

Selecting TILLING mutants

Knock-out mutations are available in most wheat genes providing an invaluable resource for characterising gene function. Targeting Induced Local Lesions In Genomes (TILLING) is a method of inducing and identifying mutations in specific genes. This section will describe how to use the *in silico* wheat TILLING resource. This resource consists of TILLING populations developed in tetraploid durum wheat cv 'Kronos' and hexaploid bread wheat cv 'Cadenza' as part of a joint project between the University of California Davis, Rothamsted Research, The Earlham Institute, and John Innes Centre ([Krasileva et al., 2017](#)). The exome sequences of 1,535 Kronos (tetraploid) and 1,200 Cadenza (hexaploid) mutants have been re-sequenced using Illumina next-generation sequencing. In the original publication ([Krasileva et al., 2017](#)) the raw data was aligned to the best available genome at the time (CSS released in 2014). This dataset is still available as legacy (described below in section e). However, given the latest genome release in 2018 (RefSeq) the raw data has been realigned to the IWGSC RefSeqv1.0 genome assembly, mutations identified, and their effects predicted based on the IWGSC RefSeqv1.1 annotation. For more information on the latest mutation calls see [Ensembl Plants](#).

The following document will explain how to select TILLING mutants for your gene(s) of interest. To begin, you will need the IWGSC RefSeq v1.1 gene model identifier for your gene(s) (see ["Finding wheat orthologs"](#) for more information).

a) Using Ensembl Plants to find TILLING mutations

1. Go to www.plants.ensembl.org.
2. Search for your gene on Ensembl Plants using the IWGSC RefSeqv1.1 gene ID (see ["Gene models"](#) and ["Using Ensembl Plants"](#) for details).
3. Click the link for your desired gene and make sure you are on the Gene tab view (Figure 1a).

4. TILLING mutants can be accessed using either the 'variant table' or 'variant image' links on the left-hand side of the page (Figure 1b, c).

The screenshot displays the Ensembl Plants interface for the gene **TraesCS5B02G236400** in *Triticum aestivum*. The top navigation bar includes links for HMMER, BLAST, BioMart, Tools, Downloads, Documentation, and Website help. The left sidebar lists various gene-based displays, with 'Genetic Variation' expanded to show 'Variant table' and 'Variant image'. The main content area shows the gene's location on Chromosome 5B (417,437,998-417,443,899) and a summary of its properties, including 65 orthologues and 3 paralogues. A genomic track at the bottom shows the gene structure and associated transcripts, with a 'Reverse strand' indicator.

Figure 1: Ensembl Plants gene tab

- The variant image is useful for looking at the position of all available mutations in your gene of interest. Mutations are coloured according to their predicted effect on the gene model. You can find further information about individual mutations by clicking on them and following the “more about...” link (Figure 2a). This link will include Zygosity and pre-designed KASP assays.

Gene: TraesCS5B02G236400

Location [Chromosome 5B: 417,437,998-417,443,899 forward strand.](#)
 About this gene [This gene has 1 transcript \(splice variant\), 65 orthologues and 3 paralogues.](#)
 Transcripts [Show transcript table](#)

Variant image

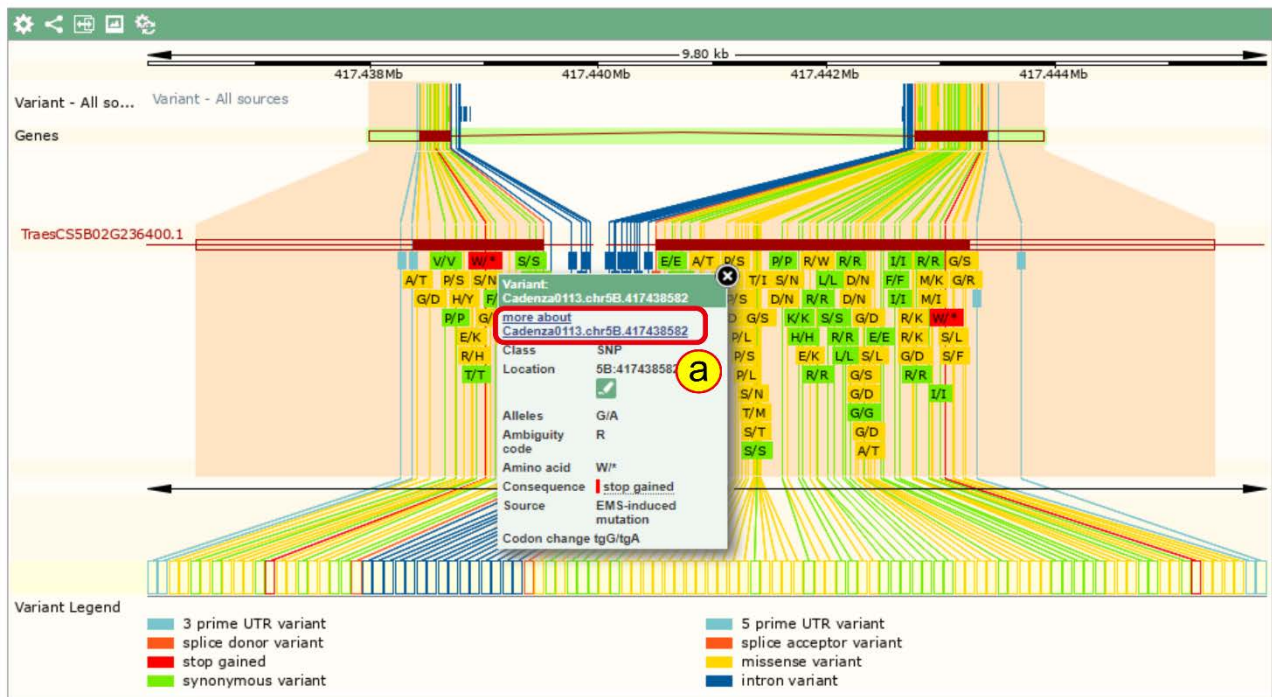


Figure 2: Ensembl Plants variant image


- The variant table contains the same information but in table format. By default, the variant table will include details about variation in your gene of interest from multiple different sources, but you can display TILLING mutations only by selecting “EMS-induced mutation” in the source filter at the top of the table (Figure 3a).

Updated April 2019

Gene: **TraesCS5B02G236400**

Location [Chromosome 5B: 417,437,998-417,443,899 forward strand.](#)
About this gene This gene has 1 transcript ([splice variant](#)), [65 orthologues](#) and [3 paralogues](#).
Transcripts [Show transcript table](#)

Variant table

Variant table 

This table shows known variants for this gene. Use the 'Consequence Type' filter to view a subset of these.

a

Filter

Show/hide columns

Variant ID	Chr: bp	Alleles	Class	Source	Conseq. Type	AA	AA coord	SIFT	Transcript
Kronos3039.chr5B.417438412	5B:417438412	G/A	SNP	EMS-induced mutation	5 prime UTR variant	-	-	-	TraesCS5B02G236400.1
Cadenza1722.chr5B.417438435	5B:417438435	G/A	SNP	EMS-induced mutation	5 prime UTR variant	-	-	-	TraesCS5B02G236400.1
Cadenza2069.chr5B.417438445	5B:417438445	G/A	SNP	EMS-induced mutation	missense variant	A/T	4	0.1	TraesCS5B02G236400.1
Cadenza1443.chr5B.417438470	5B:417438470	G/A	SNP	EMS-induced mutation	missense variant	G/D	12	0.24	TraesCS5B02G236400.1
Cadenza1474.chr5B.417438501	5B:417438501	C/T	SNP	EMS-induced mutation	synonymous variant	V	22	-	TraesCS5B02G236400.1
Cadenza1778.chr5B.417438520	5B:417438520	C/T	SNP	EMS-induced	missense variant	P/S	29	0	TraesCS5B02G236400.1

Figure 3: Ensembl Plants variant table

Understanding the variant table

Each row in the variant table corresponds to a single mutation in a single mutant line of the TILLING population with the columns containing information about the mutation:

Variant ID

The ID of the mutant line in which the mutation was found and its location in the genome, e.g. “Kronos3039.chr5B.417438412” indicates that the mutation is in line Kronos3039 located on chromosome 5B at position 417,438,412 (in base pairs) of the IWGSC Refseqv1.0 assembly. The line name indicates which population the mutation is in, either Kronos (tetraploid) or Cadenza (hexaploid) and is required for ordering seeds. In this case, Kronos3039 is the mutant ID that you would need to order seeds (see section c below).

Chr:bp

The chromosome and base pair position of the mutation in the IWGSC RefSeqv1.0 assembly.

Alleles

The wild type/reference (i.e. non-mutagenised Kronos or Cadenza depending on the population) and mutant base (i.e. mutation as a result of EMS mutagenesis) at this position (WT/mutant). For EMS-induced mutations, over 99% correspond to G/A and C/T mutations.

Class

The type of variation. For the TILLING mutants this should always be “SNP” (single nucleotide polymorphism).

Source

The source of variation. For the TILLING mutants this will always be “EMS-induced mutation”.

Updated April 2019

Conseq. Type

Consequence of the mutation on the IWGSC RefSeq v1.1 transcript as predicted by the Variant Effect Predictor (VEP) tool from Ensembl. A full detail of the different effects predicted is described [here](#).

AA

The amino acid affected in the IWGSC RefSeq v1.1 gene model based on the VEP output, if applicable. In the case of non-synonymous mutations, the display will read WT/mutant.

AA co-ord

The amino acid affected in the IWGSC RefSeq v1.1 gene model based on the VEP output, if applicable.

SIFT

Sorting Intolerant From Tolerant (SIFT) probability score that the amino acid change is tolerated. Scores <0.05 are considered deleterious whereas all others are considered tolerated. SIFT was implemented through Ensembl VEP as described in [Krasileva et al., \(2017\)](#).

Transcript

The RefSeqv1.1 transcript affected by the mutation based on the VEP output.

b)Selecting a mutant line

When selecting a mutant line, there are certain things you should take into consideration. There will be many lines that have point mutations in your gene of interest and so it may not be immediately obvious which line is the best to take forward. Every case will be slightly different and depend on the aim of the study but here we will describe some common usages.

Type of mutation

The 'Conseq Type' column in the variant table will describe the type of mutation predicted based on the gene model. As these consequence predictions are based on the gene model, it means that they are only as good as the gene model itself. Therefore, if the gene model is incorrect then the consequence prediction may also be incorrect. Even though the RefSeqv1.1 annotation is very high quality it is still important to check that your gene of interest is annotated correctly. See "[Gene models](#)" for a description of how to assess gene model quality and what to do if the gene model is incorrect.

In many cases, the most desirable consequence will be "stop_gained" (premature termination codon), which will most likely result in a loss of function of the gene of interest. It is important to consider the position of the premature stop codon in the protein as this might have implications for the phenotype. A premature stop codon early in the gene could be a good sign that the protein is non-functional, however, a premature stop codon that occurs very late in the gene could still result in a functional protein.

If there are no "stop_gained" mutations available, then a "splice_acceptor_variant" or "splice_donor_variant" could be used. These are mutations in the GT or AG sites that are located at the start and end of introns, respectively. A mutation that results in loss of one of these sites could

Updated April 2019

lead to incorrect splicing, which in turn has a high chance of resulting in a non-functional protein. As before, it is important to understand the context of the mutation and identify downstream GT or AG sequences that could potentially be used as alternative splice sites. Note that there is another category of mutations labelled as “splice_region_variant intron variant”. These are mutations occurring within 5 base pairs of the canonical splice sites, but not at the splice sites themselves. The effects of these mutations are difficult to predict. While splicing is not just controlled by the canonical splice sites themselves (GT or AG), there is no *a priori* way of identifying nucleotides relevant for correct splicing. We would discourage using these mutations without any knowledge of conserved splicing residues for a given gene.

If there are no “stop_gained” or “splice_acceptor/donor mutations” available, then another option could be a missense mutation. These are coding sequence mutations that result in a change of amino acid. The effect of these mutations on the protein is often difficult to predict and can range from no effect to disruption of an important domain resulting in either a non-functional protein or a protein with altered function. The Sorting Intolerant From Tolerant (SIFT) score can help to decide which mutation to select. The SIFT algorithm scores the probability that a particular amino acid substitution will be tolerated in the protein. It can be interpreted in the same way as a p-value, where a lower SIFT score implies a significant effect on protein function (e.g. a SIFT score of <0.05 implies that the mutation is very likely to affect protein function). More information on how the SIFT algorithm works can be found here: <http://sift.jcvi.org/>. Alternatively, you could look for missense mutations in highly conserved residues of the protein.

A “start_lost” is a loss of the ATG start codon of a gene. This could be a useful mutation, but often there is another ATG in frame that could be used so it is important to check this before selecting these mutations.

The effects of “intron_variants” and “UTR_variants” are very difficult to predict and so may not be ideal choices unless you have in depth information about the regulatory sequences of your gene.

It is important to invest time in defining the best mutations to be used in the crossing scheme. As described in [Borrill et al., \(2015\)](#), “This is especially relevant in situations of functional redundancy between homoeologs, where it is necessary to cross individual mutants to generate double or triple knockouts to observe a phenotype. The use of truncations in all homoeologs will generate a complete null across genomes, thus allowing for correct interpretation of the resulting phenotype. However, the use of missense mutations is risky if one of the mutations does not effectively abolish gene function, thereby limiting the phenotypic effect in the double/triple mutant.”

Zygoty

Mutations may be in either a homozygous or heterozygous state. Homozygous mutations are preferable as these should be present in all seeds and so fewer seeds will need to be screened to confirm mutations (see “[Designing crossing schemes](#)”, “[Designing genome specific primers](#)” and “[Wheat growth](#)” section for practical aspects of screening).

Updated April 2019

The Zygosity of the mutation can be found on Ensembl Plants. If using the Variant Image (Figure 2) you can click on the individual mutation and select “more about” (Figure 2a), or if using the Variant Table (Figure 3) press the Variant ID hyperlink itself. This will open up a page for the specific Variant with more information. Most of this information is also found on the table or pop-up, but the individual variant page also displays the zygosity and primer information for each mutant. This is provided as “Synonyms” within the Variant Page:

Synonyms

Primer metadata Cadenza2044.chr5B.417442773:snp_type=non-homoeologous;total_contigs=3;A=caccaccttccgAaaaccC;B=caccaccttccgAaaaccT;common=tcgAcacGtAtGGAAATGtACAT;primer_type=chromosome_specific;product_size=81;EMS-GT=Hom

In this example, the zygosity information is found at the end of the string (EMS-GT=Hom) meaning that the EMS mutation is homozygous (Hom). In the case of heterozygous mutations this will display as (EMS-GT=Het). This string also includes the pre-designed KASP markers for validation of the SNP (see below).

Heterozygous mutations require some additional work but have distinct advantages. The sequence information for a particular mutant comes from a single M₂ plant and the seeds you receive will be either M₄ or M₅ seed (two or three additional self-pollinated generations). Heterozygous mutations will most likely still be segregating in the M₄ or M₅ seed so a larger number of seeds will need to be screened to identify homozygous mutations. Despite this apparent drawback, the opportunity to identify both homozygous mutant and wild type plants provides a proper experimental control in case you want to assess phenotypes in this first generation. It is important to consider that some of the mutations that were heterozygous in the M₂ may have been fixed in the M₄ or M₅ seed, whereas other mutations may have been lost through genetic drift.

In all cases it is advisable that you confirm the mutations with markers (e.g. KASP assays) or alternative methods (e.g. sequencing) to ensure the presence of the mutation in plants that will be phenotyped or used for crossing. Pre-designed KASP assays are available for most mutations (see below). See “[Designing genome specific primers](#)” for more information about how to design KASP assays for mutations.

Population

As mentioned above, there are two TILLING populations available: Cadenza (hexaploid) and Kronos (tetraploid). The choice of population will depend very much on the final aim of your study. If the aim of your study is to investigate a complete knockout of the gene, then the tetraploid Kronos population might be preferred as it will be quicker to reach a complete knockout by crossing since there are only two homoeologs present (A and B). The hexaploid background (Cadenza) contains a third homoeolog (D) and will therefore require a longer and more complex crossing scheme. Alternatively, if you want to assess the effect of the mutation in elite hexaploid backgrounds then the Cadenza population would be better. Your choice of population may be influenced by the availability of mutations, for example, there may not be a Kronos line with a premature stop mutation in your gene

Updated April 2019

of interest but there might be one in the Cadenza population. Unless you have a very strict requirement for either population, it is always worth looking for mutations in both populations and then making the decision as to which lines to take forward. The genomes of both cultivars have also been recently sequenced and assembled. Both wild type assemblies are available at the [Earlham Institute Grassroots Genomics](#) blast server.

Available primers

If several mutant lines have a suitable mutation, one deciding factor could be the availability of chromosome_specific KASP genotyping primers. The pre-designed KASP genotyping markers are provided within the individual Variant page for each mutation (see Zygosity above). Once on the Variant page, the primers are presented within the “Synonyms” string:

Synonyms

Primer metadata Cadenza2044.chr5B.417442773:snp_type=non-homoeologous;total_contigs=3;A=caccaccttccgAaaaccC;B=caccaccttccgAaaaccT;common=tcgAcacGtAtGGAAATGtACAT;primer_type=chromosome_specific;product_size=81;EMS-GT=Hom

In this example, the string includes the following information:

- **snp_type:** non-homoeologous; meaning that the mutation is not found between the A, B and D genomes of wheat. This is expected for EMS-induced mutations.
- **total_contigs:** This is the number of contigs that were retrieved for the multiple alignment that helps design the genome-specific assay. In this case, there are 3 contigs, which is expected from the A, B and D genomes.
- **Primers:** the three primers from the KASP reaction are provided as A, B and common. The A primer is for the wild type/reference allele; the B primer for the EMS-induced mutation and the common primer is the one that provides genome-specificity. Letters in uppercase indicate bases that are unique to a genome and hence provide specificity to the KASP assay.
- **primer_type:** In this example the assay should be homoeolog specific (**chromosome_specific**) as PolyMarker was able to design a common primer with several homoeolog specific SNPs (indicated by the uppercase letters in the common primer). For some EMS-induced mutations, the common primer cannot be designed to be homoeolog specific and in such case the primer metadata will state “**chromosome_nonspecific**”. If primer design was not possible, this will also be indicated (“Primers not designed”).
- **product_size:** indicates the expected size of the PCR product resulting from the amplification with primers A/B against the common primer.
- **EMS-GT:** this indicates the zygosity of the mutation as either homozygous (Hom) or heterozygous (Het).

Please note:

1. The diagnostic allelic primers do not contain probe sequences for fluorescent dyes (FAM and HEX/VIC) used in a typical KASP assay. These must be added to the 5' end of the primers. Add the FAM probe sequence (GAAGGTGACCAAGTTCATGCT) to the 5' end of the wild

Updated April 2019

type/reference primer (A) and VIC/HEX probe (GAAGGTCGGAGTCAACGGATT) to the 5' end of the mutant primer (B).

2. Primers were designed using <http://www.polymarker.info/> ([Ramirez Gonzalez et al., 2015](#)).

The absence of genotyping primers, however, should not stop you from selecting a mutation if it is otherwise fit for purpose as primers for KASP and alternative genotyping methods can be designed manually. More information on genotyping can be found in the "[Designing genome specific primers](#)" document.

c) Ordering seeds

Seeds can be ordered directly through the UK Germplasm Resource Unit (GRU) website called SeedStor using the following link: <https://www.seedstor.ac.uk/shopping-cart-tilling.php>. This link will direct you to a shopping cart for the Kronos and Cadenza TILLING populations. Mutations were identified in DNA from a single M₂ plant whose M₃ seeds were pooled and bulked for distribution. Therefore, the seeds which are distributed correspond to M₄ or M₅ seed (depending on seed availability).

To order seeds you need to enter the mutant line number(s) (for example Kronos1234 or Cadenza0250) into the ordering form. Seeds from both populations can be placed within one order. You can also request seeds from the wild type Kronos or Cadenza parents. The form is sent to the GRU for processing and the order is kept confidential at all times. The developers of the populations do not have access to these orders.

Requests for seeds operate on a cost recovery basis for parties interested in research purposes only (£15 per line + £10 flat handling fee per shipment). This MTA allows the use of the mutants for research purposes only. This includes crossing between mutants and Cadenza and Kronos wild type plants. In cases where potential industrial applications are envisaged and crosses are performed to cultivars/breeding lines different to Cadenza and Kronos, there is an MTA that confers "freedom to operate" for the mutant lines. This MTA is £200 per line (+£10 handling fee per shipment) and provides a non-exclusive right to use the mutant line and its derivatives for commercial purposes. Details are described in the corresponding MTAs available on the SeedStor website (<https://www.seedstor.ac.uk/shopping-cart-tilling.php>).

d) What to do next

Germinating and growing plants:

Please see the relevant PDF documents in the "[Introduction to Wheat Growth](#)" section of the website, which has detailed explanations regarding how to germinate seeds and the basics of growing wheat under glasshouse conditions.

Designing crossing schemes:

Many of the mutations will need to be backcrossed to reduce the mutation load or inter-crossed to produce double and triple mutants of the corresponding homoeologs. We describe different crossing schemes and approaches in "[Designing crossing schemes](#)".

More information on the TILLING mutant resource can be found in Krasileva *et al.*, (2017) DOI: 10.1073/pnas.1619268114.

e) Instructions on how to use www.wheat-tilling.com (legacy website)

The www.wheat-tilling.com website is the original website used to host the data for the TILLING populations. However, this data is mapped to the CSS genome assembly, and uses the CSS gene models to predict mutation effects (see "[Genome assemblies](#)" and "[Gene models](#)" for more information). The data on this website is therefore based on less reliable gene models compared to the data hosted on Ensembl Plants and ideally should not be used on first instance. The website has been left in place as a legacy site in case this dataset is required for a specific reason. Instructions on how to use this website from a previous iteration of www.wheat-training.com are below:

Using the TILLING website

1. Go to the website www.wheat-tilling.com
2. Enter your CSS gene name(s) into the 'Search TILLING data' box (Figure 4a).
 - a. *You may also search for mutant lines using the IWGSC CSS scaffold or by blasting the sequence. Blasting the sequence may be useful for when your gene model is bad or not present in the IWGSC CSS reference (see "[Gene models](#)" for more details). You may also search for all mutations present in a specific mutant line. There is also an option to upload a ".csv" file containing your terms of choice (Figure 4d).*
3. Select the population you wish to find mutations in. You may select Cadenza (hexaploid), Kronos (tetraploid) or both (Figure 4b). The default is both populations. We will use both.
4. Click 'Search' (Figure 4c). This will open up an HTML output (Figure 6).

Wheat TILLING John Innes Centre Earlham Institute hmi Howard Hughes Medical Institute ROTHAMSTED RESEARCH

Welcome to the *in silico* wheat Target Induced Local Lesions In Genome (TILLING) website

This resource consists of TILLING populations developed in tetraploid durum wheat cv 'Kronos' and hexaploid bread wheat cv 'Cadenza' as part of a joint project between the University of California Davis, Rothamsted Research, The Earlham Institute, and John Innes Centre.

We have re-sequenced the exome of 1,535 Kronos and 1,200 Cadenza mutants using Illumina next-generation sequencing, aligned this raw data to the IWGSC Chinese Spring chromosome arm survey sequence, identified mutations, and predicted their effects based on the protein annotation available at Ensembl Plants.

Search TILLING data

Population Cadenza Kronos Both
Type in list of search terms (scaffold, line or gene)

Traes_1AL_9EC1E6F0C

Examples: Traes_1AL_9EC1E6F0C, Traes_1AL_9EC1E6F0C.1, IWGSC_CSS_2AL_scaff_6343779, 2AL_6343779, Cadenza0250

Or choose an input file containing terms:

No file chosen

BLAST Scaffold

Paste query sequence(s) or drag file containing query sequence(s) in FASTA format here ...

Nucleotide databases

IWGSC CadenzaU KronosU v1

Advanced Parameters: ?

Additional Resources
Additional resources and documentation of the website and populations

Seeds: To order seeds please visit the SeedStor site at www.seedstor.ac.uk you will need the mutant line identifier to request seed (e.g. Cadenza1111 or Kronos1111)

SeedStor

Figure 4: TILLING browser

Using BLAST to identify mutations

An alternative way to search for mutations in your gene of interest is to BLAST the gene sequence. This could be useful if, for example, your gene of interest has an incomplete gene model or no gene model present in the IWGSC CSS assembly (see “[Gene models](#)” for more information). Copy and paste your sequence into the ‘BLAST scaffolds’ box (Figure 4e) and click ‘BLAST’. The result will include:

- The visual representation of the query sequence and the hits to the database (Figure 5a)
- Table with the BLAST output (Figure 5b). Each row will be a scaffold that shows similarity to the Query input sequence. The scaffolds will either be an IWGSC CSS scaffold (e.g. IWGSC_CSS_5AL_scaff_2810281) or *de novo* assembled scaffolds with the “UCW_Kronos_U” or “TGAC_Cadenza_U” prefix. These ‘U’ scaffolds consist of *de novo* assembled reads from the exome capture and sequencing that did not align to any of the IWGSC CSS scaffolds. Clicking on the scaffold name (Figure 5c) will link to the High-scoring Segment Pair (HSP) alignment below (Figure 5d). Beside the scaffold name is a link called “Mutations” which will hyperlink to the HTML output for the EMS mutations in the subject sequence (Figure 5e). This table is the same table that would have been obtained by searching the database using that scaffold name as described above. The E value and length of the Subject sequence are also part of the output.

- c) The High-scoring Segment Pair (HSP) alignment of the Query and Subject sequences. Here the “View Mutations” button will also open up the HTML output for the EMS mutations (Figure 5f).

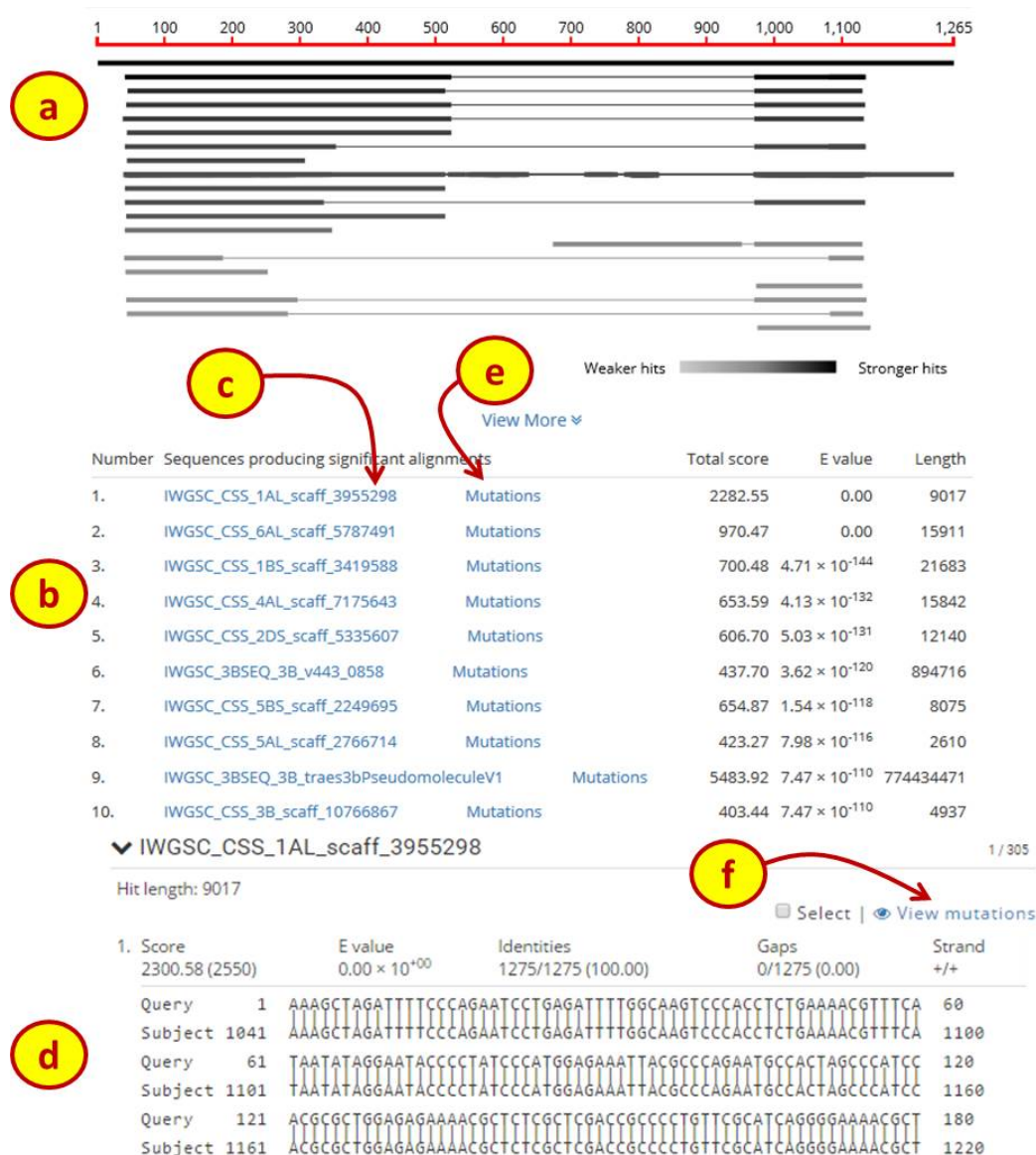


Figure 5: BLAST output
Understanding the output file

The excel file output will have 25 columns and one row for each mutation found in your gene(s) of interest. A description of the columns is in Table 1.

Table 1: Description of the TILLING output headers

Column no.	Column name	Description
1	Scaffold	This is the name of the IWGSC CSS scaffold using the Ensembl Plants nomenclature. The name has a hyperlink to the actual FASTA sequence of the scaffold which allows users to identify the position within the genomic DNA context.
2	Chr	This indicates to which chromosome an IWGSC scaffold is assigned to and is a useful filtering feature.

3	Line	Name of the mutant line, which consists of the name of the mutant population (Cadenza or Kronos) and a four-digit number (e.g Cadenza1574).
4	Category	EMS mutations were classified according to different categories which are detailed below.
5	Position	The position of the SNP relative to the start of the respective IWGSC scaffold. The actual FASTA sequence is available through the hyperlink in the scaffold name (Column 1).
6	Chromosome position	The position of the SNP relative to the chromosome pseudomolecules hosted at Ensembl Plants (Release 27; http://archive.plants.ensembl.org/Triticum_aestivum/Info/Index).
7	Rf	<i>Reference</i> : This is the base call at the position in the reference genome of Chinese Spring (CS) published by the IWGSC.
8	wt	<i>Wild type</i> : This is the base call at the particular position in a non-mutagenized wild type control of either the hexaploid variety Cadenza or the tetraploid variety Kronos.
9	mt	<i>Mutant</i> : This is the base call at the position in the EMS-mutagenized individual from either the hexaploid variety Cadenza or the tetraploid variety Kronos.
10	Het/hom	Classification of the individual as heterozygous or homozygous for the mutant allele. If the wild type coverage (WT_cov) is 0 or <15% of the total reads, then the SNP is called as homozygous mutant. Otherwise the SNP is called as heterozygous.
11	WT cov	<i>Wild type coverage</i> : Number of reads with the wild type allele at the position.
12	Mut cov	<i>Mutant coverage</i> : Number of reads with the Mutant allele at the position.
13	Gene	This is the name of the IWGSC gene model that is present on the IWGSC scaffold. More than one gene can be present on a single scaffold. The ".X" ending refers to the transcript variant which was used to predict the effect of the mutation. The gene name is hyperlinked to the Ensembl Plants gene page.
14	Consequence	Consequence of the mutation on the transcript as predicted by the Variant Effect Predictor (VEP) tool from Ensembl based on the IWGSC gene model. A full detail of the different effects predicted is described here (http://www.ensembl.org/info/genome/variation/prediction/predicted_data.html).
15	cDNA pos	Position of the mutation relative to the cDNA of the predicted IWGSC gene model.
16	CDS pos	Position of the mutation relative to the coding sequence (CDS) of the predicted IWGSC gene model.
17	Amino acids	Indicates whether a mutation leads to the change of an amino acid in the translated protein of the predicted gene model. For UTR, intron and synonymous variants this value is empty; for non-synonymous mutations the amino acids encoded by the wild type and mutant alleles are shown. For example, a value of "P/L" means that the wild type codon encoding for a proline (P) residue was mutated to a codon encoding for a Leucine (L) residue.

18	Codons	Sequence of the wild type / mutant codon in the mutant line. The uppercase letter indicates the base which was mutated. For example, cCa/cTa corresponds to a mutation in the middle position of the CCA codon (P) which was mutated to a CTA codon (L).
19	Sift score	Sorting Intolerant From Tolerant (SIFT) probability score that that the amino acid change is tolerated. Scores <0.05 are considered deleterious whereas all others are considered tolerated. SIFT was implemented through Ensembl VEP as described in Krasileva et al., 2017 .
20	Primer type	The predicted specificity of the designed KASP assays according to its ability to preferentially amplify a single target genome (specific), discriminate against at least one alternative genome (semi-specific) or not able to discriminate and hence amplifying all genomes or paralogues of the gene (non-specific). Specific and semi-specific assays are preferred where possible for downstream validation of the mutation. Primers were designed using http://www.polymarker.info (Ramirez-Gonzalez et al., 2015).
21	Orientation	Plus (+) or minus (-) according to the strand orientation of wild type and alternative primers on the IWGSC scaffold (column 1) sequence.
22	WT primer ¹	Sequence of the wild type primer. Uppercase bases within the primer sequence correspond to genome specific SNPs, whereas the 3' uppercase base corresponds to the target EMS SNP.
23	Alt primer ¹	Sequence of the alternative primer which amplifies the mutant allele.
24	Common	Common primer sequence for the KASP assay.
25	Multimap	Additional scaffolds to which a mutation could be defined based on the multi-map pipeline (see below).

¹ The diagnostic allelic primers do not contain probe sequences for fluorescent dyes (FAM and HEX/VIC) used in a typical KASP assay. These must be added to the 5' end of the primers. Add the FAM probe sequence (GAAGGTGACCAAGTTCATGCT) to the 5' end of the WT primer and VIC/HEX probe (GAAGGTGCGGAGTCAACGGATT) to the 5' end of the Alt primer.

Exporting results:

The HTML output can also be exported as an Excel compatible “.csv” file by clicking on the “Export” button (Figure 6a). This allows you to download the results for a given scaffold, gene or mutant line and perform additional analysis and filtering steps in a local environment.

Filtering:

The results can be filtered for specific terms or names by using the search box in the HTML output (Figure 6b). For example, the search term “stop” will filter the results to show only those mutations which have a “stop_gained” consequence. To remove the filtering, press the “x” at the end of the Search box.

Sorting:

The results for a given column can be sorted by clicking directly on the column header (Figure 6c). For example, to display the mutations based on their zygosity then the “Het/Hom” header can be

Updated April 2019

clicked, or the “cDNA pos” header can be clicked to sort the mutations based on their position within the target gene cDNA.

Displaying columns:

The specific column to be displayed can be modified by pressing the “show/hide columns” icon (Figure 6d) and selecting or deselecting the desired columns.

Categories of mutations

Five categories of mutations are displayed on the www.wheat-tilling.com website. The first three relate to the minimum number of reads including a mutation that were required by the SNP calling pipeline to consider a mutation real. The minimum coverage (*MC*) to call a mutation was established independently for homozygous (*Hom*) and heterozygous (*Het*) mutants using the *HomMC* and *HetMC* parameters. The optimal threshold is displayed as default: a minimum coverage of five mutant reads for heterozygous (*HetMC5*) and three for homozygous mutations (*HomMC3*) (estimated <1% potential error). Additional mutations were also identified at lower coverages *HetMC4/HomMC3* and *HetMC3/HomMC2* but at the expense of potential errors (<3% and 10% error, respectively). It is safer to use mutations identified at higher stringencies (*HetMC5/HomMC3*), but if a desired mutation is identified in the lower categories, it still has a good probability of being real.

Two additional categories were used to classify mutations.

1. *Multi-Map (MM)*: This category was used to designate regions which have high sequence similarity across multiple scaffolds and hence reads from regions mapped to multiple scaffolds. These regions correspond to recently duplicated genes, homoeologs with unusually high sequence similarity or artificially duplicated scaffolds generated during the assembly of the IWGSC reference sequence. Information of the additional putative scaffolds to which the mutation could be mapped are presented in the last column of the table output (column 25).
2. *Residual Heterogeneity (RH)*: This category corresponds to SNPs which were present in multiple individuals and are most likely the result of residual heterogeneity in the initial seed stocks used for the EMS mutagenesis.

The www.wheat-tilling.com output displays by default *HetMC5/HomMC3*, MM and RH mutations. If you cannot find an adequate mutation within these categories then click the “Show lower quality mutations” to display mutations classified as *HetMC4/HomMC3* and *HetMC3/HomMC2* (Figure 6e).

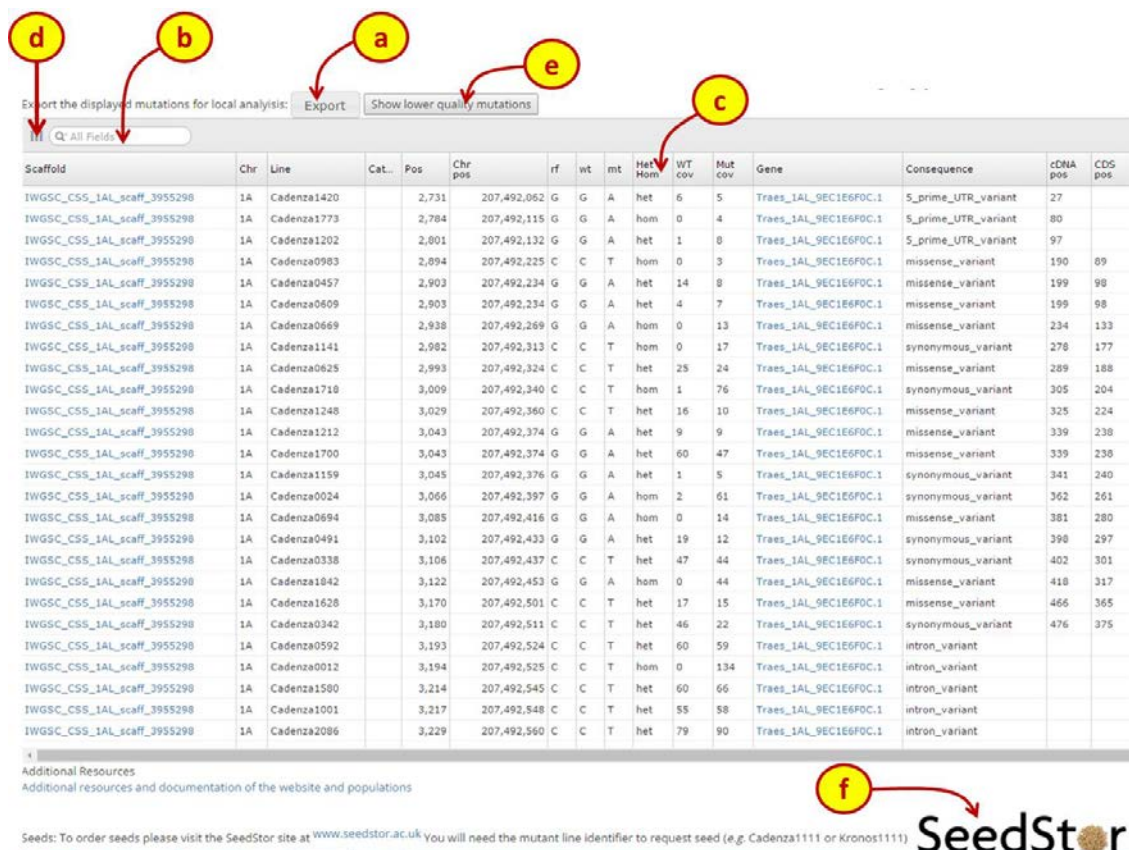


Figure 6: TILLING HTML table output