# Identification of novel mutations responsible for the low Vitamin E phenotype in *Arabidopsis thaliana*

## Introduction

The term “vitamin E” refers to a collective group of eight fat-soluble compounds - four tocopherols and four tocotrienols - virtually exclusively synthesized by photosynthetic organisms. Vitamin E is an essential nutrient that the human body cannot synthesize endogenously and, in the human diet, is usually found in vegetable oil and nuts. Although it has not yet been implicated in a vital metabolic function like most nutrients, it has been established as a potent antioxidant that can protect cell membranes from oxidative damage induced by free radicals.

The biosynthetic pathway for the production of vitamin E is very well-known. Metabolism of tyrosine from the Shikimate pathway and chlorophyll from the MEP pathway lead to the formation of tocopherols. However, not much is known about the specific physiological and molecular mechanisms for the regulation of vitamin E biosynthesis. To this end, *Arabidopsis thaliana* seeds were induced to mutate through treatment with ethyl methanesulfonate (EMS), then backcrossed with parental wild-type plants to create homozygous mutants. These mutants were then screened for a low vitamin E phenotype using normal-phase high performance liquid chromatography (NP-HPLC); six low vitamin E mutant strains were sequenced.

Our report focuses on the identification of novel mutants responsible for the low vitamin E phenotype in the LOVE2 strain. Out of the 6 candidate variants we have filtered, we submit a missense mutation in the gene AT5G34930, annotated as coding for chloroplastic arogenate dehydrogenase 1, as the mutant with the most potential to be responsible for the phenotype.

## Methods

*Sequence QC and Remapping*

Thirty-nine LOVE2 mutants were sequenced using Illumina sequencing. We received the raw sequencing data in a .fastq format. From this point on, we used the BeInf cluster provided by the University of Bern for high-performance computing. We used Trimmomatic to trim low-quality ends of reads in the raw sequencing data and used FastQC to visualize the changes performed. Once QC was performed, we remapped the reads against the *A. thaliana* TAIR10 reference genome provided by Laurent Falquet using the Burrows-Wheeler Aligner, then marked duplicates with Picard Tools.

*Sequence Base Quality Score Recalibration and Variant Calling*

After remapping was performed, we performed base quality score recalibration using GATK4 BaseRecalibrator after doing an initial variant call using HaplotypeCaller and performing a hard filter using VariantFiltration. Once the quality scores have been recalibrated, we performed HaplotypeCaller once again to produce a GVCF file.

*Variant Annotation and Filtration*

We annotated the GVCF file produced by GATK4 against the *A. thaliana* TAIR10 reference genome using SnpEff. Using SnpEff, we also removed any variants that were synonymous or were not located in coding regions, kept variants located only in less than five strains, and kept only mutations that were due to EMS (G > A and C > T mutations only). Once this was done, we used reference genomes and annotations provided to us by Laurent Falquet and used BedTools to get a list of mutations. The resulting VCF file was then visualized using IGV and GenSearch NGS for mutant analysis.

## Results

From the methods described above we analyzed several candidate variants, tabulated below.

Table Putative genes responsible for the low vitamin E phenotype in the LOVE2 strain.

|  |  |  |
| --- | --- | --- |
| **Chromosome/Position** | **Gene name (annotation)** | **Variant type** |
| **III** / 23044079 | BOR2 (Probable boron transporter 2) | Stop gained |
| **V** / 1392884 | PPR4 (Pentatricopeptide repeat-containing protein AT5G04810, chloroplastic) | Missense |
| **V** / 3264419 | AT5G10370 (ATP-dependent RNA helicase DEAH12, chloroplastic) | Missense |
| **V** / 7573327 | AT5G22760  (PHD finger family protein) | Missense |
| **V** / 13234461 | AT5G34930  (Arogenate dehydrogenase 1, chloroplastic) | Missense |
| **V** / 26552629 | PCMP-H61 AT5G66530 (Glucose-6-phosphate 1-epimerase) | Missense |

## 

## Discussion

### BOR2 variant causes a premature stop

Out of the 56 variants we filtered, only the BOR2 variant led to the gain of a stop codon. Boron is established to be an important factor to plant cell wall integrity as plants grown in boron-deficient environments have aberrant cell walls, but the mechanisms by which boron maintains the integrity of the cell wall are yet unknown (Brown et al., 2002). Deficient boron uptake leads to a strong phenotype of apical meristem cessation and brittle leaves. This was not the best candidate variant as it has no direct impact on the tocopherol synthetic pathways and its strong phenotype.

### AT5G10370 variant causes a missense mutation

AT5G10370 is annotated to be a chloroplastic ATP-dependent RNA helicase. Although its localisation makes it relevant in our LOVE2 mutant search, a missense mutation on a helicase that led to a viable phenotype suggests it may not have had an actual effect on any plant processes.

### AT5G22760 variant affects a PHD finger protein

AT5G22760 is annotated to be a PHD finger family protein. This family of proteins affects histone modification in chromatin and has been implicated in a wide range of plant functions ranging from vernalization to seed germination (Sung & Amasino, 2004). However, this specific PHD finger protein is yet uncharacterized and is unsuitable for analysis.

### Missense mutation in the AT5G66530 variant

AT5G66530is annotated to be a chloroplastic glucose-6-phosphate 1-epimerase. This specific enzyme catalyzes an equilibrium reaction between alpha-D-glucose-6-phosphate and beta-D-glucose-6-phosphate, which are intermediates in glycolysis. Due to glucose-6-phosphate not being directly involved in either the production or metabolism of tyrosine or chlorophyll, there is a very low chance a mutation of AT5G66530 having a substantial effect on vitamin E synthesis.

### PPR4 missense variant in a gene that affects chlorophyll supply

PPR4 codes for a pentatricopeptide repeat. Pentatricopeptide repeats are found widely among the eukaryotes. Although most are uncharacterized, they are thought to bind to RNA and influence RNA processing prior to translation. In a 2006 study, two PPR4 insertion mutants were shown to cause a non-photosynthetic phenotype in *Zea mays* and affect plastid ribosome biogenesis (Schmitz-Linneweber et al., 2006). The non-photosynthetic trait was caused through a lack of chlorophyll, leading to a lethal phenotype where seedlings died within three weeks. Although this variant is directly implicated in the vitamin E synthesis process, the lethality of a chlorophyll loss phenotype indicates that this PPR4 missense variant is not the best choice of the six candidates.

### 

### Missense AT5G34930 variant may be the cause for the LOVE2 phenotype

AT5G34930 was characterized in 2002 as the chloroplastic arogenate dehydrogenase TYRAAT1 (Rippert & Matringe, 2002). Arogenate dehydrogenase (ADH) catalyzes the conversion of arogenate to tyrosine and is the last step in the tyrosine biosynthetic process, indicating that a mutation in this gene can affect vitamin E biosynthesis through tyrosine availability. In a 2010 study, loss of function of arogenate dehydrogenase in maize resulted in viable plants with stunted growth and opaque endosperm, suggesting that tyrosine was not solely synthesized by this process (Holding et al., 2010). This hypothesis was confirmed in 2014 when tyrosine-insensitive prephenate dehydrogenases were characterized in legumes (Schenck et al., 2014). Originally characterized in *E. coli*, prephenate dehydrogenases synthesize tyrosine by transaminating 4-hydroxyphenylpyruvate (HPP) to form tyrosine. Thus, a loss-of-function mutation of TYRAAT1 may produce a stunted but viable phenotype that, due to the lower tyrosine supply, would have a lower vitamin E phenotype as well. It is for this reason that we select AT5G34930 as the suspect mutation for the low vitamin E phenotype in the LOVE2 *A. thaliana* strain.

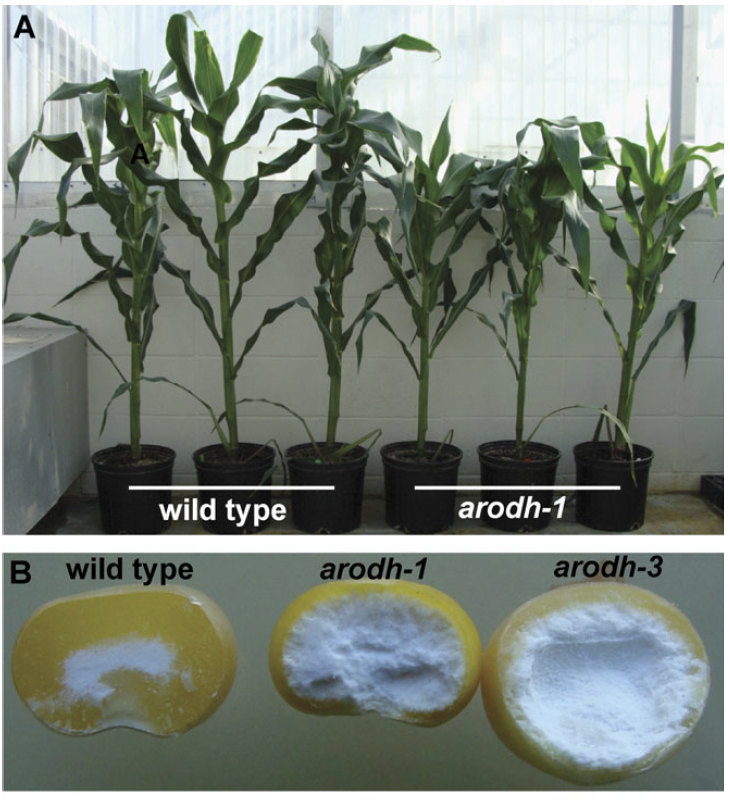
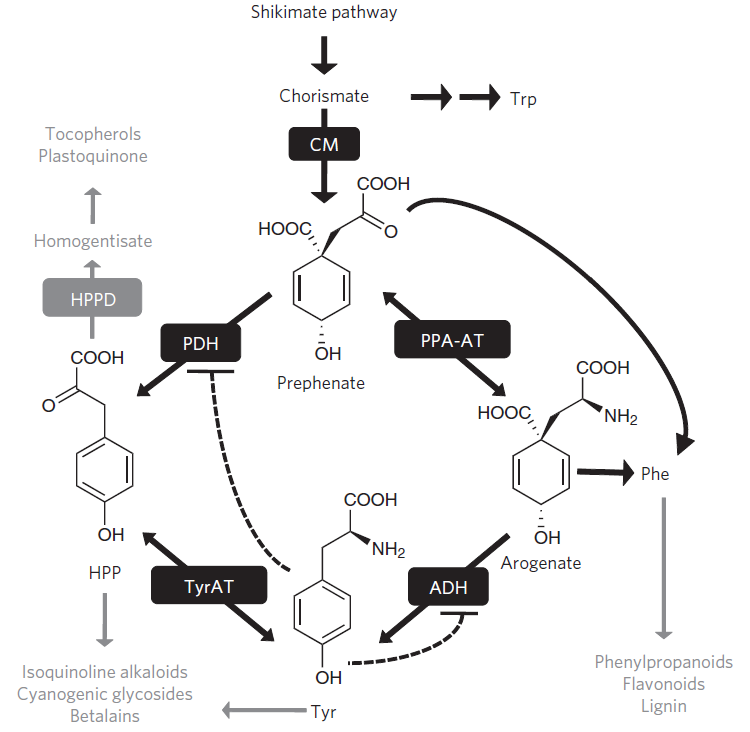


Fig. Left: Tyrosine biosynthesis pathway proposed by Schenk et al., 2014. Right: Phenotypes of ADH-deficient maize plants by Holding et al., 2010.

## References

Brown et al. 2002. Boron in Plant Biology. *Plant Biology* 4(2): 205-223.

Galli et al. 2016. Vitamin E: Emerging aspects and new directions. *Free Radical Biology and Medicine* 102: 16-36.

Holding et al. 2010. Identification and characterization of the maize arogenate dehydrogenase gene family. *Journal of Experimental Botany* 61(13): 3663-3673.

Institute of Medicine. 2000. Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids: 186-283.

Mène-Saffrané. 2018. Vitamin E Biosynthesis and its Regulation in Plants. *Antioxidants* 7(1): 2.

Mène-Saffrané and Pellaud. 2017. Current strategies for vitamin E biofortification of crops. *Current Opinion in Biotechnology* 44: 189-197.

Pellaud et al. 2017. WRINKLED1 and ACYL-COA:DIACYLGLYCEROL ACYLTRANSFERASE1 regulate tocochromanol metabolism in Arabidopsis. *New Phytologist* 217: 245-260.

Rippert & Matringe. 2002. Molecular and biochemical characterization of an Arabidopsis thaliana arogenate dehydrogenase with two highly similar and active protein domains. *Plant Molecular Biology* 48: 361-368.

Schenck et al. 2014. Non-plastidic, tyrosine-insensitive prephenate dehydrogenases from legumes. *Nature Chemical Biology* 11: 52-57.

Schmitz-Linneweber et al. 2006. A Pentatricopeptide Repeat Protein Facilitates the trans-Splicing of the Maize Chloroplast rps12 Pre-mRNA. *The Plant Cell* 18(10): 2650-2663.

Sung & Amasino. 2004. Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3. *Nature* 427: 159-164.

# Identification of Peptidoglycan Hydrolase responsible for Cheese Ripening in *Lactobacillus helveticus*

## Introduction

Bacteria are essential in cheese production. They are indispensable for texture and taste development but also help to reduce and kill pathogens as well spoilage bacteria that pose a risk for food safety. This is why, depending on the bacterial makeup within the cheese during ripening, the characteristics of the cheese will vary greatly and controlling the bacterial cultures used in cheese production is crucial (Bruggmann, 2018).

One bacterial species belonging to the group of lactic acid bacteria widely used in cheese production is *lactobacillus helveticus*. *Lactobacillus helveticus* strains have the feature to produce peptidoglycan hydrolases (PGHs), which are enzymes capable to digest the bacterial cell wall. By releasing these enzymes the bacteria kill other bacteria surrounding them but also induce their own lysis (autolysis). This feature is exploited in cheese production, for example to induce a change in bacterial makeup during ripening. *Lactobacillus helveticus* strains however show considerable genomic plasticity leading to high variation in PGH-activity from one strain to another (Bruggmann, 2018).

In a previous study the activity of a suspected PGH of an estimated size of about 30kDa was tested in six distinct *lactobacillus helveticus* strains (FAM8102c1c1, FAM23285, FAM19191, FAM22076, FAM1450, FAM1213) by zymography. The six strains showed one of two phenotypes: either they exhibited PGH activity in zymogram (phenotype A: FAM8102c1c1, FAM23285, FAM19191) or not (phenotype B: FAM22076, FAM1450, FAM1213) (Bruggmann, 2018). In order to assess the possible genomic differences underlying the two phenotypes, the six strains were sequenced using the Illumina technique. The aim of the present report was to identify the gene which is responsible for this phenotype by assembling the genomes of the six strains, annotating suspected PGHs using the annotation of a previously sequenced reference *lactobacillus helveticus* strain (DPC 4571) (Jebava et al., 2011) and comparing the genomes.

One putative PGH of about 30 kDa present only in the three strains showing activity in the zymogram could be identified. Analysis of the region surrounding the gene in question suggests that it may have been acquired by viral insertion in the ancestry of the three active strains.

## Methods

### Sequencing and de novo genome assembly

The six *lactobacillus helvetiticus* strains FAM8102c1c1, FAM23285, FAM19191, FAM22076, FAM1450, FAM1213 were sequenced by Illumina sequencing. Subsequent data analysis was performed using the server provided by the University of Berne. The quality of the sequencing reads was verified by QSUB. SICKLE as well as TRIMMOMATIC were applied to filter out bad quality reads. Two seperate softwares were applied to the reads of each strain for assembly: SOAPdenovo and SPAdes. For SOAPdenovo assemblies, k was set to: 95, 85, 75 and 65. The five assemblies generated for each strain (four by SOAPdenovo and one by SPAdes) were compared using Abyss. The maximum number of contigs was set to 1000. Amongst the five assemblies of each strain the best was chosen according to the following criteria: small total number of contigs (set to maximum 1000), big N50 and total genome size of approximately 20Mb corresponding to the expected genome size of *lactobacillus helveticus*.

### Genome annotation and pan-genome analysis

To annotate the six chosen assemblies they were submitted to the PROKKA pipeline. PROKKA is a fully automatic annotation pipeline for prokaryotic genomes (VBC Victorian Bioinformatics Consortium, 2018). In parallel the genome of the reference strain DPC 4571 obtained from external sources was also annotated by the same software. Doing so, the software was advised to blast draft genomes against the sequences of nine peptidoglycan hydrolases (PGHs) that had previously been described in the reference strain DPC 4571 in addition to the information stored within the UniProt database. Resulting annotated draft genomes were visualized with MAUVE. To get an idea about the relatedness and differences between the strains, annotated draft genomes were submitted to Roary to generate a pan-genome of the strains. The result was visualized with phandango.

### Identification of putative Peptidoglycan hydrolase described by zymography

From the files generated by Roary grep was applied to filter out all genes annotated as PGHs. The set of PGHs found within strains showing phenotype A was compared to those found in strains exhibiting phenotype B. Only those putative PGHs present solely in the three active strains were considered as candidates for the PGH in question. To further narrow down the list of candidates, the approximate protein size of each candidate was calculated from the size of the coding region and compared to the expected size of the protein found within the zymogram (30 kDa). One remaining candidate was retained after following this procedure. In the aim to learn more about the history of this gene, its genomic contexte was visualized with MAUVE. Suspecting that this gene could have been acquired by the active strains via a mobile genetic element, the genomic context was scanned for sequences that might indicate: transposons, genomic islands, plasmids and bacteriophages.

## Results

A total number of 20 putative PGHs were identified as such in the 6 *lactobacillus helveticus* strains compared to only nine PGHs that had previously been described in the reference strain. It is to be noted that, since the annotation of PGHs is based on the 9 previously described PGHs some distinct genes received the same annotation. 6 PGHs were found to be common to all 6 strains (annexes).

Genes present only in the B phenotype strains or the ones that are common between phenotype A and B are not discussed here, since they are no candidate-genes underlying phenotype A. Genes that are present only in active phenotype strains are represented in Table 2.

Table Genes that are present only in active phenotype strains

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group number** | **Gene name in ref.** | **Size of the gene and proteins** | **Presence in FAM19191** | **Presence in FAM23285** | **Presence in FAM8102** |
| group\_2319 | Lhv\_2053 Lysin(L.crispatus) pseudogene inL.helveticus | 1121 nucleotides / 41 kDa | **X** | **X** | **X** |
| group\_2343 | Lhv\_2053 Lysin (L.crispatus) pseudogene in L.helveticus | 893 nucleotides/ 33 kDa | **X** | **X** | **X** |
| group\_2663 | Lhv\_2053 Lysin(L.crispatus) pseudogene inL.helveticus | 1121 nucleotides /41 kDa |  | **X** | **X** |

The putative PGH present in group 2663 does not appear in all three strains and is therefore rejected as a candidate. Then, the gene common to group 2319 translated to a protein of a calculated size of about 42 kDa, which does not correspond to the size of the PGH identified within the zymogram. Thus, the last remaining candidate is the gene common to group 2343 that meets the protein size and presence criteria. Note that all three genes, even though different, were annotated the same.

In order to highlight the conservation of the amino acid sequence of the candidate PGH, an alignment of the region between the three active strains is shown in Fig. 2.

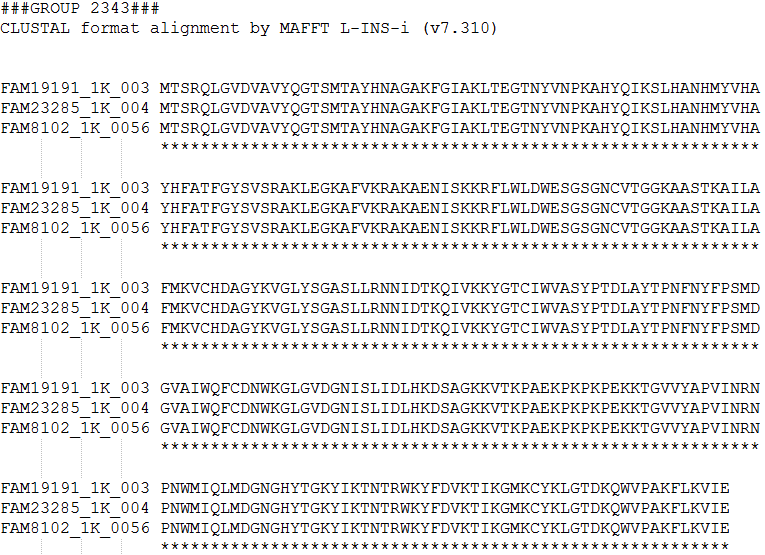


Fig. Alignment of group 2343 amino acids between the three active strains (FAM19191, FAM23285, FAM8102)

Indeed, the three strains have exactly the same amino acid sequence, which reflects a good conservation between the strains.

To understand more about the history of the candidate gene, its genomic context was analyzed with MAUVE. (see Fig. 3)

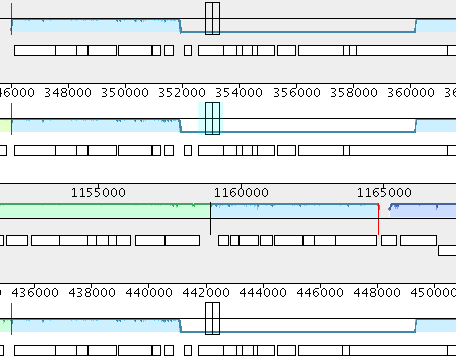


Fig. DNA sequences represented by MAUVE of FAM23285, FAM8102 BALBLABKL and FAM19191, respectively.

The two first and last sequences represent FAM23285, FAM8102 and FAM19191 genomes, respectively. The location of the cursor shows the candidate gene. Before and after that, there is a whole series of genes annotated as “hypothetical proteins” identical between these three active strains but absent in the others.

## Discussion

Since PGHs were present in all strains, it is not possible that phenotype B is a consequence of the absence of PGH. The difference between these two phenotypes is therefore likely due to a specific PGH.

The Roary result identified a gene that was only present in strains exhibiting phenotype A and whose molecular weight matched that of the gel's PGH. This is the group 2343 gene annotated as a lhv\_2053 lysine.

Since the genomic context of the group 2343 revealed a whole series of identical genes annotated as “hypothetical proteins”, it suggests that the sequence was inserted by a phage. In order to test this theory, a further analyze could be performed with Phaster, which allows to identify and annotate prophage sequences within bacterial genomes or plasmids (Arndt et al., 2016). Also, if the this theory would prove to be true, it would be very interesting to analyze how actively this gene is transcribed.

Apart from the PGH responsible for the phenotype A within the zymogram, several other new putative PGHs were found in the draft genomes of the 6 *lactobacillus helveticus* strains studied in the present report. For further investigation, these genes should be characterized in more detail: are they pseudogenes or actively transcribed and translated to functional proteins? Why do *lactobacillus helveticus* strains have so many genes apparently encoding for PGHs? Where do all of these PGHs come from? The answers to these questions could lead the way to a more comprehensive understanding of the bacteria used in cheese production.

## References

Bruggmann (2018). DNA Sequence Analysis Course Introduction Cheese Manufacturing.

Jebava I, Plockova M, Lortal S, Valence F (2011) The nine peptidoglycan hydrolases genes in Lactobacillus helveticus are ubiquitous and early transcribed. International Journal of Food Microbiology 148(2011):1–7. doi:10.1016/j.ijfoodmicro.2011.04.015.

VBC Victorian Bioinformatics Consortium (November 2018). Retrieved from <http://www.vicbioinformatics.com/software.prokka.shtml>.

[Arndt, D., Grant, J., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res., 2016 May 3.](http://www.ncbi.nlm.nih.gov/pubmed/27141966)

## Annexes

Table : PGHs in common with all strains.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **group\_1002** | **Lhv\_0191 Amidase** | **FAM1213\_1K\_00726** | **FAM1450\_1K\_00303** | **FAM19191\_1K\_00506** | **FAM22076\_1K\_00064** | **FAM23285\_1K\_00566** | **FAM8102\_1K\_00638** |
| **group\_1107** | Lhv\_0549 N-acetylmuramidase | FAM1213\_1K\_01208 | FAM1450\_1K\_00785 | FAM19191\_1K\_01147 | FAM22076\_1K\_00934 | FAM23285\_1K\_01072 | FAM8102\_1K\_01185 |
| **group\_1221** | Lhv\_1433 Lysin | FAM1213\_1K\_01854 | FAM1450\_1K\_00044 | FAM19191\_1K\_01884 | FAM22076\_1K\_01582 | FAM23285\_1K\_01903 | FAM8102\_1K\_01986 |
| **group\_3368** | Lhv\_0649 Lysozyme | FAM1213\_1K\_00886 | FAM1450\_1K\_00838 | FAM19191\_1K\_01232 | FAM22076\_1K\_00917 | FAM23285\_1K\_01191 | FAM8102\_1K\_01268 |
| **group\_850** | Lhv\_1295 Enterolysin M23 family peptidase | FAM1213\_1K\_00072 | FAM1450\_1K\_01113 | FAM19191\_1K\_00150 | FAM22076\_1K\_00164 | FAM23285\_1K\_00217 | FAM8102\_1K\_00225 |
| **group\_873** | Lhv\_1059 LysM peptidoglycan-binding domain-containing protein | FAM1213\_1K\_00184 | FAM1450\_1K\_00238 | FAM19191\_1K\_00248 | FAM22076\_1K\_00274 | FAM23285\_1K\_00308 | FAM8102\_1K\_00381 |
| **group\_969** | Lhv\_1433 Lysin | FAM1213\_1K\_00563 | FAM1450\_1K\_01203 | FAM19191\_1K\_01800 | FAM22076\_1K\_00088 | FAM23285\_1K\_01748 | FAM8102\_1K\_01891 |
| **group\_1833** | Lhv\_2053 Lysin (L.crispatus) pseudogene in L.helveticus |  | FAM1450\_1K\_00045 | FAM19191\_1K\_01885 | FAM22076\_1K\_01583 | FAM23285\_1K\_01904 | FAM8102\_1K\_01987 |
| **group\_1870** | Lhv\_2053 Lysin (L.crispatus) pseudogene in L.helveticus |  | FAM1450\_1K\_00267 | FAM19191\_1K\_00615 | FAM22076\_1K\_00716 | FAM23285\_1K\_00607 | FAM8102\_1K\_00746 |
| **group\_1335** | Lhv\_1307 Enterolysin M23 family peptidase |  |  | FAM19191\_1K\_00162 | FAM22076\_1K\_00152 | FAM23285\_1K\_00229 | FAM8102\_1K\_00237 |
| **group\_1336** | Lhv\_0190 N-acetylmuramidase |  |  | FAM19191\_1K\_00507 | FAM22076\_1K\_00063 | FAM23285\_1K\_00565 | FAM8102\_1K\_00639 |