Project Number: PZW-1982

Computational Analysis of 1-Deoxy-D-xylulose-5-phosphate reductoisomerase in 15 Plant Species

A Major Qualifying Project Report:

submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

_____

Pui Kwan Fung

Date: November 16, 2005

Approved:

_____

Professor Pamela Weathers

Acknowledge

I would like to thank Jeffrey Blanchard (UMass, Amherst), Olof Emmanuelsson (Yale University), Julia Krushkal (University of Tennessee) and Frederic F. Souret (Delaware Biotechnology Institute) for their assistance with information and bioinformatics tools. Most importantly, I am very grateful to Professor Pamela Weathers for her guidance throughout the project and her assistance with the writing of this report.

Abstract

The enzyme, 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), is a key regulatory step in the non-mevalonate terpenoid biosynthetic pathway in plastids. To investigate the molecular evolution of the enzyme and to predict its location in the chloroplast, a computational analysis was performed on 15 plant DXR sequences that have a full-length cDNA. Results revealed that DXR has an N-terminal transit domain that is likely bipartite, consisting of a chloroplast transit peptide (cTP) and a lumen transit peptide (lTP). Several features were observed in the lTP which suggest that while DXR is targeted to the chloroplast, it is in fact localized to the thylakoid lumen. These features include a twin arginine motif, a hydrophobic region and a proline-rich region. In addition, the functional domain of DXR is found to be highly conserved between prokaryotic and eukaryotic species.

Contents

1 Background

1.1 Biochemistry of terpenoid synthesis

1.1.1 Terpenoids

Terpenoids are one of the most structurally diverse groups of natural plant products where they play significant roles in pollinator attraction, defense and plant-plant communication. In addition to their ecological roles in plants, terpenoids are extensively used in the food and cosmetic industries as flavoring agents, since they are constituents of natural essential oils and floral scents. Researchers have also found pharmaceutical importance in terpenoids, with their potential use as antimicrobial agents and anticarcinogens (Mahmoud and Croteau, 2002).

1.1.2 Biosynthesis of terpenoids

Despite the great diversity in structure and function, all terpenoids are derived from two common precursors, isopentenyl diphosphate (IPP) and its allylic isomer, dimethylallyl diphosphate (DMAPP) (Mahmoud and Croteau, 2002). The repetitive head-to-tail addition of IPP units to DMAPP yields the immediate precursors to most of the terpenoid classes, though some less common terpenes are produced by non-head-to tail joining of the two building blocks or by rearrangement of a regular structure (Mahmoud and Croteau, 2002). In either case, the resultant diphosphate skeletons undergo subsequent enzymatic modifications (mostly redox reactions). This is the final and crucial step in terpenoid synthesis which contributes to the structural and functional diversity of the terpenoid family (Croteau *et al.*, 2000). Figure 1 illustrates the general biosynthesis of the various terpenoid classes from IPP and DMAPP.

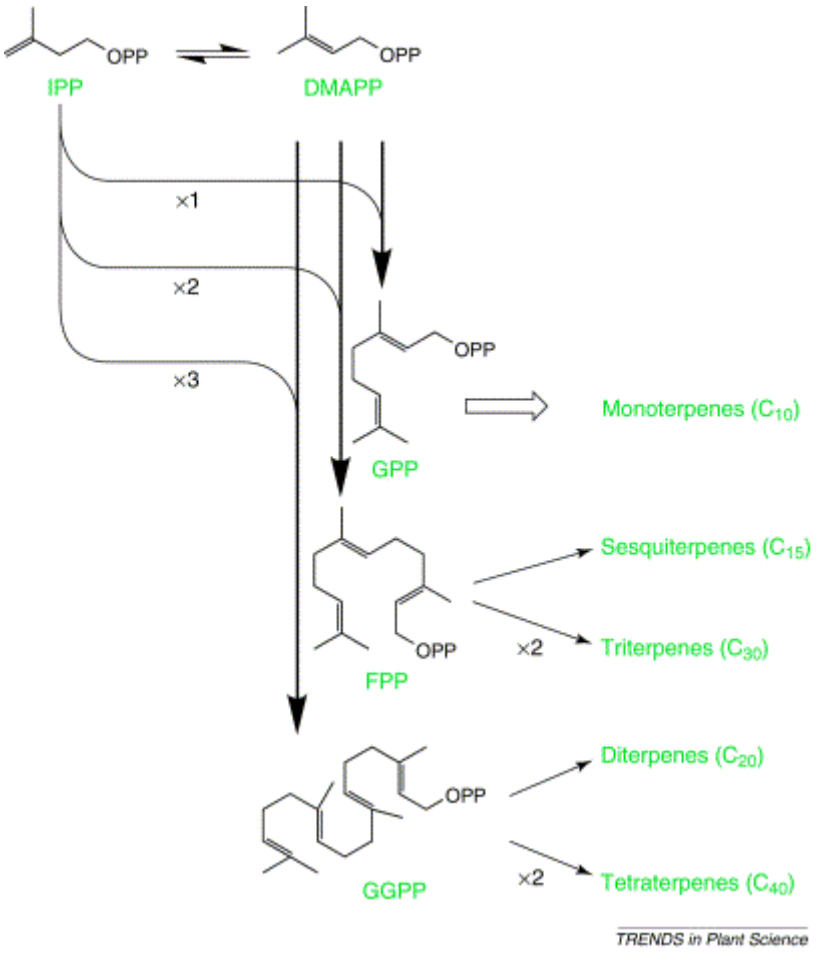


Figure 1. Terpenoid biosynthesis from IPP and DMAPP. (Taken from Mahmoud and Croteau, 2002.)

1.1.3 Biosynthesis of IPP

Although terpenoids can be found in animals and microorganisms, their synthesis is shown to be the most complex in plants, as reflected by their production of a large and diverse range of terpenoid products (Croteau *et al.*, 2000). In plants, terpenoids are produced via two IPP generating pathways: a mevalonate dependent pathway and a mevalonate-independent pathway (Figure 2). When the mevalonate dependent pathway was first discovered in the 1950s, it was commonly accepted as the only route for

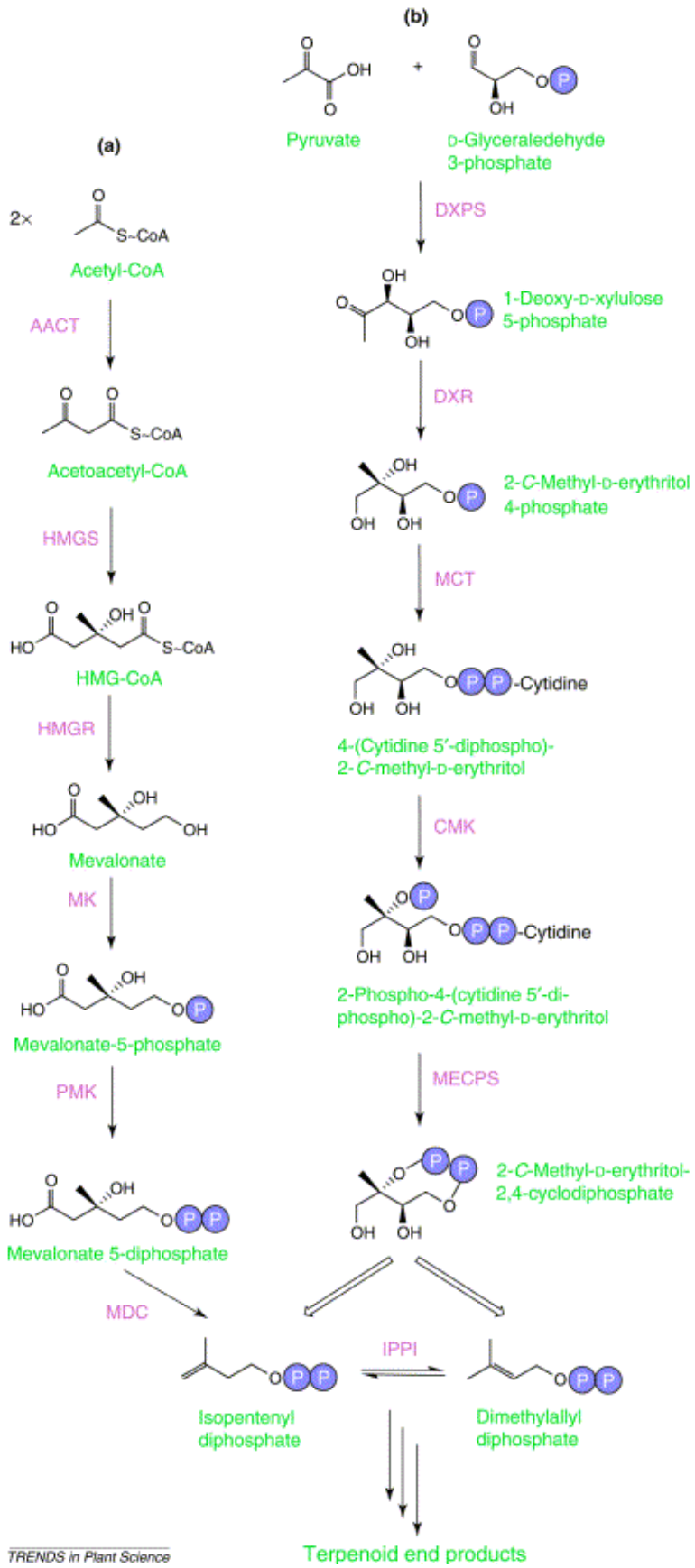


Figure 2. The mevalonate pathway (a) and the mevalonate-independent pathway (b) for the biosynthesis of IPP and DMAPP. (Taken from Mahmoud and Croteau, 2002.)

terpenoids to be produced, occuring via the condensation of acetyl CoA (Takahashi *et al.*, 1998). In the late nineties, however, a novel non-mevalonate pathway, first identified in bacteria, was found in the plastid (Dubey *et al.*, 2003). While it has been indicated that the mevalonate pathway operates mainly in the cytoplasm and mitochondria, and the non-mevalonate pathway operates in the plastid, the two pathways are not completely independent. It is possible for metabolites to be exchanged between the two pathways in their different compartments (Dubey *et al.*, 2003). Figure 3 illustrates the compartmentalization of the two pathways in higher plants, suggesting the exchange of IPP between cytosol and plastid.

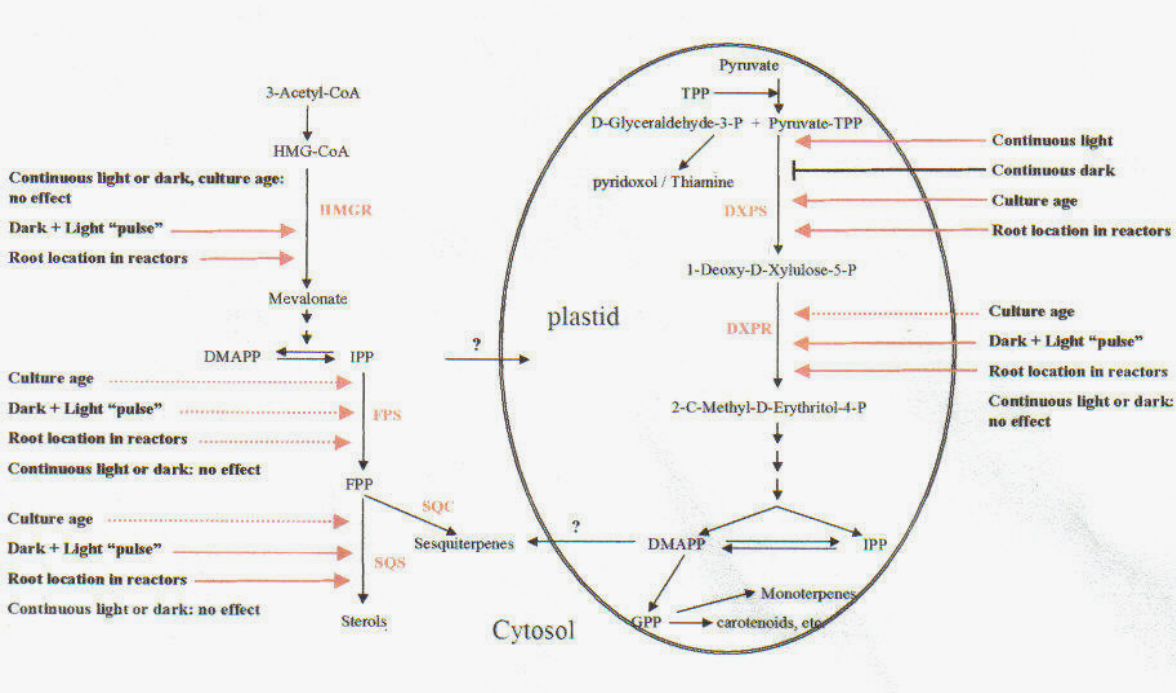


Figure 3. Compartmentation of IPP and isoprenoid biosynthesis in higher plants between cytosol(mevalonate pathway) and plastids (non-mevalonate pathway). (Taken from Souret, 2002)

1.1.4 Mevalonate (MVA) pathway

The classical acetate/mevalonate pathway (Figure 2a) begins with the condensation of three units of acetyl CoA by thiolase and hydroxymethylglutaryl-CoA (HMG-CoA) synthase to form 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA). HMG-CoA is then reduced to form mevalonic acid (MVA) which is transformed to IPP via two steps of phosphorylation and the final step of decarboxylation. HMG-CoA reductase (HMGR), which reduces HMG-CoA to MVA, is a key regulatory enzyme that has been extensively studied. This highly conserved enzyme is known to catalyze the rate-limiting step of IPP biosynthesis in animals, and possibly in the synthesis of cytosolic terpenes in plants (Dubey *et al.*, 2003).

1.1.5 Mevalonate independent (non-MVA) pathway

The mevalonate independent pathway (Figure 2b) is proposed to be the IPP biosynthetic pathway in plastids (Dubey *et al.*, 2003). Although the enzymes in the pathway have all been identified, their characteristics have yet to be fully understood. The first step of the pathway involves a transketolase-type condensation of pyruvate with glyceraldehyde-3-phosphate to form 1-deoxy-D-xylulose-5-phosphate (DXP), catalyzed by DXP synthase (DXS). DXP is then reduced to 2-C-methyl-D-erythritol 4-phosphate (MEP) by DXP reductoisomerase (DXR) and subsequently transformed to yield IPP. Besides being an intermediate for IPP, DXP is also the precursor for the biosynthesis of thiamin (Vitamin B_1) and pyridoxol (Vitamin B_6). Hence the conversion of DXP to MEP, catalyzed by DXR, is actually the first committed step in the non-MVA pathway (Carretero-Paulet *et al.*, 2002). Consequently, both DXS and DXR could both play significant roles in the regulation of DXP formation, which in turn controls the synthesis

of IPP and terpenoids. A study of these two enzymes would provide a better understanding of the regulation of terpenoid production in the non-MVA pathway.

1.2 Previous studies on DXS and DXR

1.2.1 1-deoxy-D-xylulose-5-phosphate synthase (DXS)

The regulatory role of DXS has been confirmed by experimental results, where an over-expression of DXS in transgenic *Arabidopsis* has led to an increase production of IPP (Estévez *et al.*, 2001). With DXS catalyzing one of the rate-limiting steps in the non-MVA pathway, it becomes important to fully understand its structural and functional properties. In addition, predicting the protein's subcellular localization would also provide insights into its biological functions.

A previously performed sequence analysis of DXS investigated the evolutionary changes in its structure and function (Krushkal *et al.*, 2003). Phylogenetic inference of 11 plant sequences revealed that DXS is divided into two distinct classes (DXS1 and DXS2). Analysis of the transit domain suggests that DXS is likely bipartite and targeted to the thylakoid lumen by the delta pH pathway. Although the transit peptide domain was not conserved, a consistent set of common features was identified, such as the same hydrophobic slope, hydrophobic region in residues 35-45, and a highly conserved Pro-Pro-Thr motif at the C-terminal of the domain. The functional region was, on the other hand, well conserved among the tested plant species. Secondary structure prediction using Gov IV and HNN showed five regions of conserved secondary structures in the functional domain. In particular, region III is believed to play an important role in the diversification between DXS1 and DXS2.

1.2.2 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR)

The participation of DXR in the regulation of IPP production is also evident in many of the plants tested. Overexpresion of DXR in transgenic peppermint plants led to an increase in production of monoterpenes in leaf tissue (Mahmoud and Croteau, 2001); a positive correlation was detected between the accumulation of DXR transcript and apocarotenoids in mycorhizal roots from monocots (Walter *et al.*, 2000), and also with terpenoid indole alkaloids in periwinkle cell suspension culture (Veau *et al.,* 2000). Thus far, both DXR and DXS have displayed regulatory roles in terpenoid biosynthesis.

The expression and structure of DXR has been examined by both bioinformatics analysis and experimental methods. In Carretero-Paulet *et al.* (2002), a sequence alignment of 14 plant DXR sequences as well as the *E. coli* DXR was performed. Results revealed an extension of 73 to 80 residues at the N-terminal side which was absent in the prokaryotic *E. coli* sequence. This region was predicted by the ChloroP software (Emanuelson *et al.,* 1999) to contain the transit peptide, with the cleavage site at about 50 residues from the N-terminus, before a conserved Cys-Ser-X motif. The N-terminal end of the transit peptide region was found to be poorly conserved but enriched in Ser residues, while the C-terminal end was more highly conserved and a consensus motif P(P/Q)PAWPG(R/T)A was identified. The function of this Pro-rich motif was demonstrated by complementing an *E. coli* mutant defective in DXR activity with either (1) a short derivative of *Arabidopsis* DXR which lacked the entire N-terminal extended region or (2) a longer version which included the Pro-rich motif. Both forms of *Arabidopsis* protein exhibited DXR activity. In particular, *E. coli* rescued by the longer version led to a more vigorous growth, suggesting that the Pro-rich region is likely

important for protein activity or stability. Subcellular localization of plant DXR was also examined by immunofluroscent assay, which showed DXR to be targeted to plastids and localized in chloroplasts of leaf cells (Carretero-Paulet *et al.,* 2002).

Souret (2002) further identified other conserved motifs and their possible functional implication in DXR (Souret, 2002). He predicted the transit peptide to be the first 50-60 residues at the N-terminal. Using the GORIV analytical tool, he suggested that the transit domain has a random coil secondary structure. Further, the domain was found to be positively charged, lacking acidic amino acids, and generally rich in Ser, Thr and small hydrophobic amino acids. The first interesting motif mentioned was the presence of a putative phosphorylation site at Thr^{19}, within a motif (P/G)XXX(R/K)XX(S/T)XXX(S/T) (residues 8-19) highly similar to the binding site of a 14-3-3- protein involved in plastid transport (Waegemann and Soll, 1996). A study of the functional domain of DXR revealed a potential NADPH binding site at a position between aa 81-87, consistent with the requirement of an NADPH cofactor for DXR enzymatic conversion of DXP to MEP. This motif GSTGSIG was homologous to the NADPH binding site found in ketol acid transketolase (Rane and Calvo, 1997). In addition, four potentially catalytic amino acid residues, Glu^{294} and three histidines ($His^{226, 272, 320}$) (Kuzuyama *et al.,* 2000) were found to be highly conserved among the five plant and *E. coli* DXR sequences studied by Souret (2002).

1.3 Protein localization processes

Both DXS and DXR are nuclear-encoded proteins originally synthesized in a precursor form. The precursor contains a plastid targeting peptide called the chloroplast transit peptide (cTP) at its N-terminus, responsible for directing the protein into the chloroplast. Post-translational events take place once the precursor crosses the outer and inner chloroplast membranes into the stroma, where the cTP will be cleaved leaving behind the functional part of the protein. For proteins that are targeted to the thylakoid or the thylakoid lumen, their precursors will contain another transit peptide called the lumen transit peptide (lTP), making the overall transit peptide bipartite. Cleavage of the cTP reveals the lTP domain which, though one of four distinct pathways, continues the translocation of the intermediate precursor to its final destination into either the thylakoid membrane or its lumen (Fig. 4). The SRP and spontaneous pathways transport proteins into the thylakoid membrane, while the Sec and ΔpH pathways transport proteins through the thylakoid membrane and into the thylakoid lumen. Once localized, the lTP will usually be cleaved to expose the functional protein (Mori and Cline, 2001).

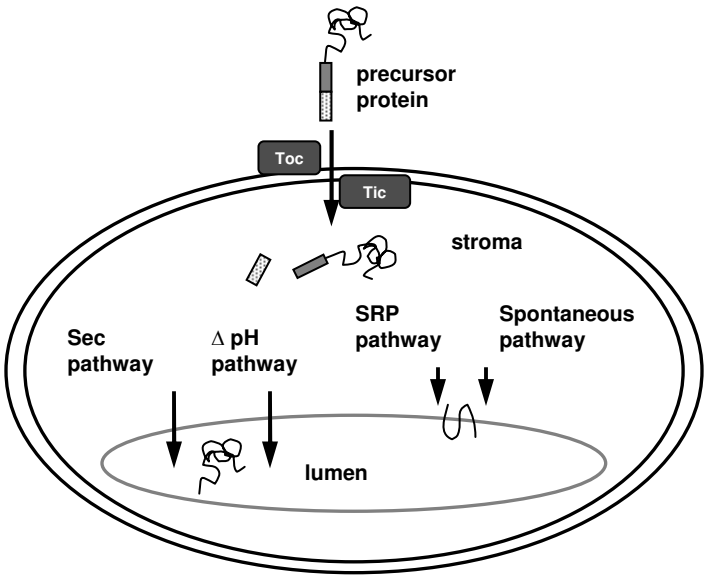


Figure 4. Localization pathways of nuclear-encoded proteins. (Adapted from Mori and Cline, 2001)

Prediction of subcellular localization can help provide important functional information about a protein, and as such has received much attention in the bioinformatics field. Most bioinformatics tools base their prediction on the presence of conserved motifs and characteristics already identified in known cTP and lTP. Thus far, the (-3, -1) rule (Emanuelsson *et al.,* 2001) is commonly used to detect the cTP cleavage site. According to this rule, the residue in position -1 must be small and neutral (e.g. Ala, Gly, Ser, Cys etc) and the residue in position -3 must be hydrophobic, polar, small and neutral (e.g. Ile, Leu, Val, Ala, Cys). The cTP is also recognized as a region enriched in hydroxylated residues containing a relatively low content of acidic residues (Emanuelsson *et al.,* 2001).

In the lTP, both the Sec and ΔpH pathways are characterized by four distinct domains: 1) an acidic N-terminal domain (A); 2) a charged N-terminal domain (N) with a twin arginine motif common in the ΔpH pathway; 3) a hydrophobic core domain (H) of 8-12 residues and 4) a polar C-terminal domain (C) which contains basic residues for the ΔpH but not the Sec pathway. The function of the acidic A-domain is unknown since its deletion does not impair transport *in vitro* (Mori and Cline, 2001). In contrast to the ΔpH pathway which transports folded proteins, the Sec pathway transports unfolded proteins.

Another distinct feature of the transit domain is the previously mentioned motif (P/G)XXX(R/K)XX(S/T)XXX(S/T), conserved in most of the chloroplast-targeting precursors (Waegemann and Soll, 1996). The motif is related to the binding of 14-3-3 proteins and contains a phosphorylation site on either a Ser or Thr residue in the motif. The 14-3-3 proteins belong to the chaperone family, molecules that are likely to be the

sites of ATP hydrolysis required for precursors to bind to and be transported across the chloroplast membranes (Jackson-Constan *et al*., 2001). Waegemann and Soll (1996) proposed a model for the role of the 14-3-3 protein during chloroplast import as follow: (1) before entry to the chloroplast, the precursor is first phosphorylated by a kinase (either on a Ser or Thr); (2) the phosphorylated precursor forms a complex with the 14-3-3 protein and possibly other chaperones; (3) the complex binds to the complementary receptor on the chloroplast outer membrane; and (4) dephosphorylation of the complex by a phosphatase is required for the precursor to be translocated across the chloroplast membranes (Waegemann and Soll, 1996). In thylakoid-targeting precursors which possess bipartite transit sequences, the consensus motif for the 14-3-3 protein and the phosphorylation site are believe to reside within the chloroplast transit domain (Waegemann and Soll, 1996). A later study, however, revealed the presence of 14-3-3 proteins in the chloroplast stroma despite their lack of an import signal, and suggested that the 14-3-3 chaperone may instead be binding to the lumen transit domain (Sehnke *et al.*, 2000).

1.4 Objectives

In this experiment, I will first predict the sub-localization of DXR in chloroplast and the transport pathway used. Subsequently, the structural and functional evolution of DXR among plant species will be studied. Phylogenetic tree construction will provide some insights to the evolutionary path of DXR, while alignment of the various plant DXR sequences will help identify regions of the sequence that had either remained highly conserved or undergone mutations. Comparison of these results with the secondary structure prediction and hydrophobicity analysis of the protein will reveal the amino acids residues that are important in the proper functioning of DXR. In particular, these residues maybe crucial for DXR to function as a regulatory enzyme in the terpenoid synthesis pathway.

2 Materials and Methods

2.1 Sequence Selection

DXR sequences from different species were collected from GenBank via a BLAST (blastn) search, using *Artemisia annua* as the query sequence. The cutoff point of selection is set to a score of >= 50. In addition, only species with complete DXR cDNA were included. A total of fifteen species satisfied these criteria. To compare the evolution of DXR between plant and bacteria, a DXR amino acid sequence from *E. coli* is also included for analysis. Table I lists details of the included species.

Table I. Selected DXR sequences for analysis.

Organism Name	Common Name	GenBank Accession #
Dicots		
Artemisia annua	Sweet wormwood	AF182287
Stevia rebaudiana	Stevia	AJ429233
Catharanthus roseus	Madagascar periwinkle	AF250235
Lycopersicon esculentum	Tomato	AF331705
Antirrhinum majus	Snapdragon	AY770406
Arabidopsis thaliana	Thale cress	AF148852
Populus alba x Populus tremula	Gray popular	AJ574852
Linum usitatissimum	Flax	AJ623266
Pueraria montana var. lobata	Kudzu	AY315651
Mentha x piperita	Peppermint	AF116825
Monocots		
Oryza sativa	Japanese rice	AF367205
Zea mays	Maize	AJ297566
Hordeum vulgare	Barley	AJ583446
Gymnosperm		
Ginkgo biloba	Maindenhair tree	AY494186
Taxus cuspidata	Japanese yew	AY575140
Bacteria		
Escherichia. coli		NP_414715

2.2 Sequence Alignment

Sequences were aligned using ClustalX version 1.81. (Thompson *et al.*, 1997). Nucleotide sequence alignment was performed on the fifteen plant species. Full amino acid sequence alignment was performed on the plant species as well as *E. coli*. The alignment data can be found in Appendix I.

2.3 Prediction of transit peptide sequences

Predotar (http://genoplante-info.infobiogen.fr/predotar/predotar.html) was used to predict the presence of a plastid transit sequence. ChloroP (http://www.cbs.dtu.dk/services/ChloroP/) and Target P (http://www.cbs.dtu.dk/services/TargetP/) were used to determine if the protein is chloroplast-targeted and to predict the potential cTP cleavage site (Emanuelsson and von Heijne, 2001). Lastly, LumenP (v.1.3, courtesy of Olof Emanuelsson) was also used to determine if the DXR protein is targeted to the thylakoid lumen and to predict the lTP cleavage site (Westerlund *et al.*, 2003).

In addition, manual prediction of the transit peptide was performed by matching the cTP cleavage site motif and the characteristics of the four domains in lTP to the amino acid sequences.

2.4 Phylogenetic Tree Inference

To investigate the molecular evolution of DXR, phylogenetic trees were inferred using MEGA v. 2.1 (http://www.megasoftware.net/) (Kumar *et al.*, 2001).

2.5 Secondary Structure Analysis

Secondary structure of the amino acid sequences were predicted using GORIV (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html) (Garnier *et al.*, 1996) and Hierarchical Neural Network (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_nn.html) (Guermeur, 1997), provided by the Network Protein Sequence Analysis (NPS@) web server at the Institute for the Biology and Chemistry of Proteins (IBCP).

2.6 Hydrophobicity analysis

To determine which amino acids may be critical for the chemical and physical properties of the DXR protein, we compared the hydrophobity plots of the 15 selected species. Hydrophobicity analysis was performed using ProtScale (Gasteiger *et al.,* 2005), available on the ExPASy WWW server (http://www.expasy.org/tools/protscale.html). This program uses the hydrophobicity plotting tool of Kyte and Doolittle (1982). Hydrophobicity analyses were done and compared separately for the DXR transit peptide and the functional domain. To study similarities between species, each hydrophobicity plot was also analyzed using linear regression.

3 Results and discussion

3.1 Prediction of transit peptide sequence

3.1.1 Prediction using ChloroP, TargetP and LumenP

The result of transit peptide sequence prediction using Predotar, ChloroP and TargetP are summarized in Table II.

Table II. Prediction of chloroplast transit peptide (cTP) in DXR proteins.

		ChloroP		TargetP	
	Predotar[a]	Prediction[b]	Length[c]	Prediction[b]	Length[c]
A. annua	none	C	46	C	46
S. rebaudiana	plastid	-	2	O	-
C. roseus	possibly plastid	C	83	O	-
L. esculentum	Plastid	C	67	C	67
A. majus	plastid	C	42	C	42
A. thaliana	plastid	C	86	C	86
P. tremula x alba	possibly plastid	-	45	C	45
L. usitatissimum	plastid	C	50	C	50
P. montana	plastid	C	44	C	44
M. piperita	plastid	C	51	O	-
O. sativa	possible plastid	C	49	C	49
Z. mays	none	C	48	C	48
H. vulgare	possibly plastid	C	59	C	59
T. cuspidata	plastid	C	55	C	55

[a] None indicates no targeted sequence is present

[b] C, cholorplast localization; O, other localization besides chloroplast, mitochondria and secretory pathway; - indicates no prediction of a cTP

[c] Predicted length of the transit peptide; - indicates no length predicted

Most of the DXR sequences were predicted by Predotar to be plastid localized. The putative cleavage site given by ChloroP and TargetP also agreed in all cases. Only

two distinct disagreements were found in *Artemisia* and *Zea*: ChloroP and TargetP predicted the two species to be chloroplast targeted, while Predotar failed to detect any targeting peptide present (Table II). Problems with predicting putative cleavage site using neural network-based methods have been reviewed by Emanuelsson and von Henine (2001). It is also important to verify the prediction results experimentally.

Table III presents the result from LumenP (courtesy of Olof Emmanuelsson) in the prediction of an lTP. A sequence is predicted to contain an lTP if it scores above the cutoffs in both "score" and "CSscore". The "score" measures how ITP-like the N-terminal part of the protein is, while the "CSscore" is the cleavage site motif score. The lTP length indicates the predicted length of the lTP.

Table III. Prediction of lumen transit peptide (lTP) in DXR proteins.

	LumenP			
	Score (cutoff: >=0.47)	CS score (cutoff: >=6.8)	lTP length	Prediction
A. annua	0.626	7.562	59	Y
S. rebaudiana	0.682	8.267	108	Y
C. roseus	0.575	9.469	67	Y
L. esculentum	0.649	9.469	68	Y
A. majus	0.699	12.129	59	Y
A. thaliana	0.426	8.267	113	N
P. tremula x alba	0.728	7.512	108	Y
L. usitatissimum	0.39	7.512	112	N
P. montana	0.658	7.138	57	Y
M. piperita	0.409	8.203	112	N
O. sativa	0.365	8.267	109	N
Z. mays	0.372	9.469	65	N
H. vulgare	0.341	8.267	119	N
G. biloba	0.177	8.267	114	N
T. cuspidata	0.681	7.562	66	Y

All the proteins in this study have relatively high "CS score", and most of them have "score" either close to or above the cutoff. The results gave a strong implication that DXR is likely lumen-targeted. Once again, this has to be verified experimentally.

3.1.2 Prediction using manual analysis

Several known cTP and lTP motifs were successfully located in the DXR transit domain (Fig. 5). Through manual analysis, the cTP is predicted to be approximately 40 amino acids long, with small and neutral residues at the -1 and -3 positions relative to the cleavage site (Fig. 5, ▼) (Emanuelsson *et al.*, 1991). Abundant hydroxylated residues (e.g. S) are observed in the proposed cTP domain (Emanuelsson *et al.*, 1991). The putative phosphorylation site as described in Souret (2002) is also identified in the cTP

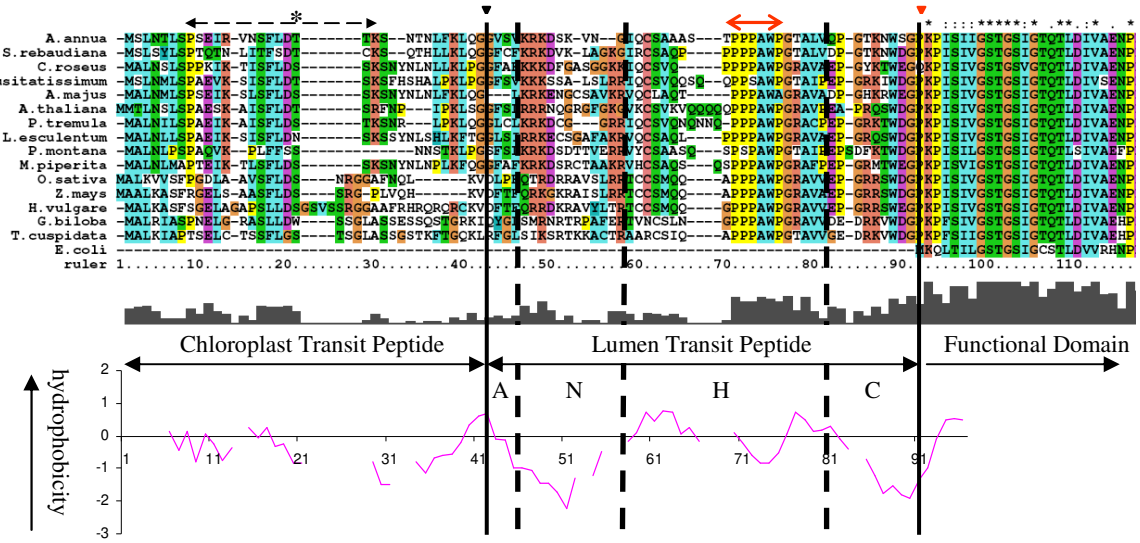


Figure 5. (Top) Alignment of amino acid sequences of DXR tp from the 15 plants. (Bottom) Hydrophobicity plot of DXR amino acid sequence from *Artemisia annua.* Plots for all 15 species can be found in Fig. 12.
(**In the cTP:** ▼: cTP cleavage site; ▼: lTP cleavage site.; * : putative phosphorylation site at Thr; ←------→ : motif containing binding site of the 14-3-3 chaperon
In the lTP: putative acidic A-domain, charged N-domain with twin arginine motif (RR), hydrophobic H-domain with conserved motif "PPPAWPG" (⟷) and polar C-domain containing basic residues).

domain (Fig. 5, *), within a motif similar to the binding site of a 14-3-3 protein involved in plastid transport (Fig. 5, ←– – – –→) (Waegemann and Soll, 1996). In addition, hydrophobicity analysis revealed that the cTP cleavage site coincides with a hydrophobic peak in the dicots (Fig. 5, with *A. annua* plot as example). Comparison to the previous study on DXS revealed a similar hydrophobic peak between amino acid positions 35-45 (Krushal *et al.*, 2003). This peak possibly corresponds to a transmembrane region, where the cTP is anchored in the chloroplast membranes and was subsequently cleaved off.

The remaining of the transit peptide strongly suggests the presence of a ΔpH type of a thylakoid lumen targeting domain. A relatively short *A* domain continues after the cTP cleavage site, a domain known to be of variable length and usually containing acidic residues. Following this is the *N* domain which holds the distinct twin arginine motif. As mentioned earlier, this motif is a characteristic of all precursors that use the ΔpH pathway (Mori and Cline, 2001), hence providing evidence that DXR is likely lumen targeted. In several of the species, the substitution of the RR motif to KR, RK and KK may hinder proper transportation of the precursor, but there is no indication that this would induce the lTP to convert to the Sec pathway (Mori and Cline, 2001). The next 12-18 residues resemble the hydrophobic *H* domain. This domain contains the Pro-rich motif "PPPAWPG" (Fig. 5, ⟷) as described in Carretero-Paulet *et al*. (2002). Interestingly, a Pro-rich motif "PP(T/I)P" is also found in a similar region (amino acid positions 78-81) in the DXS plant sequence (Krushkal *et al.*, 2003). A well-known function of protein domains rich in Pro residues is to mediate protein-protein interactions (Kay *et al.*, 2000). As such, we speculate that this Pro-rich region may play an essential role in the assembly of the ΔpH-dependent pathway translocon (Mori and Kline, 2002) Finally, the lTP ends

with the polar *C* domain, which is enriched in basic residues. The The cleavage site of the lTP can be easily deduced from sequence alignment with *E. coli*. Since bacteria do not contain the transit peptide domain, the start of the *E. coli* alignment would also indicate the start of the functional domain of the protein, hence revealing the likely lTP cleavage site (Fig. 5, ▼).

3.2 Phylogenetic Analysis

3.2.1 Established taxonomy

Taxonomy data for the 16 selected species were retrieved from NCBI (Table IV). The corresponding species tree is plotted in Fig. 6. Comparison of these established taxonomy data with the inferred trees from the 15 DXR sequences provide information on the evolutionary path of DXR.

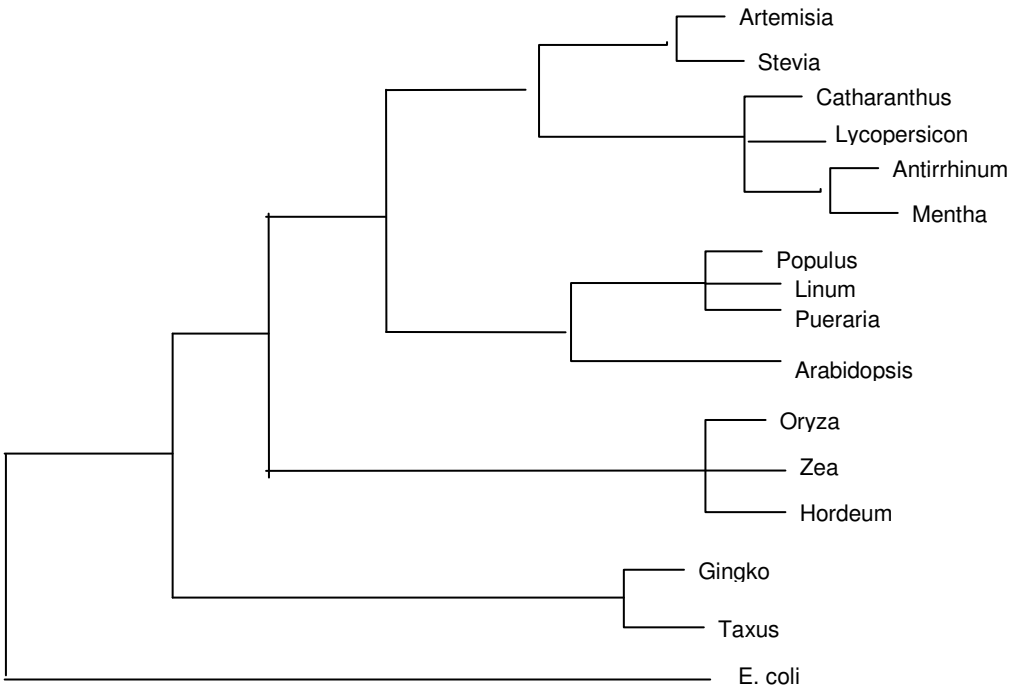


Figure 6. Established species tree based on NCBI data.

Table IV. Taxonomy data retrieved from NCBI for species researched in this project.

Species Name	Taxonomy
Dicots **(* = Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons)**	
Artemisia annua	*; Asterids; campanulids; Asterales; Asteraceae; Asteroideae; Anthemideae; Artemisia.
Stevia rebaudiana	*; Asterids; campanulids; Asterales; Asteraceae; Asteroideae; Eupatorieae; Stevia.
Catharanthus roseus	*; Asterids; lamiids; Gentianales; Apocynaceae; Rauvolfioideae; Vinceae; Catharanthus.
Lycopersicon esculentum	*; Asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
Antirrhinum majus	*; Asterids; lamiids; Lamiales; Plantaginaceae; Antirrhineae; Antirrhinum.
Arabidopsis thaliana	*; Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
Populus alba x Populus tremula	*; Rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
Linum usitatissimum	*; Rosids; eurosids I; Malpighiales; Linaceae; Linum.
Pueraria montana var. lobata	*; Rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Pueraria.
Mentha x piperita	*; Asterids; lamiids; Lamiales; Lamiaceae; Nepetoideae; Nepeteae; Mentha.
Monocots **(* = Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida)**	
Oryza sativa	*; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
Zea mays	*; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
Hordeum vulgare	*; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
Gymnosperm **(* = Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta)**	
Ginkgo biloba	*; Ginkgophyta; Ginkgoales; Ginkgoaceae; Ginkgo.
Taxus cuspidate	*; Coniferopsida; Coniferales; Taxaceae; Taxus.
Bacteria	
E. coli	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

(* species names stem from these points)

3.2.2 Evolution of the functional amino acid domain

Table V: Pairwise amino acid distances among functional regions of DXR from 16 species.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Artemisia		0.013	0.018	0.019	0.019	0.019	0.023	0.019	0.026	0.024	0.019	0.019	0.019	0.020	0.022	0.057
2. Stevia	0.066		0.017	0.018	0.018	0.018	0.023	0.019	0.026	0.023	0.018	0.017	0.019	0.020	0.022	0.058
3. Catharanthus	0.121	0.109		0.014	0.015	0.016	0.021	0.018	0.026	0.024	0.015	0.015	0.017	0.019	0.020	0.058
4. Lycopersicon	0.133	0.112	0.074		0.016	0.016	0.021	0.018	0.025	0.022	0.014	0.013	0.016	0.018	0.018	0.057
5. Antirrhinum	0.133	0.112	0.086	0.097		0.014	0.021	0.018	0.025	0.021	0.016	0.016	0.016	0.020	0.021	0.058
6. Arabidopsis	0.127	0.112	0.092	0.089	0.071		0.020	0.019	0.025	0.022	0.014	0.015	0.016	0.019	0.019	0.057
7. Populus	0.180	0.186	0.148	0.148	0.155	0.139		0.022	0.027	0.027	0.020	0.020	0.021	0.023	0.024	0.062
8. Linum	0.124	0.124	0.118	0.115	0.121	0.133	0.170		0.027	0.025	0.019	0.019	0.019	0.020	0.022	0.058
9. Pueraria	0.225	0.228	0.232	0.215	0.219	0.219	0.245	0.239		0.030	0.025	0.024	0.025	0.028	0.028	0.060
10. Mentha	0.199	0.183	0.189	0.173	0.155	0.161	0.239	0.209	0.297		0.023	0.023	0.023	0.025	0.026	0.062
11. Oryza	0.133	0.115	0.083	0.074	0.089	0.069	0.145	0.127	0.212	0.180		0.012	0.015	0.018	0.019	0.057
12. Zea	0.130	0.109	0.083	0.066	0.092	0.080	0.142	0.124	0.192	0.180	0.049		0.014	0.019	0.019	0.057
13. Hordeum	0.133	0.130	0.109	0.095	0.095	0.089	0.155	0.127	0.209	0.177	0.083	0.069		0.020	0.020	0.058
14. Taxus	0.136	0.139	0.133	0.112	0.145	0.124	0.186	0.139	0.252	0.215	0.121	0.127	0.145		0.016	0.058
15. Gingko	0.173	0.161	0.142	0.121	0.152	0.130	0.192	0.161	0.252	0.225	0.130	0.133	0.142	0.092		0.058
16. E. coli	0.796	0.814	0.814	0.796	0.814	0.808	0.889	0.820	0.863	0.889	0.802	0.802	0.814	0.814	0.814	

Lower left corner: pairwise distances values
Upper right corner: standard errors

A distance matrix for the conserved functional part of the DXR amino acids sequence shows that, as expected, the bacterium, *E. coli,* has the largest pair-wise distance value from the other plant species (Table V). A phylogenetic tree was then inferred from the distance matrix using the neighbor-joining method, with amino acid distance adjusted by the Poisson correction (Fig. 7).

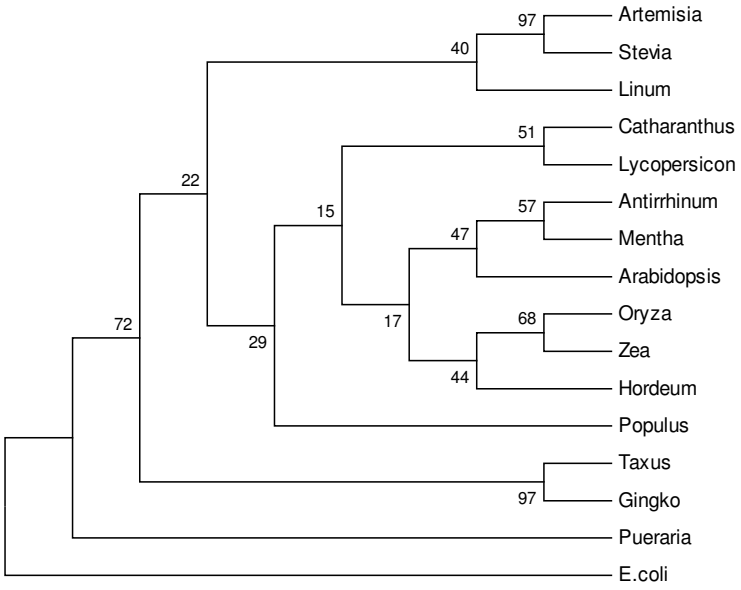


Figure 7. Tree inferred from the functional amino acid sequence of DXR using neighboring-joining method.

Compared to the NCBI tree (Figure 6), several misclusterings were present in the inferred tree using the neighbor-joining method (Fig. 7). The first is the misclustering in *Pueraria, Populus and Linum.* Though they all belong to the *Eurosid I* family, none of them were clustered together in the inferred tree (Fig. 7). Moreover, *Pueraria* was separated from the other Eukaryotes and clustered with *E. coli*, with a bootstrap value of 72%. Another misclustering was with *Arabidopsis,* which was not clustered with the other Rosids. The monocots, though being clustered together, formed a clade within the other dicots which violates the established plant taxonomy. Most of these misclusterings, however, were supported by weak bootstrap values of less than 75%. High bootstrap values were only seen in the correct clustering of the dicots *Artemisia* with *Stevia*, and the gymnosperms *Taxus* with *Gingko,* both supported by a bootstrap value of 97%. Trees inferred by the other methods derived similar conclusions (Fig. 8).

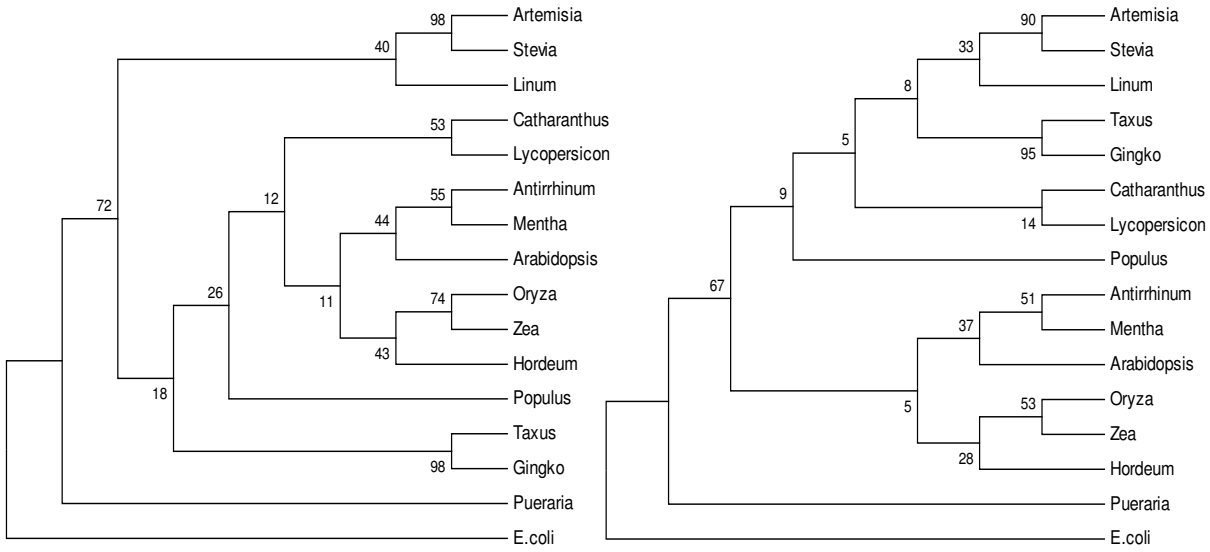


Figure 8. Phylogenetic trees inferred from the functional amino acid sequence of DXR using two different methods.
(**Left:** minimum evolution method; **Right:** maximum parsimony method)

3.2.3 Evolution of the transit peptide domain

As mentioned earlier, the full amino acid sequence contains an N-terminal transit peptide domain which is highly mutated and rather unconserved. To investigate if the apparent mutational events leading to the observed transit peptide diversity are related to the evolutionary path of the plant species, phylogenetic trees were inferred from the transit peptide along with the first ten residues of the functional domain (Fig. 9). Interestingly, using the "maximum parsimony" method, a tree was produced in which the species were correctly clustered into dicots, monocots and gymnosperms (Fig. 9 top), with high bootstrap values of 98%. As such, one can imply that although mutation is a random event, mutation in species within close taxonomy does follow a similar trend and rate, at least in the case of the transit peptide.

Trees inferred using neighbor-joining and minimum evolution methods, (Fig. 9 bottom) produced similar results, except for the violation seen in *Arabidopsis*. It is the only member of the *eurosids II* in the list and this may, thus, be a reason for its distance from the other dicots.

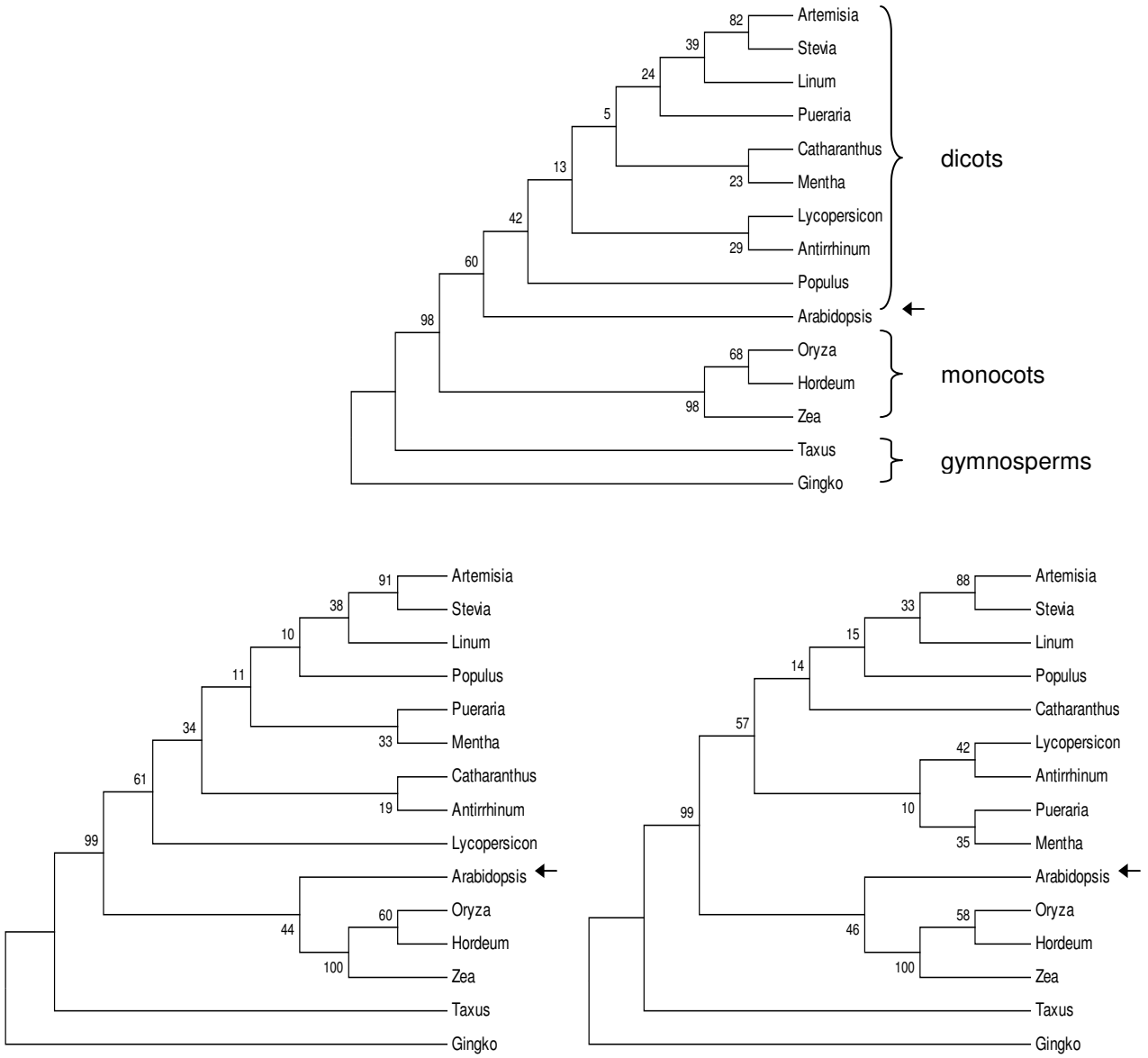


Figure 9. Phylogenetic trees inferred from the transit domain of DXR.
(**Top:** maximum parsimony method; **Bottom left:** neighboring-joining method;
Bottom right: minimum evolution method; *Arabidopsis* indicated by ←)

3.2.4 Evolution of the full nucleotide sequence

When phylogenetic trees are inferred from the full nucleotide sequence, i.e. transit peptide and functional part (Fig. 10; Fig. 11), they showed a topology almost identical to the established species tree from NCBI (Fig. 6). The maximum parsimony method, in particular, produced a tree most closely aligned with the NCBI tree among the three methods used. The tree was also supported by high bootstrap values in most of the nodes (>80%). The only misclustering was in *Linum*, which should be clustered with *Populus* and *Pueraria*, as members of the *eurosids I*. Nevertheless, the misclustering is only supported by a low bootstrap value of 33%. With the resemblance observed between the DXR gene tree and the NCBI species tree, there is evidence that the evolution of DXR sequences follows closely with the divergence of the plant species in this study.

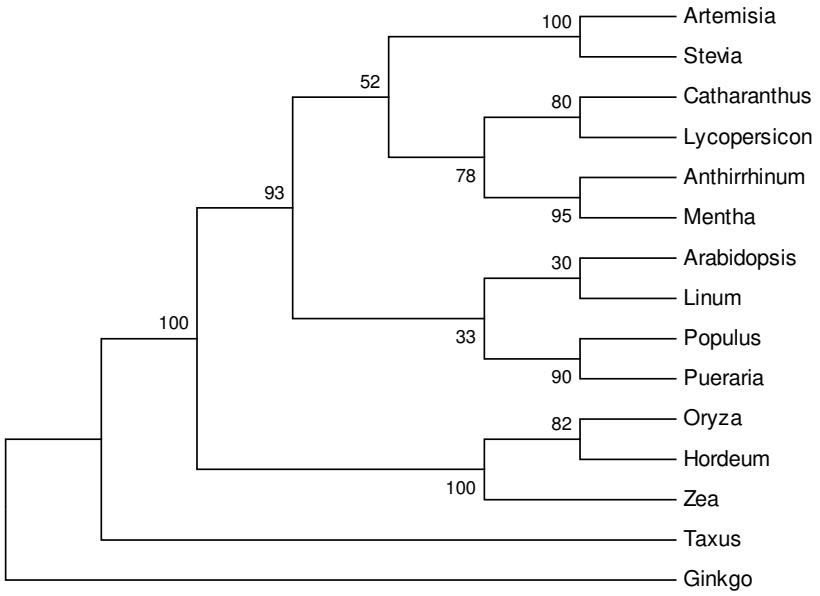


Figure 10. Phylogenetic tree inferred from the full nucleotide sequence of DXR using the maximum parsimony method.

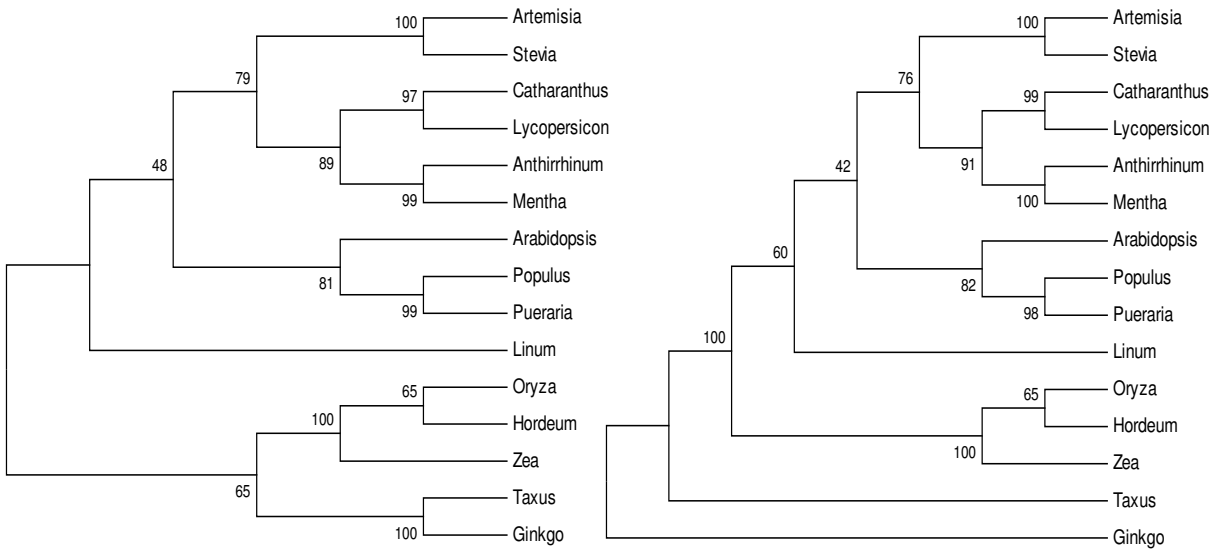


Figure 11. Phylogenetic trees inferred from the full nucleotide sequence.
(**Left:** neighbor-joining method; **Right:** minimum evolution method)

Unlike our previous study on DXS (Krushkal *et al.*, 2003) where phylogenetic analysis revealed the separation of DXS into DXS 1 and DXS 2, there is no strong indication that DXR is also separated into two classes based on the inferred trees.

3.3 Secondary Structure Prediction

Using the methods GOR IV and HNN, secondary structure prediction was performed on the conserved functional region of the DRX amino acid sequences form 15 plant species. These analyses will help to identify any conserved structural regions which may be altered in certain plant species due to mutations. Six regions of such interest were identified (Appendix II).

Region I (positions 1-25 of the protein alignment) was predicted to be structurally conserved by both GOR IV and HNN, with exception in positions 16-21. In this short region, the structure predicted by GOR IV showed extended strands in *Catharanthus, Linum* and the gymnosperms, while the other species were predicted as alpha helixes. Several amino acid substitutions were identified to support this structural difference. In *Catharanthus,* there is at position 12 an amino acid substitution from a larger isoleucine (I) to a smaller valine (V). *Linum* has a substitution from non-polar alanine (A) to polar, hydroxylated serine (S) at position 21. The gymnosperms had a substitution from small, neutral asparagine (N) to basic histidine (H) at position 23. It is interesting to note that all these substitutions correspond to a change in the residue size. Therefore, although none of these substitutions were located directly within positions 16-21, substituting to a smaller or larger residue may still have an effect on the secondary structure prediction of nearby regions.

In region II (positions 165-175), the structure predicted using GOR IV showed extended strands in *Anthirrhinum, Linum, Pueraria, Hordeum* and the two gymnosperms. The other species were all predicted in this region to be alpha helixes. Both *Anthirrhinum* and *Hordeum* had an amino acid substitution at position 170, while *Pueraria* had a

substitution at position 171. There were no significant substitutions in the gymnosperms that set them apart from the rest.

Region III (positions 181-200) contains a conserved chain of alpha helixes as predicted by both GOR IV and HNN. The only exception is in *Ginkgo*, whose structure is predicted as random coils and extended strands. Amino acid substitutions were found at position 185 (basic glutamic acid (E) to neutral glycine (G)) and at position 193 (non-polar alanine (A) to polar serine (S)); both positions may contribute to the predicted structural difference in *Ginkgo*.

Region IV (positions 241-260) can be grouped into two classes based on the amino acid at position 250. Plants with valine (V) at this position had an extended strand motif as predicted by GOR IV, while those with isoleucine (I) were predicted to have an alpha helix motif. Although exceptions were seen in *Pueraria* and *Mentha,* it appears that the amino acid at position 250 plays a crucial role in affecting the secondary structure of this region.

Region V (positions 310 -327) is predicted by both GORIV and HNN as a highly conserved region of random coils and helixes. The only exceptions were in *Populus* and *Pueraria,* which were predicted to contain extended strands in positions 315-322 by both methods. These positions correspond to a conserved motif "M(X)LAY(X)A", which is highly mutated in both *Populus* and *Pueraria*.

Region VI (positions 370-400) is located towards the C-terminal end of the DXR protein. Both GOR IV and HNN predicted the presence of a continuous chain of alpha helixes. In *Catharanthus,* however, this chain was disrupted by some random coil. The variation was likely due to an amino acid substitution at position 383.

The structural differences discussed in the above six regions are summarized in Table I. These differences were predicted using the GOR IV methods. Predictions using the HNN method were not significantly different for the six regions. Furthermore, the gymnosperms have consistently similar secondary structure in regions I to VI, which corresponds to the high bootstrap value of 97%. (Fig. 7). Likewise, both *Artemisia* and *Stevia* exhibit similar structures in the six regions, and are also supported by a bootstrap value of 97% (Fig. 7).

Table VI. Six regions of the DXR functional domain with structural differences predicted by GOR IV among 15 plant species.

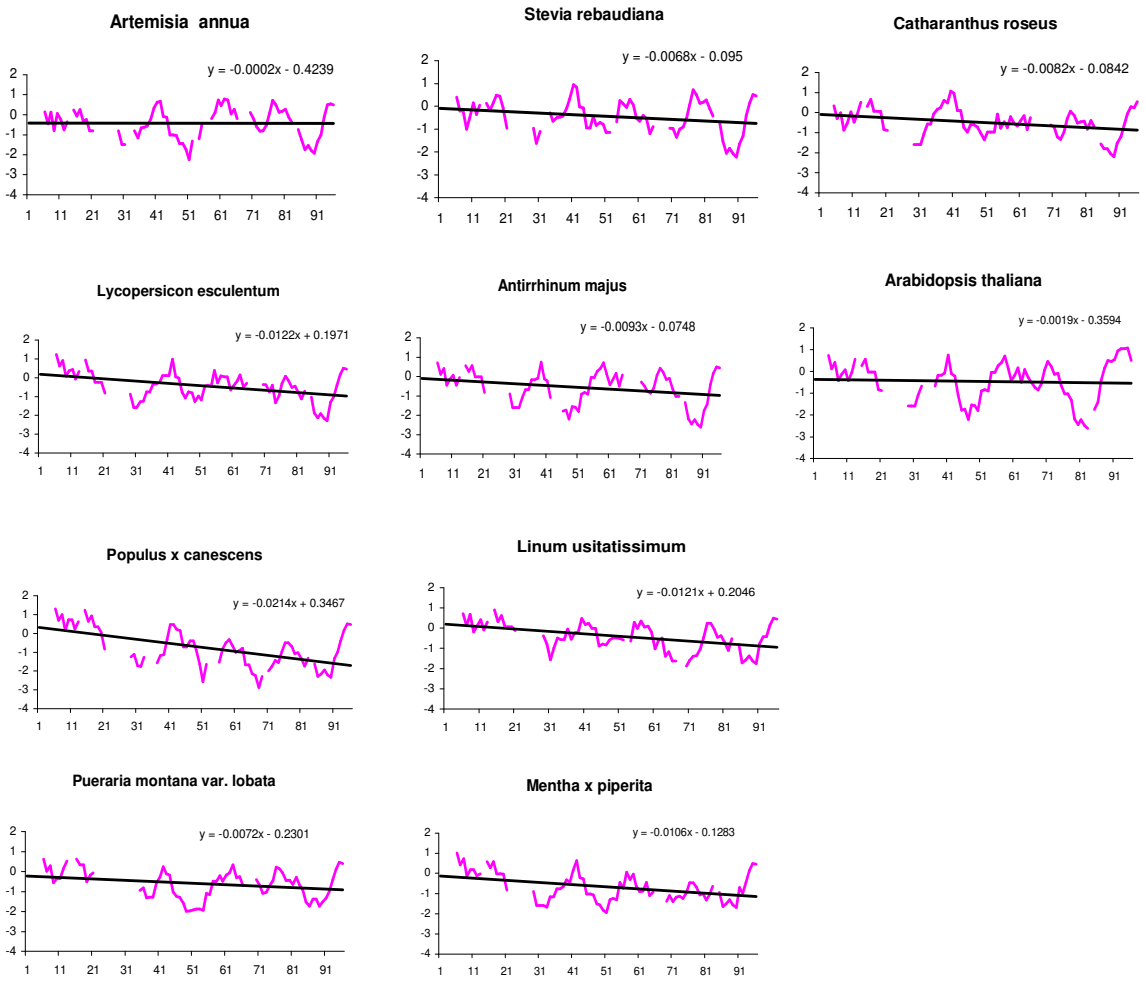
Region	**Extended strand**	**Alpha helix**	**Random coil**
I	*Catharanthus, Linum, Taxus, Ginkgo*	Others	-
II	*Anthirrhinum, Linum, Pueraria, Hordeum, Taxus, Ginkgo*	Others	-
III	*Gingko*	Others	-
IV	*Artemisia, Stevia, Catharanthus, Lycopersicon, Mentha, Zea, Taxus, Gingko*	Others	-
V	*Populus, Pueraria*	Others	-
VI	-	Others	*Catharanthus*

3.4 Hydrophobicity analysis

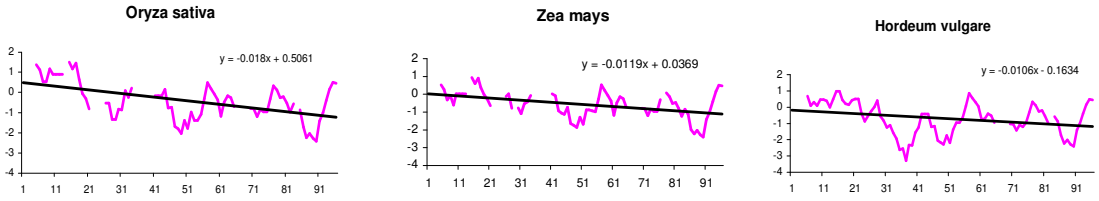
3.4.1 Transit peptide domain of DXR

Hydrophobicity plots for the fifteen selected DXR sequences are presented in Fig. 12 for the transit peptide region and regression analysis was performed. Although some species have significantly steeper negative slopes than others, the slopes of the resulting regression lines are negative for all species – a feature also observed in the DXS transit peptide (Krushkal *et al.*, 2003). The plots also have similar shapes and share some distinct features. For example, a hydrophobic peak can be observed between amino acid positions 35-45 in the dicot family, though the peak is less pronounced in the monocots and gymnosperms. As mentioned earlier, this peak corresponds to the putative cTP cleavage site. A corresponding peak was located at similar positions in the transit peptide of DXS (Krushkal *et al.*, 2003), suggesting that this is likely a hydrophobic trans-membrane region which may serve a role in the translocation of the peptide. Another distinct feature is the sharp rise in hydrophobicity between amino acid positions 91-100 found in all species, which had been identified as a feature of the H domain in the lTP. The presence of these two hydrophobic peaks likely corresponds to two trans-membrane regions, suggesting that the DXR transit peptide is bipartite.

Dicots



Monocots



Gymnosperm

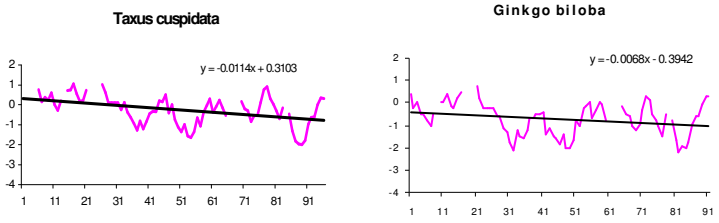


Figure 12.
Hydrophobicity plots of the transit peptide of DXR

3.4.2 Functional domain of DXR

Hydrophobicity plots of the functional DXR can be found in Appendix III. Comparison of the hydrophobicity plots for the functional protein revealed high similarity among all 15 species (Fig. 14), in accordance with the fact that the functional domain of DXR is highly conserved. A total of six prominent hydrophobic and hydrophilic peaks were identified (Table VII). In addition, there were two distinct hydrophilic peaks present only in selective species (Table VIII). These eight peaks were illustrated using *Artemisia* as an example (Fig. 12).

Three hydrophobic and three hydrophilic peaks were found common among all 15 species, with the exception of *Pueraria* at positions 24, 97, and 201 of the functional domain. Position 24 falls into region I of the predicted secondary structure, which shows an amino acid substitution from hydrophilic asparagine (N) to hydophobic phenylalanine (F) in *Pueraria*. Similarly, position 201 falls into region III of the predicted secondary structure, and have a hydrophilic asparagine (A) to hydrophobic isoleucine (I) substitution in the *Pueraria* sequence. No amino acid variation is observed at position 97.

Both *Arabidopsis* and *Mentha* were found to contain an additional hydrophilic peak at position 308. Another hydrophilic peak is also observed at position 362 but only in *Arabidopsis, Pueraria, Taxus* and *Ginkgo*. However, no amino acid variations are observed at these two positions to explain the absence of these peaks in the other species.

Comparison of Table VII and VIII showed that the hydrophobicity characteristic of *Pueraria* is different from the other dicots in most instances. Discrepancies are also present in *Arabidopsis* and *Mentha*. These three species were also the most distanced from the other dicots in the tree inferred using the maximum parsimony method (Fig. 8).

Table VII. Shared hydrophilic and hydrophobic peaks in the functional domain of DXR.

Position	*Ar*	*St*	*Ca*	*Ly*	*An*	*Ar*	*Po*	*Li*	*Pu*	*Me*	*Or*	*Ze*	*Ho*	*Ta*	*Gi*	Peak
24	-1.233	-1.3	-1.3	-1.3	-1.3	-1.3	-1.233	-1.522	**-0.6**	-1.3	-1.3	-1.3	-1.3	-1.2	-1.267	hydrophilic
31	1.789	1.722	1.722	1.722	1.722	1.722	1.789	1.789	1.5	1.756	1.722	1.722	1.722	1.789	1.722	hydrophobic
47	-1.2	-1.2	-1.511	-1.622	-1.544	-1.411	-1.578	-1.544	-1.478	-1.544	-1.556	-1.556	-1.556	-1.333	-1.933	hydrophilic
97	2.444	2.444	2.367	2.367	2.367	2.367	2.356	2.1	**1.533**	2.367	2.367	2.367	2.367	2.356	2.356	hydrophobic
131	2.056	2.056	2.046	2.056	2.056	2.056	2.056	2.056	2.089	2.056	2.056	2.056	2.056	2.056	2.056	hydrophobic
201	-1.878	-1.811	-2.111	-2.111	-2.111	-2.111	-1.811	-1.811	**-0.778**	-2.022	-2.111	-2.178	-1.811	-2.111	-2.111	hydrophilic

Table VIII. Unique hydrophilic peaks in the functional domain of DXR.

Position	*Ar*	*St*	*Ca*	*Ly*	*An*	*Ar*	*Po*	*Li*	*Pu*	*Me*	*Or*	*Ze*	*Ho*	*Ta*	*Gi*	Peak
309	-1.478	-1.478	-1.756	-1.478	-1.211	**-2.111**	0	-1.544	-1.178	**-2.389**	-1.478	-1.544	-1.478	-0.811	-0.811	hydrophilic
362	-1.133	-0.833	-0.544	-1.244	-1.311	**-1.833**	-1.133	-1.244	**-1.722**	-1.533	-1.2	-1.2	-1.411	**-2.689**	**-1.978**	hydrophilic

(Ar = Artemisia; St =Stevia; Ca =Catharanthus; Ly = Lycopersicon; An = Anthirrhinum; Ar = Arabidiposis; Po = Populus; Li = Linum; Pu = Pueraria; Me = Mentha; Or = Oryza; Ze = Zea; Ho = Hordeum; Ta = Taxus; Gi = Ginkgo)

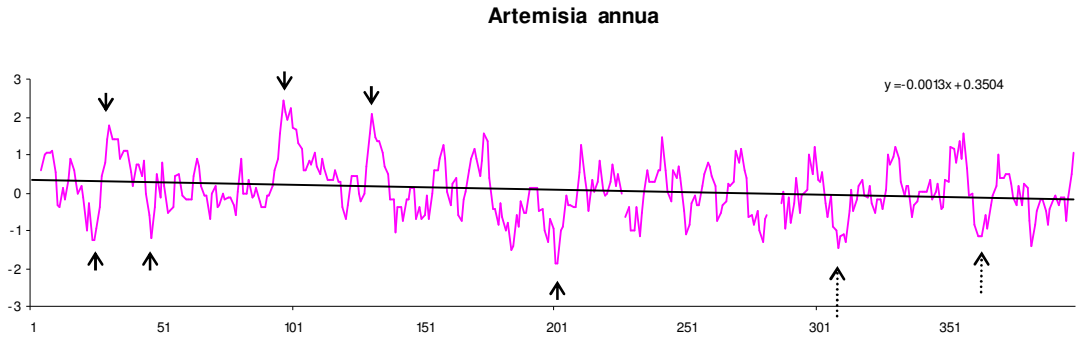


Figure 13. The eight hydrophobic and hydrophilic peaks in the functional domain of DXR, as indicated in Table VII (solid arrow) and Table VIII (dash arrow).

4 Conclusion

These results showed that DXR sequences from the 15 plant species are possibly bipartite, containing features unique to the use of the ΔpH transit pathway. To my understanding, this is the first suggestion that DXR proteins are not only targeted to the chloroplast, but also to the thylakoid lumen. In addition, I have identified several amino acids residues from the functional domain of DXR, whose mutations have led to a significant change in either the secondary structure or hydrophobicity behavior of the region. Furthermore, phylogenetic analyses revealed only a single class of DXR, in contrast to the separation of two classes in DXS.

REFERENCES

Carretero-Paulet, L., Ahumada, I., Cunillera, N., Rodriguez-Concepcion, M., Ferrer, A., Boronat, A., Campos, N. 2002. Expression and molecular analysis of the Arabidopsis DXR gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase, the first committed enxyme of the 2-C-methyl-D-erythritol 4-phosphate pathway. *Plant Physiology* 129: 1581-1591.

Croteau, R., Kutchan, T.M. and Lewis, N.G. 2000. Natural Products (Secondary Metabolites). In: *Biochemistry & Molecular Biology of Plants*, B.Croteau, W. Gruissem, R. Jones, Eds. American Society of Plant Physiologists: 150-1268.

Dubey, V.S., Bhalla, R. and Luthra, R. 2003. An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants. *Journal of Biosciences* 28: 637-646.

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

Emanuelsson, O., von Heijne, G., 2001. Prediction of organellar targeting signals. *Biochimica et Biophysica Acta* 1541:114-119.

Estévez, C.M., Cantero, A., Reindl, A., Reichler, S. and Leon, P. 2001. 1-deoxy-D-xylulose 5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *Journal of Biological Chemistry* 276: 22901-22909.

Garnier, J., Gibrat, J-F and Robson, B. 1996. GOR secondary structure prediction method version IV. *Method in Enzymology* 226: 540-553.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: *The Proteomics Protocols Handbook,* Walker, J.M., Eds. Humana Press: 571-607.

Guermeur, Y. 1997. Combinaison de classifieurs statistiques, Application a la prediction de structure secondaire des proteines. PhD. Thesis. University of Paris.

Jackson-Constan, D., Akita, M. and Keegstra, K. 2001. Molecular chaperones involved in chloroplast protein import. *Biochimica et Biophysica Acta* 1541: 102-113.

Kay, B.K., Williamson, M.P. and Sudol, M. 2000. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *The FASEB Journal* 14: 231-241.

Krushal, J., Pistilli, M., Ferrell, K.M., Souret, F.F. and Weathers, P.J. 2003. Computational analysis of the evolution of the structure and function of 1-deoxy-D-xylulose 5-phosphate synthase, a key regulator of the mevalonate-independent pathway in plants. *Gene* 313: 127-138.

Kumar S., Tamura K., Jakobsen I.B., & Nei M. 2001. MEGA 2: Molecular Evolutionary Genetics Analysis Software. *Bioinformatics* 17 (12): 1244-1245.

Kuzuyama, T., Takahashi, S., Takagi, M. and Seto, H. 2000. Characterization of 1-Deoxy-D-xylulose 5-Phosphate Reductoisomerase, an Enzyme Involved in Isopentenyl Diphosphate Biosynthesis, and Identification of Its Catalytic Amino Acid Residues. *Journal of Biological Chemistry* 275(26): 19928-19932.

Kyte, J. and Dolittle, R. 1982. A simple method for displaying the hydropathic character of proteins. *Journal of Molecular Biology* 157: 105-132.

Mahmoud, S.S. and Croteau, R.B. 2001. Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. *Proceedings of the National Academy of Science of the United States of America* 98(15): 8915-8920.

Mahmoud, S.S. and Croteau, R.B. 2002. Strategies for transgenic manipulation of monoterpene biosynthesis in plants. *Trends in Plant Science* 7(8): 366-373.

Mori, H., Cline, K. 2001. Post-translational protein translocation into thylakoids by the Sec and ΔpH-dependent pathways. *Biochimica et Biophysica Acta* 1541: 80-90.

Rane, M.J., Calvo, K.C. 1997. Reversal of the nucleotide specificity of ketol acid reductoisomerase by site-directed mutagenesis identifies the NADPH binding site. *Archives of Biochemistry and Biophysics* 338(1): 83-89.

Sehnke, P.C., Henry, R., Cline, K., Ferl, R.J. 2000. Interaction of a plant 14-3-3 protein with the signal peptide of a thylakoid-targeted chloroplast precursor protein and the presence of 14-3-3 isoforms in the chloroplast stroma. *Plant Physiology* 122: 235-241.

Souret, F.F. 2002. Transcriptional regulation of terpenoid genes in transformed roots of *Artemisia annua*. Ph.D. Thesis in Biotechnology, Worcester Polytechnic Institute, Worcester, MA.

Takahashi, S., Tomohisa, K., Hiroyuki, W. and Seto, H. 1998. A 1-deoxy-D-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-D-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis. *Proceedings of the National Academy of Sciences* 95: 9879-9884.

Thompson,J.D., Gibson,T.J., Plewniak,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.

Veau, B., Courtois, M., Oudin, A., Chenieux, J-C., Rideau, M., Clastre, M. 2000. Cloning and expression of cDNAs encoding two enzymes of the MEP pathway in *Catharanthus roseus. Biochimica et Biophysica Acta* 1517(1):159-63.

Walter, M.H., Fester, T., Stract, D. 2000. Arbuscular mycorrhizal fungi induce the non-melvalonate methylerythritol phosphate pathway of isoprenoid biosynthesis correlated with accumulation of the "yellow pigment"and other apocarotenoids. *Plant Journal* 21: 571-578.

Waegemann, K., Soll, J. 1996. Phosphorylation of the transit sequence of chloroplast precursor protein. *The Journal of Biological Chemistry* 271(11): 6545-6554.

Westerlund, I., von Heijne, G. and Emanuelsson, O. 2003. LumenP- A neural network predictor for protein localization in the thylakoid lumen. *Protein Science* 12: 2360-2366.

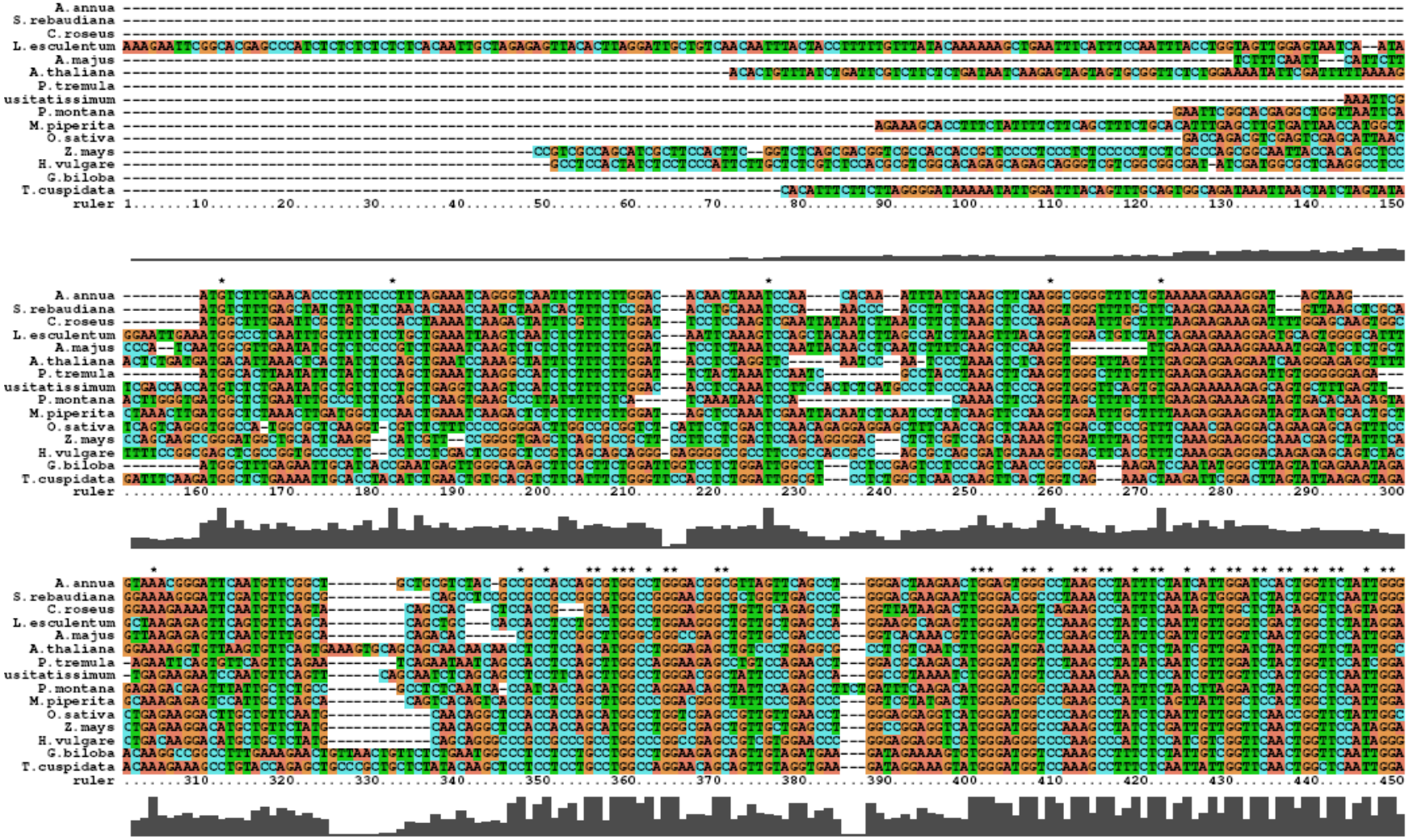
Appendix I

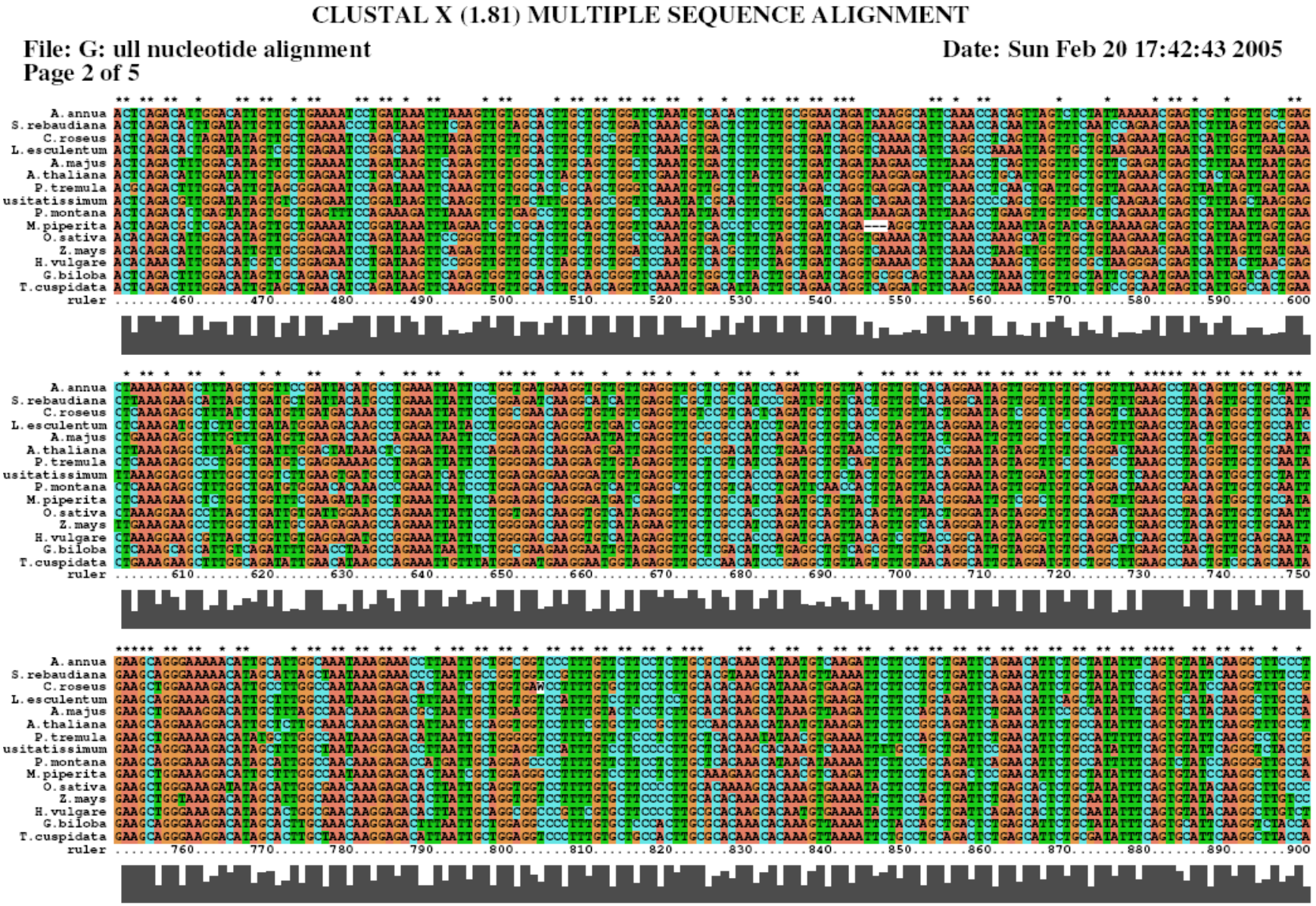
Sequence alignment of nucleotide sequences and amino acid sequences of DXR

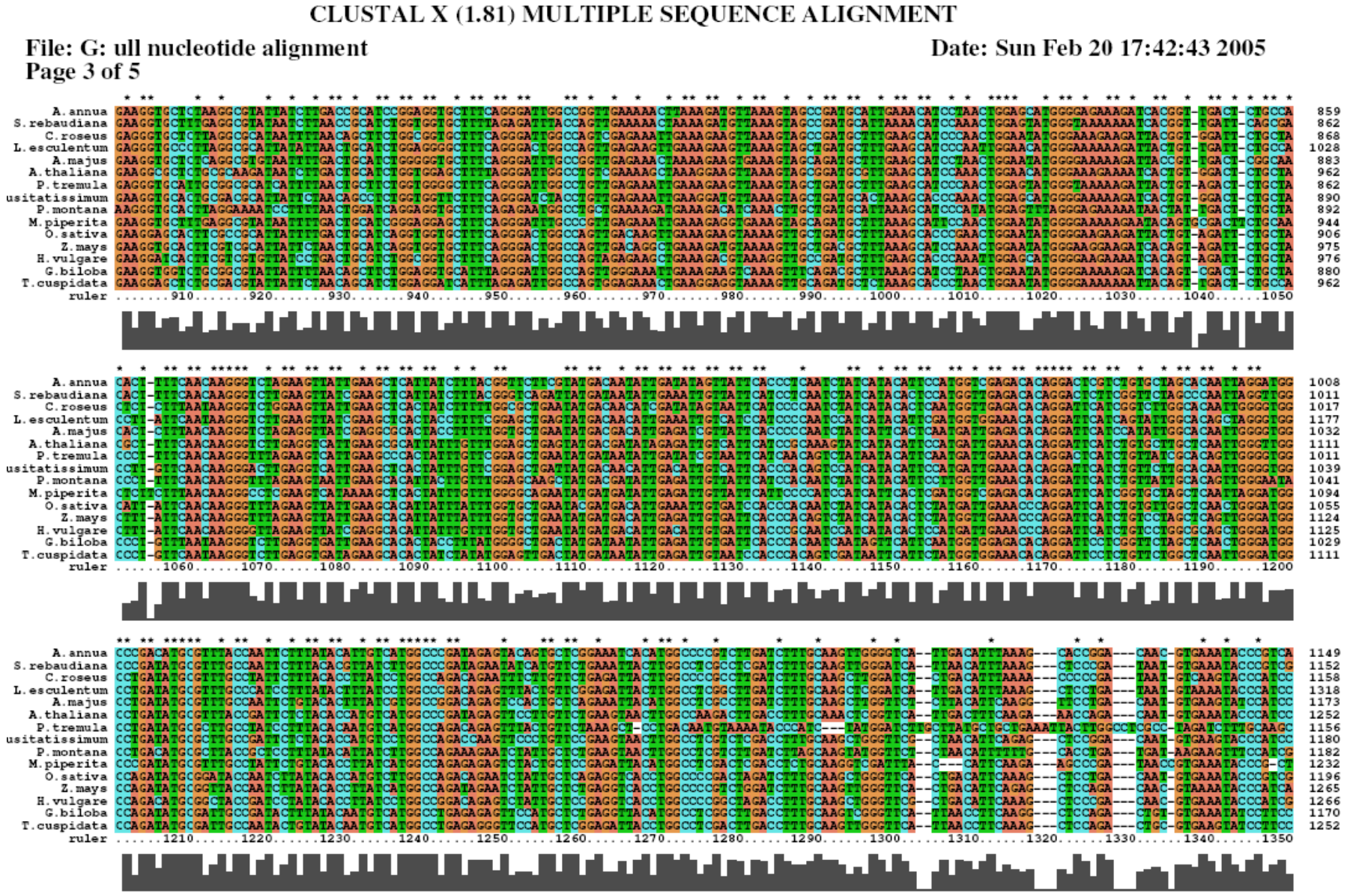
CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

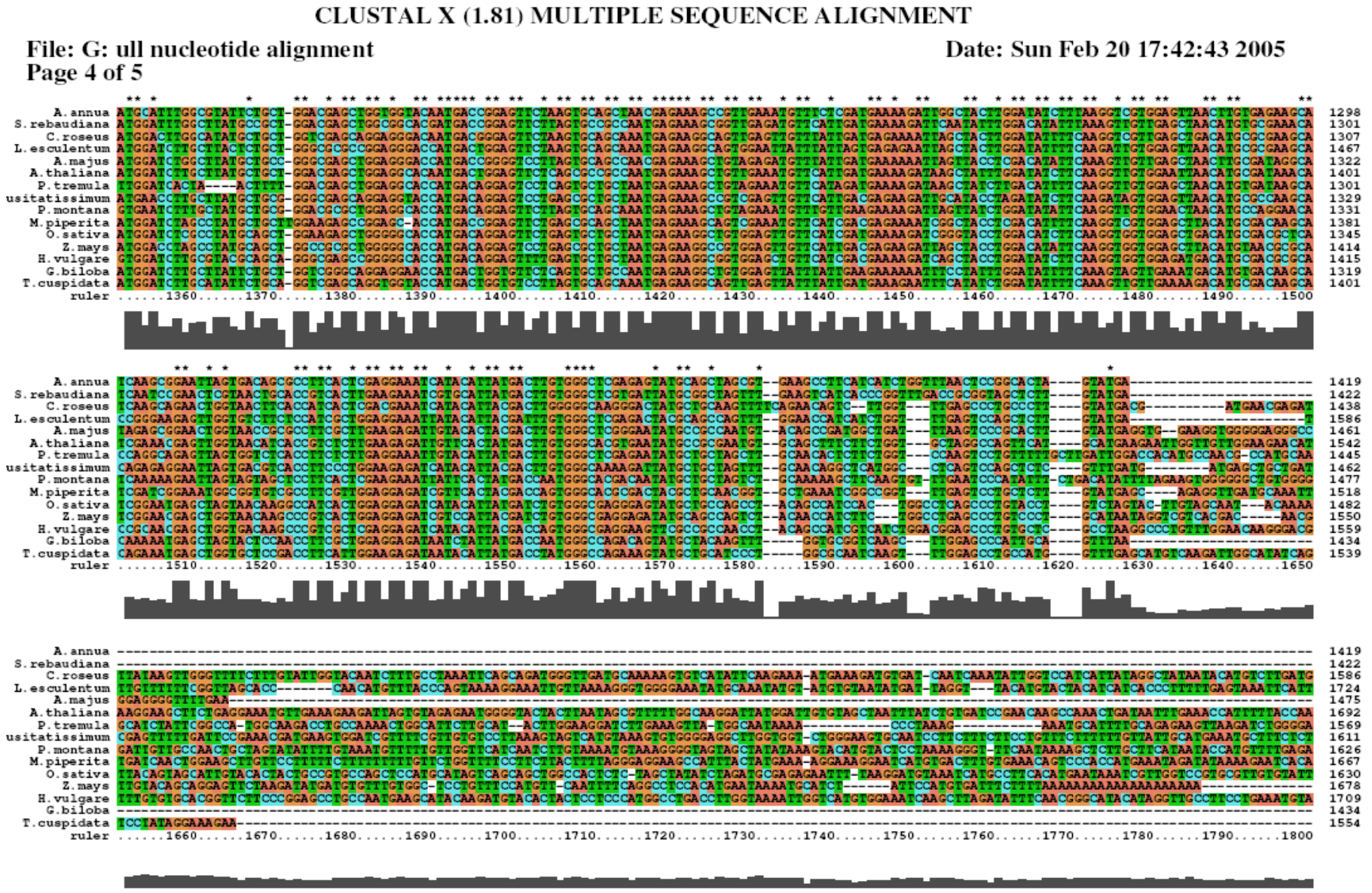
File: G: ull nucleotide alignment **Date: Sun Feb 20 17:42:43 2005**

Page 1 of 5




CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT
File: G: ull nucleotide alignment
Date: Sun Feb 20 17:42:43 2005
Page 2 of 5
A.annua
S.rebaudiana
C.roseus
L.esculentum
A.majus
A.thaliana
P.tremula
usitatissimum
P.montana
M.piperita
O.sativa
Z.mays
H.vulgare
G.biloba
T.cuspidata
ruler

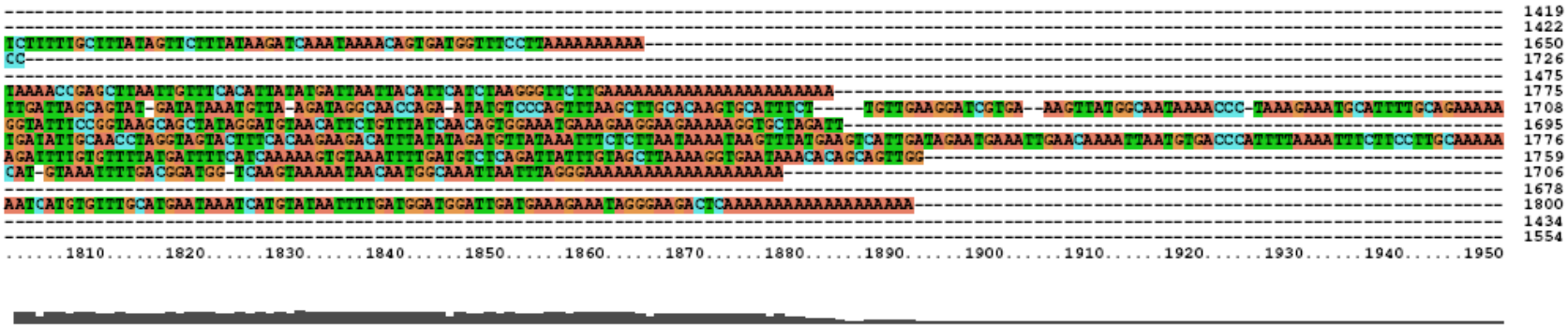

CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT
File: G: ull nucleotide alignment
Date: Sun Feb 20 17:42:43 2005
Page 3 of 5
A.annua
S.rebaudiana
C.roseus
L.esculentum
A.majus
A.thaliana
P.tremula
usitatissimum
P.montana
M.piperita
O.sativa
Z.mays
H.vulgare
G.biloba
T.cuspidata
ruler

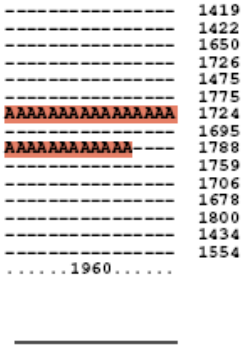

CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT
File: G: ull nucleotide alignment
Date: Sun Feb 20 17:42:43 2005
Page 4 of 5
A.annua
S.rebaudiana
C.roseus
L.esculentum
A.majus
A.thaliana
P.tremula
usitatissimum
P.montana
M.piperita
O.sativa
Z.mays
H.vulgare
G.biloba
T.cuspidata
ruler

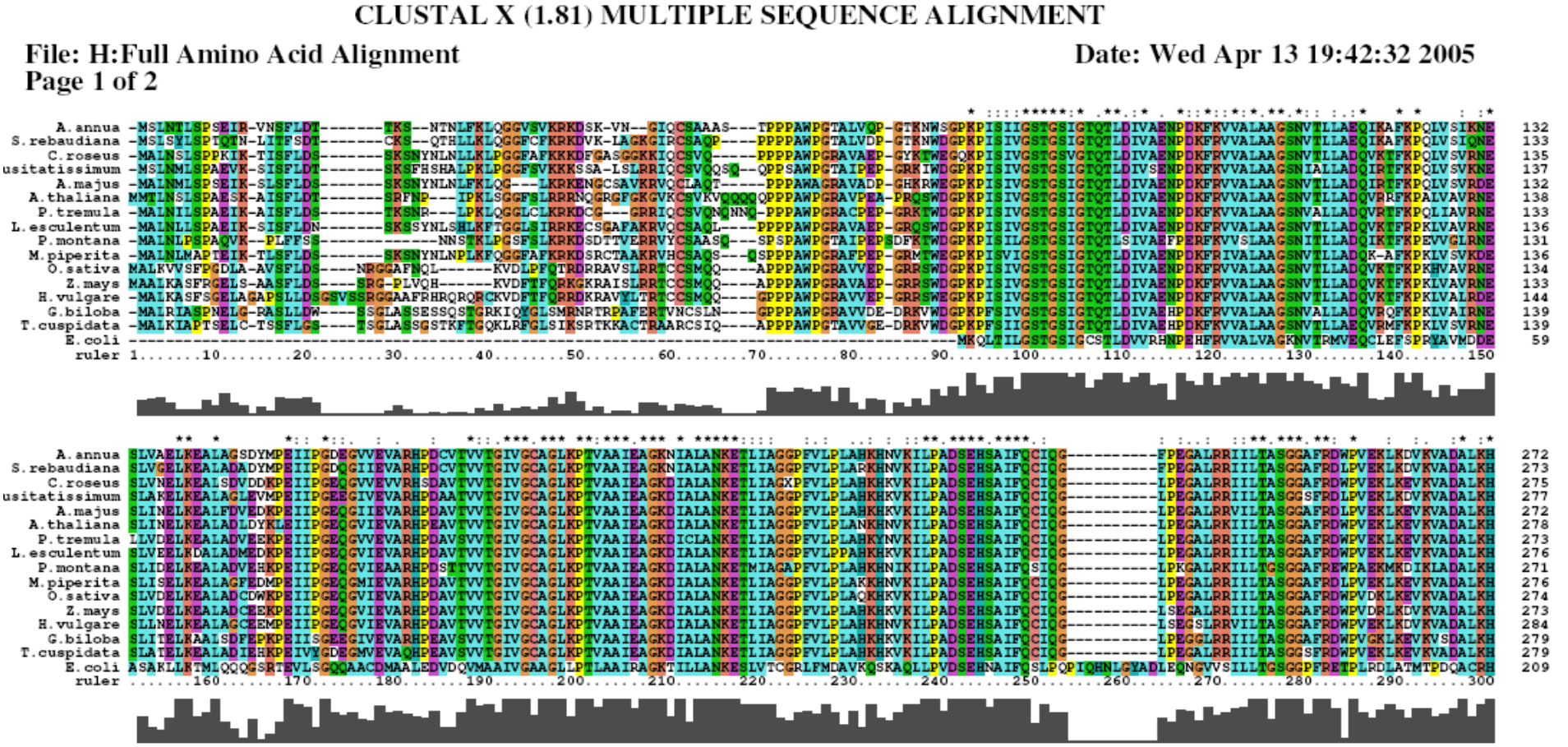
CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: G: ull nucleotide alignment **Date: Sun Feb 20 17:42:43 2005**

Page 5 of 5



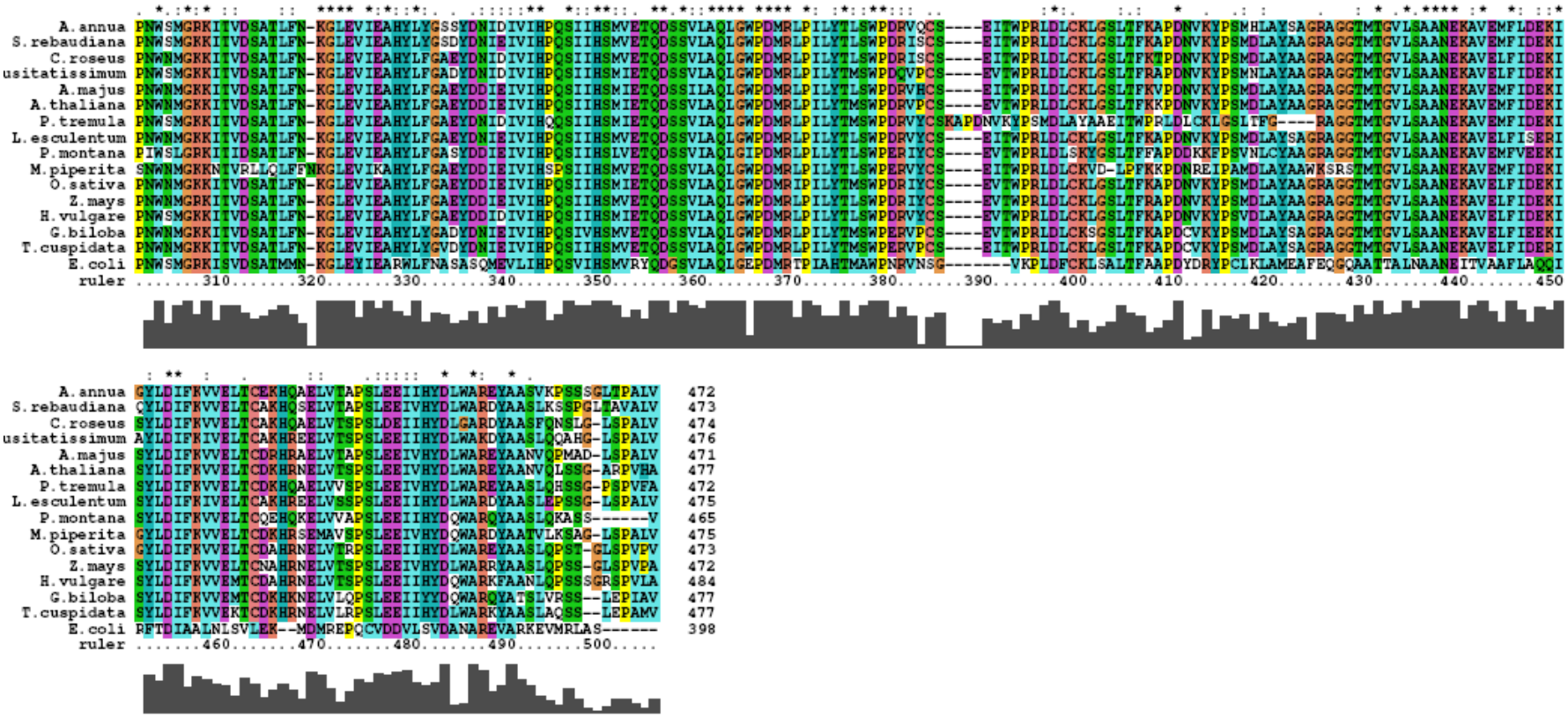



CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT
File: H:Full Amino Acid Alignment
Date: Wed Apr 13 19:42:32 2005
Page 1 of 2
A.annua
S.rebaudiana
C.roseus
usitatissimum
A.majus
A.thaliana
P.tremula
L.esculentum
P.montana
M.piperita
O.sativa
Z.mays
H.vulgare
G.biloba
T.cuspidata
E.coli
ruler

CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: H:Full Amino Acid Alignment **Date: Wed Apr 13 19:42:32 2005**

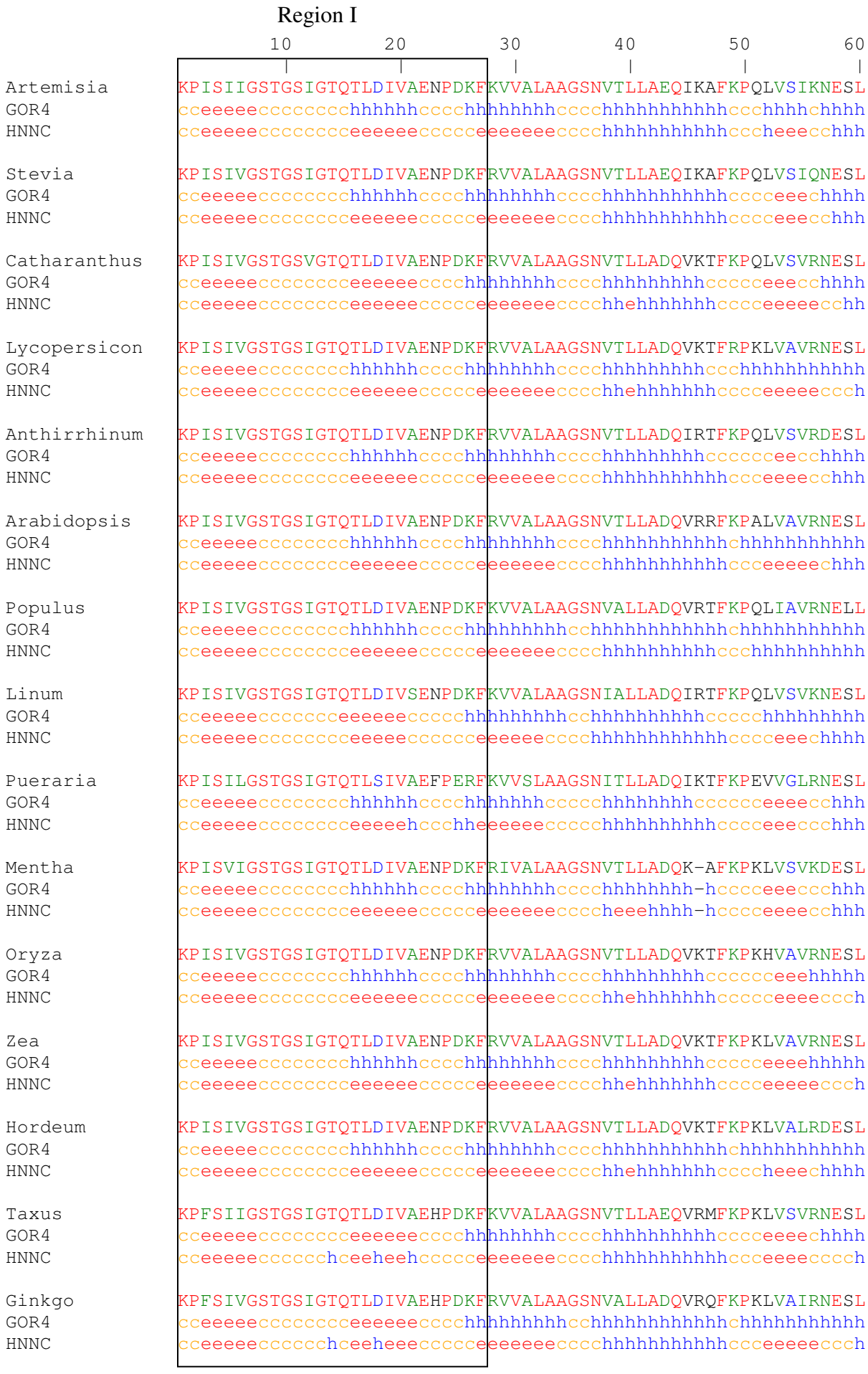
Page 2 of 2

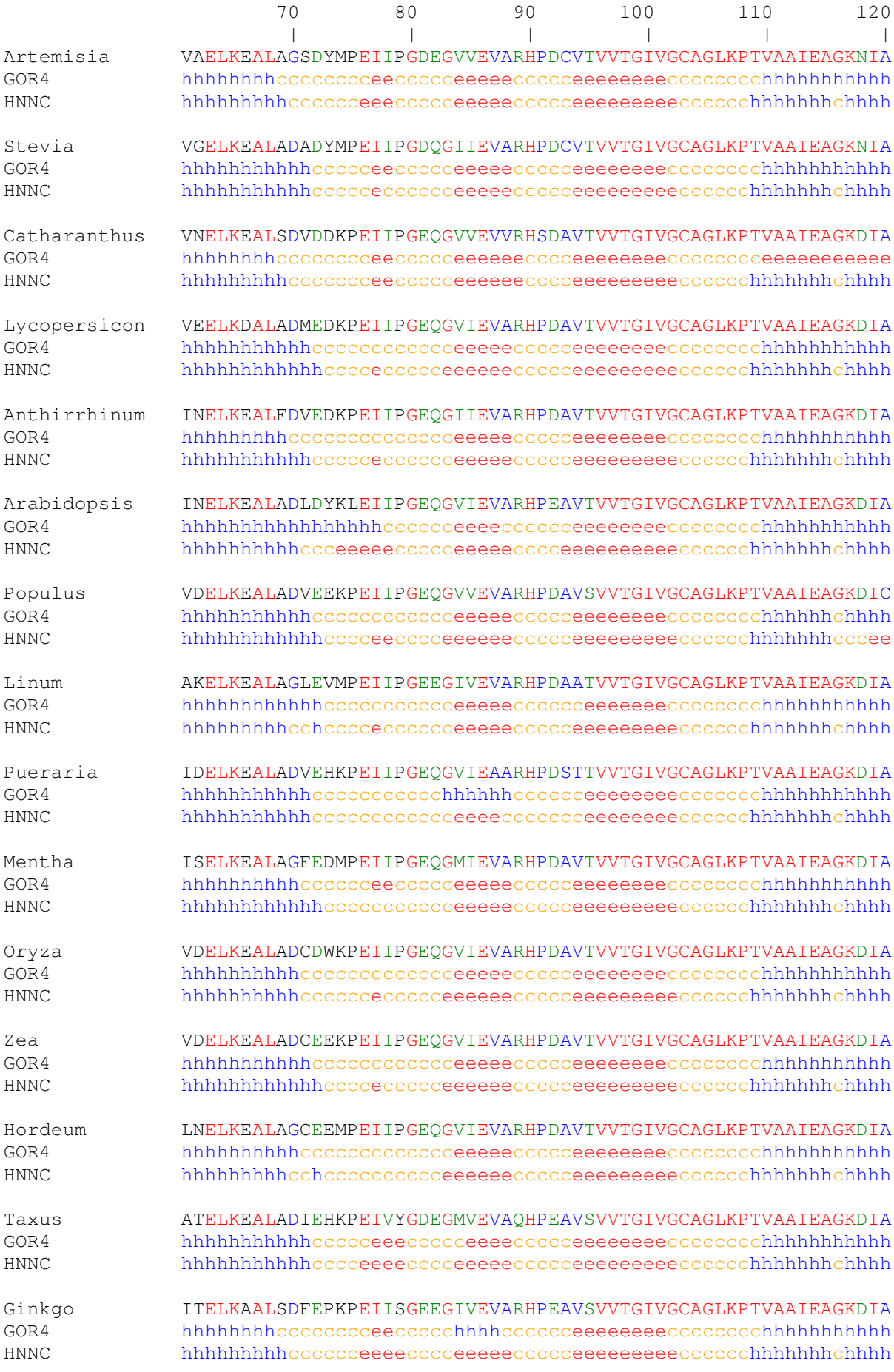


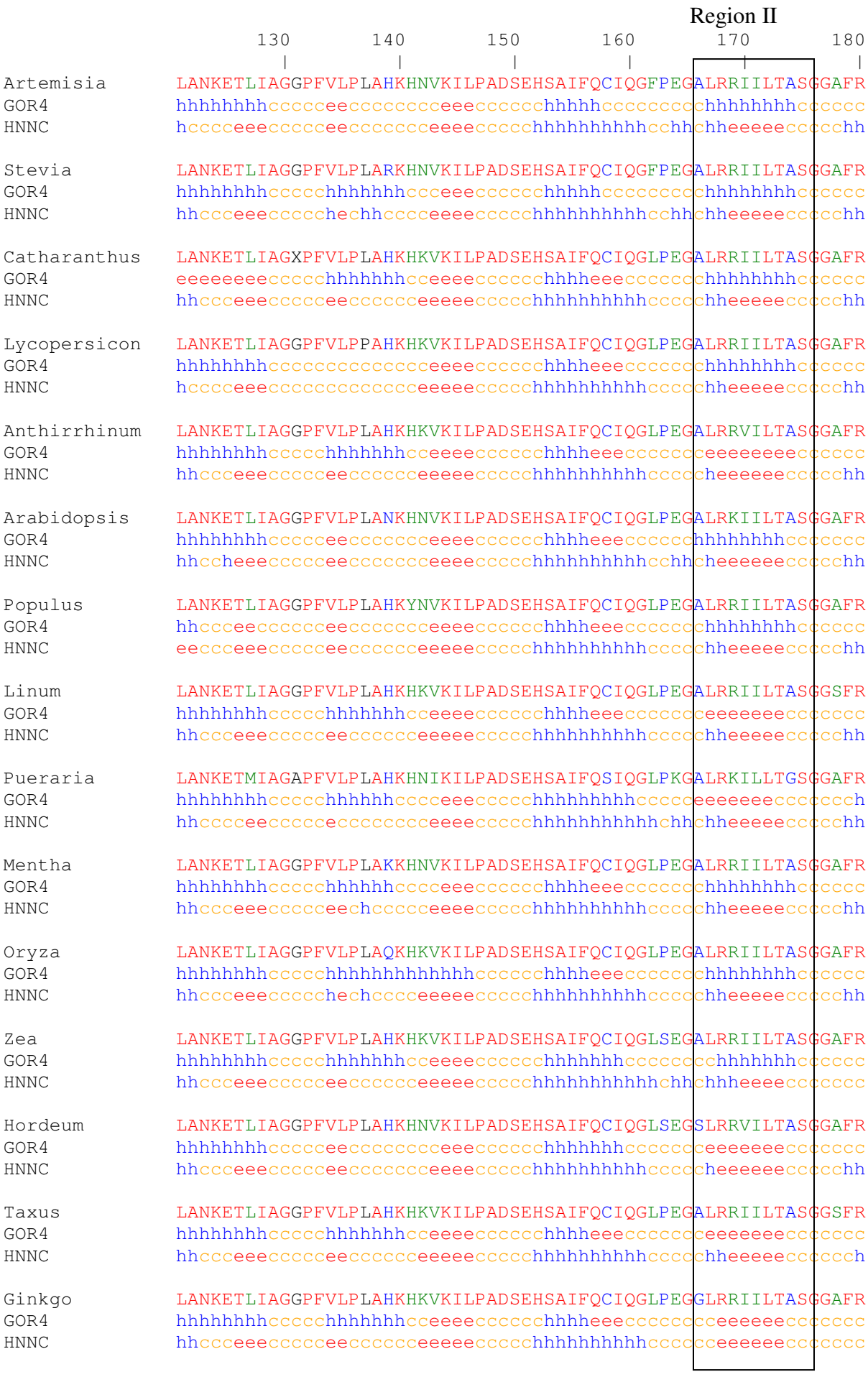
Appendix II

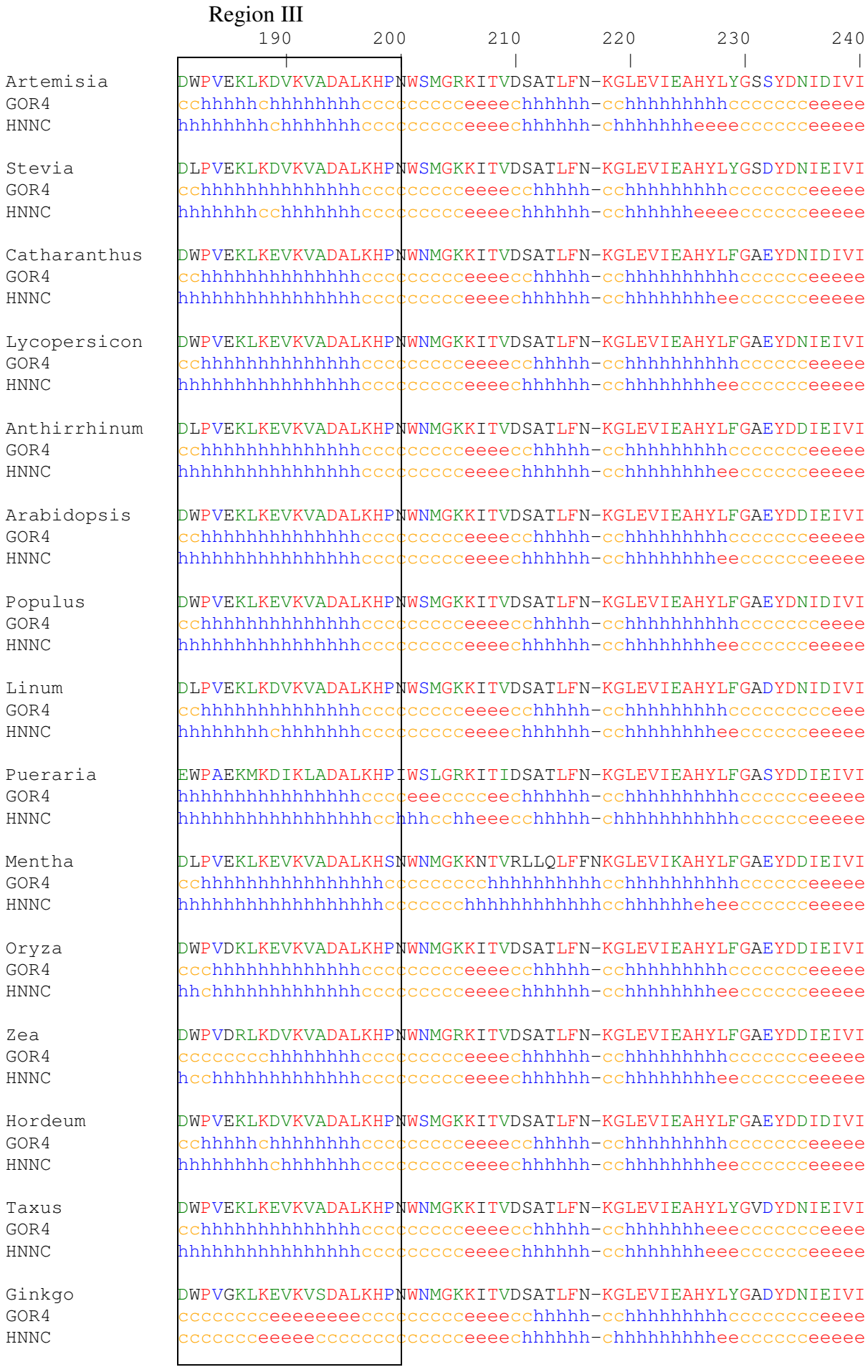
Secondary structure prediction of the functional domain in DXR (Boxed regions indicates where differences occurs)

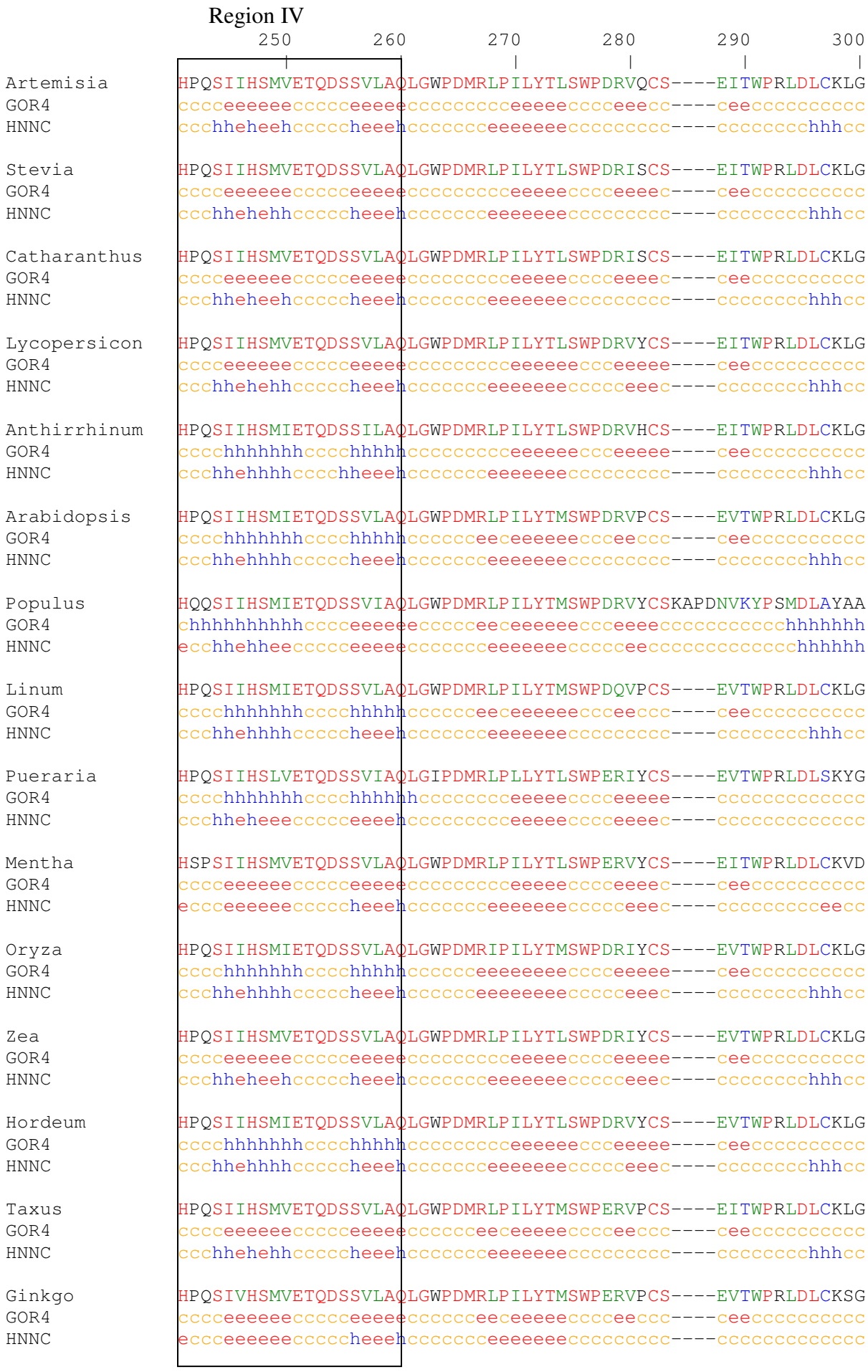
```
GOR IV and HNN:
   Alpha helix      (h)
   Extended strand (e)
   Random coil      (c)
```

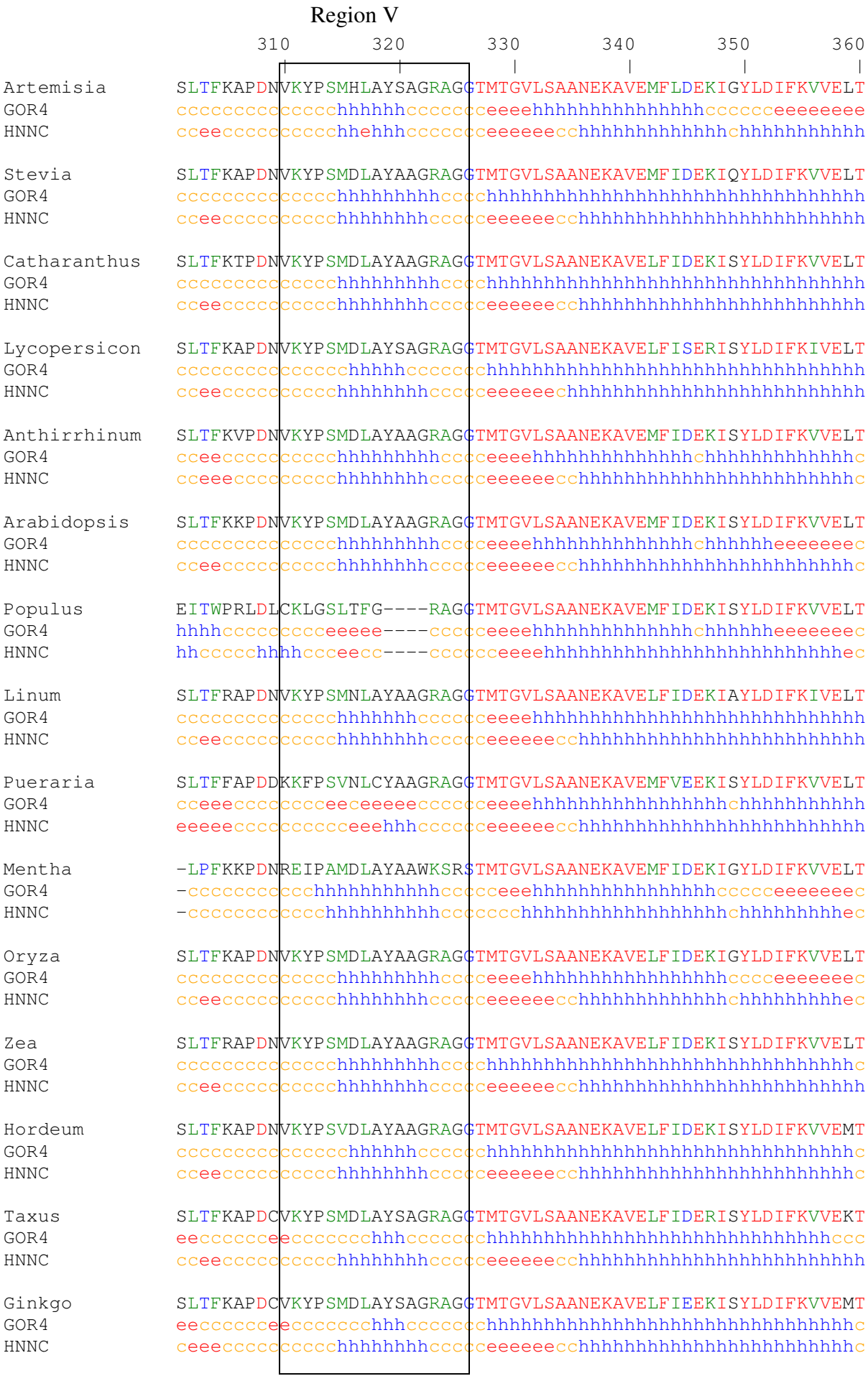

Region I
10
20
30
40
50
60
Artemisia
KPISIIGSTGSIGTQTLDIVAENPDKFKVVALAAGSNVTLLAEQIKAFKPQLVSIKNESL
GOR4
cceeeeecccccccchhhhhhccccchhhhhhhhccccchhhhhhhhhhhhccchhhhchhhh
HNNC
cceeeeecccccccceeeeeeccccceeeeeeeccccchhhhhhhhhhhhccchheeecchhh
Stevia
KPISIVGSTGSIGTQTLDIVAENPDKFRVVALAAGSNVTLLAEQIKAFKPQLVSIQNESL
GOR4
cceeeeecccccccchhhhhhccccchhhhhhhhccccchhhhhhhhhhhhcccceeechhhh
HNNC
cceeeeecccccccceeeeeeccccceeeeeeeccccchhhhhhhhhhhhcccceeecchhh
Catharanthus
KPISIVGSTGSVGTQTLDIVAENPDKFRVVALAAGSNVTLLADQVKTFKPQLVSVRNESL
GOR4
cceeeeecccccccceeeeeeccccchhhhhhhhccccchhhhhhhhhhccccceeecchhhh
HNNC
cceeeeecccccccceeeeeeccccceeeeeeeccccchhehhhhhhhccccceeeeecchh
Lycopersicon
KPISIVGSTGSIGTQTLDIVAENPDKFRVVALAAGSNVTLLADQVKTFRPKLVAVRNESL
GOR4
cceeeeecccccccchhhhhhccccchhhhhhhhccccchhhhhhhhhcccchhhhhhhhhhh
HNNC
cceeeeecccccccceeeeeeccccceeeeeeeccccchhehhhhhhhccccceeeeeccch
Anthirrhinum
KPISIVGSTGSIGTQTLDIVAENPDKFRVVALAAGSNVTLLADQIRTFKPQLVSVRDESL
GOR4
cceeeeecccccccchhhhhhccccchhhhhhhhccccchhhhhhhhhccccccceecchhhh
HNNC
cceeeeecccccccceeeeeeccccceeeeeeeccccchhhhhhhhhhhhccceeeecchhh
Arabidopsis
KPISIVGSTGSIGTQTLDIVAENPDKFRVVALAAGSNVTLLADQVRRFKPALVAVRNESL
GOR4
cceeeeecccccccchhhhhhccccchhhhhhhhccccchhhhhhhhhhhhchhhhhhhhhhh
HNNC
cceeeeecccccccceeeeeeccccceeeeeeeccccchhhhhhhhhhhhccceeeeechhh
Populus
KPISIVGSTGSIGTQTLDIVAENPDKFKVVALAAGSNVALLADQVRTFKPQLIAVRNELL
GOR4
cceeeeecccccccchhhhhhccccchhhhhhhhhcchhhhhhhhhhhhhchhhhhhhhhhh
HNNC
cceeeeecccccccceeeeeeccccceeeeeeeccccchhhhhhhhhhhcccchhhhhhhhhh
Linum
KPISIVGSTGSIGTQTLDIVSENPDKFKVVALAAGSNIALLADQIRTFKPQLVSVKNESL
GOR4
cceeeeecccccccceeeeeeccccchhhhhhhhhcchhhhhhhhhhcccccchhhhhhhhh
HNNC
cceeeeecccccccceeeeeccccccceeeeeeccccchhhhhhhhhhhhcccceeeechhhh
Pueraria
KPISILGSTGSIGTQTLSIVAEFPERFKVVSLAAGSNITLLADQIKTFKPEVVGLRNESL
GOR4
cceeeeecccccccchhhhhhccccchhhhhhhccccchhhhhhhhcccccceeeeecchhh
HNNC
cceeeeecccccccceeeeehcccchheeeeeeccccchhhhhhhhhhcccceeecccchhh
Mentha
KPISVIGSTGSIGTQTLDIVAENPDKFRIVALAAGSNVTLLADQK-AFKPKLVSVKDESL
GOR4
cceeeeecccccccchhhhhhccccchhhhhhhhccccchhhhhhhh-hccccceeecchhh
HNNC
cceeeeecccccccceeeeeeccccceeeeeeeccccheeehhhh-hccccceeeecchhh
Oryza
KPISIVGSTGSIGTQTLDIVAENPDKFRVVALAAGSNVTLLADQVKTFKPKHVAVRNESL
GOR4
cceeeeecccccccchhhhhhccccchhhhhhhhccccchhhhhhhhhcccccceeehhhhh
HNNC
cceeeeecccccccceeeeeeccccceeeeeeeccccchhehhhhhhhccccceeeeeccch
Zea
KPISIVGSTGSIGTQTLDIVAENPDKFRVVALAAGSNVTLLADQVKTFKPKLVAVRNESL
GOR4
cceeeeecccccccchhhhhhccccchhhhhhhhccccchhhhhhhhhccccceeeehhhhh
HNNC
cceeeeecccccccceeeeeeccccceeeeeeeccccchhehhhhhhhccccceeeeeccch
Hordeum
KPISIVGSTGSIGTQTLDIVAENPDKFRVVALAAGSNVTLLADQVKTFKPKLVALRDESL
GOR4
cceeeeecccccccchhhhhhccccchhhhhhhhccccchhhhhhhhhhhchhhhhhhhhhh
HNNC
cceeeeecccccccceeeeeeccccceeeeeeeccccchhehhhhhhhccccheeechhhh
Taxus
KPFSIIGSTGSIGTQTLDIVAEHPDKFKVVALAAGSNVTLLAEQVRMFKPKLVSVRNESL
GOR4
cceeeeecccccccceeeeeeccccchhhhhhhhccccchhhhhhhhhhccccceeeechhhh
HNNC
cceeeeecccccchceeheehcccccceeeeeeeccccchhhhhhhhhhhhcccceeeecccch
Ginkgo
KPFSIVGSTGSIGTQTLDIVAEHPDKFRVVALAAGSNVALLADQVRQFKPKLVAIRNESL
GOR4
cceeeeecccccccceeeeeeccccchhhhhhhhhcchhhhhhhhhhhhchhhhhhhhhhh
HNNC
cceeeeecccccchceeheeeccccceeeeeeeccccchhhhhhhhhhhhcccceeeeeccch

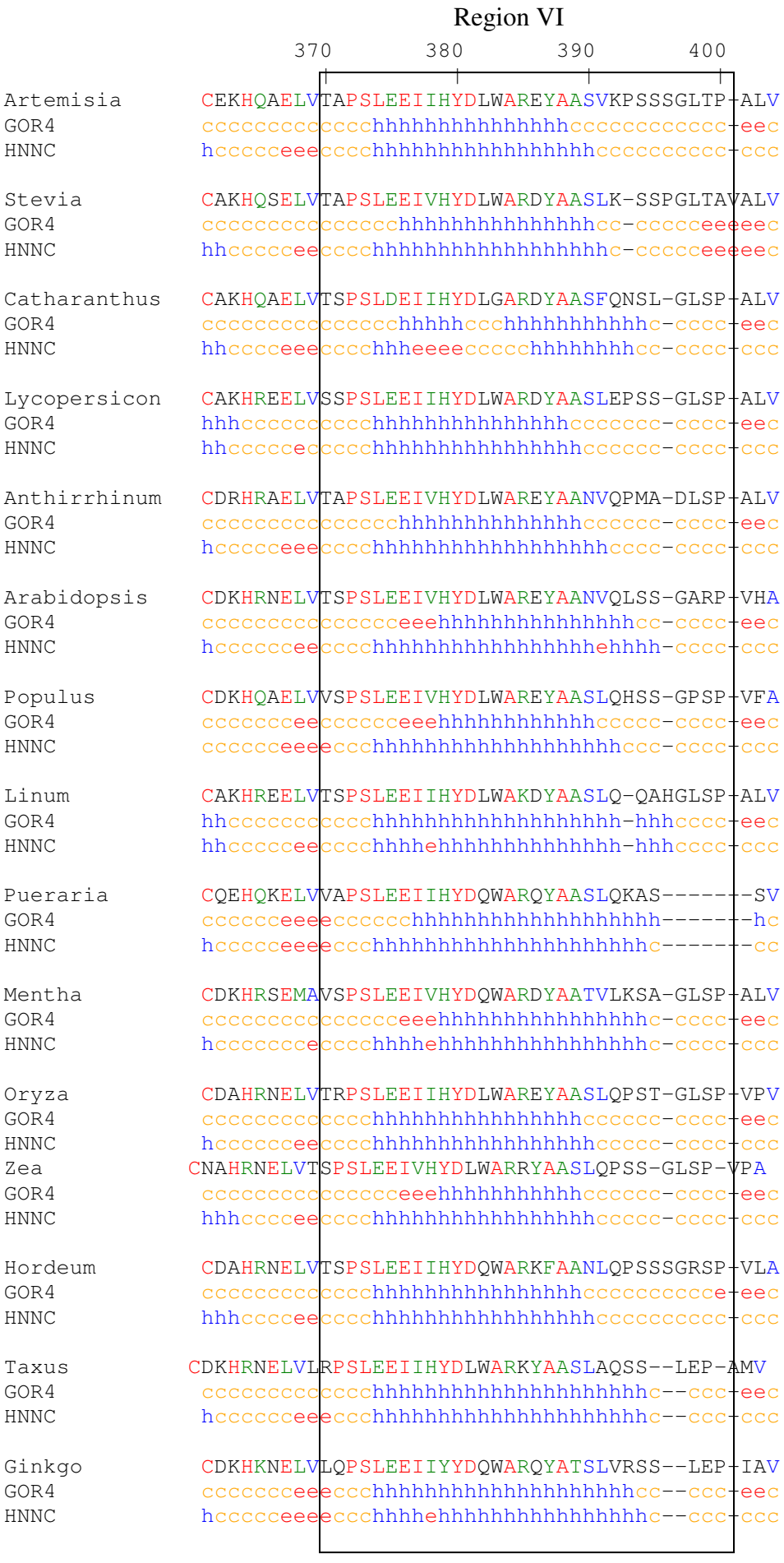

70 80 90 100 110 120
Artemisia VAELKEALAGSDYMPEIIPGDEGVVEVARHPDCVTVVTGIVGCAGLKPTVAAIEAGKNIA
GOR4 hhhhhhhhccccccceeccccceeeeeccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhccccceeeccccceeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Stevia VGELKEALADADYMPEIIPGDQGIIEVARHPDCVTVVTGIVGCAGLKPTVAAIEAGKNIA
GOR4 hhhhhhhhhhhccccceeccccceeeeeccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhhhcccccecccccceeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Catharanthus VNELKEALSDVDDKPEIIPGEQGVVEVVRHSDAVTVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhccccccceeccccceeeeeeccccceeeeeeeeccccccceeeeeeeeeee
HNNC hhhhhhhhhccccccceeccccceeeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Lycopersicon VEELKDALADMEDKPEIIPGEQGVIEVARHPDAVTVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhhhhccccccccccceeeeeccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhhhhcccceccccceeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Anthirrhinum INELKEALFDVEDKPEIIPGEQGIIEVARHPDAVTVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhhccccccccccccceeeeeccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhhhcccccecccccceeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Arabidopsis INELKEALADLDYKLEIIPGEQGVIEVARHPEAVTVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhhhhhhhhhhcccccceeeecccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhhccceeeeecccccceeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Populus VDELKEALADVEEKPEIIPGEQGVVEVARHPDAVSVVTGIVGCAGLKPTVAAIEAGKDIC
GOR4 hhhhhhhhhhhcccccccccccceeeeeccccceeeeeeeeccccccchhhhhhchhhh
HNNC hhhhhhhhhhhhcccceecccceeeeeeccccceeeeeeeeccccccchhhhhhhccccee
Linum AKELKEALAGLEVMPEIIPGEEGIVEVARHPDAATVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhhhhhcccccccccccceeeeeccccccceeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhcchccccecccccceeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Pueraria IDELKEALADVEHKPEIIPGEQGVIEAARHPDSTTVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhhhhcccccccccchhhhhhccccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhhhcccccccccccceeeeccccccceeeeeeeeccccccchhhhhhhchhhh
Mentha ISELKEALAGFEDMPEIIPGEQGMIEVARHPDAVTVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhhhcccccceeccccceeeeeccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhhhhhcccccccccccceeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Oryza VDELKEALADCDWKPEIIPGEQGVIEVARHPDAVTVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhhhcccccccccccceeeeeccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhhcccccecccccceeeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Zea VDELKEALADCEEKPEIIPGEQGVIEVARHPDAVTVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhhhhcccccccccccceeeeeccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhhhhhcccceccccceeeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Hordeum LNELKEALAGCEEMPEIIPGEQGVIEVARHPDAVTVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhhhcccccccccccceeeeeccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhcchcccccccccceeeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Taxus ATELKEALADIEHKPEIVYGDEGMVEVAQHPEAVSVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhhhhcccccceeecccccceeeeeccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhhhcccceeeecccceeeeeccccceeeeeeeeccccccchhhhhhhchhhh
Ginkgo ITELKAALSDFEPKPEIISGEEGIVEVARHPEAVSVVTGIVGCAGLKPTVAAIEAGKDIA
GOR4 hhhhhhhhccccccccceecccccchhhhcccccceeeeeeeeccccccchhhhhhhhhhh
HNNC hhhhhhhhhcccccceeeeccccceeeeeccccceeeeeeeeccccccchhhhhhhchhhh


Region II
130
140
150
160
170
180
Artemisia
LANKETLIAGGPFVLPLAHKHNVKILPADSEHSAIFQCIQGFPEGALRRIILTASGGAFR
GOR4
hhhhhhhhccccceecccccccceeeccccccchhhhhccccccccchhhhhhhhccccccc
HNNC
hcccceeeccccceeccccccceeeeccccchhhhhhhhhhhcchhchheeeeeccccchh
Stevia
LANKETLIAGGPFVLPLARKHNVKILPADSEHSAIFQCIQGFPEGALRRIILTASGGAFR
GOR4
hhhhhhhhccccchhhhhhhccceeeccccccchhhhhccccccccchhhhhhhhccccccc
HNNC
hhccceeeccccchechhccccceeeeccccchhhhhhhhhhhcchhchheeeeeccccchh
Catharanthus
LANKETLIAGXPFVLPLAHKHKVKILPADSEHSAIFQCIQGLPEGALRRIILTASGGAFR
GOR4
eeeeeeeeccccchhhhhhhhccceeeccccccchhhheeeccccccchhhhhhhhccccccc
HNNC
hhccceeeccccceeccccccceeeeeccccchhhhhhhhhhhccccchheeeeeccccchh
Lycopersicon
LANKETLIAGGPFVLPPAHKHKVKILPADSEHSAIFQCIQGLPEGALRRIILTASGGAFR
GOR4
hhhhhhhhccccccccccccccceeeecccccchhhheeeccccccchhhhhhhhccccccc
HNNC
hcccceeeccccccccccccceeeeeccccchhhhhhhhhhhccccchheeeeeccccchh
Anthirrhinum
LANKETLIAGGPFVLPLAHKHKVKILPADSEHSAIFQCIQGLPEGALRRVILTASGGAFR
GOR4
hhhhhhhhccccchhhhhhhhccceeeccccccchhhheeeccccccceeeeeeeeccccccc
HNNC
hhccceeeccccceeccccccceeeeeccccchhhhhhhhhhhccccchheeeeeccccchh
Arabidopsis
LANKETLIAGGPFVLPLANKHNVKILPADSEHSAIFQCIQGLPEGALRKIILTASGGAFR
GOR4
hhhhhhhhccccceeccccccceeeeccccccchhhheeeccccccchhhhhhhhccccccc
HNNC
hhcchheeeccccceeccccccceeeeccccchhhhhhhhhhhcchhchheeeeeccccchh
Populus
LANKETLIAGGPFVLPLAHKYNVKILPADSEHSAIFQCIQGLPEGALRRIILTASGGAFR
GOR4
hhccceecccccceecccccccceeeeccccccchhhheeeccccccchhhhhhhhccccccc
HNNC
eeccceeeccccceecccccceeeeeccccchhhhhhhhhhhccccchheeeeeccccchh
Linum
LANKETLIAGGPFVLPLAHKHKVKILPADSEHSAIFQCIQGLPEGALRRIILTASGGSFR
GOR4
hhhhhhhhccccchhhhhhhhccceeeccccccchhhheeeccccccceeeeeeeccccccc
HNNC
hhccceeeccccceeccccccceeeeeccccchhhhhhhhhhhccccchheeeeeccccchh
Pueraria
LANKETMIAGAPFVLPLAHKHNIKILPADSEHSAIFQSIQGLPKGALRKILLTGSGGAFR
GOR4
hhhhhhhhccccchhhhhhhccccceeeccccchhhhhhhhhccccceeeeeeeccccccch
HNNC
hhccccecccccceccccccccceeeeccccchhhhhhhhhhhhchhchheeeeeccccchh
Mentha
LANKETLIAGGPFVLPLAKKHNVKILPADSEHSAIFQCIQGLPEGALRRIILTASGGAFR
GOR4
hhhhhhhhccccchhhhhhhccccceeeccccccchhhheeeccccccchhhhhhhhccccccc
HNNC
hhccceeeccccceechccccceeeeccccchhhhhhhhhhhccccchheeeeeccccchh
Oryza
LANKETLIAGGPFVLPLAQKHKVKILPADSEHSAIFQCIQGLPEGALRRIILTASGGAFR
GOR4
hhhhhhhhccccchhhhhhhhhhhhhhccccccchhhheeeccccccchhhhhhhhccccccc
HNNC
hhcccecccccchechcccceeeeeccccchhhhhhhhhhhccccchheeeeeccccchh
Zea
LANKETLIAGGPFVLPLAHKHKVKILPADSEHSAIFQCIQGLSEGALRRIILTASGGAFR
GOR4
hhhhhhhhccccchhhhhhhhccceeeccccccchhhhhhhccccccccchhhhhhhccccccc
HNNC
hhcccecccccceecccccceeeeeccccchhhhhhhhhhhhchhchheeeeeccccccc
Hordeum
LANKETLIAGGPFVLPLAHKHNVKILPADSEHSAIFQCIQGLSEGSLRRVILTASGGAFR
GOR4
hhhhhhhhccccceecccccccceeeccccccchhhhhhhccccccceeeeeeeccccccc
HNNC
hhcccecccccceecccccccceeeeccccchhhhhhhhhhhccccchheeeeeccccchh
Taxus
LANKETLIAGGPFVLPLAHKHKVKILPADSEHSAIFQCIQGLPEGALRRIILTASGGSFR
GOR4
hhhhhhhhccccchhhhhhhhccceeeccccccchhhheeeccccccceeeeeeeccccccc
HNNC
hhcccecccccceecccccceeeeeccccchhhhhhhhhhhccccchheeeeeccccccch
Ginkgo
LANKETLIAGGPFVLPLAHKHKVKILPADSEHSAIFQCIQGLPEGGLRRIILTASGGAFR
GOR4
hhhhhhhhccccchhhhhhhhccceeeccccccchhhheeeccccccccceeeeeeccccccc
HNNC
hhcccecccccceecccccceeeeeccccchhhhhhhhhhhccccccceeeeeeccccccc


Region III
190
200
210
220
230
240
Artemisia
DWPVEKLKDVKVADALKHPNWSMGRKITVDSATLFN-KGLEVIEAHYLYGSSYDNIDIVI
GOR4
ccHHHHHcHHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHHHccccccceeeee
HNNC
HHHHHHHHcHHHHHHHcccccccccceeeeeHHHHHH-cHHHHHHHeeeeccccccceeeee
Stevia
DLPVEKLKDVKVADALKHPNWSMGKKITVDSATLFN-KGLEVIEAHYLYGSDYDNIEIVI
GOR4
ccHHHHHHHHHHHHHHcccccccccceeeeeccHHHHH-ccHHHHHHHHHccccccceeeee
HNNC
HHHHHHHccHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHeeeeccccccceeeee
Catharanthus
DWPVEKLKEVKVADALKHPNWNMGKKITVDSATLFN-KGLEVIEAHYLFGAEYDNIDIVI
GOR4
ccHHHHHHHHHHHHHHcccccccccceeeeeccHHHHH-ccHHHHHHHHHHccccccceeeee
HNNC
HHHHHHHHHHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHHeeccccccceeeee
Lycopersicon
DWPVEKLKEVKVADALKHPNWNMGKKITVDSATLFN-KGLEVIEAHYLFGAEYDNIEIVI
GOR4
ccHHHHHHHHHHHHHHcccccccccceeeeeccHHHHH-ccHHHHHHHHHHccccccceeeee
HNNC
HHHHHHHHHHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHHeeccccccceeeee
Anthirrhinum
DLPVEKLKEVKVADALKHPNWNMGKKITVDSATLFN-KGLEVIEAHYLFGAEYDDIEIVI
GOR4
ccHHHHHHHHHHHHHHcccccccccceeeeeccHHHHH-ccHHHHHHHHHccccccceeeee
HNNC
HHHHHHHHHHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHHeeccccccceeeee
Arabidopsis
DWPVEKLKEVKVADALKHPNWNMGKKITVDSATLFN-KGLEVIEAHYLFGAEYDDIEIVI
GOR4
ccHHHHHHHHHHHHHHcccccccccceeeeeccHHHHH-ccHHHHHHHHHccccccceeeee
HNNC
HHHHHHHHHHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHHeeccccccceeeee
Populus
DWPVEKLKEVKVADALKHPNWSMGKKITVDSATLFN-KGLEVIEAHYLFGAEYDNIDIVI
GOR4
ccHHHHHHHHHHHHHHcccccccccceeeeeccHHHHH-ccHHHHHHHHHHccccccceeeee
HNNC
HHHHHHHHHHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHHeeccccccceeeee
Linum
DLPVEKLKDVKVADALKHPNWSMGKKITVDSATLFN-KGLEVIEAHYLFGADYDNIDIVI
GOR4
ccHHHHHHHHHHHHHHcccccccccceeeeeccHHHHH-ccHHHHHHHHHccccccccceee
HNNC
HHHHHHHHcHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHHeeccccccceeeee
Pueraria
EWPAEKMKDIKLADALKHPIWSLGRKITIDSATLFN-KGLEVIEAHYLFGASYDDIEIVI
GOR4
HHHHHHHHHHHHHHHHcccceeecccceeeHHHHHH-ccHHHHHHHHHHccccccceeeee
HNNC
HHHHHHHHHHHHHHHHHHccHHHccHHeeeccHHHHH-cHHHHHHHHHHHccccccceeeee
Mentha
DLPVEKLKEVKVADALKHSNWNMGKKNTVRLLQLFFNKGLEVIKAHYLFGAEYDDIEIVI
GOR4
ccHHHHHHHHHHHHHHHHccccccccchHHHHHHHHHHccHHHHHHHHHHccccccceeeee
HNNC
HHHHHHHHHHHHHHHHHHHccccccchHHHHHHHHHHHHccHHHHHHeheeccccccceeeee
Oryza
DWPVDKLKEVKVADALKHPNWNMGKKITVDSATLFN-KGLEVIEAHYLFGAEYDDIEIVI
GOR4
cccHHHHHHHHHHHHHcccccccccceeeeeccHHHHH-ccHHHHHHHHHccccccceeeee
HNNC
hhcHHHHHHHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHHeeccccccceeeee
Zea
DWPVDRLKDVKVADALKHPNWNMGRKITVDSATLFN-KGLEVIEAHYLFGAEYDDIEIVI
GOR4
ccccccccHHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHHHccccccceeeee
HNNC
hccHHHHHHHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHHeeccccccceeeee
Hordeum
DWPVEKLKDVKVADALKHPNWSMGKKITVDSATLFN-KGLEVIEAHYLFGAEYDDIDIVI
GOR4
ccHHHHHcHHHHHHHHcccccccccceeeeeccHHHHH-ccHHHHHHHHHccccccceeeee
HNNC
HHHHHHHHcHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHHeeccccccceeeee
Taxus
DWPVEKLKEVKVADALKHPNWNMGKKITVDSATLFN-KGLEVIEAHYLYGVDYDNIEIVI
GOR4
ccHHHHHHHHHHHHHHcccccccccceeeeeccHHHHH-ccHHHHHHHeeeccccccceeee
HNNC
HHHHHHHHHHHHHHHHcccccccccceeeeeHHHHHH-ccHHHHHHHeeeccccccceeeee
Ginkgo
DWPVGKLKEVKVSDALKHPNWNMGKKITVDSATLFN-KGLEVIEAHYLYGADYDNIEIVI
GOR4
cccccccceeeeeeeeccccccccccceeeeeccHHHHH-ccHHHHHHHHHccccccceeeee
HNNC
cccccccceeeeeccccccccccccccceeeeeHHHHHH-cHHHHHHHHHeeccccccceeeee


Region IV
250
260
270
280
290
300
Artemisia
HPQSIIHSMVETQDSSVLAQLGWPDMRLPILYTLSWPDRVQCS----EITWPRLDLCKLG
GOR4
cccceeeeeeccccceeeeecccccccceeeeecccceeecc----ceecccccccccc
HNNC
ccchheheehccccchheehcccccceeeeeeecccccccc----cccccccchhhcc
Stevia
HPQSIIHSMVETQDSSVLAQLGWPDMRLPILYTLSWPDRISCS----EITWPRLDLCKLG
GOR4
cccceeeeeeccccceeeeecccccccceeeeecccceeeec----ceecccccccccc
HNNC
ccchheheehhccccchheehcccccceeeeeeecccccccc----cccccccchhhcc
Catharanthus
HPQSIIHSMVETQDSSVLAQLGWPDMRLPILYTLSWPDRISCS----EITWPRLDLCKLG
GOR4
cccceeeeeeccccceeeeecccccccceeeeecccceeeec----ceecccccccccc
HNNC
ccchheheehccccchheehcccccceeeeeeecccccccc----cccccccchhhcc
Lycopersicon
HPQSIIHSMVETQDSSVLAQLGWPDMRLPILYTLSWPDRVYCS----EITWPRLDLCKLG
GOR4
cccceeeeeeccccceeeeecccccccceeeeeeccceeeee----ceecccccccccc
HNNC
ccchheheehccccchheehcccccceeeeeeeccccceeec----cccccccchhhcc
Anthirrhinum
HPQSIIHSMIETQDSSILAQLGWPDMRLPILYTLSWPDRVHCS----EITWPRLDLCKLG
GOR4
ccccchhhhhhcccchhhhhcccccccceeeeeeccceeeee----ceecccccccccc
HNNC
ccchheheehhccccchheehcccccceeeeeeecccccccc----cccccccchhhcc
Arabidopsis
HPQSIIHSMIETQDSSVLAQLGWPDMRLPILYTMSWPDRVPCS----EVTWPRLDLCKLG
GOR4
ccccchhhhhhcccchhhhhcccccceeceeeeeeccceeccc----ceecccccccccc
HNNC
ccchheheehhccccchheehcccccceeeeeeecccccccc----cccccccchhhcc
Populus
HQQSIIHSMIETQDSSVIAQLGWPDMRLPILYTMSWPDRVYCSKAPDNVKYPSMDLAYAA
GOR4
chhhhhhhhhhcccceeeeeecccccceeceeeeeecccceeeecccccccccchhhhhhh
HNNC
ecchheheheecccccceeeeecccccccceeeeeeecccccceecccccccccccchhhhhh
Linum
HPQSIIHSMIETQDSSVLAQLGWPDMRLPILYTMSWPDQVPCS----EVTWPRLDLCKLG
GOR4
ccccchhhhhhcccchhhhhccccccceecceeeeeecccceeccc----ceecccccccccc
HNNC
ccchheheehhccccchheehcccccceeeeeeecccccccc----cccccccchhhcc
Pueraria
HPQSIIHSLVETQDSSVIAQLGIPDMRLPLLYTLSWPERIYCS----EVTWPRLDLSKYG
GOR4
ccccchhhhhhcccchhhhhhcccccccceeeeecccceeeee----ccccccccccccc
HNNC
ccchheheeecccccceeeehcccccccceeeeecccceeeec----ccccccccccccc
Mentha
HSPSIIHSMVETQDSSVLAQLGWPDMRLPILYTLSWPERVYCS----EITWPRLDLCKVD
GOR4
cccceeeeeecccccceeeeecccccccceeeeecccceeeec----ceecccccccccc
HNNC
ecccceeeeeecccccheeehcccccceeeeeeecccccceec----ccccccccceecc
Oryza
HPQSIIHSMIETQDSSVLAQLGWPDMRIPILYTMSWPDRIYCS----EVTWPRLDLCKLG
GOR4
ccccchhhhhhcccchhhhhccccccceeeeeeecccceeeee----ceecccccccccc
HNNC
ccchheheehhccccchheehcccccceeeeeeecccccceec----cccccccchhhcc
Zea
HPQSIIHSMVETQDSSVLAQLGWPDMRLPILYTLSWPDRIYCS----EVTWPRLDLCKLG
GOR4
cccceeeeeecccccceeeeecccccccceeeeecccceeeee----ceecccccccccc
HNNC
ccchheheehccccchheehcccccceeeeeeecccccceec----cccccccchhhcc
Hordeum
HPQSIIHSMIETQDSSVLAQLGWPDMRLPILYTLSWPDRVYCS----EVTWPRLDLCKLG
GOR4
ccccchhhhhhcccchhhhhcccccccceeeeeecccceeeee----ceecccccccccc
HNNC
ccchheheehhccccchheehcccccceeeeeeecccccceec----cccccccchhhcc
Taxus
HPQSIIHSMVETQDSSVLAQLGWPDMRLPILYTMSWPERVPCS----EITWPRLDLCKLG
GOR4
cccceeeeeecccccceeeeecccccceeceeeeecccceeccc----ceecccccccccc
HNNC
ccchheheehhccccchheehcccccceeeeeeecccccccc----cccccccchhhcc
Ginkgo
HPQSIVHSMVETQDSSVLAQLGWPDMRLPILYTMSWPERVPCS----EVTWPRLDLCKSG
GOR4
cccceeeeeecccccceeeeecccccceeceeeeecccceeccc----ceecccccccccc
HNNC
ecccceeeeeecccccheeehcccccceeeeeeecccccccc----ccccccccccccc


Region V
310
320
330
340
350
360
Artemisia
SLTFKAPDNVKYPSMHLAYSAGRAGGTMTGVLSAANEKAVEMFLDEKIGYLDIFKVVELT
GOR4
cccccccccccccchhhhhhccccccceeeehhhhhhhhhhhhhhhcccccceeeeeeee
HNNC
cceeccccccccchhehhhccccccceeeeeecchhhhhhhhhhhhhchhhhhhhhhhh
Stevia
SLTFKAPDNVKYPSMDLAYAAGRAGGTMTGVLSAANEKAVEMFIDEKIQYLDIFKVVELT
GOR4
cccccccccccccchhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
HNNC
cceeccccccccchhhhhhhhhccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhhh
Catharanthus
SLTFKTPDNVKYPSMDLAYAAGRAGGTMTGVLSAANEKAVELFIDEKISYLDIFKVVELT
GOR4
cccccccccccccchhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
HNNC
cceeccccccccchhhhhhhhhccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhhh
Lycopersicon
SLTFKAPDNVKYPSMDLAYSAGRAGGTMTGVLSAANEKAVELFISERISYLDIFKIVELT
GOR4
ccccccccccccccchhhhhccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
HNNC
cceeccccccccchhhhhhhhhccccceeeeeechhhhhhhhhhhhhhhhhhhhhhhhhh
Anthirrhinum
SLTFKVPDNVKYPSMDLAYAAGRAGGTMTGVLSAANEKAVEMFIDEKISYLDIFKVVELT
GOR4
cceeccccccccchhhhhhhhhccccceeeehhhhhhhhhhhhhhchhhhhhhhhhhhhc
HNNC
cceeeccccccccchhhhhhhhhccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhhc
Arabidopsis
SLTFKKPDNVKYPSMDLAYAAGRAGGTMTGVLSAANEKAVEMFIDEKISYLDIFKVVELT
GOR4
ccccccccccccchhhhhhhhhccccceeeehhhhhhhhhhhhhhchhhhhheeeeeeec
HNNC
cceeccccccccchhhhhhhhhccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhhc
Populus
EITWPRLDLCKLGSLTFG----RAGGTMTGVLSAANEKAVEMFIDEKISYLDIFKVVELT
GOR4
hhhhccccccccceeeee----ccccceeeehhhhhhhhhhhhhhchhhhhheeeeeeec
HNNC
hhccccchhhhccceecc----cccccceeeehhhhhhhhhhhhhhhhhhhhhhhhhec
Linum
SLTFRAPDNVKYPSMNLAYAAGRAGGTMTGVLSAANEKAVELFIDEKIAYLDIFKIVELT
GOR4
ccccccccccccchhhhhhhccccccceeeehhhhhhhhhhhhhhhhhhhhhhhhhhhh
HNNC
cceeccccccccchhhhhhhhhccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhhh
Pueraria
SLTFFAPDDKKFPSVNLCYAAGRAGGTMTGVLSAANEKAVEMFVEEKISYLDIFKVVELT
GOR4
cceeeccccccccceceeeeeccccccceeeehhhhhhhhhhhhhhhhhchhhhhhhhhh
HNNC
eeeeeccccccccccceeehhhccccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhh
Mentha
-LPFKKPDNREIPAMDLAYAAWKSRSTMTGVLSAANEKAVEMFIDEKIGYLDIFKVVELT
GOR4
-ccccccccccchhhhhhhhhhhccccceeehhhhhhhhhhhhhhhhccccceeeeeeec
HNNC
-ccccccccccchhhhhhhhhhccccccchhhhhhhhhhhhhhhhhhhchhhhhhhhec
Oryza
SLTFKAPDNVKYPSMDLAYAAGRAGGTMTGVLSAANEKAVELFIDEKIGYLDIFKVVELT
GOR4
ccccccccccccchhhhhhhhhccccceeeehhhhhhhhhhhhhhhhhhcccceeeeeeec
HNNC
cceeccccccccchhhhhhhhhccccceeeeeecchhhhhhhhhhhhhhchhhhhhhhhec
Zea
SLTFRAPDNVKYPSMDLAYAAGRAGGTMTGVLSAANEKAVELFIDEKISYLDIFKVVELT
GOR4
ccccccccccccchhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhc
HNNC
cceeccccccccchhhhhhhhhccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhhh
Hordeum
SLTFKAPDNVKYPSVDLAYAAGRAGGTMTGVLSAANEKAVELFIDEKISYLDIFKVVEMT
GOR4
ccccccccccccccchhhhhhccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhc
HNNC
cceeccccccccchhhhhhhhhccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhhhc
Taxus
SLTFKAPDCVKYPSMDLAYSAGRAGGTMTGVLSAANEKAVELFIDERISYLDIFKVVEKT
GOR4
eecccccceeccccccchhhccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccc
HNNC
ccceeccccccccchhhhhhhhhccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhh
Ginkgo
SLTFKAPDCVKYPSMDLAYSAGRAGGTMTGVLSAANEKAVELFIEEKISYLDIFKVVEMT
GOR4
eecccccceeccccccchhhccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhc
HNNC
ceeeccccccccchhhhhhhhhccccceeeeeecchhhhhhhhhhhhhhhhhhhhhhhhc

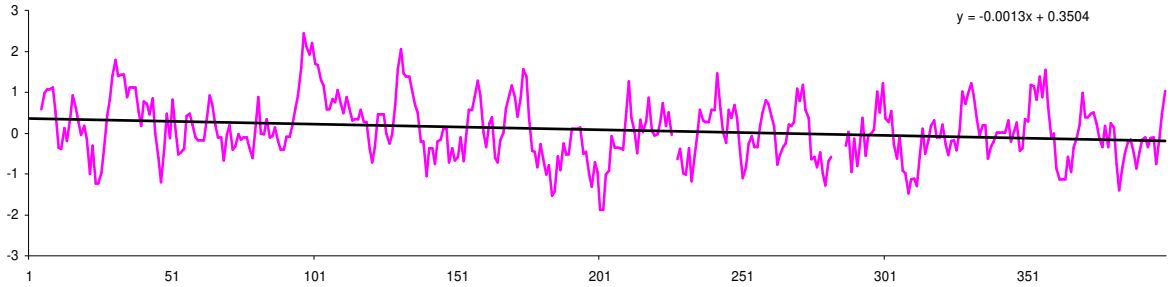

Region VI
370
380
390
400
Artemisia CEKHQAELVTAPSLEEIIHYDLWAREYAASVKPSSSGLTP-ALV
GOR4 ccccccccccccchhhhhhhhhhhhhhhhccccccccccccc-eec
HNNC hccccceeeccccchhhhhhhhhhhhhhhhhhhccccccccccc-ccc
Stevia CAKHQSELVTAPSLEEIVHYDLWARDYAASLK-SSPGLTAVALV
GOR4 ccccccccccccccchhhhhhhhhhhhhhhhcc-ccccceeeeec
HNNC hhccccceeccccchhhhhhhhhhhhhhhhhhhc-ccccceeeeec
Catharanthus CAKHQAELVTSPSLDEIIHYDLGARDYAASFQNSL-GLSP-ALV
GOR4 ccccccccccccccchhhhhccchhhhhhhhhhhc-cccc-eec
HNNC hhccccceeccccchheeeeeccccchhhhhhhhcc-cccc-ccc
Lycopersicon CAKHREELVSSPSLEEIIHYDLWARDYAASLEPSS-GLSP-ALV
GOR4 hhhcccccccccchhhhhhhhhhhhhhhhcccccccc-cccc-eec
HNNC hhccccccccccchhhhhhhhhhhhhhhhhhcccccc-cccc-ccc
Anthirrhinum CDRHRAELVTAPSLEEIVHYDLWAREYAANVQPMA-DLSP-ALV
GOR4 ccccccccccccccchhhhhhhhhhhhhhhcccccc-cccc-eec
HNNC hccccceeccccchhhhhhhhhhhhhhhhhhhcccc-cccc-ccc
Arabidopsis CDKHRNELVTSPSLEEIVHYDLWAREYAANVQLSS-GARP-VHA
GOR4 ccccccccccccccceeehhhhhhhhhhhhhhhcc-cccc-eec
HNNC hccccccceccccchhhhhhhhhhhhhhhhhhehhhh-cccc-ccc
Populus CDKHQAELVVSPSLEEIVHYDLWAREYAASLQHSS-GPSP-VFA
GOR4 cccccccceccccccceeehhhhhhhhhhhhhcccccc-cccc-eec
HNNC ccccccceeeeccchhhhhhhhhhhhhhhhhhhhhccc-cccc-ccc
Linum CAKHREELVTSPSLEEIIHYDLWAKDYAASLQ-QAHGLSP-ALV
GOR4 hhcccccccccchhhhhhhhhhhhhhhhhhhh-hhhcccc-eec
HNNC hhccccceeccchhhhehhhhhhhhhhhhhh-hhhcccc-ccc
Pueraria CQEHQKELVVAPSLEEIIHYDQWARQYAASLQKAS-------SV
GOR4 ccccccceeeccccccchhhhhhhhhhhhhhhhhhhh-------hc
HNNC hccccceeeeccchhhhhhhhhhhhhhhhhhhhhc-------cc
Mentha CDKHRSEMAVSPSLEEIVHYDQWARDYAATVLKSA-GLSP-ALV
GOR4 ccccccccccccccceeehhhhhhhhhhhhhhhhhc-cccc-eec
HNNC hccccccccecccchhhhehhhhhhhhhhhhhhhc-cccc-ccc
Oryza CDAHRNELVTRPSLEEIIHYDLWAREYAASLQPST-GLSP-VPV
GOR4 ccccccccccccchhhhhhhhhhhhhhhhhcccccc-cccc-eec
HNNC hccccccceecccchhhhhhhhhhhhhhhhhhcccccc-cccc-ccc
Zea CNAHRNELVTSPSLEEIVHYDLWARRYAASLQPSS-GLSP-VPA
GOR4 ccccccccccccccceeehhhhhhhhhhhhcccccc-cccc-eec
HNNC hhhccccceecccchhhhhhhhhhhhhhhhhhcccccc-cccc-ccc
Hordeum CDAHRNELVTSPSLEEIIHYDQWARKFAANLQPSSSGRSP-VLA
GOR4 ccccccccccccchhhhhhhhhhhhhhhhhcccccccccce-eec
HNNC hhhccccceecccchhhhhhhhhhhhhhhhhhcccccccccc-ccc
Taxus CDKHRNELVLRPSLEEIIHYDLWARKYAASLAQSS--LEP-AMV
GOR4 ccccccccccccchhhhhhhhhhhhhhhhhhhhhhhc--ccc-eec
HNNC hccccccceeeccchhhhhhhhhhhhhhhhhhhhhhc--ccc-ccc
Ginkgo CDKHKNELVLQPSLEEIIYYDQWARQYATSLVRSS--LEP-IAV
GOR4 cccccccceeeccchhhhhhhhhhhhhhhhhhhhhcc--ccc-eec
HNNC hccccceeeeccchhhhehhhhhhhhhhhhhhhc--ccc-ccc

Appendix III

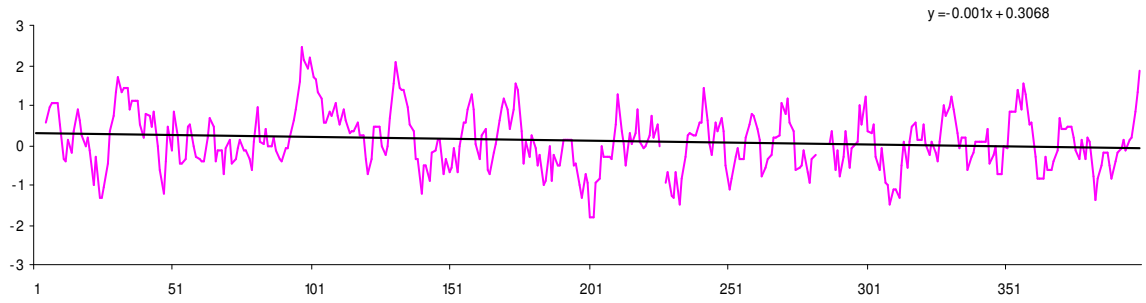
Hydrophobicity plots for the functional domain of DXR

Dicots

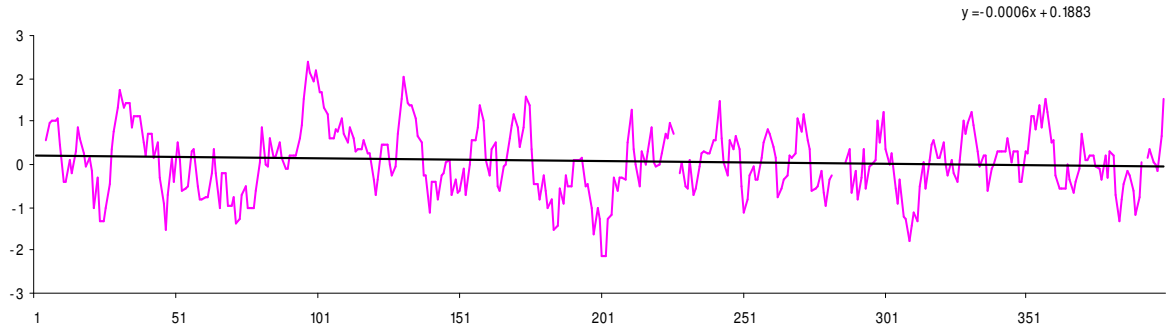
Artemisia annua



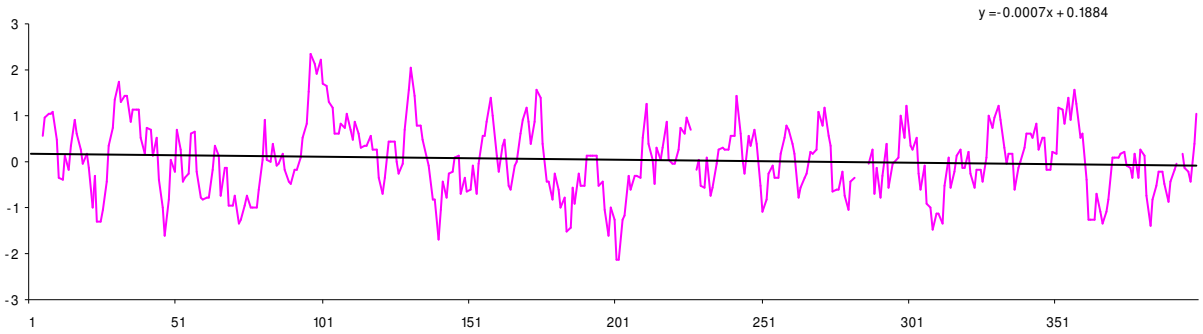
Stevia rebaudiana



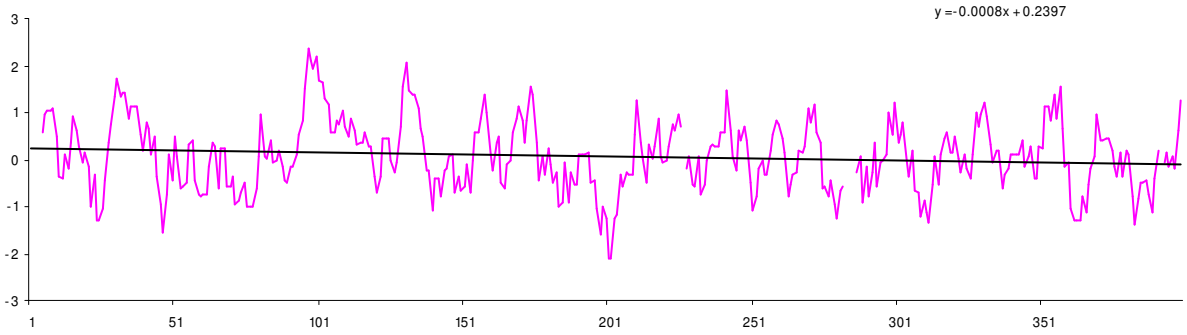
Catharanthus roseus



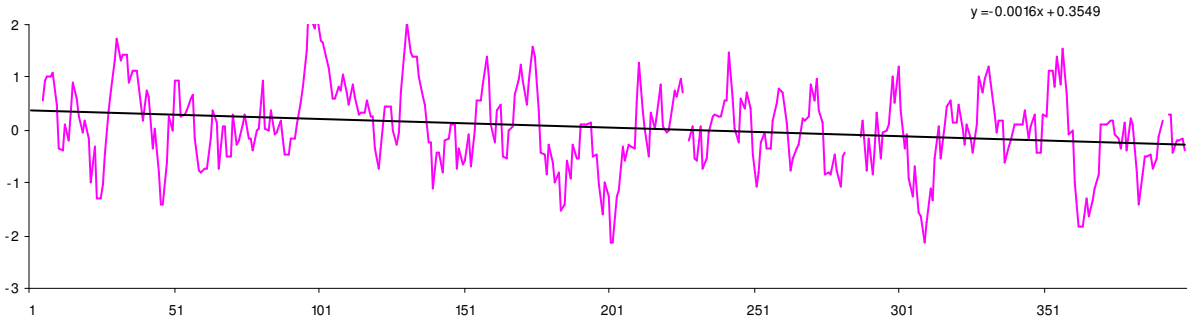
Lycopersicon esculentum



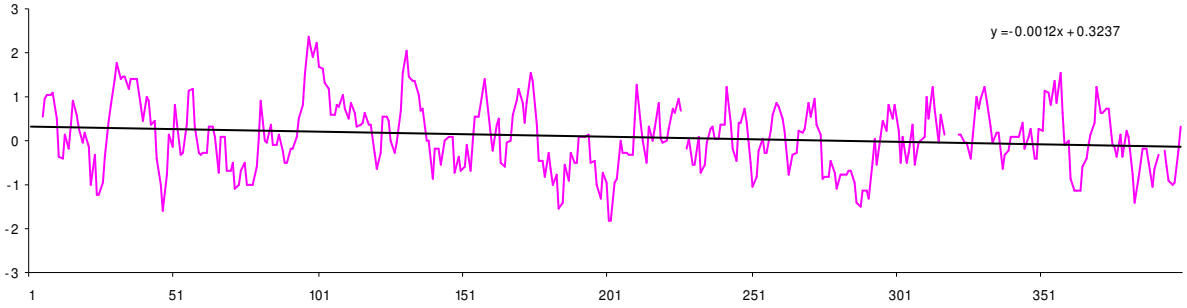
Antirrhinum majus



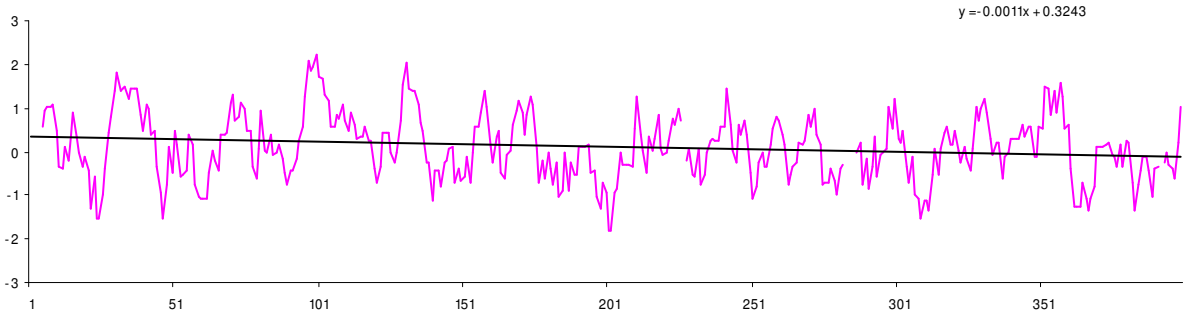
Arabidopsis thaliana



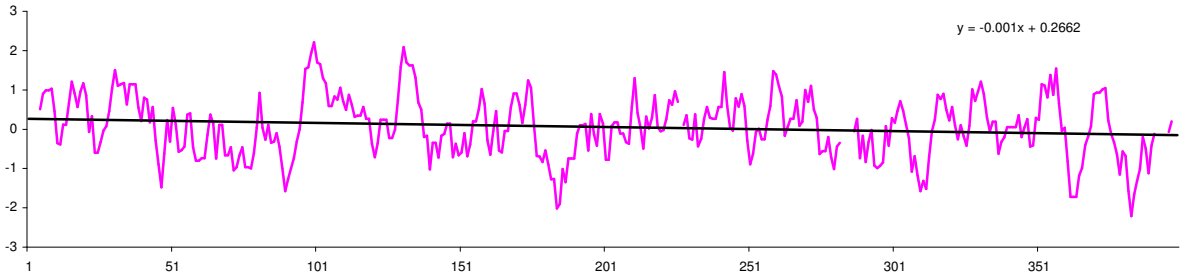
Populus alba x Populus tremula



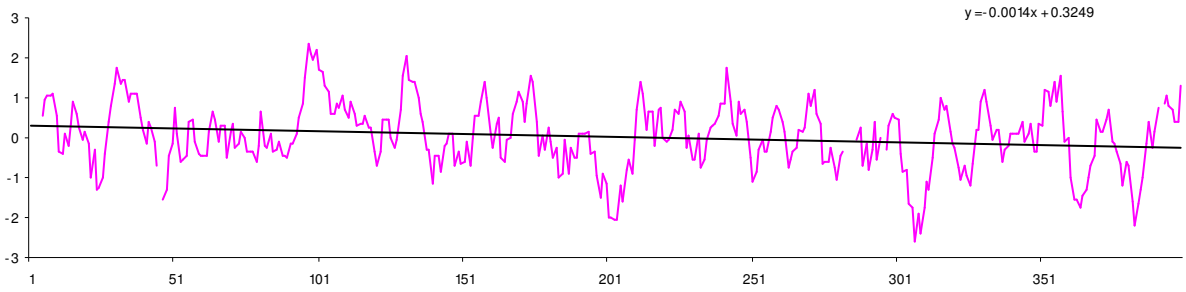
Linum usitatissimum



Pueraria montana var. lobata

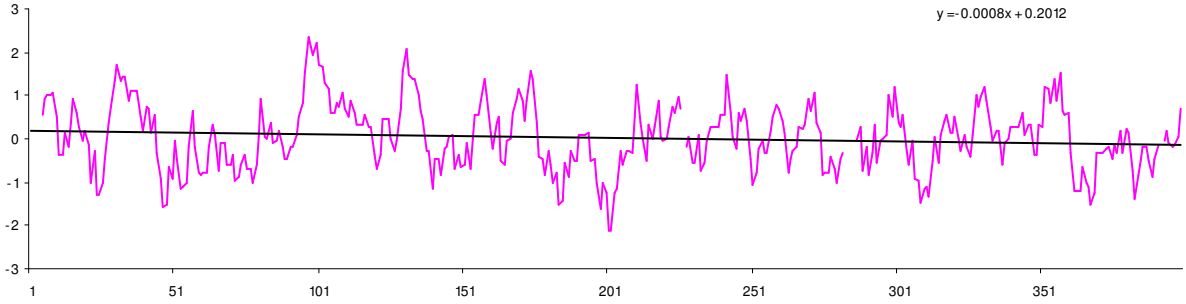


Mentha x piperita

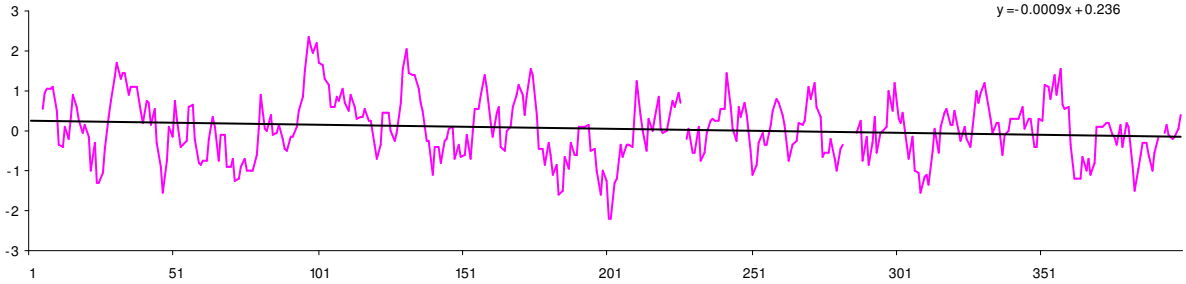


Monocots

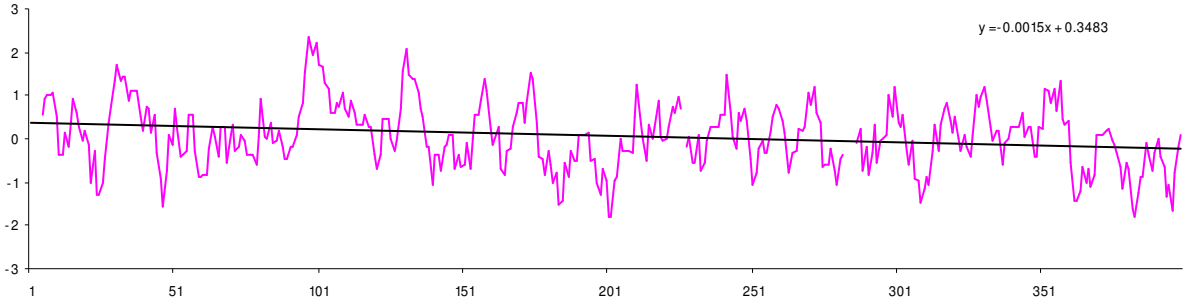
Oryza sativa



Zea mays

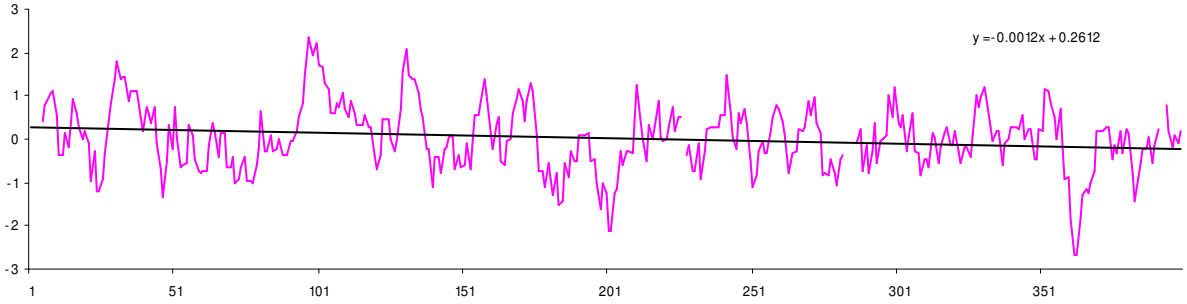


Hordeum vulgare



Gymnosperms

Taxus cuspidata



Ginkgo biloba