A STUDY OF THE CONSERVATION OF *PUN1* MEDIATED LOSS OF PUNGENCY IN *CAPSICUM* SPECIES AND THE EVOLUTION AND SUBCELLULAR LOCALIZATION OF THE *PUN1* LOCUS AND ITS GENE PRODUCT IN THE SOLANACEAE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Giulia Marina Stellari

August 2009

A STUDY OF THE CONSERVATION OF *PUN1* MEDIATED LOSS OF PUNGENCY IN *CAPSICUM* SPECIES AND THE EVOLUTION AND SUBCELLULAR LOCALIZATION OF THE *PUN1* LOCUS AND ITS GENE PRODUCT IN THE SOLANACEAE

Giulia Marina Stellari, Ph.D.

Cornell University 2009

Capsaicinoids are valuable plant secondary metabolites, arguably among the world's most widely consumed products of a plant secondary metabolic pathway. Their distribution is restricted to the genus *Capsicum* (family Solanaceae). Studies documenting the inheritance of pungency have revealed that mutations at a single locus, *Pun1*, are responsible for loss of pungency in cultivars of the species *C. annuum* and *C. chinense*. This study presents the identification of a previously unreported null allele of *Pun1* from a non-pungent accession of *C. frutescens*, the third species in the *annuum-chinense-frutescens* complex. The loss of pungency phenotype in *C. frutescens* maps to *Pun1* and co-segregates with a molecular marker developed to detect this allele of *Pun1, pun1*^{3}. While this mutation is allelic to *pun1* and *pun1*^{2}, the mutation causing loss of pungency in the undomesticated *C. chacoense*, *pun2*, is not allelic to the *Pun1* locus as demonstrated by mapping and complementation studies. *Pun1* encodes *AT3*, an acyltransferase belonging to the BAHD family of acyltransferases. Sequence from the coding region of *AT3* shows a high level of conservation throughout Solanaceae. A tandem duplication of this gene, designated *AT3-1* and *AT3*-2 respectively, predates the diversification of the Solanaceae. The *AT3-2* locus, now a pseudogene, retains regions of amino acid conservation relative to

AT3-1. Both Bayesian and maximum parsimony methods demonstrate that the paralogous gene lineages *AT3-1* and *AT3-2* form well supported phylogenetic clades. We find support for the species complex comprising *C. annuum*, *C. chinense* and *C. frutescens*, widely reported in the literature and which includes *Lycianthes*. Additionally, in *C. rhomboideum*, a recombination event has occurred between *AT3-1* and *AT3-2*. In this species, a deletion in *AT3-1* eliminates the highly conserved motif DFGWG and is correlated with non-pungency. Finally, this study finds that AT3-1 localizes to the endoplasmic reticulum (ER) suggesting that AT3-1 is not capsaicinoid synthase and that the DFGWG motif may be important for this localization. Treatment with Latrunculin A disrupts AT3-1 localization providing further evidence that it associates with the ER. This analysis sheds light on the putative role of AT3-1 in this intriguing plant biosynthetic pathway.

BIOGRAPHICAL SKETCH

Giulia Marina Stellari was born in Brooklyn, NY. She attended Harvard University where she studied organismal and evolutionary biology. While working in the laboratory of Elena Kramer, she became interested in the study of plant biology and the evolution of flower morphology and development. After graduation she worked briefly at the New York Botanical Garden where she studied the development of the megagametophyte in *Ginkgo biloba*. In 2004, she began her graduate studies at Cornell University in the department of Plant Biology under the supervision of Molly Jahn. Since 2007, she has conducted her research at SUNY Stony Brook in the laboratory of Vitaly Citovsky.

This work is dedicated to Dr. Maurice Pechet. His love of science, scholarship, and the pursuit of academic and personal moral excellence have been my inspiration since we met, almost ten years ago, in Harvard's Lowell House. His belief that one should investigate only those questions that are both personally interesting and that are relevant to the welfare of humankind have guided his years of tireless service and mentorship. His constructive and ardent support of my early attempts at research led me to pursue graduate study, and his continued support throughout my graduate years, to the completion of this work.

ACKNOWLEDGMENTS

I gratefully acknowledge my thesis advisor, Molly Jahn, my committee members, Tadhg Begley, Stephen Kresovich, and my advisor at SUNY Stony Brook, Vitaly Citovsky, whose continuous support of my research made this work possible. I am indebted to Shanna Moore Fellman and Michael Mazourek who guided and supported my studies and aspirations with kind words, encouraging suggestions and terrific insight. I acknowledge the invaluable contributions made to this work by the past and present members of the Jahn Lab, especially Jeffrey Gordon, Inwha Yeam, Kari Perez, and Jason Cavatorta who provided technical assistance with experimental techniques and the interpretation of results. A very special thank you to Mary Kreitinger whose cheerful words of encouragement were always helpful during the disappointing moments. Many thanks are extended to the entire Citovsky Lab, especially Shoko Ueki, Lisa Zalepa King, Adi Zaltsman, Benoit Lacroix, and Alex Krichevsky for generously giving of their time and materials and for their assistance in the execution of the subcellular localization experiments. Thanks to George Moriarty, Matt Falise, Brynda Beeman, Maryanne Fink, Mike Axelrod and John Klum whose assistance in the field and in the greenhouse was critical to the success of these experiments. I am deeply appreciative of the tireless support that I have received from my parents, Giovanna and Umberto Stellari, who contributed to this work with their love and unfailing encouragement of my ambitions and passions. They have given their utmost to make my education possible, and to them I will be forever grateful. Lastly my deepest gratitude goes to Arnout van de Rijt, my dedicated partner in life and in science.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AT3	*Acyltransferase3*
AT3-1	*Acyltransferase3-1*
AT3-2	*Acyltransferase3-2*
BAC	bacterial artificial chromosome
BSA	bulk segregant analysis
cM	centiMorgan
dpa	days post anthesis
ER	endoplasmic reticulum
F81	Felsenstein 1981
GFP	green fluorescent protein
GTR	general time reversible
HKY	Hasegawa Kishino and Yano
I-PCR	inverse PCR
JC	Jukes Cantor
KAS	β-ketoacyl-[acyl carrier protein] synthase
LARD	Likelihood Assisted Recombination Detection
MP	maximum parsimony
ndhF	NADH dehydrogenase subunit 5
PAL	phenylalanine ammonia lyase
ppm	parts per million
QTL	quantitative trait locus
RPD3	recombination detection program 3
TBR	tree bisection reconnection
trnLF	intergenic spacer between tRNA(L) and tRNA(F)

TRPV1	transient receptor potential vanilloid subtype 1

CHAPTER 1: INTRODUCTION

The aim of this introduction is to familiarize the reader with historical and current research on the evolution, ecology, biochemistry and genetics of the genus *Capsicum*. It provides the foundation necessary to place this study on the loss of capsaicinoid biosynthetic ability in domesticated and wild *Capsicum* species, the evolutionary history of the candidate gene for capsaicinoid biosynthesis and its subcellular localization in their correct framework. This introduction highlights the recent taxonomic and archeological discoveries that have refined our understanding of the relationship between humans and this important crop species. Special attention is paid to the ecological context within which capsaicinoid biosynthesis evolved. The capsaicinoid biosynthetic pathway is discussed so as to provide the reader with an understanding of the complexity of the phenotype that is investigated in this study. Finally this introduction will focus on recent research efforts to identify structural genes in the capsaicinoid biosynthetic pathway and on the genetic basis for loss of pungency in the two most widely cultivated species, *Capsicum annuum* and *Capsicum chinense*.

Domestication, evolution, and center of origin of the genus Capsicum

The genus *Capsicum* is one of the 95 to 100 genera recognized within the family Solanaceae and is known for the production of capsaicin and its analogs, collectively known as capsaicinoids, which are responsible for the sensation of burning and pain experienced when consuming hot peppers (Govindarajan, 1985; Nee, 2004; Olmstead *et al*, 2008; Suzuki, 1984). Humans have a long history of pungent pepper consumption and *Capsicum*, which is mostly neo-tropical in its distribution,

with the exception of a single species, *C. anomalum*, is now identified as one of the earliest domesticated plants of the Americas, alongside *Zea maize*, *Phaesolus spp*. and *Cucurbita spp*. (Eshbaugh, 1980; Perry *et al*, 2007; Perry and Flannery, 2007; Pickersgill, 2007; Purugganan and Fuller, 2009). These latter three species are together known as the "milpa" system of agriculture, which is characteristic of the Mesoamerican landscape (Colunga-GarciaMarin and Zizumbo-Villarreal, 2004). Identification of starch granules specific to the domesticated forms of the species *C. annuum* at excavation sites in Ecuador suggests an earliest domestication date at least 6000 yr. BP (Perry *et al*, 2007). As Ecuador is not thought to be one of the centers of *Capsicum* domestication, this places the domestication event for at least *C. annuum* at approximately 2000 years earlier than previous estimates (Pickersgill, 1971). Furthermore, the identification of *Capsicum* starch granules alongside maize starch granules suggests that human culture has held these two culinary staples in close association since ancient times, predating the adoption of pottery at some sites (Perry *et al*, 2007). Pre-Columbian fossils from the Oaxaca valley in Mexico, which is near the putative location for the domestication of *C. annuum*, show that numerous cultivated varieties of *C. annuum* and *C. frutescens*, as determined by evidence of fruit retention on the pedicle and other traits characteristic of cultivated varieties, were already present as early as 500AD (Lippert *et al*, 1966; Loaizafigueroa *et al*, 1989; Perry and Flannery, 2007).

Capsicum is the only genus where the accumulation of capsaicinoids has been documented, and the presence of these molecules in ripe fruit has been proposed as a defining character for the genus (Eshbaugh, 1980). This genus includes between 23 and 30 species, and is traditionally divided between the purple and the white flowered groups, each of which has a subset of species that have either only domesticated forms (*C. pubescens* and possibly *C. chinense*) or both wild and domesticated types (*C.*

annuum, *C. baccatum,* and *C. frutescens*) (Eshbaugh, 1980; Pickersgill, 1971). Much of the debate surrounding the number of species to include in the genus has centered on the relationship between the wild species, the cultivated varieties and the weedy forms of cultivated varieties (Eshbaugh, 1980; Pickersgill, 1971). The aspect of the debate that is relevant to the present study concerns the relationship *C. annuum*, *C. chinense* and *C. frutescens*, which are the three most widely cultivated species by some accounts (Eshbaugh, 1980; Pickersgill, 1971). Whereas morphology shows marked distinctions between these three species and data from interspecific crosses show that they are only partially interfertile, electrophoretic studies of isozyme variation seem to suggest that the three species are indistinguishable (Jensen *et al*, 1979; Lippert *et al*, 1966; Loaizafigueroa *et al*, 1989; Pickersgill, 1971). Genetic studies reveal that there is little genetic variation among the large fruited cultivated varieties of *C. annuum*; a similar result emerged from a study of the semi-cultivated accessions within Mexico, which suggests the presence of a genetic bottleneck (Lefebvre *et al*, 1993; Loaizafigueroa *et al*, 1989). Studies have also shown karyotypic variation within species of *Capsicum* whereas total DNA content is more variable between than within species (Moscone *et al*, 2003; Pickersgill, 1971). Taken together these data seem to suggest that morphological distinctions between cultivated species represent a complex set of differences that have as their basis some genetic variation. This genetic variability is not always easily discernible when sampling random genetic loci (van Tienderen *et al*, 2002). Unlike the scenario in other crop species, where domestication has either favored or disfavored certain secondary metabolic products over others, in the case of *Capsicum*, both the presence and the absence of capsaicinoids has played a critical role in domestication (Lewinsohn and Gijzen, 2009). Given humans' long history of interaction with *Capsicum*, it seems a difficult task to consider the evolution of the genus without taking into account the fact that

intersection with human activities has forever altered the genetics, biochemistry, and distribution of this family of plants.

Within the Solanaceae, *Capsicum* is somewhat more distantly related to the Solanum clade, which includes the widely cultivated vegetable crops tomato, potato and eggplant and even more distantly related to petunia and tobacco, which have become model systems within the Solanaceae (Bohs and Olmstead, 1997; Martins and Barkman, 2005; Mueller *et al*, 2005; Olmstead *et al*, 2008). The basic chromosome number of 12 defines the sub-family Solanaoideae and is shared between *Capsicum* and the rest of the x=12 Solanaceae (Olmstead *et al*, 2008). A recent phylogeny of the Solanaceae based on the chloroplast DNA regions *ndhF* and *trnLF* suggests that *C. rhomboideum* is basal in the genus, although this study did not sample *C. chacoense* (Olmstead *et al*, 2008). Interestingly, *C. rhomboideum* is the only species that is non-pungent (Eshbaugh, 1980). A comprehensive phylogeny of the Solanaceae also found a close association between *Capsicum* and *Lycianthes*, further suggesting that *Lycianthes* may either be paraphyletic or may need to be revised such that some of the species now belonging to this genus are included within *Capsicum* (Olmstead *et al*, 2008). This is somewhat in contrast to previous studies that showed a monophyletic origin for both *Capsicum* and *Lycianthes* (Walsh and Hoot, 2001).

Although most of the species of *Capsicum* are of New World origin, difficulties in drawing taxonomic boundaries within the genus have led to difficulties in identifying its center of origin. While it is thought that the domestication of *Capsicum* has taken place in multiple locations in the Americas, analysis of enzyme variation suggests that the center of origin of the genus is located in the highlands of Bolivia (McLeod *et al*, 1982). *Capsicum chacoense*, is endemic to this region, and patterns of allozyme variation suggest that it may be basal in the genus having given origin to both the white flowered and purple flowered types (Jensen *et al*, 1979).

Studies of the ecology of *Capsicum* and capsaicinoids have thus centered on Bolivia and have focused on its endemic species, *C. chacoense*. In Bolivia, researchers have found natural populations of *C. chacoense* that are polymorphic for pungency and have shown that pungency and the percentage of plants that are pungent increase along a clinal gradient (Tewksbury *et al*, 2006). The presence of non-pungent plants in natural populations, suggests that the ecological role played by capsaicinoids may vary as a result of the different selection pressures present in the environment (Tewksbury *et al*, 2006).

Ecological role of capsaicinoids

A major area of research on capsaicinoids is focused on defining the ecological role played by this family of molecules. The history of the study of plant secondary metabolism, recently reviewed in Hartmann (2007), indicates that the prevailing perspective on plant secondary metabolites has shifted over the years, evolving from the canonical view of plant secondary metabolites as the waste products of primary metabolism to the modern understanding that secondary metabolites have an ecological role with a potential benefit to the plant (Hartmann, 2007). Several hypotheses may be invoked to explain the presence of secondary metabolites in fruit; among them directed deterrence has been tested for capsaicinoids (Tewksbury, 2002; Tewksbury and Nabhan, 2001). Under a hypothesis of directed deterrence, capsaicinoids would promote the dispersal of *Capsicum* seeds by attracting or modifying the behavior of seed dispersal agents, while deterring fruit predation by agents who would decrease seed viability (Tewksbury and Nabhan, 2001). Directed deterrence is an extremely attractive hypothesis to test the ecological role of capsaicinoids, as these molecules cause a burning sensation in mammals, the putative

predators of *Capsicum* seeds, and at the same time are undetectable by birds, the putative dispersal agents of *Capsicum* seeds (Tewksbury and Nabhan, 2001). While it was demonstrated that capsaicinoids exert an effect on the retention time of seeds in the bird gut, their ability to deter seed predation by mammals has been more difficult to demonstrate (Levey *et al*, 2006). Thus evidence for the argument that capsaicinoids function through a mechanism of directed deterrence is lacking. Another study has shown that the effect of capsaicinoids in the bird gut appears to be matched to the primary dispersal agents and other fruit characteristics, such as seed thickness (Tewksbury *et al*, 2008a). Whereas in *C. annuum*, capsaicinoids increase seed retention time in the bird gut and have a positive effect on seed germination, in pungent *C. chacoense* they decrease seed viability after passage through the bird gut; however the primary dispersal agents for pungent *C. chacoense* seeds have shorter gut retention times, suggesting that, overall, capsaicinoids have little effect (Tewksbury *et al*, 2008a). Taken together these results suggest that capsaicinoids play a complex ecological role and one that is not easily explained by one hypothesis or by assigning a single ecological function to this family of molecules. The effect of capsaicinoids is not limited to vertebrates as these molecules are known to have an antiseptic effect, challenging the growth of microorganisms such as *Bacillus subtilis* and *Zygosaccharomyces spp*. (Suzuki, 1984). Capsaicinoid presence in fruit seems to deter hemipteran insect predation and at the same time reduce *Fusarium* infection in seeds pierced by hemipteran insects, thus demonstrating in an ecological setting their antifungal action (Tewksbury *et al*, 2008b). Capsaicinoids may thus represent an interesting case of a secondary metabolite whose ecological role is the same as one of the functional roles it serves in human culture.

The capsaicin receptor

The foundation for the hypothesis of directed deterrence is the preferential perception of capsaicinoids by mammals and the apparent avian insensitivity to this molecule. Cloning of the mammalian receptor for capsaicin, over one decade ago, has provided a molecular basis for such a difference (Caterina *et al*, 1997). The capsaicinoid receptor, transient receptor potential vanilloid subtype 1 (TRPV1), was cloned by screening a cell line derived from embryonic kidney cells, which are normally insensitive to capsaicin, transformed with a neuronal derived cDNA library, for the acquired ability to respond to capsaicin (Caterina *et al*, 1997). TRPV1 is a calcium gated ion channel which responds to capsaicin as well as to heat stimuli by depolarization through the release of internal calcium stores (Caterina *et al*, 1997). It was long known that application of capsaicin as well as other capsaicin receptor agonists such as resiniferatoxin cause hypothermia, and studies in knockout mice confirmed that this mode of action is mediated through TRPV1 (Caterina *et al*, 2000; Szallasi and Blumberg, 1989; Szallasi and Blumberg, 1990). One recent review has even suggested that TRPV1's primary function is in the regulation of body temperature in mammals (Gavva, 2008). In fact, cloning of a number of homologues of TRPV1 has shown that mammals have numerous TRPV1 receptors with different sensitivities to temperature and which respond to different sets of plant derived molecules such as camphor, menthol and allicin (Szallasi *et al*, 2007). Mammals and birds have different sensitivities to capsaicin, as noted in ecological studies, and the molecular basis for this difference was discovered when the chicken ortholog of TRPV1 was cloned and shown to lack eight key residues which confer capsaicin sensitivity to the receptor (Jordt and Julius, 2002). However, the avian ortholog of

TRPV1 is not entirely insensitive to capsaicin and seems to respond to heat stimuli with the same level of sensitivity as the mammalian receptor (Jordt and Julius, 2002). The discovery of TRPV1 provided a channel to connect the relationship between application or ingestion of capsaicin and treatment of pain and various medical disorders, as capsaicin strongly activates a pain nociceptor. Use of capsaicinoids to treat a host of human ailments, such as tonsillitis, diphtheria, hoarseness, flatulence, loss of appetite, cholera, swelling, rheumatism and gout, is ancient and widespread, having been practiced in India, among the ancient Maya and Aztecs, in the West Indies, in Africa, in England and in Spain (Dasgupta and Fowler, 1997). Currently one can find a number of topical creams that use capsaicin as an active ingredient and several clinical trials are underway to test capsaicin's ability to lessen pain in cases of neuropathy and neuralgia (Szallasi *et al*, 2007). Remarkably, capsaicin has been shown to be effective in targeting the delivery of lidocaine to pain sensing neurons, suggesting that the potential uses of this molecule are far from being fully exploited (Binshtok *et al*, 2007).

Structure and biosynthesis of capsaicin

Understanding the molecular structure of the agent causing the pungent sensation in hot peppers has been an active field of research for the past two centuries. Capsaicin was first isolated in crystalline form by Tresh in 1876 (Suzuki, 1984). Elaboration of the chemical formula took over fifty years finally arriving at N-(4-hydroxy-3-methoxybenzyl)-8-methynon-6-*trans*-enamide in 1955 by Crombie who identified the critical *trans* double bond (Crombie *et al*, 1955). The chemical components of the pungent oleoresin extract were further separated and identified as a mixture of two closely related compounds, capsaicin and dihydrocapsaicin, the latter

named for the lack of the *trans* double bond (Kosuge *et al*, 1961a; Kosuge *et al*, 1961b). Biosynthesis of these molecules was explored by two different groups using crude extracts from *C. annuum* and *C. frutescens* respectively (Bennett and Kirby, 1968; Leete and Louden, 1968). Bennett and Kirby showed that the vanillyamine moieties of capsaicin and dihydrocapsaicin were derived directly from phenylalanine and not from tyrosine (Bennett and Kirby, 1968). They determined that capsaicinoid biosynthesis proceeds via the phenylpropanoid pathway by using radiolabeled phenylpropanoid pathway intermediates to show incorporation into capsaicin. Interestingly, incorporation was highest in the presence of radiolabeled phenylalanine and lower in the presence of other radiolabeled pathway intermediates. Leete and Louden confirmed Bennett and Kirby's result that capsaicin was derived via phenylalanine. They showed that the methyl group on the methoxy moiety of vanillylamine is derived from methoinine and that valine is the precursor for the branched chain fatty acid moiety (Leete and Louden, 1968). Furthermore, they were able to recover radioactive capsaicin months after their initial feeding experiment, thus concluding that capsaicin and dihydrocapsaicin were stable end products of the biosynthetic pathway. A major restructuring of the phenylpropanoid biosynthetic pathway occurred with the discovery that many of the pathway intermediates can be found as their shikimate conjugated esters rather than in the free acid form (Humphreys and Chapple, 2002). The influence of this new perspective on the phenylpropanoid pathway has led to a revision of the capsaicinoid biosynthetic pathway to include the incorporation of enzymes that catalyze the conversion of shikimate conjugated pathway intermediates (Stewart *et al*, 2005; Suzuki, 1984). The fatty acid branch of the capsaicinoid biosynthetic pathway has benefited from research showing that premature chain termination by acyl carrier protein (ACP) leads to the biosynthesis of medium chain fatty acids, which are the precursors for the fatty acid

moieties of capsaicinoids (Harwood, 1996). It has also been shown that the fatty acid biosynthetic enzyme complex, rather than α-ketoacid elongation is responsible for the biosynthesis of the medium branched chain fatty acids that are the precursors to capsaicinoids (Kroumova *et al*, 1994; Markai *et al*, 2002).

The genetics of loss of pungency and the identification of a candidate gene for capsaicinoid biosynthesis

Given capsaicinoids' important medicinal uses and the prominent culinary role played by these molecules in certain cultures, a strong emphasis has been placed on pepper breeding and on understanding the genetic basis for capsaicinoid accumulation. Early experiments determined that a single genetic locus is responsible for loss of pungency among cultivated varieties of bell pepper (Deshpande, 1934; Webber, 1911). This locus, known as *C*, was mapped to pepper chromosome 2 in a region shown to be syntenic with tomato chromosome 2 (Blum-Eyal *et al*, 2002; Lefebvre *et al*, 1995; Lippert *et al*, 1966; Livingstone *et al*, 1999; Tanksley *et al*, 1988; Wu, 2009). Concomitant with later experiments which mapped a molecular marker to within 0.4cM of *C*, researchers identified a number of co-regulated putative structural genes in the capsaicinoid biosynthetic pathway, suggesting that coordination at the level of gene regulation was a feature of the capsaicinoid biosynthetic pathway (Blum-Eyal *et al*, 2002; Curry *et al*, 1999). Furthermore, the analysis comparing transcription of genes from pungent and non-pungent varieties of *Capsicum* identified the target tissue, the placental dissepiment, and the developmental stage, prior to fruit ripening, where such a candidate gene might be identified.

Screening of a selective subtractive hybridization library derived from immature fruit of the highly pungent cultivar *C. chinense* 'Habanero' yielded one

differentially expressed candidate cDNA fragment, *SB2-66*, which was later shown to perfectly co-segregate with pungency and map to the *C* locus (Kim *et al*, 2001; Stewart *et al*, 2005). The gene fragment *SB2-66* was cloned and discovered to encode an acyltransferase belonging to the BAHD family of acyltransferases. The locus *C*, was then renamed *Pun1* and its candidate gene, *AT3*, for *Acyltransferase3* (Stewart *et al*, 2005). In non-pungent varieties of *C. annuum*, a 2.5kb deletion comprising the upstream promoter region, transcription start site and the first exon of *Pun1*, entirely eliminates transcription of *AT3* and translation of its gene product and perfectly co-segregates with loss of pungency. AT3 belongs to a class of plant specific acyltransferases, the BAHD acyltransferases, which has only recently identified and was named for the first four characterized members of the family (D'Auria, 2006; St-Pierre and De Luca, 2000). BEAT is involved in floral scent production in *Clarkia breweri* and catalyzes the formation of benzylacetate, a volatile ester (Dudareva, D'Auria et al. 1998). AHCT from *Gentiana triflora* exhibits O-hyrdroxycinnamoyltransferase activity on acylated anthocyanins (Fujiwara *et al*, 1998). HCBT from *Dianthus caryophyllus* catalyzes the formation of dianthramides (Yang *et al*, 1997). DAT from *Catharantus roseus* catalyzes the biosynthesis of vindoline from acetyl co-A deacetylvindoline (St-Pierre *et al*, 1998). Members of this enzyme family identified in the *Arabidopsis thaliana* genome seem to play a role in plant secondary metabolism (D'Auria, 2006; Dudareva *et al*, 1998; St-Pierre and De Luca, 2000). The gene family was defined by the presence of two motifs, HXXXDG and DFGWGRP (St-Pierre and De Luca, 2000). The crystal structure of vinorine synthase showed that HXXXDG, the putative catalytic motif, is in the reaction center while DFGWG (or DFGWGRP) motif, which in some cases is indispensable for function, is not located in the catalytic domain, but may be involved in maintaining structural stability (Ma *et al*, 2005). Based on phylogenetic analysis, AT3 is grouped

with BAHD acyltransferases of unknown function and/or substrate specificity (D'Auria, 2006). In any case, it is a challenging task to predict substrate specificity or mode of action of an enzyme from sequence homology alone, thus assuming substrates and/or biochemical targets for AT3 may be perilous (Sterner *et al*, 2007).

Given *AT3*'s critical role in the capsaicinoid biosynthetic pathway and its sequence similarity to BAHD acyltransferases, it seems parsimonious to assume that AT3 may be one of the enzymes in the capsaicinoid biosynthetic pathway. In fact, the terminal biosynthetic step would require an acyltransferase function (Stewart *et al*, 2005). So far, AT3 has not been expressed in soluble form, thus impairing a traditional assay of enzymatic function and requiring the use of alternate methods to test its function in the capsaicinoid biosynthetic pathway. Virus induced gene silencing (VIGS) of *AT3* was used to show that in plants treated with the silencing construct, capsaicinoid accumulation decreased significantly, confirming its functional role in the capsaicinoid biosynthetic pathway (Stewart *et al*, 2005).

Testing whether *AT3*'s critical role in the capsaicinoid biosynthetic pathway is conserved has relied heavily on using traditional genetic tests to identify its association with capsaicinoid biosynthesis in capsaicinoid biosynthetic mutants. The *pun1* locus, which has been widely used in the breeding of non-pungent peppers, is the only locus known to date to have a qualitative effect on pungency accumulation (Andrews, 1984; Boswell, 1937; Stewart *et al*, 2005). Although it is known that capsaicinoid accumulation is variable at the level of individual fruit and plants and has a strong environmental component, quantitative trait loci (QTL) analysis has shown that four QTL control total levels of capsaicinoids in fruit, accounting together for 37% of the total variance in capsaicinoid accumulation (Ben-Chaim *et al*, 2006; Zewdie-Yayeh and Bosland, 2000). Taken together these results suggest that a single major locus, *Pun1*, with a number of modifying QTL, control capsaicinoid

accumulation in pungent peppers. These studies did not, however, determine whether the *Pun1* locus and the 2.5kb deletion identified in *C. annuum*, is a unique route to loss of pungency or whether selection for the *pun1* locus during variety development and breeding accounts for its widespread presence among non-pungent peppers.

A study investigating loss of pungency in the non-pungent variety *C. chinense* 'NMCA30036' identified a novel mutation in *AT3* that co-segregated with loss of pungency in a cross with *C. chinense* 'Habanero' (Stewart *et al*, 2007). In *C. chinense* 'NMCA30036', a 4-base pair deletion in *AT3* leads to a frameshift mutation in the first exon, a decrease in transcription and absence of translation of the *Pun1* locus. This mutation, named $pun1^2$, is allelic to the *Pun1* mutation (Stewart *et al*, 2007). The identification of two mutations within the closely related species *C. annuum* and *C. chinense* seems to suggest that among domesticated species of pepper, the *Pun1* locus may be the only route to loss of pungency. It remains unclear, however, if these mutations occurred during domestication following selection for an agronomically desirable trait or whether they were already present in the population of wild ancestors that later gave rise to the domesticated species *C. annuum* and *C. chinense*.

Accumulation of capsaicinoids in the fruit and in specific plant organelles

Earlier studies investigating the gross morphology of the placenta of pungent peppers showed that capsaicinoids seem to accumulate in the intracellular space between the epidermis and the cuticle, forming raised protrusions, called "receptacles of capsaicin" or "blisters", which filled with a capsaicin containing oleoresin (Ohta, 1962). Using ultra-thin section microscopy of samples taken from developing pungent and non-pungent pepper placentas it was found that in plants carrying the $pun1^2$ mutation, no capsaicinoids accumulate in the epidermal layer of the placenta (Stewart

et al, 2007). Using immunohistochemistry, absence of capsaicinoids in this tissue layer was also correlated with absence of β-ketoacyl-[acyl carrier protein] synthase (KAS) accumulation, one of the differentially expressed capsaicinoid structural genes identified in developing pepper fruit (Curry *et al*, 1999; Kim *et al*, 2001). This confirmed previously published results, which showed that KAS accumulation in the placenta of developing pepper fruits was positively correlated with pungency (Aluru *et al*, 2003).

Ultrastructural studies of the placenta of pungent peppers have also shown the presence of numerous electron dense granules in the vacuole, on the plasma membrane directly subtending the receptacle and in the cytoplasm of the epidermal cells of the placental dissepiment (Suzuki *et al*, 1980). These electron dense granules were not observed in parenchyma cells directly beneath the epidermis. Furthermore it was shown that dense granular structures were visible when illuminated with light at a wavelength of 280nm, corresponding to the absorption maxima for capsaicinoids. This suggested that capsaicinoids accumulate specifically in the placental dissepiment of pungent peppers and that their accumulation may involve specific organelles. It was soon after determined that the vacuole was the intracellular site of capsaicinoid accumulation (Fujiwake *et al*, 1980). The enzymatic function capsaicinoid synthetase was purified from vacuolar fractions, strongly suggesting that this organelle is involved in capsaicinoid accumulation and biosynthesis (Fujiwake *et al*, 1982). Another study that used both scanning electron microscopy (SEM) and TEM to investigate differences in the development of pungent and non-pungent placentas did not find capsaicinoids in the vacuole (Zamski *et al*, 1987). The study instead showed that capsaicinoids, identified as osmiophilic granules in TEM sections, accumulated in vesicles and in the endoplasmic reticulum (ER) and that the plasma membrane contained several protrusions, plasmalemmosomes, which were associated with

capsaicinoid containing granules. These studies have greatly clarified the structures within the cell that accumulate capsaicinoids, however such an analysis has not been extended to studying the subcellular localization of the candidate gene for capsaicinoid biosynthesis, *AT3*.

Recent advances in imaging with confocal laser microscopy as well as the ubiquitous use of green fluorescent protein (GFP) labeling have allowed for an unprecedented explosion in the field of subcellular localization of enzymatic activities, identification of interactions among pathway proteins, and organelle dynamics and functions (Brandizzi *et al*, 2004; Citovsky *et al*, 2008; Di Sansebastiano *et al*, 2007; Runions *et al*, 2006). Interactions among members of the phenylpropanoid pathway, especially on the flavonoid branch of the pathway are well characterized. Interactions between chalcone synthase (CHS) and chalcone isomerase (CHI) and the co-localization of these two proteins to the ER has been demonstrated in Arabidopsis roots (Saslowsky and Winkel-Shirley, 2001). The enzymes phenylalanine ammonia lyase (PAL) and cinnamate-4-hydroxylase (C4H), both of which are components of the phenylpropanoid pathway, have been shown to associate on the ER membrane (Achnine *et al*, 2004). ER localization of C4H from poplar and of PAL from *Primula kewensis* has also been demonstrated suggesting that the subcellular localization for several enzymes in the phenylpropanoid biosynthetic pathway may be conserved among plants (Ro *et al*, 2001; Schopker *et al*, 1995). The subcellular localization of only a few BAHD acyltransferases has been studied in detail and an initial analysis of localization sequences present in the family suggests that the members of this family of proteins would localize to the cytosol (D'Auria, 2006). Where investigated, BAHD acyltransferases have shown a different pattern of subcellular localization. CER2 is a BAHD acyltransferase involved in epicuticular wax elongation in maize that has been identified as a nuclear protein (Xia *et al*, 1997). A malonyltransferase from soybean

shows cytosolic and nuclear localization, despite having no obvious nuclear localization signal, as does another malonyltransferase from *Medicago truncatula* (Dhaubhadel *et al*, 2008; Yu *et al*, 2008). In the latter case, it was also demonstrated that the DFGWG motif, although not conclusively a targeting signal, is important for subcellular localization of the protein (Yu *et al*, 2008). Since subcellular targeting can help predict protein function and/or interacting partners, an investigation of the subcellular localization of AT3 is crucial in overcoming the challenges that poor recombinant protein expression has posed to the study of its role in the capsaicinoid biosynthetic pathway (Brandizzi *et al*, 2004).

The present study

In the first section of this study I have investigated the genetic basis for loss of pungency in *C. frutescens*, the third species in the *annuum-chinense-frutescens* complex. I show that loss of pungency in *C. frutescens* is allelic to the known loss of pungency alleles in *C. annuum* and *C. chinense*, suggesting a canalization of the capsaicinoid biosynthetic pathway in the domesticated species. This differs from the genetic basis for loss of pungency in a wild variety of *Capsicum*, *C. chacoense*. The relationship between the two loci controlling loss of pungency is explored and conclusions are drawn about the relationship of the known *Pun1* alleles to each other. The second section of this study details the evolution of our candidate gene for capsaicinoid biosynthesis, *AT3*, and uncovers an ancient tandem gene duplication event that predates the diversification of the Solanaceae. An in-frame recombination event is identified between *AT3-1* and *AT3-2* in the basal species in the genus, *C. rhomboideum*. A frameshift mutation in the second exon of *AT3-1* in *C. rhomboideum* is associated with non-pungency in this accession. The third section of this study

attempts to shed light on the function of AT3-1 using confocal microscopy with fluorescent protein tags to examine AT3-1's subcellular localization in a heterologous expression system. I show that AT3-1 localizes to the ER suggesting that AT3-1 may be involved in the phenylpropanoid branch of the capsaicinoid biosynthetic pathway. Furthermore, the DFGWG motif, which seems to be important for the enzymatic function and correct subcellular localization of BAHD family acyltransferases, is shown to be essential for AT3-1 localization to the ER.

WORKS CITED

Achnine L, Blancaflor EB, Rasmussen S, Dixon RA (2004). Colocalization of L-phenylalanine ammonia-lyase and cinnarnate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell* **16**(11)**:** 3098-3109.

Aluru MR, Mazourek M, Landry LG, Curry J, Jahn M, O CMA (2003). Differential expression of fatty acid synthase genes, *Acl*, *Fat* and *Kas*, in *Capsicum* fruit. *Journal of Experimental Botany* **54**(388)**:** 1655-1664.

Andrews J (1984). *Peppers: The Domesticated Capsicums*. University of Texas Press.

Ben-Chaim A, Borovsky Y, Falise M, Mazourek M, Kang BC, Paran I *et al* (2006). QTL analysis for capsaicinoid content in *Capsicum*. *Theoretical and Applied Genetics* **113**(8)**:** 1481-1490.

Bennett DJ, Kirby GW (1968). Constitution and Biosynthesis of Capsaicin. *Journal of the Chemical Society C-Organic* **4:** 442-446.

Binshtok AM, Bean BP, Woolf CJ (2007). Inhibition of nociceptors by TRPV1-mediated entry of impermeant sodium channel blockers. *Nature* **449**(7162)**:** 607-U613.

Blum-Eyal, Liu K, Mazourek M, Yoo Eun Y, Jahn M, Paran I (2002). Molecular mapping of the *C* locus for presence of pungency in Capsicum. *Genome* **45**(4)**:** 702-705.

Bohs L, Olmstead RG (1997). Phylogenetic relationships in *Solanum* (Solanaceae) based on *ndhF* sequences. *Systematic Botany* **22**(1)**:** 5-17.

Boswell V (1937). Impovement and Genetics of tomatoes, peppers and eggplant. *Yearbook of Agriculture*.

Brandizzi F, Irons SL, Johansen J, Kotzer A, Neumann U (2004). GFP is the way to glow: bioimaging of the plant endomembrane system. *Journal of Microscopy* **214:** 138-158.

Caterina MJ, Leffler A, Malmberg AB, Martin WJ, Trafton J, Petersen ZKR *et al* (2000). Impaired nociception and pain sensation in mice lacking the capsaicin receptor. *Science* **288**(5464)**:** 306-313.

Caterina MJ, Schumacher MA, Tominaga M, Rosen TA, Levine JD, Julius D (1997). The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**(6653)**:** 816-824.

Citovsky V, Gafni Y, Tzfira T (2008). Localizing protein-protein interactions by bimolecular fluorescence complementation in planta. *Methods* **45**(3)**:** 196-206.

Colunga-GarciaMarin P, Zizumbo-Villarreal D (2004). Domestication of plants in Maya Lowlands. *Economic Botany* **58:** S101-S110.

Crombie L, Dandegaonker SH, Simpson KB. (1955). *Journal of the Chemical Society*, pp. 1025-1027.

Curry J, Aluru M, Mendoza M, Nevarez J, Melendrez M, O CMA (1999). Transcripts for possible capsaicinoid biosynthetic genes are differentially accumulated in pungent and non-pungent *Capsicum* spp. *Plant Science* **148**(1)**:** 47-57.

D'Auria JC (2006). Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology* **9**(3)**:** 331-340.

Dasgupta P, Fowler CJ (1997). Chillies: from antiquity to urology. *British Journal of Urology* **80**(6)**:** 845-852.

Deshpande RB (1934). Studies on Indian Chillies (4) Inheritance of Pungency in *Capsicum annuum* L. *Indian Journal of Agricultural Science* **5**(IV)**:** 513-516.

Dhaubhadel S, Farhangkhoee M, Chapman R (2008). Identification and characterization of isoflavonoid specific glycosyltransferase and malonyltransferase from soybean seeds. *Journal of Experimental Botany* **59**(4)**:** 981-994.

Di Sansebastiano GP, Renna L, Gigante M, De Caroli M, Piro G, Dalessandro G (2007). Green fluorescent protein reveals variability in vacuoles of three plant species. *Biologia Plantarium* **51**(1)**:** 49-55.

Dudareva N, D'Auria JC, Nam KH, Raguso RA, Pichersky E (1998). Acetyl-CoA:benzylalcohol acetyltransferase - an enzyme involved in floral scent production in *Clarkia breweri*. *Plant Journal* **14**(3)**:** 297-304.

Eshbaugh WH (1980). The taxonomy of the genus *Capsicum* (Solanaceae). *New Phytologist* **47**(2)**:** 153-166.

Fujiwake H, Suzuki T, Iwai K (1980). Intracellular localization of capsaicin and its analogs in *Capsicum* fruit II. The vacuole as the intracellular accumulation site of capsaicinoid in the protoplast of *Capsicum* fruit. *Plant and Cell Physiology* **21**(6)**:** 1023-1030.

Fujiwake H, Suzuki T, Iwai K (1982). Intracellular distributions of enzymes and intermediates involved in biosynthesis of capsaicin and its analogs in *Capsicum* fruits. *Agricultural and Biological Chemistry* **46**(11)**:** 2685-2690.

Fujiwara H, Tanaka Y, Yonekura-Sakakibara K, Fukuchi-Mizutani M, Nakao M, Fukui Y *et al* (1998). cDNA cloning, gene expression and subcellular localization of anthocyanin 5-aromatic acyltransferase from *Gentiana triflora*. *Plant Journal* **16**(4)**:** 421-431.

Gavva NR (2008). Body-temperature maintenance as the predominant function of the vanilloid receptor TRPV1. *Trends in Pharmacological Sciences* **29**(11)**:** 550-557.

Govindarajan VS (1985). *Capsicum*: Production, Technology, Chemistry and Quality. 1. History, Botany, Cultivation and Primary Processing. *Crc Critical Reviews in Food Science and Nutrition* **22**(2)**:** 109-176.

Hartmann T (2007). From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry* **68**(22-24)**:** 2831-2846.

Harwood JL (1996). Recent advances in the biosynthesis of plant fatty acids. *Biochimica and Biophysica Acta-Lipids Lipid Metab* **1301**(1-2)**:** 7-56.

Humphreys JM, Chapple C (2002). Rewriting the lignin roadmap. *Current Opinion in Plant Biology* **5**(3)**:** 224-229.

Jensen RJ, McLeod MJ, Eshbaugh WH, Guttman SI (1979). Numerical taxonomic analyses of allozymic variation in *Capsicum* (Solanaceae) *Taxon* **28**(4)**:** 315-327.

Jordt SE, Julius D (2002). Molecular basis for species-specific sensitivity to "hot" chili peppers. *Cell* **108**(3)**:** 421-430.

Kim M, Kim S, Kim S, Kim B-D (2001). Isolation of cDNA clones differentially accumulated in the placenta of pungent pepper by suppression subtractive hybridization. *Molecules and Cells* **11**(2)**:** 213-219.

Kosuge S, Inagaki Y, Okumura H (1961a). Studies on pungent principles of red pepper. 8. On chemical constitutions of pungent principles (5) on chemical constitution of pungent principle 2. *Nippon Nogei Kagaku Kaishi* **35**(10)**:** 923-&.

Kosuge S, Kimura K, Inagaki Y (1961b). Studies on pungent principles of red pepper. 7. On chemical constitutions of pungent principles. 4. On chemical constituion of pungent principle 1. *Nippon Nogei Kagaku Kaishi* **35**(6)**:** 596-&.

Kroumova AB, Xie ZY, Wagner GJ (1994). A pathway for the biosynthesis of straight and branched, odd- and even length, medium-chain fatty acids in plants. *Proceedings of the National Academy of Sciences of the United States of America* **91**(24)**:** 11437-11441.

Leete E, Louden MCL (1968). Biosynthesis of capsaicin and dihydrocapsaicin in *Capsicum frutescens*. *Journal of the American Chemical Society* **90**(24)**:** 6837-6841.

Lefebvre V, Palloix A, Caranta C, Pochard E (1995). Construction of an intraspecific integrated linkage map of pepper using molecular markers and doubled haploid progenies. *Genome* **38**(1)**:** 112-121.

Lefebvre V, Palloix A, Rives M (1993). Nuclear RFLP between pepper cultivars (*Capsicum annuum* L.). *Euphytica* **71**(3)**:** 189-199.

Levey DJ, Tewksbury JJ, Cipollini ML, Carlo TA (2006). A field test of the directed deterrence hypothesis in two species of wild chili. *Oecologia* **150**(1)**:** 61-68.

Lewinsohn E, Gijzen M (2009). Phytochemical diversity: The sounds of silent metabolism. *Plant Science* **176**(2)**:** 161-169.

Lippert LF, Smith PG, Bergh BO (1966). Cytogenetics of vegetable crops. Garden pepper (*Capsicum sp*.). *Botanical Review* **32**(1)**:** 24-55.

Livingstone KD, Lackney VK, Blauth JR, Van Wijk R, Jahn MK (1999). Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. *Genetics* **152**(3)**:** 1183-1202.

Loaizafigueroa F, Ritland K, Cancino JAL, Tanksley SD (1989). Patterns of genetic variation of the genus *Capsicum* (Solanaceae) in Mexico. *Plant Systematics and Evolution* **165**(3-4)**:** 159-188.

Ma XY, Koepke J, Panjikar S, Fritzsch G, Stockigt J (2005). Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *Journal of Biological Chemistry* **280**(14)**:** 13576-13583.

Markai S, Marchand Patrice A, Mabon F, Baguet E, Billault I, Robins Richard J. (2002). *Chembiochem, Vol. 3*: Germany: Germany, Federal Republic of, pp 212-218.

Martins TR, Barkman TJ (2005). Reconstruction of Solanaceae phylogeny using the nuclear gene *SAMT*. *Systematic Botany* **30**(2)**:** 435-447.

McLeod MJ, Guttman SI, Eshbaugh WH (1982). Early evolution of chili pepper (*Capsicum*) *Economic Botany* **36**(4)**:** 361-368.

Moscone E-A, Baranyi M, Ebert I, Greilhuber J, Ehrendorfer F, Hunziker A-T (2003). Analysis of nuclear DNA content in *Capsicum* (Solanaceae) by flow cytometry and Feulgen densitometry. *Annals of Botany* **92**(1)**:** 21-29.

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J *et al* (2005). The SOL Genomics Network. A comparative resource for Solanaceae biology and beyond. *Plant Physiology* **138**(3)**:** 1310-1317.

Nee M (2004). Solanaceae. In: Smith N, Mori, S. A., Henderson, A., Steveson, D.W., and Heald, S. V. (ed) *Flowering Plants of the Neotropics*. Princeton University Press: Princeton and Oxford, pp 355-357.

Ohta Y (1962). Physiological and genetical studies on the pungency of *Capsicum*. 3. Development of capsaicin in *Capsicum* species. *Japanese Journal of Genetics* **37**(1)**:** 86-90.

Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia VF, Collier SM (2008). A molecular phylogeny of the Solanaceae. *Taxon* **57**(4)**:** 1159-1181.

Perry L, Dickau R, Zarrillo S, Holst I, Pearsall DM, Piperno DR *et al* (2007). Starch fossils and the domestication and dispersal of chili peppers (*Capsicum spp*. L.) in the Americas. *Science* **315**(5814)**:** 986-988.

Perry L, Flannery KV (2007). Precolumbian use of chili peppers in the Valley of Oaxaca, Mexico. *Proceedings of the National Academy of Sciences of the United States of America* **104**(29)**:** 11905-11909.

Pickersgill B (1971). Relationships between Weedy and Cultivated Forms in Some Species of Chili Peppers (Genus *Capsicum*). *Evolution* **25**(4)**:** 683-691.

Pickersgill B (2007). Domestication of plants in the americas: Insights from mendelian and molecular genetics. *Annals of Botany* **100**(5)**:** 925-940.

Purugganan MD, Fuller DQ (2009). The nature of selection during plant domestication. *Nature* **457**(7231)**:** 843-848.

Ro DK, Mah N, Ellis BE, Douglas CJ (2001). Functional characterization and subcellular localization of poplar (*Populus trichocarpa* x *Populus deltoides*) cinnamate 4-hydroxylase. *Plant Physiology* **126**(1)**:** 317-329.

Runions J, Brach T, Kuhner S, Hawes C (2006). Photoactivation of GFP reveals protein dynamics within the endoplasmic reticulum membrane. *Journal of Experimental Botany* **57**(1)**:** 43-50.

Saslowsky D, Winkel-Shirley B (2001). Localization of flavonoid enzymes in Arabidopsis roots. *Plant Journal* **27**(1)**:** 37-48.

Schopker H, Kneisel M, Beerhues L, Robenek H, Wiermann R (1995). Phenylalanine ammonia-lyase and chalcone synthase in glands of *Primula kewensis* (W. Wats): immunofluorescence and immunogold localization. *Planta* **196**(4)**:** 712-719.

St-Pierre B, De Luca V (2000). Evolution of acyltransferase genes: origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. *Recent Advances in Phytochemistry* (Evolution of Metabolic Pathways) **34:** 285-315.

St-Pierre B, Laflamme P, Alarco AM, De Luca V (1998). The terminal O-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A-dependent acyl transfer. *Plant Journal* **14**(6)**:** 703-713.

Sterner B, Singh R, Berger B (2007). Predicting and annotating catalytic residues: An information theoretic approach. *Journal of Computational Biology* **14**(8)**:** 1058-1073.

Stewart C, Kang BC, Liu K, Mazourek M, Moore SL, Yoo EY *et al* (2005). The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. *Plant Journal* **42**(5)**:** 675-688.

Stewart C, Mazourek M, Stellari GM, O'Connell M, Jahn M (2007). Genetic control of pungency in *C. chinense* via the *Pun1* locus. *Journal of Experimental Botany* **58**(5)**:** 979-991.

Suzuki T, Fujiwake H, Iwai K (1980). Intracellular localization of capsaicin and its analogs capsaicinoid in *Capsicum* fruit 1. microscopic investigation of the structure of the placenta of *Capsicum annuum* var. *annuum* cultivar Karayatsubusa. *Plant and Cell Physiology* **21**(5)**:** 839-853.

Suzuki T, Iwai, K. (1984). Constituents of Red Pepper Species: Chemistry, Biochemistry, Pharmacology, and Food Science of the Pungent Principle of *Capsicum* species *The Alkaloids*. The Academic Press. Vol. XXIII.

Szallasi A, Blumberg PM (1989). Resiniferatoxin, a phorbol related diterpene acts as an ultrapotent analog of capsaicin, the irritant constituent in red pepper. *Neuroscience* **30**(2)**:** 515-520.

Szallasi A, Blumberg PM (1990). Resiniferatoxin and its analogs provide novel insights into the pharmacology of the vanilloid (capsaicin) receptor. *Life Sciences* **47**(16)**:** 1399-1408.

Szallasi A, Cortright DN, Blum CA, Eid SR (2007). The vanilloid receptor TRPV1: 10 years from channel cloning to antagonist proof-of-concept. *Nature Reviews Drug Discovery* **6**(5)**:** 357-372.

Tanksley SD, Bernatzky R, Lapitan NL, Prince JP (1988). Conservation of gene repertoire but not gene order in pepper and tomato. *Proceedings of the National Academy of Sciences of the United States of America* **85**(17)**:** 6419-6423.

Tewksbury JJ (2002). Fruits, frugivores and the evolutionary arms race. *New Phytologist* **156**(2)**:** 137-139.

Tewksbury JJ, Levey DJ, Huizinga M, Haak DC, Traveset A (2008a). Costs and benefits of capsaicin-mediated control of gut retention in dispersers of wild chilies. *Ecology* **89**(1)**:** 107-117.

Tewksbury JJ, Manchego C, Haak DC, Levey DJ (2006). Where did the chili get its spice? Biogeography of capsaicinoid production in ancestral wild chili species. *Journal of Chemical Ecology* **32**(3)**:** 547-564.

Tewksbury JJ, Nabhan GP (2001). Seed dispersal. Directed deterrence by capsaicin in chilies. *Nature* **412**(6845)**:** 403-404.

Tewksbury JJ, Reagan KM, Machnicki NJ, Carlo TA, Haak DC, Penaloza ALC *et al* (2008b). Evolutionary ecology of pungency in wild chilies. *Proceedings of the National Academy of Sciences of the United States of America* **105**(33)**:** 11808-11811.

van Tienderen PH, de Haan AA, van der Linden CG, Vosman B (2002). Biodiversity assessment using markers for ecologically important traits. *Trends in Ecology and Evolultion* **17**(12)**:** 577-582.

Walsh BM, Hoot SB (2001). Phylogenetic relationships of *Capsicum* (Solanaceae) using DNA sequences from two noncoding regions: the chloroplast *atpB-rbcL* spacer region and nuclear *waxy* introns. *International Journal of Plant Sciences* **162**(6)**:** 1409-1418.

Webber H (1911). *American Breeders Association Annual* **V7:** 188.

Wu F, Eannetta, N., Xu, Y., Durrett, R., Mazourek, M., Jahn, M., Tanksley, S. (2009). A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. *Theoretical and Applied Genetics* **In press.**

Xia YJ, Nikolau BJ, Schnable PS (1997). Developmental and hormonal regulation of the Arabidopsis *CER2* gene that codes for a nuclear-localized protein required for the normal accumulation of cuticular waxes. *Plant Physiology* **115**(3)**:** 925-937.

Yang Q, Reinhard K, Schiltz E, Matern U (1997). Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, *Dianthus caryophyllus* L. *Plant Molecular Biology* **35**(6)**:** 777-789.

Yu XH, Chen MH, Liu CJ (2008). Nucleocytoplasmic-localized acyltransferases catalyze the malonylation of 7-O-glycosidic (iso)flavones in *Medicago truncatula*. *Plant Journal* **55**(3)**:** 382-396.

Zamski E, Shoham O, Palevitch D, Levy A (1987). Ultrastructure of capsaicinoid-secreting cells in pungent and nonpungent red pepper *Capsicum annuum* L. cultivars. *Botanical Gazette* **148**(1)**:** 1-6.

Zewdie-Yayeh, Bosland Paul W (2000). Pungency of Chile (*Capsicum annuum* L.) fruit is affected by node position. *Hortscience* **35** (6): 1174.

CHAPTER 2: CONTRASTING MODES FOR LOSS OF PUNGENCY BETWEEN CULTIVATED AND WILD SPECIES OF *CAPSICUM*

(Giulia Marina Stellari[1], Michael Mazourek[2], and Molly Jahn[3†‡])

ABSTRACT

Studies documenting the inheritance of pungency or "heat" in pepper (*Capsicum spp.*) have revealed that mutations at a single locus, *Pun1*, are responsible for loss of pungency in cultivars of the two closely related species *C. annuum* and *C. chinense*. Here we present the identification of a previously unreported null allele of *Pun1* from a non-pungent accession of *C. frutescens*, the third species in the *annuum-chinense-frutescens* complex of domesticated *Capsicums*. The loss of pungency phenotype in *C. frutescens* maps to *Pun1* and co-segregates with a molecular marker developed to detect this allele of *Pun1, pun1*^{3}. Loss of transcription of *pun1*^{3} is correlated with loss of pungency. While this mutation, is allelic to *pun1* and *pun1*^{2}, the mutation causing loss of pungency in the undomesticated *C. chacoense*, *pun2*, is not allelic to the *Pun1* locus as demonstrated by mapping and complementation studies. The different origins of non-pungency in pepper are discussed in the context of the phylogenetic relationship of the known loss of pungency alleles.

---

[1] Present address: Biochemistry Department, 424 Life Sciences Building, SUNY Stony Brook, Stony Brook NY, 11794.
[2] Present address: Plant Breeding Department, Bradfield Hall Room 316, Cornell University, Ithaca NY, 14853.
[3] Present address: Department of Agronomy and Genetics, 140 Agriculture Hall, 145 Linden Drive, Madison, WI 53706.
[†] Author contributions: MM identified germplasm, designed and performed crosses, GMS performed mapping, sequencing, RT-PCR, and phylogenetic analysis; GMS and MM wrote manuscript, all authors read and reviewed the manuscript before submission.
[‡] Submitted for publication: Stellari GM, Mazourek MM, Jahn M., *Heredity*.

INTRODUCTION

Capsaicin, an alkaloid derived from the phenylpropanoid and fatty acid biosynthetic pathways, is responsible for the sensation of burning and pain experienced when consuming pungent ("hot") peppers (Bennett and Kirby, 1968; Leete and Louden, 1968; Nelson, 1919a; Nelson, 1919b). Peppers are a widely traded spice, perhaps the most commonly consumed spice in the world, making capsaicin one of the most commonly ingested plant secondary metabolites (Govindarajan, 1985; Hadacek, 2002). At the same time, non-pungent peppers are a valuable vegetable crop. This makes understanding the capsaicinoid biosynthetic pathway an extremely important task as both pungent and non-pungent peppers are valuable targets for vegetable crop improvement through plant breeding. Beyond value for applied purposes, the evolution of capsaicinoid biosynthesis is an intriguing example of the origin of a novel metabolism within the well developed comparative system of the *Solanaceae*.

Biosynthesis of capsaicin is taxonomically restricted to the genus *Capsicum* and furthermore is confined to the developing pepper fruit, specifically the placental dissepiment (Eshbaugh, 1980; Stewart *et al*, 2007; Suzuki *et al*, 1980). This confines the study of the genetic underpinnings of this important biosynthetic pathway to either naturally occurring genetic variation in capsaicinoid biosynthesis present in the genus *Capsicum* or to variation, which has arisen in the course of artificial selection on this plant trait. As there are only rare examples of the successful generation of transgenic pepper, current study of capsaicinoid biosynthesis rests upon the exploitation of breeding lines, cultivars and wild accessions. The utilization of plant collections to further our understanding of the genetic control of agronomically important traits is a critical component of a forward genetics approach. The *Capsicum* collection in the

Plant Genetic Resources Conservation Unit at Griffin, Georgia is especially rich, housing over 4,000 accessions and is the source of the mutants used in this study (USDA-ARS database).

It has been known for almost a century that a mutation at a single genetic locus, *Pun1* formerly *C*, underlies loss of pungency in the widely cultivated bell pepper, the fruit of the plant *Capsicum annuum*, and that this mutation has for an even longer period of time been exploited in the breeding of non-pungent peppers (Deshpande, 1934; Webber, 1911). Through candidate gene analysis, we have identified the mutation that results in this loss of pungency as a deletion in the gene *AT3*. *AT3* encodes an acyltransferase protein belonging to the BAHD family of acyltransferases (Stewart *et al*, 2005). *Pun1* is the only locus known to date to have a qualitative effect on pungency accumulation (Blum *et al*, 2003). Capsaicinoid biosynthesis is environmentally variable, modulated by differences in growing environment, as well as fruit position on the plant (Zewdie and Bosland, 2000). Nonetheless, studies of the quantitative inheritance of capsaicinoid biosynthesis have shown that two major quantitative trait loci (QTL), *caps7.1* and *caps7.2* modulate quantities of capsaicinoid accumulation in red ripe pepper fruit (Ben-Chaim *et al*, 2006).

An integration of molecular genetic studies on plant secondary metabolites and their ecological role is somewhat rare. Seminal work has been performed in *Mimulus* where it has been shown that a QTL controlling plant color is responsible for differences in the pollination syndrome seen between species (Bradshaw *et al*, 1995; Bradshaw and Schemske, 2003). In *Capsicum*, recent studies have furthered our understanding of the ecological role of capsaicin in natural populations. In Bolivia, researchers have identified populations of *C. chacoense* and *C. eximium*, two species thought to be basal in the genus, which are polymorphic for levels of capsaicinoid

accumulation (Tewksbury *et al*, 2006). In these populations, it has been shown that the differences in capsaicinoid accumulation are stable and heritable and that both levels of capsaicinoid accumulation and the percentage of plants that are pungent increases along a clinal gradient (Tewksbury et al, 2006).

The goal of our study is twofold. First we aim to determine the allelic relationship between a non-pungent accession of *C. frutescens*, PI594141, identified in the ARS-GRIN repository, and the other previously characterized null alleles of *Pun1* identified in the cultivated species *C. annuum* (*pun1*) and *C. chinense* (*pun1*^{2}). Second we aim to understand whether a different genetic locus may underlie loss of capsaicinoid biosynthesis in a non-cultivated species of *Capsicum*, *C. chacoense*, PI260433. This work builds upon the previously developed genomic and genetic resources available in *Capsicum* and furthers our understanding of capsaicinoid biosynthesis by identifying novel alleles and loci with a qualitative effect on pungency and by describing their relationship to knowledge about the ecology and evolution of the genus.

MATERIALS AND METHODS

Plant Growth and Tissue collection

Capsicum frutescens PI594141 pungent, herein referred to as *C. frutescens* PI594141-P, *Capsicum frutescens* PI594141 non-pungent, herein referred to as *C. frutescens* PI594141-np, *Capsicum chacoense* PI260433 pungent, herein referred to as *C. chacoense* PI260433-P, *Capsicum chacoense* PI260433 non-pungent, herein referred to as *C. chacoense* PI260433-np, *Capsicum chinense* 'Habanero', *Capsicum chinense* 'NMCA30036', *Capsicum frutescens* BG2814-6, F_1 hybrids from all crosses

performed (see Tables 2.1 and 2.4) and select F_2 individuals (see Table 2.2) were grown from seed in the greenhouse in the Guterman facility at Cornell University, Ithaca, NY and SUNY Stony Brook, Stony Brook, NY. Growing conditions were approximately 27°/18°C (day/night) with supplemental lighting. Plants were fertilized weekly with Blossom Booster (J.R. Peters Inc., Allentown, PA) at Stony Brook and continuously at Ithaca with Excel (200ppm; The Scotts Company, Marysville, Ohio). For genetic studies in the field, plants were sown in the Ithaca greenhouse and then transplanted to fields located in Varna, NY during the summers of 2007 and 2008.

Inverse PCR

Inverse PCR (I-PCR) was performed to identify the novel mutation at *Pun1* in *C. frutescens* PI594141-np based on the protocol by Pham *et. al.* with some modifications (Pham *et al*, 1999). Genomic DNA from *C. frutescens* PI594141 accessions was extracted using a QIAGEN Plant DNA extraction kit (QIAGEN, Valencia, CA) as per manufacturer's instructions. Approximately 250ng of genomic DNA were digested with *Xho*1 (NEB, Ipswich, MA). Following heat inactivation of the enzyme, DNA was then purified using a GFX PCR DNA and Gel Band Purification Kit (GE, Healthcare Biosciences Inc., Piscataway, NJ). The purified DNA was self-ligated with T4 ligase incubated at 16°C overnight and ethanol precipitated. Nested I-PCR amplification primers, anchored in the known region of *pun1*[3] in *C. frutescens* PI594141-np were designed using the online module from Primo Inverse 3.4 available through Chang BioScience (http://www.changbioscience.com/primo/primoinv.html). A first round of PCR was performed as per the following protocol: 5µl 10X PCR Buffer, 4µl 2.5mM dNTPs, 2µl 10µM I-PCR 1F (TACCCAACCCCCAAACTATAGG), 2µl 10µM I-PCR 1R

(ACTTGTAGTTTTTCGGAAATGAAAAG) and in a separate reaction, 2µl 10µM I-PCR 1F with 2µl 10µM I-PCR 2R (GTAGTTTTTCGGAAATGAAAAGTACTG), 0.25µl ExTaq (Takara Bio Inc. Otsu, Japan), 2µl ligated DNA and H_2O to a final volume of 50µl. A touch-down PCR protocol was used for amplification with cycles as follows: 3min at 94°C, 94°C for 30 seconds, starting at 60°C for 1 minute, and reduced 1°C for 9 cycles to reach an annealing temperature of 51°C, 72°C for 4 minutes 30 seconds, followed by 27 cycles of 94°C for 30 seconds, 52°C for 1 minute, 72°C for 4 minutes 30 seconds, with a final extension at 72°C for 15 minutes. 2µl of the primary PCR was amplified with primers I-PCR 2F (TAGTATCAACATCACACCTAGAAGATG), I-PCR 2R (GTAGTTTTTCGGAAATGAAAAGTACTG) using the same PCR conditions as stated above. The two reverse primers produced PCR products of roughly the same size. The approximately 8kb PCR products were sent for direct sequencing at the SUNY Stony Brook DNA Sequencing Facility (http://www.osa.sunysb.edu/dna/). A series of primers designed to sequence overlapping fragments were designed sequentially in order to complete the primer walk of the PCR product.

RNA extraction

Tissue was collected and frozen immediately in liquid nitrogen, ground to a powder, and stored at -80°C. RNA was extracted from all tissue types using a Qiagen RNeasy kit (Qiagen, Valencia CA) according to the manufacturer's instructions. On column digestion of genomic DNA was performed using the Qiagen RNAse-Free DNAse set following the manufacturer's protocol. Approximately 100 mg of frozen ground tissue was utilized for each extraction. RNA was denatured and visualized on an agarose gel to assess quality. Quantity was assessed using a NanoDrop (Nanodrop, Wilmington, DE).

RT-PCR

First strand cDNA was synthesized using the Protoscript First Strand cDNA Synthesis Kit from New England Biolabs (New England Biolabs, Ipswich, MA) as per manufacturer's instructions. 500ng of total RNA was used as starting material for each reaction. First strand cDNA was diluted in nuclease free water at a 1 to 10 dilution and used for RT-PCR. *Ubiquitin* Conjugating Enzyme (E2) F and R primers were designed as positive controls for cDNA synthesis and amplification based on an EST sequence (accession number DQ924970) identified in GenBank (Benson *et al*, 2009). *pun1*^{3} specific PCR primers that span the intron-exon boundaries were designed based on an alignment of the *C. frutescens* PI594141-P and *C. frutescens* PI594141-np *pun1*^{3} sequences in order to exclude both genomic DNA contamination and contamination from *catf2*, a homolog to *AT3-1* known to be expressed in developing pepper fruits but that does not co-segregate with pungency (Garces-Claver *et al*, 2007). A PCR reaction designed to simultaneously amplify both the coding region of *pun1*^{3} (680bp band) and

Ubiquitin (129bp band) was performed as per the following protocol: 5µl 10X Buffer, 4µl 2.5 mM dNTPs, 1µl 10µM *pun1*^{3} RT-PCR forward (TGGCAGTTTCCCTTCTCTC), 1µl 10µM *pun1*^{3} RT-PCR reverse (GGGAATAGCCATCAGTGTATGCTTTTCG), 1µl 10µM *Ubiquitin* RT-PCR forward (TGTGTCTCAACATTCTTCGTGA), 1µl 10µM *Ubiquitin* RT-PCR reverse (ATACAGCAGCTGCGTCGT), 0.25µl ExTaq (Takara Bio Inc. Otsu, Japan), 1µl 1:10 cDNA dilution, H_2O to a final volume of 50µl, and cycling with the following conditions: 94°C for 3 minutes, 29 cycles of 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute 30 seconds, with a final extension at 72°C for 15 minutes. Reactions were performed using a PTC 225 Peltier Thermal Cycler (MJ Research, Watertown, MA).

Genotyping of C. frutescens BG2814-6 x C. frutescens PI594141 non-pungent population

DNA from F_2 segregating populations was extracted using a modified CTAB protocol (Doyle, 1990). Polymorphic simple sequence repeats (SSR) markers known to flank the *Pun1* locus on pepper chromosome 2 (http://www.sgn.cornell.edu) were used for genotyping: CA514272 (forward: ATCTATTTTCCTCCGGCGAC, reverse: CGGTAAGCTGCCTTGATCTC) and CA514621 (forward :GTCGAACAAAATGGGGTTTG, reverse: GCTGGAGAGTGCTGGTGG). PCR conditions were 94°C for 3 minutes; 29 cycles of 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds; with a final extension at 72°C for 5 minutes. Polymorphic bands were separated using denaturing polyacrylamide gel electrophoresis (PAGE) with 6% acrylamide and visualized using silver staining.

A co-dominant marker specific to the *pun1*^{3} mutation was designed in order to

genotype the segregating F_2 populations for both the greenhouse and field populations (see Figure 2.2). The *pun1*^{3} specific reverse primer was anchored in the inactivating mutation found in *C. frutescens* PI594141-np. A common forward primer was used that would enable amplification of both mutant and wild type alleles. This forward primer was paired with a reverse primer anchored in the second exon of *C. frutescens* PI594141-P that is located in the deleted region of *C. frutescens* PI594141-np. PCR conditions were as follows: 2.5µl 10X PCR buffer, 2µl 2.5 mM dNTPs, 1µl 10µM *Pun1* forward (GTAGTTTTTCGGAAATGAAAAGTACTG), 1µl 10µM *Pun1* 6R reverse (CACGCCTTGCCCAGCTTTGTAATCTTTC), 1µl 10µM *pun1*^{3} reverse (TCATGTCCATTCGGCCAAACAGTG), 0.25µl ExTaq (Takara Bio Inc. Otsu, Japan), 2µl genomic DNA solution and H_2O to a final volume of 25µl. PCR cycles were as follows: 94°C for 3 minutes, 34 cycles of 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute 30 seconds, with a final extension at 72°C for 15 minutes. Products were visualized on a 1.2% agarose gel.

Genotyping of C. chacoense PI260433 non-pungent x C. chinense 'Habanero'

Polymorphic markers co-segregating with the non-pungent phenotype were identified using bulk segregant analysis (BSA) (Michelmore *et al*, 1991). Approximately 30 markers (http://sgn.cornell.edu) sampling each of the 12 pepper chromosomes and both chromosome arms were screened using pooled F_2 DNA.

A single marker, HPMS1-172 (Forward primer: GGGTTTGCATGATCTAAGCATTTT, Reverse primer: CGCTGGAATGCATTGTCAAAGA) co-segregated with the non-pungent bulk. This marker had been previously shown to map to Pepper Chromosome 7 in a region with a known QTL affecting capsaicinoid accumulation (Ben-Chaim *et al*, 2006; Blum *et al*,

2003). This marker was scored on the entire population along with COSII markers that are known to be located on the same arm of chromosome 7 (Wu, 2009) according to protocols described therein. In that study polymorphisms were selected that discriminated *C. frutescens* and *C. annuum* haplotypes, so the markers needed to be adapted to a cross between *C. chinense* and *C. chacoens*e. Amplified fragments were excised and gel purified using the GFX PCR DNA and Gel Band Purification Kit (GE, Healthcare Biosciences Inc., Piscataway, NJ), cloned into a pCR4 TOPO TA vector and used to transform chemically competent TOP10 cells according to manufacturer instructions (Invitrogen Corporation, Carlsbad, CA) Individual colonies were prepared by Miniprep using the Bioneer Plasmid Extraction Kit (Bioneer, Alameda, CA) and sequenced at the OSA Sequencing Facility at SUNY Stony Brook (http://www.osa.sunysb.edu/dna/).

Restriction enzyme site differences were used to distinguish the *C. chacoense* PI260433-np allele from the *C. chinense* 'Habanero' allele for each marker screened. PCR was performed as above and restriction enzyme digests were performed as follows: 20µl PCR product, 3µl of the appropriate buffer, 1µl enzyme, 3µl 10X BSA and H_2O to a final volume of 30µl, incubated at 37°C for 3 hours and visualized by agarose gel electrophoresis. One COSII marker polymorphism was found that was linked to pungency, At2g24270, The PCR product obtained from COSII At2g24270 showed a restriction enzyme polymorphism difference when digested with AflII and therefore was used to genotype the *C. chacoense* PI260433-np x *C. chinense* 'Habanero' population.

A co-dominant marker specific to the *Pun1* locus was designed in order to genotype the *C. chacoense* PI260433-np x *C. chinense* 'Habanero' F_2 population. PCR was performed as stated above using the following primer pair *Pun1* F (GGTCTAGCGTTACTCGTGATCATACG) and *Pun1* R

(TCAAACACCACAAAAGACTTGGA) followed by digestion of the PCR product with ScaI.

Map location identification

MapMaker/EXP v. 3.0b was used to identify map locations and chromosomal distances and results were compared to the pepper FA03 map (http://www.sgn.cornell.edu) Map distances for the *C. frutescens* BG2814-6 x *C. frutescens* PI594141-np population were calculated using the following set of commands: "group," "pair" and "sequence" and as only 2 most likely map orders were identified, both of which were symmetric, the command "map" was used to identify the best possible map distances.

For the population *C. chacoense* PI260433-np x *C. chinense* 'Habanero' the command 'LOD table' was used to identify even marginal linkage between the chromosome 2, chromosome 7 markers and the capsaicinoid accumulation phenotype. As linkage was found between a COSII marker located on chromosome 7 and the pungency phenotype, the parameters LOD 2, 30 cM were chosen for the "group" command. Two linkage groups were identified which were then sequenced using the command "sequence" followed by the command "map" in order to identify the best possible map distances.

Capsaicinoid detection

Capsaicinoids were measured in mature dry fruit using a capsaicinoid detecting ELISA kit from Beacon Analytical (Beacon Analytical, Portland, ME). Capsaicinoids were extracted and measured as per manufacturer's instructions. When feasible each sample was measured multiple times.

Initial Alignment

Novel *Pun1* sequences identified in the course of this study, *C. frutescens* PI594141-np and *C. chacoense* PI260433-P were deposited in GenBank (accession numbers FJ871985 and FJ871984 respectively). An initial alignment of the 21 *Pun1* sequences, which include those from *Capsicum spp.* as well as from representative taxa in the Solanaceae and the outgroup sequence from *C. annuum*, *catf2*, was compiled using the program DIALIGN (http://bibiserv.techfak.uni-bielefeld.de/dialign/) and refined by hand (Morgenstern, 2004). Despite the absence of full-length sequence, *C. annuum catf2* was chosen as the outgroup for the dataset due to its high degree of nucleotide sequence similarity with *AT3*. *catf2* expression does not co-segregate with pungency and does not map to the same genomic region as *Pun1* (Garces-Claver *et al*, 2007). This first alignment of 21 sequences and 1876 characters contained the 5' UTR of *AT3* and the coding regions of *AT3* from the putative start codon to approximately 100 base pairs upstream of the stop codon, including the intron that is highly conserved and can be easily aligned. The alignment was truncated in the second exon at the site of the mutation where the *C. frutescens* PI594141-np sequence lost all similarity to the remaining sequences.

ModelTest

Prior to generating the Bayesian phylogenies, the best model of evolution was determined using the program ModelTest v3.7 (Posada and Crandall, 1998). Three separate best-fit models were estimated: the first for the first exon of *AT3*, the second for the intron, and the third for the second exon. ModelTest v3.7 estimated three best-fit models for the data. The first exon best fit the model with six rate classes, General Time Reversible (GTR) with a gamma distribution of the rate variation among sites and no invariable sites (Tavare, 1986). The intron and second exon sequences best fit the model with a single rate class, F81, equal rates of nucleotide frequencies and the proportion of invariable sites equal to zero (Felsenstein, 1981).

Bayesian Analysis

Bayesian analysis of the alignment including the introns was performed in MrBayes v3.1.2 by creating three data partitions in the alignment, one for the first exon, one for the second exon and one for the intron (Ronquist and Huelsenbeck, 2003). Each partition was assigned its best-fit model as per ModelTest v3.7 and rates were allowed to vary independently between partitions. *catf2* was set as the outgroup. A total of four chains of the Markov Chain Monte Carlo were run, sampling one tree every 100 generations. The analysis was allowed to proceed until the standard deviation of split frequencies fell below 0.01. A fixed proportion of trees comprising 25% of all trees sampled were discarded as the "burn in" fraction. A majority rule consensus tree of all trees sampled, excluding the burn in faction, was computed using PAUP v. 4.0b10 (Swofford, 2002).

RESULTS

Tests for linkage between Pun1 and loss of pungency in C. frutescens PI59414 and in C. chacoense PI260433

Two non-pungent accession of *Capsicum* were identified in the germplasm collections and in previous publications (Tewksbury *et al*, 2006; Votava and Bosland, 2002). PI594141 was originally identified as *C. eximium*, but recently reassigned to *C. frutescens*. A careful inspection of the fruit and flowers agree with this revision (M. Nee and L. Bohs, personal communication). Tasting of ripe fruit in an initial planting of both *C. frutescens* PI594141 and *C. chacoense* PI260433 showed that these accessions contained both pungent and non-pungent individuals as has been noted for *C. chacoense* populations in South America (Tewksbury *et al*, 2006). A single pungent and non-pungent plant was selected from each accession and propagated by single seed descent for 3 generations. The *C. frutescens* selection progeny had no detectable capsaicinoids by taste or ELISA assay for capsaicinoids, but the *C. chacoense* plants consistently had trace levels of capsaicinoids. Capsaicinoids produce a sensation of pungency at concentrations as low as 10ppm, but are detectable on the tongue at concentrations as low as 1ppm (Andrews, 1984). One part per million is also the limit of detection for the ELISA kit used in this study. The *C. chacoense* line consistently accumulated between 1 and 10ppm capsaicinoids in ripe fruit as determined by this assay, a measurement that correlated with samples of fruit that were tasted. Therefore, while we refer to the *C. chacoense* line as "non-pungent," it accumulates trace amounts of capsaicinoids, thus we deem it to be a dramatic knock-down capsaicinoid mutant. These plants represent the starting germplasm used for the characterization of the loss of pungency mutation in *C. frutescens* PI594141 and in *C. chacoense*

PI260433.

In order to establish the number of loci responsible for loss of pungency and to identify map location(s) for this trait, *C. frutescens* PI594141-np was crossed to a highly pungent *C. frutescens*, BG2814-6, which has been used extensively in mapping experiments (Ben Chaim et al., 2006, www.sgn.cornell.edu). Despite performing numerous manual cross-pollinations between *C. frutescens* BG2814-6 and *C. frutescens* PI594141-np, only a single F_1 plant germinated from this cross, likely related to the extremely low fruit set and seed content of PI594141-np. Fruits from this F_1 plant were uniformly pungent and comparable to the levels of pungency seen in the pungent parent, revealing that loss of pungency in *C. frutescens* PI594141 is recessive (Table 2.1). Segregation of pungency in two F_2 populations grown under field and greenhouse conditions showed that the ratio of pungent to non-pungent plants is consistent with a 3:1 ratio suggesting Mendelian inheritance of a single recessive locus causing loss of pungency in *C. frutescens* PI594141 (Table 2.1).

As an initial test of the hypothesis that *C. frutescens* PI594141 is non-pungent due to an allele of *pun1*, a map location for loss of pungency was established for the cross between *C. frutescens* BG2814-6 x *C. frutescens* PI594141-np. Simple sequence repeat (SSR) markers known to flank the *Pun1* locus in the cross between *C. frutescens* BG2814-6 and *C. annuum* 'R Naky' were used to show that loss of pungency in *C. frutescens* PI594141-np maps to the same interval as *Pun1* (Figure 2.1). Further, a marker based on a polymorphism within the *Pun1* locus perfectly cosegregated with loss of pungency in this cross as described below. This suggested that the previously unreported loss of pungency event in *C. frutescens* PI594141-np is due to a mutation at the *Pun1* locus and was therefore tentatively designated *pun1*^{3}.

Table 2.1. Single locus goodness of fit test for loss of pungency in *C. frutescens*.

Generation	*Total*	*# Pungent*	*# Non-pungent*	*Exp. Ratio*	χ^2	*df*	*p*
C. frutescens BG2814-6 x *C. frutescens* PI594141-np field grown							
P_aS_1	4	4 (3121±1680)[1]	0				
P_bS_1	4	0	4 (ND)[2]				
F_1	1	1 (3294)	0				
F_2	60	48 (5220±2708)	12 (ND)	3:1	0.8	1	0.37
C. frutescens BG2814-6 x *C. frutescens* PI594141-np greenhouse grown							
P_aS_1	1	1 (12794)	0				
P_bS_1	5	0	5 (ND)				
F_1	1	1 (3294)	0				
F_2	41	33 (5934±4555)	8 (ND)	3:1	0.7	1	0.40
C. frutescens BG2814-6 x *C. frutescens* PI594141-np pooled							
P_aS_1	5	5 (5506±4569)	0				
P_bS_1	9	0	9 (ND)				
F_1	1	1 (3294)	0				
F_2	101	81 (5511±3568)	20 (ND)	3:1	1.46	1	0.22

[1] Numbers in parentheses represent the average capsaicinoid accumulation in parts per million ± the standard deviation.
[2] Values below the detection limit of the assay (1ppm) were designated non-detectable (ND)

A test of genetic linkage between loss of pungency in *C. chacoense* PI260433-np and the *Pun1* locus was attempted in a cross to *C. frutescens* BG2814-6. Several F_1 plants based on this cross germinated but produced only parthenocarpic fruit. *C. chacoense* PI260433-np was then crossed to the highly pungent *C. chinense* ''Habanero''. Several F_1 individual plants were obtained all of which were pungent, indicating that loss of pungency in this accession is also recessive (Table 2.2). Although numerous seeds were obtained from these F_1 plants, few germinated (less than 20%) and of those that germinated, some did not set fruit in the F_2 generation. Despite the difficulty in crossing a wild non-pungent *Capsicum* species to a cultivated variety within the *annuum*, *chinense* and *frutescens* species complex, segregation of

loss of pungency in the F_2 population showed that loss of pungency in *C. chacoense* PI260433-np is consistent with a 3:1 ratio, suggesting that a single recessive mutation controls loss of pungency in this accession (Table 2.2).

Table 2.2. Single-locus goodness of fit test for loss of pungency in *C. chacoense*.

Generation	*Total*	*# Pungent*	*# Non-pungent*	*Ratio*[2]	χ^2	*df*	*p*
C. chacoense PI260433-np x *C. chinense* 'Habanero'							
P_aS_1	4	0	4 (5.28±4.53)[1]				
P_bS_1	0	5(23904±2564)	0				
F_1	7	7 (6212±4207)	0				
F_2	51	42 (7848±8644)	9 (7.67±6.07)	3:1	1.47	1	0.23

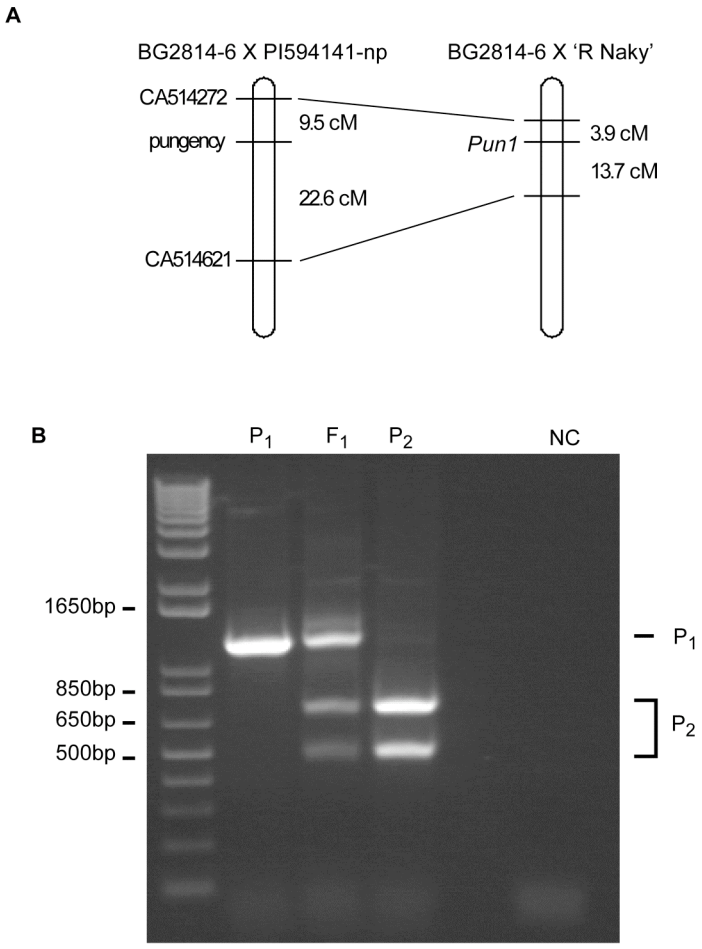
[1] Numbers in parentheses represent the average capsaicinoid accumulation in parts per million ± the standard deviation.
[2] Ratio represents the expected ratio

In this population, segregation of loss of pungency with the *Pun1* locus was tested using a CAPS marker designed based on a restriction site polymorphism in the intron of *Pun1* (Figure 2.1B) and the same SSR markers known to flank the *Pun1* locus (Figure 2.1A). The CAPS marker designed to distinguish between the parental alleles of *Pun1* was not linked with the pungency phenotype, nor were the flanking SSR markers, establishing that loss of pungency in *C. chacoense* PI260433-np maps to a region other than the *Pun1* locus and was tentatively designated *pun2*.

Figure 2.1. Map position for loss of pungency in *C. frutescens* PI594141-np

A. The loss of pungency phenotype in an F_2 population derived from a cross between *C. frutescens* BG2814-6 and *C. frutescens* PI594141-np mapped between SSR markers CA514272 and CA514621 on pepper chromosome 2. This map order was colinear with the corresponding region in a cross between *C. frutescens* BG2814-6 and *C. annuum* 'R Naky' (www.sgn.cornell.edu, FA03 map).

B. CAPS marker used to genotype the *Pun1* locus in the *C. chacoense* PI260433-np x *C. chinense* ''Habanero'' population and in the *C. chacoense* PI260433-np x *C. chinense* 'NMCA30036' F_2 populations is shown.


A
BG2814-6 X PI594141-np
BG2814-6 X 'R Naky'
CA514272
pungency
CA514621
9.5 cM
22.6 cM
Pun1
3.9 cM
13.7 cM
B
P1
F1
P2
NC
1650bp
850bp
650bp
500bp
P1
P2

Allelism of loss of pungency mutations in C. frutescens PI594141 and in C. chacoense PI260433

The results of the *Pun1* linkage analysis suggest that loss of pungency in *C. frutescens* PI594141 was due to a mutation at the *Pun1* locus whereas in *C. chacoense* PI260433 the mutation was attributed to a locus other than *Pun1*. This hypothesis was tested by using a complementation test including all known loss of pungency mutations in pepper, all of which are monogenic and recessive (Stewart *et al*, 2005; Stewart *et al*, 2007). Crosses were performed in all possible combinations, but due to known interspecific sterility barriers only some hybrids were recovered (Onus and Pickersgill, 2004; Pickersgill, 1971). Two complementation groups were defined, *pun1* and *pun2* (Table 2.3). As no significant or measurable levels of capsaicinoid accumulation were detected in fruit from F_1 plants resulting from the crosses between *C. frutescens* PI594141-np and the other reported non-pungent accessions, the results of the experiment shown in Table 2.3 demonstrate that the mutation in the accession *C. frutescens* PI594141-np fails to complement the *pun1* and $pun1^2$ mutations (Table 2.3). This establishes that loss of pungency in *C. frutescens* PI594141-np is allelic to the other mutations at the *Pun1* locus and this novel allele of *Pun1* as $pun1^3$. Loss of pungency in *C. chacoense* PI260433-np complements the *pun1*, $pun1^2$ and $pun1^3$ mutations. This confirms our hypothesis that loss of pungency in *C. chacoense* PI260433-np is due to a mutation at a novel loss of pungency locus, which we name *pun2*.

Table 2.3. Complementation test results between *Capsicum* pungency mutants. Each cell shows the average value and standard deviation in ppm of capsaicinoid levels of fruit from duplicate measurements of at least 3 plants unless 3 plants were not available.

		Female parent			
	Genotype	*C. annuum* ECW	*C. chinense* 'NMCA30036'	*C. frutescens* PI594141-P	*C. chacoense* PI260433-np
Male parent	*C. annuum* ECW	N.D.			
	C. chinense 'NMCA30036'	N.D.*	N.D.	N.D.	
	C. frutescens PI594141-np			N.D.	
	C. chacoense PI260433-np	3032 ± 754	5122 ± 1299	6114 ± 1868	1.83 ± 1.59

(*) Cells left blank indicate crosses that could not be performed.
N.D. indicates capsaicinoids not detectable.

Both *Pun1* and *Pun2* are required for pungency therefore in an F_2 population with their mutant alleles segregating we would expect a 9:7 ratio that is characteristic of complementary gene action. An F_2 population resulting from the cross between *C. chacoense* PI260433-np (*pun2*) and *C. chinense* 'NMCA30036' ($pun1^2$), which is allelic to the mutation in *C. frutescens* PI594141-np, was grown in the field, phenotyped for capsaicinoid accumulation and genotyped at the *Pun1* locus (Table 2.4). The $pun1^2$ mutation is found in a *Capsicum chinense* background, and was used as the *Pun1* mutant for this cross because in our hands, *C. chinense* has better interspecific fertility with *C. chacoense* PI260433 than *C. frutescens* which harbors the $pun1^3$ allele. The population that was obtained was too small to rigorously test a genetic ratio, but two pungency classes were observed with roughly equivalent numbers of individuals. Within the non-pungent class, individuals were observed of both *Pun1* and $pun1^2$ genotypes (Table 2.4), suggesting that *pun2* indeed has a recessive epistatic interaction with *Pun1*.

Table 2.4. Epistatic interactions between the *Pun1* and *Pun2* loci. Each row shows the genotype at the *Pun1* locus, the number of plants in each genotypic category (n) and the capsaicinoid accumulation in parts per million (ppm) with standard deviation values.

Genotype	n	Pungency (ppm)
Pun1/___	10	1,215 ± 1,696
Pun1/___	4	4.9 ± 4.4
$pun1^2/pun1^2$	5	3.0 ± 4.9

Genetic mapping of pun2

In order to identify the map location of *pun2*, bulk segregant analysis (BSA) was then used to screen over 30 loci distributed throughout the pepper genome, sampling each chromosome arm. By BSA, one marker, Hpms1-172, was identified, but could only be genotyped as a dominant marker as more than one locus was amplified in this cross. Data from other mapping populations showed that the marker Hpms1-172 maps to the upper arm of pepper chromosome 7 on the FA03 map (www.sgn.cornell.edu). Additional markers were sought for this region to confirm this map location for *pun2*. As few SSR markers were available for this genomic region that were conserved or polymorphic between the two parents, conserved orthologous sequence (COSII) markers known to map near Hpms1-172 were screened (Mueller *et al*, 2005; Wu, 2009). Despite the use of an interspecific cross for this population, polymorphism was limiting. However, linkage between COSII marker C2_At2g24270 and loss of pungency was observed (19.8 cM, LOD 2.3). A major QTL controlling levels of capsaicinoid accumulation has been mapped to this region in a segregating population between *C. frutescens* BG2814-6 x *C. annuum* 'R Naky' (Ben-Chaim *et al*, 2006) suggesting that *pun2* may therefore represent an allele of that locus.

Sequence of the Pun1 locus in C. frutescens PI594141 non-pungent and in C. chacoense PI260433 non-pungent

Having established that loss of pungency in *C. frutescens* PI594141 mapped to the same genomic region as the *Pun1* locus, *AT3*, the candidate gene underlying the *Pun1* locus, was sequenced in order to identify any possible sequence polymorphisms that could be associated with loss of pungency in this accession (Stewart *et al*, 2007). Inverse PCR revealed that in *C. frutescens* PI594141-np *AT3* shows a large sequence deletion when compared to the wild-type sequence obtained in the sister line *C. frutescens* PI594141-P and other known *Pun1* sequences. *at3* from *C. frutescens* PI594141-np has a deletion of terminal 70 amino acids which are replaced by a repetitive region showing numerous SSRs and long repetition of the bases adenine and thymine (Figure 2.2). The repetitive sequence identified modifies the putative amino acid sequence at this locus and introduces a stop codon 64 amino acids from the position of the wild-type stop codon. This would lead to a truncated protein of 376 amino acids and lacking the DFGWGKP which is characteristic for the BAHD family of acyltransferases and is present in the wild type sequence from sister line *C. frutescens* PI594141-P (St-Pierre and De Luca, 2000). Despite obtaining over 3kb of sequence downstream from the beginning of the interruption from *C. frutescens* PI594141-np *AT3*, we were not able to recover the final 210 basepairs for this sequence, thus we were not able to detect the origin of the genomic interruption causing the mutation of this novel allele of *Pun1*.

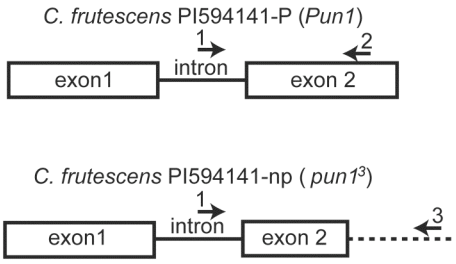
Figure 2.2. *pun1*^{3} genotyping

A. The *pun1*^{3} mutation is a large insertion or deletion that results in a truncated second exon. Genotyping primers were designed based on this mutation that discriminated wild-type *Pun1* and *pun1*^{3}. Primers 1 and 2 are anchored in exons, but primer 3 is located in the genomic sequence past the mutation site in *pun1*^{3}.

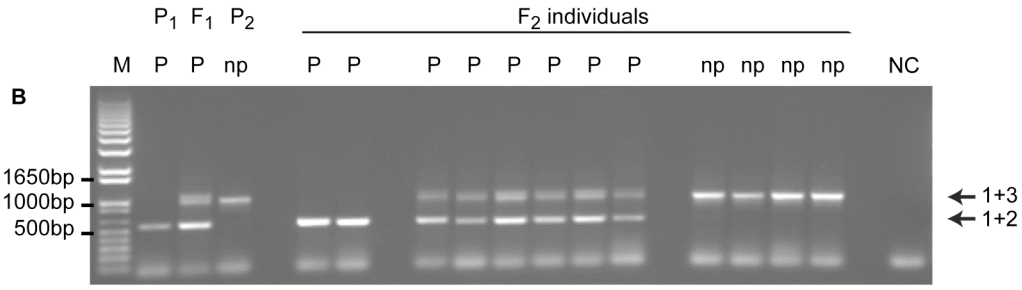
B. The primer set shown in A allows for codominant genotyping of populations segregating for *pun1*^{3}. Primer pair 1 and 3 produces a band of 1033bp in the presence of the *pun1*^{3} allele, while primer pair 1 and 2 produces a band of 586bp in the presence of a wild-type *Pun1* allele. The parental genotypes of *C. frutescens* BG2814-6 (P_1), *C. frutescens* PI594141-np (P_2) and their F_1 hybrid used as the parent of the F_2 can be distinguished. F_2 segregants are shown: *Pun1*/*Pun1* and pungent (P), *Pun1*/*pun1*^{3} and pungent and *pun1*^{3}/*pun1*^{3} and non-pungent (NP). No bands are found in the negative control (NC).

C. Capsaicinoid accumulation was measured by ELISA in red ripe fruits of the plants shown in B. No capsaicinoids were observed in *C. frutescens* PI594141-np or *pun1*^{3}/*pun1*^{3} F_2 individuals. *Pun1*/*Pun1* and *Pun1*/*pun1*^{3} individuals had a wide range of capsaicinoid levels from less than 200 ppm to more than 1,000 ppm.

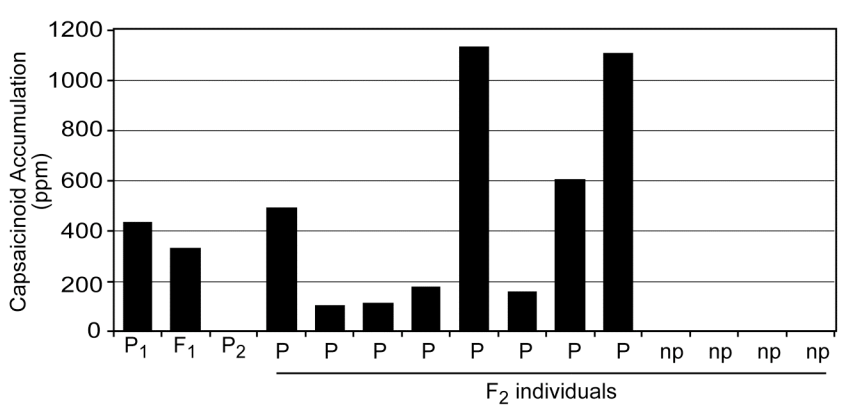
A


C. frutescens PI594141-P (Pun1)
1
2
exon1
intron
exon 2
C. frutescens PI594141-np (pun1^3)
1
3
exon1
intron
exon 2

B


P1 F1 P2
F2 individuals
M P P np P P P P P P P np np np np NC
1650bp
1000bp
500bp
1+3
1+2

C


Capsaicinoid Accumulation (ppm)
1200
1000
800
600
400
200
0
P1 F1 P2 P P P P P P P P np np np np
F2 individuals

Through comparative sequence alignment with the wild-type sequence in *C. frutescens* PI594141 pungent as well as other wild type *Pun1* alleles, we were able to design a reverse primer anchored in the mutation in *C. frutescens* PI594141-np that in combination with primers anchored in conserved region of *Pun1*, allows for codominant genotyping (Figure 2.2). It further demonstrates that loss of pungency in *C. frutescens* PI594141 can be attributed to a deletion in the coding region of the gene, consistent with the previously identified mutations at the *Pun1* locus. In contrast, the coding region of *Pun1* in *C. chacoense* PI260433-np was established to be wild type and no obvious sequence polymorphisms were identified other than a 36 base pair insertion in the intron, which is conserved between the pungent and non-pungent *C. chacoense* PI260433 lines, further suggesting that the *Pun1* locus is not responsible for loss of pungency in *C. chacoense* PI260433-np.

It is known that expression of the *Pun1* locus as well as other capsaicinoid structural genes mirrors the location and developmental time course of capsaicinoid accumulation in pepper fruit (Aluru *et al*, 2003; Curry *et al*, 1999; Stewart *et al*, 2005; Stewart *et al*, 2007). The beginning of *Pun1* expression coincides with the beginning of capsaicinoid accumulation, peaking at around 20 days post anthesis (dpa). *Pun1* expression begins to decline around 40dpa when capsaicinoid accumulation has arrived at its maximum. RT-PCR using primers specific to the *Pun1* locus shows no $pun1^3$ expression in *C. frutescens* PI594141-np (Figure 2.3).

In contrast, expression of *Pun1* in *C. frutescens* PI594141-P is prominent around 20dpa, declines significantly by 40dpa, and is overall consistent with the pattern of *Pun1* expression previously reported (Stewart *et al*, 2005). This is in contrast with what has been demonstrated for the $pun1^2$ mutation in *C. chinense* 'NMCA30036'. In this non-pungent cultivar $pun1^2$ is weakly expressed as detected by northern blot hybridization (Stewart *et al*, 2007). However, translation of the locus appeared to be

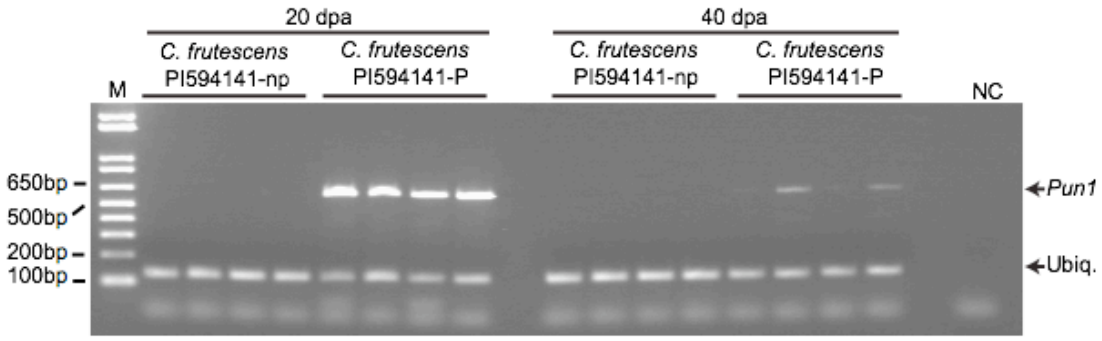
absent as no protein accumulation was detectable by western blot hybridization.

Genetic architecture of known Pun1 alleles

Although cultivars of *C. annuum*, *C. chinense* and *C. frutescens* all share mutations at the same locus that are responsible for loss of pungency, the nature and putative origin of each mutation appears to be quite different (Figure 2.4A). In *C. annuum* ECW (and other non-pungent cultivars), a 2.5kb deletion removes a majority of the first exon, and the putative promoter region thereby eliminating transcription and translation of the locus. In *C. chinense* 'NMCA30036', a 4bp deletion causes a frameshift mutation thereby reducing transcription and eliminating translation of the locus. Similarly, in *C. frutescens* PI594141-np, the deletion of part of the second exon, including the stop codon introduces a frameshift mutation that correlates with loss of transcription of the locus. This mutation also perfectly co-segregates with loss of pungency in segregating populations contrasting for the trait. Although it is possible that the mutation in *C. annum* was caused by secondary deletion following the mutation in *C. chinense*, the mutation in *C. frutescens* is clearly distinct from either of the other two as it occurs in the second exon, thus cannot have arisen from either the *pun1* or $pun1^2$ mutations, except by an unlikely process of reversion.

Figure 2.3. *Pun1* expression in fruit of wild-type and mutant peppers

Pepper fruit were collected 20 and 40 days post anthesis (dpa) and RT-PCR was performed using primers that provided PCR amplification of both *Pun1* and *pun1*^{3} alleles from genomic DNA. Robust expression of *Pun1* was observed from 20dpa fruit that were pungent, but no expression was observed in *pun1*^{3}, non-pungent fruit. Expression diminished in 40 dpa pungent fruit. An internal *ubiquitin* conjugating enzyme (E2) control was equivalent in all samples. Neither band was produced in a negative control without cDNA (NC).


20 dpa
40 dpa
C. frutescens PI594141-np
C. frutescens PI594141-P
C. frutescens PI594141-np
C. frutescens PI594141-P
M
NC
650bp
500bp
200bp
100bp
Pun1
Ubiq.

To test whether the *Pun1* null alleles all share a common origin despite showing different sequence interruptions, Bayesian phylogenetic analysis of the mutations was conducted using a dataset of full-length *Pun1* sequences assembled in the course of a different study (Fellman, Stellari and Jahn unpublished data). The phylogeny suggests that each null *Pun1* allele has a distinct phylogenetic origin and that each mutation arose independently (Figure 2.4B). At the same time, the *Pun1* locus in *C. chacoense* PI260433-np, which is wild-type in sequence is shown to be most closely related to the *Pun1* sequence from its pungent sister line *C. chacoense* PI260433-np, thereby suggesting gene flow between the pungent and non-pungent *C. chacoense* populations for this allele. This is consistent with the observation of intermixed populations of these plants in Bolivia (Tewksbury *et al*, 2006). The placement of *C. frutescens* PI594141-np in the phylogeny suggests instead that this null-allele may not be most closely related to the allele identified in the pungent sister line. Furthermore, its phylogenetic placement implies that the allele arose sometime before the origin of the *annuum*, *chinense* and *frutescens* species complex and may have been maintained in *C. frutescens* PI594141 by lineage sorting. Although we have not thoroughly sampled from the wild *Capsicum* species and thus cannot explore in more detail the evolution of the genus, the phylogeny suggests that the *pun1* and $pun1^2$ alleles are of ancient origin and predate the diversification of some of the species we currently recognize in the genus *Capsicum*.

DISCUSSION

The *Pun1* locus had been the only locus identified that had a qualitative effect on pungency (Blum-Eyal *et al*, 2002). Identification of two instances in sister taxa where loss of pungency is caused by a mutation at the *Pun1* locus suggested that all

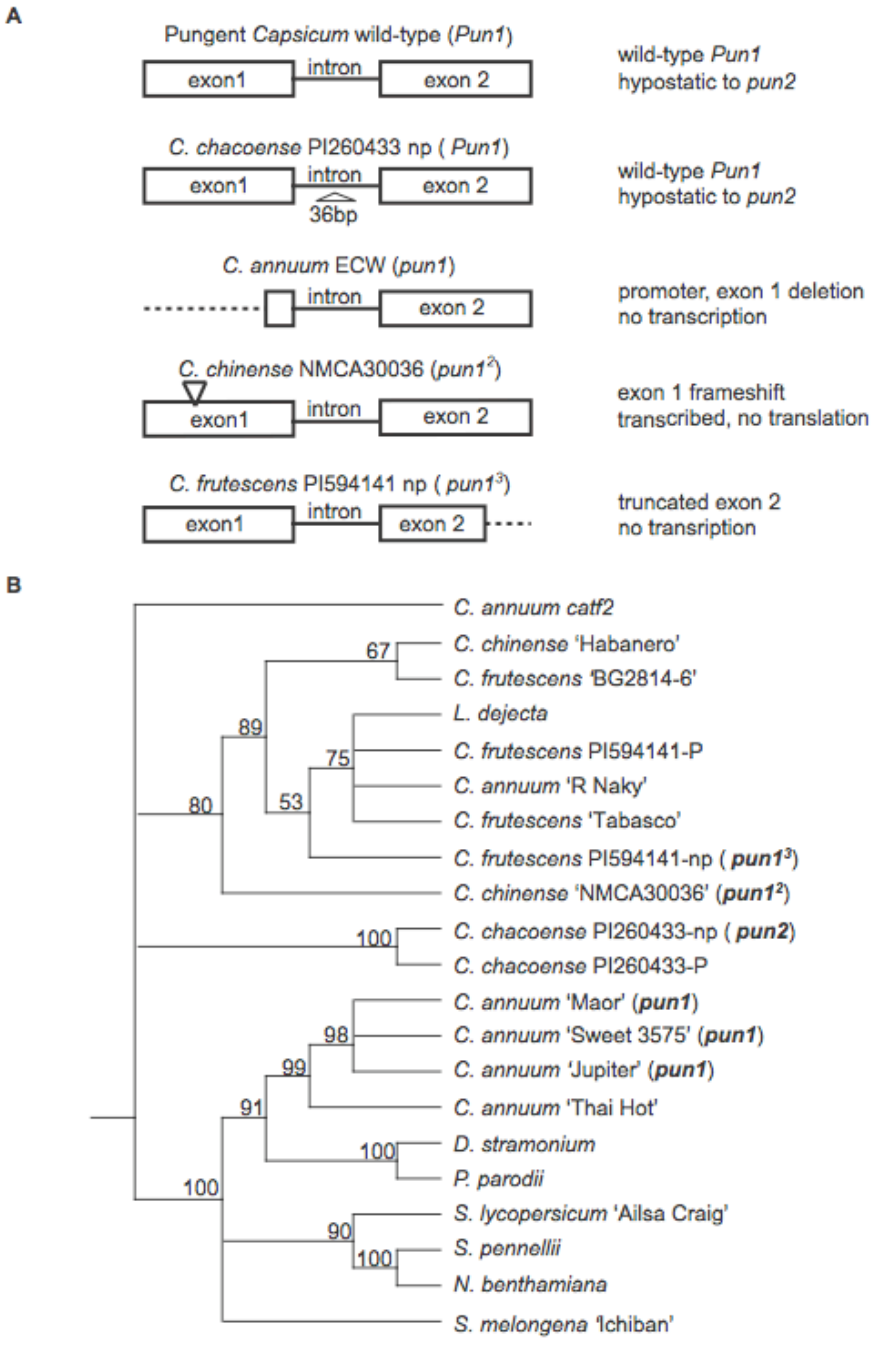
naturally occurring loss of pungency mutations in domesticated *Capsicum* would be mutations in the *Pun1* locus (Stewart *et al*, 2005; Stewart *et al*, 2007). In the effort to complete the characterization of loss of pungency in known non-pungent *Capsicum* accessions, we described the genetic basis for this trait in two non-pungent accessions. One of these was first identified in this study from the ARS-GRIN repository, *Capsicum frutescens* PI594141. The other, *Capsicum chacoense* PI260433, was known from the literature to be the subject of extensive studies investigating the ecological function of capsaicin (Levey *et al*, 2006; Tewksbury *et al*, 2006; Votava and Bosland, 2002). We hypothesized that there would be a difference in the genetic basis for loss of pungency in cultivated versus non-cultivated species in the genus *Capsicum* consistent with the phylogenetic groupings of the species.

Our results demonstrate the mutation in *C. frutescens* PI594141-np is allelic to the other two previously identified loss of pungency loci found in the cultivated species of pepper whereas the mutation in *C. chacoense* PI260433-np is due to a mutation at a different genetic locus (Table 2.3). Interestingly, *C. annuum*, *C. chinense* and *C. frutescens* have a close phylogenetic relationship, dubbed the *annuum-chinense-frutescens* complex (Olmstead *et al*, 2008; Pickersgill, 1971; Walsh and Hoot, 2001). The propensity for these three domesticated species to share mutations at the *Pun1* locus leading to loss of pungency may reflect differences in the nature of selection that distinguishes wild populations from cultivated varieties or a genomic factor within the domesticated clade that predisposes mutations at the *Pun1* locus.

Figure 2.4. Relationship of *Pun1* alleles in *Capsicum*

A. *Pun1* alleles from *C. annuum*, *C. chinense*, and *C. frutescens* differ by SNPs scattered throughout the gene. The *Pun1* allele recovered in *C. chacoense* PI260433-np contains an intron insertion and several SNPs in the intron that were used for genotyping. As observed in *C. chacoense* and *C. chinense* F_2 progeny, pungency requires both *Pun1* and *Pun2* alleles. A large 5' deletion produced $pun1^1$, a frameshift mutation in the second exon is found in $pun1^2$ and an indel truncates the second exon of $pun1^3$. The $pun1^1$ and $pun1^3$ mutations are not transcribed, while transcription but not translation has been observed for $pun1^2$.

B. Phylogeny of *Pun1* alleles of selected *Solanaceae*. Bayesian phylogenetic analysis of 21 *Pun1* alleles from a diverse sampling of species and cultivars within *Capsicum* as well as from domesticated and representative genera in the *Solanaceae* shows that the *pun1* series of null alleles are scattered not clustered together into a single clade suggesting multiple origins of the loss of pungency alleles followed by lineage sorting. *Pun1* from *C. chacoense* PI260433-np is most closely related to the sister line *C. chacoense* PI260433-P suggesting possible gene flow between the two populations for this allele.


A
Pungent Capsicum wild-type (Pun1)
exon1
intron
exon 2
wild-type Pun1
hypostatic to pun2
C. chacoense PI260433 np (Pun1)
exon1
intron
exon 2
36bp
wild-type Pun1
hypostatic to pun2
C. annuum ECW (pun1)
intron
exon 2
promoter, exon 1 deletion
no transcription
C. chinense NMCA30036 (pun1²)
exon1
intron
exon 2
exon 1 frameshift
transcribed, no translation
C. frutescens PI594141 np (pun1³)
exon1
intron
exon 2
truncated exon 2
no transription
B
C. annuum catf2
C. chinense 'Habanero'
C. frutescens 'BG2814-6'
L. dejecta
C. frutescens PI594141-P
C. annuum 'R Naky'
C. frutescens 'Tabasco'
C. frutescens PI594141-np (pun1³)
C. chinense 'NMCA30036' (pun1²)
C. chacoense PI260433-np (pun2)
C. chacoense PI260433-P
C. annuum 'Maor' (pun1)
C. annuum 'Sweet 3575' (pun1)
C. annuum 'Jupiter' (pun1)
C. annuum 'Thai Hot'
D. stramonium
P. parodii
S. lycopersicum 'Ailsa Craig'
S. pennellii
N. benthamiana
S. melongena 'Ichiban'
67
89
75
53
80
100
98
99
91
100
100
90
100

Alternatively, the *pun2* mutation could be part of another allelic series transferred to domesticated peppers. Loss of pungency in *C. chacoense* PI260433-np may be related to the *caps7.1* QTL as evidenced by the preliminary map data presented in this study (Ben-Chaim *et al*, 2006). *C. chacoense* has been used extensively in the breeding of modern pepper cultivars as a source of genetic resistance to common pathogens. The *Bs2* gene on the lower arm of pepper chromosome 9 that confers resistance to bacterial spot caused by *Xanthomonas euvesicatoria* has been introgressed into *C. annuum* from *C. chacoense* PI 260435 (Tai *et al*, 1999). The *L4* gene from *C. chacoense* PI260429 has been similarly used for resistance to several tobamoviruses including *Tobacco mosaic virus* and *Pepper mild mottle virus*. *L4* is thought to be a member of the NB-LRR class of plant resistance genes and occur in a cluster of paralogs on the bottom of pepper chromosome 11(Tomita *et al*, 2008).

The *caps7.1* QTL could therefore be an ortholog of *pun2* with different expressivity or the result of a historical introgression of the *pun2* allele that has a qualitative rather than quantitative behavior in a 'Habanero' background. Furthermore, the results of our experiments suggest that the *Pun2* locus is epistatic to the *Pun1*. The trace levels of capsaicinoids detectable in the *pun2* mutant would be consistent with this hypothesis of being part of an allele series with different strengths of expression depending on the genetic background. The identification of the genetic basis for loss of pungency in *C. chacoense* PI260433-np and the *caps7.1* QTL should resolve this relationship.

This study shows that there is a difference in the genetic basis for loss of pungency in a "wild" *Capsicum* species compared to that seen in non-pungent accessions from cultivated species. The resources described here and in previous

studies will allow the trends we observed regarding the association between species and loss of pungency mutations to be tested in other populations of un-cultivated peppers that are known to be polymorphic for pungency, such as *C. eximium* whose range overlaps with that of *C. chacoense* PI260433-np (Stewart *et al*, 2007; Tewksbury *et al*, 2006; Stewart *et al*, 2005). It remains intriguing to speculate as to why the *Pun1* locus seems to be favored in generating null capsaicinoid biosynthetic mutants in cultivated species. Similar allelic series that affect anthocyanin biosynthesis have been found to be naturally occurring in grape (This *et al*, 2007) and have been artificially generated by transposon mutagenesis to affect carotenoid accumulation in maize (Singh *et al*, 2003).

The history of artificial selection by humans in *C. annuum*, *C. chinense* and *C. frutescens* should not be discounted as a factor influencing *Pun1* either. Selection for fruit or seed characteristics could have affected the *Pun1* locus by virtue of their genetic linkage on pepper chromosome 2 (Ben Chaim *et al*, 2001; Ben-Chaim *et al*, 2006). The capsaicinoid biosynthetic pathway may also be metabolically linked to a trait that is under strong selective pressure for an agronomically desirable trait. There is evidence suggesting that natural selection on secondary metabolites has a heritable effect on plant biomass accumulation, highlighting that a tradeoff between carbon allocation to plant secondary metabolites and biomass exists in natural populations (Han and Lincoln, 1994). One could imagine a similar tradeoff existing under artificial selection for agronomically important traits at the expense of secondary metabolite accumulation. Interestingly, it has been found that the fruits of both pungent and non-pungent *C. chacoense* plants have a higher proportion of lipids which suggests avian frugivory, than cultivars of *C. annuum* (Levey *et al*, 2006); one moiety of the capsaicinoid molecule is a fatty acid. It is possible that the relationship between the capsaicinoid biosynthetic pathway and other traits of agronomic importance are

mediated either through genes with pleiotropic function or by selection on related traits (Purugganan and Fuller, 2009).

CONCLUSIONS

The *Pun1* locus has a qualitative effect on capsaicinoid biosynthesis in cultivated varieties belonging to the species *C. annuum*, *C. chinense* and *C. frutescens*, which are thought to form a closely related species complex. In the course of this study we have shown that loss of pungency in the accession *C. frutescens* PI594141 non-pungent is allelic to the mutations *pun1* and $pun1^2$, and thus designated $pun1^3$, whereas the mutation in *C. chacoense* PI260433 non-pungent is not allelic to the other known *Pun1* mutations, and thus represents the identification of *pun2*. The loss of function mutation in $pun1^3$ is a deletion of the terminal region of the second exon, including the canonical stop codon and the DFGWG (or DFGWGKP) site, characteristic of the BAHD enzyme family to which *Pun1* belongs. $pun1^3$ perfectly co-segregates with loss of pungency and furthermore is associated with a loss of transcription of the *Pun1* locus. A phylogeny of the null alleles of *Pun1* suggests that each mutation has an independent origin predating the diversification of the species complex and that some alleles may have been maintained by lineage sorting.

ACKNOWLEDGMENTS

We would like to thank the members of the Jahn and Citovsky labs for access to prepublication data and useful discussions, especially Shanna Moore Fellman, Alex Krichevsky, Adi Zaltsman, and Lisa Zalepa. Maryann Fink, Brynda Beeman, Matt Falise, George Moriarty, John Klum and Mike Axelrod provided assistance with plant

populations. We thank Ilan Paran for help interpreting the results of the mapping study, and Feinan Wu for access to the COSII marker set. Eli Borrego, Moira Sheehan and Judy Kolkman shared expertise with I-PCR. Joshua Tewksbury generously provided the *C. chacoense* material and the *C. frutescens* seed was obtained from the USDA/ARS *Capsicum* collection, courtesy of Robert Jarret, Griffin, GA. This work was supported by an NSF0417056 to M.J. and M.M. and an NSF GRF to G.M.S. as well as additional support from the Dean's fund at SUNY Stony Brook.

WORKS CITED

Aluru MR, Mazourek M, Landry LG, Curry J, Jahn M, O'Connell MA (2003). Differential expression of fatty acid synthase genes, *Acl*, *Fat* and *Kas*, in *Capsicum* fruit. *Journal of Experimental Botany* **54**(388)**:** 1655-1664.

Andrews J (1984). *Peppers: The Domesticated Capsicums*. University of Texas Press.

Ben Chaim A, Paran I, Grube RC, Jahn M, van Wijk R, Peleman J (2001). QTL mapping of fruit-related traits in pepper (Capsicum annuum). *Theoretical and Applied Genetics* **102**(6-7)**:** 1016-1028.

Ben-Chaim A, Borovsky Y, Falise M, Mazourek M, Kang BC, Paran I *et al* (2006). QTL analysis for capsaicinoid content in *Capsicum*. *Theoretical and Applied Genetics* **113**(8)**:** 1481-1490.

Bennett DJ, Kirby GW (1968). Constitution and Biosynthesis of Capsaicin. *Journal of the Chemical Society C-Organic* **4:** 442-446.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2009). GenBank. *Nucleic Acids Research* **37:** D26-D31.

Blum E, Mazourek M, O'Connell M, Curry J, Thorup T, Liu KD *et al* (2003). Molecular mapping of capsaicinoid biosynthesis genes and quantitative trait loci analysis for capsaicinoid content in *Capsicum*. *Theoretical and Applied Genetics* **108**(1)**:** 79-86.

Blum-Eyal, Liu K, Mazourek M, Yoo Eun Y, Jahn M, Paran I (2002). Molecular mapping of the *C* locus for presence of pungency in Capsicum. *Genome* **45**(4)**:** 702-705.

Bradshaw HD, Wilbert SM, Otto KG, Schemske DW (1995). Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* **376**(6543)**:** 762-765.

Curry J, Aluru M, Mendoza M, Nevarez J, Melendrez M, O CMA (1999). Transcripts for possible capsaicinoid biosynthetic genes are differentially accumulated in pungent and non-pungent *Capsicum spp*. *Plant Science* **148**(1)**:** 47-57.

Deshpande RB (1934). Studies on Indian Chillies (4) Inheritance of Pungency in *Capsicum annuum* L. *Indian Journal of Agricultural Science* **5**(IV)**:** 513-516.

Doyle JJ, Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus* **12:** 13-15.

Eshbaugh WH (1980). The taxonomy of the genus *Capsicum* (Solanaceae). *New Phytologist* **47**(2)**:** 153-166.

Felsenstein J (1981). Evolutionary trees from DNA sequences - a Maximum-Likelihood approach. *Journal of Molecular Evolution* **17**(6)**:** 368-376.

Garces-Claver A, Fellman SM, Gil-Ortega R, Jahn M, Arnedo-Andres MS (2007). Identification, validation and survey of a single nucleotide polymorphism (SNP) associated with pungency in *Capsicum* spp. *Theoretical and Applied Genetics* **115**(7)**:** 907-916.

Govindarajan VS (1985). *Capsicum*: Production, Technology, Chemistry and Quality. 1. History, Botany, Cultivation and Primary Processing. *Crc Critical Reviews in Food Science and Nutrition* **22**(2)**:** 109-176.

Hadacek F (2002). Secondary metabolites as plant traits: Current assessment and future perspective. *Critical Reviews in Plant Sciences* **21**(4)**:** 273-322.

Han KP, Lincoln DE (1994). The evolution of carbon allocation to plant secondary metabolites - a genetic analysis of cost in *Diplacus aurantiacus*. *Evolution* **48**(5)**:** 1550-1563.

Leete E, Louden MCL (1968). Biosynthesis of capsaicin and dihydrocapsaicin in *Capsicum frutescens*. *Journal of the American Chemical Society* **90**(24)**:** 6837-6841.

Levey DJ, Tewksbury JJ, Cipollini ML, Carlo TA (2006). A field test of the directed deterrence hypothesis in two species of wild chili. *Oecologia* **150**(1)**:** 61-68.

Michelmore RW, Paran I, Kesseli RV (1991). Identification of markers linked to disease resistance genes by bulked segregant analysis - a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America* **88**(21)**:** 9828-9832.

Morgenstern B (2004). DIALIGN: Multiple DNA and Protein Sequence Alignment at BibiServ. *Nucleic Acids Research* **32:** W33-W36.

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J *et al* (2005). The SOL Genomics Network. A comparative resource for Solanaceae biology and beyond. *Plant Physiology* **138**(3)**:** 1310-1317.

Nelson EK. (1919a). *Journal of the American Chemical Society, Vol. 41*, pp 1115-1121.

Nelson EK (1919b). Vanillylacyl amides. *Journal of the American Chemical Society* **41:** 2121-2130.

Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia VF, Collier SM (2008). A molecular phylogeny of the Solanaceae. *Taxon* **57**(4)**:** 1159-1181.

Onus AN, Pickersgill B (2004). Unilateral incompatibility in *Capsicum* (Solanaceae): Occurrence and taxonomic distribution. *Annals of Botany* **94**(2)**:** 289-295.

Pham HS, Kiuchi A, Tabuchi K (1999). Methods for rapid cloning and detection for sequencing of cloned inverse PCR-generated DNA fragments adjacent to known sequences in bacterial chromosomes. *Microbiology and Immunology* **43**(9)**:** 829-836.

Pickersgill B (1971). Relationships between Weedy and Cultivated Forms in Some Species of Chili Peppers (Genus *Capsicum*). *Evolution* **25**(4)**:** 683-691.

Posada D, Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**(9)**:** 817-818.

Purugganan MD, Fuller DQ (2009). The nature of selection during plant domestication. *Nature* **457**(7231)**:** 843-848.

Ronquist F, Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**(12)**:** 1572-1574.

Bradshaw HD, Schemske DW (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* **426**(6963)**:** 176-178.

Singh M, Lewis PE, Hardeman K, Bai L, Rose JKC, Mazourek M *et al* (2003). Activator mutagenesis of the *pink scutellum1*/*viviparous7* locus of maize. *Plant Cell* **15**(4)**:** 874-884.

St-Pierre B, De Luca V (2000). Evolution of acyltransferase genes: origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. *Recent Advances in Phytochemistry* (Evolution of Metabolic Pathways) **34:** 285-315.

Stewart C, Kang BC, Liu K, Mazourek M, Moore SL, Yoo EY *et al* (2005). The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. *Plant Journal* **42**(5)**:** 675-688.

Stewart C, Mazourek M, Stellari GM, O'Connell M, Jahn M (2007). Genetic control of pungency in *C. chinense* via the *Pun1* locus. *Journal of Experimental Botany* **58**(5)**:** 979-991.

Suzuki T, Fujiwake H, Iwai K (1980). Intracellular localization of capsaicin and its analogs capsaicinoid in *Capsicum* fruit 1. microscopic investigation of the structure of the placenta of *Capsicum annuum* var. *annuum* cultivar Karayatsubusa. *Plant and Cell Physiology* **21**(5)**:** 839-853.

Swofford DL. (2002). Sinauer Associates: Sunderland, MA.

Tai T, Dahlbeck D, Stall RE, Peleman J, Staskawicz BJ (1999). High-resolution genetic and physical mapping of the region containing the *Bs2* resistance gene of pepper. *Theoretical and Applied Genetics* **99**(7-8)**:** 1201-1206.

Tavare S (1986). Some probabilistic and statistical problems on the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* **17:** 57-86.

Tewksbury JJ, Manchego C, Haak DC, Levey DJ (2006). Where did the chili get its spice? Biogeography of capsaicinoid production in ancestral wild chili species. *Journal of Chemical Ecology* **32**(3)**:** 547-564.

This P, Lacombe T, Cadle-Davidson M, Owens CL (2007). Wine grape (*Vitis vinifera* L.) color associates with allelic variation in the domestication gene *VvmybA1*. *Theoretical and Applied Genetics* **114**(4)**:** 723-730.

Tomita R, Murai J, Miura Y, Ishihara H, Liu S, Kubotera Y *et al* (2008). Fine mapping and DNA fiber FISH analysis locates the tobamovirus resistance gene L-3 of Capsicum chinense in a 400-kb region of R-like genes cluster embedded in highly repetitive sequences. *Theoretical and Applied Genetics* **117**(7)**:** 1107-1118.

USDA A, National Genetic Resources Program. National Germplasm Resources Laboratory, Beltsville, Maryland.

Votava E, J, Bosland Paul W (2002). Novel sources of non-pungency in *Capsicum* species. *Capsicum and Eggplant Newsletter* **21:** 66-68.

Walsh BM, Hoot SB (2001). Phylogenetic relationships of *Capsicum* (Solanaceae) using DNA sequences from two noncoding regions: the chloroplast *atpB-rbcL* spacer region and nuclear *waxy* introns. *International Journal of Plant Sciences* **162**(6)**:** 1409-1418.

Webber H (1911). *American Breeders Association Annual* **V7:** 188.

Wu F, Eannetta, N., Xu, Y., Durrett, R., Mazourek, M., Jahn, M., Tanksley, S. (2009). A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. *Theoretical and Applied Genetics* **In press.**

Zewdie Y, Bosland P, W. (2000). Pungency of Chile (*Capsicum annuum* L.) fruit is affected by node position. *Hortscience* **35**(6)**:** 1174.

CHAPTER 3: IDENTIFICATION AND CHARACTERIZATION OF A CONSERVED TANDEM GENE DUPLICATION IMPLICATED IN CAPSAICINOID BIOSYNTHESIS

(Giulia Marina Stellari[4], Shanna Moore Fellman[5], and Molly Jahn[6†])

ABSTRACT

Capsaicinoids are synthesized exclusively in the genus *Capsicum* and are responsible for the burning sensation experienced when consuming hot pepper fruits. *AT3*, which based on homology belongs to the BAHD family of acyltransferases, has been implicated in capsaicinoid biosynthesis. The recessive forms of this gene lead to the absence of capsaicinoids and coordinated transcriptional down-regulation of phenylpropanoid and fatty acid structural genes. Sequence information from representative taxa in the Solanaceae revealed that the coding region of *AT3* is highly conserved throughout the family and also uncovered a tandem duplication of this gene, designated *AT3-1* and *AT3-2* respectively. This duplication predates the diversification of the Solanaceae. Alignment of the *AT3-1* and *AT3-2* sequences shows that the *AT3-2* locus, now a pseudogene, retains regions of amino acid conservation relative to *AT3-1*. Both Bayesian and maximum parsimony methods demonstrated that the paralogous

[4] Present address: Biochemistry Department, 424 Life Sciences Building, SUNY Stony Brook, Stony Brook NY, 11794.
[5] Present address: Plant Breeding Department, Bradfield Hall Room 316, Cornell University, Ithaca NY, 14853.
[6] Present address: Department of Agronomy and Genetics, 140 Agriculture Hall, 145 Linden Drive, Madison, WI 53706.
[†] Author contributions: SMF identified germplasm, performed northern blot, designed sequencing and RT-PCR primers and amplified *AT3-1* and *AT3-2* sequences; GMS designed AT3-2 genome walking primers, RT-PCR primers, amplified *AT3-1* and *AT3-2* from *C. chacoense* PI260433 and from *C. frutescens* PI594141 and performed phylogenetic analysis and recombination detection; GMS and SMF wrote manuscript; all authors read and reviewed the manuscript before submission.

gene lineages of *AT3-1* and *AT3-2* form well supported phylogenetic clades. We find support for the species complex comprising *C. annuum*, *C. chinense* and *C. frutescens*, widely reported in the literature. We also find that this species complex includes *Lycianthes*. Additionally, in *C. rhomboideum*, we identify a recombination event between *AT3-1* and *AT3-2*. We demonstrate *C. rhomboideum* to be non-pungent, and show that a frame-shift mutation in the second exon eliminates the critical DFGWG site and may account for the absence of capsaicin biosynthesis. Our data suggest that duplication of the original *AT3* representative, in combination with divergence, and pseudogene degeneration may account for the pattern of sequence divergence and punctuated amino acid conservation observed.

INTRODUCTION

The diversity of plant secondary metabolism is astounding, with broad implications for humankind. It is estimated that over 500,000 secondary metabolites, or 'natural products' exist in plants, with just 20% having been identified (Hadacek, 2002). Production and accumulation of specific secondary compounds is often highly regulated and may require controls at both the cellular and developmental levels. Elucidation of the biosynthesis of many compounds is further complicated by the contribution of multiple pathways to a single end product. One of the most intriguing characteristics of natural products is the often seen taxonomic isolation, with many compounds produced only by a single family or genus (e.g. vanillin, opium, digitalin, taxol) (Hadacek, 2002). An example of this taxonomic isolation is capsaicinoid biosynthesis, the alkaloids that confer pungency, or 'heat', to pepper pods or fruit. Capsaicinoids are produced exclusively by species of the genus *Capsicum*, which belongs to the family Solanaceae, and the best-known cultivated varieties are *C.*

annuum (e.g. bell types and hot peppers such as jalapeno and cayenne), *C. frutescens* (Tabasco) and *C. chinense* ('Habanero'). Pungency is one of the most important characteristics of *Capsicum* fruit. Capsaicinoids presumably evolved to deter mammalian herbivory (Jordt and Julius, 2002; Tewksbury and Nabhan, 2001), but have also been utilized historically as an analgesic, and more recently to treat conditions such as neurological disorders, bladder and digestive syndromes, and cancer (Han et al, 2001).

Recent efforts to further clarify the biosynthesis of capsaicinoids resulted in the cloning and characterization of the *Pun1* locus. *Pun1* (formerly known as *C*) was first documented in the early 1900s and in its recessive form is epistatic to all other pungency related loci (Stewart et al, 2005; Webber, 1911). Stewart et. al. (2005) demonstrated that in *C. annuum*, *pun1* represented a large 2.5 kb deletion in the 5' end of the gene that resulted in the absence of any detectable transcript or protein and this state correlated with the absence of capsaicinoids and the ultimate loss of pungency. *Pun1* has significant homology to the BAHD family of acyltransferases and thus is also known as *AT3* (Stewart et al, 2005). An additional allele of *Pun1*/*AT3* was subsequently reported in a cultivar of *C. chinense*, also resulting in the loss of pungency in this variety (Stewart et al, 2007). *Pun1*/*AT3* is currently the only locus known to have a qualitative effect on pungency. The substrate, specificity, and enzymatic function of AT3 in the capsaicinoid biosynthetic pathway is still unknown, however all evidence acquired to date suggests that *AT3* is essential for capsaicinoid biosynthesis.

As mentioned, capsaicinoids are the result of a unique biochemical capacity in peppers. This type of taxonomic isolation is often seen with interesting and valuable secondary metabolites and highlights the quandary of how these novel biosynthetic capacities evolve. It has been suggested that novel metabolites emerge as a result of

the activation of latent biosynthetic capacity, which is mostly maintained in a background state of inactivity (Lewinsohn and Gijzen, 2009). *AT3*'s pivotal role in capsaicinoid biosynthesis led us to investigate its evolution within the genus *Capsicum* as well as in other genera in the Solanaceae. We hypothesized that the phylogenetic relationship between *AT3* and its known orthologs might elucidate our understanding of the taxonomic restriction of capsaicin biosynthesis and *AT3's* potential role in the regulation of the capsaicinoid biosynthetic pathway. To this end, we sequenced *AT3* from multiple pepper species and cultivars, with additional representative species from a diverse panel of solanaceous genera. In our survey of *Pun1*/*AT3*, we identified a tandem duplication event.

Tandem gene duplication events are a well-known phenomenon characterizing genome evolution in eukaryotes (Taylor and Raes, 2004). Studies of duplicated genes have shown that diverse outcomes are possible following the duplication event. Ohno posited that duplicated genes constituted the raw material for evolution of novel traits as relaxed selection on one of the copies could lead to the evolution of novel functions, or neo-functionalization (Ohno, 1970). In the three decades since that hypothesis has been put forth, it has been discovered that the products of gene duplication events can undergo many different outcomes: neofunctionalization, subfunctionalization, and also sub-neo functionalization, where initial preservation of gene function by partitioning of expression leads to a subsequent acquisition of novel function (He and Zhang, 2005; Lynch and Force, 2000).

We present here the cloning and initial characterization of this *Pun1*/*AT3* paralog, which we have designated *AT3-2*. We demonstrate that these tandemly duplicated genes represent ancient paralogous gene lineages whose duplication predates the diversification of the Solanaceae. Characterization of a recombination event between *AT3-1* and *AT3-2* in *C. rhomboideum* reveals a putative basis for non-

pungency in what is often seen as the basal species of the *Capsicum* genus. This new information contributes to our understanding of the evolution of the genus *Capsicum* as well as to the origin of pungency. It adds to the knowledge and resources available in *Capsicum* and may facilitate efforts to manipulate the capsaicinoid pathway to optimize capsaicinoid production for specific and perhaps novel uses.

MATERIALS AND METHODS

Plant Growth and Tissue collection

All accessions from the species *Capsicum annuum*, *Capsicum frutescens*, *Capsicum chinense*, *Capsicum chacoense*, *Capsicum rhomboideum*, *Lycianthes dejecta*, *Solanum lycopersicum*, *Solanum pennellii*, *Solanum melongena*, *Nicotiana benthamiana*, *Datura stramonium*, *Petunia parodii* (see Table 3.1), were grown from seed in the greenhouse. Growing conditions were approximately 27°/18°C (day/night) and a daily fertilization with Excel solution (200ppm; The Scotts Company, Marysville, Ohio).

Green (21 days post anthesis) and ripe fruit (50 days post anthesis) were collected for all pepper varieties. Mature green (35 days post anthesis) and ripe fruit (breaker plus 10 days) were collected for *S. lycopersicum* and age-matched fruit were collected for *S. pennellii* as well as immature and mature fruit for *S. melongena*. Upon harvest, all fruits were acclimated for 4-8 h in the laboratory. Seeds were harvested, placental tissue was removed and pericarp tissue was immediately frozen in liquid nitrogen and stored at –80°C.

RNA extraction

Tissue was collected and frozen immediately in liquid nitrogen, ground to a powder, and stored at -80°C. RNA was extracted from all tissue types using Qiagen RNeasy (Qiagen, Valencia CA) according to the manufacturers instructions for isolation of plant and fungal RNA. For each extraction, approximately 50mg of frozen, ground tissue was utilized and the final elution volume was 35 μl of RNase free water. RNA was denatured and visualized on an agarose gel to assess quality.

DNA Isolation

Expanding leaf tissue was either used immediately for DNA extraction or frozen in liquid nitrogen and stored at -80°C for later extraction. Genomic DNA was extracted as described previously (Tanksley et al, 1992).

Isolation of AT3-2 and sequence analysis

Ten Bacterial Artificial Chromosomes (BACs) from a *C. frutescens* 2814 pBeloBACII DH10B library were identified previously as screening positive for *AT3-1* (K. Liu and M Jahn, unpublished results). These BACs were further screened for the presence of the additional AT3-2 intron fragment. Of those containing the additional AT3-2 intron fragment one, 159M5 was selected at random, and a large-scale DNA prep was done using Qiagen Midi Prep Kit (Qiagen, Valencia, CA) following the manufacturer's instructions for very low copy plasmids.

AT3-2 unique primers were designed from the intron sequence (*AT3-2* INTRON F 5' – AAGTAAACTGAATTTGTTTCAAAA-3'; *AT3-2* INTRON R 5'

ATTTACCCTACATTATTATCGGTC – 3', 364bp product) and these primers were used in conjunction with Bio S&T APAgene Gold Genome Walking Kit (BioS&T, Montreal, Canada) to isolate a full-length genomic sequence according to the manufactures instructions. Two separate sets of reactions were undertaken using either *AT3-2* INTRON F or *AT3-2* INTRON R as the GSPa primer.

Products were excised and gel purified using the Qiagen QIAquick Kit (Qiagen, Valencia, CA) according to the manufacturers instructions. The blunt end products were prepared for TA cloning as follows: 19.55µl of gel purified PCR product, 1.5µl 25mM $MgCl_2$, 1.25µl 10mM dNTPs, 2.5µl 10x Buffer, 0.2µl Roche Taq (Roche, Indianapolis, IN). This was incubated at 72°C for 20 minutes and then utilized in a ligation reaction using the Invitrogen TOPO TA Cloning Kit (Invitrogen Corporation, Carlsbad, CA) for sequencing with the PCR4 vector and TOPO10 electrocompetant cells. Ligations were performed as follows: 2µl of gel purified PCR product, 3µl water, 0.5µl dilute salt (1:4 dilution of Salt Solution provided with kit), 0.5µl TOPO vector. Incubated at RT for 15 minutes and stored at -20°C. 2µl of ligation product was combined with TOPO10 cells, electroporated, combined with 500µl 2xLB, incubated at 37°C for 30 minutes and 300µl was plated on LB/agar/kanamyacin plates and incubated at 37°C overnight. Resulting colonies/plates were stored at 4°C.

Multiple clones from each ligation were selected into 3mL of LB/kanamyacin and incubated with 200rpm shaking at 37°C overnight. Plasmids were purified for sequencing using the QIAGEN QIAprep Kit (Qiagen, Valencia, CA) according to the manufacturers instructions. Sequencing was conducted by the Bioresources Center, Cornell University (www.brc.cornell.edu). Reactions were submitted to the sequencing facility as follows: 3µl of resulting miniprep DNA, 1µl 10mM M13F or R, 14µl of water.

Cloned sequences were trimmed of vector and aligned with other sequences from the same ligation reaction using Seqman II 6.1 of DNASTAR suite of analysis software to generate consensus sequences. Overlapping sequences from the 5' and 3' end of AT3-2 were assembled and aligned with existing AT3-1 sequence for comparison.

Subsequent *AT3-2* sequence was amplified utilizing two separate primer combinations to ensure amplification of *AT3-2* and not *AT3-1*: *AT3-2* F (5' – ATGGCTTTTGCATTGGTATCATCACCAT -3') / *AT3-2* INTRON F resulting in a 1.1 kb product and AT3-2 INTRON R/ *AT3-2* R (5'-CGGTATACTCATTCTTACAGGTTT-3') resulting in a 860bp product. All amplifications were done in a 15µl total volume (8.65µl water, 1.5µl 10x buffer, 1.0µl 10mM dNTPs, 0.3µl 10mM Primer F, 0.3µl 10mM Primer R, 0.5µl Stratagene Easy A Taq) and under the following conditions: 95°C for 5 minutes, 35 cycles of 95°C for 30 seconds, 65°C for 30 seconds 72°C for 1 minute, 72°C for 10 minutes, 4°C. Reactions were conducted using a PTC 225 Peltier Thermal Cycler (MJ Research, Watertown, MA).

Initial Alignment

An initial alignment of 28 *AT3-1* and *AT3-2* cloned sequences, which included those from *Capsicum spp*. as well as from representative taxa in the Solanaceae and the outgroup sequence from *C. annuum*, *catf2*, was compiled using the program DIALIGN (http://bibiserv.techfak.uni-bielefeld.de/dialign/) and refined by hand (Morgenstern, 2004). Despite the absence of full length sequence, *C. annuum catf2* was chosen as the outgroup for the dataset due to its high degree of nucleotide sequence similarity with *AT3-1* and *AT3-2*. *Catf2* expression does not co-segregate

with pungency and does not map to the same genomic region as *AT3-1* and *AT3-2* thereby excluding it as related to the tandem duplication event that created *AT3-1* and *AT3-2* (Garces-Claver et al, 2007). This first alignment of 28 sequences and 1815 characters contained the 5' UTR of *AT3-1* and the coding regions of *AT3-1* and *AT3-2* from the putative start codon to approximately 100 base pairs upstream of the stop codon. The introns of *AT3-1* and *AT3-2* were initially excluded from the alignment, as they are non-homologous and caused a breakdown of the alignment of the flanking sequence when included. As of this study the authors were unable to find any reports documenting non-homologous introns flanked by homologous exons. The introns of *AT3-1* and *AT3-2* were defined using the intron-exon boundary known from cloning the mRNA of *AT3-1* from *C. chinense* ''Habanero'' (Stewart et al, 2005). After refining the DIALIGN generated alignment by hand, taking into account both the nucleotide and amino acid alignments, the intron sequences from both paralogs as well as from the outgroup sequence, *catf2*, were inserted manually and aligned by eye. All subsequent alignments generated were derived from this initial alignment. This alignment was used to detect recombination in *C. rhomboideum AT3-1* as well as for the maximum parsimony analysis once non-parsimonious characters had been excluded. The initial alignment, excluding sequence information for the 5' UTR, was used for the Bayesian analysis. Removal of the introns and any intervening stop codons generated an alignment that was used to identify codons subject to positive selection in the *AT3-1* and *AT3-2* dataset.

Maximum Parsimony

Maximum parsimony (MP) analysis of the full length *AT3-1* and *AT3-2* alignment was performed using PAUP v. 4.0b10 (Swofford, 2002). MP trees were

generated by heuristic searches with 1,000 random step-wise additions, with tree bisection-reconnection (TBR) branch swapping and the option MULTREES. Gaps were encoded as missing data and all characters were given equal weights. Using the same criteria as the initial search, bootstrap support was estimated by performing 1,000 heuristic searches with 10 addition sequence replicates per bootstrap. In order to test the suitability of *catf2* as the outgroup a truncated dataset was created where the first 386 characters that are missing in *catf2* were deleted from all sequences in the alignment. The majority rule consensus maximum parsimony tree generated using this alignment was not significantly different from the one obtained using the full-length alignment (data not shown). The results of the maximum parsimony analysis on this data set suggest that the missing characters in *catf2* are not detrimental to the analysis. When included, *catf2* was used as the outgroup in all subsequent analyses. Similarly, a truncated dataset was created that excluded the intron sequences from both *AT3-1* and *AT3-2*. This alignment was analyzed as described previously. The relationships identified at the internal nodes were the same as those seen with the full length dataset, but less resolution was obtained in the terminal nodes, therefore the full length sequence, including the introns for both *AT3-1* and *AT3-2* was used in the final analysis.

ModelTest

Prior to generating the Bayesian phylogenies, the best model of evolution was determined using the program ModelTest v3.7 (Posada and Crandall, 1998). Three separate best-fit models were estimated: the first for the coding region of *AT3-1* and *AT3-2*, the second for the intron, and the third for the 5'UTR sequences. ModelTest v3.7 could not resolve, within reasonable computing time, the best model for the 5'

UTR therefore this sequence information was excluded from the subsequent Bayesian analysis. ModelTest v3.7 estimated two best-fit models for the data. The coding region best fit the model with two rate classes, Hasegawa, Kishino and Yano (HKY) with a proportion of invariable sites and a gamma distribution of the rate variation among sites (Hasegawa et al, 1985). The intron sequences best fit the model with a single rate class, Jukes-Cantor (JC), and the proportion of invariable sites equal to zero (Jukes, 1969).

Bayesian Analysis

Bayesian analysis of the *AT3-1*/*AT3-2* alignment including the introns was performed in MrBayes v3.1.2 by creating two data partitions in the alignment that allowed one partition to include both the first and the second exons and the second partition to include only the intron (Ronquist and Huelsenbeck, 2003). Each partition was assigned its best-fit model as per ModelTest v3.7 and rates were allowed to vary independently between partitions. *catf2* was set as the outgroup in this analysis as in the maximum parsimony analysis. A total of four chains of the Markov Chain Monte Carlo were run, sampling one tree every 100 generations. The analysis was allowed to proceed until the standard deviation of split frequencies fell below 0.01. A fixed proportion of trees comprising 25% of all trees sampled were discarded as the "burn in" fraction. A majority rule consensus tree of all trees sampled, excluding the burn in faction, was computed using PAUP v. 4.0b10 (Swofford, 2002).

Identification of Recombination and robustness of result

Identification of recombinant sequences in the alignment was performed using the program RDP3 using standard initial program settings (Martin et al, 2005; Padidam et al, 1999; Posada and Crandall, 2001; Smith, 1992). The full-length alignment, containing the coding regions of *AT3-1* and *AT3-2*, both intron sequences, and the 5' UTR of *AT3-1*, was scanned for recombinant sequences using the RDP3 package. The outgroup sequence of *C. annuum catf2* was excluded from this analysis. Within the RDP3 package, recombination scans using the programs GENECONV, MaxChi, Chimaera and LARD were also implemented in order to test the position of the recombination breakpoints identified in the intitial scan (Holmes et al, 1999; Padidam et al, 1999; Posada and Crandall, 2001; Smith, 1992).

To test the robustness of the RDP3 results, separate phylogenies were estimated for each region identified as recombinant in *C. rhomboideum AT3-1* sequence using both MP and Bayesian methods. Two separate alignments were created: the first comprising base pairs 1 to 646 plus basepairs 1268 to 1815 from the initial alignment, the second alignment comprising solely basepairs 646 to 1268. Separate MP phylogenies were estimated for each new alignment as described previously. Bootstrap support was estimated as described previously except that branch swapping was not allowed. Modeltest v3.7 was used to estimate the best model given the data for each new alignment. Both alignments best supported the model HKY with no invariant sites and a gamma distribution of the rate variation among sites. Bayesian phylogenies were subsequently estimated as outlined previously with each dataset comprising a single data partition.

Capsaicinoid detection

Capsaicinoids were measured in mature dry fruit of *C. rhomboideum* using a capsaicinoid detecting ELISA kit available from Beacon Analytical (Beacon Analytical, Portland, ME). Capsaicinoids were extracted and measured as per manufacturers instructions.

RESULTS AND DISCUSSION

Alignment

In the course of this study, ten previously unreported *AT3-1* sequences of approximately 1630 base pairs comprising the 5' UTR, the single intron, and the near full-length coding region, were cloned and sequenced from *Capsicum* (three previously unreported), *Nicotiana* (one previously unreported), *Solanum* (three previously unreported), *Petunia* (one previously unreported), *Datura* (one previously unreported) and *Lycianthes* (one previously unreported) (Table 3.1). All sequences showed a high degree of conservation at both the nucleotide and amino acid levels (Figure 3.1). Without exception, the intron-exon boundary of *AT3-1* was conserved as well as the intronic sequence itself, somewhat of an unusual finding for nuclear gene sequences from the Solanaceae (Martins and Barkman, 2005; Walsh and Hoot, 2001). *N. benthamiana AT3-1* revealed a two base pair deletion resulting in a frame-shift mutation that we predict would produce a truncated *AT3-1* protein (Figure 3.1). Sequences having previously reported mutations were excluded from this analysis (Martins and Barkman, 2005; Stewart et al, 2005; Stewart et al, 2007).

Table 3.1. List of NCBI accessions for new sequences. All genera belong to the family Solanaceae.

	Family	*Species*	Cultivar/Ecotype	NCBI accession number
AT3-1	Solanaceae	*Capsicum annuum*	Thai Hot	AY819029
		Capsicum chinense	'Habanero'	AY819027
		Capsicum annuum	RNaky	FJ755173
		Capsicum frutescens	BG2614-6	AY819026
			Tabasco	FJ755174
			PI594141-P	FJ755175
		Capsicum chacoense	PI260433-np	FJ755176
		Capsicum rhomboideum		FJ755165
		Lycianthes dejecta		FJ755172
		Solanum lycopersicum	Ailsa Craig	FJ755166
		Solanum pennellii		FJ755168
		Solanum melongena	Ichiban	FJ755170
		Datura stramonium		FJ755167
		Petunia parodii		FJ755169

Table 3.1. Continued

		Nicotiana benthamiana		FJ755171
AT3-2		*Capsicum annuum*	RNaky	FJ687524
			Maor	FJ755161
			Thai Hot	FJ687530
		Capsicum frutescens	BG2614-6	FJ687526
			Tabasco	FJ755160
			PI594141-p	FJ687527
		BAC sequence	BG2814-6	FJ755162
		Capsicum chinense	'Habanero'	FJ755163
		Capsicum chacoense	PI260433-np	FJ755164
		Capsicum rhomboideum		FJ687529
		Lycianthes dejecta		FJ687528
		Solanum pennellii		FJ687531
		Nicotiana benthamiana		FJ687525

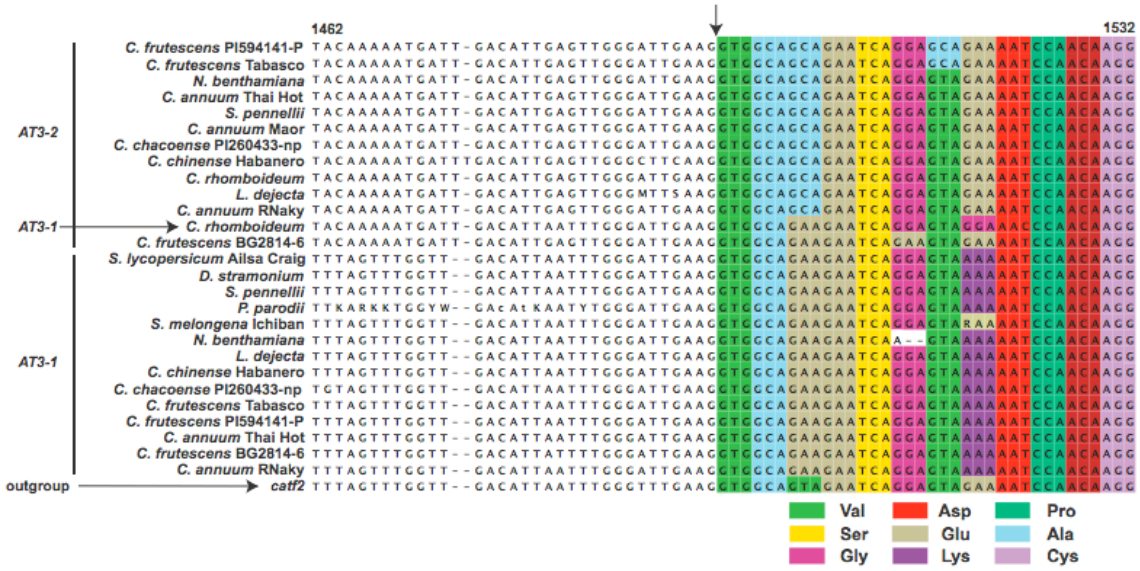


Figure 3.1. Selection of *AT3-1* and *AT3-2* nucleotide alignment. Vertical arrow marks the conservation of the 3' intron/exon boundary. Unshaded nucleotides are intronic, shaded nucleotides are part of the putative coding region. Different shading in boxes indicates putative amino acid conservation between the *AT3-1* and *AT3-2* coding region.

In the course of cloning and characterizing *AT3-1* from representative species in the Solanaceae, a novel intronic sequence was discovered that had no recognizable homology to *AT3-1* and yet shared the same intron-exon boundaries. This novel sequence was identified along with *AT3-1* in multiple clones of a Bacterial Artificial Chromosome (BAC) library made from *C. frutescens* BG2814-6. Genome walking across the intron out to the 5' and 3' ends of the gene revealed DNA sequence showing strong similarity to *AT3-1*. This tandem duplicate, named *AT3-2*, was approximately 1600 base pairs in length and was subsequently cloned and sequenced from *Capsicum* (nine previously unreported), *Nicotiana* (one previously unreported), *Datura* (one previously unreported), *Solanum* (one previously unreported) and *Lycianthes* (one previously unreported).

Our study revealed that *AT3-2* is also highly conserved throughout the Solanaceae. The initial alignment of *AT3-2* with *AT3-1* was compiled using DIALIGN and excluded the highly dissimilar *AT3-1* and *AT3-2* intron sequences (Morgenstern, 2004). Subsequently, the intron sequences were inserted into the alignment and the alignment was refined by hand. Numerous non-synonymous mutations relative to *AT3-1* are shared among all *AT3-2* sequences suggesting the possibility that the coding region of *AT3-2* diverged before the diversification of the solanaceous species represented in our study. In the first exon, the alignment between *AT3-1* and *AT3-2* also includes in/dels not in multiples of three that disturb the open reading frame of *AT3-2* while maintaining punctuated regions of amino acid conservation relative to *AT3-1*. This raises the possibility that *AT3-2* is a pseudogene. A majority of the inactivating mutations in *AT3-2* are shared among all sequenced *AT3-2* alleles suggesting that the pseudogenization of *AT3-2* occurred before the diversification of the Solanaceae. Taken together, these data suggest that the *AT3* progenitor duplicated, diverged and was pseudogenized in the ancestor of the species included in this study.

Expression

AT3 expression in *Capsicum* is confined to the developing pepper placenta, as has been shown previously by RNA gel-blot (Stewart et al, 2005). Based on the extensive sequence homology seen throughout the Solanaceae, we further examined expression in solanaceous species with similar fleshy fruit production. Utilizing a full-length cDNA as a probe, *AT3* expression was detected in 35 dpa fruit in *S. pennellii* but not in age-matched *S. lycopersicum* or immature *S. melongena* (Figure 3.2). Due to the sequence homology between *AT3-1* and *AT3-2*, extensive efforts were made to distinguish distinct gene copy expression through RT-PCR. Despite utilization of multiple restriction enzymes (individually and in combinations), only *AT3-1* expression was detected and we therefore attribute the expression seen by RNA gel-blot to *AT3-1* (RT-PCR data not shown). It is tempting to speculate that outside of *Capsicum*, *AT3-1* expression is confined to the fleshy fruit of non-domesticated solanaceous species however a more thorough sampling of solanaceous fruit is needed to support this assertion. The similarities in temporal and spatial expression of *AT3-1* between *Capsicum* and *S. pennellii* do, however, raise the intriguing possibility that control of *AT3-1* expression is linked to a trait that outside the clade *Capsicum* has been selected against during domestication.

AT3-1 has been identified as a member of the BAHD family of acyltransferases (named for the first four characterized members of the family BEAT, AHCT, HCBT, DAT) (D'Auria et al, 2002). Acylation is a common and biochemically important modification of numerous plant metabolites and members of the BAHD catalyze the transfer of the acyl moiety to a wide range of acceptor molecules and are essential in the biosynthesis of a large array of natural plant compounds such as lignin, phenolics, alkaloids, anthocyanins and volatile esters (D'Auria, 2006).

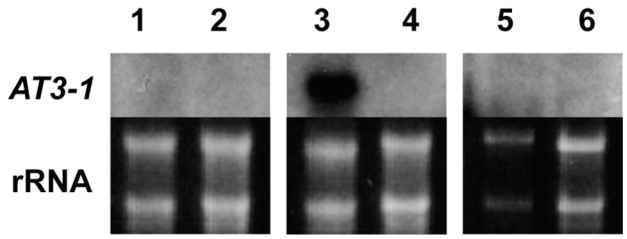


Figure 3.2. RNA gel-blots of a full-length *AT3* cDNA in solanaceous fruit tissue. Tissues utilized are as follows: **1** *S. lycopersicum* Ailsa Craig, fruit 35 dpa; **2** *S. lycopersicum* Ailsa Craig, fruit 48 dpa; **3** *S pennellii*, fruit 35 dpa; **4** *S. pennellii*, fruit 48 dpa; **5** *S. melongena* Ichiban, fruit 21 dpa; **6** *S. melongena* Ichiban, fruit 50 dpa. EtBr stained rRNA is shown as a loading control.

Specifically, AT3-1 lies within clade III of the BAHD family. Many of the representative genes in this clade can accept a diverse range of alcohol substrates, however most utilize acetyl-CoA as the major acyl donor. Members of this clade are involved in modifications of compounds such as vindoline, thebaine, as well as gerinol and other volatile esters important in flowers and ripening fruits (D'Auria, 2006).

Overall, BAHD family acyltransferases utilize a wide range of substrates, however wide or narrow substrate specificities are seen when individual enzymes are examined. Thus the products formed in planta may be determined by the availability of appropriate substrates. This may account for the high level of conservation seen in *AT3-1* across the Solanaceae despite its clear involvement in the biosynthesis of a novel compound found only in the *Capsicum*. *AT3-1* may be involved in the formation of fruit aroma or flavor compounds and have a broad range of substrate specificity in the Solanaceae, however modifications in substrate availability coupled with expression patterns confined to the placenta in the *Capsicum* may have resulted in the formation of a novel metabolite.

The similar expression pattern of *AT3-1* seen in the wild tomato relative *S. pennellii*, but absent in domesticated tomato *S. lycopersicum* Ailsa Craig could point toward its role in the synthesis of an unpalatable aroma or flavor compound that is undesirable to consumers and was selected against in domestication.

Phylogenetic analysis

Maximum parsimony (MP) and Bayesian phylogenies were estimated to test the relationship between *AT3-1* and *AT3-2* using the alignment discussed previously. Five MP trees were identified and the strict consensus of these was chosen to illustrate the results of this analysis (Figure 3.3). All bootstrap support values greater than 50

are included to indicate a measure of support for the relationships depicted in the phylogeny (Figure 3.3). The analysis suggests that *AT3-1* and *AT3-2* are ancient paralogs whose duplication predates the diversification of the Solanaceae. Separate *AT3-1* and *AT3-2* clades receive strong bootstrap support whereas internal relationships are less resolved and received lower levels of statistical support with the exception of the clade containing the major cultivated *Capsicum* varieties plus *Lycianthes* that receives moderate bootstrap support of 77 percent (Figure 3.3). Our analysis supports previous studies finding a species complex that includes the three major cultivated species *Capsicum*, *C. annuum*, *C. chinense* and *C. frutescens* (Walsh and Hoot, 2001). The lone exception is *C. annuum* 'Thai Hot' which does not fall within this clade in our analysis. Interestingly our data support the previously identified relationship between *Capsicum* and *Lycianthes* (Bohs and Olmstead, 1997; Olmstead et al, 2008). In contrast to what has already been reported in the literature, where diverse phylogenies show that *Capsicum* is embedded within *Lycianthes*, our analysis would suggest that Lycianthes is embedded within *Capsicum*. However, given the small sampling of *Lycianthes* species available to us in our study, it seems likely that this result is attributable to incomplete sampling of the genus *Lycianthes*.

A previously unreported sequence of *AT3-1* obtained from a non-pungent cultivar of *C. chacoense* PI260433-np along with *C. annuum* 'Thai Hot' are not grouped with the other *Capsicum* sequences. While the placement of *C. chacoense* PI260433-np receives very weak bootstrap support, *C. annuum* 'Thai Hot' is more strongly supported and associates with the clade containing *Solanum*, *Nicotiana*, *Datura* and *Petunia*, perhaps reflecting lineage sorting in this species.

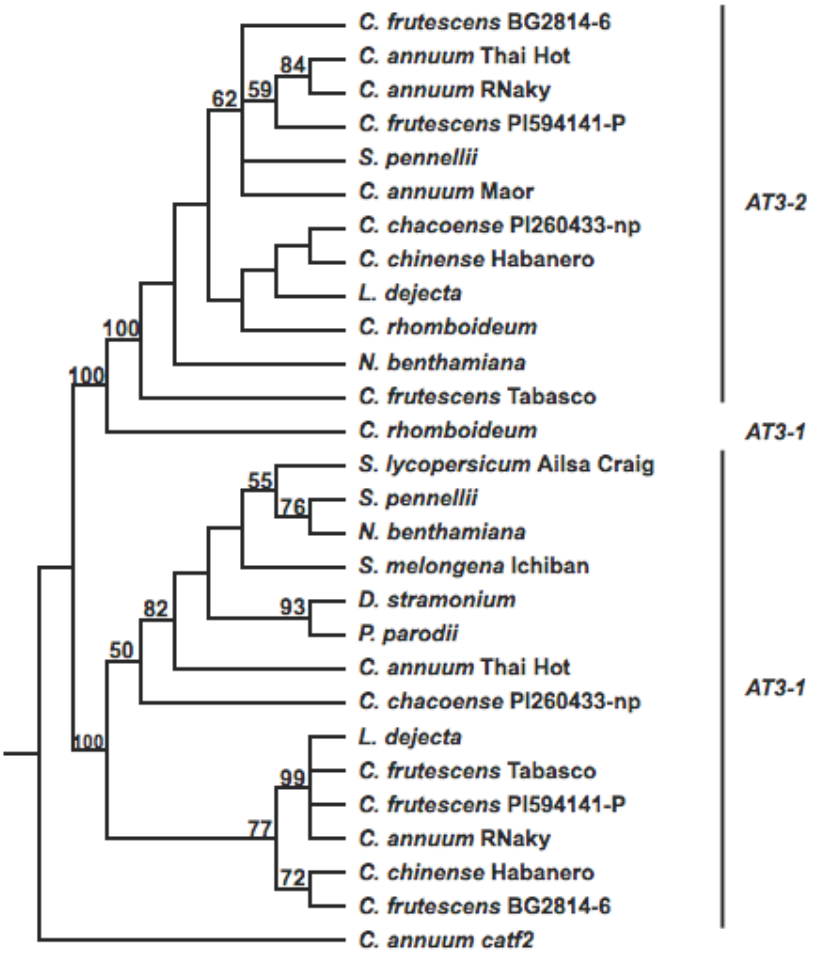


Figure 3.3. Maximum parsimony phylogeny of the *AT3-1* and *AT3-2* alignment. Phylogeny shows the strict consensus of 5 most parsimonious trees. Numbers represent bootstrap proportions; only bootstrap support values >50% are shown.

In our analysis, weak bootstrap values support the *AT3-1* clade, containing *S. lycopersicum*, *S. pennellii* and *S. melongena*, which also includes *N. benthamiana*. *Datura* stramonium and *Petunia* parodii form a subclade distinct from the *Solanum* complex. Although some of the generic level relationships identified in this study conflict with those identified in previous studies (Olmstead et al, 2008), it seems likely that these differences can be attributed to our focus on agronomically important *Capsicum* cultivars and our limited sampling of wild *Capsicums* and other solanaceous species.

Many of the relationships highlighted in the *AT3-1* clade are mirrored in the *AT3-2* clade with the notable exception of *C. annuum* 'Thai Hot'. 'Thai Hot' *AT3-2* grouped with strong bootstrap support with the cultivated *C. annuum* and *C. frutescens* sequences. Interestingly, *Lycianthes* again is placed in a clade containing the *Capsicum* sequences, however this time it falls outside the species complex comprising a majority of the cultivated *C. annuum* and *C. frutescens* varieties. In this case, *Lycianthes* is nested within a clade including *C. chacoense* PI260433-np and the basal species in the genus, *C. rhomboideum*. Surprisingly, *C. chinense* 'Habanero' *AT3-2* groups with *Lycianthes* and the non-cultivated *Capsicums*, a relationship that is supported by, among others, a 34 basepair insertion in the intron (data not shown) and this may represent lineage sorting or the unique breeding history of this cultivar. *AT3-2* was not amplified from *S. melongena* and *S. lycopersicum* suggesting gene loss or complete degeneration of *AT3-2* in these lineages. The internal order of the nodes in the *AT3-2* clade received less support than in the *AT3-1* clade making it difficult to infer deep relationships within the Solanaceae based on this gene sequence alone.

Notably, the *AT3-1* sequence from *C. rhomboideum* is basal to all other *AT3-2* sequences with 100 percent bootstrap support (Figure 3.3). The anomalous placement

of *C. rhomboideum AT3-1* among the *AT3-2* sequences suggested that this sequence might show evidence for a recombination event that confounds the phylogenetic signal and prevents its placement with the other *AT3-1* sequences. As expected *C. rhomboideum AT3-2* falls with the other *Capsicum* sequences embedded within the *AT3-2* clade.

Bayesian analysis of the *AT3-1* and *AT3-2* alignment recapitulates many of the relationships identified in the maximum parsimony analysis (Figure 3.4). Posterior probability values are much higher in the Bayesian analysis than the bootstrap support obtained for the MP phylogenies. Simulation studies have found that the relationship between the posterior probability and non-parametric bootstrap support are generally congruent and differences between the two depend on the underlying assumptions of the methods (Alfaro and Holder, 2006). Bayesian estimation of the *AT3-1* and *AT3-2* phylogeny recovers the deep split between *AT3-1* and *AT3-2* establishing them as paralogous sequences. The MP and Bayesian phylogenies are roughly congruent, with many of the same relationships recovered using both methods. Bayesian estimation of the *AT3-1* and *AT3-2* phylogeny identifies the species complex comprising *C. annuum*, *C. chinense* and *C. frutescens* with a posterior probability value of 94. Again, the clade showing *Lycianthes AT3-1* embedded with the *Capsicum* receives a high measure of support with a posterior probability value of 94.

C. chacoense PI260433-np falls outside the main clade comprising the other *Capsicum* sequences as it did in the MP phylogeny, but this time that relationship receives no support unlike in the MP analysis. In contrast, the relationship associating *C. annuum* 'Thai Hot' with the clade containing *Datura* and *Petunia* receives strong support. The clade containing the *Solanum* and *Nicotiana* sequences is not well supported, however the relationships between *S. pennellii* and *N. benthamiana* as well as the *S. pennellii*, *N. benthamiana* sub-clade which includes *S. lycopersicum* both

have a high posterior probability value of 94.

In general, the relationships depicted in the *AT3-2* clade receive much higher levels of support than those shown in the *AT3-2* clade of the MP phylogeny. Such differences have been document in other studies and have been attributed to the level of congruence between the underlying assumptions and the data used in the analysis (Alfaro and Holder, 2006). The cultivated *Capsicum* species form a distinct group within the *AT3-2* clade, but *S. pennellii* is included with a moderate level of support with these sequences. The non-cultivated *Capsicums* also receive moderate bootstrap support and include *Lycianthes*, as was already seen in the MP phylogeny. *C. frutescens* 'Tabasco' *AT3-2* falls in a clade distinct from the remaining *Capsicum* sequences and in strong association with *N. benthamiana*. The former sequence was basal in the *AT3-2* clade of the MP tree but that relationship received no bootstrap support. It seems likely that incomplete sampling of the genera outside *Capsicum* may be affecting the placement of the *N. benthamiana* and *S. pennellii* sequences in both the MP and Bayesian estimations of the phylogeny.

Recovery of the association of *Lycianthes* with *Capsicum spp*. in both the *AT3-1* and *AT3-2* clades of the MP and Bayesian phylogenies suggests a strong association between *Lycianthes* and *Capsicum*, a relationship that has received increasing support in the literature (Olmstead et al, 2008; Olmstead and Palmer, 1997). Nevertheless, the placement of *Lycianthes* relative to the other *Capsicums* cannot be ultimately determined within this study.

Bayesian estimation of the *AT3-1* and *AT3-2* phylogenies again identifies *C. rhomboideum AT3-1* as basal to all other *Capsicum AT3-2* sequences with posterior probability support of 100. This curious relationship, recovered both in the MP and Bayesian phylogenies was investigated further and tested for the presence of a recombination event in this sequence.

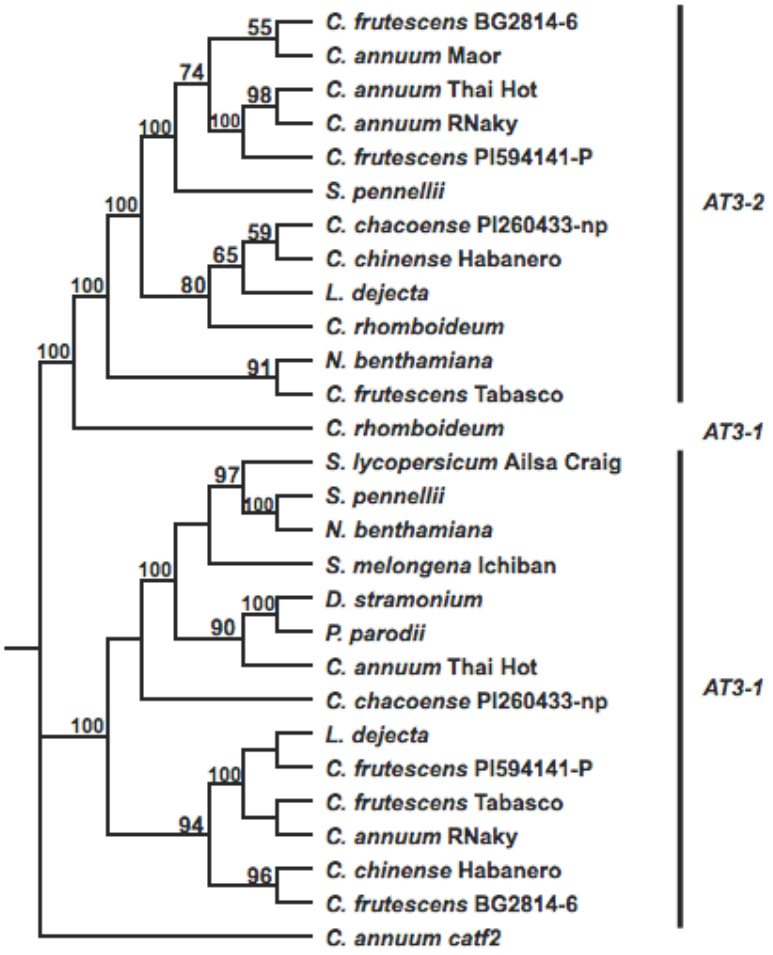


Figure 3.4. Bayesian phylogeny of the partitioned *AT3-1* and *AT3-2* alignment. Phylogeny shows the 50% majority rule consensus of trees sampled after the "burn-in". Numbers next to each node indicate the posterior probability support for branching shown.

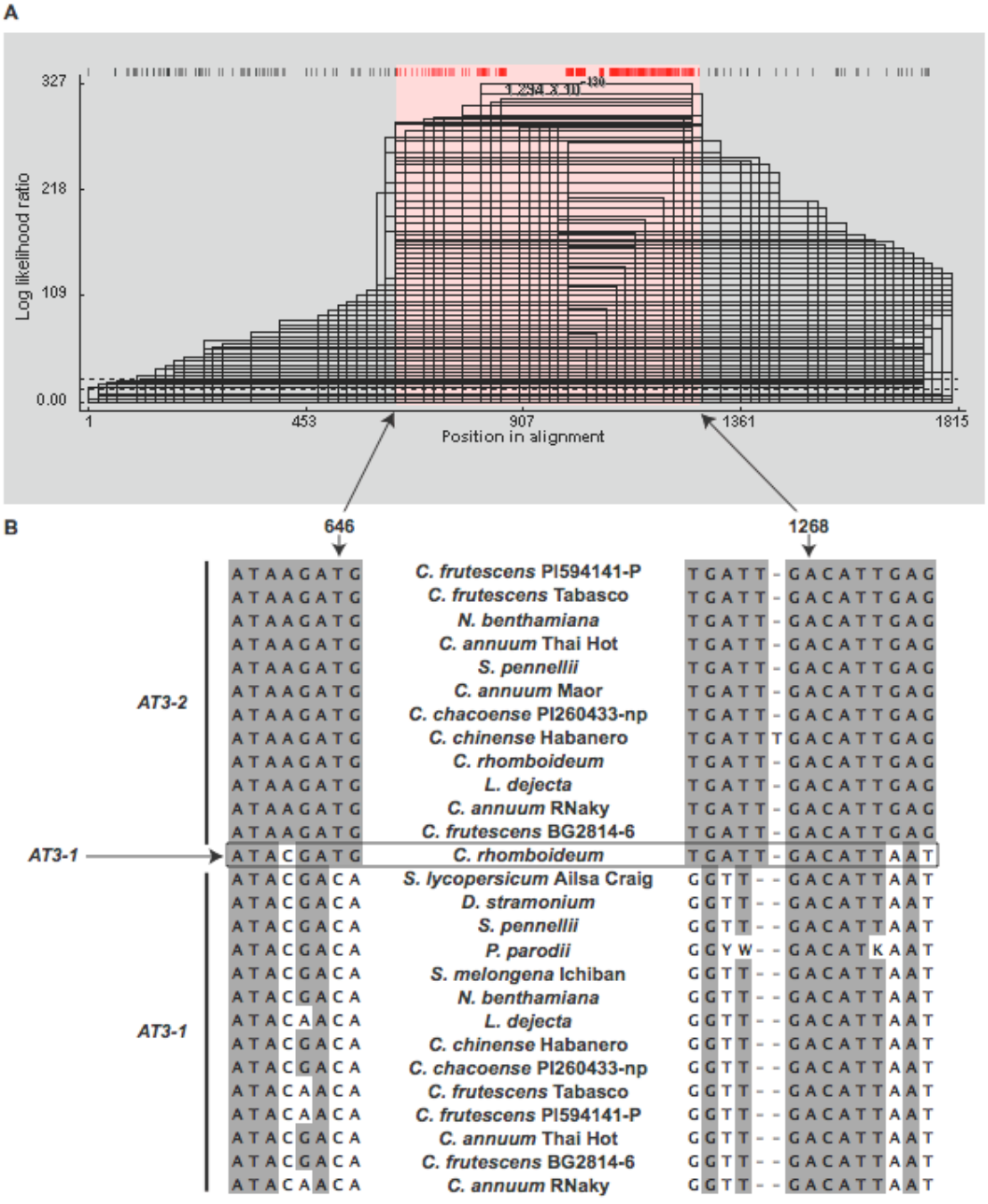
Recombination detection

Given the unexpected association between *C. rhomboideum AT3-1* with the *AT3-2* clade identified in the Bayesian and maximum parsimony analyses, the alignment was tested for the presence of recombination events using the program RDP3 (Martin et al, 2005). The full length *AT3-1* and *AT3-2* alignment was used as in the previous analyses except that the outgroup sequence *C. annuum catf2* was excluded. A scan for recombination in the alignment identified one recombination event in *C. rhomboideum AT3-1* that received strong support from a number of measures (all P-values greater than 10^{-6}) (Table 3.2). Although the parents of the recombination event could not be stated with certainty, the boundaries of the recombination event could be identified and tested further using Likelihood Assisted Recombination Detection (LARD) (Holmes et al, 1999). This program has been successfully used to identify recombination breakpoints in a gorilla mitochondrial sequence dataset (Anthony et al, 2007). Base pair 646 in the alignment was chosen as the left-most recombination breakpoint and base pair 1268 as the right-most recombination breakpoint based on the results of the initial scan for recombination performed with RDP3. Analysis of these two recombination breakpoints with LARD showed high support for a recombination event in *C. rhomboideum AT3-1* at these sites as evidenced by the high plateau corresponding to the light shaded region in Figure 3.5A. A scan of the recombination region reveals that the breakpoints for the recombination event in *C. rhomboideum AT3-1* can roughly be identified by eye as is evident in Figure 3.5B.

Table 3.2. RDP3 support for the recombination event identified in the *AT3-1* and *AT3-2* alignment.

Program	Number of recombination events detected	Average P-value
RDP	1	1.225×10^{-21}
Geneconv	1	9.411×10^{-25}
Bootscan	1	2.326×10^{-26}
MaxChi	1	2.174×10^{-14}
Chimaera	1	8.389×10^{-06}
Siscan	1	1.808×10^{-19}
PhyPro	-	-
LARD	-	-
3Seq	1	8.930×10^{-10}

Figure 3.5. RDP3 scan illustrating the recombination event in *C. rhomboideum AT3-1*. **A** shows the LARD plot obtained when base pairs 646 and 1268 in the alignment are selected as recombination breakpoints. **B** shows detail of alignment of left and right borders of the presumed recombination event *C. rhomboideum AT3-1*. Vertical arrows indicate the position of base pairs 646 and 1268 in the alignment and in the LARD plot.

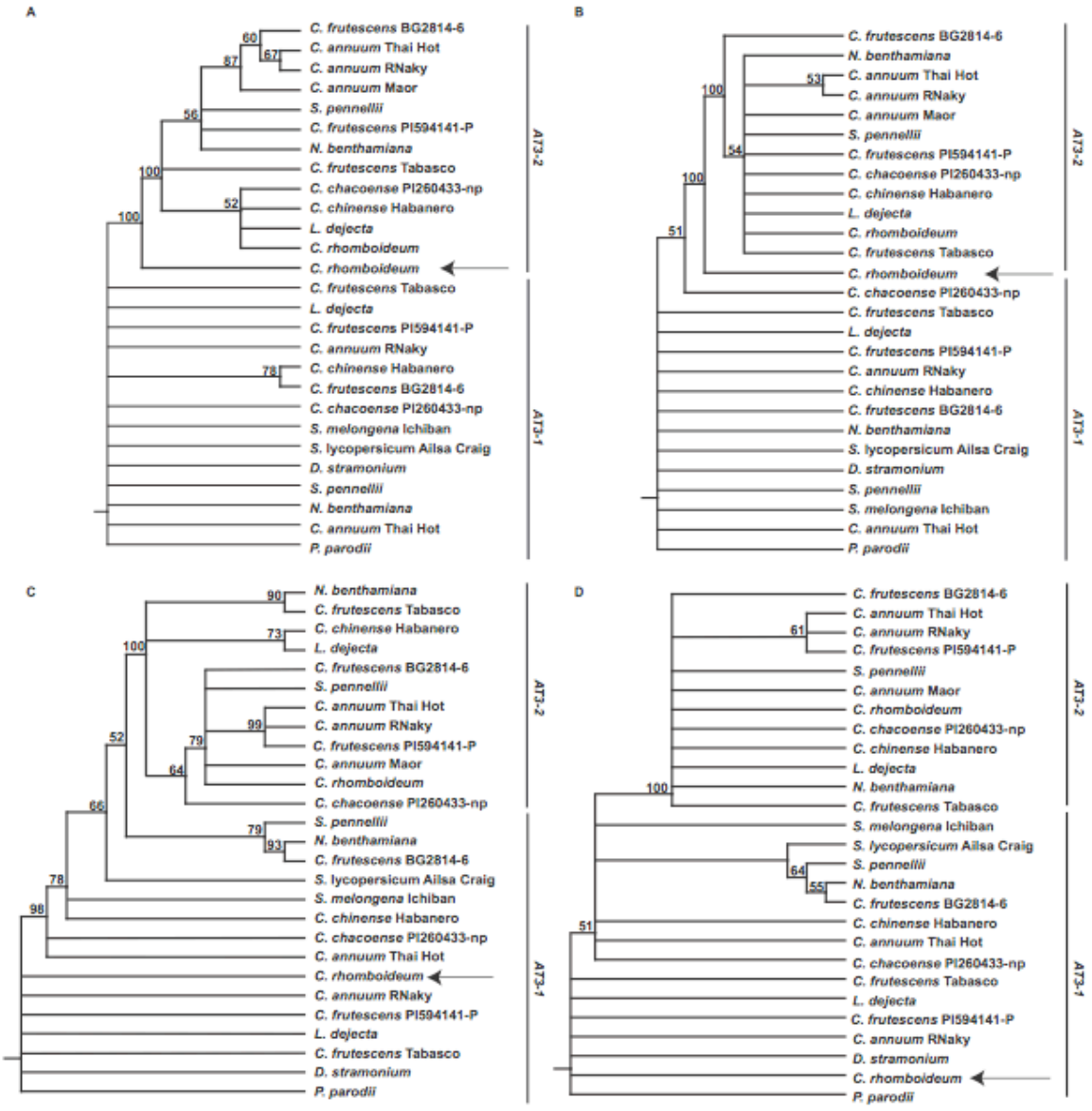

A
327
218
109
0.00
Log likelihood ratio
1
453
907
1361
1815
Position in alignment
B
646
1268
AT3-2
AT3-1
AT3-1
C. frutescens PI594141-P
C. frutescens Tabasco
N. benthamiana
C. annuum Thai Hot
S. pennellii
C. annuum Maor
C. chacoense PI260433-np
C. chinense Habanero
C. rhomboideum
L. dejecta
C. annuum RNaky
C. frutescens BG2814-6
C. rhomboideum
S. lycopersicum Ailsa Craig
D. stramonium
S. pennellii
P. parodii
S. melongena Ichiban
N. benthamiana
L. dejecta
C. chinense Habanero
C. chacoense PI260433-np
C. frutescens Tabasco
C. frutescens PI594141-P
C. annuum Thai Hot
C. frutescens BG2814-6
C. annuum RNaky

To test the robustness of the recombination event identified using RDP3 and LARD, we generated new alignments that either excluded the region identified as recombinant in *C. rhomboideum AT3-1* or were solely comprised by the region identified as recombinant in *C. rhomboideum AT3-1*. Phylogenies were then estimated using maximum parsimony and Bayesian approaches, as was done for the full-length phylogeny with the sole change that branch swapping was not allowed in estimating bootstrap support for the MP phylogeny. Taken together, the data show that in *C. rhomboideum AT3-1* the region between base pairs 646 and 1268 falls within a clade containing all *AT3-2* sequences and the placement of all *AT3-2* sequences together receives strong bootstrap and posterior probability support (Figure 3.6A and 3.6B). In contrast, the region excluding base pairs 646 to 1268 is excluded from the clade containing all other *AT3-2* sequences and groups instead with the remaining *AT3-1* sequences. The *AT3-1* clade that was strongly recovered in both the MP and Bayesian analyses of the full-length alignment (Figures 3.3 and 3.4) was not recovered using the truncated alignment possibly due to a loss of information associated with the truncation of the dataset (Figure 3.6). Overall the newly estimated phylogenies show lower levels of support and less resolution than those generated with the full length alignment; again this is likely due to the loss of phylogenetic signal in the new alignment. Nevertheless, the alignment that excludes the region between base pairs 646 to 1268 shows that a clade that includes all *AT3-2* sequences can be identified and that *C. rhomboideum AT3-1* falls outside this clade when the putative recombinant base pairs are excluded. The data shown in Figure 3.6 support the results of the RDP3 scan for recombination events by confirming that the region identified as recombinant shows different phylogenetic placement than the non-recombinant base pairs. Furthermore, this phylogenetic relationship is different from that seen with the full-length alignment (Figures 3.3, 3.4 and 3.6). This indicates that the anomalous

placement of *C. rhomboideum AT3-1* is due to the contradicting phylogenetic signal contained in this sequence caused by the identified recombination event.

Given the results of these analyses, the recombination event is likely to have occurred exclusively in *C. rhomboideum* and after the diversification of *Capsicum*. Interestingly, the recombination event preserves the open reading frame by inserting into the coding region of *C. rhomboideum AT3-1* the coding region of *AT3-2*, the pseudogenized paralog for which we found no evidence of expression. The recombination event occurs immediately following the putative active site of the enzyme denoted by the sequence HXXXDG. Furthermore, the sequence of *C. rhomboideum AT3-1* appears to be truncated by a 2-basepair deletion that causes a frameshift mutation leading to a stop codon in the second exon. This deletion occurs just before the conserved DFGWG motif (St-Pierre and De Luca, 2000). This motif may contribute to enzyme stability and has been shown to be necessary for enzyme function, wherein a single amino acid change in the motif leads to either a decrease or complete loss of function in 5-O-glucoside-6‴-O-malonyltransferase (Bayer et al, 2004; Suzuki et al, 2003). The presence of this deletion suggests that the *C. rhomboideum AT3-1* transcript would lack the terminal 70 amino acids.

Figure 3.6. Bayesian and Maximum Parsimony (MP) phylogenies of recombinant region chosen as per the results of RDP analysis. Panels **A** and **B** show Bayesian and MP phylogenies of the gene region comprising base pairs 646 to 1268 in the *AT3-1* and *AT3-2* alignment respectively. Panels **C** and **D** show Bayesian and MP phylogenies of the gene region excluding base pairs 646 to 1268 in the *AT3-1* and *AT3-2* alignment respectively. Numbers represent posterior probabilities (A/C) or bootstrap proportions (B/D). Arrow highlights the placement of the recombinant *AT3-1* sequence from *C. rhomboideum*.


A
C. frutescens BG2814-6
C. annuum Thai Hot
C. annuum RNaky
C. annuum Maor
S. pennellii
C. frutescens PI594141-P
N. benthamiana
C. frutescens Tabasco
C. chacoense PI260433-np
C. chinense Habanero
L. dejecta
C. rhomboideum
C. rhomboideum
C. frutescens Tabasco
L. dejecta
C. frutescens PI594141-P
C. annuum RNaky
C. chinense Habanero
C. frutescens BG2814-6
C. chacoense PI260433-np
S. melongena Ichiban
S. lycopersicum Ailsa Craig
D. stramonium
S. pennellii
N. benthamiana
C. annuum Thai Hot
P. parodii
AT3-2
AT3-1
B
C. frutescens BG2814-6
N. benthamiana
C. annuum Thai Hot
C. annuum RNaky
C. annuum Maor
S. pennellii
C. frutescens PI594141-P
C. chacoense PI260433-np
C. chinense Habanero
L. dejecta
C. rhomboideum
C. frutescens Tabasco
C. rhomboideum
C. chacoense PI260433-np
C. frutescens Tabasco
L. dejecta
C. frutescens PI594141-P
C. annuum RNaky
C. chinense Habanero
C. frutescens BG2814-6
N. benthamiana
S. lycopersicum Ailsa Craig
D. stramonium
S. pennellii
S. melongena Ichiban
C. annuum Thai Hot
P. parodii
AT3-2
AT3-1
C
N. benthamiana
C. frutescens Tabasco
C. chinense Habanero
L. dejecta
C. frutescens BG2814-6
S. pennellii
C. annuum Thai Hot
C. annuum RNaky
C. frutescens PI594141-P
C. annuum Maor
C. rhomboideum
C. chacoense PI260433-np
S. pennellii
N. benthamiana
C. frutescens BG2814-6
S. lycopersicum Ailsa Craig
S. melongena Ichiban
C. chinense Habanero
C. chacoense PI260433-np
C. annuum Thai Hot
C. rhomboideum
C. annuum RNaky
C. frutescens PI594141-P
L. dejecta
C. frutescens Tabasco
D. stramonium
P. parodii
AT3-2
AT3-1
D
C. frutescens BG2814-6
C. annuum Thai Hot
C. annuum RNaky
C. frutescens PI594141-P
S. pennellii
C. annuum Maor
C. rhomboideum
C. chacoense PI260433-np
C. chinense Habanero
L. dejecta
N. benthamiana
C. frutescens Tabasco
S. melongena Ichiban
S. lycopersicum Ailsa Craig
S. pennellii
N. benthamiana
C. frutescens BG2814-6
C. chinense Habanero
C. annuum Thai Hot
C. chacoense PI260433-np
C. frutescens Tabasco
L. dejecta
C. frutescens PI594141-P
C. annuum RNaky
D. stramonium
C. rhomboideum
P. parodii
AT3-2
AT3-1

A similar mutation has been identified in the non-pungent cultivar *C. frutescens* PI594141-np, which has a large deletion immediately preceding the DFGWG site and no detectable transcription of *AT3-1* (Stellari, unpublished data). Furthermore, *C. rhomboideum AT3*-1 shows many more non-synonymous mutations than those seen in the remaining *AT3-1* and *AT3-2* sequences. It seems possible that the inactivating mutation in this species may have released *AT3-1* from the selective constraint which may be enforcing conservation of the *AT3-1* sequences in the other solanaceous genera. In general, *C. rhomboideum* is considered non-pungent, and the accession used in this study was determined to have no detectable capsaicinoids by ELISA (see materials and methods; Eshbaugh, 1980). This suggests that the deletion identified in *C. rhomboideum AT3-1* either alone or in combination with the recombination event between *AT3-1* and *AT3-2* is causal for loss of pungency in this accession, which is consistent with the pivotal role of *AT3-1* in capsaicin biosynthesis (Stewart et al, 2005; Stewart et al, 2007). *C. rhomboideum* germplasm is not readily available, hampering a more thorough sampling of capsaicinoid accumulation and sequence diversity at *AT3-1* in this species. However, the data presented here represent only the second documentation of non-pungency in the species and further support our understanding of the critical role of *AT3-1* in allowing capsaicinoid biosynthesis (Eshbaugh, 1980).

While conservation of the gene duplicate is one outcome of the duplicative process, loss of the duplicated copy by the process of pseudogene formation is another, ostensibly more likely outcome. Evidence gathered in the course of this study suggests that *AT3-2* may represent a functional pseudogene. A comprehensive review of pseudogene formation and fate has identified that eukaryotic pseudogenes often exhibit characteristics of functional genes including transcription, an excess of non-synonymous mutations, and the preservation of open reading frames (Balakirev and

Ayala, 2003). This suggests that some pseudogenes may be under selective pressure maintaining either an aspect of their original function or a novel function. Strict conservation of the intron-exon boundary of *AT3-2* along with partial conservation of the amino acid alignment between *AT3-1* and *AT3-2*, especially in the second exon, seems to suggest that *AT3-2* is under selective constraint that prevents the degeneration of its sequence despite its status as a pseudogene. In fact, *AT3-2* appears to show the signature of both positive and negative selection along its putative coding region (Stellari, data not shown). This would implicate some form of selective pressure on the coding region of *AT3-2* possibly preceding its pseudogenization and which may have been maintained despite the presence of the pseudogene creating mutations that we have identified. Furthermore, the mutations responsible for *AT3-2*'s pseudogene status are broadly conserved throughout the Solanaceae, suggesting that its divergence from *AT3-1* at the amino acid level occurred early in the evolution of this plant family.

Recent studies have uncovered that tandem pseudogenes can regulate the expression of the duplicate from which they were derived (Tam et al, 2008). To date few studies outside of model organisms have tracked the evolutionary history of a tandem gene duplication event (Gottlieb and Ford, 1997; Voelker et al, 1986). In contrast, most studies have focused on related organisms for which a whole genome sequence is available (Johnson and Thomas, 2007; Tasma et al, 2008). Although functional pseudogenes have been identified in the past, the evidence linking the pseudogene to its function has only been identified more recently (Balakirev and Ayala, 2003; Gottlieb and Ford, 1997; Tam et al, 2008). Although we found no evidence for the expression of *AT3-2*, the possibility exists that this pseudogene is expressed at low levels or only partially such that it was missed in the course of our study. Inclusion of the sequence of the *AT3* homolog predating the duplication

uncovered in the course of this study is necessary to conclusively describe the relationship between *AT3-1* and *AT3-2*.

CONCLUSIONS

In the course of this study we have identified a tandem gene duplication event for *AT3-1*, a gene that has been shown to be pivotal for the expression of the pungency trait in the genus *Capsicum*. The duplication event uncovered in the course of this study is broadly conserved through the Solanaceae and likely predates the diversification of the family. The strong degree of sequence conservation seen between *AT3-1* and *AT3-2* suggest that *AT3-2* and *AT3-1* diverged before the diversification of the Solanaceae. Despite showing nucleotide level divergence, the paralogous gene lineages retain regions of amino acid conservation. At the same time, inactivating mutations would suggest that *AT3-2* is now a pseudogene, an event which also seems to predate the diversification of the Solanaceae, raising the question as to the reason for the high degree of conservation seen among this tandemly duplicated gene pair. We identify an in-frame recombination event between *AT3-1* and *AT3-2* in a non-pungent accession of *C. rhomboideum*. Furthermore, in this same accession, a frameshift mutation in the second exon eliminates the canonical DFGWG site suggesting that a truncated transcript would lead to impaired AT3-1 function, thereby explaining lack of capsaicinoid accumulation found in *C. rhomboideum* in this study. Further investigation into the conservation of this recombination event in a wider sampling of *C. rhomboideum* germplasm may demonstrate the mutation that we have uncovered as the event causing loss of pungency species-wide in this basal species in the genus.

ACKNOWLEDGMENTS

We would like to thank the entire Jahn lab for their assistance in the preparation of this manuscript. Special thanks to Lynn Bohs for providing *C. rhomboideum* and *Lycianthes* seeds. Special thanks also to Lukas Mueller, Jeff Doyle, R. Geeta, Ashley Egan, Boris Igic, Anna Savage, and Shannon Straub for their helpful comments and suggestions in refining the analysis of the data presented herein. This work was supported by a USDA NRI (CREES #2004-35300-14596) to S.M.F. and an NSF GRF to G.M.S. as well as additional support from the Dean's fund at SUNY Stony Brook.

WORKS CITED

Alfaro ME, Holder MT (2006). The posterior and the prior in Bayesian phylogenetics. *Annual Review of Ecology Evolution and Systematics* **37:** 19-42.

Anthony NM, Clifford SL, Bawe-Johnson M, Abernethy KA, Bruford MW, Wickings EJ (2007). Distinguishing gorilla mitochondrial sequences from nuclear integrations and PCR recombinants: Guidelines for their diagnosis in complex sequence databases. *Molecular Phylogenetics and Evolution* **43**(2)**:** 553-566.

Balakirev ES, Ayala FJ (2003). Pseudogenes: Are they "Junk" or functional DNA? *Annual Review of Genetics* **37:** 123-151.

Bayer A, Ma XY, Stockigt J (2004). Acetyltransfer in natural product biosynthesis - functional cloning and molecular analysis of vinorine synthase. *Bioorganic and Medicinal Chemistry* **12**(10)**:** 2787-2795.

Bohs L, Olmstead RG (1997). Phylogenetic relationships in *Solanum* (Solanaceae) based on *ndhF* sequences. *Systematic Botany* **22**(1)**:** 5-17.

D'Auria JC (2006). Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology* **9**(3)**:** 331-340.

D'Auria JC, Chen F, Pichersky E (2002). Characterization of an acyltransferase capable of synthesizing benzylbenzoate and other volatile esters in flowers and damaged leaves of *Clarkia breweri*. *Plant Physiology* **130**(1)**:** 466-476.

Eshbaugh WH (1980). The taxonomy of the genus *Capsicum* (Solanaceae). *New Phytologist* **47**(2)**:** 153-166.

Garces-Claver A, Fellman SM, Gil-Ortega R, Jahn M, Arnedo-Andres MS (2007). Identification, validation and survey of a single nucleotide polymorphism (SNP) associated with pungency in *Capsicum* spp. *Theoretical and Applied Genetics* **115**(7)**:** 907-916.

Gottlieb LD, Ford VS (1997). A recently silenced, duplicate PgiC locus in *Clarkia*. *Molecular Biology and Evolution* **14**(2)**:** 125-132.

Hadacek F (2002). Secondary metabolites as plant traits: Current assessment and future perspective. *Critical Reviews in Plant Sciences* **21**(4)**:** 273-322.

Han SS, Keum YS, Seo HJ, Chun KS, Lee SS, Surh YJ (2001). Capsaicin suppresses phorbol ester-induced activation of NF-kappaB/Rel and AP-1 transcription factors in mouse epidermis. *Cancer Letters* **164**(2)**:** 119-126.

Hasegawa M, Kishino H, Yano TA (1985). Dating of the human ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**(2)**:** 160-174.

He XL, Zhang JZ (2005). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**(2)**:** 1157-1164.

Holmes EC, Worobey M, Rambaut A (1999). Phylogenetic evidence for recombination in dengue virus. *Molecular Biology and Evolution* **16**(3)**:** 405-409.

Johnson DA, Thomas MA (2007). The monosaccharide transporter gene family in Arabidopsis and rice: A history of duplications, adaptive evolution, and functional divergence. *Molecular Biology and Evolution* **24**(11)**:** 2412-2423.

Jordt S-E, Julius D (2002). Molecular basis for species-specific sensitivity to "hot" chili peppers. *Cell* **108**(3)**:** 421-430.

Jukes T, Cantor, C. (1969). Evolution of protein molecules. In: Munro H (ed) *Mammalian Protein Metabolism*. Academic Press: New York, pp 21-132.

Lewinsohn E, Gijzen M (2009). Phytochemical diversity: The sounds of silent metabolism. *Plant Science* **176**(2)**:** 161-169.

Lynch M, Force A (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**(1)**:** 459-473.

Martin DP, Williamson C, Posada D (2005). RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**(2)**:** 260-262.

Martins TR, Barkman TJ (2005). Reconstruction of Solanaceae phylogeny using the nuclear gene *SAMT*. *Systematic Botany* **30**(2)**:** 435-447.

Morgenstern B (2004). DIALIGN: Multiple DNA and Protein Sequence Alignment at BibiServ. *Nucleic Acids Research* **32:** W33-W36.

Ohno S (1970). *Evolution by Gene Duplication*. Springer-Verlag: New York.

Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia VF, Collier SM (2008). A molecular phylogeny of the Solanaceae. *Taxon* **57**(4)**:** 1159-1181.

Olmstead RG, Palmer J D (1997). Implications for the phylogeny, classification, and biogeography of Solanum from cpDNA restriction site variation. *Systematic Botany* **22**(1): 19 29.

Padidam M, Sawyer S, Fauquet CM (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**(2)**:** 218-225.

Posada D, Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**(9)**:** 817-818.

Posada D, Crandall KA (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences of the United States of America* **98**(24)**:** 13757-13762.

Ronquist F, Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**(12)**:** 1572-1574.

Smith JM (1992). Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* **34**(2)**:** 126-129.

St-Pierre B, De Luca V (2000). Evolution of acyltransferase genes: origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. *Recent Advances in Phytochemistry* (Evolution of Metabolic Pathways) **34:** 285-315.

Stewart C, Kang BC, Liu K, Mazourek M, Moore SL, Yoo EY *et al* (2005). The Pun1 gene for pungency in pepper encodes a putative acyltransferase. *Plant Journal* **42**(5)**:** 675-688.

Stewart C, Mazourek M, Stellari GM, O'Connell M, Jahn M (2007). Genetic control of pungency in *C. chinense* via the *Pun1* locus. *Journal of Experimental Botany* **58**(5)**:** 979-991.

Suzuki H, Nakayama T, Nishino T (2003). Proposed mechanism and functional amino acid residues of malonyl-CoA : anthocyanin 5-O-glucoside-6 '"-O-malonyltransferase from flowers of *Salvia splendens*, a member of the versatile plant acyltransferase family. *Biochemistry* **42**(6)**:** 1764-1771.

Swofford DL. (2002). Sinauer Associates: Sunderland, MA.

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S *et al* (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**(7194)**:** 534-U538.

Tanksley SD, Ganal MW, Prince JP, Vicente MC, de Bonierbale MW, Broun P *et al* (1992). High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**(4)**:** 1141-1160.

Tasma IM, Brendel V, Whitham SA, Bhattacharyya MK (2008). Expression and evolution of the phosphoinositide-specific phospholipase C gene family in Arabidopsis thaliana. *Plant Physiology and Biochemistry* **46**(7)**:** 627-637.

Taylor JS, Raes J (2004). Duplication and divergence: The evolution of new genes and old ideas. *Annual Review of Genetics* **38:** 615-643.

Tewksbury JJ, Nabhan GP (2001). Seed dispersal. Directed deterrence by capsaicin in chilies. *Nature* **412**(6845)**:** 403-404.

Voelker TA, Staswick P, Chrispeels MJ (1986). Molecular analysis of 2 phytohemagglutin genes and their expression in *Phaesolus vulgaris* cv. Pinto, a lecitin deficient cultivar of the bean. *European Molecular Biology Organization* **5**(12)**:** 3075-3082.

Walsh BM, Hoot SB (2001). Phylogenetic relationships of *Capsicum* (Solanaceae) using DNA sequences from two noncoding regions: the chloroplast *atpB-rbcL* spacer region and nuclear *waxy* introns. *International Journal of Plant Sciences* **162**(6)**:** 1409-1418.

Webber H (1911). *American Breeders Association Annual* **V7:** 188.

CHAPTER 4: THE SUBCELLULAR LOCALIZATION OF AT3-1 IN *N. BENTHAMIANA* LEAVES

(Giulia Marina Stellari[7], Vitaly Citovsky[8], and Molly Jahn[9†])

ABSTRACT

Capsaicinoids are synthesized on the placental dissepiment of pungent peppers of endproducts derived from the fatty acid and phenylpropanoid biosynthetic pathways. Despite ample evidence showing that the *Pun1* locus and the gene that it encodes, the putative BAHD acyltransferase, *AT3-1*, is critical for capsaicinoid biosynthesis in cultivated species in the genus *Capsicum*, its role in the capsaicinoid biosynthetic pathway has remained elusive. This study uses fluorescent protein tagging to show that AT3-1 localizes to a thin reticulated network whose pattern is associated with the endoplasmic reticulum (ER). Treatment of cells with Latrunculin A disrupts subcellular localization of AT3-1 supplying further evidence that this protein associates with the ER. The DFGWG motif, which is conserved among BAHD acyltransferases, is shown to be important for protein localization. AT3-1 localization to the ER suggests that AT3-1 is not capsaicinoid synthase and its role in the capsaicinoid biosynthetic pathway is discussed in light of this evidence.

---

[7] Present address: Biochemistry Department, 424 Life Sciences Building, SUNY Stony Brook, Stony Brook NY, 11794.
[8] Present address: Biochemistry Department, 414 Life Sciences Building, SUNY Stony Brook, Stony Brook NY, 11794.
[9] Present address: Department of Agronomy and Genetics, 140 Agriculture Hall, 145 Linden Drive, Madison, WI 53706.
[†] Author contributions: GMS designed and executed experiments; all authors read and reviewed the manuscript before submission.

INTRODUCTION

Capsaicinoids are a group of structurally related secondary metabolites whose biosynthesis is taxonomically restricted to the genus *Capsicum* (Eshbaugh, 1980). Capsaicinoid biosynthesis involves the joining of end products of two major plant biosynthetic pathways, the phenylpropanoid and fatty acid pathways (Suzuki, 1984). The terminal enzyme in the capsaicinoid biosynthetic pathway has remained elusive and has never been purified using traditional biochemical means (Suzuki, 1984). It is thought to catalyze the reaction joining vanillylamine, derived from the phenylpropanoid pathway, to 8-methyl-6-nonenoyl-CoA, from the fatty acid pathway, and is known as capsaicinoid synthase (Bennett and Kirby, 1968; Leete and Louden, 1968; Stewart *et al*, 2005). Both the fatty acid and the phenylpropanoid biosynthetic pathways are ubiquitous and contribute end products and intermediates to the more than 500,000 secondary metabolites which plants are known to produce (Hadacek, 2002). Furthermore, plant secondary metabolites in general and capsaicinoids specifically, are of great interest and value to the flavoring and pharmaceutical industries (Govindarajan, 1986; Schwab *et al*, 2008).

Capsaicinoids are one of the most widely consumed spices and as such understanding the tissue specific localization for their biosynthesis is important to the oleoresin and spice industries which wish to maximize the efficiency of capsaicinoid extraction (Govindarajan, 1986). Ohta showed that capsaicinoids accumulate in the placental dissepiment of pungent peppers, although sporadic reports have identified capsaicinoids in the pericarp, the leaves and the stems of pepper plants (Estrada-Berta *et al*, 2002; Iwai *et al*, 1979; Ohta, 1962; Tandon *et al*, 1964). Recent investigations that combined the immunolocalization of structural genes putatively in the capsaicinoid biosynthetic pathway, with an analysis of gross fruit morphology, have

shown that capsaicinoids and the enzymes involved in their biosynthesis, accumulate predominantly in the epidermal cells of the placental dissepiment of pungent peppers (Aluru *et al*, 2003; Stewart *et al*, 2007). During development, capsaicinoids are secreted into the extracellular space between the epidermal layer and the cuticle. Their excretion lifts the cuticle forming raised pockets that have been described as "blisters" or as "receptacles of capsaicin" which fill with a capsaicinoid containing oleoresin. The cells directly beneath the blister may be glandular in nature. In the non-pungent cultivar *C. chinense* 'NMCA30036', genetic control of blister development is linked to the *Pun1* locus, which is necessary for capsaicinoid accumulation in domesticated species in the genus *Capsicum* (Stewart *et al*, 2005; Stewart *et al*, 2007). The blisterless phenotype seen in *C. chinense* 'NMCA30036' perfectly co-segregates with loss of pungency and with a four base pair deletion in the first exon of *AT3*, the gene encoded by the *Pun1* locus (Stewart *et al*, 2007). A duplication of *AT3* is described in Chapter 3, thus for clarity and consistency, *AT3* is herein referred to as *AT3-1*.

While it has now been conclusively shown that capsaicinoids accumulate in the placental dissepiment, such clarity has not been achieved for the subcellular localization of capsaicinoid production and storage. One study that used transmission electron microscopy (TEM) and ultraviolet microscopy found the presence of electron dense granules that stained deeply with osmium tetroxide in TEM sections of the placentas of pungent peppers from the variety *C. annuum* Karyatsubusa (Suzuki *et al*, 1980). These granules were identified mainly in the vacuole, attached to the tonoplast membrane and in smaller round structures identified as vesicles. Similar granular structures preferentially absorbing light at a wavelength of 280nm, the same absorption maxima as capsaicinoids, were also identified in transverse sections of pepper placentas at the developmental stage when capsaicinoids accumulate in the fruit. Taken together these results suggest that the electron dense granules identified in

the vacuoles and vesicles of the developing placenta of pungent peppers contained capsaicinoids. In contrast, a second study again using TEM found electron dense granules only in vesicles and in the ER but not in the vacuole of epidermal cells of the placental dissepiment (Zamski *et al*, 1987). These electron dense granules were also identified in the tissue directly beneath the raised cuticle and thus they were interpreted to be capsaicinoids. Although the results strongly suggest that capsaicinoids accumulate either in the ER or the in the vacuole, these results should be interpreted with caution as the studies did not use an antibody raised against capsaicin to verify the content of the osmiophilic granules.

While the capsaicinoid biosynthetic pathway is not fully characterized at the level of enzymatic reactions and subcellular localization, a more detailed understanding of the subcellular localization of the fatty acid and phenylpropanoid biosynthetic pathways has been gained in a number of different plant species. Branched chain fatty acid biosynthesis in plants localizes to the plastid and is catalyzed by a large multisubunit enzyme, fatty acid synthase (FAS) (Harwood, 1988). The export of the products of this pathway from the plastid to the cytosol is mediated by acyl-CoA synthetase (ACS) (Harwood, 1988). In contrast, the phenylpropanoid pathway seems to be localized to the cytosolic face of the endoplasmic reticulum (ER). Phenylalanine ammonia lyase (PAL) and cinnamate-4-hydroxylase (C4H), the first and second steps in the phenylpropanoid pathway localize to the ER in *Primula kewensis* and in Poplar, respectively, suggesting that pools of pathway intermediates may occur in close proximity to each other within the cell (Ro *et al*, 2001; Schopker *et al*, 1995).

Although the identity of capsaicinoid synthase is still unknown, the subcellular localization of the enzymatic activity associated with this enzyme has been investigated. Investigation of capsaicinoid synthase activity using crude plant extracts

from protoplasts derived from the placenta of pungent peppers identified this enzymatic activity in the vacuolar fraction, which also had the highest concentration of capsaicinoids (Fujiwake *et al*, 1980; Fujiwake *et al*, 1982). This same study also investigated the subcellular localization of other enzymes in the capsaicinoid biosynthetic pathway and showed that PAL activity was associated with the cytosolic subfraction (Fujiwake *et al*, 1982). Taken together these results suggest that capsaicinoid biosynthesis likely involves multiple plant organelles and that the terminal enzyme in this biosynthetic pathway is active in the vacuole.

Recombinant expression of AT3-1 has proven difficult due to the insoluble nature of this protein; furthermore, pepper is recalcitrant to transformation, thus hampering confirmation of AT3-1's role in the capsaicinoid biosynthetic pathway using functional complementation. As a means of identifying the function of AT3-1 in the capsaicinoid biosynthetic pathway, this protein was heterologously expressed in translational fusion with a yellow fluorescent protein (YFP) tag. YFP-AT3-1 is shown to localize to the ER and this localization can be disrupted through treatment with Latrunculin-A. Localization of constructs containing various deletions of the C-terminus of AT3-1 are also tested and they identify a motif that may be important for the correct folding or subcellular localization of this protein.

MATERIALS AND METHODS

Plant growth and tissue collection

Capsicum chinense ''Habanero'' was grown in the greenhouse facility at SUNY Stony Brook, Stony Brook, NY. Growing conditions were approximately 27°/18°C (day/night) with supplemental lighting. Plants were fertilized weekly with

100ppm Blossom Booster (J.R. Peters Inc., Allentown, PA).

Nicotiana benthamiana plants used for biolistic transformation and for *Agrobacterium tumefaciens* mediated infiltration were grown under short day conditions in the growth chamber facilities at SUNY Stony Brook, Stony Brook, NY. Growth conditions were approximately 22°C (day/night) with approximately 2:1 fluorescent:incandescent illumination. Plants were fertilized weekly with 100ppm NPK Peters Mix (J.R. Peters Inc., Allentown, PA).

Identification of full length C. chinense 'Habanero' AT3-1

C. chinense 'Habanero' *AT3-1* was cloned from cDNA and used to create the constructs imaged in this study. The placental dissepiment of young green fruit from *C. chinense* ''Habanero'' was collected and frozen immediately in liquid nitrogen, ground to a powder, and stored at -80°C. RNA was extracted from this tissue using a Qiagen RNeasy kit (Qiagen, Valencia CA) according to the manufacturer's instructions. Approximately 100mg of frozen ground tissue was utilized for each extraction. RNA was denatured and visualized on 1% TAE agarose gel to assess quality. Quantity was assessed using NanoDrop (Nanodrop, Wilmington, DE). First strand cDNA was synthesized using the Protoscript First Strand cDNA Synthesis Kit from New England Biolabs (New England Biolabs, Ipswich, MA) as per manufacturer's instructions and reverse transcribed using polyT_{23}VN. 500ng of total RNA was used as starting material for the cDNA synthesis reaction. Residual RNA remaining after the cDNA synthesis was digested using 1µl of RNAseH incubating at 37°C for 15 minutes followed by heat inactivation of the enzyme at 95°C for 5 minutes.

C. chinense ‘‘Habanero’’ *AT3-1* was isolated using gene specific primers designed to anneal to the 5’UTR and spanning the stop codon. PCR was performed as follows: 5µl of 10X PCR Buffer, 4µl of 2.5mM dNTPs, 1µl of 10mM *AT3-1* F (ATGTCAACCGGCCAGCAGCAT), 1µl of 10mM *AT3-1* R (CGCGGATCCTCCAAAACTTGGAATTAATTAGGC), .25µl ExTaq (Takara Bio Inc., Otsu, Japan) and cycling with the following conditions: 94°C for 3 minutes, 29 cycles of 94°C for 30 seconds, 65°C for 30 seconds, 72°C for 3 minutes, with a final extension at 72°C for 20 minutes. Reactions were performed using a PTC 225 Peltier Thermal Cycler (MJ Research, Watertown, MA).

The PCR products were prepared for TA cloning by adding to each PCR reaction 0.2µl Roche Taq (Roche, Indianapolis, IN) and incubating at 72°C for 20 minutes. 2µl of this reaction product were utilized for ligation with the Invitrogen TOPO TA PCR 4 Cloning Kit (Invitrogen Corporation, Carlsbad, CA). Ligations were performed as follows: 4µl of gel purified PCR product, 1µl Salt Solution, 1µl TOPO vector. Reactions were incubated at room temperature for 20 minutes. 2µl of the ligation product was combined with TOPO10 cells, heat shocked at 42°C for 30 seconds, and recovered at 37°C using 250µl SOC for 1 hour. The entire transformation was plated LB/agar/kanamyacin/XGal plates and incubated at 37°C overnight. Resulting colonies were stored at 4°C. At least three white clones were selected for sequencing, and sequenced at the OSA Sequencing Facility at SUNY Stony Brook (http://www.osa.sunysb.edu/dna/). A wildtype full-length clone was identified based on comparisons with the *C. chinense AT3-1* sequence deposited in GenBank (accession number AY819027), and was used for further experiments.

Plasmid construction

The subcellular localization of AT3-1 was visualized using confocal microscopy. A clone containing a yellow fluorescent protein (YFP) tag in translational fusion to the N-terminus of AT3-1 was created by amplifying the full-length cDNA of *AT3-1* with gene specific primers adapted with endonuclease restriction sites followed by cloning into the pSAT6 vector YFP-C1 (Tzfira *et al*, 2005). PCR amplification was performed as follows: 10µl of 5X PCR High Fidelity Buffer, 1µl of 10mM dNTPs, 1µl of 10mM YFPc-AT3-1 F (GGAAGATCTATGGCTTTTGCATTACCATCATCAC), 1µl of 10mM YFPc-AT3-1 R (ACGCGTCGACGGCAATGAACTCAAGGAG), 0.5µl Phusion (Finnzymes, Woburn, MA) and cycling with the following conditions: 98°C for 30 seconds, 34 cycles of 98°C for 10 seconds, 55°C for 30 seconds, 72°C for 1 minute 30 seconds, with a final extension at 72°C for 20 minutes. Reactions were performed using a PTC 225 Peltier Thermal Cycler (MJ Research, Watertown, MA).

PCR products were visualized on a 1.0% TAE agarose gel, excised and gel purified using the GFX PCR DNA and Gel Band Purification Kit (GE, Healthcare Biosciences Inc., Piscataway, NJ) as per manufacturer's instructions. Purified gel products were simultaneously digested in 3µl Buffer 3, 3µl 10X BSA, 22µl purified gel product, 1µl BglII, 1µl SalI (New England Biolabs, Ipswich, MA) for 3 hours at 37°C. Approximately 500ng of pSAT6-YFP-C1 was digested as per the protocol stated above. Digested PCR products and vector backbone were visualized and purified from a 1% TAE agarose gel. Digested PCR products and pSAT6 YPF-C1 linearized backbone were ligated overnight at 16°C in 2.0µl 10X T4 DNA Ligase Buffer, 1µl T4 DNA Ligase (New England Biolabs, Ipswich, MA) and a mixture of digested PCR products and backbone to a final volume of 20µl. Ligation products were transformed into DH5α chemically competent cells (New England Biolabs,

Ipswich, MA) by heat shocking at 42°C for 30 seconds followed by a 1 hour recovery at 37°C in 900µl SOC. Positive colonies were identified by plating on LB/agar/Ampicillin incubating overnight at 37°C. At least three clones were selected and sequenced at the OSA Sequencing Facility at SUNY Stony Brook (http://www.osa.sunysb.edu/dna/). A single full-length wildtype clone based on sequence comparison with *C. chinense* ''Habanero'' *AT3-1* sequence and with the predicted map of the YFP-AT3-1 construct, was selected for biolistic transformation.

Constructs bearing deletions of the C terminus of AT3-1 were created as per the protocol stated above. For the construct bearing a deletion of the terminal 70 amino acids of AT3-1 the following primer pair, adapted with the restriction endonuclease sites BglII/SalI, was used: YFPc-AT3-1 F (GGAAGATCTATGGCTTTTGCATTACCATCATCAC) with AT3-1 Δ70 R (ACGCGTCGACTCCCCATCCAAAATCTACAGTGTAGTACGG). For the construct bearing a deletion of the terminal 65 amino acids of AT3-1 the same forward primer stated above with the same restriction endonuclease digestion sites were used in combination with the reverse primer AT3-1 Δ65 R (ACGCGTCGACTACAGTGTAGTACGGATATTTGC). PCR cycling, gel purification, ligation and cloning were performed as stated above.

Biolistic transformation

Gold particles were prepared for biolistic transformation as follows. Approximately 5µg of plasmid DNA (described above) were prepared for transformation and affixed to 1µM gold particles using 1M $CaCl_2$ (BioRad, Hercules, CA) followed by 3 washes in 100% Ethanol and final resuspension in 3ml 100% Ethanol with the addition of 7.5µl of a 20M PVP solution. Gold particles were

biolistically bombarded onto the abaxial surface of 6 to 8 week old *N. benthamiana* leaves. Particles were accelerated at 1200psi at a distance of approximately 2cm from the surface of the leaves. Leaves were incubated in a humid chamber kept at room temperature in the dark overnight.

Agrobacterium mediated transient expression

Agrobacterium tumefaciens strain EHA105 was used to deliver the binary vector containing the YFP-AT3-1 expression cassette. The binary vector pPZP250, based on the pPZP series of vectors, was used for this experiment and all expression cassettes were cloned using the homing endonuclease site PI-PSPI (Hajdukiewicz *et al*, 1994). Approximately 500ng of pPZP250 and 500ng of plasmid containing the target insert were digested in separate reactions using 3µl Buffer PI-PSPI, 3µl 10X BSA, 1µl PI-PSPI (New England Biolabs, Ipswich, MA), 500ng of vector in a total reaction volume of 30µl incubated at 37°C for 3 hours. Digested products were cleaned using the GFX PCR DNA and Gel Band Purification Kit, and ligated with 1µl T4 DNA ligase at 16°C overnight as described above. Plasmids were first transformed into DH5α chemically competent cells (New England Biolabs, Ipswich, MA) by heat shocking at 42°C for 30 seconds followed by a 1 hour recovery at 37°C in 900µl SOC. Positive colonies were identified by plating on LB/agar/spectinomycin incubating overnight at 37°C. At least three clones were selected and sequenced at the OSA Sequencing Facility at SUNY Stony Brook (http://www.osa.sunysb.edu/dna/). Colonies containing the insert were also confirmed by restriction endonuclease digestion.

Plasmids bearing the expression cassette were then transformed into *Agrobacterium tumefaciens* strain EHA105 by heat shocking at 37°C for 5 minutes

followed by a 2 hour recovery at 28°C in 900µl LB. Bacteria were plated on LB/agar/spectinomycin and incubated at 28°C until colonies appeared. Colonies were inoculated in LB/spectinomycin media and incubated with continuous shaking until the cultures reached an OD of 0.6. Cultures were then collected and resuspended in 10mM MES and 20µM acetosyringone and treated at room temperature for at least 3 hours with continuous shaking. Infiltration was performed by gently appressing a flat tip syringe to the abaxial epidermis of the leaf. Plants were maintained in growth chambers for 3 days until fluorescence was visualized.

Confocal imaging

Bombarded cells and/or infiltrated cells were imaged using a Zeiss Laser Scanning Microscope LSM 5 Pascal with objective lens C-Aprochromat 63x/1.2W. Channel specification was as follows: Single Track (YFP track): Argon laser; Excitation: Line active 514; transmission between 2-4%; Main beam splitter 1: 415/548; Beam Splitter 2: 635; BP 530-600; Detector Gain approximately 700; Amplifier Offset: approximately -1.2.

Time lapsed images of particle movement was obtained using the above criteria and setting the time lapse for 2 seconds recording continuously for 20 seconds with 0.00 seconds between scan intervals.

Latrunculin A and Brefeldin A treatments

Leaf sections were treated with either a 2µM Latrunculin A solution in 10mM MES, or with a 50µM Brefeldin A solution in 10mM MES, delivered by infiltration using a flat tipped syringe as per the protocol described by Haupt *et. al*. (Haupt, 2008).

Cells were imaged approximately 2 hours after infiltration. Negative control was infiltration using 10μL DMSO in 1mL 10mM MES.

RESULTS

Subcellular localization of AT3-1 in Nicotiana benthamiana leaves

A series of vectors was created to test the subcellular localization of AT3-1 (Figure 4.1). Confocal imaging of a fusion construct expressing AT3-1 in translational fusion to the C terminus of full-length YFP reveals that AT3-1 is expressed in the highly reticulated pattern associated with the endoplasmic reticulum (ER) (Figure 4.2.E, .F, .G, using the construct from Figure 4.1.B). YFP-AT3-1 expression was observed concentrated around the nucleus, possibly on the nuclear envelope, which is known to be continuous with the ER membrane (Figure 4.2.C). Protein expression was not always uniform and small protein particles were observed in motion within the cell. Time lapsed images of protein particle movement reveal that YFP-AT3-1 moves in the pattern associated with endomembrane reorganization (Figure 4.2.H to .O) (Runions *et al*, 2006). Particles moved at a velocity of approximately 0.626 $\mu Msec^{-1}$ with standard deviation of ± 0.169 $\mu Msec^{-1}$, consistent with what has previously been observed for ER membrane movement visualized with photoactivatable GFP (Haupt, 2008).

Figure 4.1. Constructs used in this study.

(A) Features of the vector used in this study include the 2X CaMV 35S promoter sequence, the TEV leader sequence that enhances translational efficiency, and the CaMV 35S transcriptional terminator sequence. The multicloning site including the restriction endonucleases used for cloning is denoted by an inverted triangle. The location of the DFGWG site is noted with an arrow.
(B) YFP-AT3-1
(C) YFP-AT3-1 lacking terminal 65 amino acids (YFP-AT3-1Δ65) denoted by an inverted triangle
(D) YFP-AT3-1 lacking terminal 70 amino acids (YFP-AT3-1Δ70) denoted by an inverted triangle

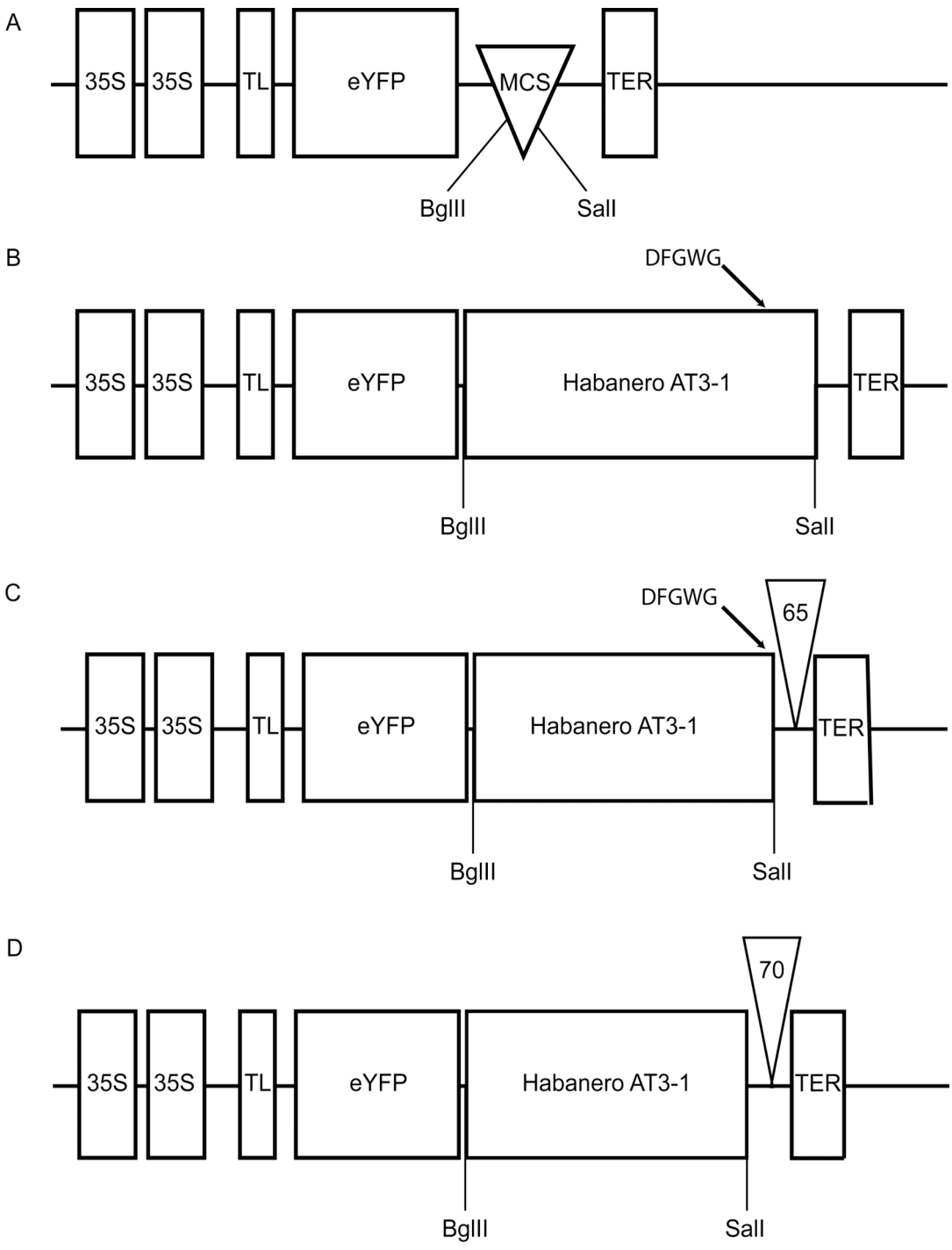
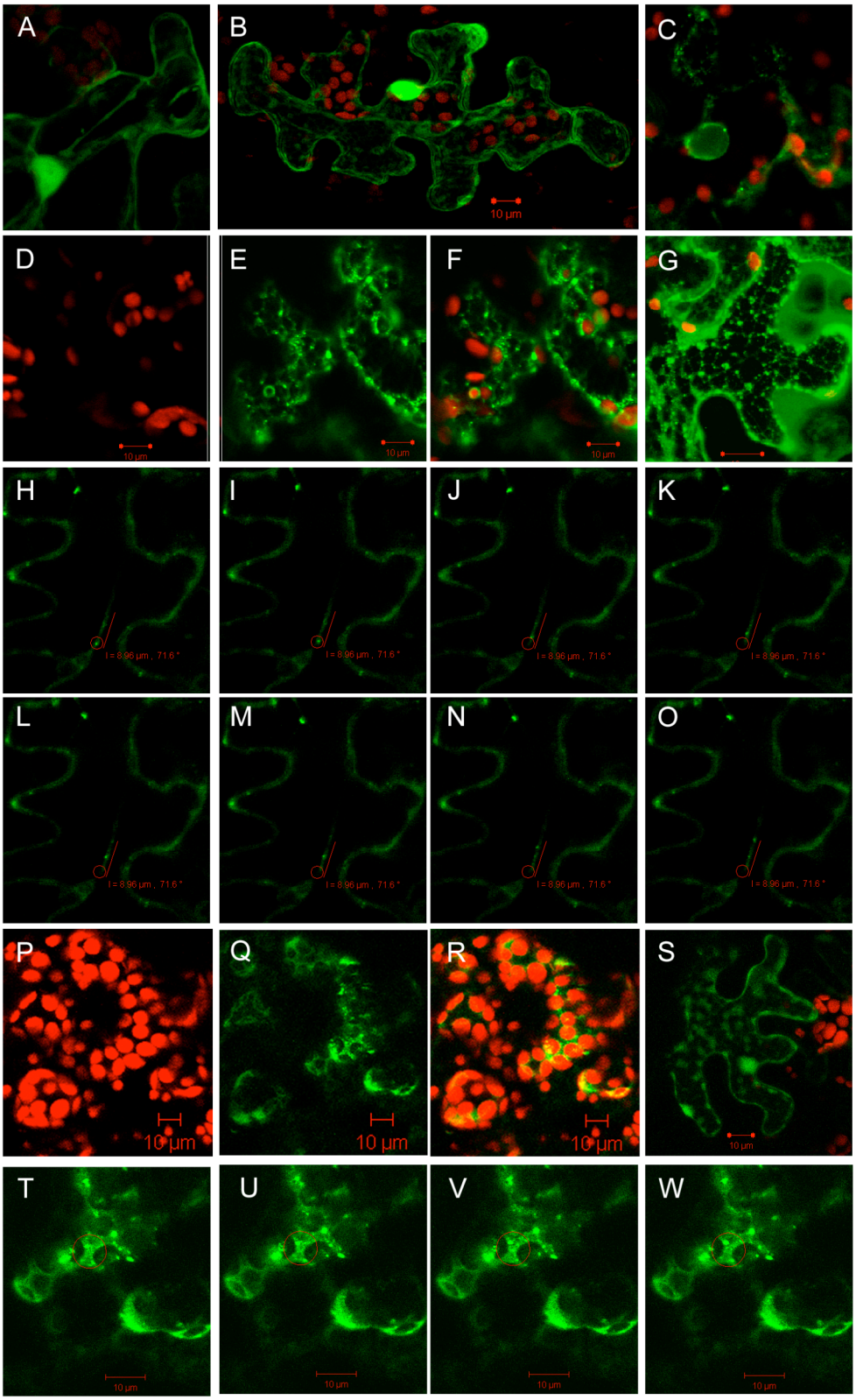

A
35S
35S
TL
eYFP
MCS
TER
BglII
SalI
B
DFGWG
35S
35S
TL
eYFP
Habanero AT3-1
TER
BglII
SalI
C
DFGWG
65
35S
35S
TL
eYFP
Habanero AT3-1
TER
BglII
SalI
D
70
35S
35S
TL
eYFP
Habanero AT3-1
TER
BglII
SalI

Figure 4.2. Subcellular localization of YFP-AT3-1 in *N. benthamiana* leaves.

All panels in this figure show the expression in young *N. benthamiana* leaves of the construct from figure 4.1.B cloned into binary vector pPZP250 and transformed using *Agrobacterium tumefaciens* mediated infiltration. YFP fluorescence is false colored green. Chloroplast autofluorescence is false colored red. Red line or bracket is 10μM in length. In panels (H-O and T-W) red circle denotes starting location for time lapsed images of showing protein particle movement. Line indicates direction of particle movement; angle and distance traveled are noted as well.
(A) Free-YFP control. Detail of nucleus showing uniform accumulation of YFP in the nucleus.
(B) Free YFP control showing cytosolic and nuclear subcellular localization.
(C) YFP-AT3-1 single confocal section showing protein expression around the nuclear envelope. Note the presence of particles in the upper left corner of the image.
(D-F) YFP-AT3-1 single confocal section showing reticulated pattern associated with ER localization. (D) autofluorescence, (E)YFP-AT3-1, (F) merged image.
(G) Projection of an 8μM section illustrating YFP-AT3-1 localization in a reticulated network pattern characteristic of ER localization.
(H-O) Time lapsed images of YFP-AT3-1 particle movement. (H) indicates starting location, (O) final particle location. Total time elapsed is approximately 8 seconds.
(P-W) Images illustrating YFP-AT3-1 appearance after treatment with Latrunculin A.
(P-Q) Single confocal section showing YFP-AT3-1 localization in lamellar sheets. (P) autofluorescence, (Q) YFP-AT3-1, (R) merged image.
(S) Projection of a 10μM section illustrating loss YFP-AT3-1 reticulated pattern.
(T-W) Time lapsed images showing loss of YFP-AT3-1 particle movement. (T) indicates starting location, (W) final particle location. Total time elapsed is approximately 16 seconds.


A
B
10 µm
C
D
10 µm
E
10 µm
F
10 µm
G
H
I = 8.96 µm , 71.6 °
I
I = 8.96 µm , 71.6 °
J
I = 8.96 µm , 71.6 °
K
I = 8.96 µm , 71.6 °
L
I = 8.96 µm , 71.6 °
M
I = 8.96 µm , 71.6 °
N
I = 8.96 µm , 71.6 °
O
I = 8.96 µm , 71.6 °
P
10 µm
Q
10 µm
R
10 µm
S
10 µm
T
10 µm
U
10 µm
V
10 µm
W
10 µm

In order to confirm AT3-1 localization to the ER, leaves expressing YPF-AT3-1 were treated with Latrunculin A, which inhibits actin polymerization and has been shown to cause endomembrane reorganization (Haupt, 2008; Runions *et al*, 2006; Spector *et al*, 1983). Following a 2 hour treatment with Latrunculin A cells no longer showed the reticulated pattern characteristic of the ER (Figure 4.2 Q, R, S). Fluorescence was more diffuse appearing in lamellar sheets rather than in thin strand-like structures. The effect of Latrunculin A on YFP-AT3-1 expression is consistent with the pattern observed for other ER associated proteins (Krishnamurthy *et al*, 2003). Furthermore particle movement was no longer evident in Latrunculin A treated cells, indicating that movement associated with endomembrane reorganization was no longer active as might be expected following actin depolymerization (Figure 4.2 T, U, V, W). These effects were not observed when cells were treated with the negative control (data not shown). Furthermore, particle movement was not halted when cells were treated with Brefeldin A, which disrupts Golgi body formation indicating that the protein particles observed are not associated with vesicles but are part of a dynamic membrane system (data not shown) (Staehelin and Driouich, 1997).

The role of the DFGWG motif

The DFGWG motif is characteristic of the BAHD acyltransferase family; it is putatively involved in maintaining protein stability and has been shown to affect the subcellular localization of a BAHD malonylacyltransferase from *Medicago truncatula* (Ma *et al*, 2005; St-Pierre and De Luca, 2000; Yu *et al*, 2008). To test whether the conserved DFGWG motif plays a role in the subcellular localization of AT3-1, deletions constructs were created which lacked the terminal 65 amino acids of AT3-1 but retained the DFGWG motif; constructs were also created which lacked the

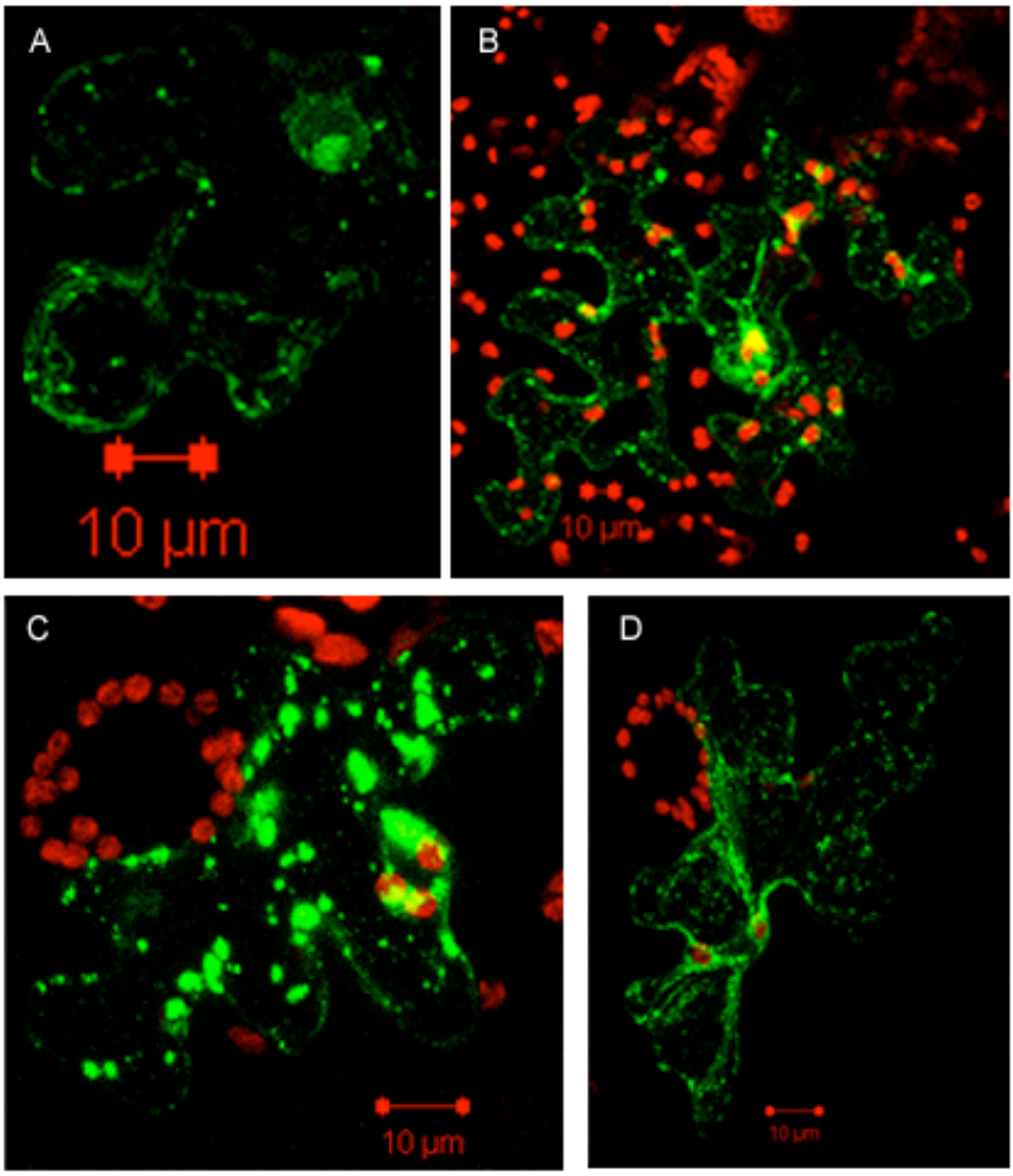
terminal 70 amino acids of AT3-1 and thus included a deletion of DFGWG motif (Figure 4.1.C, .D). Localization of AT3-1 delivered by particle bombardment for the construct lacking the terminal 65 amino acids but retaining the DFGWG motif, is similar to the pattern seen with full-length AT3-1 (compare Figure 4.3.A, with .B, and .D). In contrast AT3-1 lacking the terminal 70 amino acids, which results in a deletion of the DFGWG motif, results in the appearance of non-specific protein aggregates within the cell (Figure 4.3.C). The protein aggregates observed in Figure 4.3.C were immobile and non-uniform in size, suggesting that they may be a result of protein misfolding or protein mislocalization. This finding is consistent with what has previously been observed for other BAHD acyltransferases (Yu *et al*, 2008).

DISCUSSION

AT3-1 is critical for capsaicinoid accumulation in cultivars belonging to the *Capsicum annuum-chinense-frutescens* complex (Stewart *et al*, 2005; Stewart *et al*, 2007, and data presented herein). However the function of this protein in the capsaicinoid biosynthetic pathway is as of yet unknown. It has been suggested that AT3-1 may be capsaicinoid synthase, as this enzyme, the terminal step in the capsaicinoid biosynthetic pathway, would catalyze an acyltransferase step (Stewart *et al*, 2005). However, capsaicinoid synthase activity has been identified in the vacuole of both pungent and non-pungent peppers cultivars of *C. annuum*, suggesting that AT3-1 is not a valid candidate for capsaicinoid synthase as no *AT3-1* mRNA or protein can be detected in plants bearing the *pun1* deletion (Fujiwake *et al*, 1982; Stewart *et al*, 2005). The results presented in this work support the latter assertion by demonstrating that AT3-1 localization is to the endoplasmic reticulum (ER) and not to the vacuole as would be expected for capsaicinoid synthase.

Figure 4.3. Subcellular localization of YFP-AT3-1 and constructs bearing C terminal deletions of AT3-1 in *N. benthamiana* leaves.

All panels in this figure show the expression of the constructs from figure 4.1.B, .C, .D in young *N. benthamiana* leaves delivered using particle bombardment. YFP fluorescence is false colored green. Chloroplast autofluorescence is false colored red. Red line or bracket is 10μM in length. All panels in this image represent projections of 8 and 10μM confocal sections.
(A) YFP-AT3-1 showing protein expression around the nuclear envelope. Note the presence of small particles in the upper left corner of the image.
(B) YFP-AT3-1 showing reticulated pattern associated with ER localization.
(C) YFP-AT3-1Δ70 showing loss of reticulated expression pattern and presence or large protein aggregates.
(D) YFP-AT3-1Δ65 showing reticulated expression pattern characteristic of ER localization.


A
10 µm
B
10 µm
C
10 µm
D
10 µm

Metabolon formation has been proposed to serve several functions in plant secondary metabolic pathways, such as allowing for the sequestration of toxic intermediates and increasing enzymatic efficiency (Jorgensen *et al*, 2005). Since their initial identification over thirty years ago, evidence for such complexes has continued to emerge from studies of the flavonoid pathway, where enzymes catalyzing consecutive biosynthetic reactions are found in close associations with each other (Winkel-Shirley, 1999). This has been demonstrated for chalcone synthase (CHS) and chalcone isomerase (CHI) which both co-localize to the ER (Saslowsky and Winkel-Shirley, 2001). Similarly, PAL and C4H also associate on the ER membrane (Achnine *et al*, 2004). Evidence has also been presented suggesting that the synthesis of the cyanogenic glucoside dhurrin, from *Sorghum bicolor*, requires metabolon formation and that in the absence of one interacting partner subcellular localization is disrupted (Nielsen *et al*, 2008)

Given these results that show AT3-1 localization to the ER membrane it seems possible that AT3-1 may be participating in an enzyme complex as ER associated membrane complexes are common in certain plant secondary metabolic pathways, including the phenylpropanoid biosynthetic pathway (Winkel-Shirley, 1999, Achnine *et al*, 2004). Although such a complex has not yet been demonstrated for capsaicinoid biosynthesis, from its presence would follow a number of implications, several of which have been confirmed. Presence of enzymatic complexes would imply that free metabolite pools would not be present in the cell as intermediate metabolites would be immediately shuttled to the next biosynthetic step, and that entry points into the metabolic pathway would be restricted to the first biosynthetic steps (Winkel-Shirley, 1999). Research on capsaicinoid biosynthesis has produced evidence confirming these two implications. Bennett and Kirby, who were among the first to investigate capsaicinoid biosynthesis, reported that highest amounts of labeled capsaicinoids were

obtained when radioactively tagged phenylalanine was used in feeding experiments; lower levels of labeled capsaicinoids were produced when tagged intermediates were used as precursors (Bennett and Kirby, 1968). Furthermore, pathway intermediates have not been identified in free acid form in any appreciable quantity in pungent peppers, although they have been identified as glycosylated intermediates (Sukrasno and Yeoman, 1993). Taken together these data suggest that capsaicinoid biosynthesis may occur via an enzyme complex possibly localized on the surface of the ER.

If AT3-1 were to participate in metabolon formation it remains entirely unclear what role this protein might play in the capsaicinoid pathway. Given its localization to the ER, it is tempting to speculate that it may catalyze an acyltransferase step in the phenylpropanoid pathway. Besides capsaicinoid synthase, which seems an unlikely role for AT3-1, one candidate biosynthetic step is the conversion of *p*-coumaroyl-CoA to *p*-coumaroyl-shikimate catalyzed by the enzyme HCT; in *N. tabacum*, this enzymatic reaction is catalyzed by a BAHD acyltransferase (Hoffmann *et al*, 2003). Another possibility is that AT3-1 is not directly involved in the acylation of capsaicinoid intermediates, but is necessary for enabling complex formation on the ER. At this point, such as assertion remains highly speculative and requires further investigation.

The BAHD family of acyltransferases was identified based on the presence of two shared motifs, the HXXXDG active site and the DFGWG motif (St-Pierre and De Luca, 2000). The latter motif is removed from the catalytic site of the enzyme as determined by the crystal structure of the protein vinorine synthase, the first BAHD acyltransferase for whom a crystal structure is available (Ma *et al*, 2005). However, its role in enzyme function is obviously critical as a deletion of even a single amino acid can significantly reduce enzyme activity (Bayer *et al*, 2004; Suzuki *et al*, 2003; Unno *et al*, 2007). Thus it was concluded that the DFGWG motif may be important for

maintaining protein conformation (Ma *et al*, 2005; Unno *et al*, 2007). Furthermore, this motif has been shown to be essential for correct localization of a malonyltransferase, MtMaT1 from *Medicago truncatula* as deletion of the motif results in protein aggregation (Yu *et al*, 2008). Our results are consistent with the results of previous studies showing that the DFGWG motif is important for correct protein localization. In the absence of the DFGWG motif, AT3-1 forms non-specific protein aggregates whereas when the terminal 65 amino acids are deleted but the DFGWG motif is retained, correct protein localization to the ER is observed.

Despite the apparent consistency with the results from Yu *et*. *al*. showing the importance of the DFGWG motif for the localization of both AT3-1 and of the malonyltransferase MtMaT1, the results of this study show a different subcellular localization than what has previously been demonstrated for BAHD acyltransferases. Based on the absence of any localization signals, St-Pierre and DeLuca suggested that BAHD acyltransferases would be cytosolically localized (St-Pierre and De Luca, 2000). However, MtMaT1 from *Medicago truncatula* localizes to the nucleus and the cytosol, whereas CER2, a BAHD acyltransferase involved in epicuticular wax elongation in maize, localizes to the nucleus (Xia *et al*, 1997; Yu *et al*, 2008). The paucity of data exploring the subcellular localization of this class of proteins precludes assertions about the uniqueness of this results presented in this study.

CONCLUSIONS

This study shows that AT3-1 localizes to the endoplasmic reticulum, suggesting that AT3-1 is not capsaicinoid synthase as vacuolar subcellular localization would be expected for the terminal enzyme in the capsaicinoid biosynthetic pathway. This localization can be disrupted through treatment of cells with Latrunculin A,

which causes actin depolymerization and a reorganization of ER associated proteins. Furthermore this study finds the DFGWG motif, which is characteristic of the BAHD family of acyltransferases, to be important for the correct localization of AT3-1. This report is in contrast with previous studies that show that BAHD acyltransferases localize to the nucleus and cytosol. In conclusion the results of this study are suggestive of AT3-1's role in the phenylpropanoid pathway, an essential component of capsaicinoid biosynthesis and which is found to be associated with the ER in many plant species.

ACKNOWLEDGMENTS

We would like to thank the members of the Jahn and Citovsky labs for access to prepublication data and useful discussions, especially Michael Mazourek, Alex Krichevsky, Adi Zaltsman, Lisa Zalepa. Shoko Ueki, Benoit Lacroix and Guowei Tian provided expertise with confocal microscopy. This work was supported by an NSF GRF to G.M.S. as well as additional support from the Dean's fund at SUNY Stony Brook.

WORKS CITED

Achnine L, Blancaflor EB, Rasmussen S, Dixon RA (2004). Colocalization of L-phenylalanine ammonia-lyase and cinnarnate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell* **16**(11)**:** 3098-3109.

Aluru MR, Mazourek M, Landry LG, Curry J, Jahn M, O'Connell MA (2003). Differential expression of fatty acid synthase genes, *Acl*, *Fat* and *Kas*, in *Capsicum* fruit. *Journal of Experimental Botany* **54**(388)**:** 1655-1664.

Bayer A, Ma XY, Stockigt J (2004). Acetyltransfer in natural product biosynthesis - functional cloning and molecular analysis of vinorine synthase. *Bioorganic and Medicinal Chemistry* **12**(10)**:** 2787-2795.

Bennett DJ, Kirby GW (1968). Constitution and Biosynthesis of Capsaicin. *Journal of the Chemical Society C-Organic*(4)**:** 442-446.

Eshbaugh WH (1980). The taxonomy of the genus *Capsicum* (Solanaceae). *New Phytol* **47**(2)**:** 153-166.

Estrada-Berta, Bernal Maria A, Diaz J, Pomar F, Merino F (2002). Capsaicinoids in vegetative organs of *Capsicum annuum* L. in relation to fruiting. *Journal of Agricultural and Food Chemistry* **50**(5): 1188 1191.

Fujiwake H, Suzuki T, Iwai K (1980). Intracellular localization of capsaicin and its analogs in *Capsicum* fruit II. The vacuole as the intracellular accumulation site of capsaicinoid in the protoplast of *Capsicum* fruit. *Plant and Cell Physiology* **21**(6)**:** 1023-1030.

Fujiwake H, Suzuki T, Iwai K (1982). Intracellular distributions of enzymes and intermediates involved in biosynthesis of capsaicin and its analogs in *Capsicum* fruits. *Agricultural and Biological Chemistry* **46**(11)**:** 2685-2690.

Govindarajan VS (1986). *Capsicum* - Production, technology, chemistry and quality. 2. Processed products, standards, world production and trade. *Crc Critical Reviews in Food Science and Nutrition* **23**(3)**:** 207-288.

Hadacek F (2002). Secondary metabolites as plant traits: Current assessment and future perspective. *Critical Reviews in Plant Sciences* **21**(4)**:** 273-322.

Hajdukiewicz P, Svab Z, Maliga P (1994). The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Molecular Biology* **25**(6)**:** 989-994.

Harwood JL (1988). Fatty-acid metabolism. *Annual Review of Plant Physiology and Plant Molecular Biology* **39:** 101-138.

Haupt SZ, Angelika, Torrance, Lesley (2008). Localization of Viral Proteins in Plant Cells: Protein Tagging. In: Foster GD, Johansen, I. E., Hong, Y., Nagy, P. D. (ed) *Methods in Molecular Biology, Vol, 451, Plant Virology Protocols: From Viral Sequence to Protein Function*. Humana Press: Totowa, NJ, pp. 463-473.

Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M (2003). Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *Journal of Biological Chemistry* **278**(1)**:** 95-103.

Iwai K, Suzuki T, Fujiwake H (1979). Formation and accumulation of pungent principle of hot pepper fruits. Capsaicin and its analogs in *Capsicum annuum* var. *annuum* cultivar Karayatsubusa at different growth stages after flowering. *Agricultural and Biological Chemistry* **43**(12)**:** 2493-2498.

Jorgensen K, Rasmussen AV, Morant M, Nielsen AH, Bjarnholt N, Zagrobelny M *et al* (2005). Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Current Opinion in Plant Biology* **8**(3)**:** 280-291.

Krishnamurthy K, Heppler M, Mitra R, Blancaflor E, Payton M, Nelson RS *et al* (2003). The Potato virus X TGBp3 protein associates with the ER network for virus cell-to-cell movement. *Virology* **309**(1)**:** 135-151.

Leete E, Louden MCL (1968). Biosynthesis of capsaicin and dihydrocapsaicin in *Capsicum frutescens*. *Journal of the American Chemical Society* **90**(24)**:** 6837-6841.

Ma XY, Koepke J, Panjikar S, Fritzsch G, Stockigt J (2005). Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *Journal of Biological Chemistry* **280**(14)**:** 13576-13583.

Nielsen KA, Tattersall DB, Jones PR, Moller BL (2008). Metabolon formation in dhurrin biosynthesis. *Phytochemistry* **69**(1)**:** 88-98.

Ohta Y (1962). Physiological and genetical studies on the pungency of *Capsicum*. 3. Development of capsaicin in *Capsicum* species. *Japanese Journal of Genetics* **37**(1)**:** 86-90.

Ro DK, Mah N, Ellis BE, Douglas CJ (2001). Functional characterization and subcellular localization of poplar (*Populus trichocarpa* x *Populus deltoides*) cinnamate 4-hydroxylase. *Plant Physiology* **126**(1)**:** 317-329.

Runions J, Brach T, Kuhner S, Hawes C (2006). Photoactivation of GFP reveals protein dynamics within the endoplasmic reticulum membrane. *Journal of Experimental Botany* **57**(1)**:** 43-50.

Saslowsky D, Winkel-Shirley B (2001). Localization of flavonoid enzymes in Arabidopsis roots. *Plant Journal* **27**(1)**:** 37-48.

Schopker H, Kneisel M, Beerhues L, Robenek H, Wiermann R (1995). Phenylalanine ammonia-lyase and chalcone synthase in glands of *Primula kewensis* (W. Wats): immunofluorescence and immunogold localization. *Planta* **196**(4)**:** 712-719.

Schwab W, Davidovich-Rikanati R, Lewinsohn E (2008). Biosynthesis of plant-derived flavor compounds. *Plant Journal* **54**(4)**:** 712-732.

Spector I, Shochet NR, Kashman Y, Groweiss A (1983). Latrunculins - novel marine toxins that disrupt microfilament organization in cultured cells. *Science* **219**(4584)**:** 493-495.

St-Pierre B, De Luca V (2000). Evolution of acyltransferase genes: origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. *Recent Advances in Phytochemistry* (Evolution of Metabolic Pathways) **34:** 285-315.

Staehelin LA, Driouich A (1997). Brefeldin A effects in plants - Are different Golgi responses caused by different sites of action? *Plant Physiology* **114**(2)**:** 401-403.

Stewart C, Kang BC, Liu K, Mazourek M, Moore SL, Yoo EY *et al* (2005). The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. *Plant Journal* **42**(5)**:** 675-688.

Stewart C, Mazourek M, Stellari GM, O'Connell M, Jahn M (2007). Genetic control of pungency in *C. chinense* via the *Pun1* locus. *Journal of Experimental Botany* **58**(5)**:** 979-991.

Sukrasno N, Yeoman MM (1993). Phenylpropanoid metabolism during growth and development of *Capsicum frutescens* fruits. *Phytochemistry* **32** (4) 839 844.

Suzuki H, Nakayama T, Nishino T (2003). Proposed mechanism and functional amino acid residues of malonyl-CoA : anthocyanin 5-O-glucoside-6 '"-O-malonyltransferase from flowers of *Salvia splendens*, a member of the versatile plant acyltransferase family. *Biochemistry* **42**(6)**:** 1764-1771.

Suzuki T, Fujiwake H, Iwai K (1980). Intracellular localization of capsaicin and its analogs capsaicinoid in *Capsicum* fruit 1. microscopic investigation of the structure of the placenta of *Capsicum annuum* var. *annuum* cultivar Karayatsubusa. *Plant and Cell Physiology* **21**(5)**:** 839-853.

Suzuki T, Iwai, K. (1984). Constituents of Red Pepper Species: Chemistry, Biochemistry, Pharmacology, and Food Science of the Pungent Principle of *Capsicum* species *The Alkaloids*. The Academic Press. Vol. XXIII.

Tandon GL, Siddappa GS, Dravid SV (1964). Oleoresin of *Capsicum* (red chilies): some technological and chemical aspects. *Journal of Food Science* **29**(1)**:** 1-&.

Tzfira T, Tian GW, Lacroix B, Vyas S, Li JX, Leitner-Dagan Y *et al* (2005). pSAT vectors: a modular series of plasmids for autofluorescent protein tagging and expression of multiple genes in plants. *Plant Molecular Biology* **57**(4)**:** 503-516.

Unno H, Ichimaida F, Suzuki H, Takahashi S, Tanaka Y, Saito A *et al* (2007). Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. *Journal of Biological Chemistry* **282**(21)**:** 15812-15822.

Winkel-Shirley B (1999). Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. *Physiologia Plantarum* **107**(1)**:** 142-149.

Xia YJ, Nikolau BJ, Schnable PS (1997). Developmental and hormonal regulation of the Arabidopsis *CER2* gene that codes for a nuclear-localized protein required for the normal accumulation of cuticular waxes. *Plant Physiology* **115**(3)**:** 925-937.

Yu XH, Chen MH, Liu CJ (2008). Nucleocytoplasmic-localized acyltransferases catalyze the malonylation of 7-O-glycosidic (iso)flavones in *Medicago truncatula*. *Plant Journal* **55**(3)**:** 382-396.

Zamski E, Shoham O, Palevitch D, Levy A (1987). Ultrastructure of capsaicinoid-secreting cells in pungent and nonpungent red pepper *Capsicum annuum* L. cultivars. *Botanical Gazette* **148**(1)**:** 1-6.

CHAPTER 5: CONCLUDING REMARKS

The purpose of this study was to explore the genetic basis of loss of pungency in domesticated and wild species in the genus *Capsicum*, and to explore the evolution and function of the candidate gene for capsaicinoid biosynthesis, *AT3-1*. A collection of different methods, techniques, and approaches were employed, and resources such as molecular markers were taken from model organism systems, such as tomato and adapted for use in pepper, a non-model system. This allowed the studies presented herein to overcome significant limitations, for example by expanding the repertoire of *Capsicum* genotypes utilized in genetic studies involving interspecific crosses beyond the narrow set of the closely related cultivated varieties *C. annuum*, *C. chinense* and *C. frutescens*. This chapter illustrates the extensions and future experimental questions that directly follow from the results presented in this work.

The continued recovery of mutations at a single locus that we have observed associated with loss of pungency among cultivated varieties of *Capsicum*, is an unusual occurrence among plant secondary metabolic pathways. More commonly mutations in plant secondary metabolic pathways occur at multiple pathway intermediate steps and are identified by the accumulation of the biosynthetic intermediates associated with each sequential enzymatic metabolic product interconversion. An analysis of fruit color loci in pepper and a comparison of their map locations to known tomato fruit color loci shows that different fruit pigmentation phenotypes can be associated with structural genes in the carotenoid biosynthetic pathway (Thorup *et al*, 2000). In contrast to the evidence from the capsaicinoid biosynthetic pathway, multiple mutations along the carotenoid biosynthetic pathway have been identified which lead to the accumulation of carotenoid precursors and different fruit pigmentations and colors (Thorup *et al*, 2000). In stokes aster, natural

variation in flower color can be associated with two sequential mutations in the flavonoid and anthocyanin biosynthetic pathways (Barb *et al*, 2008). These lead to either the absence of both flavonoid and anthocyanin endproduct accumulation or the accumulation of flavonoid biosynthetic intermediates, and therefore different flower color phenotypes (Barb *et al*, 2008). Empirical studies combined with theoretical approaches suggested that the biosynthetic pathway model where sequential enzymatic interconversions catalyzed by biosynthetic enzymes sharing interconnected pools of intermediates holds quite well for many plant secondary metabolic pathways (Kacser and Burns, 1981).

Capsaicinoid biosynthesis, instead, appears to be unusual among secondary metabolic pathways as the mutations so far identified that result in the absence of capsaicinoid biosynthesis are allelic. Furthermore, these mutations reveal no intermediate biosynthetic precursor accumulation (De Marino *et al*, 2006). The elusive nature of capsaicinoid biosynthesis and the inaccessible enzymatic steps of this pathway have led to speculation and premature conclusions about the identity of the terminal enzyme in the capsaicinoid biosynthetic pathway, capsaicinoid synthase. The terminal biosynthetic step in the capsaicinoid biosynthetic pathway has been the subject of investigation for over three decades (Fujiwake *et al*, 1982; Stewart *et al*, 2005; Stewart *et al*, 2007; Suzuki, 1984). Recently a gene, *CSY1*, was identified and reported to be capsaicinoid synthase; it also appeared to represent the identification of a novel class of plant acyltransferase proteins (Prasad *et al*, 2006). Later investigations showed that CSY1 is likely a leucine rich repeat (LRR) putative protein kinase cloned out of frame, and that the enzymatic assay characterizing this protein's function was flawed and failed to use appropriate controls to test capsaicinoid synthase activity. The limitations of the study finally resulted in the paper's retraction (Prasad *et al*, 2008). To date the *Pun1* locus and its candidate gene *AT3-1* remains the only

structural gene convincingly associated with loss of capsaicinoid biosynthetic ability in pepper.

Allelic diversity at the *Pun1* locus was investigated in this study using classical genetic methods. A basic tool which has been essential to the study of plant developmental and metabolic pathways, and which has been deployed with great success since its discovery, is the complementation test (Benzer, 1955; Jendestrid, 1993; Jurgens *et al*, 1991). The complementation test was used successfully in this study to show that loss of pungency in the wild species *Capsicum chacoense* is caused by a mutation at a different locus than the mutations responsible for loss of pungency among cultivated species in the *annuum-chinense-frutescens* species complex. This result suggests that there may be differences in selection pressures in the environments experienced by wild and domesticated species thereby resulting in the canalization of capsaicinoid biosynthesis such that only mutations at a single locus may be recovered among domesticated species.

Although the hypothesis proposed above is difficult to test, one could imagine investigating the adaptive value of the *Pun1* and *Pun2* loci in the field, thus measuring their ecological effects. Such an investigation has been carried out successfully in monkeyflower (*Mimulus spp*.) (Bradshaw *et al*, 1995, Bradshaw and Schemske, 2003). Mapping of traits associated with pollinator preferences, have revealed that in monkeyflower a number of QTL are associated with differences in flower and petal characteristics (Bradshaw *et al*, 1995). Near-isogenic lines (NILs) were then used to show that substitutions at a single locus, *YUP*, could shift the pollination syndromes in the field from bumblebee pollinated, associated with the yellow form of *YUP*, to hummingbird pollinated, associated with the red form of *YUP* (Bradshaw and Schemske, 2003). Were such near-isogenic lines to be developed for the *Pun1* and *Pun2* loci, one could conduct field tests able to discern the different adaptive values of

these alleles by measuring variables that are of ecological interest for each NIL, such as seed production and germination rates, capsaicinoid accumulation, seed dispersal range and rates of mammalian predation of pungent versus non-pungent fruits.

Continued recovery of the *Pun1* locus as the locus associated with loss of pungency among domesticated species of *Capsicum* suggests that further investigations into the genetic control of the capsaicinoid biosynthetic pathway should focus on wild species within the genus. An obvious next step is then to incorporate non-pungent individuals identified among the wild *C. eximium* populations in Bolivia, as this represents the only other documented case besides the polymorphic populations of *C. chacoense* used in the study in Chapter 2, where wild populations contain individuals which accumulate very low levels of capsaicinoids (Tewksbury *et al*, 2006). Individuals from these *C. eximium* populations should be incorporated into the complementation test series described in Chapter 2. Placement of the *C. eximium* loss of pungency allele/s into the correct complementation test groups may identify further loci allowing for loss of pungency.

Continued identification of loci involved in modulating levels of capsaicinoids in ripe fruit or allowing for capsaicinoid biosynthesis, may uncover novel regulatory and/or biosynthetic steps in the capsaicinoid biosynthetic pathway. Such an analysis followed by a detailed mapping study using crosses with greater interspecific fertility than those used in this study may allow for candidate gene analysis and cloning of the locus or QTL responsible for loss of pungency among the wild species of *Capsicum*. The identification of genes responsible for the interconversion of capsaicinoid biosynthetic intermediates could clarify the evolutionary steps leading to capsaicinoid biosynthesis and potentially allow the recreation of the pathway in a heterologous expression system.

Although naturally occurring variation is a powerful and readily available tool

that can be used with great efficacy to identify loci contributing to loss of pungency, thanks to extensive germplasm collections (available through the USDA-ARS) and international enthusiasm for cataloguing varieties of pungent peppers, this resource continues to limit studies on loss of pungency to the identification of mutations which have reached equilibrium in natural populations or which have been selected for under artificial selection for agronomically viable traits. In contrast, a saturated mutagenesis screen for loss of pungency mutants would allow for a complete analysis of the capsaicinoid biosynthetic pathway. *Capsicum's* large genome, its high levels of genomic redundancy, the late development of the pungency phenotype and the sheer size of the infrastructure required for such a screen are clear limitations to undertaking this study. These limitations might be mitigated by selection of appropriate germplasm for such a study. Ideally one might want to select a highly pungent, fast growing, compact, early flowering variety, of which *C. annuum* has several. Despite the difficulty, the potential rewards are the identification of novel loci contributing to loss of pungency that have pleiotropic phenotypes that would make their identification unlikely under other circumstances.

In Chapter 3 of this study it was demonstrated that *AT3*, the candidate gene for the *Pun1* locus, has undergone a tandem gene duplication event that predates the diversification of the solanaceous species used in the study. *AT3-1* and *AT3-2* retain punctuated regions of nucleotide conservation in their coding regions, however their introns cannot be aligned and may be non-homologous. Extensive search of the literature reveals that as of the publication of this work, no other reports are available detailing the presence of non-homologous introns whose intron-exon boundaries are nonetheless entirely conserved. Four obvious questions emerge from this research: 1) when did the *AT3-1*/*AT3-2* duplication event take place, 2) what is the origin of the *AT3-2* intron and 3) what mechanism caused the *AT3-1*/*AT3-2* duplication event and

4) is that related to the presence of the non-homologous intron?

Although sequencing of the *Arabidopsis thaliana* genome has revealed the presence of BAHD acyltransferases whose amino acid sequence is similar to AT3-1, no coding regions were identified among the available sequenced genomes that are similar to either *AT3-1* or *AT3-2* at the nucleotide level. This suggests that the duplication occurred sometime after the diversification of the euasterids and the rosids (Hilu *et al*, 2003). Sequencing of *AT3-1* and/or *AT3-2* from the family that is sister to the Solanaceae, the Convolvulaceae, might elucidate the origin of the gene duplication event which created these two sister gene lineages. Furthermore, from the sequencing of the tomato genome one would expect to identify the genomic location and physical distance separating these two tandem gene duplicates, as we have identified orthologs of both *AT3-1* and *AT3-2* from *Solanum lycopersicum*. This would assist in elucidating the origin of the intron identified in *AT3-2* possibly by revealing the presence of repetitive elements that may be responsible for the gene duplication event.

Sequencing of *AT3-1* coupled with capsaicinoid measurements from a larger number of individuals from the species *C. rhomboideum* may allow us to determine whether the two base pair deletion identified in the coding region of *AT3-1* is widely distributed in populations of this species and whether this mutation is associated with the lack of capsaicinoid accumulation that appears to be characteristic of this species (Eshbaugh, 1980). If possible, it would be instructive to test the relationship of this mutation to loss of pungency by including *C. rhomboideum* into the complementation test panel discussed in Chapter 2 of this work, although finding a suitable genotype to allow for interspecific crosses will likely be a significant limitation.

Finally, using translational fusions with fluorescent protein tags, this work shows that the subcellular localization of *C. chinense* 'Habanero' AT3-1 deviates from the expected subcellular localization for capsaicinoid synthase, and is also different

from the expected subcellular localization for BAHD acyltransferases. Although not generally believed to be capsaicinoid synthase, this result nonetheless shows that AT3-1 associates with the endoplasmic reticulum pointing to this protein's role in the phenylpropanoid pathway. As numerous problems have been encountered with the expression and purification of recombinant AT3-1, the development of transgenic *Nicotiana benthamiana* or *Nicotiana tabacum* plants heterologously expressing YFP-AT3-1, which has been shown to be soluble in this study, would significantly advance the understanding of the role of AT3-1 in the capsaicinoid biosynthetic pathway. One would then be able to test whether AT3-1 and other structural genes in the capsaicinoid biosynthetic pathway whose transcription is coordinately regulated, participate in protein-protein interactions and if this interaction occurs on the surface of the ER. Furthermore, development of such a resource would also allow one to test whether AT3-1 is capsaicinoid synthase. Should development of transgenic *N. benthamiana* or *N. tabacum* plants heterologously expressing *C. chinense* 'Habanero' *AT3-1* prove successful a logical next step would be to test whether ER localization of YPF-AT3-1 is disturbed when only the DFGWG motif is mutated. This would extend the analyses presented herein and test the relationship between the DFGWG motif and the localization of YFP-AT3-1 to the ER as has been observed in this study.

Plant secondary metabolites represent an important factor in the nutritional value of crops that must be taken into account during the breeding and improvement of domesticated plant species (Galili *et al*, 2002). Plant secondary metabolites make an essential contribution to human health and play vital roles in preventing diseases such as cancer (Verpoorte and Memelink, 2002). Micronutrient deficiencies are a significant factor contributing to undernutrition, a collection of diseases that can be as pernicious as low caloric intake, as they result in blindness, stunted growth and impaired development (Horton and Ross, 2003; Qaim *et al*, 2007). Capsaicinoids are a

valuable addition to the diets of many peoples around the world and feature prominently in many cuisines (Govindarajan, 1985). Their use in increasing the palatability of starch-based diets is ancient as revealed by the association among archeological remains between maize starch granules and fossils of domesticated *Capsicum* (Perry *et al*, 2007).

Capsaicinoids are a taxonomically restricted group of plant secondary metabolites. Further research into their biosynthesis, its regulation and the identification of allelic diversity for breeding novel non-pungent varieties of *Capsicum*, is important to the clarification of this elusive plant secondary metabolic pathway and more generally serves as a model for the study of other taxonomically restricted plant secondary metabolites. This study extends our understanding of capsaicinoid biosynthesis through the elucidation of the relationship between domestication and loss of pungency, the identification of allelic diversity at our candidate gene for capsaicinoid biosynthesis and a description of its evolutionary history, and the characterization of its subcellular localization and its relationship to capsaicinoid biosynthesis. Taken together these results suggest that evolution of the capsaicinoid biosynthetic pathway did not necessarily coincide with a coding region mutation or duplication of the sequence of *AT3-1*. Furthermore, multiple alleles of *Pun1* are associated with loss of pungency among the domesticated species of the *C. annuum*, *C. chinense* and *C. frutescens* species complex and the evolution of these alleles predates the diversification of the major domesticated species of *Capsicum* suggesting that existing genetic diversity was exploited in the breeding of non-pungent varieties of *Capsicum*.

WORKS CITED

Barb JG, Werner DJ, Griesbach RJ (2008). Genetics and biochemistry of flower color in stokes aster. *Journal of the American Society for Horticultural Science* **133**(4)**:** 569-578.

Benzer S (1955). Fine structure of a genetic region in bacteriophage. *Proceedings of the National Academy of Sciences of the United States of America* **41**(6)**:** 344-354.

Bradshaw HD, Schemske DW (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* **426**(6963)**:** 176-178.

Bradshaw HD, Wilbert SM, Otto KG, Schemske DW (1995). Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* **376**(6543)**:** 762-765.

De Marino S, Borbone N, Gala F, Zollo F, Fico G, Pagiotti R *et al* (2006). New constituents of sweet *Capsicum annuum* L. fruits and evaluation of their biological activity. *Journal of Agricultural and Food Chemistry* **54**(20)**:** 7508-7516.

Eshbaugh WH (1980). The taxonomy of the genus *Capsicum* (Solanaceae). *New Phytologist* **47**(2)**:** 153-166.

Fujiwake H, Suzuki T, Iwai K (1982). Intracellular distributions of enzymes and intermediates involved in biosynthesis of capsaicin and its analogs in *Capsicum* fruits. *Agricultural and Biological Chemistry* **46**(11)**:** 2685-2690.

Galili G, Galili S, Lewinsohn E, Tadmor Y (2002). Genetic, molecular, and genomic approaches to improve the value of plant foods and feeds. *Critical Reviews in Plant Sciences* **21**(3)**:** 167-204.

Govindarajan VS (1985). *Capsicum*: Production, Technology, Chemistry and Quality. 1. History, Botany, Cultivation and Primary Processing. *Crc Critical Reviews in Food Science and Nutrition* **22**(2)**:** 109-176.

Horton S, Ross J (2003). The economics of iron deficiency. *Food Policy* **28**(1)**:** 51-75.

Hilu KW, Borsch T, Muller K, Soltis DE, Soltis PS, Savolainen V *et al* (2003). Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* **90**(12)**:** 1758-1776.

Jendestrid B (1993). Genetic control of flavonoid biosynthesis in barley. *Hereditas* **119**(2)**:** 187-204.

Jurgens G, Mayer U, Ruiz RAT, Berleth T, Misera S (1991). Genetic analysis of pattern formation in the *Arabidopsis* embryo. *Development***:** 27-38.

Perry L, Dickau R, Zarrillo S, Holst I, Pearsall DM, Piperno DR *et al* (2007). Starch fossils and the domestication and dispersal of chili peppers (*Capsicum spp*. L.) in the Americas. *Science* **315**(5814)**:** 986-988.

Qaim M, Stein AJ, Meenakshi JV (2007). Economics of biofortification. *Agricultural Economics* **37:** 119-133.

Stewart C, Kang BC, Liu K, Mazourek M, Moore SL, Yoo EY *et al* (2005). The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. *Plant Journal* **42**(5)**:** 675-688.

Stewart C, Mazourek M, Stellari GM, O'Connell M, Jahn M (2007). Genetic control of pungency in *C. chinense* via the *Pun1* locus. *Journal of Experimental Botany* **58**(5)**:** 979-991.

Suzuki T, Iwai, K. (1984). Constituents of Red Pepper Species: Chemistry, Biochemistry, Pharmacology, and Food Science of the Pungent Principle of *Capsicum* species *The Alkaloids*. The Academic Press. Vol. XXIII.

Tewksbury JJ, Manchego C, Haak DC, Levey DJ (2006). Where did the chili get its spice? Biogeography of capsaicinoid production in ancestral wild chili species. *Journal of Chemical Ecology* **32**(3)**:** 547-564.

Thorup TA, Tanyolac B, Livingstone KD, Popovsky S, Paran I, Jahn M (2000). Candidate gene analysis of organ pigmentation loci in the Solanaceae. *Proceedings of the National Academy of Sciences of the United States of America* **97**(21)**:** 11192-11197.

USDA, A National Genetic Resources Program. National Germplasm Resources Laboratory, Beltsville, Maryland.

Verpoorte R, Memelink J (2002). Engineering secondary metabolite production in plants. *Current Opinion in Biotechnology* **13** (2): 181 187.

APPENDIX

IDENTIFICATION OF INTERACTORS WITH AT3-1 BY YEAST TWO-HYBRID ASSAY

(Giulia Marina Stellari[10], Vitaly Citovsky[11], and Molly Jahn[12†])

RATIONALE

Chapter 4 of this manuscript reports that AT3-1 tagged at its N terminus with YFP and heterologously expressed in *N. benthamiana* epidermal leaf cells localizes to the endoplasmic reticulum (ER). Several protein complexes in the phenylpropanoid pathway, which provides vanillylamine, a direct precursor of capsaicinoids, have been identified which are associated with the ER, and some of these complexes were initially characterized by yeast two-hybrid assays for protein-protein interactions suggesting that a yeast two-hybrid approach may be a fruitful way of identifying protein-protein interactions in the capsaicinoid biosynthetic pathway (Burbulis and Winkel-Shirley, 1999; Saslowsky and Winkel-Shirley, 2001). Furthermore, the cloning of an *AT3-1* ortholog from *Nicotiana benthamiana* suggests that AT3-1 may have an ancestral role unrelated to capsaicinoid biosynthesis. The high degree of sequence conservation between *AT3-1* from *Capsicum spp.* and *N. benthamiana AT3-1*

[10] Present address: Biochemistry Department, 424 Life Sciences Building, SUNY Stony Brook, Stony Brook NY, 11794.
[11] Present address: Biochemistry Department, 414 Life Sciences Building, SUNY Stony Brook, Stony Brook NY, 11794.
[12] Present address: Department of Agronomy and Genetics, 140 Agriculture Hall, 145 Linden Drive, Madison, WI 53706
[†] Author contributions: GMS designed and executed experiments; all authors read and reviewed the manuscript.

and the known pattern of conservation in gene expression among solanaceous species (Rensink *et al*, 2005), suggests that it might be possible to identify putative interacting proteins with AT3-1 by yeast-two hybrid screening of a whole young plant *N. tabacum* cDNA library. The results of this experiment are reported in this appendix.

MATERIALS AND METHODS

Bait plasmid construction

The bait plasmid used for yeast two-hybrid library screening was constructed by cloning the full length sequence of *AT3-1* from *C. chinense* ''Habanero'' into the bait vector pSTT91 (TRP1+) (Sutton *et al*, 2001) by PCR amplification using gene specific primers with restriction endonuclease adaptor sequences (underlined). PCR was performed as follows: 5µl of 10X PCR Buffer, 4µl of 2.5mM dNTPs, 1µl of 10mM pSTT91-AT3-1 F (CCGCTCGAGCATGGCTTTTGCATTACCATCATCAC), 1µl of 10mM pSTT91-AT3-1 R (ATACTGCAGTCCAAAACTTGGAATTAATTAGGC), .25µl ExTaq (Takara Bio Inc., Otsu, Japan) and cycling with the following conditions: 94°C for 3 minutes, 29 cycles of 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 3 minutes, with a final extension at 72°C for 20 minutes. Reactions were performed using a PTC 225 Peltier Thermal Cycler (MJ Research, Watertown, MA). PCR products were visualized on a 1.0% TAE agarose gel, excised and gel purified using the GFX PCR DNA and Gel Band Purification Kit (GE, Healthcare Biosciences Inc., Piscataway, NJ) as per manufacturer's instructions. Purified gel products were digested with the restriction endonucleases XhoI/PstI in 3µl 10X Buffer 3, 3µl 10X BSA, 22µl purified gel product, 1µl XhoI, 1µl PstI (New England Biolabs, Ipswich, MA) for 3 hours at 37°C.

Approximately 500ng of pSTT91 was digested as per the protocol stated above. Digested PCR products and vector were purified from a 1% TAE agarose gel as stated above. Vector and insert were ligated overnight at 16°C in 1.5µl 10X T4 DNA Ligase Buffer, 1µl T4 DNA Ligase (New England Biolabs, Ipswich, MA) and a mixture of digested PCR products and vector to a final volume of 15µl. Ligation products were transformed into DH5α chemically competent cells (New England Biolabs, Ipswich, MA) by heat shocking at 42°C for 30 seconds followed by a 1 hour recovery at 37°C in 900µl SOC. Positive colonies were identified by plating on LB/agar/Ampicillin incubating overnight at 37°C. At least three clones were selected and sequenced at the OSA Sequencing Facility at SUNY Stony Brook (http://www.osa.sunysb.edu/dna/). A single full-length wildtype clone based on sequence comparison with *C. chinense* ‘‘Habanero’’ *AT3-1* accession number AY819027 was selected for library screening.

Yeast Two-hybrid Assay

Saccharomyces cerevisiae strain L40 (MATa *his3*Δ200 *trp1*-90l *leu2*-3 112 *ade2 lys2-80lam URA3*:(lexAop)4-*HIS3 LYS2* gal80) (provided by Rolf Sternglanz) was grown in yeast extract/peptone/dextrose or the appropriate minimal media using standard conditions (Hollenberg *et al*, 1995; Kaiser, 1994). Plasmids were introduced into yeast cells using a standard lithium acetate protocol (Hollenberg *et al*, 1995). The bait plasmid was tested for self-activation by assaying growth on tryptophan and histidine deficient minimal media. No growth was observed on selective media in the presence of 3-amino-triazole (AT) at a concentration of 0.0 mM, 2.5mM, 5mM, 10mM and 25mM, indicating that *AT3-1* does not self-activate *HIS3* expression.

The prey library was an *N. tabacum* (cv. Turk) cDNA library, which was cloned into the pGAD424 (LEU2+, Clontech) activation domain 2-hybrid cloning

vector (V. Citovsky). Both young and mature leaves were used for the isolation of polyA(+) RNA from healthy plants harvested just prior to flowering. The library was size-fractionated to include clones of 400 bp and greater. The library contains over 2.0 x 10^6 independent clones with an average size of 1.3 kilobases and more than 90% recombinants. The cDNA inserts were directionally cloned as EcoRI-XhoI fragments into EcoRI-SalI sites of pGAD424.

The *N. tabacum* cv. Turk cDNA library in pGAD424 (LEU2+, Clontech) was then screened with *AT3-1* in pSTT91 (TRP1+) (Sutton *et al*, 2001) as bait as previously described (Ballas and Citovsky, 1997; Hollenberg *et al*, 1995). Library transformation efficiency was estimated by plating serial dilutions of yeast co-transformed with bait plasmid and prey library on tryptophan and leucine deficient media. Positive clones were selected on a histidine deficient selective medium and confirmed by β-galactosidase assay (Durfee *et al*, 1993).

False positives were eliminated by testing the specificity of the interaction using pSTT91 (TRP1+) (Sutton *et al*, 2001) empty vector, and pBTM116 (TRP1+) (Hollenberg *et al*, 1995) expressing either human lamin C or topoisomerase I which are known to non-specifically activate reporter gene expression in the yeast two hybrid system (assay A) (Bartel, 1993; Park and Sternglanz, 1998).

Due to the high number of false positive interactions identified using assay A, a secondary method (assay B) of determining the specificity of the interaction of the bait plasmid with the prey library clone was also attempted. In this method, colonies that grew on histidine deficient minimal media and activated the β-galactosidase assay were purged of the bait construct containing *AT3-1* in pSTT91 (TRP1+) (Sutton *et al*, 2001) by repeated rounds of culturing on leucine deficient minimal media followed by assaying colonies' ability to activate the β-galactosidase assay. Once colonies were identified that could not grow on tryptophan deficient media and that could no longer

activate the β-galactosidase assay, the bait construct was re-introduced using a standard lithium acetate protocol (Hollenberg *et al*, 1995). β-galactosidase activity and growth on histidine deficient media was then assayed as described previously. Only those colonies subsequently able to activate the β-galactosidase assay in the presence of the bait construct but not in the presence of pSTT91 (TRP1+) (Sutton *et al*, 2001) empty vector or in the presence of pBTM116 (TRP1+) (Hollenberg *et al*, 1995) expressing either human lamin C or topoisomerase I were selected for subsequent analysis.

Bi-molecular fluorescence complementation

In planta interaction between AT3-1 and two prey plasmids identified as specific interactors by assay B (described above) was tested using Bi-Molecular Fluorescence complementation (BiFC). BiFC vector containing *C. chinense* ‘‘Habanero’’ *AT3-1* fused in frame to the C terminus of the N-terminal fragment of YFP was created by PCR amplification of the full length cDNA clone with primers containing restriction enzyme adaptor sites (BglII/SalI) as described in Chapter 3. The split YFP tag was designed based on the pSAT series of vectors (Tzfira *et al*, 2005).

Full length cDNA clones of the prey sequences identified in the yeast two-hybrid assay were introduced into plasmids for BiFC by PCR amplification using gene specific primers containing restriction endonuclease adaptor sites (underlined). For Histone H3, PCR was performed as per the following protocol: 10µl of 5X PCR High Fidelity Buffer, 1µl of 10mM dNTPs, 1µl of 10mM Histone H3 F (GGAAGATCTATGGCTTGTACCAAGCAAACTGC), 1µl of 10mM Histone H3 R (ACGCGTCGACAGCACGCTCGCCCCTGATACGCCTAGC), 0.5µl Phusion (Finnzymes, Woburn, MA) and cycling with the following conditions: 98°C for 30

seconds, 34 cycles of 98°C for 10 seconds, 55°C for 30 seconds, 72°C for 1 minute 30 seconds, with a final extension at 72°C for 20 minutes. Reactions were performed using a PTC 225 Peltier Thermal Cycler (MJ Research, Watertown, MA). For Histone HTC, PCR was performed as above except that the following primer pair was used: Histone HTC F (GGAAGATCTATGGCTCGTACAAAGCAAACCG) and Histone HTC R (ACGCGTCGACTGGCCTCTCACCCCTAATTCTCC).

PCR products were visualized on a 1.0% TAE agarose gel, excised and gel purified using the GFX PCR DNA and Gel Band Purification Kit (GE, Healthcare Biosciences Inc., Piscataway, NJ) as per manufacturer's instructions. Purified gel products were simultaneously digested in 3µl 10X Buffer 3, 3µl 10X BSA, 22µl purified gel product, 1µl BglII, 1µl SalI (New England Biolabs, Ipswich, MA) for 3 hours at 37°C. Approximately 500ng of pSAT6-N-YFPc1 was digested as per the protocol stated above. Digested PCR products and vector backbone were visualized and purified from a 1% TAE agarose gel. Digested PCR products and vector were ligated overnight at 16°C in 2.0µl 10X T4 DNA Ligase Buffer, 1µl T4 DNA Ligase (New England Biolabs, Ipswich, MA) and a mixture of digested PCR products and backbone to a final volume of 20µl. Ligation products were transformed into DH5α chemically competent cells (New England Biolabs, Ipswich, MA) by heat shocking at 42°C for 30 seconds followed by a 1 hour recovery at 37°C in 900µl SOC. Positive colonies were identified by plating on LB/agar/Ampicillin selective media incubating overnight at 37°C. At least three clones were selected and sequenced at the OSA Sequencing Facility at SUNY Stony Brook (http://www.osa.sunysb.edu/dna/). Full-length wild type clones for both Histone H3 and Histone HTC were selected based on sequence comparison with the NCBI Accessions AB355993.1 and AK246733.1 respectively.

Biolistic transformation

Gold particles were prepared for biolistic transformation as follows. Approximately 2.5μg of each plasmid to be tested by BiFC was affixed to 1μM gold particles (BioRad, Hercules, CA) using 1M $CaCl_2$ followed by 3 washes in 100% Ethanol and final resuspension in 3ml 100% Ethanol with the addition of 7.5μl of a 55.5μM PVP Ethanol solution. Gold particles were biolistically bombarded onto the abaxial epidermis of 6 to 8 week old *N. benthamiana* leaves. Particles were accelerated at 1200psi at a distance of approximately 2cm from the surface of the leaves. Leaves were incubated in a humid chamber kept at room temperature in the dark overnight.

Confocal imaging

Bombarded cells were imaged using a Zeiss Laser Scanning Microscope LSM 5 Pascal with objective lens C-Aprochromat 63x/1.2W. Channel specification was as follows: Single Track (YFP track): Argon laser; Excitation: Line active 514; transmission between 2-4%; Main beam splitter 1: 415/548; Beam Splitter 2: 635; BP 530-600; Detector Gain approximately 700; Amplifier Offset: approximately -.8.

RESULTS AND DISCUSSION

A yeast two-hybrid screen was undertaken in order to isolate potential protein interactors with AT3-1. Screening of more than 4,000,000 colonies resulted in an

approximately 4X coverage of library content (S. Ueki, personal communication). A total of 400 positive colonies were recovered that were positive for the β-galactosidase assay. Repeated rounds of single colony isolation followed by the β-galactosidase assay reduced the number of positive colonies to 40. Of these colonies, none screened were found to be specific interactors with AT3-1 using Assay A (see methods) suggesting that this screen was compromised by a large number of false positive interactions which possibly masked the identification of *bona fide* interactions between AT3-1 and candidate protein partners. Using Assay B, 5 interacting proteins were identified and of these two, PI5-2 and PI236, specifically interacted with AT3-1 and not with human lamin C or topoisomerase I (Table A.1). Their predicted subcellular localization based on analysis by pSORT (http://psort.ims.u-tokyo.ac.jp/) and TargetP (http://www.cbs.dtu.dk/services/TargetP/) is also reported (Emanuelsson *et al*, 2000). Surprisingly the subcellular localization for Histone H3 is not nuclear despite the fact that histones interact with DNA and localize to the nucleus (Lindsey *et al*, 1991). One possible explanation for this discrepancy is that the sequence showing highest similarity to our putative interacting protein is not known as a full-length sequence and is lacking the nuclear targeting information.

In order to test whether the interaction identified by yeast two-hybrid assay was reproducible *in planta*, bi-molecular fluorescence complementation vectors were created for these two clones based on full-length cDNA sequences cloned from the same cDNA library used in the screen. As no fluorescence was observed by BiFC, the interaction was deemed non-specific and not pursued further.

Several reasons may explain why the yeast two-hybrid assay used to identify interacting partners with AT3-1 failed to identify *bona fide* interaction partners. It is possible that the tissue from which the library was derived does not express the interacting partners of AT3-1. Subcellular localization was performed in 6-8 week old

N. benthamiana plants whereas the library was derived from *N. tabacum* plants of the same age. It has previously been found that orthologous sequences identified in *N. benthamiana* may not be present or may not be expressed in *N. tabacum* (S. Ueki, personal communication). This would lead to a null result upon library screening. Furthermore, the interacting partners with AT3-1 may be unstable or transiently expressed so that their presence in the cDNA library is greatly diminished making them difficult to identify despite the extensive number of colonies screened in this experiment.

Table A.1. Yeast two-hybrid putative interactors

	NCBI BLASTn highest similarity match	**NCBI accession number**	**pSORT localization**	**TargetP localization**	**A[1]**	**B[2]**
PI5-1	*S. lycopersicum* HTC in fruit	AK246733.1	N[3]	C[4]	No	Yes
PI5-2	*N. tabacum* Histone H3.2	AB355993.1	N	M[5]	No	Yes
PI236	*S. lycopersicum* HTC in fruit	AK246733.1	N	C	No	Yes
PI295	*N. tabacum* cDNA AFLP fragment[8]	AJ717999.1	N	M	No	No
PI296	*N. sylvestris* RUBISCO small subunit	X53426.1	E[6]	O[7]	No	No
AT3-1			C	O		
LexA-AT3-1			N	O		

[1]Interaction deemed specific by assay A
[2]Interaction deemed specific by assay B
[3]Nuclear localization
[4]Chloroplast localization
[5]Mitochondrial localization
[6]Extracellular localization
[7]Localization other than nuclear, chloroplast or mitochondrial
[8]Result based on tBlastX similarity search

CONCLUSIONS

To isolate putative interacting partners with AT3-1, a yeast two-hybrid screen against a *N. tabacum* cv. Turk cDNA library was undertaken. The assay was not successful in identifying *bona fide* protein interacting partners of AT3-1, and bi-molecular fluorescence complementation of two putative interacting proteins did not reveal a specific interaction pattern. A new screen should be attempted using a library derived from the placenta of young *C. chinense* ‘‘Habanero’’ fruit so as to maximize the chance of identifying AT3-1 interacting proteins.

WORKS CITED

Ballas N, Citovsky V (1997). Nuclear localization signal binding protein from Arabidopsis mediates nuclear import of Agrobacterium VirD2 protein. *Proceedings of the National Academy of Sciences of the United States of America* **94**(20)**:** 10723-10728.

Bartel P, Chien, C. T., Sternglanz, R., and Fields, S. (1993). In: Hartley DA (ed) *Cellular Interactions in Development: A Practical Approach*. IRL: New York, pp 153-179.

Burbulis IE, Winkel-Shirley B (1999). Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. *Proceedings of the National Academy of Sciences of the United States of America* **96**(22)**:** 12929-12934.

Durfee T, Becherer K, Chen PL, Yeh SH, Yang YZ, Kilburn AE *et al* (1993). The retinoblastoma protein associates with the protein phosphatase type-1 catalytic subunit. *Genes & Development* **7**(4)**:** 555-569.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**(4)**:** 1005-1016.

Hollenberg SM, Sternglanz R, Cheng PF, Weintraub H (1995). Identification of a new family of tissue-specific basic helix-loop-helix proteins with a 2-hybrid system. *Molecular and Cellular Biology* **15**(7)**:** 3813-3822.

Kaiser C, Michaelis S., and Michell, A. (1994). *Methods in Yeast Genetics*. Cold Spring Harbor Lab Press: Plainview, NY.

Lindsey GG, Orgeig S, Thompson P, Davies N, Maeder DL (1991). Extended C-terminal tail of wheat histone H2A interacts with DNA of the linker region. *Journal of Molecular Biology* **218**(4)**:** 805-813.

Park H, Sternglanz R (1998). Two separate conserved domains of eukaryotic DNA topoisomerase I bind to each other and reconstitute enzymatic activity. *Chromosoma* **107**(3)**:** 211-215.

Rensink WA, Lee Y, Liu J, Iobst S, Ouyang S, Buell CR (2005). Comparative analyses of six solanaceous transcriptomes reveal a high degree of sequence conservation and species-specific transcripts. *BMC Genomics* **6:** 14.

Saslowsky D, Winkel-Shirley B (2001). Localization of flavonoid enzymes in Arabidopsis roots. *Plant Journal* **27**(1)**:** 37-48.

Tzfira T, Tian GW, Lacroix B, Vyas S, Li JX, Leitner-Dagan Y *et al* (2005). pSAT vectors: a modular series of plasmids for autofluorescent protein tagging and expression of multiple genes in plants. *Plant Molecular Biology* **57**(4)**:** 503-516.