5-15-2008

Evidence in the Carotenoid Biosynthetic Pathway for Variation in Evolutionary Rates

Stephanie A. Anderson
Trinity University

Evidence in the Carotenoid Biosynthetic Pathway for

Variation in Evolutionary Rates

Stephanie A. Anderson

A departmental honours thesis submitted to the

Department of Biology at Trinity University

in partial requirements for graduation with departmental honours

May 14th, 2008

Thesis Advisor

Department Chair

Associate Vice President
for Academic Affairs

EVIDENCE IN THE CAROTENOID BIOSYNTHETIC PATHWAY FOR VARIATION IN EVOLUTIONARY RATES

STEPHANIE A. ANDERSON

Evidence in the Carotenoid Biosynthetic Pathway for

Variation in Evolutionary Rates

Stephanie A. Anderson

A departmental honors thesis submitted to the

Department of Biology at Trinity University

in partial requirements for graduation with departmental honors

May 15th, 2008

_____ _____

Thesis Advisor Department Chair

_____

Associate Vice President
for Academic Affairs

Acknowledgements

I would like to thank Kevin Livingstone, Ph.D. for providing me with my thesis topic and for his assistance throughout the two years I spent doing research. Without his expertise, this project would not have been possible. I would like to extend my appreciation to David Ribble, Ph.D. and Natasa Macura, Ph.D. for serving on my thesis advisory committee, as well as Peter Olofsson, Ph.D. for his help with the statistical tests. I would also like to thank my advisor of four years, Thomas Koppenheffer, Ph.D., for his belief in my abilities and guidance from my first day of university.

Table of Contents

List of Figures

List of Tables

Abstract

Many cellular reactions occur in linear metabolic pathways where gene products act sequentially to produce needed compounds. The interrelationships between the products of these loci raises the question of whether they evolve in concert or independently. Previous research addressing this question indicated that in the anthocyanin pathway, which produces important secondary metabolites in plants, the genes encoding downstream enzymes show an accelerated rate of evolution when compared to upstream loci. The hypothesized cause of these differences has been attributed to relaxed selective constraint. This pattern and process has not, however, been tested in other systems. The carotenoid biosynthetic pathway, which also produces important colored secondary metabolites in plants, presents an appropriate system for an additional test. To produce a dataset suitable for this test, known mRNA sequences from four carotenoid biosynthetic enzymes of *Solanum lycopersicum* (tomato) were used to identify homologous sequences in taxa representing a broad range of angiosperms. Comparisons between *Phytoene desaturase*, *Zeta-carotene desaturase*, *Lycopene beta-cyclase* and *Zeaxanthin epoxidase* show that the downstream enzymes in the pathway have greater nucleotide diversity, nonsynonymous substitution rates and synonymous substitution rates. Evidence for selective constraint and an increase in the proportion of nucleotide sites under selective constraint has also been observed.

Introduction

Evolutionary rates vary among species, organisms, and proteins (Nei 1987). Understanding protein evolution and the forces behind it is a fundamental concept of molecular evolution. Studying protein evolution helps us better understand the relative importance of selection and genetic drift, as well as the types of forces that cause protein evolution to occur. Protein evolution requires two events: mutation of the nucleotides encoding amino acids and the fixation of these changes within the population (Pál et al., 2006). Mutations are random events that occur frequently in nature, but are typically corrected before they manifest themselves in the next generation. Fixation is dependent on the fitness effects of the mutation (neutral, deleterious, or advantageous).

The Neutral Theory of Molecular Evolution has been the foundation for molecular evolution research for nearly four decades (Hahn 2008), providing the basis for statistical models which help researchers distinguish natural selection from random genetic drift (Nielsen 2001; Hahn 2007). Despite the fact that recent research has challenged two major assumptions of the Neutral Theory, it is still used to understand molecular evolution. The first principle, direct selection, claims that most of the polymorphisms within species and the fixed differences between species do not affect the fitness of the individual, consequently they are neutral. The concept that synonymous changes are neutral is a small expansion on this idea. However, many people argue the neutrality of mutations that can cause phenotypic changes. Whereas the nucleotide mutation might not result in amino acid change, a disparity in matching tRNAs and codons could cause a change in translational efficacy. Lethal or harmful mutations rarely accumulate and are therefore not observed. Similarly, adaptive mutations are fixed relatively quickly therefore are not

observed while they are still polymorphic (Hahn 2008). DNA can be constrained or unconstrained under direct selection as long as observed changes are neutral. Several studies, including the results from Begun et al. (2007), have found evidence for direct selection on regulatory and coding mutations, which manifest themselves as divergence and excessive polymorphism in different species. While recent studies acknowledge that neutral mutations still do occur, they may represent the minority of changes, which is against the direct selection claim (Hahn 2008).

The second principle of the Neutral Theory is linked selection. This concept states that linked selection does not affect most loci, therefore variation can be attributed to neutral models of molecular evolution. Variation at these linked loci will be seen as neutral polymorphisms (Hahn 2008). The level of variation will decrease and alter the mutations rates observed relative to expectations within the neutral model. It is also assumed that genetic diversity will have a positive, linear relationship to population size. However, differences in nucleotide diversity between prokaryotes and vertebrates consistently averages to approximately two orders of magnitude, yet these populations vary over many more orders of magnitude (Lynch 2006). Recent models of linked selection predict either no relationship or a weak relationship with diversity and population size (Gillespie 2001; Charlesworth et al., 1993). Negative relationships between polymorphism and divergence have even been found (Hahn 2008).

Because of the conflicting evidence in regards to the Neutral Model, selection models of molecular evolution have been investigated. These models correctly predict the negative correlation between polymorphism and divergence described above. If a mutation is advantageous, this mutation will become fixed and polymorphism will decrease as the effect of the advantageous mutation becomes more evident. Increased rates of nonsynonymous

substitutions show decreased polymorphism (Hahn 2008). Even though there seems to be mounting evidence against Neutral Models, they are still by far preferred over the newly developed selection theories. The Neutral Theory is the most overarching model that we have to date. If one assumes that mutations are neutral, then the implications of selection can be ignored. The mathematics involved in the neutral model is more tractable, giving testable hypotheses and making models easier to work with.

One way by which molecular evolution can be observed and measured under the Neutral Theory is by the presence of elevated nonsynonymous to synonymous substitution rate ratios known as d_N/d_S, or ω (Purugganan and Wessler, 1994; Purugganan et al., 1995). The rates represented by d_N and d_S are defined as the number of nonsynonymous and synonymous substitutions per amino acid site, respectively (Yang 2007). Omega is a good indicator of protein-level selection, where $\omega = 1$ indicates neutral evolution, $\omega < 1$ indicates purifying selection, and $\omega > 1$ indicates positive, diversifying selection (Yang 2007). Where $\omega > 1$, one assumes that nonsynonymous substitutions result in increased fitness of the organism and have a greater likelihood to remain intact than synonymous substitutions (Yang et al., 2000; Yang 2000). Because the ratio is averaged over the entire protein sequence, and since adaptations cause changes in very few amino acids, ω will rarely be over 1 for the whole protein (Yang 2007). Most proteins have a ratio of less than one and are therefore considered to be under purifying selection, but often specific domains have ratios approaching 1, while other regions of the same protein are highly conserved (Li, 1997; Lu and Rausher, 2003). Measuring evolutionary rates using the above methods is the standard in molecular evolution; however, these methods are not without their problems. There is evidence for selection on synonymous sites, therefore the rate of evolution at synonymous sites is not a fully accurate measure of

mutation rate. In order to solve this problem, mutation rates across an organism's genome would have to be measured experimentally, which is beyond the capabilities of modern biology within a reasonable time frame for most organisms. Therefore calculations of synonymous and nonsynonymous substitutions are our best measures to date (Pál et al., 2006).

Codon usage bias is important in molecular evolution because it is an example of weak selection at the molecular level (Novembre 2002) therefore it is often used as a measure of selective constraint. The link between codon bias and selection is important in development of alternatives to the Neutral Theory (Ohta and Gillespie, 1996). Due to the redundancy of the genetic code, there is variation at synonymous sites in the possible combinations of nucleotide sequences that can code for the same amino acid sequence. The only exceptions are the start codon, AUG, the stop codons, UAA, UAG and UGA, and tryptophan, UGG. Not all 61 codons are found in equal amounts, and different organisms express different preferences for certain codons. Codon bias is often attributed to a result of mutation bias, translational optimization, GC composition and replication strand bias (Suzuki et al., 2007). Optimal codons allow for higher translational rates and greater translational accuracy, therefore highly expressed genes are often assumed to have greater codon bias. This theory is based on the concept that highly expressed genes are typically more critical to an organism's functions, and thus would be under greater selective constraint (Lu and Rausher, 2003). There are two classes of statistics that measure the codon usage across all amino acids. The first summarizes preferred codon usage, which requires prior knowledge of preferred codons. The second compares codon usage to a null distribution, or uniform usage (Novembre 2002). Effective number of codons, ENC, is the best summary statistic that does not require awareness of preferred codons (Novembre 2002). ENC is inversely proportional to nonuniform codon usage. Third position GC content (GC3) composition is one

way to measure mutational bias, as the nucleotides cytosine and guanine are preferred at the third codon position for translational efficacy.

At present, there is not a definitive answer as to whether increased rates of nonsynonymous substitutions are due to positive selection or relaxed selective constraint. Positive selection has been found to contribute to 20-45% of all amino-acid substitutions (Fay et al., 2002). Experiments testing the two theories have shown that typically it is relaxed selective constraint that causes increased rates of evolution. Upon further insight, support for this theory is largely based on subjective deductions instead of quantifiable data. Proteins associated with reproduction often undergo positive selection pressures that cause increased rates of protein evolution (Swanson and Vacquier, 2002).

Despite the debate between positive selection and relaxed selective constraint, there are several factors that are acknowledged as contributing to protein evolution. These include, but are not limited to, mutation rate, gene dispensability, expression level and number of protein interactions (Pál et al, 2006). Mutation rate is inherent in protein evolution; the higher the mutation rate, the higher the rate of evolution. This is the fundamental idea of molecular evolution as explained at the outset.

Gene dispensability is the measure of the relative importance of a gene within a genome. It is thought that high gene dispensability results in higher rates of protein evolution. One of the first theories that has been proposed is that a protein's genomic sequence will evolve in relation to the proportion of sites within the sequence involved in specific functions. This hypothesis is known as functional density, a concept particularly applicable to enzymes (Zuckerkandl 1976). The change in fitness is scaled to relate to the importance of the protein within cellular functions (Pál et al., 2006). This theory can be problematic given the fact that it is not only the primary

amino acid sequence, but the secondary and tertiary structures of proteins that have a large influence on function, a concept that is often unpredictable or extremely difficult to analyze. Mutation rate and efficiency of selection varies systematically, indicating a component of evolutionary rate that is not associated with the individual protein's function (Pál et al., 2006). As well, it is assumed that fitness effects are additive, but this assumption is not always true. A compensatory mutation might occur that prevents the original mutation from being deleterious. The probability of a second substitution which interacts with the initial substitution has up to a five times higher chance of fixation than the initial substitution (Shim Choi et al., 2005). It is also thought that the greater the functional importance, the greater the cost of mutations will be if they decrease the fitness of the organism. However, only ~5% or less of the evolutionary rate variation has been quantified by this theory (Drummond et al., 2005), as determined by analyzing the squared correlation coefficient of this relationship. Also, our measures of fitness are not in perfect correlation with natural conditions, especially since fitness is often measured with gene knockouts, which rarely occur in nature.

It has been found that highly expressed proteins evolve more slowly, accounting for ~34% of the evolutionary rate variation in yeast (Pál et al., 2001). If these substitution rates are favorable, repeated positive selection could be the cause for the elevated ω at specific domains, and therefore phenotypic diversification (Streisfeld and Rausher, 2007). If the substitutions provide no positive adaptive significance, the regulatory genes' high rates of evolution have no contribution to morphological divergence.

Lastly, protein evolution is affected by protein interactions. The greater the number of interactions a protein has with other proteins, in general the lower the rate of protein evolution. Residues of proteins found at the interface of complexes tend to evolve more slowly in

comparison to other residues (Mintseris and Weng, 2005). However, most research does not indicate protein interactions are a strong evolutionary force (Pál et al., 2006).

The reason why there is such little conclusive evidence regarding evolutionary rate variation could be due to the fact that individual proteins that are unrelated in function are the primary source of data for most of these experiments (Rausher et al., 1999). While protein evolution has been studied intensely using the above concepts, variation in evolutionary rates in biochemical pathways is not very well understood. Investigation of proteins that directly affect other protein's functions can shed more light onto this subject. This idea suggests that the study of enzymes in biochemical pathways may help us better understand protein evolution. The proteins involved in these pathways have large, direct influences on one another, and many are also hypothesized to function in close proximity to decrease diffusion differences of products from one enzyme to the next in the pathway. An enzyme's affect on one product in the pathway inevitably influences other enzymes' abilities to function, either positively or negatively.

The concept of evolutionary rate variation within biochemical pathways has been investigated using the anthocyanin biosynthetic pathway, a series of six enzyme-catalyzed steps which produce the blue and violet anthocyanin plant pigments found in all plant cells. Anthocyanins serve an important role in photoinhibition and as antioxidants, and expression of these pigments is often amplified under high-light stress conditions. It has been shown that the downstream enzymes in the anthocyanin biosynthetic pathway are subject to a higher evolutionary rate than those found earlier in the pathway due to greater specificity in substrates in the later enzymes, and therefore fewer mutation restrictions, or reduced selective constraint (Lu and Rausher, 2003). However this analysis was limited to the *Ipomoea* genus. The enzymes later in pathways have more limited effects than those upstream. The broad scale effects of

upstream enzymes has caused researchers to hypothesize that these enzymes are under greater selective constraint, as changes in these enzymes could change the fitness of the organism more than mutations farther down the pathway (Lu and Rausher, 2003).

Testing this hypothesis in another plant pigment pathway could determine if this is an isolated finding or has the potential to be a wide spread phenomenon. In addition to the anthocyanins, plants produce carotenoids, pigments which are also synthesized in a biochemical pathway, presenting an appropriate model system for such a test. Carotenoids are naturally occurring pigments ranging in color from pale yellow to dark red and are found in plants, some algae, fungi and bacteria. They are typically 40-carbon isoprenoids with up to 15 conjugated double bonds within the polyene chains (Figure 1). All are derived from phytoene (Howitt and Pogson, 2006). The two subgroups of carotenoids are carotenes, which are deoxygenated, and xanthophylls, which are oxygenated. They are an essential physiological component of photosynthesis when found in chloroplasts. In higher plants, β-carotene binds to both photosystem I and II through the reaction center subunits. Xanthophylls serve as accessory pigments and as structural elements to light-harvesting complexes. Light-harvesting complex II, the major component in photosystem II, binds to lutein, violaxanthin and neoxanthin, as do the minor complexes. Only zeaxanthin is bound in times of excessive light stress to the minor complexes. Genetic manipulations of photosystem protein biosynthesis results in plasticity of the photosynthetic membrane to preserve the structure and function of PSII-LHCII, including varying levels of carotenoids essential for photosynthesis (Ruban et al., 2006). Xanthophylls act in concert with β-carotene as chromophores, possessing the unique ability to absorb a more extensive range of blue and ultraviolet light than chlorophyll to then transfer the energy for photosynthetic electron transport. It appears as though xanthophylls are highly conserved in

plant species due to the fact that they possess very similar spectral properties (Dall'Osto et al., 2007). Xanthophylls also serve as photoprotectants and possess antioxidant properties important to plant cells (Howitt and Pogson, 2006). They prevent membrane damage by quenching reactive oxygen species (ROS) and triplet chlorophyll, both caused by excessive exposure to light energy. Specifically, lutein quenches triplet chlorophyll by binding at the L1 site, preventing ROS formation (Dall'Osto et al., 2006). Substitution of lutein with violaxanthin decreases triplet chlorophyll quenching and results in greater amounts of ROS in cells. Lutein mutants use zeaxanthin in photoprotection, but zeaxanthin is less effective (Dall'Osto et al., 2006).

Carotenoids color flowers and fruits through accumulation in all types of plastids including chromoplasts, amyloplasts, elaioplasts, leucoplasts and etioplasts, with the exception of proplastids (Figure 2) (Howitt and Pogson, 2006). Carotenoids are localized in plastoglobuli or as crystalline structures during the transition from chloroplasts to chromoplasts (Tevini and Steinmuller, 1985). While the enzymes in the carotenoid biosynthetic pathway do not interact directly, it is thought that these enzymes are located in close proximity to one another in order shuttle products from one enzymatic reaction to another rapidly. The coloration that they provide attracts animals and insects to aid in pollination and seed dispersal. Different plants accumulate different varieties and concentrations of plastids. The leaves and stems of plants also contain carotenoids, but the color is masked by chlorophyll. Carotenoids found in chromoplasts are considered to be secondary metabolites and are precursors for abscisic acid (ABA) synthesis (Hirschberg 2001) and many scents (Howitt and Pogson, 2006). ABA is a plant hormone primarily responsible for abscission, embryo development, seed dormancy, and stomatic regulation. The significance of ABA in plant processes is another indication of the importance

of carotenoid biosynthesis. In addition to their role in plants, recent studies on lycopene and lutein, two common carotenoids, have indicated their potential benefits to human health, including powerful anticancer properties. β-carotene is a precursor for vitamin A synthesis in the human body and is thus an essential molecule. Lutein and zeaxanthin also have been found to prevent macular damage with age (Howitt and Pogson, 2006).

The carotenoid biosynthetic pathway is the series of 11 enzyme-catalyzed steps which synthesize the approximately 700 known carotenoids (Figure 1) (Britton 1998). The first recognized step in carotenoid biosynthesis is the condensation reaction of two dimethylallyl pyrophosphates, a 10-carbon compound, to produce geranylgeranyl pyrophosphate (GGPP), a 20-carbon compound, by catalysis of *Geranylgeranyl pyrophosphate synthase*. A condensation reaction of two molecules of GGPP forms 15-*cis*-phytoene, a 40-carbon compound, and the first carotenoid produced in the pathway, via the enzyme *Phytoene synthase* (*PSY*). Duplicate copies of this enzyme have been identified in a variety of species. Phytoene rarely accumulates in plant tissues. *Phytoene desaturase* (*PDS*) catalyzes two symmetrical desaturation reactions to produce di-*cis*-ζ-carotene. *ζ-carotene desaturase* (*ZDS*) catalyzes another two symmetrical desaturation reactions, which produces tetra-*cis*-lycopene. The next reactions use enzymes that preferentially bind all-*trans*-lycopene. Formation of all-*trans*-lycopene requires the catalysis of the enzyme *Carotenoid isomerase* (*CRTISO*). *CRTISO* is a relatively newly discovered enzyme. Its function is not entirely known; some researchers theorize that it is required in conjunction with the *ZDS* and *PDS* for carotenoid desaturation (Park et al., 2002).

From here, there is a split in the pathway and cyclization of the lycopene derivatives. The β,β branch produces β-carotene, while the β,ε produces α-carotene along with the respective derivatives. Both branches include the use of *Lycopene-β-cyclase* (*βLCY*) to form β rings on

both ends of lycopene, producing β-carotene in the β,β branch, and with the additional catalysis of *Lycopene-ε-cyclase* (*εLCY*), the β,ε branch produces α-carotene. α and β -carotene are modified by *β-hydroxylase* (*βOH*) to produce the xanthophylls. The addition of a hydroxyl group on the C3 of each α- and β-carotene molecule's β ring produces zeinoxanthin and β-cryptoxanthin, respectively. *βOH* works on β-cryptoxanthin again to hydroxylate the other β ring, producing zeaxanthin, while *ε-hydroxylase* (*εOH*) hydroxylates the ε ring of zeinoxanthin to produce then end product of the pathway, lutein. *εOH* is a little understood P450 cytochrome monooxygenase, whereas *βOH* is well-studied and categorized as a non-haeme diiron monooxygenase (Tian et al., 2004).

Zeaxanthin in the β,β branch undergoes epoxidation by *Zeaxanthin epoxidase* (*ZEP*). In high light intensity conditions, *Violaxanthin de-epoxidse* (*VDE*) reverses this reaction to revert back to zeaxanthin. *Neoxanthin synthase* (*NXS*) converts 9-cis-violaxanthin to 9-*cis*-neoxanthin, the last carotenoid of the β,β branch. Both violaxanthin and neoxanthin can be cleaved by *9-cis-epoxycarotenoid dioxygenase* (*NCED*) to ultimately produce abscisic acid (ABA).

There are some minor variations in the carotenoid biosynthetic pathway among species. Yellow-flesh tomatoes have a loss-of-function mutation within the *PSY1* gene (Ronen et al., 2000). Duplicates of *PSY* have also been found in numerous plant species including wheat, rice and maize (Gallagher et al., 2004), but only the *PSY1* transcript accurately correlates with carotenoid content.

The present study uses the same principles from the studies on the anthocyanin biosynthetic pathway evolutionary rate variation and applies them to the carotenoid biosynthetic pathway. The genes for the enzymes in the carotenoid biosynthetic pathway are evolutionarily

and functionally conserved. There are several notable differences between the anthocyanin and carotenoid biosynthetic pathways. The carotenoid biosynthetic pathway has more intermediates and more enzymes than the anthocyanin biosynthetic pathway. Carotenoids serve a direct physiological function within photosystems, therefore their synthesis directly affects photosynthesis, unlike the anthocyanins. Carotenoids are not modified by glycosylation like anthocyanins are. The differences between the anthocyanin and carotenoid biosynthetic pathways should be noted because while the same principles from the study of anthocyanin evolutionary rates will be applied to the carotenoid biosynthetic pathway, the methods and the interpretation of the results may differ slightly. By creating sequence alignments for a representative sample of both plant species and enzymes within pathway, one can perform a variety of statistical tests to examine the cause and types of evolutionary rate variation. These methods could be applied to other biosynthetic pathways in future research. Evidence to support the hypothesis would be cause to propose this evolutionary observation as a trend, rather than an isolated phenomenon.

Materials and Methods

Sequences and Alignments

The genes used for this analysis were selected based on their location in the pathway, and were those enzymes without evidence of duplication, with the exception of *βLCY*. Four genes were chosen for analysis: one gene can be considered to be upstream (*PDS*), two are midstream (*ZDS* and *βLCY*), and one is downstream (*ZEP*) in the carotenoid biosynthetic pathway (Figure 1). *Solanum lycopersicum* (tomato) was chosen to be the representative species based on previous research, and complementary DNA (cDNA) sequences of each of the enzymes listed above were obtained from GenBank. A BLAST search was performed against all plants in the TIGR Gene Indices (Quackenbush et al., 2001), PlantGDB (Dong et al., 2004) and GenBank databases. Sequences were determined to be homologous based on percent identities and the length of the match sequence.

The selected sequences were initially trimmed down to coding sequences and then translated using BioEdit 7.0.9.0 (Hall 1999). Edited peptide sequences were imputed into ClustalX 2.0.5 (Larkin et al., 2007; Thompson et al., 1997) to create multiple sequence alignments. Predicted chloroplast leader sequences were identified using Chlorop1.1 Server (Emanuelsson et al., 1999) and subjective analysis was used to identify the portion of the alignment that appeared to represent the predicted leader sequences from the peptide alignment. The DNA sequences corresponding to the predicted leader sequences were then removed from the DNA alignment. Manual adjustment of aligned DNA sequences was then performed as necessary. A concatenated sequence for each species was also created using each enzyme's DNA sequence alignment.

Phylogenetic Tree Construction

The maximum likelihood phylogeny of the seven species analyzed was constructed using the concatenated DNA sequences and the program dnaml, a part of the PHYLIP package (Felsenstein, 2005). The phylogeny was then tested by 1000 bootstrap replicates.

Genetic Distance and Nucleotide Diversity

Genetic distances were calculated from the concatenated gene sequence alignment using the branch lengths from the phylogenetic tree. Nucleotide diversity, π, was calculated by DnaSP 4.50.1 (Rozas et al., 2003) using the methods of Nei (1987) and Jukes and Cantor (1969).

Comparing Site Models To Determine Selective Constraint and Calculation of d_N and d_S

Two models of molecular evolution were fit to the data for each gene using PAML 4.0 (Yang 2007) to test for evidence of selective constraint. The M0 model assumes all positions across a given set of sequences have the same value for ω and estimates this value. The M1a model uses two sites classes: Class 1 is a proportion of sites with $0 < \omega < 1$, and the remainder of codons form the other class with fixed $\omega = 1$. Model M1a estimates both the proportion and ω of Class 1 sites. The fits of these two models for each genes were compared using likelihood ratio tests. Pairwise d_N and d_S values were calculated for each gene using PAML.

Codon Bias Estimation

The effective number of codons (ENC) was calculated using DnaSP implementing the methods of Wright (1990). Because codon bias is often associated with GC content at the third position in a codon (GC3 content), a graphical comparison of ENC versus GC3 content was used

to control for possible mutation bias. The observed values of ENC versus GC3 content were plotted using the Nc-plot technique of Wright (1990). Expected values were calculated using the methods of Novembre (2002).

Results

Sequences and Alignments

A large number of sequences for each enzyme analyzed were found across a diverse set of taxa. The analysis required, however, that each alignment included the same set of species, so a final sample of only 6 species was used in this study. The species analyzed include the following: *Solanum lycopersicum, Capsicum annuum, Gentiana lutea, Chrysanthemum x morifolium, Citrus spp.,* and *Daucus carota* (Table 1). Sequences for *Oryza sativa* were also obtained and used as an outgroup for phylogenetic tree construction. The Chlorop 1.1 Server is only able to approximate the length of the chloroplast leader sequence for each sequence obtained, so subjectivity was used in determining where sequences alignments were trimmed. For some sequences, that meant that some of the sequence not predicted to be in the chloroplast leader region was removed for analysis purposes. The chloroplast leader sequence region depicted in Appendix 1 is representative of how a typical leader sequence appeared. Also, because the alignments have to contain the same number of base pairs for each sequence, all sequences were trimmed at the same point in the alignment on the 5' and 3' ends. Accession numbers and base pair ranges of the sequences used are given in Table 1. A minimum of 73% of the *Solanum lycopersicum* coding sequence was used for analysis, with values up to 89% (Table 1). Both the DNA and protein multiple sequence alignments showed strong conservation among the plant species. The DNA alignments used for analysis are located in Appendices 2 - 5.

Phylogenetic Tree Construction

The phylogeny constructed from the concatenated sequence alignment (Figure 2) shows evolutionary relationships between the selected species that are in accordance with known phylogenetic data (Soltis and Soltis, 2004). Bootstrap analysis revealed strong support for this tree.

Nucleotide Diversity

Nucleotide diversity (π) increased noticeably for downstream enzymes in the pathway (Figure 3). The Jukes and Cantor (1969) method produced an overall higher value of nucleotide diversity for each enzyme, but with the same positive relationship observed using the methods of Nei (1987). These data are consistent with the hypothesis that that the downstream enzymes are evolving at a higher rate than the upstream enzymes.

Comparing Site Models To Determine Selective Constraint and Calculation of d_N and d_S

The χ^2 test to determine the fit of Models M0 and M1a for each gene rejected M0, indicating Model M1a was a better fit across all the data ($p < 0.005$). The graph of enzyme position in the carotenoid biosynthetic pathway versus ω values generated by Models M0 and M1a, as well as the graph of percent codons with $0 < \omega <$ from model M1a are shown in Figure 4. If the ω calculated using M0 are used as a general guide, it can be seen that *PDS, ZDS* and *βLCY* have similar ω values, but *ZEP* shows a large increase in ω over the other genes. Using Model M1a, ω is relatively stable across the enzymes. In considering the parameters from Model M1a, the fraction of sites subject to purifying selection appears to be relatively stable across *PDS, ZDS* and *βLCY*, but smaller for *ZEP*. Therefore, it can be concluded that the strength of selective

constraint is constant across the sites in these enzymes subject to constraint, but the proportion of sites subject to selective constraint is noticeably reduced for *ZEP*.

While it is recognized that the complete set of pairwise comparisons of d_N and d_S are not statistically independent, graphs of d_N and d_S versus genetic distance can be used to infer general trends. Linear regression showed a positive relationship across all four enzymes between d_N and the overall genetic distance calculated from the branch lengths of the phylogenetic tree ($p < 0.005$). The rate of d_N increase with genetic distance was similar for *ZDS*, *PDS* and *βLCY*, and highest for *ZEP* (Table 2; Figure 5), again showing the rate of d_N versus genetic distance values increased from upstream to downstream in the carotenoid biosynthetic pathway. Interestingly, the linear fits of d_N versus genetic distance were also variable, as measured by the squared regression coefficient, with comparisons more downstream in the carotenoid biosynthetic pathway, showing greater heterogeneity (*PDS* $r^2 = 0.873$; *ZDS* $r^2 = 0.766$; *βLCY* $r^2 = 0.586$; *ZEP* $r^2 = 0.528$).

The relationships of d_S versus genetic distance were similar to the pattern seen in the d_N versus genetic distance analysis (Table 2; Figure 6). There was a positive, linear relationship between d_S and genetic distance ($p < 0.005$) and the rate of increase of d_S versus genetic distance increased the farther downstream the enzyme is located in the carotenoid biosynthetic pathway. The rate of d_S increase with genetic distance was lowest for *PDS* and *ZDS*, intermediate for *ZEP*, and highest for *βLCY* (Figure 6). The regression correlations were higher for each gene in comparison to the d_N analysis, and overall individual species values were less variable than in the d_S versus genetic distance analysis. (*PDS* $r^2 = 0.904$; *ZDS* $r^2 = 0.804$; *βLCY* $r^2 = 0.760$; *ZEP* $r^2 = 0.767$).

Codon Use Bias

ENC measures the extent to which codon usage deviates from the possible use of the 61 sense codons in the universal genetic code. Values range from 20, where codon bias is at a maximum and only one codon is used for each amino acid, to 61, where there is no codon bias and all codons are used. Codon bias was found to be relatively uniform for *PDS, ZDS, βLCY* and *ZEP*, but *βLCY* did have slightly lower values (Table 3). This may be due to the fact that *βLCY* is under greater selection for codon use bias as compared to the other three genes. Alternatively, this difference could be attributed to other reasons, such as mutation bias. In order to distinguish between these possible explanations, Figure 7 depicts expected ENC as a function of third position GC content (Wright 1990). All of the enzymes for each species lie either at or only slightly below their expected values, with the exception of the *ZDS* gene for *Capsicum annuum*. These data therefore indicate that the four enzymes analyzed in the carotenoid biosynthetic pathway are under at most only mild selective constraint, including *βLCY*. The general lack of variation amongst species and between genes implies that codon bias has been relatively stable during evolution in the carotenoid biosynthetic pathway.

Discussion

The present research parallels the results from the Rausher, Miller and Tiffin (1999) and Lu and Rausher (2003) studies. Rausher, Miller and Tiffin (1999) looked at anthocyanin biosynthetic pathway genes by comparing monocots and dicots and found that regulatory genes were found to evolve more rapidly, identified by the reduction in nonsynonymous substitution rates of upstream genes, than the structural genes that they regulate. Lu and Rausher's (2003) study found the same nonsynonymous substitution rate pattern to be consistent within the *Ipomoea* genus between upstream and downstream genes, indicating that the genes for the upstream enzymes evolve more slowly than the genes for downstream enzymes. This difference was attributed to greater selective constraint on upstream enzymes. We approached the same scientific problem with a more restricted taxonomic range than the Rausher, Miller and Tiffin (1999) study, but more broad than Lu and Rausher's (2003) study. Our research indicates that the upstream enzymes in the carotenoid biosynthetic pathway studied across a wide range of angiosperms have overall evolved more slowly than the downstream enzymes. Reduced rates of nonsynonymous substitutions were also observed, giving indication for selective constraint, but the variance in selection constraint amongst the enzymes differed between this research and previous studies.

Nucleotide diversity can serve as a general indicator to determine whether or not there is an overall difference in diversity between genes. In this sample of genes, nucleotide diversity increases as you go farther downstream in the carotenoid biosynthetic pathway. While the standard error values are relatively high, the pattern of increasing nucleotide diversity is still conspicuous and is strong enough to discount high variability in data as a reason for the

differences seen. The high standard error values could be attributed to the fact that *Solanum lycopersicum* and *Capsicum annuum* are much more closely related in comparison to the other species therefore their nucleotide diversity values are much more similar than the other species.

Two models were fit to the data to determine the types of evolutionary pressures causing the observed increase in nucleotide diversity downstream in the carotenoid biosynthetic pathway. Model M0, which assumes a constant ω across all sites, did not fit the data as well as model M1a, indicating that ω was variable across sites. The fits of model M1a showed the fraction of sites subject to purifying selection, as measured by $0 < \omega < 1$, is relatively stable across *PDS, ZDS* and *βLCY*, with a clear decrease for *ZEP*. For those sites under selective constraint, the estimated values of ω were similar across all four genes. These results indicate that this pattern may be attributable to the fact that fewer codons within the *ZEP* gene are under selective constraint, rather than a decrease in the strength of selective constraint acting across the whole gene. *ZEP* could also just be an outlier; additional research investigating more of the enzymes in the carotenoid biosynthetic pathway could confirm whether the downstream enzymes overall show a decrease in the proportion of sites under selective constraint. It is important to note that the presence of positive selection is not ruled out by this analysis, and additional site models could also be tested to determine if these genes are under positive selection.

π does not take into account synonymous or nonsynonymous positions within the genetic code when measuring DNA polymorphism, therefore the specific location within a codon that these nucleotide differences occurred was measured through the analysis of nonsynonymous and synonymous substitution rates. Nonsynonymous substitution rates increased the farther downstream an enzyme is in the carotenoid biosynthetic pathway (Table 2; Figure 5). The d_N value for *PDS* is higher than for *ZDS*, despite *PDS*'s most upstream position in the pathway

studied, however, the values are very similar, and the standard error estimates overlap. Because these enzymes function one immediately after the other in the pathway, the d_N values for these enzymes are expected to be very similar. Also, the *Solanum lycopersicum* and *Capsicum annuum* comparison is an outlier as shown by the points to the extreme left in Figure 6. This comparison could have reduced, or possibly increased, the differences in d_N versus genetic distance values for each enzyme. The outliers could have caused the large increase in the d_N versus genetic distance regression for *ZEP*. The other enzymes analyzed are not positioned one after the other in the pathway, therefore there is greater variance in their d_N values. Further research could use data from more species from a variety of angiosperms to increase the size of the sample used. This could reduce the effect of the outlier pairwise comparison and could provide more conclusive data. By analyzing more enzymes within the pathway, it is possible to determine if this trend holds for the complete carotenoid biosynthetic pathway, or if *ZEP* is indeed an outlier. Variability in individual species values also increased the farther downstream in the pathway an enzyme was located, as measured by standard error values of squared regression coefficients. This observation is consistent with the idea that less selective constraint is acting on downstream enzymes. Reduced selective constraint would theoretically allow genes to vary more, especially concerning nonsynonymous substitution rates.

Synonymous substitutions were shown to have higher rates than nonsynonymous substitutions, which is expected due to the fact that nonsynonymous substitutions directly cause amino acid changes within a given sequence, whereas synonymous substitutions do not. Synonymous substitutions also generally increased the farther down the enzyme was found in the carotenoid biosynthetic pathway, with the exception of *βLCY* (Table 2; Figure 6). *βLCY* had the highest rate of synonymous substitutions as compared to genetic distance, but it was not the most

downstream enzyme analyzed. The squared regression coefficient for βLCY is low at a value of 0.586, therefore the individual species values for this enzyme were highly variable. βLCY also had the largest standard error value of the squared regression coefficient, which also indicates a high degree of variability between individual species. The low regression coefficient value could account for this discrepancy. In addition, standard error rates for these squared regression coefficients increased the farther downstream an enzyme was located in the carotenoid biosynthetic pathway, similar to nonsynonymous substitution rates. The differences in d_S versus genetic distance values were more evenly separated across enzymes in comparison to the d_N values. It is expected that d_S values will be relatively uniform across enzymes due to their reduced effect on protein changes in comparison to nonsynonymous substitutions. Because the d_S and d_N graphs as a function of genetic distance are not statistically independent for each enzyme, there is not a statistical test that can be used to quantitatively measure the differences between enzymes. We did not find that the most upstream enzyme analyzed had the highest rate of synonymous substitutions, unlike Lu and Rausher's (2003) observations. Lu and Rausher (2003) ruled out elevated mutation rates as a cause for higher rates of amino-acid substitutions in downstream genes because they found the most upstream gene analyzed had the highest synonymous substitution rates. We therefore did not rule out higher mutation rates as causing higher substitution rates in downstream genes.

The estimates of ω from M0 and M1a were both well under 1, indicating that positive selection does not appear to be occurring on any of the genes (Table 2). This observation leads to the idea that it is reduced selective constraint that is allowing for elevated differences rates in the downstream genes. This observation is also consistent with the differences in the degree of codon bias observed amongst the genes analyzed.

It is assumed that genes under greater selective constraint will exhibit greater codon bias. There was no observable trend in effective number of codons in comparison to the position of the gene in the carotenoid biosynthetic pathway (Table 3; Figure 7). GC3 content was also similar amongst the genes, with the exception of the slightly lower value for *βLCY* (Table 3; Figure 7). All of the mean effective number of codon values were very similar, with standard errors providing considerable overlap between genes with the exception of *βLCY*. The gene with the apparent highest codon bias, *βLCY*, had the highest rate of synonymous substitutions, as well as the second highest nonsynonymous substitution rates. A high synonymous substitution rate in the presence of high codon bias is not a typical result. It is expected that at codon-bias equilibrium, greater selection for codon bias would cause a decrease in synonymous substitution rates. However, Lu and Rausher (2003) observed the same apparent anomaly in their study of anthocyanin biosynthetic enzymes. They found that the most upstream gene had the greatest codon bias and the highest rate of synonymous substitutions. This concept assumes that all genes analyzed have a similar underlying synonymous mutation rate, regardless of selection pressures. If this assumption is not true and the underlying mutation rate is higher for *βLCY* than the other enzymes, the synonymous substitution rate will be relatively higher, but the equilibrium codon bias should not be affected by this variation in substitution rates. Therefore, it is feasible that *βLCY* is under high selective constraint, but its baseline mutation rate is higher, which could account for the apparent discrepancy in the data.

One possible reason for the differences in evolutionary rate in upstream and downstream enzymes is that the enzymes downstream directly produce the molecules that serve as structural components in photosystem II. Lutein, violaxanthin, neoxanthin and zeaxanthin all bind to PSII-LHCII. PSII is the newly evolved photosystem found only in chloroplasts and non-sulfur purple

bacteria. Therefore the latter half of the carotenoid biosynthetic pathway, which can be characterized as from the cyclization reaction and downwards, has more recent functions within photosynthesis. Since *βLCY* is responsible for this important branch point in the pathway, it is plausible that the discrepancies within the *βLCY* data could be attributed to this fact. Perhaps its expression varies depending on its location within a plant, which could cause greater variation in its evolution. Some research has indicated that *βLCY* and *εLCY* evolved from a common ancestor in bacteria due to a duplication event (Krubasik and Sandmann, 2000; Tao et al., 2004). It is known that duplicated genes often show greater mutation rates and diversity, which could also account for the discrepencies in the *βLCY* data. Further analysis of *βLCY* could show similar evolutionary trends with this enzyme as well, which could indicate that the differences in evolutionary rate for these two enzymes are due to their position at the split in the carotenoid biosynthetic pathway.

It is also important to note that photosystem I evolved much earlier than photosystem II (Amunts and Nelson, 2008). Therefore it is plausible that the enzymes that produce molecules directly involved with photosystem I, such as the molecule *β*-carotene, may have reached an evolutionary standpoint. Photosystem II could be more variable, and thus allow for more variation in the evolution of the carotenoids in the lower half of the pathway.

Overall, it appears that these data confirm the hypothesis that the downstream enzymes evolve more quickly than the upstream enzymes in the carotenoid biosynthetic pathway due to relaxed selective constraint. However, evidence of positive selection has not been fully examined, therefore it cannot be eliminated as a possible reason for variation in evolutionary rates of the studied enzymes. In the future, codon substitution models could be used to detect evidence for positive selection or not. A more detailed analysis of the pathway, including the β,ε

branch and the rest of the enzymes in the β,β branch could give further insight into the evolutionary rate trends. Anomalies found in the analysis could be further investigated to determine the reasoning for these differences. Conversely, these irregularities could instead be proven to be trends within the pathway.

References

Al-Babili, S., Hugueney, P., Schledz, M., Welsch, R., Frohnmeyer, H., Laule, O., Beyer, P. 2000. Identification of a novel gene coding for neoxanthin synthase from *Solanum tuberosum*. FEBES Letters. **485**:168-172.

Amunts, A., Nelson, N. 2008. Functional organization of a plant Photosystem I: Evolution of a highly efficient photochemical machine. Plant Physiol. Biochem. **46**: 228-237.

Begun, D.J., Holloway, A.K., Stephens, K., Hillier, L.W., Poh, W.-P., Hahn, M.W., Nsta, P.M., Jones, C.D., Kern, A.D., Dewey, C., et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophilia simulans*. PLoS Biol. **5**: e310.

Bouvier, F., D'Harlingue, A., Backhaus, R.A., Kumagai, M.H., Camara, B. 2000. Indentification of neoxanthin synthase as a carotenoid cyclase paralog. Eur. J. Biochem. **267**: 6346-6352.

Britton, G. 1998. Overview of carotenoid biosynthesis. In *Carotenoids* (eds G. Britton, S. Liaaen Jensen and H. Pfander), pp. 13-147. Birkhauser, Basel, Switzerland.

Charlesworth, B., Morgan, M.T., Charlesworth, D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics. **134**: 1289-1303.

Cunningham, F.X. 2002. Regulation of carotenoid synthesis and accumulation in plants. Pure and Applied Chemistry. **74**: 1409-1417.

Dall'Osto, L., Fiore, A., Cazzaniqa, S., Giuliano, G., Bassi, R. 2007. Different roles of alpha- and beta-branch xanthophylls in photosystem assembly and photoprotection. J. Biol. Chem. **282**: 35056 – 35068.

Dall'Osto, L., Lico, C., Alric, J., Giuliano, G., Havaux, M., Bassi, R. 2006. Lutein is needed for efficient chlorophyll triplet quenching in the major LHCII antenna complex of higher plants and effective photoprotection *in vivo* under strong light. BMC Plant Biol. **6**: 32.

Drummond, D.A., Bloom, J.D., Adamci, C., Wilke, C.O., Arnold, F.H. 2005. Why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. U.S.A. **102**: 14338-43.

Dong, Q., Schlueter, S.D., Brendel, V. 2004. PlantGDB, plant genome database and analysis tools. Nucleic Acids Res. **32**: D3354-D359.

Emanuelsson, O., Nielsen, H., Von Heijne, G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. **8**: 978-984.

Fay, J.C.. Wyckoff, G.J., and Wu, C.I. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosphilia*. Nature. **415**: 1024-1026.

Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. *Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.*

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. **125**:1-15.

Gallagher, C.E., Matthews, P.D., Li, F., Wurtzel, E.T. 2004. Gene Duplication in the Carotenoid Biosynthetic Pathway Preceded Evolution of the Grasses. Plant Physiology. **135**: 1776-1783.

Gillespie, J.H. 2001. Is the population size of a species relevant to its evolution? Evolution. **55**: 2161-2169.

Hahn, M. 2008. Toward a Selection Theory of Molecular Evolution. Evolution. **62**: 255-265.

Hahn, M. 2007. Detecting natural selection on *cis*-regulatory DNA. Genetica. **129**: 7-18.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. **41**:95-98.

Hartmann, S., Lu, D., Philips, J., Vision, T.J. 2006. Phytome: a platform for plant comparative genomics. Nucleic Acids Research. **34**:D724-D730.

Hirschberg, J. 2001. Carotenoid biosynthesis in flowering plants. Current Opinion in Plant Biology. **4**: 210-218.

Howitt, C.A., Pogson, B.J. 2006. Carotenoid accumulation and function in seeds and non-green tissues. Plant, Cell and Environment. **29**: 435-445.

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

Just., B.J., Santos, C.A.F., Fonseca, M.E.N., Boiteux, L.S., Oloizia, B.B., Simon, P.W. 2007. Carotenoid biosynthesis structural genes in carrot (*Daucus carota*): isolation, sequence-characterization, single nucleotide polymorphism (SNP) markers and genome mapping. Theor. Appl. Genet. **114**:693-704.

Krusbasik, P., Sandmann, G. 2000. Molecular evolution of lycopene cyclases involved in the formation of carotenoids with ionone end groups. Biochem. Soc. Trans. **28**: 806-10.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thomspon, J.D., Gibson, T.J., Higgins, D.,G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. **23**: 2947-2948.

Li, W.-H., 1997. Molecular evolution. Sinauer Associates. Sunderland, Mass.

Lu, Y., Rausher, M.D. 2003. Evolutionary Rate Variation in Anthocyanin Pathway Genes. Mol. Biol. Evol. **20**: 1844-1853.

Lynch, M. 2006. The origins of eukaryotic gene structure. Mol. Biol. Evol. **23**: 450-468.

Mintseris, J, Weng, Z. 2005. Structure, function and evolution of transient and obligate protein-protein interactions. Proc. Natl. Acad. Sci. USA. **102**: 10930-10935.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418-426.

Nielsen, R. 2001. Statistical test of selective neutrality in the age of genomics. Heredity. **86**: 641-647.

Novembre, J.A. 2002. Accounting for Background Nucleotide Composition When Measuring Codon Usage Bias. Mol. Biol. Evol. **19**:1390-1394.

Ohta, T., Gillespie, J. 1996. Development of neutral and nearly neutral theories. Theor. Popul. Biol. **49**: 128-142.

Pál, C., Papp, B., Hurst, L.D. 2001. Highly expressed genes in yeast evolve slowly. Genetics. **158**: 927-31.

Pál, C., Pap, B., Lercher, M.J. 2006. An integrated view of protein evolution. Nature Reviews Genetics. **7:** 337-348.

Park, H., Kreunen, S.S., Cuttriss, A.J., DellaPenna, D., Pogson, B.J. 2002. Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis. Plant Cell. **14**: 321-332.

Purugganan, M.D., Rounsley, S.D., Schmidy R.J., Yanofsky, M.F. 1995. Molecular evolution of flower development: diversification of the plant MADS-BOX regulatory gene family. Genetics. **140**: 345-356.

Purugganan, M.D., Wessler, S.R. 1994. Molecular evolution of the plant *R* regulatory gene family. Genetics. **138**: 849-854.

Quackenbush, J., Cho, J., Lee, D., Liang, F., Holt, I., Karamycheva, S., Parvizi, B., Pertea, G., Sultana, R., White, J. 2001. The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. **29**: 159-164.

Rausher, M.D., Miller, R.E., Tiffin, P. 1999. Patterns of Evolutionary Rate Variation Among Genes of the Anthocyanin Biosynthetic Pathway. Mol. Biol. Evol. **16**: 266-274.

Ronen, G., Carmel-Goren, L., Zamir, D., Hirschberg, J. 2000. An alternative pathway to β-carotene formation in plant chromoplasts discovered by map-based cloning of *Beta* and *old-gold* color mutations in tomato. Proc. Natl. Acad. Sci. **97**: 11102-11107.

Rozas, J., J.C. Sánchez-DelBarrio, X. M., Rosaz, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**: 2496-2497.

Ruban, A.V., Solovieva, S., Lee, P.J., Ilioaia, C., Wentworth, M., Ganeteq, U., Klimmek, F., Chow, W.S., Anderson, J.M., Jansson, S., Horton, P. 2006. Plasticity in the composition of the light harvesting antenna of higher plants preserves structural integrity and biological function. J. Biol. Chem. **281**: 14981 - 14990.

Saitou, N., Nei, M. 1987. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. Mol. Biol. Evol. **4**: 406-425.

Sharp, P.M., W.H. Li. 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. **4**:222-230.

Shim Choi, S., Li, W., Lahn, B.T. 2005. Robust signals of coevolution of interacting residues in mammalian proteomes identified by phylogeny-aided structural analysis. Nature Genet. **37**: 1367-1371.

Soltis, P. S., Soltis, D. E**.** 2004. The Origin and Diversification of Angiosperms. Amer. J. Bot. **91**: 1614-1626.

Streisfeld, M.A., Rausher, M.D. 2007. Relaxed constraint and evolutionary rate variation between basic helix-loop-helix floral anthocyanin regulators in Ipomoea. Mol. Biol. Evol. **24**: 2816-26.

Suzuki, H., Saito, R., Tomita, M. 2007. Variation in the Correlation of G + C Composition with Synonymous Codon Usage Bias among Bacteria. J Bioinform Syst Biol.: 61374.

Swanson, W.J., Vacquier, V.D. 2002. The rapid evolution of reproductive proteins. Nat. Rev. Genet. **3**: 137-144.

Swanson, W. J., Z. Yang, M. F. Wolfner, and C. F. Aquadro. 2001. Positive Darwinian selection derives the evolution of several female reproductive proteins in mammals. Proc. Natl. Acad. Sci. USA **98**:2509–2514.

Tao, L., Picataqquio, S., Roviere, P.E., Cheng, Q. 2004. Asymmetrically acting lycopene beta-cyclases (CrtLm) from non-photosynthetic bacteria. Mol. Genet. Genomics. **271**: 180-188.

Tevini, M., Steinmuller, D. 1985. Composition and function of plastoglobuli. Planta. **163**: 1432-2048.

Thompson, J.D., Gibson, T.J., Flewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools Nucleic Acids Research. **25**: 4876-4882.

Tian, L., Musetti, V., Kim, J., Magallanes-Lundback, M., DellaPenna, D. 2004. The Arabidopsis LUT1 locus encodes a member of the cytochrome p450 family that is required for carotenoid epsilon-ring hydroxylation activity. Proc. Natl. Acad. Sci. USA. **101**: 402 - 407.

Wright, F. 1990. The "effective number of codons" used in a gene. Gene **87**:23-29.

Wurtzel, E.T. 2007. *Psy3*, a new member of the phytoene synthase gene family conserved in the Poaceae and regulator of abiotic-stress-induced root carotenogenesis. Plant Physiology Preview. DOI.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. **24**: 1586 – 1591.

Yang, Z., Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. **17**: 32-43.

Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics. **155**: 431-449.

Zhang, L., Vision, T.J., Gaut, B.S. 2002. Patterns of Nucleotide Substitution Among Simultaneously Duplicated Gene Pairs in *Arabidopsis thaliana*. Mol. Biol. Evol. **19**: 1464-1473.

Zuckerkandl, E. 1976. Evolutionary processes and evolutionary noise at the molecular level I. Functional density in proteins. J. Mol. Evol. **9**: 167-183.

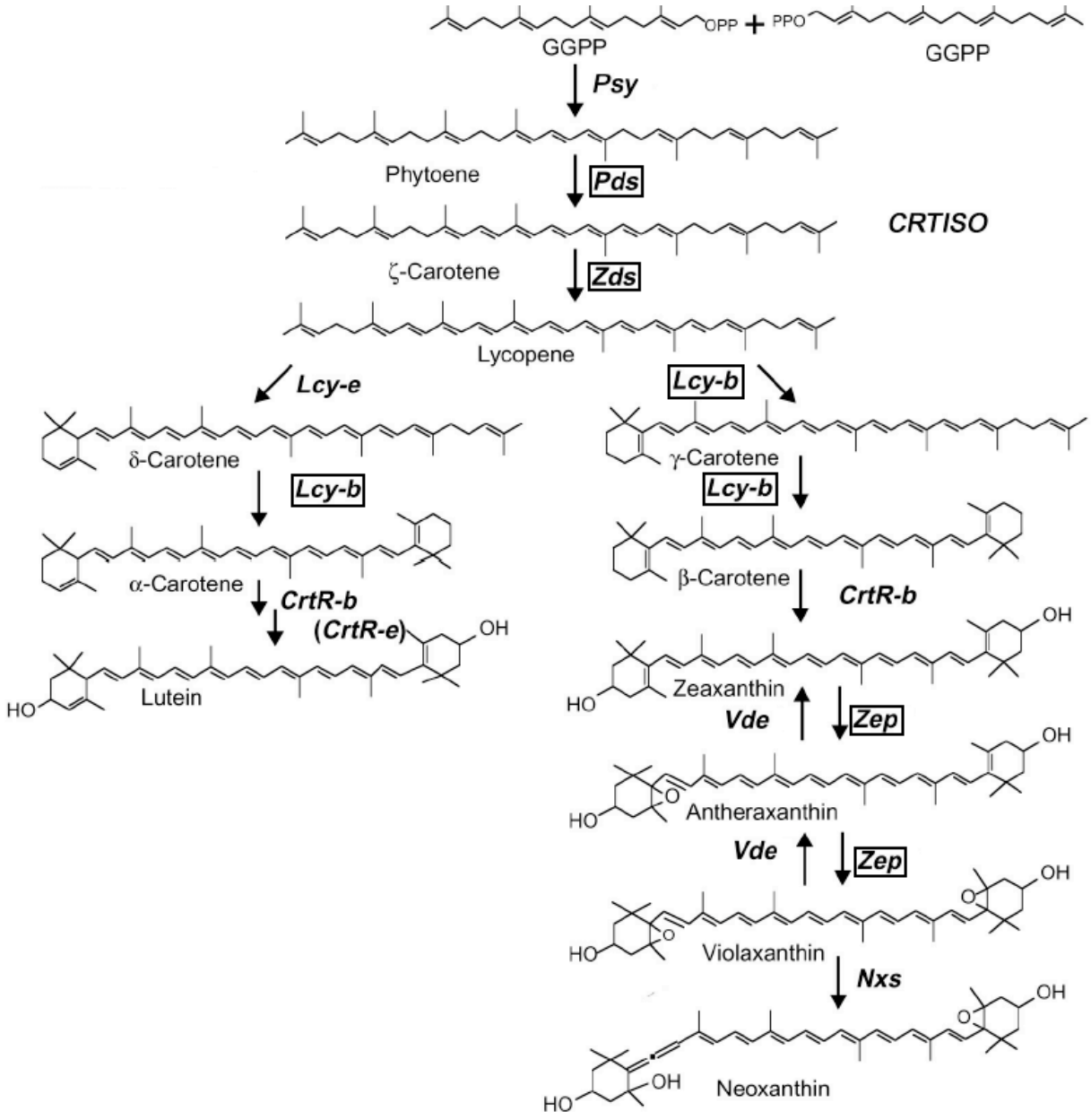


Figure 1. The carotenoid biosynthetic pathway with the carotenoid structures. The enzymes used in this analysis are boxed [Hirschberg (2001)].

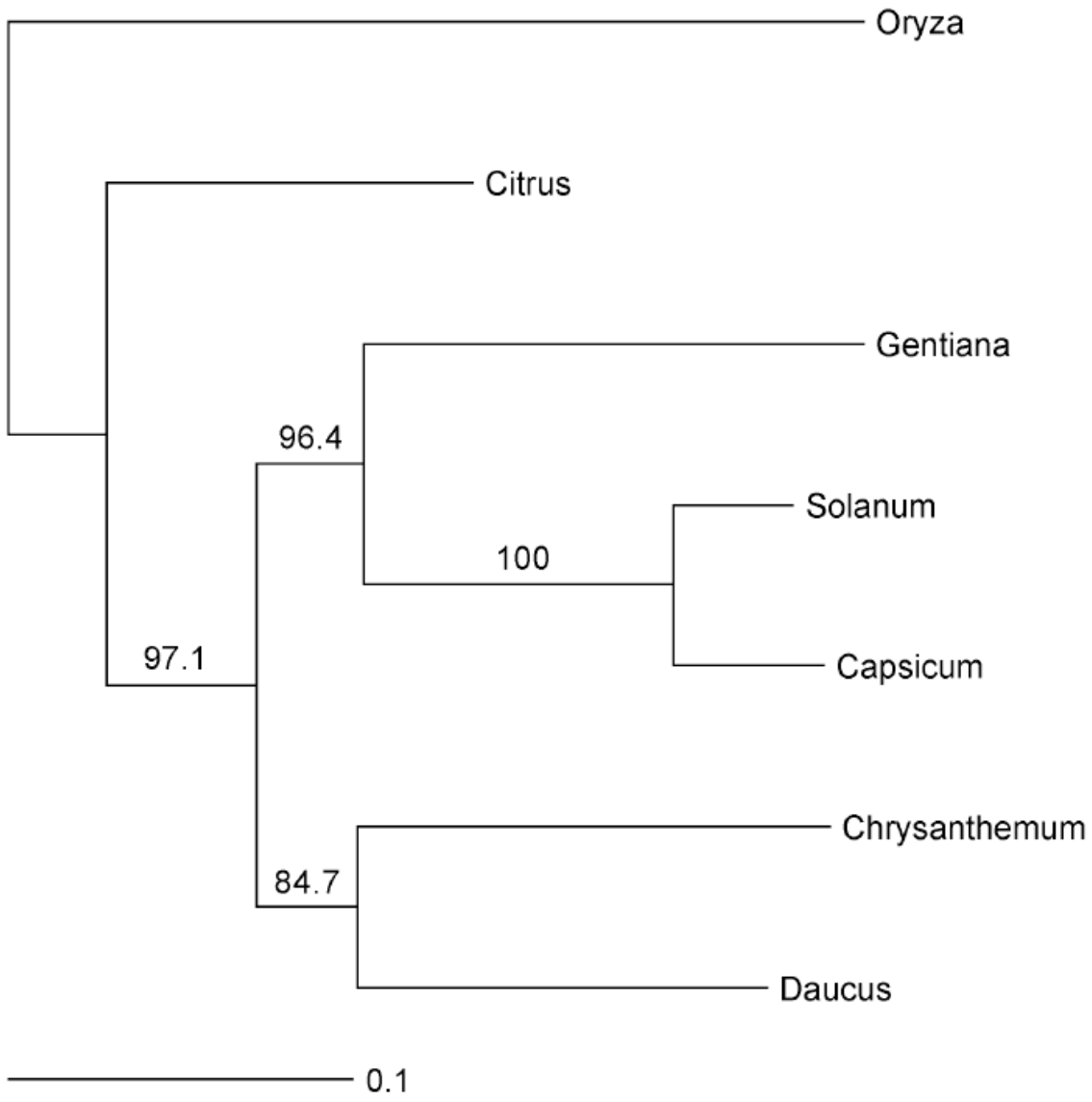


Figure 2. The maximum likelihood tree obtained using the concatenated coding sequences of *PDS, ZDS, βLCY* and *ZEP*. Numbers above the branches are the bootstrap percentages, and the scale bar shows the expected fraction of substitutions. *Oryza sativa* was used as an outgroup.

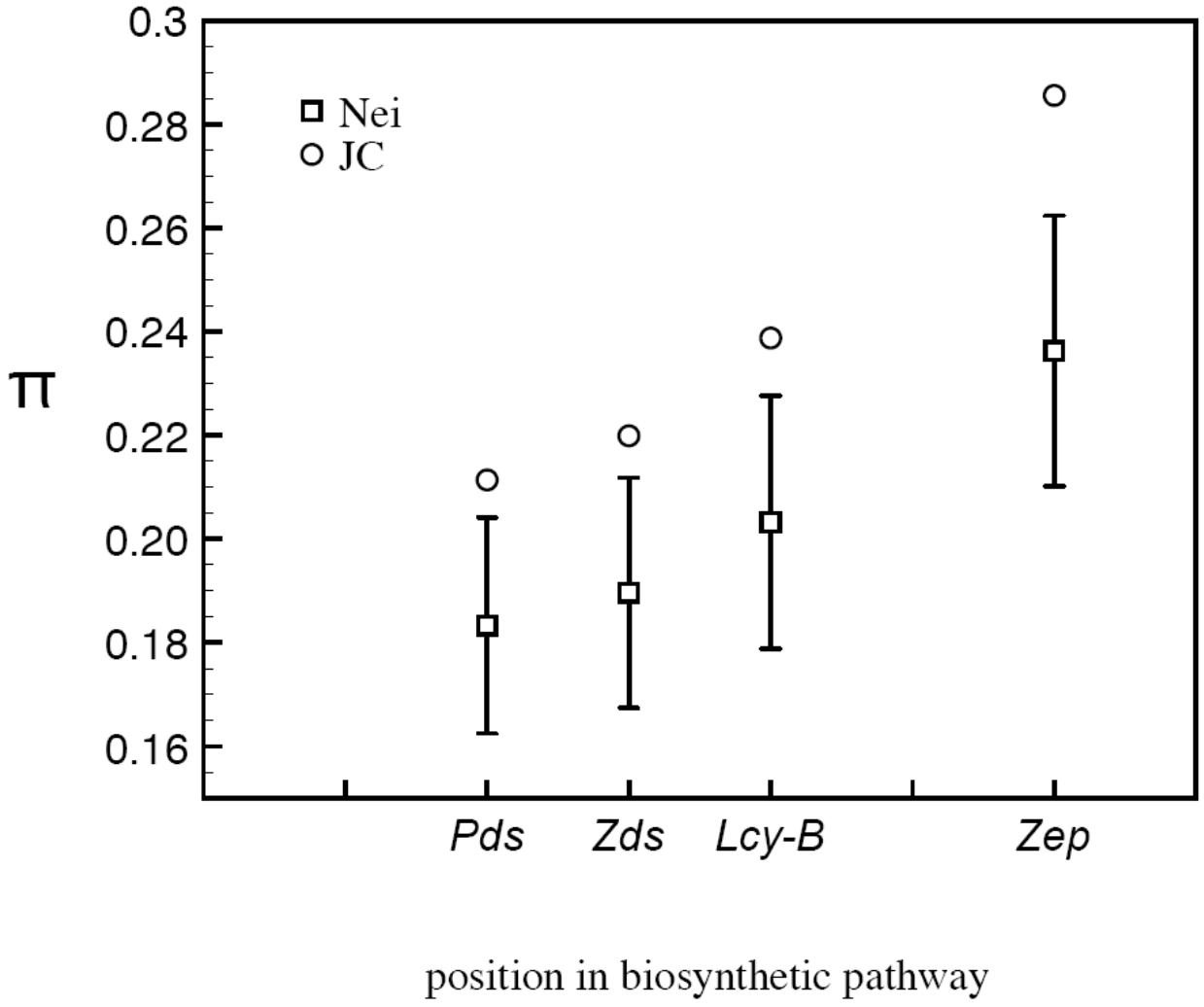


Figure 3. Nucleotide diversity, π, as a function of enzyme position in the carotenoid biosynthetic pathway. Both the Jukes and Cantor (1969) and Nei (1987) methods were used to estimate π, but only the method of Nei allows for calculation of the standard deviation of the estimate.

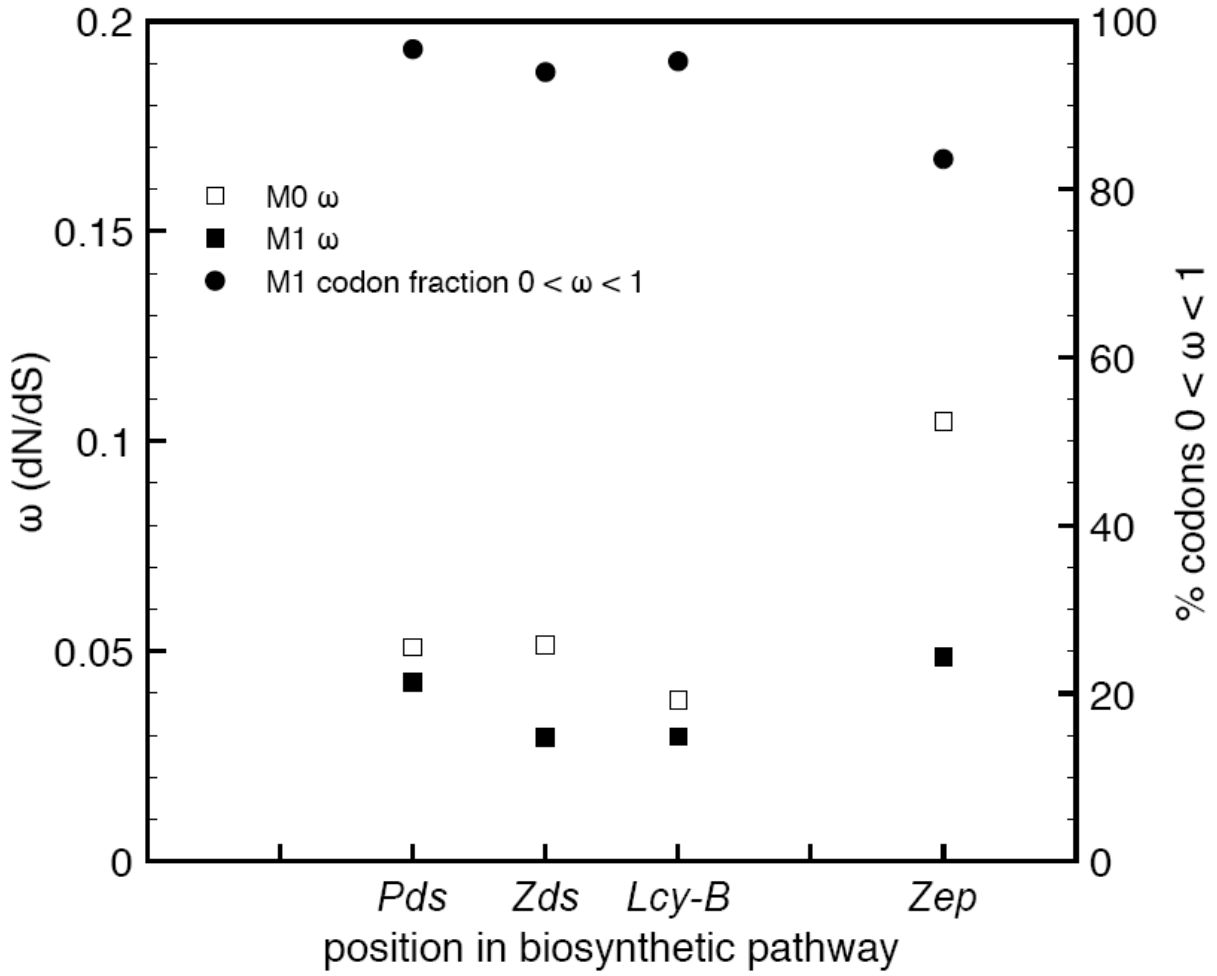


Figure 4. ω as a function of enzyme position in the carotenoid biosynthetic pathway using models M0 and M1a from PAML (Yang 2007). Percent codons with 0< ω <1 as a function of enzyme position in the carotenoid biosynthetic pathway is also shown for the M1a model.

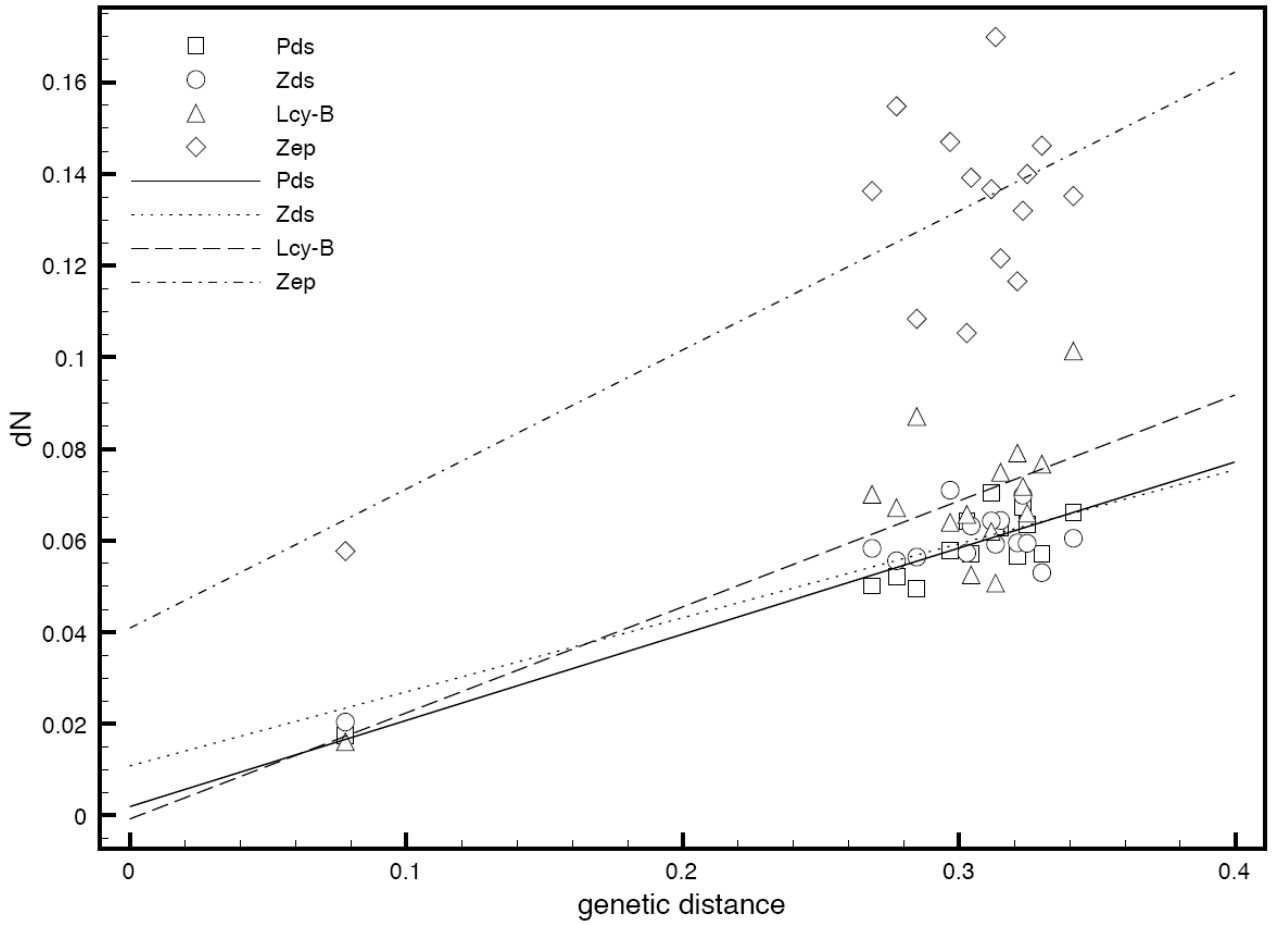


Figure 5. d_N versus genetic distance for each enzyme. Genetic distances were determined using the branch lengths from the maximum-likelihood phylogeny seen in Figure 1. Each point represents one of 15 comparisons between species pairs from the six species analyzed.

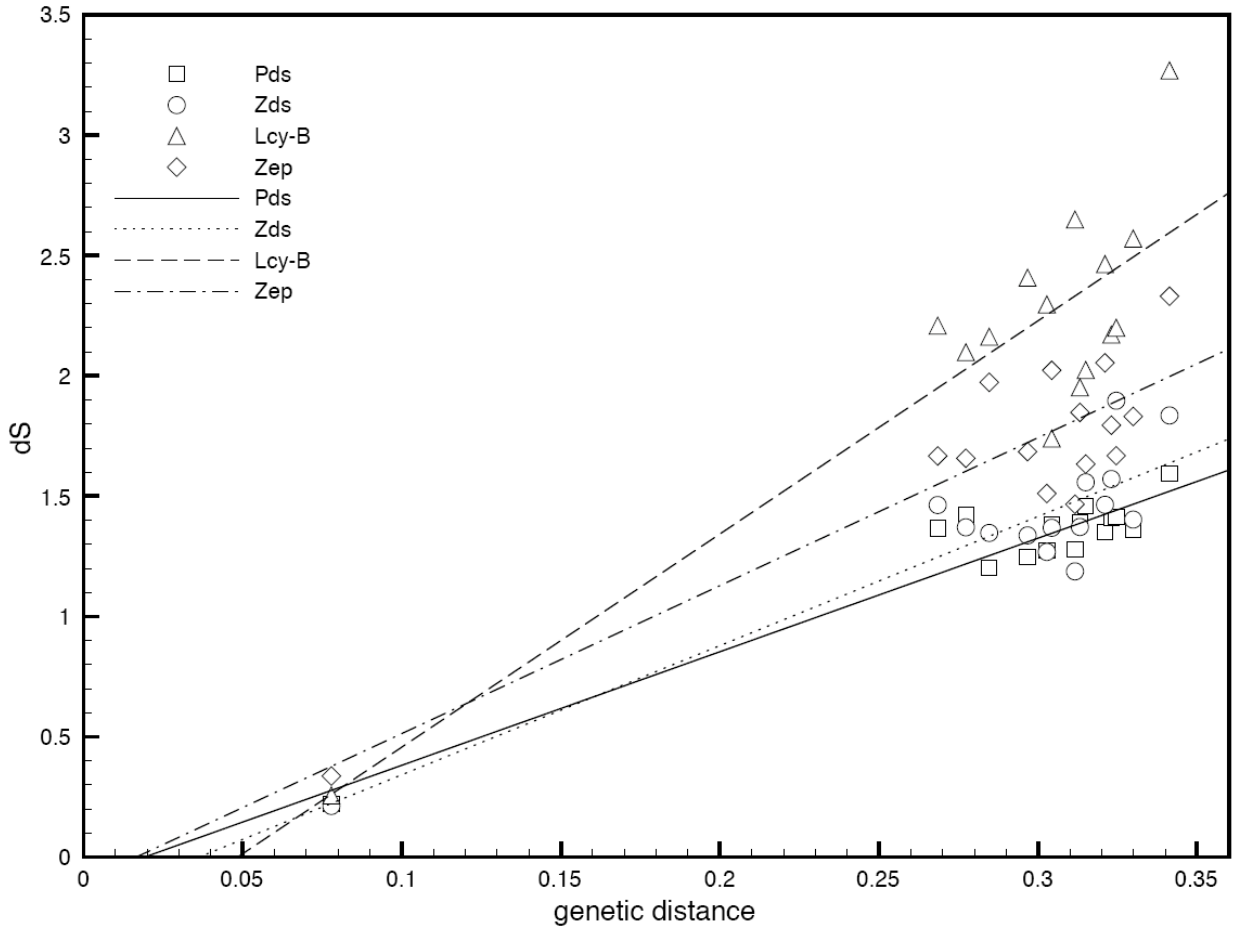


Figure 6. d_S versus genetic distance for each enzyme. Genetic distances were determined using the branch lengths from the maximum-likelihood phylogeny seen in Figure 1. Each point represents one of 15 comparisons between species pairs from the six species analyzed.

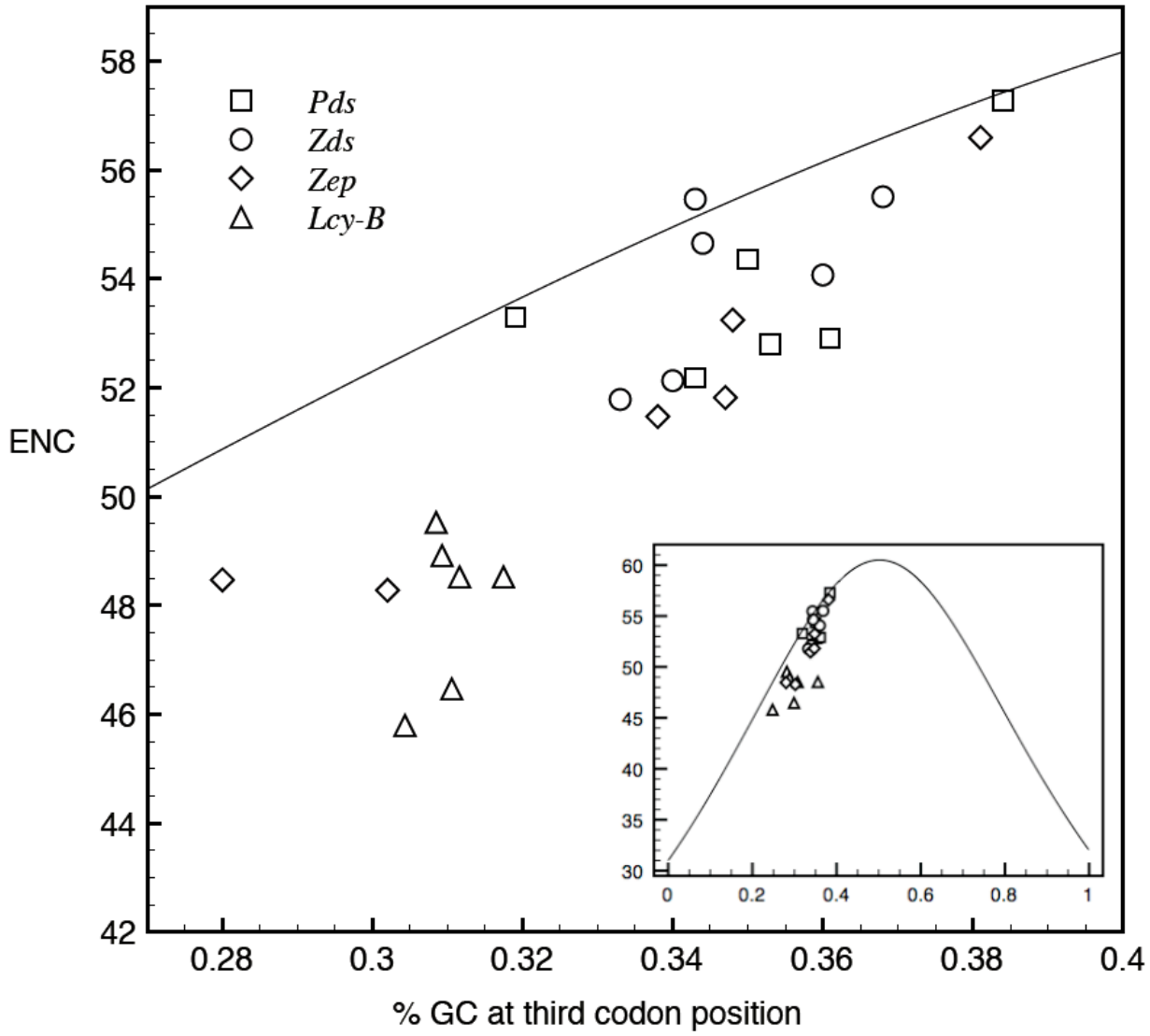


Figure 7. Graph of the effective number of codons (ENC) versus the percent GC at third codon positions. ENC under the assumption of no selection on codon usage (codon bias) is given by the solid curve.

Table 1. Species with base pair ranges used in analysis.

Gene	*PDS*		*ZDS*		*βLCY*		*ZEP*	
Species	ID	bases	ID	bases	ID	bases	ID	bases
S. lycopersicum	M88683	328 – 1611	AF195507	199 – 1694	X86452	160 - 1500	EU004202	250 - 1930
C. spp.	AJ319761	235 – 1518	AJ319762	142 – 1637	AY240787	172 - 1512	AB075547	226 - 1906
C. annuum	X68058	325 – 1608	X89897	199 – 1694	TC916392	419 - 1759	X91491	226 - 1906
G. lutea	AB028665	316 – 1599	AB017370	196 – 1691	EF062505	184 - 1524	AB017368	238 - 1918
D. carota	DQ222429	298 – 1581	DQ222430	169 – 1664	DQ192190	178 - 1518	DQ192197	244 - 1924
Chrysanthemum x morifolium	AB205049	298 - 1581	AB205052	199 – 1694	AB205041	154 - 1494	AB205053	289 - 1969

PDS = 73% of *S. lycopersicum* sequence
ZDS = 85% of *S. lycopersicum* sequence
βLCY = 89% of *S. lycopersicum* sequence
ZEP = 81% of *S. lycopersicum* sequence

Table 2. Synonymous and nonsynonymous substitution rates.

Slope	*PDS*	*ZDS*	*βLCY*	*ZEP*
d_N vs. genetic distance	0.188(0.020)	0.162(0.025)	0.231(0.054)	0.304(0.0780)
d_S vs. genetic distance	4.721(0.426)	5.371(0.735)	8.863(1.380)	6.151(0.936)

First two rows are regression coefficients of d_N and d_S versus genetic distance.
Standard errors are in parentheses.

Table 3. Mean Effective Number of Codons (ENC) and GC percentage at the third codon position values.

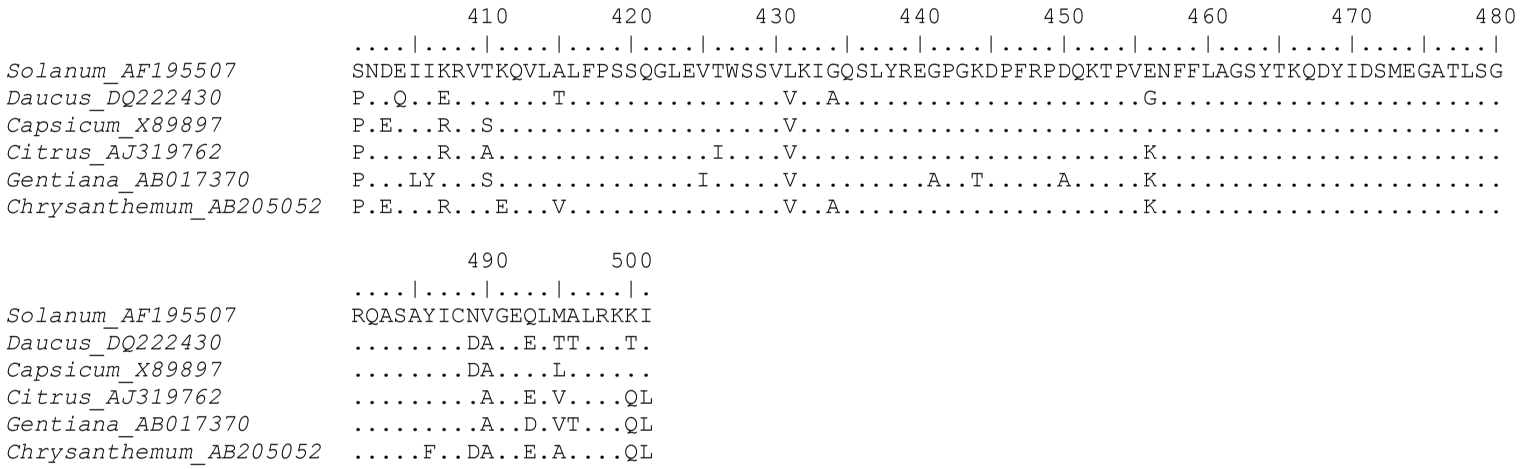
	PDS	***ZDS***	***βLCY***	***ZEP***
ENC	53.8(1.36)	53.9(1.23)	47.9(1.22)	51.8(1.76)
GC3	0.352(0.14)	0.348(0.11)	0.310(0.17)	0.333(0.19)

Standard errors are in parentheses. Points that lie on or above the expected curve fall within the null Model prediction. Points below the expected curve predict deviation from the null model of no codon preferences (Novembre 2002).

Appendices

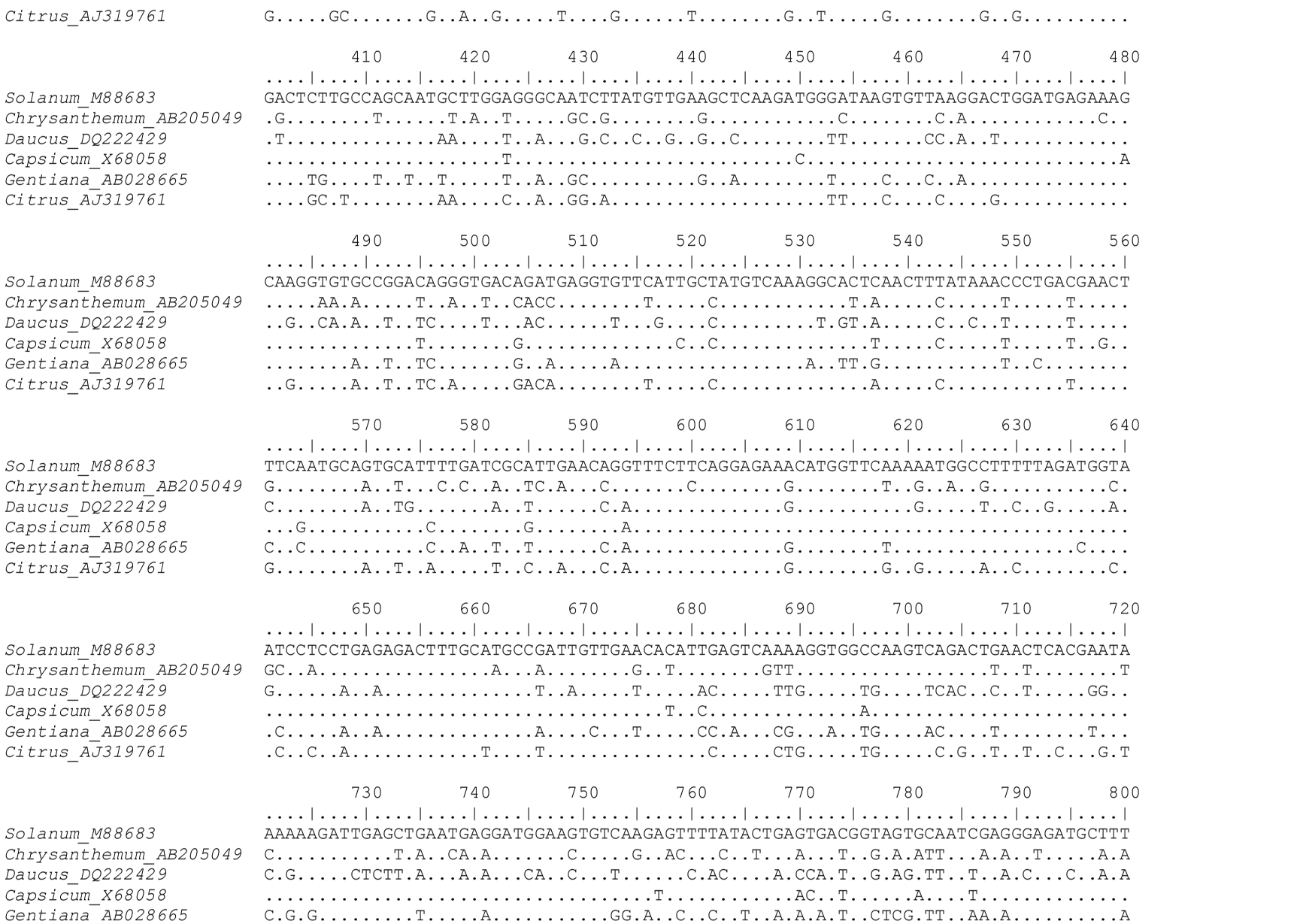
```
                                10        20        30        40        50        60        70        80
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_AF195507       LFPPEPEHYRGPKLKVAIIGAGLAGMSTAVELLDQGHEVDIYESRTFIGGKVGSFVDRRGNHIEMGLHVFFGCYNNLFRL
Daucus_DQ222430        ......QL.....................................P.........T.K......................
Capsicum_X89897        ..............................................................K.....................
Citrus_AJ319762        .....................................................S.....K.....................
Gentiana_AB017370      .....................................................P.....KG....................
Chrysanthemum_AB205052 ..........................................................KG....................

                                90       100       110       120       130       140       150       160
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_AF195507       LKKVGAEKNLLVKEHTHTFVNKGGEIGELDFRFPVGAPLHGINAFLSTNQLKIYDKARNAVALALSPVVRALVDPDGALQ
Daucus_DQ222430        ......D.....................................T.....T.....................MK
Capsicum_X89897        M.....................................................T.....................
Citrus_AJ319762        M.....D......D.......Q............I.......R.........T.......L........K.........K
Gentiana_AB017370      M............D.........................R...T.....P.....................R
Chrysanthemum_AB205052 ............D.......R.....................T.....T.....................MT

                               170       180       190       200       210       220       230       240
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_AF195507       QIRDLDNVSFSEWFLSKGGTRASIQRMWDPVAYALGFIDCDNMSARCMLTIFALFATKTEASLLRMLKGSPDVYLSGPIK
Daucus_DQ222430        D..N...I.............K.........................S...................T......R
Capsicum_X89897        ......S....D..M.........................I.....................................
Citrus_AJ319762        D.....SI...D.........QT.....................I.....................................R
Gentiana_AB017370      N.....SI...N.........................I.....................................
Chrysanthemum_AB205052 ...S..........M.R.....................I.........S.....................R

                               250       260       270       280       290       300       310       320
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_AF195507       KYIMDKGGRFHLRWGCREVLYETSSDGSMYVSGLAMSKATQKKIVKADAYVAACDVPGIKRLVPQKWRELEFFDNIYKLV
Daucus_DQ222430        D..TQ.............I...K....QT.I..I.........................L.SQ...W..........
Capsicum_X89897        ...I.....................................................V..................G.....I
Citrus_AJ319762        ...T.............I..DKAANAET..K........D..V.Q.................L.SS...MKI.N...A..
Gentiana_AB017370      D..IN.W............R.D..DN.DT..T............IH...................RQ...WD.........
Chrysanthemum_AB205052 D..IE.............I...K.DN.DT..T..................I...........L.SS..DW....D.....

                               330       340       350       360       370       380       390       400
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_AF195507       GVPVVTVQLRYNGWVTELQDLERSRQLKRAAGLDNLLYTPDADFSCFADLALASPDDYYIEGQGSLLQCVLTPGDPYMPL
Daucus_DQ222430        .................M....K....RQ.........................E...L.....................
Capsicum_X89897        ............................S...T.........................E.....................
Citrus_AJ319762        ..........................R..L.....................T..E...R.....................
Gentiana_AB017370      ..........................RK.....................S..EN.....................
Chrysanthemum_AB205052 .................M..I..A..SRK.T.....................T..E....D...................
```

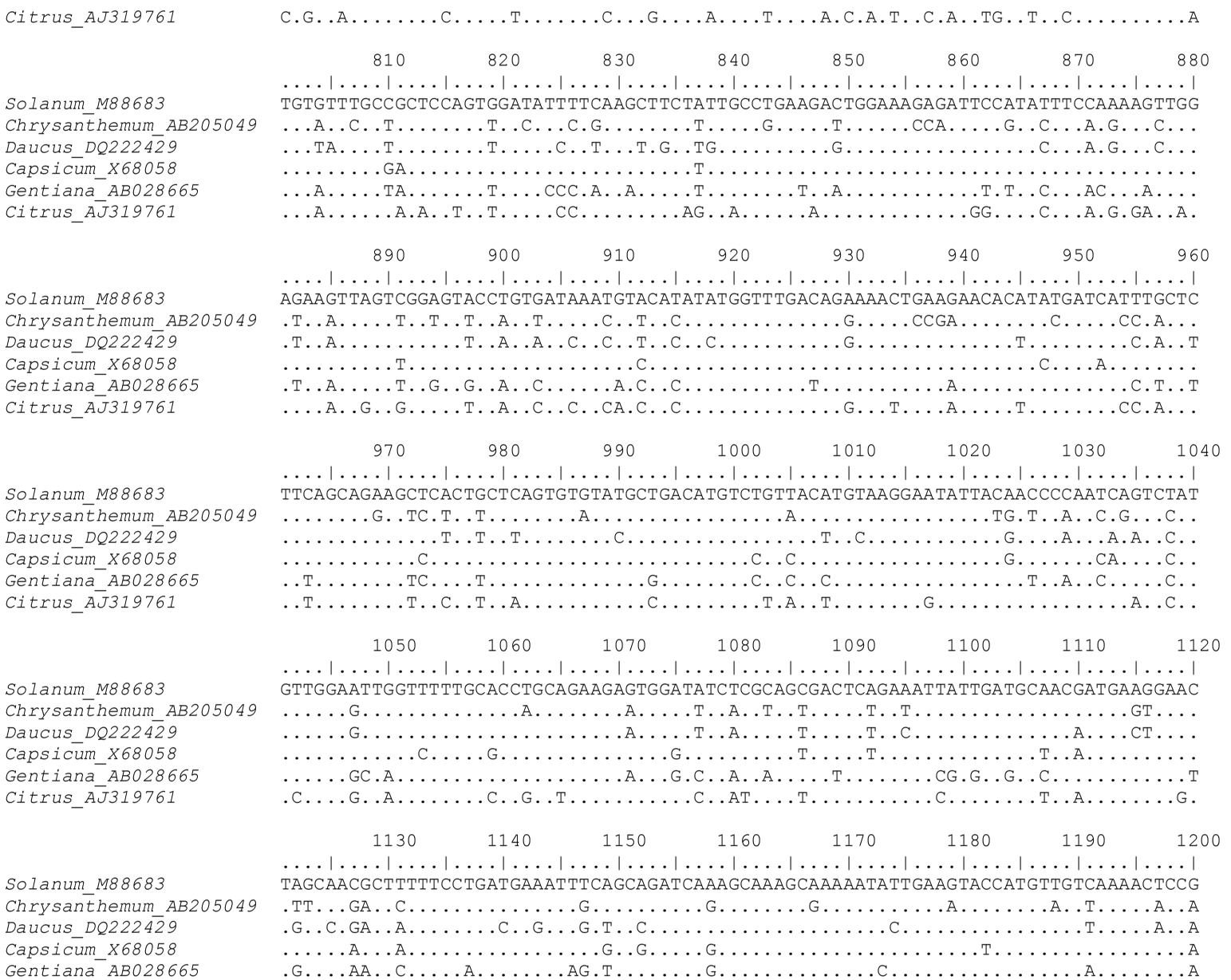
Appendix 1. ClustalX peptide sequence alignment for *ZDS*. Genus names and accession numbers are at left.


410 420 430 440 450 460 470 480
Solanum_AF195507 SNDEIIKRVTKQVLALFPSSQGLEVTWSSVLKIGQSLYREGPGKDPFRPDQKTPVENFFLAGSYTKQDYIDSMEGATLSG
Daucus_DQ222430 P..Q..E.......T.............V..A.....................G.....................
Capsicum_X89897 P.E...R..S.................V...
Citrus_AJ319762 P.....R..A.............I....V.....................K.....................
Gentiana_AB017370 P...LY...S............I.....V.........A..T.....A.....K.....................
Chrysanthemum_AB205052 P.E...R...E...V.............V..A.....................K.....................
490 500
Solanum_AF195507 RQASAYICNVGEQLMALRKKI
Daucus_DQ222430 ........DA..E.TT...T.
Capsicum_X89897 ........DA....L......
Citrus_AJ319762 .........A..E.V....QL
Gentiana_AB017370 .........A..D.VT...QL
Chrysanthemum_AB205052 .....F..DA..E.A....QL

```
                          10        20        30        40        50        60        70        80
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_M88683      AAACCATTGGAGATTGTTATTGCTGGTGCAGGTTTGGGTGGTTTGTCTACAGCAAAATATTTGGCAGATGCTGGTCACAA
Chrysanthemum_AB205049 ..G..G...A.C..C..C..............C.C.C......A..A..T.....G........T.............
Daucus_DQ222429     ........AA.TG....C...........T.......C......A.....T.........................
Capsicum_X68058     ......C.....................................................................
Gentiana_AB028665   ........A...G.G..C..C..G............C.......A....T.....G..C..A..G.....A.....T..
Citrus_AJ319761     ..G..G...A.AG....A..................C......A..A..T...................A..C.....

                          90       100       110       120       130       140       150       160
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_M88683      ACCGATACTGCTGGAGGCAAGGGATGTTCTAGGTGGAAAGGTAGCTGCATGGAAAGATGATGATGGAGATTGGTACGAGA
Chrysanthemum_AB205049 ...A..CT..T.A..A.....A..C.....C...........G..G..T........CA.G..................
Daucus_DQ222429     G..C...T..T....A..G..A........T..C........G.....T.............C.................
Capsicum_X68058     ...A.....................................................................T....
Gentiana_AB028665   ...T..T...T....A.....A......T............T...........G...........G..C.....T..A.
Citrus_AJ319761     G..TT.GT.A.....A.....A.........................C..........GG..C..GA.C.....T....

                         170       180       190       200       210       220       230       240
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_M88683      CTGGTTTGCATATATTCTTTGGGGCTTACCCAAATATTCAGAACCTGTTTGGAGAATTAGGGATTAACGATCGATTGCAA
Chrysanthemum_AB205049 .G.....A..C.....T.....A............G.A.................G.....C.....T...A.......G
Daucus_DQ222429     ....A........T.....................G.....................GC....A........CA.......G
Capsicum_X68058     ..........C...........................G........A..................A..T............
Gentiana_AB028665   .A..C..A.................A.........G.G......T....C.....GC....A.................G
Citrus_AJ319761     .A..CC.T.....T.....C.................A...................C.T..T.....T.....G.....G

                         250       260       270       280       290       300       310       320
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_M88683      TGGAAGGAACATTCAATGATATTTGCAATGCCAAGCAAGCCAGGAGAATTCAGCCGCTTTGATTTCTCCGAAGCTTTACC
Chrysanthemum_AB205049 ........G.....T.....................A......T........T..T........C...C.A..T.T.C.G..
Daucus_DQ222429     ........G.....T........C..T.....T.A.........G.....T..T..A........TC.G....T.C....
Capsicum_X68058     ..............G.....................A..................................C.............
Gentiana_AB028665   .....A........T...........T.......AT..A..T...........T..A........TG.A..G.T......
Citrus_AJ319761     ........G..C..T.....T.............A......C..............A........TC.T....T.C.T..

                         330       340       350       360       370       380       390       400
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_M88683      CGCTCCTTTAAATGGAATTTTAGCCATCTTAAAGAATAACGAAATGCTTACATGGCCAGAGAAAGTCAAATTTGCAATTG
Chrysanthemum_AB205049 T..A..A..G........A.GG..T...C.G.G...C..T........G........G........A..G..........
Daucus_DQ222429     T..A...C..........A.GG..T...........C..T........A........T...........G.....GT.G.
Capsicum_X68058     T...................G..A...C.......C..T....................A...................
Gentiana_AB028665   T..A..A.....C.....A.GG.....T..G.....C..T...........T.....T........T..G..........
```

Appendix 2. ClustalX mRNA sequence alignment for *PDS*. Genus names and accession numbers are at left

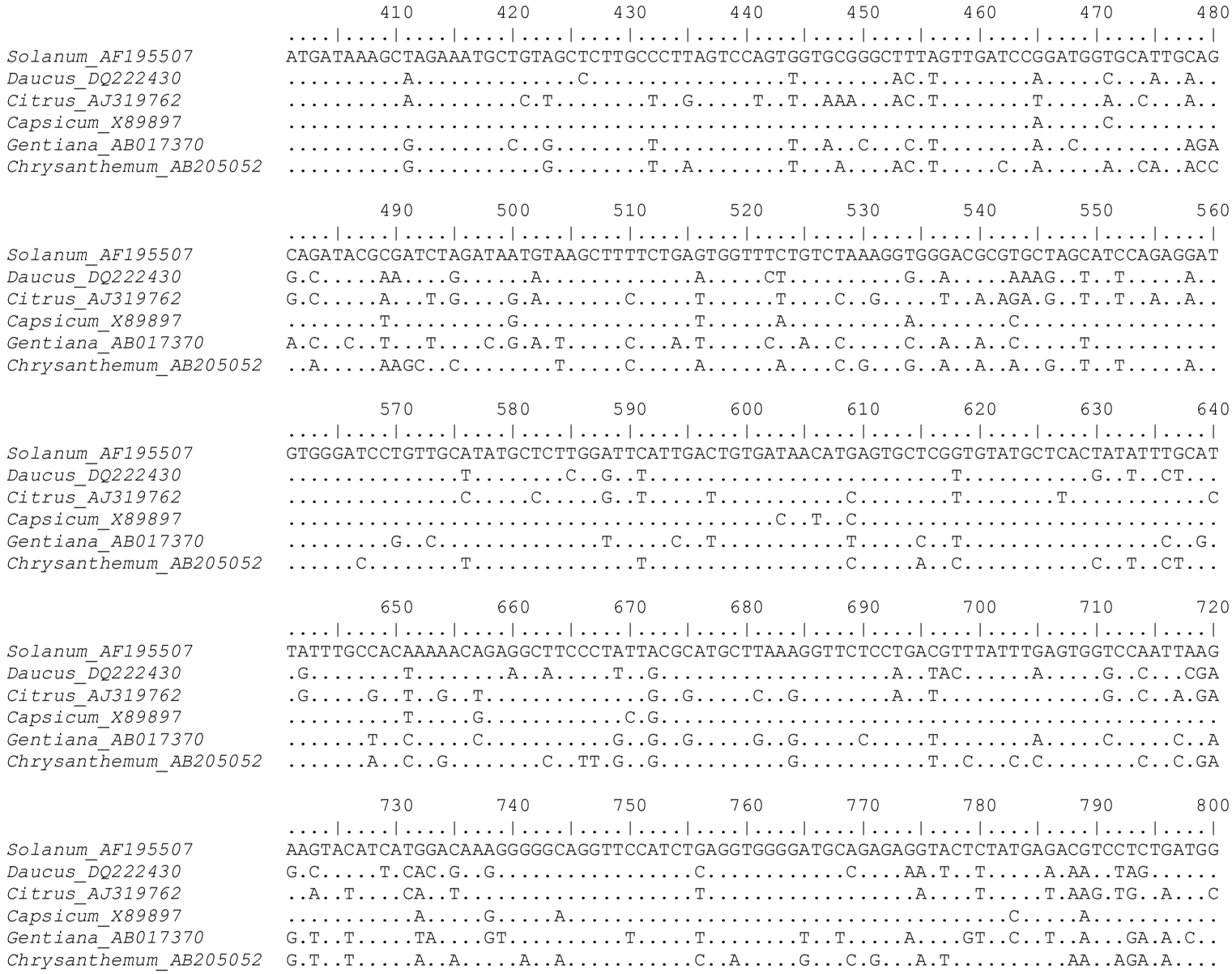

Citrus_AJ319761 G.....GC.......G..A..G.....T....G......T........G..T.....G........G..G.....
410 420 430 440 450 460 470 480
Solanum_M88683 GACTCTTGCCAGCAATGCTTGGAGGGCAATCTTATGTTGAAGCTCAAGATGGGATAAGTGTTAAGGACTGGATGAGAAAG
Chrysanthemum_AB205049 .G........T......T.A..T.....GC.G........G............C........C.A............C..
Daucus_DQ222429 .T..............AA....T..A...G.C..C..G..G..C.........TT.......CC.A..T.....
Capsicum_X68058 ......................T.........................C.............................A
Gentiana_AB028665 ....TG....T..T..T.....T..A..GC..........G..A........T....C...C..A.....
Citrus_AJ319761 ....GC.T........AA....C..A..GG.A.....................TT...C....C....G.....
490 500 510 520 530 540 550 560
Solanum_M88683 CAAGGTGTGCCGGACAGGGTGACAGATGAGGTGTTCATTGCTATGTCAAAGGCACTCAACTTTATAAACCCTGACGAACT
Chrysanthemum_AB205049 .....AA.A.....T..A..T..CACC.........T.....C............T.A.....C.....T.....T.....
Daucus_DQ222429 ..G..CA.A..T..TC....T...AC......T...G....C..........T.GT.A.....C..C..T.....T.....
Capsicum_X68058 ..............T........G.............C..C..............T.....C.....T.....T..G..
Gentiana_AB028665 ........A..T..TC.......G..A.....A..................A..TT.G...........T..C........
Citrus_AJ319761 ..G.....A..T..TC.A.....GACA........T.....C..............A.....C...........T.....
570 580 590 600 610 620 630 640
Solanum_M88683 TTCAATGCAGTGCATTTTGATCGCATTGAACAGGTTTCTTCAGGAGAAACATGGTTCAAAAATGGCCTTTTTAGATGGTA
Chrysanthemum_AB205049 G........A..T...C.C..A..TC.A...C.......C.........G.........T..G..A..G...........C.
Daucus_DQ222429 C........A..TG.......A..T......C.A..............G............G.....T..C..G.....A.
Capsicum_X68058 ...G...........C........G........A...
Gentiana_AB028665 C..C...........C..A..T..T......C.A..............G........T..................C....
Citrus_AJ319761 G........A..T..A.....T..C..A...C.A..............G........G..G.....A..C........C.
650 660 670 680 690 700 710 720
Solanum_M88683 ATCCTCCTGAGAGACTTTGCATGCCGATTGTTGAACACATTGAGTCAAAAGGTGGCCAAGTCAGACTGAACTCACGAATA
Chrysanthemum_AB205049 GC..A................A...A........G..T........GTT..................T..T.........T
Daucus_DQ222429 G......A..A..............T..A.....T.....AC.....TTG.....TG....TCAC..C..T.....GG..
Capsicum_X68058 ....................................T..C..............A.....................
Gentiana_AB028665 .C.....A..A..............A....C...T.....CC.A...CG...A..TG....AC....T........T...
Citrus_AJ319761 .C..C..A............T....T..............C.....CTG.....TG.....C.G..T..T..C...G.T
730 740 750 760 770 780 790 800
Solanum_M88683 AAAAAGATTGAGCTGAATGAGGATGGAAGTGTCAAGAGTTTTATACTGAGTGACGGTAGTGCAATCGAGGGAGATGCTTT
Chrysanthemum_AB205049 C............T.A..CA.A........C.....G..AC...C..T...A...T..G.A.ATT...A.A..T.....A.A
Daucus_DQ222429 C.G.....CTCTT.A...A.A...CA..C...T......C.AC....A.CCA.T..G.AG.TT..T..A.C...C..A.A
Capsicum_X68058 ....................................T.............AC..T......A....T..............
Gentiana_AB028665 C.G.G.........T.....A...........GG.A..C..C..T..A.A.A.T..CTCG.TT..AA.A..........A

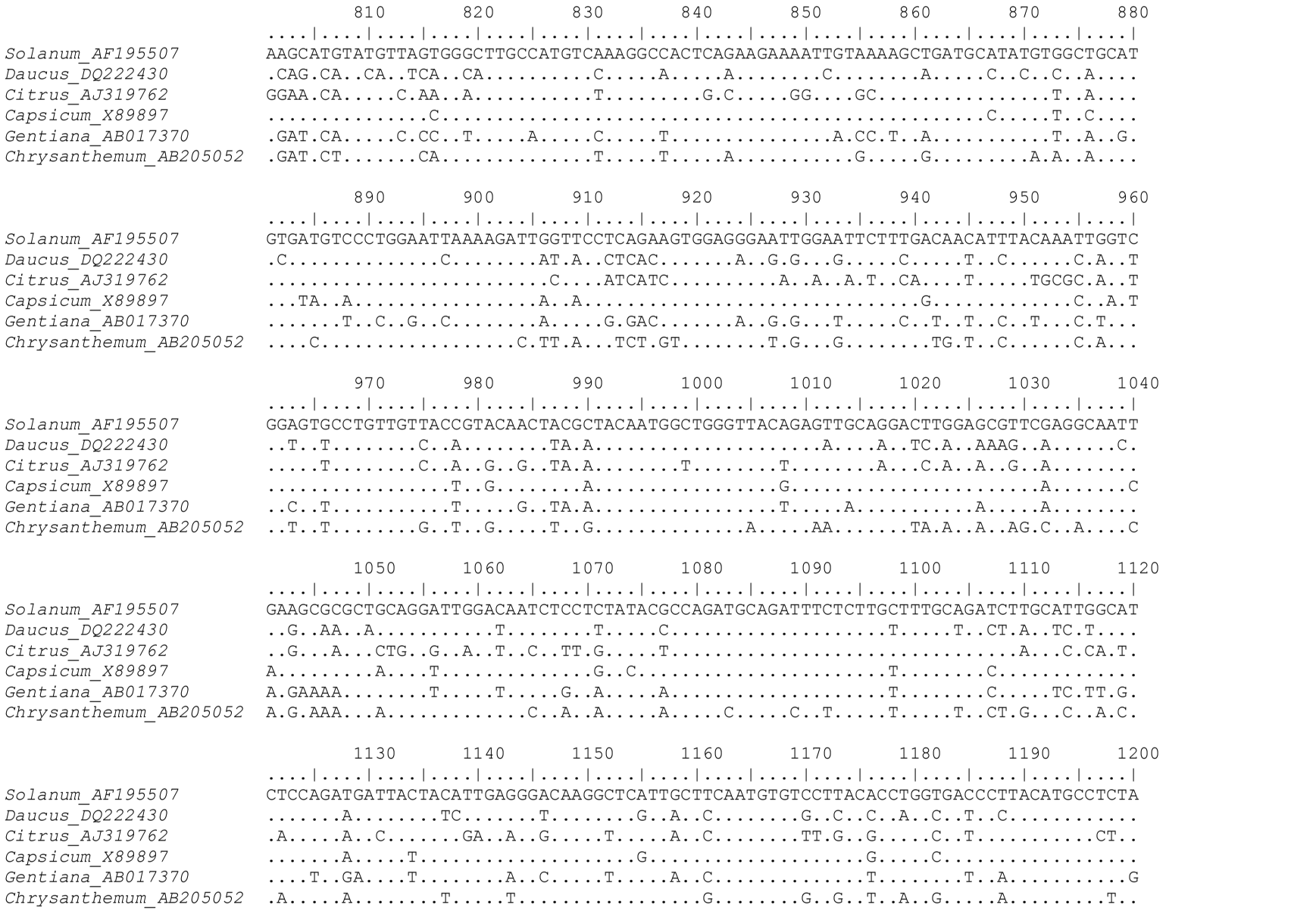

Citrus_AJ319761 C.G..A.....C.....T.....C...G....A....T....A.C.A.T..C.A..TG..T..C.....A
810 820 830 840 850 860 870 880
Solanum_M88683 TGTGTTTGCCGCTCCAGTGGATATTTTCAAGCTTCTATTGCCTGAAGACTGGAAAGAGATTCCATATTTCCAAAAGTTGG
Chrysanthemum_AB205049 ...A..C..T.....T..C...C.G.....T.....G.....T.....CCA.....G..C...A.G...C...
Daucus_DQ222429 ...TA....T.....T.....C..T...T.G..TG.....G.....C...A.G...C...
Capsicum_X68058 .....GA.....T.....
Gentiana_AB028665 ...A.....TA.....T....CCC.A..A.....T.....T..A.....T.T..C...AC...A....
Citrus_AJ319761 ...A.....A.A..T..T.....CC.....AG..A.....A.....GG....C...A.G.GA..A.
890 900 910 920 930 940 950 960
Solanum_M88683 AGAAGTTAGTCGGAGTACCTGTGATAAATGTACATATATGGTTTGACAGAAAACTGAAGAACACATATGATCATTTGCTC
Chrysanthemum_AB205049 .T..A.....T..T..T..A..T.....C..T..C.....G.....CCGA.....C.....CC.A...
Daucus_DQ222429 .T..A.....T..A..A..C..C..T..C..C.....G.....T.....C.A..T
Capsicum_X68058 .....T.....C.....C....A.....
Gentiana_AB028665 .T..A.....T..G..G..A..C.....A.C..C.....T.....A.....C.T..T
Citrus_AJ319761 ....A..G..G.....T..A..C..C..CA.C..C.....G...T....A.....T.....CC.A...
970 980 990 1000 1010 1020 1030 1040
Solanum_M88683 TTCAGCAGAAGCTCACTGCTCAGTGTGTATGCTGACATGTCTGTTACATGTAAGGAATATTACAACCCCAATCAGTCTAT
Chrysanthemum_AB205049 .....G..TC.T..T.....A.....A.....TG.T..A..C.G...C..
Daucus_DQ222429 .....T..T..T.....C.....T..C.....G....A...A.A..C..
Capsicum_X68058 .....C.....C..C.....G.....CA....C..
Gentiana_AB028665 ..T.....TC....T.....G.....C..C..C.....T..A..C.....C..
Citrus_AJ319761 ..T.....T..C..T..A.....C.....T.A..T.....G.....A..C..
1050 1060 1070 1080 1090 1100 1110 1120
Solanum_M88683 GTTGGAATTGGTTTTTGCACCTGCAGAAGAGTGGATATCTCGCAGCGACTCAGAAATTATTGATGCAACGATGAAGGAAC
Chrysanthemum_AB205049 ......G.....A.....A.....T..A..T..T.....T..T.....GT....
Daucus_DQ222429 ......G.....A.....T..A.....T.....T..C.....A....CT....
Capsicum_X68058 .....C.....G.....G.....T.....T.....T..A.....
Gentiana_AB028665 ......GC.A.....A...G.C..A..A.....T.....CG.G..G..C.....T
Citrus_AJ319761 .C....G..A.....C..G..T.....C..AT....T.....C.....T..A.....G.
1130 1140 1150 1160 1170 1180 1190 1200
Solanum_M88683 TAGCAACGCTTTTTCCTGATGAAATTTCAGCAGATCAAAGCAAAGCAAAAATATTGAAGTACCATGTTGTCAAAACTCCG
Chrysanthemum_AB205049 .TT...GA..C.....G.....G.....G.....A.....A..T.....A..A
Daucus_DQ222429 .G..C.GA..A.....C..G...G.T..C.....C.....T.....A..A
Capsicum_X68058 ......A...A.....G..G.....G.....T.....A
Gentiana_AB028665 .G....AA..C.....A.....AG.T.....G.....C.....A.....A

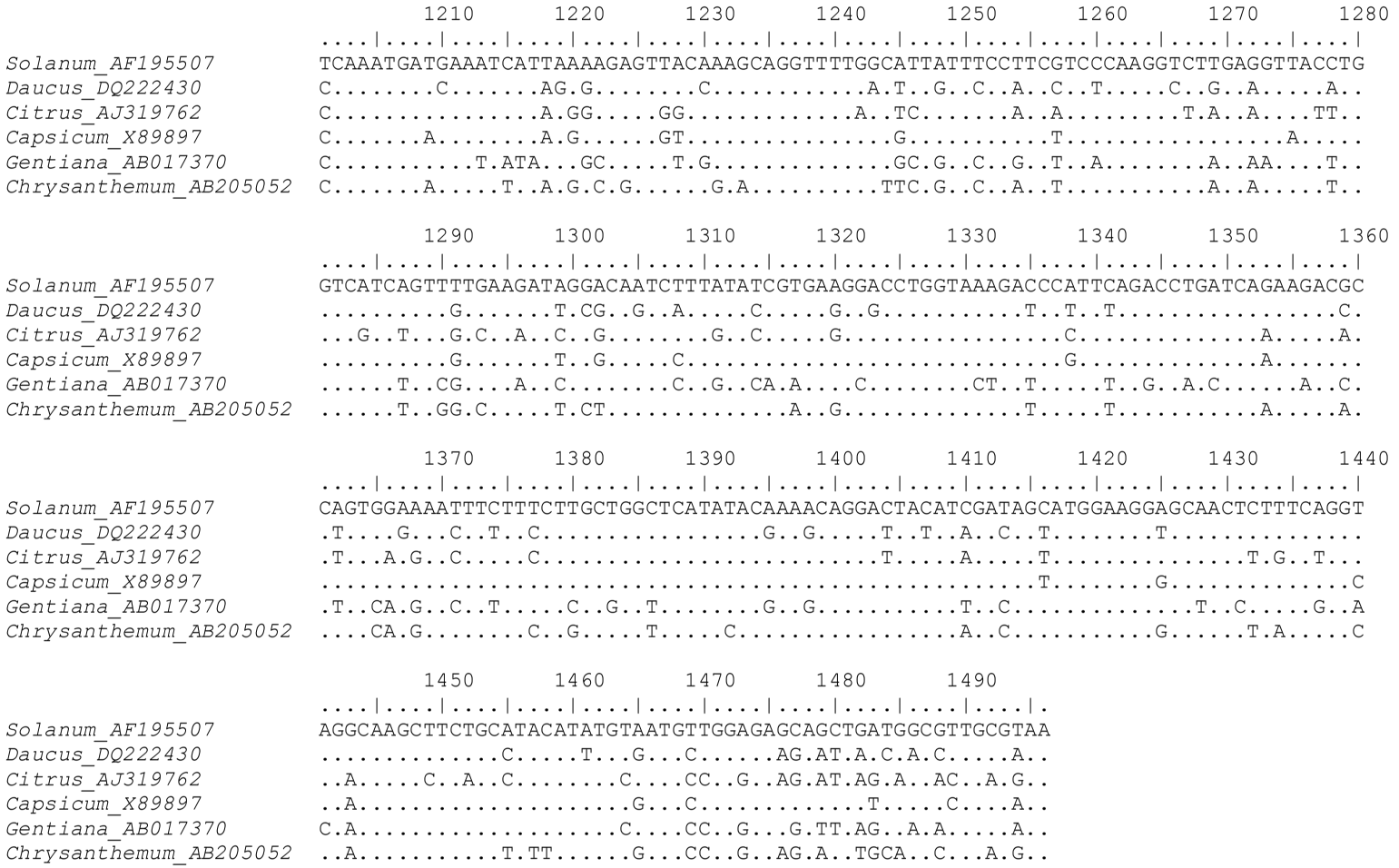
```
Citrus_AJ319761          .T....AA..A.................T..T.....G..........G..TG............C........G..A

                                 1210      1220      1230      1240      1250      1260      1270      1280
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_M88683           AGGTCTGTTTATAAAACTGTGCCAGGTTGTGAACCCTGTCGGCCTTTACAAAGATCCCCAATAGAGGGGTTTTATTTAGC
Chrysanthemum_AB205049   ...........C.....A..A....AC........T..C..T..A..G........T........A..A..C........
Daucus_DQ222429          ................A.A....AC..........C..T.....G....A...T..C.....A..A..........
Capsicum_X68058          ........A.....................................TC..G...........TG..................
Gentiana_AB028665        .....A.............C..G..CACG.....T..C..T..........G..T......A.A..A..C........
Citrus_AJ319761          ........A........CA.C...AA.........T..C..T..C........G..T..TG....A..............


                         ....
Solanum_M88683           CGGT
Chrysanthemum_AB205049   T..A
Daucus_DQ222429          T...
Capsicum_X68058          T...
Gentiana_AB028665        T...
Citrus_AJ319761          ...G
```

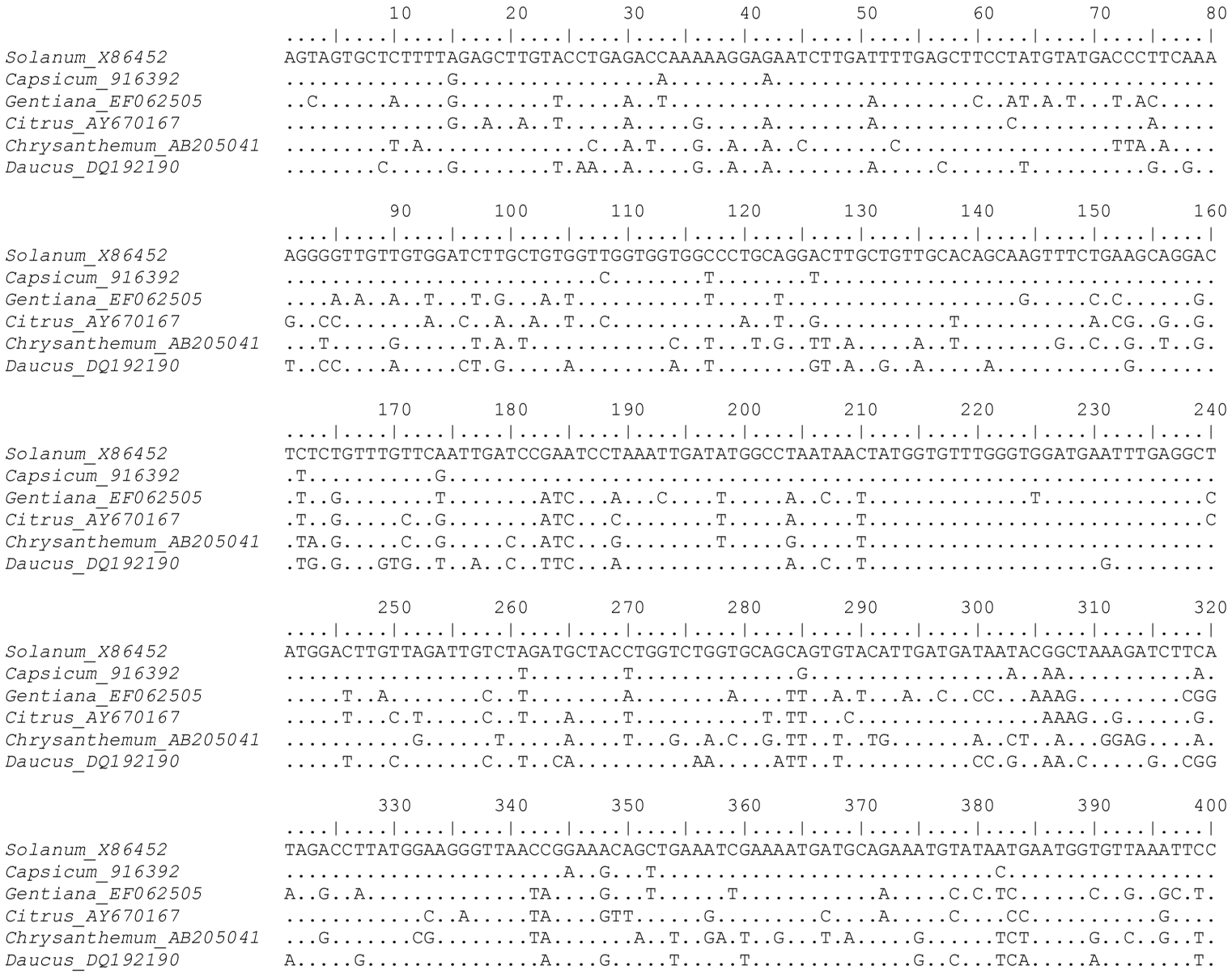
```
                                10        20        30        40        50        60        70        80
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_AF195507               CTATTTCCACCCGAGCCTGAACATTATCGGGGGCCAAAGCTGAAAGTAGCTATTATTGGAGCTGGGCTTGCAGGCATGTC
Daucus_DQ222430                T.......G..G..A...C.G.T......T..A..G.....A.....G........A..............G..A.....
Citrus_AJ319762                T.G........A..A..A.........A.A..A..............G.....................G........
Capsicum_X89897                ..........T..A.....................................................C.............
Gentiana_AB017370              T.G........G..A.................A..C...T.A.....T.....................T.....
Chrysanthemum_AB205052         T.............................T.....C...T.......G.....C..A..T.....C.....T..T.....

                                  90       100       110       120       130       140       150       160
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_AF195507               GACTGCTGTGGAGCTCTTGGATCAAGGACATGAGGTGGATATATACGAATCAAGGACTTTTATTGGTGGGAAAGTGGGTT
Daucus_DQ222430                T..C.....T.....TC.A.....G..C..C..............T.........C.............A..........
Citrus_AJ319762                A..G..A.....AT.G...........C..C.................G......T.............T..........
Capsicum_X89897                ..........................................T...........C..C...................
Gentiana_AB017370              A.....A........TC.A..............................T...C....C..............C..A.
Chrysanthemum_AB205052         A.....G........T.............................T..G........G..C........A..........

                                 170       180       190       200       210       220       230       240
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_AF195507               CTTTTGTTGATAGACGTGGGAACCACATTGAAATGGGACTGCACGTGTTCTTTGGTTGTTATAATAATCTGTTCCGTCTG
Daucus_DQ222430                ....CAC.....A...C..A..T..T..A...........T..T..A...........C..C........T........C
Citrus_AJ319762                .A.....C....A......A.....T...........C........T........A..C..C..............AT..
Capsicum_X89897                ............A.....................................................C...........A.........
Gentiana_AB017370              ....C.......A.G....A..T.................A..T..C...........C..C..C.....T.........
Chrysanthemum_AB205052         .G..........A.G....A.....T..............A..T..T...........C..C..C.....T..T.....T

                                 250       260       270       280       290       300       310       320
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_AF195507               TTGAAAAAGGTGGGTGCTGAAAAAAATCTGCTAGTGAAGGAGCATACTCACACATTTGTAAATAAAGGGGGTGAAATAGG
Daucus_DQ222430                ...........T..G.....T........T..C.....................T........C..............T..
Citrus_AJ319762                A..........C.....G..C......T.A..T........T........T............C.G..T........T..
Capsicum_X89897                A.....................................................................................
Gentiana_AB017370              A..........A...........G.................T..C........T..............A..G.....C..
Chrysanthemum_AB205052         .....G.....T........G..G..C..T..T..C.....T...........T.....G..C.G...A..A.....T..

                                 330       340       350       360       370       380       390       400
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_AF195507               GGAACTTGATTTCCGCTTTCCAGTTGGAGCACCCTTACATGGAATTAATGCATTTCTGTCTACTAATCAGTTAAAGATTT
Daucus_DQ222430                T..G........T..T..C..............A.....C..C..A.............A....C.......C.C....C..
Citrus_AJ319762                T..............G..C...A........T..G........G...CGC......T....G..A......C.T....C..
Capsicum_X89897                ...G.............................G.....................T.............AC......C..
Gentiana_AB017370              C...........T..A..C..T.....T.....GC.T.........ACG...T...T..A....A..C..AC.T...CC..
Chrysanthemum_AB205052         C.....................C...........A........G.....C..T..CT.AA....A......C.T....CA.
```

Appendix 3. ClustalX mRNA sequence alignment for *ZDS*. Genus names and accession numbers are at left.

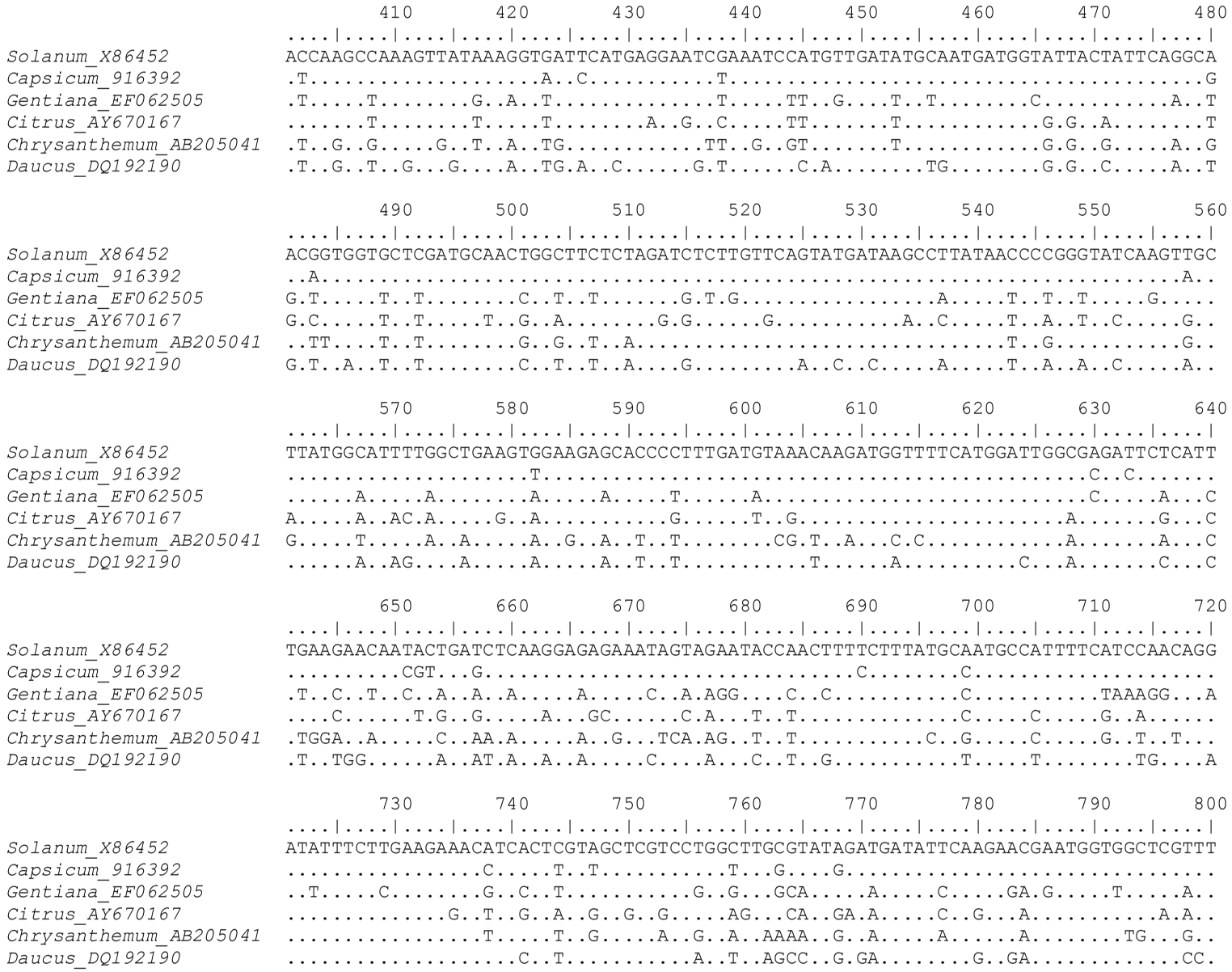

410 420 430 440 450 460 470 480
Solanum_AF195507 ATGATAAAGCTAGAAATGCTGTAGCTCTTGCCCTTAGTCCAGTGGTGCGGGCTTTAGTTGATCCGGATGGTGCATTGCAG
Daucus_DQ222430 ..........A.............C.................T........AC.T........A.....C...A..A..
Citrus_AJ319762 ..........A.........C.T........T..G.....T..T..AAA...AC.T........T.....A..C...A..
Capsicum_X89897 ..........A.....C.........
Gentiana_AB017370 ..........G........C..G........T...........T..A..C...C.T........A..C.........AGA
Chrysanthemum_AB205052 ..........G...........G........T..A........T...A....AC.T.....C..A.....A..CA..ACC
490 500 510 520 530 540 550 560
Solanum_AF195507 CAGATACGCGATCTAGATAATGTAAGCTTTTCTGAGTGGTTTCTGTCTAAAGGTGGGACGCGTGCTAGCATCCAGAGGAT
Daucus_DQ222430 G.C.....AA....G......A.............A.....CT..........G..A.....AAAG..T..T.....A..
Citrus_AJ319762 G.C.....A...T.G....G.A.......C.....T......T....C..G.....T..A.AGA.G..T..T..A..A..
Capsicum_X89897 .........T..........G...............T......A..........A.........C..................
Gentiana_AB017370 A.C..C..T...T....C.G.A.T.....C...A.T.....C..A..C.....C..A..A..C.....T...........
Chrysanthemum_AB205052 ..A.....AAGC..C.........T.....C.....A......A....C.G...G..A..A..A..G..T..T.....A..
570 580 590 600 610 620 630 640
Solanum_AF195507 GTGGGATCCTGTTGCATATGCTCTTGGATTCATTGACTGTGATAACATGAGTGCTCGGTGTATGCTCACTATATTTGCAT
Daucus_DQ222430 ...............T.........C..G..T............................T...........G..T..CT...
Citrus_AJ319762 ...............C.....C.....G..T.....T...........C.........T.........T............C
Capsicum_X89897 .....................................C..T..C.............................
Gentiana_AB017370 .........G..C..............T.....C..T...........T.....C..T.................C..G.
Chrysanthemum_AB205052 ......C.........T..............T..................C.....A..C...........C..T..CT...
650 660 670 680 690 700 710 720
Solanum_AF195507 TATTTGCCACAAAAACAGAGGCTTCCCTATTACGCATGCTTAAAGGTTCTCCTGACGTTTATTTGAGTGGTCCAATTAAG
Daucus_DQ222430 .G........T.........A..A.....T..G.....................A..TAC......A.....G..C...CGA
Citrus_AJ319762 .G.....G..T..G..T..............G..G.....C..G........A..T..............G..C..A.GA
Capsicum_X89897 ..........T.....G.............C.G.....
Gentiana_AB017370 .......T..C.....C...........G..G..G.....G..G.....C.....T........A.....C.....C..A
Chrysanthemum_AB205052 .......A..C..G........C..TT.G..G...........G...........T..C...C.C........C..C.GA
730 740 750 760 770 780 790 800
Solanum_AF195507 AAGTACATCATGGACAAAGGGGGCAGGTTCCATCTGAGGTGGGGATGCAGAGAGGTACTCTATGAGACGTCCTCTGATGG
Daucus_DQ222430 G.C.....T.CAC.G..G..................C............C....AA.T..T.....A.AA..TAG......
Citrus_AJ319762 ..A..T....CA..T.....................T.....................A....T.....T.AAG.TG..A...C
Capsicum_X89897 ...........A.....G.....A.....C.....A.....
Gentiana_AB017370 G.T..T.....TA....GT..........T.....T........T..T.....A....GT..C..T..A...GA.A.C..
Chrysanthemum_AB205052 G.T..T.....A..A.....A..A...........C..A.....G...C.G...A.T..........AA..AGA.A....

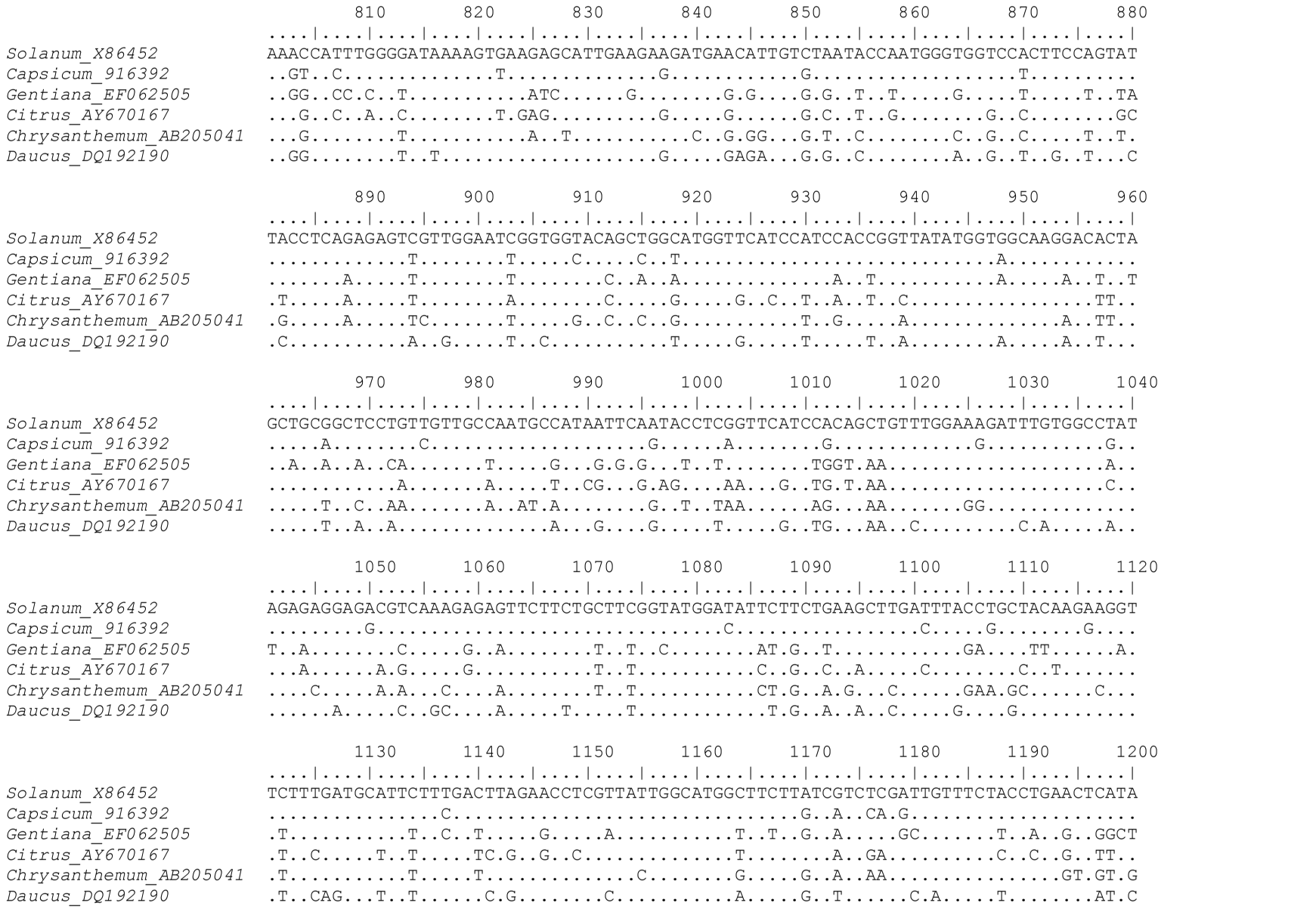

810 820 830 840 850 860 870 880
Solanum_AF195507 AAGCATGTATGTTAGTGGGCTTGCCATGTCAAAGGCCACTCAGAAGAAAATTGTAAAAGCTGATGCATATGTGGCTGCAT
Daucus_DQ222430 .CAG.CA..CA..TCA..CA.....C.....A.....A.....C.....A.....C..C..C..A....
Citrus_AJ319762 GGAA.CA.....C.AA..A.....T.....G.C.....GG....GC.....T..A....
Capsicum_X89897 .....C.....C.....T..C....
Gentiana_AB017370 .GAT.CA.....C.CC..T.....A.....C.....T.....A.CC.T..A.....T..A..G.
Chrysanthemum_AB205052 .GAT.CT.....CA.....T.....T.....A.....G.....G.....A.A..A....
890 900 910 920 930 940 950 960
Solanum_AF195507 GTGATGTCCCTGGAATTAAAAGATTGGTTCCTCAGAAGTGGAGGGAATTGGAATTCTTTGACAACATTTACAAATTGGTC
Daucus_DQ222430 .C.....C.....AT.A..CTCAC.....A..G.G...G.....C.....T..C.....C.A..T
Citrus_AJ319762 .....C....ATCATC.....A..A..A.T..CA....T.....TGCGC.A..T
Capsicum_X89897 ...TA..A.....A..A.....G.....C..A.T
Gentiana_AB017370 .....T..C..G..C.....A.....G.GAC.....A..G.G...T.....C..T..T..C..T...C.T...
Chrysanthemum_AB205052 ....C.....C.TT.A...TCT.GT.....T.G...G.....TG.T..C.....C.A...
970 980 990 1000 1010 1020 1030 1040
Solanum_AF195507 GGAGTGCCTGTTGTTACCGTACAACTACGCTACAATGGCTGGGTTACAGAGTTGCAGGACTTGGAGCGTTCGAGGCAATT
Daucus_DQ222430 ..T..T.....C..A.....TA.A.....A....A..TC.A..AAAG..A.....C.
Citrus_AJ319762 .....T.....C..A..G..G..TA.A.....T.....T.....A...C.A..A..G..A.....
Capsicum_X89897 .....T..G.....A.....G.....A.....C
Gentiana_AB017370 ..C..T.....T.....G..TA.A.....T.....A.....A.....A.....
Chrysanthemum_AB205052 ..T..T.....G..T..G.....T..G.....A.....AA.....TA.A..A..AG.C..A....C
1050 1060 1070 1080 1090 1100 1110 1120
Solanum_AF195507 GAAGCGCGCTGCAGGATTGGACAATCTCCTCTATACGCCAGATGCAGATTTCTCTTGCTTTGCAGATCTTGCATTGGCAT
Daucus_DQ222430 ..G..AA..A.....T.....T.....C.....T.....T..CT.A..TC.T....
Citrus_AJ319762 ..G...A...CTG..G..A..T..C..TT.G.....T.....A...C.CA.T.
Capsicum_X89897 A.....A....T.....G..C.....T.....C.....
Gentiana_AB017370 A.GAAAA.....T.....T.....G..A.....A.....T.....C.....TC.TT.G.
Chrysanthemum_AB205052 A.G.AAA...A.....C..A..A.....A.....C.....C..T.....T.....T..CT.G...C..A.C.
1130 1140 1150 1160 1170 1180 1190 1200
Solanum_AF195507 CTCCAGATGATTACTACATTGAGGGACAAGGCTCATTGCTTCAATGTGTCCTTACACCTGGTGACCCTTACATGCCTCTA
Daucus_DQ222430 .....A.....TC.....T.....G..A..C.....G..C..C..A..C..T..C.....
Citrus_AJ319762 .A.....A..C.....GA..A..G.....T.....A..C.....TT.G..G.....C..T.....CT..
Capsicum_X89897 .....A.....T.....G.....G.....C.....
Gentiana_AB017370 ....T..GA....T.....A..C.....T.....A..C.....T.....T..A.....G
Chrysanthemum_AB205052 .A.....A.....T.....T.....G.....G..G..T..A..G.....A.....T..


1210 1220 1230 1240 1250 1260 1270 1280
Solanum_AF195507 TCAAATGATGAAATCATTAAAAGAGTTACAAAGCAGGTTTTGGCATTATTTCCTTCGTCCCAAGGTCTTGAGGTTACCTG
Daucus_DQ222430 C.....C.....AG.G.....C.....A.T..G..C..A..C..T.....C..G..A.....A..
Citrus_AJ319762 C.....A.GG.....GG.....A..TC.....A..A.....T.A..A....TT..
Capsicum_X89897 C.....A.....A.G.....GT.....G.....T.....A.....
Gentiana_AB017370 C.....T.ATA...GC.....T.G.....GC.G..C..G..T..A.....A..AA....T..
Chrysanthemum_AB205052 C.....A.....T..A.G.C.G.....G.A.....TTC.G..C..A..T.....A..A.....T..
1290 1300 1310 1320 1330 1340 1350 1360
Solanum_AF195507 GTCATCAGTTTTGAAGATAGGACAATCTTTATATCGTGAAGGACCTGGTAAAGACCCATTCAGACCTGATCAGAAGACGC
Daucus_DQ222430 .....G.....T.CG..G..A.....C.....G..G.....T..T..T.....C.
Citrus_AJ319762 ...G..T...G.C..A..C..G.....G..C.....G.....C.....A.....A.
Capsicum_X89897 .....G.....T..G.....C.....G.....A.....
Gentiana_AB017370 .....T..CG....A..C.....C..G..CA.A....C.....CT..T.....T..G..A.C.....A..C.
Chrysanthemum_AB205052 .....T..GG.C.....T.CT.....A..G.....T.....T.....A.....A.
1370 1380 1390 1400 1410 1420 1430 1440
Solanum_AF195507 CAGTGGAAAATTTCTTTCTTGCTGGCTCATATACAAAACAGGACTACATCGATAGCATGGAAGGAGCAACTCTTTCAGGT
Daucus_DQ222430 .T....G...C..T..C.....G..G.....T..T..A..C..T.....T.....
Citrus_AJ319762 .T...A.G..C.....C.....T.....A.....T.....T.G..T...
Capsicum_X89897 .....T.....G.....C
Gentiana_AB017370 .T..CA.G..C..T.....C..G..T.....G..G.....T..C.....T..C.....G..A
Chrysanthemum_AB205052 ....CA.G.....C..G.....T.....C.....A..C.....G.....T.A.....C
1450 1460 1470 1480 1490
Solanum_AF195507 AGGCAAGCTTCTGCATACATATGTAATGTTGGAGAGCAGCTGATGGCGTTGCGTAA
Daucus_DQ222430 .....C.....T...G...C.....AG.AT.A.C.A.C.....A..
Citrus_AJ319762 ..A.....C..A..C.....C....CC..G..AG.AT.AG.A..AC..A.G..
Capsicum_X89897 ..A.....G...C.....T.....C....A..
Gentiana_AB017370 C.A.....C....CC..G...G.TT.AG..A.A.....A..
Chrysanthemum_AB205052 ..A.....T.TT.....G...CC..G..AG.A..TGCA..C...A.G..



Appendix 4. ClustalX mRNA sequence alignment for *βLCY*. Common names and accession numbers are at left.

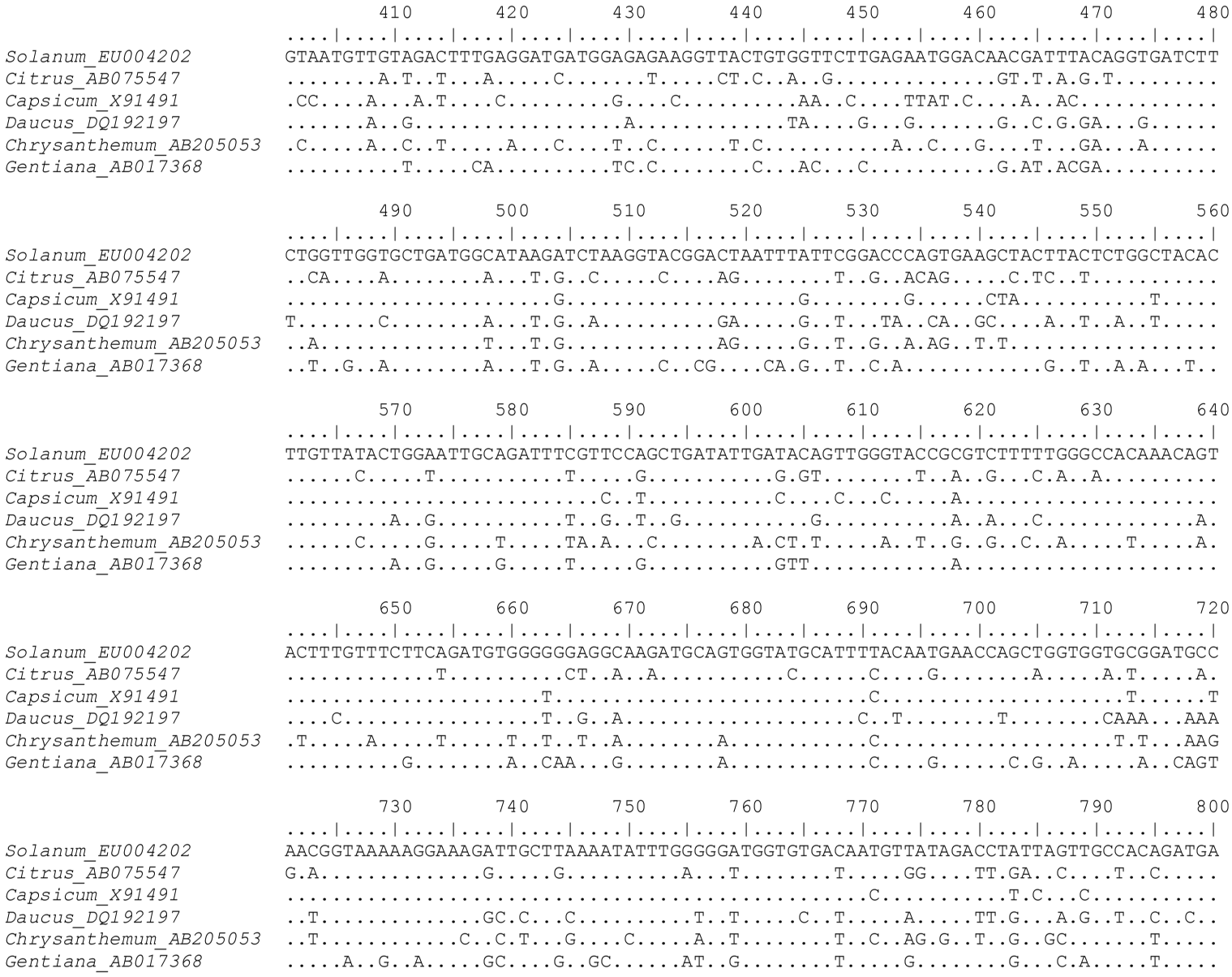

410 420 430 440 450 460 470 480
Solanum_X86452 ACCAAGCCAAAGTTATAAAGGTGATTCATGAGGAATCGAAATCCATGTTGATATGCAATGATGGTATTACTATTCAGGCA
Capsicum_916392 .T.....A..C.....T.....G
Gentiana_EF062505 .T.....T.....G..A..T.....T.....TT..G....T..T.....C.....A..T
Citrus_AY670167 .....T.....T.....T.....A..G..C.....TT.....T.....G.G..A.....T
Chrysanthemum_AB205041 .T..G..G.....G..T..A..TG.....TT..G..GT.....T.....G.G..G.....A..G
Daucus_DQ192190 .T..G..T..G...G....A..TG.A..C.....G.T.....C.A.....TG.....G.G..C.....A..T
490 500 510 520 530 540 550 560
Solanum_X86452 ACGGTGGTGCTCGATGCAACTGGCTTCTCTAGATCTCTTGTTCAGTATGATAAGCCTTATAACCCCGGGTATCAAGTTGC
Capsicum_916392 ..A.....A..
Gentiana_EF062505 G.T.....T..T.....C..T..T.....G.T.G.....A.....T..T..T.....G.....
Citrus_AY670167 G.C.....T..T.....T..G..A.....G.G.....G.....A..C.....T..A..T..C.....G..
Chrysanthemum_AB205041 ..TT....T..T.....G..G..T..A.....T..G.....G..
Daucus_DQ192190 G.T..A..T..T.....C..T..T..A....G.....A..C..C.....A.....T..A..A..C.....A..
570 580 590 600 610 620 630 640
Solanum_X86452 TTATGGCATTTTGGCTGAAGTGGAAGAGCACCCCTTTGATGTAAACAAGATGGTTTTCATGGATTGGCGAGATTCTCATT
Capsicum_916392 .....T.....C..C.....
Gentiana_EF062505 .....A.....A.....A.....A.....T.....A.....C.....A...C
Citrus_AY670167 A.....A..AC.A.....G..A.....G.....T..G.....A.....G...C
Chrysanthemum_AB205041 G.....T.....A..A.....A..G..A..T..T.....CG.T..A...C.C.....A.....A...C
Daucus_DQ192190 .....A..AG....A.....A.....A..T..T.....T.....A.....C...A.....C...C
650 660 670 680 690 700 710 720
Solanum_X86452 TGAAGAACAATACTGATCTCAAGGAGAGAAATAGTAGAATACCAACTTTTCTTTATGCAATGCCATTTTCATCCAACAGG
Capsicum_916392 .....CGT...G.....C.....C.....
Gentiana_EF062505 .T..C..T..C..A..A..A.....A.....C..A.AGG....C..C.....C.....TAAAGG...A
Citrus_AY670167 ....C.....T.G..G.....A...GC.....C.A...T..T.....C.....C.....G..A.....
Chrysanthemum_AB205041 .TGGA..A.....C..AA.A.....A..G...TCA.AG..T..T.....C..G.....C.....G..T..T...
Daucus_DQ192190 .T..TGG.....A..AT.A..A..A.....C....A...C..T..G.....T.....T.....TG....A
730 740 750 760 770 780 790 800
Solanum_X86452 ATATTTCTTGAAGAAACATCACTCGTAGCTCGTCCTGGCTTGCGTATAGATGATATTCAAGAACGAATGGTGGCTCGTTT
Capsicum_916392 .....C.....T..T.....T...G....G.....
Gentiana_EF062505 ..T.....C.....G..C..T.....G..G...GCA.....A.....C.....GA.G.....T.....A..
Citrus_AY670167 .....G..T..G..A..G..G..G.....AG...CA..GA.A.....C..G...A.....A.A..
Chrysanthemum_AB205041 .....T.....T..G.....A..G..A..AAAA..G..A.....A.....A.....TG...G..
Daucus_DQ192190 .....C..T.....A..T..AGCC..G.GA.....G..GA.....CC.

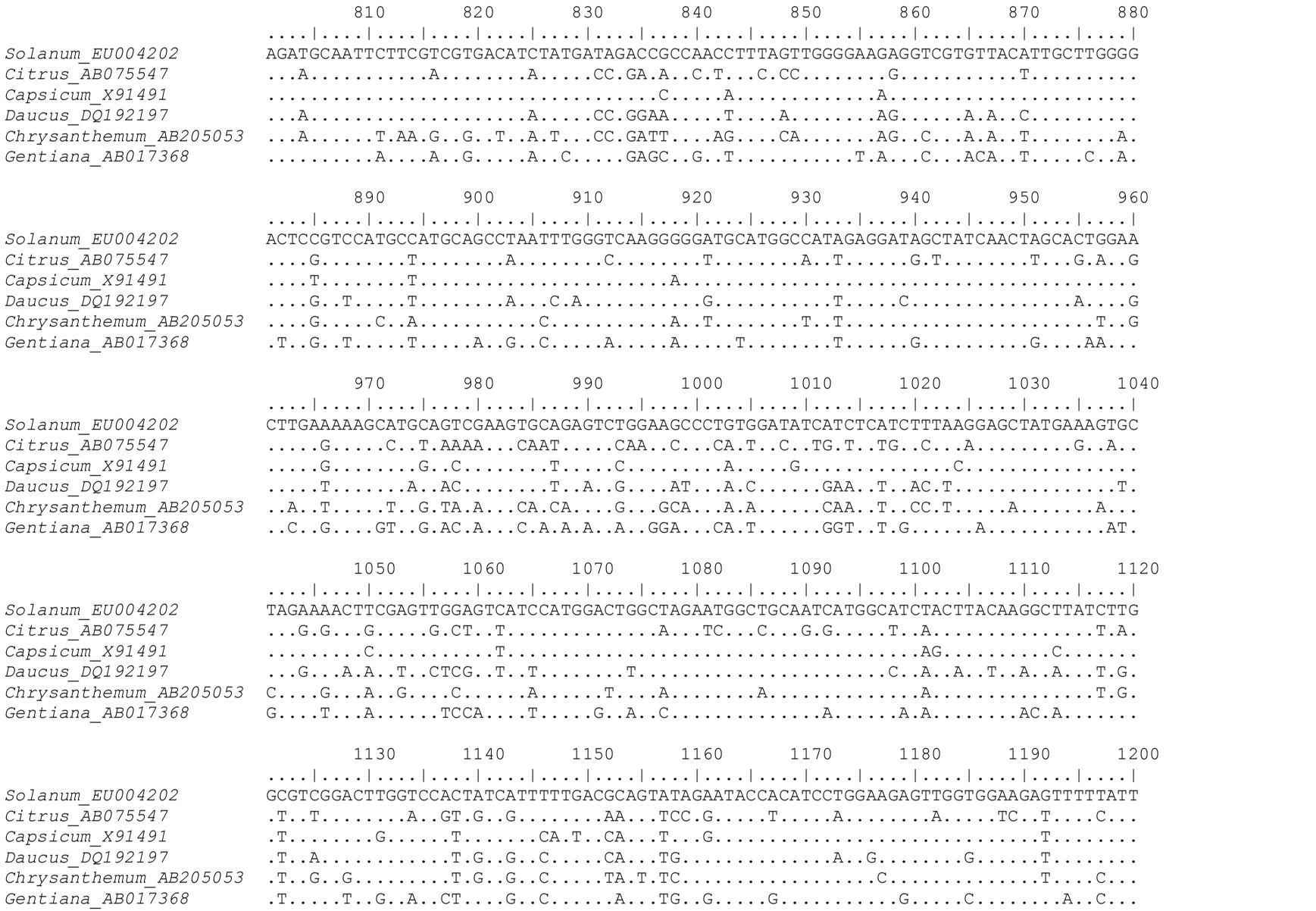

810 820 830 840 850 860 870 880
Solanum_X86452 AAACCATTTGGGGATAAAAGTGAAGAGCATTGAAGAAGATGAACATTGTCTAATACCAATGGGTGGTCCACTTCCAGTAT
Capsicum_916392 ..GT..C.....T.....G.....G.....T.....
Gentiana_EF062505 ..GG..CC.C..T.....ATC.....G.....G.G....G.G..T..T.....G.....T.....T..TA
Citrus_AY670167 ...G..C..A..C.....T.GAG.....G.....G.....G.C..T..G.....G..C.....GC
Chrysanthemum_AB205041 ...G.....T.....A..T.....C..G.GG...G.T..C.....C..G..C.....T..T.
Daucus_DQ192190 ..GG.....T..T.....G.....GAGA...G.G..C.....A..G..T..G..T...C
890 900 910 920 930 940 950 960
Solanum_X86452 TACCTCAGAGAGTCGTTGGAATCGGTGGTACAGCTGGCATGGTTCATCCATCCACCGGTTATATGGTGGCAAGGACACTA
Capsicum_916392 .....T.....T.....C.....C..T.....A.....
Gentiana_EF062505 .....A.....T.....T.....C..A..A.....A..T.....A.....A..T..T
Citrus_AY670167 .T.....A.....T.....A.....C.....G.....G..C..T..A..T..C.....TT..
Chrysanthemum_AB205041 .G.....A.....TC.....T.....G..C..C..G.....T..G.....A.....A..TT..
Daucus_DQ192190 .C.....A..G.....T..C.....T.....G.....T.....T..A.....A.....A..T...
970 980 990 1000 1010 1020 1030 1040
Solanum_X86452 GCTGCGGCTCCTGTTGTTGCCAATGCCATAATTCAATACCTCGGTTCATCCACAGCTGTTTGGAAAGATTTGTGGCCTAT
Capsicum_916392 .....A.....C.....G.....A.....G.....G.....G..
Gentiana_EF062505 ..A..A..A..CA.....T.....G...G.G.G...T..T.....TGGT.AA.....A..
Citrus_AY670167 .....A.....A.....T..CG...G.AG....AA...G..TG.T.AA.....C..
Chrysanthemum_AB205041 .....T..C..AA.....A..AT.A.....G..T..TAA.....AG...AA.....GG.....
Daucus_DQ192190 .....T..A..A.....A...G....G.....T.....G..TG...AA..C.....C.A.....A..
1050 1060 1070 1080 1090 1100 1110 1120
Solanum_X86452 AGAGAGGAGACGTCAAAGAGAGTTCTTCTGCTTCGGTATGGATATTCTTCTGAAGCTTGATTTACCTGCTACAAGAAGGT
Capsicum_916392 .....G.....C.....C.....G.....G....
Gentiana_EF062505 T..A.....C.....G..A.....T..T..C.....AT.G..T.....GA....TT.....A.
Citrus_AY670167 ...A.....A.G.....G.....T..T.....C..G..C..A.....C.....C..T.....
Chrysanthemum_AB205041 ....C.....A.A...C....A.....T..T.....CT.G..A.G...C.....GAA.GC.....C...
Daucus_DQ192190 .....A.....C..GC....A.....T.....T.....T.G..A..A..C.....G....G.....
1130 1140 1150 1160 1170 1180 1190 1200
Solanum_X86452 TCTTTGATGCATTCTTTGACTTAGAACCTCGTTATTGGCATGGCTTCTTATCGTCTCGATTGTTTCTACCTGAACTCATA
Capsicum_916392 .....C.....G..A..CA.G.....
Gentiana_EF062505 .T.....T..C..T.....G.....A.....T..T..G..A.....GC.....T..A..G..GGCT
Citrus_AY670167 .T..C.....T..T.....TC.G..G..C.....T.....A..GA.....C..C..G..TT..
Chrysanthemum_AB205041 .T.....T.....T.....C.....G.....G..A..AA.....GT.GT.G
Daucus_DQ192190 .T..CAG...T..T.....C.G.....C.....A.....G..T.....C.A.....T.....AT.C

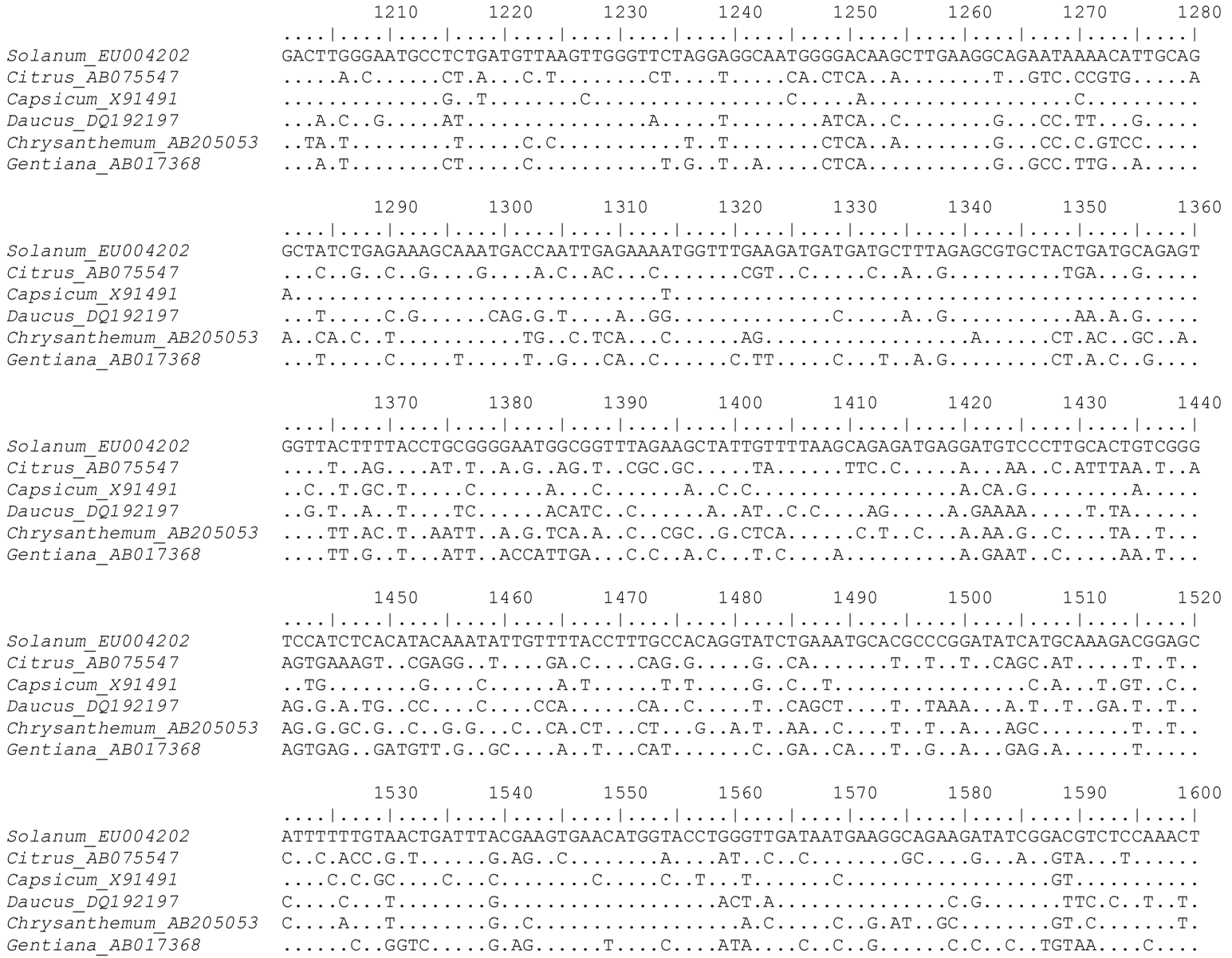

1210 1220 1230 1240 1250 1260 1270 1280
Solanum_X86452 GTTTTTGGGCTGTCTCTATTCTCTCATGCTTCAAATACTTCTAGATTTGAGATAATGACAAAGGGAACTGTTCCATTAGT
Capsicum_916392 ...C..A..T...A...C.
Gentiana_EF062505 T.A.....T.....T.....A.C..C..C..A..GC....A.....T.....AC....TC.TA.
Citrus_AY670167 ...T.....A.....C..T.....GC.A.....C...G.....C....T..G..
Chrysanthemum_AB205041 AC.....T..T.....T..TG.....C.T.....A.A.....T...G....A.....C.C..G..G.C
Daucus_DQ192190 T....C.....T.....T.....AA....C..C..C.....A.A.....C...G....A..C..G.....T..G..
1290 1300 1310
Solanum_X86452 AAATATGATCAACAATTTGTTACAGGATAAAGAA
Capsicum_916392 .C.....
Gentiana_EF062505 .....A.....C..A..G.G...C.....
Citrus_AY670167 T..C.....C...G....A....C...T
Chrysanthemum_AB205041 ...C.....C..AG.T..A...CG...T
Daucus_DQ192190 ...C...G.A.....C..A.CA.A....G....

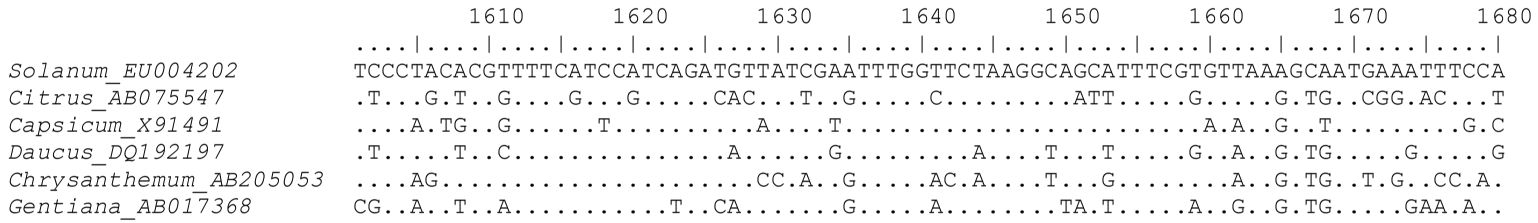
```
                                10        20        30        40        50        60        70        80
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_EU004202       AAGAAGTTGAAAGTACTTGTTGCGGGTGGTGGGATTGGTGGGTTAGTGTTTGCTTTGGCAGCAAAGAAAAGGGGGTTTGA
Citrus_AB075547        .....A....GGA.TT.G..G..T.............A..AC.T..A..........T..T..A.GG.AA..A.....
Capsicum_X91491        ...................G..A.............A..A.....T............G......G...........
Daucus_DQ192197        ...........GA.T..G..G..T..A..G........G..C..G..T..C..C.....G..G.G..GG.A.........
Chrysanthemum_AB205053 ......A.A.GG...T.G..G..A........A........A...........A..A..T..T..A.G..AA..A.....
Gentiana_AB017368      ..A..AA...GGA.TT.A..A..T.............G..AC.T..T........A..G..........A.........

                                90       100       110       120       130       140       150       160
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_EU004202       TGTATTGGTGTTTGAGAGGGATTTAAGTGCTATCAGAGGAGAGGGACAATATAGAGGTCCAATACAGATACAGAGCAATG
Citrus_AB075547        A..GC............AA...A.G........A.....T...........C...........T........A.......
Capsicum_X91491        ...G..A...........A...A..........A.....T.....G.................T.................
Daucus_DQ192197        A..GG.....................G.C......T.....T.....G..G.....G..A..T..T........A..T....
Chrysanthemum_AB205053 A..TG..........A.AA.....G........T.....T..A..T...........A..T..T........A..T....
Gentiana_AB017368      G..G.............A......G.....G..T.....G...........C..G.....T..T...........T....

                               170       180       190       200       210       220       230       240
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_EU004202       CATTGGCTGCTTTAGAAGCAATTGATTTGGATGTTGCTGAAGACATTATGAATGCTGGCTGCATCACTGGTCAAAGGATT
Citrus_AB075547        .............G.....T.....................G..AG......GA.....A..TG.T..C...G.T..A...
Capsicum_X91491        .............G............A................G..C..............T..................
Daucus_DQ192197        .T...........G.....T........G............T..GG.......A.....T..T..T......G.C..A...
Chrysanthemum_AB205053 .T...........G.....T........A...........T..AG.......A.....T..T..T........G......
Gentiana_AB017368      .T.....G.....G.....T......A.......A......A.G......GG......T..T..T..A...G.TC....C

                               250       260       270       280       290       300       310       320
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_EU004202       AATGGCTTGGTTGATGGTATTTCTGGCAACTGGTATTGCAAGTTTGATACGTTCACTCCAGCAGTGGAACGTGGACTTCC
Citrus_AB075547        .....T...........G........TTCT.....CAT............T........T....CA...AAG.....G..
Capsicum_X91491        ...........C.....................................................T..................
Daucus_DQ192197        .....A............G..........T.....C....................A..T....CA.....G..C.....
Chrysanthemum_AB205053 ..................C........T..T......AT............A..............C...A.A..T.....
Gentiana_AB017368      .....A........C..A........A..T.....CAT............C........T....C......G..G.....

                               330       340       350       360       370       380       390       400
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Solanum_EU004202       AGTGACAAGAGTTATCAGTCGCATGACTTTGCAGCAGATCCTTGCACGTGCTGTTGGTGAGGAGATAATTATGAATGAAA
Citrus_AB075547        ...C.........A..T..C..A............A..A..A..A..TAAA..G.....G..T..A..C...T.........
Capsicum_X91491        T...........C.....C..............A.....T.........CTGCAG..G.....TG................
Daucus_DQ192197        ...C.........G.A............GC....AA.A..T..G..CACC........G..T..A..T..ATGT....GT.
Chrysanthemum_AB205053 G..T..T........A..C..A........A..AA.A..T..A..TGA......A..C..C..A..T...T......GT.
Gentiana_AB017368      .......C....AG.A..C..T.....C..A..A..A........TTC......C..A.....T..T..C.AA.......
```

Appendix 5. ClustalX mRNA sequence alignment for *ZEP*. Genus names and accession numbers are at left.


410 420 430 440 450 460 470 480
Solanum_EU004202 GTAATGTTGTAGACTTTGAGGATGATGGAGAGAAGGTTACTGTGGTTCTTGAGAATGGACAACGATTTACAGGTGATCTT
Citrus_AB075547 ........A.T..T...A.....C.......T.....CT.C..A..G..............GT.T.A.G.T.........
Capsicum_X91491 .CC....A...A.T....C.........G....C..........AA..C....TTAT.C....A..AC............
Daucus_DQ192197 .......A..G.....................A.............TA....G...G.......G..C.G.GA...G......
Chrysanthemum_AB205053 .C.....A..C..T.....A...C....T..C......T.C...........A..C...G....T...GA...A......
Gentiana_AB017368 ..........T.....CA..........TC.C........C...AC...C...........G.AT.ACGA..........
490 500 510 520 530 540 550 560
Solanum_EU004202 CTGGTTGGTGCTGATGGCATAAGATCTAAGGTACGGACTAATTTATTCGGACCCAGTGAAGCTACTTACTCTGGCTACAC
Citrus_AB075547 ..CA....A........A...T.G..C.....C....AG........T..G..ACAG.....C.TC..T...........
Capsicum_X91491 ........................G.....................G........G......CTA...........T.....
Daucus_DQ192197 T.......C........A...T.G..A..........GA.....G..T...TA..CA..GC....A..T..A..T.....
Chrysanthemum_AB205053 ..A..............T...T.G.............AG.....G..T..G..A.AG..T.T..................
Gentiana_AB017368 ..T..G..A........A...T.G..A.....C..CG....CA.G..T..C.A............G..T..A.A...T..
570 580 590 600 610 620 630 640
Solanum_EU004202 TTGTTATACTGGAATTGCAGATTTCGTTCCAGCTGATATTGATACAGTTGGGTACCGCGTCTTTTTGGGCCACAAACAGT
Citrus_AB075547 ......C.....T...........T.....G...........G.GT........T..A..G...C.A..A..........
Capsicum_X91491 ............................C..T...........C....C...C.....A.....................
Daucus_DQ192197 .........A..G...........T..G..T..G...........G...........A..A...C.............A.
Chrysanthemum_AB205053 ......C.....G.....T.....TA.A...C........A.CT.T.....A..T..G..G..C..A.....T.....A.
Gentiana_AB017368 .........A..G.....G.....T.....G...........GTT............A.....................
650 660 670 680 690 700 710 720
Solanum_EU004202 ACTTTGTTTCTTCAGATGTGGGGGGAGGCAAGATGCAGTGGTATGCATTTTACAATGAACCAGCTGGTGGTGCGGATGCC
Citrus_AB075547 .............T..........CT..A..A...........C......C....G........A.....A.T.....A.
Capsicum_X91491 ......................T.........................C.....................T......T
Daucus_DQ192197 ....C..................T..G..A.....................C..T........T........CAAA...AAA
Chrysanthemum_AB205053 .T.....A.....T.....T..T..T..A........A............C.....................T.T...AAG
Gentiana_AB017368 ..........G........A..CAA...G........A............C....G......C.G..A.....A..CAGT
730 740 750 760 770 780 790 800
Solanum_EU004202 AACGGTAAAAAGGAAAGATTGCTTAAAATATTTGGGGGATGGTGTGACAATGTTATAGACCTATTAGTTGCCACAGATGA
Citrus_AB075547 G.A..............G.....G..........A...T........T.....GG....TT.GA..C....T..C.....
Capsicum_X91491 ....C.............T.C...C...........
Daucus_DQ192197 ..T..............GC.C...C..........T..T.....C..T.....A.....TT.G...A.G..T..C..C..
Chrysanthemum_AB205053 ..T............C..C.T...G....C.....A..T........T..C..AG.G..T..G..GC.......T.....
Gentiana_AB017368 .....A..G..A.....GC....G..GC......AT..G........T.....G........G...C.A.....T.....


810 820 830 840 850 860 870 880
Solanum_EU004202 AGATGCAATTCTTCGTCGTGACATCTATGATAGACCGCCAACCTTTAGTTGGGGAAGAGGTCGTGTTACATTGCTTGGGG
Citrus_AB075547 ...A.....A.....A.....CC.GA.A..C.T...C.CC.....G.....T.....
Capsicum_X91491 .....C.....A.....A.....
Daucus_DQ192197 ...A.....A.....CC.GGAA.....T....A.....AG.....A.A..C.....
Chrysanthemum_AB205053 ...A.....T.AA.G..G..T..A.T...CC.GATT....AG....CA.....AG..C...A.A..T.....A.
Gentiana_AB017368 .....A....A..G.....A..C.....GAGC..G..T.....T.A...C...ACA..T.....C..A.
890 900 910 920 930 940 950 960
Solanum_EU004202 ACTCCGTCCATGCCATGCAGCCTAATTTGGGTCAAGGGGGATGCATGGCCATAGAGGATAGCTATCAACTAGCACTGGAA
Citrus_AB075547 ....G.....T.....A.....C.....T.....A..T.....G.T.....T...G.A..G
Capsicum_X91491 ....T.....T.....A.....
Daucus_DQ192197 ....G..T.....T.....A...C.A.....G.....T.....C.....A....G
Chrysanthemum_AB205053 ....G.....C..A.....C.....A..T.....T..T.....T..G
Gentiana_AB017368 .T..G..T.....T.....A..G..C.....A.....A.....T.....T.....G.....G....AA...
970 980 990 1000 1010 1020 1030 1040
Solanum_EU004202 CTTGAAAAAGCATGCAGTCGAAGTGCAGAGTCTGGAAGCCCTGTGGATATCATCTCATCTTTAAGGAGCTATGAAAGTGC
Citrus_AB075547 .....G.....C..T.AAAA...CAAT.....CAA..C...CA.T..C..TG.T..TG..C...A.....G..A..
Capsicum_X91491 .....G.....G..C.....T.....C.....A.....G.....C.....
Daucus_DQ192197 .....T.....A..AC.....T..A..G....AT...A.C.....GAA..T..AC.T.....T.
Chrysanthemum_AB205053 ..A..T.....T..G.TA.A...CA.CA....G...GCA...A.A.....CAA..T..CC.T.....A.....A...
Gentiana_AB017368 ..C..G....GT..G.AC.A...C.A.A.A..A..GGA...CA.T.....GGT..T.G.....A.....AT.
1050 1060 1070 1080 1090 1100 1110 1120
Solanum_EU004202 TAGAAAACTTCGAGTTGGAGTCATCCATGGACTGGCTAGAATGGCTGCAATCATGGCATCTACTTACAAGGCTTATCTTG
Citrus_AB075547 ...G.G...G.....G.CT..T.....A...TC...C...G.G.....T..A.....T.A.
Capsicum_X91491 .....C.....T.....AG.....C.....
Daucus_DQ192197 ...G...A.A..T..CTCG..T..T.....T.....C..A..A..T..A..A...T.G.
Chrysanthemum_AB205053 C....G...A..G....C.....A.....T....A.....A.....A.....T.G.
Gentiana_AB017368 G....T...A.....TCCA....T.....G..A..C.....A.....A.A.....AC.A.....
1130 1140 1150 1160 1170 1180 1190 1200
Solanum_EU004202 GCGTCGGACTTGGTCCACTATCATTTTTGACGCAGTATAGAATACCACATCCTGGAAGAGTTGGTGGAAGAGTTTTTATT
Citrus_AB075547 .T..T.....A..GT.G..G.....AA...TCC.G.....T.....A.....A.....TC..T....C...
Capsicum_X91491 .T.....G.....T.....CA.T..CA...T...G.....T.....
Daucus_DQ192197 .T..A.....T.G..G..C.....CA...TG.....A..G.....G.....T.....
Chrysanthemum_AB205053 .T..G..G.....T.G..G..C.....TA.T.TC.....C.....T....C...
Gentiana_AB017368 .T.....T..G..A..CT....G..C.....A...TG..G.....G.....G.....C.....A..C...


1210 1220 1230 1240 1250 1260 1270 1280
Solanum_EU004202 GACTTGGGAATGCCTCTGATGTTAAGTTGGGTTCTAGGAGGCAATGGGGACAAGCTTGAAGGCAGAATAAAACATTGCAG
Citrus_AB075547 .....A.C.....CT.A...C.T.....CT...T.....CA.CTCA..A.....T..GTC.CCGTG.....A
Capsicum_X91491 .....G..T.....C.....C.....A.....C.....
Daucus_DQ192197 ...A.C..G.....AT.....A.....T.....ATCA..C.....G...CC.TT...G.....
Chrysanthemum_AB205053 ..TA.T.....T.....C.C.....T..T.....CTCA..A.....G...CC.C.GTCC.....
Gentiana_AB017368 ...A.T.....CT.....C.....T.G..T..A.....CTCA.....G..GCC.TTG..A.....
1290 1300 1310 1320 1330 1340 1350 1360
Solanum_EU004202 GCTATCTGAGAAAGCAAATGACCAATTGAGAAAATGGTTTGAAGATGATGATGCTTTAGAGCGTGCTACTGATGCAGAGT
Citrus_AB075547 ...C..G..C..G....G....A.C..AC...C.....CGT..C.....C..A..G.....TGA...G.....
Capsicum_X91491 A.....T.....
Daucus_DQ192197 ...T.....C.G.....CAG.G.T....A..GG.....C.....A..G.....AA.A.G.....
Chrysanthemum_AB205053 A..CA.C..T.....TG..C.TCA...C.....AG.....A.....CT.AC..GC..A.
Gentiana_AB017368 ...T.....C.....T.....T..G...CA..C.....C.TT.....C...T..A.G.....CT.A.C..G....
1370 1380 1390 1400 1410 1420 1430 1440
Solanum_EU004202 GGTTACTTTTACCTGCGGGGAATGGCGGTTTAGAAGCTATTGTTTTAAGCAGAGATGAGGATGTCCCTTGCACTGTCGGG
Citrus_AB075547 ....T..AG....AT.T..A.G..AG.T..CGC.GC.....TA.....TTC.C.....A...AA..C.ATTTAA.T..A
Capsicum_X91491 ..C..T.GC.T.....C.....A...C.....A..C.C.....A.CA.G.....A.....
Daucus_DQ192197 ..G.T..A..T....TC.....ACATC..C.....A..AT..C.C....AG.....A.GAAAA.....T.TA.....
Chrysanthemum_AB205053 ....TT.AC.T..AATT..A.G.TCA.A..C..CGC..G.CTCA.....C.T..C...A.AA.G..C....TA..T...
Gentiana_AB017368 ....TT.G..T...ATT..ACCATTGA...C.C..A.C...T.C....A.....A.GAAT..C.....AA.T...
1450 1460 1470 1480 1490 1500 1510 1520
Solanum_EU004202 TCCATCTCACATACAAATATTGTTTTACCTTTGCCACAGGTATCTGAAATGCACGCCCGGATATCATGCAAAGACGGAGC
Citrus_AB075547 AGTGAAAGT..CGAGG..T....GA.C....CAG.G.....G..CA.....T..T..T..CAGC.AT.....T..T..
Capsicum_X91491 ..TG.....G....C.....A.T.....T.T.....G..C..T.....C.A...T.GT..C..
Daucus_DQ192197 AG.G.A.TG..CC....C....CCA.....CA..C.....T..CAGCT....T..TAAA...A.T..T..GA.T..T..
Chrysanthemum_AB205053 AG.G.GC.G..C..G.G...C..CA.CT...CT...G..A.T..AA..C....T..T..A...AGC.....T..T..
Gentiana_AB017368 AGTGAG..GATGTT.G..GC....A..T...CAT.....C..GA..CA...T..G..A...GAG.A.....T.....
1530 1540 1550 1560 1570 1580 1590 1600
Solanum_EU004202 ATTTTTTGTAACTGATTTACGAAGTGAACATGGTACCTGGGTTGATAATGAAGGCAGAAGATATCGGACGTCTCCAAACT
Citrus_AB075547 C..C.ACC.G.T.....G.AG..C.....A....AT..C..C.....GC....G...A..GTA...T.....
Capsicum_X91491 ....C.C.GC....C...C.....C.....C..T...T.....C.....GT.....
Daucus_DQ192197 C....C...T.....G.....ACT.A.....C.G.....TTC.C..T..T.
Chrysanthemum_AB205053 C....A...T.....G..C.....A.C.....C..G.AT..GC.....GT.C.....T.
Gentiana_AB017368 .....C..GGTC.....G.AG.....T....C....ATA....C..C..G.....C.C..C..TGTAA....C....


1610 1620 1630 1640 1650 1660 1670 1680
Solanum_EU004202 TCCCTACACGTTTTCATCCATCAGATGTTATCGAATTTGGTTCTAAGGCAGCATTTCGTGTTAAAGCAATGAAATTTCCA
Citrus_AB075547 .T...G.T..G....G...G.....CAC...T..G.....C.........ATT.....G.....G.TG..CGG.AC...T
Capsicum_X91491 ....A.TG..G......T..........A....T.........................A.A..G..T.........G.C
Daucus_DQ192197 .T.....T..C..............A......G.........A....T...T.....G..A..G.TG.....G.....G
Chrysanthemum_AB205053 ....AG.....................CC.A..G.....AC.A....T...G........A..G.TG..T.G..CC.A.
Gentiana_AB017368 CG..A..T..A...........T..CA.......G.....A........TA.T.....A..G..G.TG.....GAA.A..