European Commission

Bioinformatics evaluation of GMO reference genes and their amplification systems

Colaiacovo, M.*, Bonfini, L., Mazzara, M., Savini, C.

* Seidor srl

2024

Joint Research Centre

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC136933

Geel: European Commission, 2024

How to cite this report: European Commission, Joint Research Centre, Colaiacovo, M., Bonfini, L., Mazzara, M. and Savini, C., *Bioinformatics evaluation of GMO reference genes and their amplification systems*, European Commission, Geel, 2024, JRC136933.

The work was carried out for the JRC in its operation of the EURL-GMFF at the JRC Ispra site..

Contents

Abstract

The present report summarises the bioinformatics analyses performed by the European Union Reference Laboratory for Genetically Modified Food and Feed (EURL GMFF) on the reference genes commonly used for detecting GM events in maize, soybean, cotton and rapeseed. The analysis had two objectives: first, to estimate the copy number of these genes in the available genomes of these species to ensure their low and constant copy number in the varieties possibly present on the market; second, to predict expected and unexpected amplifications using *in silico* PCR simulations to verify the specificity of the related taxon-specific methods. In maize, the *adh1* gene was found in three copies in one genome and in a single copy in all others. All of the copies are predicted to be amplified by the QT-TAX-ZM-001 method. In the same species, one copy of the *hmg* gene was found in all genomes, and is predicted to be amplified by the QT-TAX-ZM-002 method. In soybean, one copy of the *le1* gene was found in all genomes, and is predicted to be amplified by the QT-TAX-GM-002 and QT-TAX-GM-003 methods. In cotton, two copies of the *sah7* gene were found in all genomes; both copies are predicted to be amplified by the QT-TAX-GH-021 method. In the same species, two copies of the *adhC* gene were found in all genomes, however only one gene copy is predicted to be amplified by the QT-TAX-GH-019 method. In rapeseed, two copies of the *fatA(A)* gene were found in 5 out of 6 genomes, but only one gene copy is predicted to be amplified by the QT-TAX-BN-001 method. In the remaining genome, four copies of the *fatA(A)* gene were found, but only two of them are predicted to be amplified by the QT-TAX-BN-001 method.

1 Introduction

The present report summarises the bioinformatics analyses performed by the European Union Reference Laboratory for Genetically Modified Food and Feed (EURL GMFF) on the reference genes commonly used to detect GM events in maize, soybean, cotton and rapeseed. The goal of this analysis was to estimate the copy number of these genes in the available genomes of these species to ensure a low and constant copy number in the cultivated varieties possibly present on the market, and to predict expected and unexpected amplifications using *in silico* PCR simulations to verify the specificity of the related taxon-specific methods.

Both analyses were conducted on the same set of genomes downloaded from NCBI[1] on 4th December 2023 following the procedure below:

1. The *assembly_summary.txt* tabular file listing all the available plant genomes was downloaded from the following URL: https://ftp.ncbi.nlm.nih.gov/genomes/genbank/plant/assembly_summary.txt
2. The genome assemblies of the species of interest were selected
3. The genome assemblies were restricted to genome_rep equal to "Full", and assembly_level equal to "Complete" or "Chromosome" to exclude incomplete or fragmented genomes

A total of 100 genome assemblies were selected, covering several cultivars for each species. The number of genome assemblies for each species is reported in Table 1. The full list of assemblies with additional details is shown in Annex 1.

Table 1. Number of genome assemblies selected for each species of interest.

Species (common name)	Species (scientific name)	Number of genome assemblies
Maize	*Zea mays*	66
Soybean	*Glycine max*	19
Cotton	*Gossypium hirsutum*	9
Rapeseed	*Brassica napus*	6

The PCR simulations were performed *in silico* using the e-PCR software (version 2.3.12), developed by NCBI (Rotmistrovky et al., 2004) and the primers of the methods listed in Table 2. The software was run with the following parameters: -n 2 (maximum number of mismatches in each primer), -g 2 (maximum number of gaps in each primer), -f 3 (number of discontiguous words), -t 4 (tabular output with alignments), -m 1000 (margin), -d 20-1000 (amplicon size range). The probes were aligned to the predicted amplicons using the EMBOSS Matcher tool[2] (version 6.6.0; Huang et al., 1991). The latter was run with the following parameters: -auto (turn off prompts), -stdout (write first file to standard output), -aformat 3 srspair (srs pair format). The probe alignment was checked

[1] https://www.ncbi.nlm.nih.gov/

[2] https://emboss.sourceforge.net/apps/release/6.6/emboss/apps/matcher.html

in both orientations (parameter -sreverse1), and only alignments with at least 80% identity were accepted. Finally, the predictions were manually inspected.

Table 2. PCR amplification methods considered in this study.

GMOMETHODS ID ([1])	**Oligos**	**Target gene**	**Target species**
QT-TAX-ZM-001	Fwd: CGTCGTTTCCCATCTCTTCCTCC	*adh1*	Maize
	Rev: CCACTCCGAGACCCTCAGTC		
	Probe: FAM-AATCAGGGCTCATTTTCTCGCTCCTCA-TAMRA		
QT-TAX-ZM-002	Fwd: TTGGACTAGAAATCTCGTGCTGA	*hmg*	Maize
	Rev: GCTACATAGGGAGCCTTGTCCT		
	Probe: FAM-CAATCCACACAAACGCACGCGTA-TAMRA		
QT-TAX-GM-003	Fwd: CTTTCTCGCACCAATTGACA	*le1*	Soybean
	Rev: TCAAACTCAACAGCGACGAC		
	Probe: JOE-CCACAAACACATGCAGGTTATCTTGG-BHQ1		
QT-TAX-GM-002	Fwd: CCAGCTTCGCCGCTTCCTTC	*le1*	Soybean
	Rev: GAAGGCAAGCCCATCTGCAAGCC		
	Probe: FAM-CTTCACCTTCTATGCCCCTGACAC-TAMRA		
QT-TAX-GH-021	Fwd: AGTTTGTAGGTTTTGATGTTACATTGAG	*sah7*	Cotton
	Rev: GCATCTTTGAACCGCCTACTG		
	Probe: FAM-AAACATAAAATAATGGGAACAACCATGACATGT-TAMRA		
QT-TAX-GH-019	Fwd: CACATGACTTAGCCCATCTTTGC	*adhC*	Cotton
	Rev: CCCACCCTTTTTTGGTTTAGC		
	Probe: VIC-TGCAGGTTTTGGTGCCACTGTGAATG-TAMRA		
QT-TAX-BN-001	Fwd: ACAGATGAAGTTCGGGACGAGTAC	*fatA(A)*	Rapeseed
	Rev: CAGGTTGAGATCCACATGCTTAAATAT		
	Probe: FAM-AAGAAGAATCATCATGCTTC-MGBNFQ		

([1]) In the GMOMETHODS Database[3], detection methods with identical oligos but different reporter dyes are indicated with a different ID. However, since the bioinformatics analyses described in this report do not take into account the reporter dyes, the same conclusions remain valid for all methods sharing the same primers and probe, i.e.: QT-TAX-ZM-003 and QT-TAX-ZM-014 for QT-TAX-ZM-001; QT-TAX-ZM-005 and QT-TAX-ZM-008 for QT-TAX-ZM-002; QT-TAX-GM-020 for QT-TAX-GM-003; QT-TAX-GM-005 for QT-TAX-GM-002; QT-TAX-GH-018 and QT-TAX-GH-020 for QT-TAX-GH-019.

[3] https://gmo-crl.jrc.ec.europa.eu/gmomethods/

The similarity searches in the plant genomes were performed using the blastn[4] software (version 2.12.0; Altschul et al., 1990) with the following parameters: -dust no (low complexity filter off) -soft_masking no (soft masking off). The results were manually inspected and filtered to keep only the alignments covering at least 80% of the query sequence. Table 3 lists the Genbank record IDs that were used as queries to perform the similarity searches for each reference gene. Since a complete record was not available for the *sah7* gene in *Gossypium hirsutum*, the homologous gene in *Gossypium barbadense* (a different cotton species accounting for about 5% of the total cotton production) was used instead (i.e. Genbank ID AY117068.1). The primer alignments for the reference genomes are reported in Annex 2.

Table 3. Genbank records used to perform similarity searches in the plant genomes.

Target gene	**Target species**	**Genbank record ID**	**Sequence length (bp)**
adh1	Maize	X04050.1	5,249
hmg	Maize	AJ131373.1	2,133
le1	Soybean	K00821.1	2,152
sah7	Cotton	AY117068.1	837
adhC	Cotton	AF036569.1	1,651
fatA(A)	Rapeseed	X87842.1	3,007

The results of the analysis are reported in the following sections according to species.

[4] https://blast.ncbi.nlm.nih.gov/

2 Maize

The maize genome is about 2.4 Gbp in size. Two reference genes were analysed for this species: *adh1* and *hmg*, respectively targeted by PCR methods QT-TAX-ZM-001 and QT-TAX-ZM-002. A total of 66 genome assemblies were analysed, including the RefSeq representative genome GCA_902167145.1 (cultivar B73).

2.1 Analysis of the *adh1* reference gene

According to the similarity searches, the *adh1* gene was found in all genome assemblies, therefore it is likely present in most if not all maize cultivars. However, while most genomes have only one gene copy, genome assembly GCA_021307875.1 (cultivar A188) appears to have three gene copies. Considering GCA_902167145.1 as the reference assembly, the *adh1* gene is located on chromosome 1 (Genbank ID LR618874.1, coordinates 280,590,357-280,594,023 reverse strand). It can be found on the Gramene website with ID Zm00001eb056510. Interestingly, GCA_902167065.1 (cultivar Il14H) and GCA_902167135.1 (cultivar CML69) have rearrangements of the gene (partial translocation) or an error in the sequence assembly.

The PCR simulations confirmed that the *adh1* gene can be amplified by the QT-TAX-ZM-001 primers in all genome assemblies. A perfect alignment with primers and probe was obtained for all genome assemblies, including GCA_902167065.1 (cultivar Il14H) and GCA_902167135.1 (cultivar CML69), where rearrangements of the *adh1* gene were noticed. The predicted amplicon size was 135 bp in all genome assemblies.

A peculiar case is genome assembly GCA_021307875.1 (cultivar A188), where three different amplifications are predicted, which is in agreement with the fact that three copies of this gene are present in this genome. The genome of A188 was obtained by single molecule sequencing (Pacific Biosciences technology) and it can be considered of good quality. Cultivar A188 is an inbred maize line that shows great phenotypic difference from other inbred lines, such as in plant height, days to pollination, oil concentration, protein concentration or starch concentration (Ge et al., 2022). Its agronomical traits are poor, as it is shorter and has lower yield compared to other varieties, but its efficient regeneration from stem cells make it an interesting model for maize research. Moreover, line Hi-II, which is commonly used for maize transformation, was derived from the A188 cultivar[5].

2.2 Analysis of the *hmg* reference gene

According to the similarity searches, the *hmg* gene was found in all genome assemblies, therefore it is likely present in most if not all maize cultivars and always in a single copy. Considering GCA_902167145.1 as the reference assembly, the *hmg* gene is located on chromosome 5 (Genbank ID LR618878.1, coordinates 65,694,795-65,696,975 reverse strand). It can be found on the Gramene[6] website with ID Zm00001d014858.

The PCR simulations confirmed that the *hmg* gene can be amplified by the QT-TAX-ZM-002 primers in all genome assemblies. In 50 assemblies out of 66, the alignment with the primers was perfect;

[5] https://botany.one/2021/10/new-maize-reference-genome-may-help-grow-genetically-modified-lines-of-corn-more-easily/

[6] https://www.gramene.org/

in 15 assemblies an alignment gap was found in one of the primers, whereas in one assembly a mismatch was found in one of the primers. In all genome assemblies, the probe was predicted to anneal to the amplicon, although one mismatch was found in GCA_902167105.1 (cultivar KI3). The amplicon size was 79 bp.

3 Soybean

The soybean genome is about 1.1 Gbp in size. The *le1* reference gene (lectin) was analysed for this species, targeted by two different PCR methods (QT-TAX-GM-002 and QT-TAX-GM-003). A total of 19 genome assemblies were analysed, including the RefSeq representative genome GCA_000004515.5 (cultivar Williams 82).

3.1 Analysis of the *le1* reference gene

According to the similarity searches, the *le1* gene was found in all genome assemblies, suggesting that it may be present in most if not all soybean cultivars. Moreover, only one gene copy is present in each genome. Considering GCA_000004515.5 as the reference assembly, the *le1* gene is located on chromosome 2 (Genbank ID CM000835.4, coordinates 1,123,507-1,125,658 forward strand). It can be found on the Gramene website with ID GLYMA_02G012600.

The PCR simulations confirmed that the *le1* gene can be amplified by the QT-TAX-GM-002 and QT-TAX-GM-003 methods in all genome assemblies. The alignment with primers and probe was perfect in all genome assemblies, with an amplicon size of 74 bp and 102 bp respectively for the QT-TAX-GM-002 and QT-TAX-GM-003 methods.

4 Cotton

The cotton genome is about 2.3 Gbp in size. Two reference genes were analysed for this species: *sah7* and *adhC*, respectively targeted by PCR methods QT-TAX-GH-021 and QT-TAX-GH-019. A total of 9 genome assemblies were analysed, including the RefSeq representative genome GCA_007990345.1 (unknown cultivar).

4.1 Analysis of the *sah7* reference gene

According to the similarity searches, the *sah7* gene was found in all genome assemblies, therefore it is likely that most cotton cultivars contain this gene. Moreover, two gene copies are present in each genome. Considering GCA_007990345.1 as the reference assembly, the *sah7* gene was found in two genome locations:

- Chromosome A2 (Genbank ID CM017663.1, coordinates 15,958,024-15,959,338 forward strand). Found on Phytozome[7] with ID Gohir.A02G075700
- Chromosome D2 (Genbank ID CM017676.1, coordinates 14,221,520-14,222,762 forward strand). Found on Phytozome with ID Gohir.D02G082800.

The PCR simulations confirmed that the *sah7* gene can be amplified by the QT-TAX-GH-021 primers in all genome assemblies, with a perfect alignment of the primers. However, since the gene is duplicated in two different genomic locations, two PCR products are predicted (amplicon size of 115 and 123 bp). In both cases, the probe aligns perfectly to the amplicon.

4.2 Analysis of the *adhC* reference gene

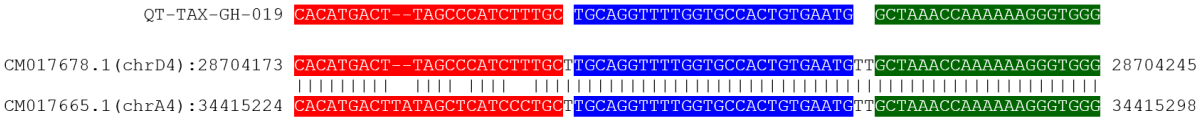
According to the similarity searches, the *adhC* gene was found in all genome assemblies, therefore it is likely that most cotton cultivars contain this gene. Moreover, two gene copies are present in each genome. Considering GCA_007990345.1 as the reference assembly, the *adhC* gene was found in the following two genome locations:

- Chromosome D4 (Genbank ID CM017678.1, coordinates 28,702,584-28,706,281 forward strand). Found on Phytozome with ID Gohir.D04G092650.
- Chromosome A4 (Genbank ID CM017665.1, coordinates 34,413,696-34,416,940 reverse strand). Found on Phytozome with ID Gohir.A04G057600.

The PCR simulations confirmed that the *adhC* gene can be amplified by the QT-TAX-GH-019 primers in all genome assemblies, with a perfect alignment of primers and probe. Interestingly, although the gene appears to be duplicated in two different chromosomes, only the first copy of the gene in chromosome D4 could be amplified, with an amplicon size of 73 bp. As shown in Figure 1, the second copy of the gene has three mismatches and one insertion of two base pairs in the region that should be annealed by the forward primer, and should therefore not be amplified.

[7] https://phytozome-next.jgi.doe.gov/

Figure 1. Alignment of the predicted PCR amplicons on chromosome D4 (Genbank ID CM017678.1) and on chromosome A4 (Genbank ID CM017665.1). The regions bound by the forward primer, the probe and the reverse primer of QT-TAX-GH-019 are highlighted in red, blue and green, respectively.



5 Rapeseed

The rapeseed genome is about 1.0 Gbp in size. The *fatA(A)* reference gene was analysed for this species, targeted by PCR method QT-TAX-BN-001. A total of 6 genome assemblies were analysed, including the RefSeq representative genome GCA_020379485.1 (cultivar Da-Ae).

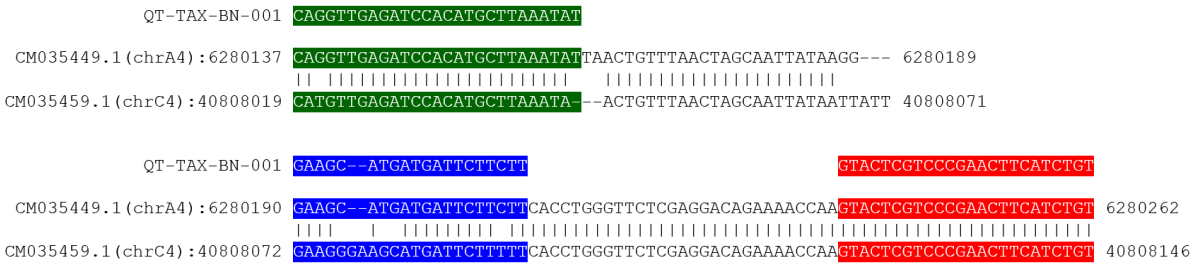
5.1 Analysis of the *fatA(A)* reference gene

According to the similarity searches, the *fatA(A)* gene was found in all genome assemblies, therefore it is likely that most rapeseed cultivars contain this gene. However, while most genomes have only one gene copy, genome assembly GCA_000686985.2 (cultivar ZS11) appears to have two gene copies. Considering GCA_020379485.1 as the reference assembly, the *fatA(A)* gene is located on chromosome C4 (Genbank ID CM035459.1, coordinates 40,807,478-40,809,230 reverse strand).

The PCR simulations confirmed that the *fatA(A)* gene can be amplified by the QT-TAX-BN-001 primers, in all genome assemblies. Differently from what was expected, two PCR products were predicted for each genome assembly, with the exception of GCA_000686985.2 (cultivar ZS11), where four amplifications were predicted. We further investigated the results of the Blast alignment by manual inspection and found another copy of gene *fatA(A)* on chromosome A4 (Genbank ID CM035449.1, coordinates 6,279,586-6,281,411 reverse strand). This second copy was initially missed due to a divergent segment in the central part of the gene, compared to the query sequence; because of this difference, the BLAST software detected two distinct alignments for two gene segments, instead of one contiguous alignment for the entire gene.

As shown in Figure 2, the copy located on chromosome A4 is the one providing the best amplification (perfect alignment with primers and probe, amplicon size 126 bp), while the copy located on chromosome C4 has two mismatches in the region bound by the reverse primer, as well as four mismatches and an insertion of 2 bp in the region bound by the probe, and is therefore unlikely to be detected with the QT-TAX-BN-001 method. This result is in agreement with the data reported in Henderson et al., 2016, where the method QT-TAX-BN-001 was first described. The authors developed this method to target specifically canola-quality *Brassica* species, i.e. the diploid *B. rapa* (turnip rape, AA) and the allotetraploids *B. napus* (rape, AACC) and *B. juncea* (brown mustard, AABB). These species share the sub-genome A, which is not present in the other *Brassica* species, i.e. the allotetraploid *B. carinata* (BBCC) and the diploids *B. nigra* (black mustard, BB) and *B. oleracea* (black mustard, CC). As our results show, QT-TAX-BN-001 can only amplify the copy of *fatA* gene from the A sub-genome, thus confirming the predictions of Henderson and colleagues. The same conclusion is reported in Jacchia et al., 2018, where the authors measured the λ values with ddPCR for several reference gene assays.

Figure 2. Alignment of the predicted PCR amplicons on chromosome A4 (Genbank ID CM035449.1) and on chromosome C4 (Genbank ID CM035459.1). The regions bound by the forward primer, the probe and the reverse primer of QT-TAX-BN-001 are highlighted in red, blue and green, respectively.



6 Conclusions

The similarity searches and PCR simulations performed on the genomes of the species of interest allowed us to draw the following conclusions.

Maize:

- One copy of the *adh1* gene was found in 65 out of 66 genome assemblies. In one genome assembly (cultivar A188), three copies of the gene were found. Method QT-TAX-ZM-001 should detect this gene in all the assemblies with a perfect alignment of primers and probe. Three PCR amplifications were predicted for the genome assembly of cultivar A188.
- One copy of the *hmg* gene was found in all genome assemblies. Method QT-TAX-ZM-002 should detect this gene in all the assemblies. In 50 out 66 assemblies, the alignment of the primers is perfect, while in the remaining assemblies one gap or one mismatch was found for one of the primers, although this should not prevent the PCR amplification. In one assembly, a mismatch was found in the alignment of the probe, while in the remaining assemblies the alignment was perfect.

Soybean:

- One copy of the *le1* gene was found in all genome assemblies. Methods QT-TAX-GM-002 and QT-TAX-GM-003 should detect this gene in all the assemblies. In all cases, the alignment of primers and probe is perfect.

Cotton:

- Two copies of the *sah7* gene were found in all genome assemblies. Both gene copies can be amplified by method QT-TAX-GH-021, with a perfect alignment of primers and probe.
- Two copies of the *adhC* gene were found in all genome assemblies. However, only one copy of this gene can be amplified by method QT-TAX-GH-019, with a perfect alignment of primers and probe.

Rapeseed:

- Two copies of the *fatA(A)* gene were found in 5 out of 6 genome assemblies. In one genome assembly (cultivar ZS11), four gene copies were found. However, only one copy of this gene can be amplified by method QT-TAX-BN-001 (two copies for cultivar ZS11), with a perfect alignment of primers and probe.

References

Altschul, S.F. et al., 'Basic local alignment search tool', *Journal of Molecular Biology*, Vol. 215, Issue 3, 1990, pp. 403-410.

Ge, F. et al., 'Genome assembly of the maize inbred line A188 provides a new reference genome for functional genomics', *The Crop Journal*, Vol. 10, Issue 1, 2022, pp. 47-55.

Henderson, N. et al., 'PCR-based detection and quantification of a transgenic glyphosate-tolerant canola using a novel reference gene system', *Food Analytical Methods*, Vol. 9, 2015, pp. 353–361.

Huang, X. and Miller, W., 'A time-efficient, linear-space local similarity algorithm', *Advances in Applied Mathematics*, Vol. 12, Issue 13, 1991, pp. 337-357.

Jacchia, S. et al., 'Identification of single target taxon-specific reference assays for the most commonly genetically transformed crops using digital droplet PCR', *Food Control*, Vol. 93, 2018, pp. 191-200.

Rotmistrovsky, K. et al., 'A web server for performing electronic PCR', *Nucleic Acids Research*, Vol. 32, Issue suppl_2, 2004, pp. W108–W112.

List of abbreviations and definitions

Abbreviations	Definitions
Gbp	Giga base pairs (1 Gbp = 1,000,000,000 bp)
EURL GMFF	European Union Reference Laboratory for Genetically Modified Food and Feed
NCBI	National Center for Biotechnology Information
PCR	Polymerase Chain Reaction

List of figures

List of tables

Annexes

Annex 1. Genome assemblies used for the PCR simulations and the similarity searches

assembly accession	species	cultivar	assembly level
GCA_000987745.1	*Gossypium hirsutum*	TM-1	Chromosome
GCA_006980745.1	*Gossypium hirsutum*	TM1	Chromosome
GCA_006980775.1	*Gossypium hirsutum*	ZM24	Chromosome
GCA_007990345.1 *	*Gossypium hirsutum*	not available	Chromosome
GCA_018997965.1	*Gossypium hirsutum*	NDM8	Chromosome
GCA_021461695.1	*Gossypium hirsutum*	not available	Chromosome
GCA_021461685.1	*Gossypium hirsutum*	BARBREN-713	Chromosome
GCA_024704785.1	*Gossypium hirsutum*	JBM	Chromosome
GCA_024600755.1	*Gossypium hirsutum*	Zhongzhimian No.2	Chromosome
GCA_000686985.2	*Brassica napus*	ZS11	Chromosome
GCA_020379485.1 *	*Brassica napus*	Da-Ae	Chromosome
GCA_026770265.1	*Brassica napus*	Zy821	Chromosome
GCA_026770255.1	*Brassica napus*	GH06	Chromosome
GCA_905183035.1	*Brassica napus*	not available	Chromosome
GCA_033556945.1	*Brassica napus*	NTS57	Chromosome
GCA_002905335.2	*Glycine max*	Lee	Chromosome
GCA_003349995.2	*Glycine max*	Zhonghuang 13	Chromosome
GCA_014282185.1	*Glycine max*	Williams 82	Chromosome
GCA_014282065.1	*Glycine max*	Hefeng 25	Chromosome
GCA_014282095.1	*Glycine max*	Jinyuan	Chromosome
GCA_014282075.1	*Glycine max*	Zhonghuang 35	Chromosome
GCA_014282145.1	*Glycine max*	Huaxia 3	Chromosome
GCA_014282035.1	*Glycine max*	Wenfeng 7	Chromosome
GCA_014282085.1	*Glycine max*	Zhonghuang 13	Chromosome
GCA_015227745.1	*Glycine max*	Tianlong1	Chromosome
GCA_000004515.5 *	*Glycine max*	Williams 82	Chromosome
GCA_019321705.1	*Glycine max*	PI594527	Chromosome
GCA_012273815.2	*Glycine max*	EMBRAPA BRS 537	Chromosome
GCA_020497155.1	*Glycine max*	Hwangkeum	Chromosome
GCA_021733175.1	*Glycine max*	Jidou 17	Chromosome
GCA_022114995.1	*Glycine max*	ZN6	Chromosome
GCA_029216145.1	*Glycine max*	ENREI	Chromosome
GCA_030864155.1	*Glycine max*	Williams 82	Complete Genome
GCA_033623075.1	*Glycine max*	Jack	Complete Genome
GCA_000005005.6	*Zea mays*	B73	Chromosome
GCA_003185045.1	*Zea mays*	inbred line Mo17	Chromosome
GCA_003709335.1	*Zea mays*	DK105	Chromosome
GCA_003704525.1	*Zea mays*	PE0075	Chromosome
GCA_009176585.1	*Zea mays*	K0326Y	Chromosome
GCA_902166995.1	*Zea mays*	M162W	Chromosome

GCA_902167005.1	*Zea mays*	NC350	Chromosome
GCA_902167045.1	*Zea mays*	Ky21	Chromosome
GCA_902167075.1	*Zea mays*	B97	Chromosome
GCA_902167135.1	*Zea mays*	CML69	Chromosome
GCA_902167145.1 *	*Zea mays*	B73	Chromosome
GCA_902167095.1	*Zea mays*	CML322	Chromosome
GCA_902167185.1	*Zea mays*	TZI8	Chromosome
GCA_902166985.1	*Zea mays*	CML277	Chromosome
GCA_902166975.1	*Zea mays*	CML247	Chromosome
GCA_902167035.1	*Zea mays*	NC358	Chromosome
GCA_902167105.1	*Zea mays*	KI3	Chromosome
GCA_902167155.1	*Zea mays*	CML228	Chromosome
GCA_902167375.1	*Zea mays*	MS71	Chromosome
GCA_902166965.1	*Zea mays*	P39	Chromosome
GCA_902167015.1	*Zea mays*	OH43	Chromosome
GCA_902167065.1	*Zea mays*	Il14H	Chromosome
GCA_902167085.1	*Zea mays*	Hp301	Chromosome
GCA_902167175.1	*Zea mays*	M37W	Chromosome
GCA_902167205.1	*Zea mays*	Tx303	Chromosome
GCA_902166955.1	*Zea mays*	OH7B	Chromosome
GCA_902167025.1	*Zea mays*	CML333	Chromosome
GCA_902167055.1	*Zea mays*	CML103	Chromosome
GCA_902167115.1	*Zea mays*	KI11	Chromosome
GCA_902167165.1	*Zea mays*	Mo18W	Chromosome
GCA_902373975.1	*Zea mays*	CML52	Chromosome
GCA_902714155.1	*Zea mays*	B73 Ab10	Chromosome
GCA_014529475.1	*Zea mays*	not available	Chromosome
GCA_905067065.1	*Zea mays*	not available	Chromosome
GCA_016432965.1	*Zea mays*	Ia453-sh2	Chromosome
GCA_019096015.1	*Zea mays*	PHB47	Chromosome
GCA_019095995.1	*Zea mays*	PHJ40	Chromosome
GCA_019095975.1	*Zea mays*	LH145	Chromosome
GCA_019096025.1	*Zea mays*	NKH8431	Chromosome
GCA_019095955.1	*Zea mays*	B84	Chromosome
GCA_910593975.1	*Zea mays*	B104	Chromosome
GCA_021307875.1	*Zea mays*	A188	Chromosome
GCA_022117705.1	*Zea mays*	Mo17-2021	Complete Genome
GCA_024505845.1	*Zea mays*	inbred line Dan340	Chromosome
GCA_024718635.1	*Zea mays*	not available	Chromosome
GCA_026226675.1	*Zea mays*	S37	Chromosome
GCA_026226635.1	*Zea mays*	PH207	Chromosome
GCA_026228945.1	*Zea mays*	huangzaosi	Chromosome
GCA_026226085.1	*Zea mays*	Zheng58	Chromosome
GCA_026226705.1	*Zea mays*	Jing724	Chromosome
GCA_026229665.1	*Zea mays*	Chang7-2	Chromosome
GCA_026226095.1	*Zea mays*	Xu178	Chromosome

GCA_026226105.1	*Zea mays*	Ye478	Chromosome
GCA_026226695.1	*Zea mays*	Jing92	Chromosome
GCA_026226685.1	*Zea mays*	OH43	Chromosome
GCA_026229635.1	*Zea mays*	Dan340	Chromosome
GCA_026229685.1	*Zea mays*	A632	Chromosome
GCA_029775835.1	*Zea mays*	LT2357	Chromosome
GCA_025175245.1	*Zea mays* subsp. *parviglumis*	Ames21814	Chromosome
GCA_001644905.2	*Zea mays* subsp. *mays*	W22	Chromosome
GCA_001984235.2	*Zea mays* subsp. *mays*	EP1	Chromosome
GCA_001990705.1	*Zea mays* subsp. *mays*	F7	Chromosome
GCA_002237485.1	*Zea mays* subsp. *mays*	PH207	Chromosome
GCA_002682915.2	*Zea mays* subsp. *mays*	CML247	Chromosome
GCA_027171705.3	*Zea mays* subsp. *mays*	CIMBL55	Chromosome
GCA_030347615.1	*Zea mays* subsp. *mays*	TPD BC8S3	Chromosome

* the genome assemblies with an asterisk have been chosen as RefSeq representative genomes.

Annex 2. Primer alignments in the reference genomes

Maize, method QT-TAX-ZM-001 (gene *adh1*), genome assembly GCA_902167145.1

```
########################################################################
#   STS maize_adh1   CCACTCCGAGACCCTCAGTC...92...GGAGGAAGAGATGGGAAACGACG
#                    ||||||||||||||||||||   92   |||||||||||||||||||||||
#   Seq LR618874.1 atCCACTCCGAGACCCTCAGTC...92...GGAGGAAGAGATGGGAAACGACGgg
LR618874.1      maize_adh1      -       280593784       280593918       135/135-135     0       0
```

Maize, method QT-TAX-ZM-002 (gene *hmg*), genome assembly GCA_902167145.1

```
########################################################################
#   STS  maize_hmg   GCTACATAGGGAGCCTTGTCCT...34...TCAGCACGAGATTTCTAGTCCAA
#                    ||||||||||||||||||||   34   |||||||||||||||||||||||
#   Seq LR618878.1 tgGCTACATAGGGAGCCTTGTCCT...34...TCAGCACGAGATTTCTAGTCCAAtc
LR618878.1      maize_hmg       -       65695943        65696021        79/79-79        0       0
```

Soybean, method QT-TAX-GM-002 (gene *le1*), genome assembly GCA_000004515.5

```
########################################################################
#   STS soybean_lectin_2   CCAGCTTCGCCGCTTCCTTC...31...GGCTTGCAGATGGGCTTGCCTTC
#                          ||||||||||||||||||||   31   |||||||||||||||||||||||
#   Seq       CM000835.4 tgCCAGCTTCGCCGCTTCCTTC...31...GGCTTGCAGATGGGCTTGCCTTCtt
CM000835.4      soybean_lectin_2        +       1124938 1125011 74/74-74        0       0
```

Soybean, method QT-TAX-GM-003 (gene *le1*), genome assembly GCA_000004515.5

```
########################################################################
#   STS soybean_lectin_1   CTTTCTCGCACCAATTGACA...62...GTCGTCGCTGTTGAGTTTGA
#                          ||||||||||||||||||||   62   |||||||||||||||||||||
#   Seq       CM000835.4 ttCTTTCTCGCACCAATTGACA...62...GTCGTCGCTGTTGAGTTTGAca
CM000835.4      soybean_lectin_1        +       1125011 1125112 102/102-102     0       0
```

Cotton, method QT-TAX-GH-021 (gene *sah7*), genome assembly GCA_007990345.1

```
########################################################################
#   STS cotton_sah7   AGTTTGTAGGTTTTGATGTTACATTGAG...66...CAGTAGGCGGTTCAAAGATGC
#                     ||||||||||||||||||||||||||||   66   |||||||||||||||||||||
#   Seq  CM017663.1 gtAGTTTGTAGGTTTTGATGTTACATTGAG...66...CAGTAGGCGGTTCAAAGATGCaa
CM017663.1      cotton_sah7     +       15958442        15958556        115/115-115     0       0

########################################################################
#   STS cotton_sah7   AGTTTGTAGGTTTTGATGTTACATTGAG...74...CAGTAGGCGGTTCAAAGATGC
#                     ||||||||||||||||||||||||||||   74   |||||||||||||||||||||
#   Seq  CM017676.1 gtAGTTTGTAGGTTTTGATGTTACATTGAG...74...CAGTAGGCGGTTCAAAGATGCaa
```

```
CM017676.1      cotton_sah7     +       14221954        14222076        123/115-115     0       0
```

Cotton, method QT-TAX-GH-019 (gene *sah7*), genome assembly GCA_007990345.1

```
#######################################################################
#   STS cotton_adhc   CACATGACTTAGCCCATCTTTGC...29...GCTAAACCAAAAAAGGGTGGG
#                     ||||||||||||||||||||||   29   ||||||||||||||||||||
#   Seq  CM017678.1 aaCACATGACTTAGCCCATCTTTGC...29...GCTAAACCAAAAAAGGGTGGGtt
CM017678.1      cotton_adhc     +       28704173        28704245        73/73-73        0       0
```

Rapeseed, method QT-TAX-BN-001 (gene *fatA(A)*), genome assembly GCA_020379485.1

```
#######################################################################
#   STS rapeseed_fata   CAGGTTGAGATCCACATGCTTAAATAT...75...GTACTCGTCCCGAACTTCATCTGT
#                       ||||||||||||||||||||||||||   75   ||||||||||||||||||||||||
#   Seq    CM035449.1 aaCAGGTTGAGATCCACATGCTTAAATAT...75...GTACTCGTCCCGAACTTCATCTGTaa
CM035449.1      rapeseed_fata   -       6280137 6280262 126/126-126     0       0
```

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

EU Science Hub
joint-research-centre.ec.europa.eu