

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Chloroplast genome data of *Luffa acutangula* and *Luffa aegyptiaca* and their phylogenetic relationships



Chutintorn Yundaeng^a, Wanapinun Nawae^a, Chaiwat Naktang^a, Jeremy R. Shearman^a, Chutima Sonthirod^a, Duangjai Sangsrakru^a, Thippawan Yoocha^a, Nukoon Jomchai^a, John R. Sheedy^b, Supat Mekiyanon^b, Methawat Tuntaisong^a, Wirulda Pootakham^{a,*}, Sithichoke Tangphatsornruang^{a,*}

^a National Omics Center, National Science and Technology Development Agency, 111 Thailand Science Park, Paholyothin Road, Khlong Nueng, Khlong Luang, Pathum Thani, 12120, Thailand
^b Chia Tai Company Limited, Phra Khanong District, Bangkok, Thailand

ARTICLE INFO

Article history: Received 22 June 2020 Revised 21 October 2020 Accepted 23 October 2020 Available online 28 October 2020

Keywords: Luffa acutangula Luffa aegyptiaca chloroplast genome PacBio sequencing comparative analysis

ABSTRACT

Luffa acutangula and Luffa aegyptiaca are domesticated plants in the family Cucurbitaceae. They are mainly cultivated in the tropical and subtropical regions of Asia. The chloroplast genomes of many Cucurbitaceae species were sequenced to examine gene content and evolution. However, the chloroplast genome sequences of L. acutangula and L. aegyptiaca have not been reported. We report the first complete sequences of L. acutangula and L. aegyptiaca chloroplast genomes obtained from Pacific Biosciences sequencing and use them to infer evolutionary relationships. The chloroplast genomes of L. acutangula and L. aegyptiaca are 157,202 and 157,275 bp, respectively. Both genomes possessed the typical quadripartite structure and contained 131 genes, including 87 coding genes, 36 tRNA genes and 8 rRNA genes. We identified simple sequence repeats (SSR) and single nucleotide polymorphisms (SNP) from both chloroplast genomes. Polycistronic mRNA was examined in L. acutangula and L. aegyptiaca using RNA sequences from Isoform sequencing to identify

* Corresponding author.

E-mail addresses: wirulda@alumni.stanford.edu (W. Pootakham), sithichoke.tan@nstda.or.th (S. Tangphatsornruang).

https://doi.org/10.1016/j.dib.2020.106470

^{2352-3409/© 2020} The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

co-transcribed genes. IR size and locations were compared to other species and found to be relatively unchanged. Phylogenetic analysis confirmed the close relationship between *L. acutangula* and *L. aegyptiaca* in the Cucurbitaceae lineage and showed separation of the *Luffa* monophyletic clade from other species in the subtribe Sicyocae. The results obtained from this study can be useful for studying the evolution of Cucurbitaceae plants.

> © 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Specifications Table

Subject	Plant Science
Specific subject area Type of data	Genomic Tables Graph Figures Raw data Sequences
How data were acquired	Pacific Biosciences sequencing (PacBio RSII sequencing)
Data format	Chloroplast raw sequence data in FASTQ format Complete chloroplast genome sequence in FASTA format
Parameters for data collection	Genomic DNA was extracted from fresh leaves of <i>L. acutangula</i> and <i>L. aegyptiaca</i> plants to derive from Chia Tai Company Limited.
	Leaves of 61 accessions of <i>L. acutangula</i> and 23 accessions of <i>L. aegyptiaca</i> seedlings (Chia Tai Co, Ltd) were harvested and genomic DNA isolated.
Description of data collection	PacBio libraries were prepared to sequence on the PacBio RSII sequencing for complete chloroplast genomes assembly.
	Illumina Hiseq X ten libraries with 150 bp pair-end were constructed and sequenced for simple sequence repeats (SSR) and single nucleotide polymorphism (SNP) identifications
Data source location	Institution: National Science and Technology Development Agency, Region: Khlong Luang, Pathum Thani Country: Thailand
Data accessibility	All data in this article are available at NCBI, BioProject number PRJNA639390. Chloroplast raw sequence data with this article are accessible under SRA accession number SRR12011300 (<i>L. acutangula</i>) and SRR12011301 (<i>L. aegyptiaca</i>). Direct URL to data: https://www.ncbi.nlm.nih.gov/sra/?term=SRR12011300 https://www.ncbi.nlm.nih.gov/sra/?term=SRR12011301 Complete chloroplast sequence data are accessible at NCBI under GenBank accession number MT381996 (<i>L. acutangula</i>) and MT381997 (<i>L. aegyptiaca</i>). Direct URL to data: https://www.ncbi.nlm.nih.gov/genome/?term=MT381996 https://www.ncbi.nlm.nih.gov/genome/?term=MT381997 Isoform sequencing (Iso-seq) data of <i>L. acutangula</i> [SRA accession number: SRR11445640] and <i>L. aegyptiaca</i> [SRA accession number: SRR11445640] Direct IIL to data: https://www.ncbi.nlm.nib.gov/sra/?term=SR811445640
	https://www.ncbi.nlm.nih.gov/sra/?term=SRR11452010

Value of the Data

• *L. acutangula* and *L. aegyptiaca* chloroplast genomes are sources of molecular data that confirm complex evolutionary relationships and support the need for phylogenetic research in various plant groups.

- The complete chloroplast genome data could be utilized in the genetics, biotechnology, plant breeding, and ecology fields.
- The sequence variation among the chloroplast genomes of *Luffa* sp. and other representatives of the family Cucurbitaceae enhances the understanding of their phylogenetic relationships.
- Polymorphisms in the chloroplast genome (e.g., simple sequence repeats (SSRs) or single nucleotide polymorphisms (SNPs)) can be used to develop potential molecular markers and study evolutionary patterns of *Luffa* sp. and closely related species.

1. Data Description

The complete chloroplast genomes of L. acutangula and L. aegyptiaca were assembled using long read sequences obtained from PacBio sequencing and annotated for gene content. The chloroplast genome sequences and annotated genes are available through NCBI accession number MT381996 (L. acutangula) and MT381997 (L. aegyptiaca). Both chloroplast genomes had the typical quadripartite structure, which consists of a small single-copy region (SSC) and a large single-copy region (LSC), separated by a pair of inverted repeats (IRs) (Fig. 1, Table 1). Both chloroplast genomes encoded 131 genes, including 87 protein-coding genes, 36 tRNA genes and 8 rRNA genes (Table 2, Table 3). The codon-usage frequencies were calculated for the proteincoding genes and tRNA genes of the L. acutangula and L. aegyptiaca chloroplast genomes (Fig 2, Table 4). Length and position of the LSC and SSC regions and genetic variation the chloroplast genomes were examined among L. acutangula, L. aegyptiaca and other species in the family Cucurbitaceae (Fig. 3 and 4). Simple sequence repeats (SSR) (Fig. 5, supplementary Table S1), single nucleotide polymorphisms (SNP) (Table 5) and RNA editing events (Table 6) in both L. acutangula and L. aegyptiaca chloroplast genomes were identified. Polycistronic transcript sequences were similar in L. acutangula and L. aegyptiaca chloroplast genomes (Table 7, supplementary Table S2). Furthermore, a phylogenetic analysis of Luffa and several Cucurbitaceae species placed L. acutangula and L. aegyptiaca closely related to Tricosanthes and Hodgsonia in the Sicyoeae tribe (Fig. 6).

Table 1				
Chloroplast genome	features	among	Cucurbitaceae	species.

	L. acutangula	L. aegyptiaca	C. lanatus	C. melo	C. sativus	С. реро
Genome size (bp)	157,202	157,275	156,906	156,017	155,293	157,343
LSC size (bp)	86,226	86,310	86,846	86,335	86,689	87,828
SSC size (bp)	18,402	18,393	17,898	18,090	18,209	18,169
IRs size (bp)	26,280	26,286	26,081	25,796	25,199	25,678
GC content (%)	37.14	37.12	37.18	36.92	37.08	37.16
LSC GC content (%)	34.96	34.93	34.94	34.67	34.85	34.91
SSC GC content (%)	31.02	31.04	31.54	30.94	31.83	31.44
IRs GC content (%)	42.86	42.86	42.84	42.79	42.83	43.05
No. of genes	131	131	124	135	133	131
No. of CDS	87	87	87	90	89	86
No. of tRNA	36	36	29	37	37	37
No. of rRNA	8	8	8	8	8	8
No. of CDS with intron	15	15	10	16	15	15
Gene coding density (%)	50.08	50.04	49.74	51.74	50.06	46.60
Genbank accession number	MT381996	MT381997	NC_032008	NC_015983	NC_007144	NC_038229



Fig. 1. The chloroplast genomes of *L. acutangula* and *L. aegyptiaca*. Genes shown outside of the circle are transcribed counterclockwise, while those inside are transcribed clockwise, as shown by the arrows. The functions of genes are grouped by color. Asterisks indicate intron-containing genes.

List of genes present in L. acutangula and L. aegyptiaca chloroplast genomes.

Photosynthesis Photosystem I (5) Photosystem II (15) psaA, psaB, psaC, psaI, psaJ psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbJ, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ Cytochome b6/f complex (6) ATP synthase (6) petA, petB*, petO*, petC, petL, petN ATP synthase (6) atpA, atpB, atpE, atpF*, atpH, atpI Rubisco large subunit (1) rbcl NADH dehydrogenase (12) ndhA*, ndhB (× 2)*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK Self-replication Large subunit Ribosomal protein (11) rpl2 (× 2)*, rpl34, rpl16*, rpl20, rpl22, rpl23 (× 2), rpl32, rpl33, rpl36 Small subunit ribosomal protein (14) rps2 rps3, rps4, rps7 (× 2) rps8, rps14, rps12 (× 2)*	Category	Gene groups	Gene name
Photosystem II (15) psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbJ, psbJ, psbK, psbL, psbN, psbT, psbZ Cytochome b6/f complex (6) petA, petB*, petD*, petG, petL, petN ATP synthase (6) atpA, atpB, atpE, atpF*, atpH, atpI Rubisco large subunit (1) rbcl NADH dehydrogenase (12) ndhA*, ndhB (× 2)*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhJ, ndhJ, ndhK Self-replication Large subunit Ribosomal protein (11) Small subunit ribosomal protein (14) rps2 rps3, rps4 rps7 (× 2) rps8 rps1 rps12 (× 2)*	Photosynthesis	Photosystem I (5)	psaA, psaB, psaC, psaI, psaJ
Self-replication Large subunit ribosomal protein (11) psbK, psbL, psbN, psbT, psbZ Self-replication Cytochome b6/f complex (6) petA, petB*, petD*, petG, petL, petN ATP synthase (6) atpA, atpB, atpE, atpF*, atpH, atpl Rubisco large subunit (1) rbcl NADH dehydrogenase (12) ndhA*, ndhB (× 2)*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhJ, ndhJ, ndhK Self-replication Large subunit Ribosomal protein (11) Self-replication Small subunit ribosomal protein (14)		Photosystem II (15)	psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ,
Cytochome b6/f complex (6) petA, petB*, petD*, petG, petL, petN ATP synthase (6) atpA, atpB, atpE, atpF*, atpH, atpl Rubisco large subunit (1) rbcl NADH dehydrogenase (12) ndhA*, ndhB (× 2)*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhJ, ndhJ, ndhK Self-replication Large subunit Ribosomal protein (11) Small subunit ribosomal protein (14) rps2 (x 2)*, rp13, rp136			psbK, psbL, psbM, psbN, psbT, psbZ
ATP synthase (6) atpA, atpB, atpE, atpF*, atpH, atpl Rubisco large subunit (1) rbcl NADH dehydrogenase (12) ndhA*, ndhB (× 2)*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhJ, ndhJ, ndhK Self-replication Large subunit Ribosomal protein (11) Small subunit ribosomal protein (14) rp32, rp133, rp136		Cytochome b6/f complex (6)	petA, petB*, petD*, petG, petL, petN
Rubisco large subunit (1) rbcl NADH dehydrogenase (12) ndhA*, ndhB (× 2)*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK Self-replication Large subunit Ribosomal protein (11) Small subunit ribosomal protein (14) rpl2 (× 2)*, rpl3, rpl36		ATP synthase (6)	atpA, atpB, atpE, atpF*, atpH, atpI
NADH dehydrogenase (12) $ndhA^*$, $ndhB$ (\times 2)*, $ndhC$, $ndhD$, $ndhF$, $ndhF$, $ndhG$, $ndhH$, $ndhI$, $ndhI$, $ndhI$, $ndhI$, $ndhI$, $ndhK$ Self-replicationLarge subunit Ribosomal protein (11) $rpl2(\times 2)^*$, $rpl13^*$, $rpl20$, $rpl22$, $rpl23$ (\times 2), $rpl32$, $rpl33$, $rpl36$ Small subunit ribosomal protein (14) $rps2$, $rps3$, $rps4$, $rps7$ (\times 2), $rps8$, $rps11$, $rps12$ (\times 2)*		Rubisco large subunit (1)	rbcl
Self-replication Large subunit Ribosomal protein (11) rpl2, rpl3, rpl36 rpl2, rpl27, rpl23 (× 2), rpl32, rpl37, rpl36 rps1 rps12 (× 2)*		NADH dehydrogenase (12)	$ndhA^*$, $ndhB$ (\times 2)*, $ndhC$, $ndhD$, $ndhE$, $ndhF$, $ndhG$,
Self-replication Large subunit Ribosomal protein (11) $rpl2(\times 2)^*$, $rpl14$, $rpl16^*$, $rpl20$, $rpl22$, $rpl23$ ($\times 2$), $rpl32$, $rpl32$, $rpl33$, $rpl36$ Small subunit ribosomal protein (14) $rps2$, $rps3$, $rps4$, $rps7(\times 2)$, $rps8$, $rps11$, $rps12(\times 2)^*$			ndhH, ndhJ, ndhJ, ndhK
Small subunit ribosomal protein (14) $rs^2 rs^3 rs^4 rs^7 (\times 2) rs^8 rs^{11} rs^{12} (\times 2)^*$	Self-replication	Large subunit Ribosomal protein (11)	$rpl2 (\times 2)^*$, $rpl14$, $rpl16^*$, $rpl20$, $rpl22$, $rpl23 (\times 2)$, rpl32, $rpl33$, $rpl36$
p_{2} , p_{2} , p_{3} , p		Small subunit ribosomal protein (14)	rps2, rps3, rps4, rps7 (\times 2), rps8, rps11, rps12 (\times 2)*,
$\frac{1}{1} \frac{1}{1} \frac{1}$		BNA polymorphic (4)	$1p_{514}, 1p_{515}, 1p_{510}, 1p_{516}, 1p_{519}$
Rive polyinerase (4) $Ipor, Ipor, I$		Rive polymerase (4) Ribecomal RNAs (8)	$1pOA, 1pOB, 1pOC1^{\circ}, 1pOC2$
$\begin{array}{ccc} \text{Ribosoniai Rvis}(6) & \text{Inter.} (\times 2), \text{Into} (\times 2), \text$		Transfer RNAs (36)	$trn 4.J (\times 2), trn 5 (\times 2), trn 10 (\times 2), trn 5 (\times 2)$
trnE-GA4 trn/L-GC trn/L-GT trn/L-GT			trnF-GAA trnfM-CAT trnG-GCC trnH-GTG trnI-CAT
$(\times 2)$ trul-CAI $(\times 2)^*$ truk-IIIIII* trul-CAA $(\times 2)$			$(\times 2)$ trnl-GAU $(\times 2)^*$ trnK-UUUU* trnl-CAA $(\times 2)$
trnL-TAG, trnL-UAA, trnM-CAT, trnN-GTT (× 2),			trnL-TAG, trnL-UAA*, trnM-CAT, trnN-GTT (\times 2),
trnP-TGG, trnQ-TTG, trnR-ACG (\times 2), trnR-TCT,			trnP-TGG, trnQ-TTG, trnR-ACG (\times 2), trnR-TCT,
trnS-GCU, trnS-GGA, trnS-TGA, trnT-GGT, trnT-TGT,			trnS-GCU, trnS-GGA, trnS-TGA, trnT-GGT, trnT-TGT,
trnV-GAC (× 2), $trnV$ -UAC*, $trnW$ -CCA, $trnY$ -GUA			trnV-GAC (× 2), trnV-UAC*, trnW-CCA, trnY-GUA
Other genes Acetyl-CoA carboxylase gene (1) accD	Other genes	Acetyl-CoA carboxylase gene (1)	accD
c-type cytochrome biogenesis (1) ccsA		c-type cytochrome biogenesis (1)	ccsA
ATP-dependent protease subunit (1) $clpP^*$		ATP-dependent protease subunit (1)	clpP*
Maturease (1) matK		Maturease (1)	matK
Membrane protein (1) cemA		Membrane protein (1)	cemA
Proteins of unknown function (7) $ycf1, ycf2 (\times 2), ycf3^*, ycf1, ycf1 (\times 2)$		Proteins of unknown function (7)	ycf1, ycf2 (\times 2), ycf3 * , ycf4, ycf15 (\times 2)
Translation-related gene (1) infA		Translation-related gene (1)	infA

* Gene with intron(s)



Amino acid

Fig. 2. Amino acid frequencies in L. acutangula and L. aegyptiaca protein-coding sequences.

Table 3 Genes with intron(s) in L. acutangula and L. aegyptiaca chloroplast genomes.

		Species										
		L. acutangı	ıla				L. aegyptiaca					
Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)	
rps16	LSC	42	855	213	-	-	45	856	213	-	-	
atpF	LSC	144	755	411	-	-	144	757	411	-	-	
rpoC1	LSC	432	753	1611	-	-	432	756	1611	-	-	
ycf3	LSC	126	740	228	743	153	126	740	228	740	156	
clpP	LSC	69	847	288	613	228	69	835	297	615	225	
petB	LSC	6	783	642	-	-	9	780	642	-	-	
petD	LSC	9	727	474	-	-	9	732	474	-	-	
rpl16	LSC	9	1100	402	-	-	9	1098	402	-	-	
rpl2	IRb	390	665	435	-	-	390	665	435	-	-	
ndhB	IRb	777	686	756	-	-	777	686	756	-	-	
rps12	IRb	114	28918	234	537	27	114	28346	234	537	27	
ndhA	SSC	552	1155	540	-	-	552	1146	540	-	-	
rps12	IRa	114	71157	234	537	27	114	71136	234	537	27	
ndhB	IRa	786	677	756	-	-	777	686	756	-	-	
rpl2	IRa	390	665	435	-	-	393	662	435	-	-	

The codon-anticodon recognition pattern and codon usage for L. acutangula and L. aegyptiaca chloroplast genomes.

acid Codon Lansmula Lasymina Lasymina Lasymina Lasymina Phe UUC 530 529 0.71 0.71 0.71 Phe UUC 530 529 0.71 0.71 0.71 Leu UUC 556 556 1.22 1.22 trmL-CA Leu CUC 190 189 0.42 0.41 trmL-CA Leu CUC 190 189 0.42 0.41 trmL-CA Leu CUC 174 175 0.38 0.38 trmL-CAU lie AUD 44 8.3 0.51 0.52 trmL-CAU Miet AUG 624 625 1 trmM-CAU trmM-CAU Val GUU 508 507 1.43 1.43 trmM-CAU Val GUU 508 507 1.43 1.43 trmM-CAU Val GUU 189 189 0.56 0.5	Amino		Frequency ^a		RSCU		
Phe UUU 957 129 129 129 170 $rmF-CAA$ Leu UUA 860 860 138 188 $rmF-CAA$ Leu UUA 860 860 138 188 $rmF-CAA$ Leu UUU 556 555 122 122 $rmF-CAA$ Leu CUC 190 189 0.42 0.41 $rmF-CAA$ Leu CUG 177 176 0.38 0.38 $rmF-CAA$ Leu CUG 174 176 0.38 0.38 $rmF-CAA$ lie AUC 474 472 0.63 0.63 $rmF-CAA$ lie AUG 624 625 1 $rmF-CAA$ $rmF-CAA$ Val GUU 508 507 1.43 $rmF-CAA$ $rmF-CAA$ Val GUU 510 1.51 $rmF-CAA$ $rmF-CAA$ $rmF-CAA$ Val GUU 573 513	acid	Codon	L. acutangula	L. aegyptiaca	L. acutangula	L. aegyptiaca	trn ^b
Phe UUC 530 529 0.71 0.71 Leu UUA 850 556 1.22 1.22 trnL-CM Leu CUU 555 555 1.28 trnL-CM Leu CUC 190 189 0.42 0.41 Leu CUC 174 176 0.38 0.33 Leu CUG 174 176 0.38 0.33 Leu CUG 174 176 0.38 0.33 Leu CUG 474 472 0.63 0.63 trnl-CAT Met AUC 625 1 1 trnd-CAT trnd-CAT Val CUG 508 507 1.43 1.45 trnd-CAC Val CUG 518 1.51 1.52 trnd-CAC Val CUG 503 531 1.51 1.57 trnd-CAC Val CUG 198 188 0.56 0.56	Phe	UUU	957	957	1.29	1.29	trnF-GAA
Leu UUA 850 860 1.88 1.88 1.88 rml-LAA Leu CUU 585 555 1.22 1.23 init-CAA Leu CUC 190 189 0.42 0.41 init-CAI Leu CUG 174 176 0.33 0.33 init-CAI Ile AUC 474 472 0.63 0.63 init-CAI Ile AUC 624 625 1 1 irml-CAT Val CUC 181 183 0.51 0.52 irml-CAT Val CUC 183 186 0.56 irml-CAT irml-CAT Val CUC 181 183 0.51 0.52 irml-CAT Val CUC 189 186 0.56 irml-CAT irml-CAT Val CUC 171 155 irml-CAT irml-CAT irml-CAT Val CUC 181 181 181 </td <td>Phe</td> <td>UUC</td> <td>530</td> <td>529</td> <td>0.71</td> <td>0.71</td> <td></td>	Phe	UUC	530	529	0.71	0.71	
Leu UUG 556 556 1.22 1.22 rmL-TAG Leu CUU 189 0.42 0.41 mL-TAG Leu CUG 174 176 0.38 0.33 Leu CUG 174 176 0.38 0.33 Leu CUG 174 176 0.33 0.63 lie AUG 474 472 0.63 0.63 lie AUG 628 687 0.92 0.92 rmL-CAT Met AUG 624 625 1 1 rmM-CAT Val CUG 508 507 143 143 143 rmV-CAC Val CUG 198 188 0.56 0.56 rmV-CAC Val CUG 198 188 0.56 0.56 rmV-CAC Ser UCA 428 429 1.27 1.27 rmV-CAC Ser UCA 428 429 </td <td>Leu</td> <td>UUA</td> <td>860</td> <td>860</td> <td>1.88</td> <td>1.88</td> <td>trnL-UAA</td>	Leu	UUA	860	860	1.88	1.88	trnL-UAA
Leu CUU 585 585 1.28 1.28 III.AG Leu CUC 190 189 0.42 0.41 Leu CUG 174 176 0.33 0.33 Ile AUC 474 472 0.63 0.63 Ile AUC 474 472 0.63 0.63 Ile AUC 624 625 1 1 trmM-CAT Val GUU 508 507 1.43 1.43 trm/-CAU Val GUU 150 513 1.5 1.5 trm/-CAU Val GUU 150 531 1.5 1.5 trm/-CAU Val GUU 139 302 0.94 0.55 trm/-CAU Val GUU 139 302 0.94 0.55 trm/-CAU Val GUU 313 1.52 trm/-CAU trm/-CAU trm/-CAU Val GUC 131	Leu	UUG	556	556	1.22	1.22	trnL-CAA
Leu CUC 190 189 0.42 0.41 Leu CUG 174 176 0.38 0.38 Leu CUG 174 176 0.38 0.38 Ile AUU 84 83 1.45 1.45 Introduction Ile AUG 624 625 1 1 mmA-CAT Val GUU 508 507 1.43 1.43 1.43 introduction Val GUU 508 507 1.43 1.45 introduction introduction Val GUU 518 138 0.51 0.52 introduction introduction Ser UCU 571 566 1.69 1.68 introduction Ser UCU 571 566 1.69 1.68 introduction Ser UCC 183 410 1.53 1.52 introduction Ser UCC 189 188 0.55	Leu	CUU	585	585	1.28	1.28	trnL-TAG
Leu CUA 377 379 0.82 0.83 Leu CUG 174 176 0.33 0.33 Ile AUC 474 472 0.63 0.63 Ile AUC 688 687 0.92 0.92 0.92 Met AUG 624 625 1 1 rm/ACAT Val GUU 508 507 1.43 1.43 rm/V-CAC Val GUC 181 183 0.51 0.52 rm/V-CAC Val GUC 181 183 0.56 0.56 model Ser UCC 319 322 0.94 0.95 model Ser UCC 319 322 0.94 0.95 model Ser UCC 133 141 1.17 1.17 model model Ser UCC 135 314 1.31 1.61 1.61 infl-CGU model <td< td=""><td>Leu</td><td>CUC</td><td>190</td><td>189</td><td>0.42</td><td>0.41</td><td></td></td<>	Leu	CUC	190	189	0.42	0.41	
Leu CUG 174 176 0.38 0.38 mathematical states of the states of	Leu	CUA	377	379	0.82	0.83	
lieAUU84831.451.451.45rml-CATlieAUA6886870.920.92rml-CATrmM-CATrmM-CATrmM-CATrmM-CATrmM-CATValGUU5085071.431.43rmM-CATValGUU5085071.431.43rmM-CATValGUU5085071.431.43rmM-CATValGUU5085071.431.43rmM-CATValGUU5085071.431.43rmM-CATValGUU5085071.510.52rmM-CATValGUU5191580.560.56rmM-CATSerUCC3193220.940.95rmM-CATSerUCA4284291.271.27rmS-UGAProCCC1314101.531.52rmP-UGCProCCC3153141.171.17rmProCCC1501510.560.56TmP-UGCThrACC2482480.750.75TmP-UGCThrACC3442351.611.61rmP-UGCThrACC3442351.611.61rmP-UGCAlaCCU6346351.811.81rmA-UCCAlaCCU6346351.641.64rmY-GUAlaCCC2312320.66 <td>Leu</td> <td>CUG</td> <td>174</td> <td>176</td> <td>0.38</td> <td>0.38</td> <td></td>	Leu	CUG	174	176	0.38	0.38	
lie AUA 688 687 0.92 0.92 rml-CAT Met AUG 624 625 1 1 rml/CAT Val GUU 508 507 1.43 1.43 rml/CAT Val GUC 181 1833 0.51 0.52 Val GUG 198 198 0.56 0.56 Ser UCC 319 322 0.94 0.95 Ser UCA 428 429 1.27 1.27 mrl-CAT Ser UCA 428 429 1.27 1.27 mrl-CAT Pro CCC 201 203 0.75 0.75 mrl-CAT Pro CCC 315 314 1.17 1.17 mrl-CAT Thr ACC 248 248 0.75 0.75 mrl-CAT Thr ACG 149 147 0.45 0.44 mrl-CAT Thr ACG <	Ile	AUU	84	83	1.45	1.45	trnI-GAU
lie AUA 688 687 0.92 0.92 $rml-CAT$ $rmlACT$ Val GUU 508 507 1.43 1.43 $rmlACT$ Val GUU 508 507 1.43 1.43 $rmlACT$ Val GUU 530 531 1.5 $rmVCAC$ Val GUG 198 198 0.56 0.56 7 Ser UCU 371 566 1.69 1.68 $rmS-CAT$ Ser UCA 428 429 1.27 1.27 $rmS-CAT$ Ser UCA 428 429 1.27 1.27 $rmT-CAT$ Pro CCA 135 314 1.17 1.17 $rmT-CAT$ Pro CCG 135 314 1.17 1.17 $rmT-GGT$ Thr ACA 397 399 1.2 1.2 $rmT-GT$ Thr ACA 397 399 1.2 1.2 $rmT-GT$ Thr ACA 397 399 1.2 1.2	Ile	AUC	474	472	0.63	0.63	
Met AUG 624 625 1 1 trm/M-CAT trm/M-CAT trm/M-CAT Val GUC 181 183 0.51 0.52 Val GUA 530 531 1.5 1.5 tm/M-CAT Val GUA 530 531 1.5 tm/M-CAT Val GUA 530 531 1.5 tm/M-CAT Ser UCC 319 322 0.94 0.95 Ser UCA 428 429 1.27 1.27 tm/P-UGA Pro CCC 201 203 0.75 0.75 Pro CCC 201 203 0.75 0.75 Pro CCC 315 314 1.17 1.17 Pro CCG 150 151 0.56 0.56 Thr ACG 349 232 0.66 0.66 0.66 Thr ACG 149 147 0.45 0.44 0.4 DuA 54 535 1.81 1.81 1.81 tm/-C	Ile	AUA	688	687	0.92	0.92	trnI-CAT
Val GUU 508 507 1.43 1.43 ITMV-CAU Val GUC 181 183 0.51 0.52 Val GUG 198 198 0.56 0.56 Ser UCU 571 566 1.69 1.68 trmV-CAC Ser UCA 428 429 1.27 1.27 trmS-UG Ser UCA 133 1.52 trmP-UG CG 107 117 Pro CCC 201 203 0.75 0.75 trmV-CG Pro CCCA 315 314 1.17 1.17 117 Pro CCG 150 151 0.56 0.56 1.01 Thr ACU 534 535 1.61 1.61 trmT-UCU Thr ACC 248 248 0.75 0.75 1.02 Thr ACG 149 147 0.45 0.43 0.43 <td< td=""><td>Met</td><td>AUG</td><td>624</td><td>625</td><td>1</td><td>1</td><td>trnM-CAI</td></td<>	Met	AUG	624	625	1	1	trnM-CAI
Val CuC 181 183 0.51 0.52 Val CUC 181 183 0.51 0.52 Val CUG 181 183 0.51 0.52 Val CUG 198 198 0.56 0.56 Ser UCC 319 322 0.94 0.95 Ser UCA 428 429 1.27 1.27 tm7-UGA Ser UCG 189 188 0.56 0.56 0.56 Pro CCC 201 203 0.75 0.75 0.75 Pro CCC 315 314 1.17 1.17 1.7 Pro CCC 315 314 1.16 1.6 1.6 1.6 1.6 1.7 1.7 Pro CCC 315 314 1.17 1.7 1.7 1.7 1.6 1.6 1.6 1.6 1.6 1.6 1.6 1.6 1.6 1	Val	CUU	E09	507	1.40	1.40	trnjM-CAI
Val GUC 181 183 0.51 0.52 tmV-UAC Val GUG 198 198 0.56 0.56 15 tmV-UAC Ser UCU 571 566 1.69 1.68 tmS-GGA Ser UCA 428 429 1.27 tmS-UGA 127 Ser UCA 428 429 1.27 tmS-UGA 127 Pro CCU 413 410 1.53 1.52 tmP-UGA Pro CCC 201 203 0.75 0.75 tmT-UGA Pro CCG 150 151 0.56 0.56 tmT-UGU Thr ACU 534 535 1.61 1.61 tmT-GUU Thr ACC 248 248 0.75 0.75 tmT-UCU Thr ACG 149 147 0.45 0.44 tmT-UCU Thr ACG 149 150 0.43 0.43 <td>Val</td> <td>GUU</td> <td>101</td> <td>307</td> <td>1.45</td> <td>1.45</td> <td>UNV-GAC</td>	Val	GUU	101	307	1.45	1.45	UNV-GAC
Val GOA J30 J31 L3 L3 L1 L3 L1 L3 L1 L3 L1 L3 L1 L3 L1 L30 L1 L10 L10 L10 L10 L10 L10 L10 L10 L10 <thl10< th=""> <thl10< th=""> <thl10< th=""></thl10<></thl10<></thl10<>	Val	GUC	101	100 501	1.5	0.52	trol UAC
val GCU 571 566 1.69 1.68 tmS -GA Ser UCC 319 322 0.94 0.95 Ser UCA 428 429 1.27 1.27 tmS-GA Ser UCA 428 429 1.27 1.27 tmS-UGA Ser UCA 428 429 1.27 0.75 0.75 Pro CCC 201 203 0.75 0.75 0.75 Pro CCC 201 203 0.75 0.75 0.75 Thr ACU 534 535 161 161 161 tmT-GUU Thr ACC 248 248 0.75 0.75 0.75 Thr ACC 248 248 0.75 0.44 0.75 Thr ACG 147 0.45 0.44 0.41 0.41 Ala GCU 634 635 1.81 1.81 ms////////////////////////////////////	Val	CUC	108	108	0.56	0.56	univ-one
Str UCC 319 323 0.04 0.05 1.05 Gold Ser UCA 428 429 1.27 1.27 mS-UGA Ser UCG 189 188 0.56 0.56 Pro CCC 201 203 0.75 0.75 Pro CCC 201 203 0.75 0.75 Pro CCC 201 203 0.75 0.75 Pro CCC 150 151 0.56 0.56 Thr ACC 248 248 0.75 0.75 Thr ACC 248 248 0.75 0.75 Thr ACG 149 147 0.45 0.44 Ala GCC 231 232 0.66 0.66 Ala GCC 384 383 1.1 109 Ala GCC 134 153 151 mM-GUA Tyr UAC 184 <td< td=""><td>Ser</td><td></td><td>571</td><td>566</td><td>1.69</td><td>168</td><td>trnS_CCA</td></td<>	Ser		571	566	1.69	168	trnS_CCA
Carton UCA 428 429 127 127 tmS-UGA Ser UCG 189 188 0.56 0.56 Pro CCU 413 410 1.53 1.52 tmP-UGG Pro CCC 201 203 0.75 0.75 0.75 Pro CCA 315 314 1.17 1.17 1.17 Pro CCA 150 151 0.56 0.56 0.56 Thr ACU 534 535 1.61 1.61 tm1'-GU Thr ACC 248 248 0.75 0.75 m1'-UCU Thr ACG 149 147 0.45 0.44 m1'-UCU Thr ACG 149 150 0.43 0.43 0.43 Ala GCC 194 194 0.4 0.4 0.4 Yr UAC 194 194 0.4 0.4 0.4 Tyr UAC 194 194 0.4 0.4 0.4 STOP	Ser		319	300	0.94	0.95	1115-00/1
Ser UCG 189 188 0.56 0.56 0.56 Pro CCU 413 410 1.53 1.52 $tmP-UGG$ Pro CCC 201 203 0.75 0.75 0.75 Pro CCG 150 151 0.56 0.56 0.56 Thr ACC 248 248 0.75 0.75 0.75 Thr ACC 248 248 0.75 0.75 0.75 Thr ACC 248 248 0.75 0.75 0.75 Thr ACC 149 147 0.45 0.44 0.40 Ala GCU 634 635 1.81 1.81 $tmA-UGC$ Ala GCC 231 232 0.66 0.66 0.66 Ala GCG 149 150 0.43 0.43 0.43 Tyr UAC 184 193 1.93 1.51 1.54 Tyr UAC 194 194 0.44 0.4 0.4	Ser	UCA	428	429	127	127	trnS-UGA
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Ser	UCG	189	188	0.56	0.56	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Pro	CCU	413	410	1.53	1.52	trnP-UGG
Pro CCA 315 314 117 117 117 Pro CCG 150 151 0.56 0.56	Pro	CCC	201	203	0.75	0.75	
Pro CCG 150 151 0.56 0.56 Thr ACU 534 535 1.61 1.61 trnT-GGU Thr ACC 248 248 0.75 0.75 Thr ACG 397 399 1.2 1.2 trnT-UGU Ala CCU 634 635 1.81 1.81 trnA-UGC Ala CCC 231 232 0.66 0.66 0.43 Ala CCG 149 150 0.43 0.43 1.41 Tyr UAC 194 194 0.4 0.4 0.4 Tyr UAC 194 194 0.4 0.4 0.4 STOP UAA 54 54 1.93 1.93 1.93 STOP UAG 16 16 0.57 0.57 TrnH-GTG His CAU 477 1.53 1.53 trnH-GTG Gln CAG 215	Pro	CCA	315	314	1.17	1.17	
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Pro	CCG	150	151	0.56	0.56	
Thr ACA 397 399 1.2 1.3 1.3 1.3 1.3 1.3 1.3 1.3 1.3 1.3 1.3 1.3 1	Thr	ACU	534	535	1.61	1.61	trnT-GGU
ThrACA ACG3973991.21.21.2trThrACG1491470.450.44AlaGCU6346351.811.81trmA-UGCAlaGCC2312320.660.66AlaGCG1491500.430.43TyrUAU7827841.61.6TyrUAC1941940.40.4TyrUAC16160.570.57TOPUAC16160.570.57HisCAC4171460.440.4ChnCAA7197201.541.53trmP-GUAClnCAA7197201.541.53trmN-GTTAsnAAC2932980.460.47trmD-GTCLysAAA48421.51.5trmC-UULysAAG3503480.50.5TrmD-GTCAspGAU8738711.611.61trmD-GTCCysUGU2162161.471.47trmC-GCACysUGU2162161.471.49trmR-ACGArgCGC131000.390.38trmR-ACGCysUGU2162161.411.41trmR-ACGArgCGC131000.390.38trmR-ACGCysUGU2162161.411.41trm	Thr	ACC	248	248	0.75	0.75	
Thr ACG 149 147 0.45 0.44 Ala GCU 634 635 1.81 1.81 trnA-UGC Ala GCC 231 232 0.66 0.66 0.66 Ala GCC 231 232 0.66 0.66 0.66 Ala GCG 149 150 0.43 0.43 0.43 Tyr UAC 184 194 0.4 0.4 0.4 0.4 STOP UAA 54 54 1.93 1.93 1.93 1.93 STOP UAG 16 16 0.57 0.57 1.53 trnH-GTG His CAU 477 146 0.47 0.47 .41 Gln CAG 215 216 0.46 0.46 .41 <td>Thr</td> <td>ACA</td> <td>397</td> <td>399</td> <td>1.2</td> <td>1.2</td> <td>trnT-UGU</td>	Thr	ACA	397	399	1.2	1.2	trnT-UGU
Ala GCU 634 635 1.81 1.81 $trnA-UGC$ Ala GCC 231 232 0.66 0.66 0.66 Ala GCA 384 383 1.1 109 Ala GCG 149 150 0.43 0.43 Tyr UAU 782 784 1.6 1.6 trnY-GUA Tyr UAC 194 194 0.4 0.4 3.3 1.33 STOP UAG 16 16 0.57 0.57 0.57 1.53 1.53 trnH-GTG His CAU 477 146 0.47 0.47 0.47 1.7G 1.61 1.64 1.93 1.93 1.93 1.93 1.93 1.93 1.93 1.93 1.93 1.7G 1.53 1.51 trnH-GTG 1.53 1.53 trnH-GTG 1.53 1.51 trnQ-TG 1.61 1.64 1.53 trnN-GTT 3.83 9.82 1.54 1.53 trnN-GTG 1.53 trnK-UUU 1.55 1.5 trnK-UUU 1	Thr	ACG	149	147	0.45	0.44	
Ala GCU 634 635 1.81 1.81 trnA-UGC Ala GCC 231 232 0.66 0.66 Ala GCC 149 150 0.43 0.43 Tyr UAU 782 784 1.6 1.6 trnY-GUA Tyr UAC 194 194 0.4 0.4 0.4 STOP UAA 54 1.93 1.93 53 trnY-GUA STOP UAA 54 1.93 1.93 trnY-GUA 1.6 1.6 STOP UAA 54 1.53 1.53 trnH-GTG 1.53 1.53 trnH-GTG His CAC 147 146 0.47 0.47 .71 GIA 6.46 .46 .48 .53 trnQ-TTG 1.54 1.53 trnQ-TTG .53 trnN-GTT .53 trnQ-TG .54 .53 1.51 trnQ-TG .53 .53 .53 .54 .53 .54 .53 .55 .55 .55 .54 .54 .53 .53 </td <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>							
Ala CCC 231 232 0.66 0.66 Ala GCA 384 383 1.1 109 Ala GCG 149 150 0.43 0.43 Tyr UAU 782 784 1.6 1.6 trnY-GUA STOP UAA 54 54 1.93 1.93 STOP UAG 16 16 0.57 0.57 His CAU 475 477 1.53 1.53 trnH-GT Gln CAA 719 720 1.54 1.54 trnQ-TTG Gln CAA 719 720 1.54 1.53 trnN-GTT Asn AAU 983 982 1.54 1.53 trnW-GTT Lys AAG 350 348 0.5 0.5 0.5 Lys AAG 350 348 0.5 0.5 0.5 Clu GAA 20 22 1.49 1.49 trnE-TTC Glu GAG 348 349 0.51	Ala	GCU	634	635	1.81	1.81	trnA-UGC
Ala CCA 384 383 1.1 1.09 Ala CCG 149 150 0.43 0.43 Tyr UAU 782 784 1.6 1.6 trny-GUA Tyr UAC 194 194 0.4 0.4 0.4 STOP UAG 16 16 0.57 0.57 1.53 trnH-GTG His CAU 475 477 1.53 1.53 trnH-GTG Gln CAA 719 720 1.54 1.54 trnQ-TTG Gln CAG 215 216 0.46 0.47 0.47 Lys AAA 48 42 1.5 1.5 trnN-GTT Asn AAC 293 298 0.46 0.47 0.5 Lys AAG 350 348 0.5 0.5 0.5 Asp GAU 873 871 1.61 1.61 trnE-TTC Glu GAG 348 349 0.51 0.51 CCAC CS 1.7	Ala	GCC	231	232	0.66	0.66	
Ana CCG 149 150 0.43 0.43 Tyr UAU 782 784 1.6 1.6 tr Tyr UAC 194 194 0.4 0.4 0.4 STOP UAA 54 54 1.93 1.93 . STOP UAA 54 54 1.93 1.93 . STOP UAA 54 54 1.53 1.53 trm/-GUA His CAU 475 477 1.53 1.53 trm/-GTG His CAC 147 146 0.46 0.46 . Gln CAG 215 216 0.46 0.47 . Lys AAA 48 42 1.5 1.5 trn/-CTC Asp GAU 873 871 1.61 1.61 trn/-D-GTC Asp GAU 873 871 1.61 1.61 trnE-TTC Clu	Ala	GCA	384	383	1.1	1.09	
Iyi DAO 782 784 1.5 1.6 117-COA Tyr UAC 194 194 0.4 0.4 0.4 STOP UAG 16 16 0.57 0.57 His CAU 477 1.53 1.53 trnH-GTG His CAC 147 146 0.47 0.47 Gln CAA 719 720 1.54 1.53 trnN-GTG Asn AAU 983 982 1.54 1.53 trnN-GTT Asn AAC 293 298 0.46 0.47	Ald	GCG	149	150	0.45	0.45	tun V CUA
Tyr Drc 194 194 0.4 STOP UAG 16 16 0.57 0.57 His CAU 475 477 1.53 1.53 trnH-GTG His CAU 475 477 1.53 1.53 trnH-GTG Gln CAA 719 720 1.54 1.54 trnQ-TTG Gln CAG 215 216 0.46 0.47	Tyr	UAC	10/	104	0.4	0.4	UNIT-GOA
STOP UAG 16 16 0.57 0.57 His CAU 475 477 1.53 1.53 trnH-GTG His CAC 147 146 0.47 0.47 Gln CAG 215 216 0.46 0.46 Asn AAU 983 982 1.54 1.53 trnV-GTT Asn AAC 293 298 0.46 0.47 0.47 Lys AAA 48 42 1.5 1.5 trnK-UUU Lys AAG 350 348 0.5 0.5 348 0.5 0.5 Asp GAU 873 871 1.61 1.61 trnD-GTC Asp GAU 873 878 0.53 0.53 0.53 0.53 0	STOP	LIAA	54	54	1.93	1.93	
Arise Arise Arise Arise Arise Arise His CAU 475 477 1.53 1.53 trnH-GTC His CAC 147 146 0.47 0.47 Cln CAA 719 720 1.54 1.54 trnQ-TTG Cln CAA 719 720 1.54 1.53 trnQ-TTG Asn AAU 983 982 1.54 1.53 trnN-GTT Asn AAC 293 298 0.46 0.47	STOP	LIAG	16	16	0.57	0.57	
InitI	His	CAU	475	477	153	153	trnH-GTG
ChCATipTipTipTipTipChCAG2152160.460.46AsnAAU9839821.541.53 $trnN-GTT$ AsnAAU9839821.541.53 $trnN-GTT$ LysAAA48421.51.5 $trnV-GTC$ AspGAU8738711.611.61 $trnD-GTC$ AspGAU8738711.611.61 $trnD-GTC$ AspGAU8738711.611.61 $trnD-GTC$ CluGAA20221.491.49 $trnE-TTC$ CluGAG3483490.510.51CCysUGC78780.530.53STOPUGA14140.50.5TrnW-CCAArgCGU3543541.341.34 $trnR-ACG$ ArgCGC1031000.390.38 $trnR-ACG$ ArgCGA3683701.41.41AtrnR-ACGArgCGA3683701.41.41AtrnR-ACGArgCGA131120.430.43AtrnR-ACGSerAGU4013991.191.18 $trnG-GCU$ SerAGC1211220.360.36AtraArgAGA4744781.81.82ArgArgAGA1660.640.63AtraAtra </td <td>His</td> <td>CAC</td> <td>147</td> <td>146</td> <td>0.47</td> <td>0.47</td> <td></td>	His	CAC	147	146	0.47	0.47	
Cln CAG 215 216 0.46 0.46 Asn AAU 983 982 1.54 1.53 trnN-GTT Asn AAC 293 298 0.46 0.47 Lys Lys AAA 48 42 1.5 1.5 trnK-UUU Lys AAG 350 348 0.5 0.5 - Asp GAU 873 871 1.61 1.61 trnD-GTC Asp GAC 211 209 0.39 0.39 - Clu GAA 20 22 1.49 1.49 trnC-GTC Asp GAU 873 871 0.51 0.51 - Clu GAA 20 22 1.49 1.49 trnC-GTC Cys UGC 78 78 0.53 0.53 - Trp UGG 464 462 1 1 trnR-ACG Arg CGC	Gln	CAA	719	720	1.54	1.54	trnO-TTG
Asn AAU 983 982 1.54 1.53 trnN-GTT Asn AAC 293 298 0.46 0.47	Gln	CAG	215	216	0.46	0.46	
Asn AAC 293 298 0.46 0.47 Lys AAA 48 42 1.5 1.5 trnK-UUU Lys AAG 350 348 0.5 0.5 Asp GAU 873 871 1.61 1.61 trnD-GTC Asp GAC 211 209 0.39 0.39 Clu GAA 20 22 1.49 1.49 trnE-TTC Clu GAG 348 349 0.51 0.51 Cys UGU 216 216 1.47 1.47 trnC-GCA Cys UGA 14 14 0.5 0.53 Trp UGA 464 462 1 1 trnR-ACG Arg CGU 354 1.34 1.34 trnR-ACG Arg CGC 103 100 0.39 0.38 trnR-ACG Arg	Asn	AAU	983	982	1.54	1.53	trnN-GTT
Lys AAA 48 42 1.5 1.5 trnK-UUU Lys AAG 350 348 0.5 0.5 Asp GAU 873 871 1.61 1.61 trnD-GTC Asp GAU 211 209 0.39 0.39 0.39 Glu GAA 20 22 1.49 1.49 trnE-TTC Clu GAG 348 349 0.51 0.51 0.51 Cys UGC 78 78 0.53 0.53 0.53 STOP UGA 14 14 0.5 0.5 0.5 Trp UGG 464 462 1 1 trnW-CCA Arg CGU 354 1.34 1.34 trnR-ACG Arg CGC 103 100 0.39 0.38 trnR-TCT Arg CGA 368 370 1.4 1.41 1.4 1.41 Arg	Asn	AAC	293	298	0.46	0.47	
Lys AAG 350 348 0.5 0.5 Asp GAU 873 871 1.61 1.61 trnD-GTC Asp GAC 211 209 0.39 0.39 - Clu GAA 20 22 1.49 1.49 trnE-TTC Clu GAG 348 349 0.51 0.51 - Cys UGU 216 216 1.47 1.47 trnC-GCA Cys UGG 78 78 0.53 0.53 - Trp UGG 464 462 1 1 trnR-CCA Arg CGU 354 354 1.34 1.34 trnR-ACC Arg CGC 103 100 0.39 0.38 trnR-TCT Arg CGG 113 112 0.43 0.43	Lys	AAA	48	42	1.5	1.5	trnK-UUU
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Lys	AAG	350	348	0.5	0.5	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Asp	GAU	873	871	1.61	1.61	trnD-GTC
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Asp	GAC	211	209	0.39	0.39	
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Glu	GAA	20	22	1.49	1.49	trnE-TTC
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Glu	GAG	348	349	0.51	0.51	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Cys	UGU	216	216	1.47	1.47	trnC-GCA
S10P UGA 14 14 0.5 0.5 Trp UGG 464 14 0.5 0.5 Arg CGU 354 354 1.34 1.34 tmR-ACG Arg CGC 103 100 0.39 0.38 tmR-TCT Arg CGG 113 112 0.43 0.43	Cys	UGC	78	78	0.53	0.53	
Irp UGG 464 462 I I I I It It<	STOP	UGA	14	14	0.5	0.5	
Arg CGU 534 1.54 1.54 1.73 1.75 1.77 1.73 1.73 1.77 1.73 1.71 1.71 1.77 1.73 1.73 1.71 1.71 1.72 0.36 0.36 0.36 1.75 <t< td=""><td>1rp Ang</td><td>UGG</td><td>464</td><td>462</td><td>1 24</td><td>1 24</td><td>trmP_ACC</td></t<>	1rp Ang	UGG	464	462	1 24	1 24	trmP_ACC
Arg CGC 105 100 0.59 0.58 1111 Arg CGA 368 370 1.4 1.41 Arg CGG 113 112 0.43 0.43 Ser AGU 401 399 1.19 1.18 tmS-GCU Ser AGC 121 122 0.36 0.36	Aig	CGO	334 102	334 100	1.54	1.34	true D TCT
Arg CGA 365 376 1.4 1.41 Arg CGG 113 112 0.43 0.43 Ser AGU 401 399 1.19 1.18 tmS-GCU Ser AGC 121 122 0.36 0.36 Arg AGA 474 478 1.8 1.82 Arg AGG 168 166 0.64 0.63 Gly GGU 606 606 1.35 1.35 tmG-GCC Gly GGA 727 727 1.62 1.62 Gly GGG 295 292 0.66 0.65	Arg	CCA	368	370	14	1.30	UNK-ICI
Arg AGU 401 399 1.19 1.18 tmS-GCU Ser AGU 401 399 1.19 1.18 tmS-GCU Ser AGC 121 122 0.36 0.36 Arg AGA 474 478 1.8 1.82 Arg AGG 168 166 0.64 0.63 Gly GGU 606 606 1.35 1.35 trnG-GCC Gly GGC 166 167 0.37 0.37 GIV GGG 295 292 0.66 0.65	Aro	CGG	113	112	0.43	0.43	
Arg AGC 121 122 0.36 0.36 Arg AGA 474 478 1.8 1.82 Arg AGG 168 166 0.64 0.63 Gly GGU 606 606 1.35 1.35 tmG-GCC Gly GGC 166 167 0.37 0.37 Gly GGG 295 292 0.66 0.65	Ser	AGU	401	399	119	118	trnS-CCU
Arg AGA 474 478 1.8 1.82 Arg AGG 168 166 0.64 0.63 Cly GGU 606 606 1.35 1.35 tmG-GCC Cly GGC 166 167 0.37 0.37 Gly GGG 295 292 0.66 0.65	Ser	AGC	121	122	0.36	0.36	1110-000
Arg AGG 163 162 Arg AGG 168 166 0.63 Gly GGU 606 606 1.35 1.35 Cly GGC 166 167 0.37 0.37 Gly GGA 727 727 1.62 1.62 Gly GGG 295 292 0.66 0.65	Arg	AGA	474	478	1.8	1.82	
Gly GGU 606 606 1.35 1.35 tmG-GCC Gly GGC 166 167 0.37 0.37 Gly GGA 727 727 1.62 1.62 Gly GGG 295 292 0.66 0.65	Arg	AGG	168	166	0.64	0.63	
Gly GGC 166 167 0.37 0.37 Gly GGA 727 727 1.62 1.62 Gly GGG 295 292 0.66 0.65	Gly	GGU	606	606	1.35	1.35	trnG-GCC
Gly GGA 727 727 1.62 1.62 Gly GGG 295 292 0.66 0.65	Gly	GGC	166	167	0.37	0.37	
Gly GGG 295 292 0.66 0.65	Gly	GGA	727	727	1.62	1.62	
	Gly	GGG	295	292	0.66	0.65	

*RSCU (Relative synonymous codon usage) value \geq 1.00

^a Frequency of codon usage in 23,224 and 23,220 codons in all potential protein-coding genes of *L. acutangula* and *L. aegyptiaca*, respectively;

^b Gene encoding transfer RNA



Fig. 3. Comparison of the chloroplast genome borders of the LSC, SSC, and IR regions among six species, ψ partial fragment of the ycf1 gene.



Fig. 4. Alignment of chloroplast genome sequences, showing percent similarity, among six species using L. acutangula as a reference.



Fig. 5. Simple sequence repeat (SSR) analysis in *L. acutagula* and *L. acgyptiaca* chloroplast genomes. (a) SSR percentage in the LSC, SSC and IR regions, (b) Number of SSR per motif size.



Fig. 6. Phylogenetic relationship of 17 species within Cucurbitaceae family based on 66 protein-coding chloroplast genes. *O. sativa* and *A. thaliana* are outgroups. Numbers above the node are the bootstrap values of maximum likelihood (ML) analysis.

2. Experimental Design, Materials and Methods

2.1. DNA extraction, sequencing and assembly

Young leaves of *L. acutangula* (ridge gourd) and *L. aegyptiaca* (smooth gourd) plants from Chia Tai Company Limited were collected at National Omics Center, Thailand Science Park, Pathum Thani, Thailand in March 2019 for DNA extraction. Genomic DNA was extracted using a CTAB method [2]. Total DNA was examined using a NanoDrop One spectrophotometer (Thermo Scientific, Wilmington, USA) and visualized by pulsed-field gel electrophoresis (PFGE). High quality DNA was used to construct PacBio libraries according to the 'Procedure & Checklist–20 Kb Template Preparation Using Bluepippin Size Selection System' protocol and sequenced on the PacBio RSII system. The short PacBio reads were used to correct the long PacBio reads and the corrected long reads were assembled using CANU version 1.4 software [3]. The resulting contigs were blasted against the plastid genome database to identify any chloroplast contigs, which were used to construct full chloroplast genomes.

Candidate single nucleotide polymorphisms (SNPs) identified in CDS between the reference (*L. Acutangula*) and *L. aegyptiaca*.

Position	Reference	L. aeg	Sustitutionsa	Gene	Function
1973	Т	С	NS	matK	Maturease K
3132	G	Т	S	matK	Maturease K
5299	Т	G	NS	rps16	30S ribosomal protein S16
8127	С	А	NS	psbK	Photosystem II reaction center protein K
8217	С	А	NS	, psbK	Photosystem II reaction center protein K
12059	G	Т	S	atpA	ATP synthase subunit alpha
13328	G	Т	S	atpF	ATP synthase subunit b
17060	G	Т	S	rps2	30S ribosomal protein S2
17982	С	А	NS	rpoC2	DNA-directed RNA polymerase subunit beta
18665	С	А	NS	rpoC2	DNA-directed RNA polymerase subunit beta
19148	С	Т	S	rpoC2	DNA-directed RNA polymerase subunit beta
19540	С	А	NS	rpoC2	DNA-directed RNA polymerase subunit beta
20274	G	Т	NS	rpoC2	DNA-directed RNA polymerase subunit beta
20678	А	G	S	rpoC2	DNA-directed RNA polymerase subunit beta
20777	А	G	S	rpoC2	DNA-directed RNA polymerase subunit beta
25097	G	Т	S	rpoB	DNA-directed RNA polymerase subunit beta
26705	С	Т	S	rpoB	DNA-directed RNA polymerase subunit beta
27002	C	Т	S	rpoB	DNA-directed RNA polymerase subunit beta
35125	G	С	NS	psbD	Photosystem II D2 protein
51601	G	Т	NS	ndh]	NAD(P)H-quinone oxidoreductase subunit [
52335	G	Т	S	ndhK	NAD(P)H-quinone oxidoreductase subunit K
55091	A	Т	S	atpE	ATP synthase epsilon chain
55260	Т	G	NS	atpB	ATP synthase subunit beta
55588	С	A	S	atpB	ATP synthase subunit beta
56576	G	А	NS	atnB	ATP synthase subunit beta
57691	T	G	NS	rbcL	Ribulose bisphosphate carboxylase large chain
59684	A	c	NS	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
59876	С	A	NS	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
59878	C	G	NS	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
59913	G	c	S	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
60037	Ā	G	S	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
60042	Т	G	S	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
60169	Т	C	NS	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
60287	С	A	S	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
60384	G	С	S	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
60417	С	G	S	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
60615	C	G	S	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
60665	G	Т	S	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
60914	G	С	NS	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
60921	Т	G	S	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
60963	А	G	S	accD	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
62698	С	A	S	vcf4	Proteins of unknown function
63405	C	А	S	cemA	Chloroplast envelope membrane protein
63691	A	С	NS	cemA	Chloroplast envelope membrane protein
64793	G	A	S	netA	Cytochrome f
67969	Т	G	S	petG	Cytochrome b6-f complex subunit 5
112795	Т	G	NS	ndhF	NAD(P)H-quinone oxidoreductase subunit 5
112868	С	G	NS	ndhF	NAD(P)H-quinone oxidoreductase subunit 5
112869	C	A	NS	ndhF	NAD(P)H-quinone oxidoreductase subunit 5
113666	C	А	S	ndhF	NAD(P)H-quinone oxidoreductase subunit 5
114616	C	G	NS	ndhF	NAD(P)H-quinone oxidoreductase subunit 5
114678	G	Ā	NS	ndhF	NAD(P)H-quinone oxidoreductase subunit 5
117774	Т	С	S	ccsA	Cytochrome c biogenesis protein

Note: L. aeg, Luffa aegyptiaca; a Ns: Non-synonymous, S: Synonymous

Young leaves of *L. acutangula* and *L. aegyptiaca* seedlings (Chia Tai Co, Ltd) were harvested and genomic DNA isolated using the High Pure PCR Template Preparation kit of Roche. Genomic DNA was examined using a NanoDrop One spectrophotometer (Thermo Scientific, Wilmington, USA). High quality DNA was used to prepare Illumina Hiseq X Ten libraries and 150 bp pair-end sequencing was performed by Novogene, Singapore according to standard Illumina protocols.

Table 6
Comparison of RNA editing patterns in L. acutangula and L. aegyptiaca chloroplast genomes with other species.

Location	Gene	AA position	Codon conversion	AA Change	Substitution	L. acutangula	L. aegyptiaca	C. sativus	C. pepo	A. thaliana	N. tabacum
LSC	atpA	258	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(-)	(+)	(-)	(-)	(-)	(-)
		305	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(-)	(+)	(-)	(-)	(-)	(-)
		383	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(-)	(+)	(-)	(-)	(-)	(-)
	atpF	31	cCa→cUa	$P \rightarrow L$	Nonsynonymous	(+)	(+)	(+)	(+)	(+)	(+)
	rps2	83	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(-)	(+)	(+)	(+)	-	(+)
	rpoC2	1,245	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(+)	(+)	(+)	(+)	-	(+)
	rpoB	809	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(-)	(+)	(+)	(+)	(+)	(+)
	ndhK	22	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(+)	(+)	(-)	(-)	(-)	(-)
	petA	273	$Cag \rightarrow Uag$	Q→Q	Synonymous	(-)	-	(-)	(-)	(-)	(-)
		276	$gCg \rightarrow gUg$	$A \rightarrow S$	Nonsynonymous	(-)	(+)	(-)	(-)	(-)	(-)
		279	guC→guU	$V \rightarrow V$	Synonymous	(-)	-	(-)	(-)	(-)	(-)
	psbJ	20	cCu→cUu	$P \rightarrow L$	Nonsynonymous	(+)	(+)	(-)	(-)	(-)	(-)
	psbF	26	uCu→uUu	$S \rightarrow F$	Nonsynonymous	(+)	(+)	(+)	(-)	(+)	(+)
	rpoA	67	uCu→uUu	$S \rightarrow F$	Nonsynonymous	(+)	(+)	(-)	(-)	(-)	(-)
		277	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(+)	(+)	(+)	(+)	-	(+)
	rps11	36	uuC→uuU	$F \rightarrow F$	Synonymous	-	-	(-)	(-)	(-)	(-)
IRb	rpl23	24	uCu→uUu	$S \rightarrow F$	Nonsynonymous	(-)	(+)	(-)	(-)	(-)	(-)
SSC	ndhD	97	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(+)	(-)	(-)	(-)	(-)	(-)
		194	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(+)	(+)	(-)	(-)	(-)	(-)
		262	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(-)	(+)	(-)	(-)	(-)	(-)
		265	uCg→uUg	$S \rightarrow L$	Nonsynonymous	(+)	(-)	(-)	(-)	(-)	(-)
	ndhE	77	cCa→cUa	P→L	Nonsynonymous	(+)	(+)	(-)	(-)	(-)	(-)
	ndhA	114	uCa→uUa	$S \rightarrow L$	Nonsynonymous	(+)	(+)	(-)	(-)	(+)	(+)
	ndhH	169	Cau→Uau	$H \rightarrow Y$	Nonsynonymous	(+)	(+)	(-)	(-)	(-)	(-)

Capital letters in codon triplets indicate target nucleotides; AA, Amino acid; (+), editing; (-), no editing; -, U encoded in the DNA (no editing); Blank space, Silent mutation

Polycistronic gene clusters in *L. acutangula* and *L. aegyptiaca* chloroplast genomes.

Function	Gene cluster	Luffa acutangula Genes	Position	Length (bp)	Luffa aegyptiaca Genes	Position	Length (bp)
ATP synthase	atp-1	atpI+atpH	16,50714,566	1,942	atpI+atpH	16,51114,570	1,942
Ribosomal protein, ATP synthase	atp-2	rps2+atpI+atpH	17,42214,566	2,857	rps2+atpl	17,43215,768	1,665
NADH oxidoreductase	ndh-1	ndhC+ndhK+ndhJ	52,89451,215	1,680	ndhC+ndhK+ndhJ	52,97051,292	1,679
NADH oxidoreductase	ndh-2	ndhE+psaC+ndhD	120,578118,128	2,451	ndhE+psaC+ndhD	120,668118,224	2,445
Photosystem II	psb-1	psbE+psbF+psbL+psbJ	66,38865,615	774	psbE+psbF+psbL+psbJ	66,49365,721	773
Ribosomal protein	rpl-1	rpl14+rps8+infA+ rpl36+rps11	82,93680,856	2,081	rpl16+rpl14+rps8+ infA+rpl36+rps11	84,67880,945	3,734
Ribosomal protein	rpl-2	-	-	-	rpl22+rps3	85,96384,819	1,145
Ribosomal protein	rpl-3	-	-	-	rpl23+rpl2+rps19	88,16386,033	2,131
Ribosomal protein	rps-1	-	-	-	rps12+rpl20	71,65270,393	1,260
Ribosomal protein	rps-2	-	-	-	rps19+rpl22+rps3	86,31184,819	1,493
Ribosomal protein, NADH oxidoreductase	rps-3	rps15+ndhH	126,075124,517	1,559	rps15+ndhH	126,156124,599	1,558
Ribosomal RNAs	rrn-1	rrn23+rrn4.5+rrn5	106,587.109,977	3,391	rrn23+rrn4.5+rrn5	106,675110,065	3,391

2.2. Chloroplast genome annotation

The assembled chloroplast genomes of *L. acutangula* and *L. aegyptiaca* were annotated using GeSeq MPI-MP CHLOROBOX tool [4], specifically HMMER, tRNAscan and ARAGORN. An annotated genome map was generated using Organellar Genome DRAW (OGDRAW) [5]. Finally, the preliminary annotations were corrected manually to ensure that the correct start and stop positions were reported.

2.3. Codon usage analysis

L. acutangula and *L. aegyptiaca* coding sequences were used to calculate relative synonymous codon usage (RSCU) value using CodonW version 1.4.2 software [6]. Codon usage frequency was calculated and expressed as the number of codons encoding the same amino acid divided by the total number of codons [7].

2.4. Comparative structure analysis

IR regions in the chloroplast genomes of *L. acutangula, L. aegyptiaca, Cucumis melo* (NC_015983), *Cucumis sativus* (NC_007144), *Citrullus lanatus* (NC_032008), and *Cucurbita pepo* (NC_038229) were compared using IRscope software [8]. Sequences of all analyzed chloroplast genomes were aligned using LAGAN mode of mVISTA alignment software [9] (http://genome.lbl. gov/vista/mvista/submit.shtml).

2.5. Simple sequence repeat (SSR) analysis

L. acutangula and *L. aegyptiaca* chloroplast genomes were scanned for simple sequence repeats (SSRs) using MIcroSAtellite (MISA) identification tool [10]. The length threshold of minimum repetitive units were set to ten repeats for mono-nucleotide repeats, four repeats for diand tri-nucleotide repeats, and three repeats for tetra-, penta- and hexa-nucleotide repeats according to the method of Ivanova and co-workers [11].

2.6. Single nucleotide polymorphism (SNP) identification

Illumina sequences were mapped to the chloroplast genomes using Burrows-Wheeler Aligner (BWA-MEM) software [12]. SNPs were identified from *L. acutangula* and *L. aegyptiaca* using Genome Analysis Toolkit (GATK) software v 4.1.2.0 [13]. All SNPs were filtered with criteria of read depth \geq 20 and missing data \leq 10%.

2.7. RNA editing analysis and polycistronic mRNA in chloroplast genomes

RNA sequencing of *L. acutangula* [SRA accession number: SRR11445640] and *L. aegyptiaca* [SRA accession number: SRR11452010] from isoform sequencing (Iso-seq) were obtained from a previous study of Pootakham et al. (2020) [1]. These long-read sequences were mapped to their corresponding chloroplast genomes using BWA-MEM software [12]. Subsequently, RNA editing sites were checked by calling SNPs using GATK and comparing to the genomic SNP data [13]. The RNA reads were mapped against their respective chloroplast genome sequence using blastN version 2.2.28 to identify single reads that spanned more than one gene to identify gene clusters that are co-transcribed.

2.8. Phylogenetic analysis

The chloroplast genomes of *L. acutangula* and *L. aegyptiaca*, together with 13 chloroplast genomes in the lineage of the Cucurbitaceae family were selected to analyze phylogenetic relationships. The 13 other species were *Cucumis melo* (NC_015983), *Cucumis sativus* (NC_007144), *Coccinia grandis* (NC_031834), *Citrullus lanatus* (NC_032008), *Lagenaria siceraria* (NC_036808), *Cucurbita maxima* (NC_036505), *Cucurbita moschata* (NC_036506), *Cucurbita pepo* (NC_038229), *Trichosanthes kirilowii* (NC_041088), *Hodgsonia macrocarpa* (NC_039628), *Momordica charantia* (NC_036807), *Siraitia grosvenorii* (NC_043881), and *Gynostemma pentaphyllum* (NC_029484). *Oryza sativa* (NC_031333) and *Arabidopsis thaliana* (NC_000932) were also included as outgroups. Sixty-six protein coding genes, conserved among these 17 species (Table S3), were aligned using Kalign software [14], and a phylogenetic tree was constructed using MEGA-X software [15] with the maximum likelihood (ML) method. Bootstrap analysis was calculated by 1000 replications for correction.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

This work was supported by the National Science and Technology Development Agency (NSTDA) [grant number 2000025], Thailand.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2020.106470.

References

- W. Pootakham, C. Sonthirod, C. Naktang, W. Nawae, T. Yoocha, W. Kongkachana, D. Sangsrakru, N. Jomchai, S. Uthoomporn, J.R. Sheedy, J. Buaboocha, S. Mekiyanon, S. Tangphatsornruang, De novo assemblies of *Luffa acutangula* and *Luffa cylindrica* genomes reveal an expansion associated with substantial accumulation of transposable elements, Mol. Ecol. Resour. 1755–0998 (2020) 13240, doi:10.1111/1755-0998.13240.
- [2] M.A. Lodhi, G.N. Ye, N.F. Weeden, B.I. Reisch, A simple and efficient method for DNA extraction from grapevine cultivars and Vitis species, Plant Mol. Biol. Rep. 12 (1994) 6–13, doi:10.1007/BF02668658.
- [3] S. Koren, B.P. Walenz, K. Berlin, J.R. Miller, N.H. Bergman, A.M. Phillippy, Canu: scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation, Genome. Res. 27 (2017) 722–736, doi:10.1101/gr. 215087.116.
- [4] M. Tillich, P. Lehwark, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock, S. Greiner, GeSeq Versatile and accurate annotation of organelle genomes, Nucleic Acids Res. 45 (2017) W6–W11, doi:10.1093/nar/gkx391.
- [5] S. Greiner, P. Lehwark, R. Bock, OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes, nucleic acids research. 17 (2019) W59–W64. 10.1093/nar/gkz238.
- [6] J.F. Peden, Analysis of codon usage, Biosystems 106 (1999) 45-50, doi:10.1016/j.biosystems.2011.06.005.
- [7] S. Tangphatsornruang, D. Sangsrakru, J. Chanprasert, P. Uthaipaisanwong, T. Yoocha, N. Jomchai, S. Tragoonrung, The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships, DNA Res. (2010), doi:10.1093/dnares/dsp025.
- [8] A. Amiryousefi, J. Hyvönen, P. Poczai, IRscope: an online program to visualize the junction sites of chloroplast genomes, Bioinformatics (Oxford, England). 34 (2018) 3030–3031. 10.1093/bioinformatics/bty220.
- [9] M. Brudno, C.B. Do, G.M. Cooper, M.F. Kim, E. Davydov, E.D. Green, A. Sidow, S. Batzoglou, LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA, Genome. Res. 13 (2003) 721–731, doi:10.1101/gr. 926603.

- [10] T. Thiel, W. Michalek, R.K. Varshney, A. Graner, Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.), Theor. Appl. Genet. 106 (2003) 411–422, doi:10.1007/ s00122-002-1031-0.
- [11] Z. Ivanova, G. Sablok, E. Daskalova, G. Zahmanova, E. Apostolova, G. Yahubyan, V. Baev, Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response, Front. Plant Sci. 8 (2017) 204, doi:10.3389/fpls.2017.00204.
- [12] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (2009) 1754–1760, doi:10.1093/bioinformatics/btp324.
- [13] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo, The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (2010) 1297–1303, doi:10.1101/gr.107524.110.
- [14] T. Lassmann, E.L.L. Sonnhammer, Kalign, Kalignvu and Mumsa: web servers for multiple sequence alignment, Nucleic Acids Res. 34 (2006) W596–W599, doi:10.1093/nar/gkl191.
- [15] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549, doi:10.1093/molbev/msy096.