MOLECULAR GENETICS AND GENOMICS OF EFFECTIVE PATHOGEN RESPONSE IN *B. VULGARIS* L.

L. David Kuykendall*, Jonathan Shao, Kate Burdekin and Tatsiana Kurushko
Molecular Plant Pathology Laboratory, USDA-ARS, 10300 Baltimore Avenue, Beltsville, MD 20705

Introduction:

McGrath et al., (2004) constructed a Bacterial Artificial Chromosome (BAC) genomic library of sugar beet hybrid H20. In collaboration with Dr. McGrath at MSU, our laboratory successfully identified, sequenced and annotated about 39 Kb (Kuykendall et al., 2007) of a BAC clone carrying *NPR1*, a key gene controlling disease resistance in plants. The NPR1 protein is required for induced systemic resistance in sugar beet in much the same manner as in *Arabidopsis thaliana* (Bargabus-Larson and Jacobsen, 2007). Evolutionary conservation of *Beta vulgaris FLOWERING LOCUS C* that mediates vernalization response was reported (Reeves et al., 2007).

Complete gene content of the entire 130-Kb *NPR1*-carrying BAC was reported (Kuykendall et al., 2008). Functions of predicted protein products of four core plant genes were predicted and conserved microsynteny established in *B. vulgaris*, *Medicago truncatula* and *Populus trichocarpa*, except in *M. truncatula* where *HSF* is lacking (Kuykendall et al., 2008). *Coe1*, a DNA transposase gene within a LTR-retrotransposon, a novel arrangement, was discovered upstream from *CaMP* on the *NPR1* BAC (Kuykendall et. al., 2008). A nest of long terminal repeat LTR-retrotransposons, just upstream of the *NPR1* gene, contains both a *copia*-like and a *gypsy*-like transposon, *SCHULTE* and *SCHMIDT* respectively, within a much older retroelement (Kuykendall et al., in press).

In this study, *Solanum lycopersicum* and *Vitis vinifera* were assessed for micro-colinearity of *HSF*, *NPR1*, *CaMP* and *CK1PK* genes. HSF, identified by BLAST as a developmental heat shock factor-like protein, and NPR1 proteins activate expression of genes whose products are required for either developmental processes or effective defense against pathogen attack, respectively. CK1 activates expression of genes whose products are involved with chromosome partitioning and circadian rhythm. A role of CaMP protein encoded by *CaMP* is as yet undefined, but results of experiments herein reported demonstrate that transcription of *CaMP* in *B. vulgaris* is upregulated by *Erwinia betavasculorum* challenge, and it can therefore be inferred that the CaMP protein plays a role in an effective pathogen response.

Materials and Methods:

Expression Analysis - RNA purification and RT PCR analysis. Seed of sugar beet germplasm C69, obtained from Bob Lewellen (Salinas, CA), were germinated to grow plants in the greenhouse. Biotic stress challenge responsiveness of C69 plants was tested by experimental infection using *E. betavasculorum* strain EB4 [described previously (Kuykendall and Hunter, 2008)]. Sugar beet plants were inoculated by first puncturing two leaf stems per relatively mature plant/leaves and then applying about 100µL of a

fresh 48-hr broth culture of *E. betavasculorum* containing about 100,000,000 colony-forming units (10^8 CFU), whereas controls received sterile LB broth medium. Inoculated and control wounds were both immediately wrapped with Parafilm to prevent desiccation. Inoculated and uninoculated control plants were incubated for seven days in a greenhouse.

Sampling of mid-sized leaves consisted of several about 0.5 cm leaf squares cut adjacent the mid vein, avoiding primary and secondary veins. *B. vulgaris* leaf extracts were obtained by liquid nitrogen freeze & thaw and maceration in a microcentrifuge tube as a minimortar and an autoclavable cone-shaped plastic pestle. RNA purifications from *B. vulgaris* leaf extracts were done using Qiagen (Santa Clarita, CA) RNeasy, adhering closely to the recommendations of the manufacturer. NanoDrop (Wilmington, DE) was used to measure uv absorption to estimate RNA concentration. It is important to note that RNA samples were carefully adjusted to a uniform concentration of 5 or 8 ng per µl then diluted 1:20 to provide a limiting RNA concentration and subjected to reverse transcriptase (RT)-PCR using MasterAmp (Epicenture, Madison, WI) following the manufacturer's specified protocol. A thermal cycler (MJ Research, Inc. PTC 100 Thermocycler, Waltham, MA) was used. The following primers were designed for DNA amplification of an about 400 basepair internal segment of the first exon of *CaMP:*

5' TTGTTCACATTAGTCTGCAATGG 3' (forward) and
5' GTTTCTTCGAGTCCGGTAACTCT 3' (reverse).

Amplification cycle conditions were programmed as follows: 20 min. 60^0C, 50 cycles [(94^0C for 1 min.) (62^0C for 1 min.) (72^0C for 3 min.)], stabilization at 72^0C and 4^0C hold. PCR products, following 2% agarose gel electrophoresis, were visualized by 310 nm *uv* illumination after staining with dilute ethidium bromide. Sequencing reactions were performed on RT-PCR products at the University of Maryland Center for Agricultural Biotechnology (College Park, MD) using an automated DNA sequencer (ABI Prism model 377, Foster City, CA), and data examined to confirm DNA sequence.

Bioinformatic Protein Analysis. Simple Modular Architecture Research Tool, or SMART (Letunic et al., 2009), was used to determine protein domains by examination of the predicted protein sequence (http://smart.embl-heidelberg.de/index2.cgi). Specifically, protein domains within *B. vulgaris* CaMP protein (ABM55247), 525 amino acids long, were determined. Conserved motifs were analyzed by Motif Scan (http://hits.isb-sib.ch/cgi-bin/PFSCAN). The most likely calmoduling-binding site in a peptide sequence was identified with the use of Calmodulin Target (http://calcium.uhnres.utoronto.ca/ctdb/ctdb/home.html). The hmmtop server (http://www.enzim.hu/hmmtop) was used to predict transmembrane helices, and a signal peptide was detected using the SignalP 3.0 and TargetP 1.0 server (http://www.cbs.dtu.dk/services).

Identification of Microcolinearity. BLAST searches, done for protein products of the predicted ORFs of the *B. vulgaris NPR1* BAC (EF101866), were also performed against databases for *V. vinifera* and *S. lycoperscium* (databases: http://www.sgn.cornell.edu/ and http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis). High scoring pairs, the predicted protein products with highly significant matches, were viewed as products of orthologous genes. Corresponding DNA regions are considered microcolinear when two or more orthologous genes are present in close physical proximity, in the same order, and

transcribed in the same relative direction. GenBank accession number CU459230 is representative of chromosome XI of *V. vinifera*, and the subset of particular interest are coordinates 1544959 to 1599363. Genbank accession numbers AC210362 and AC212660 are representative of chromosome VII of *S. lycoperscium*. AC210362 and AC212660 were aligned using SeqMan (DNASTAR, Madison, WI). Colinearity in *B. vulgaris*, *M. truncatula*, and *P. trichocarpa* was previously described (Kuykendall et. al., 2008). Colinearity was determined of four core genes in *V. vinifera* and *S. lycopersicum.. lycoperscium.* Putative genes in these genomic regions were annotated using FgenesH program (Gene finding in Eukaryota, http://linux1.softberry.com/). BLASTP searches were performed on putative genes at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/).

Results and Discussion:

This is the first report of conserved microsynteny of *HSF*, *NPR1*, *CaMP* and *CK1PK* genes in *Vitis vinifera* and *Solanum lycopersicum* (Figure 1). Conserved microsynteny of these four core genes in *V. vinifera* and *S. lycopersicum* (Figure 1) closely resembles the conserved microsynteny discovered in *B. vulgaris*, *M. truncatula* and *P. trichocarpa* (Kuykendall et al., 2008). Conservation of a cluster of orthologous genes (COG) consisting of these four core plant genes in *B. vulgaris* as well as four other eudicots reflects positive natural selection for close physical proximity. The hypothesis that this gene arrangement may contribute to coordinated expression of two or more genes whose products are critical for managing biotic and other environmental stresses was supported. Increased understanding of molecular mechanisms underlying effective, genetically-programmed responses to pathogens is sought with the goal of contributing to full realization of plant genome potential for effectively resisting microbial invasion.

Detailed analyses of gene arrangements are required to define functional relationships, if any. Similar small gaps between core plant genes *NPR1*, *CaMP* and *CK1PK* occur in *M. truncatula*, *P. trichocarpa* and *V. vinifera,* whereas in *B. vulgaris* and *S. lycopersicum* larger gaps are found between these genes. This difference in core gene spacing is largely, if not entirely, due to mobile, repetitive DNA elements, but it could also reflect an underlying higher degree of molecular genetic and genomic similarity between *S. lycopersicum* and *B. vulgaris* than with other plant species (see Dohm *et al.*, 2009).

Achieving a direct understanding of the significance of gene arrangement and spacing, as seen in Figure 1, is daunting. Tighter gene clustering could lead to better coordinated gene expression in *M. truncatula*, *P. trichocarpa* and *V. vinifera* since all three species exhibit tighter gene spacing than in *B. vulgaris or S. lycopersicum*, but a relationship is unclear. In all five species of eudicots listed, however, *CaMP* and *CK1* are less than 16 Kb apart while *CaMP* and *NPR1* are less than 20 Kb apart in *M. truncatula*, *P. trichocarpa* and *V. vinifera*. Greater conservation of tight spacing, only 3 Kb or less apart, is found between *NPR1* and *HSF* genes in four of the five eudicot species examined [*M. truncatula* lacks *HS*F].

While close proximity of certain core genes in diverse eudicots suggests coordinated expression, genes encoding proteins determining steps in a signaling pathway tend to be clustered in contrast to unrelated genes which tend to be either in a

more distant position or on a distinct linkage group (Wei et. al., 2006). In sugar beet, the relatively enormous 82-Kb gap between *NPR1* and *CaMP* genes is largely accounted for by retroelements (Kuykendall et al., in press). Multiple or single LTR-retrotransposons may modify transcription downstream, but one can deduce that a negative affect on the expression of essential genes would have been counter-selected. Most likely, insertion of a repetitive DNA element would have a survival advantage if and only if it optimized expression of downstream essential genes, such as *NPR1*.

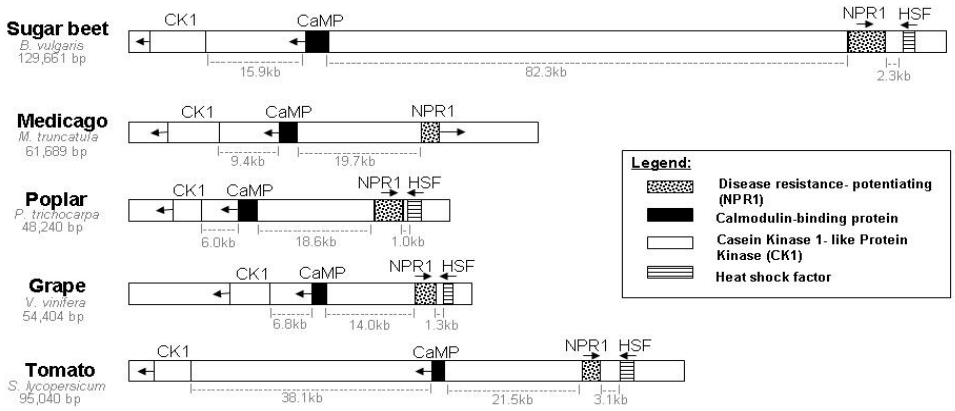

Figure 1: Conserved clustering of orthologous genes, or COGs, in five plant species. Shown above is a schematic representation of a microsyntenous, or closely physically linked, *NPR1* genomic region: (Top) *B. vulgaris* (Genbank EF101866), (Second) *M. truncatula* (Genbank AC124609), (Third) a subset of *P. trichocarpa* NC_008472, (Fourth) *V. vinifera* (Genbank CU459230), (Bottom) *S. lycopersicum* (Genbank AC212660 overlapped with AC210362). Genes are indicated by boxes: HSF (lined box), *NPR1* (gray or dotted box), *CaMP* (black box), *CK1PK* (white box). Arrows indicate the direction of transcription. Gray bars indicate Kb between genes from the end exon of one gene to the end exon of the next gene. DNA regions under comparison are in base pairs. As of this writing, the TIGR-annotated *NPR1* region for *M. truncatula* is only a draft one-half BAC sequence which has a gap in the center.

Bioinformatic analyses performed on the 525-amino acid sequence of CaMP protein product of *B. vulgaris CaMP* gene revealed molecular features. An N-terminal signal peptide with a cleavage site between glutamine and lysine at amino acid positions 17 and 18 was predicted by SignalP 3.0 (Figure 2). SMART analysis found an IQ domain from amino acid positions 131 to 153 as well as four regions of intrinsic disorder. Four phosphorylation sites found by Motif Scan (not shown) are roughly correlated with regions of intrinsic disorder (Iakoucheva et al., 2004). The most likely calmodulin-binding site on the BvCamP was predicted by Calmodulin Target Database at amino acid positions 201-209.

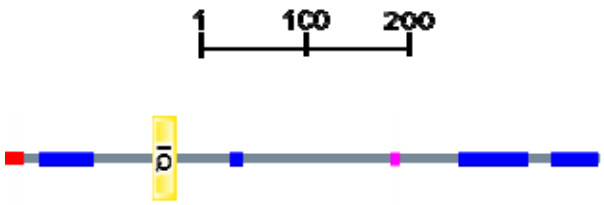

Figure 2: Bioinformatic SMART analysis of predicted CaMP protein of *B. vulgaris*.(ABM55247), 525 amino acids, confidently predicted domains, motifs and features shown here: a signal peptide at positions 1-17; an IQ domain from amino acid positions 131 to 153 (yellow box); four disordered regions at positions 31-79, 199-210, 401-463, 483-525; and a segment of low compositional complexity at positions 341-349.

The 525-amino acid CaMP precursor protein matures into a 508-amino acid protein as cleavage of the 17-amino acid N-terminal signal peptide occurs coincident with targeted passage into chloroplast. Interestingly, Cucumber Mosaic Virus Coat protein also functions as a chloroplast transit protein in infected plants (Xiang et al., 2006); the viral protein has an N-terminal amino acid sequence resembling a signal peptide for chloroplast transit (TP) in the following respects: an alanine-rich uncharged region near the N-terminus followed by a short region rich in basic amino acids, a conserved chloroplast TP phosphorylation motif, and specific proteolytic cleavage upon import into the chloroplast. In summary, bioinformatic analyses of the *B. vulgaris* CaMP classify it as an IQ-domain calmodulin-binding protein with a chloroplast-targeted N-terminal peptide.

CaMP expression analysis using RT PCR of RNA extracted from leaves sampled from greenhouse-grown sugar beets, consistently revealed an approximately 400-bp DNA product with the sequence expected given the *CaMP*-specific primer set employed. In other words, DNA sequence analysis confirmed that the amplified DNA is an internal portion of the first exon of *CaMP*. Thus, plants under biotic stress, i.e., experimental *E. betavasculorum* infection, showed enrichment of a higher level of *CaMP*-specific RNA transcript compared with uninoculated control plants (Figure 3). This new finding suggests that CaMP, a chloroplast protein, may be involved in conferring disease resistance as is *NPR1,* both in the same conserved genomic neighborhood, consistent with the hypothesis that at least two microsyntenous genes may be critically involved in plant adaptation to biotic stress.

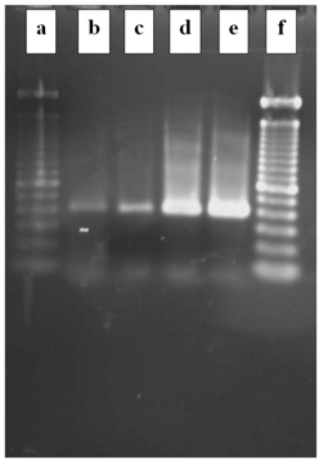

Figure 3. *CaMP* transcript enriched under biotic stress (*E. betavasculorum* challenge). A uv photograph of an ethidium bromide-stained 2% agarose gel showing relative amounts of RT-PCR-amplified DNA product formed from RNA isolated from leaves of *B. vulgaris* plants left uninoculated as controls (b, c) compared with DNA amplified from RNA from plants subjected a week earlier to biotic stress, i.e., experimental *E. betavasculorum* infection, after one week (d, e). Lanes a and f contain "ladder" i.e., 100-bp incremental molecular size, markers at 1 and 1.5X concentrations, respectively.

Oxidative stress is said to stimulate rapid flooding of calcium into the cytosol, which in turn activates enzymes involved in signal transduction cascades and triggers expression of defense-response genes. Infected bean leaves undergoing a hypersensitive response expressed 26 putative calmodulin-binding proteins representing as many as 8 genes up-regulated in an incompatible (i.e., resistant) plant-microbe interaction (Ali *et al.*, 2003). As many as five bean genes encoding calmodulin-binding proteins may be involved in a successful defense response in common bean and these genes produce homologs of previously identified calmodulin-binding proteins in plants. Chloroplast-targeted lipoxygenase genes in common bean are upregulated by wounding (Porta et. al., 2007). A *TomLoxD* gene, in *S. lycopersicum*, a wound-inducible gene, encodes a protein product with a cleavable N-terminal stretch of 60 amino acid residues constituting a chloroplast-targeted transit peptide (Heitz et. *al.*, 2005).

A calmodulin-binding chloroplast-targeted protein is involved in adaptation to biotic stress in *B. vulgaris*. Chloroplast import involves a two multi-subunit protein complex in the outer and inner envelope membranes of the chloroplast, termed Toc and Tic, respectively. A subunit of the Tic translocon, Tic32, is a calmodulin binding protein which, like CaMP in *B. vulgaris*, reacts to calmodulin, a calcium sensor (Chigri, et al., 2006). Most precursor proteins entering the chloroplast are transited by this import pathway and these entering proteins require an N-terminal cleavable transit sequence, or signal peptide, like the precursor molecule for CaMP in *B. vulgaris.*

Despite their endosymbiont origin, chloroplasts have evolved a need to import much of their protein complements post-translationally from the cytosol. Hundreds of nuclearly-encoded proteins are imported into the chloroplast from the cytosol through a mechanism involving two translocons (Soll and Schleiff, 2004). Calmodulin-binding proteins (CaMPs) in plants are very diverse, exhibit various motifs and perform a

correspondingly wide variety of functions (Rhoads et al., 1997; Bahler et al., 2002; Reddy et al., 2002; Charpenteau et al., 2004). For example, an ethylene-upregulated calmodulin-binding protein in *Arabidopsis* triggers senescence and death (Yang *et. al.,* 2000). Another example in *Arabidopsis* is a calmodulin-binding protein with an amino-terminal DNA-binding domain (Reddy et al., 2000). Although the diverse nature of calmodulin-binding proteins has somewhat obscured their vital importance to the cell in responding to environmental challenges, CaMPs are critical for maintaining cell viability. Pivotal roles played by calmodulin-binding proteins include instantaneous molecular switches that activate enzymatic signal transduction cascades and timely transcriptional activators of gene expression. Some calmodulin-binding proteins play a role in the cascade of events triggered in response to plant pathogen attack (Bouché *et. al*., 2005), consistent with upregulated *CaMP* expression in response to *E. betavasculorum* infection discovered in *B. vulgaris*. Conserved microsynteny of the biotic stress-responsive *CaMP* gene with *NPR1*, responsible for induced resistance, is not mere coincidence.

References Cited:

Ali GS, Reddy VS, Lindgren PB, Jakobek JL, Reddy ASN. "Differential expression of genes encoding calmodulin-binding proteins in response to bacterial pathogens and inducers of defense responses," *Plant Molecular Biology*, vol.51, pp.803-815, 2003.

Bahler M., Rhoads A.,"Calmodulin signaling via the IQ motif," *Federation of European Biochemical Societies*, vol. 513, pp. 107-113, 2002.

Bargabus-Larson, RL and BJ Jacobsen. "Biocontrol-elicited systemic resistance in sugarbeet is salicylic acid independent and NPR1 dependent," *Journal of Sugar Beet Research*, vol. 44, pp. 17–33, 2007.

Bouché N, Yellin A, Snedden WA, Fromm H, "Plant-specific calmodulin-binding proteins," *Annu. Rev. Plant Biology*, vol.56, pp. 435-466, 2005.

Charpenteau MK, Jaworski BC, Ramirez AT, Ranjeva R, Ranty B, "A receptor-like kinase from *Arabidopsis thaliana* is a calmodulin-binding protein," *Biochemistry*, vol. 379, pp. 841-848, 2004.

Chigri, F. Hörmann, A. Stamp, D.K. Stammers, B. Bölter, J. Soll, U.C. Vothknecht, "Calcium regulation of chloroplast protein translocation is mediated by calmodulin binding to Tic32," *Plant Biology*, vol.103, no.43, pp.16051-16056, 2006.

Dohm JC, Lange C, Reinhardt R, Himmelbauer H. "Haplotype divergence in *B. vulgaris* and microsynteny with sequenced plant genomics," *The Plant Journal,* vol. 57, pp.14-27, 2009. http://dx.doi.org/10.1111/j.1365-313X.2008.03665.

Heitz T, Bergey DR, Ryan C. "A Gene Encoding a Chloroplast-Targeted Lipoxygenase in *Solancea lycopersicum* Leaves is Transiently Induces by Wounding, Systemin, and Methyl Jasmonate." *Plant Physiology*, vol. 114, pp. 1085-1093, 1997.

Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z and Dunker AK. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acid Research*, vol. 32, No. 3, pp. 1037-1049, 2004.

Kuykendall, LD, J. Shao and T. Murphy. Conserved microsynteny of *NPR1* with genes encoding a signal calmodulin-binding protein and a CK1- class protein kinase in *B. vulgaris* and two other eudicots. *International Journal of Plant Genomics,* vol. 2008, Article ID 391259, 8 pages, 2008. http://dx.doi.org/10.1155/2008/391259.

Kuykendall, LD and WJ Hunter. Carbohydrate release and ethane formation during necrosis induced *Erwinia betavasculorum. Current Microbiology* 56: 128-133, 2008.

Kuykendall, LD. J. Shao and K. Trimmer. *Coe1,* a DNA transposon within a LTR retrotransposon in *B. vulgaris* L. *International Journal of Plant Genomics 2008. 360874.* http://dx.doi.org/10.1155/2008/360874.

Kuykendall, LD, TS Murphy, J Shao and JM McGrath. Nucleotide sequence analyses of a sugar beet genomic *NPR1*- class disease resistance gene. *Journal of Sugar Beet Research* 44: 35-49, 2007.

Kuykendall, LD, J Shao and K Trimmer. A Nest of LTR Retrotransposons Adjacent the Disease Resistance-Priming Gene *NPR1* in *B. vulgaris* L. U.S. Hybrid H2O. *International Journal of Plant Genomics* (in press).

Letunic I, Doerks T, Bork P. SMART 6: recent updates and new developments. *Nucleic Acids Research*, vol. 37, 2009. Database issue D229–D232. http://dx.doi.org/10.1093/nar/gkn808

McGrath, JM, RS Shaw, BG de los Reyes, JJ Weiland. Construction of a sugarbeet BAC library from a hybrid that combines diverse traits," *Plant Molecular Biology Reporter*, vol. 22, pp. 23-28, 2004.

Porta H, Figueroa-Balderas RE, Rocha-Sosa M. Wounding and pathogen infection induce a chloroplast-targeted lipoxygenase in the common bean (*Phaseolus vulgaris* L.). *Planta*, vol. 227, pp. 363-373, 2007.

Reddy VS, Ali GS, Reddy ASN, "Genes encoding calmodulin-binding proteins in the *Arabidopsis* genome," *Journal of Biological Chemistry*, vol. 277, no. 12, pp. 9840-9852, 2002.

Reddy ASN, ReddyVS, Golovkin M. "A calmodulin binding protein from *Arabidopsis* is induced by ethylene and contains a DNA-binding motif," *Biochemical and Biophysical Research Communications*, vol. 279, pp. 762-769, 2000.

Reeves PA, Y He, RJ Schmitz, RM Amasino, LW Panella and CM Richards. Evolutionary conservation of the *FLOWERING LOCUS C*-mediated vernalization response: evidence from the sugarbeet (*B. vulgaris*)," *Genetics*, vol. 176, pp. 295-307, 2007.

Rhoads AR, Friedberg F. "Sequence motifs for calmodulin recognition," *Federation of American Societies for Experimental Biology Journal*, vol. 11, pp. 331-340, 1997.

Soll J, Schleiff E. “Protein import into chloroplasts,” *Nature Review Molecular Cell Biology*, vol.5, pp.198-208, 2004.

Wei H, Persson S, Mehta T, Srinivasasainagendra V, Chen L, Page GP, Somerville C, and Loraine A. Transcriptional Coordination of the Metabolic Network in Arabidopsis. *Plant Physiology*, vol. 142, pp. 762-774, 2006.

Xiang, Y., K. Kakani, R. Reade, E. Hui and D. Roshon. “A 38-amino acid sequence encompassing the arm domain of *Cucumber Necrosis Virus* coat protein functions as a chloroplast transit peptide in infected plants.” *Journal of Virology* vol. 80, no.16, pp.7952-7964, 2006.

Yang T, Poovaiah BW, “An early ethylene up-regulated gene encoding a calmodulin-binding protein involved in plant senescence and death.” *Journal of Biological Chemistry*, vol. 275, no. 49, pp. 38467-38473, 2000.

.