JOURNAL OF UNDERGRADUATE RESEARCH

Understanding the Browning Reactions in Sweet Basil: Gene Cloning and RNAi Construct Development for Polyphenol Oxidase

Annelise Vieira
University of Florida

Faculty Mentor: David G. Clark, Environmental Horticulture Department

Abstract

Polyphenol oxidase (PPO) is a well-characterized enzyme that is conserved across plant species and is responsible for catalyzing the rapid polymerization of o-quinones that results in tissue browning. As these oxidation reactions result in significant reductions in food crop commercial viability, nutritional content, and general desirability, an opportunity exists to improve a variety of crops via genetic regulation of PPO expression. The main objective of this project was to clone the DNA sequence of genes encoding PPO. Using mRNA as a template to synthesize DNA, two putative PPO sequences were cloned from *Ocimum basilicum* and *Ocimum americanum*, two basil species of commercial relevance. A secondary objective of this project was to insert a cloned PPO sequence into an RNAi vector for the purpose of making transgenic basil plants with reduced PPO expression. A hairpin RNAi construct was successfully developed to reduce expression of both PPO genes as their DNA sequences were highly homologous. The next step for this project will depend on the development of transformation and regeneration protocols for these species, which have proven to be difficult. Once the RNAi PPO constructs are introduced into basil cells and successfully regenerated into plants, reduced browning of basil leaf tissue when it is wounded or cut is expected.

Keywords: Ocimum basilicum, Lamiaceae, polyphenol oxidase, plant browning, RNAi

Introduction

A wide range of plants undergo oxidation reactions that result in tissue browning when cut, wounded, or processed (Martinez & Whitaker, 1995). This process along with other deteriorative reactions that occur postharvest are responsible for half of fruit and vegetable loss worldwide, making their effects significantly economically and environmentally impactful (Friedman, 1996). These large-scale reductions in food crop nutritional content, safety, and commercial viability thus necessitate efforts to mitigate these reactions via genetic modification (Caspi et al., 2016). Such efforts have already been undertaken in several crops; for instance, browning-resistant apple and potato varieties are emerging commercially (Waltz, 2015). The present work pursues a similar goal in two commercially relevant basil species, *Ocimum basilicum* and *Ocimum americanum*. The ‘Caesar’ cultivar of *O. basilicum* was selected for study due to its status as an industry-standard cultivar. The sequence of the wild African basil species *O. americanum* was

also determined due to its possession of genetic resistance to basil downy mildew, a disease currently generating major crop losses across continents (Ben-Naim, Falach, & Cohen, 2018). The creation of a genetically modified, browning-resistant basil variety is appropriate due to basil's position among the most prominent and economically significant herbs used in the United States and Europe for culinary purposes (Putievsky & Galambosi, 1999). Further, the heavy processing of basil leaf tissue involved in making pesto makes it particularly susceptible to browning.

These oxidation reactions are attributed to polyphenol oxidase (PPO), a well-characterized enzyme that is conserved across plant species and is responsible for catalyzing the rapid polymerization of o-quinones that results in tissue browning, as shown in Figure 1 (Queiroz, Lopes, Fialho, & Valente-Mesquita, 2008). While this enzyme's activity can be controlled by environmental factors such as low pH and cold temperatures, the feasibility and efficacy of these approaches for large-scale production and distribution are limited (Martinez & Whitaker, 1995). To achieve the most effective control of oxidation, a genetically modified plant that will not undergo browning reactions can be generated by silencing or knockdown of PPO gene expression via a variety of established genetic modification techniques, such as RNA interference (RNAi). The RNAi technique (also known as post-transcriptional gene silencing) derives from a cellular mechanism acting on double-stranded RNA that has been harnessed for sequence-specific gene silencing; RNAi has been used extensively in plant tissue culture experiments focused on reducing expression of mRNA transcripts (Hannon, 2002). This technique was used in the development of the browning-resistant apples and potatoes mentioned above (Waltz, 2015).

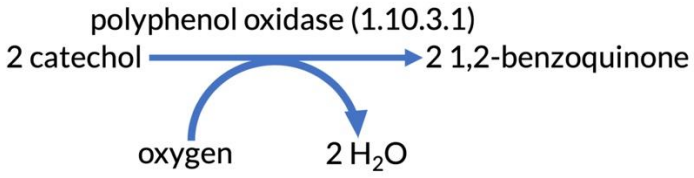

Figure 1: Polyphenol oxidase's catalytic role in the dehydrogenation of o-dihydroxyphenols to o-quinones (Caspi et al., 2016)

Knowledge of the PPO gene sequence is a prerequisite to venture toward its control via genetic modification. Because a basil genome has not been published at this time, the objective of the present work was to determine the nucleotide sequence via molecular cloning in the two

basil species proposed. Following the elucidation of these gene sequences, a second objective of this project was to develop long hairpin RNAi constructs to facilitate future efforts in developing a browning-resistant basil variety by interfering with the gene's expression.

This study reports the nucleotide sequences of two genes encoding PPO in *O. bacilicum* and one gene in *O. americanum*. Analysis of the homology of these newly identified basil genes with published sequences for PPO in other plant species revealed that they are most similar to a PPO sequence of *Erythranthe guttatus* (seep monkeyflower). One of the cloned gene sequences of *O. basilicum* (ObPPO1) was subsequently used in the development of a construct to reduce PPO experession via RNAi.

Materials & Methods

Cloning of PPO

isolation of RNA, creation of cDNA libraries, and PCR amplification. RNA was extracted from *O. basilicum* and *O. americanum* leaf tissue using Trizol® and was purified using TURBO™ DNase treatment. The purified mRNAs (Fig. 2, left) then served as templates for conventional generation of cDNA libraries (Fig. 2, right).

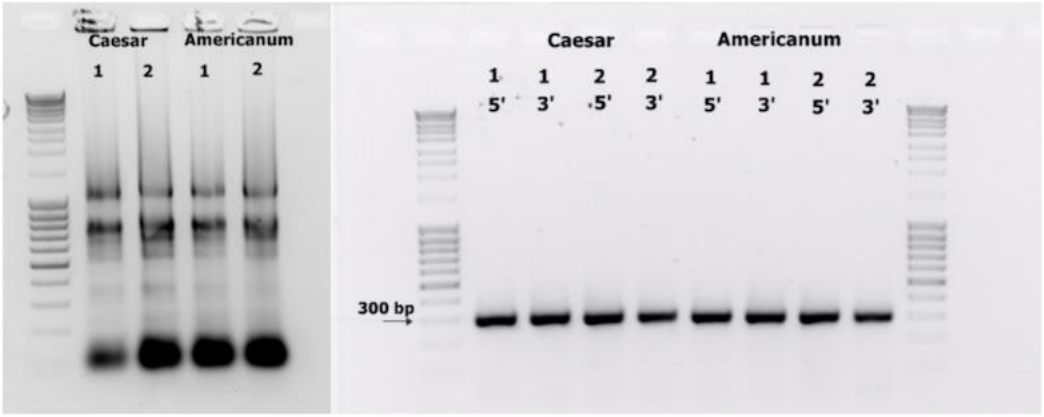

Figure 2. Confirmation of basil RNA (left) and cDNA (right) quality via gel electrophoresis

Primers for Rapid Amplification of cDNA Ends (RACE) were designed to allow for PCR amplification of the putative sequences from cDNA of both basil species of interest (Frohman, 1993). Given the absence of a published basil genome, the initial primers were designed by identifying conserved domains in PPO genes of other species in the Lamiaceae family published on NCBI via Clustal Omega multiple alignment (Sievers et al., 2011). These initial primers were

used to identify partial sequences that then allowed the development of the primers shown in Table 1 to amplify the complete PPO coding sequence.

Table 1. Primers Used in the Final Cloning Trials to Isolate the Complete Coding Sequences in *O. basilicum* and *O. americanum*.

Number	Orientation	Melting temperature	GC Content	Nucleotide Sequence
Primer #1795	F	61.10°C	43.48	GGGGATCACTTCCCATAACAATA
Primer #1796	R	60.01°C	43.48	CACACATGCACATTACACAACAG

These final primers were developed based on the PPO partial sequences obtained in initial cloning trials.

Using these primers, the PCR was run on a Veriti® 96-Well Thermal Cycler with a 53°C annealing temperature and 1:30 second extension-time for 40 cycles. The candidate PPO PCR products were isolated via gel electrophoresis and extracted using the QIAquick® Gel Extraction Kit.

ligation to vector and transformation of E. coli with PPO. The PCR products were ligated to the pGEM®-T Easy cloning vector via incubation with T4 DNA ligase, and the resultant vector was used to transform TOP10® competent *E. coli* cells via heat shock. A 50 μL aliquot of the *E. coli* cells was spread on Lysogeny Broth (LB) agar plates containing carbenicillin as a selectable marker for successfully transformed *E. coli* containing carbenicillin-resistance conferred by the pGEM vector. Plates were incubated overnight at 37°C.

screening for inserts and sequencing. Select *E. coli* colonies were cultured in 3 mL of LB containing carbenicillin overnight in a shaking incubator at 37°C. The plasmids were then extracted using a miniprep protocol and digested with the restriction enzyme NOTI and visualized via gel electrophoresis to confirm PCR product insertion. Samples containing the insert were sequenced by Genewiz Sanger Sequencing at the UF-ICBR sequencing core facility.

sequence analysis and bioinformatics. Preliminary consensus sequences were generated by separating all raw sequences into groupings by similarity, performing multiple alignments for each group, and calling each base based on prevalence among the replicates. These consensus sequences were then entered into Blastn, and those that were found to contain either of the two conserved PPO copper-binding domains or showed homology to tyrosinase superfamily proteins were identified as candidates (Zhang, Schwartz, Wagner, & Miller, 2000). From these working sequences, PCR primers were designed to clone the remainder of each gene. Sequencing data

from later trials yielded a full gene sequence containing a putative start codon and stop codon, at which point an open reading frame could be identified and Blastn could be used to identify similarities between the full cloned sequences and sequences in the NCBI databases.

Development of RNAi Constructs

The elucidation of these sequences then allowed for RNAi construct development using the *O. basilicum* PPO gene sequence to derive the hairpin. Two pairs of primers were constructed from this sequence such that the resultant amplified segments would form a hairpin secondary structure following transcription when ligated into a vector in the appropriate orientation (represented diagrammatically in Figure 3). As there is currently no published report of RNAi construct development in sweet basil, modifications were made to an established RNAi protocol for petunia, which allowed for insertion of a hairpin sequence with a pFMV promoter to the pHK1001 vector to permit a high level of expression (Dexter et al., 2007).

The major steps involved in the development of the constructs are summarized as follows: Primers were designed to amplify two fragments from the coding sequence: both primer sets had the same forward primer. One reverse primer was approximately 300 bases downstream of the forward primer, and the other reverse primer was approximately 700 bases downstream. The objective was thus to generate two fragments that when ligated, inserted into a vector, and transcribed would have a complementary region allowing for the formation of a hairpin secondary structure. PCR was performed to amplify the two fragments and the products were purified. The fragments and pFMV vector were digested with restriction enzymes, purified, and ligated to generate a pFMV-insert vector. It was found that it was essential that pFMV undergo a treatment with alkaline phosphatase before the ligation procedure to maintain a linear form capable of ligation to the insert. The inserted segment was confirmed via heat shock transformation of *E. coli* and culturing on selective media, miniprep to isolate the vector, digestion to isolate the insert, and gel electrophoresis to visualize the insert. Samples indicating that that ligation had occurred correctly were sequenced via Genewiz Sanger Sequencing for confirmation. The pFMV construct and the pHK vector were digested with the NOTI restriction enzyme such that the hairpin with FMV promoter could be inserted into the pHK vector. pHK was also treated with alkaline phosphate prior to ligation. Finally, another confirmation step was undertaken. The construct was again introduced into competent E. coli cells via heat shock

transformation, followed by culturing on selective media, miniprep to isolate the vector, digestion to isolate the insert, and gel electrophoresis to confirm the insert.

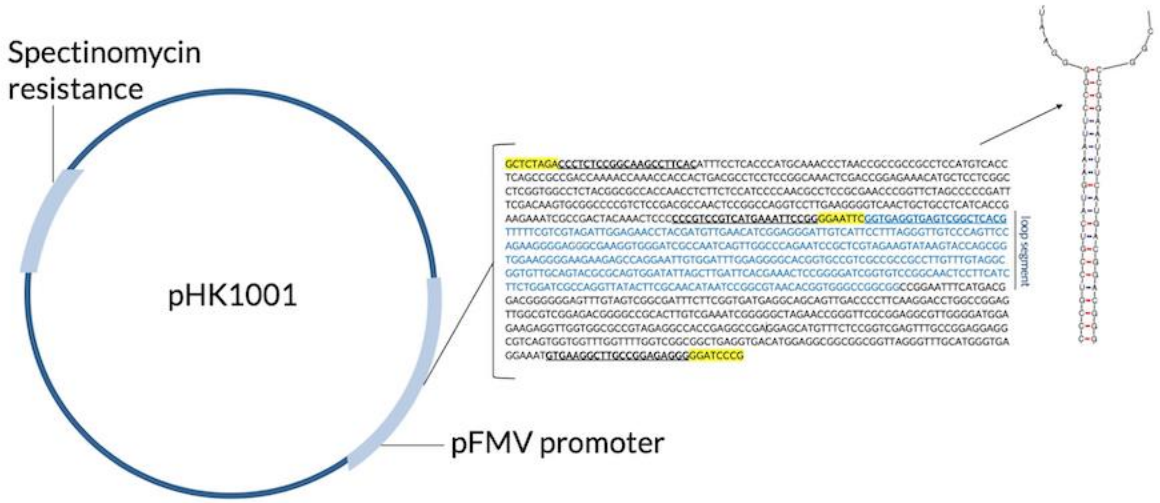

Figure 3. A simplified representation of the RNAi construct developed. The ObPPO1 hairpin sequence (center) and the pFMV promoter were inserted into the pHK1001 vector (left). The figure on the right is a representation of a portion of the hairpin structure that the mRNA will form when the sequence is transcribed.

Results

Nucleotide sequences of the putative PPO genes in *O. basilicum* and *O. americanum* were identified. In both species, the majority of the cloning trials acted as replicates for a single consensus sequence, which was named ObPPO1 in *O. basilicum* and OaPPO2 in *O. americanum.* In addition, a similar but distinct variant sequence was observed in *O. basilicum* in several trials, which was named ObPPO2.

NCBI Blast searches indicated that the cloned PPO sequences showed high identity to the PPO sequences of other species (Table 2). ObPPO1, ObPPO2, and OaPPO1 were most similar to the PPO sequence of *Erythranthe guttatus* (seep monkeyflower) with identities of 74.08% (E value = 3e-129) 73.36% (E value = 1E-95), and 73.11% (E value = 5e-137) respectively. While the percent identity between the basil sequences and those listed in Table 2 were consistently high with low E values, only the *E. guttatus* sequence demonstrated a high query cover. The additional sequences showed significantly lower query cover, from 4-11%. The full nucleotide sequences of the cloned sequences and the PPO sequences of *E. guttatus* and *P. avium*, as an example of a sequence showing low query cover, are shown in multiple alignment in the Figure 4. The lack of a greater number of high query cover matches is likely due to the lack of published

PPO sequences of species in the Lamiaceae family. Putative amino acid sequences for each gene could be generated via open reading frame analysis, and alignment of these sequences demonstrates very strong identity between the PPO1 sequences of the two *Ocimum* species (Fig. 5). The two copper binding domains characteristic of the enzyme are also clearly present in all cloned sequences, an indication that the data is reliable (Malviya, Srivastava, Diwakar, & Mishra, 2011).

Table 2. The Species and Accessions of the NCBI Nucleotide Collection (nt) Database Sequences with the Highest Homology to the Cloned PPO Sequences.

Species	Sequence Accession	Most Similar Sequence	Identity (%)	E Value	Query Cover (%)
Erythranthe guttatus	XR_001173111.1	OaPPO1	73.11	6.00E-137	83
Prunus avium	XM_021975103.1	OaPPO1	77.67	4.00E-24	11
Prunus mume	XM_008226776.1	OaPPO1	84.68	5.00E-23	6
Prunus mexicana	DQ851217.1	OaPPO1	84.68	5.00E-23	6
Prunus persica	XM_007211330.2	OaPPO1	83.87	2.00E-21	6
Pryus x bretschneideri	XM_009354696.2	ObPPO1	84.762	5.00E-18	5
Eriobotrya japonica	JX025010.1	ObPPO1	84.762	5.00E-18	5
Morus notabilis	XM_010102272.2	OaPPO1	83.76	4.00E-19	6
Morus alba var. *multicaulis*	KT371993.1	OaPPO1	83.76	4.00E-19	6
Malus domestica	XM_029103209.1	ObPPO1	83.81	2.00E-16	5
Lyonothamnus floribundus	DQ851210.1	OaPPO1	82.26	5.00E-18	6
Prunus salicina	AY866432.1	OaPPO1	83.06	1.00E-19	6
Cucubita moschata	XM_023071312.1	ObPPO2	84.62	8.00E-14	8
Musa acuminata	XM_009416120.2	ObPPO2	83.51	3.00E-13	8
Sorghum bicolor	XM_002443941.2	ObPPO2	93.33	1.00E-06	4

All of the above sequences are annotated as encoding polyphenol oxidase.

```
XM          GGTGTTAGTGATTGGCGATCATATTTTATAATTCTTTTCCGTCAACTAGATCGGGCTGTC    60
XR          ------------------------------------------------------------    0
ObPPO2      ------------------------------------------------------------    0
ObPPO1      ------------------------------------------------------------    0
OaPPO1      ------------------------------------------------------------    0

XM          TGCATGGCTATTGCCTATTGGATAGAAAAGTAAAACCCAAAATCCTCTATAAAAAGCAAC    120
XR          ------------------------------CCCATCC-CAAAATACATATTTTGCAATGA    29
ObPPO2      ------------------------------------------------------------    0
ObPPO1      ----------------------GGGGATCACTTCCCATAACAATATTATTCATCATGG      36
OaPPO1      ----------------------GGGGATCACTTCCCATAACAATATTTTTCATCATGG      36

XM          CCTCCTTTGGGGTTTATACCATCCAACCCAACTAAGCCAGAGAGCTAACCATGGCTTCTC    180
XR          CATCTCTTT----ATCTATCCTGCGCCACCACCAACA----CCGCCAT------------    69
ObPPO2      ------------------------------------------------------------    0
ObPPO1      CTTCTCTTC----TCATTCCCCCCTCCACCACCTCCG----CCG----------------    72
OaPPO1      CTTCTCTTC----TCATTCCCCCTTCCACCACCTCCG----CCGGCGCC-----------    77

XM          TGTCACCTCCACCAGCCATTCCCACTACCTCCACAAAACCTCTCTCTCCCTTCACTCAAA    240
XR          ------CTCCGCCACCCCCTCCTCC------CACCCCCGCCCACTCCTCCCTAAGCCATC    117
ObPPO2      ------------------------------------------------------------    0
ObPPO1      -----CAGCCACCGCCCCGTCCCGC------CGGACCACCACTCACCCCCTCTCCGGCAA    121
OaPPO1      -GCCGGCGCCACCGCCCCGTCCCGC------CGGACCACCACCCATCCCCTCTCCGCCAA    130

XM          ACAGCTCCCAAGTTTCCCTACTCTCAAACCCCAAGCG--------TTCCTTTTCACCAAG    292
XR          --ACATTCACATTTCATCACCCATGCAAAGCTCACCAATCGCCTCCACCATGTTTCTTGC    175
ObPPO2      ------------------------------------------------------------    0
ObPPO1      --GCCTTCACATTTCCTCACCCATGCAAACCCTAACCGCCGCCGCCTCCATGTCACCTCA    179
OaPPO1      --GCCTTCACATTTCCTCACCCATGCAAACCCTAACCGCCGCCGCATCCACGTCACCTCT    188
```

```
XM          GGTTTCATGCAAAGCCAAAAACAGTGATAACCAAAATGATCAAGCACAGACTAAACTAGA    352
XR          TCCTCCACCC-CTCAAAACCAAAACGACGAGACGAAGAAACCAACCGCAAAGGTCGTCGA    234
ObPPO2      ------------------------------------------------------------    0
ObPPO1      GCCGC----------CGACCAAAACCAAACCACCACTGACGCCTCCTCCGGCAAACTCGA    229
OaPPO1      GCCGC----------CGACCAAAACCAAACCACCACTGACGCCTCCGCCGGCAAACTCGA    238

XM          TAGGAGAAATGTGCTTCTTGGTCTCGGAGGTCTATATGGCGTGGCTGGTCTGGGCACAGA    412
XR          CAGGAGGAACATGCTCGTCGGCATGGGCGGACTTTACGGCGCCACCACCCTAGCCTCCAC    294
ObPPO2      ------------------------------------------------------------    0
ObPPO1      CCGGAGAAACATGCTCCTCGGCCTCGGTGGCCTCTACGGCGCCACCAACCTCTTCTCCAT    289
OaPPO1      CCGGAGAAACATGCTCCTCGGCCTCGGCGGCCTCTACGGCGCCACCAACCTCTTCTCCAT    298

XM          CTCGTTCGCCTTTGCCAAGCCCGTGGCCCCACCAGACGTATCTAAATGTGGAGCTGCCGA    472
XR          TCCATCCGCCGCGGCGAACCCCATACAGGCGCCGGAGCTGAACAAGTGCGGCACGGCCAC    354
ObPPO2      ------------------------------------------------------------    0
ObPPO1      CCCCAACGCCTCCGCGAACCCGGTTCTAGCCCCCGATTTCGACAAGTGCGGCCCCGTCTC    349
OaPPO1      CCCCACCGCCTCCGCGAACCCCGTTCTAGCCCCCGATTTCGACAAGTGCGGCCCCGTCTC    358

XM          CTTGCC---AACCG---GGGCCACCCCCACCGACTGCTGCCCACCTCCGGCCACCAAAAT    526
XR          CAACTTGAACAACGGGGAGAAACTCGACATCAACTGCTGCCCCCCGATATCCGACCGAAT    414
ObPPO2      ------------------------------------------------------------    0
ObPPO1      CGACGCCAACTCCGGCCAGGTCCTTGAAGGGGTCAACTGCTGCCTCATCACCGAAGAAAT    409
OaPPO1      CGACGCCAACTCCGGTCAGCTCCTCGAAGGGGTCGACTGCTGCCTCATCACCGAAGAAAT    418

XM          CATAGACTTTAAACTCCCCGCCCCAGGACAACTTCGCGTCAGGCTGGCGGCTCACGCCGT    586
XR          CATCGACTACAAACTCCCCCCCGTCTTCCAGATGAAAATCCGTCCCTCCGCCCACCGCGT    474
ObPPO2      ------------------------------------------------------------    0
ObPPO1      CGCCGACTACAAACTCCCCC---CGTCCGTCATGAAATTCCGGCCGCCGGCCCACCGTGT    466
OaPPO1      CGAAGACTACAAACTCCCCT---CATCCGTCATGAAATTCCGGCCGCCGGCCCACCGTGT    475
```

```
XM          TGACCAAGCCTACATAGAGAAATACTCCAAAGCCATCGAGCTCATGAAAGCCCTCCCCGA    646
XR          ATCCCCCGAGTACATGTTCAAGTTCAACACCGCCATCGACCGCATGAAGCGCCTCCCGAA    534
ObPPO2      ------------------------------------------------------------    0
ObPPO1      TACGCCGGATTATGTTGCGAAGTATAACCTGGCGATCCAGAAGATGAAGGAGTTGCCGGA    526
OaPPO1      TACGCCAGATTATGTTTTGAAGTATAACCTGGCGATCCAGAAGATGAAGCAGCTTCCGGA    535

XM          TGACGATCCACGCAGCTTCAAGCAACAAGCCAACGTGCATTGCGCTTATTGCGACGGCGC    706
XR          AGACGATCCGCGTAACTTCATGCAGCAAGCGAACATCCACTGCGCGTACTGCAACGGCGC    594
ObPPO2      ------------------------------------------------------------    0
ObPPO1      CACCGATCCCCGGAGTTTCGTGAATCAAGCTAATATCCACTGCGCGTACTGCAACACCGC    586
OaPPO1      CACCGATCCCCGGAGTTTCATGAATCAAGCTAATATCCACTGCGCGTACTGCAACACCGC    595

XM          CTACGACCAAGCC---GGGTTCCCGGACCTGGAGCTCCAAATCCACAACTCCTGGCTCTT    763
XR          GTACGATCAGCCG---GGGCAGGGCACACTCGACCTCCAAATCCACAACTCATGGCTCTT    651
ObPPO2      ------------------------------------------------------------    0
ObPPO1      CTACAAACAAGGCGGCGGCGACGGCACCGTGCCCCTCCAAATCCACAATTCCTGGCTCTT    646
OaPPO1      CTACAAACAAGGCGGCGGCGACGGCACCGTGCCACTCCAAATCCACAATTCCTGGCTCTT    655

XM          CTTCCCCTTCCACCGCTACTACCTATACTTCTACGAAAAGATCTTGGGCAAGCTCATCAA    823
XR          CTTCCCCTTCCACAGATGGTATCTATACTTCTACGAAAGAATCCTCGGGAAACTCATCGG    711
ObPPO2      ------------------------------------------------------------    0
ObPPO1      CTTCCCCTTCCACCGCTGGTACTTATACTTCTACGAGCGGATTCTGGGCCAACTGATTGG    706
OaPPO1      CTTCCCCTTCCACCGCTGGTACTTATACTTCTACGAGCGGATTCTGGGGCAACTGATCGG    715

XM          CGACCCCACATTCGCTTTGCCCTTTTGGAACTGGGACTCGCCAGCTGGCATGCAGCTGCC    883
XR          AGATCCCACCTTCGCCTTGCCCTTCTGGAACTGGGACAACCCTAAAGGCATGACAATGCC    771
ObPPO2      --------CCTTCGCCATGCCCTTCTGGAACTGGGACAACCCTAAAGGAATGACAATCCC    52
ObPPO1      CGATCCCACCTTCGCCCTCCCCTTCTGGAACTGGGACAACCCTAAAGGAATGACAATCCC    766
OaPPO1      TGATCCCACCTTCGCCATGCCCTTCTGGAACTGGGACAACCCTAAAGGAATGACAATCCC    775
                   * *****  * ***** ************   **    ** ***    * **
```

```
XM          CGCTCTTTTCGCCAACCCAAAGTCCCCTCTCTACGACAAGTTCCGTGCCGCCAGCCACCA    943
XR          CCCCATGTTCGTTGACCCCAAATCCTCCCTCTACGACAAGAAGCGCAACCAGGAACACCT    831
ObPPO2      TCCGATGTTCAACATCGTAGGTTCTCCAATCTACGACGAGAAACGCGAGCCGACTCACCT    112
ObPPO1      TCCGATGTTCAACATCGTAGGTTCTCCAATCTACGACGAAAAACGTGAGCCGACTCACCT    826
OaPPO1      TCCCATGTTCAACATCGTAGGTTCTCCAATCTACGACGAAAAGCGCGAGCCGACTCACCT    835
              *  * ***     *      **  *  ******** *    **          ****

XM          GCCGCCGACCCTCATCGATCTCGACTACAACGGCACCGACGAAAAAGTGTCGAACGAAAC    1003
XR          GCCGCCGGCGGTGGTGGATCTCGGCCTTACGAACAGCACCGACACTCTCCAGGTGGTGGC    891
ObPPO2      ---CACCTCCATCGTCGACCTCGGACGTACCGGCAGCACCGACCCTCTGCAAGTCGTGGC    169
ObPPO1      ---CACCTCCATCGTCGACCTCGGCCTCACCGGCAGCACCGACCCTCTCCAAGTCGTCGC    883
OaPPO1      ---CACCTCCATCGTCGACCTCGGACTCACCGGCAGCACCGACCCTCTGCAAGTCGTCTC    892
                 *  *  *  * ** ****     *    ** *  ***     *       *   *

XM          CCAAATCAACGCCAACCTCAAGATCATGTACAGGCAAATGGTGTCCAACGCCAAGAACCC    1063
XR          C---------AACAACCTCACCATCATGTACAGCGAGATGATCCGCGGCAACTCCGACGC    942
ObPPO2      C---------AACAATCTCACCATCATGTACTCCGAAATGGTCAGAGGAAACAACGATGT    220
ObPPO1      C---------AACAATCTCACCGTCATGTACTCCGAAATGGTCAGAGGAAACAACGATGT    934
OaPPO1      C---------AACAATCTCACCGTCATGTACTCCGAAATGGTCAGAGGAAACAACGATGT    943
            *           *** ****   ********    * *** *         *    *

XM          TCAGCTATTCTTTGGCAACCCCTACAGGGCCGGTGACGAGCCGGATCCGGG---TGCCGG    1120
XR          CAACGACTTCATGGGGCAGGCTTACCGCGAGGGTGACGAGAGCTCGGCCAAGGCGGCGGG    1002
ObPPO2      GTTCGATTTCATGGGACAACCTTACCGCCTTGGAACTCCGGTGAGCCCCGG---CGCCGG    277
ObPPO1      GTTCGATTTCATGGGACAACCTTACCGCCTCGGAACTCCAGTCAGCCCTGG---CGCTGG    991
OaPPO1      ATTCGATTTCATGGGACAACCTTACCGCTTGGGAACTCCTGTGAGCCCCGG---CGCCGG    1000
                   *** * **  *  * *** *    **              *       ** **

XM          TTCGATCGAATCGACCCCACACGGGCCGGTTCATCTCTGGA---CCGGTGACAACACCCA    1177
XR          TTCCTCGGAGCGTGGCTCCCACACATCCGTCCACGCGTGGGTTGGAGATTTCAAGGATCA    1062
ObPPO2      AGCCTCCGAGCGCGGGTCCCACACCTCTATCCACATCTTCGTCG---GAGACAGCCGCCA    334
ObPPO1      AGCTTCCGAGCGCGGTTCCCACACCTCGATCCACATCTTCGCCG---GCGACACGCGCCA    1048
OaPPO1      AGCCTCCGAGCGCGGGTCCCACACCTCGATCCACATCTTCGCCG---GCGACTCCCGCCA    1057
              *    **        * ***    *  * **    *             *      **
```

```
XM          GCCCAATTTTGAGGACATGGGGAACTTTTATTCCGCTGCCAGGGACCCCATATTTTTCTC    1237
XR          GCCCAGCGGTGAAGACATGGGGAACTTCTACTCCGCCGGGCGGGACCCGCTATTCTACAG    1122
ObPPO2      GCCGAGGAAGGAGAACATGGGCAACTTCTACTCCGCGGGGCGGGACCCGCTTTTCTACTG    394
ObPPO1      GCCAAGGAGGGAGAACATGGGCAACTTCTACTCTGCGGGGCGGGACCCGCTTTTCTACTG    1108
OaPPO1      GCCGAGGAGGGAGAACATGGGCAACTTCTACTCCGCAGGGCGGGACCCGCTCTTCTACTG    1117
            *** *     **  ******* ***** ** ** ** *   *******  * ** * *

XM          GCATCACTCGAATGTGGATCGGATGTGGAGCATATGGAAAACACTGGGAGGCAAAAG---    1294
XR          CCACCACGCCAACGTGGACCGCATGTGGACGCTGTGGCAGTACTTCCTCCCGAGCAACAA    1182
ObPPO2      CCACCACGCCAACGTCGACCGCATGTGGACGGTCTGGCAGAAGCTCCCGTCAACCG---T    451
ObPPO1      CCACCACGCCAACGTTGACCGCATGTGGACGATCTGGCAGAAGATCCCGTCGACTG---T    1165
OaPPO1      CCACCACGCCAACGTCGACCGCATGTGGACGATCTGGCAGAAGATCCCGTCGACCG---T    1174
             ** *** * ** ** ** ** *******   * *** *     *       *

XM          ---------AACTGATATTTCTGAGTCGGACTGGTTAGACTCCGGGTTCTTGTTTTACAA    1345
XR          AGTCCCCGACAAGAGGATAACCGACACTGATTTCTTGAACACTGCTTTCTTATTCTACGA    1242
ObPPO2      CATCCCGAAGAAAACGATTGACGATCCCGATTTCCTTAATGCCACCTTCCTCCTCTACGA    511
ObPPO1      CATTCCGAAGAAAACGATCGACGATCCTGATTTCCTCAACGCCACCTTCCTCCTCTACGA    1225
OaPPO1      AATTCCCAAAAAAACAATCGACGATCCCGATTTCCTCAACGCCTCCTTCCTCCTCTACGA    1234
                      *     **    **  * ** *   *  *  *    *** *  * *** *

XM          CGAGAACGCCGAGTTAGTCCGAGTTAAAGTTCGTGACTCTCTGGAGTCTAAGAGACTAGG    1405
XR          CGAGAACGCGCAGTTGGTGCGCGTTACGGTGAAGGACTGCCTGGACAACTTACGAATGGG    1302
ObPPO2      CGAGAATGGCAAGCTCGTCCGCGTCTCCGTCAAAGACACCATCGACAACCGCAAAATGGG    571
ObPPO1      CGAGAACGGCCAGCTTGTCCGCGTCTCCGTGAAAGACAGCGTCGACAATCGCAAAATGGG    1285
OaPPO1      CGAGTACGGCAAACTCGTCCGCGTCTCCGTCAAAGACACCATCGACAACCGCAAAATGGG    1294
            **** * *   *  * ** ** **    **    ***    * **         * * **

XM          GTATGTATACCAAGAGGTTGACATTCCATGGCTGCAGTCTAAGCCAACCCCGCGAAGGGC    1465
XR          ATACGATTTCGAGAGGATCGATCTTCCGTGGCTGGACTACAGACCGCCGCGGCAGTCGGC    1362
ObPPO2      GTACGACTTCGAGAGGATCGACTTGCCGTGGCAGGACTACCGCCCGCCACGGCAGACTGC    631
ObPPO1      ATACGACTTCGAGAGGATCGACCTGCCGTGGCAGGACTACCGCCCGCCACGACAGACTGC    1345
OaPPO1      ATACGACTTCGAGAGGATCGACCTGCCGTGGCAGGACTACCGCCCGCCACGACAGACTGC    1354
             ** *  * * *   * * **  * ** **** * * *     **  * *  *     **
```

```
XM          AAAGCTGGCAATAAGCAAAATAGCGAAGAAGCTGGGAGTTGCGCACGCGGCTGAGAGCTC    1525
XR          TAGGGCGAAGATCAACAGGGCAAGCTCCACCGCCGCACCC----------------------    1402
ObPPO2      GAAGGCGAAGATAAACCGCGCCTCGGCGCCTAAGCTGCCG----------------------    671
ObPPO1      AAAGGCGAAGATAAACCGCACCTCGGCGCCCAGGCCGCCT----------------------    1385
OaPPO1      AAAGGCGAAGATAAACCGCACCTCGGCGCCTAGGCCGCCT----------------------    1394
             * *  *   ** * *

XM          CACCAAGATCGTGGCGGGGAGGGCGTTCCCAATAAATCTGGAGACCAAGATAAGCACGGT    1585
XR          ----------AAGGCGGCGGCGCTGTTCCCTCTGACGCTTGACGGAGTGGTGAGG-----    1447
ObPPO2      ----------AAGGCGCGGTCTCTCTTCCCGCTAAAACTCGATAAGGTGGTGAGG-----    716
ObPPO1      ----------AAGGCGAAGTCTCTCTTCCCGCTAAAACTCGATAAGGTGGTGAGG-----    1430
OaPPO1      ----------AAGGCGAAGTCTCTCTTCCCGCTAAAACTCGATAAGGTGGTGAGG-----    1439
                      ****  *      *****  * *  ** **      * * **

XM          GGTGCCGAGGCCGAAGCAGAAGAAGAGGAACAAGAAGGAGAAGGAGGAGGAGGAGGAGAT    1645
XR          ------------------TTCCAGGTGGACAAGACGAAGAAGGGGAAAGCAGACGAGTC    1488
ObPPO2      ------------------TTTGAAGTTGACAAGACCACGAAGGGGGTGGCCGACGAGTC    757
ObPPO1      ------------------TTCGAAGTCGACAAGACGGCAAAGGGGGTGGCGGATGAGTC    1471
OaPPO1      ------------------TTCGAAGTCGACAAGACGGCAAAGGGGGTGGCGGACGAGTC    1480
                               *     ******     **** *   *  ** ***

XM          ATTGGTGATCGAAGGGATTGAGTTTGACAGGGACGTGGCGGTGAAGTTTGACGTGTATGT    1705
XR          TCTGGTGCTCGAGGATATCAAGGTCGACACGACGAAGCTGCTCAAGTTTGACGTGTTTGT    1548
ObPPO2      GATCTTGCTGGAGAACATCATCGTCGACACGTCGAAGTTCTTGAAGTTCGACGTGTTTGT    817
ObPPO1      GATCTTGCTAGAGAACATCACCGTCGACACGTCGAAGTTCTTGAAGTTCGACGTGTTTGT    1531
OaPPO1      GATCTTGCTGGAGAACATCACCGTCGACACGTCGAAGTTCTTGAAGTTCGACGTGTTCGT    1540
              *  ** * **    **     * **** *     *    * ***** *******  **

XM          GAATGACGTGGACGACTTGCCGAGTGGGCCCGACAAGACGGAGTTTGCCGGAAGCTTTGT    1765
XR          GAACGACGAGGACGATAACCCCGGCGAGCTCGACAAGGCCGCGTACTTGGGCACGTACGC    1608
ObPPO2      GAACGACGAGGACGACGCGCCGAATGAGTTGGACAAGGCAGCATACGCGGGGACGTACGC    877
ObPPO1      GAACGACGAGGACGACGCTCCGAACGAGCTGGACAAGGCGGCGTACGCGGGGACGTACGC    1591
OaPPO1      GAACGACGAGGACGACGCCCCGAACGAGCTGGACAAGGCGGCGTACGCAGGTACCTACGC    1600
            *** **** ******    **    * *   ****** * *  *     ** *  *  *
```

```
XM          GAGCGTGCCGCACAGGCATAAGCACAAGAAGAAGATCAACACGATCTTGAGGTTGGGGTT    1825
XR          GCAGGTGCCGCATAAATCTGCGAACAATAATAGTTCGACATCTTCCATTAAGCTGAAGCT    1668
ObPPO2      GCAGGTGCCGCACAAGAGTGACAACGGG---AAGGCGACGTCATCGATAAAGCTGAGGTT    934
ObPPO1      ACAGGTGCCGCACAAGAGCGACAACGGG---AAGGCTACGTCGTCGATAAAGCTGAGGCT    1648
OaPPO1      ACAGGTGCCGCACAAGAGCGACAACGGG---AAGGCTACGTCGTCGATAAAGCTGAGGCT    1657
                ******** *         **      *     *   *     * * * **  * *

XM          GACGGACTTGTTGGAGGACATTGAAGCCGAGGATGATGAGAGCGTGGTGGTGACTTTGAT    1885
XR          GACTGAGCTCTACGATGACATGGATATCGATGACGACGACACAATCGTGGTGACGCTGGT    1728
ObPPO2      GACGGAGCTGTACGAGGATATGGACATCGAGGACGACGACTCGGTCGTGGTGACGATCGT    994
ObPPO1      GACGGAGCTGTACGAGGATATGGATATCGACGACGACGACTCGATCGTGGTGACCATCGT    1708
OaPPO1      GACGGAGCTGTACGAGGATATGGATATCGACGACGACGACTCGATCGTGGTGACGATCGT    1717
            *** **  * *  ** ** ** **   *** ** ** **     * ********  *  *

XM          ACCCAAGTATGGG------GCTGTCAAGATTGGCGGTGTCAAGATTGAATTTGCTTCTTA    1939
XR          GCCGAGGCATCAAGGCCCGGT-GTTACCATTGGTGGTATCAAGATTGTCGAGAATCCACC    1787
ObPPO2      GCCGCGCCACGAGGGGCCCGGGGTCACCATTGGTGGTATCAAAATCGTCGCCAACTGATC    1054
ObPPO1      GCCGCGCCACAAGGGGCCGGGGGTCACCATTGGTGGTATCAAGATCGTGGCCAACTGATC    1768
OaPPO1      GCCGCGCCACAAGGGGCCGGGGGTCACCATTGGTGGTATCAAGATCGTCGCCAGCTGATC    1777
             **     *          *  ** *  ***** *** **** ** *

XM          GA-----------------------------------TTAATTGAAGCATAAACCAAAT    1963
XR          GAAACCGGCTGCCGGTAGTACTTGATCATTAGTTTTAATTAATTGATGCATGATTTCCAC    1847
ObPPO2      GGTCGACGGCGGTGTTGGGAGTAATTAATCTGTTGA-----TGGG--------------    1094
ObPPO1      GGTCGACGGCGGTGTTTGAATTGATTAGTTGATTAA-----TCTGTTGTGTAATGTGCAT    1823
OaPPO1      GGTCGACGACGGTATTGG-ATTGATTAGTTGATTAA-----TCTGTTGTGTAATGT----    1827
            *                                              *

XM          TTCCATTCATCATCAAATATGTTTTAGTCCTTTAATTTGGTTTTTTTTTTCTTTCTTCGT    2023
XR          ATGCA----TGCTCGTGTTTGC----TAGCTTCCGATCGTTTTCCTATGTTTCTGATCGA    1899
ObPPO2      ------------------------------------------------------------    1094
ObPPO1      GTG------TGAAT-----------AAGATTTCTGTGGGAAATCTGTGTGA--ATGTGA    1863
OaPPO1      ------------------------------------------------------------    1827
```

Figure 4. Clustal Omega multiple alignment of cloned PPO nucleotide sequences and PPO sequences of *E. guttatus* (XR) and *P. avium* (XM).

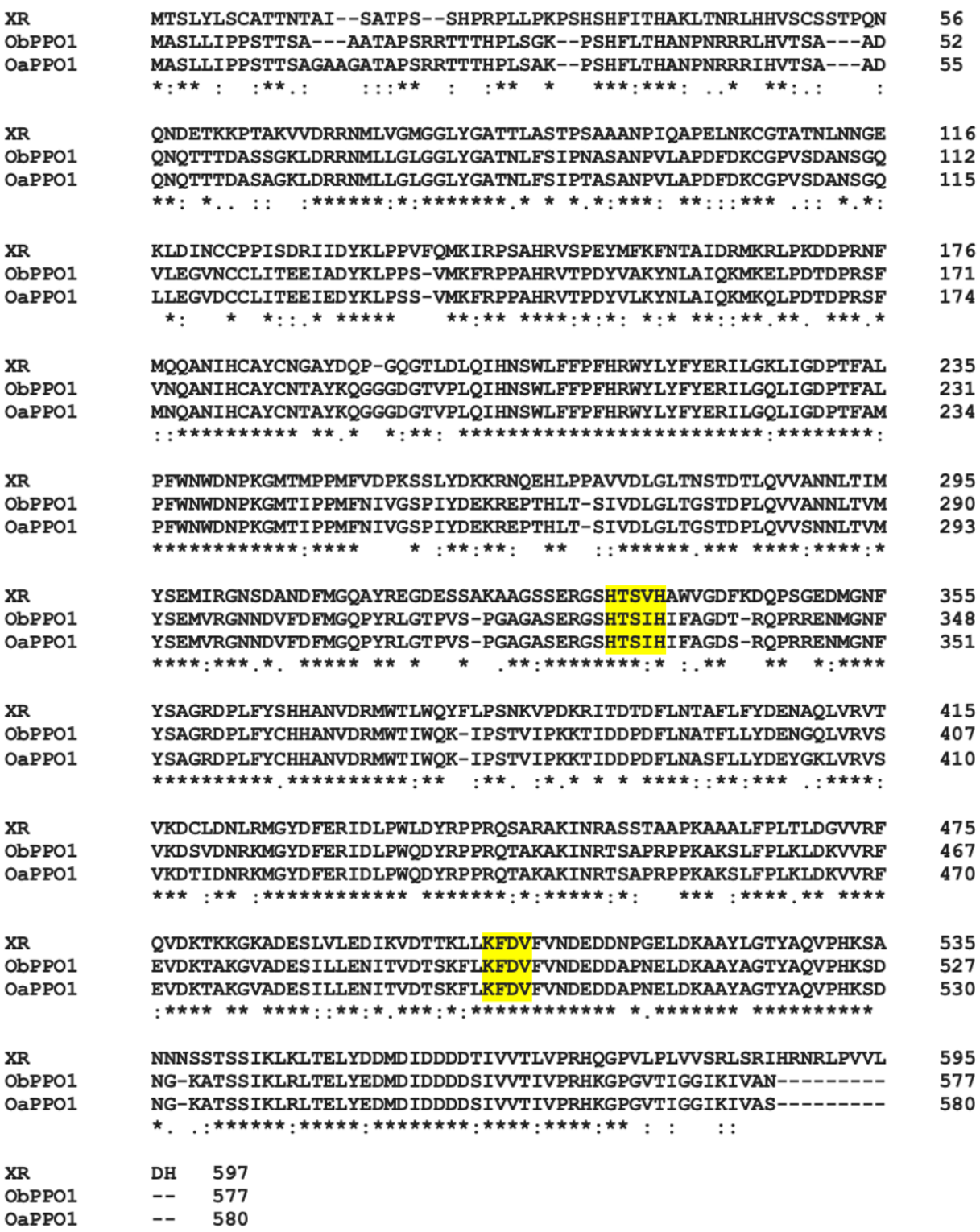

Figure 5. Clustal Omega multiple alignment of the putative polyphenol oxidase amino acid sequences of *O. basilicum* (ObPPO1), *O. americanum* (OaPPO1), and *E. guttatus* (XR) (Sievers et al., 2011). The two basil sequences demonstrate 86.5% identity to one another, and the two yellow boxes represent the enzyme's copper binding sites that are largely conserved among plants.

Development of an RNAi construct was also undertaken using the ObPPO1 gene. While sequencing results indicate that the attempt was successful, experimental verification of knockdown *in planta* has not yet been carried out. Although the RNAi construct was developed from *O. basilicum* cDNA, it is theoretically possible that the construct would also induce PPO knockdown in *O. americanum* due to the high level of PPO sequence identity between the two species. A diagrammatic representation of the construct and the hairpin mRNA molecule is presented in Figure 3. Figure 6 shows the modified ObPPO1 sequence inserted into the vector and identifies primers, restriction enzyme sites, and the complementary segments relevant for the formation of the mRNA secondary structure. This secondary structure can be visualized by *in silico* modelling based on identification of hybridizing regions of the sequence (Fig. 7).

GCTCTAGACCCTCTCCGGCAAGCCTTCACATTTCCTCACCCATGCAAACCCTAACCGCCGCCGCCTCCATGTCACC
TCAGCCGCCGACCAAAACCAAACCACCACTGACGCCTCCTCCGGCAAACTCGACCGGAGAAACATGCTCCTCGGC
CTCGGTGGCCTCTACGGCGCCACCAACCTCTTCTCCATCCCCAACGCCTCCGCGAACCCGGTTCTAGCCCCCGATT
TCGACAAGTGCGGCCCCGTCTCCGACGCCAACTCCGGCCAGGTCCTTGAAGGGGTCAACTGCTGCCTCATCACCG
AAGAAATCGCCGACTACAAACTCCCCCCGTCCGTCATGAAATTCCGGGGAATTCGGTGAGGTGAGTCGGCTCACG
TTTTTCGTCGTAGATTGGAGAACCTACGATGTTGAACATCGGAGGGATTGTCATTCCTTTAGGGTTGTCCCAGTTCC
AGAAGGGGAGGGCGAAGGTGGGATCGCCAATCAGTTGGCCCAGAATCCGCTCGTAGAAGTATAAGTACCAGCGG
TGGAAGGGGAAGAAGAGCCAGGAATTGTGGATTTGGAGGGGCACGGTGCCGTCGCCGCCGCCTTGTTTGTAGGC
GGTGTTGCAGTACGCGCAGTGGATATTAGCTTGATTCACGAAACTCCGGGGATCGGTGTCCGGCAACTCCTTCATC
TTCTGGATCGCCAGGTTATACTTCGCAACATAATCCGGCGTAACACGGTGGGCCGGCGGCCGGAATTTCATGACG
GACGGGGGGAGTTTGTAGTCGGCGATTTCTTCGGTGATGAGGCAGCAGTTGACCCCTTCAAGGACCTGGCCGGAG
TTGGCGTCGGAGACGGGGCCGCACTTGTCGAAATCGGGGGCTAGAACCGGGTTCGCGGAGGCGTTGGGGATGGA
GAAGAGGTTGGTGGCGCCGTAGAGGCCACCGAGGCCGAGGAGCATGTTTCTCCGGTCGAGTTTGCCGGAGGAGG
CGTCAGTGGTGGTTTGGTTTTGGTCGGCGGCTGAGGTGACATGGAGGCGGCGGCGGTTAGGGTTTGCATGGGTGA
GGAAATGTGAAGGCTTGCCGGAGAGGGGGATCCCG

Figure 6. The modified ObPPO1 DNA sequence inserted in the construct to generate a hairpin when transcribed. The bolded and underlined segments represent the primers used to amplify the initial fragments from the ObPPO1 cDNA. The highlighted segments are the restriction enzyme sites used in development of the construct. The black regions represent the complementary segments that allow the hairpin to form, and the blue segment comprises the loop that forms between them.

Figure 7. An *in silico* prediction of the construct insert's secondary structure. The long complementary region makes up the major length of the structure and the loop section also adopts its own complex secondary structures (Zuker, 2003).

Discussion

This work has established the nucleotide and amino acid sequences of three genes encoding PPO in *O. basilicum* and *O. americanum*, thus filling a gap in the literature. The PPO sequences of the two species have high sequence identity and contain the two widely conserved copper-binding domains that characterize the enzyme are present in both sequences (Malviya, Srivastava, Diwakar, & Mishra, 2011).

The identification of two PPO sequences in *O. basilicum* is notable. These distinct sequences indicate that basil produces multiple variants of the PPO enzyme, an idea that is consistent with the literature on polyphenol oxidase and its role in other species (Martinez & Whitaker, 1995). The enzymes exist in diverse families with variants having different properties and tissue localizations, and it is likely that several forms of the enzyme remain to be identified in both basil species (Malviya, Srivastava, Diwakar, & Mishra, 2011). For example, the PPO gene family of tomato contains seven different genes that are differentially expressed spatially and temporally (Newman et al., 2003). An analysis of the genomes of twenty-five land plants found the number of genes in the PPO family is generally between four and eleven in eudicotyledonous plants, though it varies considerably between species (Tran et al., 2012). The PPO sequences identified in this study showed highest similarity to the polyphenol oxidase I sequence of *E. guttatus*, a species containing nine putative PPO variants. Basil may contain a PPO gene family of a similar size, though significant variation between closely related species implies that it is difficult make such generalizations, as other dicots like cucumber and soybean contain only a single PPO sequence (Tran et al., 2012). It is also possible that the two varying sequences identified in this study instead represent different alleles or redundant genes.

Further exploration of the basil PPO gene family is warranted. Short of genome sequencing, more information can be gathered about whether these sequences represent varying alleles or distinct genes through restriction fragment length polymorphism (RFLP) mapping studies to gain insight into their localization within the genome (Newman et al., 2003). The identification of additional genes encoding PPO can be accomplished through more extensive cloning trials or by bioinformatic gene prediction from genomic or transcriptomic data. Isolation of the genes involved in the PPO gene family presents opportunities for the study of their differential expression and functions. Differential expression can be studied best through quantitative PCR (qPCR) studies, in which the expression of these sequences in RNA extracted from basil tissues

of different plant organs and developmental stages can be quantified. Such studies will rely on knowledge of the sequence differences between variants such that variant-specific PCR primers can be designed. Understanding of the spatial and temporal PPO expression will inform the identification of the variants most appropriate for knockdown in the development of a browning-resistant cultivar. For example, PPO expressed in the leaves will be much more important to regulate than PPO expressed in the roots. Therefore, further study is necessary to identify additional genes in the basil PPO family and to understand the intracies of their respective functions and localizations.

At this time, a reproducible procedure has not been published for the transformation and regeneration of these varieties of basil in tissue culture, and many trials in the [PI name] lab have suggested that they may be recalcitrant *in vitro*. These difficulties have prevented the development of a basil plant with reduced polyphenol oxidase expression using the RNAi constructs developed at this time. Future success in developing regeneration protocol for basil may encourage further studies in polyphenol oxidase knockdown using these constructs.

This study was successful in generating sequence data for three previously uncharacterized genes encoding PPO in *O. basilicum* and *O. americanum*, including two variants of the PPO gene in *O. basilicum*. This data was used for sequence alignment analysis that found notable homology to the published PPO sequence of *E. guttatus* as well as similarity between the three basil sequences. This work was also successful in using this sequence data for the development of a construct for the reduction of PPO expression via RNAi. As the project did not conclude with the development of a transformed plant, the major significance of the RNAi component of this work lies in the establishment of an adapted RNAi construct development protocol that is specialized for basil.

Acknowledgements

I would like to thank Dr. Clark for giving me the opportunity to undertake this project and for giving me the support, advising, and resources to do so. I would also like to thank my mentors Keun Cho and Samantha Burrell for their generous guidance and training that enabled me to complete this study and develop as a scientist.

References

Ben-Naim, Y., Falach, L., & Cohen, Y. (2015). Resistance against basil downy mildew in *Ocimum* species. *Phytopathology, 105*(6), 778-785. doi:10.1094/phyto-11-14-0295-r

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D. S., & Karp, P. D. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research, 44*(D1), D471-D480. doi:10.1093/nar/gkv1164

Dexter, R., Qualley, A., Kish, C. M., Ma, C. J., Koeduka, T., Nagegowda, D. A., Dudareva, N., Pichersky, E., & Clark, D. (2007). Characterization of a petunia acetyltransferase involved in the biosynthesis of the floral volatile isoeugenol. *Plant Journal, 49*(2), 265-275. doi:10.1111/j.1365-313X.2006.02954.x

Friedman, M. (1996). Food browning and its prevention: an overview. *Journal of Agricultural and Food Chemistry, 44*(3), 631-653. doi:10.1021/jf950394r

Frohman, M. A. (1993). Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs - thermal RACE. *Methods in Enzymology, 218*, 340-356.

Hannon, G. J. (2002). RNA interference. *Nature, 418*(6894), 244-251. doi:10.1038/418244a

Malviya, N., Srivastava, M., Diwakar, S. K., & Mishra, S. K. (2011). Insights to sequence information of polyphenol oxidase enzyme from different source organisms. *Applied Biochemistry and Biotechnology, 165*(2), 397-405. doi:10.1007/s12010-011-9259-2

Martinez, M. V., & Whitaker, J. R. (1995). The biochemistry and control of enzymatic browning. *Trends in Food Science & Technology, 6*(6), 195-200. doi:10.1016/s0924-2244(00)89054-8

Newman, S. M., Eannetta, N. T., Yu, H. F., Prince, J. P., Devicente, M. C., Tanksley, S. D., & Steffens, J. C. (1993). Organization of the tomato polyphenol oxidase gene family. *Plant Molecular Biology, 21*(6), 1035-1051. doi:10.1007/bf00023601

Putievsky, E. & Galambosi, Bertalan. (1999). Production Systems of Sweet Basil. Basil: The Genus Ocimum. 39-65.

Queiroz, C., Lopes, M. L. M., Fialho, E., & Valente-Mesquita, V. L. (2008). Polyphenol oxidase: Characteristics and mechanisms of browning control. *Food Reviews International, 24*(4), 361-375. doi:10.1080/87559120802089332

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W. Z., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology, 7*, 6. doi:10.1038/msb.2011.75

Tran, L. T., Taylor, J. S., & Constabel, C. P. (2012). The polyphenol oxidase gene family in land plants: Lineage-specific duplication and expansion. *Bmc Genomics, 13*, 12. doi:10.1186/1471-2164-13-395

Waltz, E. (2015). Nonbrowning GM apple cleared for market. *Nature Biotechnology, 33*(4), 326-327. doi:10.1038/nbt0415-326c

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology, 7*(1-2), 203-214. doi:10.1089/10665270050081478

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research, 31*(13), 3406-3415. doi:10.1093/nar/gkg595