphic allelic fragments. The DNA sequences showed a high similarity to *CslA* gene, among representative taro tubers compared to the reference plants. A total of three nucleotide sequences fragments from the S7 and S34 genotypes as well as two from S15, S28, S30, and S36, corresponded to *CslA* gene of *Amorphophallus konjac.* Phylogenetic analysis based on nucleotide sequences showed that S7 and S34 had distinctive characteristics, indicating specific and wide adaptation, respectively. Despite the presence of single nucleotide polymorphism, the *in-silico* transcription-translation showed that the protein constructed had a highly similar consensus motif. These results suggested the identified sequences as a potential *CslA*-encoding gene that had functioned in the biosynthesis of mannan synthase to produce glucomannan in taro plants.

Keywords: Biosynthesis; *CslA* gene; glucomannan; sequences polymorphism; taro breeding

ABSTRAK

Glukomanan ialah karbohidrat bukan-kanji dalam ubi dan digunakan secara meluas dalam industri makanan dan kesihatan. Ubi keladi diketahui mengandungi glukomanan, namun pengetahuan berkaitan perkara ini terutamanya di tahap gen masih sedikit. Tujuan penyelidikan ini untuk mengenal pasti jujukan gen selulosa seperti sintase A (*CslA*) yang mungkin memainkan peranan dalam biosintesis glukomanan pada keladi jenis eddoe. Pengasingan genom telah dijalankan pada enam genotip keladi eddoe yang mempunyai perbezaan yang ketara pada kandungan glukomanan. Sepasang primer yang direka daripada urutan gen pengekodan manan sintase yang diperoleh daripada bank gen NCBI. Jujukan produk PCR dianalisis untuk pengenalpastian dan analisis *in silico*. Hasil analisis *in silico* RFLP bagi jujukan menunjukkan bahawa enam genotip mempunyai serpihan alel polimorfik. Analisis jujukan DNA menunjukkan persamaan yang tinggi dengan gen *CslA*, antara keladi perwakilan berbanding dengan tumbuhan rujukan. Terdapat tiga serpihan jujukan nukleotida daripada genotip S7 dan S34 dan dua serpihan daripada S15, S28, S30 dan S36 yang sepadan dengan gen *CslA Amorphophallus konjac*. Analisis filogenetik berdasarkan jujukan nukleotida menunjukkan bahawa S7 dan S34 dipisahkan daripada genotip lain. Genotip S34 telah dikenal pasti sebagai adaptasi sekitaran luas manakala S7 mempunyai adaptasi lokasi khusus. Penjajaran menemui polimorfisme nukleotida tunggal tetapi terjemahan transkripsi *in silico* mendedahkan penjajaran protein dengan motif konsensus serupa yang tinggi. Ia dianggap bahawa jujukan yang ditemui ialah gen pengekodan *CslA* yang berpotensi yang telah berfungsi dalam biosintesis manan sintase untuk menghasilkan glukomanan dalam keladi.

Kata kunci: Biosintesis; gen *CslA;* glukomanan; pembiakan talas; polimorfisme jujukan

INTRODUCTION

Mannan polysaccharides are found in various types of plants, serving as hemicellulose components in cell wall structure, carbohydrate reservation, and other potentially essential functions. As a storage, mannan functions in the form of non-starch carbohydrates, which are stored in seeds and vegetative tissues (Liepman et al. 2007; Moreira & Filho 2008). Plants of the Araceae family often possess underground stem, namely corm, which serve as storage organs containing glucomannan, a polysaccharide in mannan family. Glucomannan is commonly found in the corm of Porang (*Amorphophallus muelleri* Blumei) (Wigoeno, Azrianingsih & Roosdiana 2013), iles-iles tubers (*Amorphophallus campanalatus* Blumei), taro tubers (Chairul & Chairul 2006), and *Amorphophallus konjac* tubers (Gille et al. 2011).

Glucomannan is an auspicious polysaccharide, often used in the food, health, and cosmetic industry. This polysaccharide has been examined as a pharmaceutical excipient in hydrogel, tablets, and beads, due to the ability to combine with other polymers and increased flexibility in the drug delivery field (Alonso-Sande et al. 2009). Moreover, glucomannan supplementation can reduce the body weight in overweight or obese because of the soluble fiber (Birketvedt et al. 2005; Keithley & Swanson 2005; Zalewski et al. 2015), beneficial for constipation treatment (Baucke, Miele & Staiano 2004), and alternative therapy for type 2 diabetes mellitus (Karima, Elya & Sauriasari 2023; Mashudi et al. 2023). In the food industry, it is also used as an additive, gelling agent, and cryoprotectant (Yang et al. 2017).

The taro genotype with superiority in glucomannan content has been morphologically characterized. Previous research investigated that among 14 genotypes, accession number S7, S15, and S34 were high glucomannan content. Specifically, S34 had wide location adaptation, S17, and S7 showed specific location adaptation, while S28, S30, and S36 possessed low glucomannan content (Maretta et al. 2023). According to clustering analysis, S7, S15, and S17 were shown as introduction cultivars, with S28, S30, and S36 being suggested as originally local varieties of Indonesian eddoe taro (Maretta et al. 2020).

The current advancements in molecular biology, mainly DNA sequencing technologies, have delivered big data and resulted in the occurrence of novel discoveries in biological research. The broad use, numerous applications, and *in-silico* analysis have revolutionized genetic research and increased understanding of genomic (Chauhan et al. 2023; Yadav et al. 2022). This recent advancement also facilitates the comprehension of gene role in the biosynthesis of glucomannan in plants. Generally, mannan polysaccharides are engaged in several growth and developmental processes. At the genetic level, the synthesis of glucomannan under the influence of *CslA* gene is considered throughout β–manan synthase enzyme production. This is because *CslA* gene family encodes glucomannan synthases, suggesting a role in metabolic system reserves, cell wall structure, and storage of carbohydrates.

Previous research on the expression patterns of *CslA* gene and the deposition of mannan polysaccharides suggests physiological functions beyond construction and storage (Liepman, Wilkerson & Keegstra 2005; Liepman et al. 2007). For example, in *Populus trichocarpa* plant, *CslA1* and *CslA3* encode β–manan synthase, producing β-D-glucomannan (Suzuki et al. 2006). The expression of this gene has been observed in flower tissue, stem, and embryo of Arabidopsis, including the chloronemal filament of *Physcomitrella patens* (Goubet et al. 2009). In the tuberized plant of *Amorphophallus konjac*, the high abundance of the *AkCSLA3* transcript in developing corm using the substrates GDP-D-mannose and GDP-D-glucose shows the role of *CslA1* for the synthesis of storage glucomannan (Gille et al. 2011). Although *CslA* gene has been detected in 100 plant species (Liepman et al. 2007), there is still limited knowledge regarding the identification process in the taro plant. Therefore, this research aimed to identify and characterize *CslA* gene from different genotypes of eddoe taro with the variability of glucomannan content through *in-silico* analysis using *A. konjac CslA* gene as a reference.

MATERIAL AND METHODS

PLANT MATERIAL AND GENOME EXTRACTION

A total of six accessions of Indonesian eddoe taro, namely S7, S15, S28, S30, S34, and S36, were used as plant material representatives. The young leaves of each accession were used for genome isolation through DNA extraction kit (Geneaid™, China). Plant tissue dissociation was achieved through instantaneous grinding after tissue sample freeze with liquid nitrogen. Subsequently, 70-80 mg of sample powder was added with buffer and RNAse solution provided for the process of tissue lysis. This was followed by DNA binding, washing, and elution with buffer as well as filter column tubes included in the kit package. The duration and speed of the centrifugation process were determined according to the provided protocol.

DESIGN OF cDNA BASE PRIMER

Primer design was initiated by identifying sequences of encoding gene of mannan synthase (*CslA03*) in the NCBI gene bank. The obtained sequences were the complete cDNA sequences of the putative mannan synthase mRNA of the *Amorphophallus konjac* plant, with a size of 1560 bp (NCBI sequence ID HQ833588.1) (Gille et al. 2011). Furthermore, sequences BLASTn (Basic Local Alignment Search Tool) were carried out against the whole genome shotgun sequences from the *Colocasia esculenta* cultivar, namely Longxiangyu. This confirmed that the putative sequences *CslA03* was also positioned on Chromosome

number nine of *Colocasia esculenta* cultivar Longxiangyu ecotype Jiangsu (NCBI sequence ID: CM024761.1.). The BLASTn process of ID HQ833588.1 launched four segments of nucleotide sequences with homology levels exceeding 85%, as shown in Table 1.

During the BLASTn process of ID HQ833588.1, four segments were used for online primer design on the Primer3Plus website (https://www.primer3plus.com/). The retrieved primer pairs were arranged in line with the 5000 bp nucleotide of chromosome nine sequences of *Colocasia esculenta* (NCBI ID: CM024761.1), ranging from 27,347,222 to 27,342,222. Subsequently, primer conformity with the ideal criteria was evaluated using http://www.premierbiosoft.com/netprimer. The selected primer that was used in the PCR amplification stage was Forward 5' - 3' GGGGCCAGGTAAGAGGTATC and Reverse 5' - 3' CCTGAACGAGAGCAATCTCC, with a PCR product size was 1500 bp.

PCR AMPLIFICATION

The total volume of PCR mix was 20 µL, consisting of 2 µL free nuclease water, 10 µL (2x) MyTaq™ HS Red Mix Bioline, 2 µL of diluted forward primer 10 µM, 2 µL diluted reverse primer 10 µM, and 4 µL DNA template. In this research, the amplification process carried out was divided into three main stages. The first stage was pre-denaturation for 3 min at 95 °C, followed by 30 cycles of the second stage, which included denaturation for 0.5 min at 95 °C, annealing for 0.5 min at 58.5 °C and extension for 1 min at 72 °C. Meanwhile, the last stage was a final extension for 5 min at 72 °C. The product of PCR was confirmed by visualization of gel electrophoresis (0.8 % agarose gel, 1xTBE buffer, gel staining Greensafe^{TM} Premium 4 µL/100 mL, 30 min, voltage 100 V). Positive amplification was shown by the presence of a band sized at 1500 bp on the gel electrophoresis.

NUCLEOTIDE SEQUENCING

The single band of PCR product, corresponding to the targeted size, was used as sequencing material. The examination was carried out by 1stBASE (First BASE Laboratories Sdn Bhd, Malaysia), which started through PCR product purification with Zymoclean™ Gel DNA Recovery Kit (Zymo Research, D4001). Subsequently, the purified PCR product was cloned to pTA2 Vector with Toyobo Target Clone (Toyobo, TAK-101) and transformed to *Escherichia coli* DH5α with Mix and Go Competent *E. coli* Cells (Zymo Research, T3007). PCR was conducted to the colonies using (2x) MyTaq HS Red Mix Bioline (BIO-25048), along with primer T3 and T7 promotor. Positive insert PCR results were confirmed by the presence of a DNA band measuring approximately 1500 bp on TBE 0.8% agarose gel electrophoresis. Plasmid extraction was carried out using ZR Plasmid Miniprep™-Classic (Zymo Research, D4015) and the colony of PCR product was used for bi-directional sequencing. Moreover, the nucleotide sequences obtained from six genotypes of eddoe taro were compiled as data for further analysis. These DNA sequences were analyzed to determine the degree of homology through arrangement using Geneious Prime 10.1.3 software and online NCBI.

RFLP *In-silico*

Candidate of enzyme restriction in this step was screened online http://insilico.ehu.es/restriction/. Among the tools provided, the 'Compare restriction pattern of many sequences' was selected and the FASTA formatted sequences were entered. This was followed by selecting the option 'Only Restriction Enzymes with Known Bases (no N, R, Y…)' to obtain clear restriction enzymes without N, R, or Y bases. The process was followed by option 'Get List of Restriction Enzymes' and the selected enzyme restriction was used for further analysis (Aulia et al. 2021).

In-silico RFLP was conducted virtually on the site https://www.benchling.com/, which required registration through an e-mail account. The process started by importing sequences of the six genotypes obtained from previous laboratory work in the Benchling site. Subsequently, the scissor symbol was clicked and restriction enzymes were typed into the 'Find Enzymes' menu. The procedure was continued by selecting the 'Run Digest' button for restriction and the output appeared in the 'Virtual Digest' tools in the pattern of electropherograms (Yeriska et al. 2021).

TABLE 1. The homolog segment of putative sequences of mannan synthase *Amorphophallus konjac** to chromosome 9 of *Colocasia esculenta***

Sequences Range **	Identities
Range 1: 27.343.816 to 27.343.955	127/140 (91%)
Range 2: 27.344.291 to 27.344.400	103/110 (94%)
Range 3: 27.345.695 to 27.345.953	227/259 (88%)
Range 4: 27.346.942 to 27.347.222	240/281 (85%)

Sequence query Id NCBI HQ833588.1(*); CM024761.1 (**)

In-silico PROTEIN ANALYSIS

Mannan synthase gene was evaluated based on the identification domains in the NCBI conserved database, specifically for *CslA* gene. The domains were recognized as a part of the web-based Blast interface (NCBI) according to the CDD (v3. 10-44354 PSSMs), containing an RPS-Blast search against those matrices (Marchler-Bauer et al. 2013). The retrieved sequences were further assessed through Blastp searches reciprocally, to confirm their similarity to proteins of the corresponding family. The majority of searches at NCBI were performed applying the non-redundant protein database (http://www.phytozome.net/) and protein arrangement was carried out by CLUSTALW (Thompson, Gibson & Higgins 2003), followed by editing and manual handling with BioEdit (Hall 1999).

Mannan synthase gene was affixed to the respective genus and species, followed by sequences cutting to retain only the relevant protein domains for phylogenetic analysis of conserved domains. MEGA software in the mode of neighbor joining (NJ) with 100 bootstrapping repetitions was used for phylogenetic tree construction based on sequences to assign evolutionary trends and distribution of mannan in plants. The proteins with three-dimensional structure were executed by the PHYRE^2 server and the feedback was accepted as the PDB (Protein Data Bank) configuration (Kelley & Sternberg 2009). Subsequently, the PDB format was provided to YASARA software to generate the three-dimensional structure (Krieger, Koraimann & Vriend 2002).

RESULTS AND DISCUSSION

A total of six genotypes were selected based on variations in glucomannan content from previous research. Genotypes S7, S15, and S34 were strongly identified as introduction cultivars, while S28, S30, and S36 were considered to be the original Indonesian local eddoe taro (Maretta et al. 2020). Phenotypically, accession S28, S30, and S36 had low glucomannan, with S7, S15 and S34 showing high content. Based on production, accession S34 was confirmed as genotype with wide location adaptation, while S7 had specific location adaptation (Maretta et al. 2023). In this research, genome extraction was performed on samples and the collected DNA showed a valuable concentration and pureness ($A_{260/\ A280}$) for PCR amplification. The designed primer was used to amplify DNA samples, showing compatibility with Eddoe taro genome. The PCR product was consistently sized at 1500 bp and was identical among tested accessions, as shown in Figure 1.

Electrophoresis visualization on TBE agarose confirmed the presence of a positive inserted clone in the pTA2 vector. The resultant plasmid of colonies number 2 (S7 and S34), 3 (S15 and S28), 6 (S30), and 8 (S36) were selected for the sequencing process, as shown in Figure 2. Sequences output was used for *in-silico* RFLP analysis, arrangement, as well as *in-silico* amino acid and protein translation.

In-silico RFLP on the nucleotides of *CslA* gene sequences of eddoe taro and *A. konjac* was exercised using *Msc* I restriction enzyme, which recognized specific locations between 4 and 9 nucleotides. The *Msc* I restriction enzyme was selected from the screening process and the recognition digest site was identified as TGGCCA of the PCR product sequences.

Digestion of sequences of *A. konjac CslA*03 mRNA (NCBI ID: HQ833588.1) produced three fragments sized 850, 750, and 75 bp. The completion of the digestion of six taro sequences exposed variative allelic fragments. Specifically, genotypes S36, S34, and S15 had 350 bp sized allele in length, while each of S30, S28, and S7 had 370 bp, 250 bp, and 150 bp, respectively. As shown in Figure 3, *A. konjac* had a distinct allele sized 750 bp compared to

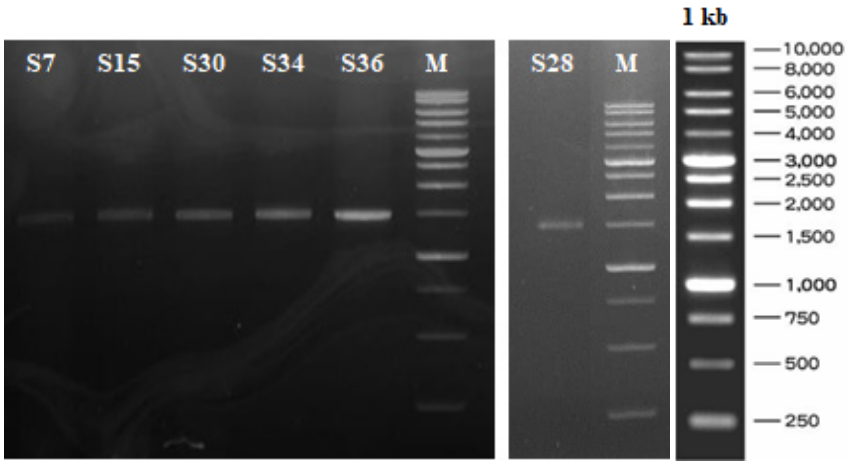

FIGURE 1. PCR product of six different genotypes of eddoe taro

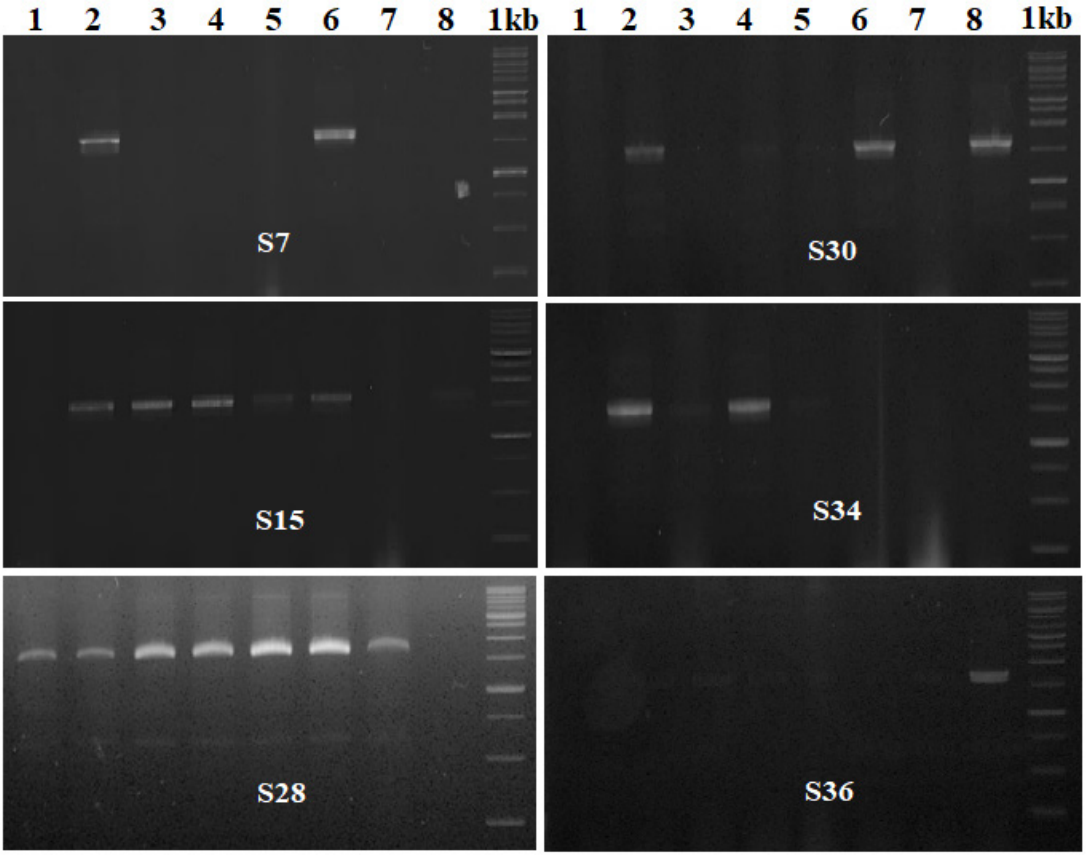

FIGURE 2. Isolation of plasmid containing *CslA* gene from six different genotypes of eddoe taro

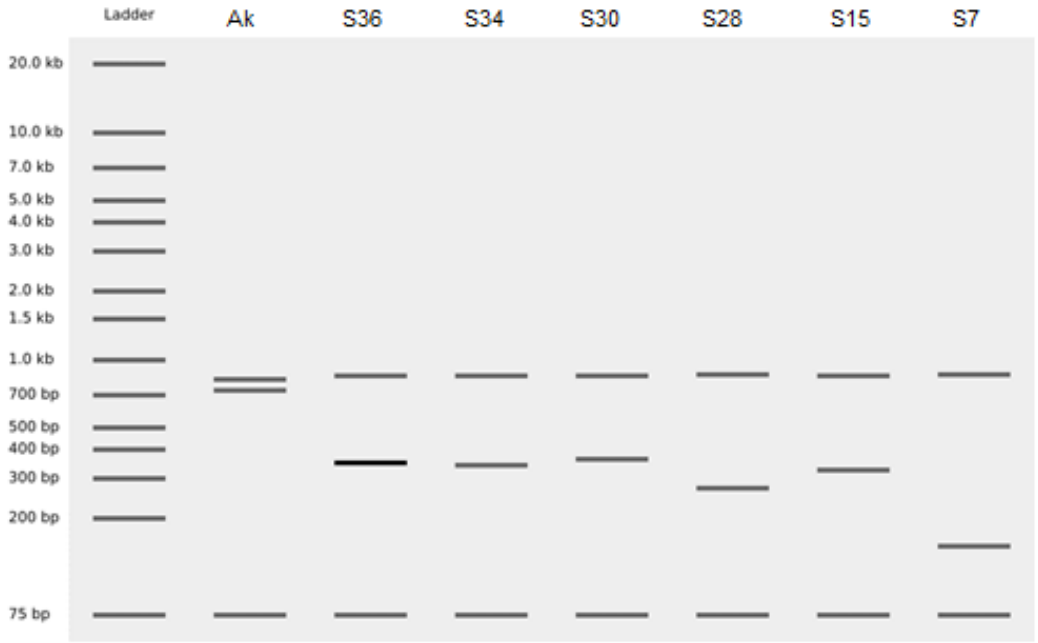

FIGURE 3. Electropherogram sequences restriction uses *Msc*l enzyme of *CslA* sequences gene; Ladder 1 kb; Ak (*A. konjac CslA03* mRNA complete cds as a reference)

the investigated eddoe taro. The RFLP method served as a codominant marker that used restriction enzymes at specific sites, resulting in varying length of fragments (Theodore 2000). This method was used to estimate DNA variation, through restriction sites at specific DNA sequences, as changes in the presence of restriction sites corresponded to differences in length of fragment. The differences in allele sizes observed in this research suggested the possibility of genetic diversity at the DNA level in eddoe taro plants.

The arrangement among genotypes showed a similarity level of more than 95%, with a distance matrix among sequences of six genotypes, as presented in Table 2. The highest similarity of 99.73% was observed between S15 and S30 due to the elevated value of the conserved area, reaching 94.1% across 1419 of 1508 nucleotides.

Neighbor joining analysis showed that the six genotypes examined were divided into two main clusters. As shown in Figure 3, genotypes S15, S28, S30, S34, and S36 were constructed in the first cluster, while S7 was in the second cluster. The phylogenetic diagram in Figure 4 was consistent with the matrix output, showing the furthest distance of 0.02 unit, specifically between S7 and the other five genotypes as well as S34 with S7 and S28 (Table 2). Previous research had established that S15, S28, S30, and S36 were landrace genotypes, while S7 and S34 were introduced accessions (Maretta et al. 2020). The phylogenetic diagram based on *CslA* gene sequences data showed a symmetrical pattern consistent with previous results. This showed that the gene sequences observed distinguished the landrace and introduced accession group. Habitat fragmentation could also influence patterns of genetic variation caused by human activities in the use of plant genetic resources, with implications for conservation strategies (Gao et al. 2017).

A. konjac is a bulbous plant of Araceae family that is often used as a model for glucomannan biosynthesis due to its high contents (Gao et al. 2022, 2017; Gille et al. 2011). In this research, the NCBI BLASTn algorithm analysis of six nucleotide sequences showed a comparison with a reference plant (accession ID: HQ833588.1) in the range of 87 - 91% (Table 3), with consensus identities of the three fragments of genotype S7 described in Supplement 1. The algorithm analysis of six nucleotide sequences of taro to the cDNA of putative mannan synthase of *A. konjac* (accession ID: HQ833588.1) obtained nucleotide similarities in length ranging from 102 to 248 bases. This resulted in the construction of three corresponding fragments of genotypes S7 and S34, including two for S15, S28, S30, and S36, as shown in Table 3. The S7 had the highest homogeneity at 90.91%, while S34 reached approximately 90.19% with putative gene *CslA03* of the comparison plant, confirming the close genetic distance reported by Gao et al. (2022, 2017) and Hu et al. (2019).

Sequences were arranged to distinguish the nucleotide polymorphism among *CslA* gene of taro plant. The arrangement output showed differing nucleotides at 21 sites for S7 and 8 sites for S34 compared to other sequences. Moreover, 10 nucleotide sites were found to be identical to S7 and S34 but distinct to S15, S28, S30, and S36 sequences, as shown in Supplement 2. The nucleotide sequences of a gene, through the medium of mRNA, were *in-silico* translated into the amino acid, as presented in Figure 5. The retrieved DNA→mRNA→protein *in-silico* transcription-translation test showed a significant amount of similarity with minimal variance. The protein arrangement constructed 98.9% similar to consensus motif, while mismatch was found in three positions according to the consensus sequences. The first was at the 22nd position namely aspartate (D), which was glycine (G) in S7, the 41st position was predominantly lysine (K) but isoleucine (I) in the S30, the 57th position was phenylalanine (F) but leucine (L) in the S34.

The protein sequences obtained were used to construct a three-dimensional structure of β–manan synthase protein. However, the mismatch did not affect the structure of the β–manan synthase protein. Retrieved from PHYRE analysis, the amino acid translation released homolog of the β-manan synthase protein across six candidate *CslA* gene, as shown in Figure 6. The homolog protein was expected to have identical glucomannan characteristics in expression level. However, further research on the expression level is recommended to understand the roles of *CslA* gene on glucomannan production in taro corm.

The differences in discovered sequences, the number of similar fragments to the *A. konjac* as comparison plant, single nucleotide polymorphism, and the allelic polymorphism of *CslA* gene can affect the level, stability, and adaptability of the produced glucomannan in eddoe taro corm. In this research, genotypes S7 and S34 had a higher percentage of homogeneity and a greater number of the corresponding fragment to putative *CslA* gene compared to other examined genotypes. Previous multilocation research (Maretta et al. 2023) showed that genotypes S7 and S34 often produced a stable and high glucomannan content in the three different experimental locations. An examination of both genotypes showed that S7 had specific location adaptation, while S34 had broad adaptation. Genotype S15 also had high glucomannan content but was unstable due to variations in environmental conditions, while S28, S30, and S36 were relatively low. Additionally, single nucleotide polymorphisms found in plant gene sequences contributed to the liability encode distinctness, uniformity, and stability of traits (Jo et al. 2022; Mahendhiran et al. 2014). These results provide valuable information for breeding program to develop high glucomannan taro variety, although further research on *CslA* gene is recommended. Variation at the nucleotide arrangement can also be exploited for plant breeding and variety protection by developing SNPs based marker technology to discover genes coding for important traits in commodities (Tasma 2015). After characterization, further evaluation should be carried out to determine the effectiveness of SNPs as a marker for taro germplasm screening or marker-assisted selection.

TABLE 2. Similarity and distance matrix of sequences of six genotypes

Genotype	S7		S15		S28		S30		S34		S36
	%										
S7	1000		-		-		-		-		-
S15	95.48	0.02	1000		-		-		-		-
S28	95.41	0.02	99.39	0.01	1000		-		-		-
S30	95.61	0.02	99.73	0	99.53	0	1000		-		-
S34	95.42	0.02	97.12	0.01	96.92	0.02	97.25	0.01	1000		-
S36	95.28	0.02	99.39	0	99.19	0	99.53	0	96.92	0.01	1000

TABLE 3. BLASTn sequences of six genotypes and putative mannan synthase of *A. konjac* mRNA gene sequences (NCBI ID: HQ833588.1)

Genotype	% identities	Arrangement length	Start to end of Genotype sequences	Start to end of *A. konjac* sequences	Number of corresponding fragments
S7	90.91	220	1243 - 1462	281 - 62	3
	87.08	240	1 - 240	616 - 377	
	90.19	102	1062 - 1163	379 - 278	
S15	87.50	248	1230 - 1477	281 - 34	2
	87.50	240	1 - 240	616 - 377	
S28	87.50	248	1231 - 1478	281 - 34	2
	87.50	240	1 - 240	616 - 377	
S30	87.50	248	1230 - 1477	281 - 34	2
	87.08	240	1 - 240	616 - 377	
S34	87.50	248	1239 - 1485	281 - 34	3
	87.08	240	1 - 240	616 - 377	
	90.19	102	1046 - 1147	379 - 278	
S36	87.50	248	1226 - 1473	281 - 34	
	87.50	240	1 - 240	616 - 377	2

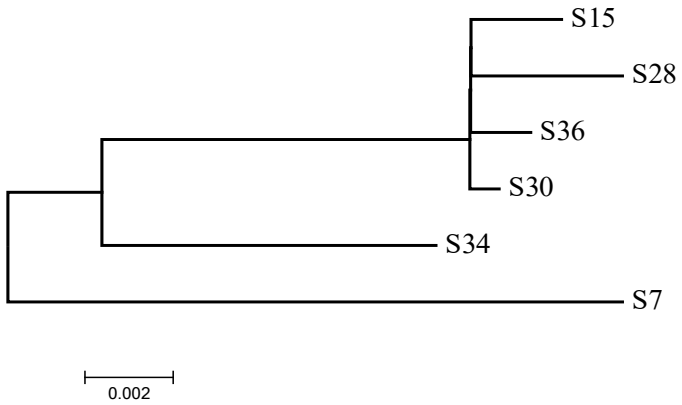

FIGURE 4. Phylogenetic diagram of six genotypes eddoe taro based on CslA nucleotide sequences

```
Consensus                          SGSPGLQEFPILLNESNLRVVDEKWCCLHKEVGFWLEVGIKDGNIVARLHISILHSFLEG  60
S7 extraction translation          SGSPGLQEFPILLNESNLRVVGEKWCCLHKEVGFWLEVGIKDGNIVARLHISILHSFLEG  60
S15 extraction translation         SGSPGLQEFPILLNESNLRVVDEKWCCLHKEVGFWLEVGIKDGNIVARLHISILHSFLEG  60
S28 extraction translation         -GSPGLQEFPILLNESNLRVVDEKWCCLHKEVGFWLEVGIKDGNIVARLHISILHSFLEG  59
S30 extraction translation         -GSPGLQEFPILLNESNLRVVDEKWCCLHKEVGFWLEVGIIDGNIVARLHISILHSFLEG  59
S34 extraction translation         SGSPGLQEFPILLNESNLRVVDEKWCCLHKEVGFWLEVGIKDGNIVARLHISILHSILEG  60
S36 extraction translation         -GSPGLQEFPILLNESNLRVVDEKWCCLHKEVGFWLEVGIKDGNIVARLHISILHSFLEG  59

Consensus                          TCFVPIPVGPGLIFYVHPRRSPFLAFFVHKI                                91
S7 extraction translation          TCFVPIPVGPGLIFYVHPRRSPFLAFFVHKI                                91
S15 extraction translation         TCFVPIPVGPGLIFYVHPRRSPFLAFFVHKI                                91
S28 extraction translation         TCFVPIPVGPGLIFYVHPRRSPFLAFFVHKI                                90
S30 extraction translation         TCFVPIPVGPGLIFYVHPRRSPFLAFFVHKI                                90
S34 extraction translation         TCFVPIPVGPGLIFYVHPRRSPFLAFFVHKI                                91
S36 extraction translation         TCFVPIPVGPGLIFYVHPRRSPFLAFFVHKI                                90
```

FIGURE 5. ClustalW multiple sequences arrangement of *CslA* translation of eddoe taro

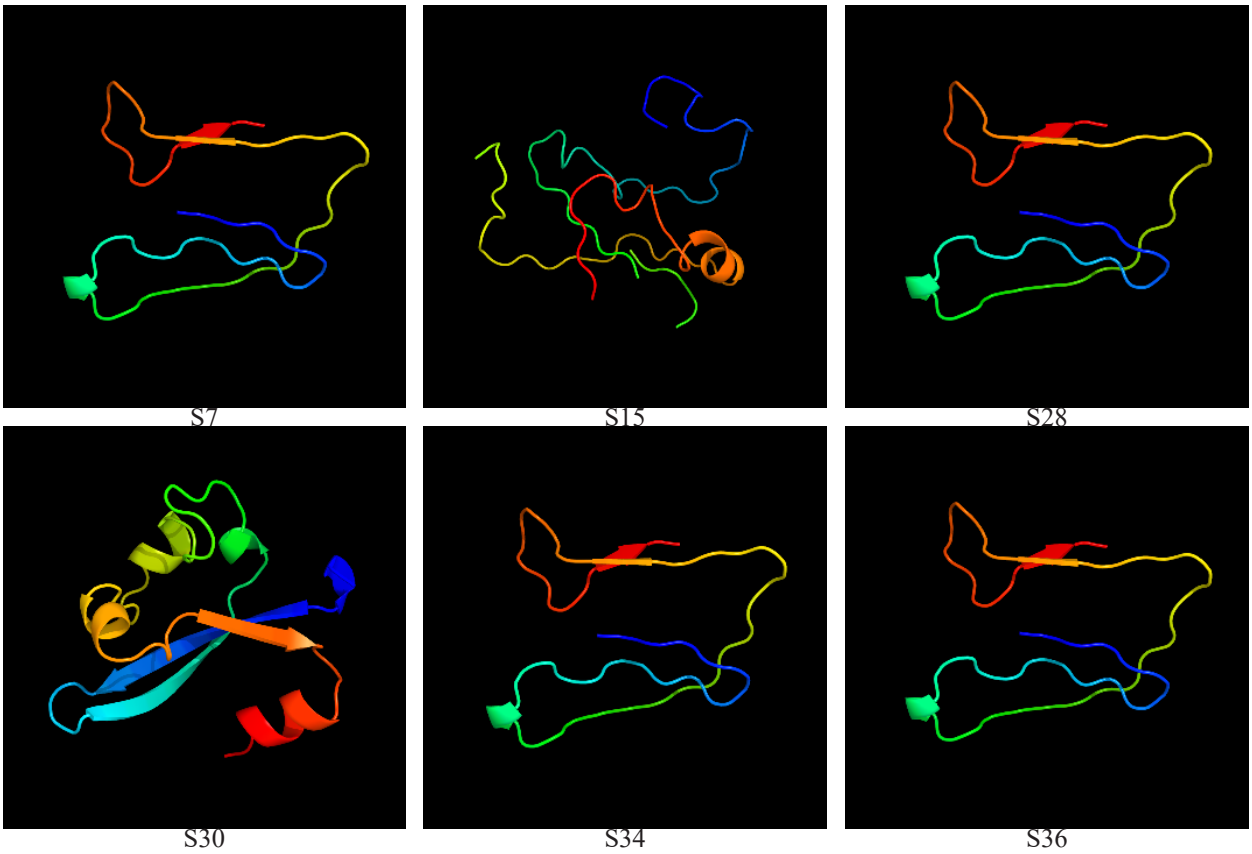

FIGURE 6. Predicted three-dimensional structure of β–manan synthase protein

CONCLUSIONS

In conclusion, this research showed that *CslA* gene sequences of representative taro were highly similar to *A. konjac*. Despite the similarity among the six genotypes examined, SNPs were detected, without significant effect on the structure of β-manan synthase protein. Sequences analysis was in line with previous research based on morphological characters, distinguishing between taro with high and low glucomannan content, which was stable in different growing environments. Based on the results, sequences found were considered as potentially *CslA* gene, playing a role in the biosynthesis of mannan synthase to produce glucomannan in taro plants. However, further research was recommended to understand the function of *CslA* gene in glucomannan production in taro tubers to support breeding programs.

ACKNOWLEDGMENTS

The authors are grateful to the Ministry of Research and Technology and The Agency of Application and Assessment of Technology Indonesia, which was established in 2021 as the National Research and Innovation Agency, for financial support through Saintek Scholarship and providing of research facilities as shown in the Decision Letter of Head of BPPT No 244/2017.

REFERENCES

Alonso-Sande, M., Teijeiro-Osorio, D., Remuñán-López, C. & Alonso, M.J. 2009. Glucomannan, a promising polysaccharide for biopharmaceutical purposes. *Journal of Pharmaceutics and Biopharmaceutics* 72(2): 453-462. doi:10.1016/j.ejpb.2008.02.005

Aulia, N., Putri, C.K., Khairiyah, L., Khairunnisa, F.A. & Achyar, A. 2021. Analysis of genetic variation of measles virus genotype and isolate M-F intergenic spacer sequences PopSet: 2105287799 using Restriction Fragment Length Polymorphism (RFLP) *in-silico*. *Prosiding Seminar Nasional Biologi* 1(2): 1139-1146. https://semnas.biologi.fmipa.unp.ac.id/index.php/prosiding/article/view/298/224

Baucke, L., Miele, E. & Staiano, A. 2004. Fiber (glucomannan) is beneficial in the treatment of childhood constipation. *Pediatrics* 113(3): e259-e264. doi:10.1542/peds.113.3.e259

Birketvedt, G.S., Shimshi, M., Erling, T. & Florholmen, J. 2005. Experiences with three different fiber supplements in weight reduction. *Medicinal Science Monitoring* 11(1): PI5-8.

Chairul, C. & Chairul, S.M. 2006. Isolasi glukomanan dari dua jenis araceae: Talas (*Colocasia esculenta* (L.) Schott) dan ıles-ıles (*Amorphophallus campanulatus* Blumei). *Berita Bio* 8(3): 171-178.

Chauhan, A., Modgil, M., Rajam, M.V.R., Sharma, J.N. & Siddappa, S. 2023. Isolation, cloning and *in-silico* analysis of β-tubulin gene from apple leaf blotch fungus *Marssonina coronaria*. *Indian Phytopathology* 76: 371-382. https://doi.org/10.1007/s42360-023-00614-z

Gao, Y., Zhang, Y., Feng, C., Chu, H., Feng, C., Wang, H., Wu, L., Yin, S., Liu, C., Chen, H., Li, Z., Zou, Z. & Tang, L. 2022. A chromosome-level genome assembly of *Amorphophallus konjac* provides insights into konjac glucomannan biosynthesis. *Computational and Structural Biotechnology Journal* 20: 1002-1011. doi:10.1016/j.csbj.2022.02.009

Gao, Y., Yin, S., Yang, H., Wu, L. & Yan, Y. 2017. Genetic diversity and phylogenetic relationships of seven Amorphophallus species in southwestern China revealed by chloroplast DNA sequences. *Mitochondrial DNA Part A* 29(5): 679-686. doi:10.1080/24701394.2017.1350855

Gille, S., Cheng, K., Skinner, M.E., Liepman, A.H., Wilkerson, C.G. & Pauly, M. 2011. Deep sequencing of voodoo lily (*Amorphophallus konjac*): An approach to identify relevant genes involved in the synthesis of the hemicellulose glucomannan. *Planta* 234(3): 515-526. doi:10.1007/s00425-011-1422-z

Goubet, F., Barton, C.J., Mortimer, J.C., Yu, X., Zhang, Z., Miles, G.P., Richens, J., Liepman, A.H., Seffen, K. & Dupree, P. 2009. Cell wall glucomannan in Arabidopsis is synthesised by *CSLA* glycosyltransferases, and influences the progression of embryogenesis. *The Plant Journal* 60(3): 527-538. doi:10.1111/j.1365-313X.2009.03977.x

Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series*. pp. 95-98.

Hu, H., Liu, J., Wang, B., An, J. & Wang, Q. 2019. Characterization of the complete chloroplast genome of *Amorphophallus konjac* (Araceae) and its phylogenetic analysis. *Mitochondrial DNA Part B* 4(1): 1658-1659. doi:10.1080/23802359.2019.1606683

Jo, J., Kang, M.Y., Kim, K.S., Youk, H.R., Shim, E.J., Kim, H., Park, J.S., Sim, S.C., Yu, B.C. & Jung, J.K. 2022. Genome-wide analysis-based single nucleotidepolymorphism marker sets to identify diverse genotypes in cabbage cultivars (*Brassica oleracea* var. capitata). *Scientific Reports* 12: 20030. https://doi.org/10.1038/s41598-022-24477-y

Karima, R., Elya, B. & Sauriasari, R. 2023. Mechanism of action of glucomannan as a potential therapeutic agent for type 2 diabetes mellitus. *Tropical Journal of Natural Product Research* 7(12): 5460-5469. DOI 10.26538/tjnpr/v7i12.15

Keithley, J. & Swanson, B. 2005. Glucomannan and obesity: A critical review. *Alternative Therapies in Health and Medicine* 11(6): 30-34.

Kelley, L.A. & Sternberg, M.J.E. 2009. Protein structure prediction on the web: A case study using the Phyre server. *Nature Protocols* 4(3): 363-371.

Krieger, E., Koraimann, G. & Vriend, G. 2002. Increasing the precision of comparative models with YASARA NOVA- A selfparameterizing force field. *Proteins: Structure, Function and Genetics* 47(3): 393-402.

Liepman, A.H., Wilkerson, C.G. & Keegstra, K. 2005. Expression of cellulose synthase-like (*Csl*) genes in insect cells reveals that *CslA* family members encode mannan synthases. *Proceedings of the National Academy of Sciences of the United States of America* 102(6): 2221-2226. doi:10.1073/pnas.0409179102

Liepman, A.H., Nairn, C.J., Willats, W.G.T., Sorensen, I., Roberts, A.W. & Keegstra, K. 2007. Functional genomic analysis supports conservation of function among cellulose synthase-like a gene family members and suggests diverse roles of mannans in plants. *Plant Physiology* 143(4): 1881-1893. doi:10.1104/pp.106.093989

Mahendhiran, M., Ramirez, P.J.H., Escobedo, G., Medrano, R.M., Canto, C.B., Tzec-Simá, M., Grijalva, A.R. & James, K.A. 2014. Single nucleotide polymorphisms in partial sequences of the gene encoding the large sub-units of ADP-glucose pyrophosphorylase within a representative collection of 10 Musa genotypes. *Electronic Journal of Biotechnology* 17(3): 137-147. doi:10.1016/j.ejbt.2014.04.004

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M.K., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Lu, S., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D. & Bryant, S.H. 2013. CDD: Conserved domains and protein three-dimensional structure. *Nucleic Acids Research* 41(Database issue): D348-D352. https://doi.org/10.1093/nar/gks1243

Maretta, D., Sobir, Helianti, I., Ridwan, D., Purwono & Santosa, E. 2023. Glucomannan content stability of Eddoe taro tuber based on parametric, non parametric, and AMMI analysis. *Sains Malaysiana* 52(11): 3135-3145. http://doi.org/10.17576/jsm-2023-5211-09

Maretta, D., Sobir, Helianti, I., Purwono & Santosa, E. 2020. Genetic diversity in eddoe taro (*Colocasia esculenta* var. Antiquorum) from Indonesia based on morphological and nutritional characteristics. *Biodiversitas* 21(8): 3525-3533. https://doi.org/10.13057/biodiv/d210814

Mashudi, S., Putri, D.R. & Hariyatmi, Aziz, Y.S. 2023. The impact of konjac glucomannan in reducing blood glucose levels. *Community Practitioner* 20(10): 387-393. DOI: 10.5281/zenodo.10060277

Moreira, L.R.S. & Filho, E.X.F. 2008. An overview of mannan structure and mannan degrading enzyme systems. *Applied Microbiology and Biotechnology* 79: 165-178.

Suzuki, S., Li, L., Sun, Y.H. & Chiang, V.L. 2006. The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. *Plant Physiology* 142(3): 1233-1245. doi:10.1104/pp.106.086678

Tasma, I.M. 2015. Utilization of genome sequencing technology to accelerate plant breeding program. *Jurnal Penelitian dan Pengembangan Pertanian* 32(2): 159-168.

Theodore, G.S. 2000. Mitochondrial DNA and the peopling of the new world genetic variations among native Americans provide further clues to who first populated the Americas and when they arrived. *American Scientist Online* https://www.sas.upenn.edu/~tgschurr/pdf/Am%20Sci%20Article%202000.pdf

Thompson, J.D., Gibson, T.J. & Higgins, D.G. 2003. Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics* Chapter 2:Unit 2.3.

Wigoeno, Y.A., Azrianingsih, R. & Roosdiana, A. 2013. Analisis kadar glukomanan pada umbi porang (*Amorphophallus muelleri* Blume) menggunakan refluks kondensor. *Jurnal Biotropika* 1(5): 231-235.

Yadav, S., Yadava, Y.K., Kohli, D., Meena, S., Kalwan, G., Bharadwaj, C., Gaikwad, K., Arora, A. & Jain, P.K. 2022. Genome-wide identification, *in-silico* characterization and expression analysis of the RNA helicase gene family in chickpea (*C. arietinum* L.). *Scientific Reports* 12: 9778. https://doi.org/10.1038/s41598-022-13823-9

Yang, D., Yuan, Y., Wang, L., Wang, X., Mu, R., Pang, J., Xiao, J. & Zheng, Y. 2017. A review on konjac glucomannan gels: Microstructure and application. *International Journal of Molecular Sciences* 18(11): 2250. doi:10.3390/ijms18112250

Yeriska, F., Umar, M.Z., Hijriah, N.W., Fadhlurrohman, R. & Achyar, A. 2021. Analisis variasi genetik sekuen gen pengkode protein spike virus MERS-CoV (PopSet: 1843801421) menggunakan RFLP secara *in-silico*. *Prosiding Seminar Nasional Biologi* 1(2): 898-906. https://semnas.biologi.fmipa.unp.ac.id/index.php/prosiding/article/view/145?articlesBySameAuthorPage=3

Zalewski, B.M., Chmielewska, A. & Szajewska, H. 2015. The effect of glucomannan on body weight in overweight or obese children and adults: A systematic review of randomized controlled trials. *Nutrition* 31(3): 437-442. https://doi.org/10.1016/j.nut.2014.09.004

*Corresponding author; email: delvi.maretta@brin.go.id

Note: Consensus identities of three nucleotide fragments of genotype S7 sequences that correspond to the *A. konjac* mRNA genes sequences based on NCBI algorithm (ID: HQ833588.1); frame of blue = 220 bp, green = 240 bp, red = 102 bp

SUPPLEMENT 1. Consesus identity genotype S7 sequences correspond to the *A. konjac* mRNA genes

```
S15             CCTGAACGAGAGCAATCTCCGGGTTGTGGATGAGAAATGGTGTTGTCTGCACAAGGAAGT
S28             CCTGAACGAGAGCAATCTCCGGGTTGTGGATGAGAAATGGTGTTGTCTGCACAAGGAAGT
S30             CCTGAACGAGAGCAATCTCCGGGTTGTGGATGAGAAATGGTGTTGTCTGCACAAGGAAGT
S36             CCTGAACGAGAGCAATCTCCGGGTTGTGGATGAGAAATGGTGTTGTCTGCACAAGGAAGT
S34             CCTGAACGAGAGCAATCTCCGGGTTGTGGATGAGAAATGGTGTTGTCTGCACAAGGAAGT
S7              CCTGAACGAGAGCAATCTCCGGGTTGTGGGTGAGAAATGGTGTTGTCTGCACAAGGAAGT
                **************************** ****************************

S15             CGGATTTTGGTTGGAAGTCGGCATCAAAGATGGCAACATAGTCGCACGCCTCCACATATC
S28             CGGATTTTGGTTGGAAGTCGGCATCAAAGATGGCAACATAGTCGCACGCCTCCACATATC
S30             CGGATTTTGGTTGGAAGTCGGCATCATAGATGGCAACATAGTCGCACGCCTCCACATATC
S36             CGGATTTTGGTTGGAAGTCGGCATCAAAGATGGCAACATAGTCGCACGCCTCCACATATC
S34             CGGATTTTGGTTGGAAGTCGGCATCAAAGATGGCAACATAGTCGCACGCCTCCACATATC
S7              CGGATTTTGGTTGGAAGTCGGCATCAAAGATGGCAACATAGTCGCACGCCTCCACATATC
                ************************* *********************************

S15             TATACTTCATTCCTTCCTTGAGGGCACCTGCTTTGTACCCATTCCTGTTGGCCCTGGTCT
S28             TATACTTCATTCCTTCCTTGAGGGCACCTGCTTTGTACCCATTCCTGTTGGCCCTGGTCT
S30             TATACTTCATTCCTTCCTTGAGGGCACCTGCTTTGTACCCATTCCTGTTGGCCCTGGTCT
S36             TATACTTCATTCCTTCCTTGAGGGCACCTGCTTTGTACCCATTCCTGTTGGCCCTGGTCT
S34             TATACTTCATTCCCTCCTTGAGGGCACCTGCTTTGTACCCATTCCTGTTGGCCCTGGTCT
S7              TATACTTCATTCCTTCCTTGAGGGCACCTGCTTTGTACCCATTCCTGTTGGCCCTGGTCT
                ************ ********************************************

S15             CATATTTTATGTTCACCCCCGTCGTAGTCCATTTCTGGCATTCTTTGTACACAAGATCCT
S28             CATATTTTATGTTCACCCCCGTCGTAGTCCATTTCTGGCATTCTTTGTACACAAGATCCT
S30             CATATTTTATGTTCACCCCCGTCGTAGTCCATTTCTGGCATTCTTTGTACACAAGATCCT
S36             CATATTTTATGTTCACCCCCGTCGTAGTCCATTTCTGGCATTCTTTGTACACAAGATCCT
S34             CATATTTTATGTTCACCCCCGTCGTAGTCCATTTCTGGCATTCTTTGTACACAAGATCCT
S7              CATATTTTATGTTCACCCCCGTCGTAGTCCATTTCTGGCATTCTTTGTACACAAGATCCT
                ************************************************************
```

```
S15             ACATGAGACCACCACCCCA-------------CCCCAAAAAAAAAATGGGTAAATTAGGGA
S28             ACATGAGACCACCACCCCA-------------CCCCAAAAAAAAAATGGGTAAATTAGGGA
S30             ACATGAGACCACCACCCCA-------------CCCCAAAAAAAAAATGGGTAAATTAGGGA
S36             ACATGAGACCACCACCCCA-------------CCCCAAAAAAAAAATGGGTAAATTAGGGA
S34             ACATGAGACCACCACCCCA-------------CCCCCCAAAAAAAAAGGGTAAATTAGGGA
S7              ACATGAGACCACCACCCCATCCCCCACCCCACCCCCCAAAAAAAAATGGGGAAATTAGGGA
                ******************              ****  ******* *** **********

S15             GCAAAGCAAAATTATCAAGAGAAAGATCAAATGTTCTGGACAGCAGAATATACTGTCATT
S28             GCAAAGCAAAATTATCAAGAGAAAGATCAAATGTTCTGGACAGCAGAATATACTGTCATT
S30             GCAAAGCAAAATTATCAAGAGAAAGATCAAATGTTCTGGACAGCAGAATATACTGTCATT
S36             GCAAAGCAAAATTATCAAGAGAAAGATCAAATGTTCTGGACAGCAGAATATACTGTCATT
S34             GCAAAGCAAAATTATCAAGAGAAAGATCAAATGTTCTGGACAGCAGTATATACTGTCATT
S7              GCAAAGCAAAATTATCGACAGAAAGATCAAATGTTCTGGACAGCGGTATATACTGTCATT
                *************** * ************************ * *************

S15             TTAATTATACATATAGAAAAATGGAGATCAGCTCTGCCCAGCACGGTATATATGGGTACT
S28             TTAATTATACATATAGAAAAATGGAGATCAGCTCTGCCCAGCACGGTATATATGGGTACT
S30             TTAATTATACATATAGAAAAATGGAGATCAGCTCTGCCCAGCACGGTATATATGGGTACT
S36             TTAATTATACATATAGAAAAATGGAGATCAGCTCTGCCCAGCACGGTATATATGGGTACT
S34             TTAATTATACATATAGAAAAATGGAGATCATAGCTGCCCAGCACGGTATATATGGGTACT
S7              TTAATTATACATATAGAGAAATGGAGATCATAGCTGCCCAGCACGGTAT----GGGTACT
                ***************** ************   ****************    *******

S15             TGAGAAAAAGTGGGCACATGGTTGGGTCCTTTTCGAAAATGTCTCATTAAGTAGGGTTCT
S28             TGAGAAAAAGTGGGCACATGGTTGGGTCCTTTTCGAAAATGTCTCATTAAGTAGGGTTCT
S30             TGAGAAAAAGTGGGCACATGGTTGGGTCCTTTTCGAAAATGTCTCATTAAGTAGGGTTCT
S36             TGAGAAAAAGTGGGCACATGGTTGGGTCCTTTTCGAAAATGTCTCATTAAGTAGGGTTCT
S34             TGAGAAAAAGTGGGCACATGGTTGGGTCCTTTTCGAAAATGTCTCATTCAGTAGGGTTCT
S7              TGAGAAAAGGTGTGCACATGGTTGGGTCCGTTTCGAAAATGTCTCATTAAGTAGGCGTCT
                ******** *** *************** ****************** ******  ***

S15             TTGTTCTTATAATATTTTTACTATAAAAGATATTCTAGTAGTACTATTAATAAGATGAAA
S28             TTGTTCTTATAATATTTTTACTATAAAAGATATTCTAGTAGTACTATTAATAAGATGAAA
S30             TTGTTCTTATAATATTTTTACTATAAAAGATATTCTAGTAGTACTATTAATAAGATGAAA
S36             TTGTTCTTATAATATTTTTACTATAAAAGATATTCTAGTAGTACTATTAATAAGATGAAA
S34             TTGTTCTTATAATATTTTTACTATAAAAGATATTCTAGTAGTACTATTAATAAGATGAAA
S7              GTGTTCTTATAATATTTTTACTATAAAAGATATTCTAGTAGTACTATTAATAAGATGAAA
                 ************************************************************

S15             AGCTTCTTTCATTGGTAGC-TATTATTAATTATAATTAAGTATAGTGTAATTAATTACAT
S28             AGCTTCTTTCATGGGTAGGGTATTATTAATTATAATTAAGTATAGGGTAATTAATTACAA
S30             AGCTTCTTTCATTGGTAGC-TATTATTAATTATAATTAAGTATAGTGTAATTAATTACAT
S36             AGCTTCTTTCATTGGTAGC-TATTATTAATTATAATTAAGTATAGTGTAATTAATTACAT
S34             AGCTTCTTTCATTGGTAGC-TATTATTAATTATAATTAAGTATAGTGTAATTAATTACAT
S7              AGCTTCTTTCATTGGTGGC-TATTATTAATTATAATTAAGTATAGTGCAATTAATTACAT
                ************ *** *  ************************ * ************

S15             GTCTTGTAATTTGTAGAAGTATCATAGCTAGGATTCAAGAGGACCACCTTTGCTTAAAGG
S28             GTCTTGTAATTTGTAGAAGTATCATAGCTAGGATTCAAGAGGACCCCCTTTGCTTAAAGG
S30             GTCTTGTAATTTGTAGAAGTATCATAGCTAGGATTCAAGAGGACCACCTTTGCTTAAAGG
S36             GTCTTGTAATTTGTAGAAGTATCATAGCTAGGATTCAAGAGGACCACCTTTGCTTAAAGG
S34             GTCTTGTAATTTGTAGAAGTATCATACCTAGGATTCAAGAGGACCACCTTTGCTTAAAGA
S7              GTCTTGTAATTTGTAGAAGTATCATAGCTAGGATTCAAGAGGACCACCTTTGCTTAAAGA
                ************************* ****************** *************

S15             TTCCATTTATATTCCCTGGTGTGTTCTAGGTAGATCCTTGCAAGCTTAGGAGTATATAGA
S28             TTCCATTTATATTCCCTGGTGTGTTCTAGGTAGATCTTTGCAAGCTTAGGAGTATATAGA
S30             TTCCATTTATATTCCCTGGTGTGTTCTAGGTAGATCTTTGCAAGCTTAGGAGTATATAGA
S36             TTCCATTTATATTCCCTGGTGTGTTCTAGGTAGATCTTTGCAAGCTTAGGAGTATATAGA
S34             TTCCATTTATATTCCCTG-TGTGTTCTAGGTAGATCTTTGCAAGCTTAGGAGTATATAGA
S7              TTCCATTTATATTCCCTGGTGTGTTCTAGGTAGATCTTTGCAAGCTTAGGAGTATATAGA
                ****************** ***************** ***********************
```

```
S15             ACAAAAAAACAAAAACAAAAA--TGTAGAGCTTGCCGTACGTTTTAAAGGCCCTTCTAGC
S28             ACAAAAAAACAAAAACAAAAA--TGTAGAGCTTGCCGTACGTTTTAAAGGCCCTTCTAGC
S30             ACAAAAAAACAAAAACAAAAA--TGTAGAGCTTGCCGTACGTTTTAAAGGCCCTTCTAGC
S36             ACAAAAAAACAAAAACAAAAA--TGTAGAGCTTGCCGTATGTTTTAAAGGCCCTTCTAGC
S34             AAAAAAAAACAAAAA-AAAAA--TGTAGAGCTTGCCGTACGTTTTAAAGGGCCTTCTAGC
S7              AAAGAAAAAAAACAAAAAAAAATTGTAGAGCTTGCCGTACGTTTTAAAGGCCCTTCTAGC
                * * ***** ** ** *****  ********** ********** *********

S15             TTTTTCAATTACCGGCTGGCCATCCTCGTGAACATCCATAGAATAGAATCCATGAGAACA
S28             TTTTTCAATTACCGGCTGGCCATCCTCGTGAACATCCATAGAATAGAATCCATGAGAACA
S30             TTTTTCAATTACCGGCTGGCCATCCTCGTGAACATCCATAGAATAGAATCCATGAGAACA
S36             TTTTTCAATTACCGGCTGGCCATCCTCGTGAACATCCATAGAATAGAATCCATGAGAACA
S34             TTTTTCAATTACCGGCTGGCCATCCTCGTGAACACCCATAGAATAGAATCCATAAGAACA
S7              TTTTTCAATTACCGGCTGGCCATCCTCGTGAACATCCATAGAATAGAATCCATGAGAACA
                ********** ********** ********** ********** ******

S15             ATTCAAGATGGCCAGGAAAGATTATATGCATACTGGCGTGCGCAGGACCAGTGCTGTGAT
S28             ATTCAAGATGGCCAGGAAAGATTATATGCATACTGGCGTGCGCAGGACCAGTGCTGTGAT
S30             ATTCAAGATGGCCAGGAAAGATTATATGCATACTGGCGTGCGCAGGACCAGTGCTGTGAT
S36             ATTCAAGATGGCCAGGAAAGATTATATGCATACTGGCGTGCGCAGGACCAGTGCTGTGAT
S34             ATTCAAGATGGCCAGGAAAGATTATATGCATACTGGCGTGCACAGGACCAGTGCTGTCAT
S7              ATTCAAGATGGCCAGGAAAGATTATATGCATACTGGCGTGCACACGACCAGTGCTGACAT
                ********** ********** ********** ** ********** **********

S15             TGAGCAATATATAGCTACGCAATGCGCGCCAGCACCTGAACTGTCCAATGGCAGCAGCAG
S28             TGAGCAATATATAGCTACGCAATGCGCGCCAGCACCTGAACTGTCCAATGGCAGCAGCAG
S30             TGAGCAATATATAGCTACGCAATGCGCGCCAGCACCTGAACTGTCCAATGGCAGCAGCAG
S36             TGAGCAATATATAGCTACGCAATGCGCGCCAGCACCTGAACTGTCCAATGGCAGCAGCAG
S34             TGAGCAATATATAGCTACGCAATGCGCGCCAGCACCTGAACTGTCCAATGGCAGCAGCAG
S7              TGAGCAATATATAGCTACGCAATGCGCGCCAGCACCTGAACTGTCCAATGGCAGCAGCAG
                ************************************************************

S15             CGGTCATACATCAAGCAATATGCATAAAGTTCTTTTTTTT-CATTTGGATGTGTTGTGTC
S28             CGGTCATACATCAAGCAATATGCATAAAGTTTTTTTTTTT-CATTTGGATGTGTTGTGTC
S30             CGGTCATACATCAAGCAATATGCATAAAGTTTTTTTTTTT-CATTTGGATGTGTTGTGTC
S36             CGGTCATACATCAAGCAATATGCATAAAGTTTTTTTTTTT-CATTTGGATGTGTTGTGTC
S34             CGGTCATACATCAAGCAATATGCATAAAGTTTTTTTTTTTTTCATTTGGATGTGTTGTGTC
S7              CGGTCATACATCAAGCAGTATGCATAAAGTTTTTTTTTTT-CATTTGGATGTGTTGTGTC
                ********** ********** ******** ********** ********** **********

S15             CGCT----TAAACCTCTTTGAGATCGGTCTTTGGGGTGCATGAATCGCCTTGATGAATGG
S28             CGCT----TAAACCTCTTTGAGATCGGTCTTTGGGGTGCATGAATCACCTTGATGAATGG
S30             CGCT----TAAACCTCTTTGAGATCGGTCTTTGGGGTGCATGAATCACCTTGATGAATGG
S36             CGCT----TAAACCCCTTTGAGATCGGTCTTTGGGGTGCATGAATCACCTTGATGAATGG
S34             CGCT----TAAACCTCTTTGAGATCGGTCTTTGGGGTGCATGAATCACCTTTATGAATGG
S7              CGCTAGCTTAAACCTCTTTGAGATCGGTCTTTGGGGTGCATGAATCACCTTGATGAATGG
                ****    ****** ****************************** **** ********

S15             GTCTGTGGAGTCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
S28             GTCTGTGGAGTCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
S30             GTCTGTGGAGTCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
S36             GTCTGTGGAGTCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
S34             GTCTGTGGAATCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
S7              GTCTGTGGAATCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
                ********* ************************************************

S15             TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACATCATTATAT----
S28             TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACATCATTATAT----
S30             TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACATCATTATAT----
S36             TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACATCATTATAT----
S34             TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACACCATTATATCATT
S7              TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACATCATTATAT----
                ********************************************* ********
```

```
S15             --ATATCATTAGAGAGAGAGAGAG----GCGCGCACTCAAACACACACACACACAGAGCT
S28             --ATATCATTAGAGAGAGAGAGAG----GCGCGCACTCAAACACACACACACACAGAGCT
S30             --ATATCATTAGAGAGAGAGAGAG----GCGCGCACTCAAACACACACACACACAGAGCT
S36             --ATATCATTAGAGAGAGAG--------GCGCGCACTCAAACACACACACACACAGAGCT
S34             ACAGAGAGTTAGAGAGAGAGAGAGAGGCGCGCACTCAAACACACACACACACAGAGCT
S7              ------CATTAGAGAGAGAGAGAGAG--GCGCGCACTCAAACACACACACACACAGAGCT
                       ************        ********************************

S15             ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGACCATGGGGTAAGCTGCGGTGCCC
S28             ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGACCATGGGGTAAGCTGCGGTGCCC
S30             ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGACCATGGGGTAAGCTGCGGTGCCC
S36             ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGACCATGGGGTAAGCTGCGGTGCCC
S34             ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGAC-ATGGGGTAAGCTGCGGTGCCC
S7              ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGACCATGGGGTAAGCTGCGGTGCCC
                ************************************ **********************

S15             AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
S28             AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
S30             AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
S36             AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
S34             AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
S7              AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
                ************************************************************

S15             ACCTTCACGTAGAGGCTCACACCACCCATGAACAGCTTCTCCGCCAGCAGCATTACCGAC
S30             TGAGCAATATATAGCTACGCAATGCGCGCCAGCACCTGAACTGTCCAATGGCAGCAGCAG
S36             TGAGCAATATATAGCTACGCAATGCGCGCCAGCACCTGAACTGTCCAATGGCAGCAGCAG
S34             TGAGCAATATATAGCTACGCAATGCGCGCCAGCACCTGAACTGTCCAATGGCAGCAGCAG
S7              TGAGCAATATATAGCTACGCAATGCGCGCCAGCACCTGAACTGTCCAATGGCAGCAGCAG
                ************************************************************

S15             CGGTCATACATCAAGCAATATGCATAAAGTTCTTTTTTTT-CATTTGGATGTGTTGTGTC
S28             CGGTCATACATCAAGCAATATGCATAAAGTTTTTTTTTTT-CATTTGGATGTGTTGTGTC
S30             CGGTCATACATCAAGCAATATGCATAAAGTTTTTTTTTTT-CATTTGGATGTGTTGTGTC
S36             CGGTCATACATCAAGCAATATGCATAAAGTTTTTTTTTTT-CATTTGGATGTGTTGTGTC
S34             CGGTCATACATCAAGCAATATGCATAAAGTTTTTTTTTTTTTCATTTGGATGTGTTGTGTC
S7              CGGTCATACATCAAGCAGTATGCATAAAGTTTTTTTTTTT-CATTTGGATGTGTTGTGTC
                ***************** ************* ******** ******************

S15             CGCT----TAAACCTCTTTGAGATCGGTCTTTGGGGTGCATGAATCGCCTTGATGAATGG
S28             CGCT----TAAACCTCTTTGAGATCGGTCTTTGGGGTGCATGAATCACCTTGATGAATGG
S30             CGCT----TAAACCTCTTTGAGATCGGTCTTTGGGGTGCATGAATCACCTTGATGAATGG
S36             CGCT----TAAACCCCTTTGAGATCGGTCTTTGGGGTGCATGAATCACCTTGATGAATGG
S34             CGCT----TAAACCTCTTTGAGATCGGTCTTTGGGGTGCATGAATCACCTTTATGAATGG
S7              CGCTAGCTTAAACCTCTTTGAGATCGGTCTTTGGGGTGCATGAATCACCTTGATGAATGG
                ****    ****** ******************************** **** ********

S15             GTCTGTGGAGTCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
S28             GTCTGTGGAGTCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
S30             GTCTGTGGAGTCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
S36             GTCTGTGGAGTCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
S34             GTCTGTGGAATCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
S7              GTCTGTGGAATCATCAAGAACTTGGACAATCAGCCGATCGGACGGCCATTGCAGACTGCA
                ********* ************************************************

S15             TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACATCATTATAT----
S28             TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACATCATTATAT----
S30             TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACATCATTATAT----
S36             TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACATCATTATAT----
S34             TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACACCATTATATCATT
S7              TGCAGCTTTGATCGACAGATGATAAACCTGCAAACGTCCCAAACACATCATTATAT----
                ********************************************* ********
```

```
S15             --ATATCATTAGAGAGAGAGAGAG----GCGCGCACTCAAACACACACACACACAGAGCT
S28             --ATATCATTAGAGAGAGAGAGAG----GCGCGCACTCAAACACACACACACACAGAGCT
S30             --ATATCATTAGAGAGAGAGAGAG----GCGCGCACTCAAACACACACACACACAGAGCT
S36             --ATATCATTAGAGAGAGAG--------GCGCGCACTCAAACACACACACACACAGAGCT
S34             ACAGAGAGTTAGAGAGAGAGAGAGAGAGGCGCGCACTCAAACACACACACACACAGAGCT
S7              ------CATTAGAGAGAGAGAGAGAG--GCGCGCACTCAAACACACACACACACAGAGCT
                       ************        ********************************

S15             ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGACCATGGGGTAAGCTGCGGTGCCC
S28             ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGACCATGGGGTAAGCTGCGGTGCCC
S30             ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGACCATGGGGTAAGCTGCGGTGCCC
S36             ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGACCATGGGGTAAGCTGCGGTGCCC
S34             ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGAC-ATGGGGTAAGCTGCGGTGCCC
S7              ACCTCTTTCTCATTGTACATGGGTATTTGGACGAGGACCATGGGGTAAGCTGCGGTGCCC
                ************************************ ********************

S15             AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
S28             AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
S30             AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
S36             AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
S34             AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
S7              AGCTCCTCGTCTCCTCCCACGGGCTCCCACATGTAGACCTTCCTCGGCCGCCGGCGGAAA
                ************************************************************

S15             ACCTTCACGTAGAGGCTCACACCACCCATGAACAGCTTCTCCGCCAGCAGCATTACCGAC
S28             ACCTTCACGTAGAGGCTCACACCACCCATGAACAGCTTCTCCGCCAGCAGCATTACCGAC
S30             ACCTTCACGTAGAGGCTCACACCACCCATGAACAGCTTCTCCGCCAGCAGCATTACCGAC
S36             ACCTTCACGTAGAGGCTCACACCACCCATGAACAGCTTCTCCGCCAGCAGCATTACCGAC
S34             ACCTTCACGTAGAGGCTCACACCACCCATGAACAGCTTCTCCGCCAGCAGCATTACCGAC
S7              ACCTTCACGTAGAGGCTCACACCACCCATGAACAGCTTCTCCGCCAGCAGCATTACCGAC
                ************************************************************

S15             ATGATGATGCACAGCAGCACGGCGAGTCTCATCGCCGGAACGAGCAAGATACCTCTTACC
S28             ATGATGATGCACAGCAGCACGGCGAGTCTCATCGCCGGAACGAGCAAGATACCTCTTACC
S30             ATGATGATGCACAGCAGCACGGCGAGTCTCATCGCCGGAACGAGCAAGATACCTCTTACC
S36             ATGATGATGCACAGCAGCACGGCGAGTCTCATCGCCGGAACGAGCAAGATACCTCTTACC
S34             ATGATGATGCACAGCAGCACGGCGAGCCTCATCGCCGGAACGAGCAAGATACCTCTTACC
S7              ATGATGATGCACAGCAGCACGGCGAGCCTCATTGCCGGAACGGGCAAGATACCTCTTACC
                ************************** ***** ********* *****************

S15             TGGCCCCA
S28             TGGCCCCA
S30             TGGCCCCA
S36             TGGCCCCA
S34             TGGCCCCA
S7              TGGCCCCA
                ********
```

Note. The arrangement output showed differing nucleotides at 21 sites for S7 ( ) and 8 sites for S34 ( ) compared to other sequences. There 10 nucleotide sites were found to be identical to S7 and S34 but distinct to S15, S28, S30, and S36 sequences ( )

SUPPLEMENT 2. Nucleotides alignment