# Improving Gene Model Accuracy for Nitrogen-Use Efficiency Genes in *Zea mays*

by

Russell Walden

A thesis presented to the Honors College of Middle Tennessee State University in partial fulfillment of the requirements for graduation from the University Honors College

Fall 2020

Improving Gene Model Accuracy for Nitrogen-Use Efficiency Genes in *Zea mays*

by Russell Walden

APPROVED:

_____

Dr. Rebecca Seipelt-Thiemann, Thesis Director
Professor, Biology Department

Dr. John Dubois, Second Reader
Professor, Biology Department

_____

Dr. Dennis Mullen, Thesis Committee Chair
Chair, Biology Department

## ACKNOWLEDGEMENTS

I'd like to thank Dr. Rebecca Seipelt-Thiemann for her guidance through the research for and the writing of this thesis. Dr. Seipelt was always incredibly patient, kind, and helpful whenever I needed guidance, and I would like to thank her for that. I'd also like to thank my wonderful parents, my brother and sister Gabriel and Katerina, my grandmother Lynn Russell (Bom Bom), my girlfriend Jerrica and her parents James and Kelly Voyles and her brother and sisters Wyatt Voyles, Katie and Michael Hinson, and Alex and Mark Vibbert, my Aunt Dana, Uncle Frank, Aunt Candy, my cousins Brandon and Jodi Russell and their sons Kohan, Walker, and Ashton, my friend Hank Clement, and my roommate and friend Jack Shreeve for all of their support as well.

## ABSTRACT

*Zea mays* is one of the most highly produced crops in the world. It plays a big role in food security, fuel production, and economic stability. A factor that contributes to the production cost and yield of *Zea mays* is its nitrogen-use efficiency. At least twenty genes in *Zea mays* are related to nitrogen-use efficiency. The purpose of this study was to improve the accuracy of existing models of nine genes relating to nitrogen-use efficiency in *Zea mays* by using data present in the Apollo gene annotation platform to inform changes required for updated gene models. These changes were made and used to produce a supertranscript for each gene. Alignment of proteins encoded by all possible transcripts was performed to identify differences in protein structure and domain presence where applicable. These models can be used to provide insight into the gene regulation and protein isoform functio

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## Introduction

Grains such as corn, rice, and wheat play important roles in food security, fuel production, and economic stability. Corn is the world's largest grain crop with world production of maize in 2013-14 at 967 million metric tons (Shah *et al.* 2016). It is considered a staple food in numerous parts of the world and is the third leading crop in the world after rice and wheat (Shah *et al.* 2016). Approximately 25% of U.S. corn croplands are used for ethanol production (Mumm *et al.* 2014). Corn distillers' oil is a by-product of starch and ethanol production in corn and this oil can be used to synthesize biodiesel (Veljković *et al.* 2018). In relation to the economy, the B73 maize reference sequence promises to advance basic research and to facilitate efforts in an era of global climate change (Schnable *et. al.* 2009).

One nutrient that is critical for development and growth of many agriculturally important plants, including *Zea mays*, is nitrogen (Sharma and Bali 2018). Corn's ability to deplete the soil of nitrogen requires addition of fertilizer to the soil, as well as crop rotation to maintain agriculture sustainability. Both of these solutions have accompanying problems. Crop rotation reduces annual yield and addition of fertilizers has been shown to have negative environmental impacts, such as ecosystem disruption and soil acidification (Singh 2018). A new approach to improve corn growth and yield without the disadvantages is to utilize new genome-level data to investigate the metabolic and physiological aspects of corn's nitrogen use with the aim of engineering new corn varieties to use less nitrogen (Simons *et al.* 2014).

To understand how corn uses nitrogen, metabolic and physiological pathways, including the genes that encode the enzymes within these pathways, must be accurately

understood at the genetic, transcript, protein, and regulatory levels. In 2009, the corn genome was sequenced, which contains over 32,000 predicted genes, and resulted in the B73 reference genome version 1 (B73 reference genome version 5 was used as reference data in this study) (Schnable *et. al.* 2009). 32,540 protein-encoding genes were predicted from assembled or improved bacterial artificial chromosome (BAC) contigs by a combination of evidence based and ab initio approaches (Schnable *et. al.* 2009). Maize was also found to exhibit extremely high levels of both phenotypic and genetic diversity (Schnable *et. al.* 2009). Extensive structural variation, including hundreds of copy number variants (CNVs) and thousands of present-absent variants (PAVs) were revealed by resequencing and array-based comparative genomic hybridization between the B73 and Mo17 inbred lines (Schnable *et. al.* 2009).

While the corn genome has been sequenced, and the genes computationally predicted and revised, transcript evidence showing how and when gene products are produced in different corn tissues is just now becoming available (Ware, personal communication) (Xu *et al.* 2009). Recent advances in RNA analyses called RNA sequencing have enable scientists to assay the entire pool of RNAs produced within a particular tissue or under a distinct condition, the transcriptome. By comparing the RNAs produced across tissues, developmental stages, and conditions, a full picture of how each gene's structure could differ at the RNA level though alternative promoter use, alternative splicing, alternative transcription termination, and alternative polyadenylation can be identified. Additionally, this information would then form the basis for regulatory studies to determine how these transcripts are generated, the protein isoforms encoded, and the functional role each plays in corn life. This, in turn, could then be used to

identify points in metabolic pathways that could be better engineered for more efficient or different use of the plant's resources.

The goal of this study was to use existing genome, gene, and tissue-specific RNA evidence to develop supertranscript gene models for nine genes that are known to be involved in nitrogen use efficiency, nitrogen assimilation, amino acid metabolism, and auxin signaling which would then be tested experimentally and corrected within the community-based Apollo genome annotation platform (Tello-Ruiz *et al.* 2017). This study contributes to research to determine the genomic structure, expression, and regulation of all corn genes, but especially the nitrogen-use efficiency (NUE) genes which may then be targeted for genetic based improvements to corn agriculture sustainability.

## Materials and Methods

### *NUE Genes*

The names and genome coordinates of a selection of 20 genes relating to nitrogen use efficiency, nitrogen assimilation, amino acid metabolism, and auxin signaling in *Zea mays* genome version 5 (Table 1) were provided by Dr. Doreen Ware of Cold Spring Harbor Laboratories (personal communication).

### *Data Viewer*

The *Zea mays* version 5 (v5) gene models (Figure 1), as well as multiple short and long-read RNA evidence tracks including: RNA sequencing data for six tissues, each with six replicates, IsoSeq RNA sequencing, and full-length RNA sequencing were visualized within the Apollo genome viewer and annotation platform. Access was provided by Cold Spring Harbor Laboratory (CSHL) (Ware, personal communication). Briefly, different gene features are represented visually in this user space (Figure 1A). Within the gene models, exons are denoted by rectangles, while introns and intergenic spaces are denoted by horizontal lines. The direction of transcription is noted with an arrowhead at the 3' end of the terminal exon. Within transcriptome data, a single read aligned to its corresponding genome sequence is positioned below the gene model in exactly the position identified by genome sequence. Multiple reads aligning to the same region are stacked vertically to show "depth of coverage" for a particular genomic region. Because alternative splicing joins non-contiguous sequences, for an individual read, the sequence that is actually present in the read is noted as a rectangle, and the implied sequence is noted as a horizontal line.

**Table 1. Summary of twenty genes involved in nitrogen-use efficiency in *Zea mays*.** Gene coordinates and functions are from Gramene (Tello-Ruiz *et al.* 2018).

| | Gene ID# | Gene Coordinates | Function/Description |
|---|---|---|---|
| 1 | Zm00001d031769 | chr1:203089608..203094203 | Nitrate reductase [NADH] 2 |
| 2 | Zm00001d049995 | chr4:58905791..58910442 | Nitrate reductase |
| 3 | Zm00001d018206 | chr5:219126820..219129945 | Nitrate reductase [NADH] 2 |
| 4 | Zm00001d048050 | chr9:152280180..152287969 | Glutamine synthetase 3 isoform 1%3B Glutamine synthetase 3 isoform 2 |
| 5 | Zm00001d028260 | chr1:27922075..27924590 | Glutamine synthetase 6 |
| 6 | Zm00001d017958 | chr5:213469469..213473069 | Glutamine synthetase root isozyme 3 |
| 7 | Zm00001d022388 | chr7:180078625..180096312 | Ferredoxin-dependent glutamate synthase%2C chloroplastic |
| 8 | Zm00001d011610 | chr8:155140627..155152027 | Glutamate synthase 1 [NADH] chloroplastic |
| 9 | Zm00001d043845 | chr3:213907196..213918549 | Glutamate synthase 1 [NADH] chloroplastic |
| 10 | Zm00001d038948 | chr6:174758495..174770022 | Glutamate synthase 1 [NADH] chloroplastic |
| 11 | Zm00001d038948 | chr4:183572732..183576346 | Nitrite reductase 2 |
| 12 | Zm00001d018161 | chr5:218273575..218276598 | Ferredoxin--nitrite reductase chloroplastic |
| 13 | Zm00001d034420 | chr1:294101398..294107663 | Glutamate dehydrogenase |
| 14 | Zm00001d025984 | chr10:136652035..136656257 | Glutamic dehydrogenase 2 |
| 15 | Zm00001d002052 | chr2:5280850..5283996 | Probable isoaspartyl peptidase/L-asparaginase 2 |
| 16 | Zm00001d028750 | chr1:44908940..44912682 | Asparagine synthetase 3 |
| 17 | Zm00001d045675 | chr9:34625498..34634881 | Asparagine synthetase 1 |
| 18 | Zm00001d022152 | chr7:174854238..174860476 | Alanine aminotransferase 9 |
| 19 | Zm00001d014258 | chr5:38654080..38663968 | Alanine aminotransferase 5 |
| 20 | Zm00001d007357 | chr2:229098421..229111535 | Protein Auxin Signaling F-Box 3 |

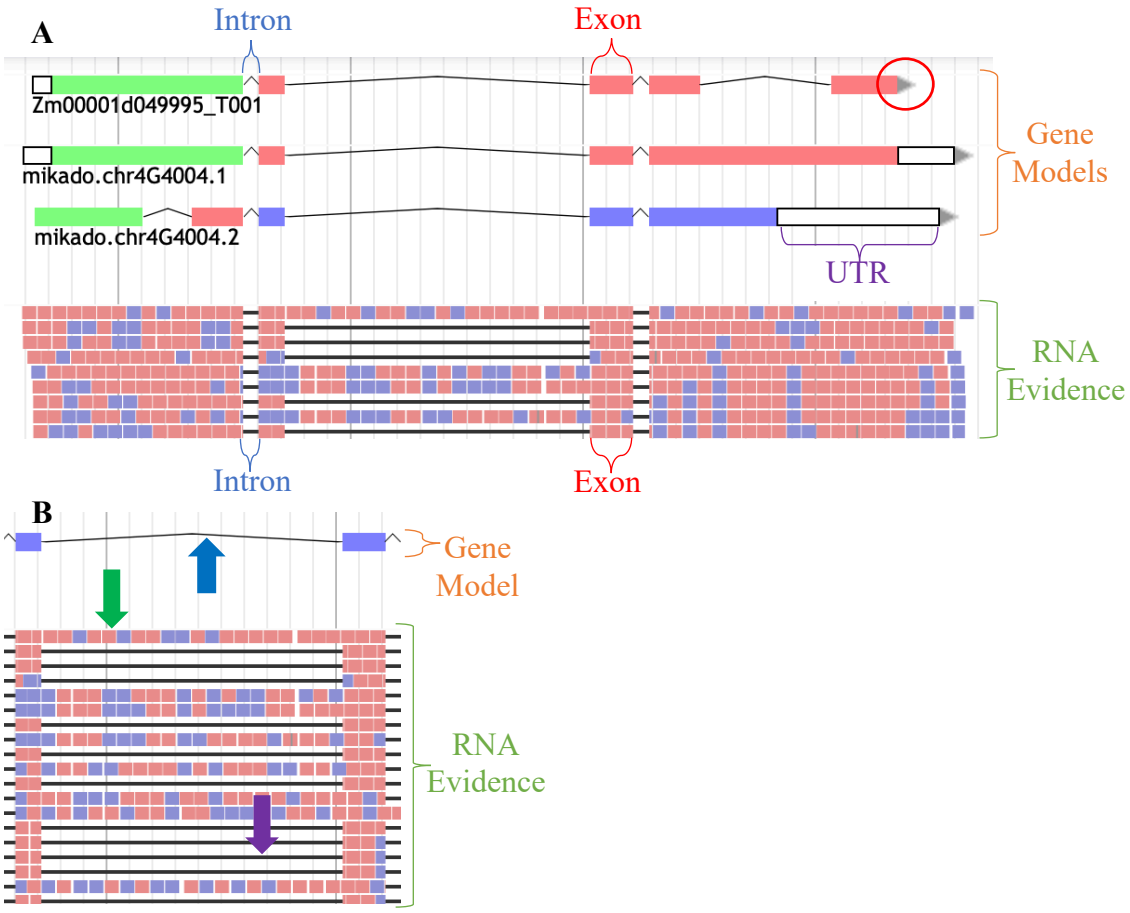
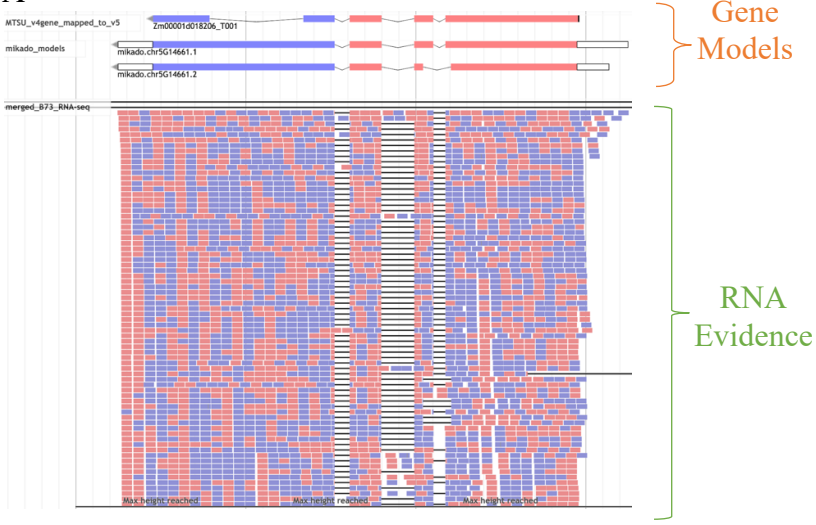



**Figure 1. Example of gene structure models and RNA sequence data in the Apollo genome annotation platform.** Panel A shows an example of gene structure models and RNA sequence data. Introns are denoted in both the gene models and in the RNA sequence data by black lines (Blue bracket). Exons are denoted by colored boxes (Red bracket). Gene models are at the top (Orange bracket) and have been produced previously using RNA evidence data (Green bracket). Boxes filled with white denote untranslated regions (UTR, purple bracket). Arrows show the direction of transcription for the model (Red circle) which means the arrow is at the end of the transcript and is at the 3' end of the model. In panel B, the model shows presence of an intron between two exons (blue arrow). However, some RNA evidence suggests the presence an exon (green arrow) and some evidence suggests the presence of an intron (purple arrow).

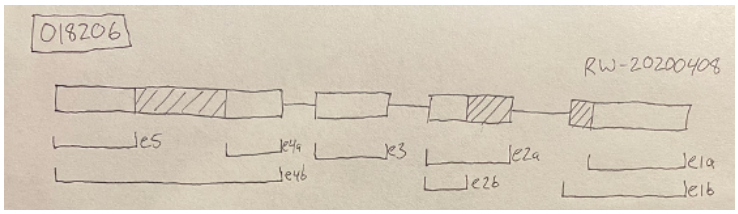
***Data Analysis and Consolidation***

A collapsed gene model (or SuperTranscript) (Davidson *et al.* 2017) was created for each gene (Figure 2). First, the PASA-informed v5 gene model was compared to the RNA evidence in the consolidated Mikado RNA evidence, which represents RNA evidence from six replicates of six different corn tissues (Figure 2A). First, existing genome data in the Apollo genome annotation platform was compared to models to first determine the model accuracy and then to look at evidence that changes were necessary. Next, aligned RNA sequencing reads were visualized one set at a time to determine gather evidence for the gene features represented in the v5 model and any features not represented in the v5 model (Figure 2B). Evidence was in the form of aligned transcript reads from RNA sequencing experiments and included large reads (IsoSeq) and short reads from six tissues, as well as full-length RNAs. Possible changes include: retained intron, alternative 5' splice site, alternative 3' splice site, alternative exon, and alternative intron. Next, the gene features with adequate evidence in any of the RNA data was collapsed visually to form the SuperTranscript with exonic features noted as rectangles, alternative exonic regions noted as hatched rectangle areas, introns noted as horizontal lines, and exons numbered beginning with 1 at the first transcribed nucleotide (Figure 2C). This SuperTranscript was a diagrammatic representation of all transcript evidence combined with the new PASA-informed v5 gene model. It essentially allowed for visualization of all genomic regions that were retained in the mature mRNA for all transcripts in all tissues where data were available. All possible transcripts were then constructed with the Apollo user space to be available to the research community (Figure 2D).

A




B




C

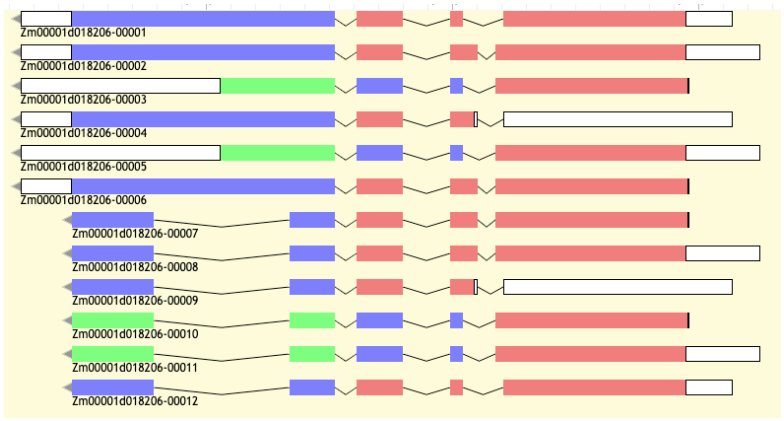



**Figure 2 – Example of workflow used to create improved gene models for gene Zm000018206** – Panels A shows some of, but not all of the data used as evidence for reannotation of gene models. Previously compiled gene models are shown (orange bracket) with RNA sequencing data underneath (green bracket). Panel B shows an example of a hand-drawn collapsed gene model, or SuperTranscript, A SuperTranscript is a visual representation of the similarities and differences between all possible transcripts. Hashed boxes denote features that differ between possible transcripts. Panel C shows an example of transcripts constructed in the Apollo user-created annotations panel. This shows all the ways RNA could be constructed by the cell and was constructed to produce the amino acid sequence of each transcript to generate protein alignments.

## Results

The first gene that was evaluated was Zm00001d031769, which encodes nitrate reductase [NADH] 2 (Tello-Ruiz *et. al.* 2017). The original gene structure was compared with the data in the Apollo annotation platform and a model was produced as noted in the methods (Figure 3). Comparison of the v5 and Mikado RNA models showed the extension of the gene with addition of an intron and exon, as well as an alternative 5' splice site (Figure 3A). RNA sequencing evidence (Figure 3B) supported both the presence and absence of an intron at that location, so it was labeled as a retained intron in the revised model (*, Figure 3C, Table 2). Additionally, the data supported the presence of the additional 3' terminal exon (Figure 3A, 3B), so this was labeled as an alternative exon in the revised model (**, Figure 3C; Table 2). Finally, a new 5' splice site was identified in the Mikado RNA model (Figure 3A) and supported in the RNA sequencing evidence, (Figure 3B) and incorporated into the SuperTranscript (***, Figure 3C).

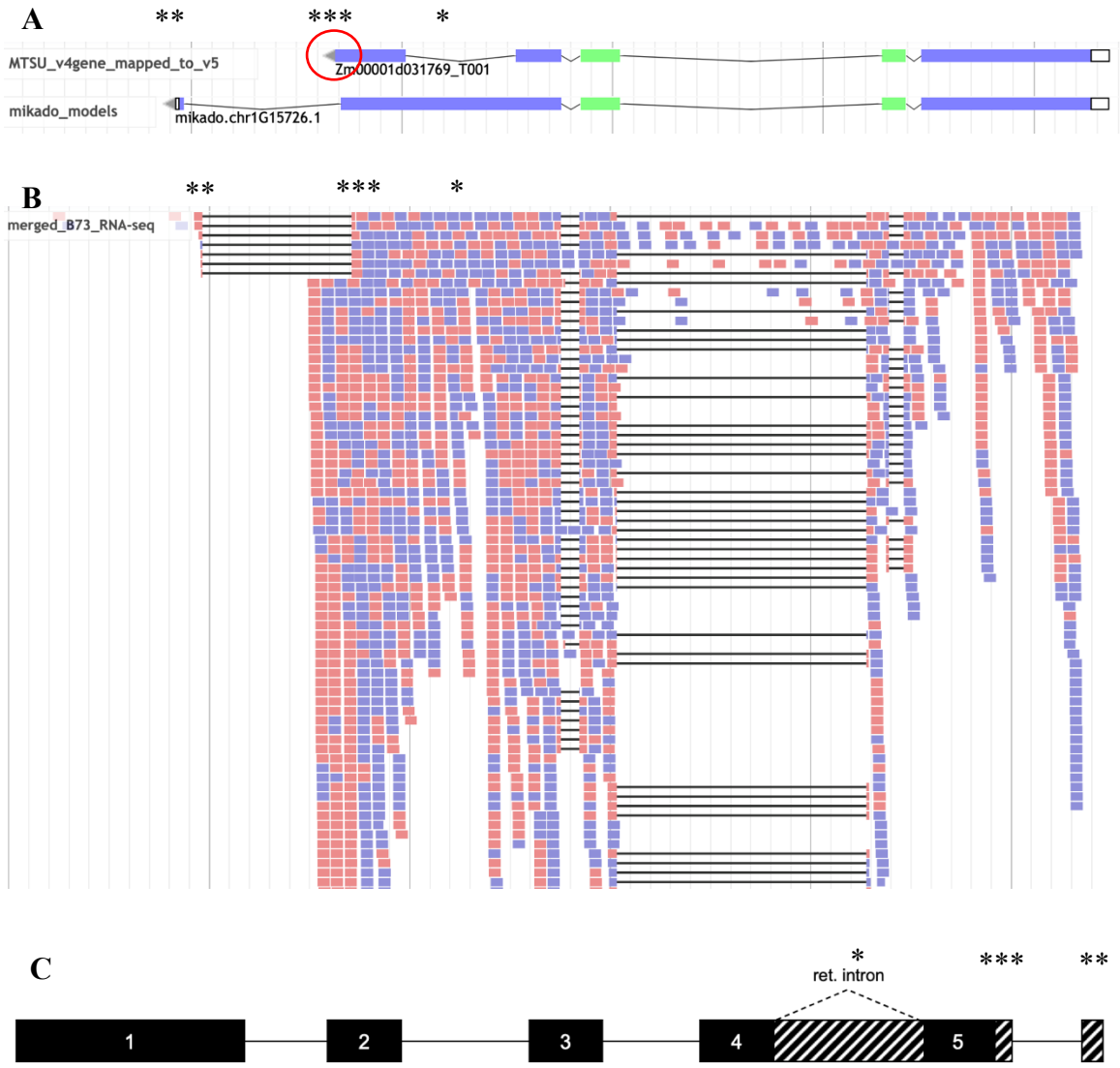



**Figure 3. Evidence for reannotation for gene Zm00001d031769.** Panel A shows the original structure of gene Zm00001d031769. Notice that the arrow on the original structure in panel A indicates the 3' end is on the left side of the gene (red circle) so the direction of the gene will be the reversed in the final model. The green and purple boxes show differentiation between the gene's reading frames. Panel B shows the data that was used to determine changes to be made to produce the final model. Panel C shows the final revised model. Note that the final model shows the RNA strand 5' to 3', which is the reverse of how the original model is displayed. The numbered black boxes indicate exons and their sequence from left to right. The lines between exons indicate introns.

**Table 2. Changed features of genes.**

| Gene name | Features changed |
|---|---|
| Zm00001d031769 | Retained intron between exons 4 and 5. Alternate 5' splice site after exon 5. Alternative exon at 3' end. |
| Zm00001d049995 | Retained introns between exons 1 and 2, 3 and 4, and 5 and 6. |
| Zm00001d018206 | Alternate 5' splice site after exon 1. Alternate 3' splice site before exon 2. Retained intron between exons 4 and 5. |
| Zm00001d017958 | Alternate 3' splice sites before exons 2, 3, 4, 6, and 10. Two alternate 3' splice sites before exon 9. Alternate 5' splice sites after exons 5, 8, and 9. |
| Zm00001d022388 | Alternative (or cassette) exons between exons 9 and 12. Alternative intron inside exon 28. Alternative 5' splice site after exon 30. |
| Zm00001d052165 | Alternate start of transcription site. Retained intron between exons 1 and 2, and 2 and 3. Alternate end of transcription site after exon 5. |
| Zm00001d018161 | Alternate 3' splice site before exon 3. |
| Zm00001d025984 | Alternate 3' splice sites before exons 2 and 8. |
| Zm00001d028750 | Retained introns between exons 1 and 2, 6 and 7, and 10 and 11. Two alternate end of transcription sites with possible retained intron. |

Using the SuperTranscript as a guide, three RNAs were constructed in the Apollo space. The encoded amino acid sequences for each transcript were then downloaded from the Apollo workspace and aligned using Clustal Omega (Sievers *et. al.* 2011) and shaded using BOXSHADE to allow visualization of similarities and differences between the proteins encoded by the different transcripts. The three transcripts produce proteins that differ in the central region (Figure 4). Transcript 3 encodes a protein missing 216 amino acids compared to transcript 1. Transcript 2 encodes a protein missing 218 amino acids compared to transcript 1. Lastly, the sequence of amino acids encoded by each transcript was analyzed for the presence of known protein domains using the SMART domain sequence analysis program (Letunic and Bork 2017). The cytochrome b5-like Heme/Steroid binding domain was present in transcript 1, but not transcript 2 or 3, suggesting transcript 1, but not transcript 2 or 3, encodes a functional protein.

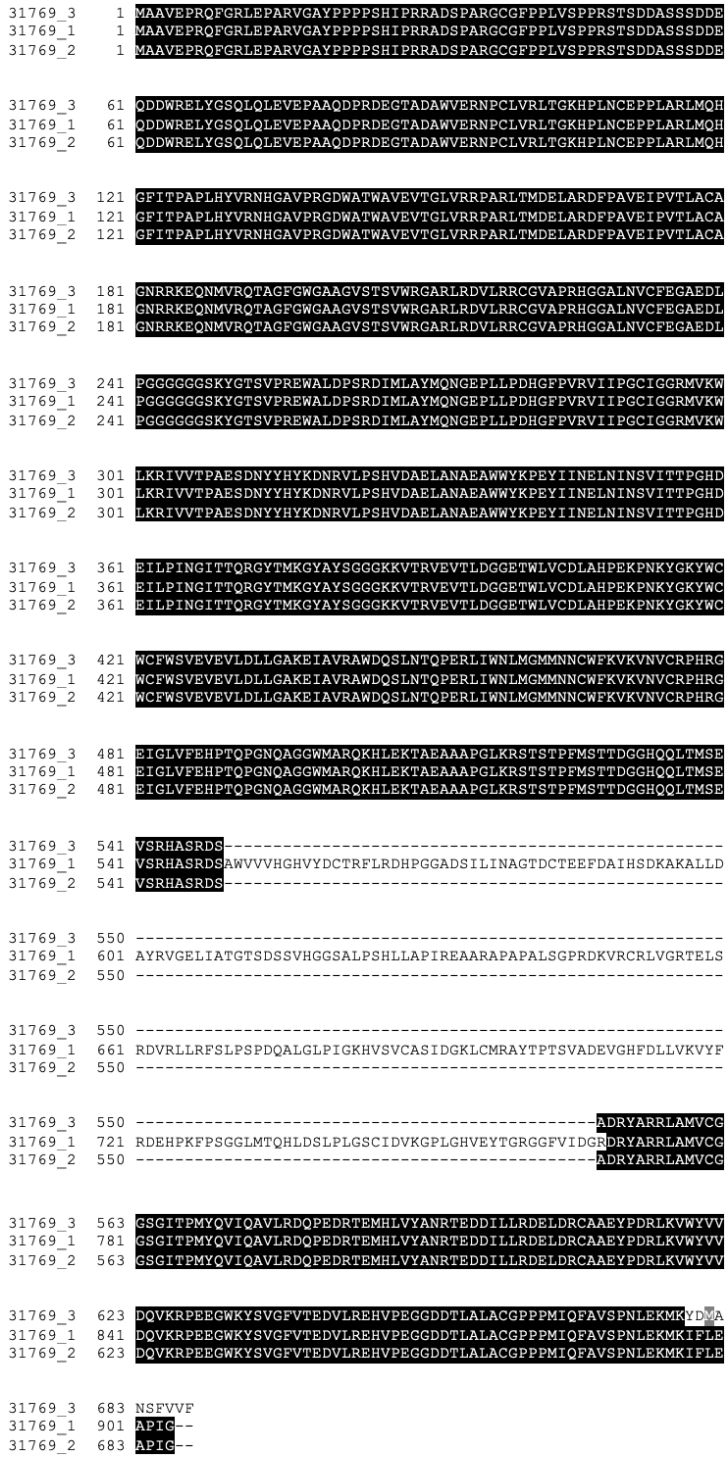



**Figure 4. Alignment of proteins encoded by Zm00001d031769.** Amino acid sequence of gene Zm00001d031769 showing three different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

The next gene that was evaluated was Zm00001d049995, which encodes nitrate reductase (Tello-Ruiz *et. al.* 2017). The original gene structure was compared with the data in the Apollo annotation platform and a model was produced as noted in the methods (Figure 5). Comparison of the v5 and Mikado RNA models showed the shortening of the gene with addition of introns (Figure 5A). Mikado models supported both the presence and absence of an intron (*, Figure 5A), and was supported by RNA sequencing data showing the absence of an intron (*, Figure 5B), so it was labeled as a retained intron in the revised model (*, Figure 5C, Table 2). Additionally, v5 evidence supported the presence of an intron (***, Figure 5A), while RNA sequencing evidence (***, Figure 5B) supported the absence of an intron at that location, so it was labeled as a retained intron in the revised model (***, Figure 5C; Table 2). Finally, RNA sequencing data suggesting both the presence and absence of an intron was found (**, Figure 5B), and was then incorporated into the SuperTranscript and labeled as a retained intron in the revised model (**, Figure 5C, Table 2).

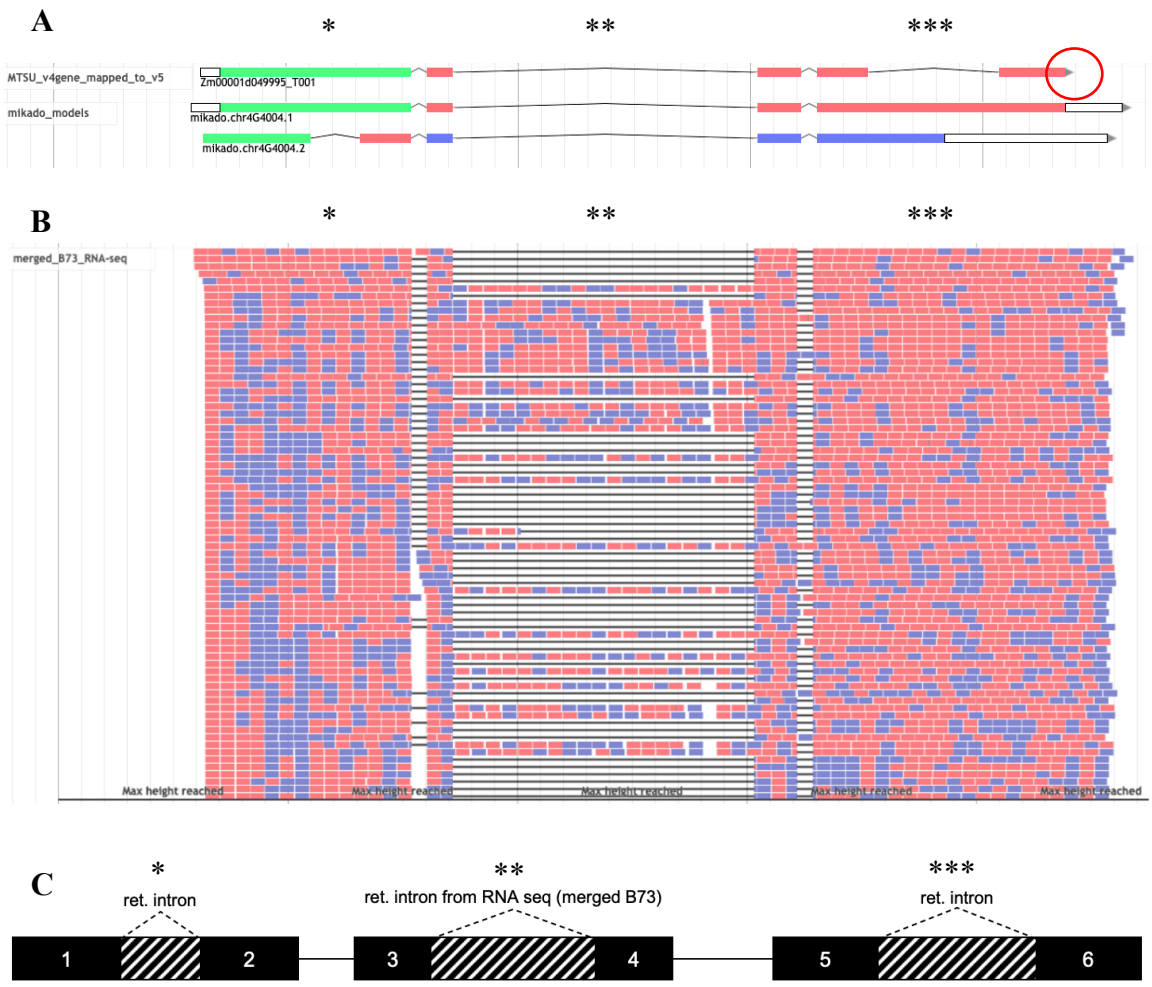



**Figure 5. Evidence for reannotation for gene Zm00001d049995.** Panel A shows the original structure of gene Zm00001d049995. Notice that the arrow on the original structure in panel A indicates the 3' end is on the right side of the gene (red circle) so the direction of the gene will be the same in the final model. The green, red, and purple boxes show differentiation between the gene's reading frames. Panel B shows the final revised model. Note that the final model shows the RNA strand 5' to 3', which is the same as how the original model is displayed. The numbered black boxes indicate exons and their sequence from left to right. The lines between exons indicate introns.

Using the SuperTranscript as a guide, eight RNAs were constructed in the Apollo space. The encoded amino acid sequences for each transcript were then downloaded from the Apollo workspace and aligned using Clustal Omega (Sievers *et. al.* 2011) and shaded using BOXSHADE to allow visualization of similarities and differences between the proteins encoded by the different transcripts. The eight transcripts produce proteins that differ in various regions (Figure 6). Transcripts 3 and 7 both encode the same protein which is 400 amino acids shorter than transcript 5. Transcript 6 encodes a protein which is 235 amino acids shorter compared to transcript 5. Transcript 2 encodes a protein missing 322 amino acids compared transcript 5. Transcript 1 encodes a protein missing 303 amino acids compared to transcript 5. Transcript 4 encodes a protein missing 608 amino acids compared to transcript 5. Lastly, the sequence of amino acids encoded by each transcript was analyzed for the presence of known protein domains using the SMART domain sequence analysis program (Letunic and Bork 2017). The cytochrome b5-like heme/steroid binding domain was found present in the transcripts.

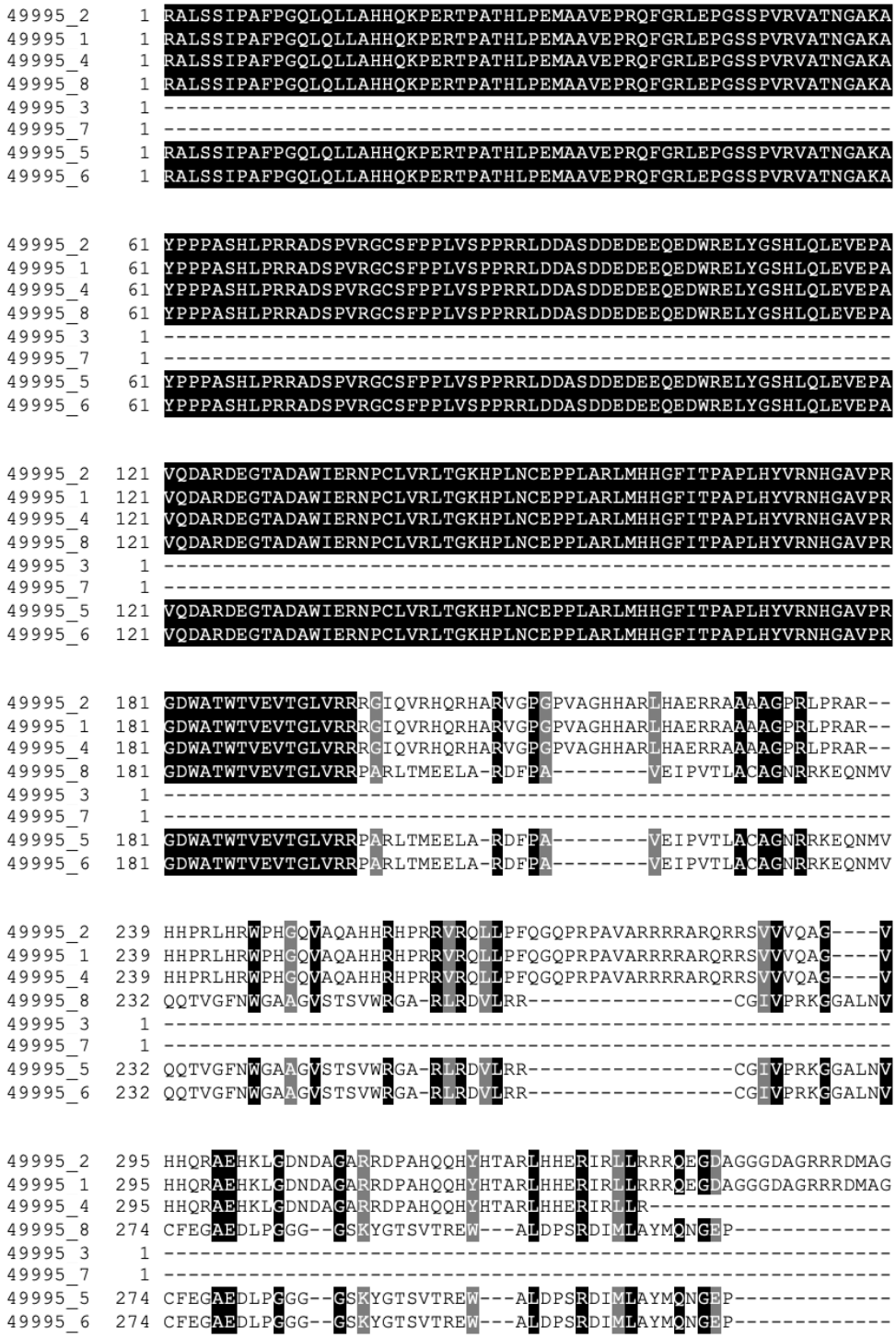



**Figure 6. Alignment of proteins encoded by Zm00001d049995 – Part 1.** Amino acid sequence of gene Zm00001d049995 showing eight different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

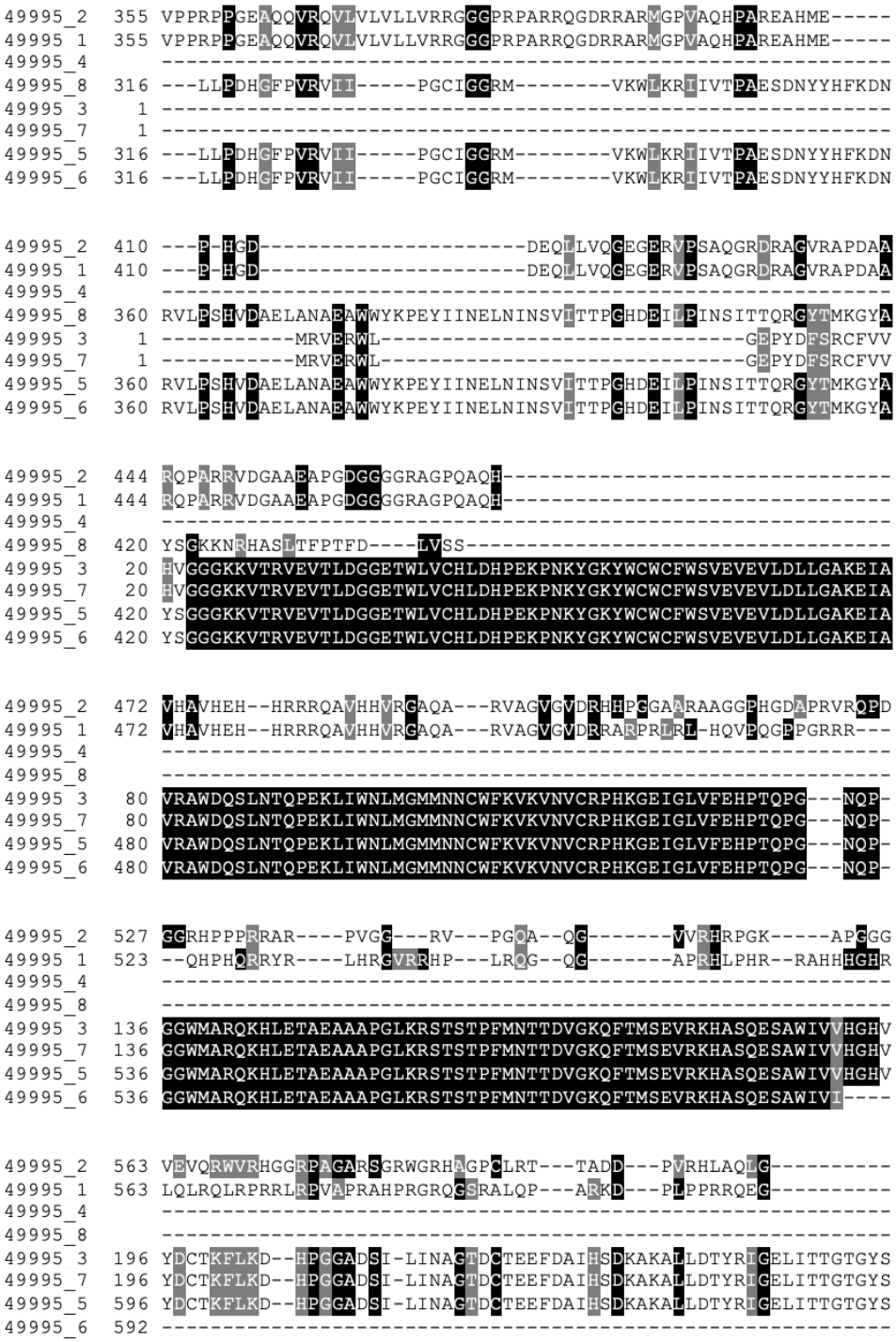


**Figure 6. Alignment of proteins encoded by Zm00001d049995 – Part 2.** Amino acid sequence of gene Zm00001d049995 showing eight different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

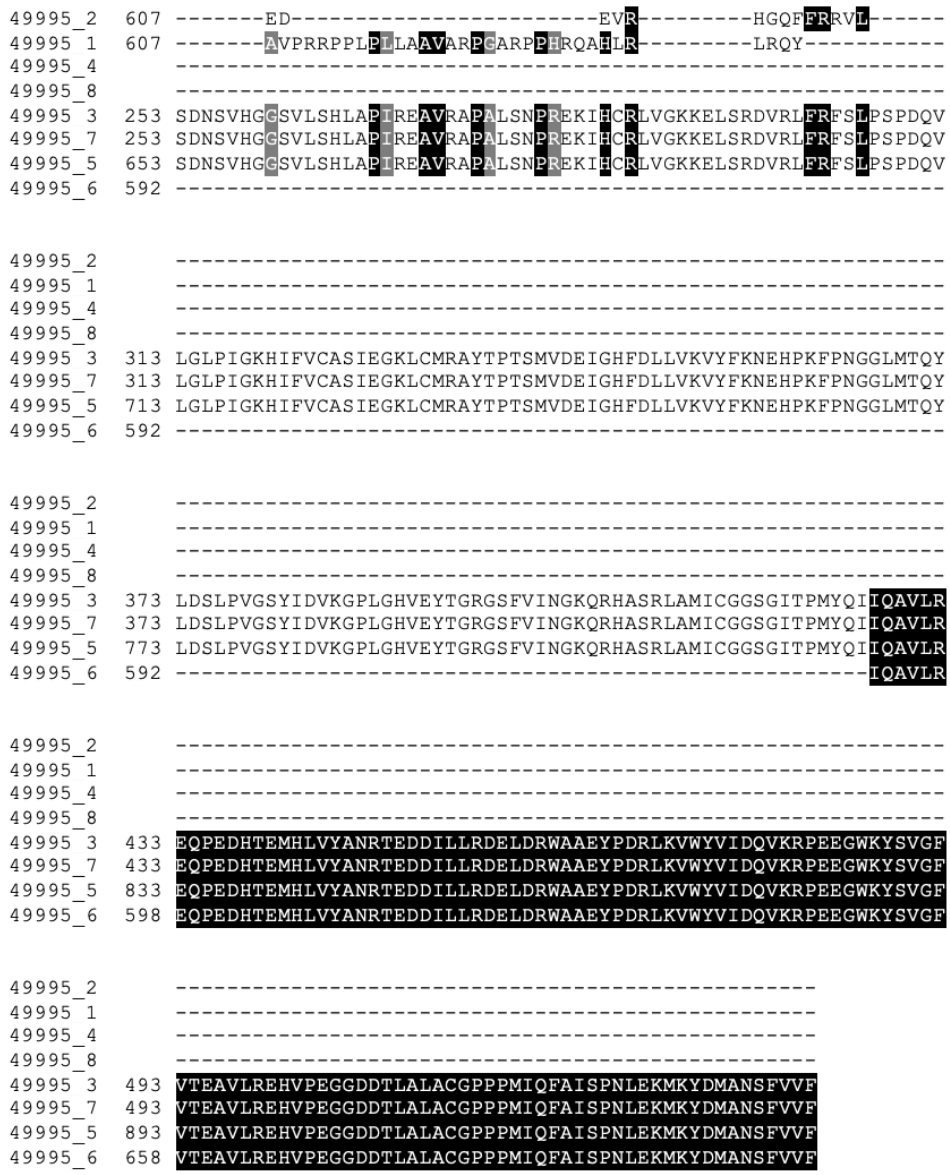



**Figure 6. Alignment of proteins encoded by Zm00001d049995 – Part 3.** Amino acid sequence of gene Zm00001d049995 showing eight different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

The next gene that was evaluated was Zm00001d018206, which encodes nitrate reductase [NADH] 2 (Tello-Ruiz *et. al.* 2017). The original gene structure was compared with the data in the Apollo annotation platform and a model was produced as noted in the methods (Figure 7). Comparison of the v5 and Mikado RNA models showed the shortening of the gene with addition of an intron, and alternate 5' and 3' splice sites (Figure 7A). The v5 model supported evidence of an intron while Mikado models supported the absence of an intron at that same location (***, Figure 7A). RNA sequencing evidence also supported the absence of an intron at that location (***, Figure 7B), so it was labeled as a retained intron in the revised model (***, Figure 7C, Table 2). Additionally, a new 5' splice site was identified in the Mikado RNA model (*, Figure 7A), supported in the RNA sequencing evidence, (*, Figure 7B) and incorporated into the SuperTranscript (*, Figure 7C; Table 2). Finally, a new 3' splice site was identified in the Mikado RNA model (**, Figure 7A) and supported in the RNA sequencing evidence, (**, Figure 7B) and incorporated into the SuperTranscript (**, Figure 7C, Table 2).

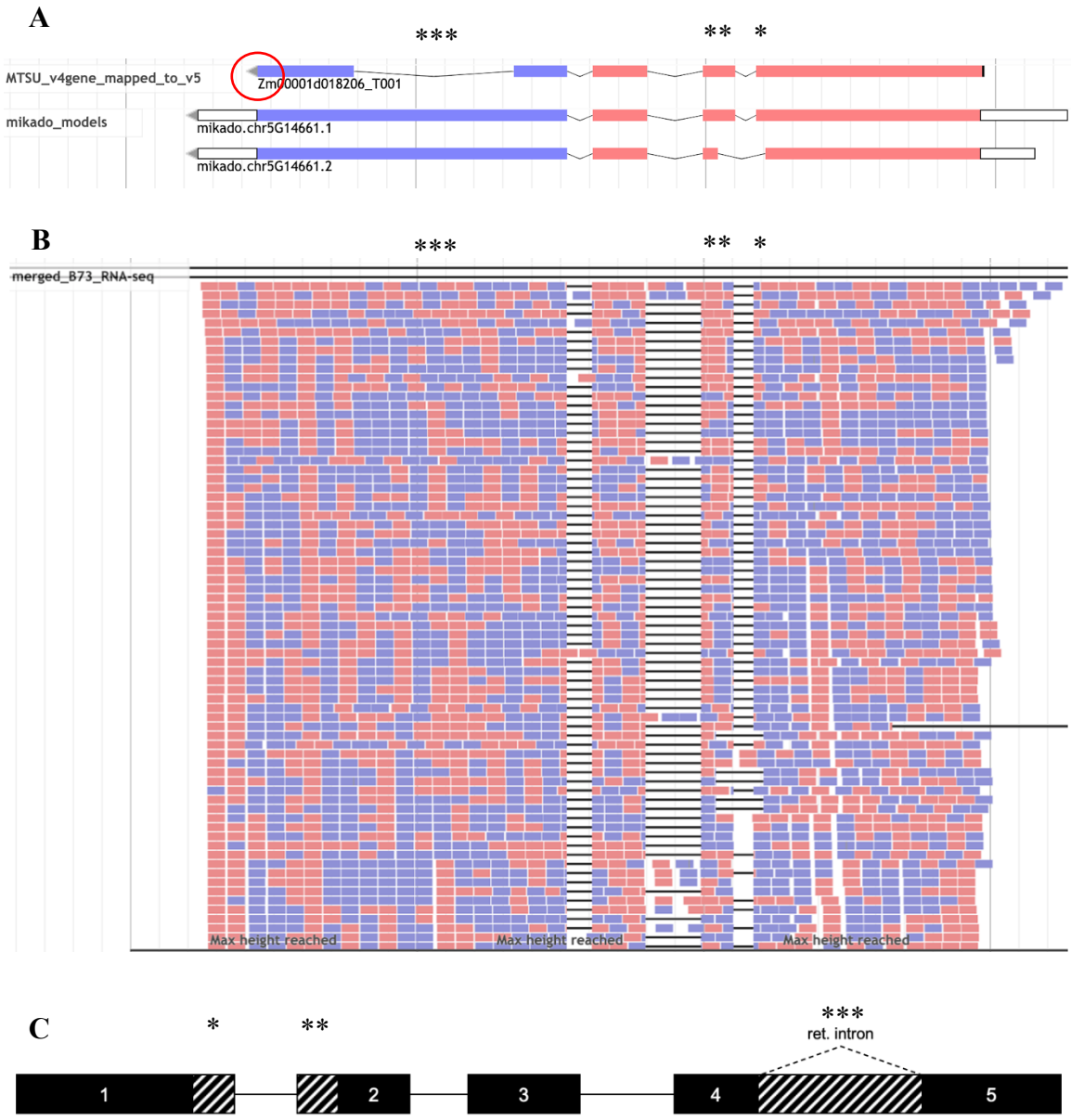



**Figure 7. Evidence for reannotation for gene Zm00001d018206.** Panel A shows the original structure of gene Zm00001d018206. Notice that the arrow on the original structure in panel A indicates the 3' end is on the left side of the gene (red circle) so the direction of the gene will be the reversed in the final model. The purple and red boxes show differentiation between the gene's reading frames. Panel B shows the final revised model. Note that the final model shows the RNA strand 5' to 3', which is the reverse of how the original model is displayed. The numbered black boxes indicate exons and their sequence from left to right. The lines between exons indicate introns.

Using the SuperTranscript as a guide, seven RNAs were constructed in the Apollo space. The encoded amino acid sequences for each transcript were then downloaded from the Apollo workspace and aligned using Clustal Omega (Sievers *et. al.* 2011) and shaded using BOXSHADE to allow visualization of similarities and differences between the proteins encoded by the different transcripts. The seven transcripts produce proteins that differ in various regions (Figure 8). Transcript 1 encodes a protein missing 44 amino acids compared to transcript 4. Transcript 7 encodes a protein missing 274 amino acids compared to transcript 4. Transcript 3 encodes a protein missing 333 amino acids compared to transcript 4. Transcript 5 encodes a protein missing 563 amino acids compared to transcript 4. Transcript 6 encodes a protein missing 255 amino acids compared to transcript 4. Transcript 2 encodes a protein missing 277 amino acids compared to transcript 4. Lastly, the sequence of amino acids encoded by each transcript was analyzed for the presence of known protein domains using the SMART domain sequence analysis program (Letunic and Bork 2017). The cytochrome b5-like heme/steroid binding domain was found present in the gene.

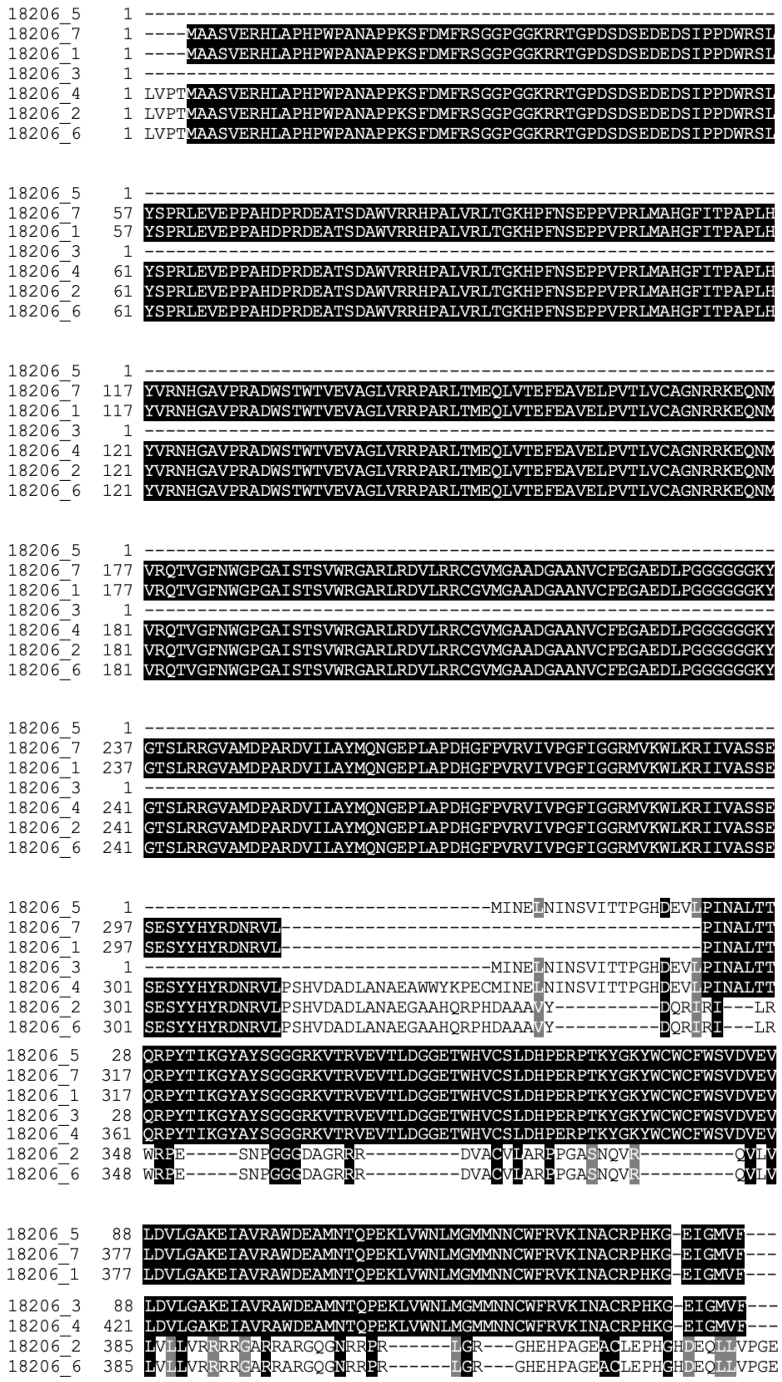


**Figure 8. Alignment of proteins encoded by Zm00001d018206 – Part 1.** Amino acid sequence of gene Zm00001d018206 showing seven different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

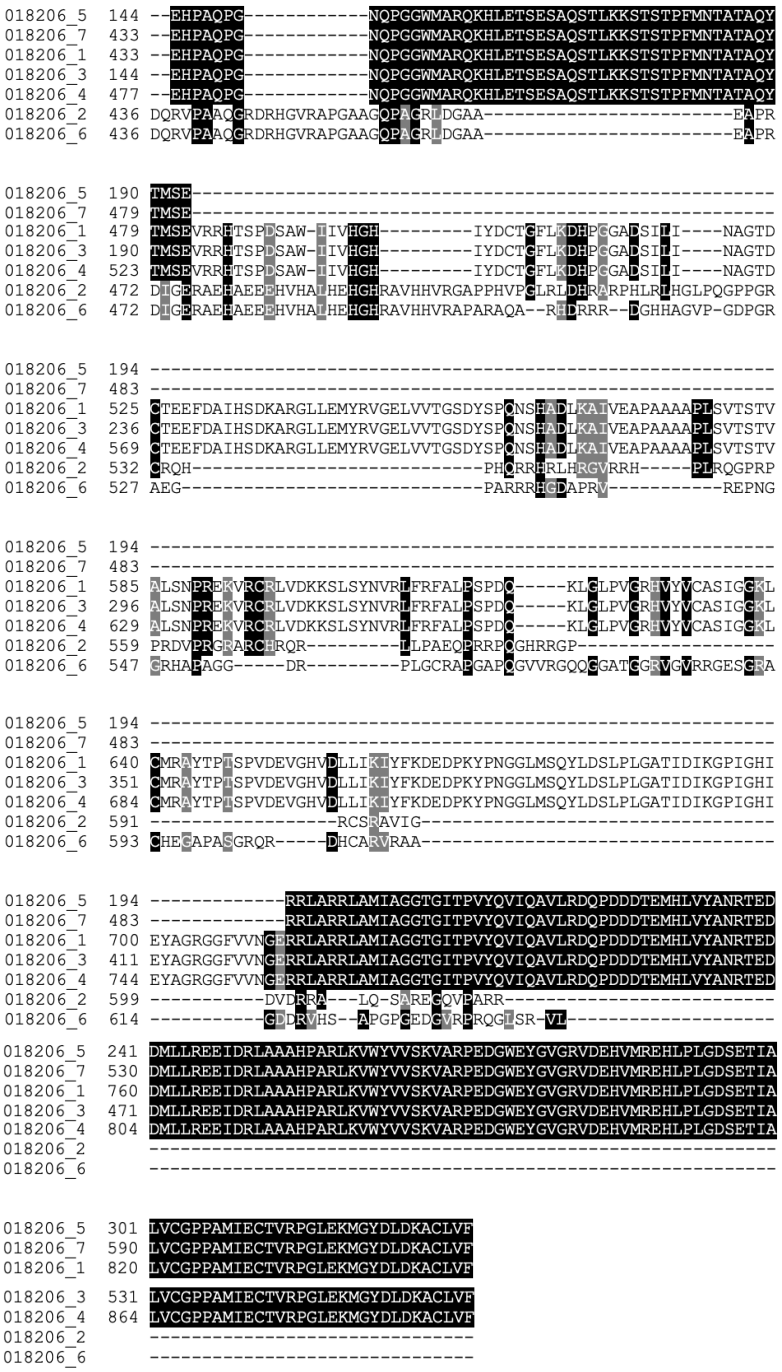


**Figure 8. Alignment of proteins encoded by Zm00001d018206 – Part 2.** Amino acid sequence of gene Zm00001d018206 showing seven different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

The next gene that was evaluated was Zm00001d017958, which encodes glutamine synthetase root isozyme 3 (Tello-Ruiz *et. al.* 2017). The original gene structure was compared with the data in the Apollo annotation platform and a model was produced as noted in the methods (Figure 9). Comparison of the v5, Mikado, IsoSeq, and full-length cDNA (flc) RNA models showed alternative 5' splice sites and alternative 3' splice sites (Figure 9A). Data from the IsoSeq model supported the presence of an alternative 3' splice site (*, Figure 9A), and its presence was supported in the RNA sequencing evidence (*, Figure 9B) so it was incorporated into the SuperTranscript (*, Figure 9C, Table 2). Data from an evidence model supported the presence of an alternative 3' splice site (**, Figure 9A), and its presence was supported in the RNA sequencing evidence (**, Figure 9B) so it was incorporated into the SuperTranscript (**, Figure 9C, Table 2). Data from the Mikado, IsoSeq, flc, and evidence models supported the presence of an alternative 3' splice site (*, Figure 9A), and its presence was supported in the RNA sequencing evidence (*, Figure 9B), so it was incorporated into the SuperTranscript (*, Figure 9C, Table 2). Data from an RNAseq Mikado model supported the presence of an alternative 5' splice site (***, Figure 9A), and its presence was supported in the RNA sequencing evidence (***, Figure 9B) so it was incorporated into the SuperTranscript (***, Figure 9C, Table 2). Data from an RNAseq Mikado model supported the presence of an alternative 3' splice site (**, Figure 9A), and its presence was supported in the RNA sequencing evidence (**, Figure 9B) so it was incorporated into the SuperTranscript (**, Figure 9C, Table 2). Data from the IsoSeq model supported the presence of an alternative 5' splice site (*, Figure 9A), and its presence was supported in the RNA sequencing evidence (*, Figure 9B) so it was incorporated into the

SuperTranscript (*, Figure 9C, Table 2). Data from the RNA sequencing evidence supported the presence of alternative 3' splice sites (***, Figure 9B), so they were incorporated into the SuperTranscript (***, Figure 9C, Table 2). Data from RNA sequencing data supported the presence of an alternative 5' splice site (**, Figure 9B), so it was incorporated into the SuperTranscript (**, Figure 9C, Table 2). Finally, a new 3' splice site was identified in the RNA sequencing data (***, Figure 9B) and incorporated into the SuperTranscript (***, Figure 9C).

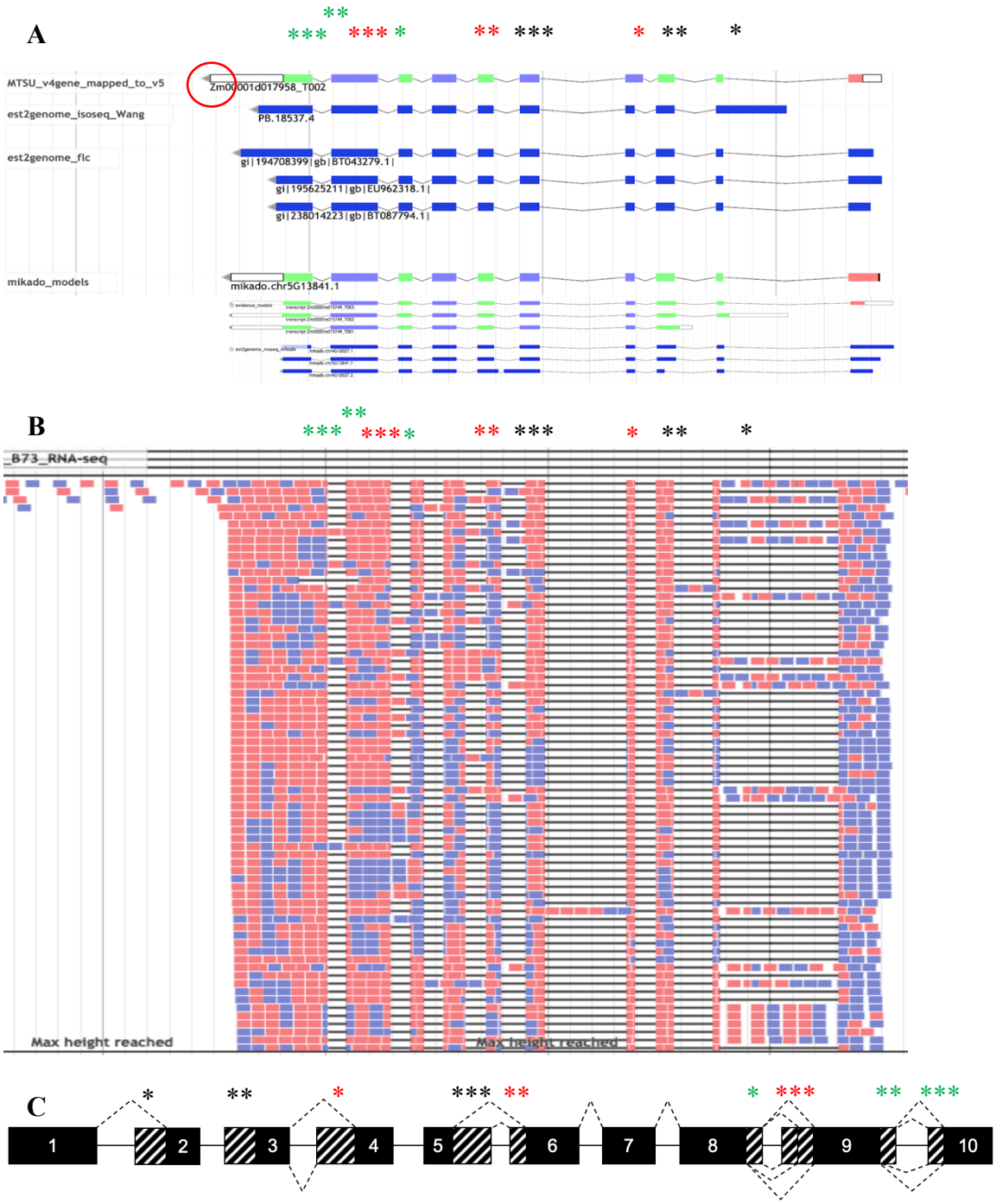



**Figure 9. Evidence for reannotation for gene Zm00001d017958.** Panel A shows the original structure of gene Zm00001d017958. Notice that the arrow on the original structure in panel A indicates the 3' end is on the left side of the gene (red circle) so the direction of the gene will be the reversed in the final model. The green and purple boxes show differentiation between the gene's reading frames. Panel B shows the final revised model. Note that the final model shows the RNA strand 5' to 3', which is the reverse of how the original model is displayed. The numbered black boxes indicate exons and their sequence from left to right. The lines between exons indicate introns.

Using the SuperTranscript as a guide, four RNAs were constructed in the Apollo space. The encoded amino acid sequences for each transcript were then downloaded from the Apollo workspace and aligned using Clustal Omega (Sievers *et. al.* 2011) and shaded using BOXSHADE to allow visualization of similarities and differences between the proteins encoded by the different transcripts. The four transcripts produce proteins that differ in a couple of regions (Figure 10). Transcripts 2 and 4 encode the same protein. Transcripts 1 and 3 also encode the same protein which is missing 17 amino acids compared to transcripts 2 and 4. Lastly, the sequence of amino acids encoded by each transcript was analyzed for the presence of known protein domains using the SMART domain sequence analysis program (Letunic and Bork 2017). The glutamine synthetase, catalytic domain was found present in the gene.


**Figure 10. Alignment of proteins encoded by Zm00001d017958.** Amino acid sequence of gene Zm00001d017958 showing four different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

The next gene that was evaluated was Zm00001d022388, which encodes ferredoxin-dependent glutamate synthase 2C chloroplastic (Tello-Ruiz *et. al.* 2017). The original gene structure was compared with the data in the Apollo annotation platform and a model was produced as noted in the methods (Figure 11). Comparison of the v5 and Mikado RNA models showed the presence of cassette exons, as well as an alternative 5' splice site (Figure 11A). The v5 model and Mikado model supported the presence of an exon, but in different locations (* and *, Figure 11A), and was supported in RNA sequencing data (* and *, Figure 11B), so they were incorporated as cassette exons in the SuperTranscript (* and *, Figure 11C, Table 2), meaning they could be present in that location or not depending on what transcript is produced. Additionally, RNA sequencing evidence supported both the presence and absence of an intron (**, Figure 11B), so it was labeled as an alternative intron in the final model (**, Figure 11C, Table 2). Finally, a new 5' splice site was identified in the Mikado RNA model (**, Figure 11A) and supported in the RNA sequencing evidence, (**, Figure 11B) and incorporated into the SuperTranscript (**, Figure 11C, Table 2).

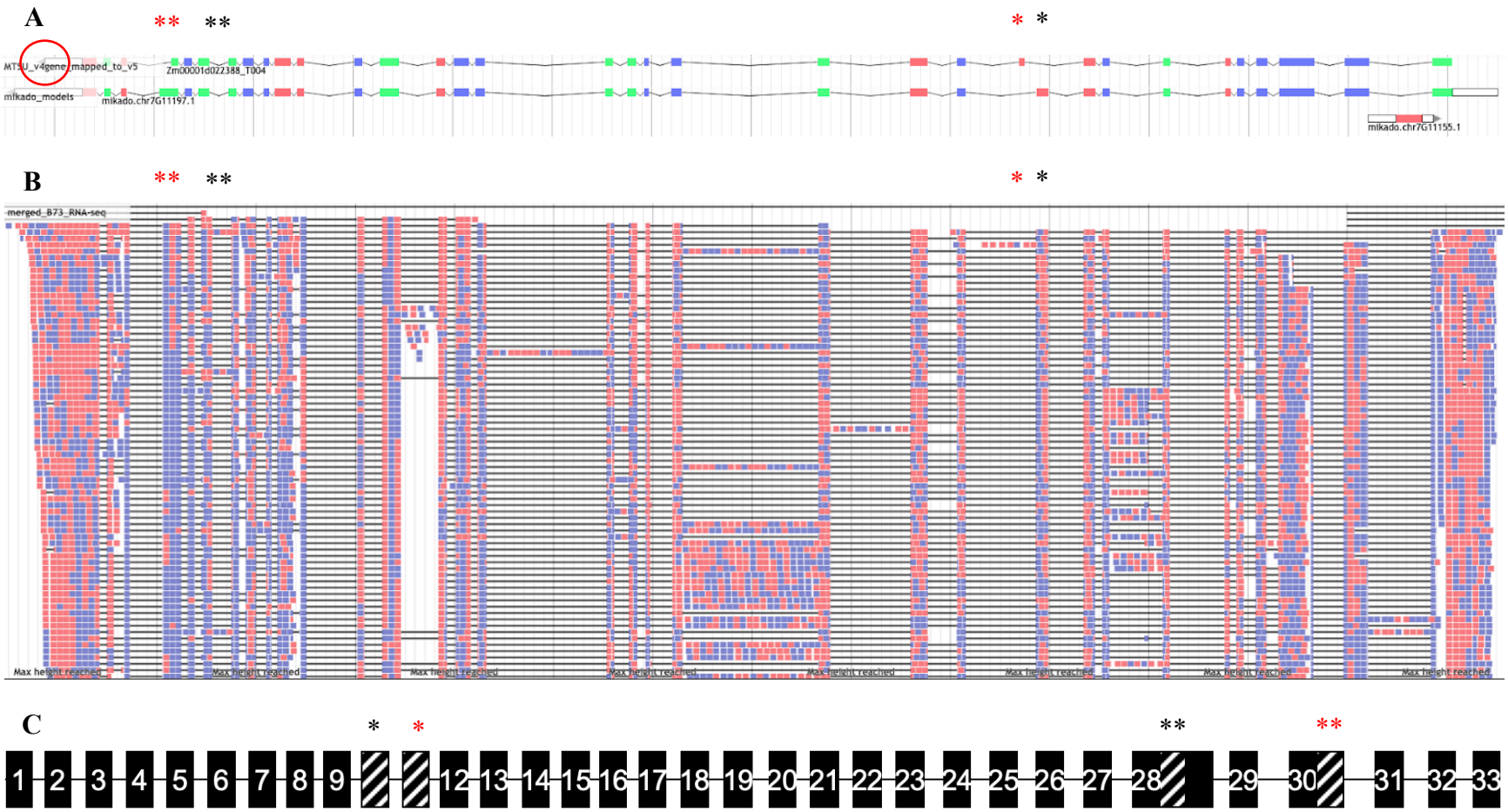



**Figure 11. Evidence for reannotation for gene Zm00001d022388.** Panel A shows the original structure of gene Zm00001d022388. Notice that the arrow on the original structure in panel A indicates the 3' end is on the left side of the gene (red circle) so the direction of the gene will be the reversed in the final model. The green and purple boxes show differentiation between the gene's reading frames. Panel B shows the final revised model. Note that the final model shows the RNA strand 5' to 3', which is the reverse of how the original model is displayed. The numbered black boxes indicate exons and their sequence from left to right. The lines between exons indicate introns.

Using the SuperTranscript as a guide, four RNAs were constructed in the Apollo space. The encoded amino acid sequences for each transcript were then downloaded from the Apollo workspace and aligned using Clustal Omega (Sievers *et. al.* 2011) and shaded using BOXSHADE to allow visualization of similarities and differences between the proteins encoded by the different transcripts. The four transcripts produce proteins that differ in a few regions (Figure 12). Transcript 4 encodes a protein missing 28 amino acids compared to transcript 2. Transcript 3 encodes a protein missing 47 amino acids compared to transcript 2. Transcript 1 encodes a protein missing 75 amino acids compared to transcript 2. Lastly, the sequence of amino acids encoded by each transcript was analyzed for the presence of known protein domains using the SMART domain sequence analysis program (Letunic and Bork 2017). No identifiable domains were found present in the transcripts.

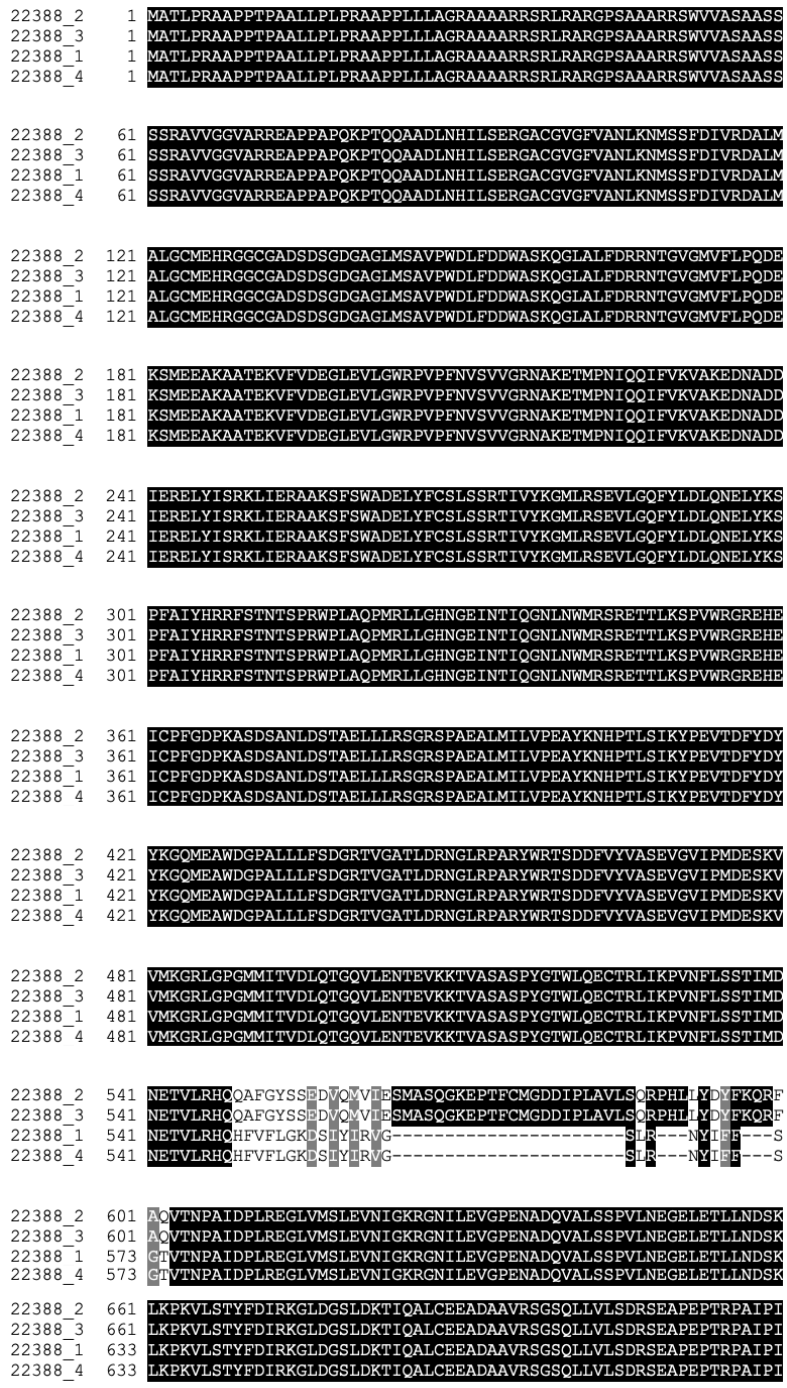



**Figure 12. Alignment of proteins encoded by Zm00001d022388 – Part 1.** Amino acid sequence of gene Zm00001d022388 showing four different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

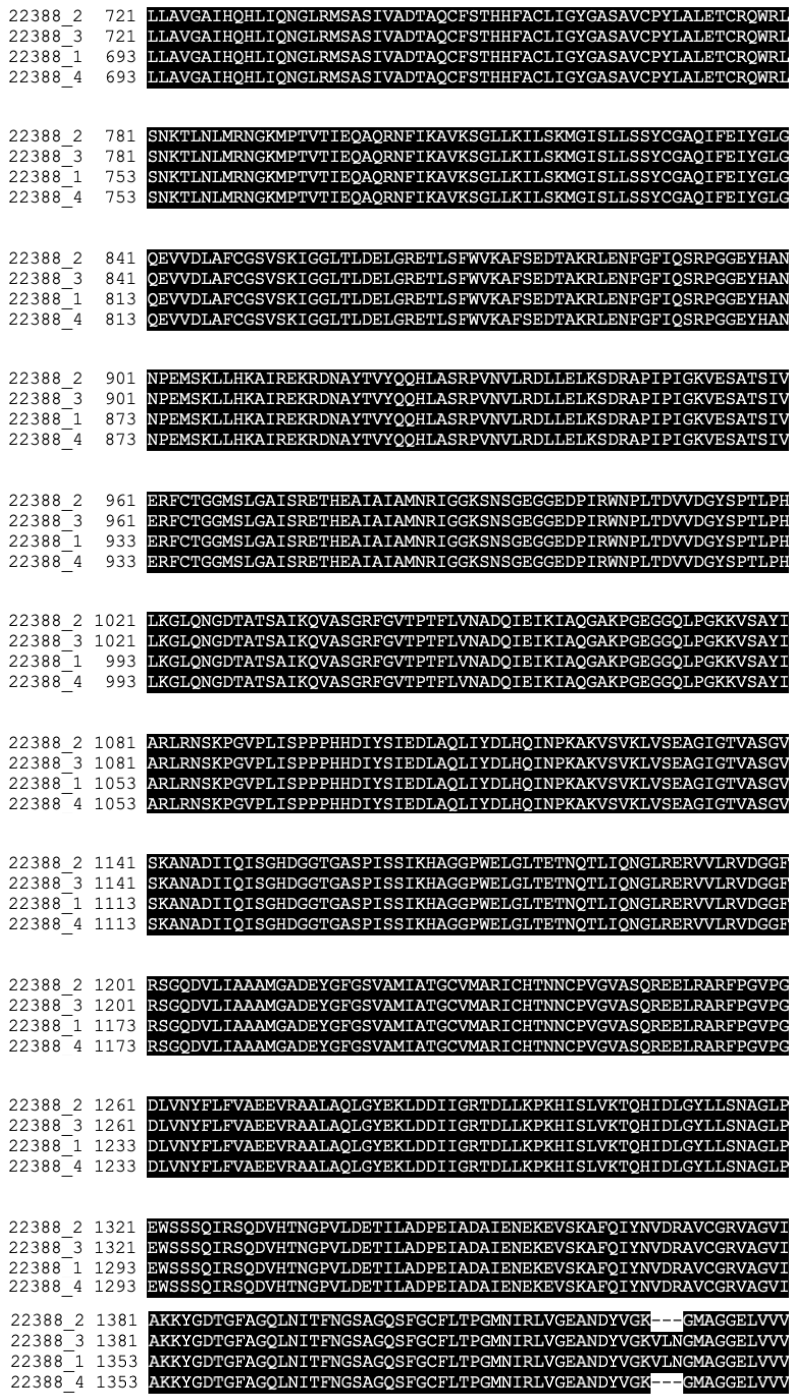



**Figure 12. Alignment of proteins encoded by Zm00001d022388 – Part 2.** Amino acid sequence of gene Zm00001d022388 showing four different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in an alternate RNA strand.


**Figure 12. Alignment of proteins encoded by Zm00001d022388 – Part 3.** Amino acid sequence of gene Zm00001d022388 showing four different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

The next gene that was evaluated was Zm00001d052165, which encodes nitrate reductase 2 (Tello-Ruiz *et. al.* 2017). The original gene structure was compared with the data in the Apollo annotation platform and a model was produced as noted in the methods (Figure 13). Comparison of the v5, IsoSeq, full-length cDNA (flc), and Mikado RNA models showed the presence of retained introns, as well as alternative start and end of transcription sites (Figure 13A). The IsoSeq and flc models supported both the presence of an alternate start of transcription (*, Figure 13A), and was supported in RNA sequencing data (*, Figure 13B), so it was incorporated into the SuperTranscript (*, Figure 13C, Table 2). IsoSeq and flc models supported the presence and absence of an intron in one location (**, Figure 13A), and RNA sequencing data supported the absence of an exon in that location (**, Figure 13B), so it was labeled as retained intron in the final model (**, Figure 13C, Table 2). RNA sequencing data also supported the presence and absence of an intron (***, Figure 13B, so it was incorporated into the SuperTranscript (***, Figure 13C, Table 2). of the additional 3' terminal exon (Figure 13A, 13B), so this was labeled as an alternative exon in the revised model (**, Figure 13C; Table 2). Finally, a new end of transcription site was identified in the IsoSeq model (****, Figure 13A), and supported in the RNA sequencing evidence, (****, Figure 13B) and incorporated into the SuperTranscript (****, Figure 13C, Table 2).

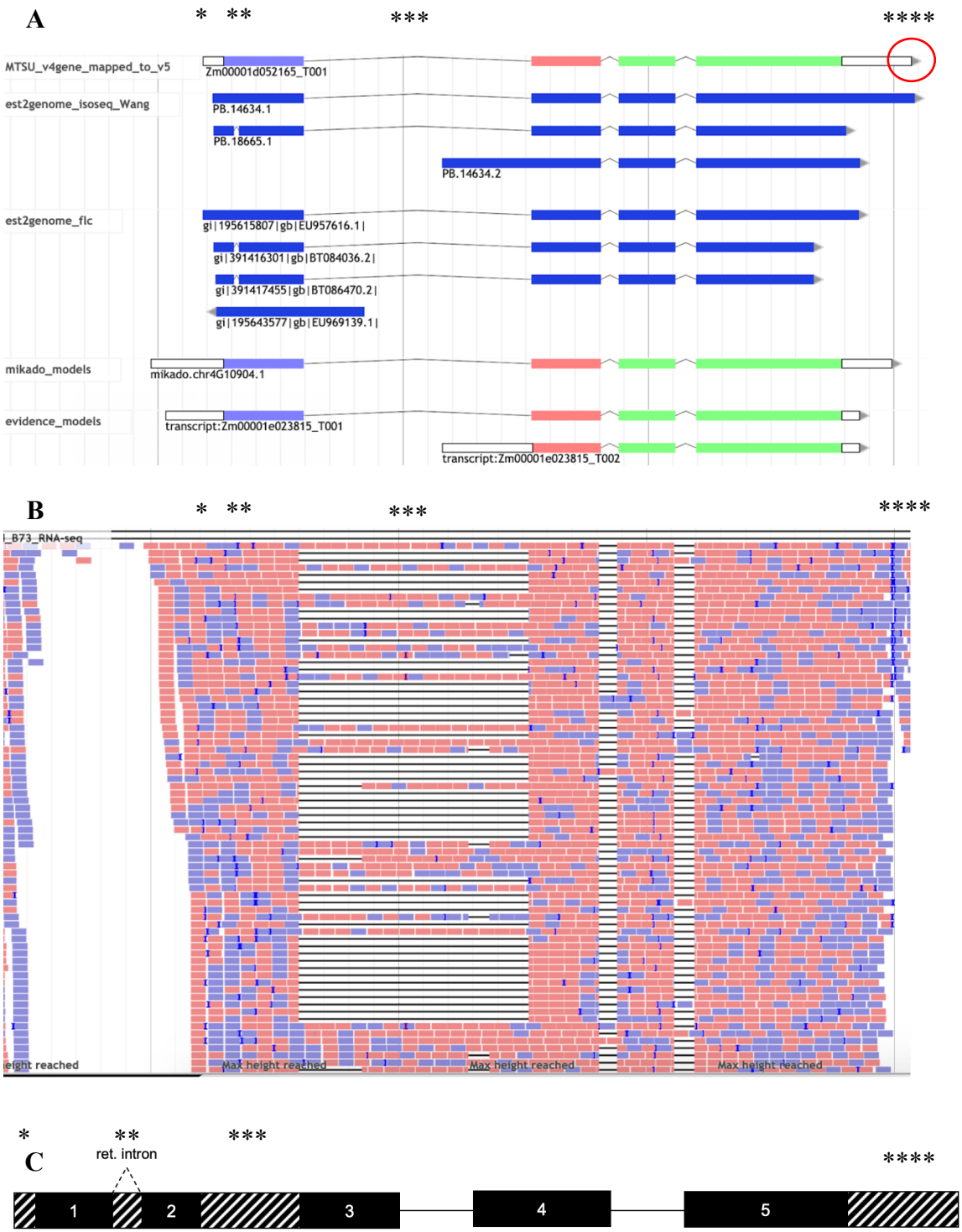



**Figure 13. Evidence for reannotation for gene Zm00001d052165.** Panel A shows the original structure of gene Zm00001d052165. Notice that the arrow on the original structure in panel A indicates the 3' end is on the left side of the gene (red circle) so the direction of the gene will be the reversed in the final model. The green and purple boxes show differentiation between the gene's reading frames. Panel B shows the final revised model. Note that the final model shows the RNA strand 5' to 3', which is the reverse of how the original model is displayed. The numbered black boxes indicate exons and their sequence from left to right. The lines between exons indicate introns.

Using the SuperTranscript as a guide, five RNAs were constructed in the Apollo space. The encoded amino acid sequences for each transcript were then downloaded from the Apollo workspace and aligned using Clustal Omega (Sievers *et. al.* 2011) and shaded using BOXSHADE to allow visualization of similarities and differences between the proteins encoded by the different transcripts. The five transcripts produce proteins that differ in various regions (Figure 14). Transcripts 1 and 3 encode the same protein. Transcript 5 encodes a protein missing 92 amino acids compared to transcripts 1 and 3. Transcript 4 encodes a protein missing 49 amino acids compared to transcripts 1 and 3. Transcript 2 encodes a protein missing 56 amino acids compared to transcripts 1 and 3. Lastly, the sequence of amino acids encoded by each transcript was analyzed for the presence of known protein domains using the SMART domain sequence analysis program (Letunic and Bork 2017). No identifiable domains were found present in the transcripts.

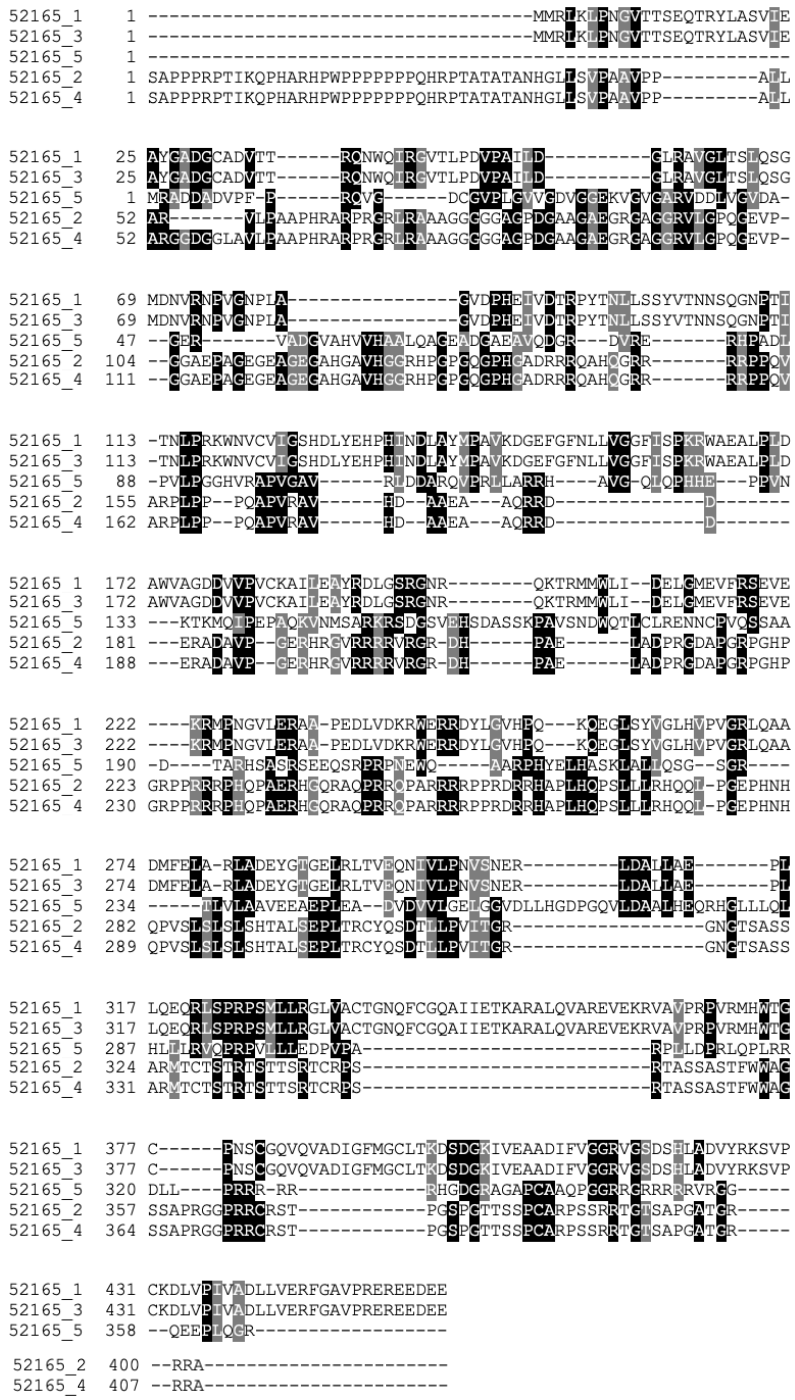


**Figure 14. Alignment of proteins encoded by Zm00001d052165.** Amino acid sequence of gene Zm00001d052165 showing five different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

The next gene that was evaluated was Zm00001d018161, which encodes ferredoxin-nitrite reductase chloroplastic (Tello-Ruiz *et. al.* 2017). The original gene structure was compared with the data in the Apollo annotation platform and a model was produced as noted in the methods (Figure 15). Comparison of the v5 and Mikado RNA models showed an alternative 3' splice site (*, Figure 15A), and RNA sequencing evidence supported presence of an alternative 3' splice site at that location (*, Figure 15B), so it was incorporated into the SuperTranscript (*, Figure 15C, Table 2).

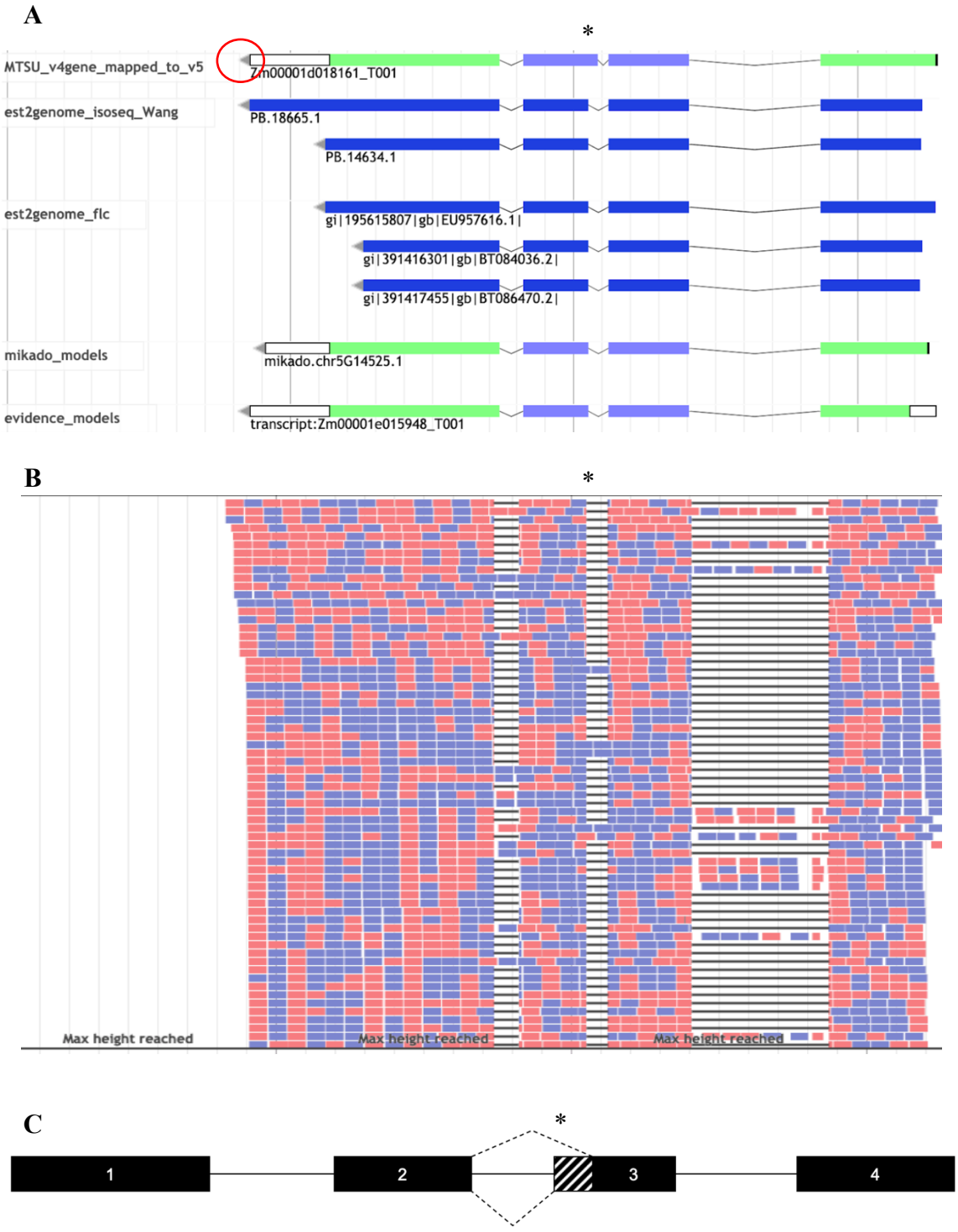



**Figure 15. Evidence for reannotation for gene Zm00001d018161.** Panel A shows the original structure of gene Zm00001d018161. Notice that the arrow on the original structure in panel A indicates the 3' end is on the left side of the gene (red circle) so the direction of the gene will be the reversed in the final model. The green and purple boxes show differentiation between the gene's reading frames. Panel B shows the final revised model. Note that the final model shows the RNA strand 5' to 3', which is the reverse of how the original model is displayed. The numbered black boxes indicate exons and their sequence from left to right. The lines between exons indicate introns.

Using the SuperTranscript as a guide, four RNAs were constructed in the Apollo space. The encoded amino acid sequences for each transcript were then downloaded from the Apollo workspace and aligned using Clustal Omega (Sievers *et. al.* 2011) and shaded using BOXSHADE to allow visualization of similarities and differences between the proteins encoded by the different transcripts. The four transcripts produce proteins that differ in the beginning and central regions (Figure 16). Transcript 2 encodes a protein missing 12 amino acids compared to transcript 1. Transcript 3 encodes a protein missing 13 amino acids compared to transcript 1. Transcript 4 encodes a protein missing 25 amino acids compared to transcript 1. Lastly, the sequence of amino acids encoded by each transcript was analyzed for the presence of known protein domains using the SMART domain sequence analysis program (Letunic and Bork 2017). No identifiable domains were found present in the transcripts.

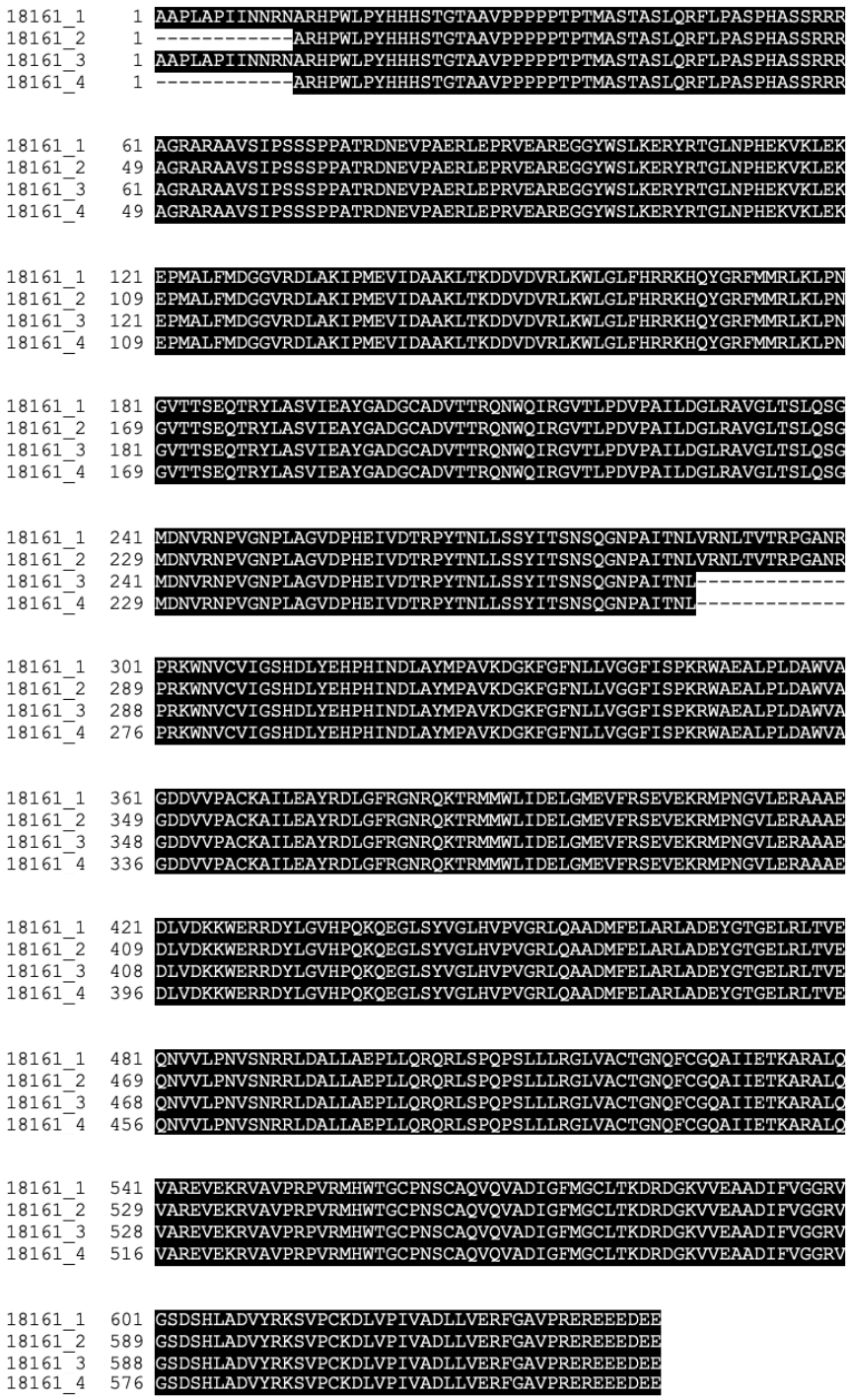



**Figure 16. Alignment of proteins encoded by Zm00001d018161.** Amino acid sequence of gene Zm00001d018161 showing four different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

The next gene that was evaluated was Zm00001d025984, which encodes glutamic dehydrogenase 2 (Tello-Ruiz *et. al.* 2017). The original gene structure was compared with the data in the Apollo annotation platform and a model was produced as noted in the methods (Figure 17). Comparison of the v5 and Mikado RNA models showed the presence of an alternative 3' splice site (*, Figure 17A) and was supported by RNA sequencing evidence (*, Figure 17B), so it was incorporated into the SuperTranscript (* Figure 17C, Table 2) Additionally, RNA sequencing data supported the presence of another 3' alternative splice site (**, Figure 17B), so this was incorporated into the SuperTranscript (**, Figure 17C; Table 2).

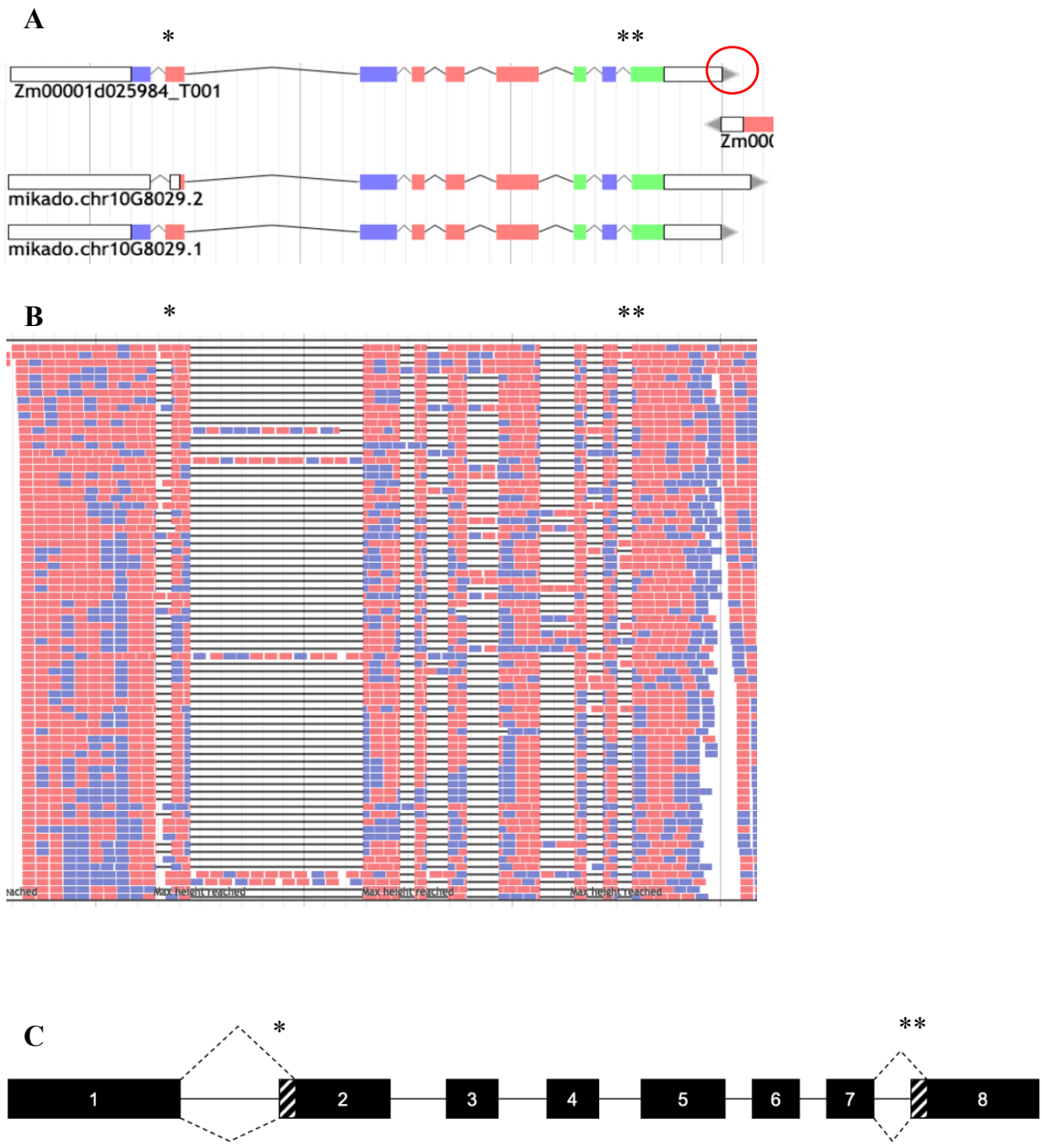



**Figure 17. Evidence for reannotation for gene Zm00001d025984.** Panel A shows the original structure of gene Zm00001d025984. Notice that the arrow on the original structure in panel A indicates the 3' end is on the left side of the gene (red circle) so the direction of the gene will be the reversed in the final model. The green and purple boxes show differentiation between the gene's reading frames. Panel B shows the final revised model. Note that the final model shows the RNA strand 5' to 3', which is the reverse of how the original model is displayed. The numbered black boxes indicate exons and their sequence from left to right. The lines between exons indicate introns.

Using the SuperTranscript as a guide, four RNAs were constructed in the Apollo space. The encoded amino acid sequences for each transcript were then downloaded from the Apollo workspace and aligned using Clustal Omega (Sievers *et. al.* 2011) and shaded using BOXSHADE to allow visualization of similarities and differences between the proteins encoded by the different transcripts. The four transcripts produce proteins that differ in a couple of regions (Figure 18). Transcript 2 encodes a protein missing 2 amino acids compared to transcript 1. Transcript 3 encodes a protein missing 62 amino acids compared to transcript 1. Transcript 4 encodes a protein missing 64 amino acids compared to transcript 1. Lastly, the sequence of amino acids encoded by each transcript was analyzed for the presence of known protein domains using the SMART domain sequence analysis program (Letunic and Bork 2017). The glutamate/leucine/phenylalanine/valine dehydrogenase domain was found present in the gene.

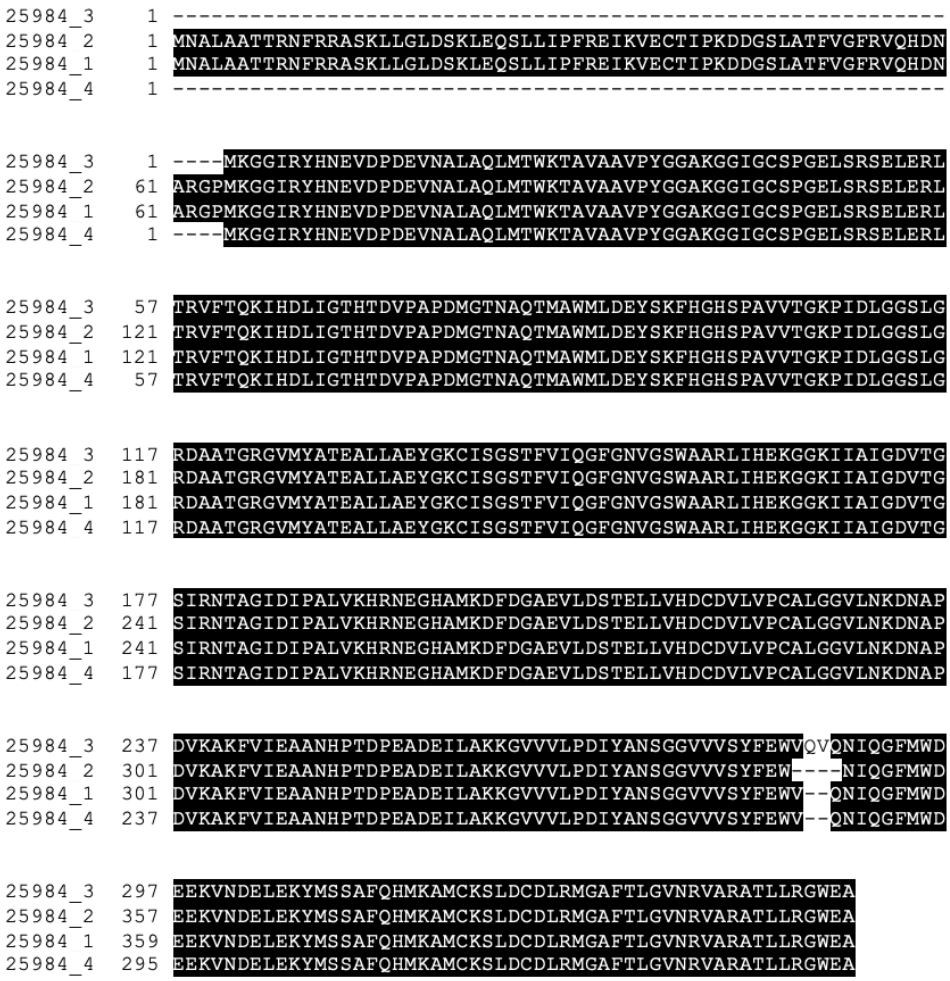



**Figure 18. Alignment of proteins encoded by Zm00001d025984.** Amino acid sequence of gene Zm00001d025984 showing four different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

The next gene that was evaluated was Zm00001d028750, which encodes asparagine synthetase 3 (Tello-Ruiz *et. al.* 2017). The original gene structure was compared with the data in the Apollo annotation platform and a model was produced as noted in the methods (Figure 19). RNAseq Mikado models supported both the presence and absence of an intron at one location (*, Figure 19A), and RNA sequencing data supported the absence of an intron in that location (*, Figure 19B), so it was labeled as a retained intron in the revised model (*, Figure 19C, Table 2). Comparison of v5 and IsoSeq models showed the presence and absence of an intron at one location (**, Figure 19A), and RNA sequencing data supported the absence of an intron in that location (**, Figure 19B), so it was labeled as a retained intron in the revised model (**, Figure 19C, Table 2). Comparison between v5 and an flc model shows both the absence and presence of an intron in the same location (***, Figure 19A), and RNA sequencing data shows the absence of that intron (***, Figure 19B), so it was incorporated into the SuperTranscript (***, Figure 19C, Table 2). Finally, two end of transcription sites were identified in the Mikado RNA models (****, Figure 19A), and supported in the RNA sequencing evidence, (****, Figure 19B) and incorporated into the SuperTranscript (****, Figure 19C).

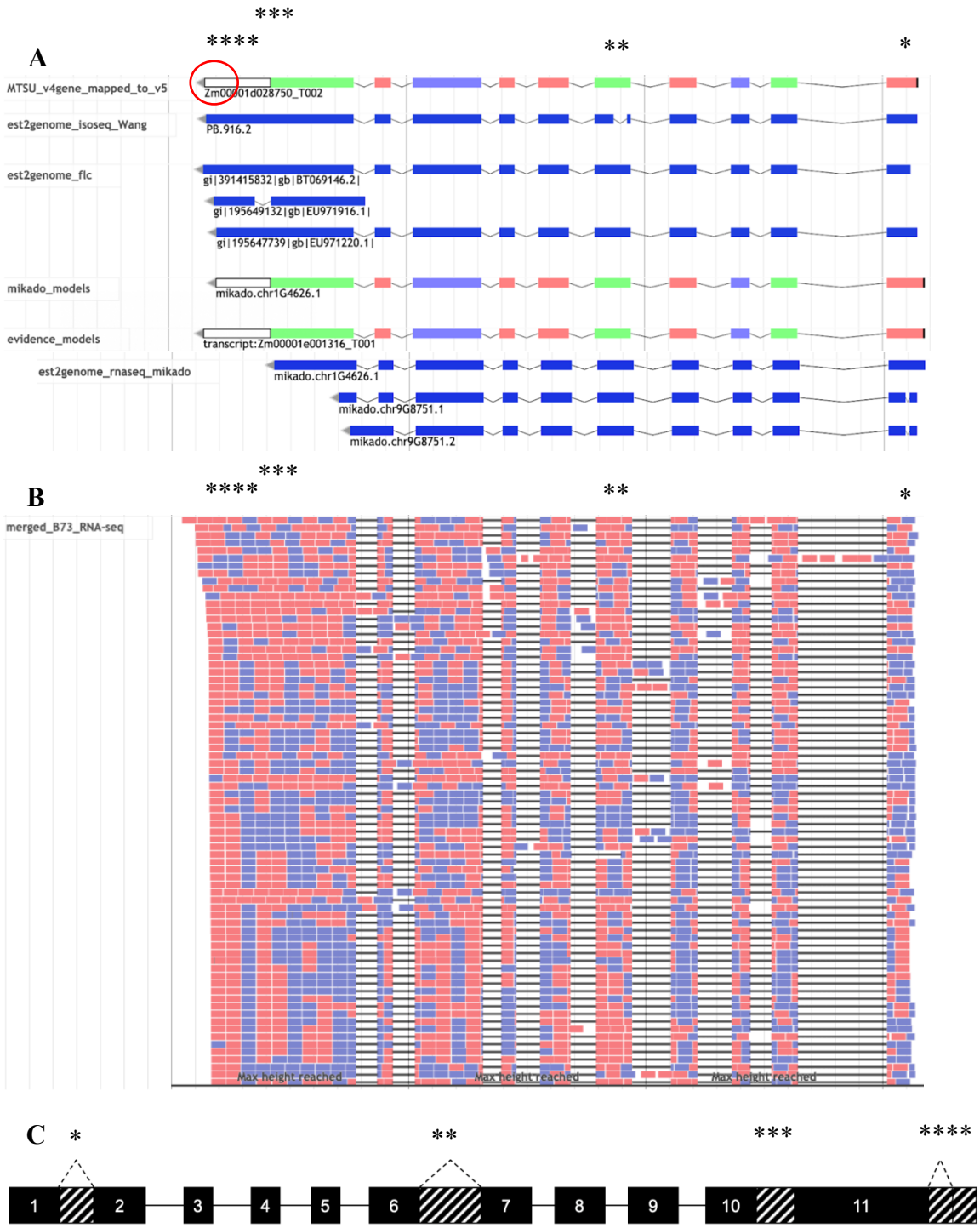



**Figure 19. Evidence for reannotation for gene Zm00001d028750.** Panel A shows the original structure of gene Zm00001d028750. Notice that the arrow on the original structure in panel A indicates the 3' end is on the left side of the gene (red circle) so the direction of the gene will be the reversed in the final model. The green and purple boxes show differentiation between the gene's reading frames. Panel B shows the final revised model. Note that the final model shows the RNA strand 5' to 3', which is the reverse of how the original model is displayed. The numbered black boxes indicate exons and their sequence from left to right. The lines between exons indicate introns.

Using the SuperTranscript as a guide, eight RNAs were constructed in the Apollo space. The encoded amino acid sequences for each transcript were then downloaded from the Apollo workspace and aligned using Clustal Omega (Sievers *et. al.* 2011) and shaded using BOXSHADE to allow visualization of similarities and differences between the proteins encoded by the different transcripts. The eight transcripts produce proteins that differ in various regions (Figure 20). Transcript 1 encodes a protein that is 12 amino acids shorter compared to transcript 2. Transcript 3 encodes a protein missing 19 amino acids compared to transcript 2. Transcript 6 encodes a protein missing 24 amino acids compared to transcript 2. Transcript 5 encodes a protein missing 36 amino acids compared to transcript 2. Transcript 4 encodes a protein missing 38 amino acids compared to transcript 2. Transcript 7 encodes a protein missing 43 amino acids compared to transcript 2. Transcript 8 encodes a protein missing 62 amino acids compared to transcript 2. Lastly, the sequence of amino acids encoded by each transcript was analyzed for the presence of known protein domains using the SMART domain sequence analysis program (Letunic and Bork 2017). No identifiable domains were found present in the transcripts.

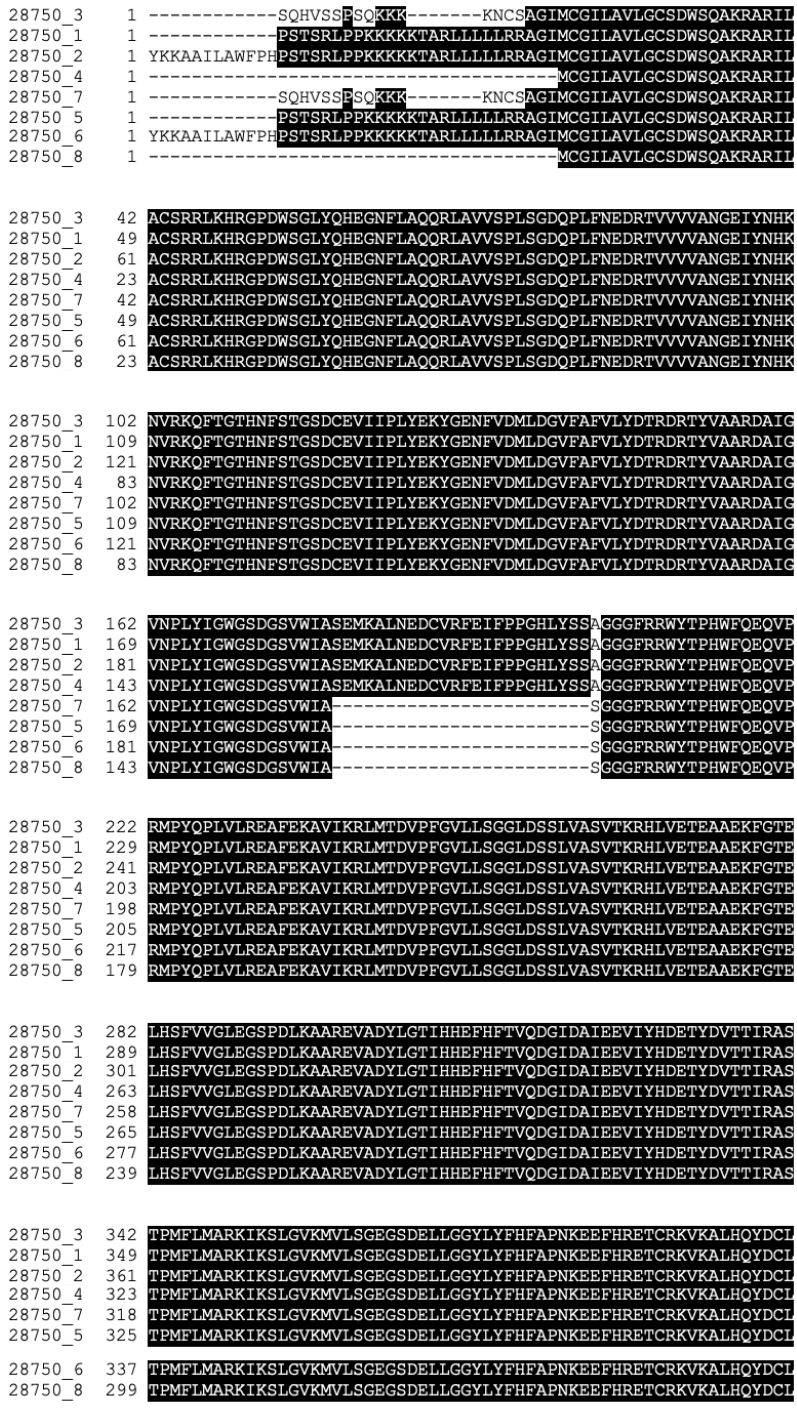



**Figure 20. Alignment of proteins encoded by Zm00001d028750 – Part 1.** Amino acid sequence of gene Zm00001d028750 showing eight different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

```
28750_3  402 RANKATSAWGLEVRVPFLDKEFINVAMGMDPEWKMYDKNLGRIEKWVMRKAFDDDEHPYL
28750_1  409 RANKATSAWGLEVRVPFLDKEFINVAMGMDPEWKMYDKNLGRIEKWVMRKAFDDDEHPYL
28750_2  421 RANKATSAWGLEVRVPFLDKEFINVAMGMDPEWKMYDKNLGRIEKWVMRKAFDDDEHPYL
28750_4  383 RANKATSAWGLEVRVPFLDKEFINVAMGMDPEWKMYDKNLGRIEKWVMRKAFDDDEHPYL
28750_7  378 RANKATSAWGLEVRVPFLDKEFINVAMGMDPEWKMYDKNLGRIEKWVMRKAFDDDEHPYL
28750_5  385 RANKATSAWGLEVRVPFLDKEFINVAMGMDPEWKMYDKNLGRIEKWVMRKAFDDDEHPYL
28750_6  397 RANKATSAWGLEVRVPFLDKEFINVAMGMDPEWKMYDKNLGRIEKWVMRKAFDDDEHPYL
28750_8  359 RANKATSAWGLEVRVPFLDKEFINVAMGMDPEWKMYDKNLGRIEKWVMRKAFDDDEHPYL

28750_3  462 PKHILYRQKEQFSDGVGYNWIDGLKSFTEQQVTDEMMNNAAQMFPYNTPVNKEAYYYRMI
28750_1  469 PKHILYRQKEQFSDGVGYNWIDGLKSFTEQQVTDEMMNNAAQMFPYNTPVNKEAYYYRMI
28750_2  481 PKHILYRQKEQFSDGVGYNWIDGLKSFTEQQVTDEMMNNAAQMFPYNTPVNKEAYYYRMI
28750_4  443 PKHILYRQKEQFSDGVGYNWIDGLKSFTEQQVTDEMMNNAAQMFPYNTPVNKEAYYYRMI
28750_7  438 PKHILYRQKEQFSDGVGYNWIDGLKSFTEQQVTDEMMNNAAQMFPYNTPVNKEAYYYRMI
28750_5  445 PKHILYRQKEQFSDGVGYNWIDGLKSFTEQQVTDEMMNNAAQMFPYNTPVNKEAYYYRMI
28750_6  457 PKHILYRQKEQFSDGVGYNWIDGLKSFTEQQVTDEMMNNAAQMFPYNTPVNKEAYYYRMI
28750_8  419 PKHILYRQKEQFSDGVGYNWIDGLKSFTEQQVTDEMMNNAAQMFPYNTPVNKEAYYYRMI

28750_3  522 FERLFPQDSARETVPWGPSIACSTPAAIEWVEQWKASNDPSGRFISSHDSAATDHTGGKP
28750_1  529 FERLFPQDSARETVPWGPSIACSTPAAIEWVEQWKASNDPSGRFISSHDSAATDHTGGKP
28750_2  541 FERLFPQDSARETVPWGPSIACSTPAAIEWVEQWKASNDPSGRFISSHDSAATDHTGGKP
28750_4  503 FERLFPQDSARETVPWGPSIACSTPAAIEWVEQWKASNDPSGRFISSHDSAATDHTGGKP
28750_7  498 FERLFPQDSARETVPWGPSIACSTPAAIEWVEQWKASNDPSGRFISSHDSAATDHTGGKP
28750_5  505 FERLFPQDSARETVPWGPSIACSTPAAIEWVEQWKASNDPSGRFISSHDSAATDHTGGKP
28750_6  517 FERLFPQDSARETVPWGPSIACSTPAAIEWVEQWKASNDPSGRFISSHDSAATDHTGGKP
28750_8  479 FERLFPQDSARETVPWGPSIACSTPAAIEWVEQWKASNDPSGRFISSHDSAATDHTGGKP

28750_3  582 AVANGGGHGAANGTVNGKDVAVAIAV
28750_1  589 AVANGGGHGAANGTVNGKDVAVAIAV
28750_2  601 AVANGGGHGAANGTVNGKDVAVAIAV
28750_4  563 AVANGGGHGAANGTVNGKDVAVAIAV
28750_7  558 AVANGGGHGAANGTVNGKDVAVAIAV
28750_5  565 AVANGGGHGAANGTVNGKDVAVAIAV
28750_6  577 AVANGGGHGAANGTVNGKDVAVAIAV
28750_8  539 AVANGGGHGAANGTVNGKDVAVAIAV
```

**Figure 20. Alignment of proteins encoded by Zm00001d028750 – Part 2.** Amino acid sequence of gene Zm00001d028750 showing eight different versions of the protein that could be produced by the DNA sequence based on the revised model. The absence sequence is denoted by dashes, meaning the next amino acid in the chain would be the next letter after the dashes. Identities between different versions of protein are denoted by the shaded boxes. Note also that the boxes may be shaded even if the sequence is "skipped" in the comparison protein.

## Conclusions

Nitrogen-use efficiency plays an important role in *Zea mays'* growth and development. Nitrogen depletion by *Zea mays*, therefore, requires additional fertilizer and crop rotation in order to maintain both croppable soil and plant health. Understanding the biochemistry and genetics of nitrogen-use efficiency may allow for development of corn varieties less dependent on financially costly fertilizer addition and time costly crop rotation.

As a first step in understanding the genetic and biochemical aspects of genes involved in corn nitrogen use, a systematic categorization of all transcripts and encoded proteins for nine NUE genes was undertaken using newly available whole transcriptome data from multiple corn tissues. Using Apollo genome annotation collaboration software, (Ware) existing v5 corn gene models and transcriptome data were used to generate a SuperTranscript for each gene. Exon skipping, intron retention, alternate start and end of transcription, and alternate 5' and 3' splice sites were observed (Table 1). Thirty-five differences were found across the nine genes including: ten intron retentions, six alternate 5' splice sites, eleven alternate 3' splice sites, three alternative exons, three alternate end of transcription sites, one alternate start of transcription site, and one alternative intron (Table 2). Using these alternative features, old transcripts were re-annotated or constructed to accurately reflect the breadth of transcript versions (isoforms) produced by each gene. Furthermore, the encoded proteins for each gene were compared by multiple alignment and computational domain analysis. While most transcripts encoded slightly different proteins, only one protein's domain analysis showed a known protein domain. Therefore, until additional domain information becomes available, or these proteins are

studied biochemically, it is unclear how the proteins encoded by the different transcripts might function differently.

In addition to the nine genes of focus in this study, there are other genes pertaining to nitrogen use efficiency, including *Zm00001d048050, Zm00001d028260, Zm00001d011610, Zm00001d043845, Zm00001d038948, Zm00001d034420, Zm00001d002052, Zm00001d045675, Zm00001d022152, Zm00001d014258,* and *Zm00001d007357* (Table 1). Once all NUE genes have been systematically studied for transcript isoforms, researchers will have a more complete view of all proteins involved in the process and can use this knowledge to investigate mechanisms to alter regulation of these proteins to generate corn varieties with specific characteristics.

**Literature Cited**

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, *et al*. 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 40(W1):W597-W603. doi:10.1093/nar/gks400

Davidson NM, Hawkins ADK, Oshlack A. 2017. SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes. Genome Biology. 18(148). doi:10.1186/s13059-017-1284-1

Kennett DJ, Thakar HB, VanDerwarker AM, Webster DL, Culleton BJ, Harper TK, Kistler L, Scheffler TE, Hirth K. 2017. High-precision chronology for central American maize diversification from El Gigante rockshelter, Honduras. PNAS. 114(34):9026-9031. doi:10.1073/pnas.1705052114

Letunic I, Bork P. 2017. 20 years of the SMART protein domain annotation resource. Nucleic Acids Research. 46(D1):D493-D496. https://doi.org/10.1093/nar/gkx922

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Research. 47(W1):W636-W641. doi:10.1093/nar/gkz268

Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, *et al*. 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations**.** Nucleic Acids Research. doi:10.1093/nar/gky1100

Shah TR, Prasad K, Kumar P. 2016. Maize—A potential source of human nutrition and health: A review. Cogent Food & Agriculture. 2(1). doi:10.1080/23311932.2016.1166995

Sharma LK, Bali SK. 2018. A Review of Methods to Improve Nitrogen Use Efficiency in Agriculture. Sustainability. 10(1). https://doi.org/10.3390/su10010051

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:539. doi: 10.1038/msb.2011.75

Simons M, Saha R, Guillard L, Clément G, Armengaud P, Cañas R, Maranas CD, Lea PJ, Hirel B. 2014. Nitrogen-use efficiency in maize (*Zea mays* L.): from 'omics' studies to metabolic modelling. Journal of Experimental Botany. 65(19):5657-5671. doi:10.1093/jxb/eru227

Singh B. 2018. Are nitrogen fertilizers deleterious to soil health?. Agronomy. 8(4). https://doi.org/10.3390/agronomy8040048

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, *et al*. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science. 326(5956):1112-1115. doi:10.1126/science.1178534

Tello-Ruiz MK, Naithani S, Stein JC, Gupta P, Campbell M, Olson A, Wei S, Preece J, Geniza MJ, Jiao Y, et al. 2017. Gramene 2018: unifying comparative genomics and pathway resources for plant research. Nucleic Acids Research. 2018; 46(D1). doi:10.1093/nar/gkx1111

Veljković VB, Biberdžić MO, Banković-Ilić IB, Djalović IG, Tasić MB, Nježić ZB, Stamenković OS. 2018. Biodiesel production from corn oil: A review. Renewable and Sustainable Energy Reviews. 2018; 91:531-548. https://doi.org/10.1016/j.rser.2018.04.024

Xu Y, Skinner DJ, Wu H, Palacios-Rojas N, Araus JL, Yan J, Gao S, Warburton ML, Crouch JH. 2009. Advances in maize genomics and their value for enhancing genetic gains from breeding. International Journal of Plant Genomics. 2009:1-30. https://doi.org/10.1155/2009/957602