



Universiteit
Leiden
The Netherlands

Opleiding Bioinformatica

The Bioinformatic Search for *ENOD40* Genes in the Rosales Order
and Possible Insertions of Transposable Elements

Niels van Belle

Supervisors:

Dr. A.P. Goultiaev & Dr. K.J. Wolstencroft

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)

www.liacs.leidenuniv.nl

05/07/2021

Abstract

The root nodule symbiosis between species of the *Fabaceae* family and rhizobia is an important relationship for biological nitrogen fixation, but the exact processes that lie behind this endosymbiosis remain unknown. One of the genes that can provide more insight, however, is the gene *ENOD40*, which was originally identified to play a role in the early root nodule formation. *ENOD40* codes for a short peptide, but also contains a stretch of non-coding functional RNA structures, where the three structures called domains 4, 5 and 6 have initially only been identified in legumes and, thus, are assumed to be directly involved in the formation of nitrogen fixing root nodules. This research has tried to shed more light on the *ENOD40* gene by annotating multiple *ENOD40* homologues in plant species of the Rosales order, which contains the only species besides legumes that were found to form a symbiosis with rhizobia. This has been done by using two earlier listed homologues and BLASTing them against multiple fully sequenced genomes of the Rosales order. In total 45 new *ENOD40* homologues have been identified and annotated in over 38 different species. For the first time domains 4, 5 and 6 have also been identified in species other than legumes and a new domain has been discovered, named domain S, upstream of the short open reading frame. Furthermore, new probable consensus structures have been introduced for Rosales species. Lastly, since some of the structures of *ENOD40* are suggested to originate from the insertions of transposable elements, this research has examined the homologues for possible transposable element insertions. This paper presents more information on the possible mechanics and evolutionary history behind *ENOD40*, presenting multiple possibilities for follow-up research.

Acknowledgments

I would like to thank my supervisor Dr. Goultiaev for his support and advice during this project.

Dr. Goultiaev has helped a lot by discovering the so called domain S in this project and by annotating some of the *ENOD40* genes in, for example, *Humulus lupulus* and *Cannabis sativa*. I also appreciated the suggestions on what to do next when I got stuck on something and the suggestions on the species I should give some more attention to, which eventually also led to the discovery of some domains in species I had not noted before.

Contents

1	Introduction	1
1.1	Thesis overview	2
2	Methods	3
2.1	GenBank and RefSeq	3
2.2	BLAST	3
2.3	RNAstructure	3
2.4	LocARNA	3
2.5	Annotating <i>ENOD40</i> homologues	4
2.6	Creating Consensus Structures	4
2.7	Identifying Transposons	4
3	Results	5
3.1	Newly annotated <i>ENOD40</i> homologues	5
3.2	New consensus structures for Rosales species	6
3.3	Identification of transposable element insertions	6
4	Discussion	16
4.1	Limitations	18
5	Conclusions and Further Research	18
	References	19
A	A list of the whole genomes from Rosales species, which were used as subject sequences in BLAST	23
B	FASTA-format input for domains 4, 5, 6 and S in LocARNA	24
C	All newly annotated Rosales <i>ENOD40</i> homologues	27

1 Introduction

Biological nitrogen fixation is an important process, that provides the building blocks for important compounds such as proteins and nucleic acids [WN05]. The process consists of converting atmospheric nitrogen to a nitrogen source that plants can use, such as ammonia [Pos82], which is also important for the nitrogen uptake by humans, either directly or indirectly, from those plants [Van01]. The fixation of nitrogen in plants is mainly done by specialized bacteria called rhizobia or bacteria from the genus *Frankia* in an endosymbiotic relationship called the nitrogen-fixing root nodule symbiosis (NFN symbiosis) [GCL+18]. Unlike the other well-known endosymbiotic relationship between mycorrhizal fungi and plants, which can be found in a major part of all the plant families [WQ06], NFN symbiosis has only been discovered in 10 plant families up until now, which belong to four orders, Fabales, Fagales, Cucurbitales and Rosales, that is known as the NFN clade [SSM+95, KP02, GCL+18]. Within this clade legume species (*Fabaceae*) belonging to the Fabales order are the only species, besides the non-legume species in the *Parasponia* genus, that were found to actually form a symbiosis with rhizobia instead of the other nitrogen fixing bacteria [vVHB+18]. This symbiosis is seen as especially critical, since it is the most important nitrogen fixer in agriculture with soybean (*Glycine Max*) being the most dominant [HPB08], therefore playing a major part in the food chain and making it essential to understand the mechanisms that lie behind the symbiosis between legumes, some non-legumes and rhizobia and the evolutionary development of nitrogen fixation in species. One of the genes that can help to get insight into this endosymbiotic root nodule symbiosis is the gene *ENOD40*.

ENOD40 is a gene that was originally found to be involved in the early formation of root nodules in legumes [YKH+93, KH93], but has later also been identified in many non-legume species, also playing a regulatory role in multiple stages of plant growth [Rut03, vdSPC+96]. Although, the *ENOD40* gene has been found in many plant species, it has been noted to be structurally as well as sequentially varying across different species [GRG+03], with sequence similarity being down to 30% between some distantly related species [Rut03]. The current global structure of *ENOD40* can be seen in the schematic of figure 1 [GR07]. As shown *ENOD40* consists of a short open reading frame (sORFI), which codes for a short peptide that is expected to be involved in the regulation of sucrose in the nitrogen fixing nodules [RSM+02]. Additionally, the remaining part of *ENOD40* consists of non-coding RNA, which is folded into multiple structures that are thought to also be actively involved in the plant development, acting as regulators [CJP+94]. Lastly, within the non-coding RNA, region II is located, which has been identified as the most conserved nucleotide sequence in the gene [vdSPC+96]. The reason, now, why *ENOD40* can help in understanding the mechanics behind NFN symbiosis between rhizobia and plants and the evolution of this relationship, is because the secondary structures called domains 1, 2 and 3 are conserved in most legumes and non-legumes, while the so called domains 4, 5 and 6 were found to be exclusively conserved in legume species, which suggests that these domains are directly involved in the formation of nitrogen fixing root nodules. Furthermore, domain 4 is almost solely found in legumes that form indeterminate nodules instead of determinate [GRG+03, GR07]. Another interesting observation about *ENOD40* is that it was found to contain insertions of transposable elements [CRA+03], which can give more evidence of transposons being directly involved in the origination of functional non-coding RNA structures, which often leads to the conservation of functions [BBG+18]. The complete functionality of *ENOD40* and the reason why the gene in legumes is structurally different

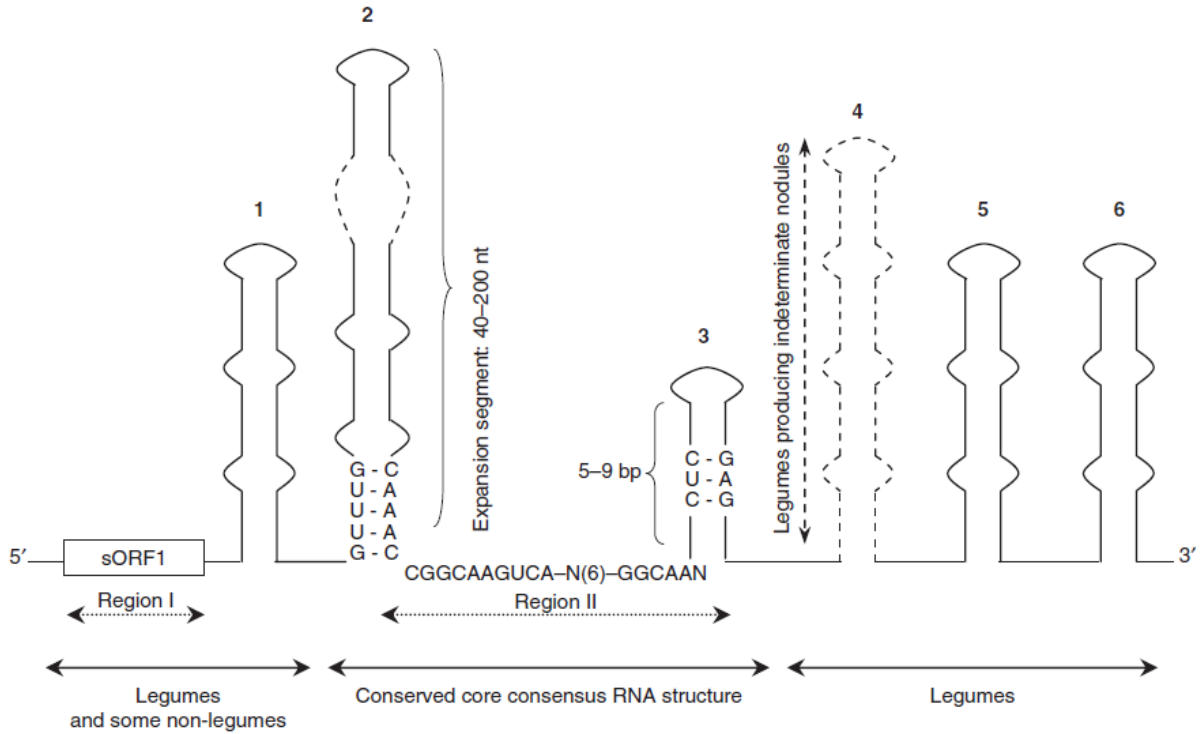


Figure 1: Schematic of the global structure of *ENOD40*, showing the different regions and domains, including the conserved sequence motifs for domains 2 and 3. Region II has been identified as the most conserved region in the gene. [GR07]

than in non-legumes is still not entirely clear, however, and requires further research.

This paper will try to shed some more light on the *ENOD40* gene by searching for new homologues in the Rosales order, that is part of the NFN clade and contains the *Parasponia* species, which are the only species besides legumes that form an endosymbiotic relationship with rhizobia. This will be done by using two earlier annotated *ENOD40* genes from the Rosales order and by making use of the conservation of region II and some of the domains. The RNA sequences and structures of the newly found homologues will be examined, especially for the presence of domains 4, 5 and 6, and annotated. Additionally, the homologues will be examined for possible transposable element insertions, which will give more evidence for the origination of structural domains from these elements. These aims will assist in answering the following research question: What is the sequential and structural composition of *ENOD40* RNA in species of the Rosales order?

1.1 Thesis overview

This section contains the introduction, which presents the topic and content of the thesis and the aim of this project. Section 2 will give an overview of the used tools and the methods used to acquire results. Section 3 will present the obtained results. Section 4 will discuss the results and the limitations of this paper. Finally, section 5 will conclude the paper and present some future research possibilities.

2 Methods

This section will give an overview of the tools and services used in this research and will go further into detail about the methods used for annotating new *ENOD40* homologues, producing consensus structures and finding transposable element insertions.

2.1 GenBank and RefSeq

ENOD40 queries were BLASTed against whole genomes of multiple species. These genomes were annotated in the Genbank (release 242.0) and Refseq (release 204.0) databases, provided by the National Center of Biotechnology information (NCBI) as part of their nucleotide database. GenBank and RefSeq are both comprehensive databases with publicly available nucleotide sequences, which are assigned unique identifiers, called accession numbers, so that they can be easily accessed [BCC⁺12, OWB⁺16].

2.2 BLAST

To find similarities between previously annotated *ENOD40* genes and nucleotide sequences of other species, the Basic Local Alignment Search Tool (BLAST) was used, specifically the nucleotide blast option for whole genomes, provided by the NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>, release 2.11.0). BLAST is a bioinformatic tool, which is used for aligning two or more sequences to identify similar regions [AGM⁺90]. Once two sequences are aligned, BLAST will return a list with regions from the target sequence that are significantly related to the query sequence, scored by a so called Expect Value (E-value), which represents the chance of finding the alignment randomly.

2.3 RNAstructure

To identify some of the domains in the newly obtained *ENOD40* homologues, the online ‘Predict a Secondary Structure’ tool was used, which is part of the RNAstructure software package [RM10] (<https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html>, version 6.0.1). This tool uses thermodynamics, specifically free energy minimization, to predict the most likely folds from the given DNA or RNA sequence [MT06]. The generated secondary structure predictions are listed in ascending order, since a lower free energy matches a structure with greater stability.

2.4 LocARNA

New consensus structures for domains S¹, 4, 5 and 6 were generated with the help of the online alignment and folding tool LocARNA (<https://rna.informatik.uni-freiburg.de/LocARNA/Input.jsp>, version 1.9.1). This tool aligns multiple sequences, while also taking into account the base-pairing between two bases within a sequence in order to maintain structural similarity. LocARNA outputs the alignment, showing the conservation of the base pairs, whilst also outputting a predicted secondary consensus structure, based on the alignment and conservation [SHR⁺10, WRH⁺07, WJH⁺12].

¹Domain S is a secondary structure, which is located upstream of sORFI and will be further mentioned and discussed in sections 3 and 4.

2.5 Annotating *ENOD40* homologues

In order to find new *ENOD40* homologues, two earlier annotated homologues listed in previous research by Gulyaev and Roussis [GR07] were BLASTed against the whole genomes of species within the Rosales order and also against the nucleotide collection of Rosales species to find *ENOD40* homologues in species without fully sequenced genomes. The two initial query sequences used, were *ENOD40* from *Prunus armeniaca* and *Malus x domestica*, accession numbers CV047471 and CN917334 respectively. The list of whole genomes that were used as subject sequences can be found in appendix A. Since *ENOD40* sequences across species can differ substantially, the blastn option was used in BLAST Genomes for somewhat similar sequences with the default parameter settings, because this allowed for smaller word sizes, which accounted for the variability in *ENOD40* genes. After getting a hit, the alignment was compared to the query by eye, especially by looking at the existence of the highly conserved region II and common sequence motifs for domain 2 and 3, to confirm that the result was indeed an *ENOD40* homologue. No cutoff was used for the E-values of the acquired hits, since most of the results only consist of an alignment with region II, which is the most conserved among the homologues in different species, sometimes resulting in high E-values, so each hit was manually checked for the possibility of being an *ENOD40* gene. Once, domains 2 and 3 and region II were identified and annotated in the new homologues, sORF I was also annotated by looking for a start- and stopcodon upstream of domain 2. Since, domains 1, 4, 5 and 6 are less conserved than the other regions and domains across species, these domains were identified and annotated by using the online secondary structure prediction tool from RNAstructure to identify and annotate these domains. To predict domain 1, the nucleotide sequence between the earlier annotated sORFI and domain 2 was used as input and for domains 4, 5 and 6 the nucleotide sequence downstream of domain 3 was used as input. The RNA folds were predicted with the default parameter settings, except for the temperature, which was adjusted to 293.15 Kelvin (20 °C), because this was found to translate better to the plants' natural conditions [GRG⁺03]. Whenever a newly found homologue was fully annotated, it was further used as a query in BLAST to once again find new homologues in the earlier defined subject sequences, repeating the process.

2.6 Creating Consensus Structures

In order to create probable, new consensus structures, which are more suited toward species within the Rosales order, in specific for domain 4, 5, 6 and S, the different domains of newly identified *ENOD40* homologues were put in separate FASTA-files (see appendix B) and used as input in the LocARNA online tool. All default parameter settings were used, except for the temperature, which was changed to 20 °C, since the default allowed for a global alignment of all input sequences that takes into account a probable structural conservation of the sequences even when the sequence similarity is low.

2.7 Identifying Transposons

In order to identify possible insertions of transposable elements in the newly found *ENOD40* homologues, the homologues were BLASTed against their own genomes. The nblast algorithm was used again with the default parameter settings, except for the species-specific repeat element mask filter, which had been deactivated, since transposons are a category of repeats [Ler10]. Whenever

BLAST returned multiple hits aligning to the same region within the *ENOD40* query sequence, this was considered a possible transposable element. No E-value cutoff was used, since the word sizes and, thus, the aligned sequences could be small in some hits, resulting in higher E-values in some cases.

3 Results

This section will present the results of the search for new *ENOD40* homologues and the creation of a new consensus structures based on these new homologues that is more consistent with Rosales species. Lastly, the results of the search for transposable element insertions will also be presented.

3.1 Newly annotated *ENOD40* homologues

In total 45 new *ENOD40* homologues have been identified and annotated in 38 different species of the Rosales order. A complete list of all discovered homologues of species within the Rosales order with their accession numbers and the locations of the different domains can be viewed in table 1². The nucleotide sequences of the newly found homologues with the different domains and regions colored, can be seen in appendix C. In 7 different species a domain 4 and 5 have been identified, while in 8 different species a domain 6 has been identified. Furthermore, in 13 different species a new domain, directly upstream of the sORFI, has been identified and exists of a stable hairpin with an almost perfect inverse repeat in some species. This new domain has been named ‘domain S’³ based on the previously used nomenclature [GRG+03]. The different domains 4, 5, 6 and S in the newly found homologues are presented in figures 2, 3, 4, 5 respectively. As shown, domain 4 is mostly around 50-60 bases long, consisting of a few internal loops with the bottom one being the biggest most of the time and a smaller hairpin loop at the end. Domain 5 is also around 50-60 bases long with the exception of *M. alba* and *R. rubrinervis*, which are about 30 bases long. The structures of domain 5 mostly consist of small internal loops and bulges with a bigger hairpin loop at the end with the exception of *R. rubrinervis*, which is almost a perfect inverse repeat. Domain 6 is around 35-40 bases long, except for *H. lupulus var. lupulus* and *R. rubrinervis*, which are around 20 bases long. The structures of domain 6 mostly consist of one or two internal loops with a hairpin loop of around 5 bases, except for *P. andersonii* and *T. orientale*, which contain a single bulge in the stem, and *H. lupulus var. lupulus* and *R. rubrinervis*, which consist of a simple stem-loop. At last, domain S is around 40-50 bases long, except for *R. rubrinervis*, *O. trinervis* and *z. jujuba*, which are around 30 bases long. In most species domain S consists of a stable hairpin with sometimes one internal loop or bulge or both, except for *A. camansi*, which contains multiple bigger loops. The sequence of domain S forms almost a perfect inverse repeat in most species. Although multiple species contain a domain 4, 5 and 6, only *Parasponia andersonii* and *Morus alba* were found to contain all regions and domains, including domain S, as the sORFI in most of these species was not conserved, because of a disrupted stopcodon. Lastly, another noticeable observation which can

²The *ENOD40* homologues in *Humulus lupulus var. lupulus* and *Cannabis sativa* were identified by Dr. Goultiaev in this research.

³This domain has been initially identified and discovered in this research by Dr. Goultiaev in the *Parasponia andersonii*, *Trema orientale*, *Humulus lupulus var. lupulus* and *Cannabis sativa*.

be made is that the two homologues found in *Ficus erecta* also have an extra domain between domains 4 and 5, which has not been identified in the other species, as can be seen in figure 6.

3.2 New consensus structures for Rosales species

The alignments of domains 4, 5, 6 and S of the newly found *ENOD40* homologues and the probable consensus structures based on these alignments are presented in figure 7. As can be seen in the figure, the alignments and structures indicate that the domains are conserved across species in the Rosales order, also containing multiple base covariations. Domains S and 5 seem to be more conserved, however, than domains 4 and 6, sequentially as well as structurally.

3.3 Identification of transposable element insertions

In none of the currently annotated *ENOD40* homologues of the Rosales order transposon insertions could be identified. However, the two species *Ficus erecta* and *Artocarpus camansi* still presented noteworthy results. The BLAST search in *Ficus erecta* resulted in multiple hits similar to a part mainly located in domain 5, sometimes extended to domain 6 and in *Artocarpus camansi* it resulted in multiple hits similar to a part in domain S, as can be seen in figure 8. The hits were mainly associated with the repeating sequence TAA in *Ficus erecta* and the repeating sequence GATTCTT in *Artocarpus camansi*. These repeating sequences will be discussed in more detail in section 4.

Species	Accession	Reverse complement	Domain S	sORF I	Domain 1	Domain 2	Domain 3	Domain 4	Domain 5	Domain 6	
<i>Prunus armeniaca</i>	CV047471	no	-	80-127	133-175	179-263	287-302	-	-	-	[Rut03]
<i>Prunus salicina</i>	CM026527	yes	-	25129958-25129911	25129905-25129863	25129859-25129775	25129751-25129736	-	-	-	
	WERZ01000077	yes	-	203144-203097	203091-203049	203045-202961	202937-202922	-	-	-	
<i>Prunus persica</i>	NC_034009	no	-	26194361-26194408	26194414-26194456	26194460-26194544	26194568-26194583	-	-	-	
<i>Prunus dulcis</i>	NC_047650	no	-	22781428-22781475	22781481-22781524	22781528-22781612	22781636-22781651	-	-	-	
<i>Prunus mume</i>	NC_024127	yes	-	18629217-18629170	18629164-18629122	18629118-18629034	18629010-18628995	-	-	-	
<i>Prunus yedoensis</i>	BJCG01002293	yes	-	2948449-2948402	2948396-2948354	2948350-2948266	2948242-2948227	-	-	-	
	BJCG01000003	no	-	1939390-1939437	1939443-1939485	1939489-1939573	1939597-1939612	-	-	-	
	BJCG01003116	yes	-	46334-46287	46281-46239	46235-46151	46127-46112	-	-	-	
<i>Prunus avium</i>	NW_018921537	yes	-	513610-513563	513557-513515	513511-513427	513403-513388	-	-	-	
<i>Malus x domestica</i>	CN917334	no	-	-	30-75	79-188	211-228	-	-	-	[GR07]
<i>Malus baccata</i>	VIEB01000769	yes	-	-	466807-466762	466758-466649	466626-466609	-	-	-	
<i>Rosa chinensis</i>	NC_037091	yes	-	63744234-63744196	63744183-63744153	63744149-63744063	63744041-63744018	-	-	-	
<i>Rosa multiflora</i>	BDJD01002838	no	-	72779-72817	72830-72860	72864-72950	72972-72995	-	-	-	
<i>Rosa luciae</i>	RQIQ01004170	no	-	2549-2587	2600-2630	2634-2720	2742-2765	-	-	-	
<i>Rosa x damascena</i>	LYNE01000270	no	-	37673-37711	37724-37754	37758-37844	37866-37889	-	-	-	
<i>Fragaria iinumae</i>	CM019120	yes	-	32931807-32931742	32931729-32931691	32931687-32931600	32931577-32931554	-	-	-	
<i>Fragaria nilgerrensis</i>	CM020980	yes	-	34980215-34980150	34980137-34980099	34980095-34980008	34979985-34979962	-	-	-	
<i>Fragaria vesca</i>	NC_020494	yes	-	18831652-18831587	18831574-18831536	18831532-18831445	18831422-18831399	-	-	-	
<i>Fragaria x ananassa</i>	BATT01307848	yes	-	1394-1329	1316-1278	1274-1187	1164-1141	-	-	-	
	BATT01197499	no	-	1248-1313	1326-1364	1368-1455	1478-1501	-	-	-	
<i>Fragaria nubicola</i>	BATW01062352	no	-	2947-3012	3025-3063	3067-3154	3177-3200	-	-	-	
<i>Fragaria nipponica</i>	BATV01076376	yes	-	11530-11465	11452-11414	11410-11323	11300-11277	-	-	-	
<i>Pyrus betulifolia</i>	CM017599	yes	-	-	5572192-5572162	5572150-5572040	5572017-5572000	-	-	-	
<i>Pyrus pyrifolia</i>	BNSU01000004	no	-	-	12849441-12849471	12849483-12849593	12849616-12849633	-	-	-	
<i>Cydonia oblonga</i>	JADOB010303350	no	-	3514-3588	3653-3683	3695-3805	3828-3845	-	-	-	

<i>Geum urbanum</i>	OEJZ01132735	no	-	-	8086-8112	8118-8201	8225-8245	-	-	-	
	OEJZ01153179	no	-	12471-12554	12568-12589	12600-12684	12708-12728	-	-	-	
<i>Purshia tridentata</i>	QANT01007693	no	-	3285-3368	-	3389-3481	3504-3527	-	-	-	
<i>Dryas drummondii</i>	QANW01002007	no	-	1510989-1511063	-	1511095-1511187	1511210-1511233	-	-	-	
<i>Ulmus americana</i>	WUAT02001391	yes	-	30262-30224	-	30080-30005	29981-29964	-	-	-	
<i>Boehmeria nivea</i>	PHNS01007792	no	-	999354-999413	-	999434-999554	999577-999594	-	-	-	
<i>Ochetophila trinervis</i>	QANX01004723	yes	5817-5790	5786-5748	5730-5687	5675-5582	5559-5542	-	-	-	
<i>Ziziphus jujuba</i>	NC_029688	no	5986153-5986179	5986183-5986221	5986241-5986279	5986291-5986385	5986408-5986425	-	-	-	
<i>Artocarpus camansi</i>	LNSY01044605	no	1171-1211	-	1271-1316	1327-1433	1456-1480	-	-	-	
	LNSY01137363	no	586-632	635-676	689-734	748-856	879-902	-	-	-	
<i>Cannabis sativa</i>	NC_044373	no	4233028-4233075	-	-	4233201-4233314	4233337-4233357	-	-	4233521-4233554	2
<i>Humulus lupulus</i> var. <i>lupulus</i>	LD164023	yes	6371-6324	-	6263-6200	6197-6101	6078-6061	-	5993-5938	5936-5916	2
<i>Rhamnella rubrinervis</i>	CM017577	no	2400987-2401015	-	2401077-2401113	2401124-2401216	2401239-2401256	2401276-2401335	2401337-2401363	2401369-2401389	
<i>Ficus carica</i>	CM019749	no	2264655-2264694	-	2264762-2264850	2264862-2264977	2265000-2265023	-	2265100-2265159	-	
<i>Ficus erecta</i>	BKCH01000918	yes	46293-46254	-	46186-46114	46102-45984	45961-45938	45933-45881	45843-45783	45782-45743	
	BKCH01000012	yes	5151121-5151082	-	5151014-5150942	5150930-5150812	5150789-5150766	5150761-5150709	5150670-5150610	5150609-5150570	
<i>Morus alba</i>	CP050237	no	2644262-2644304	2644305-2644349	2644359-2644389	2644399-2644529	2644553-2644575	2644582-2644630	2644646-2644675	2644676-2644716	
<i>Morus notabilis</i>	NW_010362482	no	1331034-1331069	1331070-1331114	1331121-1331158	1331164-1331306	1331330-1331352	1331359-1331403	-	1331430-1331470	
<i>Morus indica</i>	KF030989	no	11-53	54-98	108-138	148-278	302-324	331-379	-	-	
<i>Trema orientale</i>	JXTC01000021	yes	920687-920640	-	920584-920529	920526-920435	920411-920396	920388-920325	920322-920274	920272-920238	
<i>Parasponia andersonii</i>	JXTB01000342	yes	184947-184900	184899-184852	184843-184788	184785-184695	184671-184656	184648-184586	184582-184535	184533-184499	

Table 1: A complete list of all *ENOD40* homologues in the Rosales order with their accession numbers, showing the locations of the domains and the sORFI in each gene

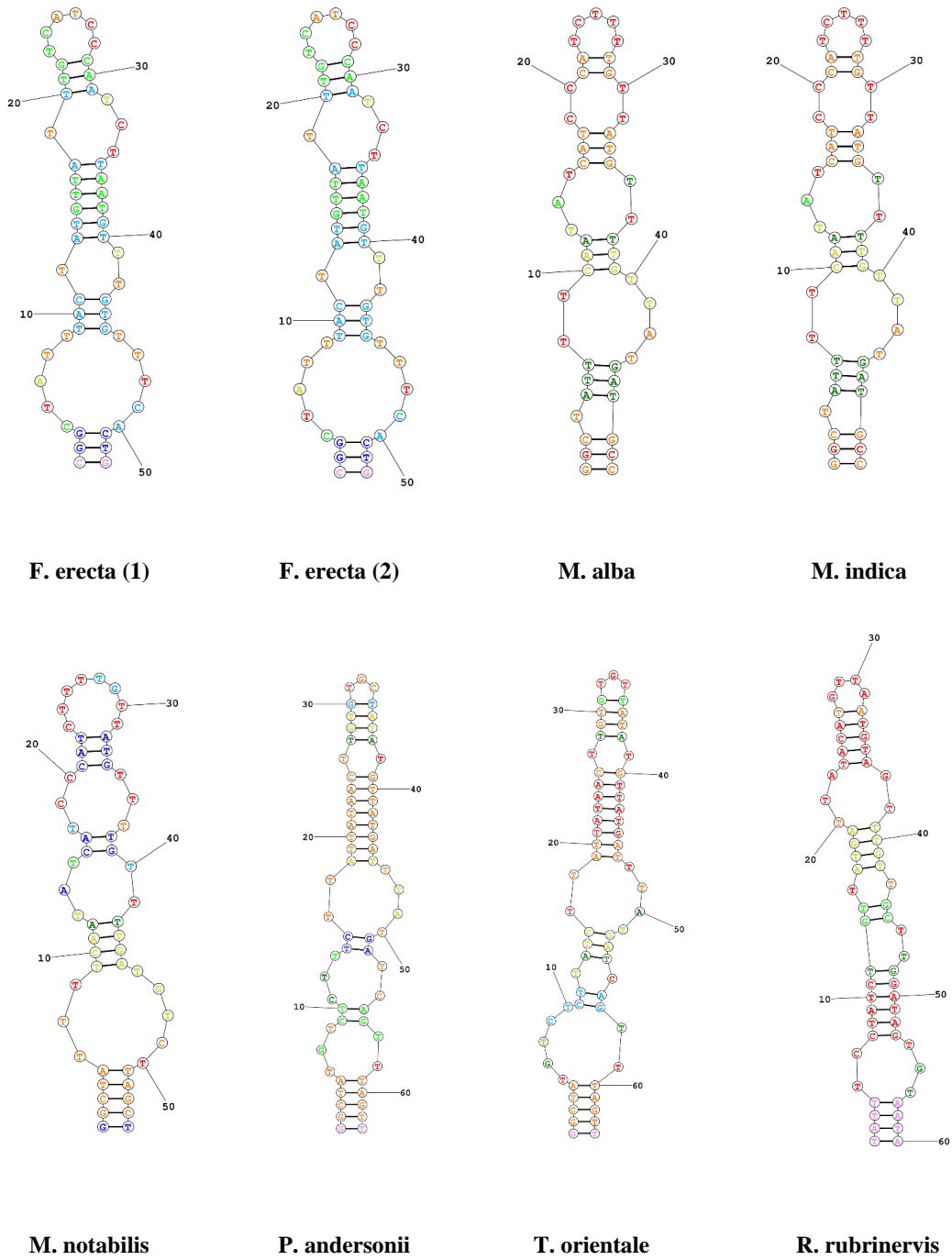


Figure 2: Predicted structures of all identified domains 4 in the *ENOD40* genes of species from the Rosales order

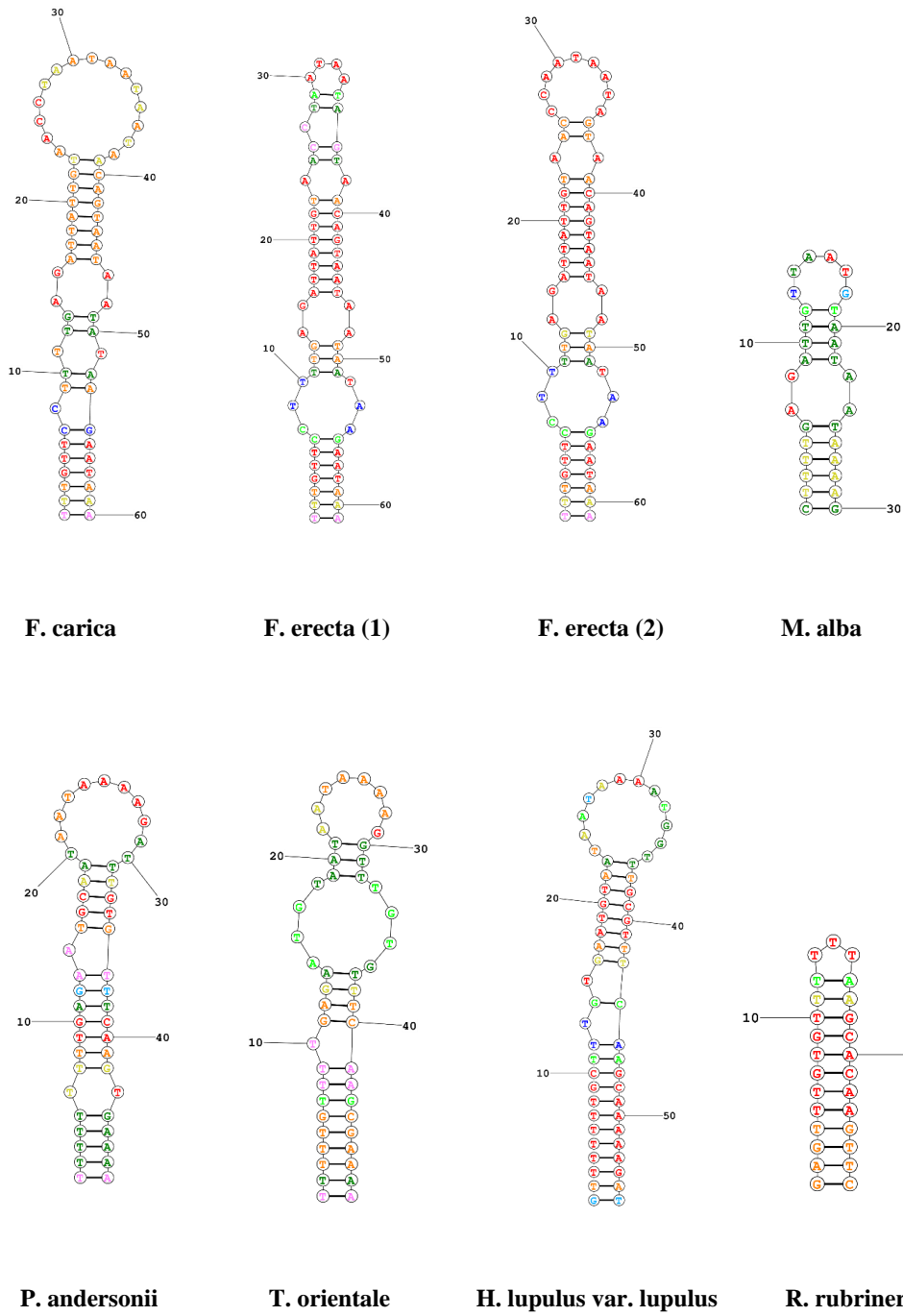


Figure 3: Predicted structures of all identified domains 5 in the *ENOD40* genes of species from the Rosales order

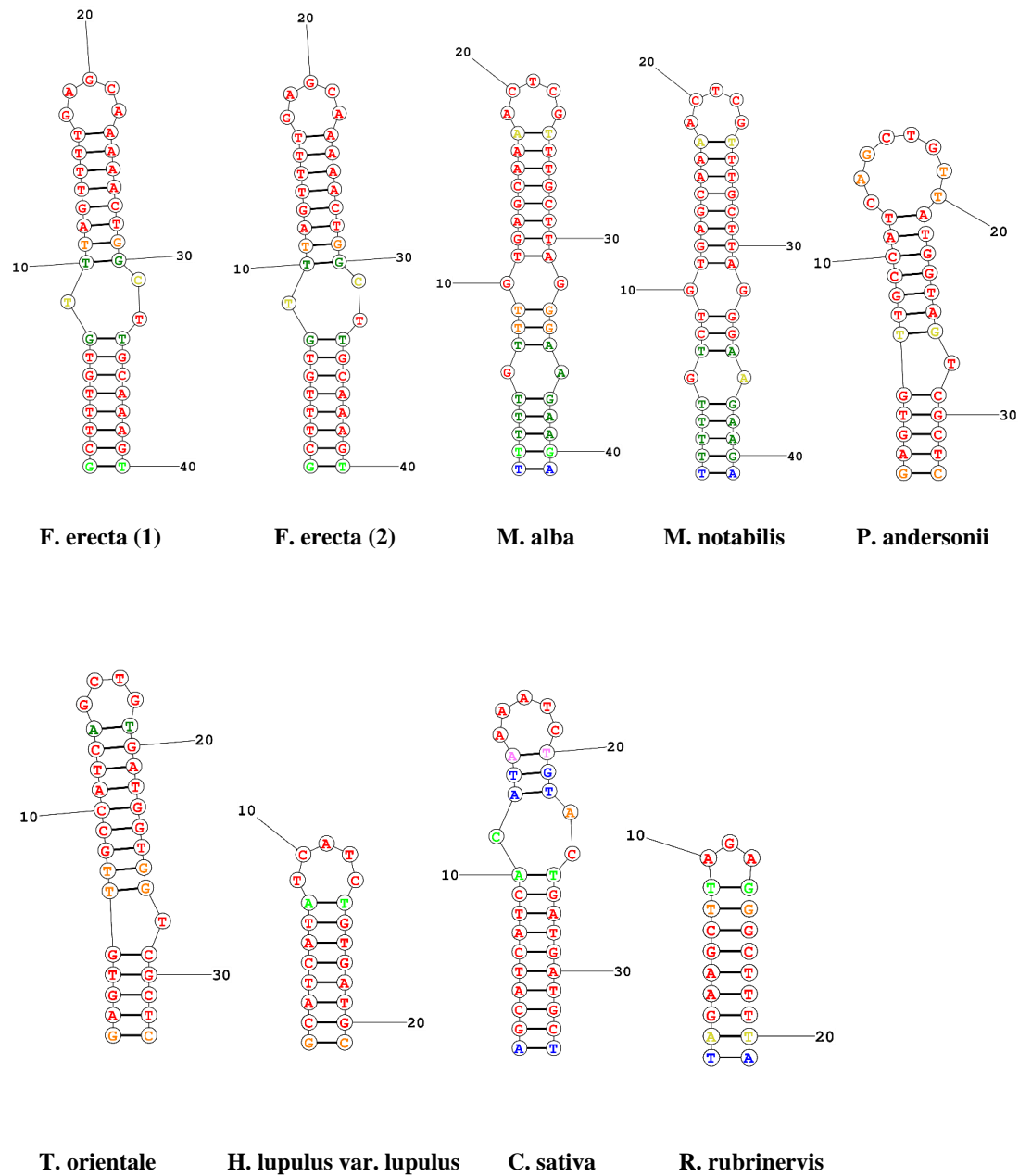


Figure 4: Predicted structures of all identified domains 6 in the *ENOD40* genes of species from the Rosales order

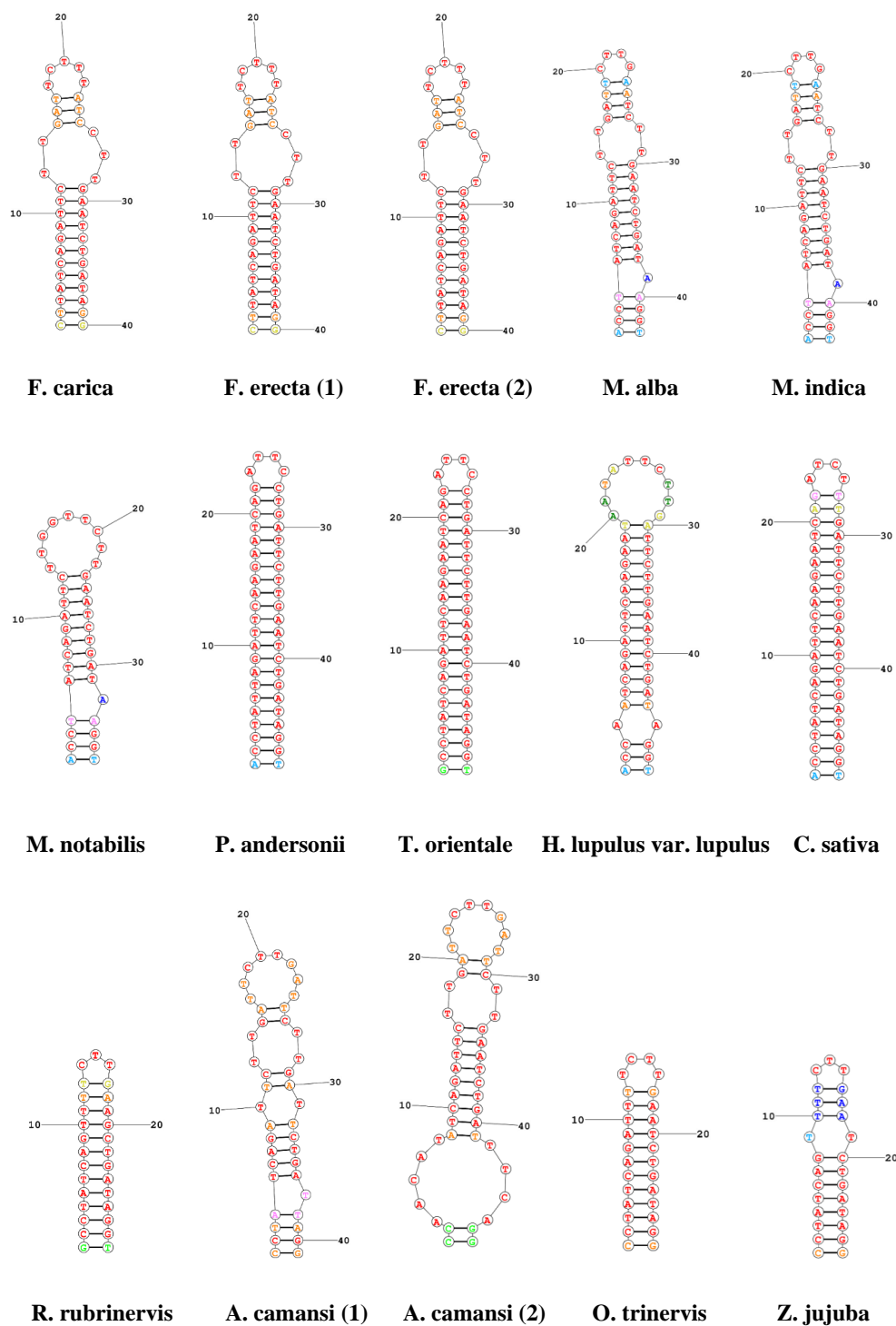


Figure 5: Predicted structures of all identified domains S in the *ENOD40* genes of species from the Rosales order

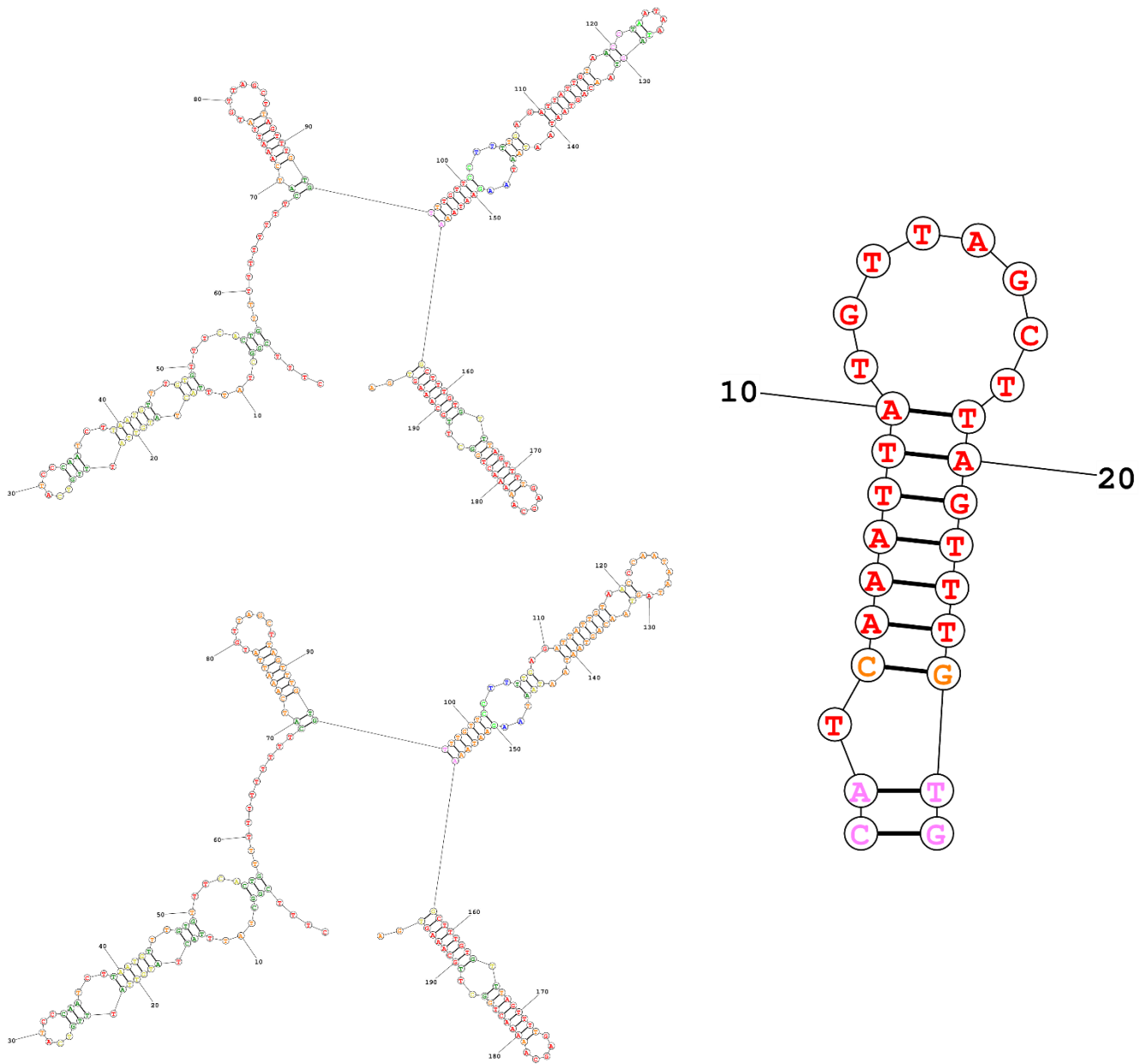


Figure 6: *Ficus erecta* is the only species in the Rosales order that was found to contain an extra domain between domains 4 and 5. The left side shows the domains 4, 5 and 6 of both *ENOD40* homologues in *Ficus erecta* and the right side shows a close up of the extra domain, which consists of a short hairpin loop with a bulge at the beginning of the stem.

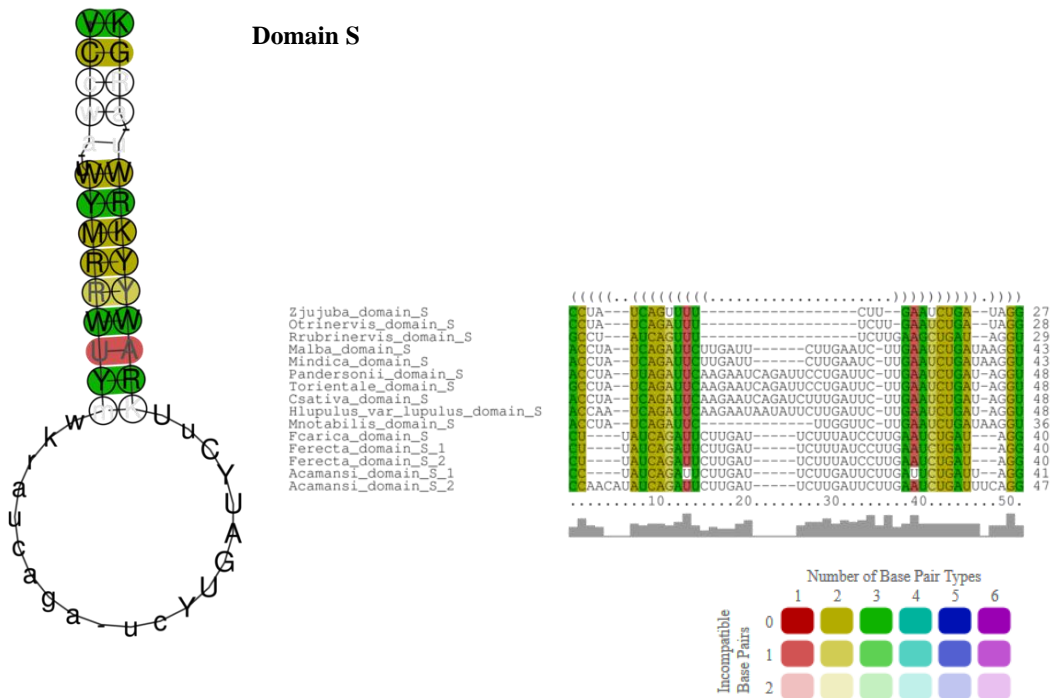
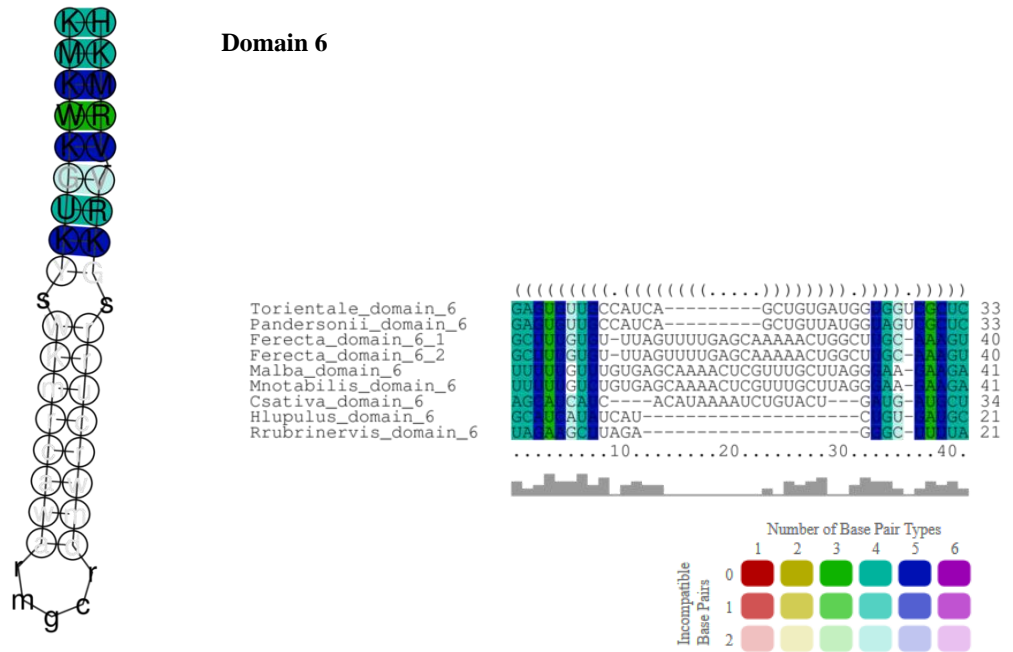


Figure 7: Probable consensus structures for domains 4, 5, 6 and S in species of the Rosales order, produced by the LocARNA tool. The results show that the domains are conserved across species. The colors of each base pair indicates sequential conservation with red being the highest conservation and the opaqueness of each base pair indicates the structural conservation with the greatest opacity being the highest conservation.

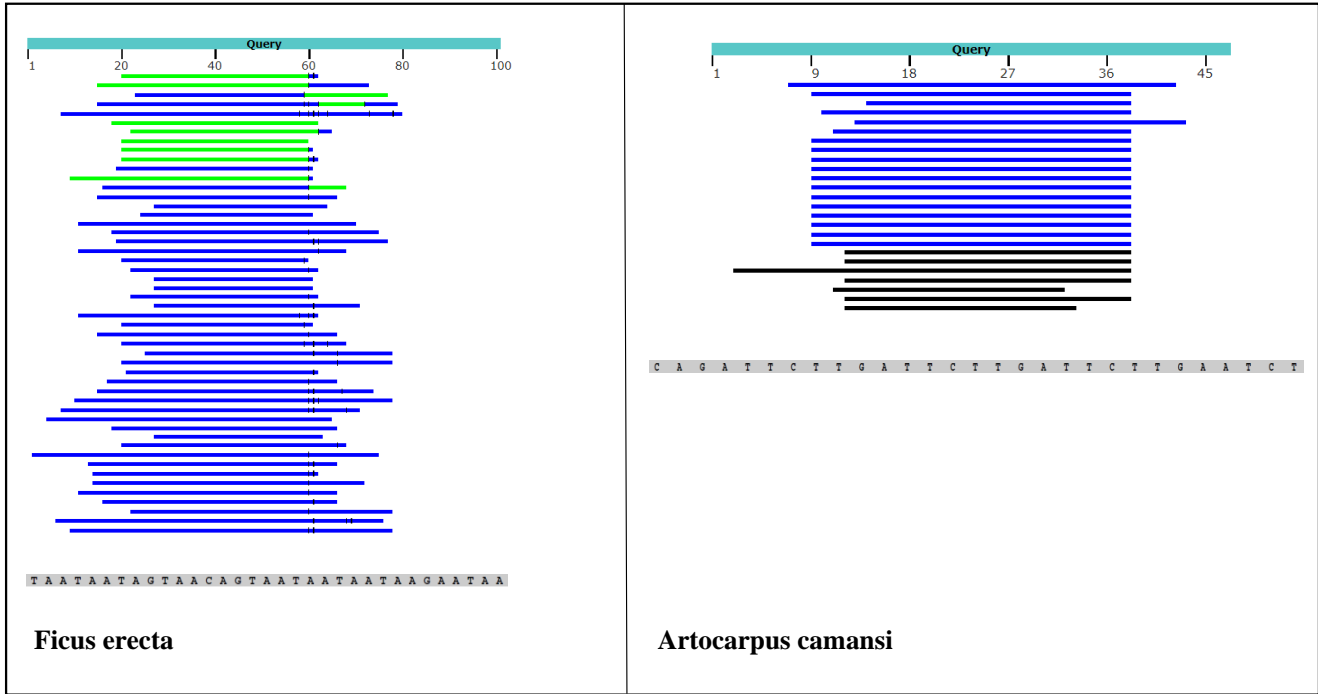


Figure 8: Schematics showing multiple hits for domain 5 in the *ENOD40* gene of *Ficus erecta* and multiple hits for domain S in *Artocarpus camansi*. The sequences underneath each schematic shows the main sequence in domain 5 and S that is associated with each result, which is the repeating sequence TAA in *F. erecta* and GATTCTT in *A. camansi*.

4 Discussion

In this paper, multiple new *ENOD40* homologues have been identified and annotated. For the first time domains 4, 5 and 6 have been identified in species other than legumes (see table 1 and figures 2, 3 and 4), indicating that these domains are not exclusive to these species. Since the species *Parasponia andersonii*, which is part of the only genus (*Parasponia*) that can form a symbiotic relationship with rhizobia besides legume species, was found to contain domains 4, 5 and 6 it could strengthen the assumption that these domains are directly involved in the formation of nitrogen fixing root nodules. Nevertheless, the other 9 species that were also found to contain at least one of these three domains are not known to form any nodules [RMR⁺20], including *Trema orientale* which is phylogenetically closely related to *Parasponia* [YvVB⁺13]. Furthermore, the three domains could also not be identified in species that do nodulate, such as *Ochetophila trinervis* and *Dryas drummondii*, albeit forming a symbiosis with other nitrogen fixing bacteria instead of rhizobia [GCL⁺18]. The reason for the lack in nodulation in some of these species can, however, probably be explained by the absence of other essential genes for nitrogen fixing root nodulation, such as the *NODULE INCEPTION (NIN)* gene [vVHB⁺18, VKF⁺15]. Additionally, the presence of the *ENOD40* gene in non-nodulating species gives more evidence that its function is not restricted to nitrogen fixing symbiosis, but that it is also active in a multitude of other processes [VGW02, SCB⁺01, FCV⁺02]. The exact functionality of *ENOD40* is still unknown, though, and requires further research by examining the pathways and processes in which the gene is expressed more extensive and by comparing its functionality to related genes.

Another noticeable finding in the newly annotated homologues is the presence of domain S in some species, directly upstream of the sORF I (see table 1 and figure 5). This domain has not been observed in other plant orders yet and it is unknown what the exact function of this domain could be. Nevertheless, a likely function for this domain can be to act as a regulator for the translation of sORF I, affecting the binding to ribosomes, for example [LZV⁺12]. Besides domain S, *Ficus erecta* was the only species containing an additional secondary structure between domains 4 and 5 (see figure 6). Although *Ficus erecta* seems to have acquired this domain independently, since it has not been found in other species, it remains to be solved what the exact function is, but it could, for instance, increase the stability of the *ENOD40* RNA [NTJ97] or assist in the regulatory functions of the gene within *Ficus erecta*.

Additionally, in some of the species the presence of sORF I was not observed in the homologues, mostly due to a disrupted stopcodon. This was especially apparent in species containing one or more of the domains S, 4, 5 or 6, since in 8 of the 15 homologues of those species the sORFI was not conserved (see table 1 and appendix C). It could be that the peptides encoded by this sORF are not necessarily needed anymore in these species for proper functioning of the *ENOD40* gene. Furthermore, in the species *Geum urbanum* and *Artocarpus camansi* two homologues of the *ENOD40* gene are present, but only one of the paralogs in each of these species contains a short open reading frame. This could indicate pseudogenization of some of these duplicates [Tut12] or it could indicate that the paralogs complete each other through subfunctionalization [LF00]. The conservation of the sORF I among and across species requires further research, however, in order to give more insight.

The newly formed consensus structures and alignments for domains S, 4, 5 and 6 for species in the Rosales order suggests that these domains are conserved across species within this order (see figure 7). The domains 4, 5 and 6 of legume species [GRG⁺03] seem to show sequential as well as structural differences in comparison to the Rosales species, also lacking a domain S. Region II and domains 2 and 3 with their sequential motifs GUUUG/CAAAC and CUC/GAG, respectively [GR07], do seem to be well conserved in Rosales species, however. This could indicate that the *ENOD40* gene originated in a common ancestor but developed independently across the different lineages. The evolution of genes involved in nodulation is still unknown, though [YH11], and there is need for a thorough comparison between *ENOD40* in legumes and Rosales species.

Since no insertions of transposable elements could be found in the homologues of the Rosales species, no further direct evidence can be given for the origination of the structural domains in *ENOD40* from these elements. The species *Ficus erecta* and *Artocarpus camansi* still presented multiple hits, however, corresponding to the repeating sequences TAA and GATTCTT (see figure 8). More of these repeating sequences can also be found, for instance, in the *Morus* genus by repeating the sequence GAA in domain 2 and the *Ficus carica* repeating TA in domain 1 (see appendix C), although these species returned less results in the BLAST search query. These kinds of simple sequence repeats seem to be evolutionary, strongly related to transposons, even showing signs that these repeats originate from transposons or parts of transposons. The process of expansion or origination of repeating elements from these transposons is, however, not clear yet and needs further research [PFVdS19, MMP⁺15]. Thus, although no direct transposon insertions could be found, the simple sequence repeats that were identified in multiple domains could give evidence of possible interactions with these elements in evolutionary history and may require more attention in

further research.

4.1 Limitations

There are two main limitations to the findings and results in this paper. One is the lack of fully sequenced genomes for species in the Rosales order. Only 39 genomes were fully sequenced, see appendix A, and BLASTing against the nucleotide collection of Rosales only added one more species. There is, thus, a high probability that multiple *ENOD40* homologues have not been sequenced yet, so these species are not available for comparison and to make more elaborated conclusions about the results. The other limitation is that the predictions of the secondary structures are performed in silico, so there is no absolute certainty that the same structures with the same conformation are also observed in vivo.

5 Conclusions and Further Research

This paper has presented a multitude of new annotated *ENOD40* homologues in species of the Rosales order. For the first time domains 4, 5 and 6 have been identified in species other than legumes, giving a new perspective on the functions of the gene and its evolutionary history. In addition, a new domain has been discovered, upstream of sORFI, named domain S. Domain S together with domains 4, 5 and 6, has led to the creation of new consensus structures, which are better fit for Rosales species. No new transposable elements have been identified, but repeating sequences, found in some of the domains, can still possibly indicate a connection with them. All things considered, this research has shed more light on the *ENOD40* gene, but there are also still a lot of open questions. Further research, involving a thorough comparison between Rosales and legume *ENOD40* and the conservation of the domains and regions between them, will probably help answer some of these questions. Lastly, once more genomes have been fully sequenced more elaborated assumptions can be made and researched.

References

- [AGM⁺90] Stephen F Altschul, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, 1990.
- [BBG⁺18] Guillaume Bourque, Kathleen H Burns, Mary Gehring, Vera Gorbunova, Andrei Seluanov, Molly Hammell, Michaël Imbeault, Zsuzsanna Izsvák, Henry L Levin, Todd S Macfarlan, et al. Ten things you should know about transposable elements. *Genome biology*, 19(1):1–12, 2018.
- [BCC⁺12] Dennis A Benson, Mark Cavanaugh, Karen Clark, Ilene Karsch-Mizrachi, David J Lipman, James Ostell, and Eric W Sayers. Genbank. *Nucleic acids research*, 41(D1):D36–D42, 2012.
- [CJP⁺94] MD Crespi, E Jurkevitch, M Poiret, Y d’Aubenton Carafa, G Petrovics, E Kondorosi, and A Kondorosi. enod40, a gene expressed during nodule organogenesis, codes for a non-translatable rna involved in plant growth. *The EMBO Journal*, 13(21):5099–5112, 1994.
- [CRA⁺03] Bert Compaan, Tom Ruttink, Cathy Albrecht, Robert Meeley, Ton Bisseling, and Henk Franssen. Identification and characterization of a zea mays line carrying a transposon-tagged enod40. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1629(1-3):84–91, 2003.
- [FCV⁺02] Bruno Favery, Arnaud Complainville, Jose Maria Vinardell, Philippe Lecomte, Danièle Vaubert, Peter Mergaert, Adam Kondorosi, Eva Kondorosi, Martin Crespi, and Pierre Abad. The endosymbiosis-induced genes enod40 and ccs52a are involved in endoparasitic-nematode interactions in medicago truncatula. *Molecular plant-microbe interactions*, 15(10):1008–1013, 2002.
- [GCL⁺18] Maximilian Griesmann, Yue Chang, Xin Liu, Yue Song, Georg Haberer, Matthew B Crook, Benjamin Billault-Penneteau, Dominique Lauressergues, Jean Keller, Leandro Imanishi, et al. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science*, 361(6398), 2018.
- [GR07] Alexander P Gulyaev and Andreas Roussis. Identification of conserved secondary structures and expansion segments in enod40 rnas reveals new enod40 homologues in plants. *Nucleic acids research*, 35(9):3144–3152, 2007.
- [GRG⁺03] Geneviève Girard, Andreas Roussis, Alexander P Gulyaev, Cornelis WA Pleij, and Herman P Spaink. Structural motifs in the rna encoded by the early nodulation gene enod40 of soybean. *Nucleic acids research*, 31(17):5003–5015, 2003.
- [HPB08] David F Herridge, Mark B Peoples, and Robert M Boddey. Global inputs of biological nitrogen fixation in agricultural systems. *Plant and soil*, 311(1):1–18, 2008.
- [KH93] Hiroshi Kouchi and Shingo Hata. Isolation and characterization of novel nodulin cdnas representing genes expressed at early stages of soybean nodule development. *Molecular and General Genetics MGG*, 238(1):106–119, 1993.

- [KP02] Catherine Kistner and Martin Parniske. Evolution of signal transduction in intracellular symbiosis. *Trends in plant science*, 7(11):511–518, 2002.
- [Ler10] E Lerat. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity*, 104(6):520–533, 2010.
- [LF00] Michael Lynch and Allan Force. The probability of duplicate gene preservation by subfunctionalization. *Genetics*, 154(1):459–473, 2000.
- [LZV⁺12] Fan Li, Qi Zheng, Lee E Vandivier, Matthew R Willmann, Ying Chen, and Brian D Gregory. Regulatory impact of rna secondary structure across the arabidopsis transcriptome. *The Plant Cell*, 24(11):4346–4359, 2012.
- [MMP⁺15] Nevenka Meštrović, Brankica Mravinac, Martina Pavlek, Tanja Vojvoda-Zeljko, Eva Šatović, and Miroslav Plohl. Structural and functional liaisons between transposable elements and satellite dnas. *Chromosome Research*, 23(3):583–596, 2015.
- [MT06] David H Mathews and Douglas H Turner. Prediction of rna secondary structure by free energy minimization. *Current opinion in structural biology*, 16(3):270–278, 2006.
- [NTJ97] Jacek Nowakowski and Ignacio Tinoco Jr. Rna structure and stability. In *Seminars in virology*, volume 8, pages 153–165. Elsevier, 1997.
- [OWB⁺16] Nuala A O’Leary, Mathew W Wright, J Rodney Brister, Stacy Ciufu, Diana Haddad, Rich McVeigh, Bhanu Rajput, Barbara Robbertse, Brian Smith-White, Danso Ako-Adjei, et al. Reference sequence (refseq) database at ncbi: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*, 44(D1):D733–D745, 2016.
- [PFVdS19] Ana Paço, Renata Freitas, and Ana Vieira-da Silva. Conversion of dna sequences: From a transposable element to a tandem repeat or to a gene. *Genes*, 10(12):1014, 2019.
- [Pos82] John Raymond Postgate. *The fundamentals of nitrogen fixation*. CUP Archive, 1982.
- [RM10] Jessica S Reuter and David H Mathews. Rnastructure: software for rna secondary structure prediction and analysis. *BMC bioinformatics*, 11(1):1–9, 2010.
- [RMR⁺20] Luuk Rutten, Kana Miyata, Yuda Purwana Roswanjaya, Rik Huisman, Fengjiao Bu, Marijke Hartog, Sidney Linders, Robin van Velzen, Arjan van Zeijl, Ton Bisseling, et al. Duplication of symbiotic lysin motif receptors predates the evolution of nitrogen-fixing nodule symbiosis. *Plant Physiology*, 184(2):1004–1023, 2020.
- [RSM⁺02] Horst Röhrig, Jürgen Schmidt, Edvins Miklashevichs, Jeff Schell, and Michael John. Soybean enod40 encodes two peptides that bind to sucrose synthase. *Proceedings of the National Academy of Sciences*, 99(4):1915–1920, 2002.
- [Rut03] Tom Ruttink. *ENOD40 affects phytohormone cross-talk*. Wageningen University, 2003.

- [SCB⁺01] Christian Staehelin, Celine Charon, Thomas Boller, Martin Crespi, and Ádám Kondorosi. Medicago truncatula plants overexpressing the early nodulin gene enod40 exhibit accelerated mycorrhizal colonization and enhanced formation of arbuscules. *Proceedings of the National Academy of Sciences*, 98(26):15366–15371, 2001.
- [SHR⁺10] Cameron Smith, Steffen Heyne, Andreas S Richter, Sebastian Will, and Rolf Backofen. Freiburg rna tools: a web server integrating intarna, exparna and locarna. *Nucleic acids research*, 38(suppl_2):W373–W377, 2010.
- [SSM⁺95] Douglas E Soltis, Pamela S Soltis, David R Morgan, Susan M Swensen, Beth C Mullin, Julie M Dowd, and Peter G Martin. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proceedings of the National Academy of Sciences*, 92(7):2647–2651, 1995.
- [Tut12] Yusuf Tutar. Pseudogenes. *Comparative and functional genomics*, 2012, 2012.
- [Van01] Carroll P Vance. Symbiotic nitrogen fixation and phosphorus acquisition. plant nutrition in a world of declining renewable resources. *Plant physiology*, 127(2):390–397, 2001.
- [vdSPC⁺96] Karin van de Sande, Katharina Pawlowski, Inge Czaja, Ursula Wieneke, Jeff Schell, Jürgen Schmidt, Richard Walden, Martha Matvienko, Joan Wellink, Ab van Kammen, et al. Modification of phytohormone response by a peptide encoded by enod40 of legumes and a nonlegume. *Science*, 273(5273):370–373, 1996.
- [VGW02] Erika Varkonyi-Gasic and Derek William Richard White. The white clover enod40 gene family. expression patterns of two types of genes indicate a role in vascular function. *Plant physiology*, 129(3):1107–1118, 2002.
- [VKF⁺15] Tatiana Vernié, Jiyoung Kim, Lisa Frances, Yiliang Ding, Jongho Sun, Dian Guan, Andreas Niebel, Miriam L Gifford, Fernanda de Carvalho-Niebel, and Giles ED Oldroyd. The nin transcription factor coordinates diverse nodulation programs in different tissues of the medicago truncatula root. *The Plant Cell*, 27(12):3410–3424, 2015.
- [vVHB⁺18] Robin van Velzen, Rens Holmer, Fengjiao Bu, Luuk Rutten, Arjan van Zeijl, Wei Liu, Luca Santuari, Qingqin Cao, Trupti Sharma, Defeng Shen, et al. Comparative genomics of the nonlegume parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences*, 115(20):E4700–E4709, 2018.
- [WJH⁺12] Sebastian Will, Tejal Joshi, Ivo L Hofacker, Peter F Stadler, and Rolf Backofen. Locarna-p: accurate boundary prediction and improved detection of structural rnas. *Rna*, 18(5):900–914, 2012.
- [WN05] Dietrich Werner and William E Newton. *Nitrogen fixation in agriculture, forestry, ecology, and the environment*, volume 4. Springer Science & Business Media, 2005.

- [WQ06] B Wang and Y-L Qiu. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza*, 16(5):299–363, 2006.
- [WRH⁺07] Sebastian Will, Kristin Reiche, Ivo L Hofacker, Peter F Stadler, and Rolf Backofen. Inferring noncoding rna families and classes by means of genome-scale structure-based clustering. *PLoS Comput Biol*, 3(4):e65, 2007.
- [YH11] Keisuke Yokota and Makoto Hayashi. Function and evolution of nodulation genes in legumes. *Cellular and molecular life sciences*, 68(8):1341–1351, 2011.
- [YKH⁺93] Wei-Cai Yang, Panagiotis Katinakis, Peter Hendriks, Arie Smolders, Floris de Vries, Johan Spee, Ab van Kammen, Ton Bisseling, and Henk Franssen. Characterization of gmenod40, a gene showing novel patterns of cell-specific expression during soybean nodule development. *The Plant Journal*, 3(4):573–585, 1993.
- [YvVB⁺13] Mei-Qing Yang, Robin van Velzen, Freek T Bakker, Ali Sattarian, De-Zhu Li, and Ting-Shuang Yi. Molecular phylogenetics and character evolution of cannabaceae. *Taxon*, 62(3):473–485, 2013.

A A list of the whole genomes from Rosales species, which were used as subject sequences in BLAST

Species	Genome	Accession
<i>Rosa chinensis</i>	RchiOBHm-V2	GCF_002994745.2
<i>Rosa multiflora</i>	RMU_r2.0	GCA_002564525.1
<i>Rosa lucieae</i>	ILVO_Rwic.1.0	GCA_006954505.1
<i>Rosa x damascena</i>	ASM166254v1	GCA_001662545.1
<i>Fragaria iinumae</i>	FIL1.0	GCA_009720345.1
<i>Fragaria nilgerrensis</i>	ASM1013465v1	GCA_010134655.1
<i>Fragaria vesca</i>	FraVesHawaii.1.0	GCF_000184155.1
<i>Fragaria x ananassa</i>	FAN_r1.1	GCA_000511835.1
<i>Fragaria nubicola</i>	FNU_r1.1	GCA_000511995.1
<i>Fragaria nipponica</i>	FNI_r1.1	GCA_000512025.1
<i>Fragaria orientalis</i>	FOR_r1.1	GCA_000517285.1
<i>Prunus salicina</i>	SCAU_Psal.1.0	GCA_014863905.1
<i>Prunus persica</i>	Prunus_persica_NCBIV2	GCF_000346465.2
<i>Prunus dulcis</i>	ALMONDv2	GCF_902201215.1
<i>Prunus mume</i>	P.mume_V1.0	GCF_000346735.1
<i>Prunus yedoensis</i>	CYE_r3.1	GCA_005406145.1
<i>Prunus avium</i>	PAV_r1.0	GCF_002207925.1
<i>Malus baccata</i>	Malus_baccata_v1.0	GCA_006547085.1
<i>Pyrus betulifolia</i>	ASM784424v1	GCA_007844245.1
<i>Pyrus pyrifolia</i>	PPY_r1.0	GCA_016587475.1
<i>Pyrus x bretschneideri</i>	Pbr_v1.0	GCF_000315295.1
<i>Cydonia oblonga</i>	ASM1570837v1	GCA_015708375.1
<i>Geum urbanum</i>	G_urb.d1	GCA_900236755.1
<i>Purshia tridentata</i>	ASM325488v1	GCA_003254885.1
<i>Dryas drummondii</i>	ASM325486v1	GCA_003254865.1
<i>Ulmus americana</i>	ASM1001500v2	GCA_010015005.2
<i>Boehmeria nivea</i>	ASM1813214	GCA_018132145.1
<i>Ochetophila trinervis</i>	ASM325497v1	GCA_003254975.1
<i>Ziziphus jujuba</i>	ZizJuj_1.1	GCF_000826755.1
<i>Artocarpus camansi</i>	Acamansi1.0	GCA_002024485.1
<i>Cannabis sativa</i>	cs10	GCF_900626175.2
<i>Humulus lupulus var. lupulus</i>	hl_SW_version_1.0.fasta	GCA_000831365.1
<i>Rhamnella rubrinervis</i>	ASM784410v2	GCA_007844105.2
<i>Ficus carica</i>	UNIPLFiCari_1.0	GCA_009761775.1
<i>Ficus erecta</i>	FER_r1.1	GCA_008635985.1
<i>Morus alba</i>	ASM1206604v3	GCA_012066045.3
<i>Morus notabilis</i>	ASM41409v2	GCF_000414095.1
<i>Trema orientale</i>	TorRG33x02_asm01	GCA_002914845.1
<i>Parasponia andersonii</i>	PanWU01x14_asm01	GCA_002914805.1

B FASTA-format input for domains 4, 5, 6 and S in LocARNA

Domain 4:

```
>Pandersonii_domain_4
ggctatgtctcttttcttattataacttgtgtgctatatgttatgatttatgatcagtttagtt
>Orientale_domain_4
ggctatgtctcttattcttattataacttgtgtggttatatgttatgatttatgatcagtttagtt
>Malba_domain_4
ggctattttcaatatcatcccatcttttgttatgttttgttatgatgcc
>Mnotabilis_domain_4
ggctattttcaatatcatcccatcttttgttatgttttgttttgatgtcttagct
>Mindica_domain_4
ggctattttcaatatcatcccatcttttgttatgttttgttatgatgcc
>Ferecta_domain_4_1
cggctatttactatgttattttgtcatcccaatcttaatgtttgtgtttcactg
>Ferecta_domain_4_2
cggctatttactatgttattttgtcatcccaatcttaatgtttgtgtttcactg
>Rrubrinervis_domain_4
tatttcctatctgttatgattatacatgttaatgtagtttggttgcttgatagtgtaata
```

Domain 5:

```
>Pandersonii_domain_5
tttttttttgagaatgcaataataaaaagatttggtgtttcaagtgaataa
>Orientale_domain_5
tttttggtttgagaatgtaataataaaaaggtttgtggtttcaagcgaaaa
>Malba_domain_5
cttttgagattgttaatgtaataataaaaag
>Fcarica_domain_5
tttgttccttttgagattattgtaacctaataataaacagtaataataagaataaa
>Ferecta_domain_5_1
tttgttccttttgagattattgtaacctaataatagtaacagtaataataagaataaa
>Ferecta_domain_5_2
tttgttccttttgagattattgtaacccaataatagtaacagtaataataagaataaa
>Hlupulus_domain_5
gtttttttgctttgtgaatgtaataataaaaatggtttgctttcaagcaaaaagat
>Rrubrinervis_domain_5
gagtttggttttttaagcacaagttc
```

Domain 6:

```
>Pandersonii_domain_6
gagtggtgcatcagctggttatggttagtcgctc
>Orientale_domain_6
```

gagtgttgccatcagctgtgatgggtgctc
>Malba_domain_6
tttttgtttgtgagcaaaactcgtttgcttagggaaga
>Mnotabilis_domain_6
tttttgtctgtgagcaaaactcgtttgcttagggaaga
>Ferecta_domain_6_1
gctttgtgttttagttttgagcaaaaactggcttgcaaagt
>Ferecta_domain_6_2
gctttgtgttttagttttgagcaaaaactggcttgcaaagt
>Hlupulus_domain_6
gcatcatatcatctgtgatgc
>Csativa_domain_6
agcatcatcacataaaatctgtactgatgatgct
>Rrubrinervis_domain_6
tagaagcttagagggtttta

Domain S:

>Pandersonii_domain_S
acctattagattcaagaatcagattcctgattcttgaatctgataggt
>Orientale_domain_S
gcctatcagattcaagaatcagattcctgattcttgaatctgataggt
>Malba_domain_S
acctatcagattcttgattcttgaatcttgaatctgataaggt
>Mindica_domain_S
acctatcagattcttgattcttgaatcttgaatctgataaggt
>Mnotabilis_domain_S
acctatcagattcttggttcttgaatctgataaggt
>Fcarica_domain_S
cttatcagattcttgattctttatccttgaatctgatagg
>Ferecta_domain_S_1
cttatcagattcttgattctttatccttgaatctgatagg
>Ferecta_domain_S_2
cttatcagattcttgattctttatccttgaatctgatagg
>Acamansi_domain_S_1
cctatcagattcttgattcttgattcttgattctgattagg
>Acamansi_domain_S_2
ccaacatatcagattcttgattcttgattcttgaatctgatttcagg
>Hlupulus_var_lupulus_domain_S
accaatcagattcaagaataatattcttgattcttgaatctgataggt
>Csativa_domain_S
acctatcagattcaagaatcagatctttgattcttgaatctgataggt
>Rrubrinervis_domain_S
gcctatcagttttcttgaagctgataggt
>Zjujuba_domain_S
cctatcagttttcttgaatctgatagg

```
>Otrinervis_domain_S  
cctatcagatcttcttgaatctgatagg
```

C All newly annotated Rosales *ENOD40* homologues

Domain/Region	Color
Domain S	X
sORFI	X
Domain 1	X
Domain 2	X
Region II	X
Domain 3	X
Domain 4	X
Domain 5	X
Domain 6	X

Rosa chinensis:

NC_037091.1

(RANGE: 63743700 to 63744305 plus/minus)

```

1 atccctccac tgactcttta aggccccaat ggattcatag agctatgatc agtttctcca
61 atctgatagg atggaggca gattgggaag aacacatcca tgggcattaa aagcattggc
121 aatggagtga ggttatacct caacctcact ttaacagttc gctgagetc agtcactggc
181 tgctgctaac tacagcaact tcatatccag aggtagcaga gggttgcta gtgcttagca
241 aaccggaag tcacaaatcg gcatggactc acactctgtg atgagctctt ctcttctgta
301 atgttattat agtctagtcc tggcatgaca ctgtcatcca tgtgttctgt ttcaagtctt
361 ttgtgtgttt ttctttgttg tgtgtagtgt gttttatgag tcggttatgt cccttgtatt
421 cttcattaat tggtttatgc attcttcatt aattggttta tgctggagta gaatatggat
481 ttagaataaaa attgaaaact gaaatctctg ctactctgat caggatcaat caataatc
541 aattcaaaaa tgcaagtaaa actgaatgac aatgtatttt tgatatcttt ggaaccataa
601 atttac

```

Rosa multiflora:

BDJD01002838.1

(RANGE: 72696 to 73301 plus/plus)

```

1 agccagagag taatccctcc actgactctt taaggcccca atggattcat agagctatga
61 tcagtttctc caatctgata ggtatggagg cagattggga agaacacatc catgggcatt
121 aaaagcattg gcaatggagt gaggttatc ctcaacctca ctttaacagt tctgtgagc
181 tcagtcactg gctgttgcta actacagcaa cttcatatcc agaggtagca gagggttggc
241 tagtgcttag caaaccgga agtcacaaat cggcatggac tcacactctg tgatgagtct
301 tttctcttct gtaatgttat tctagtctag tcttggcata acactgtcat ccattgtgtc
361 tgtttcaagt cttttgtgtg tttttctttg ttctgtgtag tgtgttttat gactcgggta
421 tgtcccttgt attcttcatt aattggttta tgctggagta gaatatggat ttagaataaa
481 attgaaaact gaaatctcgc tactctgac aggatcaatc aataatatca attcaaaaat
541 gcaagtaaaa ctgaatgaca atgtattttt gatatctttg gaaccttaaa tttacatacg
601 tgttca

```

Rosa luciae:

RQIQ01004170.1

(RANGE: 2466 to 3071 plus/plus)

```
1 agccagagag taatccctcc actgactcct taaggcccca atggattcat agagctatga
61 tcagtttctc caatctgata ggtatggagg cagattggga agaacacatc catgggcatt
121 aaagcattg gcaatggagt gaggttatac ctcaacctca ctttaacagt tgcgtctgagc
181 tcagtcactg gctgctgcta actacagcaa ctccatattc agaggtagca gagggttggc
241 tagtgcttag caaacggca agtcacaaaat cggcatggac tcacactctg tgatgagtct
301 tttctcttct gtaatgttat tctagtctct gcatgacact gtcattccatg tgttctgttt
361 caagtctttt gtgtgttttt ctttgttgtg tgtagtgtgt tttatgagtc ggttatgtcg
421 cttgtattct tcattaattg gtttatgctg gactagaata tggatttaga ataaaattga
481 aactgaaat ctctgctact ctgatcagga tcaatcaata atatcaattc aaaaatgcaa
541 gcaaaactga atgacaatgt atttttgata tctttggaac cataaattta catacgtggt
601 cactct
```

Rosa x damascena:

LYNE01000270.1

(RANGE: 37590 to 38195 plus/plus)

```
1 agccagagag taatccctcc actgactcct taaggcccca atggattcat agagctatga
61 tcagtttctc caatctgata ggtatggagg cagattggga agaacacatc catgggcatt
121 aaagcattg gcaatggagt gaggttatac ctcaacctca ctttaacagt tgcgtctgagc
181 tcagtcactg gctgctgcta actacagcaa ctccatattc agaggtagca gagggttggc
241 tagtgcttag caaacggca agtcacaaaat cggcatggac tcacactctg tgatgagtct
301 ttctcttctg taatgttatt atagtctagt cctggcatga cactgtcatc catgtgttct
361 gtttcaagtc ttttgtgtgt ttttctttgt tgtgtgtagt gtgttttatg agtcggttat
421 gtcccttgta ttcttcatta attggtttat gcattcttca ttaattgggt tatgctggag
481 tagaatatgg atttagaata aaattgaaaa ctgaaatctc tgctactctg atcaggatca
541 atcaataata tcaattcaaa aatgcaagta aaactgaaat acaatgtatt tttgatattc
601 ttgaa
```

Fragaria iinumae:

CM019120.1

(RANGE: 32931258 to 32931863 plus/minus)

```
1 agcctcagag taatccctcc actgaatcct taaggcccca atggattcat agagctatga
61 tcagtttctc caatctgata ggtattgaag cagattggga agaacacatc cattggcctt
121 aaagcattg gcaatggagt gagccgggta gatcacacct caacctcact ccaatagttt
181 gtctgagctc agtcacctgc tgctgctaac tacagaaact tcatatccag aggtagcaga
241 aggtttggct agcgcttagc aaacggcaaa gtcacaaaat cggcatggac tcacactctg
301 tgatgagtct tttatcttct gtaatattat tctactctag tcctgtcatg acaatgtaat
361 cgatgtgtat ctgtttcaag tatgtgtttt tctttttcgt gtgtagtgtg ttttattgac
421 tcttttgttc aataattggt ttatgggtga acgaaatatt gatttagaat aaaaatcgaaa
481 atatatatga atcaggatca atcgattaca tcaaattcaa aattgcaaat aaaaaagaat
541 gacattgtat ttttgatatt tttggtacca aacatgtaca tacatagctt ccctctgcat
601 aactag
```


Fragaria nilgerrensis:

CM020980.1

(RANGE: 34979666 to 34980271 plus/minus)

```
1 agcctcagag taatccctcc actgaatctt taaggtccca atggattcat agagctatga
61 tcagtttctc caatctgata ggtattgaag cagattggga agaacacatc cattggcctt
121 aaaagcattg gcaatggagt gagccagtga gatcatacct caacctcact ccaatagttt
181 gcctgagctc agtcacctgc tgctgctaac tacagaaact tcatatccaa aggtagcaga
241 aggtttgct agcgcttagc aaaccggcaa gtcacaaaat cggcatggac tcacactctg
301 tgatgagtct tttatcttct gtaatattat tctactctag tcctgtcatg acaatgtaat
361 cgatgtgat ctgtttaagt atgtgtttt ctttttggtg tgcctgtgtg tttattgagt
421 cttttgttca ataattggtt tatggtggaa tgaactgttg atttagaata aaatcgaaaa
481 tatatatgaa tcaggatcaa tcgattacat caaattcaaa attgcaata aaaaagaatg
541 acattgtatt tttgatatct ttggtaccaa acatgtacag acatacgttc actctgcata
601 actagc
```

Fragaria vesca:

NC_020494.1

(RANGE: 18831103 to 18831708 plus/minus)

```
1 agcctcagag taatccctcc actgaatctt taaggtccca atggattcat agagctatga
61 tcagtttctc caatctgata ggtattgaag cagattggga agaacacatc cattggcctt
121 aaaagcattg gcaatggagt gagccagtga gatcatacct caacctcact ccaatagttt
181 gcctgagctc agtcacctgc tgctgctaac tacagaaact tcatatccag aggtagcaga
241 aggtttgct agcgcttagc aaaccggcaa gtcacaaaat cggcatggac tcacactctg
301 tgatgagtct tttatcttct gtaatattat tctactctag tcctggcatg acaacgtaat
361 cgatgtgat ctgtttaagt atgtgtttt ctttttggtg tgcctgtgtg tttattgagt
421 cttttgttca ataattggtt tatggtggaa tgaactgttg atttagaata aaatcgaaaa
481 tatatatgaa tcaggatcaa tcgattacat caaattcaaa attgcaata aaaaagaatg
541 acattgtatt tttgatatct ttggtaccaa acatgtacag acatacgttc actctgcata
601 actagc
```

Fragaria x ananassa:

BATT01307848.1

(RANGE: 845 to 1450 plus/minus)

```
1 agcctcagag taatccctcc actgaatctt taaggtccca atggattcat agagctatga
61 tcagtttctc caatctgata ggtattgaag cagattggga agaacacatc cattggcctt
121 aaaagcattg gcaatggagt gagccagtga gatcatacct caacctcact ccaatagttt
181 gcctgagctc agtcacctgc tgctgctaac tacagaaact tcatatccag aggtagcaga
241 aggtttgct agcgcttagc aaaccggcaa gtcacaaaat cggcatggac tcacactctg
301 tgatgagtct tttatcttct gtaatattat tctactctag tcctggcatg acaacgtaat
361 cgatgtgat ctgtttaagt atgtgtttt ctttttggtg tgcctgtgtg tttattgagt
421 cttttgttca ataattggtt tatggtggaa tgaactgttg atttagaata aaatcgaaaa
481 tatatatgaa tcaggatcaa tcgattacat caaattcaaa attgcaata aaaaagaatg
541 acattgtatt tttgatatct ttggtaccaa acatgtacag acatacgttc actctgcata
601 actagc
```

BATT01197499.1

(RANGE: 1192 to 1797 plus/plus)

```
1 agcctcagag taatccctcc actgaatcct taaggtcca atggattcat agagctatga
61 tcagtttctc caatctgata ggtattgaag cagattggga agaacacatc cattggcctt
121 aaaagcattg gcaatggagt gagccagtaa gatcatacct caacctcact ccaatagttt
181 gtctgagctc agtcacctgc tgctgctaac tacagaaact tcatatccag aggtagcaga
241 aggtttgct agcgcttagc aaaccggcaa gtcacaaaat cggcatggac tcacactctg
301 tgatgagtct tttatcctt gtaatattat tctactctag tctgtcatg acaatgtaat
361 cgatgtgtat ctgtttcaag tatgtgtttt tctttgtcgt gtgtagtggt ttttatgttc
421 aataattggt ttaggggtga atgaaatatt gatttggact aaaatcgaaa atatatatga
481 atcaggatca atcgattaca tcaaattcta aattgcaaat aaaaaagaat gacattgtat
541 ttttgatc tttggtacca aacatgtaca tcatacgttc actctgcata actagcacc
601 tacata
```

Fragaria nubicola:

BATW01062352.1

(RANGE: 2891 to 3496 plus/plus)

```
1 agcctcagag tagtccctcc actgaatcct taaggtcca atggattcat agagctatga
61 tcagtttctc caatctgata ggtattgaag cagattggga agaacacatc cattggcctt
121 aaaagcattg gcaatggagt gagccagtga gatcatacct caacctcact ccaatagttt
181 gtctgagctc agtcacctgc tgctgctaac tacagaaact tcatatccag aggtagcaga
241 aggtttgct agcgcttagc aaaccggcaa gtcacaaaat cggcatggac tcacactctg
301 tgatgagtct tccttttacc ttctgtaata ttattctact ctagtctctg catgacaatg
361 taatcgatgt gtatctgttt caagtatgtg tttttctttt tgggtgtgtg tgtgttttat
421 tgagtctttt gttcaataat tggtttatgg tggaatgaaa tattgattta gaataaaatc
481 gaaaatatat atgaatcagg atcaatctat tacatcaaat tcaaaattgc aaataaaaaa
541 gaatgacatt gtatTTTTga tacctttggt accaaacatg tacatacata cgttcaactc
601 gcataa
```

Fragaria nipponica:

BATV01076376.1

(RANGE: 10981 to 11586 plus/minus)

```
1 agcctcagag taatccctcc actgaatcct taaggtcca atggattcat agagctatga
61 tcagtttctc caatctgata ggtattgaag cagattggga agaacacatc cattggcctt
121 aaaagcattg gcaatggagt gagccagtga gatcatacct caacctcact ccaatagttt
181 gtctgagctc agtcacctgc tgctgctaac tacagaaact tcatatccag aggtagcaga
241 aggtttgct agcgcttagc aaaccggcaa gtcacaaaat cggcatggac tcacactctg
301 tgatgagtct tttatcatct gtaatattat tctactctag tctgtcatg acaatgtaat
361 cgatgtgtgt ctgtttcaag tatgtgtttt tctttttcgt gtgtagtggt ttttattgac
421 tcttttggtc aataattggt ttatgggtga atgaaatatt gatttagaat aaaatcgaaa
481 atatatatga atcaggatca atcgattaca tcaaatttaa aattgtaaat aaagaagaat
541 gactttgtat ttttgatc tttggtacca aacatgtaca tacatacgtt cactctgcac
601 aactag
```

Prunus salicina:

CM026527.1

(RANGE: 25129424 to 25130029 plus/minus)

```
1 attttaagtg cccattggat tggaaatcat atcattagag caacctatca gtttagtttc
61 tccaaatctt gatgggtgcg gaaacagatt ggcaagaaga gccaatccat gggacttaaa
121 agcatggaca tggagtgagg tttgtctcaa acctctcca aaactcaaca gttcgtctga
181 gctcagtcaa tggctgctaa ctacagcaag gcaatgatct agaggtagca gaggtttggc
241 tagtgcttgg cgaacggca agtcacaaat cggcaaccgg ctcagtcacg agtc ttttaa
301 ttttaaggata tctttttatt ctggtatggt attttgagtt cttgcatgag tgtaataata
361 caagtgtttg tggttttaat atgatgattc tcttctctt atcagttcag ttgctgcaaa
421 acgtgttgag taaaaactga aatatgaaga aagactttga ggtacaattt gaatcgtttc
481 cttgaagtgt tatttgccct cacttaattg agtttgctaa agggttcttg acaaaccatt
541 tctaggatac aacaaagtct ggaatctgtg acccctactt ttttttaatc tcttaataag
601 ggagat
```

WERZ01000077.1

(RANGE: 202610 to 203215 plus/minus)

```
1 attttaagtg cccattggat tggaaatcat atcattagag caacctatca gtttagtttc
61 tccaaatctt gatgggtgcg gaaacagatt ggcaagaaga gccaatccat gggacttaaa
121 agcatggaca tggagtgagg tttgtctcaa acctctcca aaactcaaca gttcgtctga
181 gctcagtcaa tggctgctaa ctacagcaag gcaatgatct agaggtagca gaggtttggc
241 tagtgcttgg cgaacggca agtcacaaat cggcaaccgg ctcagtcacg agtc ttttaa
301 ttttaaggata tctttttatt ctggtatggt attttgagtt cttgcatgag tgtaataata
361 caagtgtttg tggttttaat atgatgattc tcttctctt atcagttcag ttgctgcaaa
421 acgtgttgag taaaaactga aatatgaaga aagactttga ggtacaattt gaatcgtttc
481 cttgaagtgt tatttgccct cacttaattg agtttgctaa agggttcttg acaaaccatt
541 tctaggatac aacaaagtct ggaatctgtg acccctactt ttttttaatc tcttaataag
601 ggagat
```

Prunus persica:

NC_034009.1

(RANGE: 26194291 to 26194896 plus/plus)

```
1 ttttaagtgc ccattggatt gaaatcata tcattagagc aacctatcag tttagtttct
61 ccaaactctt atgggtgtgg aaacagattg gcaagaagag ccaatccatg ggaacttaaaa
121 gcatggacat ggagtgagg tttgtctcaa cctctctcaa aactcaacag ttcgtctgag
181 ctcagtcagt ggctgctaac tacagcaagg caatgatcta gaggtagcag aggtttggct
241 attgcttggc gaacggcaa gtcacaaatc ggcaaccggc tcagtcatga gtc ttttaat
301 ttaaggatat ctttcaattc tgttatgtta ttttgagttc ttgcatgaga atgtaataat
361 acaagtgttt gtggttttaa tgtgatgatt ctcttctctt tatcagttca gttgctgcaa
421 aacgtgttga gtaaaaactg aatatgaag aaagactttg aggtaaaatt tgaatcgttt
481 tcttaagta ttatttgctt ccaacttaatt gagtttgcta aagggttctt ggcaaacctac
541 tttcaggata caacaaagtc tggaaatctgc aatatgcaa aactcgaaga attcccttag
601 aaattg
```

Prunus dulcis:

NC_047650.1

(RANGE: 22781358 to 22781963 plus/plus)

```
1 ttttaagtgc ccattggatt ggaaatcata tcattagagc aacctatcag tttagtttct
61 ccaaatcttg atgggtgtgg aaacagattg gcaagaagag ccaatccatg ggacttaaaa
121 gcagggacat ggagtgaggt ttgtctcaaa acctctcca aaactcaaca gttcgtctga
181 gctcagtcaa tggctgctaa ctacagcaag gcaatgatct agaggtagca gaggtttggc
241 tattgcttgg cgaaccggca agtcacaaat cggcaaccga ctcagtcatg agtcttttaa
301 ttttaaggata tctttcaatt ctggtatggt attttgagtt cttgcatgag aatgtaataa
361 tacaagtgtt tgtggtttta atatgatgat tctcttctct ttatcagttc agttgctgca
421 aaacgtgttg agtaaaaact gaaatatgaa gaaagacttt gaggtacaat ttgaatcggt
481 tccttaaagt gttatttgcc cccacttaat tgagtttgct aaagggttct tggcaaacca
541 cttccaagt acagcaaagt ctggaatctg cgaatagcaa aaactcgaag aattccotta
601 gaaatt
```

Prunus mume:

NC_024127.1

(RANGE: 18628683 to 18629288 plus/minus)

```
1 attttaagtgc ccattggat tggaaatcat atcattagag caacctatca gtttagttgc
61 tccaaatctt gatgggtgcg gaaacagatt ggcaagaaga gccaatccat gggacttaaa
121 agcagggaca tggagtgagg ttgtctcaa acctctcca aaactcaaca gttcgtctga
181 gctcagtcaa tggctgctaa ctacagcaag gcaatgatct agaggtagca gaggtttggc
241 tagtgctttg cgaaccggca agtcacaaat cggcaaccgg ctcagtcacg agtctttaaa
301 ttttaaggata tccttctatt ctggtatggt attttgagtt cttgcattag aaagtaataa
361 tacaagtgtt tgtggtttta atctgatgat tctcttctct atatcagttc agttgctgca
421 aaacgtgttg agtaaaaact gaaatatgaa gaaagacttt gaggtacaat ttgaatcggt
481 tccttaaagt gttatttgcc tccacttaat tgagtttgct aaaaggttct tggcaaacca
541 cttctaggat acaacaaagt ctggaatttg cgaatagcaa aaactcgaag aattccotta
601 gaaatt
```

Prunus yedoensis:

BJCG01002293.1

(RANGE: 2947914 to 2948519 plus/minus)

```
1 ttttaagtgc ccattggatt ggaaatcata tccttagagc aacctatcag tttagtttct
61 ccaaatcttg atgggtgtgg aaacagattg gcaagaagag ccaatccatg ggacttaaaa
121 gcagggacat ggagtgaggt ttgtctcaaa cctctctcaa cactcaacag ttcgtctgag
181 ctcagtcaat ggctgctaac tacagcaagg caatgatcaa gaggtagcag aggtttggct
241 agtgcttggc gaaccggcaa gtcacaaatc ggcaaacggc tcagtcatga gtctttttaat
301 ttaaggatat ctttctattc tgtaatgta ttttgagttc ttgcatgaga atgtaataat
361 acaagtgttt gtggttttta tctcatgatt ctcttctctt tatcagttca gttgctgcaa
421 aacgtgttga gtaaaaaaac ttttgaggta gaatttgaat cgtttcctta aagtgttatt
481 tgcccctact taattgagtt tgctaattgg ttattggcaa accacttcta ggatacaaca
541 aagtctggaa tatgtgagta gcaaaaactc gaagaattcc cttagaaatt gctgtaagga
601 aattgt
```

BJCG01000003.1

(RANGE: 1939320 to 1939925 plus/plus)

```
1 ttttaagtgc ccattggatt gaaatcata tcattagagc aacctatcag tttagtttca
61 ccaaatcttg atgggtgtgg aaacagattg gcaagaagag ccaatccatg ggacttaaaa
121 gcatggacat ggagtgaggt ttgtctcaaa cctcctcaa aactcaacag ttcgtctgag
181 ctcagtcaat ggctgctaac tatagcaagg caatgatcaa gaggtagcag aggtttggct
241 agtgcttggc gaaccggcaa gtcacaaatc ggcaaacggc tcagtcatga gtctttcaat
301 ttaaggatat ctttctattc tgtaatgtta ttttgagttc ttgcatgaga atgtaataat
361 acaagtgttt gtggttttaa tctcatgatt ctcttctctt tatcagttca gttgctgcaa
421 aacgtgttga gtaaaaaatc ttttgaggta gaatttgaat cgtttcctta aagttttatt
481 tgccccact taattgagtt tgctaattggg ttcttgca accacttcta ggatacaaca
541 aagctggaa tatgtgagca gcaaaaactc gaagaattcc cttagaagtt gctgtaaggaa
601 aattgt
```

BJCG01003116.1

(RANGE: 45799 to 46404 plus/minus)

```
1 ttttaagtgc ccattggatt gaaatcata tccttagagc aacctatcag tttagtttct
61 ccaaatcttg atgggtgtgg aaacagattg gcaagaagag ccaatccatg ggacttaaaa
121 gcatggacat ggagtgaggt ttgtctcaaa cctcctcaa cactcaacag ttcgtctgag
181 ctcagtcaat ggctgctaac tacagcaagg caatgatcaa gaggtagcag aggtttggct
241 agtgcttggc gaaccggcaa gtcacaaatc ggcaaacggc tcagtcatga gtcttttaat
301 ttaaggatat ctttctattc tgtaatgtta ttttgagttc ttgcatgaga atgtaataat
361 acaagtgttt gtggtttact catgattctc ttctcttct atcagttcag ttgctgcaaa
421 acgtgttgag taaaaaaact tttgaggtag aatttgaatc gtttccttaa agtgttattt
481 gccctactt aattgagttt gctaattggg tattggcaaa ccacttctag gatacaacaa
541 agtctggaat atgtgagtag caaaaactcg aagaattccc ttagaaattg ctgtaaggaa
601 attgtg
```

Prunus avium:

NW_018921537.1

(RANGE: 513075 to 513680 plus/minus)

```
1 ttttaagtgc ccattggatt gaaatcata tcattagagc aacctatcag tttagtttct
61 ccaaatcttg atgggtgtgg aaacagattg gcaagaagag ccaatccatg ggacttaaaa
121 gcatggacat ggagtgaggt ttgtctcaaa cctcctcaa aactcaacag ttcgtctgag
181 ctcagtcaat ggctgctaac tacagcaagg caatgatcaa gaggtagcag aggtttggct
241 agtgcttggc gaaccggcaa gtcacaaatc ggcaaacggc tcagtcatga gtcttttat
301 ttaaggatat ctttctattc tgtaatgtta ttttgagttc ttgcatgaga atgtaataat
361 acaagtgttt gtggttttaa tctcatgatt ctcttctctt tatcagttca gttgctgcaa
421 aacgtgttga gtaaaaaaac ttttgaggta gaatttgaat cgtttcctta aagagttatt
481 tgccccagt taattgagtt tgctaattggg ttcttgca accacttcta ggatacaaca
541 aagtctggaa tcggtgagta gcaaaaacac taagaattcc cttagaatt gctataaggaa
601 aattgt
```

Pyrus betulifolia:

CM017599.1

(RANGE: 5571726 to 5572331 plus/minus)

```
1 gatgggttgt caagttttca actaaagaac atatccctca cttgctgacc aagtatttaa
61 aggcaatagc cagcgagtta atcttocact ggttttaaga cccattatc caaaatatgg
121 ggcttaaaag cattggacat ggagtgagga tttgcctaaa acctcctcca aaacttgaac
181 agttagtctg agctcagtc atagctgctg ctaactacac caagcgcaat tatcacacat
241 actgatcata tgatctggag gtagcagagg tttgggctcg tgcttagcga accggaagt
301 cacaaatcgg caacggactc agtcatgagt ccctatggct ctcttttttg ttctgtaatg
361 tttatttaag ttccagtatg aaatgtaata aaacaagtgt tggtagtatt ttgggcagac
421 ggtagcact cttctgttta tcctttcatc tgctgcaaag tgagtaacac ctttggtttg
481 tctcccttct gtgaatgtac gtgttgaat ttatttttct tatgggtgta tgataaattt
541 tagttattag ttatgcagaa attggaaaaa tggtagttat tatttgttct ctttttttag
601 cgcttt
```

Pyrus pyrifolia:

BNSU0100004.1

(RANGE: 12849302 to 12849907 plus/plus)

```
1 gatgggttgt caagttttca actaaagagc atatccctca cttgctgacc aagtatttaa
61 agcaatagc cagcgagtta atcttocact ggttttaaga cccattatc caaaatatgg
121 ggcttaaaag cattggacat ggagtgagga tttgcctaaa acctcctcca aaacttgaac
181 agttagtctg agctcagtc atagctgctg ctaactacac caagtgcaat tatcatacat
241 actgatcata tgatctggag gtagcagagg tttgggctcg tgcttagcga accggaagt
301 cacaaatcgg caacggactc agtcatgagt ccctatggct ctcttttttg ttctgtaatg
361 tttatttaag taccagtatg aaatgtaata aaacaagtgt tggtagtatt ttgggcagac
421 ggtagcact cttctgttta tcctttcatc tgctgcaaag tgagtaacac ctttggtttg
481 tctcccttct tgtgaatgta cgtgttggaa tttatttttc ttatgggtg atgataaatt
541 ttatttatta gttatgcaga aattggaaaa atggtagata ttatttgttc tcttttttta
601 gcgctt
```

Cydonia oblonga:

JADOBS010303350.1

(RANGE: 3444 to 4049 plus/plus)

```
1 tttcattaaa atgaataata tcgcgaactt ttcgttaaat gccctgattt ttaagttagt
61 ggaggagttag atgggttgtc aagttttcaa ctacagagca tatcactcac ttgcagatca
121 tgtattttaa gccaatagcc agtgagttaa tcctccact gattttaaga cccattatc
181 caaaatatgg gacttaaaag cattggacat ggagtgagga tttgcctcaa acctcctcca
241 aaactcaaac agttagtctg agctcagtc atggctgctg ctaactacac caagtgcaat
301 tatcatacat actgatcata tgatctggag gtagcagagg tttgggctcg tgcttagcga
361 accggaagt cacaaatcgg caacggactc agtcatgagt ccctatggct ctctttctgt
421 tctgtaatgt ttatttaagt tccagtatga aacgtaataa aacaagtgtt ggtagtattt
481 tgagcagacg gttagcactc ttctgtttat ctgttcatct gctgcaaaat gagtaacacc
541 tttggtttgt ctcccttctt gtgaatgtac gtgttgaat ttatttttct taaggtggga
601 tgataa
```

Geum urbanum:

OEJZ01132735.1

(RANGE: 7949 to 8554 plus/plus)

```
1 ccagagagta ttgctccac tggctcttct tgaggcccca atggatttat agagatatga
61 tcaatttctc caatctgatg ggaatggaag aagattggga agaacacatc catggggact
121 taaaggcaat ggcaatagac cgagggttata cctcaacctc agtccaacag tttgttagag
181 ctcagtcaat gtgctgctgc taactacagc aacatatcca gaggtaccag aaggattggc
241 taggcttata aaccggaag tcacaaaatc agcatggact cacactagt atgagtcatt
301 ttccttctgc aatgttactc taatctagt t atgcatgaa catgtaatcc agtattttgtg
361 ttcctagtct tcatgtgtga ttgtttgtgg taagtactaa gtagtgatgt atgtgtttgt
421 tttgttgtgt tttcatcaat tggtttatgc tgggatatga ttcagagtaa aatttatatc
481 tatgcgagca tctgaaagcc taacatggat caatctgtaa ttctgtatca ttcattcaaa
541 aatttcaaat aaatatgaat tacattatat tttttatatt ggaacccaaa atgtacaaca
601 aagata
```

OEJZ01153179.1

(RANGE: 12431 to 13036 plus/plus)

```
1 ccagagagta ttgcctccat tggctcttct caaggcccca atggatctat cgagatatga
61 acagtttctc cattctgatg ggaatggaag caagatcgag aagaacacat ccatggggcc
121 ttaaaagcaa tagcaatagt atgaggttat acctcaactc actccaacag tttgtttgag
181 ttcagtcaat gtgctgctgc taactacagc aacatatcca gaggtaccaa aaggattgga
241 tagggcttac aaaccggaaa gtcataaaaat cggcattgac tcacactagt gacgagtcat
301 tttccttctg aaatgtttac tctaacttag ttttggcatg aaaatgtaat ccagtatctg
361 tgttcctagt cttcatgtgt ggtaagtact gagtagtgat gtatgtgttt gttttgttct
421 ttgtattttt catcaattgg tttatgctgg gatatgattc agaataaaat ttatatctat
481 gcgcgcgtct gaaagcttaa catggatcaa tctgtatcaa tcattcaaaa ttttcaata
541 aatatgaatt acattatatt tttgatattg gaaccaaaca tgtacaacaa agatatatcc
601 actctg
```

Purshia tridentata:

QANT01007693.1

(RANGE: 3214 to 3819 plus/plus)

```
1 gcctccactg ctattttgag aggccattgg atatcatgta gaccaaccaa cccattagtt
61 tcgccaact gatgggtgtg gcagcagatt ggaaaaagaa ccattattc atatatgggg
121 actttagcat tggaacgga gtgaggtgtc cctgaacctc ctccaaaact caacagcttg
181 cttgagcttt agtcattggc tccgctctac tacagaaaac agcgtgcac tagaggtata
241 gcagaggttt ggctagtget tagcgaaccg gcaagtcaca caacggcaat ggactcaggc
301 tcagtcatga gtctttttatg gctctatctt tctttctctt ctcttctgta atggtattgt
361 acttcttgca tgagaatgta ataccagtgt atgtgtttgt aagcagaacc ttagcagact
421 ttgacaaagc aagtgttttg atgatactcg atttgcagat gatgagcaat cttctgctct
481 tgctgcaaaa tgagtaacat cttgctctgt atttgtctgt gagcatatgt gggttggggg
541 gtgtgttttg tgtgtgcatg tacacatcta tgtgctgca aatttgtgtt gtgtgtgtgc
601 aattg
```

Malus baccata:

VIEB01000769.1

(RANGE: 466334 to 466939 plus/minus)

```
1 atgggttgtc aagttttcaa ctacagagca tatccctcac ttgctgatca agtattttaa
61 gccaatagcc agcgggttaa tctctccact gattttaaga cccattatc caaaacatgg
121 ggcttaaaag cattggacat ggagtgagga tttgcccaa acctcctcca aaactcaaac
181 agtttagctg agctcagtc atggctgctg caactacacc aagtgcaatt atcatacata
241 ctgatcatat gatctggagg tagcagaggt ttgggctcgt gcttagcgaa cgggcaagtc
301 acaaatcggc aacggactca gtcatgagtc cctatggctc tcttttttgt tctgtaatgt
361 ttatataagt tccagtatga aatgtaataa aacaagtgtt ggtagtattt tgagcagacg
421 gttgacactc ttctgtttat ctgtttatct gctgcaaaat gagtaacacc tttggtttta
481 ctcccttctt gtgaatgtac gtgttggaaat ttatttttct tatgggtgtga tgataaattt
541 tatttattag ttatgcagaa attggaaaaa tggtagatat tatttgttct cttttttttg
601 gcgctt
```

Dryas drummondii:

QANW01002007.1

(RANGE: 1510947 to 1511552 plus/plus)

```
1 atatcatgta gaccaaccaa cccattagtt tttccagac tgatgggtgt ggaagcagat
61 tggtcaaaaga accattatt catatatggg gactttaaag cattggacat ggagtgagggt
121 gtccctcaac ctcctccaaa actcaacagc ttgtctgagc tttagtcatt ggctccgctc
181 tactacagaa agcagcttgc atctagaggt atagcagagg tttggtagt gcttagcgaa
241 cgggcaagtc acacaacggc aatggactca ggctcagtc tgagtctttt atggctctat
301 ctttctttcc cttctcttct gtaatgttat tgtacttctt gcatgagaat gtaataaccag
361 tgtttggtt tgtaagcaga accttcacag actttgacaa agcaagcgtt ttgatgatac
421 tcgatttgca gatgatgagc aatcttctgt tcatcagttg ctgcaaaatg agtaatatct
481 tgctctgtat ttgtctgtga gcatatgtgg gttaggggtg gtgtttttgtg tgtgcatgta
541 cacatctatg tgcgtgcaaa tttgtgtttg gtgtgtgtgc aaattgggtgc ttgtatcatc
601 caccta
```

Ulmus americana:

WUAT02001391.1

(RANGE: 29663 to 30268 plus/minus)

```
1 aaacagatga tttatcaact tcaggttcta tataaagctg tatagccaag aagtgatttc
61 ccctctagct gtattgagag ataatttggg gtataagcat ggaaacagtg tggtggaag
121 aatctctcca tggctcttaa agcaggaatt ggaggtggag ggcaagaatt ctctccaat
181 tatcaaaaag ttgtgagttt tagccatggg ctcagttaaa cgcagcgtcg tttgaagtgt
241 gttttggcta aaggettacc aaaccggcaa gtcacagaac ggcaattggg ctcagtcacg
301 agtctttctg gctatccctt ttttttaatt ttttgttatt attatctctt ttgtgtcatg
361 ttttagtttg tttttagtgt tctttagtatt cttgttcttg cttttattat gtaataaagt
421 ttgtgatga gctattgcag gcttttgcca ggctcgattt attgataatt tactctatca
481 gtgttgaaaa atatgtgaat attaagattc tatcatctat ttgtgttggc cttgttatat
541 tgagttgacc gtgattcttt aatgggtggag aaagagttga aattccatga aaattgttta
601 tgctta
```


Boehmeria nivea:

PHNS01007792.1

(RANGE: 999292 to 999897 plus/plus)

```
1 ttctggcaaa tactaaacaa acccatcaat ggccttttaa aataacacag cacccaaaaa
61 taatggagat tggagatgag gtatttacac gactcggac acacacacaa catatcctct
121 aatcctcctc cagatttcta tagtttgtct gagtttaagt ttaaagccat aatattaggt
181 ctctgatgag agttgggcat gaagaagaag aagaagaaag aagaggcttt ttgtttggct
241 tttctctgaa ggctttggca aacggcaag tcacgaaacg gcaatggact cagaaatgag
301 tctttctcgg ctaatccatc tttaaaattt tctatgtact tttgttgttg ttgttgttcg
361 ctgtttagtt ttagtacttg ttgctttttgc agattaatgt aacaataaag aaaactgtat
421 tatgttgcac ggaaaaaaaa aaaagcgttt gtgttactca aatctttttt ttttttttat
481 tttttaaata caaattacca gtttctggga attgttttga cataacggtt tttttttttt
541 tcccctgtaa gaaactaaca cgtgctaaca cgtggggatg ctctggcaca cgtgtaatac
601 gtatta
```

Parasponia andersonii:

JXTB01000342.1

(RANGE: 184352 to 184957 plus/minus)

```
1 atatataaca acctattaga ttcaagaatc agattcctga ttcttgaatc tgataggtat
61 ggtaatggaa tcagcttgca aaggagaatc cattaatggg gtttaggaca gcaattggag
121 aaggagtgaa gggccccctc ttggcctcat ccttctatcc agatttccaa cagtttgtct
181 gagctttgag ccattagecg ggtatgaaga ggaactctc ttgaatgcaa aagaggagct
241 ttttgcctta aggctcagca aacggcaag tcacaaaacg gcaatggact cagtcacgag
301 tcatttcacg gctatgtctc tttcttatta taacttgtgt gctatatgtt atgattttatg
361 atcagtttag ttagtttttt ttttgagaat gcaataataa aagattttgtg tttcaagtga
421 aaagtgagtg ttgccatcag ctgttatggt agtcgctcag tatctttttt ctcttgtttg
481 ttcaccaatt taagtttacc tatttctcct gcaaatgat ttttttagat gaaattcgac
541 agaacgagtg aaattaaatt aagagaggaa agaaaaaag atgaaccaat tcgatggcaa
601 aaattc
```

Trema orientale:

JXTC01000021.1

(RANGE: 920092 to 920697 plus/minus)

```
1 atatataaca gcctatcaga ttcaagaatc agattcctga ttcttgaatc tgataggtat
61 ggtaatggaa tcagcttgca aaggagaatc cattaatggg gtttaggacag caattggaga
121 aggagtgaag ggtccccctc ttggcctcat ctctatcca gatttccaac agtttgtccg
181 agctttgagc cattagcagg gtatgaagag ggaactctct tgaatgcaa agaggagctt
241 tttggcttta aggcttagca aacggcaag tcacaaaacg gcaatggact cagtcacgag
301 tcatttcacg gctatgtctc ttatcttatt ataacttgtg tgttatatgt tatgattttat
361 gatcagttta gttagttttt gttttgagaa tgtaataata aaaggtttgt gtttcaagcg
421 aaagtgagtg gttgccatca gctgtgatgg tggcgcctca atatattttt gctcttgttg
481 gttaccaat ttaaagttta cctctttctt ctgcaaattg actttataca acatttttct
541 cctatctctc tctctctctc agaagaaatg tatgctggca cgtgtttatg ctcaggcgca
601 cgtata
```

Morus alba:

CP050237.2

(RANGE: 2644239 to 2644844 plus/plus)

```
1 taccatacca gaaaaaagga ccaacctatc agattcttga ttcttgaatc ttgaatctga
61 taaggatagg aatcagattt ctgggaagaa ccctttaatg ggctttcta agacagcaaa
121 ttggagaatg gatcgccaag tcctcctcca gaattcaaca gtttgtctaa gtctttcttt
181 ctctcggcca tttgggtctc agctaattgt gatgacagat taatgaagaa gaagaagaat
241 gatgaagcag agaggctttt tttggctttt tgaagaagg ctttagcaaa cggcaagtc
301 acaaaacggc aaatggactc tctcacgtac cgagtcctt ttcggctatt ttcaatatca
361 tcccatottt tgttatgttt tgttatgatg ccttagctta gttgtttctt ttgagattgt
421 taatgtaata ataaaagtgt ttgtttgtga gaaaactcg tttgcttagg gaagaagag
481 ctcttttcat gctcaaatct gttttcccg tttctgggaa attgtcttgg ctgactgcat
541 ttaattaaca aactgtcaaa cttctctctc tctctctctc cgggagaaca cgtgcttgca
601 cgtgtt
```

Morus notabilis:

NW_010362482.1

(RANGE: 1331011 to 1331616 plus/plus)

```
1 taccatacca gaaacaagga ccaacctatc agattcttgg ttcttgaatc tgataaggta
61 tggatcaga tttctgggaa gaaccctta atggccttt ctaagacagc aaattggaga
121 atggatggcc aagtcctcct ccagatttca acagtttgtc taagtcttc tttctctcgg
181 ccatttgggt ctcagctaatt tgtgatgaca gattaatgaa gaagaagaag aagaagaaga
241 agaatgatga agcagagagg ctttttgaaa gaagcttta gcaaacggc aagtcacaaa
301 acggcaaatg gactctgtca cgtaccgagt cccttttcgg ctattttcaa tatcatcca
361 tcttttgta tgttttgttt tgatgtctta gctcagttgt tttaatgtaa taataaaagt
421 ttttgtctgt gagcaaaact cgtttgctta ggaagaaga agctcttttc atgctcaaat
481 ctgtttttcc cgtttctggg aaattgtctt ggctgctgc atttaattaa caaactgtca
541 aacttctctc tctctcggg agaacacgtg cttgcacgtg ttatgctcag gcacacctgt
601 ttcgtc
```

Morus indica:

KF030989.1

(RANGE: 1 to 398 {begin to end} plus/plus)

```
1 tcaaggacca acctatcaga ttcttgattc ttgaatcttg aatctgataa ggtatggaat
61 cagatttctg ggaagaacc tttaatgggc ctttctaaga cagcaaatg gagaatggat
121 cgccaagtcc tcctccagaa ttcaacagtt tgtetaagtc tttctttctc toggccatta
181 ggtctcagc taattgtgat gacagattaa tgaagaagaa gaagaatgat gaagcagaga
241 ggcctttttt ggcctttttg aagaaggctt tagcaaacg gcaagtcaca aaacggcaaa
301 tggactctct cacgtaccga gtcccttttc ggcctatttc aatatcatcc catcttttgt
361 tatgttttgt tatgatgct tagcttagtt gtttcttt
```

Ficus carica:

CM019749.1

(RANGE: 2264638 to 2265243 plus/plus)

```
1 attc caaaaa agaccaa ctt atcagattct tgattcttta tccttgaate tgatagg gga
61 tatggaatca gatttctggc aagcaaaaag aatcctaata tgggctttct caagacagca
121 aaaa tggaga ttggagctga ggggatgta tgtcactata tatatatata tatatataaa
181 tatatatgtc cctattagtc gtcctctcct ccagatttca aacagttgtc tgagtagctt
241 cttcttaagc cattggggtc gcagctcatg agggcatgaa tgaagaaga ttttagaagc
301 aggggctctt ttgggtttt gaagaaatgc tttggcagac cggcaagtca cagaacggca
361 atggacttgc tctcgttacg aggtcctttc ggctattacc atgttatttg tcatccaat
421 cttaatgttt gtgtttcact gtttttttta gcttagtttg tg tttgttcc ttttgagatt
481 attgtaacct aataataata acagtaataa tataagaata aagggttttg tttagttttg
541 agcaaaaact ggcttgcaaa gtgaagaag aacaagaagt agtagctttt aatttgtgat
601 gtctgt
```

Ficus erecta:

BKCH01000918.1

(RANGE: 45705 to 46310 plus/minus)

```
1 attc caaaaa agaccaa ctt atcagattct tgattcttta tccttgaate tgatagg gga
61 tatggaatca gatttctggc aagaaaaaag aatcctaata tgggctttct caagacagca
121 aaaa tggaga ttggagctga ggggatgta tgtcactata aatatatata tgtccctatt
181 agtgcctctc tcctccagat ttcaaacagt ttgtctgagt agcttcttct taagccattt
241 tgggtcgcag ctcatgagg catgaatgaa agaagatttt agaagcagg gctctttttg
301 ggttttttaa gaaatgctt ggcagac cgg caagtcacag aacggcaatg gacttctct
361 cgttacgagg tccctttcgg ctatttacta tgttatttgt catccaatc ttaatgtttg
421 tgtttcactg tttttttttt catcaaatta tgttagctta gtttgtgtt tttcctttttg
481 agattattgt aacctaata tagtaacagt aataataata agaataaagc tttgtgttta
541 gttttgagca aaaactggct tgcaaaagtga aagaagaaca agaagtagta gcttgaatt
601 tgtgat
```

BKCH01000012.1

(RANGE: 5150533 to 5151138 plus/minus)

```
1 attc caaaaa agaccaa ctt atcagattct tgattcttta tccttgaate tgatagg gga
61 tatggaatca gatttctggc aagaaaaaag aatcctaata tgggctttct caagacagca
121 aaaa tggaga ttggagctga ggggatgta tgtcactata aatatatata tgtccctatt
181 agtgcctctc tcctccagat ttcaaacagt ttgtctgagt agcttcttct taagccattt
241 tgggtcgcag ctcatgagg catgaatgaa agaagatttt agaagcagg gctctttttg
301 ggttttttaa gaaatgctt ggcagac cgg caagtcacag aacggcaatg gacttctct
361 cgttacgagg tccctttcgg ctatttacta tgttatttgt catccaatc ttaatgtttg
421 tgtttcactg tttttttttt tcatcaaatt atgttagctt agtttgtgt tttcctttttg
481 gagattattg taaccaata atagtaacag taataataat aagaataaag ctttgtgttt
541 agttttgagc aaaactggc ttgcaaaagt aagaagaac aagaagtagt agcttgaatt
601 ttgtga
```

Artocarpus camansi:

LNSY01044605.1

(RANGE: 1161 to 1766 plus/plus)

```
1 aaaacaccaa cctatcagat tcttgattct tgattcttga ttctgattag gccatggaat
61 cagatttctg ggatcaacc tttaatgggc ttttcttaa tgcagcaaaa tggagcgaat
121 gggagccatt actctagtgg ctaaaatcct cctccagatt tcatcagttt gtcogagtct
181 ttctttcttt tagccatttg ggtctcagct tgacgaagaa gaatgataga gaagaggctt
241 gtttttttgg cttttggaaa gggctcagcg aac cggcaag tcacaaaacg gcaatggact
301 cttgtcacgt tacagagtcc cttttcggct attatccagt gtctattttc atcccatctt
361 ttgttgtgct tttgttatga tgtcttagct tacttttggt ttccattggg actaatattg
421 taataataaa aagcttgtgt tttaatctat ttctggagat ttacttctac tgcatataat
481 aaacttctca tctgcctctc tctctcagga aacacgtgcc tgcacgtgct ggcctcaggc
541 acacttgtgt tgtctgcat aagatcagat tttagagcat gtctgctgct taggaagatt
601 tggaac
```

LNSY01137363.1

(RANGE: 580 to 1185 plus/plus)

```
1 aaaagaccaa catatcagat tcttgattct tgattcttga atctgatttc aggccatgga
61 atcagatttc tggcaagatt taatgggctt tttttaata cagcaaaaat ggagagaatt
121 gggagccatt tccctatagt ggctaaaate ctcca ccaga tttcaacagt ttgtccagct
181 ctttctttct tttagccatt tgggtctcag tttaatgaag aagaatgatg aaaagagggc
241 ttgttttttg gcttttggaa aaaagggctt ggcgaac cgg caagtcacaa aacggcagtg
301 gaatctgtca cgttacagag tcc cttttcgg gctattgtat ctagtttcat ttcattccat
361 cttttgttgt gcttttgta tgatgtcttg gcttagtttt gtttcctttg agattaatat
421 gtaataataa aaagtttgtg tttttttgag ctaaaattcg attgcttaat ctatttctgg
481 aaatttactt ctactgcata taacaaactt ctcgtctgct tctctccgac gaaatatgctg
541 cctgcacgtg taattctcag gcacacgtgt gtcgtttgct ataaggtcag agttgtgctg
601 gctgta
```

Humulus lupulus var. lupulus:

LD164023.1

(RANGE: 5777 to 6382 plus/minus)

```
1 ataaataacc aaccaatcag attcaagaat aatattcttg attcttgaat ctgataggta
61 tgggaatgga atcagtttgc aggagaaata taaatccatt aatgggggta agacagcact
121 tggagatgga acgaggggtc cctcaggctc agctcaggcc tcaactctct atccagattc
181 caacagtttg tctgagtttt gaggccataa gctttagcag ggcgtgaaga agaagtagtc
241 tcattcaaaa cagaagcttt ttggctttta ggettcgcaa ac cggcaagt cacaaaaacgg
301 caatggactc attcacgagt ct tttcacgg ctatatttct tttcttttct tcttttgta
361 tatgttgttg atgattatca gttcagttag tttttttgct ttgtgaatgt aataataaaa
421 tggtttgcgt ttcaagcaaa aagattgc atatcatct gtgatgctca atatctttat
481 gcttttgttt tgtaaagtt ctcatctttt gatctctaca cattgacttt gtgcaacatt
541 tttctcttct ctctttgtct ctctgtgtga gaaaatgtgt gccgccagcc acgtgtttat
601 ggtcaa
```

Cannabis sativa:

NC_044373.1

(RANGE: 4233018 to 4233623 plus/plus)

```
1 ccaaaaacca acctatcaga ttcaagaatc agatctttga ttcttgaatc tgataggtat
61 gggaatggaa tcagttttca aagtgaaaaa taaatccatt aatggggttt tcagtacagc
121 acttgaagat ggaaagagga gaaaactcag ctcaagcctc actttcctat cgagatttca
181 acagtttgc tgagttttga ggccaattta ccttttgcag ggtgtcaaag aagaagaaga
241 actcactttc aattttcaaa gcaaaagggt tctatggctt ttaaagcttt accaaacggg
301 caagtcacaaa aacggcaatg gactcattca tgatgagtct tttcacggct atatttcttt
361 aatcattacc tttttttggt tgtttgtctg ttcttttgtt atgtgttttt ttttttttgt
421 gtgtgattat cagtttttagt tagttttttg ttttgtgaat gtaataataa taaaatgggt
481 tgtgttttaa gtagaagaga gttagcatca tcacataaaa tctgtactga tgatgctcaa
541 tatcttttaa tatgcttttg tttttgtaa agttgtttat cttttctttt gatctctcta
601 cacatt
```

Rhannella rubrinervis:

CM017577.2

(RANGE: 2400945 to 2401550 plus/plus)

```
1 ctgttttgag agtcccattt ggtgtcacia aaaaagagac cagcctatca gttttcttga
61 agctgatagg tatggtggtg gaagtcaatc gggaagaatc catccatggg atcttaaaag
121 ggcatatggg tattggagtc agtgaagtcc tcacggtcac accctccaac atttcaacag
181 ttgacctgag ttttttagcc attggctccg ctaactgaag cagaattctc caaaatgtag
241 agcagggttt tatggctaat tgcttagcaa acggcaagt cacagaacgg caatggactc
301 agtcatgggt ctttatggct aatatttcat ctatttccca tctgttatga ttatacatgt
361 taatgtagtt tgttgcttgg atagtgtaat agagagtttg gttttttaag cacaagttca
421 tcaatagaag cttagagggc ttttaaaccc atgcatttgt gaagatgatg gcttagtttg
481 tacatgaatt gtcaaatttg aagtttattt tagtttctgc aagttgacat ggtatgtcct
541 catctctctc tctcttcttg gcacttgtgt gcatcatgtg ttggcaaac ggcacacggt
601 tatgcc
```

Ziziphus jujuba:

NC_029688.1

(RANGE: 5986120 to 5986725 plus/plus)

```
1 gagcctcatt ggggtgtcaca catagaccaa aagcctatca gttttcttga atctgatagg
61 aaaatggaag tcaatttaga agaatacatc caagggcctt aaaatatggg catgggaaaa
121 atggatagag ggaaaggaaa cttttctcac cctctctcca aattccaaac agtttgtctg
181 agtttttag ccattgatct gcaccattcc gtgaagcaga atctccaaa atgctgcaga
241 gtttttggtt aaatgcttag gcatacgggc aagtcacaga acggcaacgg actcagtgat
301 gagtctttaa ggctatcctt ttttttctt tgtaatattg taaatgtagt ttttgtttga
361 tgaatgtaa taaaagttca ccagaagagc ctgtagacag gctttgcaa ggaatgcatt
421 tgtgattttg tgatgcttta gttttagac gaattgggaa aatttcaaaa gtttattctt
481 agtttctgca aaattcagtg aataatgact ctgcactctc atctctctct tctctgtgct
541 tctctctcca tctatttcaa tgtgtccctc cctcttttat ctatttcaaa cagaaatggt
601 aatag
```

Ochetophila trinervis:

QANX01004723.1

(RANGE: 5241 to 5846 plus/minus)

```
1  agcctggttg ggtgtcacag agaccaaagc ctatcagatt ttcttgaatc tgataggaaa
61  atggaagtta atttagaaga acacatccaa gggccttaa aggcgcatgg gaaaaatgga
121 gagagggaaa gaaaatttt tttttctca ccctcctcca aatttcaaac agtttgtctg
181 agtttttag ccattaatcc gcaccagttg atgaagcaga atcctccaaa atgcagcaga
241 gttttggct aaatgcttag cagacggca agtcacagaa cggcaatgga ctcagttatg
301 agtctttatg gctatcctct cttttctttt aatattgtta atgtagtttt tgtttgataa
361 atgtaataaa agttcagcag aagaaccttt agacaggctt tgcaaagcca tgcatttatg
421 atgctttagt ttgtagacga attggaaaat ttcaaagtt agttttattt tttatacttt
481 ctgcaatatt cagtgaataa tgattttgta tcttcgtctc tccctctatc tatttctaac
541 agacacatgt atgtttttta gttccttggg gatttgggtt atgaatcttt aatccaaaag
601 acttga
```