Phylogenetics of the genus *Erica* and anthocyanin synthesis gene expression in *Erica plukenetii*

Presented in partial fulfillment of the requirements for the degree of Doctor of Science at the Department of Biochemistry, Stellenbosch University

Nicholas Le Maitre

March 2017

Supervisor: Prof DU Bellstedt

Co-supervisor: Dr MD Pirie

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Nicholas Le Maitre

March 2017

Dedication

This thesis is dedicated to the memory of Stacey Rumble who so looked forward to its completion and tragically is no longer here to see it.

Acknowledgements

I would like to acknowledge the following people and institutions:

Prof DU Bellstedt for leadership, direction, advice and friendship

Dr MD Pirie for advice and guidance

Everyone in the Bellstedt-Botes lab for giving of their time and energy

My friends and family for putting up with me

Stellenbosch University

The National Research Foundation for funding

Abstract

The drivers of species radiations are central to questions about the evolution of diversity. The flora of the Cape Floristic Region (CFR) is particularly diverse, has exceptionally high level of endemism and may have radiated at an exceptionally high rate. Various drivers of this radiation have been proposed, including climate change, fire, niche adaptation, persistence of lineages and shifts in pollination syndrome. *Erica* is the largest genus in the CFR but its radiation has not been well studied phylogenetically. A multiple marker phylogeny would be significant in establishing its radiation rate and further elucidating the role and importance that factors such as biogeography and pollinator shifts, have played in driving its radiation specifically and in the CFR flora in general. Floral colour shifts between red and white flowers have been shown to be important in switches between pollination syndromes. The anthocyanin pathway produces coloured anthocyanins that colour the flowers of plants, also in *Erica*.

A multiple chloroplast and ITS marker region phylogeny was constructed for 597 accessions. Automated and manual alignment strategies were used to generate phylogenies and found to not be significantly different. Overall the phylogeny showed African species are descended from European species and that Mascarean and Drakensberg species may share a common ancestor with Cape species. A single Cape clade is present, sister to one anomalous species, and the sub clades reveal structure primarily related to biogeography and not morphology. Both flower colour and pollination syndrome are highly labile and multiple switches have occurred between anemophily, entomophily and ornithophily. Red flowers and ornithophily have evolved independently on at least 14 occasions.

In red flowered *Erica plukenetii* whole genome sequencing approaches using Illumina NGS sequencing were used to obtain sequences of the anthocyanin pathway genes and their *trans*-acting regulatory genes. RT-PCR and RT-qPCR were used to measure the expression of these genes in two populations of red-, pink- and white-flowered *E. plukenetii.* Expression of the CHS and the ANS genes were found to be reduced in white flowers in these populations respectively. Sequencing of the promoter regions of these genes in red-, pink- and white-flowered plants revealed mutations in the promoter binding sites of the white flowered plants that likely are the cause of anthocyanin synthesis enzyme gene down regulation and consequent loss of flower colour.

Biogeographical factors and shifts between pollination syndromes that potentially result from changes in red anthocyanin synthesis contributed to the loss of anthocyanin production are therefore likely important drivers of the radiation of *Erica* in the CFR.

Opsomming

Die aandrywers van spesiesverspryding is sentraal tot vrae oor die evolusie van diversiteit. Die flora van die Kaapse Floristiese Streek (KFS) is veral divers, het 'n baie hoë vlak van endemisme en mag versprei het teen 'n buitengewone hoë tempo. Verskeie aandrywers van hierdie verspreiding word voorgestel, insluitende klimaatsverandering, vuur-, nisaanpassing, volharding van afstammelinge en verskuiwings in bestuiwingsindroom. *Erica* is die grootste genus in die KFS, maar die verspreiding is nog nie goed bestudeer op filogenetiese vlak nie. ‘n Veelvuldige geen filogenie sou insiggewend wees i.v.m. die bepaling van verspreidingstempos, asook die rol wat biogeografie en bestuiwerverskuiwings op hierdie tempo gehad het in hierdie genus, maar ook in die KFS in geheel. Blommekleurverskuiwings tussen rooi en wit blomme is getoon om belangrik te wees vir die omskakelings tussen bestuiwingsindrome. Die antosianien padweg produseer gekleurde antosianiene wat die blomme van plante kleur, insluitende die van *Erica*.

‘n Veelvuldige chloroplast en ITS merker filogenie is gegenereer vir 597 versamelinge. Outomatiese asook oplynings met die oog oplyningsstrategieë van DNS volgordes is gebruik om filogenieë te genereer en te vergelyk. Hulle is gevind om nie beduidend te verskil nie. Die filogenie toon dat die Afrikaanse spesies afstam van die Europese spesies en dat Maskareense en Drakensberg spesies 'n gemeenskaplike voorouer met Kaapse spesies deel. 'n Enkele Kaap klade is teenwoordig, die suster van een onreëlmatige spesie, en die sub clades openbaar struktuur wat hoofsaaklik verband hou met biogeografie en nie morfologie nie. Beide blomkleur en bestuiwingsindroom is hoogs labiel en verskeie verskuiwings het plaasgevind tussen wind-, insek- en voëlbestuiwing. Rooi blomme en voëlbestuiwing het onafhanklik ontwikkel op ten minste 14 geleenthede.

‘n Heelgenoom volgorderbepaalingsbenadering was gevolg vir die rooiblom *E. plukenetii*, op die Illumina platform, in ‘n pogig om die volgordes te bepaal van die antosianien pad gene en hul *trans*-werkende regulerende gene te verkry. RT-PCR en RT-qPCR is gebruik om die uitdrukking van hierdie gene te meet in twee bevolkings van rooi-, pink- en wit-blommende *E. plukenetii*. Uitdrukking van die CHS en die ANS gene is onderskeidelik gevind om afreguleer te word in wit blomme in hierdie bevolkings. Die promotorstreke se DNS volgordes is vir hierdie monsters bepaal en daar is gevind dat die promotor bindingstreke in die witblom plante gemuteer is. Dit is waarskynlik die oorsaak van antosianiensintese ensiemgene afregulering en gevolglike verlies van blomkleur. Biogeografiese faktore en verskuiwings tussen bestuiwingsindrome wat potensieel as gevolg van veranderinge in rooi antosianien sintese bygedra het tot die verlies van antosianien produksie is dus waarskynlik belangrike drywers van die verspruiding van *Erica* in die KFS.

Contents

List of abbreviations

CFR: Cape Floristic Region

Mya: Million years ago

ITS: Internal Transcribed Spacer region

RC: Reverse Complement

CHS: Chalcone Synthase

CHI: Chalcone Isomerase

F3H: Flavanone 3-Hydroxylase

F3'H: Flavanoid 3'-Hydroxylase

F3',5'H: Flavanoid 3',5'-Hydroxylase

DFR: Dihydroflavanol 4-Reductase

ANS: Anthocyanidin Synthase

UDP-GST: Uridine diphosphate glucose-3-glucosyltransferase

MYB: myeloblastosis

bHLH: basic Helix-Loop-Helix

WDR: W-D repeat

MRE: MYB Recognition Element

BRE: bHLH Recognition Element

NGS: Next Generation Sequencing

PTB1: Paratubulin 1

RT-qPCR: Reverse Transcriptase quantitative PCR

DP: Du Toit's Kloof Pass

FP: Franschhoek Pass

Chapter 1: Structure of this thesis

This thesis is structured into four main chapters.

Chapter 2 gives a review of the literature pertaining to the genus *Erica*, radiations in the Cape Floristic Region and the phylogenetics, pollination syndromes and pollination of the genus and finally anthocyanin biosynthesis and its role in pollination.

Chapter 3 details the contribution that the author made to a paper on a multigene phylogeny of the *Erica* genus that was published in BMC Evolutionary Biology in 2016 (Appendix A). The author sequenced multiple genes of 255 *Erica* accessions (Appendix B). These sequences were manually added to an already extant existing alignment matrix of sequences generated by one of the senior authors of this research, Dr Michael Pirie, bringing the total number of accessions to 612. Manual alignments, by their sequential, rather than global nature introduce the possibility of bias. For this reason an automated alignment of the same dataset was performed so that it could be compared to the manual alignment to test for bias. Conflicts between the ITS and chloroplast trees required the removal of 15 accessions, reducing the final number in the combined phylogeny to 597. Biogeographical characters, nectar sugar composition and morphological data were plotted onto the tree to detect any evolutionary signals in these characteristics.

Chapter 4 describes an investigation into the genetic changes in the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes underlying the floral colour changes observed in *Erica plukenetii*, using Next Generation Sequencing, Reverse Transcriptase PCR and Real Time Quantitative PCR.

Chapter 5 concludes the thesis and describes perspectives for future work emanating from this research.

Chapter 2: Literature review

2.1 The *Erica* radiation in the Cape Floristic Region

2.1.1 Classification

Erica, commonly known as heathers or heath-type plants, is a genus in the subfamily Ericoideae which comprises 19 genera of acidophilic, woody plants. Revisions by Oliver[1,2] have reduced many minor genera such as *Phillipia, Blaeria* and *Ericinella*, to synonomy, thereby bringing the total number of species in the genus *Erica* to ca. 865.

The positioning of the genus within the subfamily has been confirmed by several systematic studies, most recently by Gillespie and Kron (2010)[3]. The Angiosperm Phylogeny Group (2016)[4] places the Ericoideae in the family Ericaceae and in the order Ericales (Figure 1), sister to the Caryophyllales, Campanulids and Lamiids.

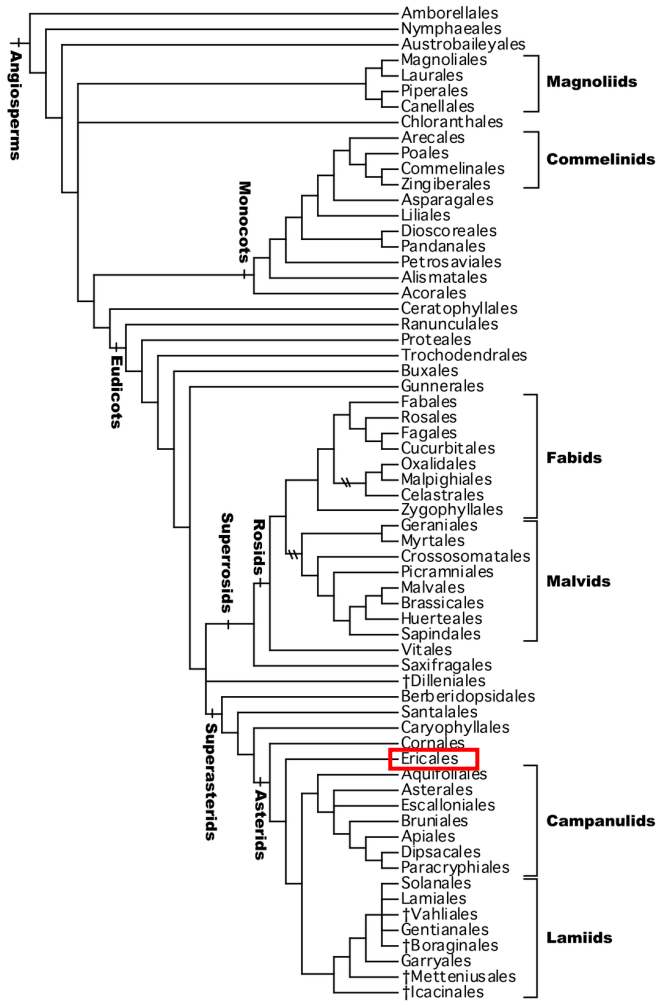


Figure 1: The Angiosperm phylogeny showing Ericales (red box) in the Asterids. From the Angiosperm Phylogeny Group (2016)[4].

Based on a molecular phylogeny generated by Bremer *et al.* (2002)[5], using three protein coding genes, rbcL, matK and ndhF and three non-coding regions, the order Ericales consists of 30 families. This shows (Figure 2) the close relationship of the family Ericaceae to the families Theaceae and Actinidiaceae which are important to the genomic analyses performed in this study.

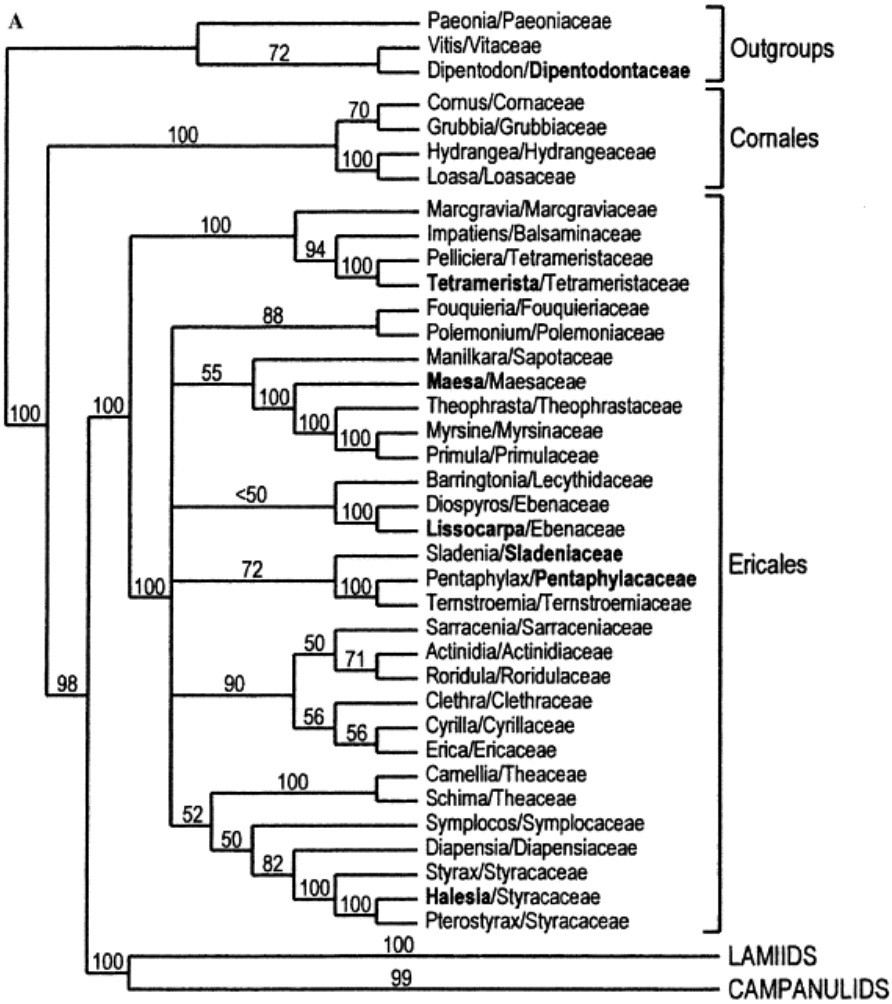


Figure 2: The phylogenetic relationships with the Ericales and closely related groups according to Bremer *et al.* (2002)[5].

Using rbcL and matK data from 450 ingroup and 29 outgroup species, covering all subfamilies, most tribes and genera, Schwery *et al.* (2014) generated a well supported and resolved phylogeny of the Ericales. The phylogeny was dated using 18 fossil calibrations and diversification rates were estimated using binary-state speciation and extinction (BiSSE) and multistate speciation and extinction (MuSSE) analyses. This revealed that multiple radiations in certain groups including Rhodoreae, *Rhododendron*, Vaccinieae and *Erica*, have occurred, indicating a predisposition to rapid radiation in the family. Furthermore, these radiations were found to predominantly occur in mountainous areas across the globe. Such a radiation occurred in *Erica* in which ca. 865 species are currently recognised making it one of the largest genera in the family[6].

2.1.2 Global distribution of the *Erica* genus

Geographically, the genus *Erica* occupies a narrow north to south distribution from northern Europe, across the African mountains, to southern Africa, the Middle East and Madagascar (Figure 3). Throughout this range the genus is widespread with populations as far east as Turkey and Lebanon, where *E. spiculifolia* occurs, and the south western tip of the Arabian Peninsula, where *E. arborea* occurs (although it possesses a much wider distribution). The westward range limit is formed by the Atlantic coast of Spain and the Pyrenees. Populations also occur on islands in the Mediterranean Sea.

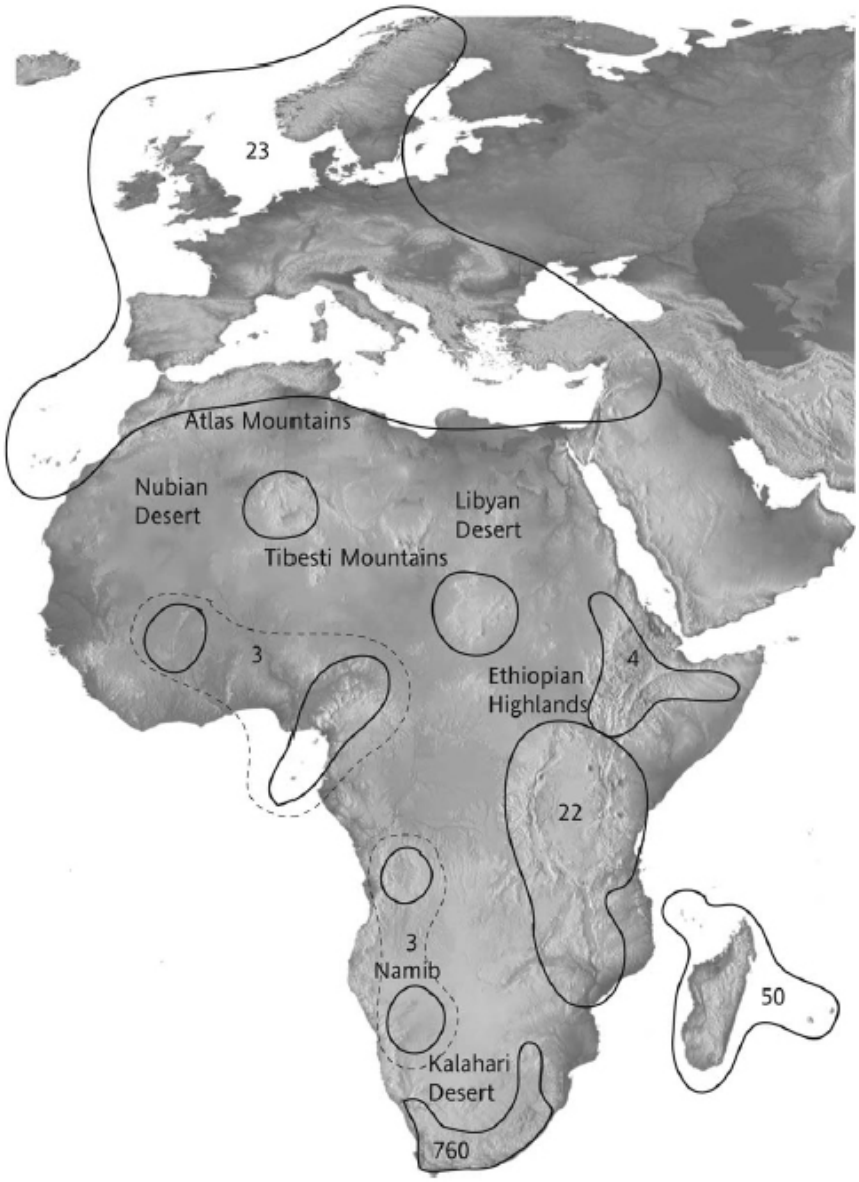


Figure 3: Distribution of Erica species in Europe and Africa, with estimates of the number of species. From McGuire and Kron (2005)[6].

In Africa, *Erica* species are typically found in grasslands and montane scrub. Populations occur on the Mediterranean coast between Morocco and Tunisia and there is an isolated population in the

Sahara on the Tibesti Mountains. In the east African mountains from Ethiopia to northern Malawi, 22 species of *Erica* occur, forming a distinct zone of high altitude vegetation and is referred to as the ericoid vegetation zone due to its dominance. Fifty species occur in Madagascar and its surrounding islands, with many restricted to higher altitudes.

The *Erica* genus has an uneven relationship between diversity and distribution. Generally, in the northern parts of the genus' range, the number of species is low, but those few species are widely distributed. Further south in the range, especially in the Cape Floristic Region (CFR) of South Africa diversity is very high, with around 680 species[7] indigenous to the region. Most of these species are confined to very small ranges[6,8–10]. The reasons for this diversity gradient are unknown, but it is speculated that the conditions in the CFR, an area of high species diversity, may be particularly conducive to radiations.

2.1.3 Species richness in the Cape Floristic Region

The southern African Flora, situated at the southern tip of the African continent is recognised as one of the richest temperate and sub tropical floras[11]. It contains multiple vegetations zones and climatic regions from desert to high altitude peaks of more than 3000m (including the Drakensberg). The CFR makes up the major part of the greater Southern Africa flora. The CFR covers approximately 90 000km^2 at the south western tip of the African continent (Figure 4). Its species richness when compared to its area is well above what would be expected (Figure 5) and is extremely high in endemism (Figure 6). The CFR is recognised as one of the 25 global biodiversity hotspots, along with regions like the Brazilian Rain Forest, Madagascar, New Guinea, the Mediterranean Basin, South-western Australia and the California Floristic Province, amongst others[12]. These hotspots contain approximately 35% of the vertebrate species and 44% of the plant species on the planet, with very high levels of endemicity[12]. The CFR ranks 12th overall but has levels of endemicity rarely seen outside of the tropics and more reminiscent of island floras, with 5682 endemic plant species of 8200 plant species in total (~70%); making up 44% of the total species in the Southern African region[13].

The CFR can be subdivided into zones of climate and topology (Figure 4)[14]. It is typified by coastal plains in the west and south, backed by the Cape Fold mountains that in places exceed 2000m. More arid conditions occur in the north west and inland in the rain shadow of the mountains. Rainfall increases to the south east and remnant forest patches occur. These zones correspond well with the phytogeographical centres identified by Oliver *et al.* (1983)[15] (Figure 7).

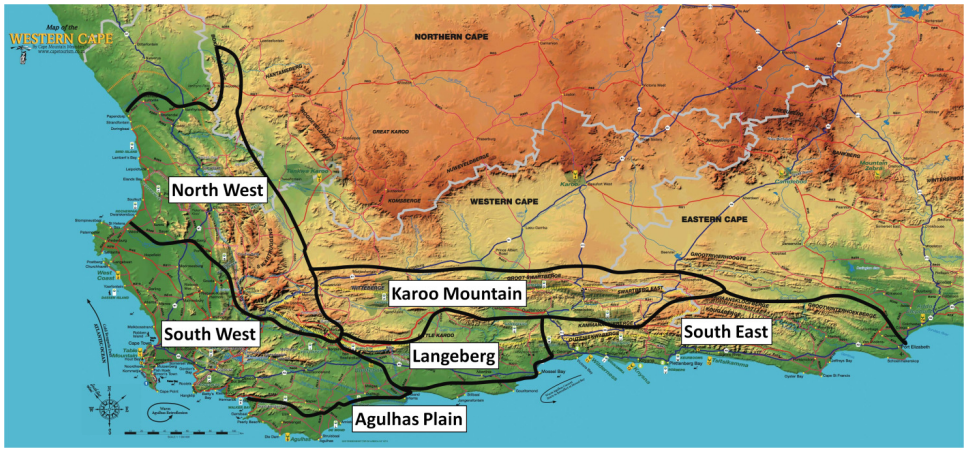


Figure 4: The Cape Floristic Region showing climatic zones and topography. Redrawn from Goldblatt (1997)[14], base image from capetourism.co.za.

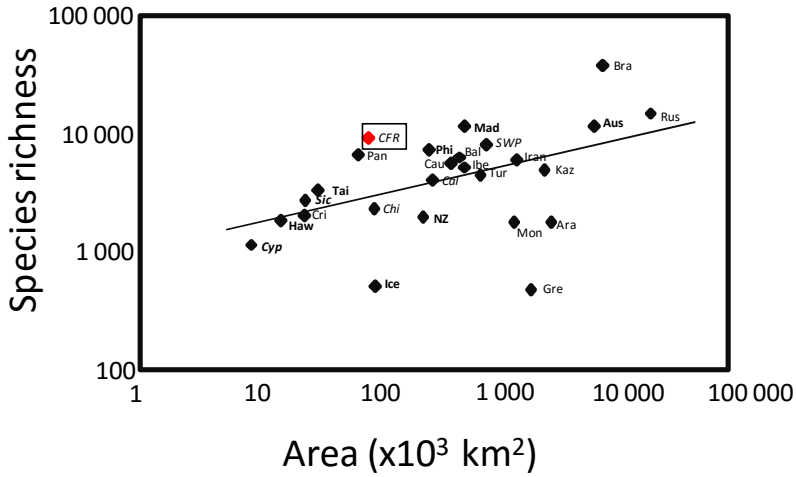


Figure 5: A comparison of species richness between the CFR and other regions. The fitted line indicates expected species richness, with countries higher than the line richer than expected and vice versa. Names in bold are islands, italics are areas with a Mediterranean climate. Axes are log scale. Arabia (Ara), Australia (Aus), Balkans (Bal), Bra (Brazil), California Floristic Province (Cal), Caucasus (Cau), Cape Floristic Region (CFR), Central Chile (Chi), Crimea (Cri), Cypress (Cyp), Greenland (Gre), Hawaii (Haw), Iberia (Ibe), Iceland (Ice), Iran (Iran), Kazakhstan (Kaz), Madagascar (Mad), Mongolia (Mon), New Zealand (NZ), Panama (Pan), Philippines (Phi), Soviet Union (Rus), Sicily (Sic), South Western Australia (SWP), Taiwan (Tai), Turkey (Tur). From Linder *et al.* (2003)[16].

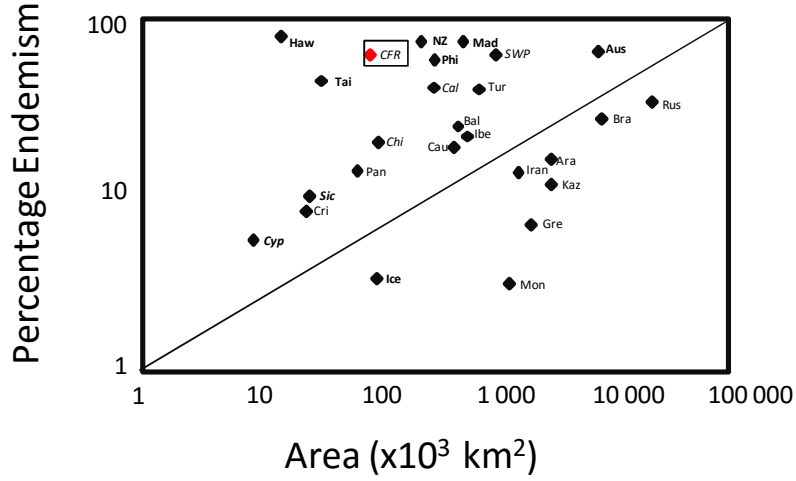


Figure 6: A comparison between the log of percentage of endemism and the log of area of a region. Island floras tend to have higher levels of endemism. Abbreviations are as for Figure 5. From Linder *et al.* (2003)[16].

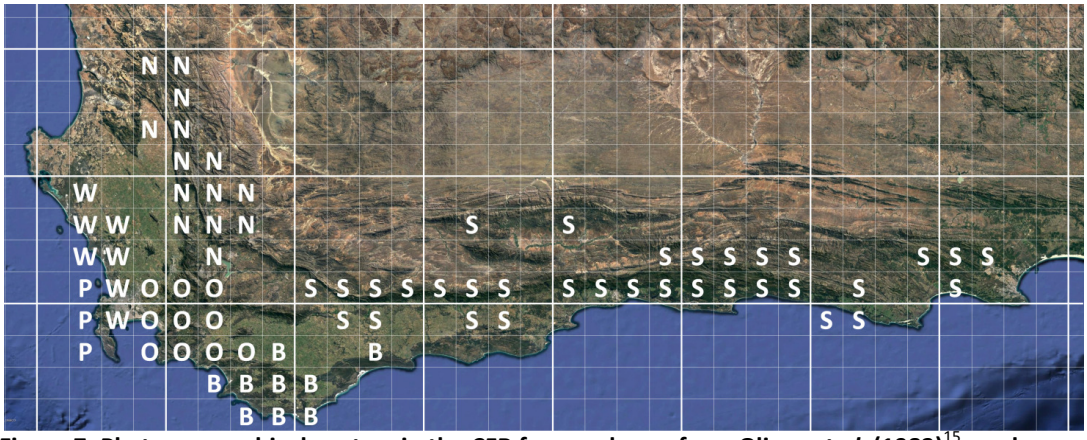


Figure 7: Phytogeographical centres in the CFR from redrawn from Oliver *et al.* (1983)[15], each square represents a quarter degree square. Base image from Google Earth. N: Northern, W: West coastal, P: Cape Peninsula, O: South Western, B: Bredasdorp, S: Southern.

2.1.4 Radiations in the CFR

Just 33 'Cape floral clades' account for nearly 50% of the species in the CFR, a number seldom seen outside of island floras such as Hawaii [16]. Of these 33 floral clades, 26 seem to have originated in the CFR with 75% of the species endemic to the CFR. The number of Iridaceae, Ericaceae, Aizoaceae, Proteaceae and Restionaceae species are higher in the CFR than anywhere else[17]. The Proteaceae have radiated exceptionally rapidly in the CFR, second only to the rate in South Western Australia[18]. The crown ages of some of the 'Cape floral clades' have been dated using molecular clock estimates (Table 1). The succulent Karoo lineages seem to have radiated more recently and the lineages endemic to the fynbos show both older and younger lineages[19,20], making the CFR both a museum of lineages that have persisted and an area of recent rapid radiations.

Table 1: 'Cape Clades' from Linder (2003)[16]with CFR radiation dates. Mean ages are given with references.

Clade	Family	Total Species	CFR species	Age (Mya)
Tetratria	Cyperaceae	40	38	18[21]
Irideae	Iridaceae	226	136	46[22]
Ixioideae	Iridaceae	900	516	42[22]
Ehrharta	Poaceae	36	20	41.9[20]
Pentaschistis-Pentameris-Prionanthium	Poaceae	81	55	13.9[23]
African Restionaceae	Restionaceae	350	340	91[20]
Relhaniinae	Asteraceae	170	131	21[24]
Erica	Ericaceae	860	658*	20[25]
Crotalarieae	Fabaceae	297	291	32.6[26]
Podalyrieae, Liparieae	Fabaceae	128	120	32.6[26]
Pelargonium	Geraniaceae	250	148	43[27]
Lampranthus	Aizoaceae	155	125	8.7[20]
Muraltia, Nylandia	Polygalaceae	117	108	18.5[28]
Proteae	Proteaceae	340	264	17.7[29]
Phyliceae	Rhamnaceae	152	134	18[30]

*As specified by Linder (2003)[16], new species descriptions increase this number to ca. 685 at present.

2.1.5 Factors driving radiations in the CFR

The current consensus of opinion on the history of the CFR is[31–35]: from the late Oligocene the globe warmed, with a peak in the mid Miocene epoch, approximately 17-15Mya. In the latter half of the Miocene, the globe began to cool possibly due to a combination of changes in the planet's orbit around and obliquity to the sun[33]. The Polar Regions also cooled and as a result, the ice sheets expanded and Antarctica was completely covered by an ice-sheet again by 10Mya. Additionally the Central American Seaway through the Isthmus of Panama closed, preventing flow between the Atlantic and Pacific Oceans, altering the currents in the Atlantic Ocean. The cold Benguela current along the west coast of the African continent became significantly stronger and Benguela Upwelling System arose along the Atlantic coast[32,35], between 15 and 5.1Mya[36,37].

During the mid Miocene, the climate of Southern Africa was characterised by all year round rainfall and tropical to sub-tropical conditions. The cooling of the oceans reduced the availability of moisture in the air necessary for the year round rainfall and the tropical climate[38].

Around 20Mya the entire southern half of the African continent was uplifted, reducing the amount of moisture from the Indian Ocean reaching the centre of the continent. The cooling, combined with the reduction in moisture and the uplift, resulted in a gradual transition to a winter rainfall regime, which continues until today. The Cape vegetation was prominently tropical with the ancestors of the modern Cape species restricted to the nutrient poor sandstone soils at higher altitudes in the Cape Fold mountains that had formed during the breakup of the Gondwana supercontinent[39].

The shift in rainfall patterns and the associated increase in aridity[38], as well as the establishment of a regular fire regime[40,41], lead to the extinction of the tropical vegetation making large areas available for colonisation by the flora from the mountains[35,42], including *Erica*. The regular fire regime had a further role in driving diversification, as it asserted a strong selective pressure for the development of reseeder and resprouter fire survival strategies in the surviving vegetation[43–45].The reseeder strategy, which is common in CFR *Erica* species[44,46], may also drive diversification as populations are subject to more intense selection as their generation time is shorter[47].

This change in vegetation types is evident in pollen recovered from seabed cores off the mouth of the Orange River[35,48], further north along the Namibian coast[49–52] and excavations at Langebaanweg[53]. Dating of these records places the change of vegetation types during the Late Miocene and into the Pleistocene. Tropical and subtropical groups declined and Ericaceae pollen appeared about 10Mya in sea bed cores. Due to the paucity of the fossil record it is not possible to

put a very exact date on to the emergence of present day Cape flora, with fossil based estimates varying between the Pliocene[14] and Late Pliocene[54].

Subsequent uplift of the East African escarpment, about 5-3Mya[55], by several hundred meters[56,57] and the potential creation of a rain shadow[16] to the west of the escarpment is unlikely to have played a major role in the interruption of the tropical climate as it only occurred long after the changes were already underway and the influence of the cold Benguela current off the southern African west coast was the dominating climatic factor.

The Cape region experienced a long period of large scale climatic stability following the establishment of the winter rainfall regime[58]: it did not experience the glacial cycles and consequent loss of species that occurred in Chile and the Northern Hemisphere[14,17]. Before 3Mya, wetter and warmer conditions favouring savannah vegetation and grasses were found. From 3.1 to 2.2Mya, there was an increase in vegetation dependent on good winter rains that were accompanied by an increase in the Benguela Upwelling System bringing colder polar water to the surface along the west coast and a northward shift of the Southern Polar Front Zone. Then between 2.7 and 2.2Mya, the polar-equator temperature gradient increased and global atmospheric circulation improved due to increased glaciations; this resulted in an increased expansion of semi-arid areas. From 2.2Mya both climatic fluctuations and aridification increased further as the Polar Front Zone moved southwards and there was a reduction in the availability of moisture from the Atlantic Ocean as the westerly winds declined as well[59]. These small, gradual, changes in the regional climate would have favoured first one group of species and then another[60–62].

The Cape Fold mountains of the CFR are steep, rugged sandstone mountains with nutrient poor quartzitic soils. Verboom *et al.*(2015)[63] studied six CFR lineages: *Elegia-Thamnochortus, Leucodendron, Protea, Tetraria, Stoebe* and *Syncarpha*. They found high altitude species were typically more restricted in range than low altitude species and are more range exclusive. They conclude that areas with complex terrains typical of the Cape Fold mountains promoted isolation at higher altitudes and allowed vicariant speciation and that the mode of speciation may differ between high and low altitude species.

The CFR has a wide diversity of soil types, altitudinal gradients and locally varied rainfall patterns that combine to produce widely divergent habitats located in close proximity to one another[15,45]. In this it strongly resembles the California Floristic Province and the Mediterranean Basin: winter rainfall regions with similarly diverse soils and steep altitudinal gradients that are species rich when compared to their neighbours but significantly less species rich than the CFR even though they are significantly larger in area[14,17,64]. South-western Australia, which has a similar number of species, is

quite similar to the CFR, although it is much larger in area and lacks the steep altitudinal gradients. Both it and the CFR have large areas of nutrient poor quartzitic soils which have been linked to higher than normal species richness[65,66].

Gene flow in plants is affected by two major components, pollen flow and seed dispersal. There is no evidence that pollen flow differs significantly in the CFR to the rest of the world, with similar availability of numbers and types of pollinator[67]. Seed dispersal restrictions are common in CFR species[14,67] as many plants have only short range seed dispersal. Myrmecochory or dispersal of seeds by ants, plays a disproportionately large role in the dispersal of seeds of CFR species, particularly on the nutrient poor soils[14]. More than 1000 CFR species' seeds have fatty elaiosomes to attract ants. Myrmecochory may be a further adaptation to fire as the ants bury seeds very successfully[14]. Most other CFR species have passive dispersal mechanisms with ranges of less than five meters. Plants with low seed dispersal distances tend to have narrow ranges and higher species diversity and tend to be localised to zones of high endemicity[15], while the few CFR species with long range dispersal mechanisms, such as birds, bats and wind, have larger ranges with less diversity[14,17]. It would seem that gene flow restrictions caused by low seed dispersal distances play a role in diversification as a result of niche occupation. Whole genome duplications do not appear to play a role in the diversification of the CFR flora[68].

It is likely the combination of the removal of the tropical vegetation, the macro stability of the climate, the small climatic fluctuations, the development of fire survival strategies, the steep altitudinal gradients, the nutrient poor soils, the availability of highly diverse biogeographic niches and the gene flow restrictions that have contributed to the diversity of plant species seen today in the CFR.

As *Erica* is the largest genus in the CFR and is distributed throughout the CFR in every habitat, identifying the causes of the radiations in *Erica* may therefore provide further insights into the drivers of other radiations in the Cape flora.

2.1.6 Phylogenetic relationships in *Erica*

The cause of the radiation of *Erica* in the CFR and conversely, the lack of diversity outside of the CFR is not well understood. *Erica* is the largest genus in the CFR with ca. 685[7] species that typically have small ranges[7]. In stark contrast, only ca. 25% of *Erica* species are found outside of the CFR, and these species have large ranges. Due to this diversity gradient, a south to north spread, the so called "Cape to Cairo" route, was suggested for *Erica*[23]. McGuire and Kron (2005)[6] and Pirie

et al. (2011)[7]could reject this hypothesis in favour of a southwards spread from Europe to southern Africa based on molecular phylogenetic analyses.

Erica is one of the least well studied CFR clades, at least phylogenetically. A morphologically based phylogeny showed poor resolution and unresolved basal polytomies[2]. Molecular phylogenetic studies have been attempted by a group led by Professor Mark Chase at the Royal Botanical Gardens at Kew, followed by another attempt by a group led by Professor Kathleen Kron at Wake Forest University in the USA but these failed, likely due to the daunting size of the task and the numerous difficulties in obtaining DNA of sufficient quality and quantity for PCR (Bellstedt, pers. comm.).

McGuire and Kron (2005)[6] studied 25 European and 14 African *Erica* species. African *Erica* species formed a monophyletic grouping within European *Erica* indicating that they have descended from the European species. Cape *Erica* species were shown to be sister to *Erica arborea*, having shared a common ancestor with the European species as can be seen in Figure 8.

Figure 8: Strict consensus tree of Erica species using nuclear and chloroplast markers showing the relationship between European and African Erica species. From McGuire and Kron (2005)[6].

With the development of a direct amplification procedure[69], Pirie *et al.* (2011)[7] were able to expand significantly on this work with a nuclear Internal Transcribed Spacer region (ITS) phylogeny,

containing approximately 45% of the species within the genus. The additional species were mostly either CFR or African species. The addition of these species to the phylogeny diminished support for the sister relationship between *E. arborea* and the African clade, showing rather a large clade which contained the Madagascan, Mascarene and African *Erica* species (Figure 9).

The support for a relationship between east African species and southern African clades is rather weak but they appear to be distantly related, possibly indicating multiple independent origins of Cape species rather than a monophyletic Cape clade as indicated by McGuire and Kron (2005)[6], albeit on a much smaller dataset (Figure 8). The CFR *Erica* species mostly sort into clades either consisting of species with a narrow distribution or species with a widespread distribution, possibly indicating that the clades of species with narrow distributions represent local radiations[7]. Floral morphology appears to be highly labile with multiple switches between floral size, -shape, and -colour. Pollination syndromes appear to have switched frequently as well, ornithophily has evolved at least eight times across multiple clades in the phylogeny[7].

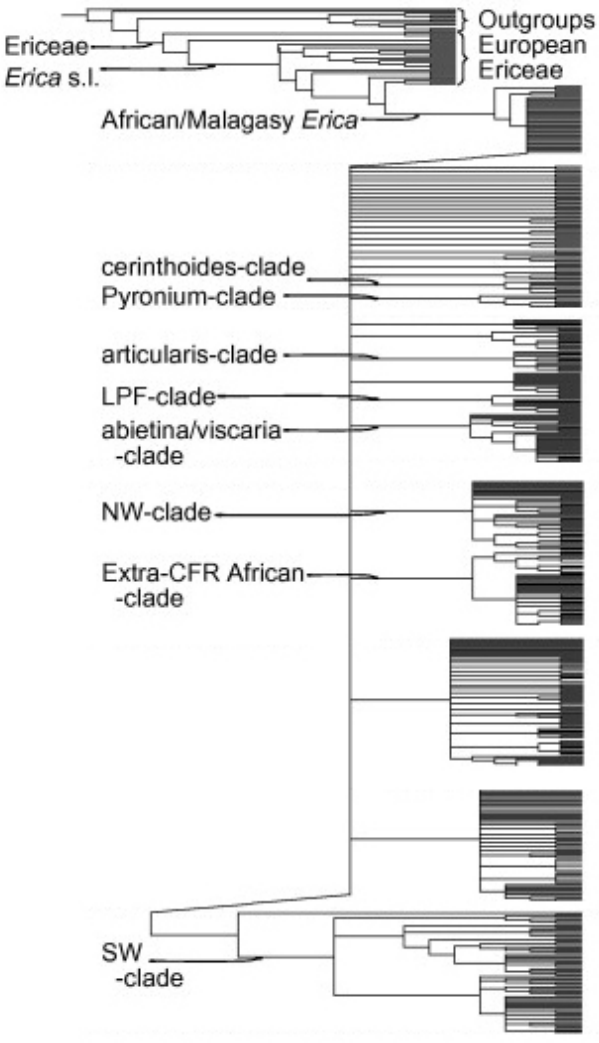


Figure 9: ITS tree from Pirie *et al* (2012)[7]. African and Malagasy *Erica* species form a single large clade.

The ITS phylogeny suggested potential mechanisms for the dramatic radiation of the *Erica* species in southern Africa with the availability of diverse biogeographical ranges in very close proximity to one another, the large scale climatic stability and consequent lack of extinctions as well as reproductive isolation due to pollinator shifts[7,70] all having played a possible role.

The ITS phylogeny raised many questions and provided few clear answers to the question of the drivers of the radiation of *Erica*. It lacked resolution and a more comprehensive, multi-gene phylogeny was required to improve resolution and perhaps answer those questions.

2.2 Pollination

2.2.1 Pollination syndromes

Angiosperm flowers occur in a multitude of different sizes, shapes, colours, nectar volumes, nectar compositions and/or positions of sex organs. Many flowers are 'generalists' that do not favour one functional group of pollinators over others[71–73]. These 'generalist' flowers undergo 'adaptive wandering'[74] where they adapt to local pollinators without becoming specialised to a particular pollination syndrome. Thompson and Wilson (2008)[73] liken this adaptive wandering to logs, each representing a trait, floating in the ocean that respond to changes in winds and currents but without overall directionality. In contrast to the generalist flowers, some flowers conform to specific pollination syndromes; they tend towards a specific set of characteristics that favour one class of pollinator over another[75]. Thompson and Wilson (2008)[73] liken these pollination syndromes to "vortices"[73]. In these "vortices", the logs cannot escape; they are trapped in a self reinforcing cycle tending towards a particular pollination syndrome. Both "positive" and "negative"[76] changes occur, where positive changes attract a particular pollinator and negative changes seek to exclude others[77]. Within each "vortex", adaptive wandering can occur, but changes between "vortices", e.g. a shift from entomophily to ornithophily, requires both destabilisation of the current "vortex" and strong pull from another "vortex"[73]. Despite the seeming complexity of this process and the combination of factors required, switches between pollination syndromes are common; with examples numbering in the hundreds just from the western region of North America[78–80].

Pollinators favouring one floral colour over another has been noted in: *Antirrhinum majus*[81,82], *Aquilegia* species [83–85], *Bixa orellana*[86], *Clarkia gracilis*[87], *Claytonia virginica*[88], *Dactylorhiza sambucina*[89,90], *Delphinium nelsonii*[91,92], *Disa ferruginea*[93], *Erica perspicua*[94], *Gentiana lutea*[95], *Heliconia caribaea*[96], *Hydrophyllum appendiculatum*[97], *Ipomoea purpurea*[98,99], *Ipomopsis* species[100–103], *Iris* species[104–106], *Keckiella* species[80], *Linanthus* species[107,108], *Linaria canadensis*[109], *Linum pubescens*[110], *Lobelia* species[111,112], *Malva moschata*[113], *Mimulus* species[114–119], *Ourisia*

glandulosa[120], *Penstemon* species[80,121], *Phlox* species[122–127], *Platystemon* species[128], *Protea* species[129], and *Raphanus* species [130–132]. Of these examples, pollinator imposed selection has been shown to be occurring in *Aquilegia caerula*[83], *Bixa orellana*[86], *Dephinium nelsonii*[91,92], *Disa ferruginea*[93], *Gentiana lutea*[95], *Ipomoea purpurea*[98,99], *Ipomopsis* species[100–102], *Iris* species[104–106], *Linanthus parrye*[107,108]and *Raphanus sativus*[132]. Pollinator imposed selection is therefore well documented but not ubiquitous to every change in floral colour.

Strongly contrasting colours differentiate the flowers, or parts thereof, from the rest of the plant as a signal for pollinators[72,75,133,134]. Destabilisation of the entomophile "vortex"[73] may occur as a result of a floral colour change from white, yellow, pink, purple or blue to red, as most insects do not perceive well in the blue end of the visual spectrum and consequently, take longer to locate red coloured flowers as they appear camouflaged[133,135]. Birds, with their greater range of colour perception, more easily find these flowers[133,135] . The floral colour change towards red would have a negative, exclusionary effect on insect pollinators and possibly a positive, attractive effect on bird pollinators, although the evidence for this positive effect is not clear[135,136]. These floral colour changes may lead to assortative mating which could influence selection[75,137].

The destabilising effect of a floral colour change to a colour more suited to ornithophily coupled to the observation that birds are typically more efficient pollinators than insects[138,139], could be sufficient to effect the change from entomophily to ornithophily[73]. Similarly, the loss of red pigmentation and physical changes to plant structures, such as branches less suitable to perching could cause a change from ornithophily to entomophily[70].

In essence, entomophily requires a suite of characters which make the plant attractive to insects and physically capable of being pollinated by insects. Ornithophily requires a different suite of characters, which make the plant both attractive to, and capable of being pollinated by, birds. This implies that switches between these pollination syndromes require multiple changes to a number of characters, followed by selection for those characters to combine them in a suite suitable for either the one or the other pollinator type.

2.2.2 Pollination syndromes in *Erica*

Rebelo and Siegfried (1985)[140] analysed the floral morphology of CFR *Erica* species, and found that the species segregated into three groups. Anemophilous species had the lowest levels of colour polymorphism, and have colours that make them inconspicuous to pollinators. Entomophilous species had intermediate levels of colour polymorphism and are coloured pink, purple or white; with smaller flowers; characteristics that suit most insect pollinators. Ornithophilous species were

the most polymorphic with respect to colour and are red, orange, yellow or green; with larger flowers; characteristics that exclude insect pollinators. More recent studies have shown that rodents are also pollinators of a few *Erica* species[141,142]. Some degree of specialisation with respect to flower corolla length and bird beak length has been observed in several CFR plant species[140,143,144] including *E. plukenetii*[70]. Rebelo[143] speculates that due to the longer ranges of bird pollinators than insect pollinators, the effective population size of ornithophilous species is much larger and allows the maintenance of more polymorphisms. In the smaller populations of entomophilous species that are more subject to limited gene flow, these polymorphisms would be lost faster. Ornithophilous *Erica* species are pollinated by the Orange Breasted Sunbird[143,145], *Anthobaphes violacea*, the Southern Double-collared Sunbird, *Cinnyris chalybea* [144,146], and the Malachite Sunbird, *Nectarinia famosa*[70]. Low numbers of pollinator species would also maintain polymorphisms in their populations[143].

Based on the pollination syndrome assignments of Rebelo and Siegfried (1985)[140] and Rebelo *et al.* (1985)[147], Barnes *et al.* (1995)[148] compared fructose, glucose and sucrose ratios in nectar between entomophilous *Erica* species and ornithophilous *Erica* species. Ornithophilous *Erica* species had predominantly high concentrations of sucrose relative to glucose and fructose; and entomophilous species had predominantly low concentrations of sucrose relative to glucose and fructose. There were also entomophilous species with high sucrose relative to glucose and fructose and ornithophilous species with low sucrose relative to glucose and fructose. Their results may be affected by sampling bias as there are 37 ornithophilous species versus only 13 entomophilous species in their dataset. They concluded that pollination syndromes in *Erica* cannot be deduced from relative sugar concentrations in nectar. Sucrose:hexose ratios in nectar may rather be a consequence of flower shape, as the high hexose nectars are less prone to evaporation (due to the increased osmolarity of the nectar) and tend to occur in open flowers. For high sucrose nectars, the opposite is true, with flowers being predominantly closed in shape[149,150]. Nectar volumes and relative concentrations of total sugar have been shown to play a role in pollinator selection by nectar feeding birds and insects[75,136,151,152]. No comparisons of nectar volumes and relative concentrations of total sugar between entomophilous *Erica* species and ornithophilous *Erica* species have been made.

2.2.3 Pollinator selection in *Erica*

Heystek *et al.* (2014)[94] found that pink- as opposed to white-flowered morphs of *E. perspicua* were preferred by the Orange-breasted Sunbird with no fitness effect, nectar volumes and concentrations of sugars did not differ significantly between the colour morphs. In *E. plukenetii*, Van der Niet *et al.*

(2014)[70], found that bird-pollinated ecotypes were red- or white-flowered, while moth-pollinated ecotypes were exclusively white-flowered. This evidence indicates that flower morphology, including anthocyanin pigmentation, is linked to pollination processes in *Erica*. The interaction between anthocyanin biosynthesis and pollinators is therefore a plausible driver of speciation in *Erica*.

2.3 Anthocyanins

2.3.1 The role of anthocyanins

The biosynthesis of anthocyanins has been studied in great detail for more than 100 years[153]. The colour changes caused by increases in the anthocyanin pigments present are highly visible and as such provide excellent markers of inheritance of linked traits. Their importance as a marker of fruit ripeness, a model for movement of metabolites both intra- and extracellularly, control of gene expression and as a model system for the first use of RNA interference has lead to the anthocyanin synthesis pathway and its regulation being very well characterised.

Anthocyanins are secondary plant metabolites that colour various plant tissues, primarily leaves, flowers and seeds. Anthocyanins are a primary determinant of floral colour[154]. They are thought to provide some protection to environmental stresses[155] due to their predictable induction at various times of the year, prevent herbivores from eating leaves by their camouflaging effect [156], prevention of photo-inhibition of photosynthesis[157,158], chelators of excess metal ions[159], prolong leaf life during starvation[159], form precursors for phytoalexins that play a role in the plant immune system[160] and protect other photolabile[161] compounds in leaves. It is thought that anthocyanins also play a role in protecting the plant from UV-B radiation, however the evidence for this is not clear with support both for and against[159], indicating that this may be a secondary role[156].

2.3.2 Synthesis of anthocyanins

Synthesis of anthocyanins in plants occurs via the anthocyanin pathway which is a branch of the phenylpropanoid pathway (Figure 10). Three molecules of malonyl-CoA are coupled with a molecule of 4-coumaroyl-CoA in the first step by chalcone synthase (CHS) to form chalcone in a polyketide folding reaction. Chalcone is then isomerised by chalcone isomerase (CHI) to form naringenin. Naringenin is then oxidised by flavanone 3-hydroxylase (F3H) to dihydrokaempferol. At this point the pathway branches and dihydrokaempferol is converted to dihydroquercetin or dihydromyricetin by flavanoid 3'-hydroxylase (F3'H), or flavanoid 3',5'-hydroxylase (F3',5'H), respectively. Dihydroflavanol 4-reductase (DFR) then reduces the dihydroquercetin to leucocyanidin, the dihydrokaempferol to leucopelargonidin and the dihydromyricetin to

leucodelphinidin, respectively. Anthocyanidin synthase (ANS) then reduces colourless leucocyanidin to cyan coloured cyanidin, colourless leucodelphinidin to purple coloured delphinidin, and colourless leucopelargonidin to yellow coloured pelargonidin, respectively. Each of the anthocyanidin molecules is then stabilised by glycosylation in a reaction catalysed by uridine diphosphate glucose-3-0-glucosyltransferase (UDP-GST), forming cyanidin-3-0-glucoside, delphinidin-3-0-glucoside and pelargonidin-3-0-glucoside, respectively[162].

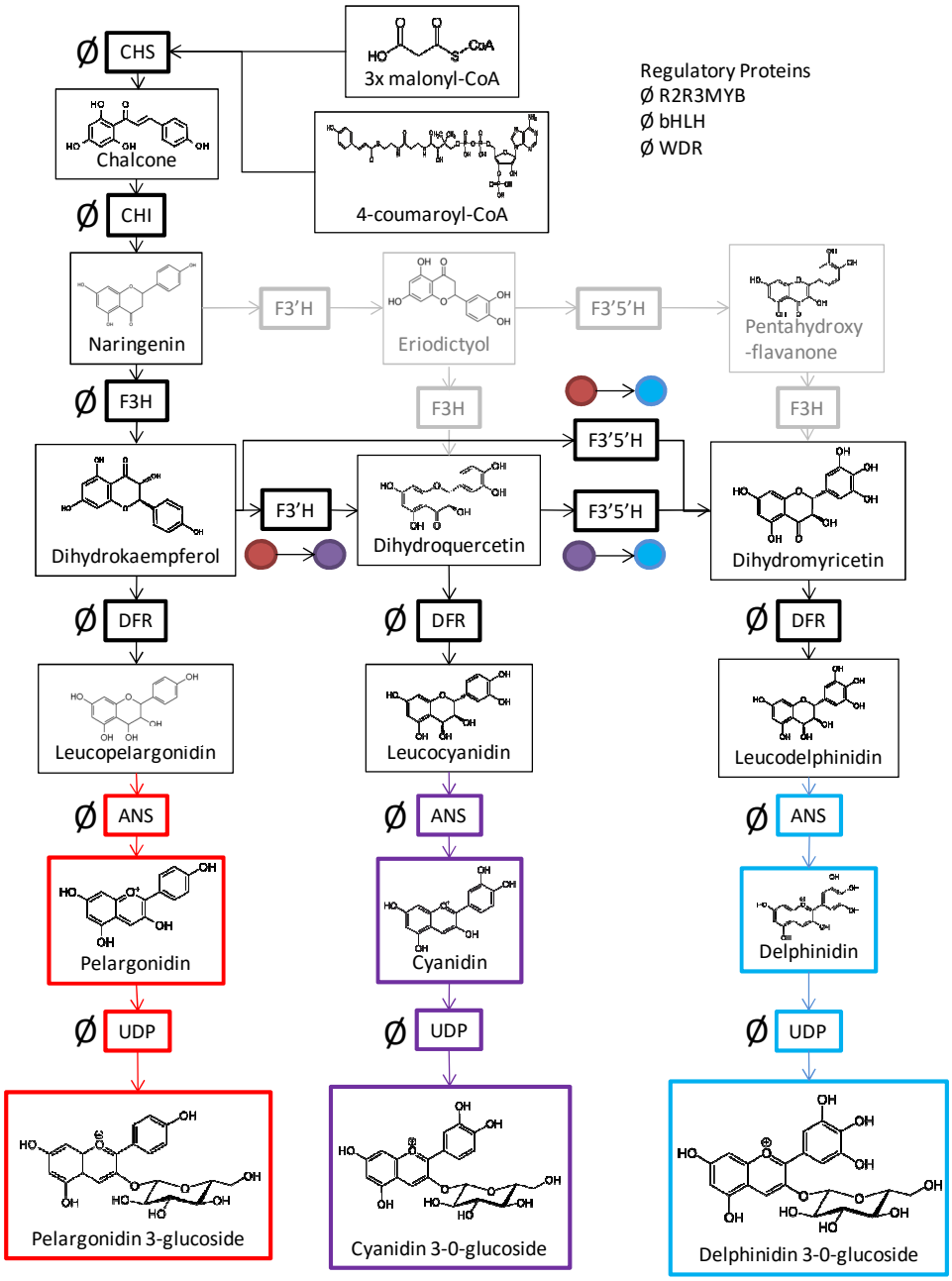


Figure 10: The anthocyanin synthesis pathway. The colour of the end products is mostly determined by the relative levels of flux through the three branches of the pathway. Inactivation of the F3'H and/or F3',5'H enzymes will force their precursor molecules into the remaining branch(es), changing flower pigmentation as indicated by the red, purple and blue circles. Transitions from coloured flowers to white flowers can be accomplished by loss of function mutations in enzymes and transcription factors and/or downregulation of expression of those same enzymes and transcription factors marked with Ø. Reactions in grey are a secondary pathway. From Wessinger and Rausher (2012)[121].

Changes in relative flux through each of the branches of the pathway are responsible for changes in the relative concentrations of the anthocyanins produced by the pathway and the resulting floral colour, with changes from red/pink to blue indicating more delphinidin than cyanidin for instance.

2.3.3 Anthocyanins in *Erica* species

The major floral pigments in *Erica* have been determined using paper chromatography, and were found to be anthocyanins: cyanidin-3-0-glucoside, delphinidin-3-0-glucoside, pelargonidin-3-0-glucoside, malvidin-3-0-glucoside and peonidin-3-0-glucoside[163]. In all 39 *Erica* species tested cyanidin-3-0-glucoside was present, with delphinidin-3-0-glucoside also common. The methylated glucosides, malvidin-3-0-glucoside and peonidin-3-0-glucoside, were less frequent[163].

2.3.4 Control of anthocyanin biosynthesis

Understanding of the mechanisms of control of the synthesis of anthocyanins has increased greatly in the past decades. The original simplistic understanding was that DFR and or CHS controlled the flux through the pathway[164]. This simplistic model has been replaced as molecular techniques have improved and the current best model is unified control of the synthesis of the enzymes in the pathway, under the control of *trans*-acting regulatory factors[154,162,165,166]. All of the anthocyanin synthesis enzymes' genes are controlled by the same highly conserved *trans*-acting regulatory factors; namely a complex of three subunits: IpMYB1-IpbHLH2-IpWDR1 in *Ipomoea purpurea*[154] or PAP1-GL3-TTG1 in *Arabidopsis thaliana*[162,166]. This complex of *trans*-acting regulatory factors binds to recognition sites upstream of the start of transcription of each gene. These recognition sites are a seven base pair motif bound by the R2R3 myeloblastosis transcription factor (MYB), with sequence ANCNNCC, known as the MYB Recognition Element (MRE); and a six base pair motif bound by the basic Helix-Loop-Helix transcription factor (bHLH), with sequence CACNHK, known as the bHLH Recognition Element (BRE). The BRE and MRE are always located within 400bp of the start of transcription and within 80-120bp of one another[154]. The BRE is always 5' of the MRE. These motifs are common to numerous species and are found in angiosperms and gymnosperms. The presence of these promoter motifs upstream of every gene of the pathway provides a mechanism for unified control of the pathway[154].

Transitions from coloured to white flowers tend to be as a result of either mutations that affect regulation of expression or loss of function mutations. These can be further divided into mutations that affect the regulatory genes and those that affect the anthocyanin synthesis genes. Amongst the spontaneous mutations that affect flower colour, the most common are transcription factor mutations, primarily in bHLH and WDR, that affect their ability to bind to the recognition sites,

followed by loss of function mutations in the anthocyanin synthesis genes and then mutations in promoter binding sites of the anthocyanin synthesis genes (46%, 42% and 12% of cases documented by Streisfeld and Rausher (2011)[167] respectively). In populations where the flower colour change has been fixed, and is heritable, the most common mutations are mutations in the transcription factors, mostly in MYB, indicating that MYB mutations are more tolerable[121].

Should a loss of anthocyanin synthesis be found, transcriptomics provides an excellent method for determining which one of either transcription factor gene mutations, loss of function mutations in the anthocyanin synthesis genes or mutation(s) in promoter binding sites of the anthocyanin synthesis genes are the cause. If none of the genes of the pathway are expressed, then the cause is probably mutation to the transcription factor gene(s). If only one of the genes of the pathway is not expressed, then the cause is probably a mutation or mutations to its transcription factor binding site(s). If all of the genes are expressed but no anthocyanins are produced, then it is probable that one of the genes has suffered a loss of function mutation i.e. a mutation in the coding region. Sequencing of the transcription factor genes, the promoter regions of the single gene or HPLC characterisation of the intermediates in the pathway would indicate where the causative mutation may have occurred.

2.4 Objectives:

The first major objective of this study was to sample most of the ca. 865 *Erica* species and sequence several chloroplast and nuclear ITS markers in the sampled *Erica* species. The objective was then to construct a multigene, maximum likelihood tree of the entire genus and to elucidate the relationships between the European, African and Cape *Erica* species, to test if there have been numerous shifts in biogeographic range or morphology, to test the hypothesis of Barnes *et al.* (1995)[148] that nectar sugar composition may be a useful source of taxonomic information, and if red flowers have a single common ancestor or have evolved independently.

The second major objective was to focus on the *Erica plukenetii* clade, where floral colour is very labile, with multiple populations showing colour polymorphisms. The objective was to locate the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes using Next Generation Sequencing (NGS) and BLAST searches in the data. The expression of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes in different coloured flowered samples was studied with the objective of determining a possible mechanism for the observed floral colour changes in *Erica plukenetii.* This methodology can then be expanded to test for the genetic changes underlying the phenotypic colour changes observed in other species in the genus.

Chapter 3: A molecular phylogeny of the genus *Erica*

3.1 Introduction

Prior to the development of NGS techniques and the related dramatic increase in sequence data volumes, it was possible for alignments of nucleic acid sequences to be done manually as the number of taxa in the groups to be analysed was usually small and length of the sequences was limited by Sanger Sequencing technology. Some computer algorithms were used but computational power was limited. In, for example, Clustal, pairwise alignments of sequences were performed and then each pair is progressively aligned with the next most similar sequence[168]. The manual approach has several pitfalls[168]. Parsimony has to be used rather than maximum likelihood, but the most parsimonious alignment is not necessarily the most likely. The analysis assumptions, i.e. gap cost, have to be used consistently and those assumptions have to be repeatedly tested for sensitivity. Finally, there is potential for bias as the alignments are not global, but rather along the sequence from one end to the other and sequences are added sequentially to the alignment, increasing the likelihood that they will be fitted to the previous data rather than evaluated holistically.

However, sequence data sets are now huge and manual alignments are not feasible. High-throughput pipelines for handling the sequence data in an automated way have had to be developed. Fortunately computational power has grown incredibly rapidly, following the predictions of Moore's Law[169], and a variety of algorithms that can be implemented to align large data sets have been developed, including Clustal[170], ClustalW[171], PROBCONS[172], MAFFT[173], PRALINE[174], Scaccato[175] and MUSCLE[176]. Choosing between these algorithms can be difficult. Fortunately benchmarking tools such as PREFAB, SABMARK and IRMBASE have been developed to make the task easier[177–179]. In these benchmarking tests MAFFT has been shown to perform as well as, if not better than, competing algorithms[180]. These alignment algorithms have several advantages. The assumptions that they are based on are applied with greater consistency. Furthermore, they implement maximum likelihood rather than parsimony, i.e. returning the most probable alignment rather than necessarily the shortest alignment. The alignments are performed globally and simultaneously rather than pairwise. Sensitivity testing is incorporated automatically. Finally, the results are reproducible by other researchers[181]. Likewise, post-alignment processing of the data to remove poorly aligned regions that could affect further analyses can also be automated and the alignments trimmed using trimAl[182].

The computational algorithms used for alignment are not without their criticisms, mostly on the method of scoring homology. Homology is a multi-level, hierarchical concept that goes beyond the simplistic direct comparison of character states and scoring them based on that comparison.

Characters can be homologous at base pair, amino acid, protein, functional and phenotypic level amongst others. Few of the DNA alignment algorithms take this into account, with the exception of MAFFT 7.0 and Staccato, that heuristically combine compositional similarity and topology[183,184]. Partitioning the data into coding, non-coding and/or other conserved regions is also used to address this issue[185].

Phylogenies in *Erica* in the past have used manual alignments of small datasets, with McGuire and Kron (2005)[6] comparing three sequence regions and 26 species and Pirie *et al.* (2011)[7]using one sequence region and 477 accessions. A more complete phylogeny of the genus has recently been published by the research group of which the author is part (Appendix A). This study included 597 accessions representing 488 of the ca. 865 species in the genus. In the published phylogeny, alignments of the multiple gene regions were performed manually, using parsimonious principles, with sequences added consecutively, with occasional reordering of the taxa to reduce the potential for bias.

In this study, RAxML maximum likelihood trees generated from the manual alignment of a number of gene regions of *Erica* species will be compared to RAxML maximum likelihood trees of a MAFFT alignment of the same dataset using TOPD/FMTS[186] to compare tree topologies and test for significant differences between the trees resulting from the manual and automated alignment. TOPD/FMTS uses a variety of algorithms, including "Nodal", which calculates the distance between the input trees, using the root-mean-square-distance (RMSD). Visual inspections of differences between the topologies were accomplished using the online tool at www.phylo.io. Characters will be mapped onto the phylogeny to identify any clades that correspond to a particular biogeographical range, have consistent morphology, are ornithophilous, and/or have red flowers. Finally, the hypothesis of Barnes *et al.* (1995)[148] that nectar sugar composition may be useful source of taxonomic information will be tested.

3.2 Materials and Methods

3.2.1 Sampling

Erica species were collected from the entire species range. Sixty percent of the CFR species, 53% of the Drakensberg species, 42% of the Madagascar and Mascarene islands species, 57% of the tropical east African species, 89% of the European and Middle Eastern species as well as six outgroups were collected.

3.2.2 DNA extraction, gene region amplification and sequencing

Direct PCR amplification of up to nine gene regions from the leaves of *Erica* species was performed using the direct amplification protocol developed by Bellstedt *et al.* (2010)[69] rather than a DNA extraction method. PCR amplifications were carried out using the Kapa Biosystems 3G Plant PCR kit. Each 25µl PCR mix contained 12.5µl Buffer, 2µl 25mM $MgCl_2$, 0.75µl 20mM Forward and 0.75µl 20mM Reverse Primers (see Table 2), 0.2µl 2.5U/µl 3G Plant Taq, 0.2µl 5% DMSO, 1µl of plant homogenate and 8.6µl milliQ H_2O in an Applied Biosystems Veriti PCR thermal cycler. The PCR programme used was 2 minutes at 95°C; followed by 35 cycles of 95°C for 30 seconds, annealing at the temperature given in Table 2 for 30 seconds, and extension at 72°C for 30 seconds; followed by a final extension step of 72°C for 5 minutes.

Primer sequences for the tabF, tabF Reverse Complement (RC) and ndhJ primers were modified from Taberlet *et al.* (1991)[187] to allow for polymorphisms present in *Erica* species. Due to the presence of a 5' poly-A and/or a 3' poly-T repeat flanking the central sequence of the trnT-L region of the some of the *Erica* species chloroplast, internal primers (trnTL int F and trnTL int R) were designed to amplify that region specifically.

The ITS and the chloroplast trnT-L, trnL intron, trnL-F, trnF-ndhJ gene regions of all species were sequenced. The chloroplast atpI-H, rpl16 intron, matK, ndhF and rbcL gene regions of species, selected to represent larger clades identified in the ITS phylogeny[7], were sequenced. All primers were designed using Geneious R9[188] unless stated otherwise.

Sanger sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fischer Scientific) in an Applied Biosystems Veriti thermal cycler using the STeP sequencing protocol[189]. Each 10µl sequencing reaction contained 5µl Sequencing Buffer, 1µl 0.8mM Primer, 1µl Big Dye, 1µl of PCR product and 2.5µl milliQ water. Sequencing electrophoresis was performed at the Central Analytical Facility at Stellenbosch University.

3.2.3 Alignments of gene regions

Forward and reverse sequences were paired in ChromasPro v1.7.5 and checked for ambiguities and miscalled bases. The sequences were added sequentially to the pre-existing data set, and aligned by hand in Mesquite by Dr Michael Pirie, one of the senior authors of this research. The same, unaligned, data set was also aligned using the MAFFT algorithm in Geneious R9[188].

PartitionFinder[185] was used to infer the substitution models and best fitting partitioning strategy of the respective gene regions for both alignments. Potential data partitions were specified; these

included coding regions, non-coding regions, codon positions and markers. TrimAl was used to trim the automated alignment to remove any poorly aligned sequences.

Table 2: Primer sequences used for amplification and sequencing of chloroplast markers and ITS of Erica species. Characters in bold and underlined indicate modifications to primer sequences from Taberlet *et al.* (1991)[187]. RC indicates where a primer sequence is the reverse complement.

Region Amplified	Name	Sequence 5'-3'	Annealing Temperature (°C)
trnT-L	tabA[187]	CAT TAC AAA TGC GAT GCT CT	59.0
	tabB[187]	TCT ACC GAT TTC GCC ATA TC	
	trnTL int F	GCA TAT TAG AAA TGT CTA ATT ACT AAA TTA	
	trnTL int R	CGT CTA AAC TTA CAC CTT TAT GAA	
trnL intron	tabC[187]	CGA AAT CGG TAG ACG CTA CG	55.0
	tabD[187]	GGG GAT AGA GGG ACT TGA AC	
trnL-F	tabE[187]	GGT TCA AGT CCC TCT ATC CC	57.5
	tabF[187]	ATT TGA ACT GGT GAC ACG AG	
	tabF new	ATT TGA ACT GG**C** GAC ACG AG	
trnF-ndhJ	tabF RC[187]	CTC GTG TCA CCA GTT CAA AT	59.5
	tabF RC new	CTC GTG TC**G** CCA GTT CAA AT	
	ndhJ - C[187]	ATG CC**C** GAA AGT TGG ATA GG	
	ndhJ - T[187]	ATG CC**T** GAA AGT TGG ATA GG	
ITS	ITS4*[190]	TCC TCC GCT TAT TGA TAT GC	54.0
	ITSL*[190]	GTC CAC TG AAC CTT ATC ATT TAG	
	ITS8P[190]	CAC GCT TCT CCA GAC TAC A	
	AB101[190]	ACG AAT TCA TGG TCC GGT GAA GTG TTC G	
atpI-H	atpI[191]	TAT TTA CAA GYG GTA TTC AAG CT	55.0
	atpH[191]	CCA AYC CAG CAG CAA TAA C	
rpl16 intron	rpl16F71[191]	GCT ATG CTT AGT GTG TGA CTC GTT G	55.0
	rpl16R1516[191]	CCC TTC ATT CTT CCT CTA TGT TG	
matK	matK1F[190]	ATG GAG GAA TTC AAA AGA AAT TTA G	55.0
	matK1600R[190]	CCT CGA TAC CTA ACA TAA TGC	
ndhF	ndhF5F[190]	ATG GAA CAK ACA TAT SAA TAT GC	55.0
	ndhF1318R[190]	CGA AAC ATA TAA AAC GCA GTT AAT CC	
	ndhF972F[190]	GTC TCA ATT GGG TTA TAT GAT G	
	ndhF2110R[190]	CCC CCT AYA TAT TTG ATA CCT TCT CC	
rbcL	rbcL1F[190]	ATG TCA CCA CAA ACA GAA AC	55.0
	rbcL724R[190]	TCG CAT GTA CCT GCA GTA GC	
	rbcL636F[190]	GCG TTG GAG AGAT CGT TTC T	
	rbcL1460r[190]	TCC TTT TAG TAA AAG ATT GGG CCG AG	

*Internal primers used only for sequencing and not amplification.

3.2.4 Comparison of phylogenetic trees generated from manual and automated alignments

Trees were generated using the RAxML[192] Maximum Likelihood analysis on the CIPRES gateway using identical methodology for both the manual and the automated alignments. The RAxML

analysis incorporates the inferred partitions and substitution models from PartitionFinder. Bootstrapping was used to determine clade support, with the bootstrapping ended, automatically, using the majority rule "autoMRE" criterion. Single gene region analyses were performed so that individual marker trees could be compared to one another separately as well as to check for experimental error. Combined analyses of the nuclear markers, the chloroplast markers and the combined nuclear and chloroplast markers were performed. Fifteen taxa that showed significant topological conflict were excluded from further analyses, reducing the number of accessions from 612 to a final total of 597 accessions in the combined tree. Tred (http://www.reelab.net/tred/) and Geneious R9[188] were used to draw phylogenetic trees from the RAxML analysis. TOPD/FMTS[186], with the Method set to "Nodal", was used to calculate the topological distance between the trees. The online tool, www.Phylo.io, was used to visually compare the manually aligned tree to the MAFFT aligned tree.

3.2.5 Character mapping

Biogeographical characters, some morphological characters, ornithophily and red flowered species were mapped on the phylogeny[147]. Nectar sugar composition from Barnes *et al.* (1995)[148] was also mapped. Pagel's λ test was used to test if nectar sugar composition of the 40 species in the tree for which data was available[148] is phylogenetically informative.

3.3 Results

3.3.1 Sampling

A complete list of all species and collections that were analysed in this study is in Appendix B.

3.3.2 DNA extraction, gene region amplification and sequencing

The nuclear and chloroplast markers of 255 accessions were sequenced. A complete list of which markers were sequenced for each accession is in Appendix B.

3.3.3 Alignments of gene regions

Sequences from 255 accessions were combined with those sequenced by other members of our group, bringing the total to 612 accessions and 489 species in the alignment.

3.3.4 Comparison of phylogenetic trees generated from manual and automated alignments

Alignment of the data matrix using the MAFFT, trimAl, PartitionFinder and RAxML automated pipeline was much faster than the manual approach and easy to accomplish.

When comparing the combined chloroplast and nuclear marker trees from the manual and the MAFFT alignments, TOPO/FMTS calculated a RMSD score of 0.03. This indicates that very little distance, i.e. difference in tree topology, exists between the tree from the MAFFT alignment and the tree from the manual alignment. A score of zero would indicate complete congruence between the trees and a score of 1 would indicate completely different trees. Visual inspection of the differences using Phylo.io found that the major topology of the trees is identical and that differences are restricted to minor rearrangements of clades within the Cape clade and to the tips of branches only.

3.3.5 Tree topology

The phylogeny in Figure 11 shows that the genus *Erica* forms a monophyletic group within the Ericaceae. A more detailed figure is in Appendix C. The genus divides into a European grade that subtends the African and Mascarene clade. The African and Mascarene clade consists of a polytomy of an African-Mascarene clade, a Cape clade, *E. trimera*, *E. pauciovulata* and *E. arborea*. Within the African-Mascarene clade Tropical East African, Mascarene and Drakensberg subclades are found. The Cape clade excludes only one Cape species (*E. pauciovulata)* and includes two species that occur in the Drakensberg and two species that are distributed from the CFR into the Drakensberg.

African *Erica* species have descended from the European species. Mascarene and Tropical East African *Erica* species may share a common ancestor with the Cape *Erica* species. The Bootstrap support for the Cape clade is rather low at 70, but if *E. pauciovulata* is removed, bootstrap support for the Cape clade rises to 89 and all other topology remains identical.

The trnT-L sequence of *E. pauciovulata* is unique in containing a region flanked by a 5' poly-T repeat and a 3' poly-A repeat that is significantly different to other species' trnT-L sequences. Poly-A and poly-T repeats are known to cause slippage during replication[193]. In other species, the 5' poly-A is broken by the presence of a T, eight to ten bases from the start of the region. This single T is sufficient to reduce slippage to the extent that the sequences from other species have undergone much less mutation in the same time frame. This slippage was also noticed in the sequencing chromatograms and this necessitated the use of internal sequencing primers as indicated in section 3.2.2.

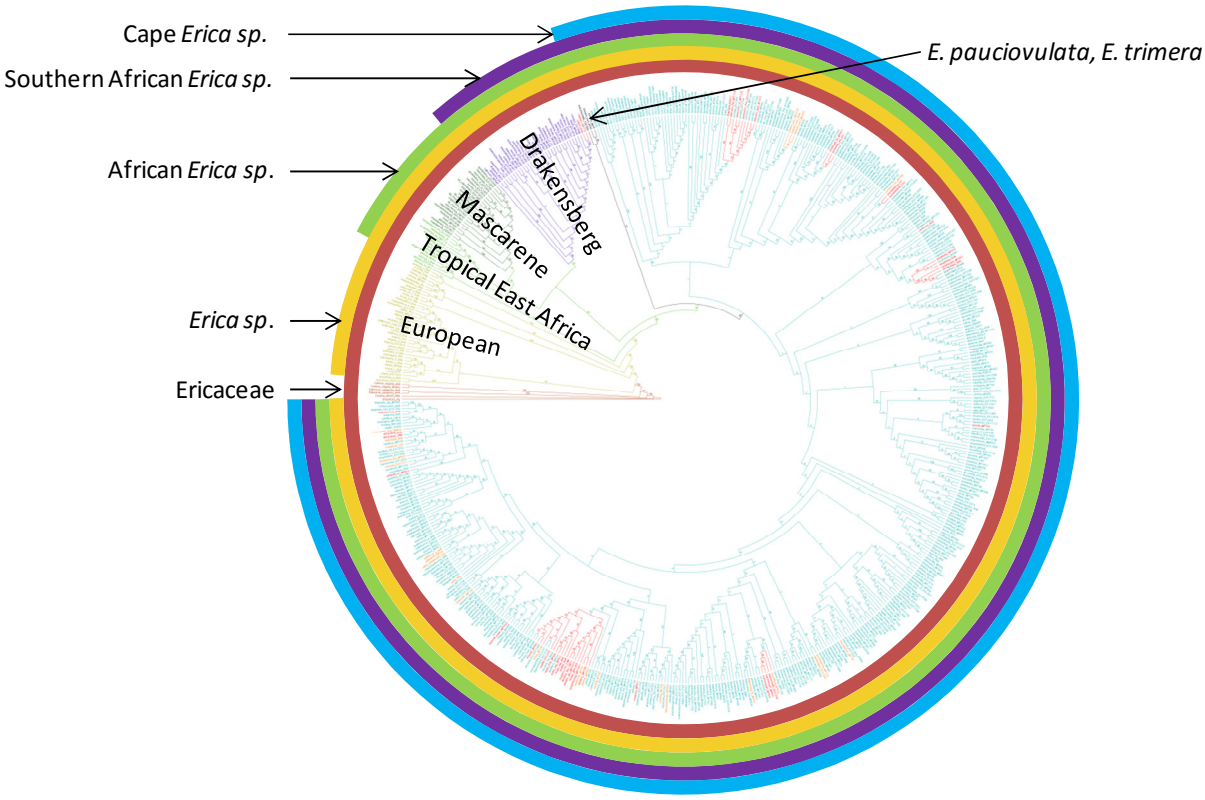


Figure 11: RAxML Tree generated by CIPRES. *Erica* species (yellow arc) form a clear monophyletic group within the *Erica*ceae (red circle). African *Erica* species (green arc) clearly descended from the European *Erica* species. Tropical East African (names in light green), Mascarean (names in dark green), Drakensberg (names in purple) and Cape clades (names in blue) are distinguishable. *E. pauciovulata and E. trimera* are indicated in black. Names of species and species clades where red colour and ornithophily have evolved are indicated in red. Species names in orange indicate ornithophilous species from Rebelo *et al.* (1985)[147].

3.3.6 Character mapping

The distribution data for *Erica* species in the CFR accumulated in the Precis database of the South African National Biodiversity Institute was incomplete and consequently, no ancestral state reconstructions could be performed. The data was therefore also not suitable for mapping onto the phylogeny. Instead, distribution data provided by Dr. EGH Oliver, the recognised expert on the *Erica* genus, was used to identify small clades that share a particular habitat or range.

Within the Cape clade, similar to the ITS phylogeny[7], species are distributed into smaller clades that are either clades of species with narrow distributions occupying a particular biogeographical range, or clades of widely distributed species that occupy a variety of diverse habitats (Figure 12). As with the ITS phylogeny, there are relatively few morphologically based groupings. The following 19 clades confined to narrow biogeographical ranges could be distinguished (Figure 12, green, numbered boxes): a clade distributed from the Langeberg to Outeniqua mountains, three Langeberg clades, a Klein Karoo clade, a coastal clade distributed from Kogelberg to Hermanus, a Southern Cape and Eden clade, two Kouebokkeveld clades, two Cape Peninsula clades, a Cederberg clade, a Riviersonderend Mountains clade, a clade of species only found on alkaline soils, a Swartland clade, a clade distributed from the Hottentots Holland Mountains to Hermanus, a Limietberg clade,

two South Western Cape clades, and a clade distributed from the Wemmershoek Mountains to the Hexrivier Mountains. Six clades showing the following morphology could be distinguished (Figure 12, red, lettered boxes): two clades of species with long tubed flowers with narrow distributions, a clade of species with exerted stamens, a clade of wind pollinated species, a clade of widespread species with long tubed flowers and a clade of species with enlarged calyces.

Red colour has independently evolved 14 times in single species or species clades namely: *E. oatsii*, *E. discolour* clade, the *E. plukenetii* clade, the *E. coccinea* clade, *E. cameronii*, the *E. cerinthoides*/*E. sparmanii* clade, *E. tumida,* the *E. mammosa*/*E. sessiliflora* clade, *E. massonii,* the *E. viscaria*/*E. abietina* clade, *E. haematocodon*, *E. pillansii*, *E. annectans* and *E. leucotrachela.* These are indicated in red in Figures 11-13 and in Appendix C.

Ornithophilous, but not red flowered, *Erica* species as designated by Rebelo *et al.* (1985)[147] are indicated in orange on Figures 9-11 and in Appendix C. Ornithophily has also evolved independently, with 21 ornithophilous, but not red flowered species indicated.

The nectar sugar composition data of the *Erica* species studied by Barnes *et al.* (1995)[148] (Figure 13) was mapped on to the phylogeny. Pagel's λ test, on a tree pruned to the branches for which nectar sugar composition data was available[148], returned a $\lambda = 0.9814624$, with $p = 0.02418609$, indicating a nearly pure Brownian motion model of evolution. A λ of one would indicate purely random evolution. It can therefore be concluded that species with similar sucrose:hexose ratios show no phylogenetic relationships with one another, clearly having evolved independently of one another.

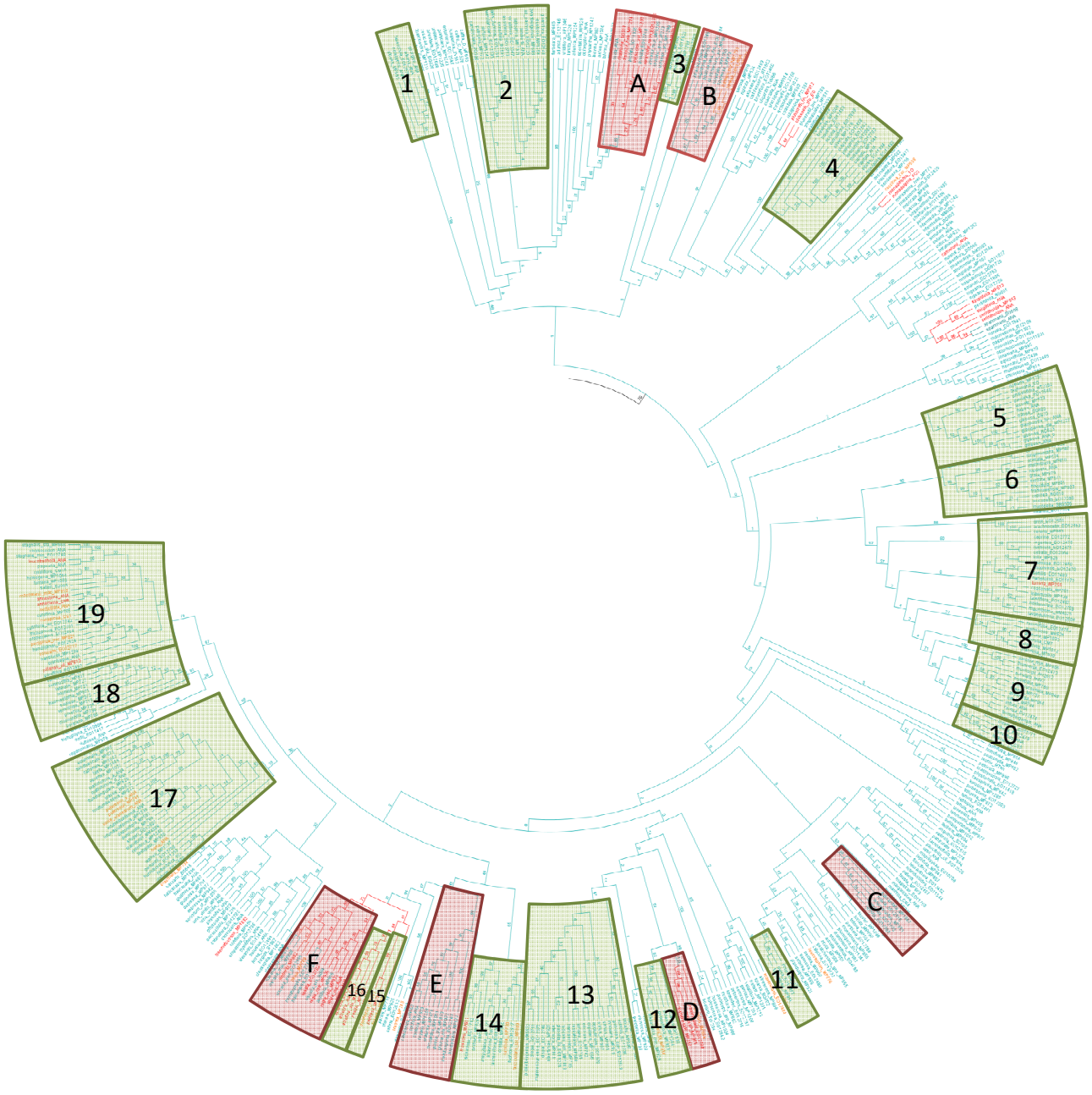


Figure 12: The Cape of Erica species. Clades where species occupy a narrow biogeographical range (green, 1-19) or share morphological characteristics (red, A-F) are indicated. Names of species and species clades where red colour has evolved are indicated in red. Species names in orange indicate ornithophilous species from Rebelo *et al.* (1985)[147]. 1: Langeberg to Outeniqua. 2: Langeberg. 3: Klein Karoo. 4: Coastal, Kogelberg to Hermanus. 5: Southern Cape, Eden. 6: Langeberg. 7: Kouebokkeveld. 8: Cape Peninsula. 9: Kouebokkeveld. 10: Cederberg. 11: Riviersonderend Mountains. 12: Alkaline soils. 13: Swartland. 14: Hottentots Holland to Hermanus. 15: Limietberg. 16: Cape Peninsula endemics. 17: South western Cape. 18: Wemmershoek to Hexrivier Mountains. 19: South western Cape. A: Long tubed flowers. B: Exerted stamens. C: Wind pollinated. D: Widespread with long tubed flowers. E: Enlarged calyx. F: Long tubed flowers.

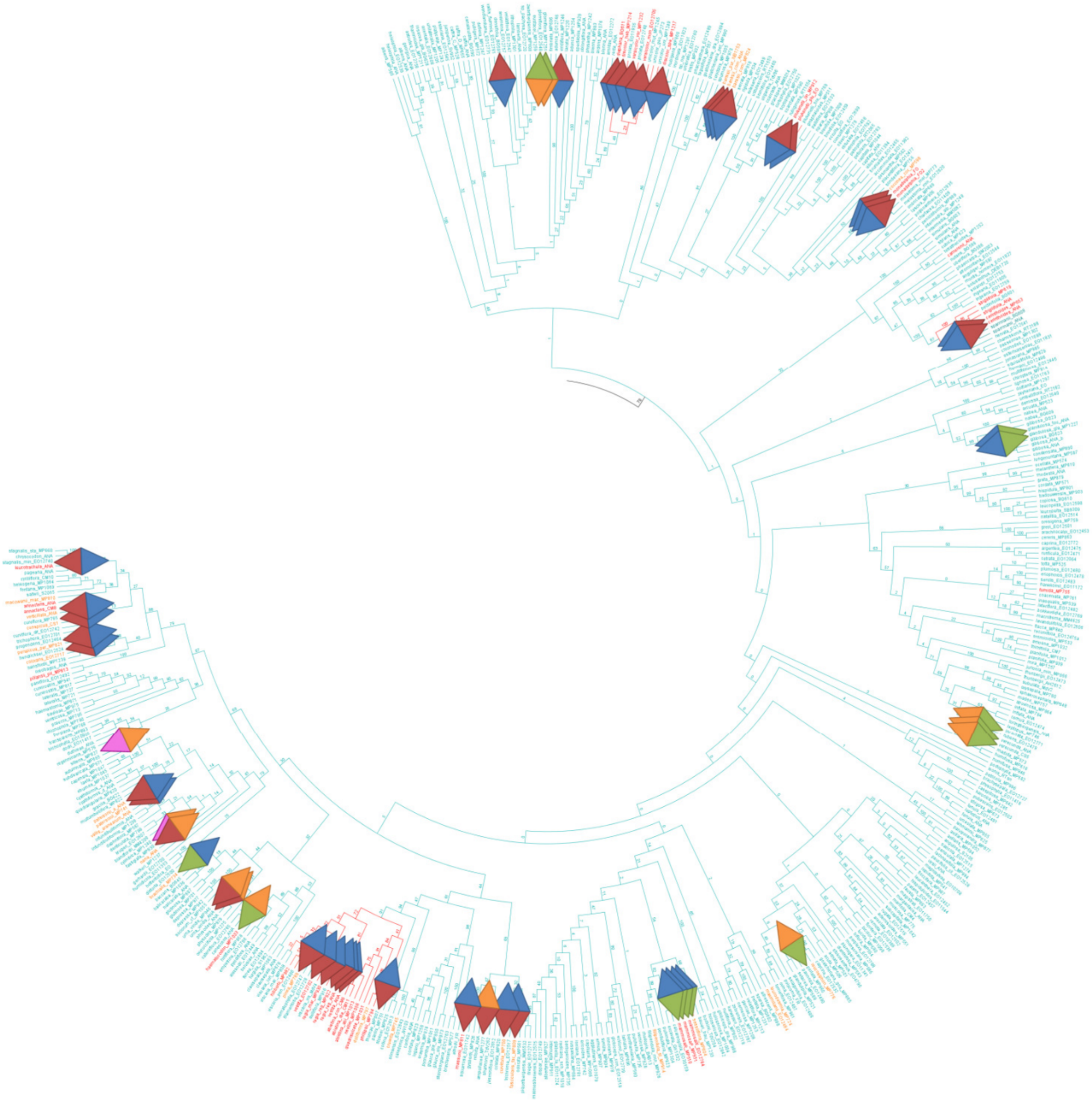


Figure 13: Nectar sugar composition of *Erica* species for which data is available. Red triangles indicate high sucrose, pink triangles indicate intermediate sucrose, green triangles indicate low sucrose. Blue triangles indicate ornithophily and orange triangles indicate entomophily. Names of species and species clades where red colour has evolved are indicated in red. Species names in orange indicate ornithophilous species from Rebelo *et al.* (1985)[147].

3.4 Discussion

The lack of significant differences between the trees generated from the manually aligned phylogeny and the MAFFT aligned phylogeny, for such a large data set, is encouraging. The use of algorithms to perform the alignments, provided that they are used correctly, is fast, efficient, removes the possibility of bias and makes the results very reproducible. It generates maximum likelihood, exhaustive, alignments rather than the parsimonious, non exhaustive, alignments resulting from the manual approach. Some manual checking of the data is inescapable but it is far less than the manual approach requires.

The conflict introduced by *E. pauciovulata*, and the subsequent increase of bootstrap support for the Cape Clade following its removal, is likely an artefact of the differences between its trnT-L sequence and those in the rest of the phylogeny. Its trnT-L sequence is unique in containing a region flanked by poly-A and poly-T repeats. These repeats likely lead to slippage during replication, increasing the mutation rate in the region they flank, significantly above what would be expected. When the sequence is then aligned with the rest, it appears to be very divergent. The substitution model for a partition is assumed to be consistent across all the sequences in the partition but this assumption does not hold for this sequence. Furthermore, the hypervariability in this region may introduce homoplasy.

E. arborea and *E.trimera* were sequenced for all the gene regions and their position in the phylogeny is a likely a true reflection of their relationships within the *Erica* genus. *E. trimera* shares a common African ancestor with the Cape clade and similarly *E. arborea* shares a common European ancestor with the African and Mascarene clade.

This phylogeny provides further, unequivocal support for the hypothesis that European *Erica* species spread southwards through the Middle East and/or crossed the Mediterranean Ocean into Africa and that the "Cape-to-Cairo" theory does not hold true for *Erica* as McGuire and Kron (2005)[6] and Pirie *et al.* (2011)[7] had also concluded.

McGuire and Kron (2005)[6] found a sister relationship between *E. arborea* and other Cape species and inferred that *Erica* had spread using the mountains of Tropical East Africa. In contrast, both Pirie *et al.* (2011)[7] and this study do not support this conclusion, with the relationships between *E. arborea* and the African *Erica* species lacking in resolution. Also given a sister-group relationship of *E. arborea* (widespread across Europe and East Africa) and *E. lusitanica* (exclusively European) implied by the ITS data[190], a better supported explanation is that the Mascarene and Tropical East African species are the result of independent colonisation events.

Resolution of the basal relationships between the clades within the Cape clade is sometimes adequate to discern the basal relationships but often is not sufficiently well resolved. Better resolution could be obtained with more sequence data for more marker genes.

Within the Cape clade, multiple sub-clades can be distinguished. Primarily these sub-clades are groups of species that share a biogeographical range. This would indicate that diversity of the Cape *Erica* species is likely as a result of multiple local, *in situ*, radiations within each biogeographic range from a single founder species in that specific range, throughout the Cape region. These radiations have primarily occurred within ecological niches (Figure 12) that correspond to altitude,

climate zones, rainfall levels, or soil types (Oliver, pers. comm.). These niches correspond well with the phytogeographical centres identified by Oliver *et al.* (1983)[15] (Figure 7).The Southern centre corresponds to clades 1, 2, 3, 5, 6 and 11; the Bredasdorp centre to clade 12; the Overberg centre to clades 4 and 14; the Cape Peninsula centre to clades 8 and 16; the South Western centre to clades 17 and 19; the Northern centre to clades 7, 9, 10, 13, 15 and 18. The fit between the phytogeographical centres from Oliver *et al.* (1983)[15] and the clades on the tree with a narrow biogeographical range would imply that the drivers of the radiations in *Erica* are likely similar to those of the other families and/or genera in these centres. The anomalous *E. pauciovulata* would be expected to fall into one of the southern Cape groups as it is endemic to the Bredasdorp area.

Mountainous areas have clearly played a role in the diversification of the CFR *Erica* species, as they have with other CFR species, such as *Protea* and *Leucodendron*[63], and with other genera in the Ericaceae[25]. Of the 19 clades in which the species share a biogeographical range, 11 correspond to areas in the Cape Fold mountains. The original founder species colonised the biogeographic range and then diversified, possibly driven by the influence of the rugged topography which generated numerous ecological niches. The cooler, wetter climate typical of the mountains has also likely provided a refuge that has allowed the persistence of lineages.

The presence of a clade endemic to the alkaline soils[194] of the Agulhas Plain in the south western Cape, shows a similar pattern to the clades containing species restricted to mountainous areas, with initial colonisation followed by diversification within that niche.

Fire survival strategy has played a role in *Erica* diversification The clades of species with narrow biogeographical ranges in the south and west of the CFR mostly contain reseeder species, whereas resprouters such as the *E. cerinthoides* clade, the *E. discolor* clade and the *E. coccinea* species complex have wide distributions across the CFR from east to west and do not fit into a particular biogeographical range, echoing the results of Ojeda (1998)[44]. The reseeder strategy has been shown to lead to greater diversity as generation times are shorter[46,47] which has likely contributed to the diversification. *E. cerinthoides*, *E. discolour* and *E. coccinea* are all also ornithophilous species and the greater range of their pollinators has likely increased gene flow between populations and reduced the effects of genetic drift.

Relatively few clades with species sharing a common morphology could be identified: most of these clades (A, B, D and F in Figure 12) have long tubed flowers, or other adaptations characteristic of insect pollination (Oliver, pers. comm.). The lack of strong morphological signal, the multiple independent evolution of red colour and the multiple independent evolution of ornithophily across the tree confirms what Pirie *et al.* (2011)[7] had concluded, namely that floral morphology in *Erica* is

highly labile and multiple shifts have occurred between anemophilous, entomophilous and ornithophilous forms. With the exception of the *E. cerinthoides* clade, the *E. discolor* clade and the *E. viscaria* clade, red flower colour is mostly confined to the tips of branches, indicating that it is a relatively recent change from the plesiomorphic state.

Barnes *et al.* (1995)[148] suggested that the nectar sugar composition of *Erica* species could be a useful source of taxonomic information. Mapping their data onto the phylogeny, and testing using Pagel's λ, showed this hypothesis could be rejected, as the species with similar sucrose:hexose ratios were randomly distributed across the phylogeny. Due to the bias in the data from Barnes *et al.* (1995)[148], with 37 of the 50 species studied being ornithophilous and only 13 entomophilous (with only 40 species represented in the phylogeny), and the lack of data for more than 90% of the other accessions in the phylogeny, further studies with a more complete data set are required before the hypothesis that nectar sugar composition is correlated to phylogeny can be rejected.

Several of the smaller clades are ideal for further in depth analysis of the root cause of the floral morphology shifts, as they contain very closely related species that differ with respect to floral colour and/or shape. Of these, the *E. plukenetii* clade (Figure 14) is particularly well suited as *E. plukenetii* is highly polymorphic with respect to flower colour with red, pink and white individuals found commonly alongside one another and the clade contains closely related yet morphologically distinct subspecies or sub-clades whose formation has been shown to have been influenced by pollinator shifts[70].

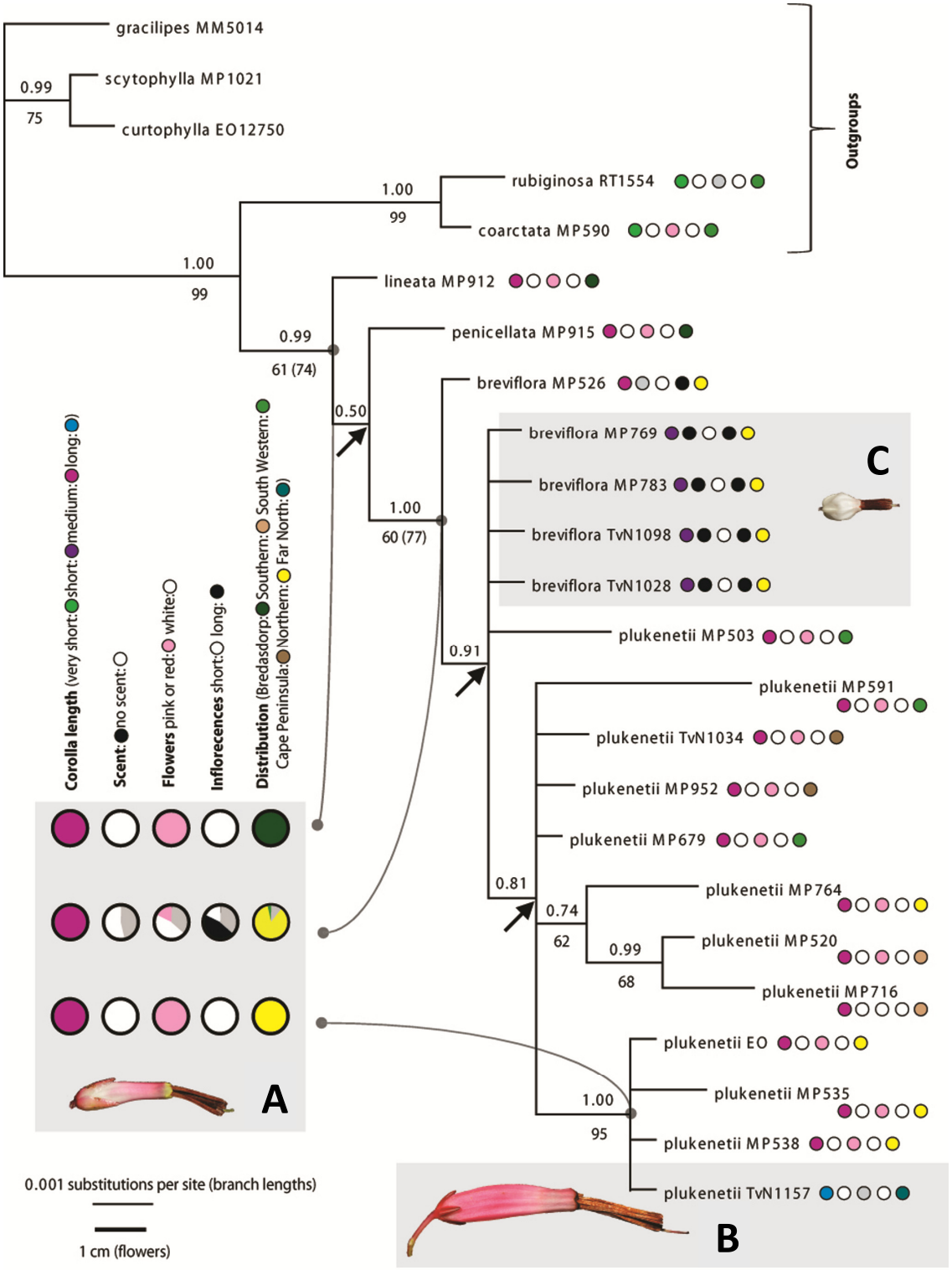


Figure 14: The *E. plukenetii* clade from van der Niet *et al.* (2014)[70]. Phylogenetic tree of Erica plukenetii subspecies and outgroups, from a MrBayes majority rule consensus based on chloroplast and nuclear DNA markers, with parsimony bootstrap values below and clade posterior probability above the branches (values in parentheses are based on the matrix without missing data). Arrows indicate clades that break down in the shortest strict consensus trees. Pie charts represent Fitch parsimony ancestral state reconstructions and coding of samples for five characters summarized over 1000 MrBayes trees: character states are as indicated by the colour code, with ambiguity or unknown states indicated by grey. A) Ancestral state, pollinated by short billed Orange Breasted Sunbird. B) Long corolla pollinated by long billed Malachite Sunbird. C) Subspecies breviflora with a short, white corolla, pollinated by moth.

Chapter 4: NGS approaches to characterising the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes and expression studies in colour polymorphic *Erica plukenetii*

4.1 Introduction

Many factors may have contributed to the radiation of the flora in the CFR. Evidence suggests that large scale climatic stability[16], small scale climate variations[38], topological gradients[63], establishment of a regular fire regime[195], and shifts between pollination syndrome[45] have driven significant speciation in the CFR. Most studies focusing on the role of pollination shifts have concentrated on geophytic plants, such as *Gladiolus*[196,197], in the CFR. Relatively few studies have focused on the role of pollinators in the speciation of the shrubby plants that make up a majority of the fynbos vegetation[70,94]. *Erica* is typical of the shrubby CFR vegetation and has been shown to have extensive floral variation and anemophilous, entomophilous and ornithophilous species are commonly found[140,147], and even two rodentophilous species[141,142]. Further, from the multigene phylogeny as outlined in Chapter 3 and the ITS phylogeny[7], it can been seen that floral morphology in *Erica* is evolutionarily very labile and has shifted from the plesiomorphic, European, form with short, pink floral tubes to forms with long floral tubes and multiple colours, including white, red, and yellow.

The determination of the causes of switches in pollination syndrome is complex and challenging. Using the terminology developed by Thompson and Wilson (2008)[73], the pollinator syndrome "vortex" that the plant in question occupies must be destabilized and additionally there must be sufficient "push" and "pull" factors exerting influences of sufficient strengths to allow it to move between "vortices"[73].

Floral colour is largely determined by the accumulation of a variety of pigments in the flowers, the anthocyanins, carotenoids and the betalains[162]. These pigments, primarily anthocyanins, colour flowers and have various protective roles in the plant tissues[154,159,198]. The anthocyanins are products of the anthocyanin biosynthesis pathway[162]. In *Arabidopsis*, the anthocyanin biosynthesis pathway synthesises anthocyanins via three branched pathway, consisting of eight enzymes and regulated by a complex of three *trans*-acting factors, from malonyl-coA and 4-coumaroyl-coA[199].

Red coloured flowers are a method of excluding insect pollinators as insects typically have poor visual perception in the red spectrum and consequently take much longer to find red flowers[136]. Birds have greater visual acuity in the red spectrum the red colour is typically more easy for birds to perceive[136]. Floral colour shifts have frequently been shown to influence the shift between pollination syndromes[72,121,167,198]. Floral colour changes from red to white have attributed to various

loss of function mutations in the anthocyanin pathway genes, their *trans*-acting regulatory genes and/or mutations in *cis* regulatory motifs upstream of the start of the gene[121]. These *cis* regulatory motifs have been shown to be conserved across angiosperms with a bHLH Recognising Element and an MYB Recognising Element present[154].

In the *E. plukenetii* clade, closely related sub-species appear to have differentiated within a pollinator syndrome and between pollinator syndromes[70]. The typical red flower, pollinated by the relatively short billed Orange Breasted Sunbird, has a corolla length of 9-22mm. In the northern range of *E. plukenetii*, where Orange Breasted Sunbirds are not present, corolla lengths have increased to 22-40mm, to accommodate Malachite Sunbirds with longer bills. *E. plukenetii* ssp. *breviflora*, with a corolla length of 0-9mm, has become entomophilous. Floral colour has changed from red to white, branches are weaker and less able support the weight of birds and floral scent has increased.

Little is known about the anthocyanin pathway in *Erica* generally, beyond that anthocyanins are produced, and specifically cyanidin is produced in *Erica plukenetii*[163]. Therefore, investigations of the role of changes in the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes in *E. plukenetii* would first require the determination of the sequences of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes.

The initial approach was to use NGS transcriptomics and sequence total mRNA from red and white flowers and compare expression levels. Initial attempts were made to isolate sufficient quantities of high quality RNA necessary for NGS transcriptomics but these failed. The presence of high levels of waxes and oils in *Erica*[200], which co-isolate with the mRNA, during phenol-guadinium thiocyanate extractions made the extracted mRNA unsuitable for further NGS transcriptomic analyses. Column extractions also failed to yield sufficient quantities of mRNA for further NGS transcriptomic studies as the waxes and oils blocked the columns. An approach in which Poly-A tailed magnetic beads were used was also unsuccessful.

Consequently it was decided to use a DNA based approach, as DNA is less susceptible to extraction problems than mRNA, and find the anthocyanin pathway genes and their *trans*-acting regulatory genes directly. Once these genes had been identified, their expression patterns could be determined.

The DNA of a red flowered *E. plukenetii* was therefore subjected to a number of NGS sequencing runs. BLAST searches of the NGS data were then used to locate the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes. Once the

contiguous sequences containing all or part of each gene of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes were found and primer walking approaches were used to attain any missing sequences. Primers were then designed to amplify their expressed sequences. Small quantities of mRNA required for Reverse Transcriptase quantitative PCR (RT-qPCR) were extracted. The role changes in anthocyanin pathway genes and/or their *trans*-acting regulatory genes expression plays in floral colour changes in samples from two populations containing colour polymorphic individuals was tested using RT-qPCR.

4.2 Materials and methods

4.2.1 *Erica plukenetii* sample collection

A red flowered *E. plukenetii* sample was collected from Table Mountain for NGS sequencing. For expression studies red-, pink- and white-flowered *E. plukenetii* samples were collected from individual bushes in a population on Du Toit's Kloof Pass (DP) and red- and white-flowered *E. plukenetii* samples from individual bushes in a population on Franschhoek Pass (FP). White flowered individuals were not *E. plukenetii* ssp. *breviflora*, but clearly white mutants of the *Erica plukenetii* belonging to the same population. Flowers at eight, arbitrary but easily distinguishable points in the floral development cycle as shown and described in Figure 15, were collected, directly frozen in liquid nitrogen and stored at -80°C.

4.2.2 Whole genome NGS

The red flowered *E. plukenetii*, which had been collected from Table Mountain, was sent to Plant Research International at Wageningen University for Illumina NGS runs. DNA was extracted using their in-house CTAB extraction protocol, the Illumina library was prepared and one Illumina HiSeq2500 paired end run with 250bp reads and two Illumina MiSeq paired end runs with 2x300bp reads were performed and the data was assembled using CLC Bio.

4.2.3 Identification and annotation of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes

A local BLAST database was created in Geneious R9[188] from the contiguous sequences. The annotated exon sequences of the anthocyanin pathway and *trans*-acting regulatory genes from the most closely related species that could be found, namely *Vitis vinifera, Camellia sinensis, Camellia nitidissima, Diospyros kaki, Vaccinium corymbosum, Actinidia chinensis* and *Rhododendron simsii*, were used in tBLASTx searches of the local database to find contiguous sequences containing anthocyanin pathway genes and *trans*-acting regulatory genes. *Cis* regulatory sites were found using the Motif Finder in Geneious R9[188] and the methodology described by Zhu *et al* (2015)[154].

4.2.4 Thermal Asymmetric Interlaced PCR

Thermal Asymmetric Interlaced PCR following the protocols of Liu and Whittier (1995)[201] and Liu *et al.* (2007)[202] was used in an attempt to obtain promoter binding sequences upstream of the first exon of a gene if the upstream sequence was not present in the relevant contiguous sequence.

4.2.5 Re-sequencing of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes

Primer pairs were designed to amplify overlapping regions from the upstream regulatory regions of each gene to the end of the last exon based on the aligned exons from the tBLASTx searches. Subsequently a chromosome walking approach was used to complete the sequences where data was missing. DNA was directly amplified from the leaves of red flowered *E. plukenetii* using the protocol developed by Bellstedt *et al.* (2010)[69]. PCR amplifications were carried out using the Kapa Biosystems 3G Plant PCR kit. Each 25µl PCR mix contained 12.5µl Buffer, 2µl 25mM $MgCl_2$, 0.75µl 20mM Forward and 0.75µl 20mM Reverse Primers (see Table 4), 0.2µl 2.5U/µl 3G Plant Taq, 0.2µl 5% DMSO, 1µl extracted DNA and 8.6µl milliQ H_2O in an Applied Biosystems Veriti PCR thermal cycler

The PCR programme was 2 minutes at 95°C; followed by 35 cycles of 95°C for 30 seconds, annealing at the temperature given in Table 3 for 30 seconds, and extension at 72°C for 30 seconds; followed by a final extension step of 72°C for 5 minutes.

Sanger sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fischer Scientific) in an Applied Biosystems PCR thermal cycler using the STeP sequencing protocol[189]. Each 10µl sequencing reaction contained 5µl Sequencing Buffer, 1µl 0.8mM Forward Primer or Reverse Primer (see Table 3), 1µl Big Dye, 1µl of PCR product and 2.5µl milliQ water. Sequencing electrophoresis was performed at the Central Analytical Facility at Stellenbosch University.

Table 3: Primer pairs used to re-sequence the genes of the anthocyanin biosynthesis pathway enzymes, their regulatory regions and their trans-acting transcription factor genes.

Name	Sequence	Annealing temperature (°C)
ANS-352F	GTCATTGACTTCCTCTTGCGC	
ANS-1,126R	TTGGGAATGGACTGGATGCC	60
ANS-1,023F	ATTTCGGGTACTGCCTGCAA	
ANS-2,269R	GAGTTCGTCCTTGCCACCA	66
ANS-2,175F	GAGGCCTAACGACCAAGGTC	
ANS-2,944R	AAGAGACAATCACAAGAGAAGTAGA	60
CHI-25F	AACCGACACAGCATCCAGAG	
CHI-878R	CGTGAAAGAAAGAACAAGAGGGT	60
CHI-632F	ACGGGCAAGCAATACTCAGA	
CHI-1,766R	CTTTGGGGAGGTCTTTGGCT	60
CHI-1,601F	ATTCCCAGTGAGCAACCACC	
CHI-2,522R	CAGGCTTTGAGTCCTCAGGG	60
CHI-2,411F	ACCACGGCAGCTTTATTCCA	
CHI-3,723R	GCCATGCTTCCCAACAATCG	60
CHI-3,646F	CCTGAGAAGGGCAAAGTGGT	
CHI-4,546R	CAAAGCTTCCACAGTATGCCA	60
CHS-9F	ATACGTTCCTGGCTACCCCT	
CHS-1,097R	ACTCCTTGATGGCCTTCACG	60
CHS-992F	AATGTGTGCGCATACATGGC	
CHS-2,020R	ACTGCAAACAACGGGCCTAA	60
DFR-2,142F	GTGGGAGTAGAGTAGCCCCA	
DFR-2,866R	TTGGTTGAGGGCACAAACCT	62
DFR-2,720F	AGGATAACGTGAACGGCTCG	
DFR-3,500R	CCGGACGGAGGGAGTAGTTA	66
DFR-3,351F	AGCTTTGACGAGGCCATTGA	
DFR-4,242R	GGTTCAAGTGCTGCCCTACT	66
DFR-4,722F	ACGATCTCAAACTCAGGGCC	
DFR-5,823R	TGCTCTTGGACATTGACGGT	56
DFR-5,714F	AAGCCGACGATCAATGGTGT	
DFR-6,920R	CCCCATCCCTTGCAACTTCT	60
DFR-6,044F	GAGCATCCTGAAGCAGAGGG	
DFR-6,996R	TTGTTTGAGACTGCTTTATATTTTCG	62
F3'5'H-261F	CGGAGATGCTCACGTACTCC	
F3'5'H-1,158R	AGCTCGTGCAAACTGACCAT	60
F3'5'H-1,077F	GGTTGTGGAAAGTGAACGCA	
F3'5'H-2,274R	ACCACCACATTGACCATCGA	60
F3'H-1,335F	TCGGGCAAATATAACCTCGGA	
F3'H-2,134R	AAGTCGTCCAAGGCCTTAGC	60
F3'H-2,009F	CTCCGGGGCCAAGCATATT	
F3'H-3,130R	ACAATAATGCTAGCCCCGGG	60
F3'H-3,050F	TGGAAAACGAACCATGCGTTC	
F3'H-4,027R	ACCGATGAGTTAAGTTGGCGA	60
F3'H-3,926F	TGGTGAACAGACTTTTCAACTGG	60

Name	Sequence	Annealing temperature (°C)
F3'H-5,083R	GGTGCACAGGTTAAGGAGCT	60
F3'H-5,000F	GAAGTAGCCATTCTGACGCG	
F3'H-6,088R	CCCATTACGGTTGAGCACGA	60
F3'H-6,070F	CGTGCTCAACCGTAATGGGA	
F3'H-6,487R	AGGAGCGCATGAAACCGTAT	60
F3'H-6,287F	CGATCTGAGACCCAACCCAA	
F3'H-7,426R	GGAGAGAGAGTGGGGTGGAT	60
F3'H-7,230F	CTCGTCCAGCACAGTGGAAT	
F3'H-8,391R	AAGTGAACTTTTCATCCCCTTTTT	60
F3H-528F	CGTTTAGACCTTCTCTCGAGCA	
F3H-1,519R	ACAATCTTAAGTTTCCCATATTGACCT	60
F3H-1,060F	CAGTCAAAGTTCGTCCGGGA	
F3H-2,139R	GCGATGCGATCCGGTTTAAC	60
F3H-2,023F	AAGCGGTCCCACAGTTTTGA	
F3H-3,740R	ATGACCATGGTCGCCCAAAT	56
F3H-2,810F	GAGCGAGAGTAAGTTGCCGT	
F3H-3,740R	ATGACCATGGTCGCCCAAAT	50
F3H-3,562F	AATGTCCACAACCCGACCTC	
F3H-4,617R	ACGTACGGGATTGGTGCTTT	60
F3H-4,448F	GCCCAGGACAGCCCAATTAT	
F3H-5,794R	TGGTGGGTAGCAAATCTCGG	60
UDP-GST-38F	TCGGGGTAGGTTTTCGTGTG	
UDP-GST-839R	GTTGAACTTTGCGGCGACTT	60
UDP-GST-720F	TTGAAGGCCACCATGATGCT	
UDP-GST-1,884R	CGTTGTCATCTCACGCCTGA	60
MYB-26F	GCGTCCACTTGTGTGTTTCC	60
MYB-1,395R	AACACGTACGGCTAACACAA	
bHLH-353F	GACCCATTGACGTGTTTGCC	60
bHLH-1,360R	ACAGACAACAGAGCTTCACA	
WDR-707F	CAGGTTGCCTTGTTTGCAGT	60
WDR-2,220R	TGTGATGAACAATGTGGGGC	

4.2.6 The expression of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes

Total RNA was extracted using a Qiagen RNeasy Plant Mini Kit from eight different corolla lengths (Figure 15) of the collected red-, pink- and white-flowered *E. plukenetii* samples from DP and FP. RNA quality was verified using gel electrophoresis. RNA was quantified using a Nanodrop ND-1000 spectrophotometer. The 260/280 absorbance ratios were greater than 1.95.

Primers were designed to amplify short (100-200bp) amplicons of the expressed sequences of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes. Paratubulin 1 (PTB1), a house keeping gene, that has been shown to be stably expressed[203] in

the closely related *C. sinensis*, was also identified in the same manner as the genes in section 4.2.1. DNA probes were designed to bind specifically to the amplicons; each probe was labelled at the 5' end with a fluorescent reporter dye and a quencher on the 3' end. Primer and probe sequences are shown in Table 4. Primer and probe binding sites are shown in Figure 16.

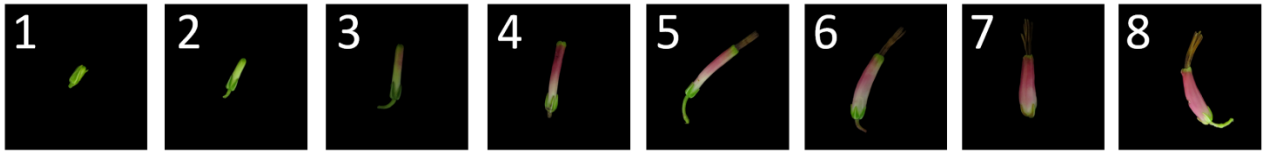


Figure 15: Photographs of eight flowers at different development points that were selected for the expression studies of the genes of the anthocyanin biosynthesis pathway enzymes and their trans-acting transcription factor genes, in red-, white- and pink-flowered (shown) E. plukenetii. 1) No corolla emergence, 2) corolla emerged 1-2mm, 3) first indication of colouration, 4) immediately prior to anther emergence, 5) anthers emerged 1-2mm, 6) basal corolla swelling, 7) flower fully coloured and 8) fully mature flower.

Extensive optimisations were conducted to establish a repeatable expression assay. Initially the suitability of the primers was tested using directly extracted DNA[69] to confirm that amplification was successful and the presence the amplified products was verified using agarose gel electrophoresis. Subsequently the Reverse Transcriptase PCR was modified for Real Time detection. The RT-qPCR assay was optimised to ensure that all the genes had similar amplification efficiency and Cq values.

In the samples collected from the DP population, the expression of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes was determined using conventional Reverse Transcriptase PCR, with primers from Table 4. Each 20µl PCR mix contained 10µl Kapa 3G Plant Taq buffer, 1.2µl 25mM $MgCl_2$, 1.0µl DTT, 0.5µl 20mM Forward primer, 0.5µl 20mM Reverse primer, 0.2µl 2.5U/µl Kapa 3G Plant Taq, 0.1 Roche SuperScript III reverse transcriptase 1.6µl template and 8.6µl ddH_2O. The PCR was run in an Applied Biosystems PCR thermal cycler for 30 minutes at 48°C, followed by 35 cycles of 95°C for 20 seconds, 60°C for 20 seconds and 72°C for 20 seconds, with a final elongation step of 72°C for 5 minutes. The PCR products were visualised using agarose gel electrophoresis and the differential expression was assessed visually. No template and no reverse transcriptase controls were run for each reaction.

In the samples collected from the FP population, the expression of the genes of the anthocyanin biosynthesis pathway enzymes was determined using RT-qPCR in triplicate, using the Kapa Probe Fast One Step RT-qPCR kit. Each multiplex RT-qPCR mix contained 10µl Buffer; 2 or 3 sets of: 0.4µl 10mM Forward and 0.4µl 10mM Reverse Primer (see Table 4), 0.4µl 10mM probe (see Table 4); 0.4µl 10mM PTB1-1,416F, 0.4µl 10mM PTB1-1,564R and 0.4µl 10mM PTB1-1500Probe; 0.2µl 2.5U/µl 3G Plant Taq; 1µl extracted RNA and sufficient milliQ H2O to make up a total of 20µl. RT-qPCR was performed in a Roche LightCycler 96, with a single preincubation step of 42°C

for 5 minutes then 95°C for 3 minutes, followed by 45 cycles of 95°C for 3 seconds, 60°C for 20 seconds and 72°C for 2 seconds. Expression of the genes of the anthocyanin biosynthesis pathway enzymes was normalised to the expression of the reference gene, PTB1. No template and no reverse transcriptase controls were run for each reaction. Expression of the *trans*-acting transcription factor genes was determined using conventional RT-PCR and gel electrophoresis as in section 4.2.5.

Table 4: RT-qPCR primers and probe sequences

Name	Label	Quencher	Sequence	Expected Product size (bp)
ANS-1,281F			AGTCCTCTCCCTAGGCTTGG	
ANS-1,328Probe	6-FAM	3' Iowa Black® FQ	AAGTTGGTGGCAAGGACGAA	167
ANS-1,448R			ATGAAGGTGAGGGCGCTTAC	
CHI-632F			ACGGGCAAGCAATACTCAGA	
CHI-657Probe	Hex	3' Iowa Black® FQ	TGGTGGAAAACTGTGTTGCC	184
CHI-816R			CTAACCGTTAGCGACCCCAG	
CHS-153F			CCGTCATGGCTATCGGGAC	
CHS-188Probe	Cy5	3' Iowa Black® RQ-Sp	TGCGTTGATCAGGCCACTTA	109
CHS-262R			CTCCTTCAACTCGGCCTTGT	
DFR-112F			AGGATAACGTGAACGGCTCG	
DFR-175Probe	Cy5	3' Iowa Black® RQ-Sp	GCTCCTGGCTCATCATGAGG	117
DFR-229R			ACGGTGGCTCGAACAACATA	
F3'5'H-261F			CGGAGATGCTCACGTACTCC	
F3'5'H-360Probe	Hex	3' Iowa Black® FQ	ACATGGTGGTGGAGCTCATG	139
F3'5'H-400R			GTTGAATAAACCGGCCGACG	
F3'H-406F			CTCCGGGGCCAAGCATATT	
F3'H-471Probe	Cy5	3' Iowa Black® RQ-Sp	GGCGGATGCTCAGGAAGATA	125
F3'H-531R			AAGTCGTCCAAGGCCTTAGC	
F3H-259F			GATATCGCTAGCCGGGATCG	
F3H-338Probe	6-FAM	3' Iowa Black® FQ	TGGGGGATATTCCAGGTGGT	127
F3H-386R			TAATCAGACCGGCATCCACG	
UDP-GST-480F			AAGTCGCCGCAAAGTTCAAC	
UDP-GST-512Probe	Cy5	3' Iowa Black® RQ-Sp	GTTTTCCACGGCATCAGCTT	135
UDP-GST-615R			GGCACCAAAAAGGGTTCGTC	
PTB1-1,416F			TTCATCAGAACCGGCTCAGG	
PTB1-1,500Probe	Texas Red	3' Iowa Black® RQ-Sp	CGCTGATGTCGCTGGAAATG	148
PTB1-1,564R			TGCTGACAAGACGTGCATCA	
MYB-998F			ATAACCCAAAGCCCACGAGG	136
MYB-1,134R			CACCCGATCAACCTCAGCTT	
bHLH-645F			AGTTGCGGAGGGATAGGCTA	151
bHLH-796R			GTCTGTTCTGGGAGGCCTTC	
WDR-1,605F			CAGGACCCCAGGTATACGGA	151
WDR-1,756R			CCTCACTCGCACTGTGGAAT	

A χ^2 test was used to test for significant differences in expression between the red- and white-flowered samples. The null hypothesis was that the expression of the enzymes was identical in red-

versus white-flowered samples. The gene expression level data for red flowers was used as the expected value and compared to the gene expression level data for white flowers within that population in the χ^2 test. A regression analysis of anthocyanin enzyme gene expression data versus floral growth point was not performed as the expression data of individual genes is independent.

4.3 Results

4.3.1 Whole genome NGS

DNA was successfully isolated from the leaves of a red flowered *E. plukenetii* and was used for whole genome NGS.

The assembled Illumina NGS data of the red flowered E. plukenetii contained 602 million reads, totalling more than eight billion bases. These reads were assembled using CLC Bio into 400 000 contiguous sequences. With a genome size of approximately 800 megabases (M Pirie, pers. comm.), that equates to 100X theoretical coverage. However, due to the presence of fungal, bacterial, chloroplast and mitochondrial DNA, detected by BLAST searches, the actual coverage of the genome was closer to 20X. Of the assembled 4 443 321 contiguous sequences, only 3.75% were larger than 1000bp, with an N50 of 1 673 (excluding scaffolded regions). Incomplete mitochondrial and chloroplast genomes were obtained. The assembly report is shown in Appendix D.

4.3.2. Identification and annotation of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes

All of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes, were found in the NGS data using the tBLASTx method. tBLASTx searches of the NGS data aligned specific contiguous sequences of the Erica NGS data to exons of closely related species. Once all of the exons of a particular gene had been aligned to the contiguous sequence(s), the annotations were transferred to the contiguous sequence(s). If the aligned exons spanned multiple contiguous sequences, the contiguous sequences were concatenated. Start codons were identified and the stop codon at the end of the final exon was assumed to be the end of the gene. The *cis* BRE and MRE motifs upstream of the transcription start site were found in all the contiguous sequences, except for CHI, where the region upstream of the start of the first exon of the gene was absent.

All of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes are highly similar to their homologs in closely related species, with sequence identity in the aligned exonic regions approaching 90%.

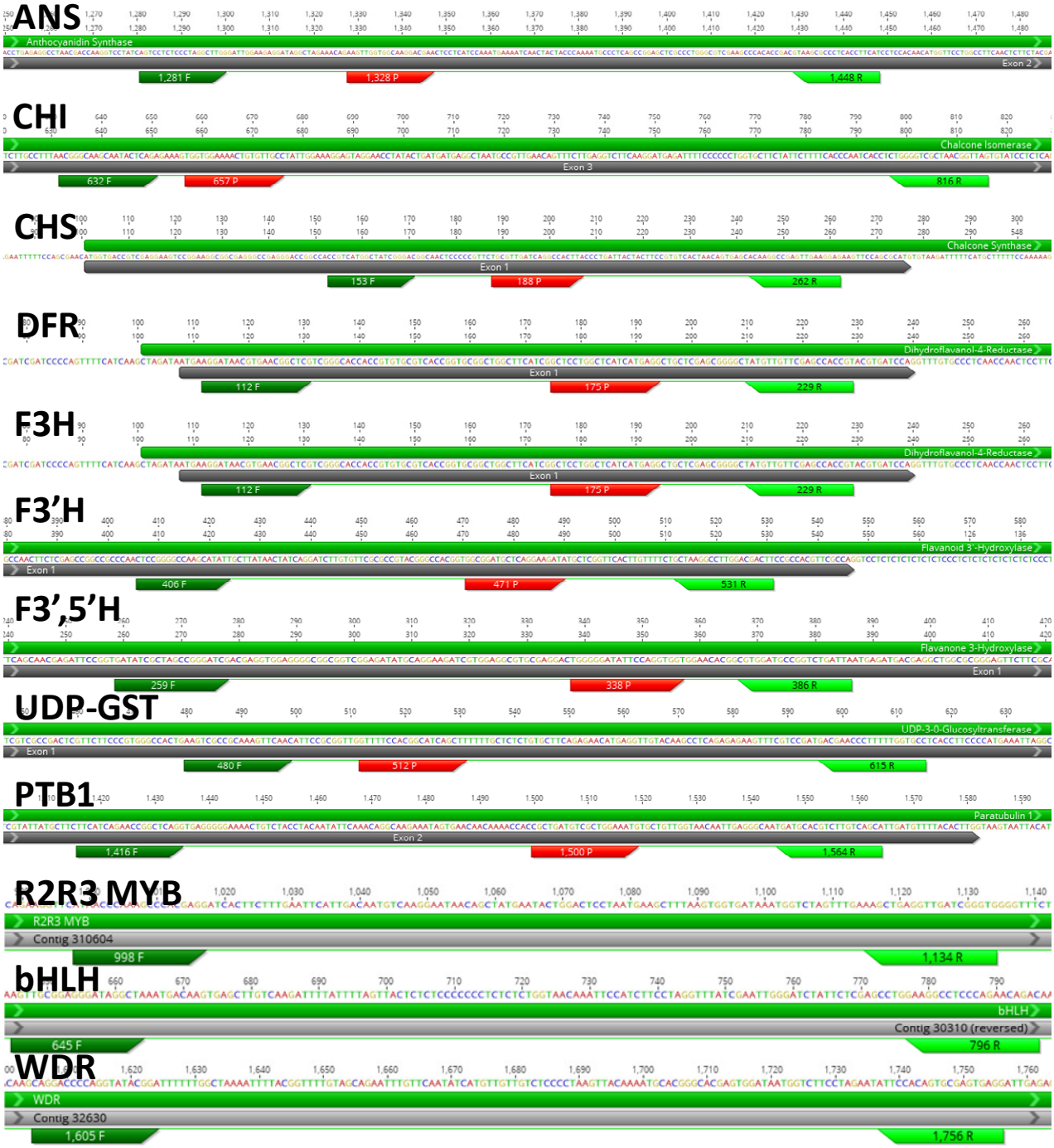


Figure 16: RT-qPCR primers and probe binding sites for all eight genes of the anthocyanin biosynthesis pathway enzymes, the reference gene PTB1 and their three trans-acting transcription factor genes.

4.3.3 Thermal Asymmetric Interlaced PCR

The contiguous sequence containing the CHI gene obtained from the NGS began only a few bases upstream of the start codon. Attempts to sequence the upstream region using Thermal Asymmetric Interlaced PCR[201] were not successful and the sequence of the regulatory region could not be obtained.

4.3.4 Re-sequencing of the anthocyanin pathway genes and trans-acting regulatory genes

Not all of the contiguous sequences containing the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes were complete, some contained stretches of unknown bases and there were gaps between concatenated contiguous sequences. The data from the re-sequencing was successfully used to fill in any gaps, replace any unknown sequences and correct any errors in the Illumina sequencing.

Key characteristics of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes are shown in Table 5. Figures 17-27 show the Sanger sequencing coverage, primer binding sites and *cis* regulatory sites, as well as approximate exon locations and lengths derived from the pairwise alignments of exons from closely related species. Pairwise alignments with the source sequences are in Appendix E.

Table 5: The key properties of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes. "Source" indicates the gene from a related species that was used in tBLASTx searches. Exons numbers are derived from the "Source" species. UDP-GST may have two exons or one longer exon as the exons are contiguous. BRE and MRE elements were found in the upstream region of all genes with the exception of CHI.

Name	Exons	Length	BRE & MRE present	Source
ANS	2	1690	Yes	*V. vinifera* NC_012008
CHI	4	3801	Upstream sequence not present	*A. chinensis* Achn328591
CHS	2	1583	Yes	*R. simsii* AJ413277
DFR	5	4796	Yes	*A. chinensis* Achn135311
F3',5'H	2	1811	Yes	*V. vinifera* AB213606
F3'H	3	7061	Yes	*C. nitidissima* NC_012023
F3H	3	4559	Yes	*V. vinifera* NC_012010
UDP-GST	2 (1)	1425	Yes	*C. sinensis* CsUGTC1
MYB	1	1095	NA	*V. corymbosum* JQ085966
bHLH	1	880	NA	*C. sinensis* HQ660376
WDR	1	1433	NA	*D. kaki* HQ880577

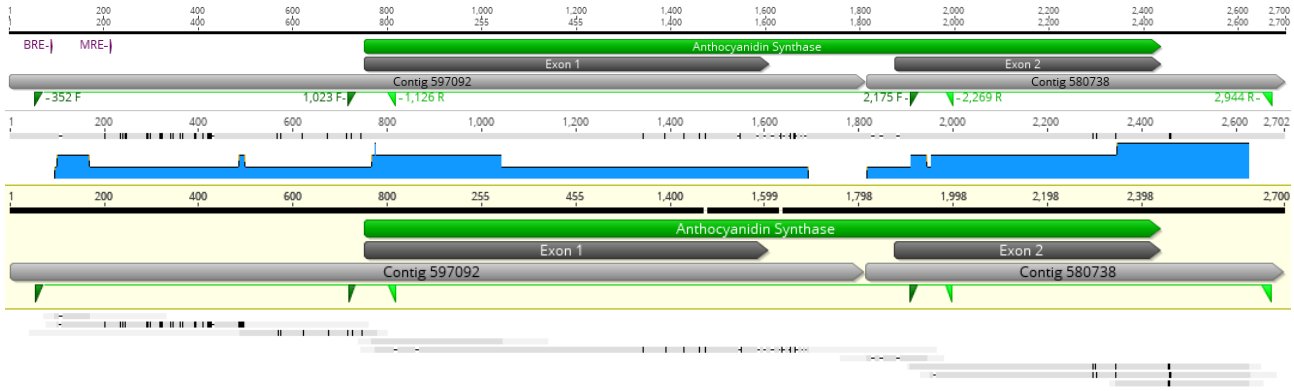


Figure 17: Anthocyanidin synthase showing *cis* promoter binding sites, primer binding sites, NGS contiguous sequences, exon locations, sequencing coverage (blue) and mismatches (black lines) between contigs and Sanger sequencing in the lowest grey bars. Exon locations were derived from *Vitis vinifera*.

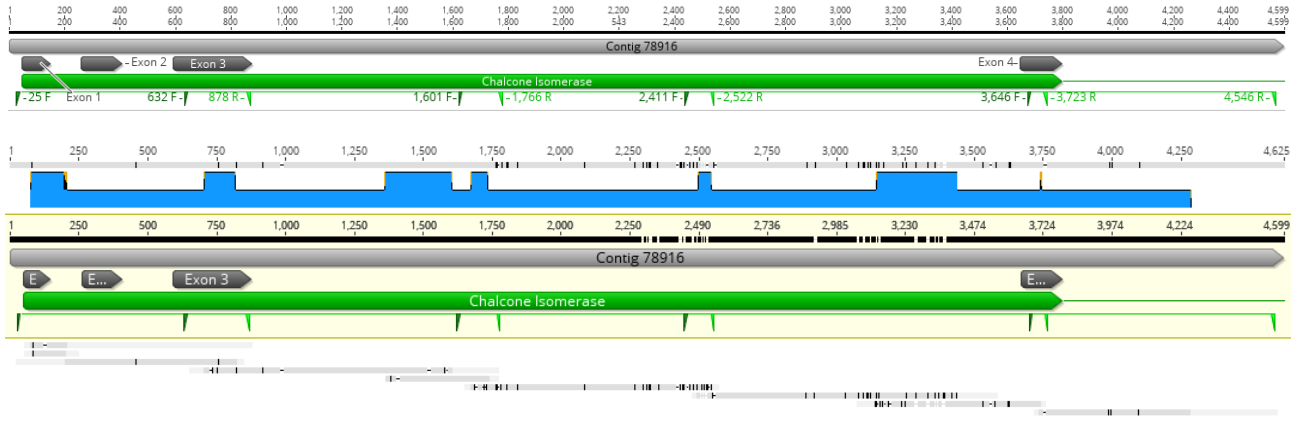


Figure 18: Chalcone isomerase showing *cis* promoter binding sites, primer binding sites, NGS contiguous sequences, exon locations, sequencing coverage (blue) and mismatches (black lines) between contigs and Sanger sequencing in the lowest grey bars. Exon locations were derived from *Actinidia chinensis*.

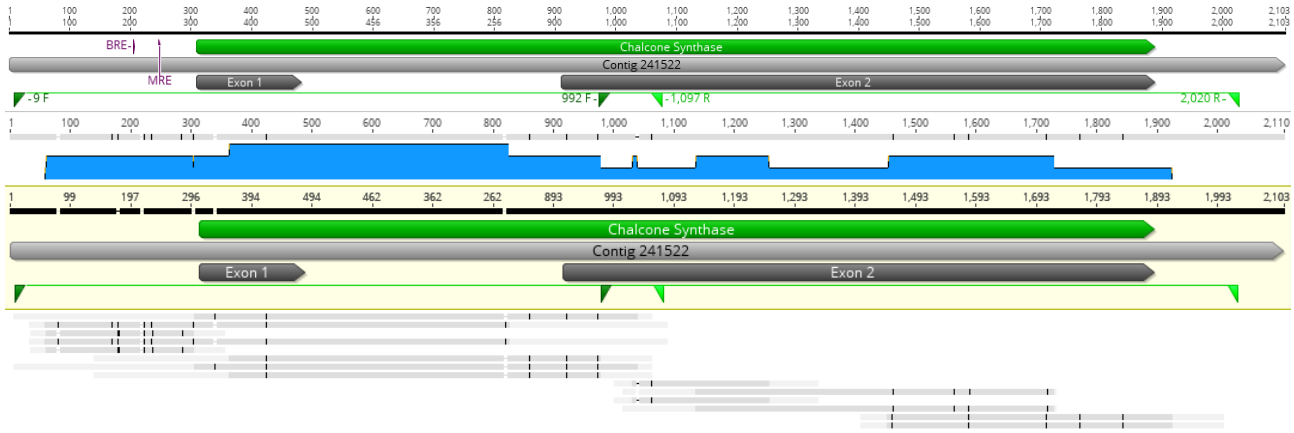


Figure 19: Chalcone synthase showing *cis* promoter binding sites, primer binding sites, NGS contiguous sequences, exon locations, sequencing coverage (blue) and mismatches (black lines) between contigs and Sanger sequencing in the lowest grey bars . Exon locations derived from *Rhododendron simsii.*

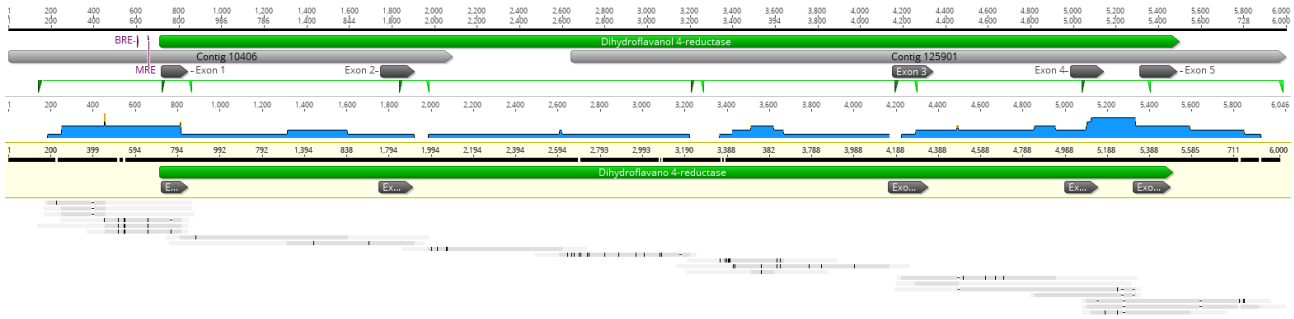


Figure 20: Dihydroflavanol 4-reductase showing *cis* promoter binding sites, primer binding sites, NGS contiguous sequences, exon locations, sequencing coverage (blue) and mismatches (black lines) between contigs and Sanger sequencing in the lowest grey bars . Exon locations were derived from *Actinidia chinensis*.

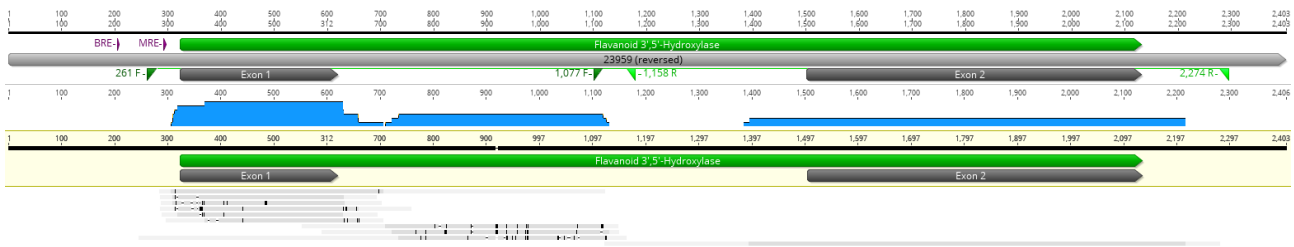


Figure 21: Flavanoid 3’5’-hydroxylase showing *cis* promoter binding sites, primer binding sites, NGS contiguous sequences, exon locations, sequencing coverage (blue) and mismatches (black lines) between contigs and Sanger sequencing in the lowest grey bars. Exon locations were derived from *Vitis vinifera.*

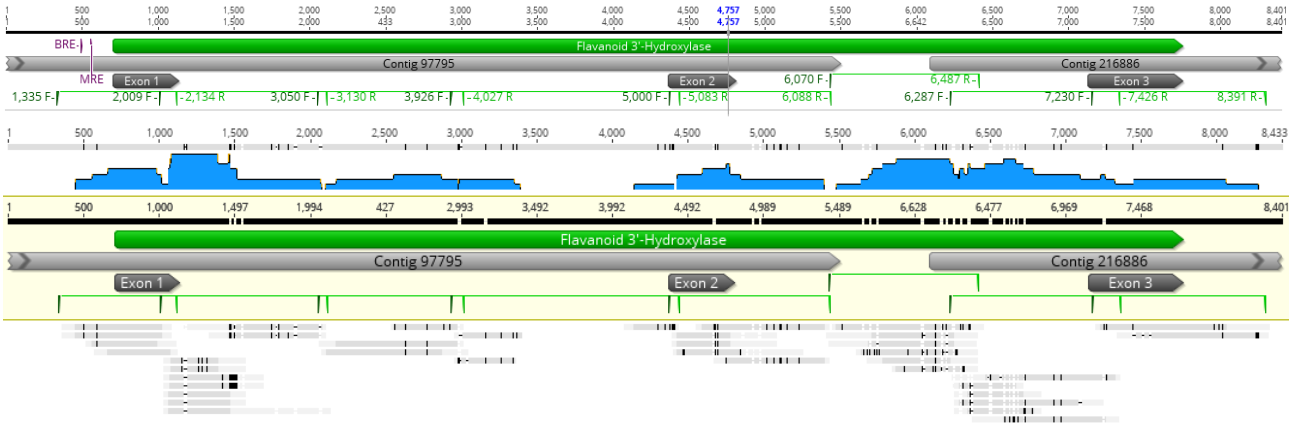


Figure 22: Flavanoid 3'-hydroxylase showing *cis* promoter binding sites, primer binding sites, NGS contiguous sequences, exon locations, sequencing coverage (blue) and mismatches (black lines) between contigs and Sanger sequencing in the lowest grey bars. Exon locations were derived from *Camillia nitidissima*.

Figure 23: Flavanone 3-hydroxylase showing *cis* promoter binding sites, primer binding sites, NGS contiguous sequences, exon locations, sequencing coverage (blue) and mismatches (black lines) between contigs and Sanger sequencing in the lowest grey bars. Exon locations were derived from *Vitis vinifera*.

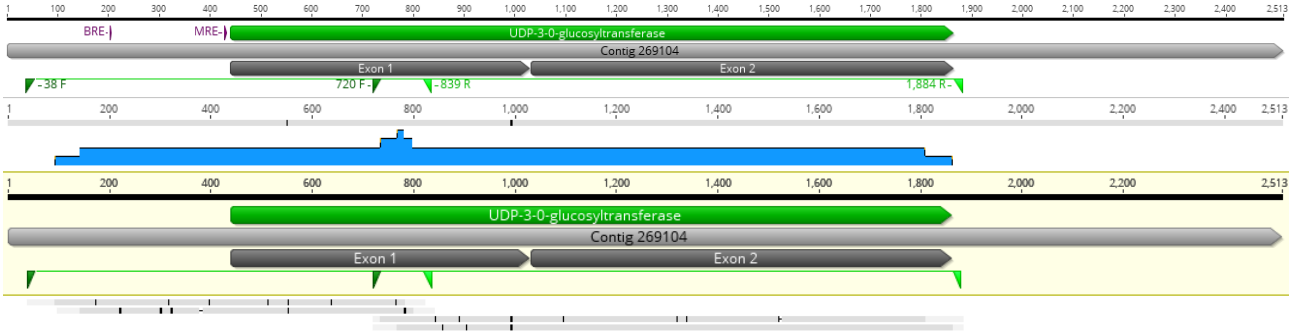


Figure 24: UDP-3-0-glucosyltransferase showing *cis* promoter binding sites, primer binding sites, NGS contiguous sequences, exon locations, sequencing coverage (blue) and mismatches (black lines) between contigs and Sanger sequencing in the lowest grey bars. Exon locations were derived from *Camillia sinensis*.

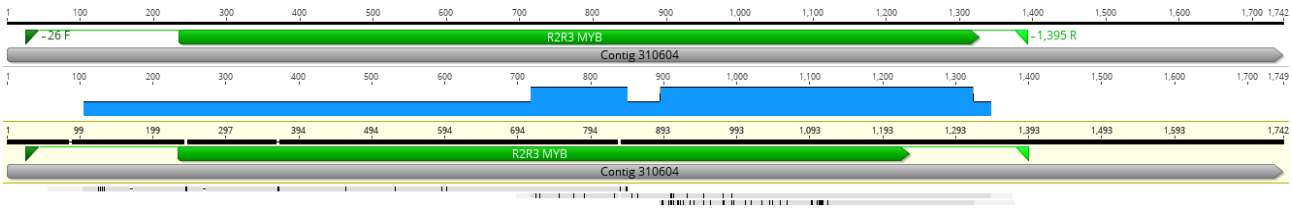


Figure 25: R2R3 MYB primer binding sites and sequencing coverage (blue) and mismatches (black lines) between contigs and Sanger sequencing in the lowest grey bars. Coding sequence derived from *Vaccinium corymbosum*.

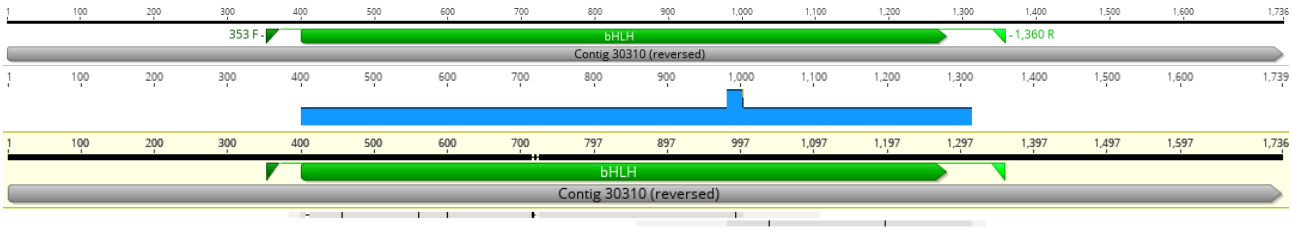


Figure 26: bHLH primer binding sites and sequencing coverage (blue) and mismatches (black lines) between contigs and Sanger sequencing in the lowest grey bars. Coding sequence derived from *Camillia sinensis*.

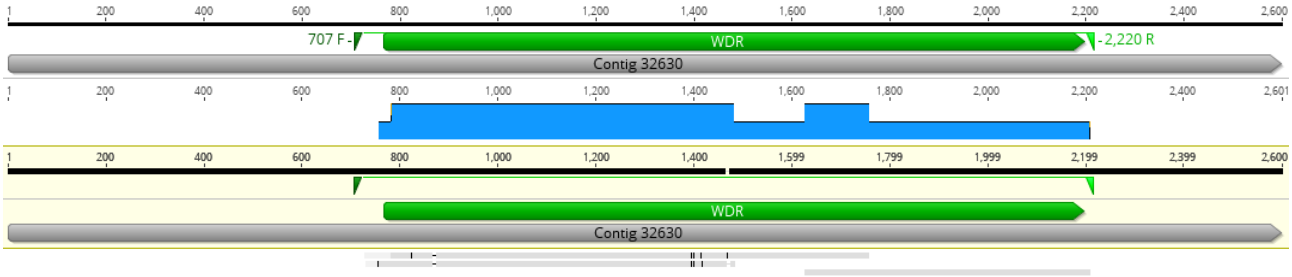


Figure 27: WDR primer binding sites and sequencing coverage (blue) and mismatches (black lines) between contigs and Sanger sequencing in the lowest grey bars. Coding sequence derived from *Diospyros kaki*.

4.3.5 The expression of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes

RNA was successfully isolated from flowers with eight different corolla growth points of red-, pink- and white-flowered *E. plukenetii* (Figure 15) that were collected from the DP population. RNA was successfully isolated from red- and white-flowered *E. plukenetii* from the FP population (no pink morph was present) at the same eight points for RT-qPCR expression studies.

RNA quality was verified using gel electrophoresis. Contamination from gDNA was minimal (Figure 28).

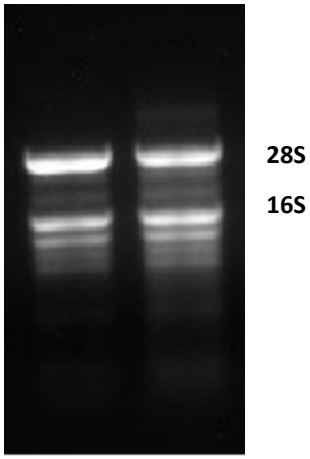


Figure 28: Typical RNA check gel showing clear bands for 28S and 16S rRNA.

In the samples collected from the DP population the expression of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes was confirmed using agarose gel electrophoresis. Bands on the gels were scored for presence or absence. No non

specific amplification was seen and the no template and no reverse transcriptase controls were blank for all samples. All the target bands were the expected size (Table 5).

The expression of PTB1 was found to be stable and consistent across all samples. Expression patterns of CHI, F3H, F3'H, F3',5'H, DFR and ANS (Figure 29) were the same across all samples with expression occurring from growth point 2. UDP-GST was expressed from growth point 3 in all samples (Figure 29). R2R3 MYB, bHLH and WDR were expressed from growth point 2. CHS was not expressed in white *E. plukenetii* (Figure 29).

In the FP population, expression of the genes of the anthocyanin biosynthesis pathway enzymes in red- (Figure 30) and white-flowered (Figure 31) *E. plukenetii* samples, was measured at eight different points (Figure 15) with three biological replicates of each. The no template and no reverse transcriptase controls did not show amplification. The assays were first run with only a single gene and the reference gene in the assay for optimization. Once all of the assays had amplification efficiencies of greater than 95%, Cq values greater than 20 but less than 30 and R^2 vales greater than 0.980, the assays were multiplexed together. If the efficiency and y-intercept values remained similar, then determinations were done. Expression of PTB1 appeared stable in each assay. Due to time constraints only a single reference gene was used. Expression was normalized to that of the reference gene using the ΔΔCq method.

Expression of R2R3 MYB, bHLH and WDR were expressed from growth point 2 in both red- and white-flowered *E. plukenetii*, with expression being noticeable from corolla growth point 2. Expression of the genes of the anthocyanin biosynthesis pathway enzymes above the baseline begins at growth point 2, increasing to a maximum at growth points 3 and 4, when red colour is first observable, and then declining gradually towards growth point 8 when the flower is fully mature.

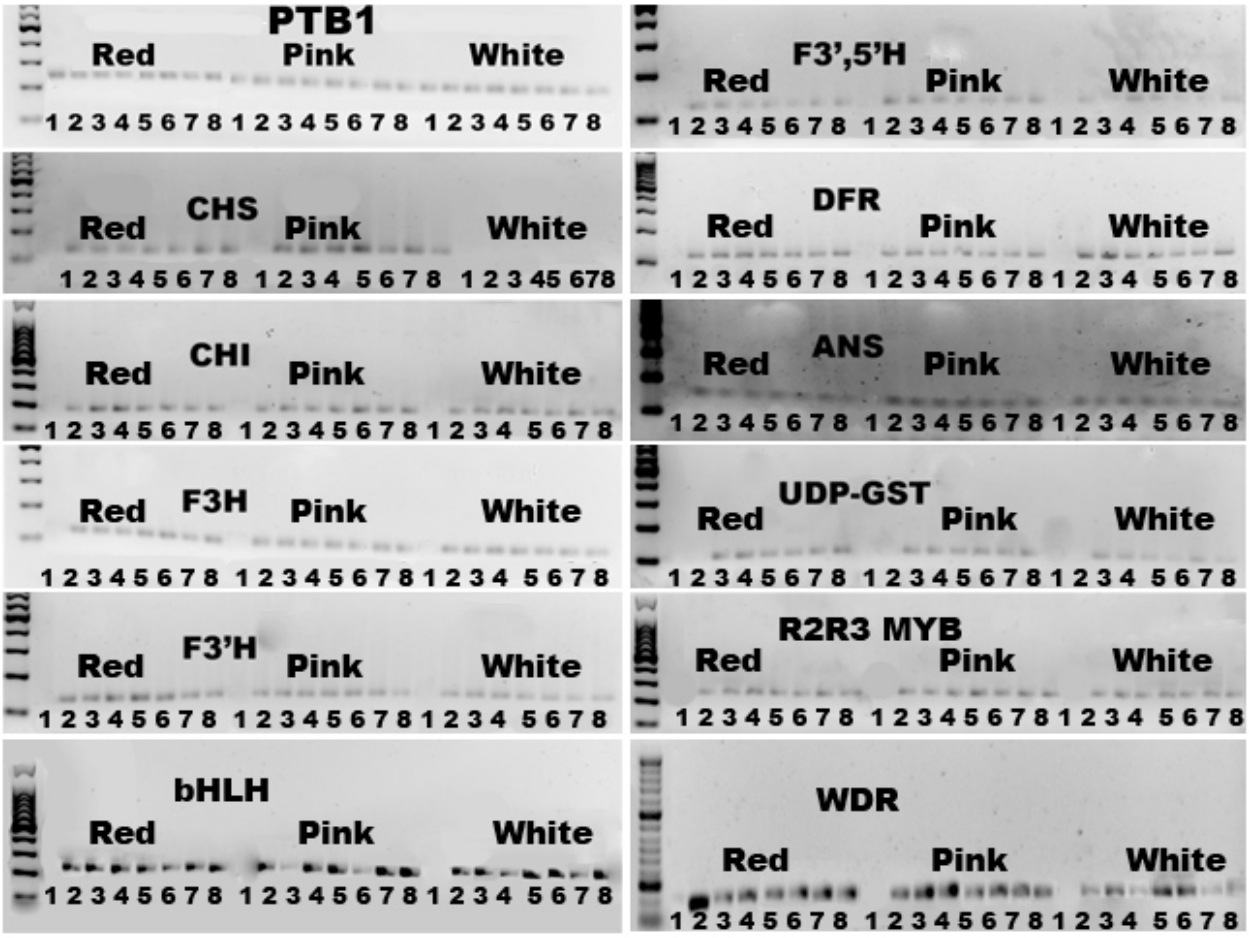


Figure 29: Expression of the genes of the anthocyanin biosynthesis pathway enzymes in eight growth points (1-8) of red-, white- and pink-flowered *E. plukenetii.* Samples were loaded on the gels in the order of corolla length. The eight samples from the red flowered plant were loaded in numerical order first, followed by the eight samples from the pink flowered plant and the eight samples from the white flowered plant last. A 1000bp ladder was used with bands in multiples of 100bp from 100bp to 1500bp.

The expression of the genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes in white flowered samples is identical to the red flowered samples, with the exception of the expression of ANS, which is significantly ($p < 0.05$) lower in white flowered samples when compared to red flowered samples at corolla growth points 3 and 4, when determined using a χ^2 test (Appendix F).

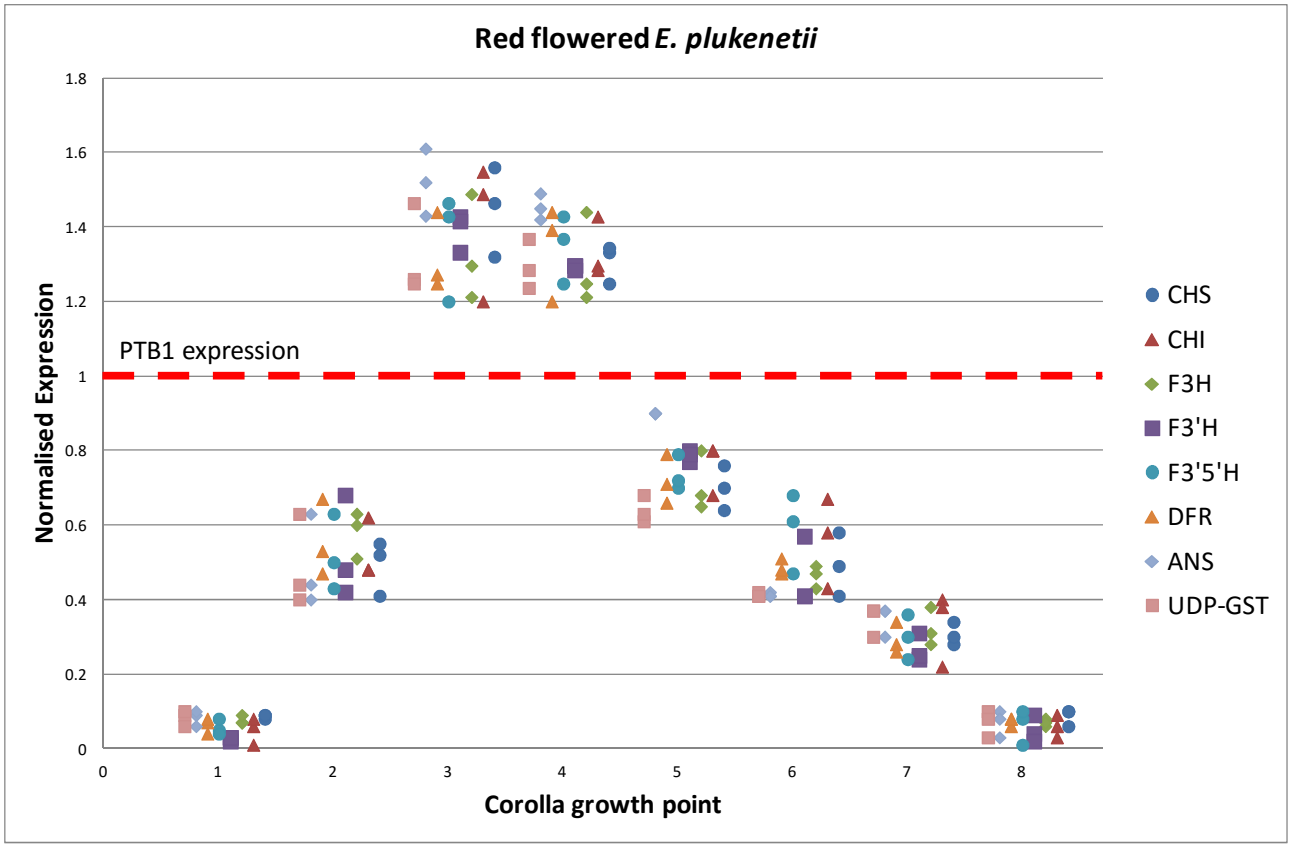


Figure 30: Expression of the genes of the anthocyanin biosynthesis pathway enzymes in red flowered *E. plukenetii.* Expression has been normalised to the expression of PTB1. Three biological replicates of each corolla growth point were tested.

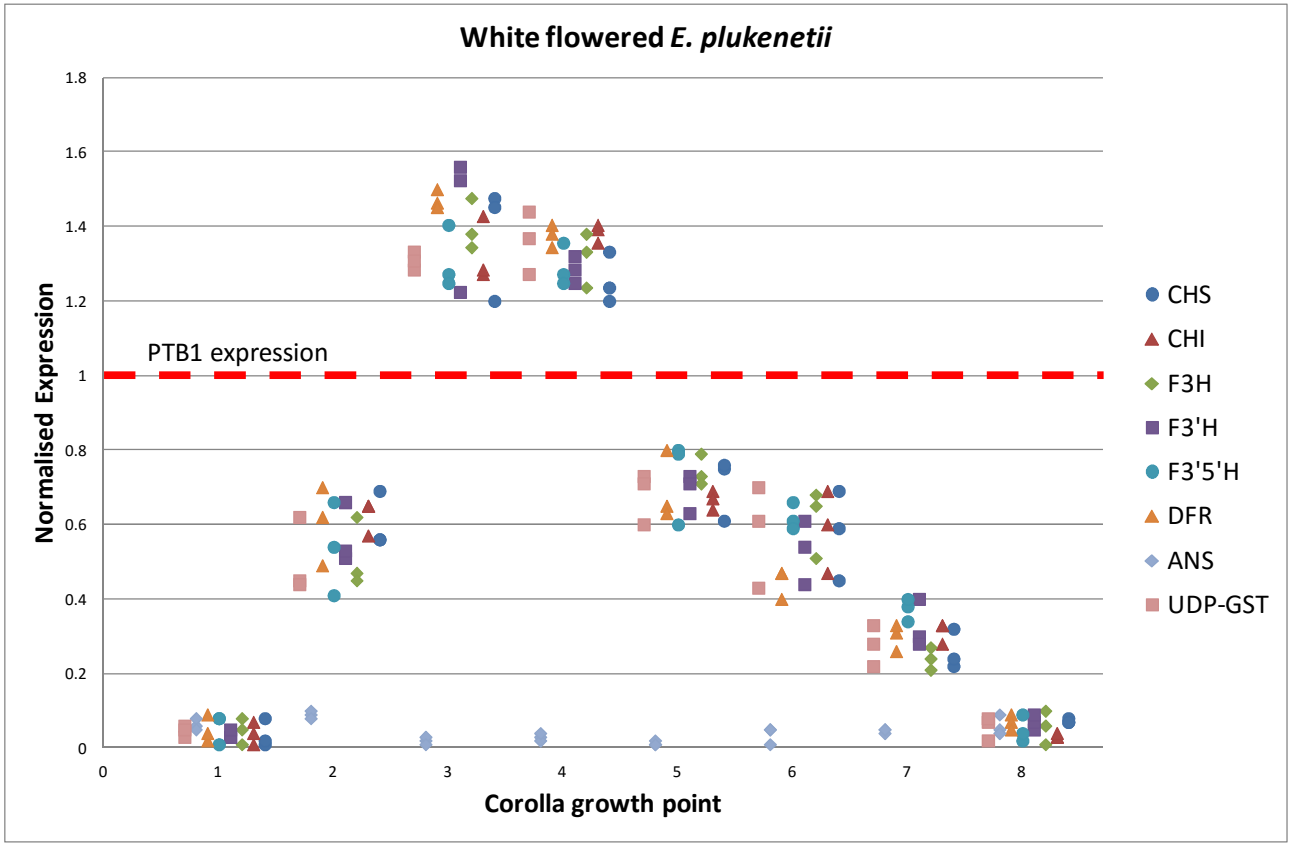


Figure 31: Expression of the genes of the anthocyanin biosynthesis pathway enzymes in white flowered *E. plukenetii.* Expression has been normalised to the expression of PTB1. Expression of ANS is significantly different at growth points 3 and 4. Three biological replicates of each corolla growth point were tested.

4.3.6 Sequencing of promoters in the DP and FP population samples

Due to the observed expression differences in CHS expression in the white flowered sample from the DP population, the promoter region of the CHS gene of the red-, pink- and white-flowered samples used in the expression studies was sequenced using the same methods as described in section 4.2.5. It was found that there was a single base pair deletion in the BRE motif (Figure 32) with CACATT becoming CA-ATT. The deletion was not present in the BRE motif of the red- or pink-flowered samples.

Figure 32: Alignment of promoter region sequences of CHS genes of *E. plukenetii* of the red-, pink- and white-flowered samples from the DP population. The deletion of a cytosine base in the BRE motif of the white flowered sample is indicated by a green square.

Due to the observed expression differences in ANS expression in the white flowered samples from the FP population, the promoter region of the ANS gene of one of the red- and white-flowered samples used in the expression studies was sequenced using the same methods as described in section 4.2.5. It was found that the MRE (Figure 33) had mutated, with the conserved ATCCTCC motif mutating to ATCCTAA in the white flowered sample. The mutation was not present in the red flowered sample.

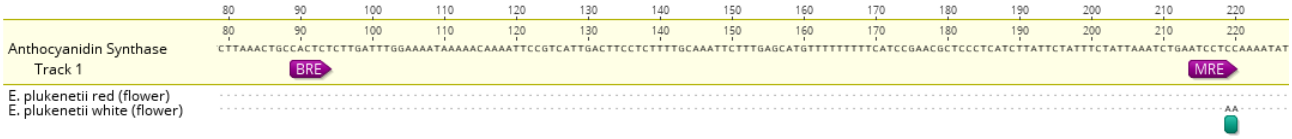


Figure 33: Mutation in the MRE of the ANS gene promoter in white flowered *E. plukenetii* from the FP population, indicated by the blue rectangle.

4.4 Discussion

The NGS sequencing approach was successful but contiguous sequence lengths were typically too short to contain entire genes and exons often had to be located on separate contigs and concatenated. Sanger sequencing was then used to fill the gaps between contigs. Coverage was not sufficient for a complete genome assembly. Further non paired end runs would be required in order to obtain complete chromosomes. Several large contigs were retrieved but they were all bacterial genomes, primarily *Pseudomonas* species.

All of the eight genes of the anthocyanin biosynthesis pathway enzymes and their three *trans*-acting transcription factor genes were found in the NGS data. The exons of the eight genes are highly

similar to the exons of those same genes in the most closely related species that could be found i.e. *Vitis vinifera, Camellia sinensis, Camellia nitidissima, Diospyros kaki, Vaccinium corymbosum, Actinidia chinensis* and *Rhododendron simsii* indicating that their functions are both conserved and important in the plant.

Despite the failure of the attempts to extract total RNA of the quantity and quality required for NGS transcriptomics, sufficient mRNA for Reverse Transcriptase and RT-qPCR was obtained. The presence of expressed mRNA of the eight genes of the anthocyanin biosynthesis pathway enzymes, the reference gene PTB1 and their three *trans*-acting transcription factor genes was confirmed.

A probe based RT-qPCR method for determining the expression of the eight genes of the anthocyanin biosynthesis pathway enzymes and their three *trans*-acting transcription factor genes was developed and utilized to quantify their expression in red- and white-flowered *E. plukenetii.* Probe based RT-qPCR has numerous advantages over PCR and gel based assays as well as over conventional Real Time assays. The use of the sensitive Real Time PCR apparatus allows the determination of the relative expression levels of the genes rather than the binary presence/absence that can be scored from the Reverse Transcriptase assay using gels. Using fluorescently labelled probes rather than a double strand DNA binding dye, such as Sybr Green, increases the accuracy of the assay as specificity is tested twice, once at primer binding prior to amplification and again with the binding of the probe to the amplicons. Furthermore, using differentially labelled probes allows the multiplexing of multiple assays together increasing throughput and efficiency. Multiplexing assays together does however require significant amounts of time and effort spent optimizing the assays to achieve equal amplification efficiencies. The time taken from RNA extraction to having usable quantification results is substantially reduced from at least four hours to only two by the use of the RT-qPCR assay as no gels are needed. The RT-qPCR probes are very expensive when compared to conventional PCR primers, by a factor of 10 to 20 times, but the gains in efficiency and accuracy outweigh this drawback.

Due to the clustering of the expression levels of all the genes at each corolla length it appears that the genes of the anthocyanin biosynthesis pathway enzymes' expression is regulated by a common method of transcriptional regulation, with expression of all the genes at each time point being roughly equal. This agrees strongly with the findings of Zhu *et al.* (2015)[154] who found coordinated expression of the same genes in *Ipomoea purpurea*. Should the loss of pigment formation in these white flowered individuals be the result of a loss of function mutation, i.e. a mutation in the coding region, then the expression of that gene would be expected to be unchanged relative to that of the red flowered individuals. However, the expression of CHS and ANS were found to be reduced in

white flowered samples from the DP and FP populations respectively, therefore excluding the possibility of loss of function mutations of these genes. The reduction appears to be linked to mutations in the *cis* motifs to which the *trans*-acting transcription factors bind.

While mutations in the *cis* motifs to which the *trans*-acting transcription factors bind are not the most common cause of the loss of red pigmentation, they have been observed in 12% of studies documented by Streisfield and Rauscher (2011)[167], and site directed mutagenesis of the BRE and MRE motifs in *I. purpurea* was found to significantly reduce expression of anthocyanin pathway genes[154].

In the FP sample, there is a poly-A region immediately adjacent to the MRE motif. Slippage on the poly-A by DNA polymerase during replication is a likely cause of the mutation to the *cis* motifs to which the *trans*-acting transcription factors bind [193]. In the DP sample a potential cause for the mutation is likely to be due to the flipping of a single base followed by slippage during replication[204–206].

In the tested *E. plukenetii* samples, the loss of the highly conserved *cis* motifs to which the *trans*-acting transcription factors bind, in the BRE of CHS in the DP samples and in the MRE of ANS in the FP samples respectively, will likely reduce the strength of or entirely disrupt the binding of the *trans*-acting MYB-bHLH-WDR complex. The disruption of the binding of the *trans*-acting factors to the motifs would lead to reduction of the expression of the CHS and ANS genes and consequently the CHS and ANS enzymes would not be produced. These mutations are likely the cause of the loss of red colour in the flowers of the plants sampled. The existence of these motifs, the loss of gene expression linked to mutations in the motifs and the coordinated expression of the genes of the anthocyanin biosynthesis pathway indicate that the mechanism of control of anthocyanin biosynthesis in *E. plukenetii* is the same as in *Ipomoea*[154] and *Arabidopsis*[162,166].

If the CHS and ANS enzymes are not produced, chalcone and the cyanidins are not synthesised, respectively[121,154]. The loss of chalcone would be rather problematic as it is not only a precursor for the anthocyanin pathway but also forms part of the plant immune system[207,208] and plays a role in reducing the effects of heat stress[99] in addition to its role as a precursor for the rest of the anthocyanin pathway. The loss of the cyanidins would be relatively less problematic for the plant as at that point in the anthocyanin synthesis pathway there are no further branches[121]. There would however be consequences for the plant beyond the loss of colour as anthocyanins have multiple roles in plant tissues including protection of photosynthetic molecules and chelation of photosynthetic by-products [157–159], and protection of other photolabile[161] compounds in leaves.

The heritability and any fitness effects of these mutations have not been tested, nor has the incidence of these mutations within the sampled populations been quantified. It is reasonable to assume that the same or similar mutations occur in the other white flowered *E. plukenetii* plants in the populations and that these mutations are tolerable across generations as white flowered plants of various ages occur in the populations but this should be empirically verified.

The causal relationships in the switch between ornithophily and entomophily in *Erica plukenetii* are difficult to determine. From the phylogeny, it appears that the plesiomorphic form is red/pink in colour with a medium length corolla and is ornithophilous[70]. It would therefore follow that the loss of red colour may have contributed to the switch to entomophily but is certainly not the sole reason. The changes in habitat from mountains to sandy plains and the associated changes to the growth form making perching more difficult would have certainly contributed[70]. Additionally, the acquisition of scented flowers as a further attractor of moth pollinators must play a role[70]. Currently the major pollinator of *E. plukenetii*, the Orange Breasted Sunbird, is present throughout the range of *E. plukenetii* ssp. *breviflora* but no fossil evidence of their range exists, making it impossible to determine if availability of typical pollinators (or the lack there of) also contributed. Which of these factors initially triggered the destabilisation of the ornithophile "vortex"[73] and which have subsequently reinforced the shift to entomophily is not clear.

Chapter 5: Conclusions and future perspectives

A multiple marker *Erica* species phylogeny was generated based on sequences 60% of the species in the genus. The phylogeny resulting from an automated alignment of the sequence data was compared to that resulting from a manual alignment of the same dataset. The resultant phylogenies were found to not differ significantly. This result demonstrates that concerns of possible bias and errors in the manual dataset were unfounded. The lack of significant differences between the phylogenies resulting from the two methods notwithstanding, the automated pipeline approach has multiple advantages, namely: speed, accuracy, reproducibility, the removal of the possibility for bias and efficiency of the process relative to the manual approach. *Erica* is one of the larger genera of plants worldwide and has offered an excellent opportunity to test the utility of automated versus manual alignments.

The overall congruence between the ITS[7] phylogeny and the phylogeny presented here, in which the same clades were again retrieved but now with greater bootstrap support, would indicate that little change to the overall clade structure of the phylogeny could be expected should more species be added. However, increased taxon sampling would likely increase resolution and bootstrap support for the existing clades and should be considered.

This phylogeny will be an invaluable tool for further elucidating the role that a variety of factors have played in driving the radiation of *Erica* in the CFR. The mapping of a variety of species data, such as fire survival strategy, rainfall requirements or altitude preference amongst others, on the phylogeny should provide insights into the influence these individual factors may have had and even continue to have on further speciation in the genus. Unfortunately the distribution data for *Erica* species in the CFR accumulated in the Precis database of the South African National Biodiversity Institute was not suitable for mapping onto the phylogeny. This data will require some reworking to make it suitable for character mapping before biogeographical analyses with ancestral state reconstructions using tools such as DIVA, Lagrange and others can be performed.

The anomalous positioning of *E. pauciovulata* basal to the Cape clade is more likely to be an artefact of its trnT-L sequence than a true reflection of its position in the phylogeny. Deletion of its central region flanked by poly-A and poly-T repeats could be considered but the missing data in the alignment may bring even further bias into its phylogenetic position. Further markers should therefore rather be sequenced to place it accurately, i.e. within the Cape clade, in which case the Cape clade will have descended from a single common ancestor.

The spectacular radiation of the *Erica* genus in the CFR and by contrast, the lack of diversity elsewhere in the range of the genus, is likely the result of a combination of factors. The transition to winter rainfall and consequent removal of the ancestral tropical vegetation followed by the large scale stability of the climate created the opportunity for the radiations and caused less subsequent extinction of species. Within the Cape clade, the existence of multiple clades with limited biogeographic ranges indicates that multiple *in situ* radiations occurred from founder species in each range. Outside of the CFR, the absence of many of these factors and a possibly higher extinction rate are likely contributors to the observed lack of similar levels of diversification[57].

The multiple, independent evolution of red colour and ornithophily in *Erica* generally and subsequent reversion to entomophily in *Erica plukenetii* ssp. *breviflora* specifically, would suggest that shifts between pollination syndromes potentially alter fitness and have been important in the diversification of *Erica* in the CFR. The multiple, independent evolution of red flower colour make the genus *Erica* an ideal model system to study the gain and loss of red colour within a genus and the role it may have played in the diversification of the *Erica* species in the CFR.

This study has shown that the mechanisms underlying the flower colour polymorphism in *E. plukenetii* can be associated with differential expression of the genes of the anthocyanin biosynthesis pathway enzymes, likely due to the mutations in the promoter binding sites. Whether the floral colour changes preceded the shift between ornithophily and entomophily or merely reinforced the effects of other changes that had already occurred is not clear and a more resolved phylogeny at subspecies level would be required to make a better determination of the temporal relationship of the changes within this species complex.

Development of a rapid, PCR based test for the characterised mutations in *E. plukenetii* would allow population and generation level testing of the prevalence of the mutations to be determined. From this it would be possible to test for Hardy-Weinberg equilibrium[209] and if the mutation is indeed subject to selective pressure.

Expansion of testing of expression of genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes to other clades or species complexes in *Erica* where red colour and ornithophily have evolved and the ancestral states are better characterised is now possible. This may be very valuable in determining the causal relationship between floral colour changes and pollinator shifts. Furthermore, it would be possible to determine if intermediate floral colours, such as pink, are the result of differential gene expression of enzymes in the pathway or dosage effects.

Testing would have to take a multifaceted approach, initially using expression studies to determine if the colour changes are the result of reduced expression of all the anthocyanin pathway genes, indicating mutations in the *trans*-acting genes (or reductions in their expression) or reduced expression of a single gene indicating mutation(s) in the *cis* promoter binding sites of that particular gene.

As with the CHI gene, where the upstream region was not present in the contiguous sequence obtained from the Illumina sequencing, sequencing of the promoter regions of the genes of the anthocyanin biosynthesis pathway enzymes in other *Erica* species may well prove problematic as the upstream primer binding sites are not conserved and techniques such as Thermal Asymmetric Interlaced PCR[201,202] or Inverse PCR[210] could be used.

As gene expression is only a proxy for the presence of an enzyme and provides no information on enzyme function, if no expression differences are seen, HPLC determinations of which of the intermediates in the anthocyanin pathway are present would indicate any possible loss of function mutations in the genes of the anthocyanin pathway enzymes themselves. The genes of the anthocyanin pathway enzymes could be sequenced to find these mutation(s).

Anthocyanins play a role in protecting the plant from environmental stresses[155], pathogens[160] and protect the photosynthetic apparatus[156] and their role in floral colour is likely of secondary importance. The plesiomorphic form of *Erica* is pink flowered as a result of anthocyanin production[163]. The shift from pink- to red-flowered and the consequent shift in pollinators would allow a new niche to be occupied, without negative consequences, as the anthocyanins are still produced. However, the loss of anthocyanin production, as seen in white flowered *Erica plukenetii,* would likely result in measureable fitness effects with the plant being more susceptible to environmental stresses and pathogens.

Fitness, pollinator visitation and pollination efficiency should be tested in red- versus white-flowered plants in order to determine if the loss of floral colour influences reproductive success and if pollinators display a preference for a particular floral colour in *Erica*. With the causal relationships better understood it will be possible to better determine the influence of pollinator shifts on diversification in *Erica* and in the CFR as a whole.

Bibliography

1. Oliver, E. G. H. The Ericoideae and the southern African heathers. *Bot. J. Linn. Soc.* **101,** 319–327 (1989).
2. Oliver, E. G. H. Systematic studies in the tribe Ericeae (Ericaceae–Ericoideae). (1999).
3. Gillespie, E. & Kron, K. Molecular phylogenetic relationships and a revised classification of the subfamily Ericoideae (Ericaceae). *Mol. Phylogenet. Evol.* **56,** 343–54 (2010).
4. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181,** 1–20 (2016).
5. Bremer, B., Bremer, K., Heidari, N., Erixon, P., Olmstead, R. G., Anderberg, A. A., Källersjö, M. & Barkhordarian, E. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. *Mol. Phylogenet. Evol.* **24,** 274–301 (2002).
6. Mcguire, A. F. & Kron, K. A. Phylogenetic relationships of European and African Ericas. **166,** 311–318 (2005).
7. Pirie, M. D., Oliver, E. G. H. & Bellstedt, D. U. A densely sampled ITS phylogeny of the Cape flagship genus *Erica* L. suggests numerous shifts in floral macro-morphology. *Mol. Phylogenet. Evol.* **61,** 593–601 (2011).
8. Quezel, P. Analysis of the Flora of Mediterranean and Saharan Africa. *Ann. Missouri Bot. Gard.* **65,** 479–534 (1978).
9. Oliver, E. G. H. *Systematics of Ericeae (Ericaceae-Ericoideae): species with indehiscent and partially dehiscent fruits*. (Bolus Herbarium, University of Cape Town, 2000).
10. Oliver, E. G. H. The Eriocoideae (Ericaceae) - A review. *Contrib. from Bolus Herb.* **13,** 158–208 (1991).
11. Linder, H. P. & Verboom, G. A. The Evolution of Regional Species Richness: The History of the Southern African Flora. *Annu. Rev. Ecol. Evol. Syst.* **46,** 393–412 (2015).
12. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403,** 853–858 (2000).
13. *Plants of Southern Africa. Names and distribution. Memoirs of the Botanical Survey South Africa 62.* (National Botanical Institute, 1993).
14. Goldblatt, P. Floristic Diversity in the Cape Flora of South Africa. *Biodivers. Conserv.* **6,** 359–377 (1997).
15. Oliver, E. G. H., Linder, H. P. & Rourke, J. P. Geographical distribution of present-day Cape taxa and their phytogeographical significance. *Bothalia* **14,** 427–440 (1983).
16. Linder, H. P. The radiation of the Cape flora, southern Africa. *Biol. Rev.* **78,** 597–638 (2003).
17. Goldblatt, P. & Manning, J. C. Plant diversity of the Cape region of Southern Africa. *Ann. Missouri Bot. Gard.* **89,** 281–302 (2002).
18. Sauquet, H., Weston, P. H., Anderson, C. L., Barker, N. P., Cantrill, D. J., Mast, A. R. & Savolainen, V. Contrasted patterns of hyperdiversification in Mediterranean hotspots. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 221–225 (2009).
19. Linder, H. P. Evolution of diversity: the Cape flora. *Trends Plant Sci.* **10,** 536–41 (2005).
20. Verboom, G. A., Archibald, J. K., Bakker, F. T., Bellstedt, D. U., Conrad, F., Dreyer, L. L., Forest, F., Galley, C., Goldblatt, P., Henning, J. F., Mummenhoff, K., Linder, H. P., Muasya, A. M., Oberlander, K. C., Savolainen, V., Snijman, D. A., Niet, T. van der & Nowell, T. L. Origin and diversification of the Greater Cape flora: Ancient species repository, hot-bed of recent radiation, or both? *Mol. Phylogenet. Evol.* **51,** 44–53 (2009).
21. Slingsby, J. A., Britton, M. N. & Verboom, G. A. Ecology limits the diversity of the Cape flora: Phylogenetics and diversification of the genus *Tetraria*. *Mol. Phylogenet. Evol.* **72,** 61–70 (2014).
22. Goldblatt, P., Savolainen, V., Porteous, O., Sostaric, I., Powell, M., Reeves, G., Manning, J. C., Barraclough, T. G. & Chase, M. W. Radiation in the Cape flora and the phylogeny of peacock irises*Moraea* (Iridaceae) based on four plastid DNA regions. *Mol. Phylogenet. Evol.*

25, 341–360 (2002).
23. Galley, C. & Linder, H. P. The phylogeny of the *Pentaschistis* clade (Danthonioideae, Poaceae) based on chloroplast DNA, and the evolution and loss of complex characters. *Evolution* **61,** 864–84 (2007).
24. Bergh, N. G. & Linder, H. P. Cape diversification and repeated out-of-southern-Africa dispersal in paper daisies (Asteraceae-Gnaphalieae). *Mol. Phylogenet. Evol.* **51,** 5–18 (2009).
25. Schwery, O., Onstein, R. E., Bouchenak-Khelladi, Y., Xing, Y., Carter, R. J. & Linder, H. P. As old as the mountains: the radiations of the Ericaceae. *New Phytol.* (2014). doi:10.1111/nph.13234
26. Galley, C. & Linder, H. P. Geographical affinities of the Cape flora, South Africa. *J. Biogeogr.* **33,** 236–250 (2006).
27. Bakker, F. T., Culham, A., Marais, E. M. & Gibby, M. in *Plant Species-level Syst.* (eds. Bakker, F. T., Chatrou, L. W., Gravendeel, B. & Pelser, P. B.) 75–100 (A.R.G. Gantner, 2005).
28. Forest, F., Nänni, I., Chase, M. W., Crane, P. R. & Hawkins, J. A. Diversification of a large genus in a continental biodiversity hotspot: Temporal and spatial origin of *Muraltia* (Polygalaceae) in the Cape of South Africa. *Mol. Phylogenet. Evol.* **43,** 60–74 (2007).
29. Valente, L. M., Reeves, G., Schnitzler, J., Mason, I. P., Fay, M. F., Rebelo, T. G., Chase, M. W. & Barraclough, T. G. Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: Equal rates in different biomes. *Evolution (N. Y).* **64,** 745–760 (2010).
30. Richardson, J. E., Weitz, F. M., Fay, M. F., Cronk, Q. C., Linder, H. P., Reeves, G. & Chase, M. W. Rapid and recent origin of species richness in the Cape flora of South Africa. *Nature* **412,** 181–3 (2001).
31. Siesser, W. G. in *Antarct. glacial Hist. world Paleoenviron.* (ed. Van Zinderen Bakker, E. M.) 105–133 (1978).
32. Siesser, W. G. Late Miocene Origin of the Benguela Upswelling System off Northern Namibia. *Science (80-.).* **208,** 283–285 (1980).
33. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science (80-.).* **292,** 686–693 (2001).
34. Linder, H. P. & Hardy, C. R. Evolution of the species-rich Cape flora. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **359,** 1623–32 (2004).
35. Dupont, L. M., Linder, H. P., Rommerskirchen, F. & Schefuß, E. Climate-driven rampant speciation of the Cape flora. *J. Biogeogr.* **38,** 1059–1068 (2011).
36. Krammer, R., Baumann, K.-H., Henrich, R. & Heinrich, R. Middle to late Miocene fluctuations in the incipient Benguela Upwelling System revealed by calcareous nannofossil assemblages (ODP Site 1085A). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **230,** 319–334 (2006).
37. Matthews, T., Measey, G. J. & Roberts, D. L. Implications of summer breeding frogs from Langebaanweg, South Africa: Regional climate evolution at 5.1 Mya. *S. Afr. J. Sci.* **112,** 1–7 (2016).
38. Marlow, J. R., Lange, C. B., Wefer, G. & Rosell-Mele, A. Upwelling intensification as part of the Pliocene-Pleistocene climate transition. *Science (80-.).* **290,** 2288–2291 (2000).
39. Partridge, T. C. & Maud, R. R. in *Cenozoic South. Africa* (eds. Partridge, T. C. & Maud, R. R.) 3–18 (Oxford University Press, 2000).
40. Bytebier, B., Antonelli, A., Bellstedt, D. U. & Linder, H. P. Estimating the age of fire in the Cape flora of South Africa from an orchid phylogeny. *Proc. R. Soc. B Biol. Sci.* **278,** 188–195 (2011).
41. He, T., Lamont, B. B. & Manning, J. A Cretaceous origin for fire adaptations in the Cape flora. *Sci. Rep.* **6,** 34880 (2016).
42. Linder, H. P., Meadows, M. E. & Cowling, R. M. in *Ecol. fynbos Nutr. fire Divers.* (ed. Cowling, R. M.) 113–134 (Oxford University Press, 1992).

43. Litsios, G., Wüest, R. O., Kostikova, A., Forest, F., Lexer, C., Linder, H. P., Pearman, P. B., Zimmermann, N. E. & Salamin, N. Effects of a fire response trait on diversification in replicated radiations. *Evolution (N. Y).* **68,** 453–465 (2014).
44. Ojeda, F. Biogeography of seeder and resprouter *Erica* species in the Cape Floristic Region-Where are the resprouters? *Biol. J. Linn. Soc.* **63,** 331–347 (1998).
45. van der Niet, T. & Johnson, S. D. Patterns of plant speciation in the Cape Floristic Region. *Mol. Phylogenet. Evol.* **51,** 85–93 (2009).
46. Segarra-Moragues, J. G. & Ojeda, F. Postfire response and genetic diversity in *Erica coccinea*: Connecting population dynamics and diversification in a biodiversity hotspot. *Evolution (N. Y).* **64,** 3511–3524 (2010).
47. Ojeda, F., van der Niet, T., Malan, M. C., Midgley, J. J. & Segarra-Moragues, J. G. Strong signature of selection in seeder populations but not in resprouters of the fynbos heath *Erica coccinea* (Ericaceae). *Bot. J. Linn. Soc.* **181,** 115–126 (2016).
48. Zhao, X., Dupont, L., Meadows, M. E. & Wefer, G. Pollen distribution in the marine surface sediments of the mudbelt along the west coast of South Africa. *Quat. Int.* 1–13 (2015). doi:10.1016/j.quaint.2015.09.032
49. Shi, N., Dupont, L. M., Beug, H.-J. & Schneider, R. Correlation between Vegetation in Southwestern Africa and Oceanic Upwelling in the Past 21,000 Years. *Quat. Res.* **54,** 72–80 (2000).
50. Dupont, L. & Behling, H. Land-sea linkages during deglaciation: High-resolution records from the eastern Atlantic off the coast of Namibia and Angola (ODP site 1078). *Quat. Int.* **148,** 19–28 (2006).
51. Dupont, L. M., Behling, H., Jahns, S., Marret, F. & Kim, J. H. Variability in glacial and Holocene marine pollen records offshore from west southern Africa. *Veg. Hist. Archaeobot.* **16,** 87–100 (2007).
52. Hoetzel, S., Dupont, L. M. & Wefer, G. Miocene-Pliocene vegetation change in south-western Africa (ODP Site 1081, offshore Namibia). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* (2015). doi:10.1016/j.palaeo.2015.02.002
53. Coetzee, J. A. Tertiary environmental changes along the south-western African coast. *Palaeontol. africana* **23,** 197–203 (1980).
54. Cowling, R. M. & Pressey, R. L. Rapid plant diversification: planning for an evolutionary future. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 5452–7 (2001).
55. Partridge, T. C. Cainozoic environmental change in southern Africa, with special emphasis on the last 200 000 years. *Prog. Phys. Geogr.* **21,** 3–22 (1997).
56. Partridge, T. C. Of diamonds, dinosaura and diastrophism: 150 Million years of landscape evolution in southern Africa. *South African J. Geol.* **101,** 167–184 (1998).
57. Partridge, T. C. & Maud, R. R. *The Cenozoic of southern Africa*. (Oxfrod University Press, 2000).
58. Cowling, R. M., Procheş, Ş. & Partridge, T. C. Explaining the uniqueness of the Cape flora: Incorporating geomorphic evolution as a factor for explaining its diversification. *Mol. Phylogenet. Evol.* **51,** 64–74 (2009).
59. Dupont, L. M. Late Pliocene vegetation and climate in Namibia (southern Africa) derived from palynology of ODP Site 1082. *Geochemistry, Geophys. Geosystems* **7,** (2006).
60. Meadows, M. E. & Sugden, J. M. A vegetation history of the last 14000 years on the Cedarberg, south-western Cape Province. *S. Afr. J. Sci.* **84,** 34–43 (1991).
61. Midgley, G. F., Hannah, L., Roberts, R., MacDonald, D. J. & Allsopp, J. Have Pleistocene climatic cycles influenced species richness patterns in the Greater Cape Mediterranean Region? *J. Biogeogr.* **2,** 137–144 (2001).
62. Midgley, G. F., Reeves, G. & Klak, C. Late Tertiary and Quaternary climate change and centres of endemism in the southern African flora. *Phylogeny Conserv.* 230–242 (2005).
63. Verboom, G. A., Bergh, N. G., Haiden, S. A., Hoffmann, V. & Britton, M. N. Topography as a driver of diversification in the Cape Floristic Region of South Africa. *New Phytol.* (2015).

doi:10.1111/nph.13342
64. Baldwin, B. G. Origins of plant diversity in the California Floristic Province. *Annu. Rev. Ecol. Evol. Syst.* **45,** 347–369 (2014).
65. Tilman, D. *Resource competition and community structure*. *Monogr. Popul. Biol.* **17,** (1982).
66. Tilman, D. in *Mediterr. Ecosyst. role Nutr.* (eds. Kruger, F. J., Mitchell, D. T. & Jarvis, J. U. M.) 322–336 (Springer Berlin Heidelberg, 1983). doi:10.1007/978-3-642-68935-2
67. Linder, H. P. in *Species Speciat.* (ed. Vrba, E. S.) 53–57 (Transvaal Museum Monographs, 1985).
68. Oberlander, K. C., Dreyer, L. L., Goldblatt, P., Suda, J. & Linder, H. P. Species-rich and polyploid-poor : Insights into the evolutionary role of whole-genome duplication from the Cape flora biodiversity hotspot. *Am. J. Bot.* **103,** 1–12 (2016).
69. Bellstedt, D. U., Pirie, M. D., Visser, J. C., de Villiers, M. J. & Gehrke, B. A rapid and inexpensive method for the direct PCR amplification of DNA from plants. *Am. J. Bot.* **97,** (2010).
70. Van Der Niet, T., Pirie, M. D., Shuttleworth, A., Johnson, S. D. & Midgley, J. J. Do pollinator distributions underlie the evolution of pollination ecotypes in the Cape shrub *Erica plukenetii*? *Ann. Bot.* **113,** 301–315 (2014).
71. Johnson, S. D. & Steiner, K. E. Generalization versus specialization in plant pollination systems. *Trends Ecol. Evol.* **15,** 140–143 (2000).
72. Fenster, C. B., Armbruster, W. S., Wilson, P., Dudash, M. R. & Thomson, J. D. Pollination syndromes and floral specialization. *Annu. Rev. Ecol. Evol. Syst* **35,** 375–403 (2004).
73. Thomson, J. D. & Wilson, P. Explaining evolutionary shifts between bee and hummingbird pollination: convergence, divergence, and directionality. *Int. J. Plant Sci.* **169,** 23–38 (2008).
74. Wilson, P. & Thomson, J. D. in *Flor. Biol. Stud. Flor. Evol. Anim. plants* 88–111 (1996). doi:10.1007/978-1-4613-1165-2
75. Cronk, Q. & Ojeda, I. Bird-pollinated flowers in an evolutionary and molecular context. *J. Exp. Bot.* **59,** 715–727 (2008).
76. Faegri, K. & Van Der Pijl, L. *The principles of pollination ecology*. **3rd,** (1979).
77. Castellanos, M. C., Wilson, P., Keller, S. J., Wolfe, A. D. & Thomson, J. D. Anther evolution: pollen presentation strategies when pollinators differ. *Am. Nat.* **167,** 288–96 (2006).
78. Stebbins, G. L. Adaptive radiation of reproductive characteristics in angiosperms, I: pollination mechanisms. *Annu. Rev. Ecol. Syst.* **1,** 307–326 (1970).
79. Grant, V. Historical development of ornithophily in the western North American flora. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10407–10411 (1994).
80. Wilson, P., Wolfe, A. D., Armbruster, W. S. & Thomson, J. D. Constrained lability in floral evolution: counting convergent origins of hummingbird pollination in Penstemon and Keckiella. *New Phytol.* **176,** 883–890 (2007).
81. Whibley, A. C., Langlade, N. B., Andalo, C., Hanna, A. I., Bangham, A., Thébaud, C. & Coen, E. Evolutionary paths underlying flower color variation in*Antirrhinum*. *Science (80-.).* **313,** 963–966 (2006).
82. Jones, K. N. & Reithel, J. S. Pollinator-mediated selection on a flower color polymorphism in experimental populations of *Antirrhinum* (scrophulariaceae). *Am. J. Bot.* **88,** 447–454 (2001).
83. Miller, R. Hawkmoths and the geographic patterns of floral variation in *Aquilegia caerulea*. *Evolution (N. Y).* **35,** 763–774 (1981).
84. Hodges, S. A. & Arnold, M. L. Floral and ecological isolation between *Aquilegia formosa* and *Aquilegia pubescens*. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 2493–2496 (1994).
85. Kramer, E. M. & Hodges, S. A. *Aquilegia* as a model system for the evolution and ecology of petals. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **365,** 477–490 (2010).
86. Joseph, N. & Siril, E. A. Floral color polymorphism and reproductive success in Annatto (*Bixa orellana* L.). *Trop. Plant Biol.* **6,** 217–227 (2013).
87. Jones, K. N. Fertility selection on a discrete floral polymorphism in *Clarkia* (Onagraceae).

Evolution (N. Y). **50,** 71–79 (1996).
88. Frey, F. M. Opposing natural selection from herbivores and pathogens may maintain floral-color variation in *Claytonia virginica* (Portulacaceae). *Evolution* **58,** 2426–2437 (2004).
89. Gigord, L. D., Macnair, M. R. & Smithson, A. Negative frequency-dependent selection maintains a dramatic flower color polymorphism in the rewardless orchid Dactylorhiza sambucina (L.) Soo. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 6253–6255 (2001).
90. Pellegrino, G., Bellusci, F. & Musacchio, A. Evidence of post-pollination barriers among three colour morphs of the deceptive orchid Dactylorhiza sambucina (L.) Soo. *Sex. Plant Reprod.* **18,** 179–185 (2005).
91. Waser, N. M. & Price, M. V. Pollinator choice and stabilizing selection for flower color in *Delphinium nelsonii*. *Evolution (N. Y).* **35,** 376–390 (1981).
92. Waser, N. M. & Price, M. V. Pollinator behaviour and natural selection for flower colour in *Delphinium nelsonii*. *Nature* **302,** 422–424 (1983).
93. Newman, E., Anderson, B. & Johnson, S. D. Flower colour adaptation in a mimetic orchid. *Proc. Biol. Sci.* **279,** 2309–13 (2012).
94. Heystek, A., Geerts, S., Barnard, P. & Pauw, A. Pink flower preference in sunbirds does not translate into plant fitness differences in a polymorphic *Erica* species. *Evol. Ecol.* **28,** 457–470 (2014).
95. Sobral, M., Veiga, T., Domínguez, P., Guitián, J. A., Guitián, P. & Guitián, J. M. Selective pressures explain differences in flower color among *Gentiana lutea* populations. *PLoS One* **10,** e0132522 (2015).
96. Temeles, E. J., Newman, J. T., Newman, J. H., Cho, S. Y., Mazzotta, A. R. & Kress, W. J. Pollinator competition as a driver of floral divergence: An experimental test. *PLoS One* **11,** e0146431 (2016).
97. Wolfe, L. M. Reproductive consequences of a flower color polymorphism in *Hydrophyllum appendiculatum*. *Am. Midl. Nat.* **129,** 405–408 (1993).
98. Coberly, L. C. The cost of white flowers: pleiotropy and the evolution of flower color. (2003).
99. Coberly, L. C. & Rausher, M. D. Analysis of a chalcone synthase mutant in *Ipomoea purpurea* reveals a novel function for flavonoids: Amelioration of heat stress. *Mol. Ecol.* **12,** 1113–1124 (2003).
100. Campbell, D. R., Waser, N. M. & Melendez-Ackerman, E. J. Analyzing pollinator-mediated selection in a plant hybrid zone: Hummingbird visitation patterns on three spatial scales. *Am. Nat.* **149,** 295–315 (1997).
101. Meléndez-Ackerman, E., Campbell, D. R. & Waser, N. M. Hummingbird behavior and mechanisms of selection on flower color in *Ipomopsis*. *Ecology* **78,** 2532–2541 (1997).
102. Melendez-Ackerman, E. & Campbell, D. R. Adaptive significance of flower color and inter-trait correlations in an *Ipomopsis* hybrid zone. *Evolution (N. Y).* **52,** 1293–1303 ST–Adaptive significance of flower co (1998).
103. Anderson, L. L. & Paige, K. N. Multiple herbivores and coevolutionary interactions in an *Ipomopsis* hybrid swarm. *Evol. Ecol.* **17,** 139–156 (2003).
104. Wesselingh, R. A. & Arnold, M. L. Pollinator behaviour and the evolution of Louisiana iris hybrid zones. *J. Evol. Biol.* **13,** 171–180 (2000).
105. Emms, S. K. & Arnold, M. L. Site-to-site differences in pollinator visitation patterns in a Louisiana iris hybrid zone. *Oikos* **91,** 568–578 (2000).
106. Wang, H., Conchou, L., Bessière, J.-M., Cazals, G., Schatz, B. & Imbert, E. Flower color polymorphism in *Iris lutescens* (Iridaceae): Biochemical analyses in light of plant–insect interactions. *Phytochemistry* **94,** 123–134 (2013).
107. Schemske, D. W. & Bierzychudek, P. Perspective: Evolution of flower color in the desert annual *Linanthus parryae*: Wright revisited. *Evolution (N. Y).* **55,** 1269–82 (2001).
108. Schemske, D. W. & Bierzychudek, P. Spatial differentiation for flower color in the desert annual *Linanthus parryae*: Was Wright right? *Evolution (N. Y).* **61,** 2528–2543 (2007).

109. Wolfe, L. M. & Sellers, S. E. Polymorphic floral traits in *Linaria canadensis* (Scrophulariaceae). *Am. Midl. Nat.* **138,** 134–139 (1997).
110. Wolfe, L. M. Associations among multiple floral polymorphisms in *Linum pubescens* (Linaceae), a heterostylous plant. *Int. J. Plant Sci.* **162,** 335–342 (2001).
111. Gómez, J. M. Phenotypic selection and response to selection in *Lobularia maritima*: Importance of direct and correlational components of natural selection. *J. Evol. Biol.* **13,** 689–699 (2000).
112. Bartkowska, M. P. & Johnston, M. O. Pollinators cause stronger selection than herbivores on floral traits in *Lobelia cardinalis* (Lobeliaceae). *New Phytol.* **193,** 1039–1048 (2012).
113. Frey, F. M., Dunton, J. & Garland, K. Floral color variation and associations with fitness-related traits in *Malva moschata* (Malvaceae). *Plant Species Biol.* **26,** 235–243 (2011).
114. Schemske, D. W. & Bradshaw, H. D. Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc. Natl. Acad. Sci. U. S. A.* **96,** 11910–11915 (1999).
115. Bradshaw, H. D. & Schemske, D. W. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* **426,** 176–178 (2003).
116. Streisfeld, M. A. & Kohn, J. R. Contrasting patterns of floral and molecular variation across a cline in *Mimulus aurantiacus*. *Evolution (N. Y).* **59,** 2548–2559 (2005).
117. Streisfeld, M. A. & Kohn, J. R. Environment and pollinator-mediated selection on parapatric floral races of *Mimulus aurantiacus*. *J. Evol. Biol.* **20,** 122–132 (2007).
118. Yuan, Y. W., Sagawa, J. M., Young, R. C., Christensen, B. J. & Bradshaw, H. D. Genetic dissection of a major anthocyanin QTL contributing to pollinator-mediated reproductive isolation between sister species of *Mimulus*. *Genetics* **194,** 255–263 (2013).
119. Handelman, C. & Kohn, J. R. Hummingbird color preference within a natural hybrid population of *Mimulus aurantiacus* (Phrymaceae). *Plant Species Biol.* **29,** 65–72 (2014).
120. Campbell, D. R., Bischoff, M., Lord, J. M. & Robertson, A. W. Flower color influences insect visitation in alpine New Zealand. *Ecology* **91,** 2638–2649 (2010).
121. Wessinger, C. A. & Rausher, M. D. Lessons from flower colour evolution on targets of selection. *J. Exp. Bot.* **63,** 5741–5749 (2012).
122. Levin, D. A. & Kerster, H. W. Natural selection for reproductive isolation in *Phlox*. *Evolution (N. Y).* **21,** 679–687 (1967).
123. Levin, D. A. The effect of corolla color and outline on interspecific pollen flow in *Phlox*. *Evolution (N. Y).* **23,** 444–455 (1969).
124. Levin, D. A. The adaptedness of corolla-color variants in experimental and natural populations of *Phlox drummondii*. *Am. Nat.* **106,** 57–70 (1972).
125. Levin, D. A. Reproductive character displacement in *Phlox*. *Evolution (N. Y).* **39,** 1275–1281 (1985).
126. Levin, D. A. & Brack, E. T. Natural selection against white petals in *Phlox*. *Evolution (N. Y).* **49,** 1017–1022 (1995).
127. Hopkins, R., Levin, D. A. & Rausher, M. D. Molecular signatures of selection on reproductive character displacement of flower color in *Phlox drummondii*. *Evolution (N. Y).* **66,** 469–485 (2012).
128. Hannan, G. L. Flower color polymorphism and pollination biology of *Platystemon californicus* Benth. (Papaveraceae). *Am. J. Bot.* **68,** 233–243 (1981).
129. Carlson, J. E. & Holsinger, K. E. Direct and indirect selection on floral pigmentation by pollinators and seed predators in a color polymorphic South African shrub. *Oecologia* **171,** 905–919 (2013).
130. Kay, Q. O. N. Preferential pollination of yellow-flowered morphs of *Raphanus raphanistrum* by *Pieris* and *Eristalis* spp. *Nature* **261,** 230–232 (1976).
131. Stanton, M. L., Snow, A. A., Handel, S. N. & Bereczky, J. The impact of a flower-color polymorphism on mating patterns in experimental populations of wild radish (*Raphanus raphanistrum* L.). *Evolution (N. Y).* **43,** 335–346 (1989).
132. Irwin, R. E. & Strauss, S. Y. Flower color microevolution in wild radish: evolutionary

response to pollinator-mediated selection. *Am. Nat.* **165,** 225–237 (2005).
133. Shrestha, M., Dyer, A. G., Boyd-Gerny, S., Wong, B. B. M. & Burd, M. Shades of red: Bird-pollinated flowers target the specific colour discrimination abilities of avian vision. *New Phytol.* **198,** 301–310 (2013).
134. Whitfield, M., Kohler, A. & Nicolson, S. W. Sunbirds increase foraging success by using color as a cue for nectar quality. *Behav. Ecol.* **25,** 328–334 (2014).
135. Lunau, K., Papiorek, S., Eltz, T. & Sazima, M. Avoidance of achromatic colours by bees provides a private niche for hummingbirds. *J. Exp. Biol.* **214,** 1607–1612 (2011).
136. Rodríguez-Gironés, M. A. & Santamaría, L. Why are so many bird flowers red? *PLoS Biol.* **2,** e350 (2004).
137. Marias, D. D. Molecular Evolution of Anthocyanin Biosynthesis in Morning Glories. 78 (2008).
138. Thomson, J. D. & Thomson, B. A. Pollen presentation and viability schedules in animal-pollinated plants: consequences for reproductive success. *Ecol. Evol. Plant Reprod. New Approaches* 1–24 (1992).
139. Janzen, D. H. When is it coevolution? *Evolution (N. Y).* **34,** 611–612 (2012).
140. Rebelo, A. G. & Siegfried, W. R. Colour and size of flowers in relation to pollination of *Erica* species. *Oecologia* **4,** 584–590 (1985).
141. Turner, R. C., Midgley, J. J. & Johnson, S. D. Evidence for rodent pollination in *Erica hanekomii* (Ericaceae). *Bot. J. Linn. Soc.* **166,** 163–170 (2011).
142. Lombardi, G., Peter, C., Midgley, J. J. & Turner, R. Evidence for rodent-pollination in _Erica langinosa_ (Ericaceae). *South African J. Bot.* **86,** 175–176 (2013).
143. Rebelo, A. G. Management implications. A preliminary synthesis of pollination biology in the Cape flora. *South African Natl. Sci. Program. Rep.* **141,** 193–211 (1987).
144. Geerts, S. & Pauw, A. Hyper-specialization for long-billed bird pollination in a guild of South African plants: the Malachite Sunbird pollination syndrome. *South African J. Bot.* **75,** 699–706 (2009).
145. Turner, R. C., Midgley, J. J., Barnard, P., Simmons, R. E. & Johnson, S. D. Experimental evidence for bird pollination and corolla damage by ants in the short-tubed flowers of *Erica halicacaba* (Ericaceae). *South African J. Bot.* **79,** 25–31 (2012).
146. Geerts, S. & Pauw, A. Easy technique for assessing pollination rates in the genus *Erica* reveals road impact on bird pollination in the Cape fynbos, South Africa. *Austral Ecol.* **36,** 656–662 (2011).
147. Rebelo, A. G., Siegfried, W. R. & Oliver, E. G. H. Pollination syndromes of *Erica* species in the south-western Cape. *S. Afr. J. Bot.* **51,** 270–280 (1985).
148. Barnes, K., Nicolson, S. W. & Van Wyk, B. E. Nectar sugar composition in *Erica*. *Biochem. Syst. Ecol.* **23,** 419–423 (1995).
149. Percival, M. S. Types of nectar in Angiosperms. *New Phytol.* **60,** 235–281 (1961).
150. Baker, H. G. & Baker, I. in *Handb. Exp. Biol.* (eds. Jones, C. E. & Little, R. J.) (Van Nostrand Reynold, 1983).
151. Dupont, Y. L., Hansen, D. M., Rasmussen, J. T. & Olesen, J. M. Evolutionary changes in nectar sugar composition associated with switches between bird and insect pollination: The Canarian bird-flower element revisited. *Funct. Ecol.* **18,** 670–676 (2004).
152. Abrahamczyk, S., Kessler, M., Hanley, D., Karger, D. N., Müller, M. P. J., Knauer, A. C., Keller, F., Schwerdtfeger, M. & Humphreys, A. M. Pollinator adaptation and the evolution of floral nectar sugar composition. *J. Evol. Biol.* (2016). doi:10.1111/jeb.12991
153. Onslow, M. W. *The anthocyanin pigments of plants*. (University Press, 1916). doi:10.5962/bhl.title.1687
154. Zhu, Z., Wang, H., Wang, Y., Guan, S., Wang, F., Tang, J., Zhang, R., Xie, L. & Lu, Y. Characterization of the cis elements in the proximal promoter regions of the anthocyanin pathway genes reveals a common regulatory logic that governs pathway regulation. *J. Exp. Bot.* **66,** 3775–3789 (2015).

155. Chalker-Scott, L. Environmental significance of anthocyanins in plant stress responses. *Photochem. Photobiol.* **70,** 1–9 (1999).
156. Gould, K. S. Nature's swiss army knife: the diverse protective roles of anthocyanins in leaves. *J. Biomed. Biotechnol.* **5,** 314–320 (2004).
157. Smillie, R. M. & Hetherington, S. E. Photoabatement by anthocyanin shields photosynthetic systems from light stress. *Photosynthetica* **36,** 451–463 (1999).
158. Gould, K. S., Kuhn, D. N., Lee, D. W. & Oberbauer, S. F. Why leaves are sometimes red. *Nature* **378,** 241–242 (1995).
159. Landi, M., Tattini, M. & Gould, K. S. Multiple functional roles of anthocyanins in plant-environment interactions. *Environ. Exp. Bot.* **119,** 4–17 (2015).
160. Shih, C.-H., Chu, I. K., Yip, W. K. & Lo, C. Differential expression of two flavonoid 3'-hydroxylase cDNAs involved in biosynthesis of anthocyanin pigments and 3-deoxyanthocyanidin phytoalexins in sorghum. *Plant Cell Physiol.* **47,** 1412–9 (2006).
161. Gould, K. S., Dudle, D. A. & Neufeld, H. S. Why some stems are red: Cauline anthocyanins shield photosystem II against high light stress. *J. Exp. Bot.* **61,** 2707–2717 (2010).
162. Grotewold, E. The genetics and biochemistry of floral pigments. *Annu. Rev. Plant Biol.* **57,** 761–780 (2006).
163. Crowden, R. K. & Jarman, S. J. Anthocyanins in the genus *Erica*. *Phytochemistry* **15,** 1796–1797 (1976).
164. Dooner, H. K. & Robbins, T. P. Genetic and developmental control of anthocyanin biosynthesis. *Annu. Rev. Genet.* **25,** (1991).
165. Irani, N. G., Hernandez, J. M. & Grotewold, E. in *Recent Adv. Phytochem.* **37,** 101–125 (2003).
166. Koes, R., Verweij, W. & Quattrocchio, F. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* **10,** 236–42 (2005).
167. Streisfeld, M. A. & Rausher, M. D. Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. *Evolution (N. Y).* **65,** 629–642 (2011).
168. Phillips, A., Janies, D. & Wheeler, W. Multiple sequence alignment in phylogenetic analysis. *Mol. Phylogenet. Evol.* **16,** 317–330 (2000).
169. Moore, G. E. Cramming More Components onto Integrated Circuits. *Electronics* **86,** 82–85 (1998).
170. Higgins, D. G. & Sharp, P. M. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73,** 237–244 (1988).
171. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).
172. Do, C. B., Mahabhashyam, M. S. P., Brudno, M. & Batzoglou, S. ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res.* **15,** 330–340 (2005).
173. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–3066 (2002).
174. Simossis, V. A. & Heringa, J. PRALINE: A multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res.* **33,** W289-94 (2005).
175. Shatsky, M., Nussinov, R. & Wolfson, H. J. Optimization of multiple-sequence alignment based on multiple-structure alignment. *Proteins Struct. Funct. Genet.* **62,** 209–217 (2006).
176. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
177. Wallace, I. M., Blackshields, G. & Higgins, D. G. Multiple sequence alignments. *Curr. Opin. Struct. Biol.* **15,** 261–266 (2005).
178. Edgar, R. C. & Batzoglou, S. Multiple sequence alignment. *Curr. Opin. Struct. Biol.* **16,** 368–373 (2006).

179. Warnow, T. in *Model. Algorithms Genome Evol.* (eds. Chauve, C., El-Mabrouk, N. & Tannier, E.) 1–90 (Springer, 2013). doi:10.1007/978-1-4471-5298-9_6
180. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
181. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* **17,** 540–552 (2000).
182. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25,** 1972–1973 (2009).
183. Morrison, D. A. Is sequence alignment an art or a science ? *Am. Soc. Plant Taxon.* **6,** 651–655 (2015).
184. Morrison, D. A., Morgan, M. J. & Kelchner, S. A. Molecular homology and multiple-sequence alignment: An analysis of concepts and practice. *Aust. Syst. Bot.* **28,** 46–62 (2015).
185. Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29,** 1695–701 (2012).
186. Puigbò, P., Garcia-Vallvé, S. & McInerney, J. O. TOPD/FMTS: A new software to compare phylogenetic trees. *Bioinformatics* **23,** 1556–1558 (2007).
187. Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. Universal primers for the amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* **17,** 1105–1109 (1991).
188. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–9 (2012).
189. Platt, A. R., Woodhall, R. W. & George, A. L. Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *Biotechniques* **43,** 58–62 (2007).
190. Mugrabi De Kuppler, A. L., Fagúndez, J., Bellstedt, D. U., Oliver, E. G. H., Léon, J. & Pirie, M. D. Testing reticulate versus coalescent origins of *Erica lusitanica* using a species phylogeny of the northern heathers (Ericeae, Ericaceae). *Mol. Phylogenet. Evol.* **88,** 121–131 (2015).
191. Shaw, J., Lickey, E. B., Schilling, E. E. & Small, R. L. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The Tortoise and the hare III. *Am. J. Bot.* **94,** 275–288 (2007).
192. Stamatakis, A., Hoover, P. & Rougemont, J. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **57,** 758–771 (2008).
193. Viguera, E., Canceill, D. & Ehrlich, S. D. Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J.* **20,** 2587–2595 (2001).
194. Thwaites, R. N. & Cowling, R. M. Soil-vegetation relationships on the Agulhas Plain, South Africa. *CATENA* **15,** 333–345 (1988).
195. Ojeda, F., Brun, F. G. & Vergara, J. J. Fire, rain and the selection of seeder and resprouter life-histories in fire-recruiting, woody plants. *New Phytol.* **168,** 155–165 (2005).
196. Johnson, S. D. The pollination niche and its role in the diversification and maintenance of the southern African flora. *Philos. Trans. R. Soc. London B Biol. Sci.* **365,** (2010).
197. Valente, L. M., Manning, J. C., Goldblatt, P. & Vargas, P. Did pollination shifts drive diversification in southern African Gladiolus? Evaluating the model of pollinator-driven speciation. *Am. Nat.* **180,** 83–98 (2012).
198. Rausher, M. D. Evolutionary transitions in floral color. *Int. J. Plant Sci.* **169,** 7–21 (2008).
199. Mol, J., Grotewold, E. & Koes, R. How genes paint flowers and seeds. *Trends Plant Sci.* **3,** 212–217 (1998).
200. van Wilgen, B. W., Higgins, K. B. & Bellstedt, D. U. The role of vegetation structure and fuel chemistry in excluding fire from forest patches in the fire-prone fynbos shrublands of

South Africa. *J. Ecol.* **78,** 210–222 (1990).
201. Liu, Y.-G. G. & Whittier, R. F. Thermal asymmetric interlaced PCR: Automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25,** 674–681 (1995).
202. Liu, Y. G., Chen, Y., Yao-Guang, L. & Yuanling, C. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. *Biotechniques* **43,** 649–656 (2007).
203. Hao, X., Horvath, D., Chao, W., Yang, Y., Wang, X. & Xiao, B. Identification and evaluation of reliable reference genes for Quantitative Real-Time PCR analysis in tea plant (*Camellia sinensis* (L.) O. Kuntze). *Int. J. Mol. Sci.* **15,** 22155–22172 (2014).
204. Banavali, N. K. Analyzing the relationship between single base flipping and strand slippage near DNA duplex termini. *J. Phys. Chem. B* **117,** 14320–14328 (2013).
205. Banavali, N. K. Partial base flipping is sufficient for strand slippage near DNA duplex termini. *J. Am. Chem. Soc.* **135,** 8274–8282 (2013).
206. Manjari, S. R., Pata, J. D. & Banavali, N. K. Cytosine unstacking and strand slippage at an insertion-deletion mutation sequence in an overhang-containing DNA duplex. *Biochemistry* **53,** 3807–3816 (2014).
207. Dao, T. T. H., Linthorst, H. J. M. & Verpoorte, R. Chalcone synthase and its functions in plant resistance. *Phytochem. Rev.* **10,** 397–412 (2011).
208. Rozmer, Z. & Perjési, P. Naturally occurring chalcones and their biological activities. *Phytochem. Rev.* **15,** 87–120 (2016).
209. Hardy, G. H. Mendelian proportions in a mixed population. *Science (80-.).* **28,** (1908).
210. Ochman, H., Gerber, A. S. & Hartl, D. L. Genetic Applicatins of an Inverse Polymerase Chain Reaction. *Genetics* **120,** 621–623 (1988).

Appendix A: The biodiversity hotspot as evolutionary hot-bed: spectacular radiation of Erica in the Cape Floristic Region

Pirie et al. BMC Evolutionary Biology (2016) 16:190
DOI 10.1186/s12862-016-0764-3

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

The biodiversity hotspot as evolutionary hot-bed: spectacular radiation of *Erica* in the Cape Floristic Region

M. D. Pirie[1,2*], E. G. H. Oliver[3], A. Mugrabi de Kuppler[4], B. Gehrke[2], N. C. Le Maitre[1], M. Kandziora[2] and D. U. Bellstedt[1]

Abstract

Background: The disproportionate species richness of the world's biodiversity hotspots could be explained by low extinction (the evolutionary "museum") and/or high speciation (the "hot-bed") models. We test these models using the largest of the species rich plant groups that characterise the botanically diverse Cape Floristic Region (CFR): the genus *Erica* L. We generate a novel phylogenetic hypothesis informed by nuclear and plastid DNA sequences of c. 60 % of the c. 800 *Erica* species (of which 690 are endemic to the CFR), and use this to estimate clade ages (using RELTIME; BEAST), net diversification rates (GEIGER), and shifts in rates of diversification in different areas (BAMM; MuSSE).

Results: The diversity of *Erica* species in the CFR is the result of a single radiation within the last c. 15 million years. Compared to ancestral lineages in the Palearctic, the rate of speciation accelerated across Africa and Madagascar, with a further burst of speciation within the CFR that also exceeds the net diversification rates of other Cape clades.

Conclusions: *Erica* exemplifies the "hotbed" model of assemblage through recent speciation, implying that with the advent of the modern Cape a multitude of new niches opened and were successively occupied through local species diversification.

Keywords: Biodiversity, Cape Floristic Region, Diversification, *Erica*, Evolution

Background

Biological diversity is spread unevenly across the globe and across the tree of life, clustered in geographic hotspots [1] and species-rich clades [2–4]. Diverse organisms with a range of life history and other traits have radiated in environments with different topographies, climates, and histories. The hyper-diverse tropical Andes set the stage for a spectacular radiation of lupins (*Lupinus*; Fabaceae) [2], the Amazon rainforest for that of *Inga* (Fabaceae) [3], the Mediterranean hotspot for that of carnations (*Dianthus*; Caryophyllaceae) [4] and Southern Africa's succulent karoo for that of ice plants (Ruschioideae; Aizoaceae) [5]. These species-rich groups present us with a rich and powerful source of data for bettering our understanding of the origins of biological diversity: we can analyse numerous speciation events in comparable biological systems within evolutionarily recent, and hence more tractable, timescales.

The mountainous landscape of South Africa's Cape Floristic Region (CFR) hotspot [1] hosts 9000 plant species, 70 % endemic [6, 7], within only c. 90,000 km^2. Thirty-three species rich "Cape clades" collectively account for around half of this remarkable richness [7], of which the genus *Erica* L. would be the largest, if the around 690 species [8] represent a single clade. *Erica* species are woody shrubs that dominate the CFR's heathland "fynbos" vegetation as well as heathland ecosystems in the western Palearctic (including the Mediterranean) and mountain "sky islands" of Tropical Africa [9] and Madagascar. However, the numbers of species in regions outside the CFR are lower by an order of magnitude. Such striking regional asymmetries in species richness within a group of notably

* Correspondence: pirie@uni-mainz.de
[1]Department of Biochemistry, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa
[2]Institut für Spezielle Botanik und Botanischer Garten, Johannes Gutenberg-Universität, Anselm-Franz-von-Bentzelweg 9a, 55099 Mainz, Germany
Full list of author information is available at the end of the article

consistent habit pose a fascinating evolutionary conundrum, the solutions for which can inform our general understanding of the assemblage of regional biotas.

Here we ask: a) Does the extraordinary diversity of *Erica* in the CFR stem from a single common ancestor in the Cape? b) Are regional asymmetries in species richness the result of shifts in rates of diversification within the *Erica* clade and in different areas? c) Does the radiation of Cape *Erica* reflect a 'museum' (low extinction) and/or 'hot-bed' (high speciation) model for the biotic assemblage of the CFR? Such an analysis demands a credible, detailed and dated phylogenetic tree of the group: we present a phylogenetic hypothesis for *Erica* based on greatly increased sampling of species and molecular markers.

Methods

Taxon and character sampling: Our phylogenetic hypothesis is informed by nuclear and plastid DNA sequences of c. 60 % of all *Erica* species, represented by 606 accessions of 488 species and 28 sub-specific taxa from across the geographic range of the clade (17 of 19 Palearctic species [89 %], 414 of 690 CFR [60 %]; 13 of 23 Tropical Africa [57 %]; 27 of 51 Drakensberg [53 %]; and 17 of c. 41 Madagascar/Mascarenes [42 %]), plus six outgroups (Additional file 1: Table S1). Specimens were collected in the field and determined by EGHO. Vouchers were lodged in herbaria (Additional file 1: Table S1), and leaf samples dried in silica gel and archived at -20 °C to preserve the DNA. Most sequences were obtained newly for this study, with some from previous work [10–12]. We obtained DNA sequences mostly using a direct PCR amplification protocol [13] with universal angiosperm primers [14, 15] as described in [12]. We employed a targeted supermatrix sampling strategy [16]: ITS and chloroplast *trnT-trnL* and *trnL-trnF-ndhJ* spacer sequences were obtained for all samples, and other plastid markers (*trnL* intron, *atpI-atpH* spacer, *trnK-matK* intron and *matK* gene, *psbM-trnH* spacer, *rbcL* gene, *rpl16* intron, *trnL-rpl32* spacer) were added for taxa selected, on the basis of preliminary analyses, as representative of early diverging lineages within each of the major subclades, in order to improve resolution of deeper nodes in the plastid tree. Sequences in general, and particularly ITS, were inspected to confirm the absence of polymorphism and (other) evidence of paralogy (e.g. indels in coding regions). An accessions table including Genbank accessions numbers is presented in Additional file 1: Table S1.

Phylogenetic inference: Individual matrices including all sequences for each marker were aligned in Mesquite [17] and imported into SequenceMatrix [18] to export concatenated matrices (excluding taxa causing topological conflict between gene trees; see below) for further analyses. A matrix of 63 phylogenetically representative taxa for which a minimum of 14 of the 20 data partitions were available was analysed using PartitionFinder [19] to infer best fitting data partitioning strategies and substitution models (heuristic search strategy 'greedy'; comparison of fit using the Bayesian information criterion). Individual markers, coding and non-coding regions within those markers, and codon positions within protein coding genes were all specified as potential data partitions. Maximum likelihood (ML) analyses were performed using RAxML on CIPRES [20, 21] incorporating the data partitions inferred using PartitionFinder. Clade support was estimated using bootstrapping halted automatically by RAxML following the majority-rule 'autoMRE' criterion. To test for experimental error, confirm congruence of individual plastid markers, and to infer and compare gene trees we performed preliminary phylogenetic analyses of individual markers separately. These were followed by final analyses of ITS, combined plastid data and combined ITS and plastid data. Fifteen taxa causing topological conflict subject to ≥70 % bootstrap support (BS) between ITS and combined plastid gene trees were excluded from analyses of the concatenated data (leaving 597) under the assumption that such conflicts reflect (apparently uncommon) incidences of reticulation or incomplete lineage sorting that violate the assumption of a bifurcating tree [22]. Further phylogenetic analyses were performed using BEAST 1.8 [23] (as below).

Molecular dating: Two dating methods were employed on the Ericeae matrix: BEAST [23], using the 63 taxa matrix but excluding the most distant outgroup, *Empetrum*; and RELTIME [24], using the 597 taxa ML tree from the RAxML concatenated data analysis, removing *Empetrum* and *Corema album*. We used the 63 taxa matrix with BEAST because of the failure of multiple runs to converge with the full supermatrix, a not unexpected phenomenon in the presence of large proportions of missing data [16]. The targeted sampling strategy meant that the same internal focal nodes are represented in both trees. For BEAST, the root age (most recent common ancestor of *Erica* and *Daboecia*) was constrained based on the results of [25] in Ericaceae-wide analyses employing multiple fossil calibrations (producing results consistent with those presented in [26]). We used a normal distribution with mean 62 Mya and SD 10, giving a 95 % prior probability distribution of 42-82 Mya reflecting uncertainty in the original analyses [25]. In a further analysis an additional prior was implemented to reflect the age of Ericaceae pollen in sediments offshore of Southern Africa [27] and thereby test the impact on age estimates assuming that this pollen record represents *Erica*. For this, we used an exponential distribution with offset of 10 Mya (a hard minimum) and mean of 2.0, giving a 95 % prior probability distribution of 10-16

Mya (i.e. a soft maximum) for the stem node of Cape *Erica*. This is to assume that the Cape *Erica* clade is at least as old as the age of the pollen record and may be older to a limited degree. Following preliminary partitioned analyses that failed to converge, the data were not partitioned; we applied a GTR + G substitution model, lognormal relaxed clock, Yule process speciation model, and otherwise default priors, and performed two runs of 10 million generations sampling every 1000 in each case. Convergence was assessed using TRACER 1.6 [28] and Are We There Yet [29], and the results summarised using programs of the BEAST package. For RELTIME we assumed local clocks and imposed age constraints by means of a point estimate for the root node (the minimum, mean and maximum ages as above).

Diversification rates analyses: To infer the net diversification rate of the *Erica* Cape clade and compare it to those of other Cape and rapidly radiating clades, we used the method of Magallón & Sanderson ([30], as implemented in GEIGER; [31]). For Cape *Erica*, we used species richness and full range of crown node ages (minimum and maximum under RELTIME and highest posterior density intervals under BEAST) as inferred here. For comparison, we performed the same calculations based on data from the literature for the recent rapid radiations of lupins [2], *Inga* [3], carnations [4], and ice plants [5]; as well as the Cape clades *Muraltia* [32], *Pentameris* [33] *Protea* [34] and Restionoideae ("African Restionaceae") [32]. The latter are examples for which detailed time-calibrated phylogenies of ancestrally CFR species – not those that also diversified in other areas – are available. We did not account for the impact on crown node age estimates of unsampled species during the calculation, and used relative extinction rates of 0.9 and zero across the board.

To test whether diversification rate heterogeneity reflects different speciation and extinction rates between geographic areas, we used MuSSE (Multiple State Speciation and Extinction) as implemented in diversitree 0.93 [35]. MuSSE uses maximum likelihood to estimate the values of different parameters under a constant birth death model: speciation (λ) and extinction (μ) rates under each of the discrete states of the character (in this case, geographic distribution), and rates of transition (q) from one state (area) to another. We compared the rates between Palearctic, Tropical African, Madagascan, Drakensberg and Cape species of *Erica*, *Calluna* and *Daboecia*. The areas are indicated in Fig. 1 and were so defined because they are often compared in the literature, are largely geographically isolated and <1 % of *Erica* species are widespread between any two of them (these limited to two species in both the Cape and Drakensberg – *E. caffra* and *E. cerinthoides* – and one in the Palearctic and Tropical Africa – *E. arborea*). We used the discrete multistate model, instead of GeoSSE, that models widespread geographic distributions, to represent multiple areas (rather than just two in GeoSSE) under the assumption that widespread distributions were rare throughout the evolutionary history of the group. We coded the three samples of widespread species according to the region in which they were collected under the assumption that effectively failing to sample such species across their wider distribution would have little impact on the results. We used the rate-smoothed 597-taxa RAxML tree, having removed multiple accessions of species and outgroups (leaving 487 terminals), and corrected for incomplete sampling by assigning region-specific sampling fractions. We did not consider phylogenetic uncertainty, as the major clades are well supported and largely restricted to single regions and thus the uncertainty regarding our question remains low. We compared maximum likelihood estimates given models considering different regions (either 5 distinct regions or combinations of Palearctic, Cape and the rest of Africa or of Europe versus Africa and or Cape versus other regions) and considering single versus multiple rates for speciation and/or extinction. For all but the unconstrained model, we constrained the transition rates to one parameter. Thereafter, for the best model, we tested whether constraining the transition rate reduces the likelihood. We compared the fit of the models to the data using the anova function in diversitree and using the AIC to compare the fit of the models. The parameters for the best fitting model were then calculated using a Bayesian MCMC approach run for 10,000 steps using an exponential probability distribution as prior for the underlying rates in the model. We assessed convergence by comparing the probability values of the sampling after excluding a burnin of 25 %.

To further determine whether there is diversification rate heterogeneity in the *Erica* dataset, we used BAMM 2.5 and Bammtools 2.1 [36, 37]. The method compares the fit of different models (a series of diversification processes) assuming different numbers of shifts based on a reversible jump MCMC to explore parameter space. We used the pruned, rate-smoothed RAxML tree, as above, and corrected for non-random species sampling by assigning regional specific proportions to the few, largely endemic, clades. We used "setBAMMpriors" to adjust the priors according to the scaling of the tree. The initial speciation rate was set to 0.18 and extinction rate to 0.111 according to inferred rates for Ericaceae [25]. Preliminary results showed that different initial speciation and extinction rate did not have a large effect on our results. The MCMC was run for 10,000,000 generations, with every 1000 generation saved. To assess convergence, the likelihood of all sampled generations was plotted in R (burnin = 10 %) and ESS values for the

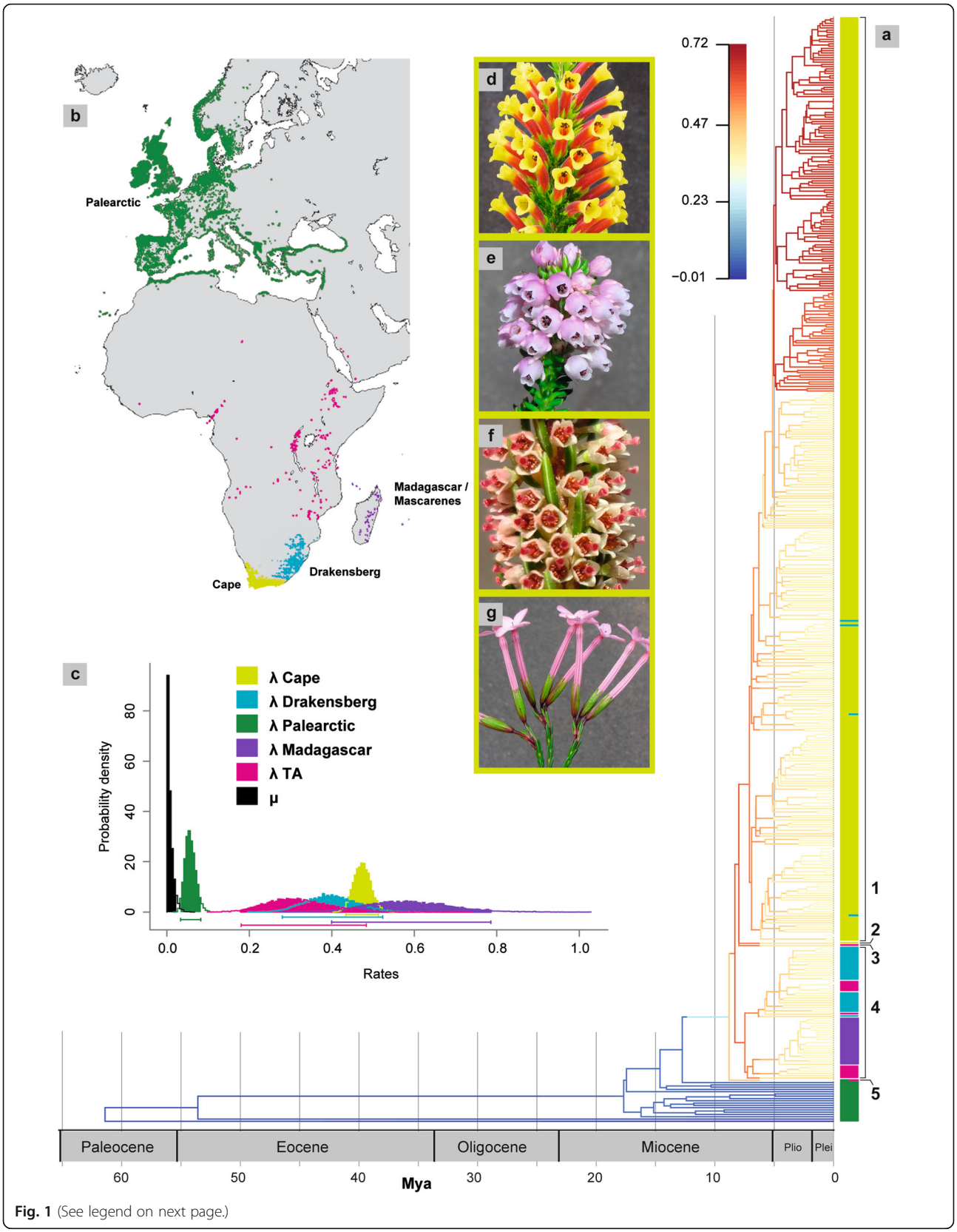


Fig. 1 (See legend on next page.)

(See figure on previous page.)
Fig. 1 The diversification of *Erica* in space and time. **a** Time-calibrated phylogenetic tree of 478 extant lineages that populated the radiation of *Erica* with branches coloured according to mean net diversification rates (scale indicates species per million years) inferred using BAMM, with regions of samples indicated by the coloured bar at the terminals and clades/species referred to in the text indicated with numbers: 1 = Cape clade; 2 = *E. pauciovulata*; 3 = *E. trimera*; 4 = Afrotemperate clade; 5 = *E. arborea*. **b** Geographic distribution of *Erica*, based on collections databased by GBIF, showing Palearctic, Tropical Africa, Madagascar, Drakensberg and Cape regions. **c** Region specific speciation rates (λ) and the single extinction rate (μ). **d**-**g** Examples of the spectacular floral diversity of Cape *Erica*: d) *E. macowanii*, e) *E. pulvinata*, f) *E. coarctata*, and g) *E. jasminiflora*

likelihood and the inferred numbers of shifts were calculated using the coda package [38]. It was not possible to compare Bayes Factors for zero rate shifts with those for given numbers of shifts (see BAMM Documentation part 7.6), but we compared the prior probability of a given number of shifts to the posterior probability to confirm that these differed. We then computed the set of credible shifts and reconstructed the mean of the marginal posterior density of speciation, extinction and net diversification rates across the tree. We sought to assess whether the BAMM results are dependent on the particular topology and branch lengths of the phylogenetic tree used above by repeating the analyses with 25 randomly selected, rate smoothed and pruned RAxML bootstrap trees.

Results

Results of preliminary analyses of individual plastid markers (TreeBase study accession URL: http://purl.org/phylo/treebase/phylows/study/TB2:S18291) showed no conflicting nodes supported at ≥70 % bootstrap support (BS). The better resolved combined plastid gene tree (Additional file 2: Figure S1a) was largely consistent with that of ITS (Additional file 2: Figure S1b). Fifteen taxa causing topological conflict subject to ≥70 % BS between ITS and combined plastid gene trees (indicated in Additional file 2: Figures S1a and b) were excluded from analyses of the concatenated data (leaving 597; Additional file 2: Figure S1c). Further exclusion of one Cape species, *E. pauciovulata*, resulted in an increase in support for a single Cape clade (see below) from 70 % BS (Additional file 2: Figure S1c) to 89 % (Additional file 2: Figure S1d).

The *Erica* phylogeny roots in a northern Palearctic grade subtending a southern African/Madagascan clade. The latter comprises a deep polytomy including a) the Cape clade, including all but one Cape species plus four found in the Drakensberg (two of which also distributed in the Cape), b) a single further Cape species, *E. pauciovulata*, c) *E. trimera* (Tropical Africa), d) the 'Extra-CFR African clade' that includes all other Drakensberg and Tropical African species (except for *E. arborea*) and a clade of all Madagascan/Mascarene species, and e) *E. arborea* (Palearctic and TA) (Fig. 1, Additional file 2: Figure S1). Our age estimates for clades within *Erica* (Fig. 1, Additional file 3: Figure S2a-c) are based on the similar results of both the two relaxed clock molecular dating methods (RELTIME, Additional file 3: Figure S2a; and BEAST, Additional file 3: Figure S2b) with secondary calibration, and additional BEAST results (Additional file 3: Figure S2c) further constrained using the Southern Africa offshore microfossil record. The crown node of *Erica* was estimated at 18 (24-12) Mya (RELTIME on the matrix of 597 taxa; Additional file 3: Figure S2a) and 27-19/31-12 Mya (95 % posterior probability ranges from BEAST, using the reduced matrix of 62 taxa, with/without microfossil evidence; Additional file 3: Figure S2b/c). The radiation of lineages within the African/Madagascan clade was estimated at 9 (12-6) Mya (RELTIME) and 14-11/17-7 Mya (BEAST). The stem node of the Cape clade was estimated at 8 (11-6) Mya (RELTIME) and 12-10/15-6 Mya (BEAST); the crown node at 7 (9-5) Mya (RELTIME) and 11-9/15-6 Mya (BEAST).

Given our dated phylogenetic trees, the net diversification rate of Cape *Erica* was 0.28-0.7 (assuming relative extinction of 0.9) or 0.39-0.97 (relative extinction zero) species per million years; estimated rates of other Cape clades and faster recent species radiations reported worldwide are presented in Table 1.

MuSSE analyses performed with diversitree identified differences in speciation rates specific to geographic regions, with the best scoring model otherwise including only single rates for extinction and for transitions (dispersals between regions; Table 2). The lowest diversification rate is in the Palearctic, while rates in all other regions are high (Fig. 1c; Table 3). Three further models scored within ≤2 of the best model according to the AIC (Table 2); these included single rates for transitions, either five or two parameters for region-specific speciation rates (the latter, Palearctic versus Africa) and two differing rates for extinction (either Palearctic versus Africa or Cape versus all other areas; Additional file 4: Table S2). The inferred rates for extinction were universally similar and low (Additional file 4: Table S2)

BAMM analyses also indicated strong support for heterogeneous diversification dynamics within *Erica*, in the form of multiple accelerations in the rate of diversification (Figs. 1 and 2, Additional file 5: Figure S3a; posterior probability [PP] of a single rate model = 0; PP density of 2-4 rate shifts = 0.74; 2-5 rate shifts = 0.87; Additional file 5: Figure S3b). Extinction rates appear to be constant through time, but speciation rates vary greatly. We inferred 14 distinct configurations within the 95 % credible shift sets. Distinct diversification regimes were associated

Table 1 Plant diversification rates in the CFR and beyond. Net diversification rates of Cape clades and other recent radiations worldwide in species per million years, estimated using species numbers and clades ages with the method of S Magallón and MJ Sanderson [30]

Clade	Species numbers	Crown age (Mya)	Reference	Rate (Species/Mya) relative extinction = 0.9	Rate (Species/Mya) relative extinction = 0.0	Note
Cape clades:						
Cape *Erica*	690	6.0–15.0	This paper	0.28–0.70	0.39–0.97	Range of estimates from RELTIME and BEAST analyses
Muraltia	124	8.6–16.4	[32]	0.15–0.29	0.25–0.48	Presented range (molecular dating)
Restionoideae	350	31.7–65.4	[32]	0.05–0.11	0.07–0.16	Presented range (molecular dating)
Pentameris	83	13.2–16.1	[33]	0.13–0.16	0.23–0.28	Presented range (molecular dating)
Protea	69	11.2–27.2	[34]	0.07–0.18	0.13–0.32	Presented range (molecular dating)
Other clades:						
Andean lupins	85	1.6–2.3	[2]	0.96–1.38	1.64–2.37	*Lupinus* stem calibrated at 21.16 Ma
		1.2–1.8	As above	1.24–1.86	2.13–3.18	*Lupinus* stem calibrated at 16.01 Ma
Ice plants (Aizoaceae: Ruschioideae)	1563	0.6–7.0	[5]	0.28–3.32	0.5–5.88	Calibrated with ITS substitution rates
		8.0–9.4	As above	0.53–0.63	0.71–0.83	Calibrated with plastid substitution rates
European *Dianthus*	200	0.61–2.4	[4]	1.23–4.90	1.9–7.55	Min. and max. ages reported
Inga	300	1.6–9.8	[3]	0.34–2.11	0.51–3.13	Calibrated with *trnL-F*/ITS substitution rates

Table 2 Comparison of different MuSSE models estimated for 478 species of *Erica*

Geographical regions	constraints	No. of λ parameters	No. of μ parameters	No. of q parameters	lnLik	AIC	AIC weights	Parameter estimates
Palearctic, Cape, Drakensberg, Madagascar, Tropical Africa	-	5	5	20	-1048.9	2157.7	6.056120e-06	
	λ, q	1	5	1	-1082.9	2179.8	9.621447e-11	
	μ, q	***5***	***1***	***1***	***-1060.8***	***2135.5***	***4.007405e-01***	Table 3
	μ	5	1	20	-1051.3	2154.7	2.714164e-05	
One region	λ, μ, q	1	1	1	-1162.5	2331.0	6.040484e-05	
Palearctic, Cape, rest of Africa	λ, q	3	5	1	-1062.4	2142.8	1.041570e-02	
	μ, q	5	3	1	-1060.8	2139.5	5.423432e-02	
	λ, μ, q	3	3	1	-1063.3	2140.6	3.129048e-02	
Palearctic, Africa	λ, q	2	5	1	-1062.4	2140.8	2.831280e-02	
	μ, q	**5**	**2**	**1**	**-1060.8**	**2137.5**	**1.474242e-01**	Additional file 4: Table S2, Model 1
	λ, μ, q	**2**	**2**	**1**	**-1063.6**	**2137.1**	**1.800643e-01**	Additional file 4: Table S2, Model 2
Cape, other	λ, q	2	5	1	-1081.1	2178.1	2.251079e-10	
	μ, q	**5**	**2**	**1**	**-1060.8**	**2137.5**	**1.474242e-01**	Additional file 4: Table S2, Model 3
	λ, μ, q	2	2	1	-1100.9	2211.7	1.138265e-17	

Abbreviations: λ – speciation rate, μ - extinction rate, q – transition rate, lnLik – logarithm of likelihood, AIC – Akaike information criterion. The best scoring model is indicated with bold italics (parameter estimates presented in Table 3); three models with AIC scores within 2 of the best scoring model are indicated in bold (parameter estimates presented in Additional file 4: Table S2)

Table 3 Parameter estimates given the best scoring MuSSE model

	λ Cape	λ Drakensberg	λ Palearctic	λ Madagascar	λ TA	μ	transition rate	*p*
Min.	0.4043	0.1971	0.01685	0.2771	0.1013	2.20E-7	2.39E-4	-1074
1st Qu.	0.4595	0.3575	0.04795	0.5179	0.2658	2.24E-3	1.15E-3	-1064
Median	0.4734	0.3978	0.05598	0.5818	0.3149	5.37E-3	1.43E-3	-1062
Mean	0.4735	0.4002	0.0569	0.5869	0.3216	7.65E-3	1.49E-3	-1062
3rd Qu.	0.4872	0.4402	0.06498	0.6527	0.3702	1.08E-2	1.77E-3	-1061
Max.	0.5536	0.6644	0.11022	1.0282	0.7366	6.40E-2	3.85E-3	-1059

with the *Erica* clade, the African/Madagascan clade (either including *E. arborea*, or not) and within the Cape clade; the former is found in only four of the nine configurations with highest PP which together sum to PP > 0.90 (Fig. 2) (and generally fewer than half of each of the individual configurations based on 25 bootstrap trees), whilst the two latter shifts are found in all of them (the African/Madagascan clade shift in almost all, and the shift within the Cape clade generally in more than half of the individual bootstrap configurations; Additional file 6: Table S3).

Discussion

Whilst the richness of *Erica* species in the CFR is renowned, our results can finally confirm *Erica* as the most species rich Cape clade. With the possible exception of a single anomalous species (*E. pauciovulata*), all *Erica* in the CFR that we analysed can be traced back to a single common ancestor that colonised the region no earlier than c. 15 Mya, and all but a handful are endemic (Fig. 1). A Late Miocene initiation of the *Erica* radiation in the CFR is consistent with the first appearance of

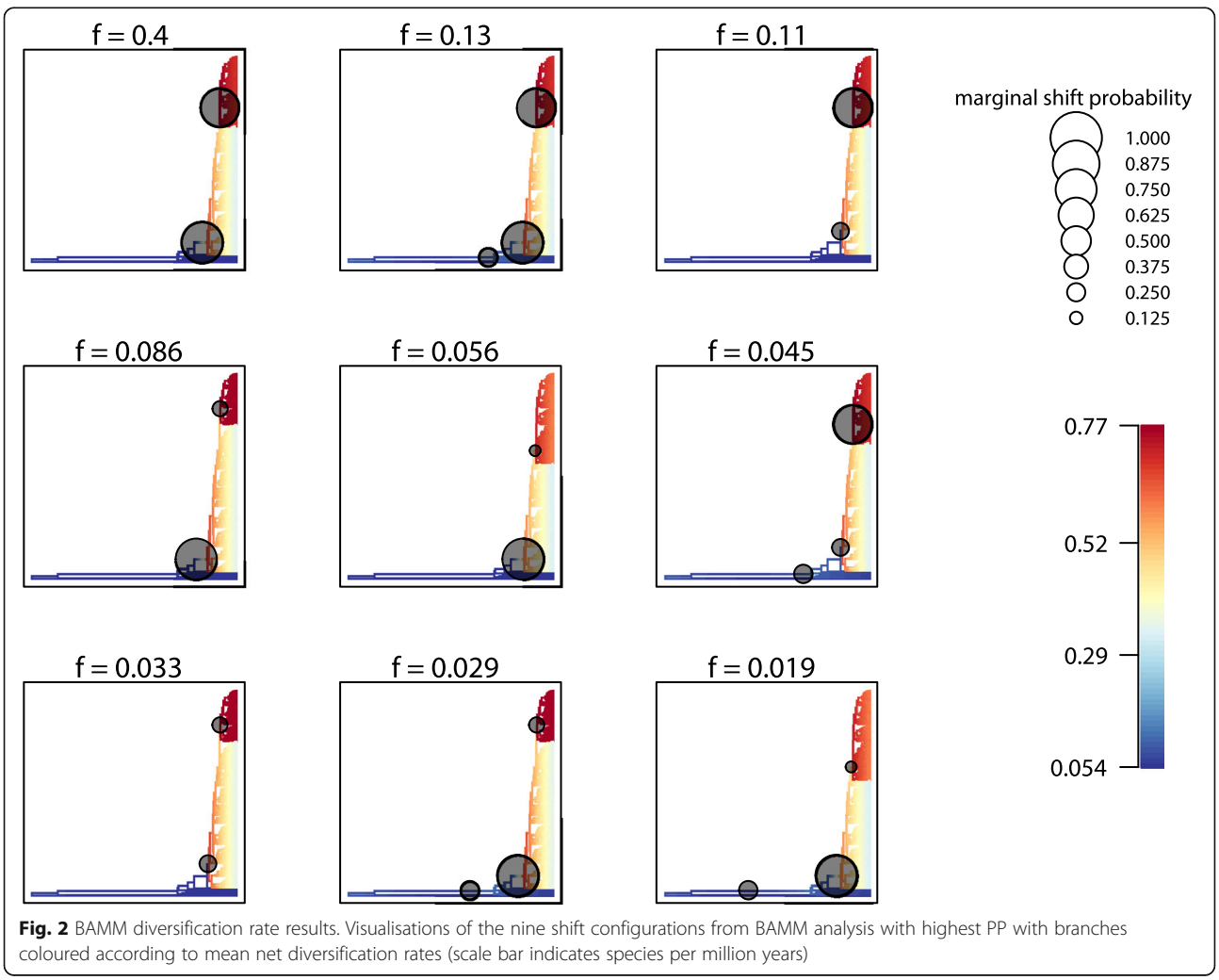


Fig. 2 BAMM diversification rate results. Visualisations of the nine shift configurations from BAMM analysis with highest PP with branches coloured according to mean net diversification rates (scale bar indicates species per million years)

pollen of Ericaceae (and various other typical fynbos groups) in the fossil record in Southern Africa after 10 Mya [27]. Cape clades differ widely in age [39], and Cape *Erica* is neither conspicuously old nor young in this context. Its net diversification rate is modest compared to the most spectacular examples in flowering plants, documented from the much greater areas of the Andean Páramo [2] and the Mediterranean [4]. The *Erica* diversification rate is more similar to those of other rapidly evolving Cape clades, although notably faster than any that we compared (Table 1).

This remarkable radiation of *Erica* in the CFR is in stark contrast to the comparatively impoverished older Palearctic lineages. The heathers originated in the Northern Hemisphere [10, 12] and northern lineages (including monotypic *Calluna* and two species of *Daboecia*) are older than the single southern *Erica* clade ('African/Malagasy Erica'; [10]). Higher diversity in Cape compared to Mediterranean clades has been attributed to lower rates of extinction [40]. Our results instead imply slower speciation in the wider western Palearctic (Fig. 1; Table 3 and Additional file 4: Table S2), although this conclusion must be qualified by the known difficulty of inferring extinction rates from molecular phylogenies [41]. Although the ranges of speciation (and hence net diversification) rates in different regions outside the Palearctic overlapped (perhaps a methodological artefact caused by the much lower species numbers outside the Cape [42]), we discovered evidence for a rate increase within the Cape clade (Figs. 1 and 2). Phylogenetic uncertainty within the Cape clade is considerable (reflecting the short internal branches typical of bursts of lineage diversification [37]), but geographically, this diversification centres on lineages of the large SW-clade [10] mostly restricted within the South-Western CFR. Irrespective of inferred shifts in diversification rates, the greater areas of equivalent habitat in Tropical Africa and the Drakensberg (for similarly distributed *Protea*, estimated at roughly 17-fold; [34]) and in Madagascar compared to the CFR represent far lower densities of species and of speciation events through time given the phylogeny and clade ages inferred here.

The CFR is one of a number of mountainous and Mediterranean climate regions with unique and hyperdiverse biotas that both coincidentally, and as the result of worldwide climatic changes, originated within similar, relatively recent timeframes [2, 25, 40, 43]. The modern CFR was shaped by globally influenced palaeoclimatic dynamics that established during the Miocene, particularly world-wide cooling that led to aridification [44], and the establishment of the cold Benguela current off the south-west African coast, that led to the development of a winter rainfall regime and frequent fires [45, 46]. The disappearance of more mesic tropical forest elements from fossil deposits [27, 47] was followed by the appearance of more arid and/or fire adapted elements such as Aizoaceae, including Ruschioideae (in the succulent karoo), and *Erica* (in the CFR) [27]; the latter with its reduced leaf area and resistant yet inflammable wax-rich cuticles [48] combined with post-fire re-sprouting and smoke-stimulated re-seeding recruitment strategies [49, 50].

As with other mountainous hotspots [43], the CFR was also influenced by local uplift, that occurred during the Miocene [51, 52]. The high species richness and local endemism of the present day Cape is plausibly a direct result of this uplift: new niches opened, with physical barriers to gene flow between them, creating a stimulus for allopatric speciation [53]. Topographical complexity also creates local temperature and moisture gradients [53], and the patchwork of soils derived from the different lithologies of the Cape [54, 55] adds a further dimension to the resulting fine-scaled mosaic of habitats. By contrast to regions of the Northern Hemisphere, the Cape was buffered from the extremes of Pleistocene glacial cycles, and by implication from resulting extinction [56]; instead (less extreme) shifts in multiple local-scale ecological gradients, acting in concert, might actually drive speciation [57]. Key innovations in particular groups are also often mooted, such as adaptations to specialised pollinator interactions [58] that might reinforce speciation [59]. The numerous apparent shifts in pollination syndrome in *Erica* and the higher diversity of different syndromes in *Erica* in the Cape than elsewhere [10] make the latter a tempting explanation for the acceleration of the Cape *Erica* radiation.

However, meta-analyses of Cape phylogenies have provided support for multiple such hypotheses, with evidence for both ecological and/or pollinator shifts [60] and distributional and phenological shifts [61] in e.g. *Muraltia*, Cape Restionaceae and *Pentameris*; each of these and others too (such as edaphic shifts apparent in *Babiana* (Iridaceae) [62]) may have played a role. Given the results presented here, it is also plausible that (combinations of) factors specific to the most species-rich SW region may be responsible for the highest rates of diversification within the CFR. The relative contribution of these different factors overall is still hotly debated, and with a phylogenetic hypothesis for the clade now available, Cape *Erica* offers the greatest single source of data for further testing their importance in the assemblage of the flora.

Conclusions

In two contrasting perspectives, the CFR is interpreted as a 'museum' of diversity [63], with persistence of pre-Miocene lineages [64] and lower extinction e.g. compared to the Mediterranean [40]; or the evolutionary 'hot-bed' of (recent) radiations [32, 39, 65, 66]. These models are not mutually exclusive [39, 67]. However, our results further

weigh the balance in favour of the latter. The largest Cape clade, *Erica*, represents more species than included in most meta-analyses of Cape clades performed to date. Much of this remarkable diversity originated within the last few million years.

Additional files

Additional file 1: Table S1. Accessions table including Genbank accession numbers. (XLSX 119 kb)

Additional file 2: Figure S1. Phylogenetic hypotheses: best trees with bootstrap support values from RAxML analyses of a) concatenated plastid data and b) from nuclear ribosomal ITS (with taxa showing conflicting positions according to the two gene trees highlighted in yellow); and c) and d) of the combined data (excluding conflicting taxa): c) with and d) without *Erica pauciovulata* (exclusion of which leads to increased support for the Cape clade from 70 % to 89 %). (ZIP 8409 kb)

Additional file 3: Figure S2. Relaxed clock molecular dating results: a) age estimates for clades within *Erica* inferred using RELTIME [24] with the best tree from RAxML (Additional file 2: Figure S1c); b) and c) phylogeny and relaxed clock molecular dating age estimates for clades within *Erica* inferred using BEAST [23] on a reduced matrix of 62 taxa b) with and c) without additional constraint based on microfossil evidence (error bars represent 95 % Posterior Probability (PP) intervals; PP clade support is indicated at nodes). (ZIP 11562 kb)

Additional file 4: Table S2. Parameter estimates given the three best scoring suboptimal MuSSE models. (DOCX 15 kb)

Additional file 5: Figures S3. BAMM diversification rate results: a) BAMM tree as presented in Fig. 1, including labels for tips and nodes referred to in the text; branches subtending *Erica*, *Calluna* and *Daboecia* are not to scale. b) Probabilities of overall numbers of diversification shifts inferred using BAMM. (ZIP 2444 kb)

Additional file 6: Table S3. Summary of BAMM results based on 25 rate-smoothed RAxML bootstrap trees. (DOCX 12 kb)

Acknowledgements
We thank F.P.D. Cotterill, T. van der Niet and J.W. Kadereit for comments on drafts of the paper; D. Franke for assistance with graphics; J. Fagúndez, A. Hitchcock, R. Turner, M. Muasya, C. Stirton, R. Clark, B. Bytebier, M. Pimentel, F. Ojeda, C. Merry, and many others for providing samples; and Cape Nature and South Africa National Parks for assistance with permits. The authors gratefully acknowledge the computing time granted on the supercomputer Mogon at Johannes Gutenberg University Mainz (https://hpc.uni-mainz.de/).

Funding
South African National Research Foundation (NRF; DUB and MDP); a postdoctoral fellowship from the Claude Leon Foundation (MDP); DFG (PI1169/1-1 to MDP); and the Ministerium für Klimaschutz, Umwelt, Landwirtschaft, Natur- und Verbraucherschutz des Landes Nordrhein-Westfalen, the Faculty of Agriculture Lehr- und Forschungsschwerpunkt „Umweltverträgliche und Standortgerechte Landwirtschaft", Bonn University; and the Landgard foundation (AMK). Any opinion, finding and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept liability in this regard.

Availability of data and material
DNA sequence data presented here will be available on publication from Genbank (accession numbers KU831550-KU833209 and KU863006-KU863021) and matrices and phylogenetic trees from TreeBase (study accession URL: http://purl.org/phylo/treebase/phylows/study/TB2:S18291).

Authors' contributions
DUB, MDP & EGHO planned and designed the research; EGHO, MDP, BG and DUB conducted fieldwork; EGHO identified specimens; NLM, AMK, & MDP: performed lab work; MDP & MK: analysed data; and MDP led the writing (to which all authors contributed). All authors read and approved the final manuscript.

Authors' information
MDP & BG are research associates/lecturers, and MK received her PhD (2015), at the Johannes Gutenberg-Universität, Mainz, Germany; MDP and EGHO are honorary research Associates, NLM PhD student, and DUB full professor, at the University of Stellenbosch, South Africa; and AMK received her PhD (2013) at the University of Bonn, Germany.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Author details
[1]Department of Biochemistry, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa. [2]Institut für Spezielle Botanik und Botanischer Garten, Johannes Gutenberg-Universität, Anselm-Franz-von-Bentzelweg 9a, 55099 Mainz, Germany. [3]Department of Botany and Zoology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa. [4]INRES Pflanzenzüchtung, Rheinische Friedrich-Wilhelms-Universität Bonn, Katzenburgweg 5, 53115 Bonn, Germany.

Received: 7 May 2016 Accepted: 8 September 2016
Published online: 17 September 2016

References

1. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J. Biodiversity hotspots for conservation priorities. Nature. 2000;403(6772):853–8.
2. Hughes C, Eastwood R. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. Proc Natl Acad Sci U S A. 2006;103(27):10334–9.
3. Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM. Rapid diversification of a species-rich genus of Neotropical rainforest trees. Science. 2001;293(5538):2242–5.
4. Valente LM, Savolainen V, Vargas P. Unparalleled rates of species diversification in Europe. Proc R Soc B. 2010;277(1687):1489–96.
5. Klak C, Reeves G, Hedderson TA. Unmatched tempo of evolution in Southern African semi-desert ice plants. Nature. 2004;427(6969):63–5.
6. Goldblatt P, Manning JC. Plant diversity of the Cape region of southern Africa. Ann Missouri Bot Gard. 2002;89(2):281–302.
7. Linder HP. The radiation of the Cape flora, southern Africa. Biol Rev. 2003; 78(4):597–638.
8. Oliver EGH: Ericaceae. In: Plants of the Greater Cape Floristic Region, Volume 1: The Core Cape Flora. Edited by Manning JC, Goldblatt P, vol. 29. Pretoria: South African National Biodiversity Institute (SANBI Publishing); 2012: 482—511.
9. Hedberg O. Vegetation belts of the East African mountains. Svensk Bot Tidsk. 1951;45:140–202.
10. Pirie MD, Oliver EGH, Bellstedt DU. A densely sampled ITS phylogeny of the Cape flagship genus *Erica* L. suggests numerous shifts in floral macro-morphology. Mol Phylogenet Evol. 2011;61(2):593–601.
11. Van der Niet T, Pirie MD, Shuttleworth A, Johnson SD, Midgley JJ. Do pollinator distributions underlie the evolution of pollination ecotypes in the Cape shrub *Erica plukenetii*? Ann Bot. 2014;113(2):301–16.
12. Mugrabi de Kuppler AL, Fagúndez J, Bellstedt DU, Oliver EGH, Léon J, Pirie MD. Testing reticulate versus coalescent origins of *Erica lusitanica* using a species phylogeny of the northern heathers (Ericeae, Ericaceae). Mol Phylogenet Evol. 2015;88:121–31.
13. Bellstedt DU, Pirie MD, Visser JC, de Villiers MJ, Gehrke B. A rapid and inexpensive method for the direct PCR amplification of DNA from plants. Amer J Bot. 2010;97(7):e65–68.
14. Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. Pl Mol Biol. 1991;17(5):1105–9.
15. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Amer J Bot. 2007;94(3):275–88.

16. Wiens JJ, Fetzner JW, Parkinson CL, Reeder TW. Hylid frog phylogeny and sampling strategies for speciose clades. Syst Biol. 2005;54(5):778–807.
17. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. Version 3.04 In.: Available from http://mesquiteproject.org; 2015.
18. Vaidya G, Lohman DJ, Meier R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 2011;27(2):171–80.
19. Lanfear R, Calcott B, Ho SYW, Guindon S. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molec Biol Evol. 2012;29(6):1695–701.
20. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22(21):2688–90.
21. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. Syst Biol. 2008;57(5):758–71.
22. Pirie MD. Phylogenies from concatenated data: Is the end nigh? Taxon. 2015;64(3):421–3.
23. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Molec Biol Evol. 2012;29(8):1969–73.
24. Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. Estimating divergence times in large molecular phylogenies. Proc Natl Acad Sci U S A. 2012;109(47):19333–8.
25. Schwery O, Onstein RE, Bouchenak-Khelladi Y, Xing Y, Carter RJ, Linder HP. As old as the mountains: the radiations of the Ericaceae. New Phytol. 2015; 207(2):355–67.
26. Popp M, Mirré V, Brochmann C. A single Mid-Pleistocene long-distance dispersal by a bird can explain the extreme bipolar disjunction in crowberries (Empetrum). Proc Natl Acad Sci USA 2011;108:6520–5.
27. Dupont LM, Linder HP, Rommerskirchen F, Schefuß E. Climate-driven rampant speciation of the Cape flora. J Biogeogr. 2011;38(6):1059–68.
28. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v1.6, 2014. Available from http://beast.bio.ed.ac.uk/Tracer.
29. Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics. 2008;24(4):581–3.
30. Magallón S, Sanderson MJ. Absolute diversification rates in angiosperm clades. Evolution. 2001;55(9):1762–80.
31. Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. GEIGER: investigating evolutionary radiations. Bioinformatics. 2008;24(1):129–31.
32. Linder HP. The evolution of African plant diversity. Front. Ecol. Evol. 2014;2: 38. doi:10.3389/fevo.2014.00038.
33. Linder HP, Antonelli A, Humphreys AM, Pirie MD, Wüest RO. What determines biogeographical ranges? Historical wanderings and ecological constraints in the danthonioid grasses. J Biogeogr. 2013;40(5):821–34.
34. Valente LM, Reeves G, Schnitzler J, Mason IP, Fay MF, Rebelo TG, Chase MW, Barraclough TG. Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. Evolution. 2010;64(3):745–60.
35. FitzJohn RG. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol Evol. 2012;3(6):1084–92.
36. Rabosky DL. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS ONE. 2014;9(2):e89543.
37. Rabosky DL, Grundler M, Anderson C, Shi JJ, Brown JW, Huang H, Larson JG. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods Ecol Evol. 2014;5(7):701–7.
38. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. R News. 2006;6(1):7–11.
39. Verboom GA, Archibald JK, Bakker FT, Bellstedt DU, Conrad F, Dreyer LL, Forest F, Galley C, Goldblatt P, Henning JF, et al. Origin and diversification of the Greater Cape flora: ancient species repository, hot-bed of recent radiation, or both? Mol Phylogenet Evol. 2009;51(1):44–53.
40. Valente LM, Vargas P. Contrasting evolutionary hypotheses between two mediterranean-climate floristic hotspots: the Cape of southern Africa and the Mediterranean Basin. J Biogeogr. 2013;40(11):2032–46.
41. Rabosky DL. Extinction rates should not be estimated from molecular phylogenies. Evolution. 2010;64(6):1816–24.
42. Davis M, Midford P, Maddison W. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. BMC Evol Biol. 2013;13(1):38.
43. Hughes CE, Atchison GW. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. New Phytol. 2015;207(2):275–82.
44. Zachos J, Pagani M, Sloan L, Thomas E, Billups K. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science. 2001;292:686–93.
45. Marlow JR, Lange CB, Wefer G, Rosell-Mele A. Upwelling intensification as part of the Pliocene-Pleistocene climate transition. Science. 2000;290(5500):2288.
46. Bytebier B, Antonelli A, Bellstedt DU, Linder HP. Estimating the age of fire in the Cape flora of South Africa from an orchid phylogeny. Proc R Soc B. 2011;278:188–95.
47. Coetzee JA, Rogers J. Palynological and lithological evidence for the miocene palaeoenvironment in the Saldanha region (South Africa). Palaeogeog Palaeoclim Palaeoecol. 1982;39(1):71–85.
48. Van Wilgen BW, Higgins KB, Bellstedt DU. The role of vegetation structure and fuel chemistry in excluding fire from forest patches in the fire-prone fynbos shrublands of South Africa. J Ecol 1990;78:210–222.
49. Brown N, Kotze G, Botha P. The promotion of seed germination of Cape *Erica* species by plant-derived smoke. Seed Sci Technol. 1993;21(3):573–80.
50. Ojeda F. Biogeography of seeder and resprouter Erica species in the Cape Floristic Region—Where are the resprouters? Biol J Linn Soc. 1998;63(3):331–47.
51. Cowling RM, Proches S, Partridge TC. Explaining the uniqueness of the Cape flora: incorporating geomorphic evolution as a factor for explaining its diversification. Mol Phylogenet Evol. 2009;51(1):64–74.
52. Hoffmann V, Verboom GA, Cotterill FPD. Dated plant phylogenies resolve neogene climate and landscape evolution in the Cape Floristic Region. PLoS ONE. 2015;10(9):e0137847.
53. Verboom GA, Bergh NG, Haiden SA, Hoffmann V, Britton MN. Topography as a driver of diversification in the Cape Floristic Region of South Africa. New Phytol. 2015;207(2):368–76.
54. Goldblatt P. An analysis of the flora of Southern Africa: its characteristics, relationships, and origins. Ann Missouri Bot Gard. 1978;65:369–436.
55. Goldblatt P. Floristic diversity in the Cape Flora of South Africa. Biodiv Cons. 1997;6(3):359–77.
56. Jansson R, Dynesius M. The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. Annu Rev Ecol Syst. 2002;33:741–77.
57. Ellis AG, Verboom GA, Van der Niet T, Johnson SD, Linder HP. Speciation and extinction in the Greater Cape Floristic Region. In: Allsopp N, Colville JF, editors. Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region. Verboom GA: Oxford University Press; 2014. p. 119–41.
58. Johnson SD. The pollination niche and its role in the diversification and maintenance of the southern African flora. Philos Trans Roy Soc Ser B. 2010; 365(1539):499–516.
59. Van der Niet T, Johnson SD, Linder HP. Macroevolutionary data suggest a role for reinforcement in pollination system shifts. Evolution. 2006;60(8): 1596–601.
60. Van der Niet T, Johnson SD. Patterns of plant speciation in the Cape floristic region. Mol Phylogenet Evol. 2009;51(1):85–93.
61. Warren B, Bakker F, Bellstedt D, Bytebier B, Claszen-Bockhoff R, Dreyer L, Edwards D, Forest F, Galley C, Hardy C, et al. Consistent phenological shifts in the making of a biodiversity hotspot: the Cape flora. BMC Evol Biol. 2011; 11(1):39.
62. Schnitzler J, Barraclough TG, Boatwright JS, Goldblatt P, Manning JC, Powell MP, Rebelo T, Savolainen V. Causes of plant diversification in the Cape biodiversity hotspot of South Africa. Syst Biol. 2011;60(3):343–57.
63. Cowling RM, Potts AJ, Bradshaw PL, Colville J, Arianoutsou M, Ferrier S, Forest F, Fyllas NM, Hopper SD, Ojeda F, et al. Variation in plant diversity in mediterranean-climate ecosystems: the role of climatic and topographical stability. J Biogeogr. 2015;42(3):552–64.
64. Linder HP, Hardy CR. Evolution of the species-rich Cape flora. Philos Trans Roy Soc Ser B. 2004;359(1450):1623–32.
65. Richardson JE, Weitz FM, Fay MF, Cronk QCB, Linder HP, Reeves G, Chase MW. Rapid and recent origin of species richness in the Cape flora of South Africa. Nature. 2001;412:181–3.
66. Mummenhoff K, Al-Shehbaz IA, Bakker FT, Linder HP, Mühlhausen A. Phylogeny, morphological evolution, and speciation of endemic Brassicaceae genera in the Cape Flora of South Africa. Ann Missouri Bot Gard. 2005;92(3):400–24.
67. Linder HP. Plant species radiations: where, when, why? Philos Trans Roy Soc Ser B. 2008;363(1506):3097–105.

Appendix B: *Erica* species sequenced

Accessions on a green background were sequenced by the author and appear in the combined tree. Accessions on a yellow background were sequenced by the author and do not appear in the combined tree due to conflicts resulting from differential positioning between the ITS and chloroplast marker trees (only 15 species), the presence of duplicates of that accession, or incomplete sequence data for that accession. Accessions on a red background were sequenced by other members of our group and appear in the combined tree. Genbank accession numbers are given where sequences have been submitted.

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
Daboecia cantabrica PPA	HQ859000	KP737380	KP737653	KP737744	KP737633		KP737495	KP737722	KP737708
E. plumigera_EO11341	HQ859214	KU832970	KU832176						
E.abietina dia CM11	KU832331	KU832579	KU831804						
E.accommodata EO11382	HQ858891	KU832580	KU831805						
E.adunca EO12746	KU832332	KU832583	KU831808						
E.agglutinans EO7679	HQ858896	KU832585	KU831810						
E.albens MP585	HQ858897	KU832587	KU831811						
E.albertyniae MP927	HQ858898	KU832588	KU831812						
E.ampullacea MP1277	KU832337	KU832596	KU831820						
E.anguliger MP597	HQ858907	KU832598_KU832599	KU831822						
E.arachnocalyx EO12453	HQ858909	KU832603	KU831825						
E.arborea EO12619	KP737520	KP737385	KP737655						
E.ardens MP1076	KU832342	KU832606	KU831828						
E.aristata EO12830	HQ858917	KU832609	KU831831						
E.axillaris MP1052	KU832345	KU832617	KU831839						
E.baroniana cf Lar8	KU832347	KU832625	KU831846						
E.baueri bau MP1233	KU832348	KU832626	KU831847						
E.benguelensis A1	KU832349	KU832627	KU831848	KU831615	KU833135	KU831724	KU831757	KU831553	KU831739
E.benguelensis be DB1174	HQ858937	KU832628	KU831849						
E.bicolor MP1098	KU832350	KU832633	KU831853						
E.blandfordii MM4208	HQ858940	KU832634	KU831854						
E.bojeri MP1087	KP737538	KP737404	KP737659	KP737760	KP737638	KP737689	KP737500	KP737727	KP737712
E.bokkeveldia EO12769	KU832352	KU832638	KU831857						
E.bolusiae var. Cyanthiformis ANA212	KU832354	KU832640	KU831859	KU831617					
E.brachysepala EO12727	KU832356	KU832643	KU831862						
E.caldonica JW103	KU832359	KU832654	KU831873	KU831623					
E.calycina EO12532	HQ858958	KU832655	KU831874						
E.cameronii ANA215	KU832360	KU832656	KU831875	KU831624					
E.canescens ANA217	KU832361	KU832658	KU831877	KU831625					
E.caprina EO12772	KU832366	KU832663	KU831881						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.chamissonis RT2188	HQ858970	KU832669	KU831887						
E.chrysocodon ANA222	KU832368	KU832674	KU831892	KU831628					
E.condensa MP890	HQ858982	KU832682	KU831898						
E.conspicua CS1	HQ858984	KU832684	KU831900						
E.cooperi EO12588	KU832374	KU832685	KU831901						
E.corifolia MP1261	HQ858988	KU832690	KU831905	KU831633	KU833143	KU831728	KU831764	KU831563	KU831743
E.corydalis AH2536	KU832377	KU832691	KU863017						
E.cristata MP820	HQ858989	KU832692	KU831906	KU831634	KU833144	KU831729	KU831765	KU831564	KU831744
E.cyathiformis a ANA	KU832383	KU832702	KU831917	KU831637					
E.cyathiformis d ANA	KU832384	KU832703	KU831918	KU831638					
E.deflexa MP1247	KU832387	KU832707	KU831922						
E.diaphana BG611	HQ859012	KU832715	KU831930						
E.dissimulans EO12596	KU832392	KU832719	KU831934	KU831639					
E.dolfiana MP1297	KU832393	KU832722	KU831937						
E.drakensbergensis DB1443	KU832394	KU832725	KU831940	KU831642_KU831643					
E.duthieae ANA241	KU832396	KU832727	KU831942	KU831644					
E.elimensis EO12843	KU832399	KU832730							
E.esteriana EO12848	HQ859032	KU832740	KU831954		KU833151		KU831769	KU831569	
E.eustacei MP1259	KU832402	KU832742	KU831956						
E.fairii_ANA	KU832403	KU832746		KU831647					
E.fausta MP663	HQ859040	KU832750	KU831963						
E.filamentosa EO12728	KU832405	KU832753	KU831966						
E.fillipendula min MP926	HQ859046	KU832755	KU831968						
E.florifera EO12536	KU832406	KU832759	KU831972						
E.fontana MP1069	KU832407	KU832760	KU831973						
E.garciae MP1253	KU832409	KU832763	KU831976						
E.georgica ANA	KU832410	KU832764	KU831977	KU831650					
E.gibbosa GS23	KU832414	KU832769	KU831980						
E.glabella la EO11224	HQ859059	KU832770	KU831981						
E.goudotiana cf DB1227	KU832421	KU832786	KU831997		KU833158			KU831572	

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.goudotiana EO12636	KU832422	KU832787	KU831998	KU831654_KU831655	KU833157				
E.hanekomii EO11172	HQ859079	KU832797	KU832009						
E.hibbertii MP982	KU832432	KU832806	KU832017						
E.holoserica EO12842	KU832434	KU832811	KU832022	KU831659					
E.holosericea MP1004	KU832435	KU832812	KU832023						
E.hottentotica EO11971	HQ859088	KU832814	KU832025						
E.incarnata EO12771	KU832438	KU832821	KU832032						
E.inflata MP784	HQ859095	KU832823	KU832033	KU831662	KU833165	KU831730	KU831775	KU831576	KU831745
E.inordinata EO11823	HQ859098	KU832828	KU832038		KU833166		KU831777	KU831578	
E.karooica MP1285	KU832448	KU832839	KU832049						
E.karwyderi EO12718	KU832449	KU832840	KU832050						
E.krugeri EO12807	KU832450	KU832842_KU832843	KU832052						
E.lachnaeifolia MP994	KU832451	KU832845	KU832054						
E.lasciva MP906	HQ859113	KU832850	KU832059	KU831667					
E.lateriflora EO12482	HQ859115	KU832853	KU832062						
E.leonis RTsn	HQ859118	KU832856	KU832065						
E.leucopelta EO12598	KU832459	KU832862	KU832070						
E.leucosiphon PW121110	HQ859125	KU832864	KU832072						
E.lignosa EO11763	HQ859127	KU832866	KU832073						
E.lithophila MP1301	KU832462	KU832867	KU832074						
E.loganii MP1258	KU832463	KU832868	KU832075						
E.lusitanica FV	KP737567	KP737433		KP737780					
E.lutea MP672	KP737564	KP737430	KP737667						
E.maderensis AH	KP737575	KP737440	KP737669						
E.magistrati EO11750	HQ859139	KU832876	KU832082						
E.mammosaP MP951	HQ859141	KU832879	KU832085						
E.mammosaW EO12784	HQ859142	KU832880	KU832086						
E.margaritaceae ANA	KU832468	KU832882_KU832883	KU832088						
E.marifolia CM4	KU832469	KU832884	KU832089						
E.mauritiensis CB2632	KU832470	KU832886	KU832091						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.mauritiensis EO12811	KU832471	KU832887	KU832092	KU831670_KU831671					
E.melastoma min EO12820	KU832472	KU832891	KU832096	KU831672					
E.monsoniana MP986	KU832479	KU832898	KU832103		KU833173				
E.nabea BG609	KU832481	KU832904	KU832109	KU831676	KU833177	KU831734	KU831787	KU831588	KU831749
E.nana ANA280	KU832482	KU832905	KU832110	KU831677					
E.oatesii SB9774	KU832486	KU832914	KU832119						
E.oblongiflora ANA283	KU832487	KU832915	KU832120	KU831678					
E.oliveri MP1278	KU832488	KU832921	KU832126						
E.oreotragus ANA284	KU832489	KU832922	KU832127	KU831679					
E.orientalis EO12608	KU832490	KU832924	KU832129						
E.oxycoccifolia MP1275	KU832491	KU832926	KU832131						
E.pageana ANA	KU832492	KU832928	KU832133	KU831680					
E.paniculata MP1274	KU832493	KU832930	KU832135						
E.passerinae MP1302	KU832495	KU832936	KU832141						
E.pauciovulata TdV258b	KU832497	KU832941	KU832145	KU831684					
E.perlata MP960	HQ859191	KU832946	KU832150						
E.perplexa EO12788	KU832499	KU832947	KU832151						
E.petrophila EO7592	HQ859195	KU832952	KU832156						
E.petrophila RT2065	HQ859196	KU832953	KU832157						
E.phacelanthera EO12489	HQ859198	KU832955	KU832159						
E.phaeocarpa SM2003	HQ859058	KU832956	KU832160						
E.plukenetii plu EOns	KF160932	KF160944	KU832175		KP737646		KP737507	KP737734	
E.polycoma FR	KU832509	KU832974	KU832180						
E.psittacina IJ1237	KU832511	KU832979	KU832185	KU831690_KU831691	KU833183	KU831736	KU831791	KU831592	KU831751
E.pubescens EO12503	HQ859223	KU832980	KU832186						
E.pulchella MP736	HQ859225	KU832982	KU832188						
E.pulvinata MP1304	KU832512	KU832983	KU832189						
E.pyxidiflora ANA295	KU832514	KU832985	KU832191	KU831692					
E.recta ANA298	KU832517	KU832990	KU832196	KU831695					
E.remota MP1263	HQ859240	KU832998	KU832204						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.rhodopis BAB13	HQ859243	KU833002	KU832207						
E.rhopalantha MP909	HQ859244	KU833003	KU832208						
E.rubiginosa RT1554	HQ859252	KF160958	KU832214						
E.scoparia EO12618	KP737594	KP737460	KP737675						
E.senilis EO12483	HQ859262	KU863014	KU832223						
E.setacea MP589	HQ859266	KU833019	KU832226						
E.silvatica CJ 11532	HQ859268	KU833022	KU832229						
E.simii SB9385	HQ859270	KU833025	KU832232						
E.simulans BG603	HQ859272	KU833028	KU832235						
E.sp mad3 EO12632	KU832528	KU833033	KU832240	KU831701_KU831702	KU833190				
E.sp nov 1 MP1291	KU832532	KU833037	KU832244						
E.spiculifolia AS57234	KP737610	KP737475	KP737678	KP737814	KP737649	KP737699	KP737510	KP737737	KP737717
E.steinbergiana EO12763	KU832536	KU833048	KU832254						
E.stokoeanthus EO4790	HQ859284	KU833049	KU832255						
E.stylaris ANA311	KU832538	KU833055	KU832260	KU831709					
E.swaziensis L1187	KU832541	KU833060	KU832265		KU833197				
E.thimifolia CM7	KU832545	KU833066	KU832270						
E.thunbergii AH2612	KU832547	KU833070_KU833071	KU832274						
E.thunbergii EO12473	HQ859298	KU833072	KU832275						
E.toringbergensis ANA315	KU832548	KU833074	KU832277	KU831713					
E.trichroma EO12517	HQ859305	KU833080	KU832283						
E.trimera MsnA	KP737625	KP737487	KP737681	KP737826	KP737651	KP737702	KP737512	KP737739	KP737719
E.tysonii EO12583	KU832550	KU833086	KU832289		KU833201			KU831602	
E.umbellata ANA061	KP737626	KP737488	KU832291	KP737827		KP737703			
E.umbelliflora RT2182	HQ859156	KU833089	KU832293						
E.unicolor geo JP273	KU832553	KU833091	KU832295		KU833203				
E.vallis-aranearum ANA323	KU832559	KU833100	KU832303	KU831717					
E.vallis-fluminis EO12761	KU832560	KU833101	KU832304						
E.velatiflora MP1213	KU863009	KU833103	KU832306						
E.verecunda CS5	HQ859325	KU833106	KU832308	KU831719					

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.verticillata ANA	KU832566	KU833110	KU832312						
E.whyteana A4	KU832572	KU833121	KU832323	KU831721	KU833208	KU831738	KU831801	KU831607	KU831753
E.alticola EO12511			trnT-L-intron						
E.amphigena MP1260	ITS	trnL-F-ndhJ	trnT-L-intron						
E.ampullacea EO11396		trnL-F-ndhJ	trnT-L-intron						
E.ampullacea MP1278		trnL-F							
E.ampullacea MP1279		ndhJ							
E.ampullacea MP1280		ndhJ							
E.arborea VC561			trnT-L						
E.azorica AV			trnT-L						
E.bakeri EO12828	ITS	trnL-F-ndhJ	trnT-L-intron						
E.betsileana DB1310			trnT-L						
E.blancheana ANA209		trnL-F-ndhJ	trnT-L						
E.bodkinii MP1281	ITS		trnT-L-intron						
E.brachyphylla EO12809	ITS		trnT-L-intron						
E.brachyphylla EO12810	ITS	trnL-F-ndhJ							
E.calcicola EO12814	ITS	trnL-F-ndhJ							
E.campanularis EO11352			trnT-L-intron						
E.cf vogelpoelii EO12815	ITS	trnL-F-ndhJ	trnT-L-intron						
E.chonantha cf EO12822		trnL-F-ndhJ	trnT-L-intron						
E.chrysocodon EO12845		trnL-F-ndhJ	trnT-L-intron						
E.ciliaris PPC		trnL-F-ndhJ	trnT-L						
E.cinerea Popp	ITS	trnL-F-ndhJ							
E.coccinea ssp. uniflora EO12849	ITS	trnL-F-ndhJ	intron						
E.coventryi EO12832	ITS		trnT-L-intron						
E.cristiflora EO11174		trnL-F-ndhJ	trnT-L						
E.dracomontana EO12601		trnL-F-ndhJ	trnT-L-intron						
E.elimensis ANA099		trnL-F-ndhJ	trnT-L						
E.eremioides ssp. Pubescens MP1271		trnL-F-ndhJ	trnT-L-intron						
E.flavicoma EO12823	ITS	trnL-F-ndhJ	trnT-L-intron						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.flexistyla EO12818	ITS		trnT-L-intron						
E.gigantea EO12821	ITS		trnT-L-intron						
E.glabella EO12817	ITS	trnL-F-ndhJ							
E.hexandra VC528			trnT-L						
E.hirta EO12812	ITS	trnL-F-ndhJ	trnT-L-intron						
E.holosericea EO12841	ITS	trnL-F-ndhJ							
E.insignis MP1290		trnL-F-ndhJ	trnT-L-intron						
E.insolitanthera ANA266		trnL-F-ndhJ	trnT-L						
E.ixanthera JP215		trnL-F-ndhJ	intron						
E.jacksoniana EO12791	ITS	trnL-F-ndhJ	trnT-L-intron						
E.klotzschii EO12762		trnL-F-ndhJ	trnT-L						
E.leucopelta SB9392			trnT-L						
E.lowryensis ANA270		trnL-F-ndhJ	trnT-L						
E.manipuliflora KK		trnL-F-ndhJ	trnT-L						
E.modesta ANA276		trnL-F-ndhJ	trnT-L						
E.niveniana MP1267	ITS	trnL-F-ndhJ							
E.obliqua RT2079		trnL-F-ndhJ	trnT-L-intron						
E.oblongiflora EO12813	ITS		trnT-L-intron						
E.oligantha MP1283	ITS	trnL-F-ndhJ	trnT-L-intron						
E.ostiara ANA285		trnL-F-ndhJ	trnT-L						
E.ovina MP1265			trnT-L						
E.oxycoccifolia CM14	ITS		trnT-L-intron						
E.paludicola AH2680	ITS	trnL-F-ndhJ	intron						
E.parvula ANA287		trnL-F-ndhJ							
E.petraea BG627		trnL-F-ndhJ	trnT-L						
E.petraea MP607		trnL-F-ndhJ	trnT-L						
E.pilosiflora pil MP1262	ITS		trnT-L						
E.pilosiflora pil RT2058		trnL-F-ndhJ	trnT-L-intron						
E.pilulifera EO12816	ITS	trnL-F-ndhJ							
E.plena EO12824	ITS		trnT-L						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.plena MP1282		trnL-F-ndhJ	trnT-L-intron						
E.plukenetii plu SH	ITS	trnL-F-ndhJ	intron						
E.polifolia MP575		ndhJ							
E.propinqua MP1019		trnL-F-ndhJ	trnT-L-intron						
E.puberuliflora MP1266	ITS		trnT-L-intron						
E.pudens JM6		trnL-F-ndhJ	trnT-L-intron						
E.purgatoriensis EO12846	ITS	trnL-F-ndhJ							
E.quadrangularis MP1264			trnT-L						
E.recta MP1288		trnL-F-ndhJ	trnT-L-intron						
E.retorta BAB1			trnT-L-intron						
E.retorta RT2096		ndhJ							
E.ribisaria RT2059		trnL-F-ndhJ	trnT-L						
E.sagittata EO12737		trnL-F-ndhJ	intron						
E.schlechteri EO12591		trnL-F-ndhJ	trnT-L						
E.seriphiifolia MP1223			trnT-L						
E.silvatica VC582			trnT-L						
E.sp. EO12847	ITS	trnL-F-ndhJ	trnT-L-intron						
E.sperata MP931			intron						
E.spinifera EO12640		trnL-F-ndhJ							
E.taxifolia MP1268	ITS		trnT-L-intron						
E.taxifolia MP692	ITS	trnL-F-ndhJ		matK					
E.tenella EO12839	ITS	trnL-F-ndhJ							
E.tenuifolia MP1006			trnT-L						
E.terminalis KK		trnL-F-ndhJ	trnT-L						
E.tetralix Popp		trnL-F-ndhJ	trnT-L-intron						
E.tragulifera BG617		trnL-F-ndhJ							
E.turneri EO12736	ITS								
E.valida EO11826		trnL-F-ndhJ	intron						
E.vallis-aranearum MJ		trnL-F-ndhJ	trnT-L						
E.venustiflora ssp. gladulosa MP1284	ITS		trnT-L-intron						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.villosa EO11396		ndhJ	intron						
E.walkeriana ANA140		trnL-F-ndhJ	trnT-L						
E.winteri ANA333		trnL-F-ndhJ	trnT-L						
E.xeranthemifolia MM5080		trnL-F-ndhJ	trnT-L-intron						
Calluna vulgaris ANA	KP737514	KP737377		KP737741					
Calluna vulgaris BG643	HQ858882	KP737378		KP737742	KP737632	KP737685		KP737721	KP737707
Daboecia cantabrica ANA	KU832375	KU832688		KU831631					
Daboecia_cantabrica_ANA	KP737515	KP737379		KP737743					
E. acuta_MP506	HQ858892	KU832581	KU831806						
E. adnata_MP501	HQ858893	KU832582	KU831807						
E. albens_ANA	KU832334	KU832586		KU831608					
E. albescens_MP898	HQ858899	KU832589	KU831813						
E. angulosa_S2105	HQ858908	KU832600	KU831823						
E. areolata_EO12502	HQ858914	KU832607	KU831829						
E. artemisioides_MP551	HQ858918	KU832610	KU831832						
E. articularis_MP750_cp	HQ858919; HQ858920								
E. australis_b_ANA	KP737534	KP737401							
E. australis_PPF	HQ858926	KP737398	KP737657		KP737636	KP737688	KP737498	KP737725	KP737711
E. australis_PPM	HQ858927	KP737399	KP737658		KP737637		KP737499	KP737726	
E. axillaris_MP594	HQ858930	KU832618	KU831840		KU833133		KU831755	KU831551	
E. bodkinii_TdV204	HQ858942	KU832637	KU831856						
E. caespitosa_MP642	HQ858952	KU832649	KU831868		KU833138		KU831760	KU831556	
E. chionodes_EO11699	HQ858972	KU832671	KU831889					KU831561	
E. chiroptera_MP814	HQ858974	KU832673	KU831891						
E. cinerea_a_ANA	KP737551	KP737417		KP737770		KP737693			
E. cinerea_MP967	KP737549	KP737415	KP737663	KP737769	KP737641	KP737692	KP737503	KP737729	KP737714
E. conferta_MP887	HQ858983	KU832683	KU831899						
E. dianthifolia_MP583	HQ859011	KU832714	KU831929						
E. equisetifolia_ANA	KU832401	KU832732	KU831946						
E. equisetifolia_MP829	HQ859023	KU832733	KU831947		KU833149		KU831767	KU831567	

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E. eugenea_EO12485	HQ859033	KU832741	KU831955						
E. fairii_CM12	KU832404	KU832747	KU831960						
E. floccifera_MP987	HQ859051	KU832758	KU831971						
E. glandulipila_MP521	HQ859061	KU832771	KU831982						
E. globiceps_gra_MP988	HQ859067	KU832777	KU831988						
E. goatcheriana_goa_MP865	HQ859072	KU832784	KU831995						
E. grisbrookii_EO12716	KU832423	KU832791	KU832003						
E. hermani_EO12498	HQ859082	KU832804	KU832015		KU833160		KU831772	KU831573	
E. hispiduloides_EO11544	HQ859087	KU832810	KU832021						
E. intervallaris_MP556	HQ859102	KU832831	KU832041						
E. jonasiana_MP985	KU832447	KU832836	KU832046						
E. lehmanni_MP625	HQ859117	KU832855	KU832064						
E. lepidota_MP541	HQ859119	KU832857	KU832066						
E. leptopus_ANA	KU832458	KU832858		KU831668					
E. leucanthera_EO12452	HQ859121	KU832860	KU832068						
E. leucodesmia_MP724	HQ859122	KU832861	KU832069						
E. lucida_MP690	HQ859129	KU832870	KU832077						
E. maderensis_b_ANA	KP737577	KP737442		KP737788					
E. magnisylvae_EO10708	KU832465	KU832877	KU832083						
E. mammosaW_EO12784	KU832467	KU832881	KU832087						
E. maximilianii_EO12484	HQ859144	KU832888	KU832093						
E. mollis_CM5	KU832476	KU832895	KU832100						
E. multiflexuosa_EO12445	HQ859150	KU832900	KU832105						
E. muscosa_MP775	HQ859154	KU832902	KU832107		KU833175		KU831785	KU831586	
E. nemerosa_MP1208	KU832484	KU832908	KU832113						
E. nervata_EO12541	HQ859158	KU832909	KU832114						
E. nudiflora_MP802	HQ859160	KU832912	KU832117						
E. ovina_EO12487	HQ859171	KU832925	KU832130						
E. pannos_EO12490	HQ859175	KU832931	KU832136						
E. parviporandra_MP877	HQ859179	KU832935	KU832140		KU833179		KU831788	KU831589	

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E. patens_EO12457	HQ859182	KU832937	KU832142						
E. paucifolia_cil_EO12528	HQ859184	KU832940	KU832144						
E. peltata_MP1231	KU832498	KU832943	KU832147						
E. petricola_MP996	HQ859194	KU832951	KU832155		KU833182		KU831790	KU831591	
E. peziza_MP719	HQ859197	KU832954	KU832158						
E. physophylla_EO11418	HQ859201	KU832961	KU832164						
E. podophylla_MP582	HQ859217	KU832972	KU832178						
E. recurvata_EO12467	HQ859230	KU832991	KU832197						
E. rugata_EO12516	HQ859253	KU833009	KU832215						
E. saxicola_EO12515	HQ859257	KU833014	KU832220						
E. selaginifolia_EO12488	HQ859261	KU863013	KU832222						
E. sicula_boc_DM	KP737598	KP737463	KP737676	KP737803	KP737648	KP737697	KP737509	KP737736	KP737716
E. sicula_sic_AM	KP737599	KP737464	KP737677						
E. sicula_sic_c_ANA	KP737602	KP737467		KP737806					
E. sparsa_BG602	HQ859277	KU833041	KU832247		KU833193			KU831597	
E. strigosa_MP673	HQ859288	KU833054	KU832259		KU833195		KU831795	KU831598	
E. subcapitata_MP1042	KU832539	KU833056	KU832261						
E. tenuis_ANA	KU832542	KU833062		KU831711					
E. tenuis_EO12491	HQ859294	KU833063	KU832267		KU833198		KU831796	KU831599	
E. tomentosa_MP961	HQ859299	KU833073	KU832276						
E. triflora_MP564	HQ859306	KU833081	KU832284						
E. umbellata_DS	KU832551	KU833088	KU832292	KU831714	KU833202	KU831737	KU831798	KU831603	KU831752
E. woodii_RC513	KU832574	KU833125	KU832326						
E. zebrensis_EO12787	KU832575	KU833126	KU832327						
E.abietina abi MP1013	HQ858885	KU832577	KU831802		KU833129		KU831754	KU831550	
E.abietina con CM6	KU832330	KU832578	KU831803						
E.adunca MP1254	KU832333	KU832584	KU831809						
E.alexandri EO12449	HQ858900	KU832590	KU831814						
E.alfredii FR	KU832335	KU832591	KU831815						
E.algida MP645	HQ858901	KU832592	KU831816		KU833130				

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.alopecurus MP630	HQ858902	KU832593	KU831817	KU831609	KU833131				
E.amidae EO12272	HQ858905	KU832594	KU831818						
E.amoena MP1032	KU832336	KU832595	KU831819						
E.aneimena EO12757	KU832338	KU832597	KU831821						
E.annectens ANA	KU832341	KU832601		KU831610					
E.annectens CM8	KU863006	KU832602	KU831824						
E.arborea l ANA	KP737529	KP737394		KP737756					
E.arborea m ANA	KP737530	KP737395		KP737757					
E.arborea Miehe	KP737521	KP737386	KP737656	KP737748	KP737635	KP737687	KP737497	KP737724	KP737710
E.arborescens FM747	KU863007	KU832604	KU831826	KU831611_KU831612	KU833132				
E.arcuata MP523	HQ858913	KU832605	KU831827						
E.argentea EO12475	HQ858915	KU832608	KU831830						
E.aspalathifolia DB1408	KU832343	KU832611	KU831833						
E.astroites EO12758	KU832344	KU832612	KU831834						
E.atherstonei EO12261	HQ858922	KU832613	KU831835	KU831613					
E.atromontana EO12544	HQ858924	KU832614	KU831836						
E.atrovinosa MP864	HQ858925	KU832615	KU831837						
E.autumnalis MP665	HQ858928	KU832616	KU831838						
E.axilliflora MP916	HQ858929	KU832619	KU831841						
E.azorica b ANA	KP737537	KP737403							
E.baccans BG645	HQ858932	KU832620	KU831842						
E.banksii com ANA	KU832346	KU832621	KU831843						
E.banksii com MP824	HQ858933	KU832622	KU831844	KU831614	KU833134		KU831756	KU831552	
E.banksii pur NB1753	HQ858934	KU832623	KU863015						
E.barbigeroides MP735	HQ858935	KU832624	KU831845						
E.benguelensis be QL12641	HQ858936	KU832629	KU831850						
E.benthamiana MP560	HQ858938	KU832630_KU832631	KU831851						
E.bergiana MP768	HQ858939	KU832632	KU831852						
E.blenna ANA	KU832351	KU832635		KU831616					
E.blenna MP883	HQ858941	KU832636	KU831855		KU833136		KU831758	KU831554	

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.bolusanthus DEB1720	KU832353	KU832639	KU831858						
E.boutonii EO12666	KU832355	KU832641	KU831860	KU831618_KU831619					
E.brachialis MP734	HQ858944	KU832642	KU831861						
E.bracteolaris MP577	HQ858945	KU832644	KU831863						
E.brevifolia EO12459	HQ858946	KU832645	KU831864						
E.brevifolia MP1008	KU832357	KU832646	KU831865						
E.bruniades EO12465	HQ858949	KU832647	KU831866						
E.bruniifolia EO12460	HQ858950	KU832648	KU831867		KU833137		KU831759	KU831555	
E.caffra ANA	KU832358	KU832650	KU831869	KU831620					
E.caffra C MP528	HQ858953	KU832651	KU831870	KU831621	KU833139	KU831725		KU831557	KU831740
E.caffra D MP655	HQ858954								
E.caffrorum MP644	HQ858956	KU832652	KU831871	KU831622		KU831726		KU831558	KU831741
E.calcareophila EO30159	HQ858957	KU832653	KU831872						
E.canaliculata BG590	HQ858962	KU832657	KU831876						
E.capensis MP1047	KU832362	KU832659	KU831878						
E.capillaris MP1066	KU832363	KU832660	KU831879						
E.capitata ANA	KU832364	KU832661	KU831880	KU831626					
E.capitata MP1044	KU832365	KU832662	KU863016						
E.carnea ATsn5	HQ858963	KP737405	KP737660	KP737761	KP737639	KP737690	KP737501	KP737728	KP737713
E.carnea b ANA	KP737540	KP737407		KP737763					
E.caterviflora EO12785	KU832367	KU832664	KU831882						
E.cereris MP863	HQ858964	KU832665	KU831883						
E.cerinthoides ANA	KP737541	KP737408		KP737764					
E.cerinthoides MP653	HQ858966	KU863010	KU831884	KU831627	KU833140	KU831727	KU831761	KU831559	KU831742
E.cernua EO12474	HQ858968	KU832667	KU831885						
E.cetrata EO12064	HQ858969	KU832668	KU831886		KU833141		KU831762	KU831560	
E.chartacea EO11408	HQ858971	KU832670	KU831888						
E.chionophila MP790	HQ858973	KU832672	KU831890						
E.ciliaris c ANA	KP737546	KP737413		KP737767					
E.ciliaris EO12622	KP737542	KP737409	KP737661						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.ciliaris KKN1	KP737543	KP737410	KP737662		KP737640		KP737502		
E.clavisepala ANA	KU832369	KU832675		KU831629					
E.clavisepala MP1063	KU832370	KU832676	KU831893		KU833142		KU831763	KU831562	
E.coacervata MP761	HQ858979	KU832677	KU831894						
E.coarctata ANA	KU832371	KU832678		KU831630					
E.coarctata MP590	HQ858980	KF160955							
E.coccinea coc MP598	HQ858981	KU832679	KU831895						
E.collina EO12613	KU832372	KU832680	KU831896						
E.colorans EO12717	KU832373	KU832681	KU831897						
E.copiosa BG610	HQ858986	KU832686	KU831902						
E.cordata MP571	HQ858987	KU832687	KU831903						
E.corifolia ANA	KU832376	KU832689	KU831904	KU831632					
E.cruenta MP745	HQ858991	KU832693	KU831907						
E.cryptoclada EO12684	KU832378	KU832694	KU831908	KU831635_KU831636	KU833145				
E.cubica MP623	HQ858992	KU832695	KU831909						
E.cumuliflora EO12699	KU832379	KU832696	KU831910						
E.curtophylla EO12750	KF160923	KF160956	KU831911						
E.curviflora dif EO12742	KU832380	KU832697	KU831912						
E.curviflora MP765	KU832381	KU832698	KU831913		KU833146		KU831766	KU831565	
E.curvifolia MP700	HQ858998	KU832699	KU831914						
E.curvirostris MP817	HQ858999	KU832700	KU831915						
E.curvirostris MP941	KU832382	KU832701	KU831916						
E.cylindrica MP1240	KU832385	KU832704	KU831919						
E.cyrilliflora CM10	KU832386	KU832705	KU831920						
E.daphniflora MP567	HQ859003	KU832706	KU831921						
E.demissa EO12540	HQ859005	KU832708	KU831923						
E.densifolia BG591	HQ859006	KU832709	KU831924		KU833147			KU831566	
E.denticulata MP799	HQ859007	KU832710	KU831925						
E.depressa MP942	HQ859009	KU832711	KU831926						
E.depressa MP977	KU832388	KU832712	KU831927						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.desmantha MP562	HQ859010	KU832713	KU831928						
E.discolor heb MP1214	KU832389	KU832716	KU831931						
E.discolor spe MP1217	KU832390	KU832717	KU831932						
E.dispar EO12749	KU832391	KU832718	KU831933						
E.distorta EO12500	HQ859014	KU832720	KU831935						
E.dodii EO11417	HQ859015	KU832721	KU831936						
E.doliiformis MP797	HQ859016	KU832723	KU831938						
E.dominans MP648	HQ859019	KU832724	KU831939	KU831640_KU831641	KU833148				
E.dregei EO12711	KU832395	KU832726	KU831941						
E.eburnea MP1037	KU832397	KU832728	KU831943						
E.ecklonii EO12739	KU832398	KU832729	KU831944						
E.empetrina EO12786	KU832400	KU832731	KU831945						
E.erasmia MP874	HQ859024	KU832734	KU831948						
E.eremioides MP533	HQ859025	KU832735	KU831949						
E.ericoides MP742	HQ859026	KU832736	KU831950						
E.erigena a ANA	KP737559	KP737425		KP737775		KP737694			
E.erigena EO12623	KP737557	KP737423	KP737665		KP737642			KP737730	
E.erigena Pim	KP737558	KP737424	KP737666						
E.erinus MP907	HQ859027	KU832737	KU831951						
E.eriophoros EO12478	HQ859030	KU832738	KU831952						
E.esteriana EO12545	HQ859031	KU832739	KU831953		KU833150		KU831768	KU831568	
E.evansii MP641	HQ859035	KU832743	KU831957	KU831645_KU831646	KU833152				
E.excavata GK1532	HQ859036	KU832744	KU831958						
E.exleeana EO12499	HQ859037	KU832745	KU831959						
E.fascicularis fac MP809	HQ859038	KU832748	KU831961						
E.fastigiata MP830	HQ859039	KU832749	KU831962						
E.ferrea EO12494	HQ859041	KU832751	KU831964						
E.filago BG93	HQ859043	KU832752	KU831965						
E.fillipendula fil MP914	HQ859044	KU832754	KU831967						
E.fimbriata MP606	HQ859047	KU832756	KU831969						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.flacca MP840	HQ859048	KU832757	KU831970						
E.frigida MP658	HQ859055	KU832761	KU831974		KU833153		KU831770	KU831570	
E.galioides FN745	KU832408	KU832762	KU831975	KU831648_KU831649	KU833154				
E.gerhardii EO12700	KU832411	KU832765	KU831978						
E.gibbosa ANA	KU832412	KU832766		KU831651					
E.gibbosa ANA b	KU832413	KU832767		KU831652					
E.gibbosa BG623	HQ859057	KU832768	KU831979						
E.glandulosa fou ANA	KU832415			KU831653					
E.glandulosa gla MP1227	KU832416	KU832772	KU831983						
E.glaphyra MP647	HQ859062	KU832773	KU831984		KU833155				
E.glauca ele MP973	HQ859063	KU832774	KU831985						
E.glauca gla MP850	HQ859064	KU832775	KU831986						
E.globiceps con EO12519	HQ859065	KU832776	KU831987						
E.glomiflora EO12548	HQ859068	KU832778	KU831989		KU833156		KU831771	KU831571	
E.glomiflora MP1251	KU832417	KU832779	KU831990						
E.glutinosa MP687	HQ859070	KU832780	KU831991						
E.glutinosa MP701	KU832418	KU832781	KU831992						
E.gnaphaloides MP511	HQ859071	KU832782	KU831993						
E.goatcheriana dra EO12694	KU832419	KU832783	KU831994						
E.goatcheriana pet MP990	KU832420	KU832785	KU831996						
E.gracilipes MM5014	HQ859073	KF160957	KU831999						
E.gracilis BG622	HQ859074	KU832788	KU832000						
E.grata MP879	HQ859076	KU832789	KU832001						
E.greyi EO12501	HQ859077	KU832790	KU832002						
E.gysbertii MP826	HQ859078	KU832792	KU832004						
E.haemastoma MP871	KU832424	KU832793	KU832005						
E.haematocodon MP1033	KU832425	KU832794	KU832006						
E.halicacaba ANA	KP737562	KP737428	KU832007	KP737778					
E.halicacaba MP1036	KU832426	KU832796	KU832008						
E.hansfordii MP1239	KU832427	KU832798	KU832010						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.hebeclada EO12686	KU832428	KU832799	KU832011	KU831656_KU831657	KU833159				
E.heleogena MP1064	KU832429	KU832800	KU863018						
E.heleophila ANA	KU832430	KU832801	KU832012	KU831658					
E.heleophila EO12539	HQ859080	KU832802	KU832013						
E.hendricksei EO12524	HQ859081	KU832803	KU832014						
E.hexandra RC465	KU832431	KU832805	KU832016						
E.hillburtii EO12593	KU832433	KU832807	KU832018		KU833161		KU831773	KU831574	
E.hirtiflora MP958	HQ859084	KU832808	KU832019						
E.hispidula MP801	HQ859086	KU832809	KU832020						
E.holtii TO	KU832436	KU832813	KU832024		KU833162				
E.humbertii EO12638	KU832437	KU832815	KU832026	KU831660_KU831661	KU833163				
E.humidicola EO11353	HQ859089	KU832816	KU832027						
E.humifusa MP616	HQ859090	KU832817	KU832028						
E.humifusa MP846	HQ859091	KU832818	KU832029		KU833164		KU831774	KU831575	
E.imbricata MP688	HQ859093	KU832819	KU832030						
E.inaequalis MP539	HQ859094	KU832820	KU832031						
E.inflata ANA	KP737563	KP737429		KP737779					
E.infundibuliformis ANA	KU832439	KU832824	KU832034	KU831663					
E.infundibuliformis MP1238	KU832440	KU832825	KU832035						
E.ingeana EO11805	HQ859096	KU832826	KU832036	KU831664		KU831731	KU831776	KU831577	KU831746
E.ingeana EO12759	KU832441	KU832827	KU832037						
E.innovans MP918	HQ859097	KU832829	KU832039						
E.intermedia alb MP1248	KU832442	KU863012	KU863021						
E.intermedia MM5082	HQ859100	KU832830	KU832040						
E.interrupta MP911	HQ859101	KU863011	KU863020						
E.ioniana EO12781	KU832443	KU832832	KU832042						
E.isaloensis DB1253	KU832444	KU832833	KU832043						
E.jasminiflora EO12612	KU832445	KU832834	KU832044		KU833167		KU831778	KU831579	
E.johnstoniana RC464	KU832446	KU832835	KU832045						
E.juniperina SV952	HQ859105	KU832837	KU832047						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.junonia min MP866	HQ859107	KU832838	KU832048		KU833168		KU831779	KU831580	
E.kingaensis rug BB2981	HQ859109	KU832841	KU832051	KU831665	KU833169	KU831732	KU831780	KU831581	KU831747
E.labialis MP696	HQ859110	KU832844	KU832053						
E.laeta MP1045	KU832452	KU832846	KU832055						
E.lambertii ANA	KU832453	KU832847	KU832056	KU831666					
E.lanata MP1220	KU832454	KU832848	KU832057						
E.lanceolifera RC463	KU832455	KU832849	KU832058						
E.lateralis MP721	KU832456	KU832851	KU832060						
E.lateralis MP727	KU832457	KU832852	KU832061						
E.lavandulifolia EO12506	HQ859116	KU832854	KU832063						
E.leptopus MP746	HQ859120	KU832859	KU832067						
E.leucopelta SB9309	KU832460	KU832863	KU832071		KU833170		KU831781	KU831582	
E.leucotrachela ANA	KU832461	KU832865		KU831669					
E.longimontana MP587	HQ859128	KU832869	KU832076						
E.mackayana b ANA	KP737571	KP737436		KP737783					
E.mackayana PPK	HQ859131	KP737434	KP737668	KP737781	KP737643	KP737695	KP737504	KP737731	KP737715
E.macowanii mac MP810	HQ859132	KU832871	KU832078						
E.macrotrema MM4625	HQ859134	KU832872	KU832079						
E.madagascariensis MP1086	KU832464	KU832873							
E.maderi MP757	HQ859137	KU832874	KU832080						
E.madida MP573	HQ859138	KU832875	KU832081		KU833171		KU831782	KU831583	
E.malmesburiensis EO12575	KU832466	KU832878	KU832084		KU833172		KU831783	KU831584	
E.manipuliflora a ANA	KP737580	KP737446		KP737789					
E.manipuliflora AJH	KP737578	KP737443	KP737670		KP737644		KP737505	KP737732	
E.manipuliflora EO12697	KP737579	KP737445	KP737672						
E.massonii MP811	HQ859143	KU832885	KU832090						
E.melanthera MP610	HQ859145	KU832889	KU832094						
E.melastoma mel MP773	HQ859146	KU832890	KU832095						
E.microdonta A5	KU832473	KU832892	KU832097						
E.mira MP1257	KU832474	KU832893	KU832098						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.modesta ANA	KU832475	KU832894	KU832099	KU831673					
E.monadelphia FO	KU832477	KU832896	KU832101	KU831674					
E.monadelphia FO2	KU832478	KU832897	KU832102						
E.montis hominis EO11827	HQ859149	KU832899	KU832104						
E.multiflora a ANA	KP737583	KP737449		KP737792		KP737696			
E.multumbellifera MP822	HQ859151	KU832901	KU832106						
E.myriadenia EO12679	KU832480	KU832903	KU832108		KU833176		KU831786	KU831587	
E.nabea ANA	KP737586	KP737452		KP737795					
E.natalitia EO12514	HQ859157	KU832906	KU832111						
E.nematophylla EO12747	KU832483	KU832907	KU832112						
E.nevillei MP1056	KU832485	KU832910	KU832115						
E.nubigena MP868	HQ859159	KU832911	KU832116						
E.nutans BG599	HQ859161	KU832913	KU832118						
E.oatesii ANA	KP737587	KP737453		KP737796					
E.obtusata EO12458	HQ859164	KU832916	KU832121						
E.occulta EO	HQ859165	KU832917	KU832122						
E.ocellata MP574	HQ859166	KU832918	KU832123						
E.odorata MP561	HQ859167	KU832919	KU832124						
E.oligantha TdV86	HQ859168	KU832920	KU832125						
E.oresigena MP759	HQ859170	KU832923	KU832128						
E.oxysepala MP780	HQ859172	KU832927	KU832132						
E.palliiflora EO12533	HQ859173	KU832929	KU832134						
E.parilis MP751	HQ859177	KU832932	KU832137						
E.parkeri EO12682	KU832494	KU832933	KU832138	KU831681_KU831682	KU833178				
E.parviflora EO12492	HQ859178	KU832934	KU832139						
E.patersonii a ANA	KU832496	KU832938		KU831683					
E.patersonii MP741	HQ859183	KU832939	KU832143						
E.pectinifolia BG601	HQ859186	KU832942	KU832146	KU831685	KU833180	KU831735	KU831789	KU831590	KU831750
E.penduliflora MP923	HQ859188	KU832944	KU832148						
E.penicilliformis MP889	HQ859190	KU832945	KU832149						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.perrieri EO12635	KU832500	KU832948	KU832152	KU831686	KU833181				
E.perspicua per MP821	HQ859192	KU832949	KU832153						
E.petiolaris EO12783	KU832501	KU832950	KU832154						
E.phillipsii MP794	HQ859199	KU832957	KU832161						
E.physantha MP1242	KU832502	KU832958	KU832162						
E.physodes ANA	KU832503	KU832959		KU831688					
E.physodes MP979	HQ859200	KU832960	KU832163						
E.pillansii pil MP813	HQ859202	KU832962	KU832165						
E.pinea MP789	HQ859204	KU832963	KU832166						
E.piquetbergensis MM5532	KU832504	KU832964	KU832167						
E.placentiflora EO12477	HQ859205	KU832965	KU832168						
E.planifolia MP1012	KU832505	KU832966	KU832169						
E.planifolia MP939	HQ859207	KU832967	KU832170						
E.platycalyx MP1243	KU832506	KU832968	KU832171						
E.platycodon mad EO12695	KP737588	KP737454	KP737673		KP737645		KP737506	KP737733	
E.platycodon pla c ANA	KP737592	KP737457		KP737799					
E.pleiotricha RC461	KU832507	KU832969	KU832172						
E.plukenetii bre MP769	HQ859208	KF160938	KU832173						
E.plukenetii lin MP912	HQ859209	KF160942	KU832174						
E.plumosa EO12480	HQ859215	KU832971	KU832177						
E.pogonanthera EO12835	KU832508	KU832973	KU832179	KU831689					
E.praecox MP795	HQ859219	KU832975	KU832181						
E.prolata EO11935	HQ859221	KU832976	KU832182						
E.prolata EO12748	KU832510	KU832977	KU832183						
E.propendens EO12464	HQ859222	KU832978	KU832184						
E.pubigera MP572	HQ859224	KU832981	KU832187						
E.pycnantha MP1011	KU832513	KU832984	KU832190		KU833184			KU831593	
E.quadrangularis MP620	HQ859227	KU832986	KU832192						
E.quadrisulcata MP1031	KU832515	KU832987	KU832193						
E.radicans sch MP1018	HQ859249	KU832988	KU832194						

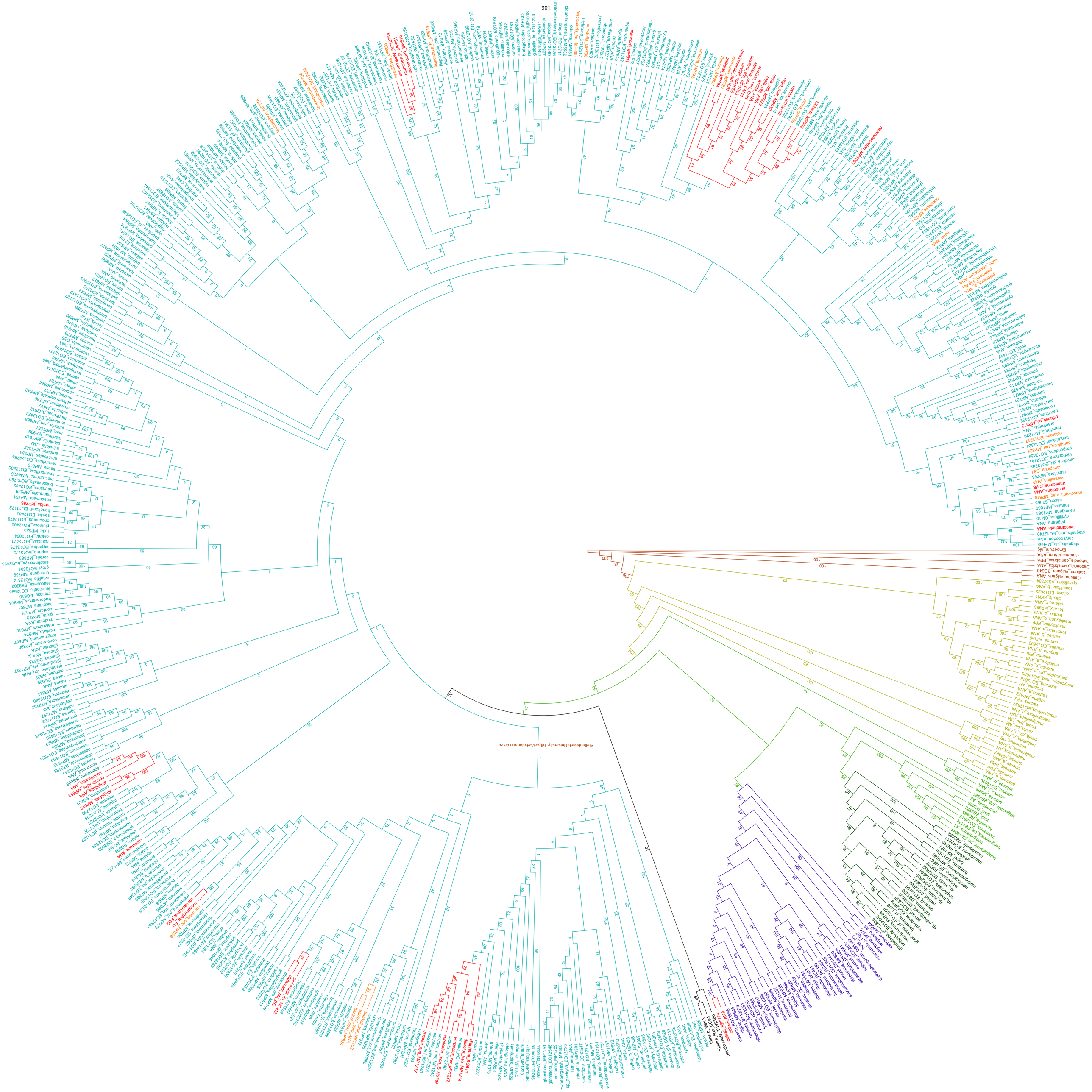
Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.rakotozafyana EO12661	KU832516	KU832989	KU832195	KU831693_KU831694	KU833185				
E.recurvifolia EO12475a	HQ859229	KU832992	KU832198						
E.reenensis MP661	HQ859233	KU832993	KU832199						
E.regerminans MP576	HQ859234	KU832994	KU832200						
E.regia mar MP930	HQ859236	KU832995	KU832201						
E.regia reg ANA	KU832518	KU832996	KU832202	KU831696					
E.regia reg MP922	HQ859237	KU832997	KU832203						
E.retorta ANA	KU832519	KU832999		KU831697					
E.reunionensis FN744	KU832520	KU833000	KU832205	KU831698_KU831699	KU833186				
E.revoluta BT13679	KU832521	KU833001	KU832206						
E.rigidula MP534	HQ859246	KU833004	KU832209		KU833187		KU831792	KU831594	
E.riparia MP908	HQ859247	KU833005	KU832210						
E.rivularis BB13936	KU863008	KU833006	KU832211						
E.rosacea gla BG628	HQ859248	KU833007	KU832212						
E.rubens EO12479	HQ859251	KU833008	KU832213						
E.russakiana MP684	HQ859254	KU833010	KU832216						
E.rusticula EO12471	HQ859255	KU833011	KU832217						
E.salteri S2065	HQ859256	KU833012	KU832218						
E.savileae MP975	KU832522	KU833013	KU832219						
E.scoparia AH	KP737593	KP737459	KP737674		KP737647		KP737508	KP737735	
E.scytophylla MP1021	HQ859260	KF160959	KU832221						
E.serrata MP818	HQ859264	KU833017	KU832224		KU833188		KU831793	KU831595	
E.sessiliflora MP604	HQ859265	KU833018	KU832225						
E.shannonii TdV262	HQ859267	KU833020	KU832227						
E.silvatica A2	KU832523	KU833021	KU832228		KU833189		KU831794	KU831596	
E.silvatica QL12629	HQ859269	KU833023	KU832230						
E.simii RC466	KU832524	KU833024	KU832231						
E.similis MP804	HQ859271	KU833026	KU832233						
E.simulans ANA	KU832525	KU833027	KU832234						
E.sitiens MP827	HQ859273	KU833029	KU832236						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.sociorum cf MP1055	KU832526	KU833030	KU832237						
E.solandri EO12753	KU832527	KU833031	KU832238						
E.sonderiana MP756	HQ859274	KU833032	KU832239						
E.sp mad5 EO12641	KU832529	KU833034	KU832241	KU831703_KU831704	KU833191				
E.sp mitrastylus2 EO12691	KU832530	KU833035	KU832242	KU831705_KU831706	KU833192				
E.sp muscosa EO12639	KU832531	KU833036	KU832243						
E.sp pachysa EO12720	KU832533	KU833038	KU832245						
E.sparrmanii ANA	KU832534	KU833039		KU831707					
E.sparrmanii BG608	HQ859276	KU833040	KU832246						
E.spectabilis MP929	HQ859278	KU833042	KU832248						
E.sphaerocephala MP848	HQ859280	KU833043	KU832249						
E.spiculifolia b ANA	KP737612	KP737477		KP737816					
E.spumosa MP978	HQ859281	KU833044	KU832250						
E.squarrosa EO11742	HQ859282	KU833045	KU832251						
E.stagnalis min EO12740	KU832535	KU833046	KU832252						
E.stagnalis sta MP668	HQ859283	KU833047	KU832253						
E.stokoei MP825	HQ859285	KU833050	KU832256						
E.straussiana MP638	HQ859286	KU833051	KU832257		KU833194				
E.strigilifolia ANA	KU832537	KU833052	KU832258	KU831708					
E.strigilifolia MP619	HQ859287	KU833053	KU863019						
E.subdivaricata MP671	HQ859289	KU833057	KU832262						
E.subulata MdV2	HQ859290	KU833058	KU832263						
E.subverticillaris EO12625	KU832540	KU833059	KU832264	KU831710	KU833196				
E.tegulifolia MP557	HQ859292	KU833061	KU832266						
E.terminalis a ANA	KP737614	KP737479		KP737818					
E.tetralix c ANA	KP737622	KP737484		KP737823					
E.tetralix MP966	KP737618	KP737481	KP737679	KP737820	KP737650	KP737700	KP737511	KP737738	KP737718
E.tetrathecoides MP1252	KU832543	KU833064	KU832268						
E.thamnoides MP1211	KU832544	KU833065	KU832269						
E.thodei MP656	HQ859296	KU833067	KU832271						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.thomae MP807	HQ859297	KU833068	KU832272	KU831712					
E.thomensis EO12615	KU832546	KU833069	KU832273		KU833199			KU831600	
E.totta MP525	HQ859300	KU833075	KU832278						
E.tradouwensis MP903	HQ859301	KU833076	KU832279						
E.transparens MP893	HQ859303	KU833077	KU832280						
E.trichophora EO12701	KU832549	KU833078	KU832281						
E.trichophylla EO10906	HQ859304	KU833079	KU832282						
E.trimera ken BG94	HQ859307	KU833082	KU832285						
E.tristis MP932	HQ859310	KU833083	KU832286		KU833200		KU831797	KU831601	
E.tumida MP755	HQ859311	KU833084	KU832287						
E.turgida S1962	HQ859312	KU833085	KU832288						
E.uberiflora BG586	HQ859313	KU833087	KU832290						
E.umbratica EO12760	KU832552	KU833090	KU832294						
E.unicolor mut MP1245	KU832554	KU833092	KU832296						
E.unicolor uni MP1249	KU832555	KU833093	KU832297						
E.unilateralis MP1205	KU832556	KU833094	KU832298						
E.urceolata MP955	HQ859315	KU833095	KU832299		KU833204				
E.urna viridis ANA	KU832557	KU833096		KU831715					
E.urna viridis MP946	HQ859316	KU833097	KU832300						
E.ustulescens RT1553	HQ859317	KU833098	KU832301						
E.uysii ANA322	KU832558	KU833099	KU832302	KU831716					
E.vagans a ANA	KP737628	KP737492		KP737830		KP737705			
E.vagans MP972	HQ859319	KP737490	KP737682						
E.vagans PPJ	HQ859320	KP737491	KP737683	KP737829	KP737652	KP737704	KP737513	KP737740	KP737720
E.velatiflora EO12547	HQ859322	KU833102	KU832305		KU833205		KU831799	KU831604	
E.ventricosa MP713	KU832562	KU833104	KU832307						
E.verecunda ANA	KU832563	KU833105		KU831718					
E.vernicosa MP928	HQ859326	KU833107	KU832309						
E.versicolor mon EO12705	KU832564	KU833108	KU832310						
E.versicolor ver MP1232	KU832565	KU833109	KU832311						

Accession	ITS	trnL-F-ndhJ	trnT-L-intron	trnK-matK	rpl32	rpl16	psbM	atpI	rbcL
E.vestita ANA	KU832567	KU833111	KU832313	KU831720					
E.vestita EO12702	KU832568	KU833112	KU832314		KU833206		KU831800	KU831605	
E.villosa EO11394	HQ859330	KU833113	KU832315						
E.viridiflora MP1246	KU832569	KU833114	KU832316						
E.viscaria lon MP678	HQ859332	KU833115	KU832317		KU833207			KU831606	
E.viscaria mac MP808	HQ859333	KU833116	KU832318						
E.viscaria pen EO12466	HQ859334	KU833117	KU832319						
E.viscaria vis MdV4	HQ859335	KU833118	KU832320						
E.walkeri MP1237	KU832570	KU833119	KU832321						
E.wendlandiana EO12731	KU832571	KU833120	KU832322						
E.whyteana BG75	HQ859337	KU833122	KU832324						
E.woodii DB1444	KU832573	KU833123_KU833124	KU832325	KU831722_KU831723	KU833209				
E.zeyheriana EO	KU832576	KU833127	KU832328						
E.zwartbergensis MP608	HQ859341	KU833128	KU832329						
Empetrum_nig	HQ858880	KP737381	KP737654	KP737745	KP737634	KP737686	KP737496	KP737723	KP737709

Appendix C: *Erica* phylogeny


Stellenbosch University https://scholar.sun.ac.za

Appendix D: CLC Bio assembly report


CLC bio

Table of contents

1. Erica_pluketnii_ATCACG_L002_R1_001 (paired) trimmed (paired) assembly summary report

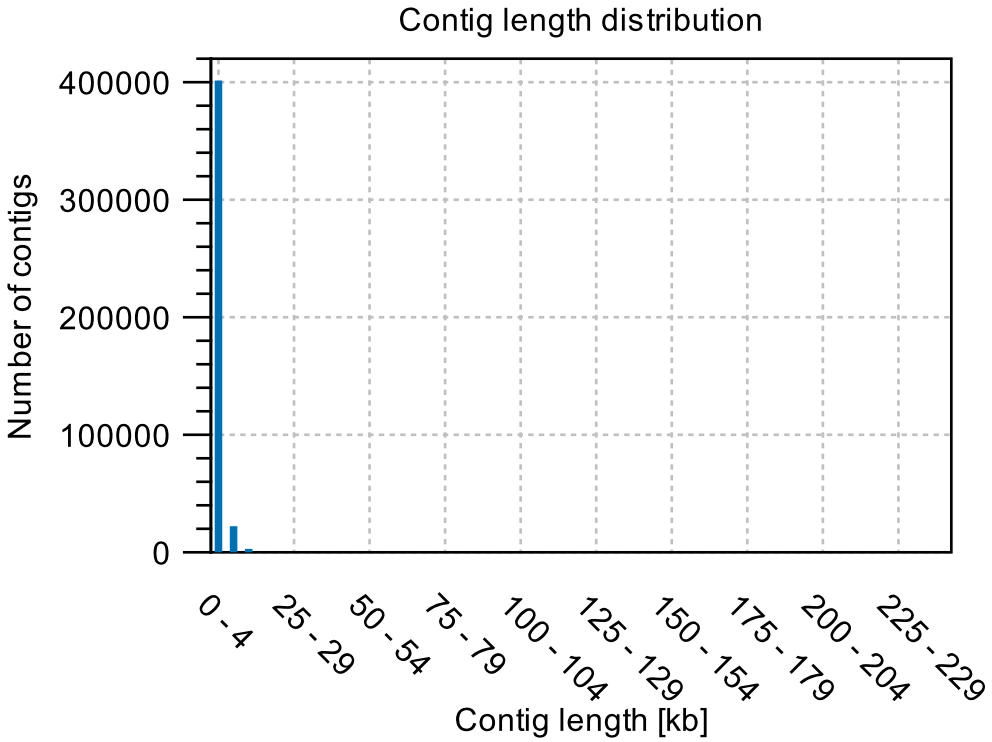
1.1 Nucleotide distribution

Nucleotide	Count	Frequency
Adenine (A)	224,390,663	28.0%
Cytosine (C)	156,691,149	19.5%
Guanine (G)	156,523,550	19.5%
Thymine (T)	224,491,981	28.0%
Any nucleotide (N)	40,110,018	5.0%

1.2 Contig measurements (including scaffolded regions)

N75	1,301
N50	2,510
N25	4,980
Minimum	500
Maximum	243,863
Average	1,872
Count	428,553

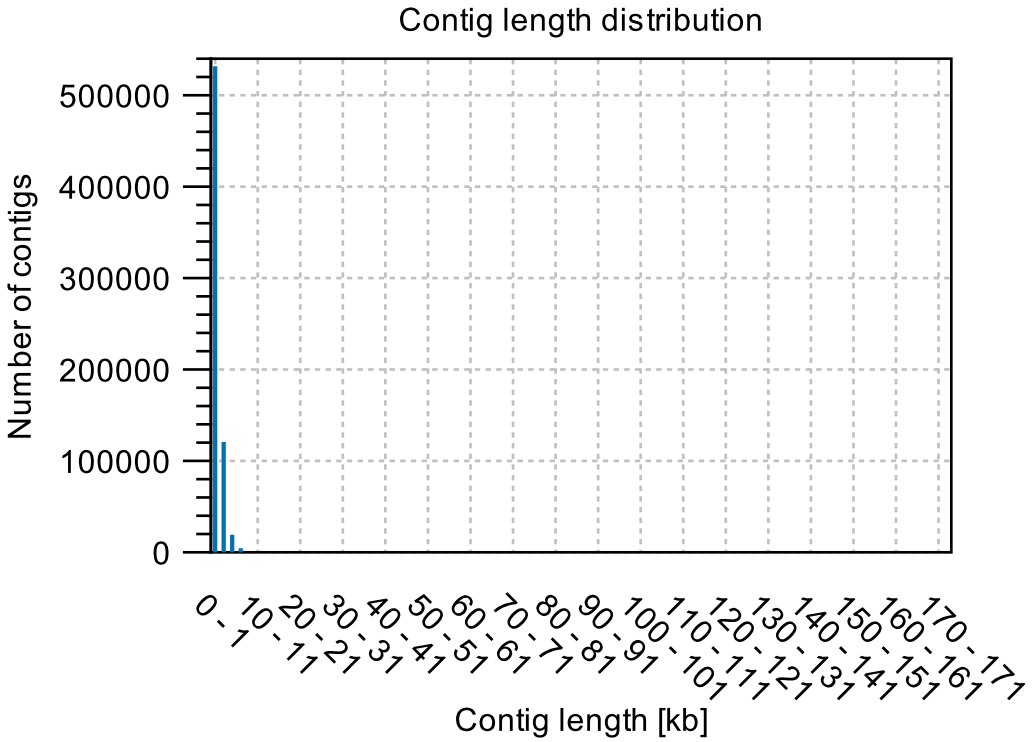
Total	802,207,361



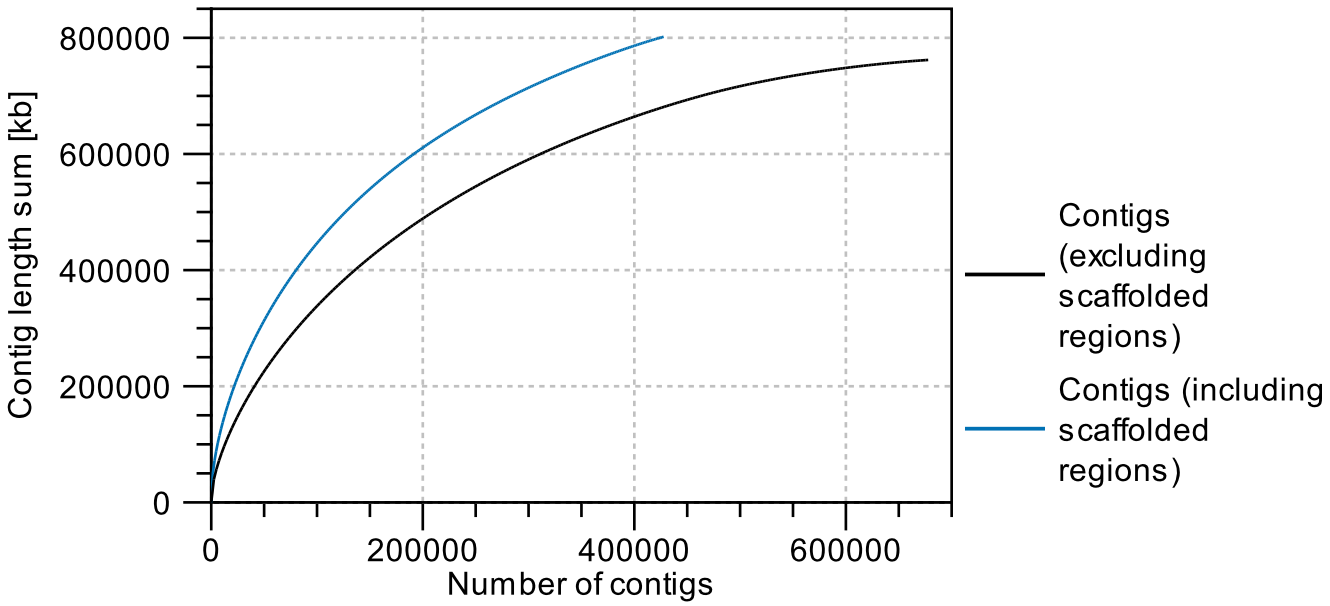
1.3 Contig measurements (excluding scaffolded regions)

N75	918
N50	1,673
N25	3,051
Minimum	27
Maximum	172,266
Average	1,122
Count	679,357

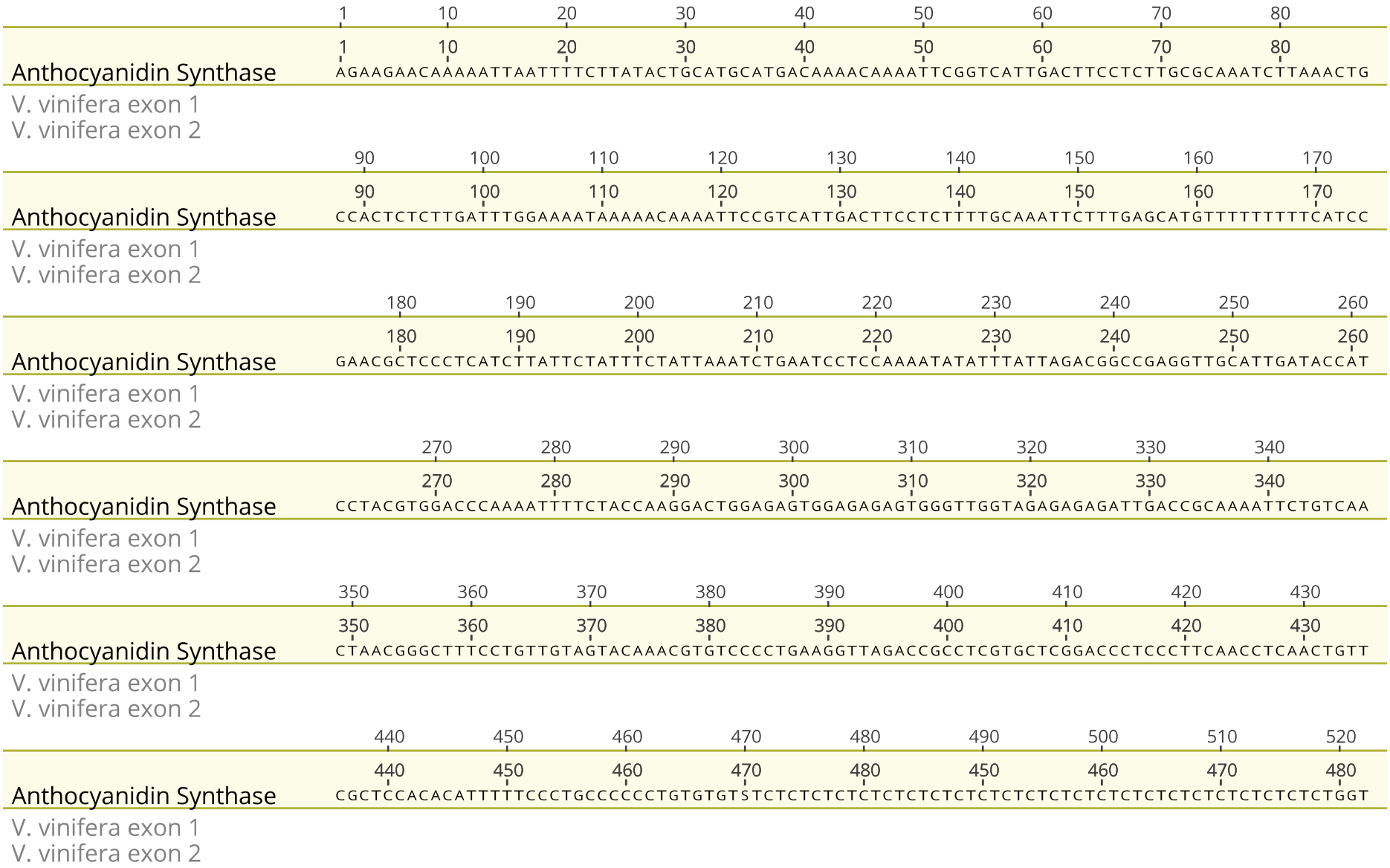
Total	762,099,485

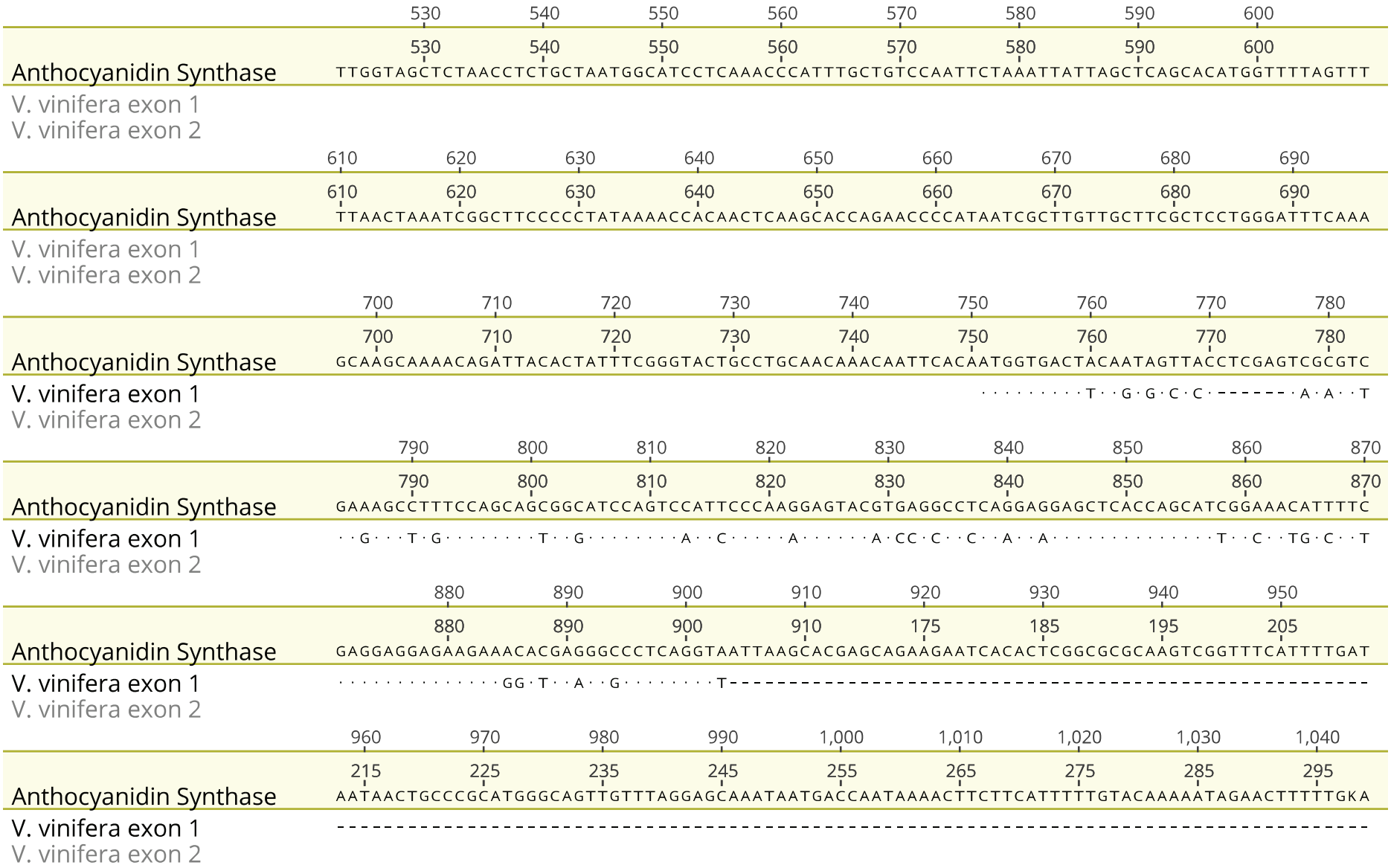


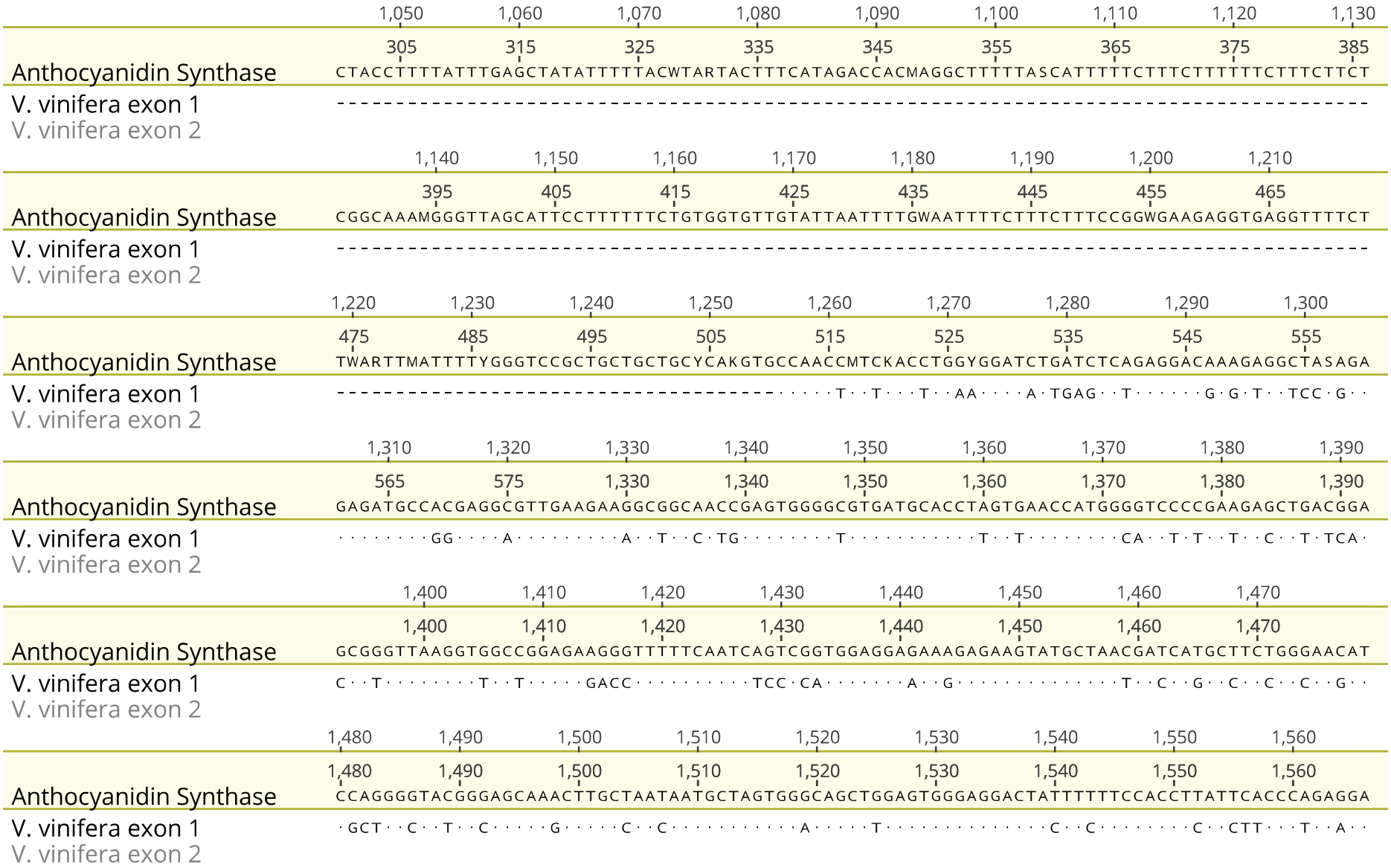
1.4 Accumulated contig lengths

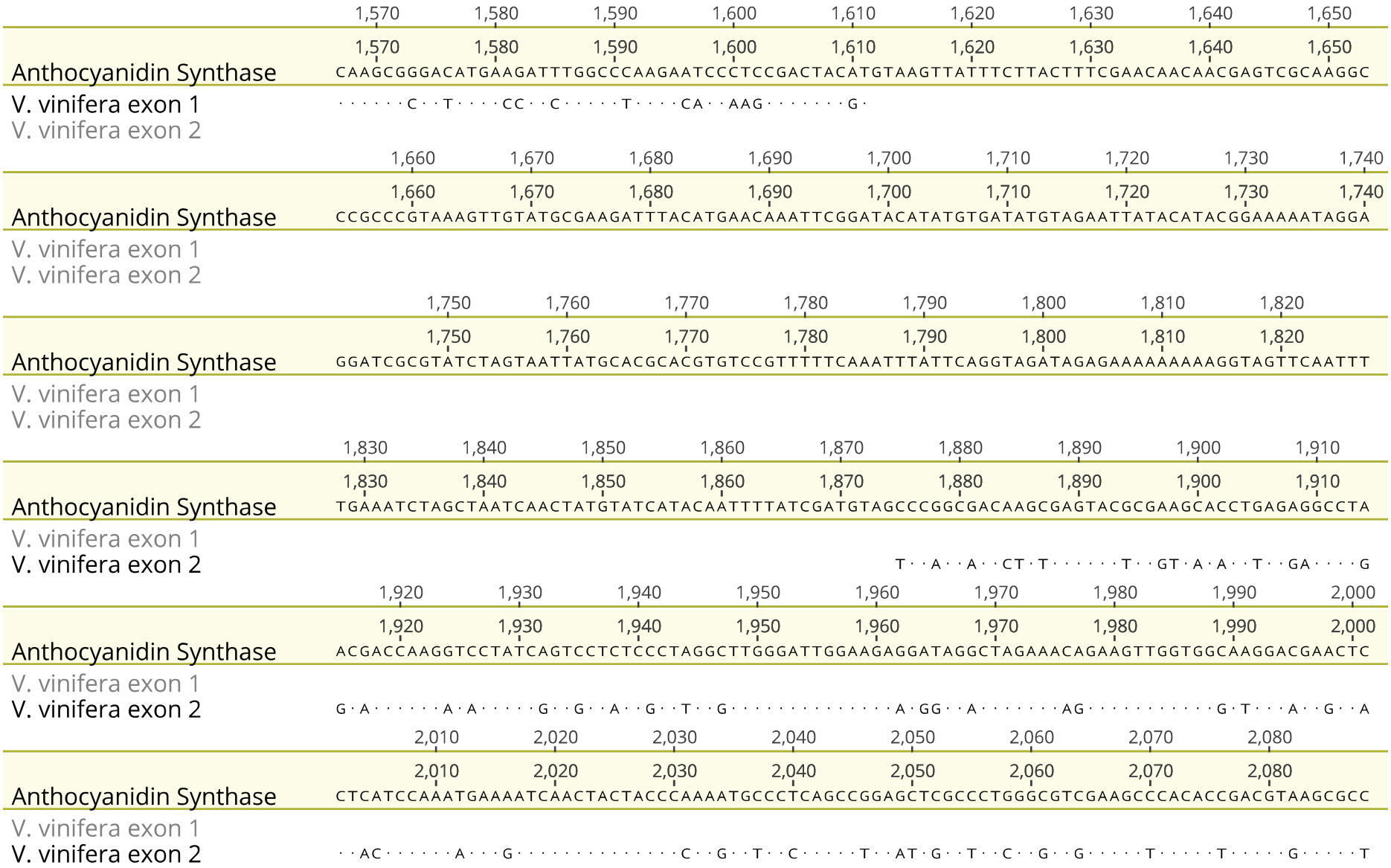


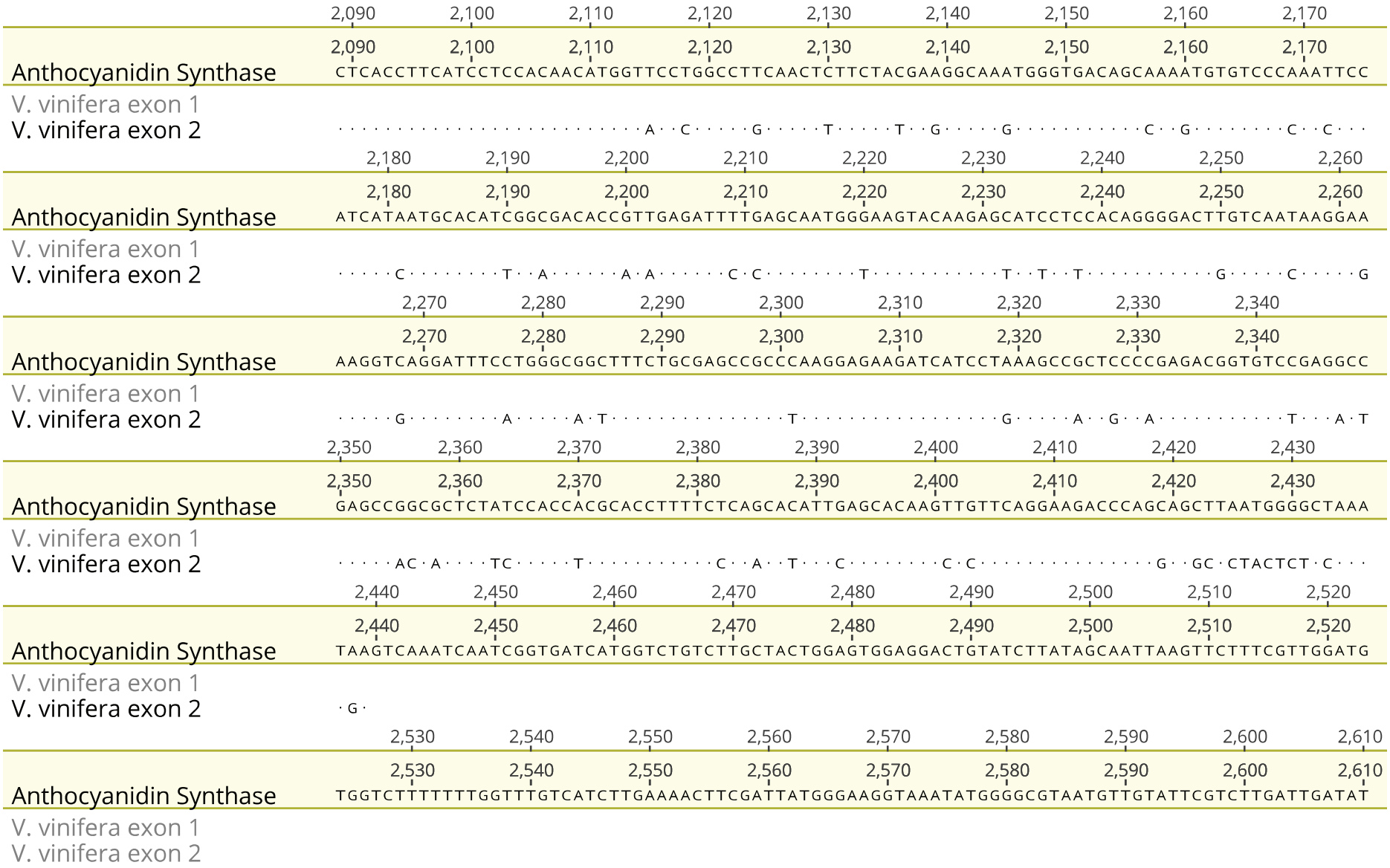
Appendix E: Sequence alignments

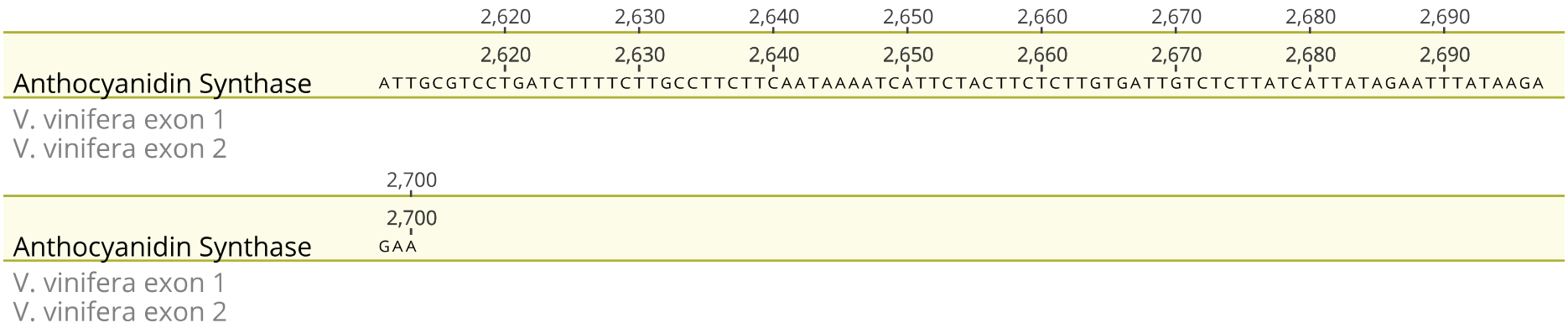


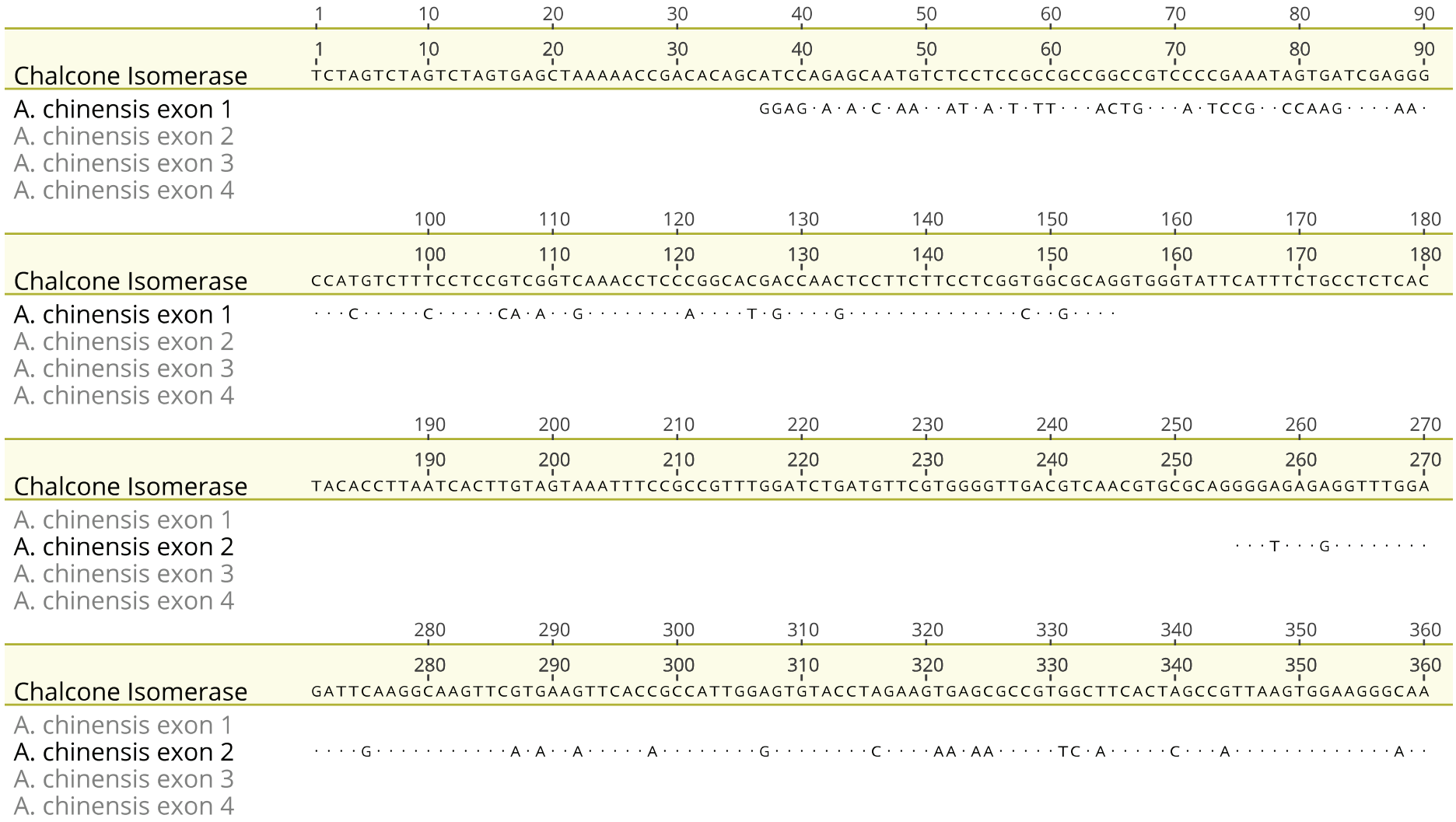


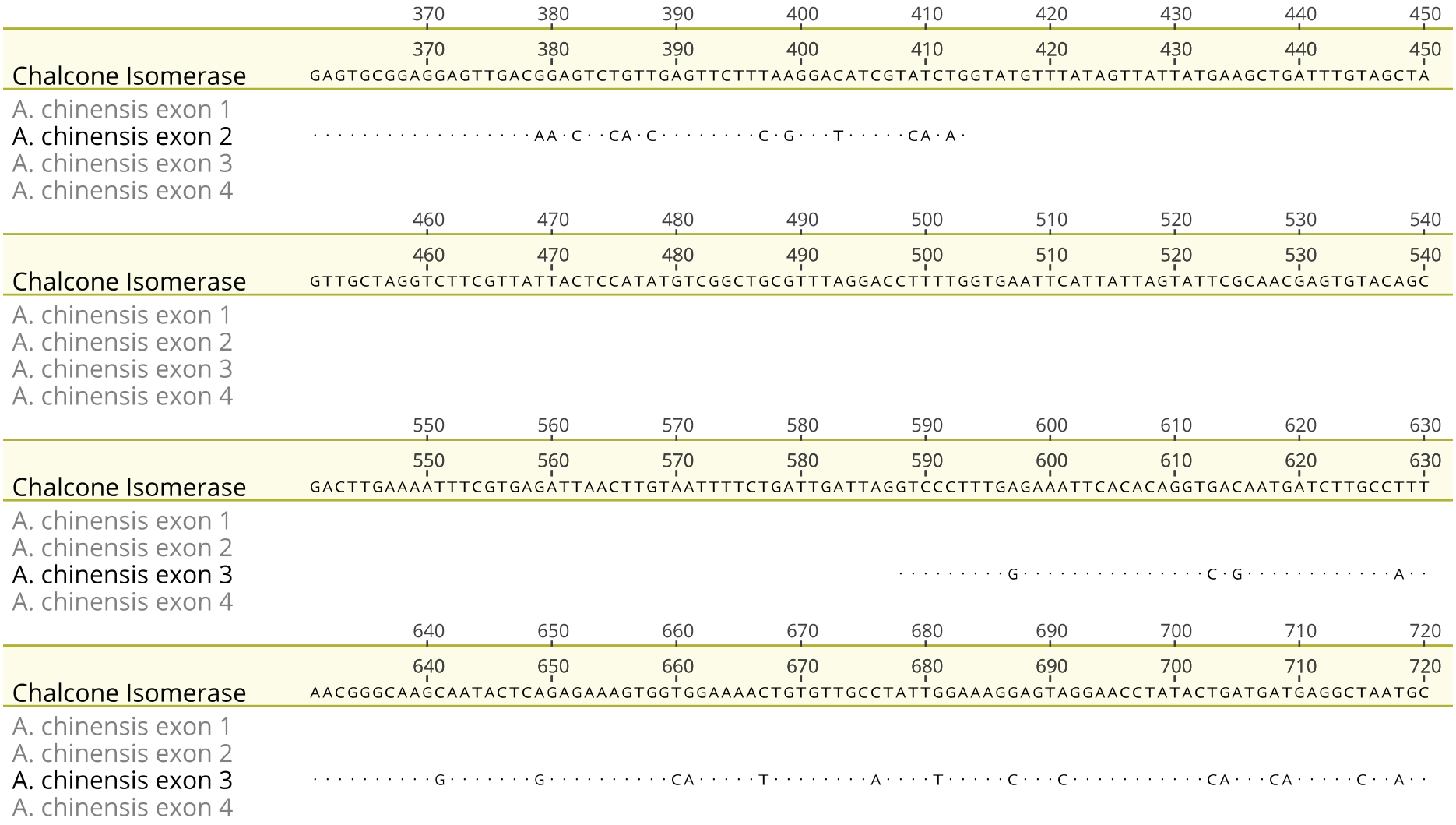


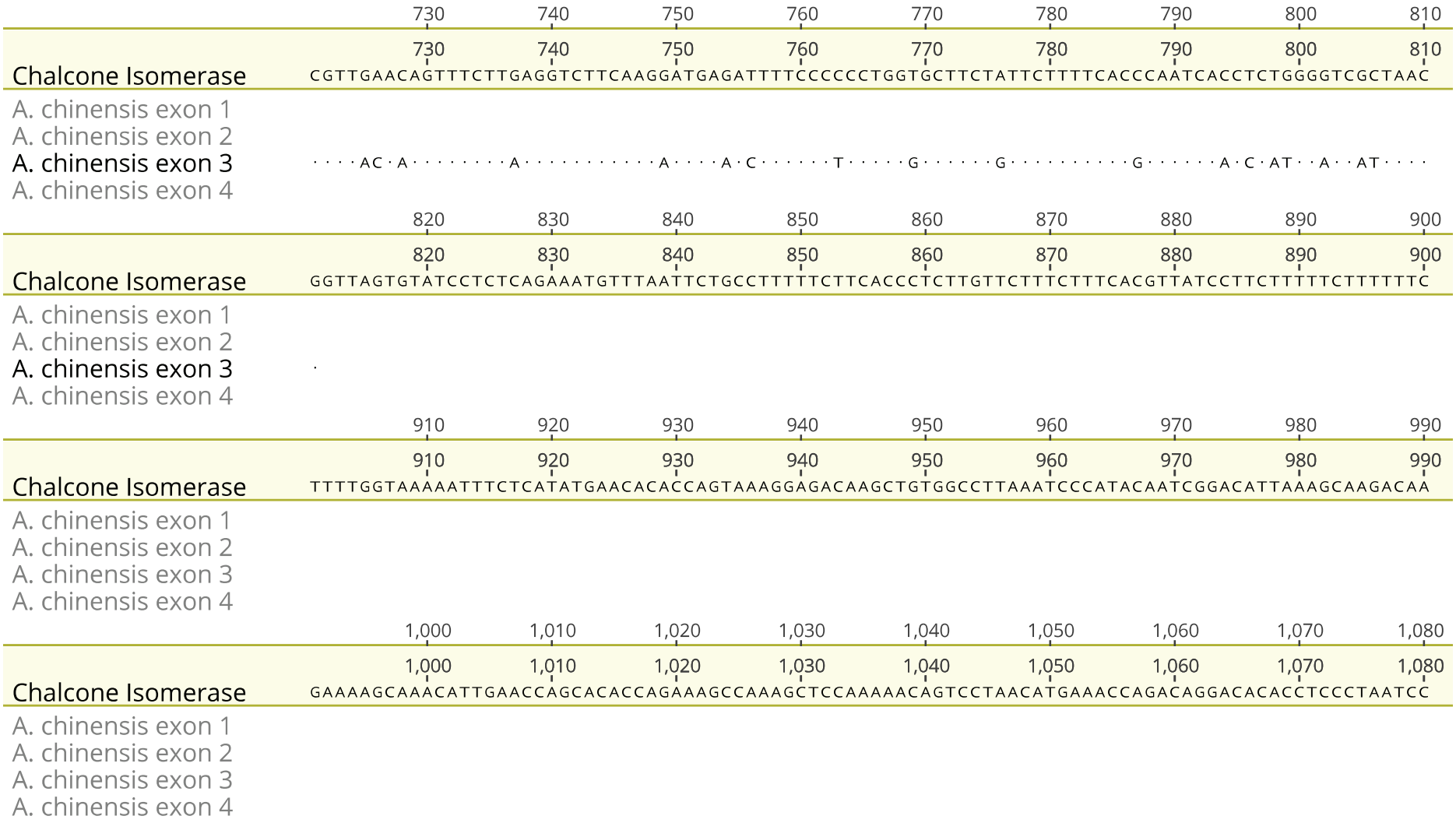


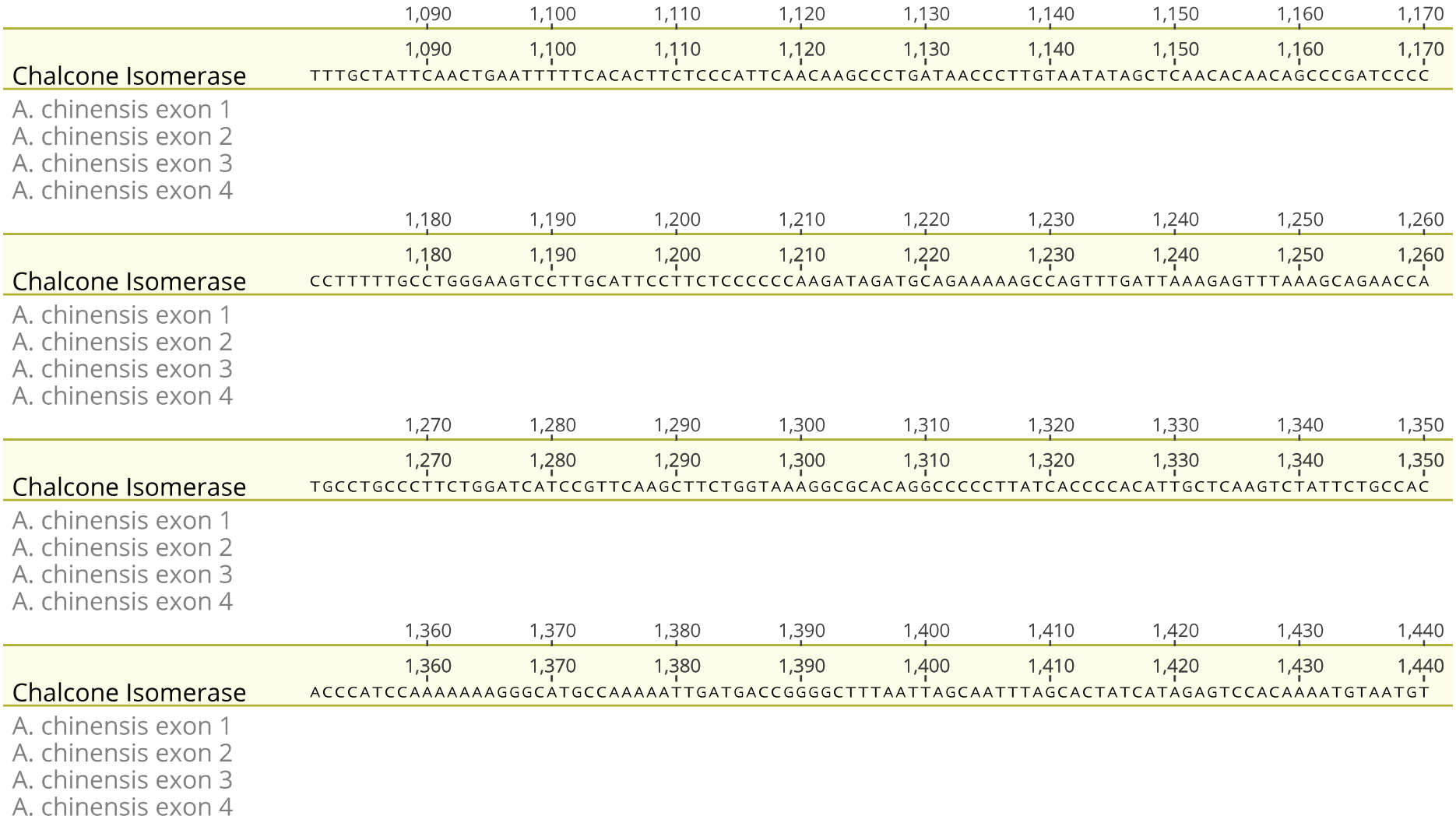


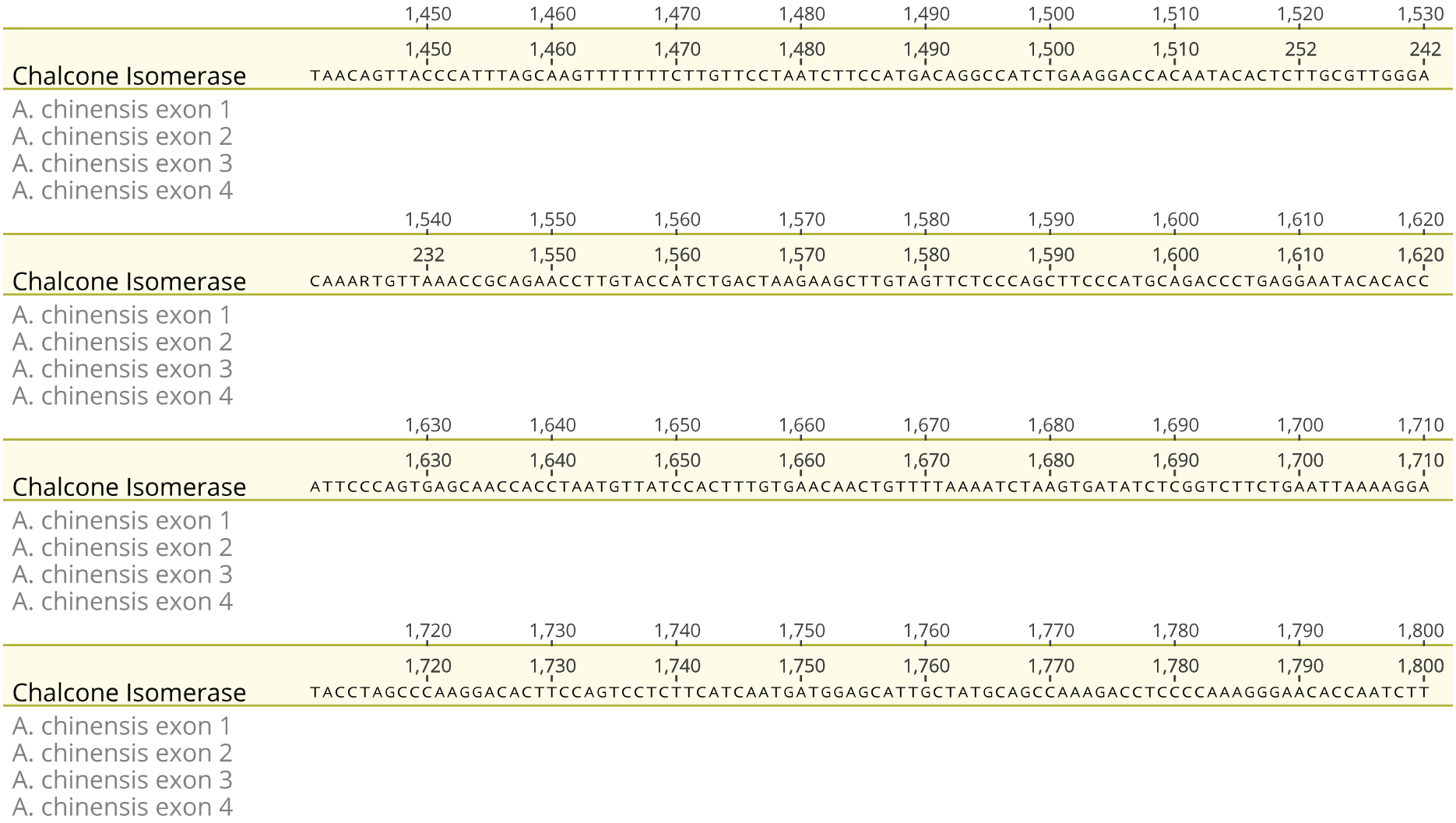


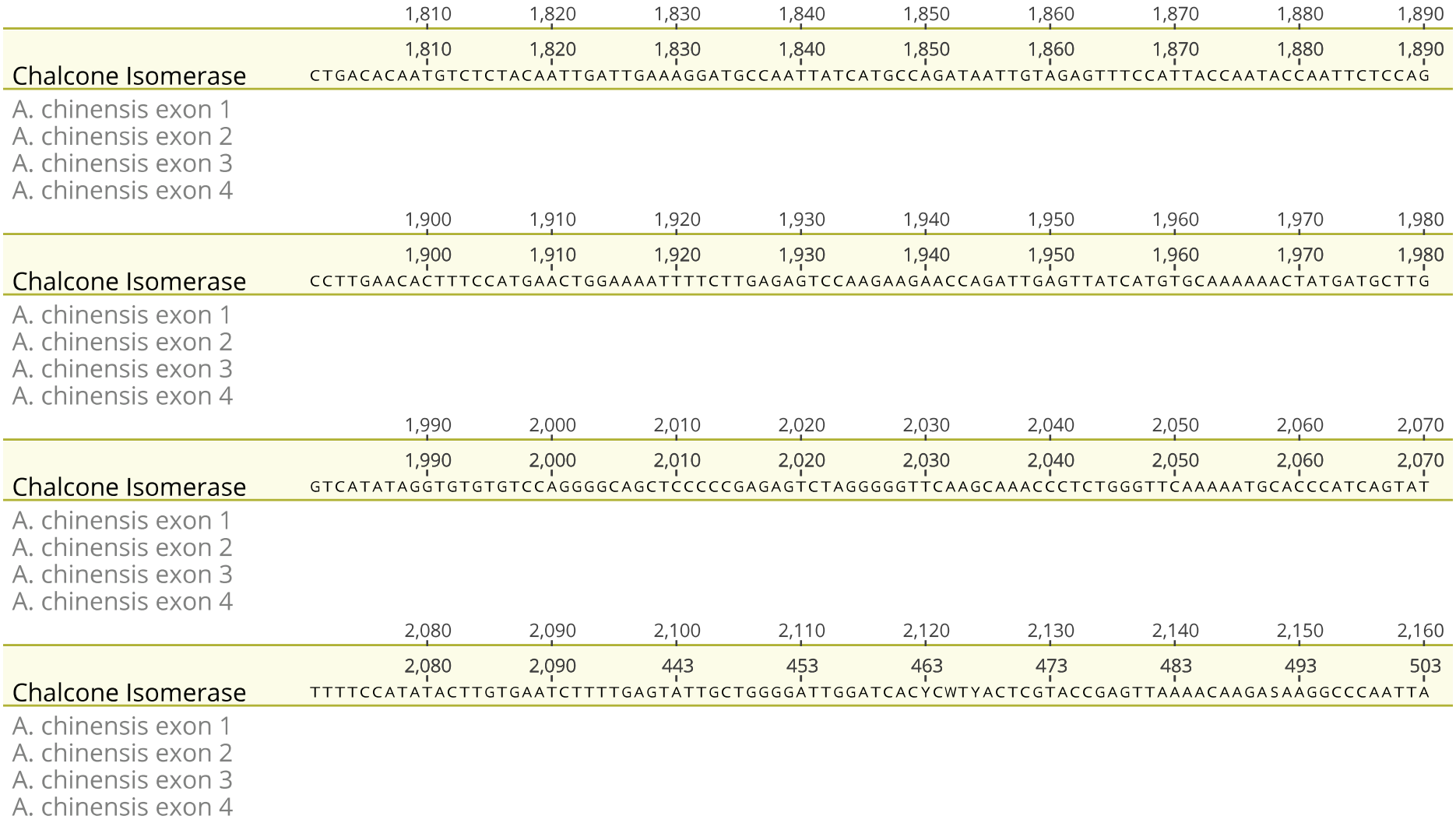


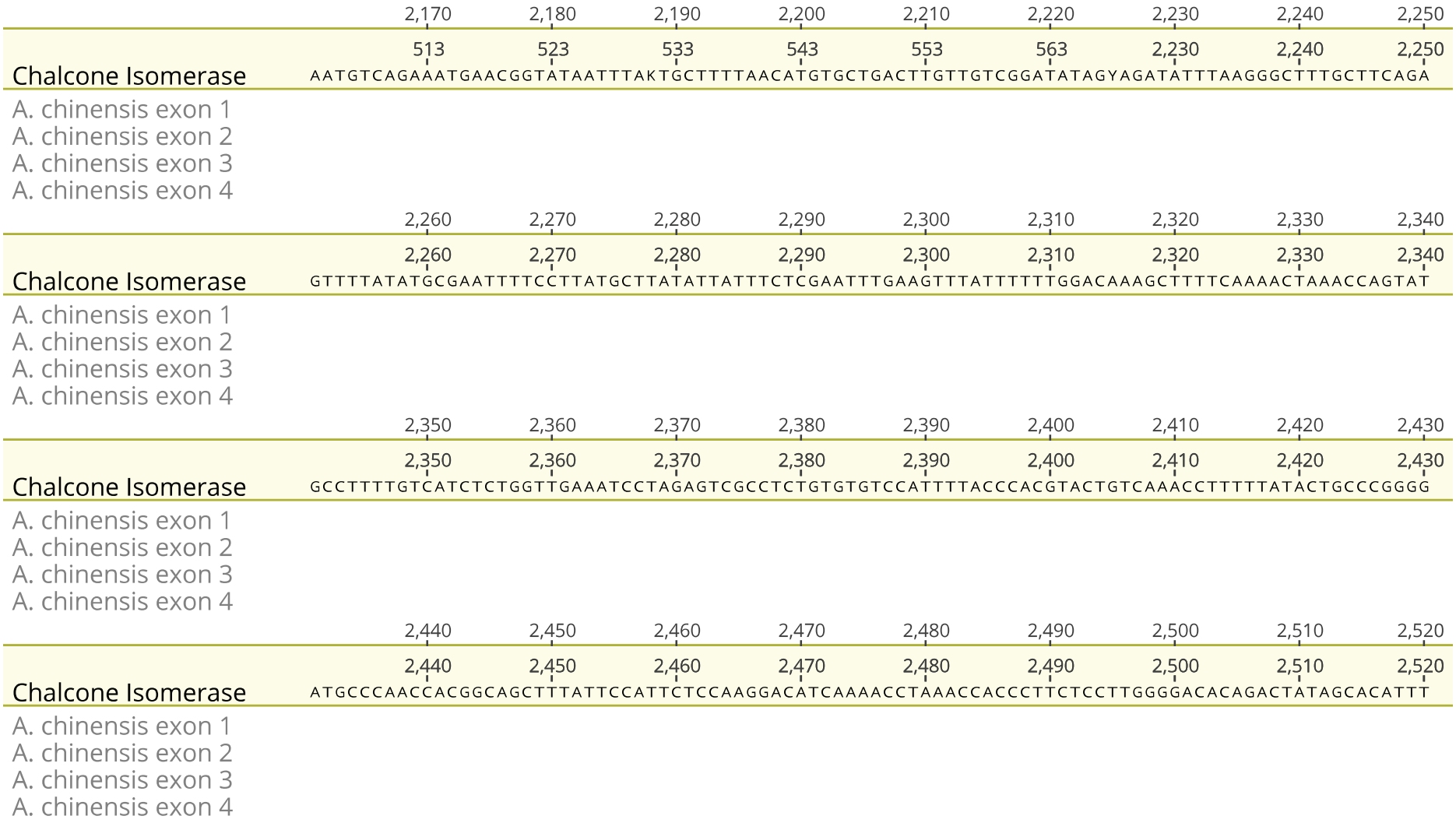


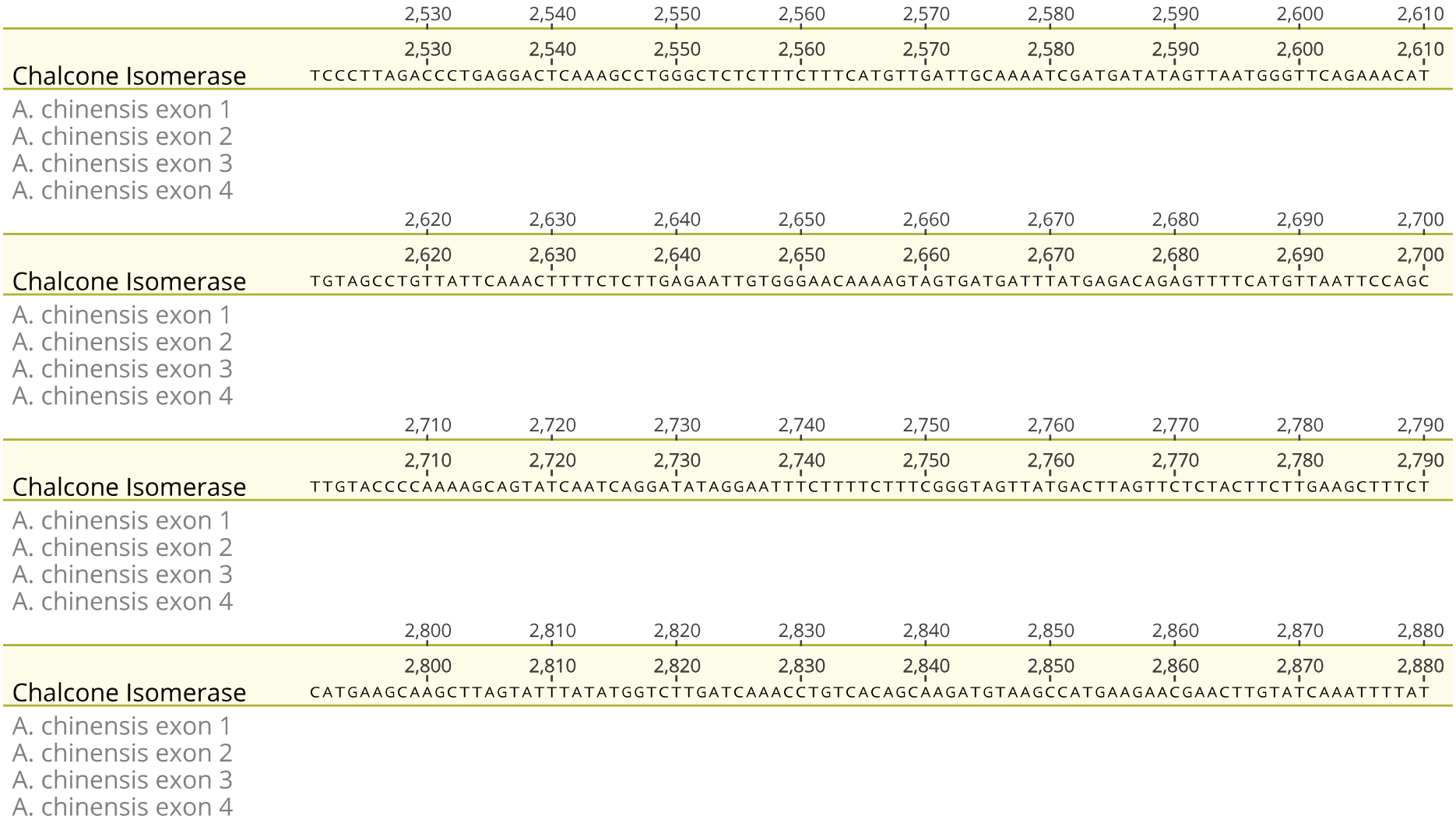


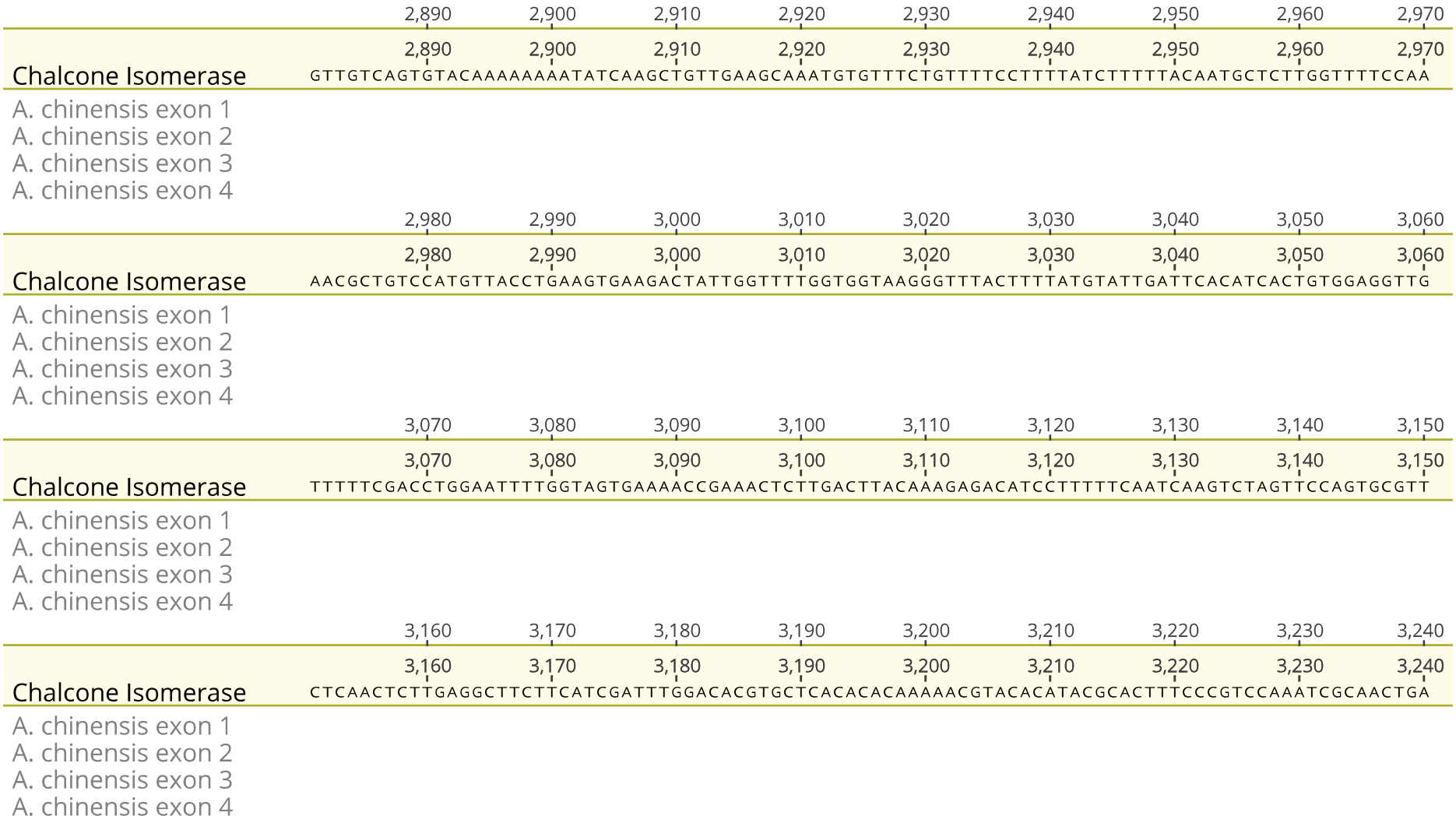


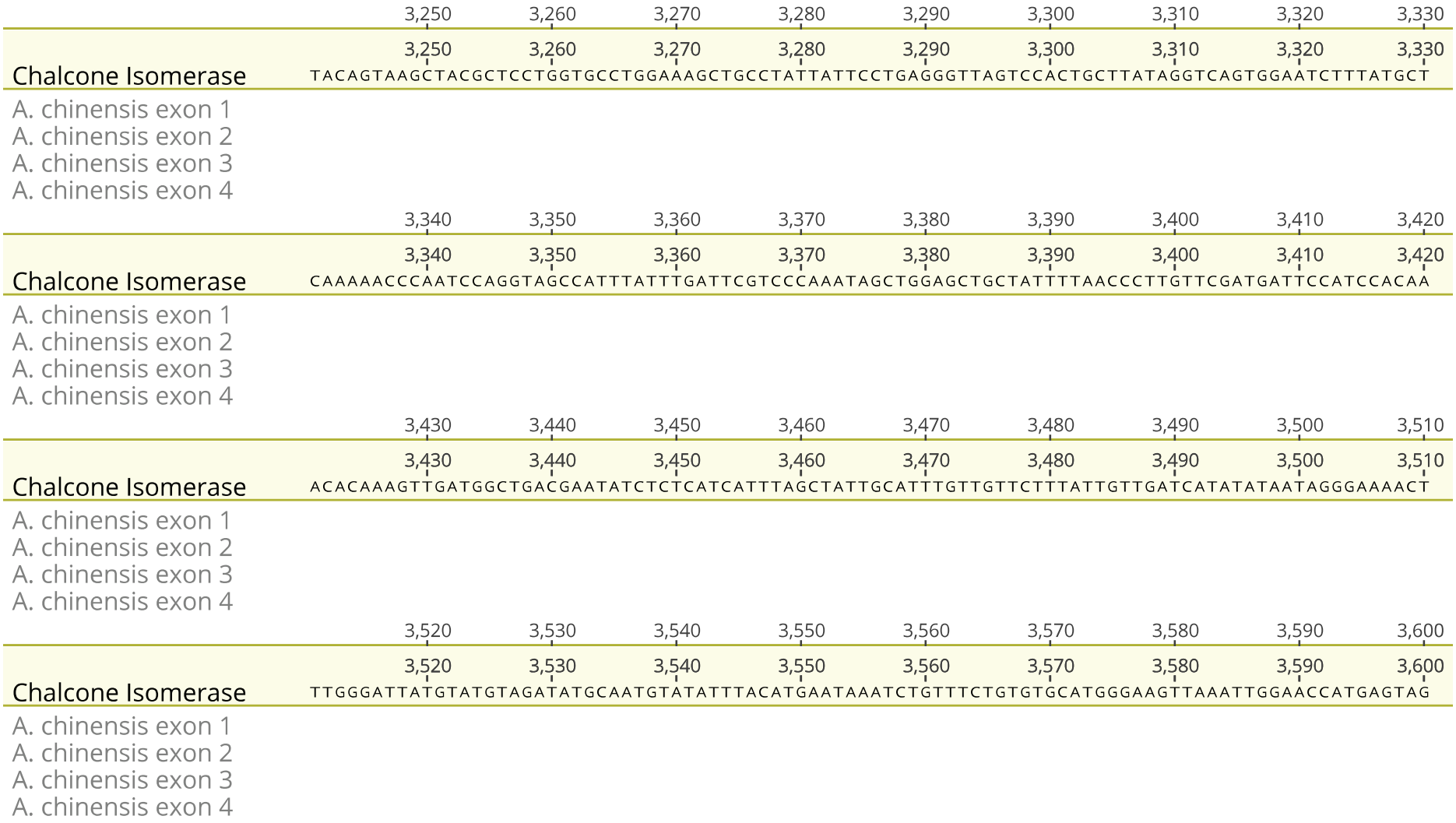


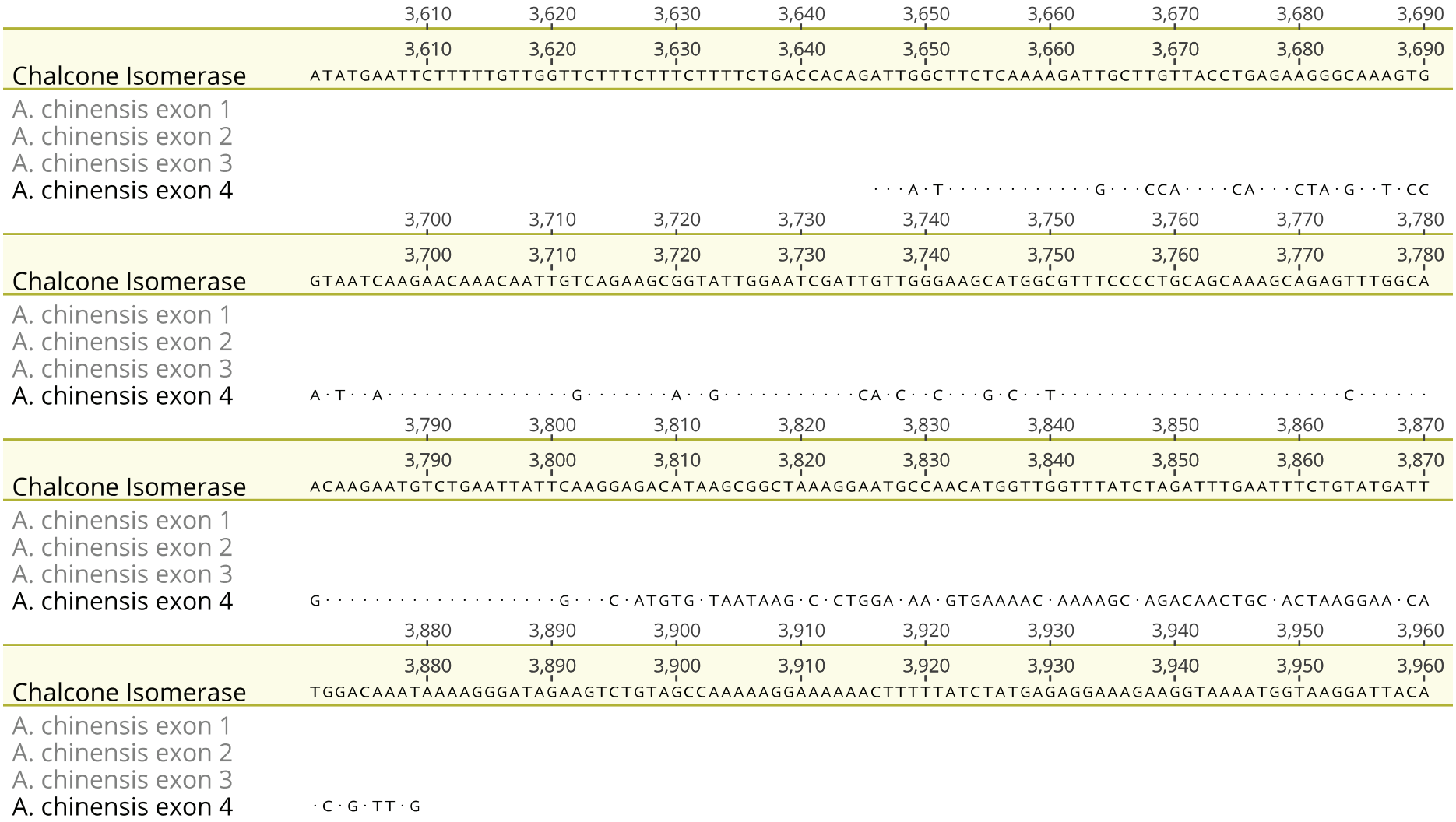


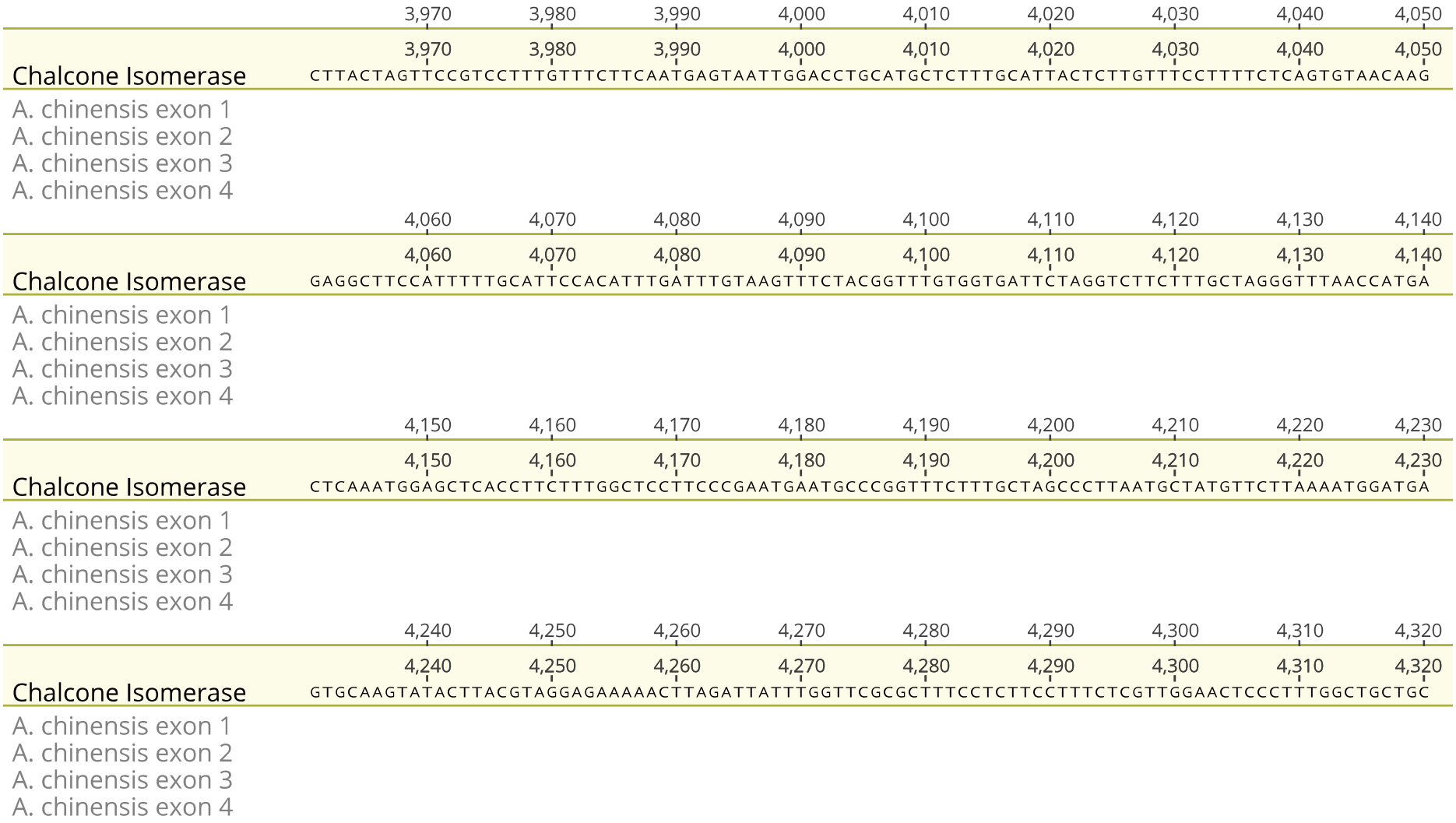


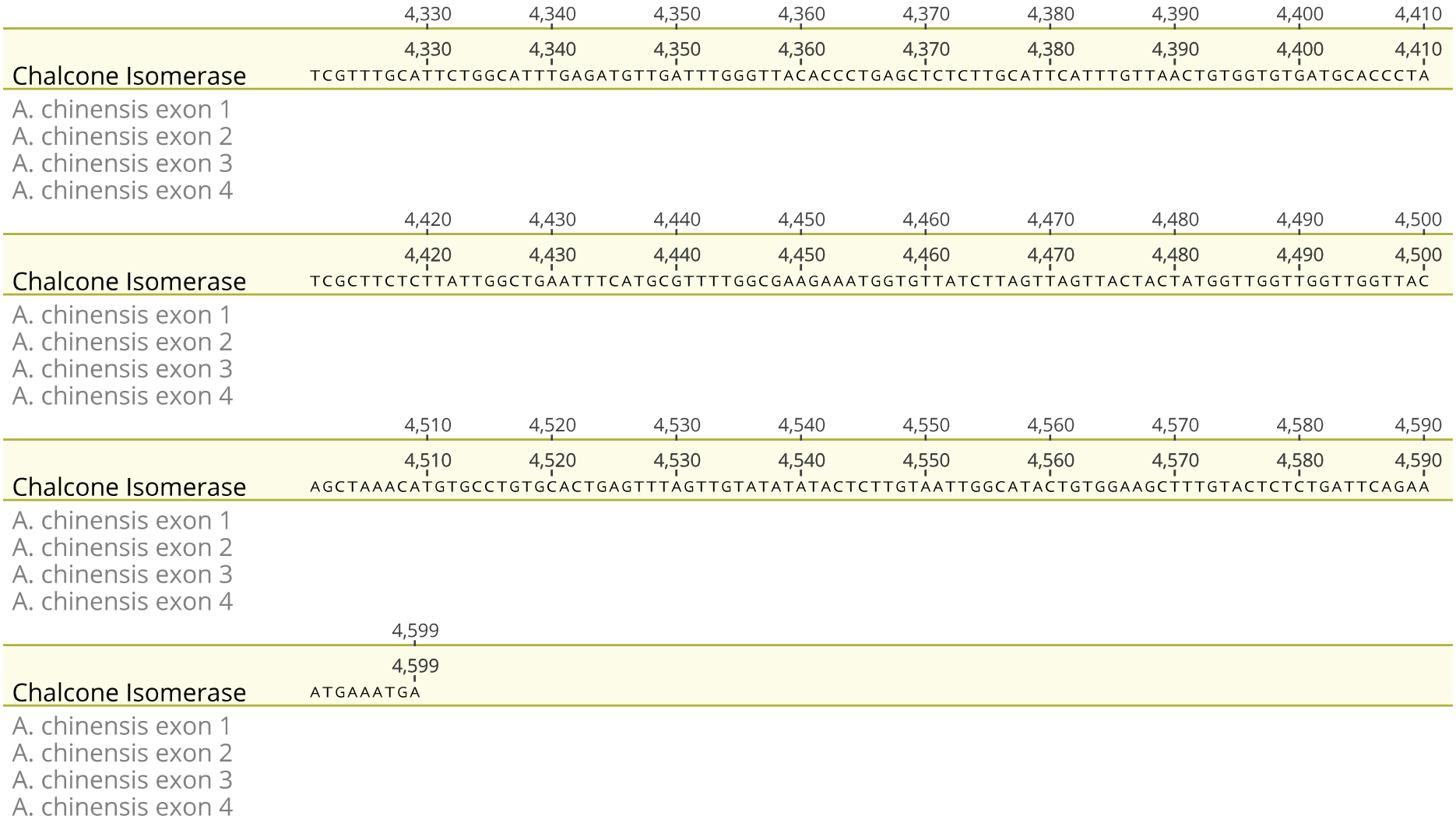


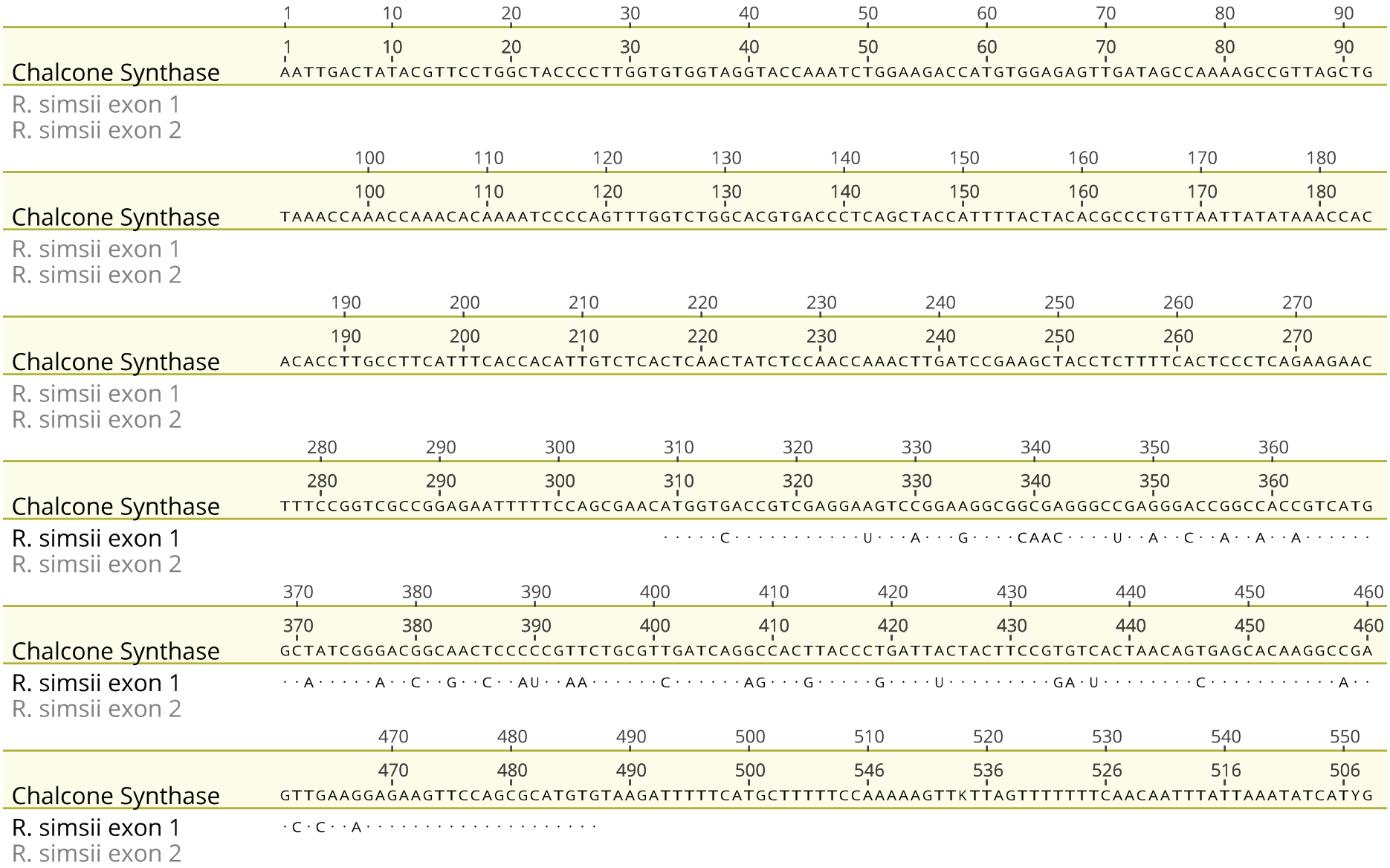


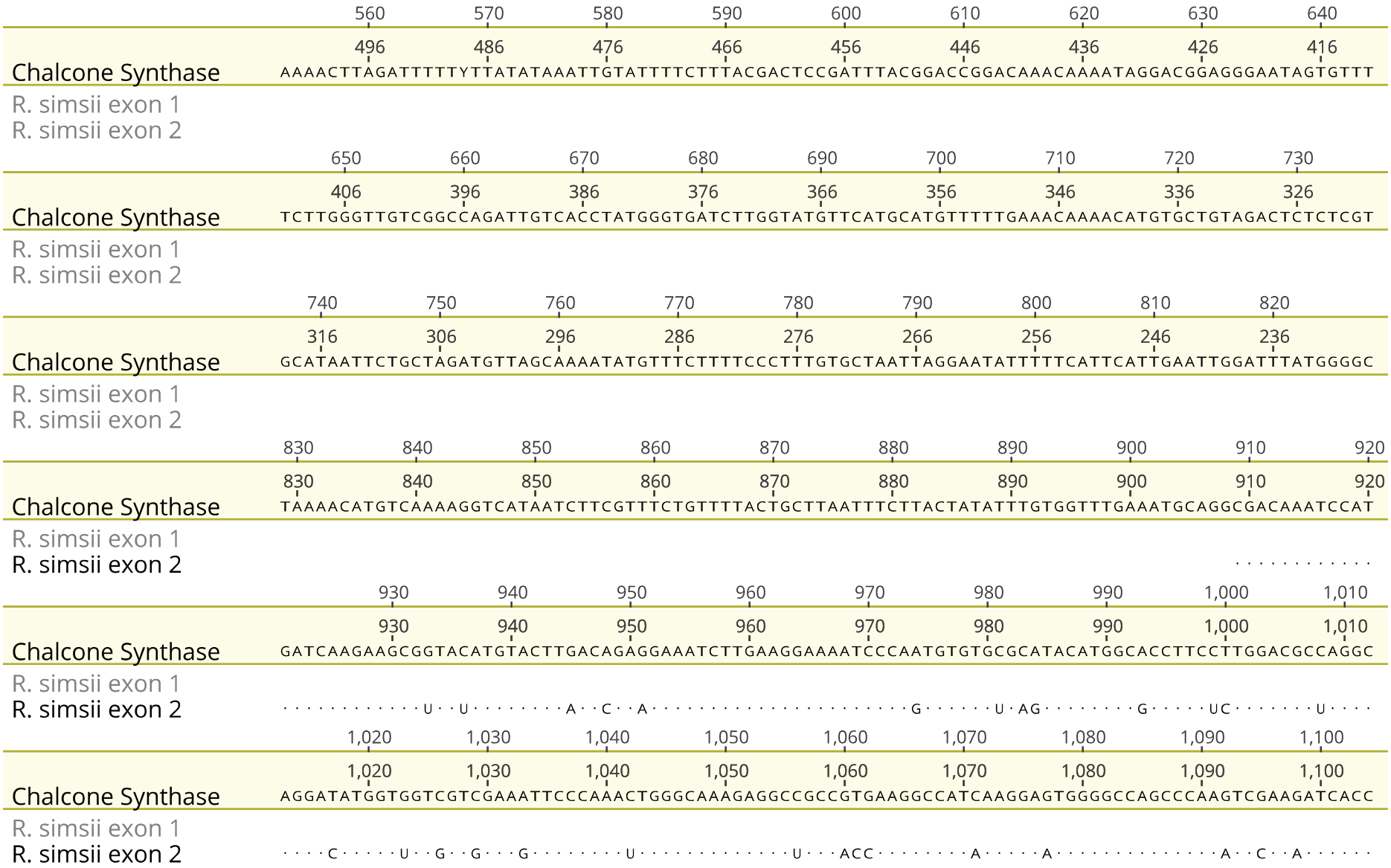


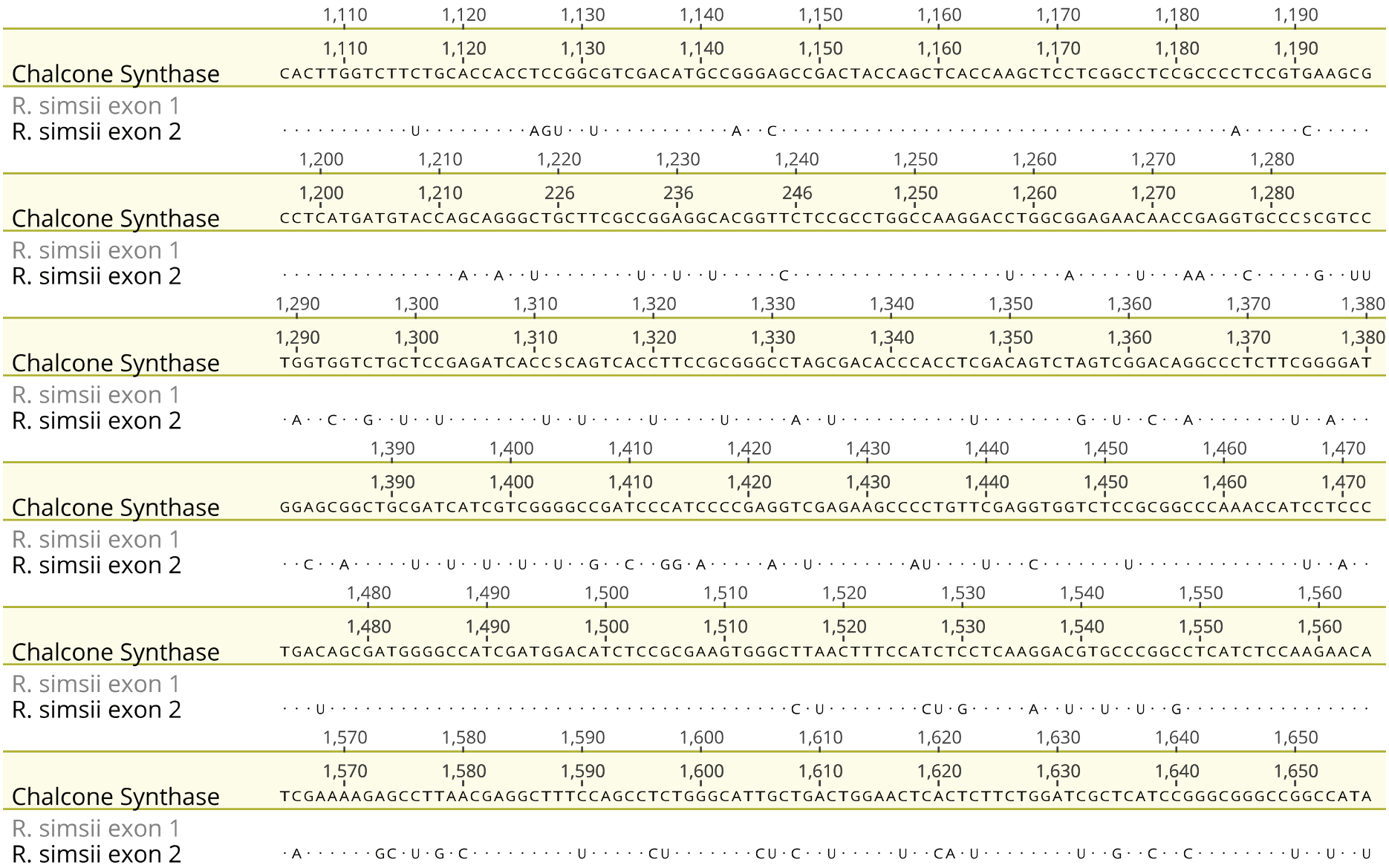


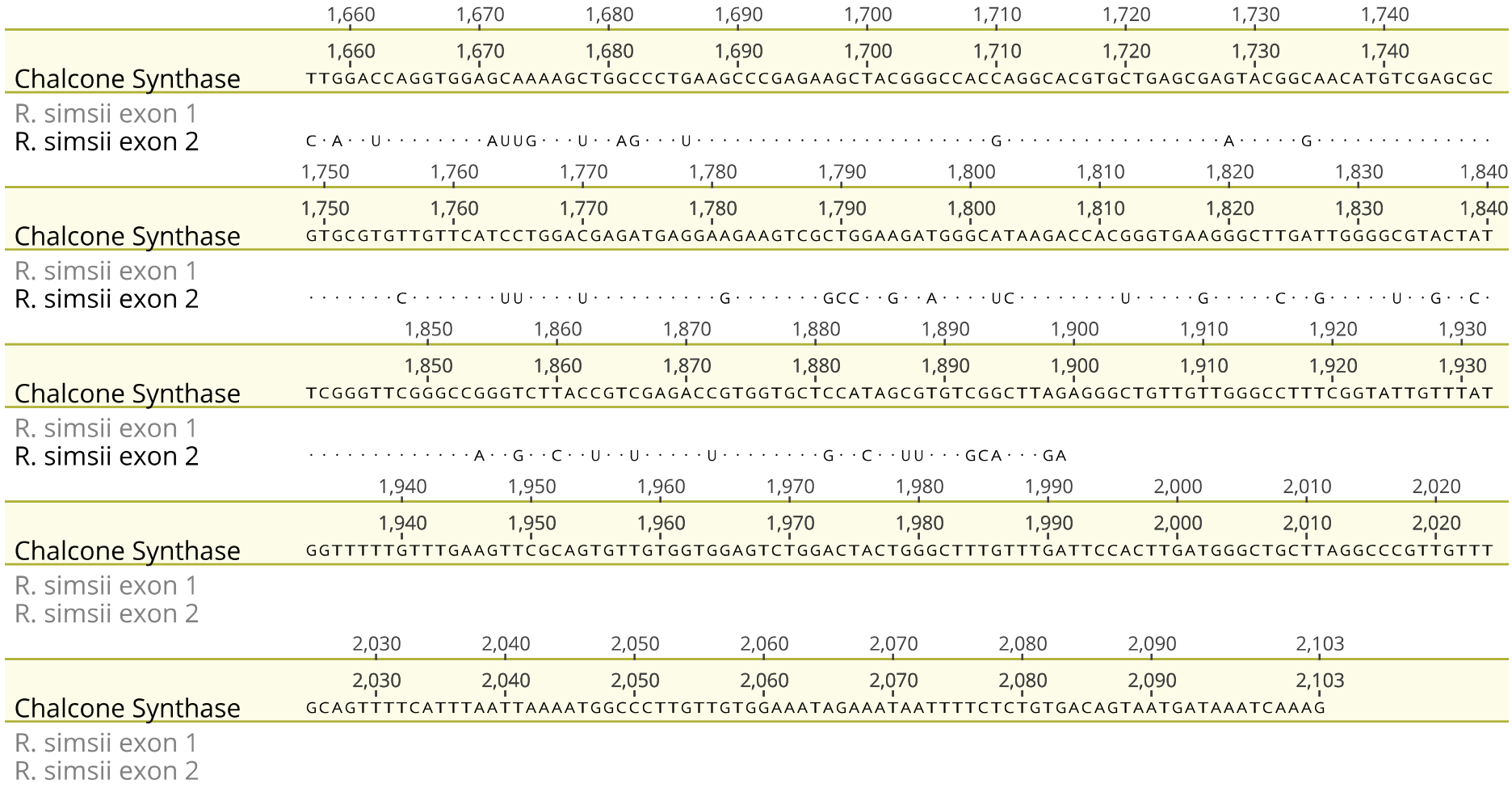


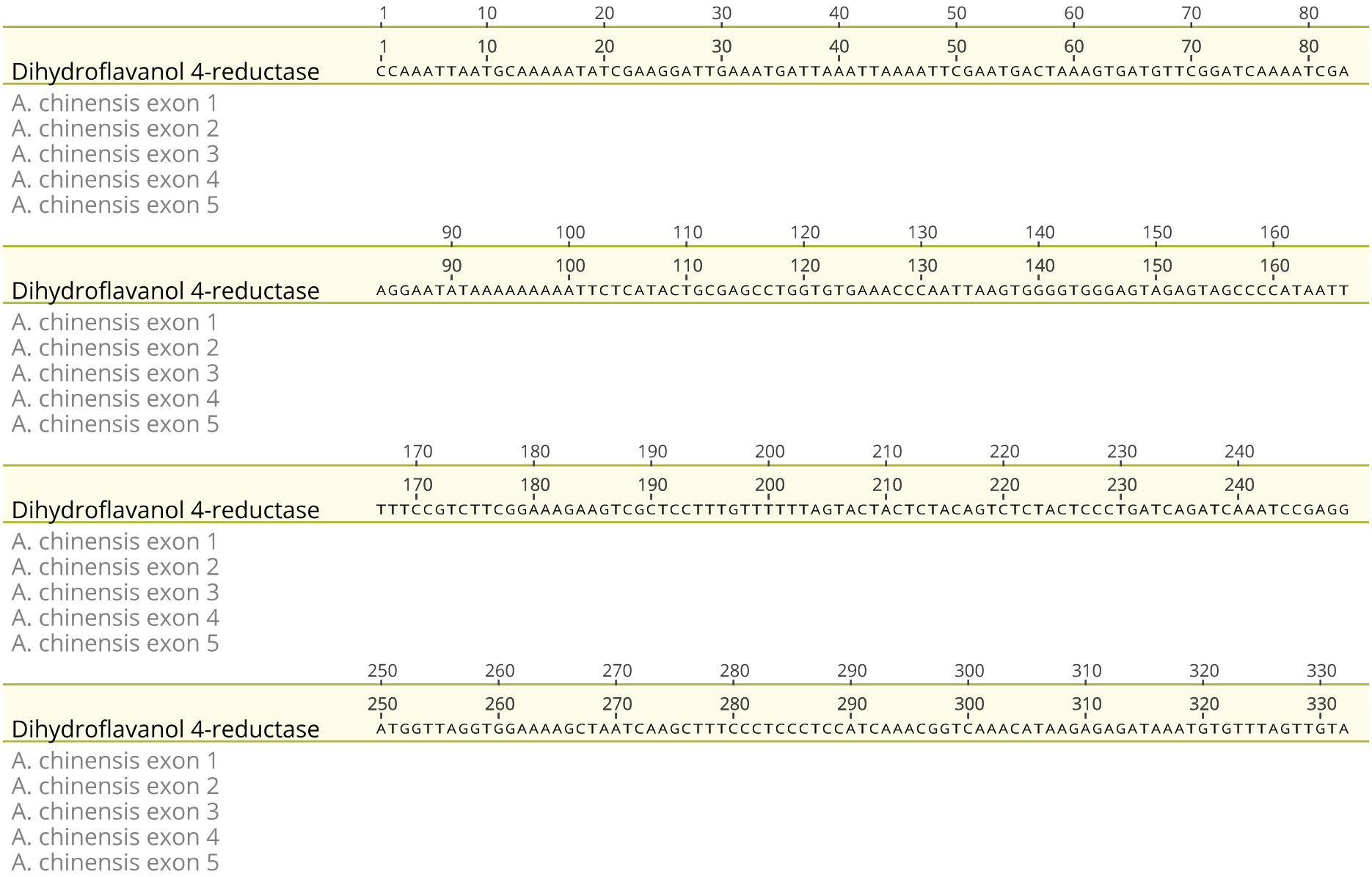


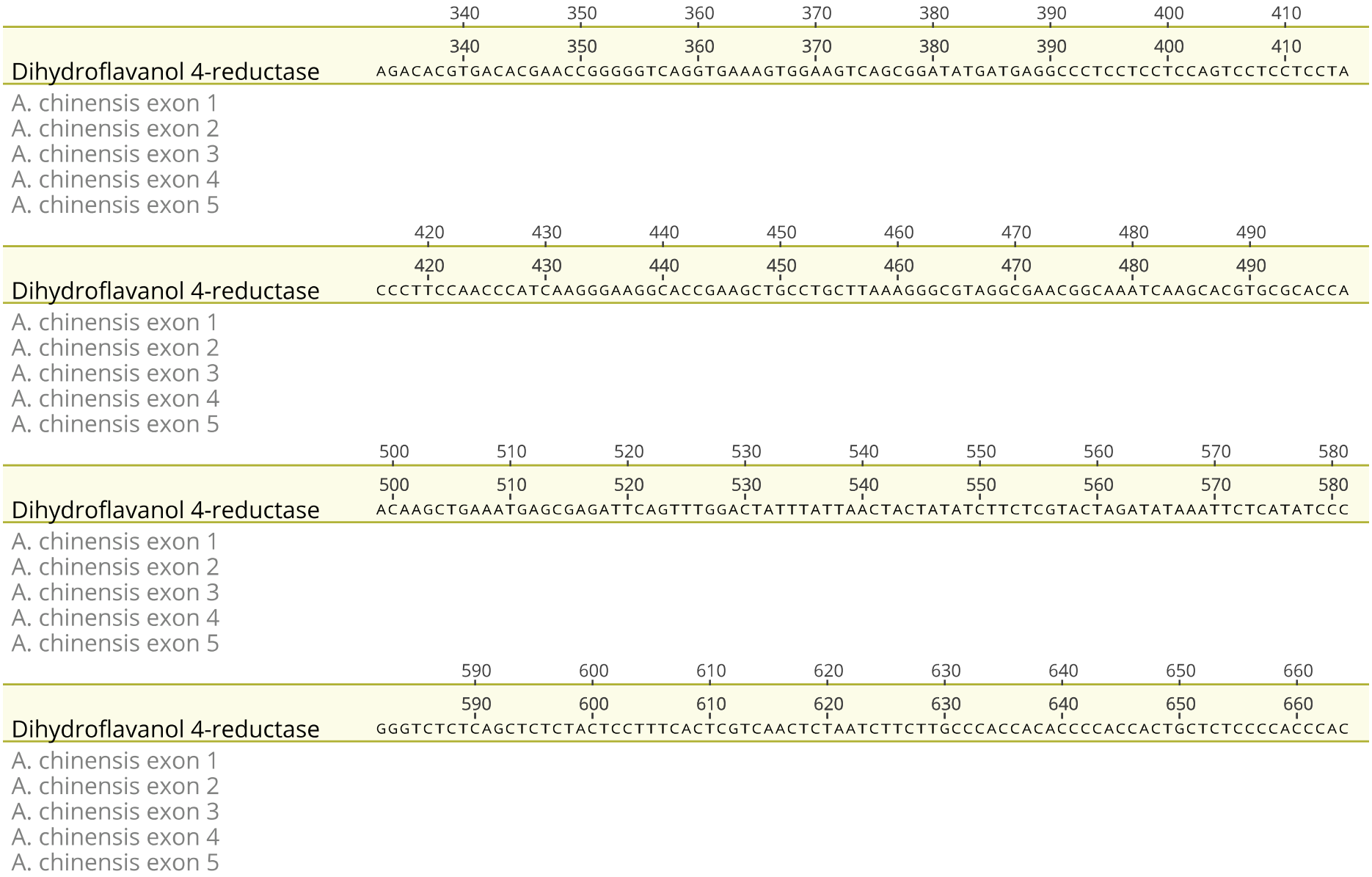


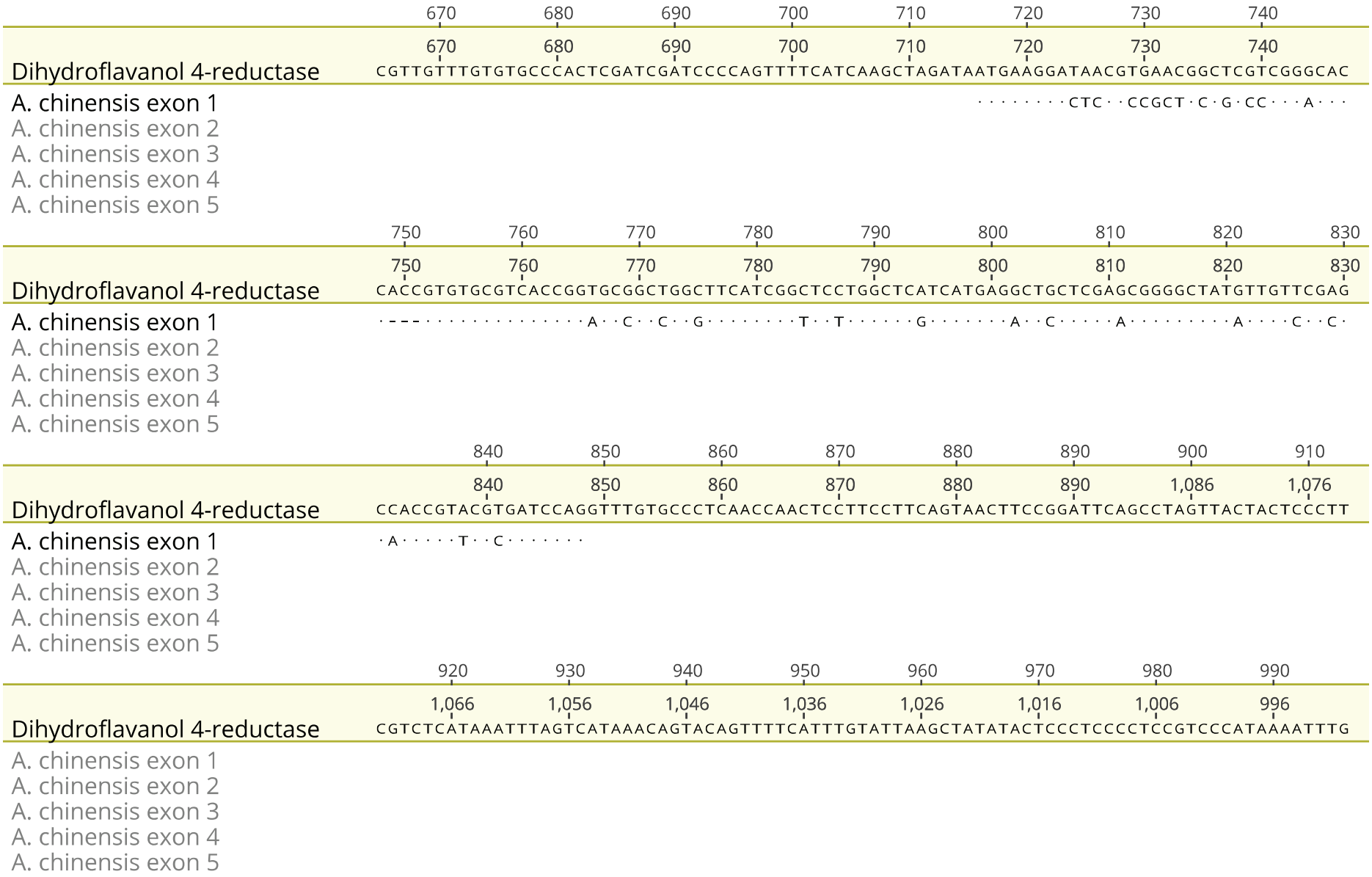


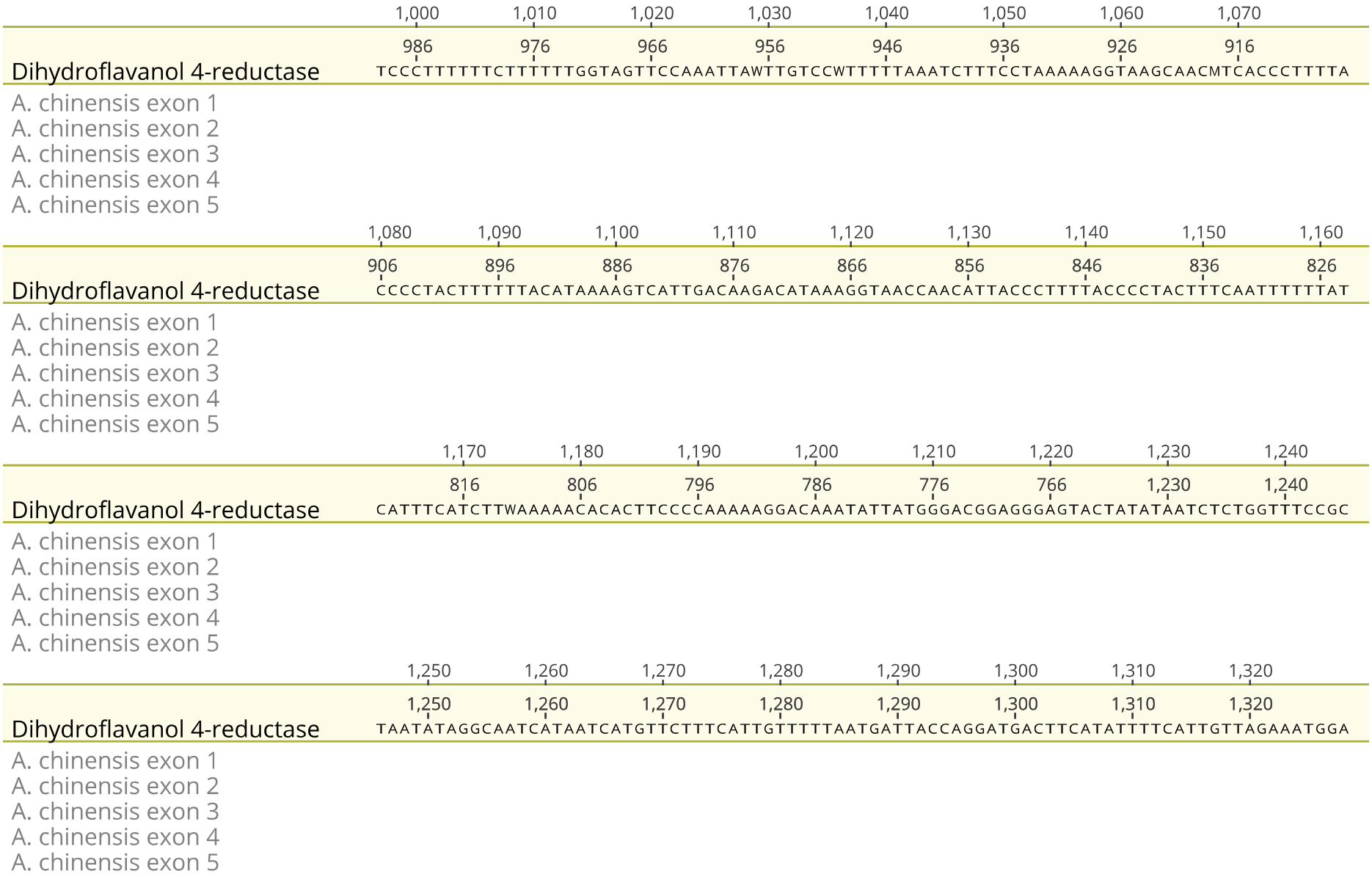


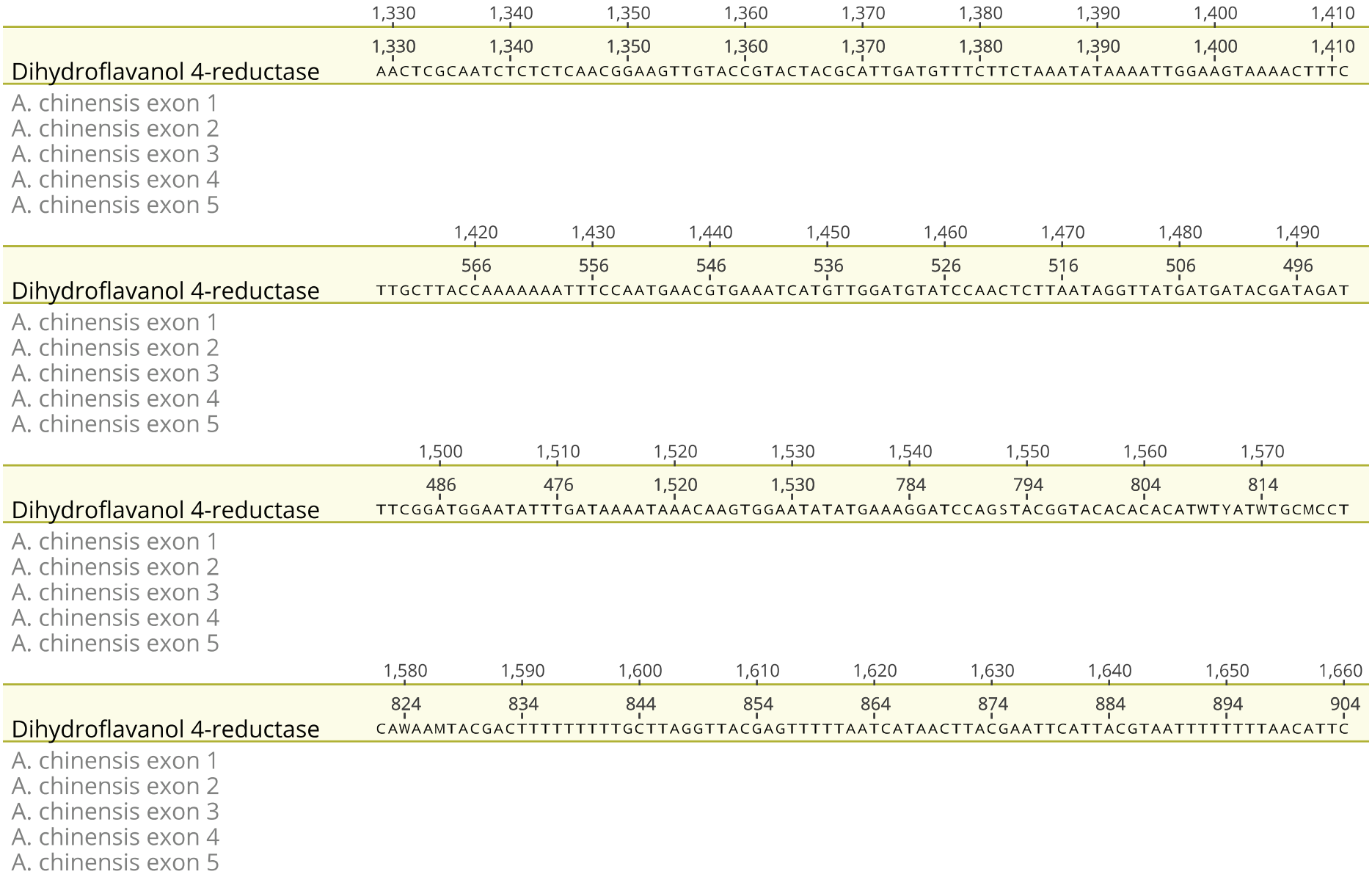


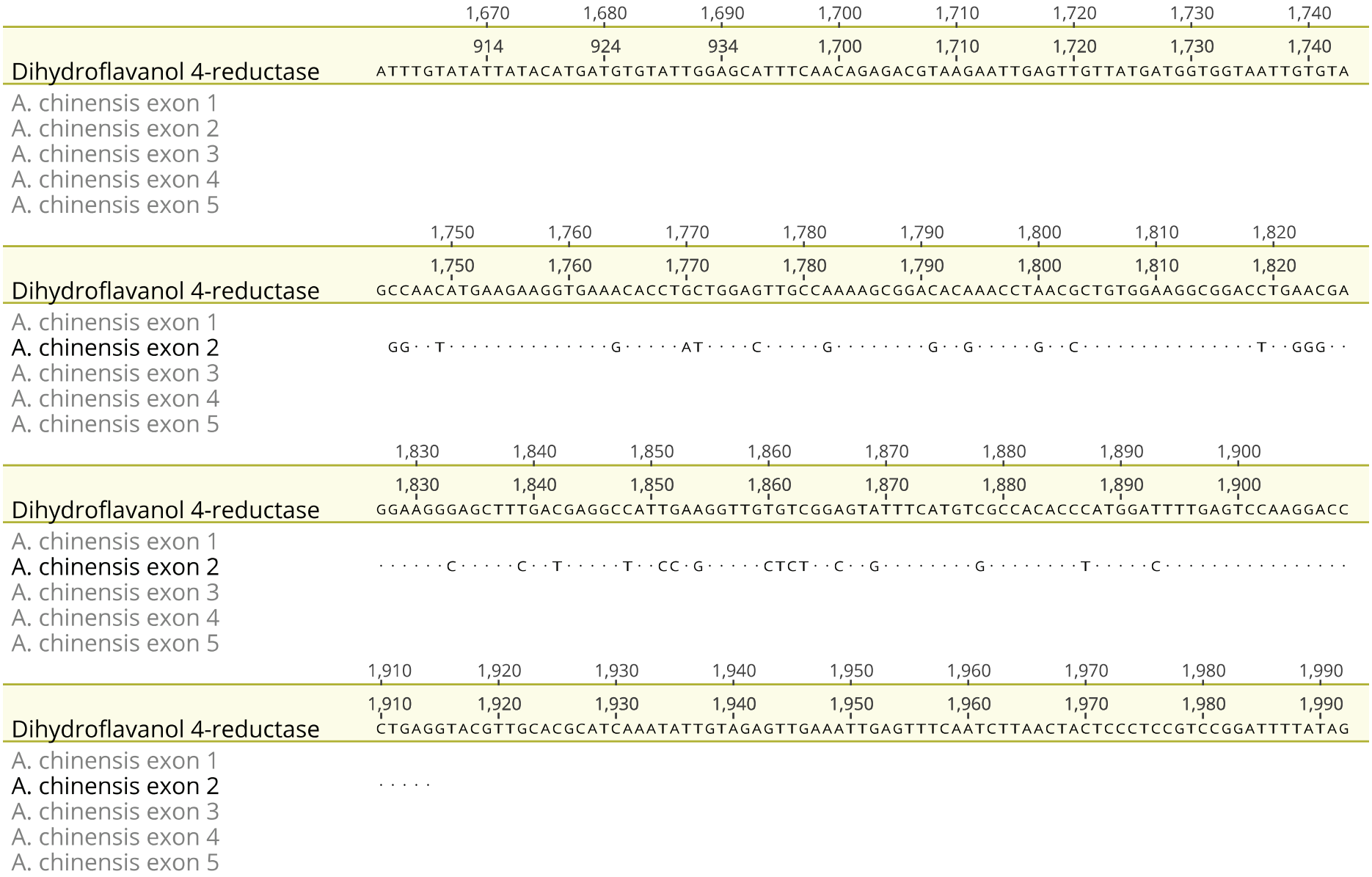


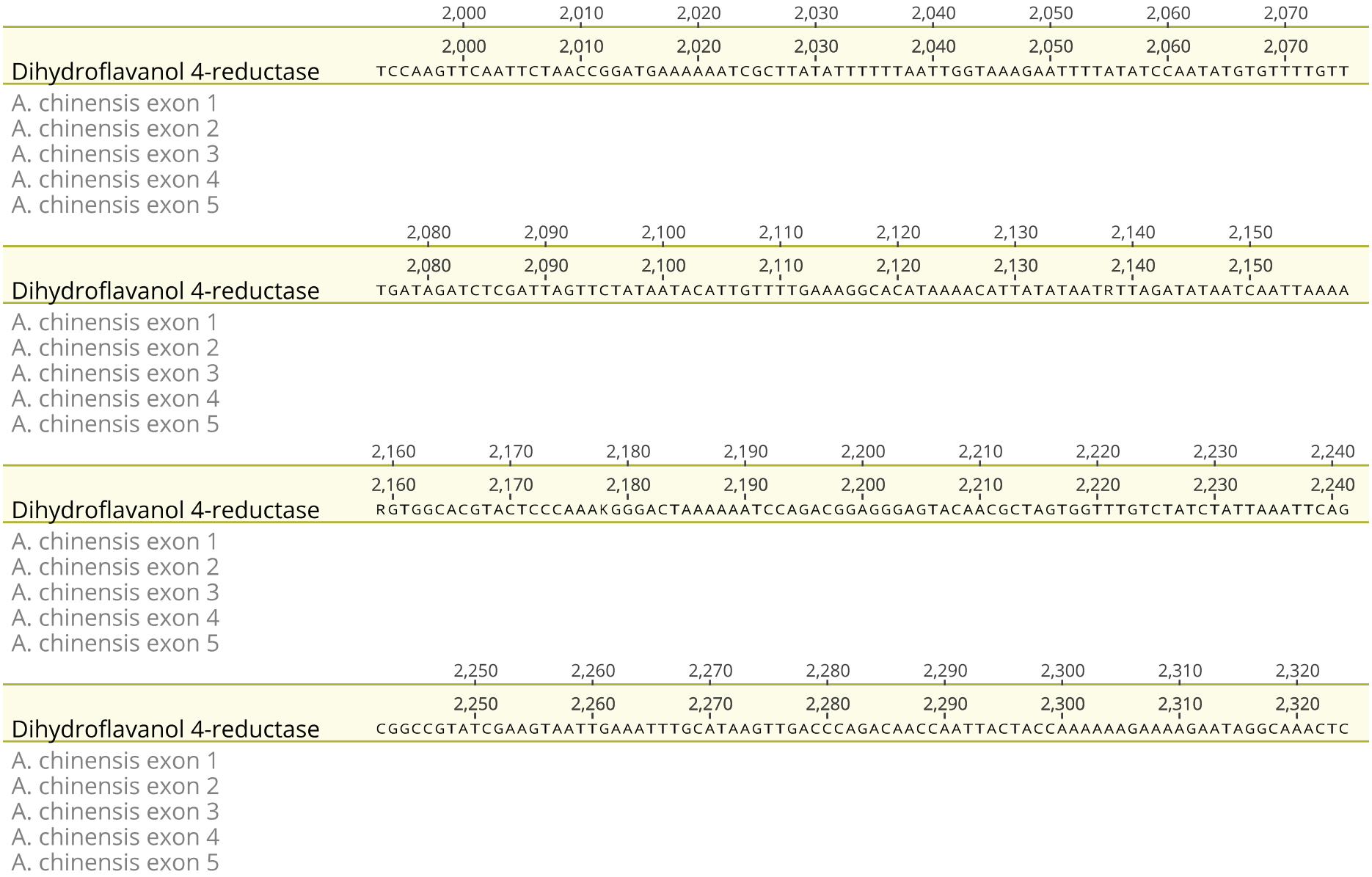


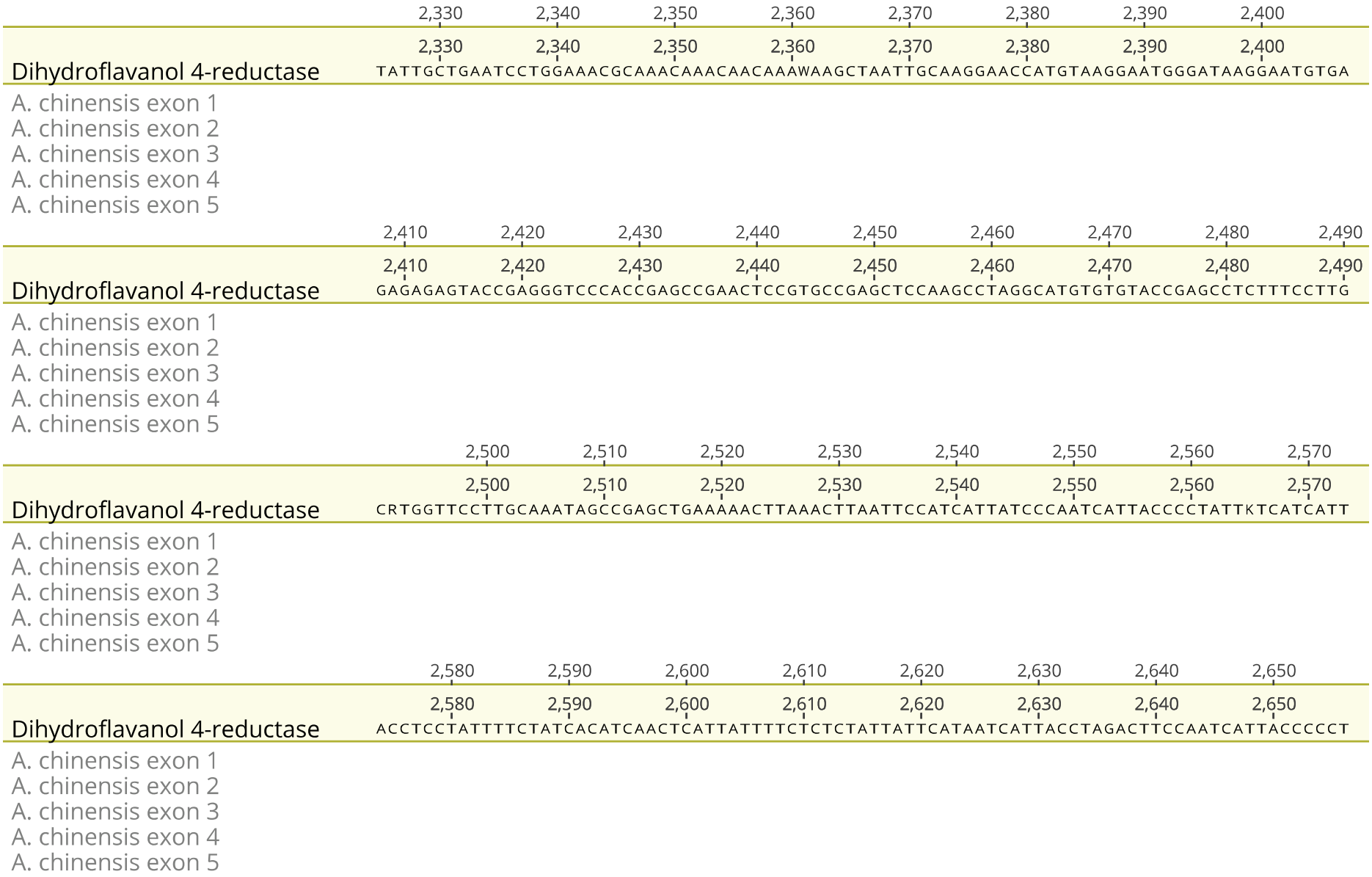


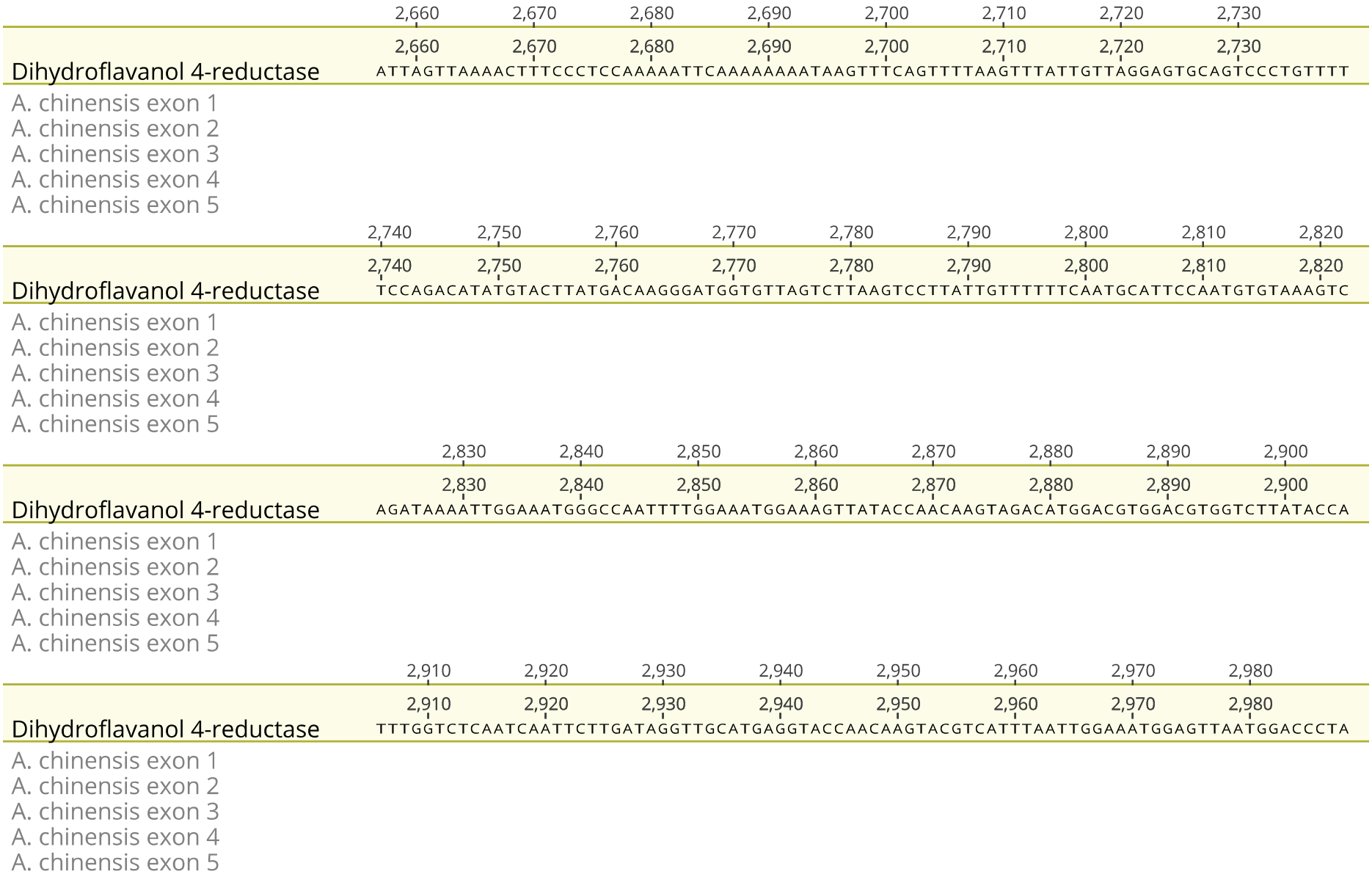


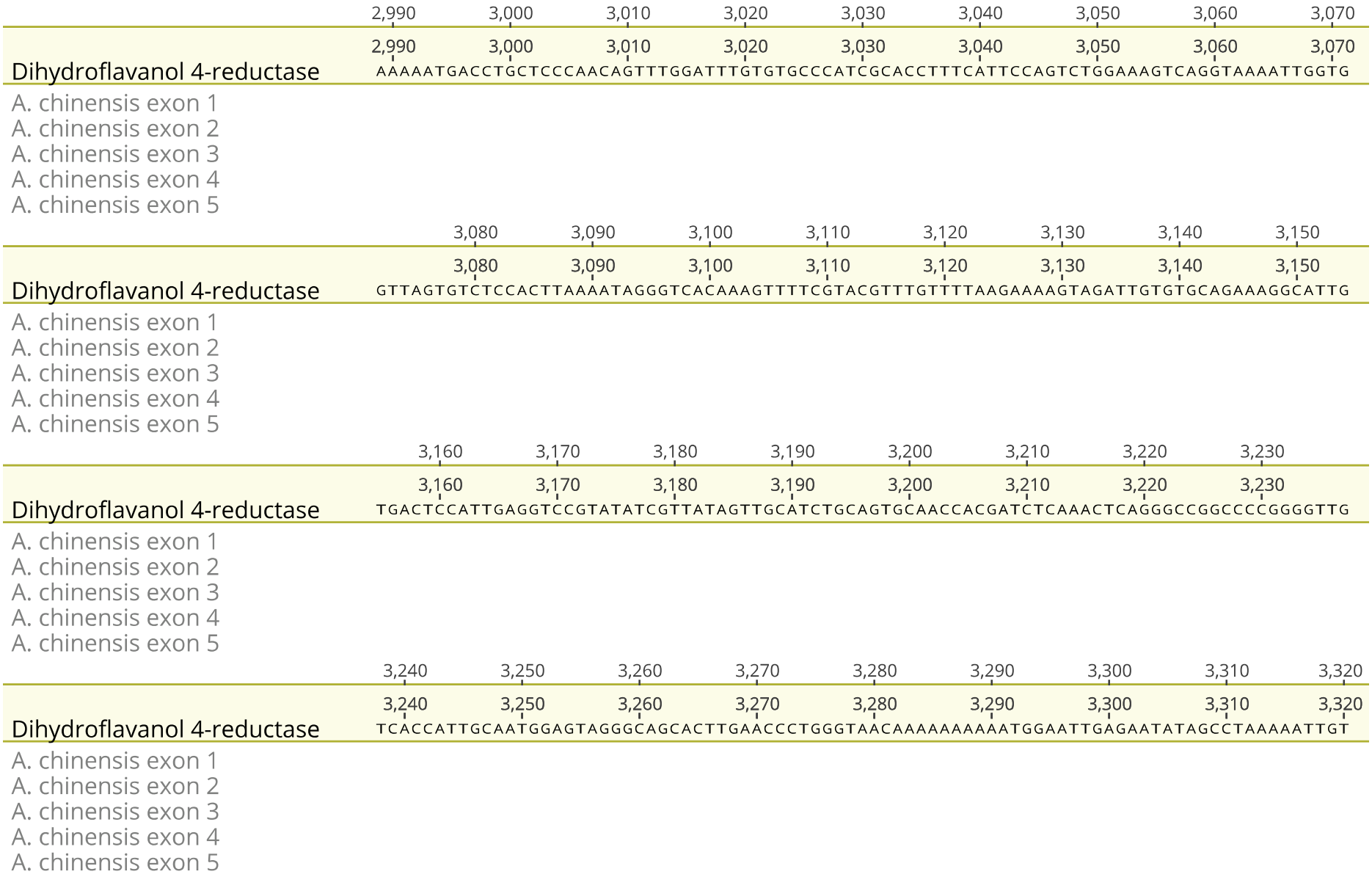


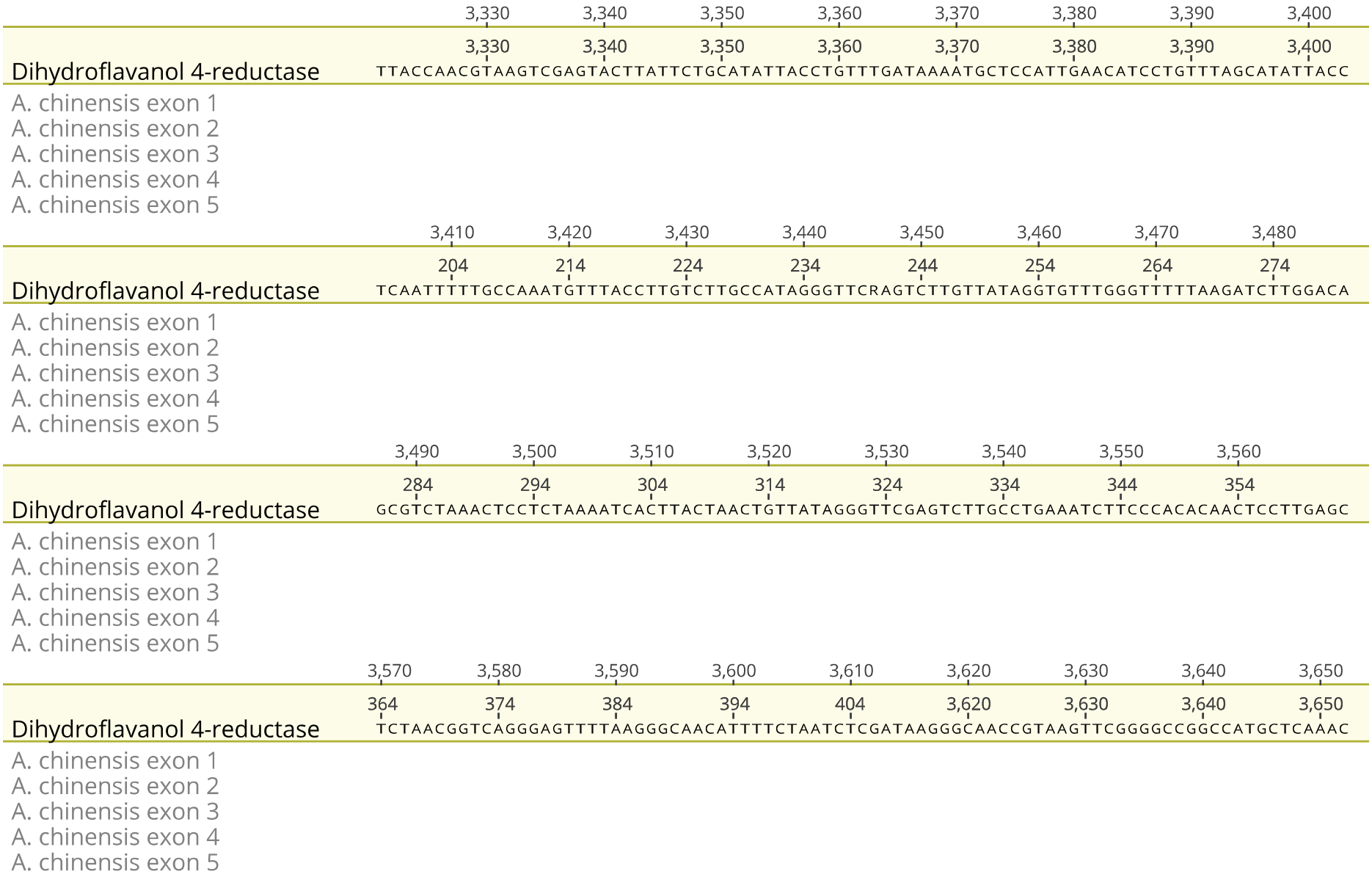


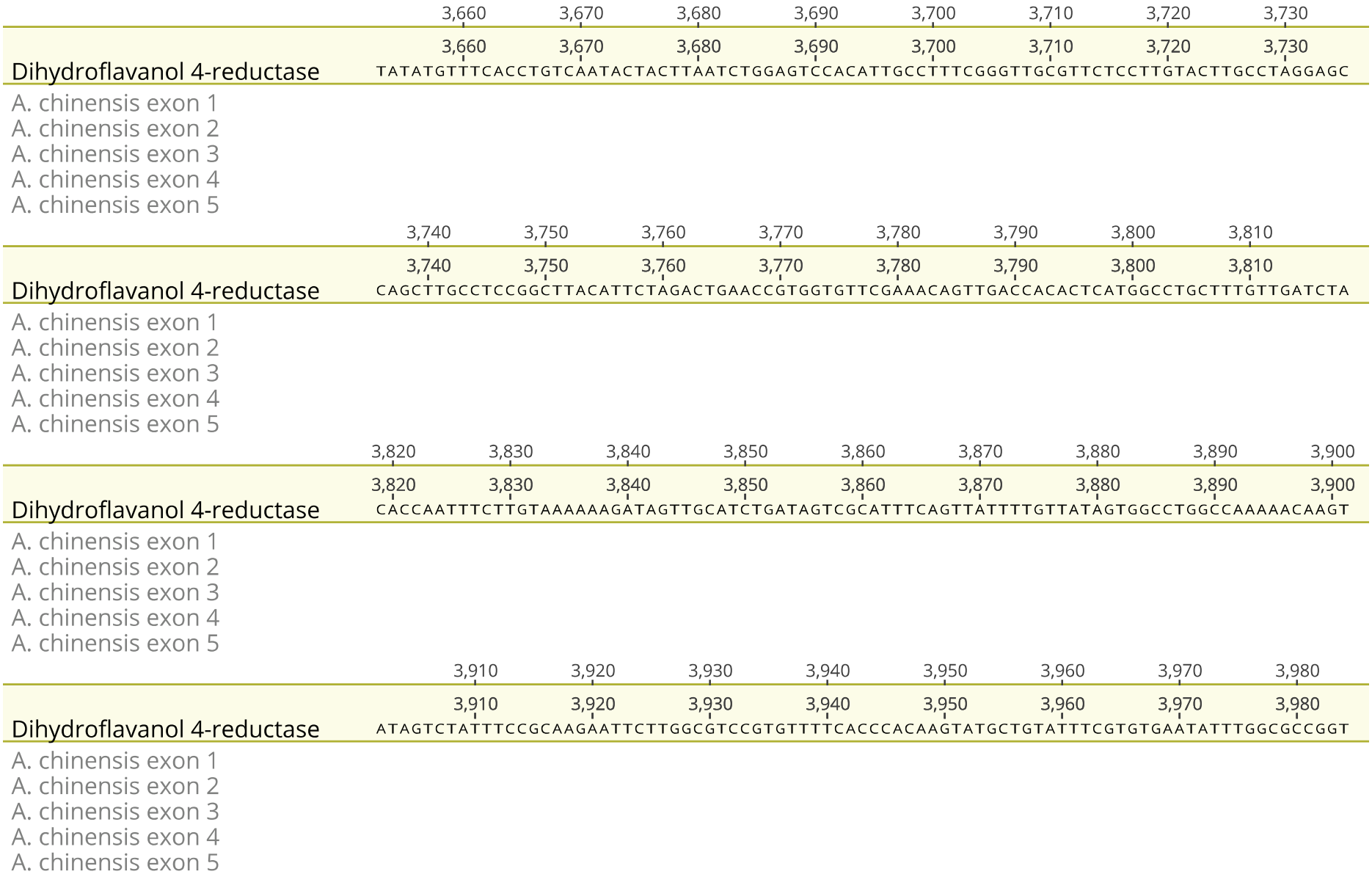


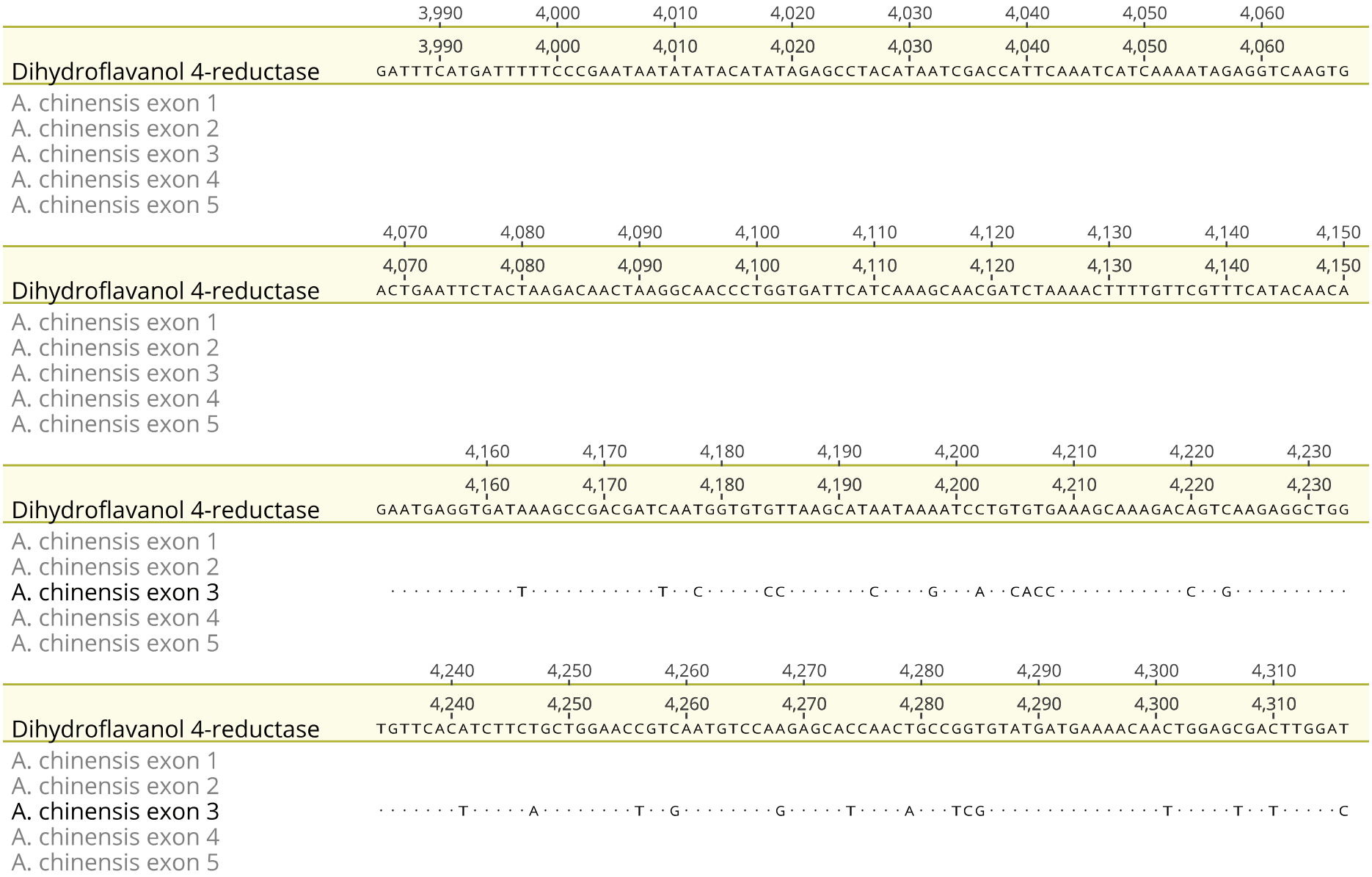


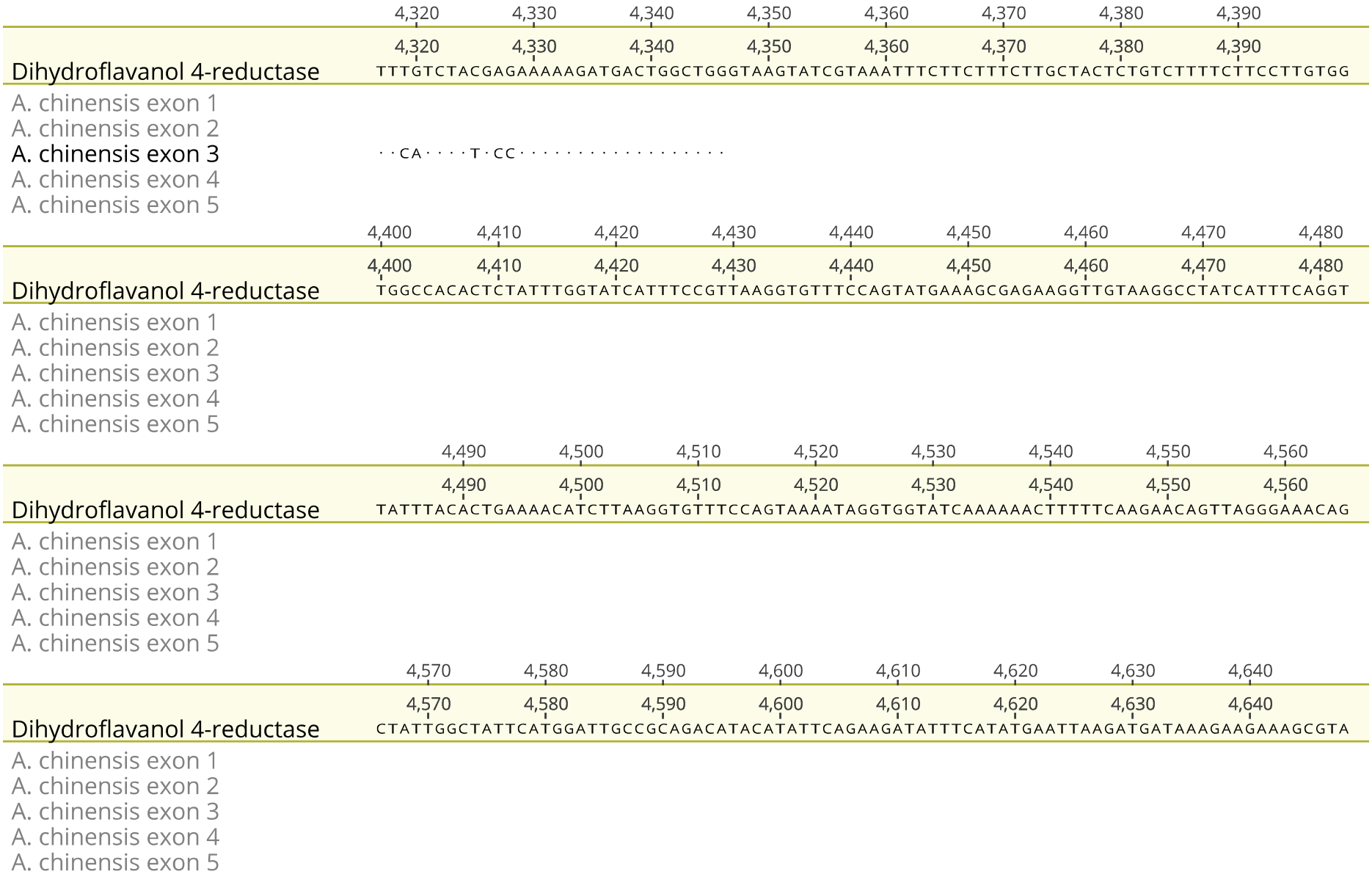


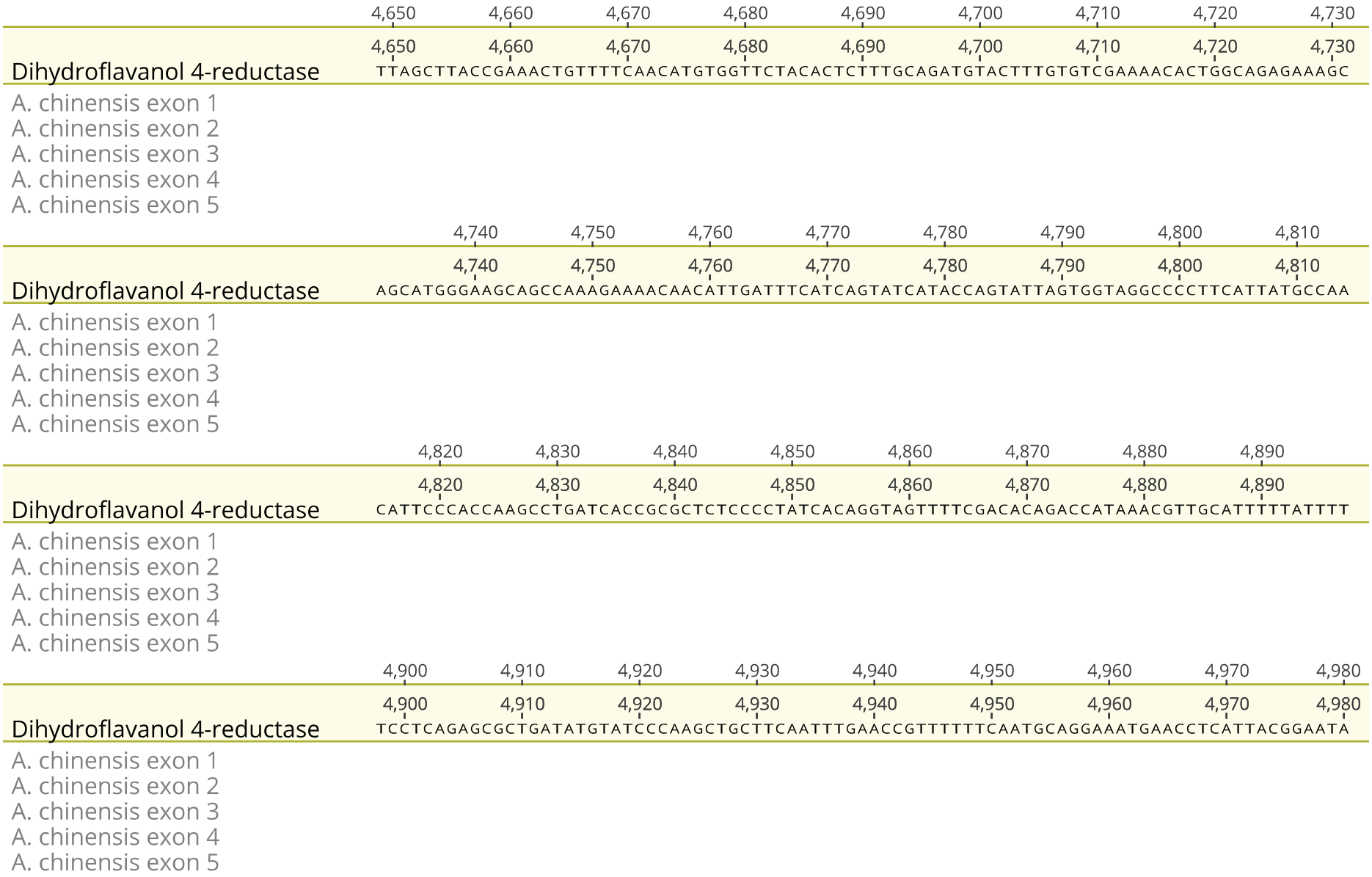


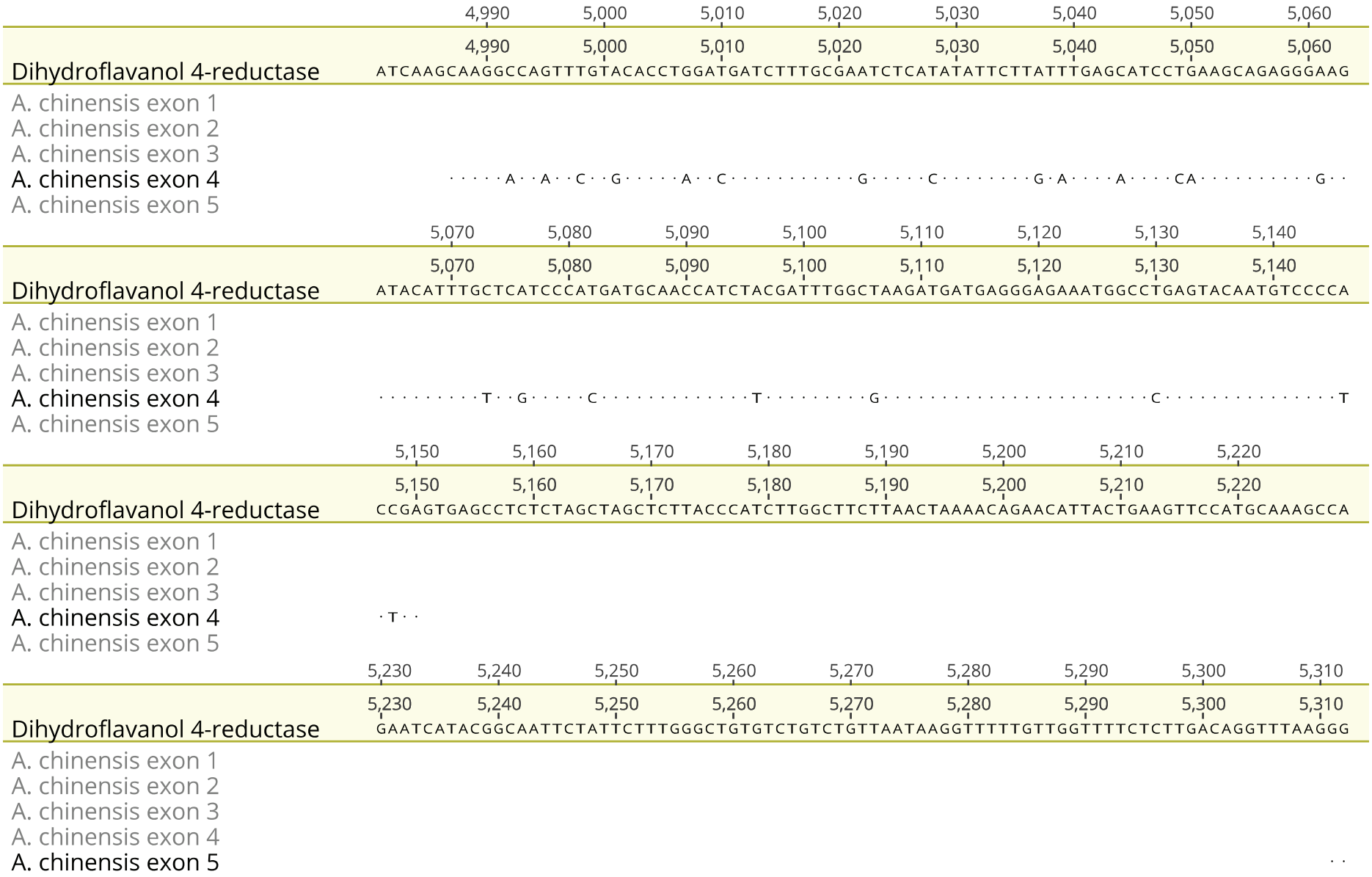


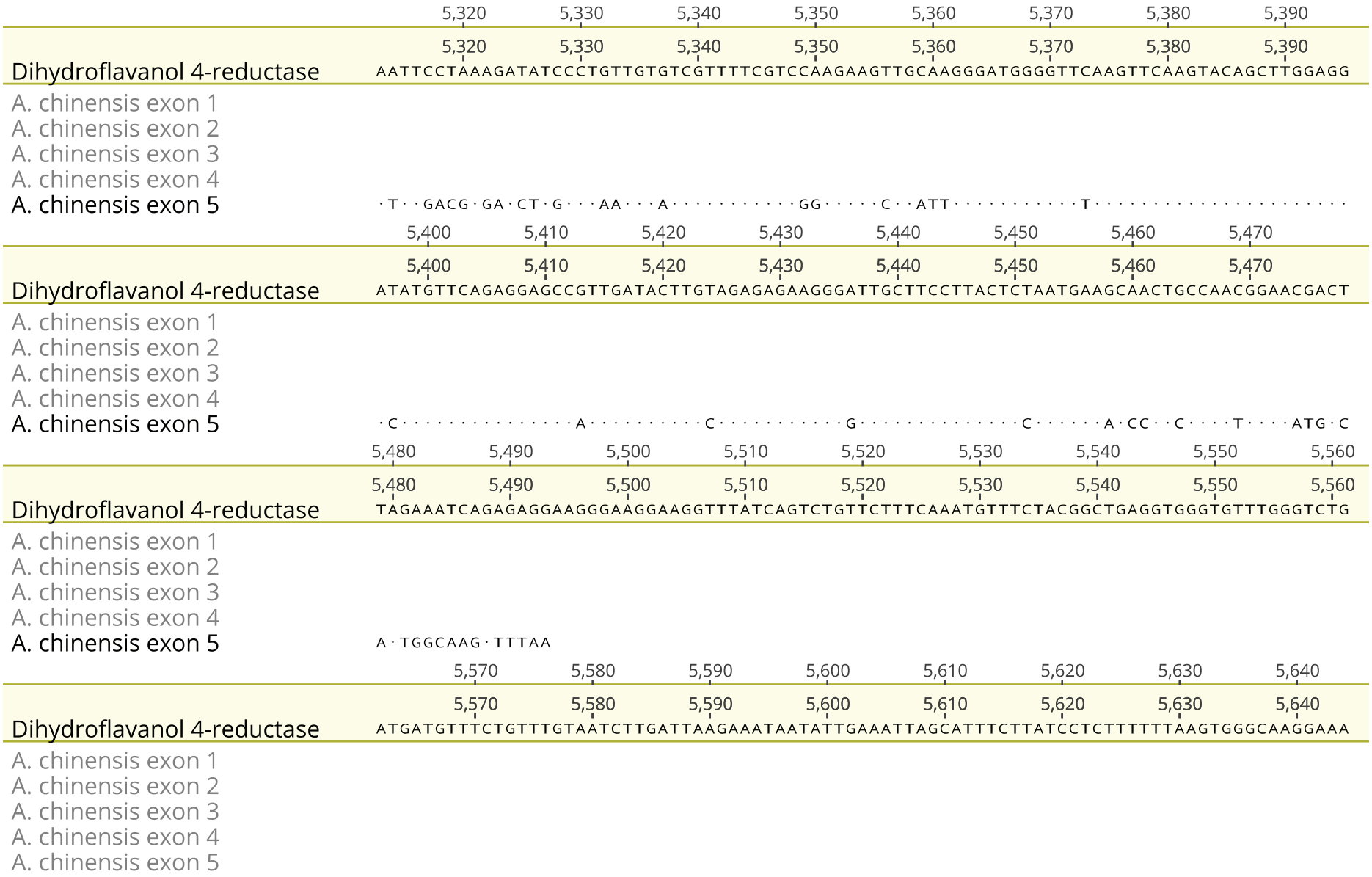


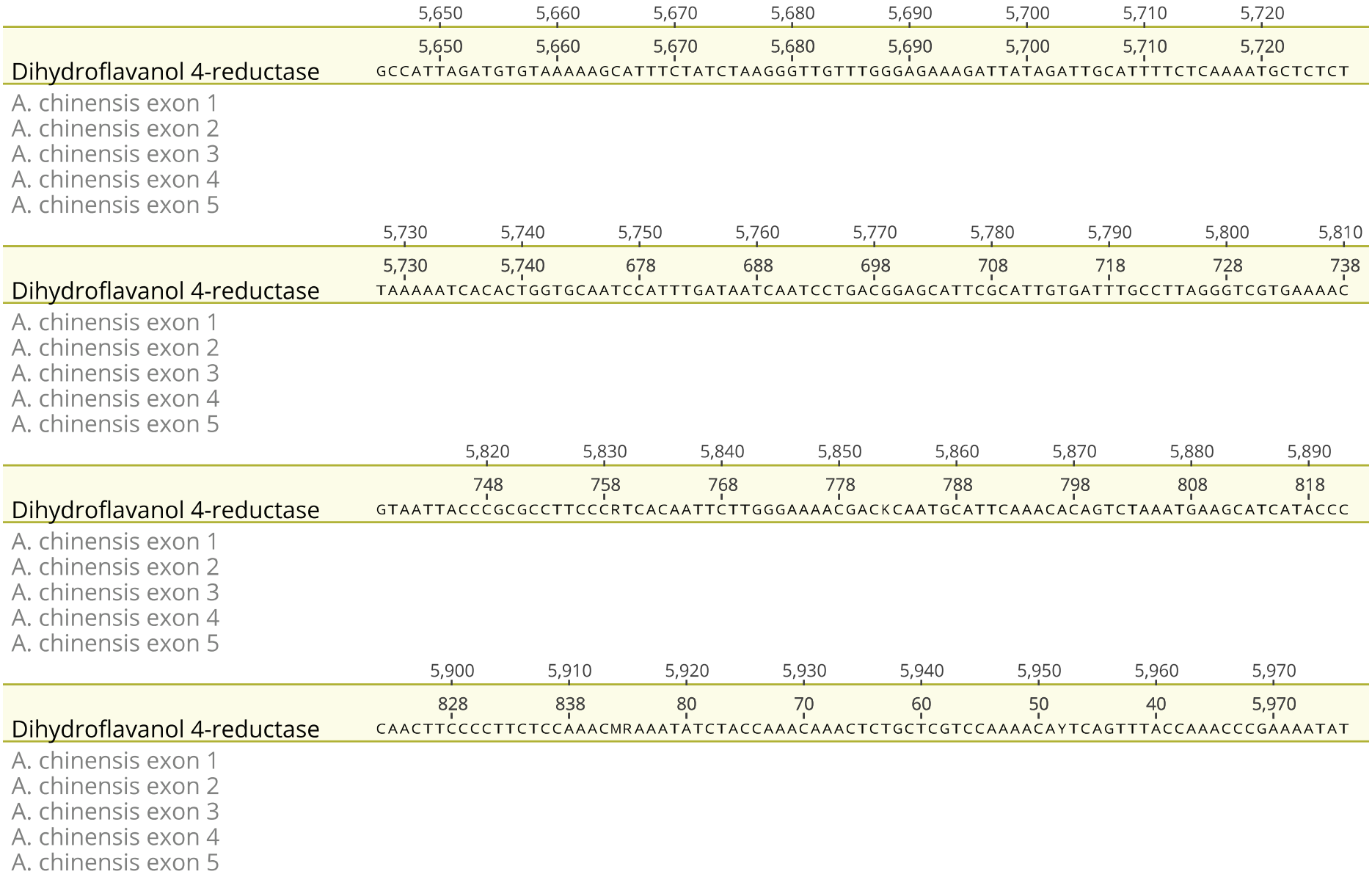


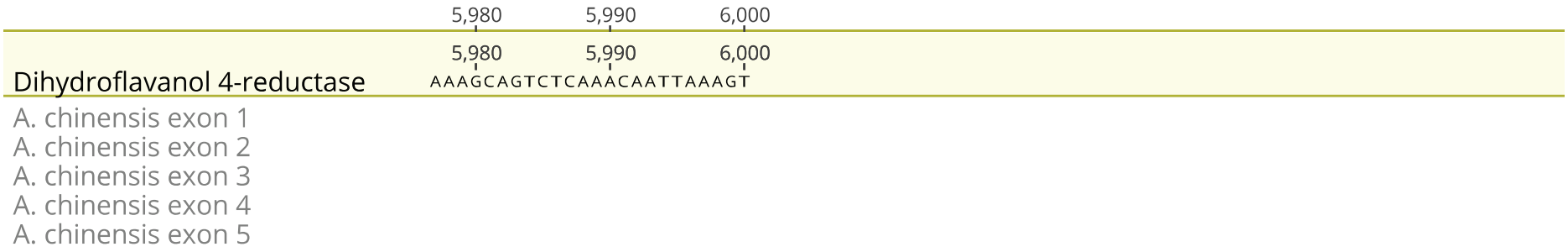


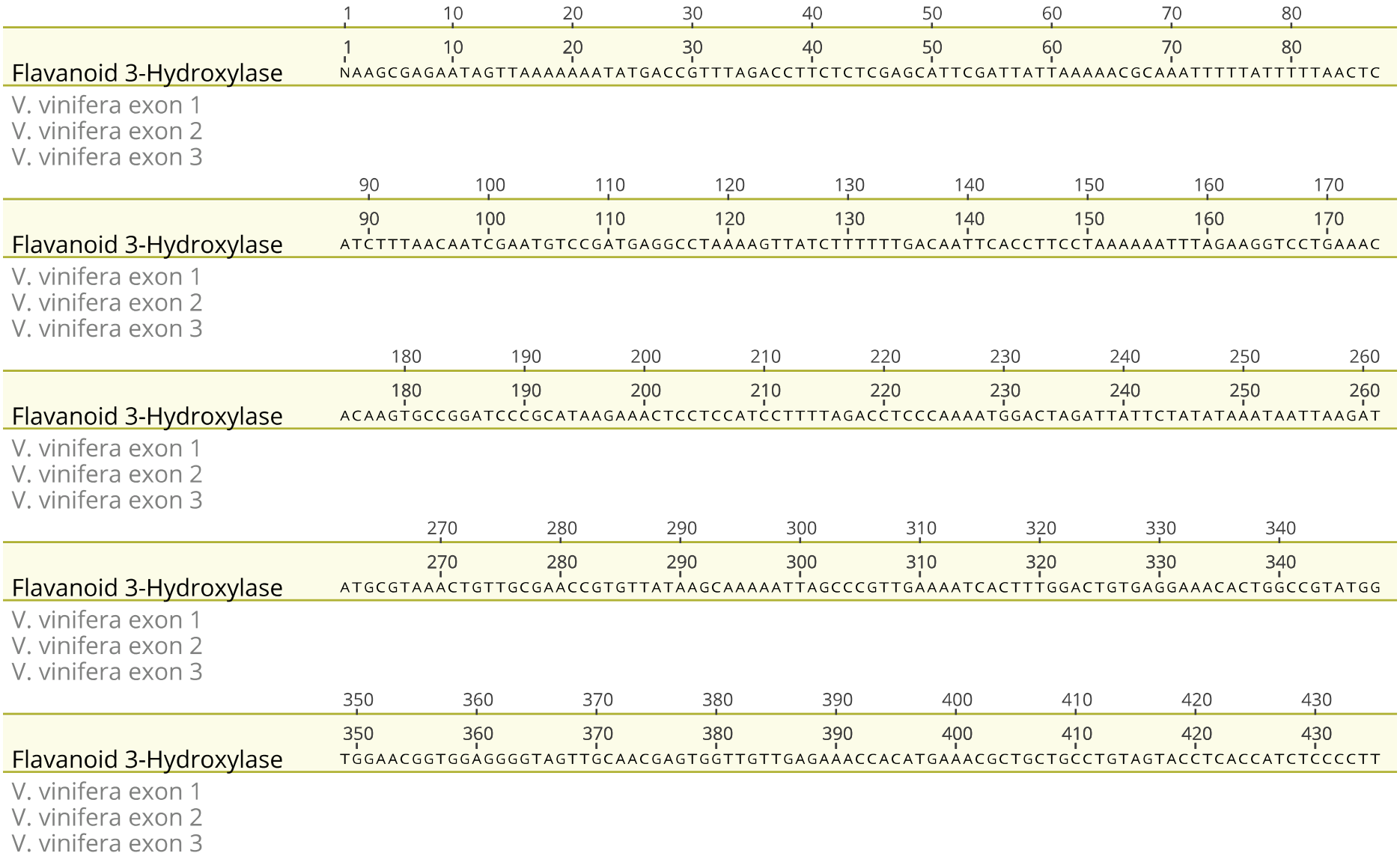


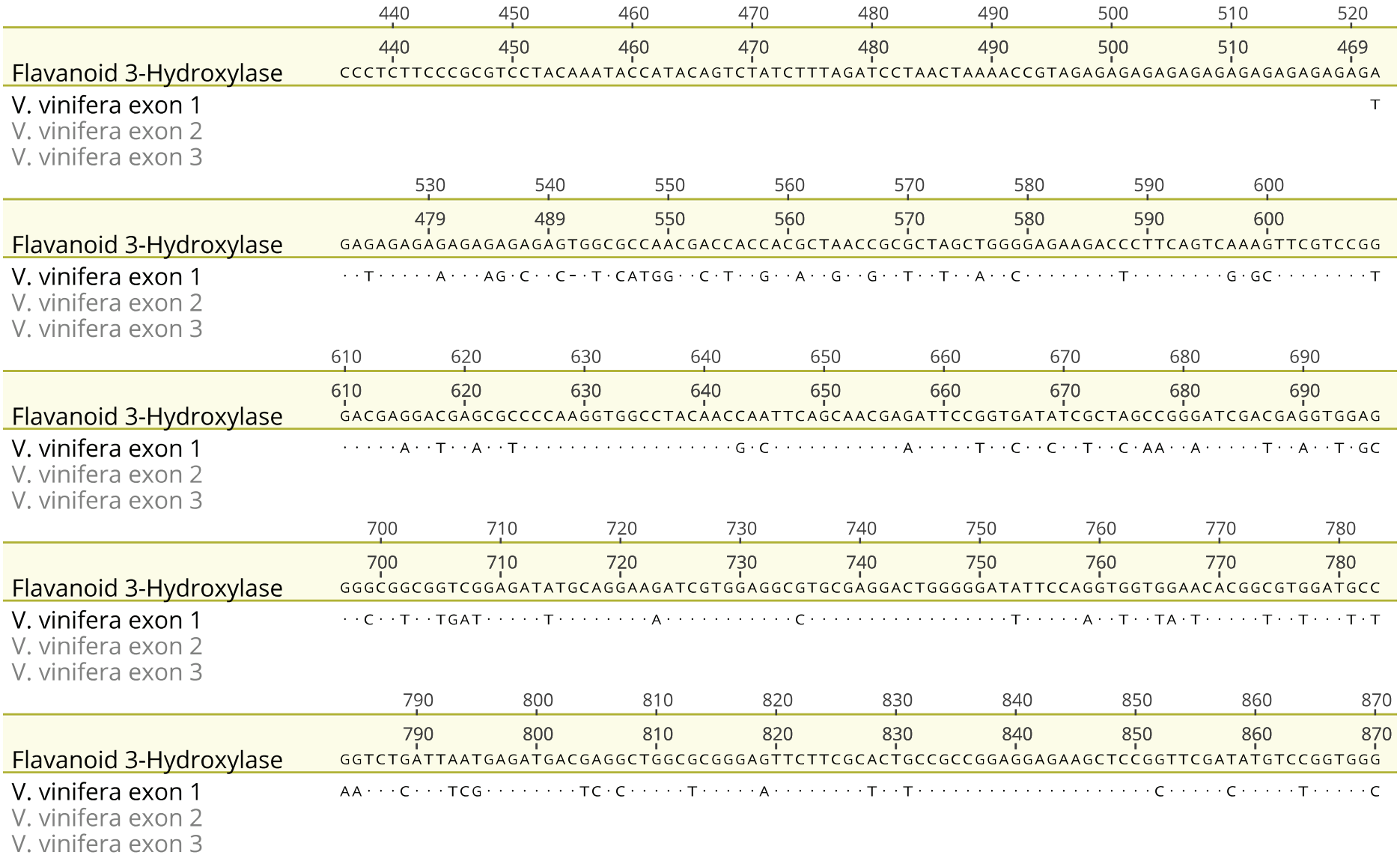


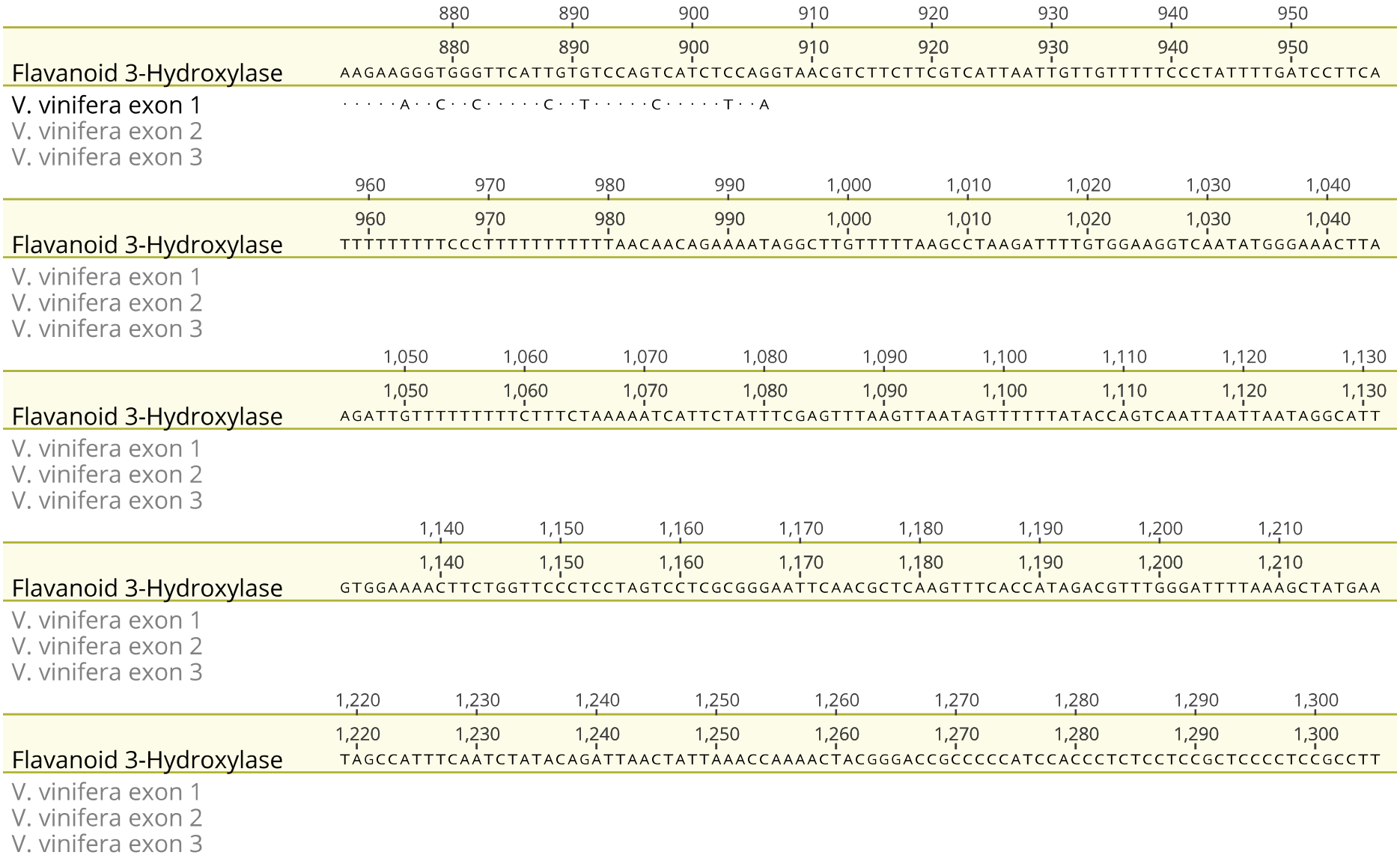


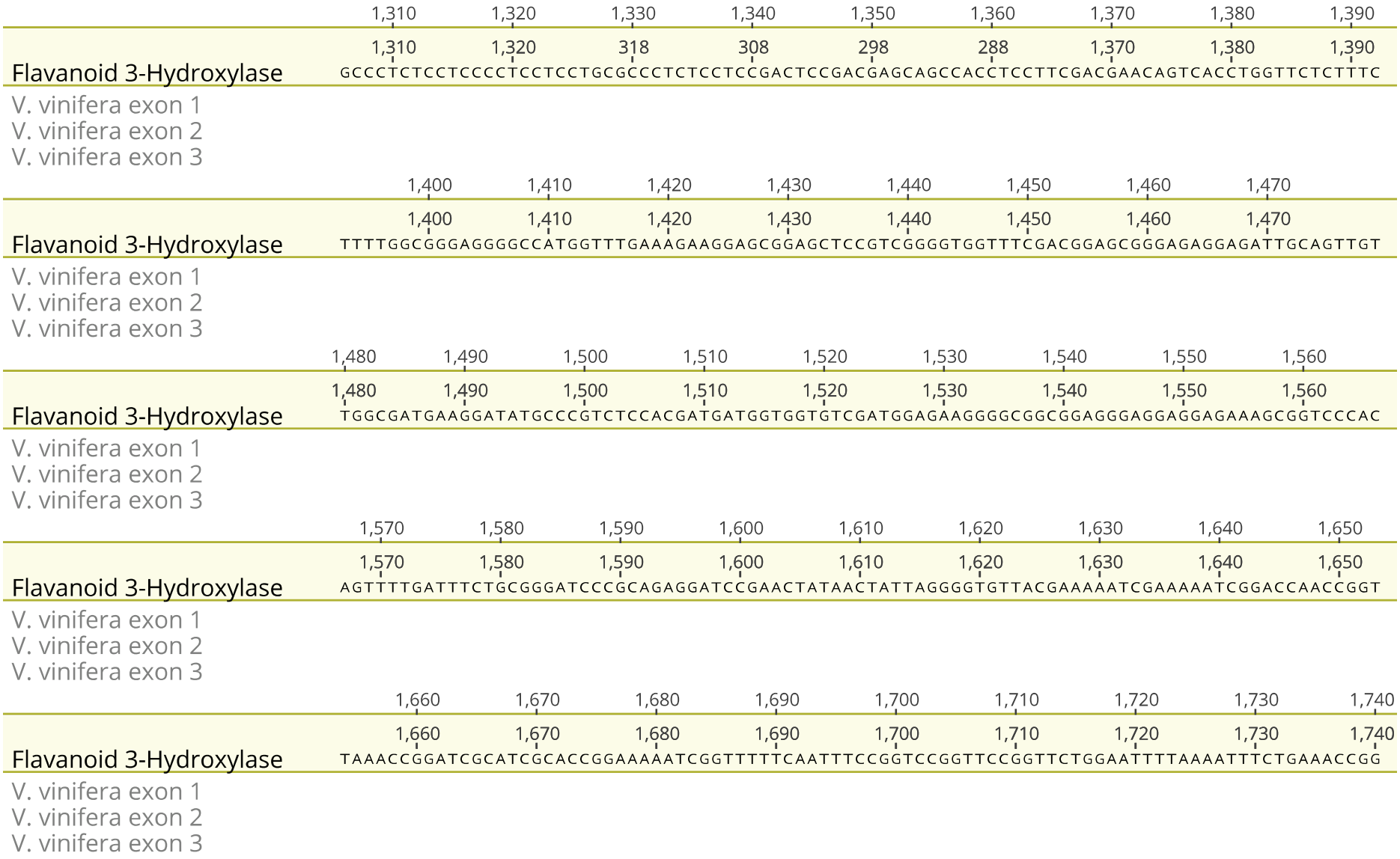


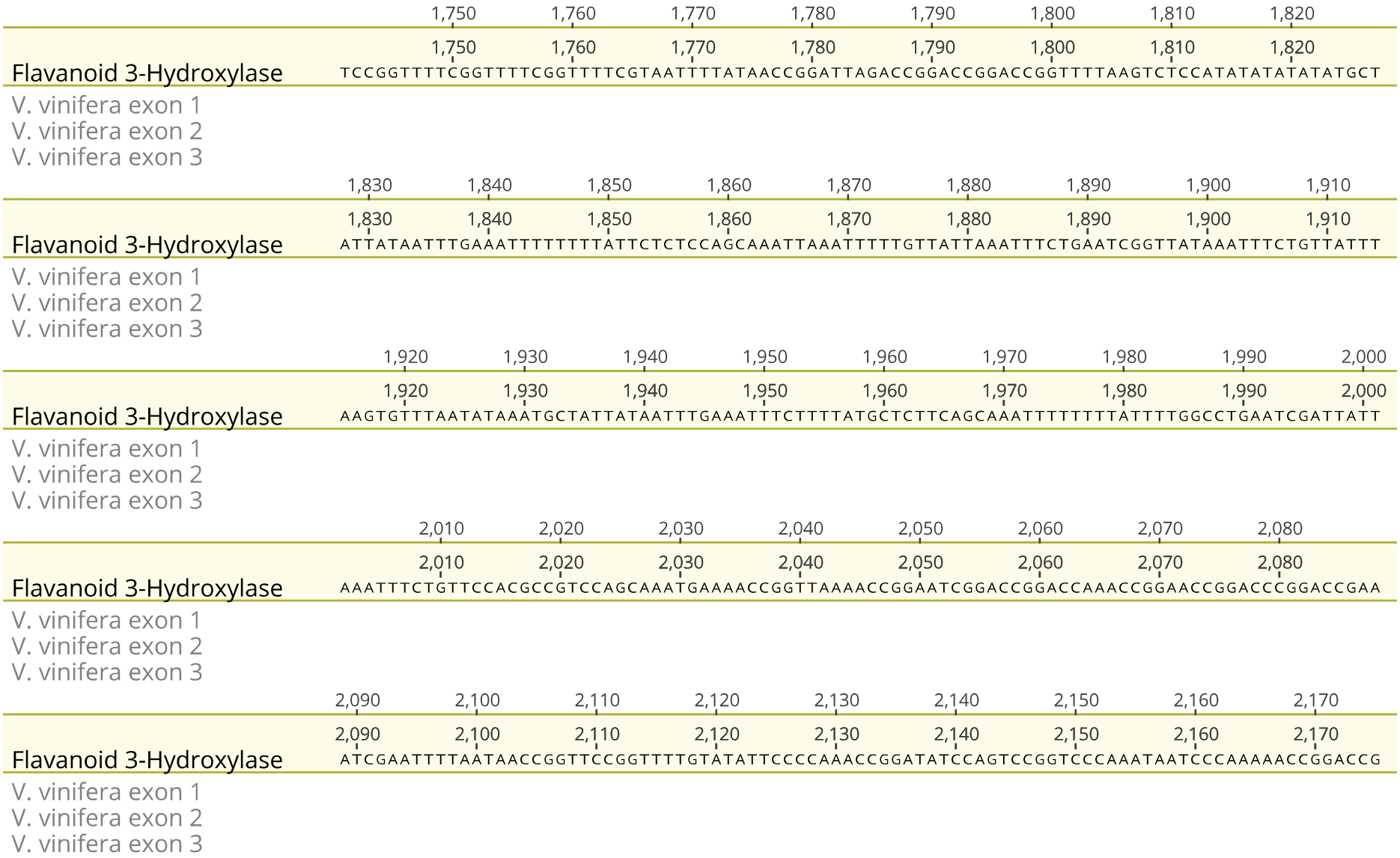


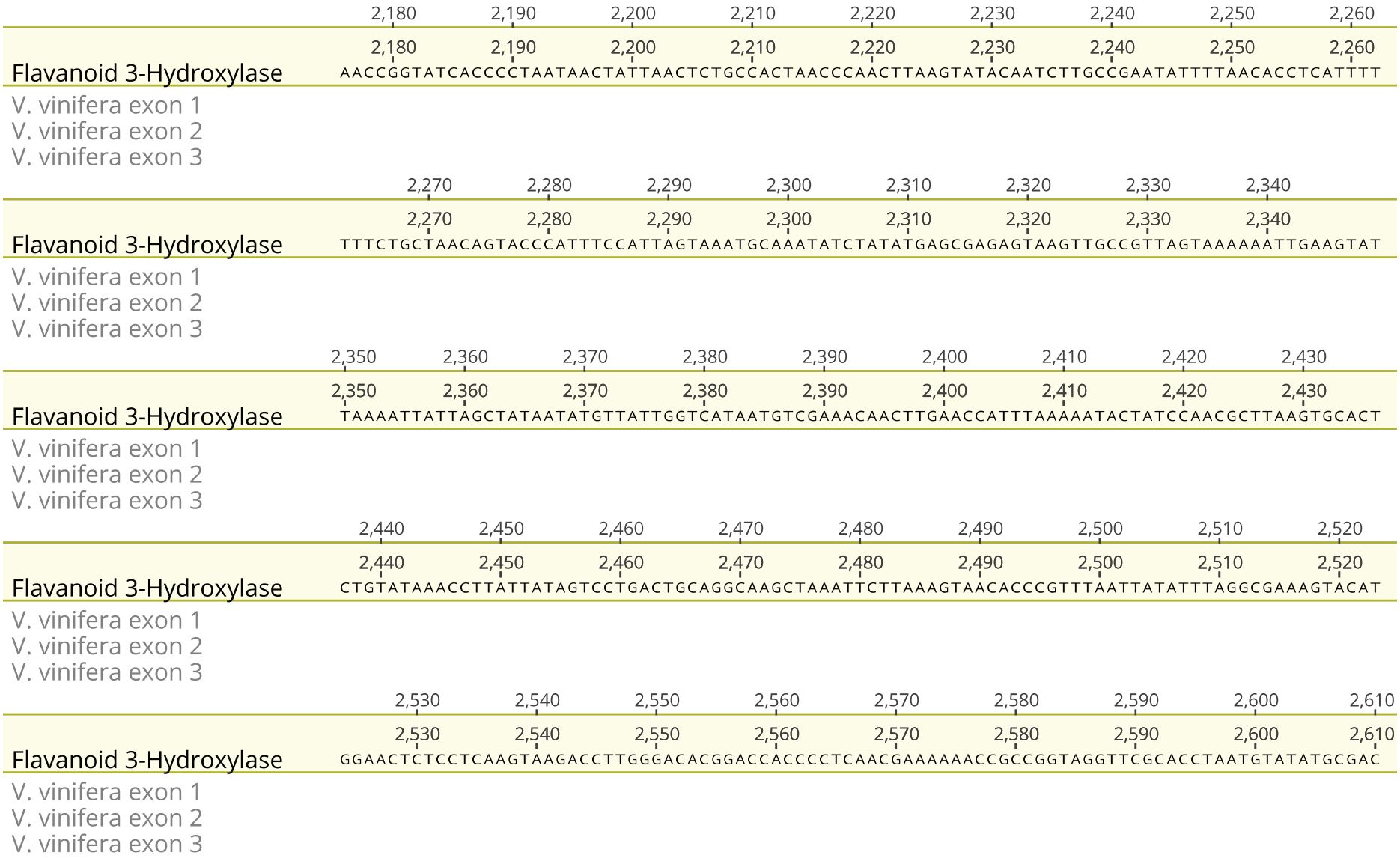


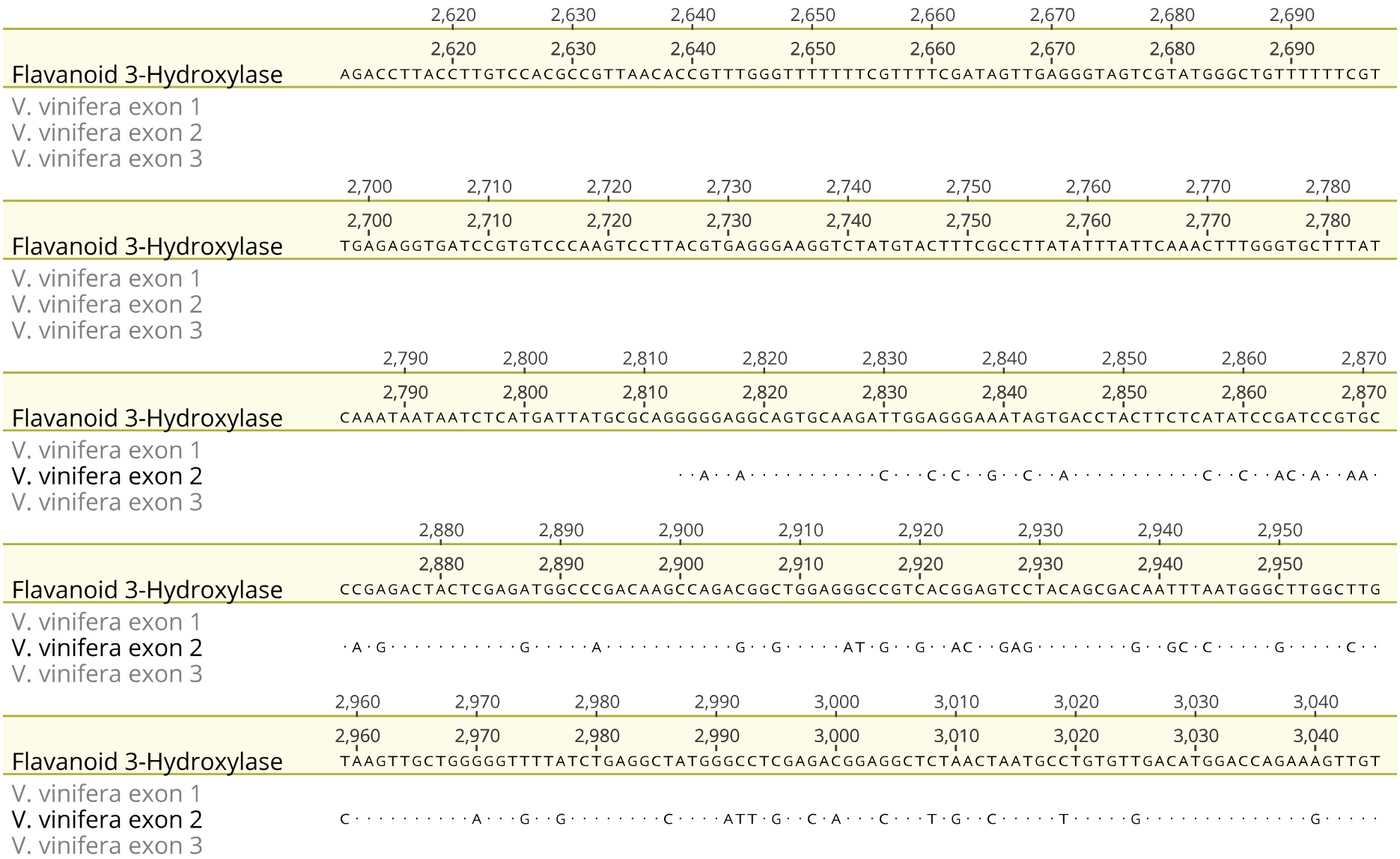


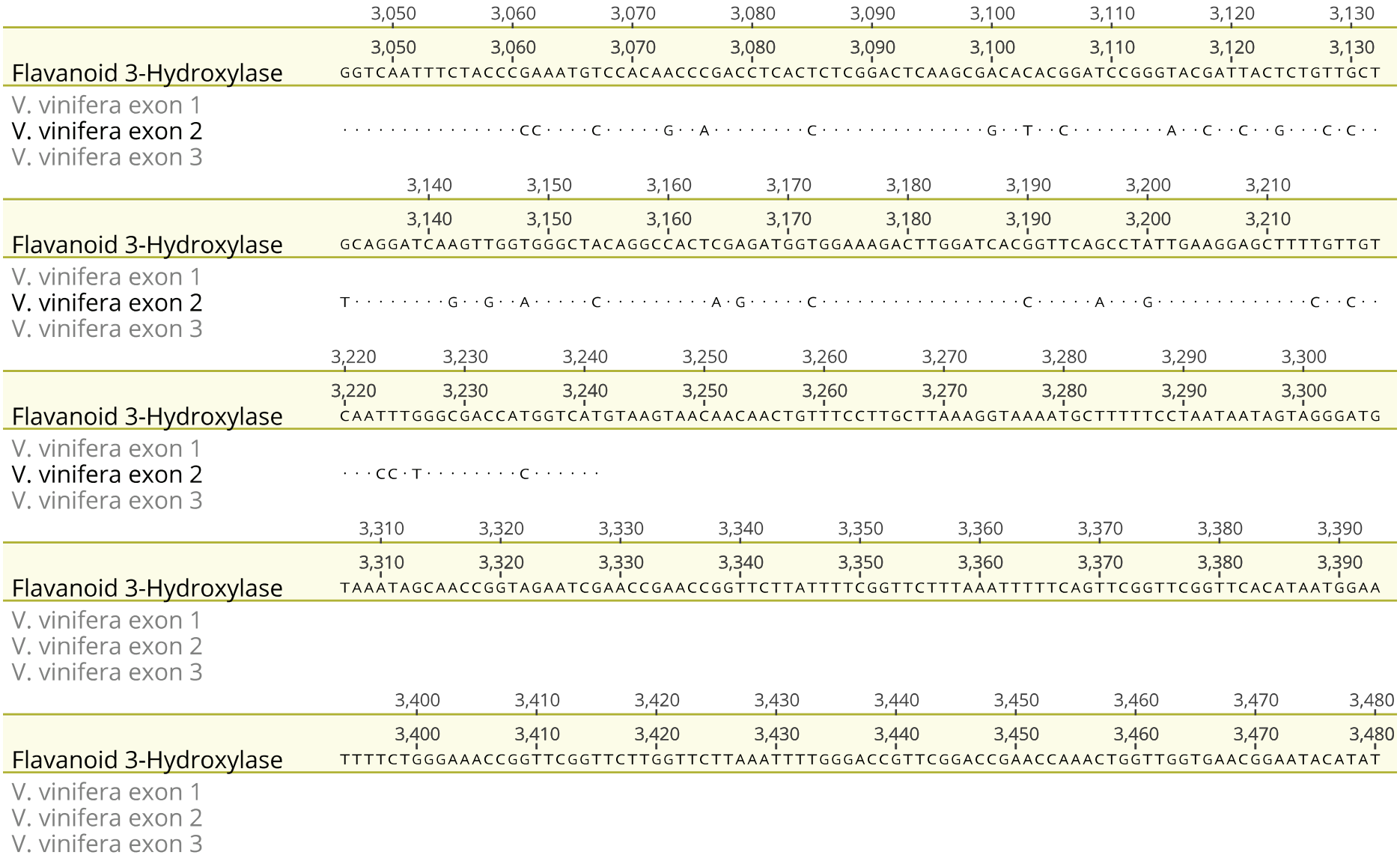


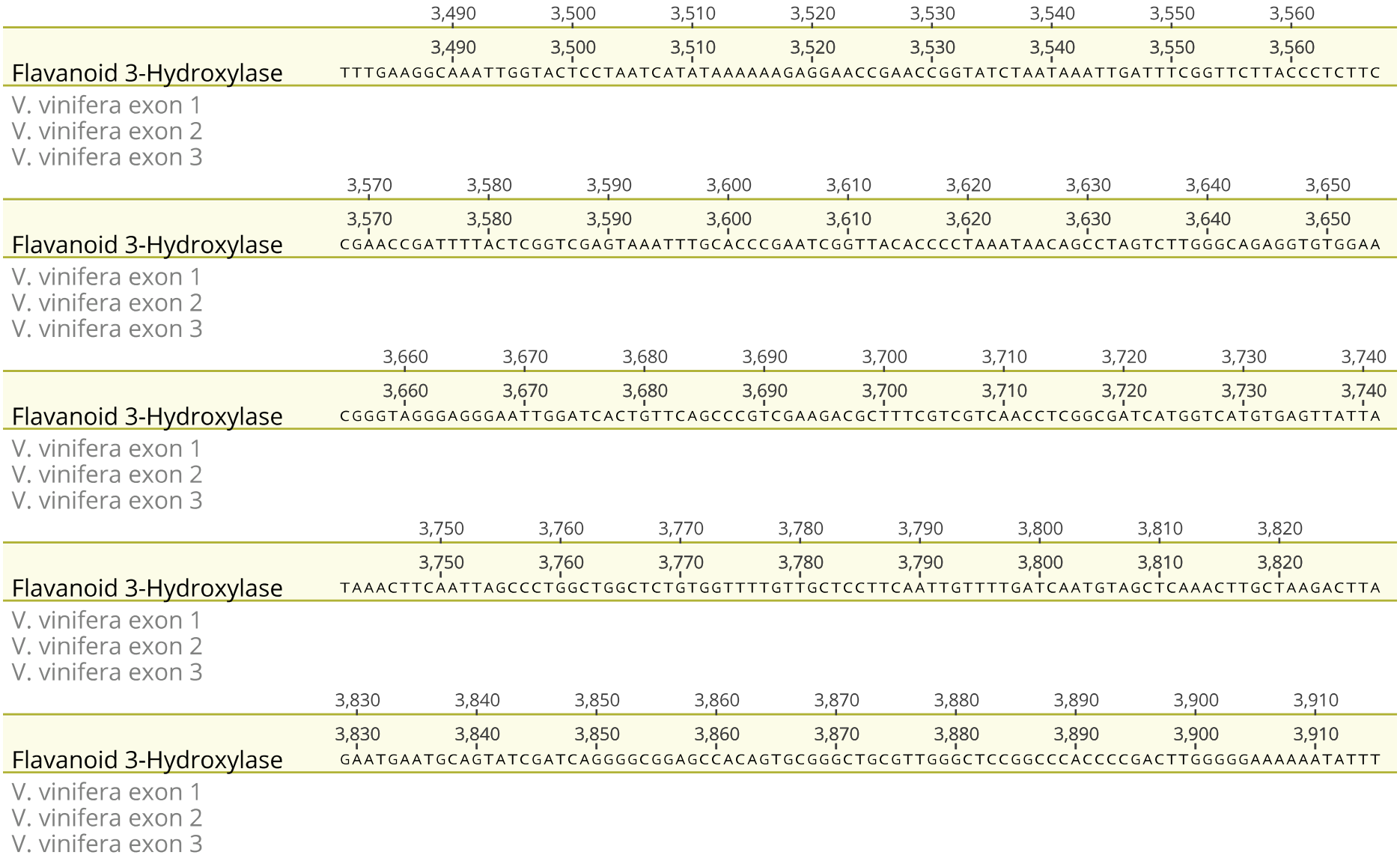


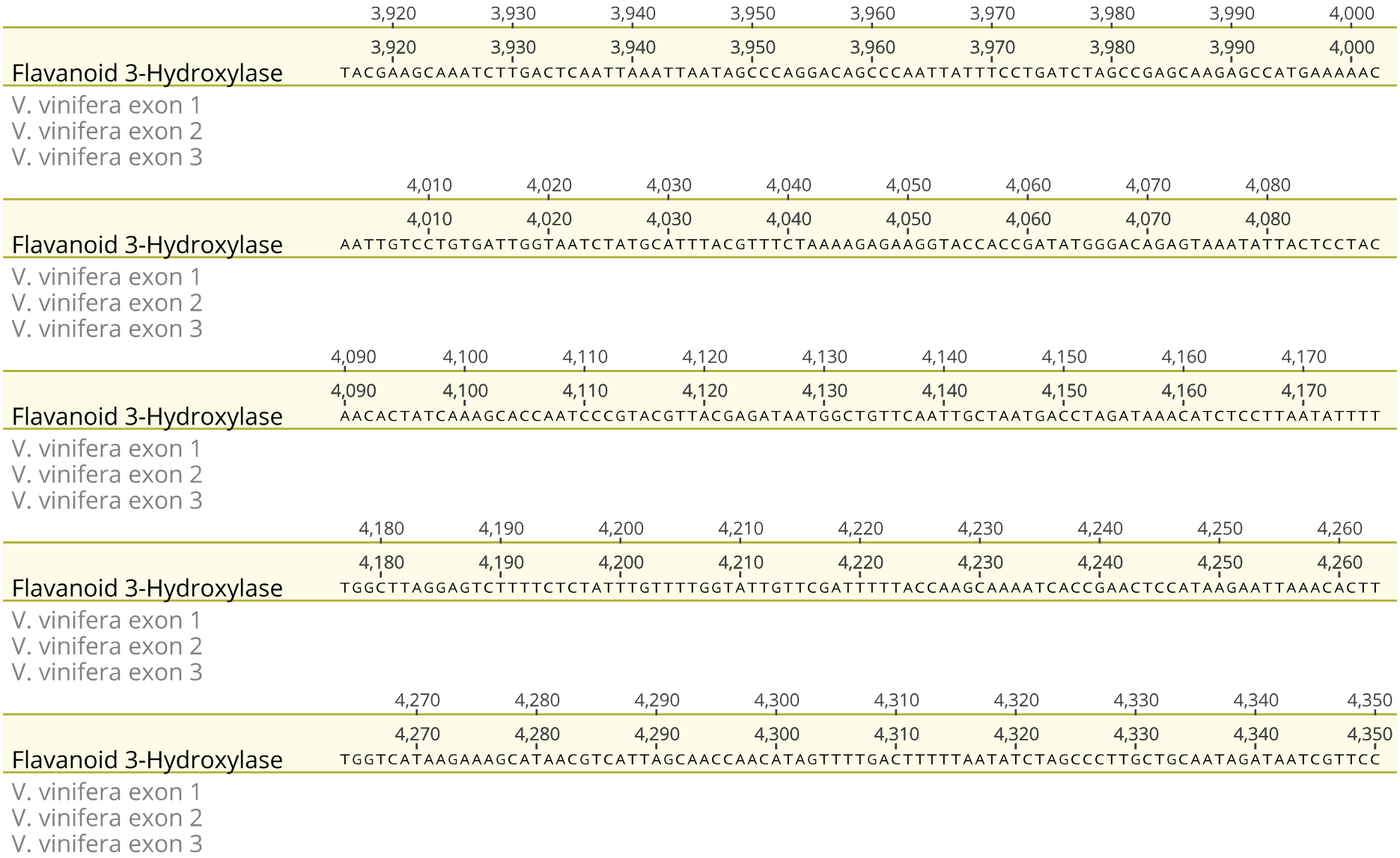


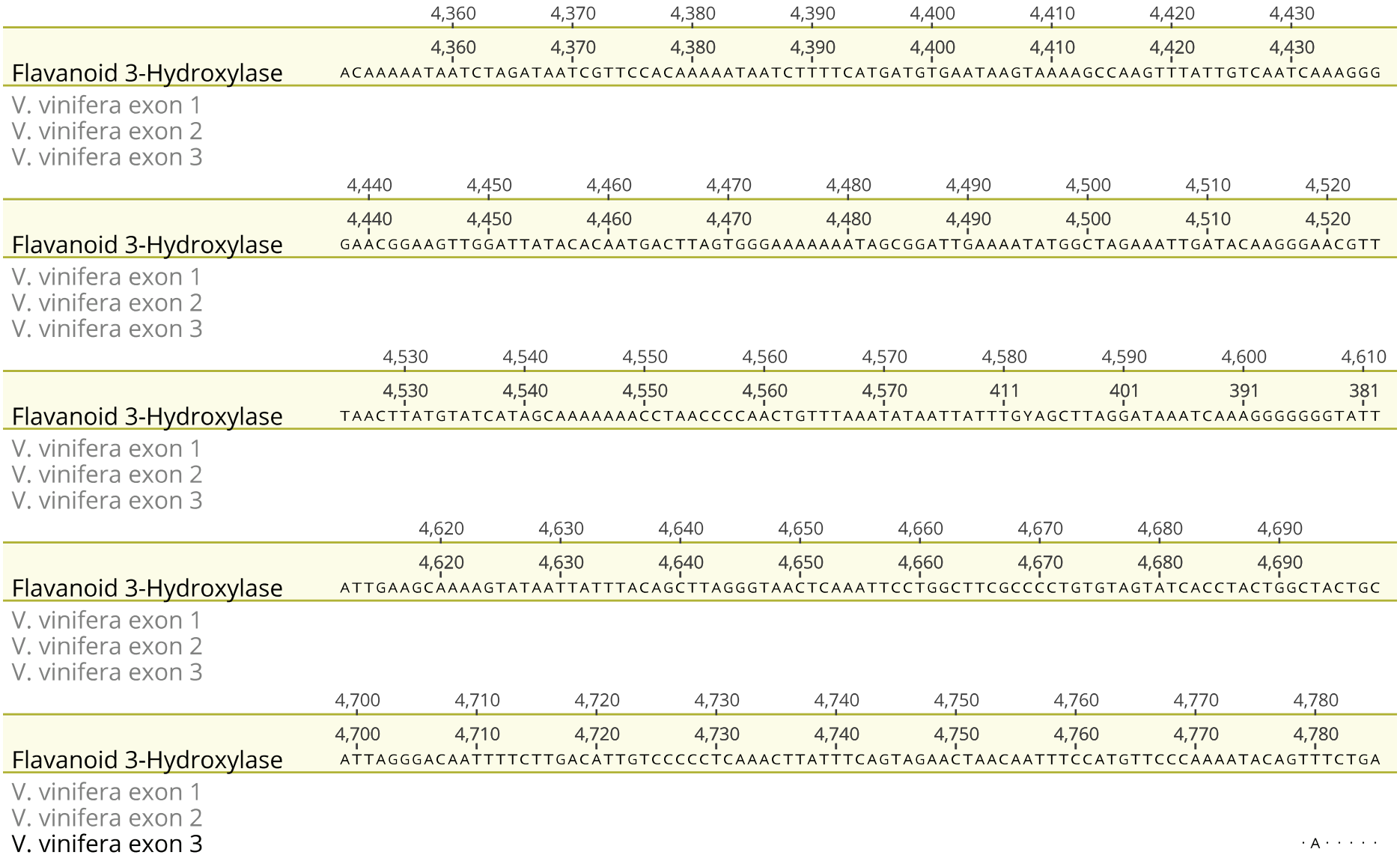


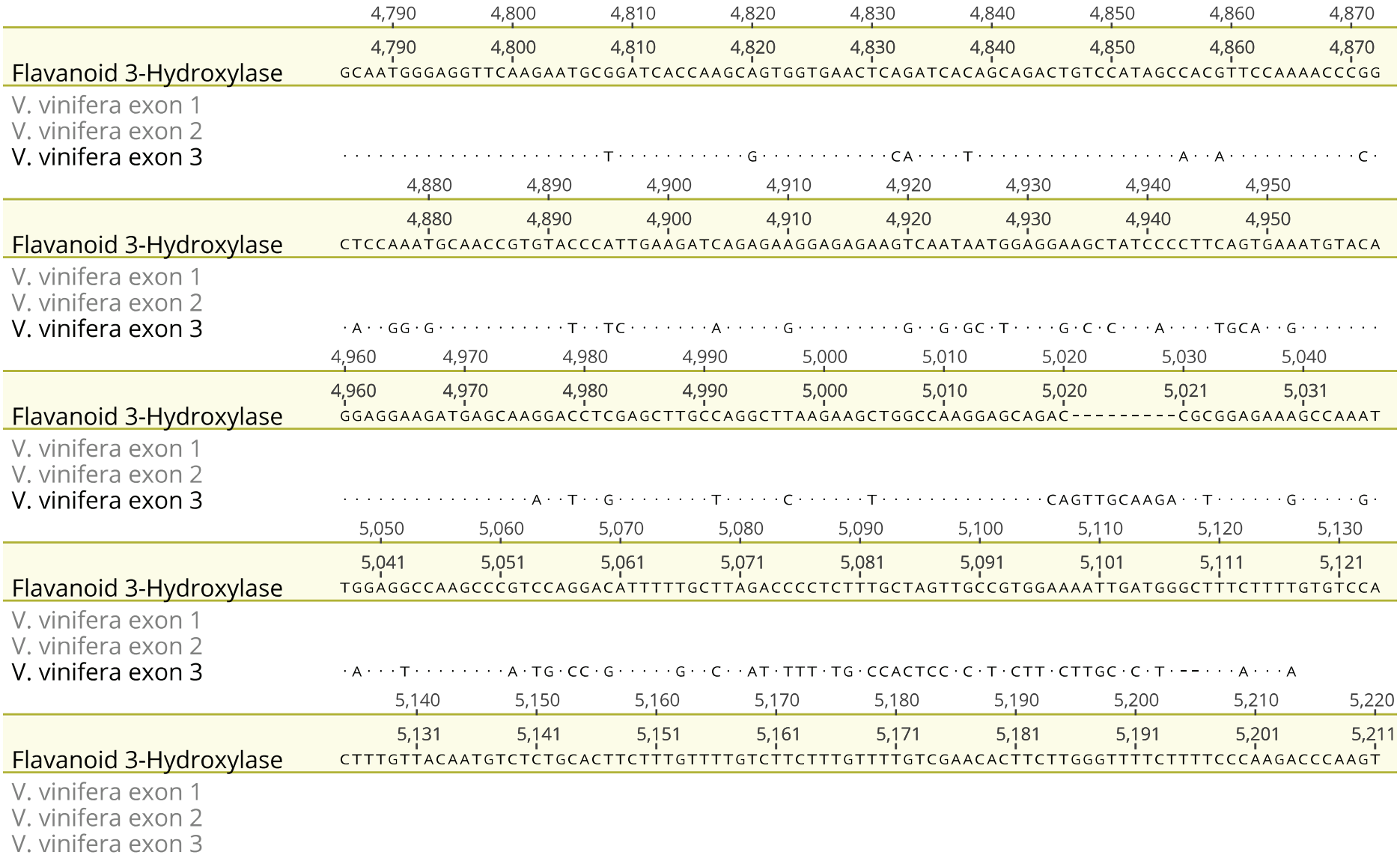


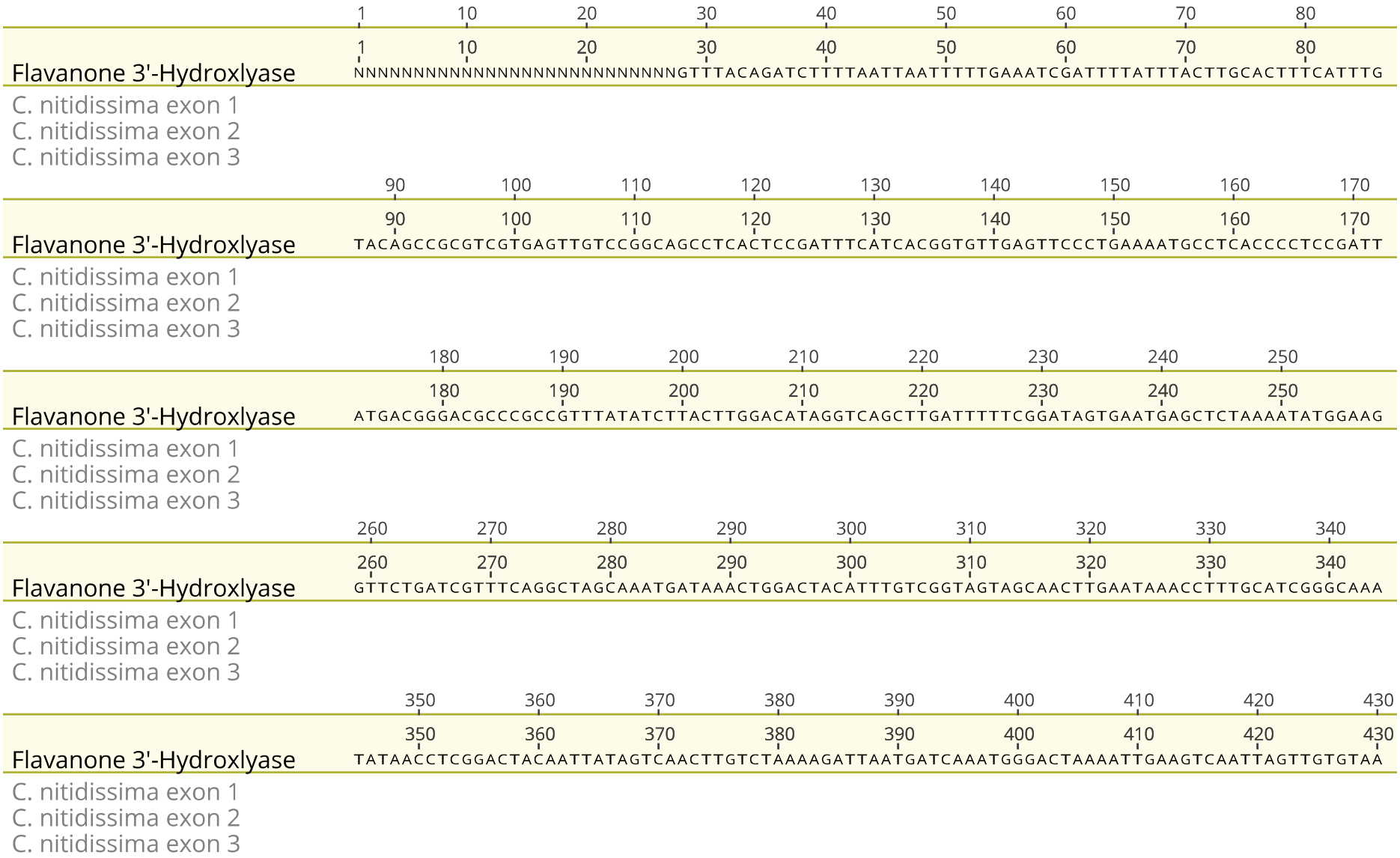


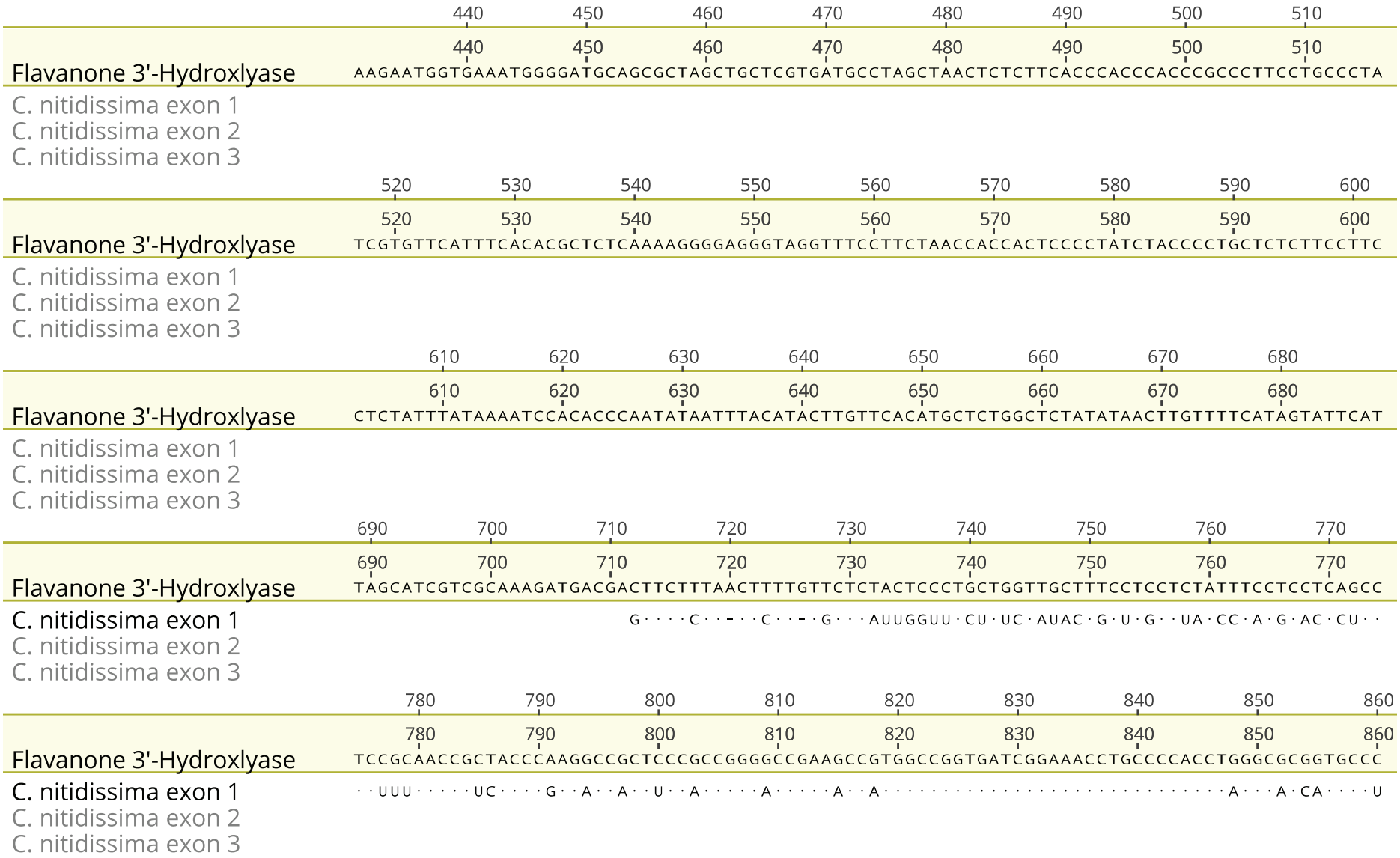


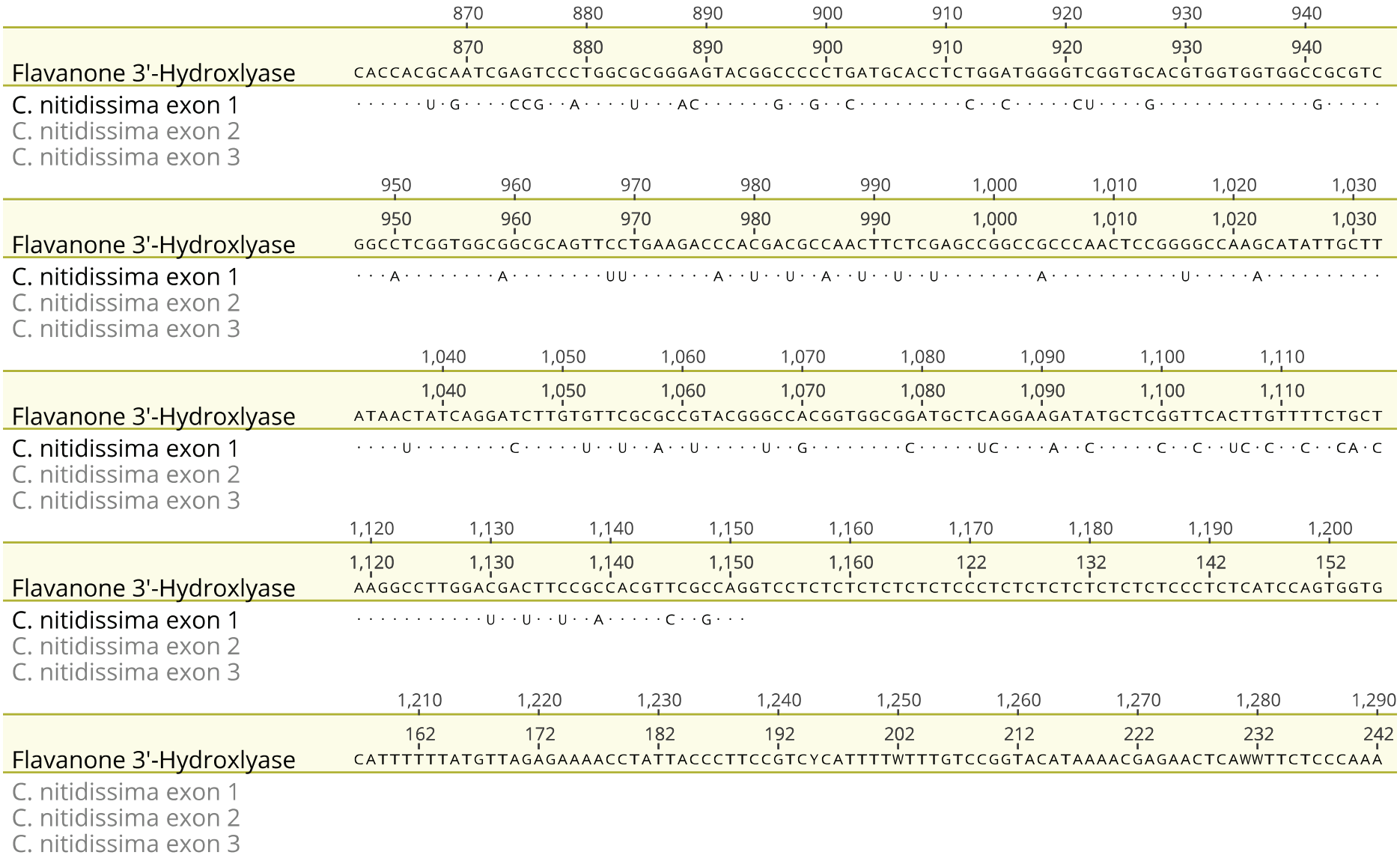


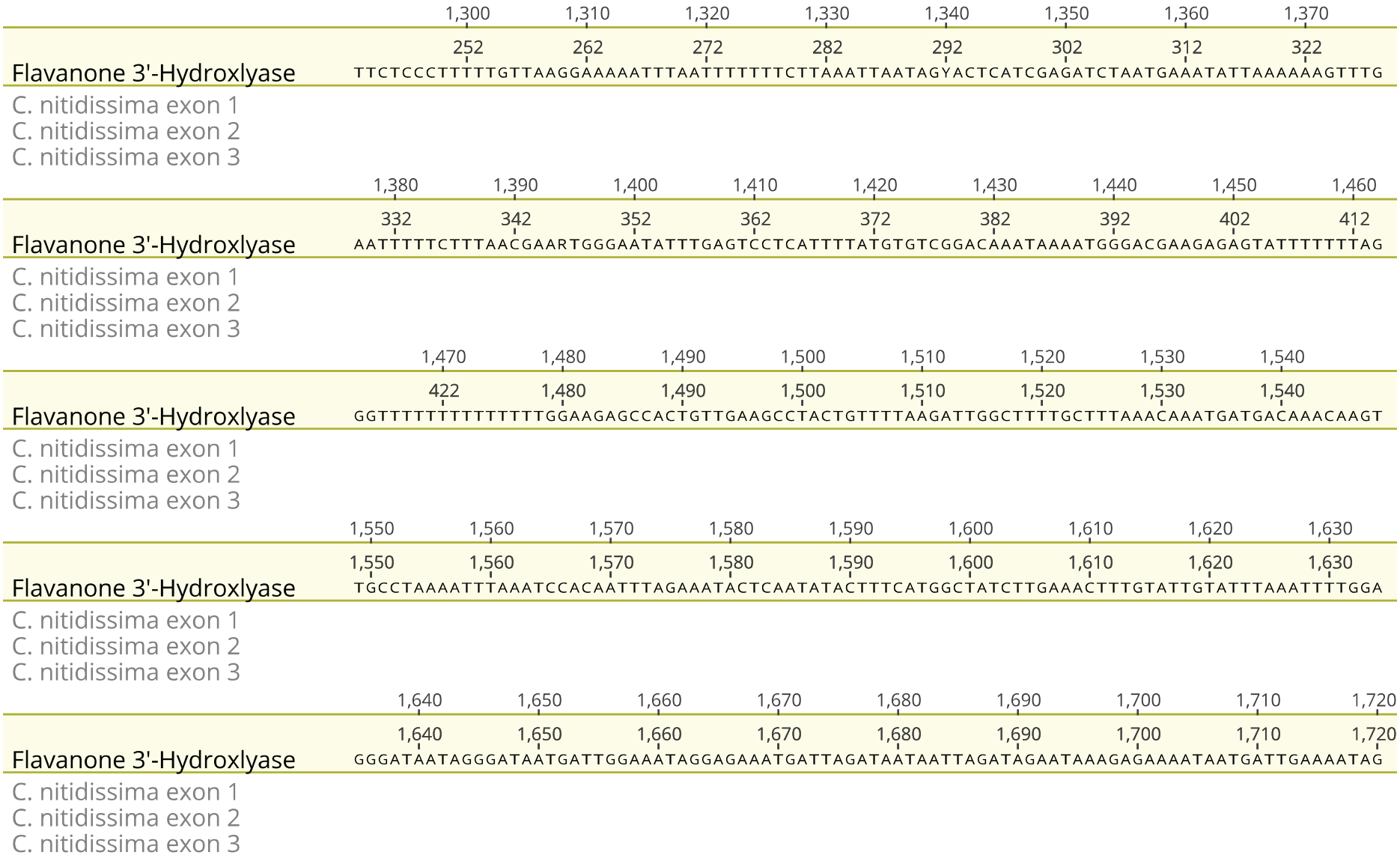


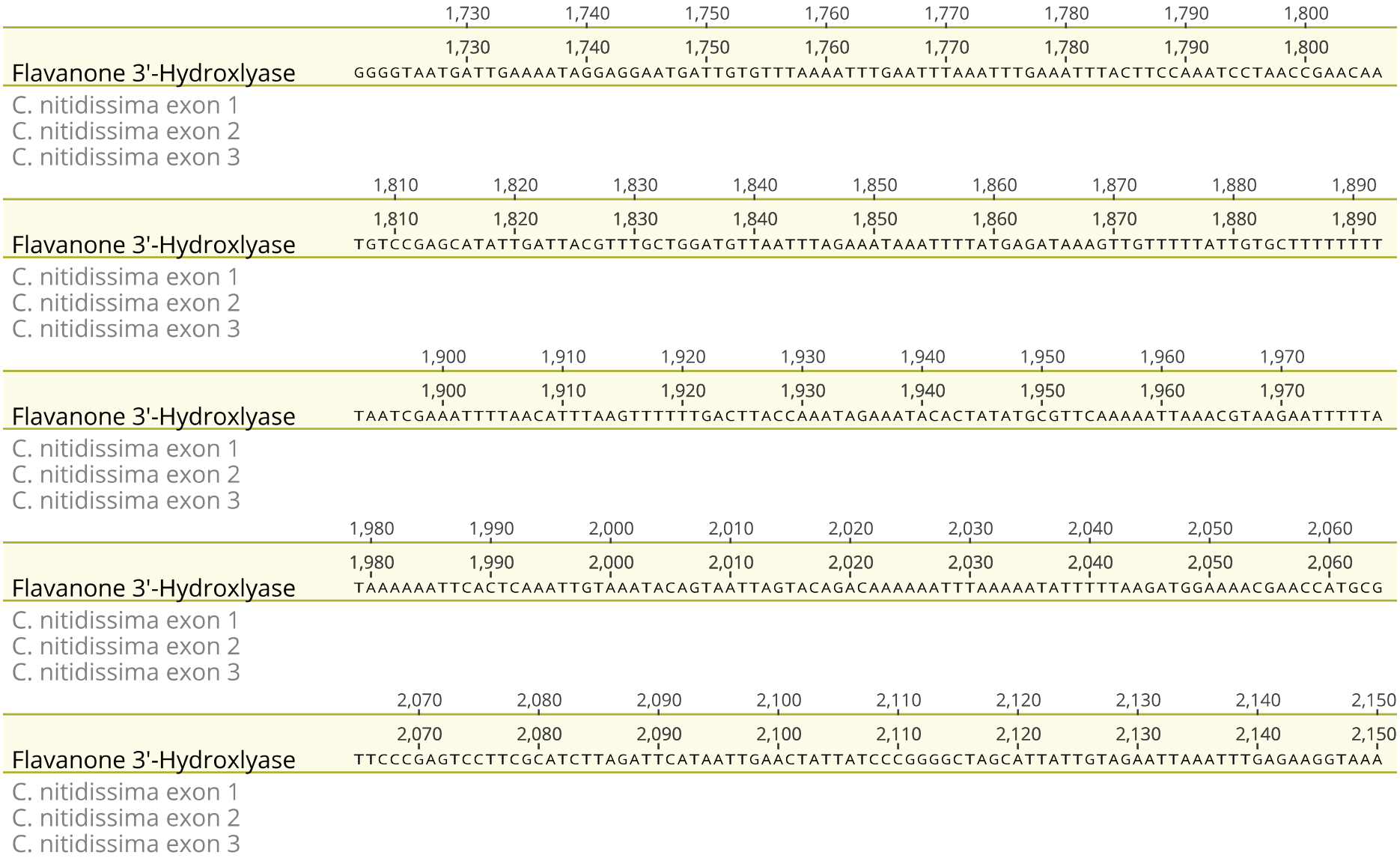


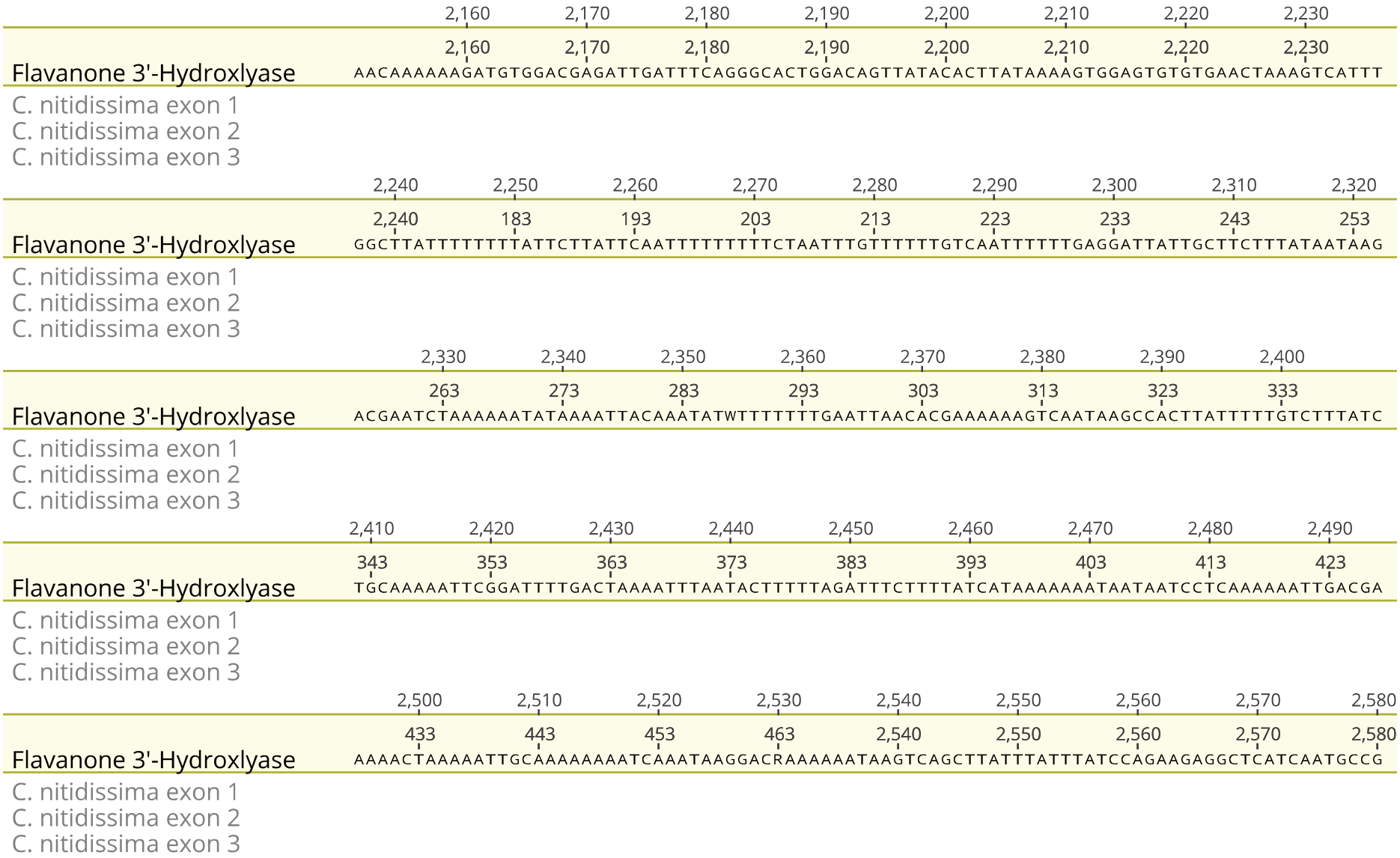


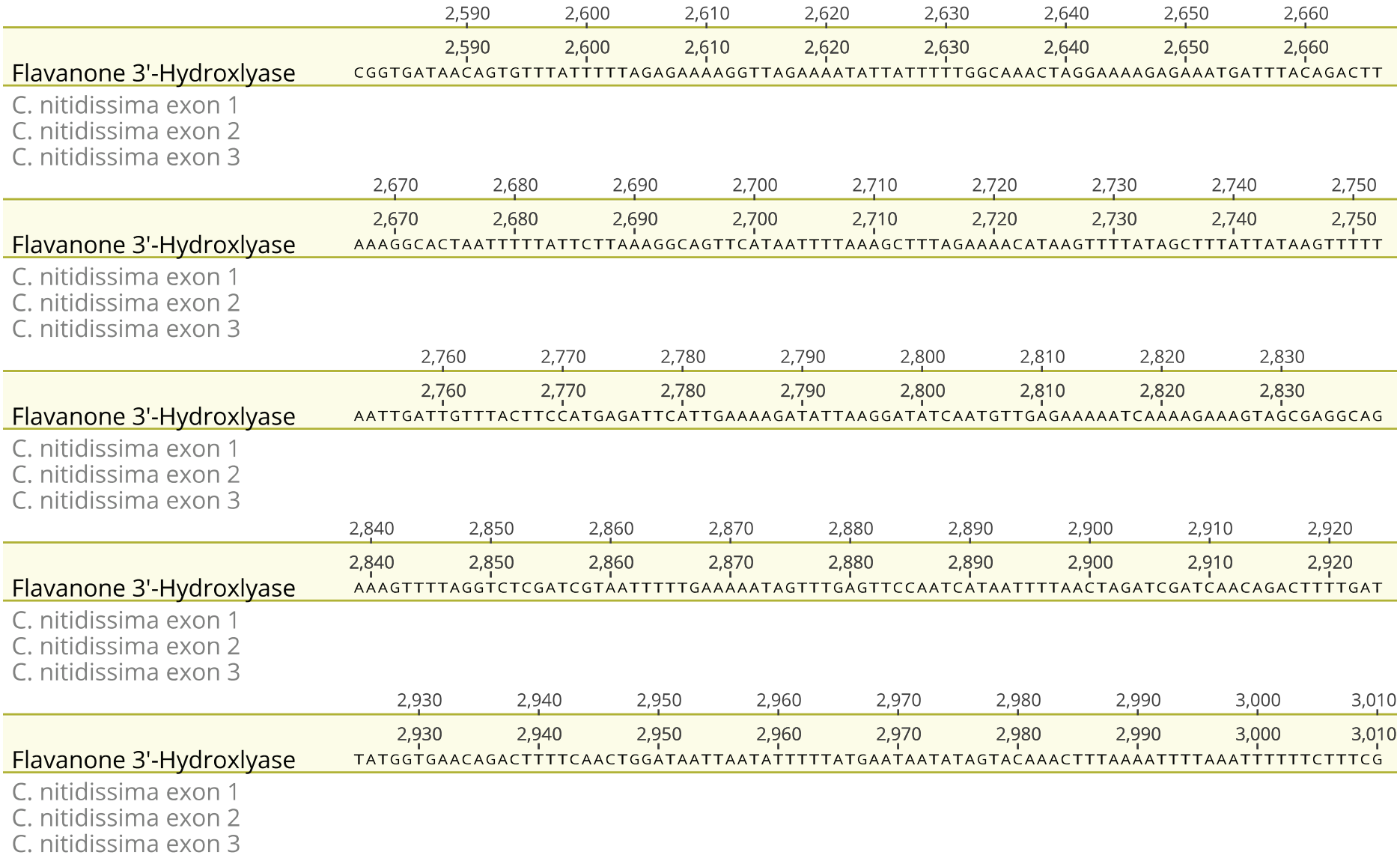


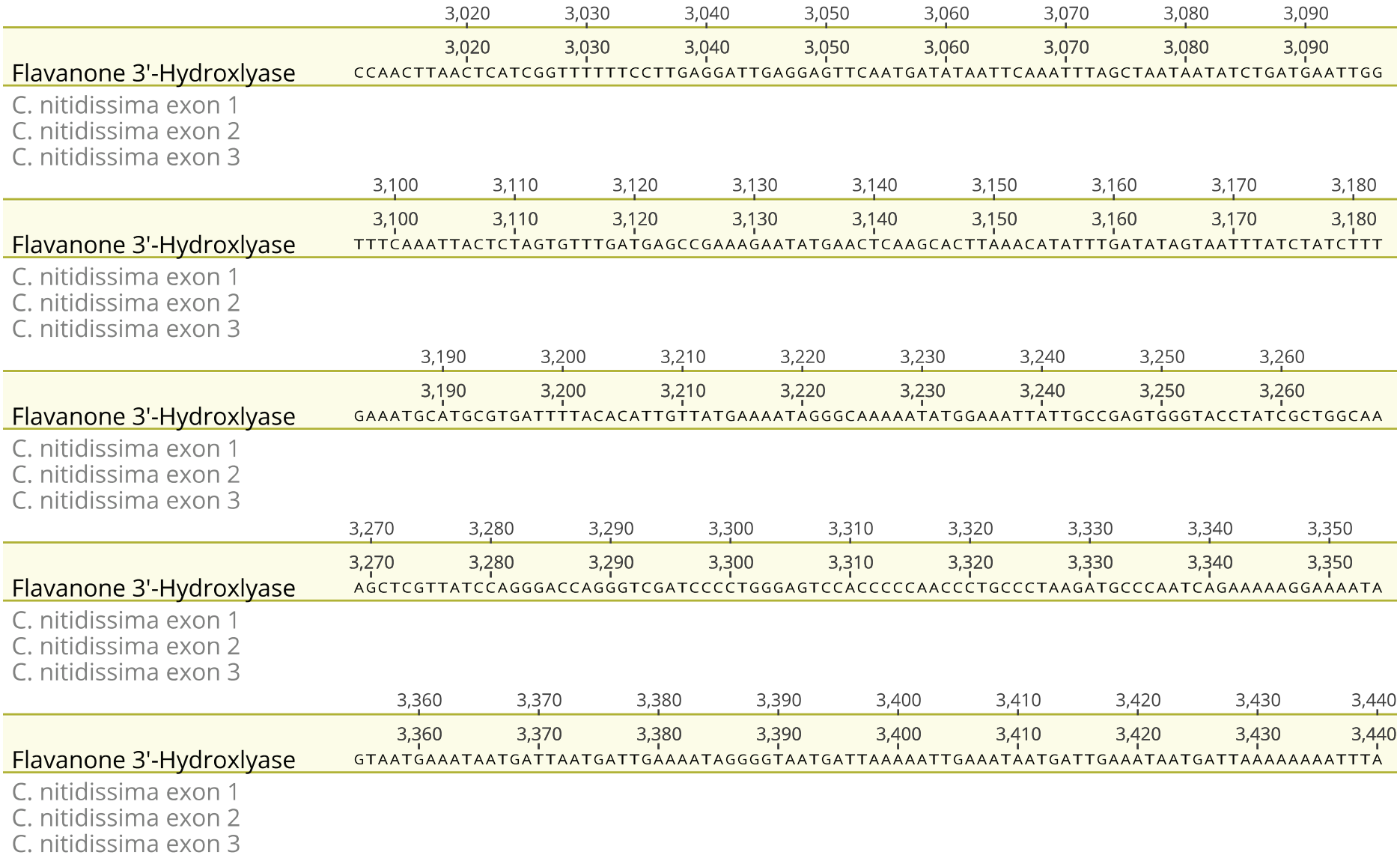


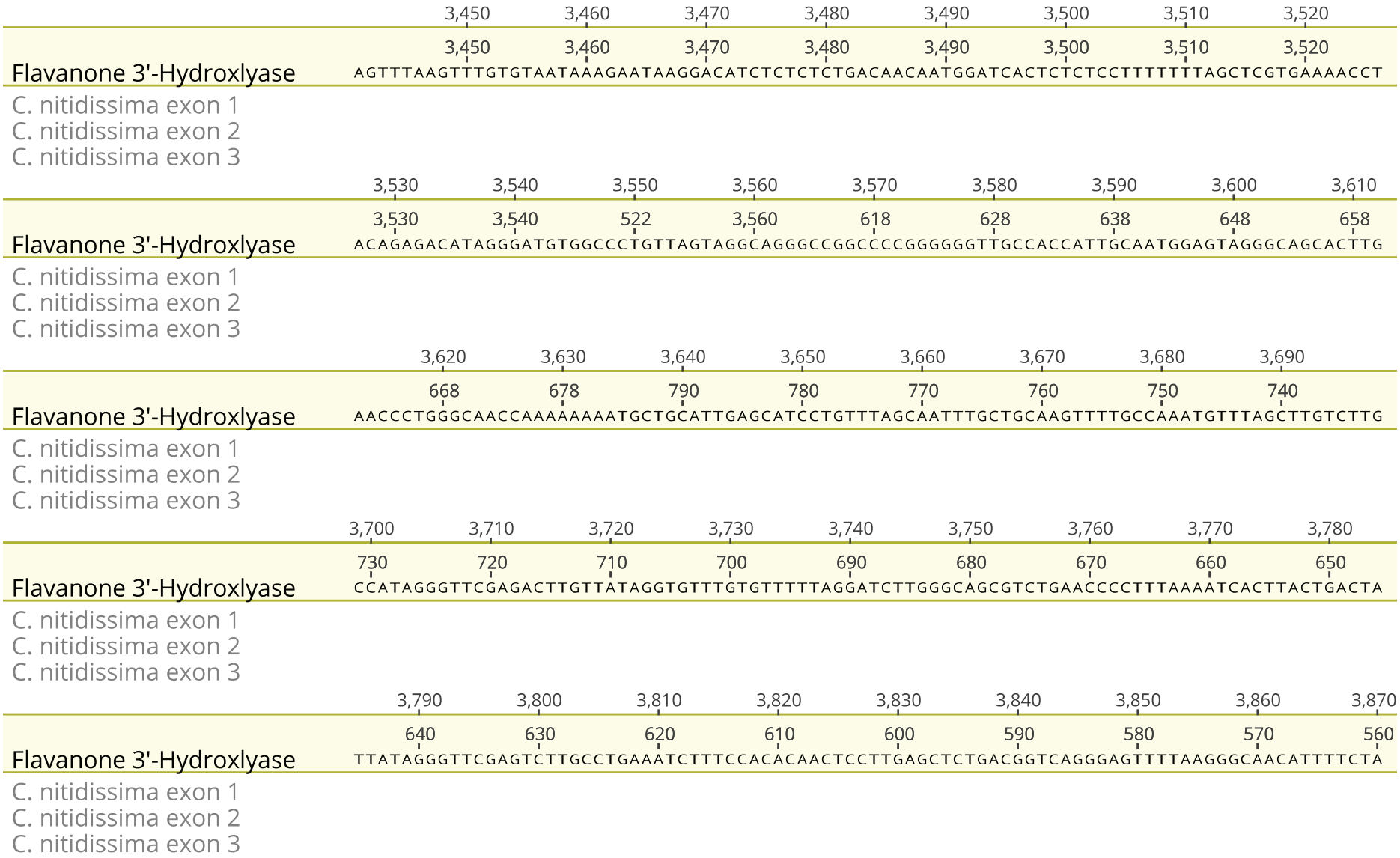


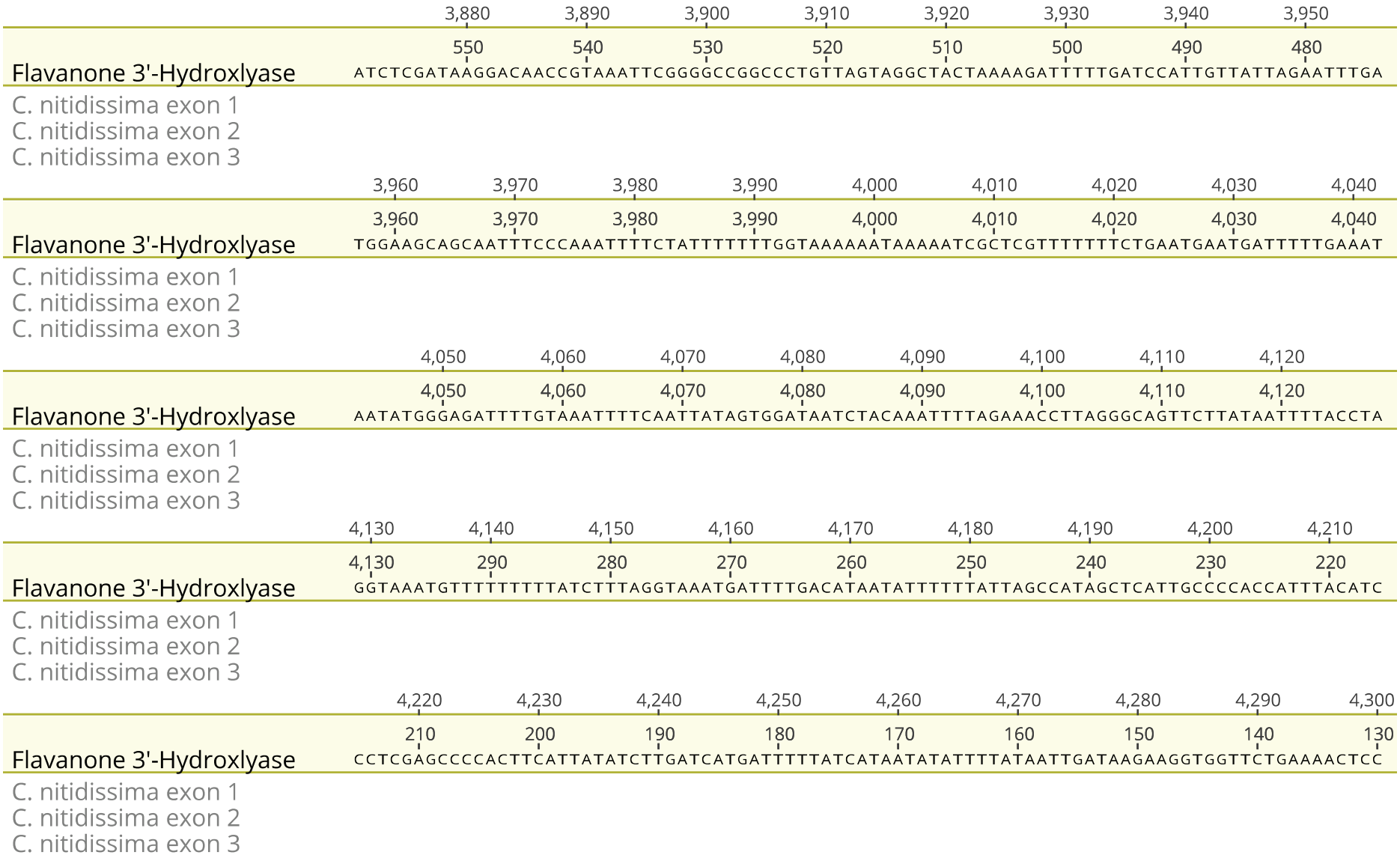


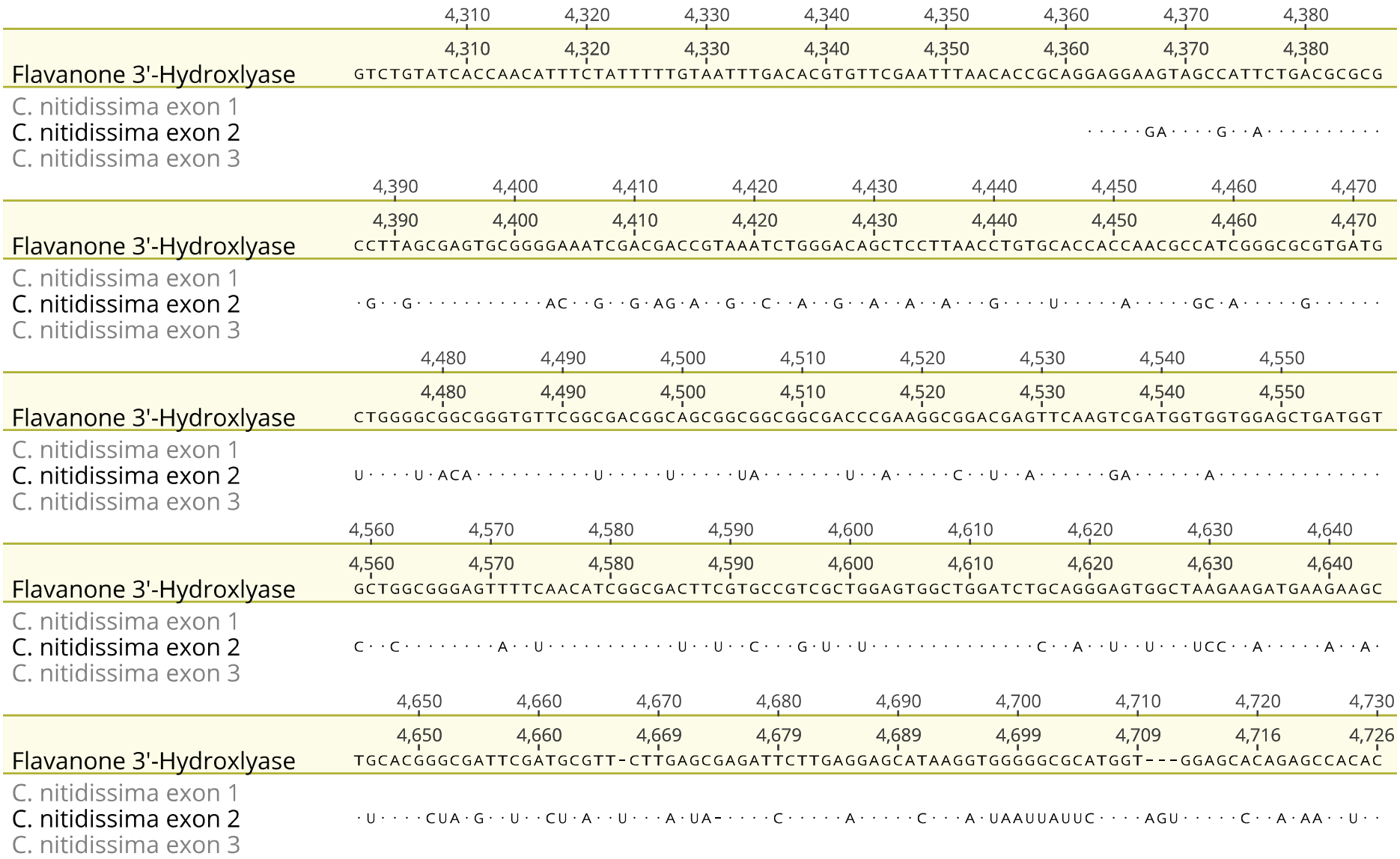


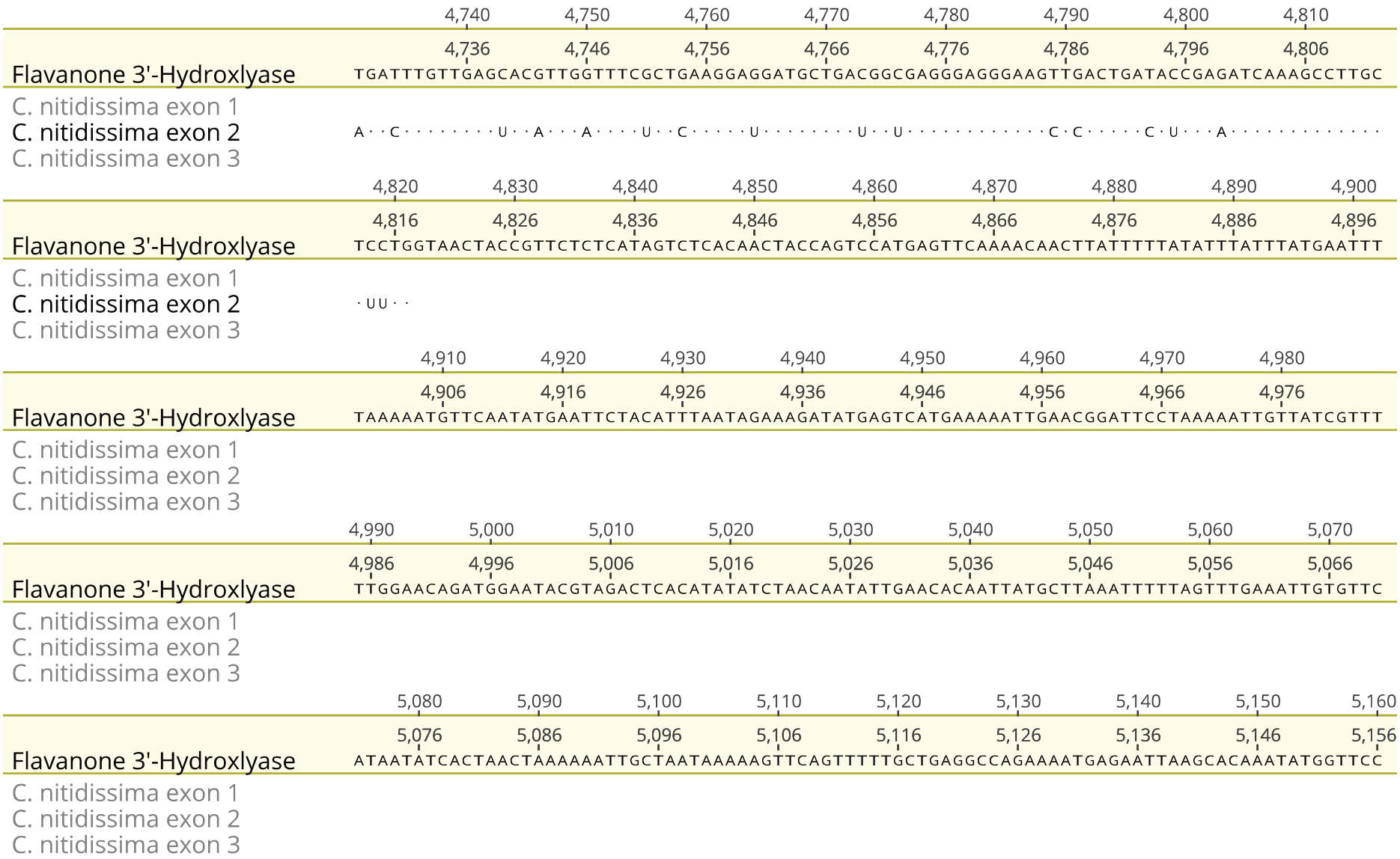


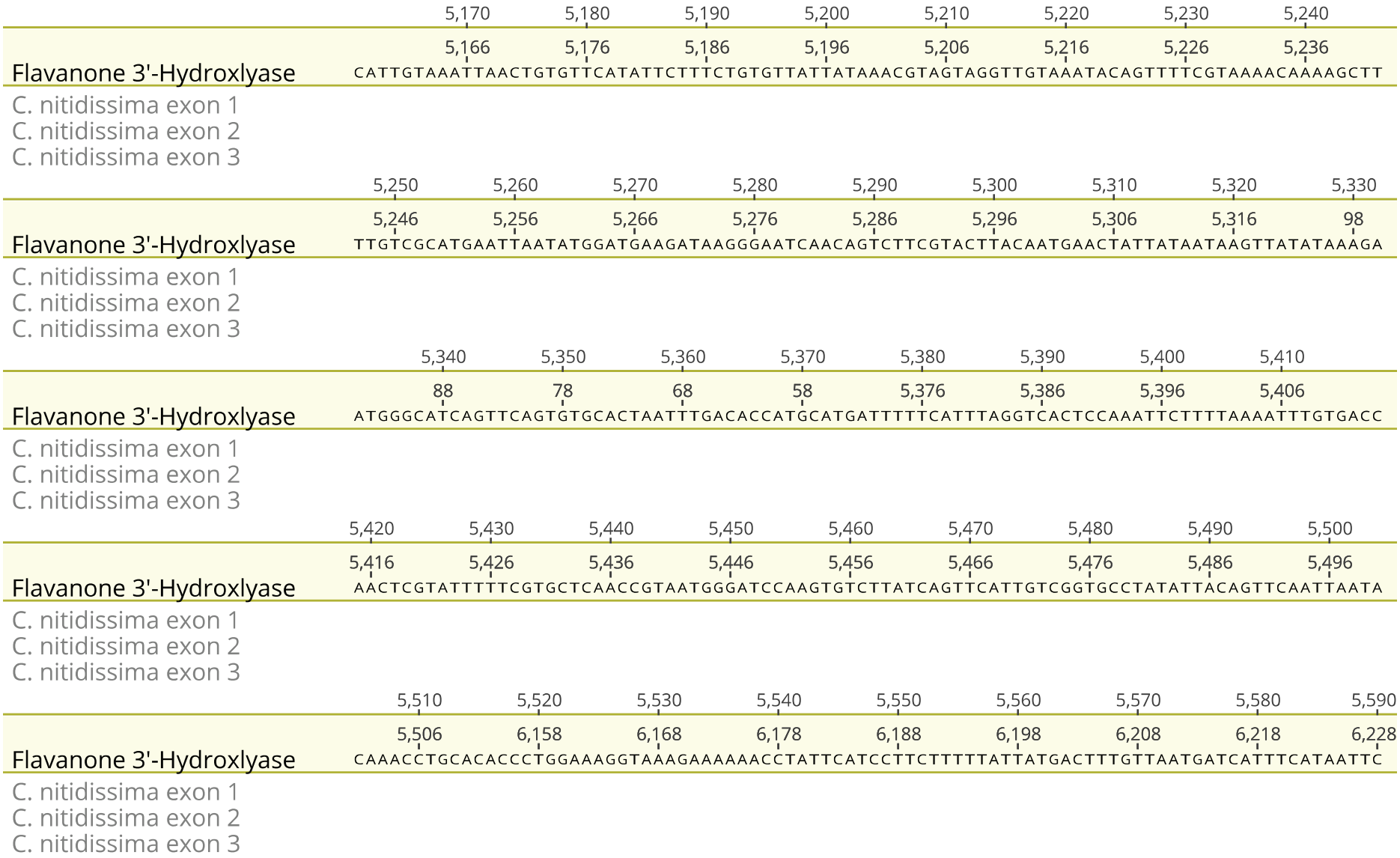


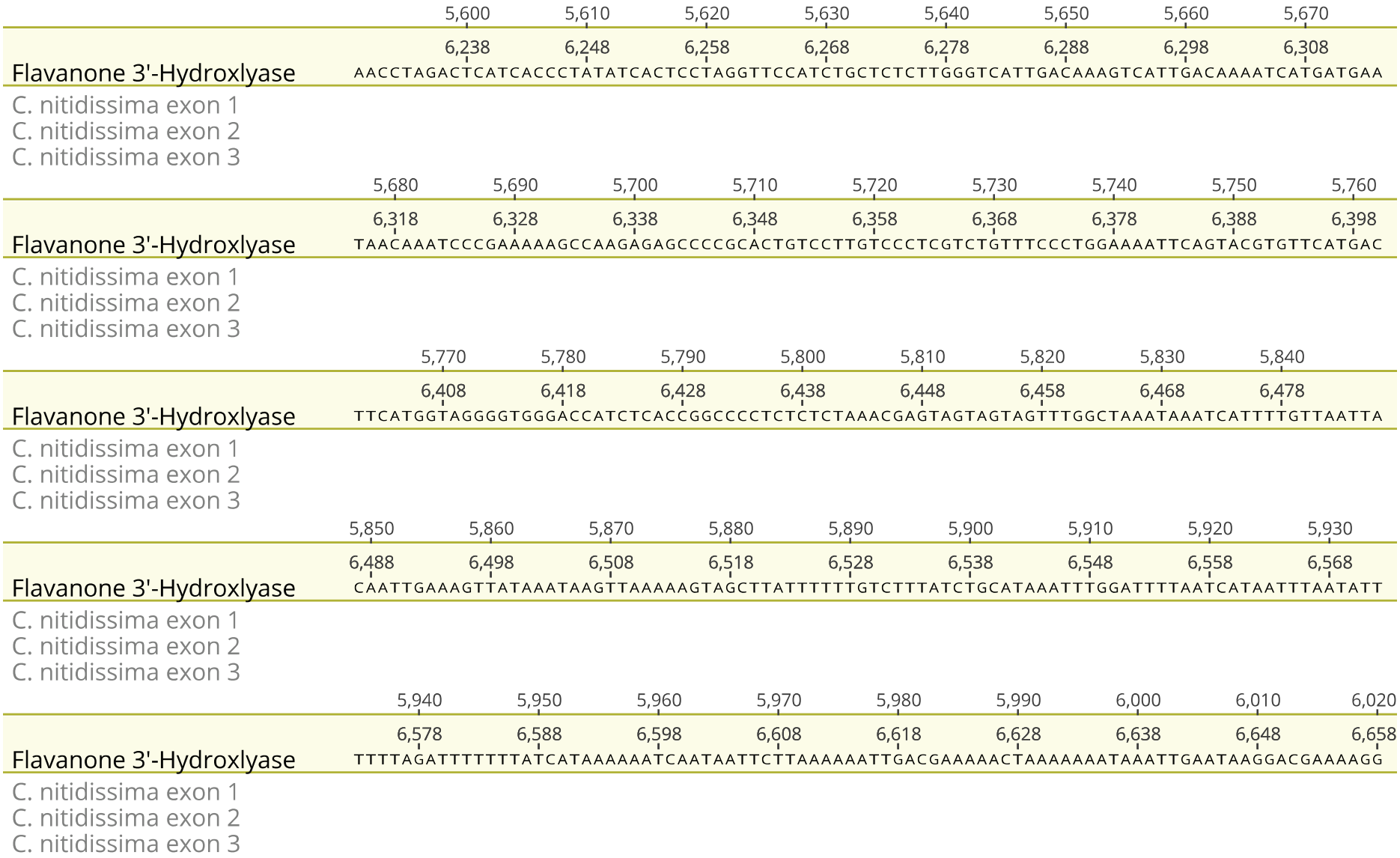


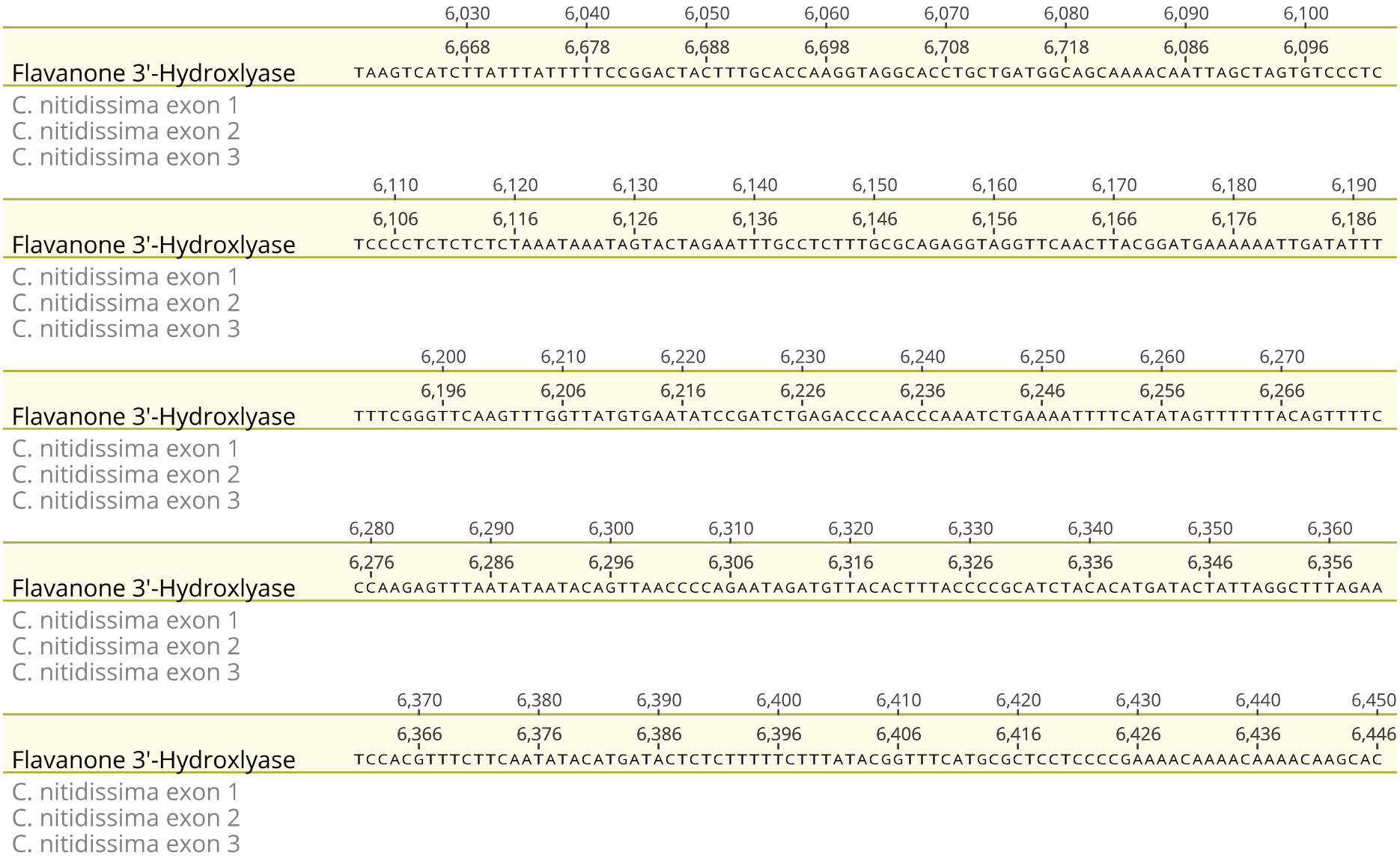


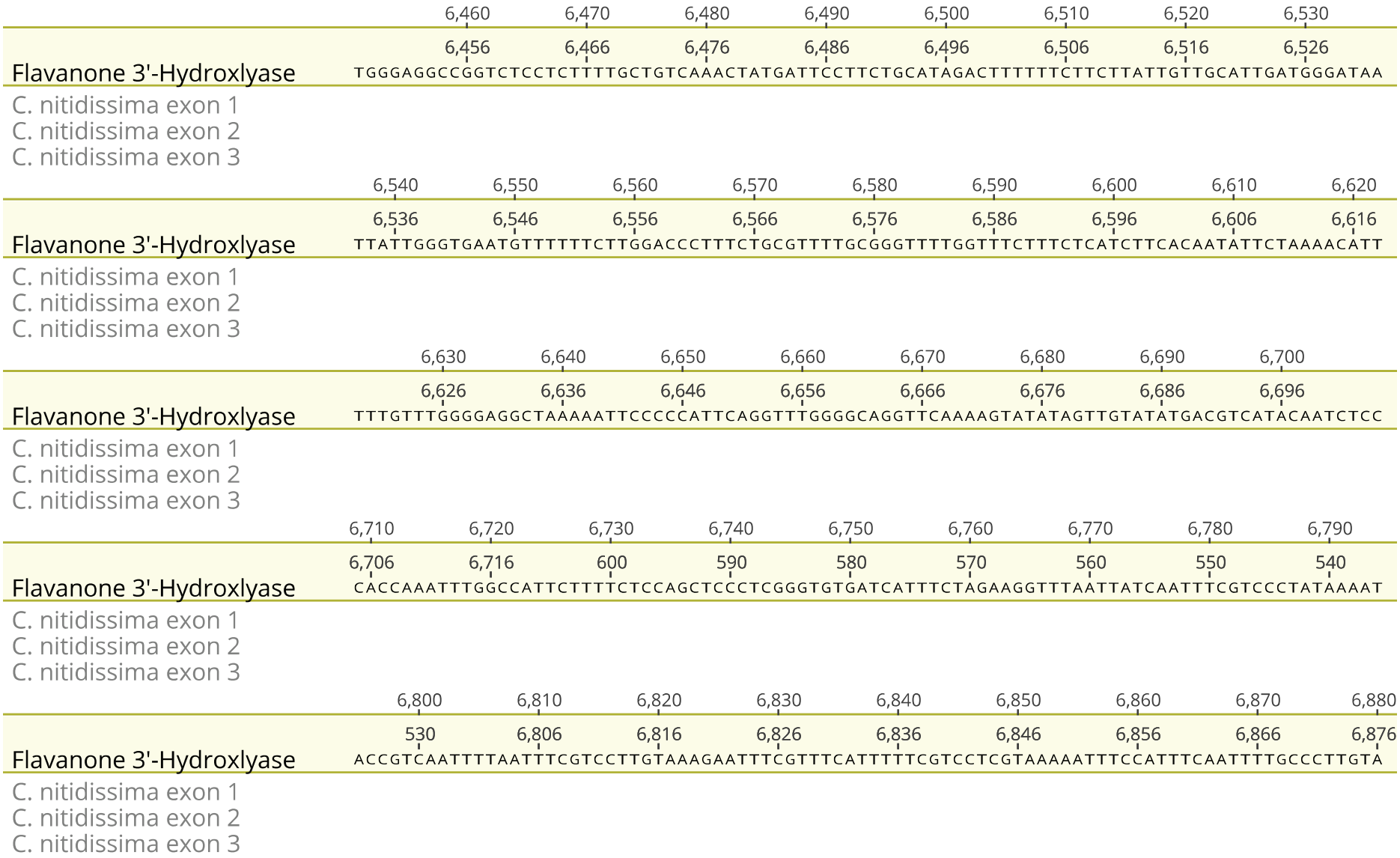


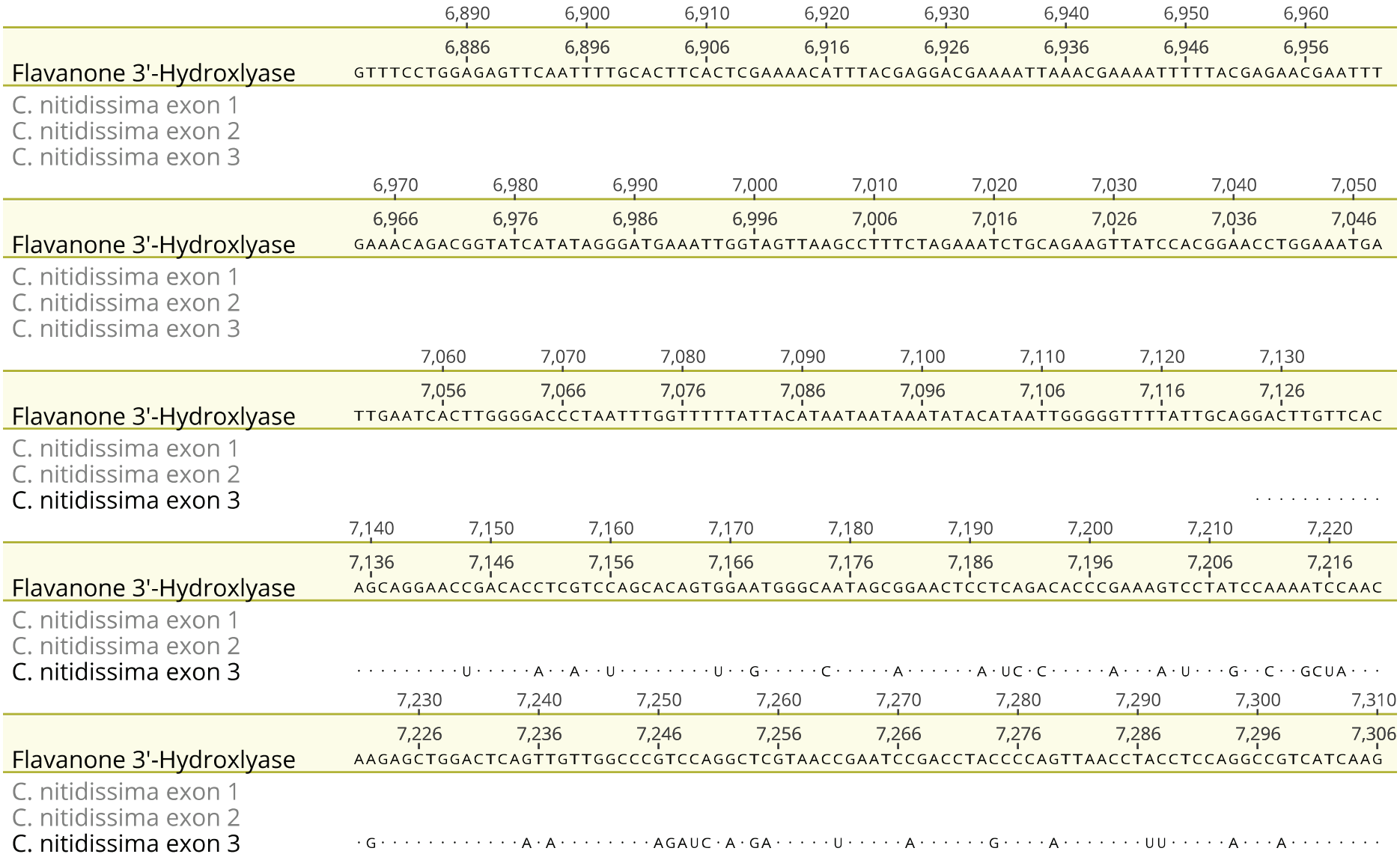


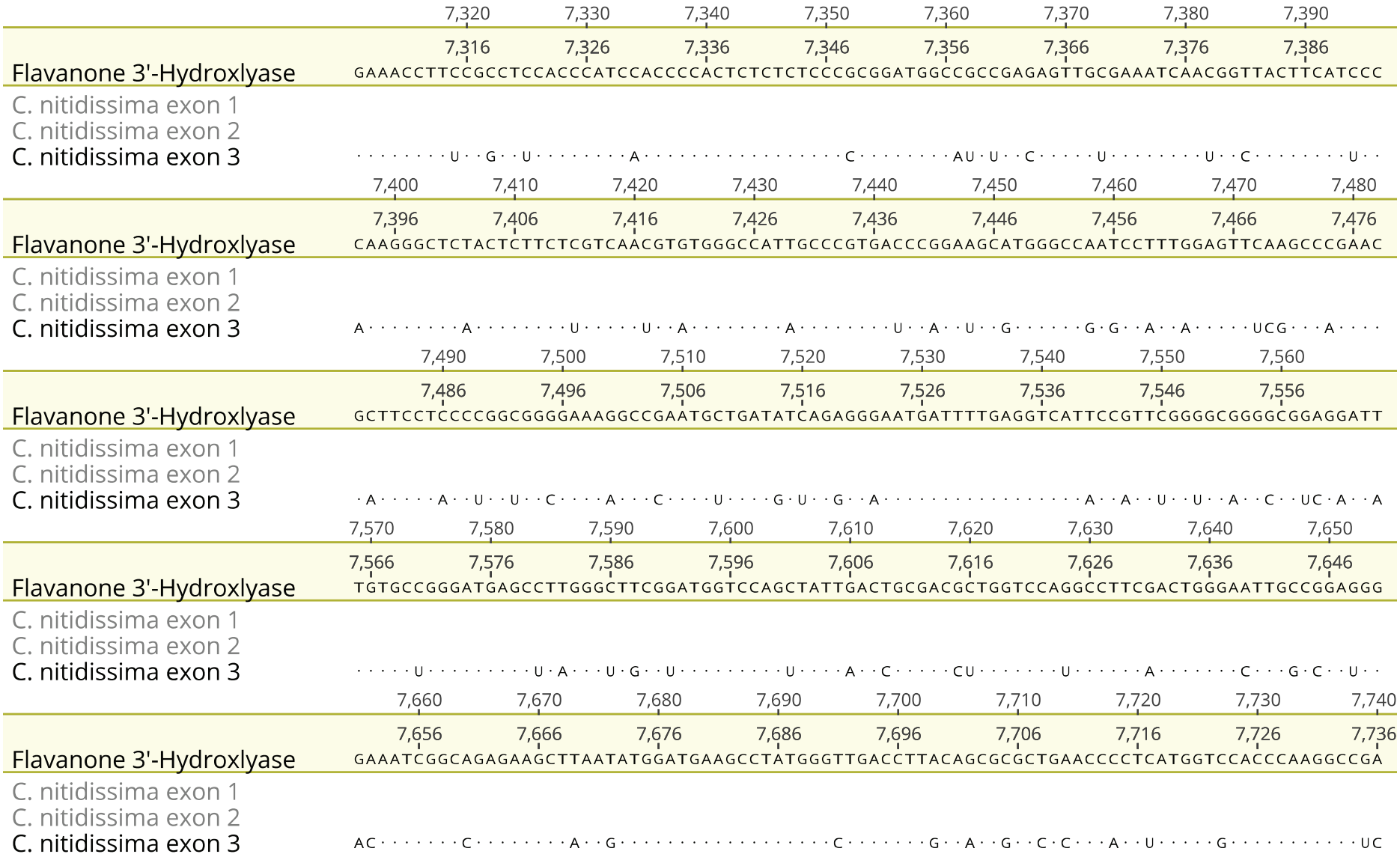


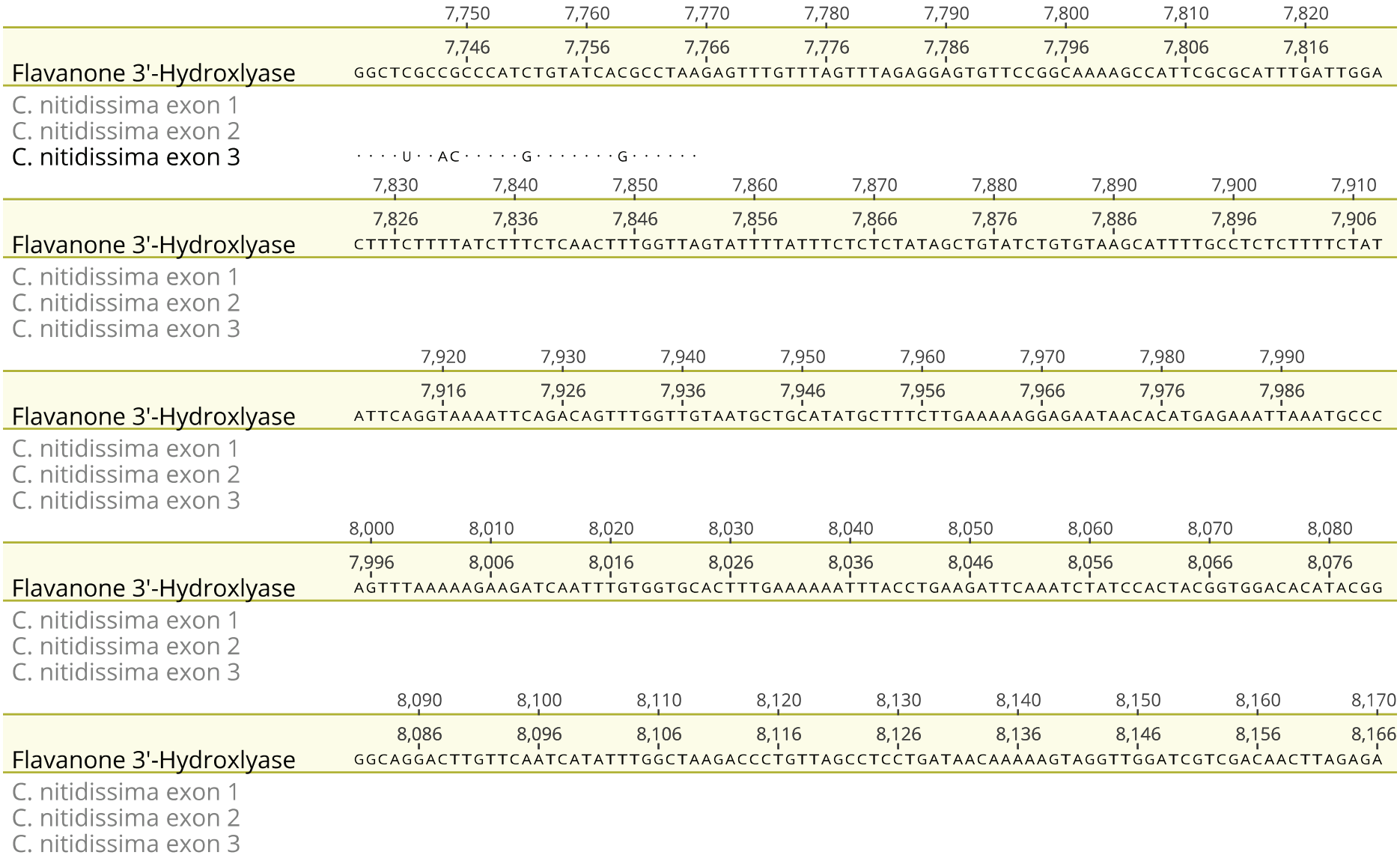


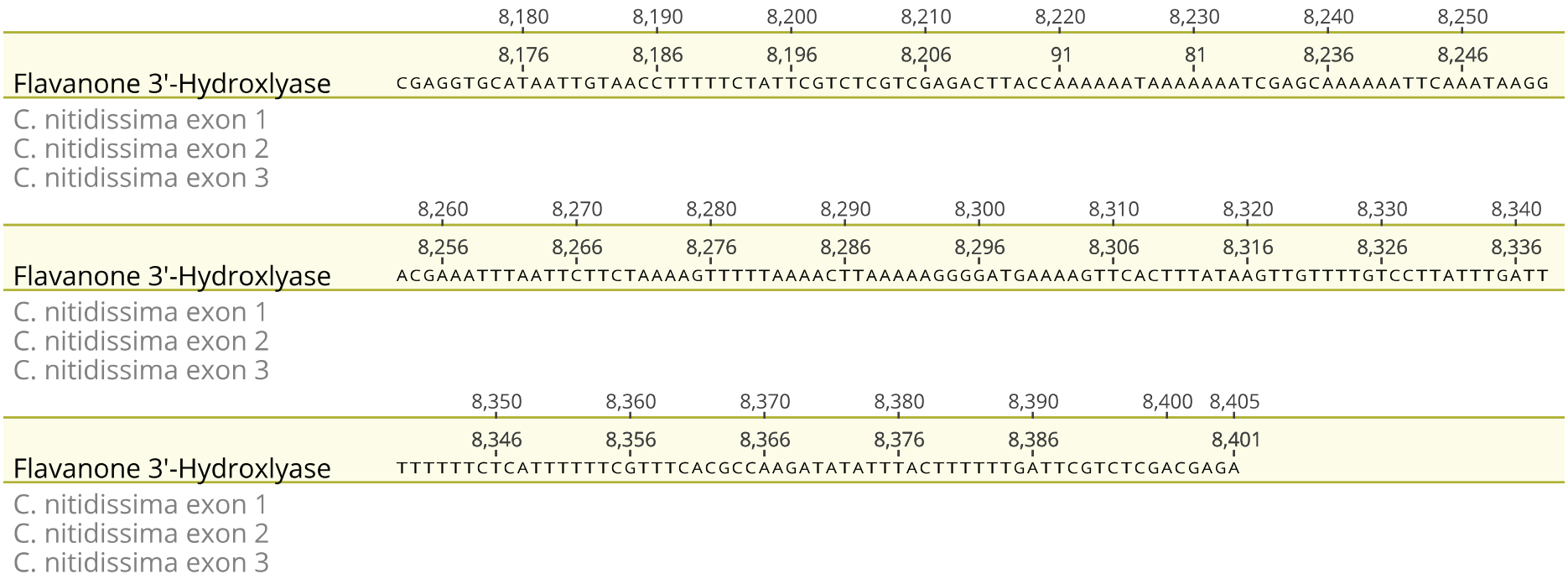


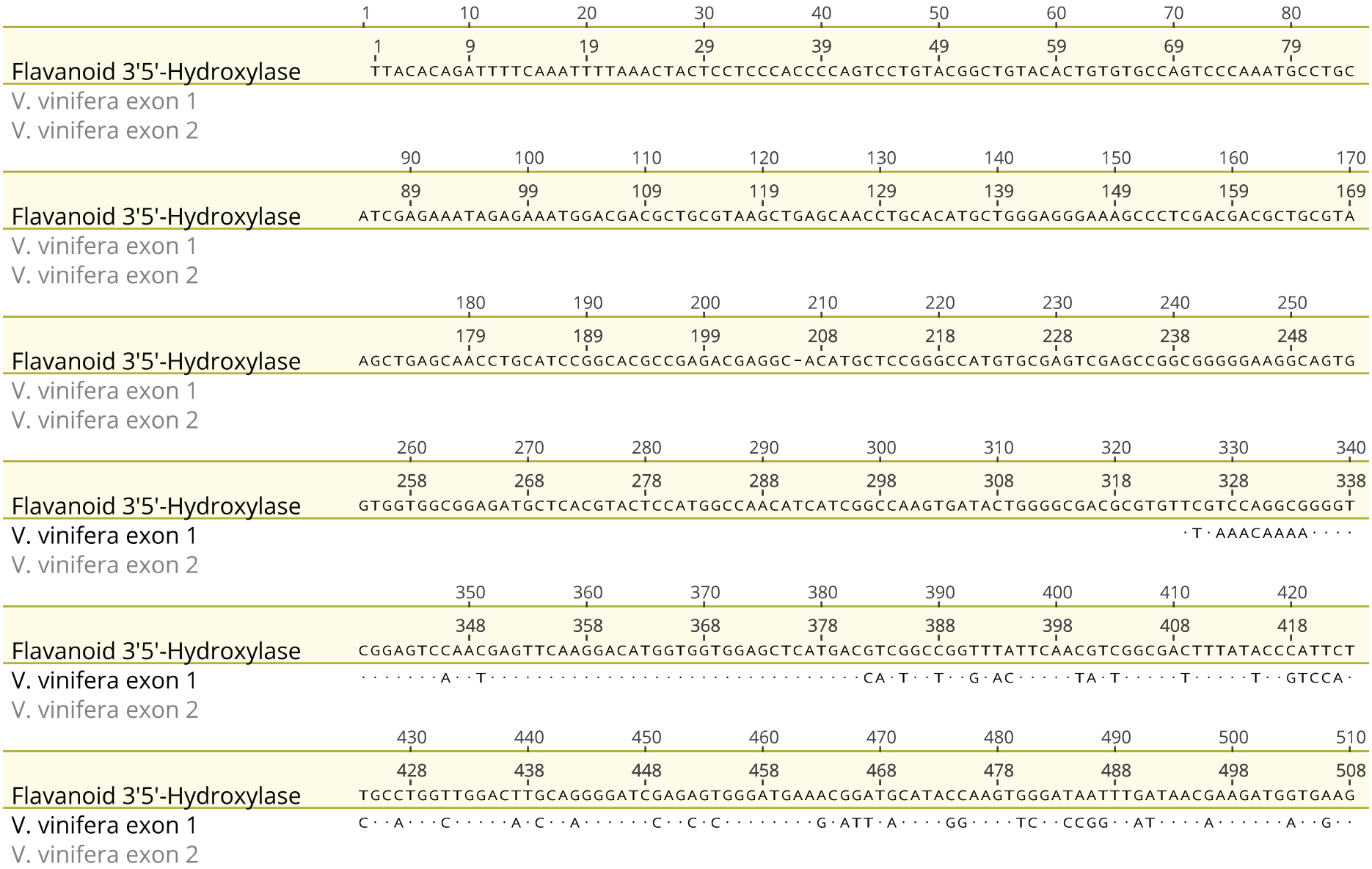


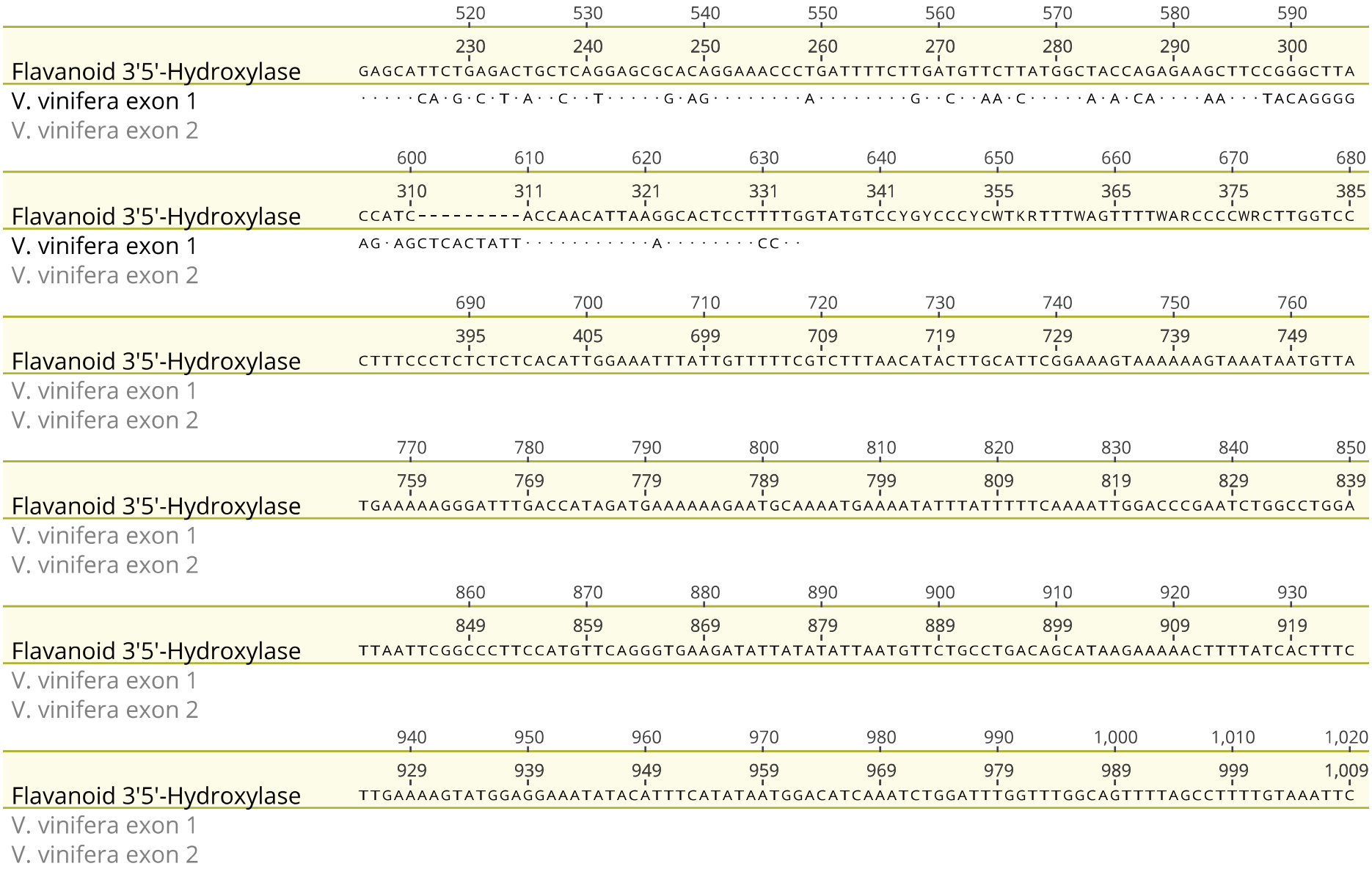












	1,030	1,040	1,050	1,060	1,070	1,080	1,090	1,100
	1,019	1,029	1,039	1,049	1,059	1,069	1,079	1,089

Flavanoid 3'5'-Hydroxylase AAGATCCAAATATAGTTAATATATATGCCCGAACTGCAAAATGAAAAAAAAATAAAAATTAGAAGGTTATTTGATGTAAAGGGTT
V. vinifera exon 1
V. vinifera exon 2

	1,110	1,120	1,130	1,140	1,150	1,160	1,170	1,180	1,190
	1,099	1,109	1,119	1,129	1,139	1,149	1,159	1,169	1,179

Flavanoid 3'5'-Hydroxylase GTTCATGGTTGTGGAAAGTGAACGCAAGAGAAATTTTGTACTCATTTTTATTCATGAGCTAGAGTTTTATGGTCAGTTTGCACGA
V. vinifera exon 1
V. vinifera exon 2

	1,200	1,210	1,220	1,230	1,240	1,250	1,260	1,270
	1,189	1,199	1,209	1,219	1,229	1,239	1,249	1,259

Flavanoid 3'5'-Hydroxylase GCTTTAATTTTAATTAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
V. vinifera exon 1
V. vinifera exon 2

	1,280	1,290	1,300	1,310	1,320	1,330	1,340	1,350	1,360
	1,269	1,279	1,289	1,299	1,309	1,319	1,329	1,339	1,349

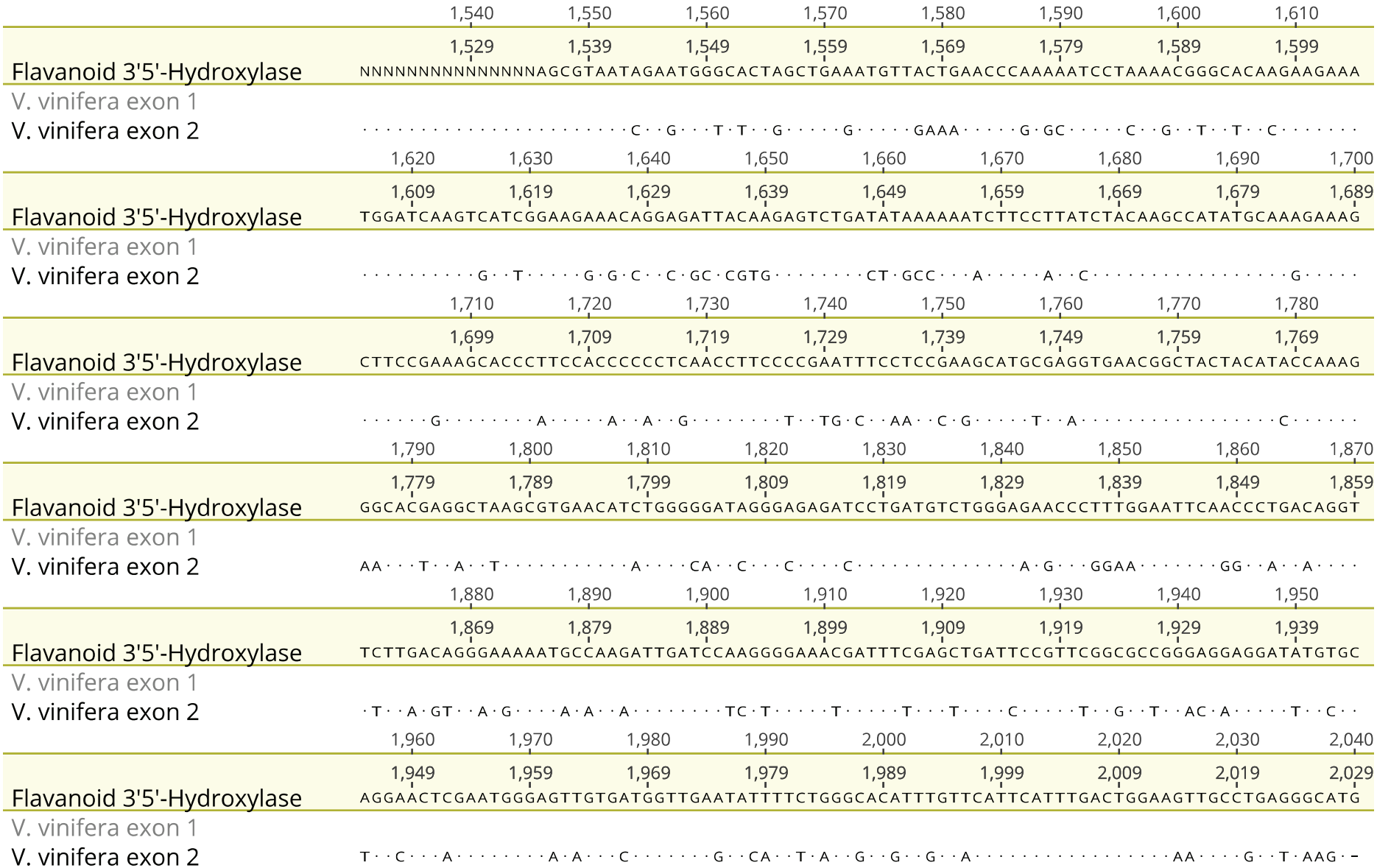
Flavanoid 3'5'-Hydroxylase NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
V. vinifera exon 1
V. vinifera exon 2

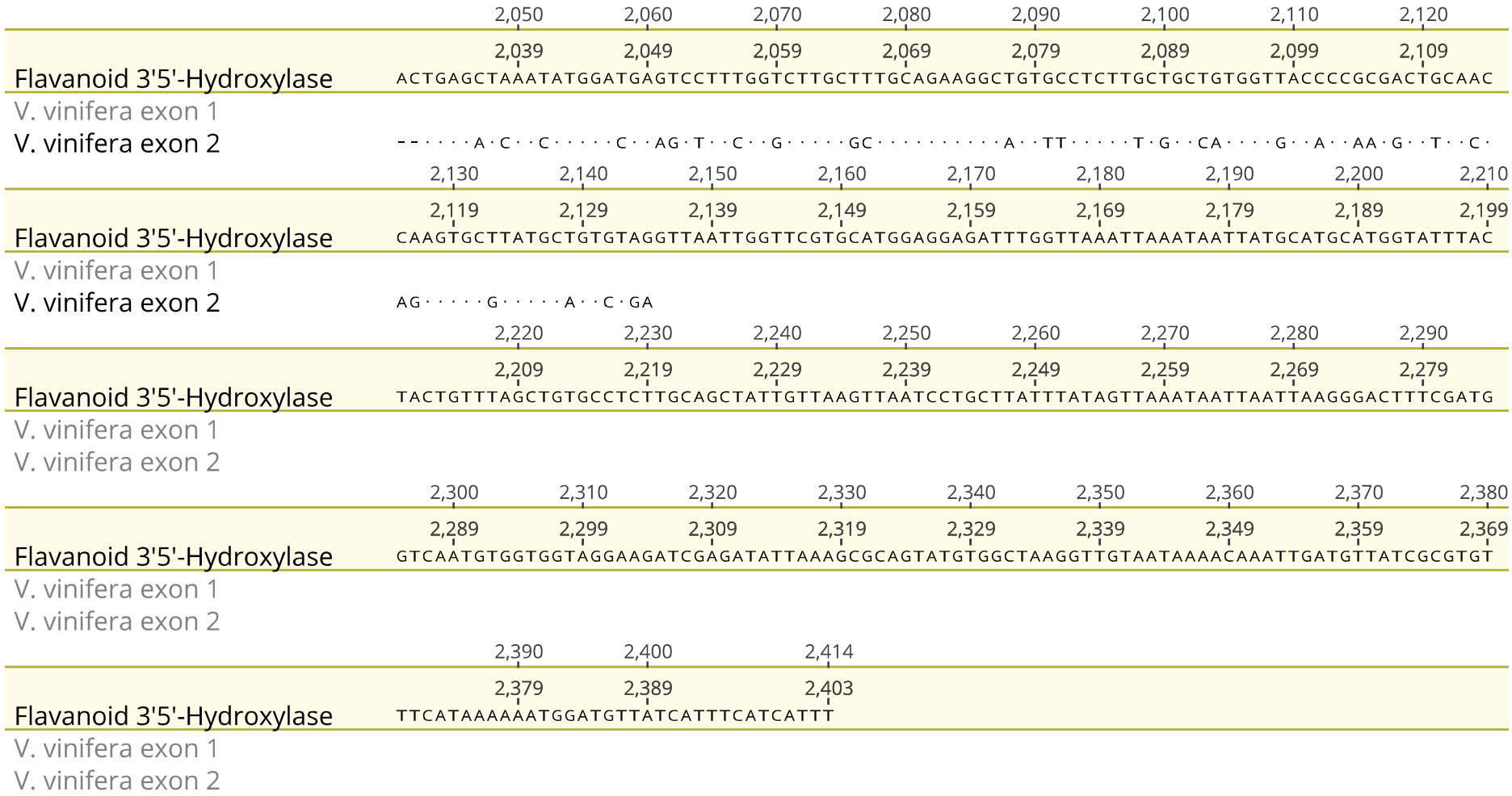
	1,370	1,380	1,390	1,400	1,410	1,420	1,430	1,440
	1,359	1,369	1,379	1,389	1,399	1,409	1,419	1,429

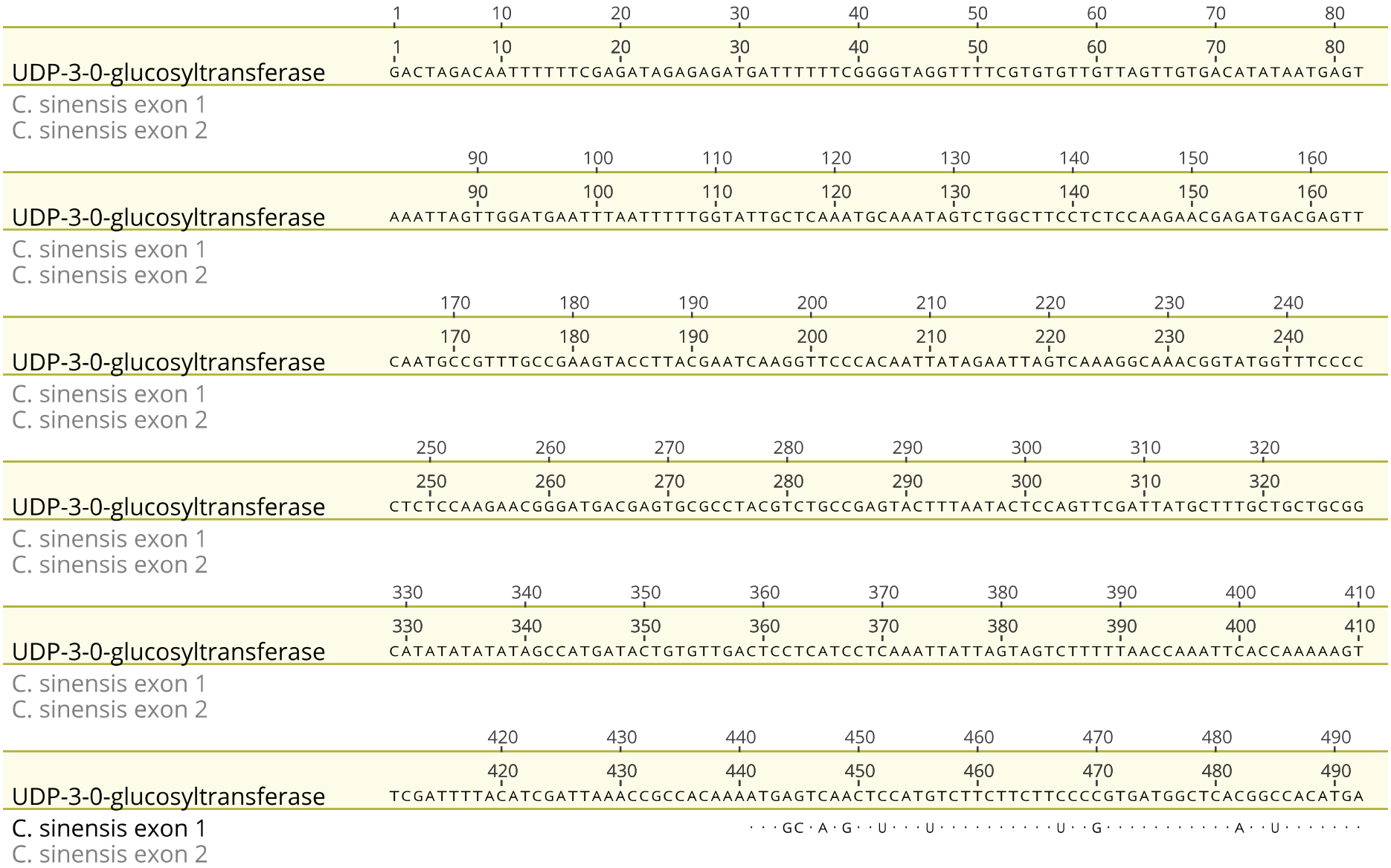
Flavanoid 3'5'-Hydroxylase NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
V. vinifera exon 1
V. vinifera exon 2

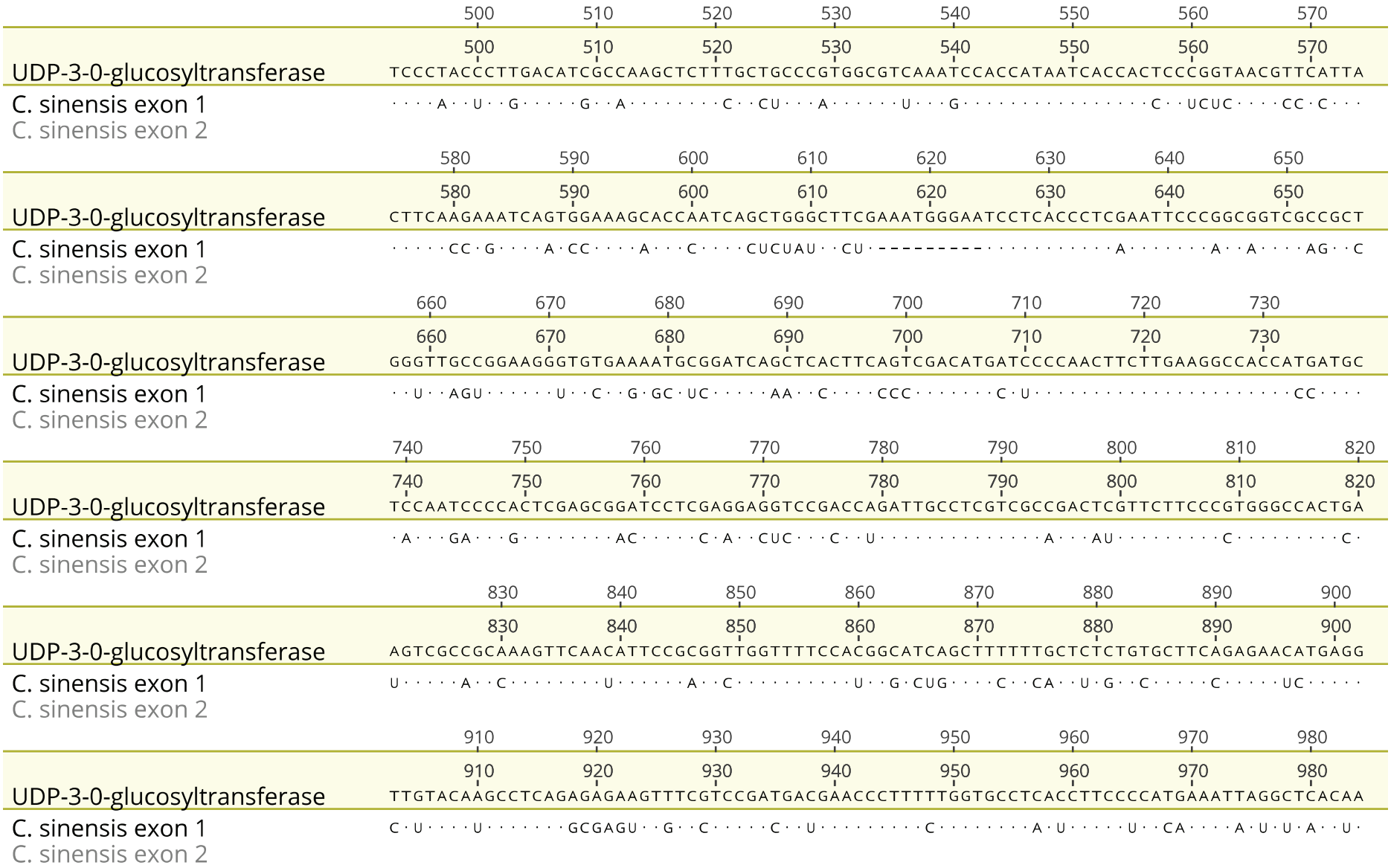
	1,450	1,460	1,470	1,480	1,490	1,500	1,510	1,520	1,530
	1,439	1,449	1,459	1,469	1,479	1,489	1,499	1,509	1,519

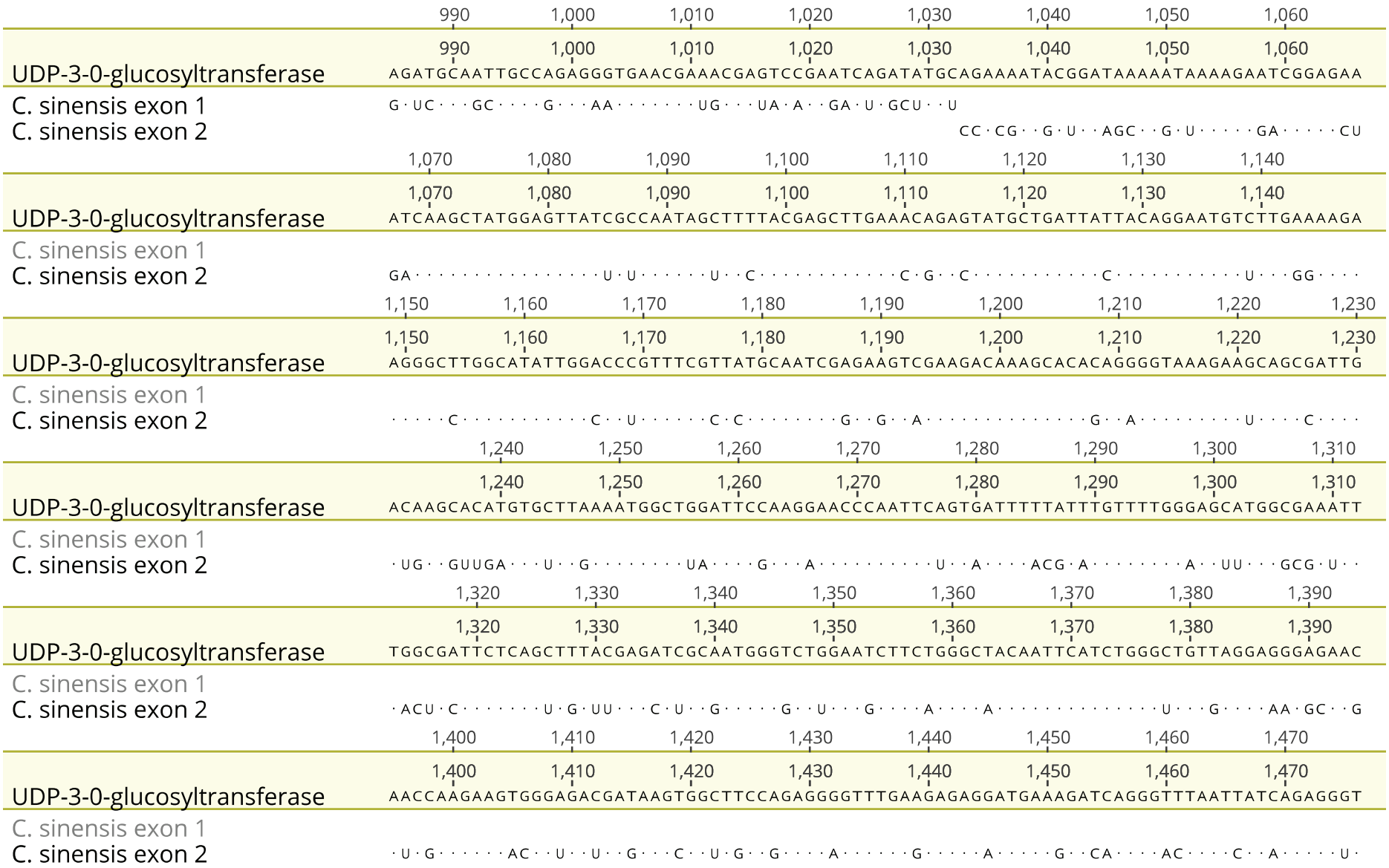
Flavanoid 3'5'-Hydroxylase NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
V. vinifera exon 1
V. vinifera exon 2 ....

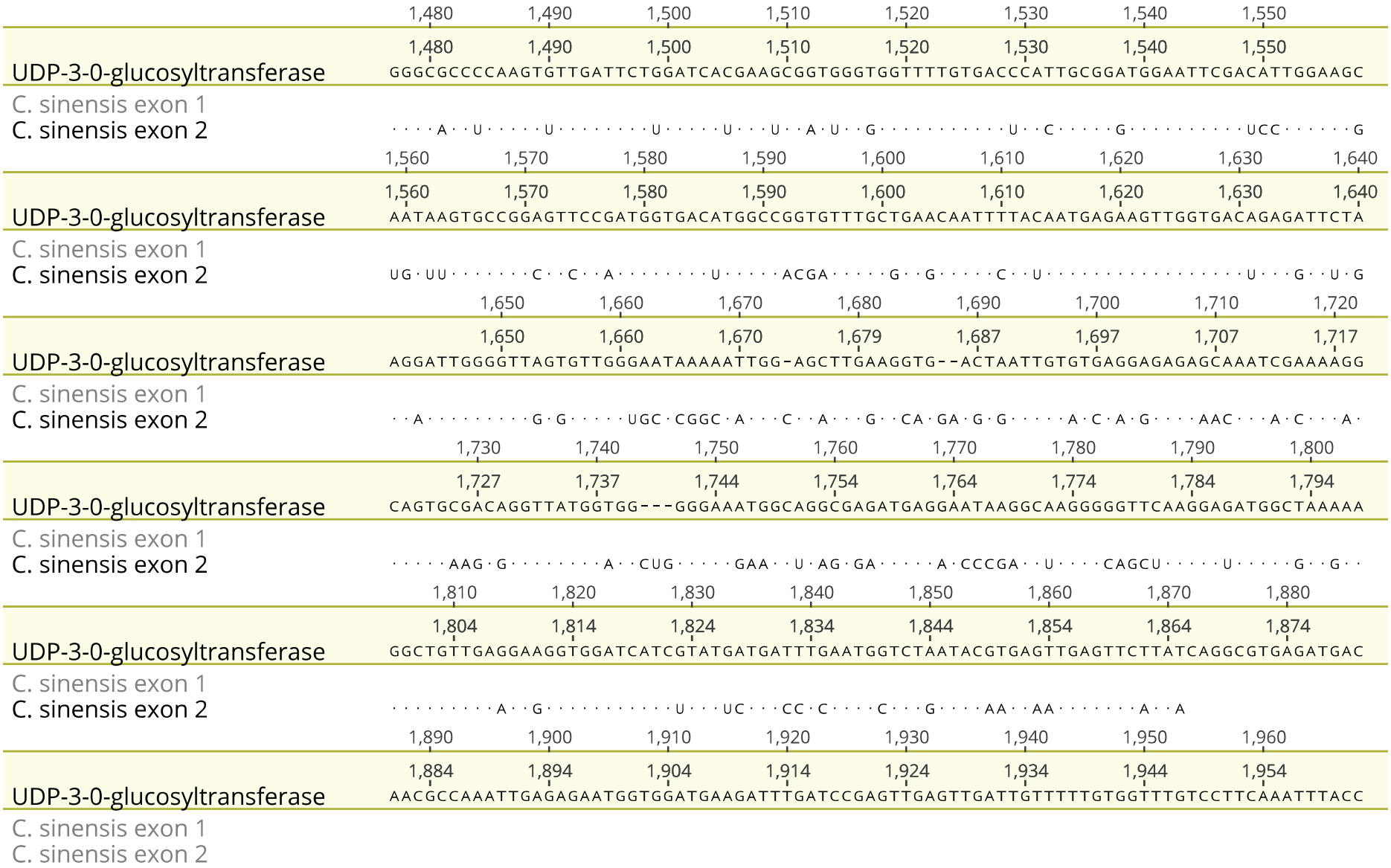


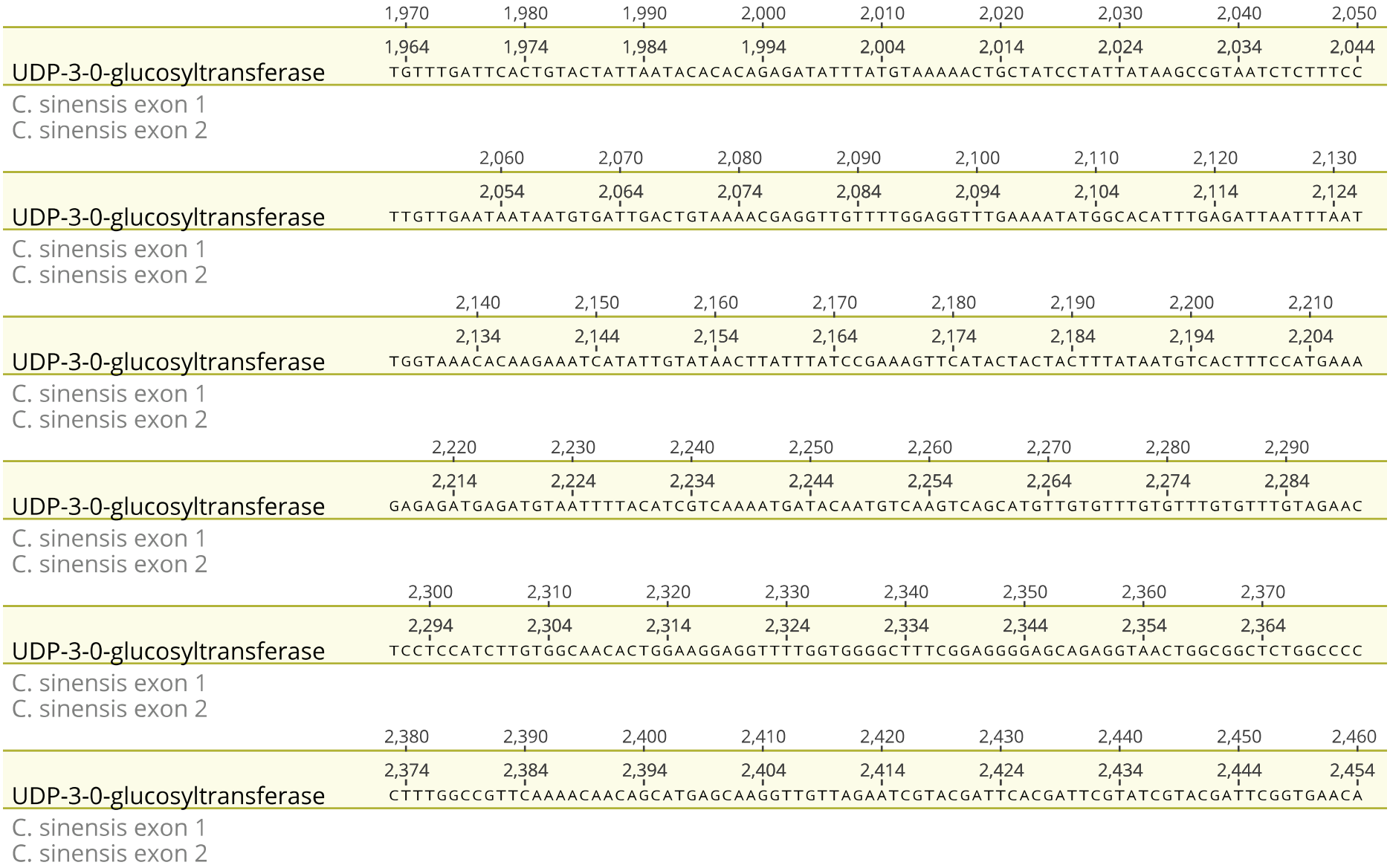


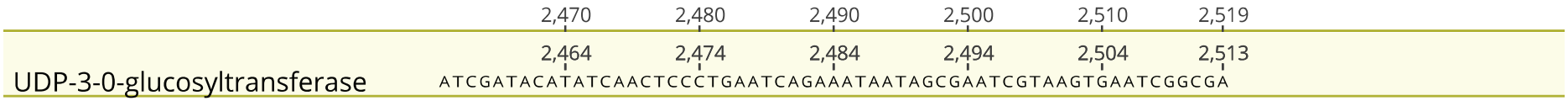


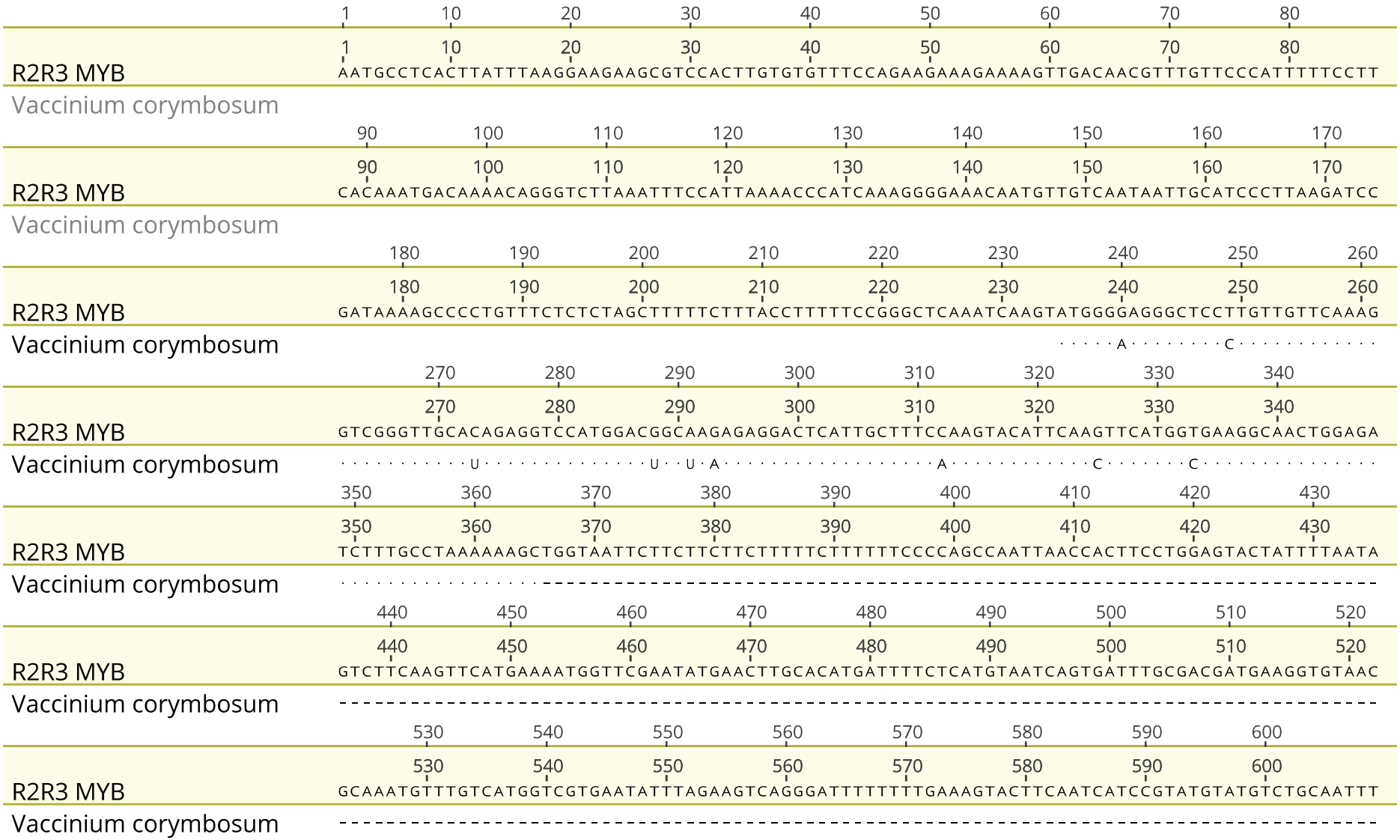


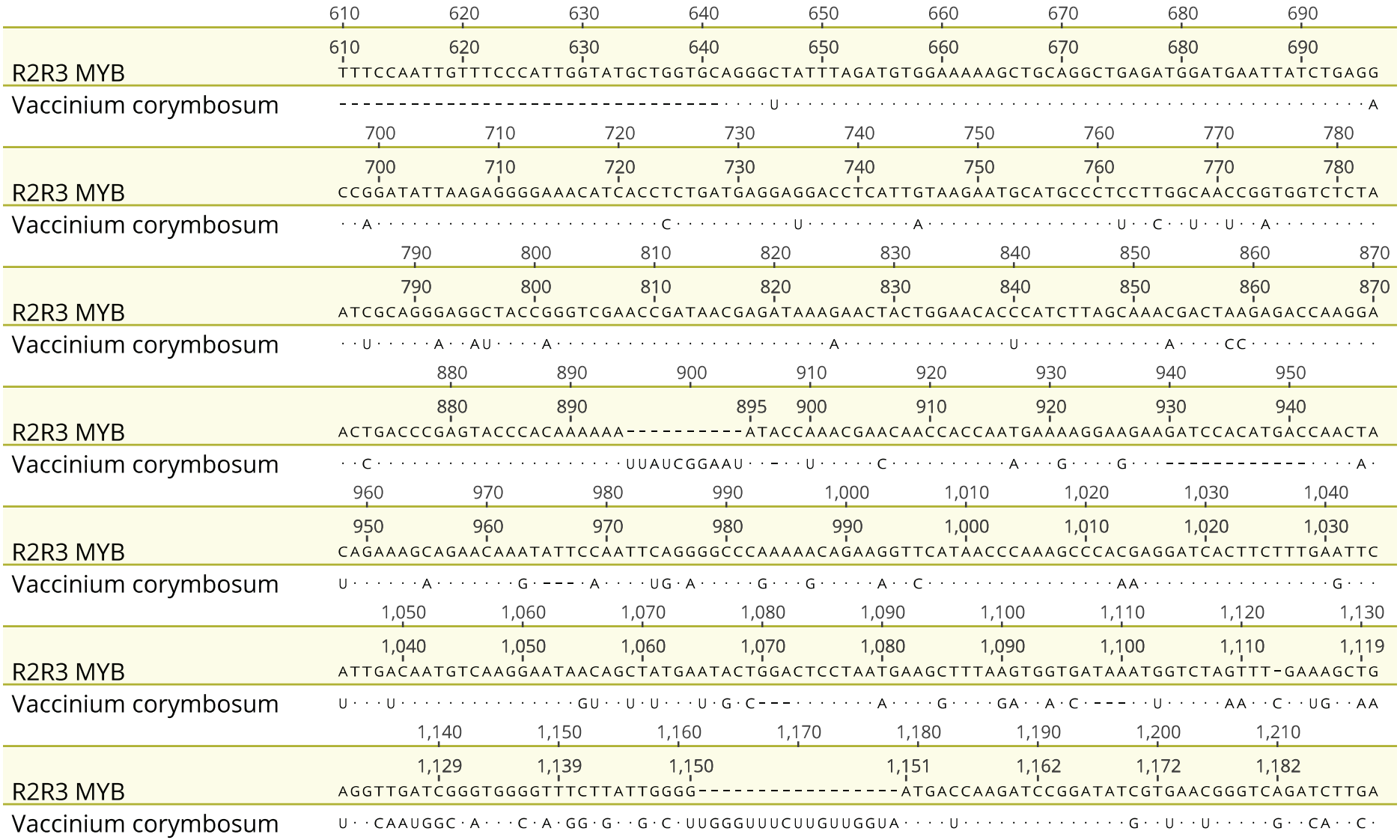


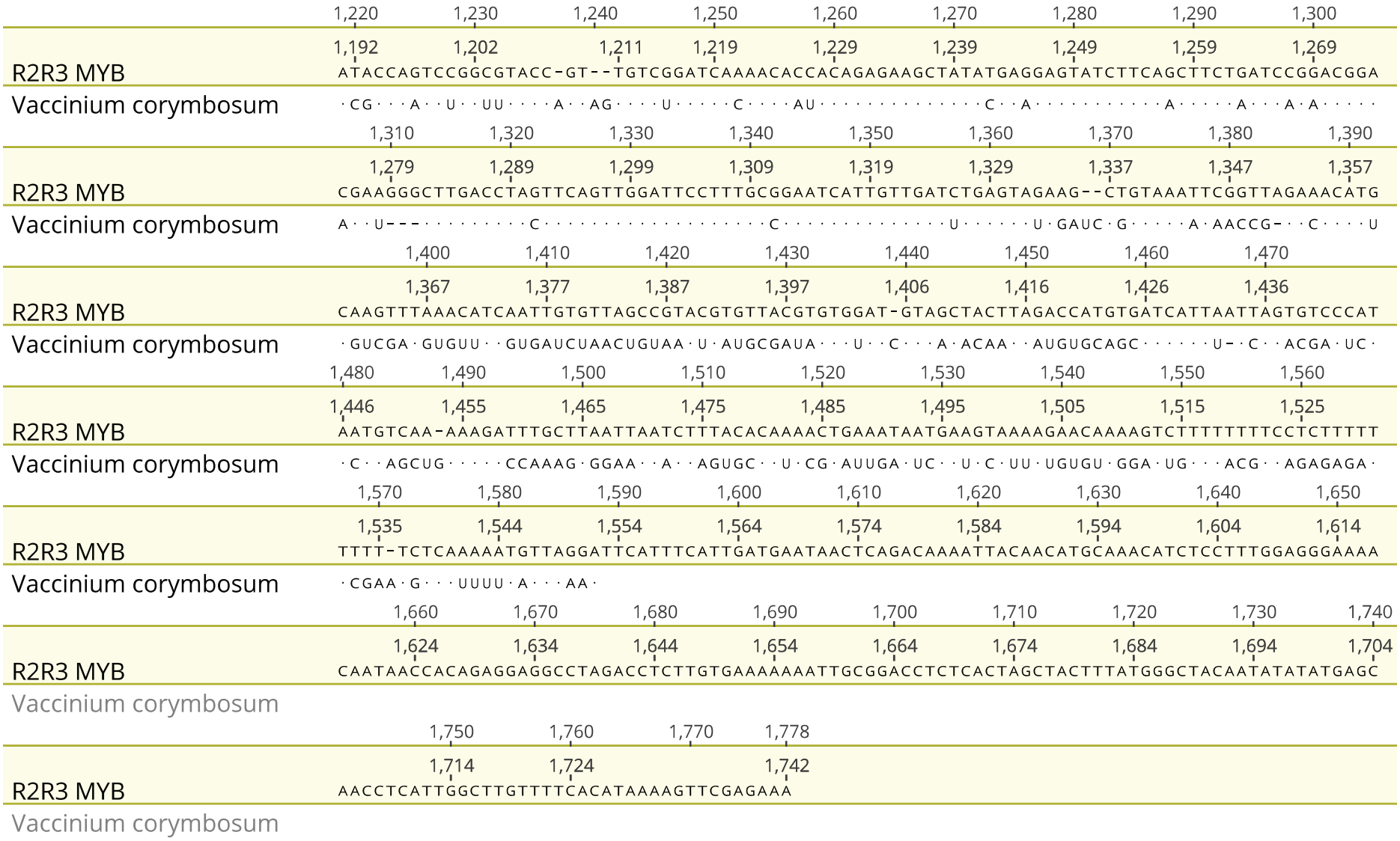


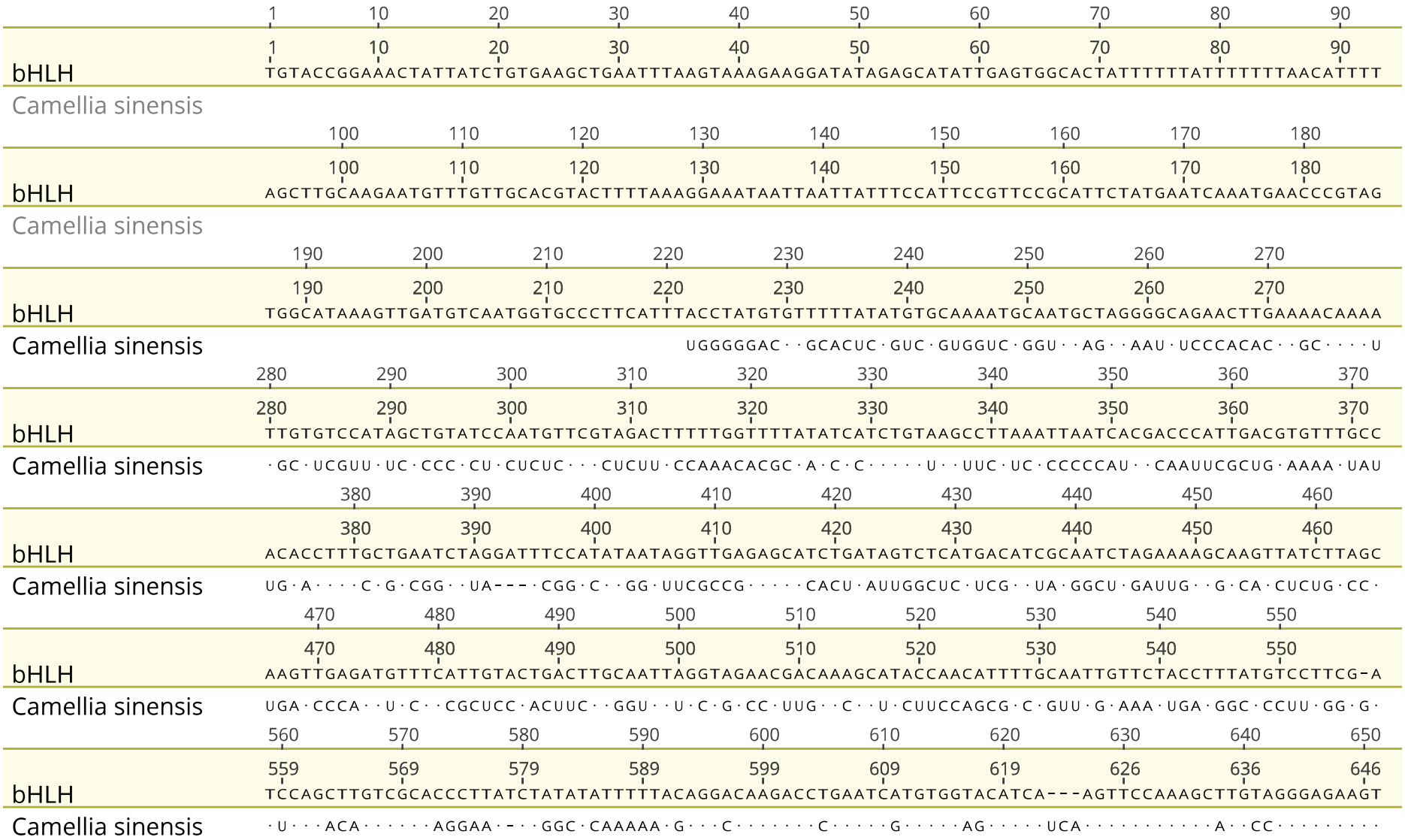


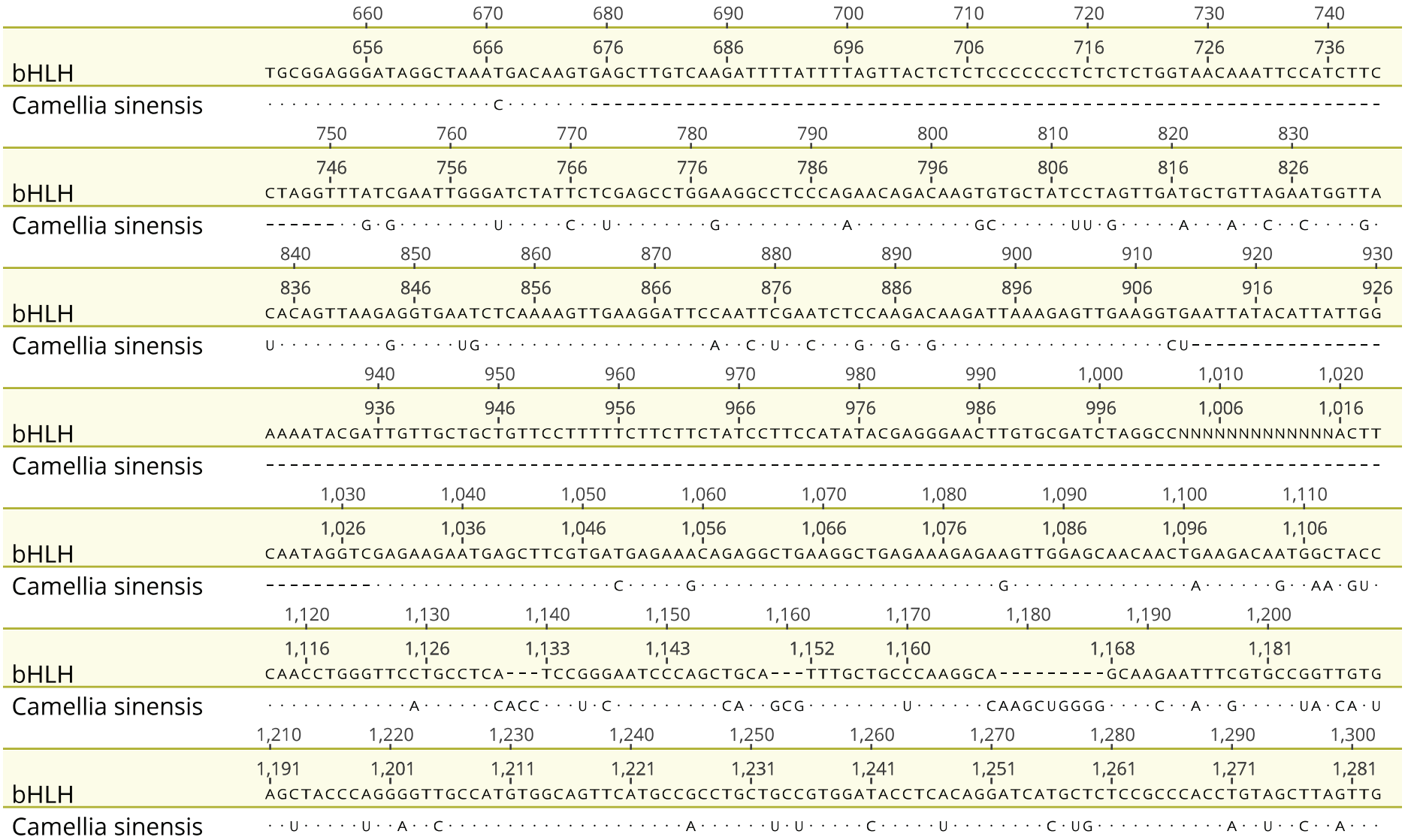


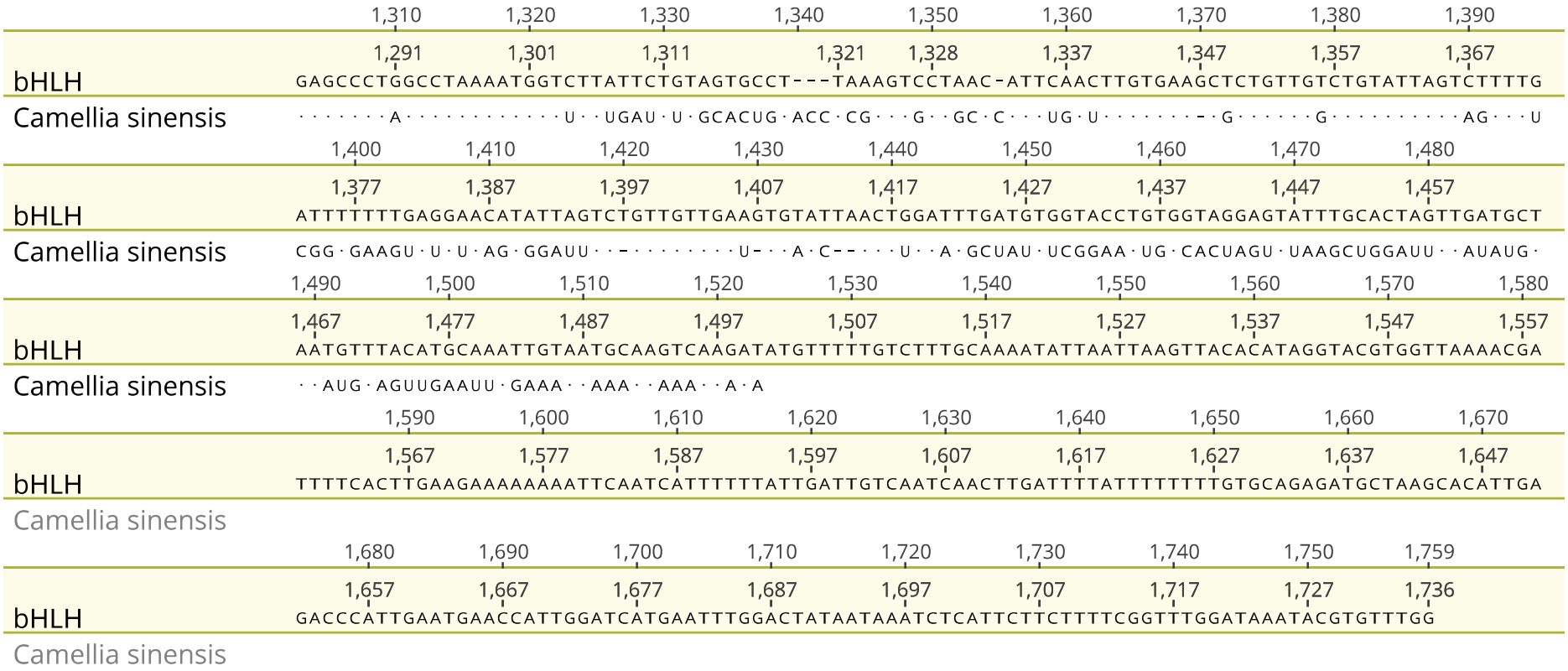


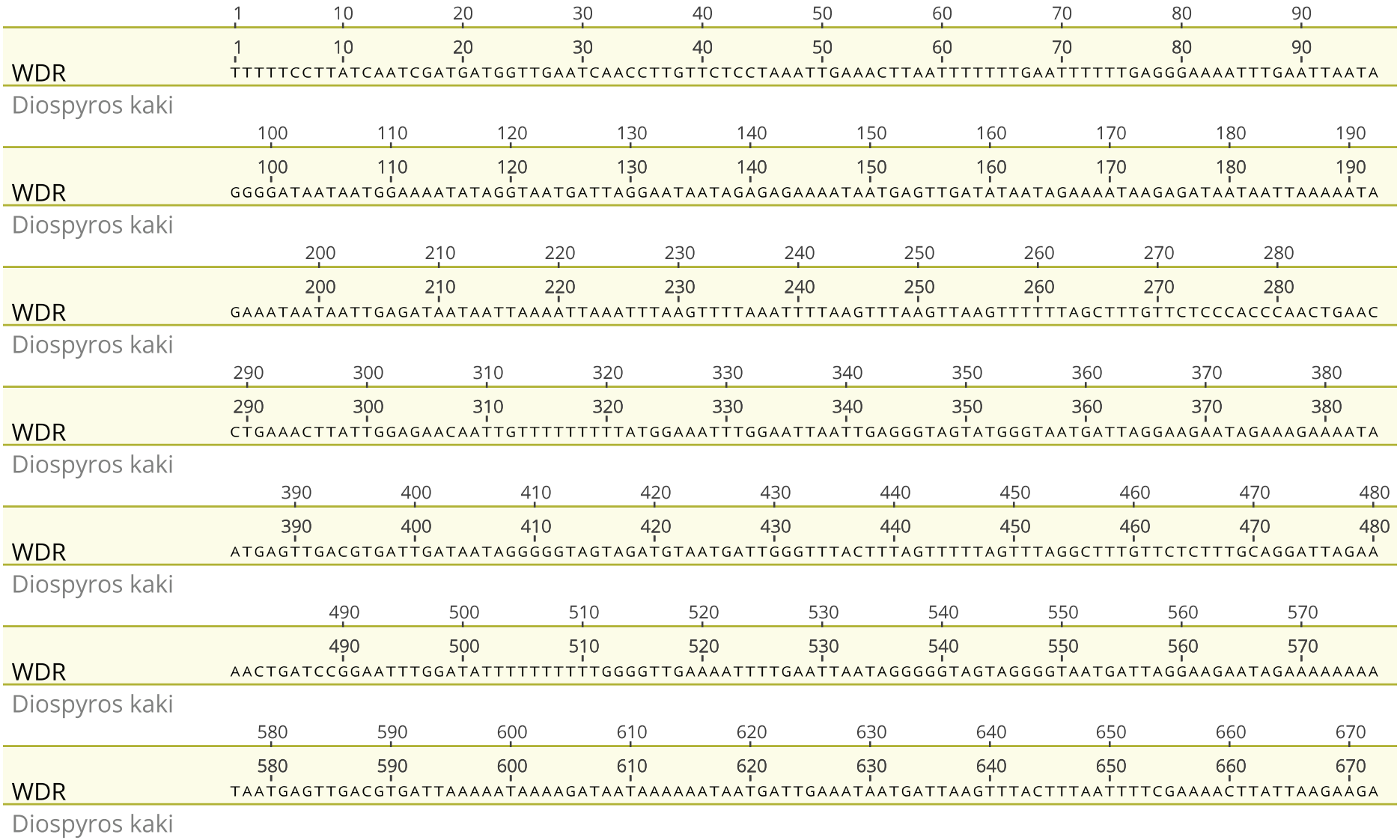


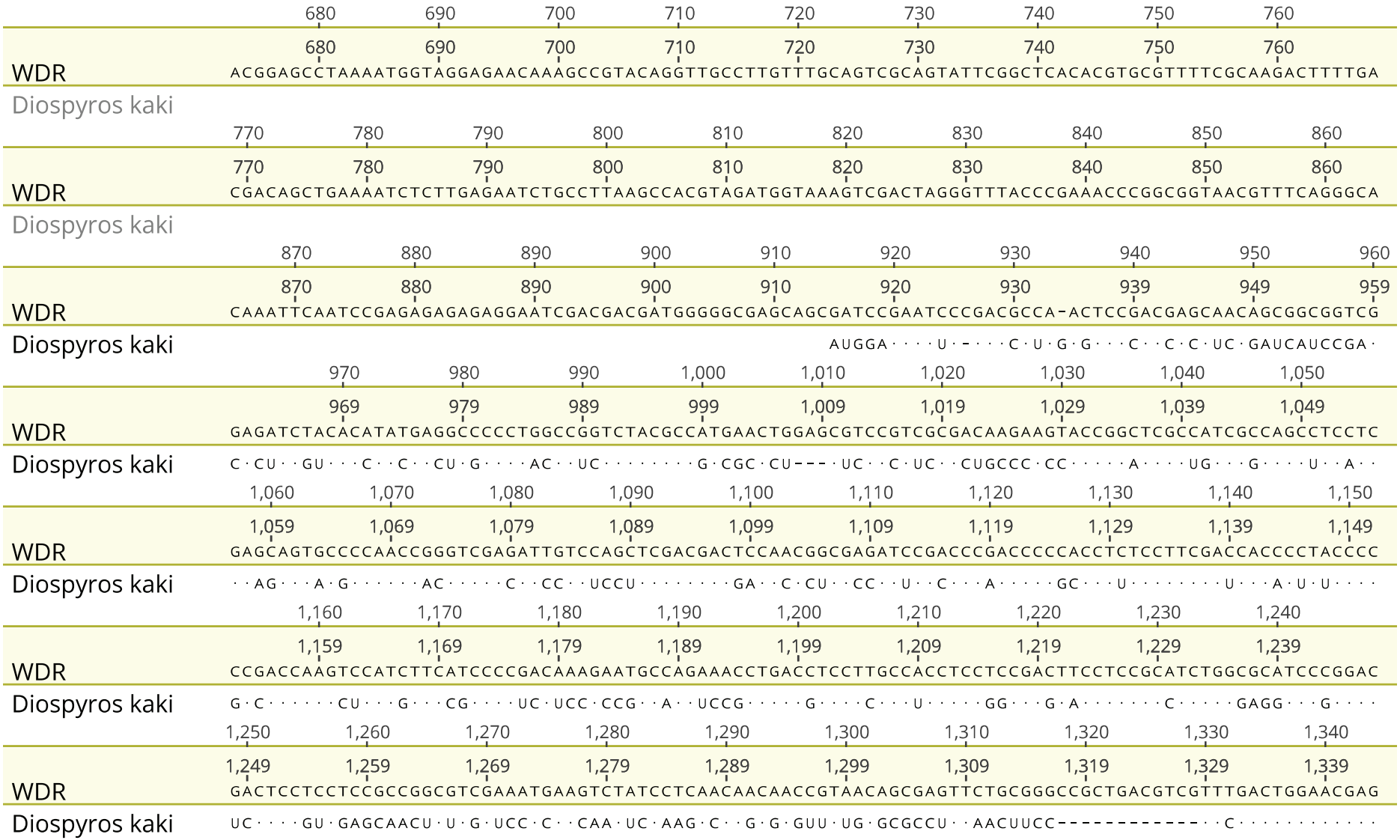


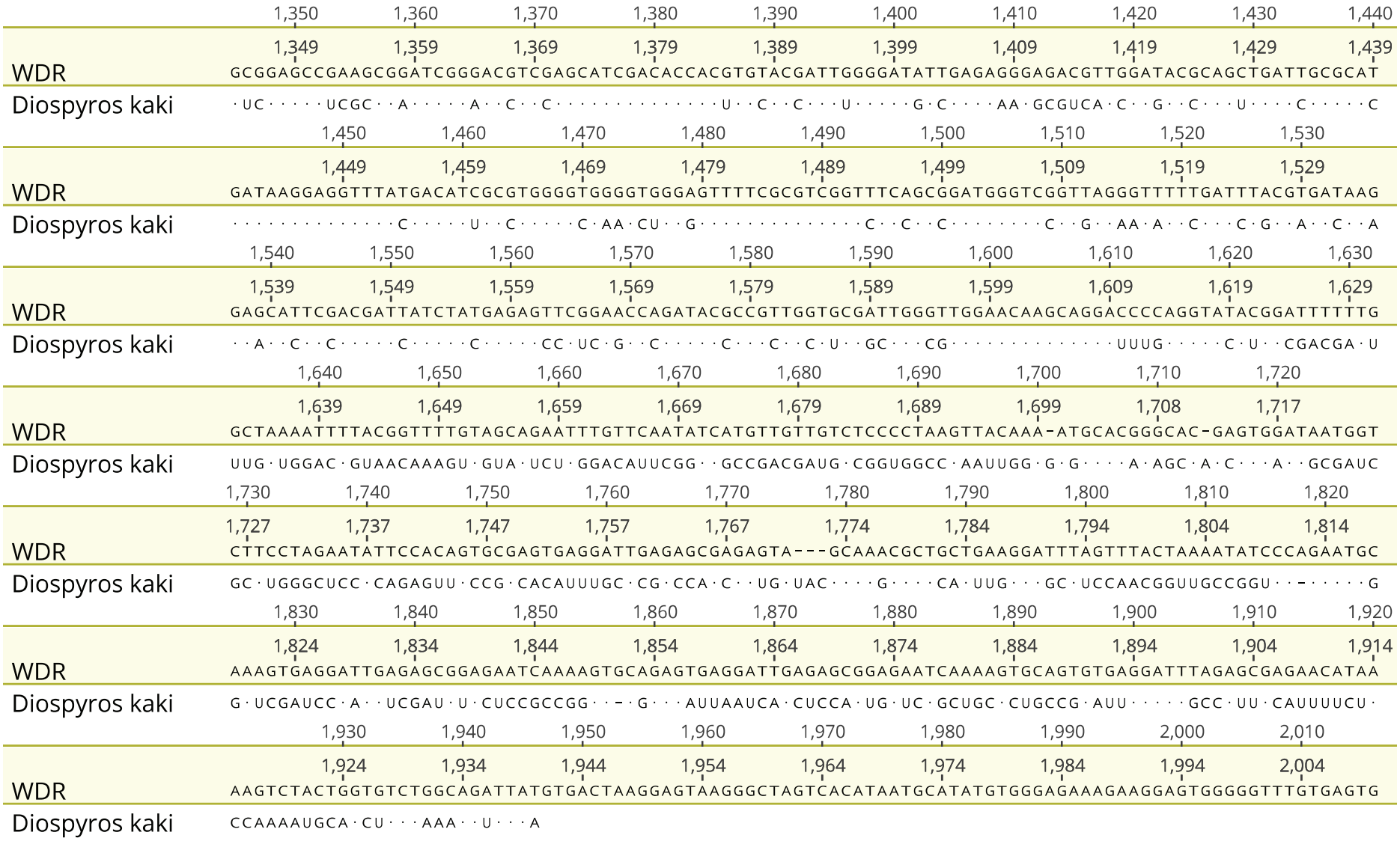


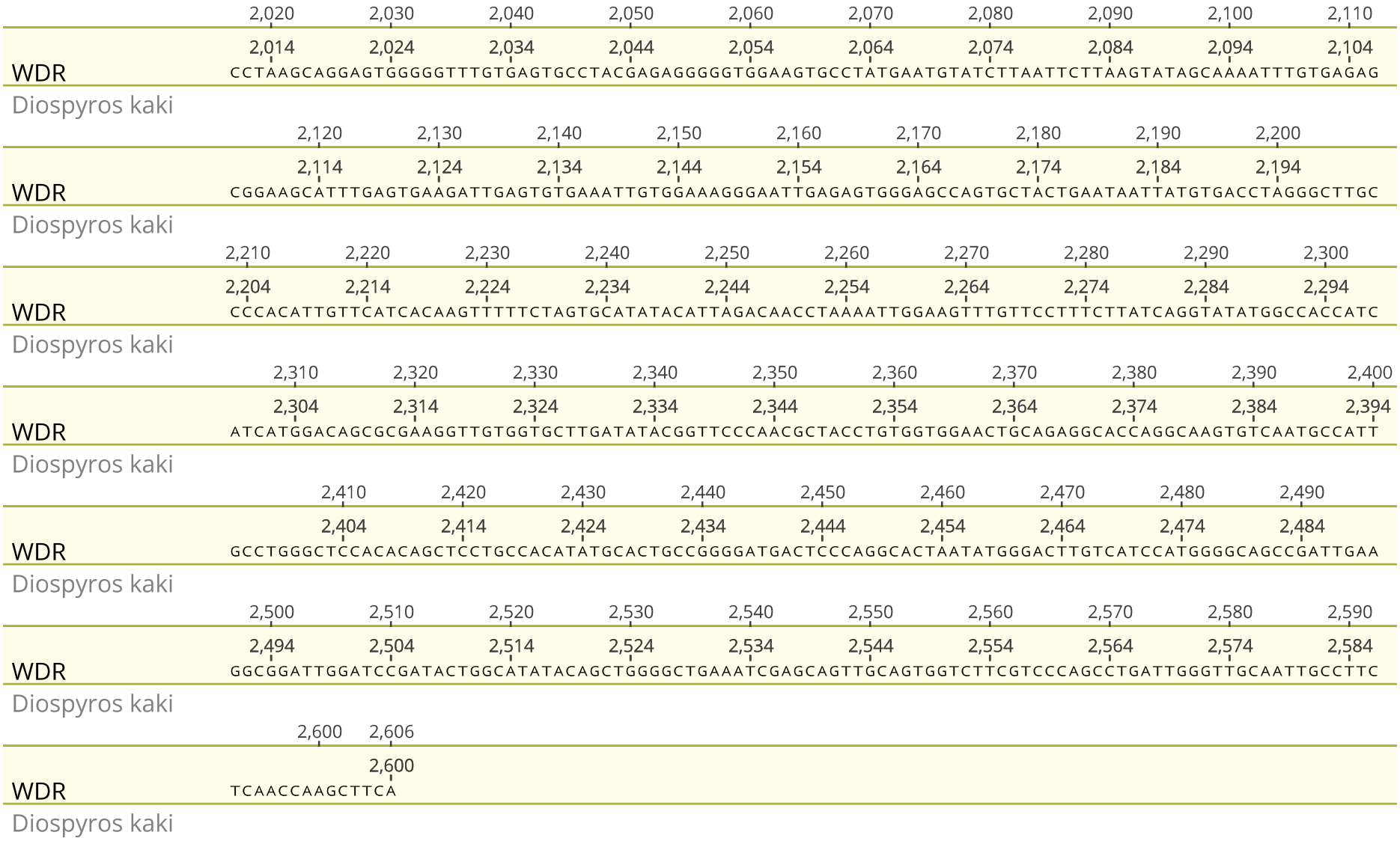












Appendix F: χ^2 test results

	ANS white	ANS red	χ2	p-value	CHI white	CHI red	χ2	p-value	CHS white	CHS red	χ2	p-value	DFR white	DFR red	χ2	p-value
1	0.06	0.09	0.01		0.01	0.06	0.041667		0.01	0.08	0.06125		0.02	0.07	0.035714	
1	0.08	0.06	0.006667		0.04	0.01	0.09		0.08	0.09	0.001111		0.04	0.08	0.02	
1	0.05	0.1	0.025	**0.838256**	0.07	0.08	0.00125	**0.715427**	0.02	0.09	0.054444	**0.732525**	0.09	0.04	0.0625	**0.730979**
2	0.09	0.44	0.278409		0.65	0.48	0.060208		0.56	0.52	0.003077		0.62	0.67	0.003731	
2	0.1	0.63	0.445873		0.57	0.62	0.004032		0.69	0.41	0.19122		0.49	0.47	0.000851	
2	0.08	0.4	0.256	**0.322129**	0.65	0.48	0.060208	**0.724258**	0.56	0.55	0.000182	**0.659215**	0.7	0.53	0.054528	**0.807907**
3	0.02	1.43	1.39028		1.428	1.2	0.04332		1.476	1.56	0.004523		1.464	1.44	0.0004	
3	0.03	1.61	1.550559		1.272	1.488	0.031355		1.452	1.464	9.84E-05		1.5	1.272	0.040868	
3	0.01	1.52	1.500066	**0.035088**	1.284	1.548	0.045023	**0.729362**	1.2	1.32	0.010909	**0.900823**	1.452	1.248	0.033346	**0.784733**
4	0.03	1.45	1.390621		1.392	1.296	0.007111		1.332	1.248	0.005654		1.404	1.2	0.03468	
4	0.02	1.49	1.450268		1.404	1.284	0.011215		1.236	1.332	0.006919		1.38	1.44	0.0025	
4	0.04	1.42	1.341127	**0.040855**	1.356	1.428	0.00363	**0.882203**	1.2	1.344	0.015429	**0.867106**	1.344	1.392	0.001655	**0.843776**
5	0.01	0.9	0.880111		0.69	0.8	0.015125		0.75	0.64	0.018906		0.65	0.71	0.00507	
5	0.02	0.9	0.860444		0.67	0.8	0.021125		0.76	0.76	0		0.8	0.79	0.000127	
5	0.01	0.9	0.880111	**0.10548**	0.64	0.68	0.002353	**0.844237**	0.61	0.7	0.011571	**0.861411**	0.63	0.66	0.001364	**0.935444**
6	0.01	0.41	0.390244		0.47	0.43	0.003721		0.59	0.58	0.000172		0.47	0.51	0.003137	
6	0.05	0.42	0.325952		0.6	0.67	0.007313		0.69	0.49	0.081633		0.4	0.47	0.010426	
6	0.05	0.41	0.316098	**0.309621**	0.69	0.58	0.020862	**0.858255**	0.45	0.41	0.003902	**0.769707**	0.47	0.48	0.000208	**0.906582**
7	0.04	0.37	0.294324		0.28	0.22	0.016364		0.22	0.28	0.012857		0.26	0.28	0.001429	
7	0.05	0.37	0.276757		0.33	0.4	0.01225		0.32	0.34	0.001176		0.31	0.34	0.002647	
7	0.05	0.3	0.208333	**0.37732**	0.33	0.38	0.006579	**0.851193**	0.24	0.3	0.012	**0.871818**	0.33	0.26	0.018846	**0.879661**
8	0.09	0.1	0.001		0.03	0.03	0		0.08	0.06	0.006667		0.09	0.08	0.00125	
8	0.05	0.03	0.013333		0.03	0.06	0.015		0.07	0.1	0.009		0.07	0.06	0.001667	
8	0.04	0.08	0.02	**0.853**	0.04	0.09	0.027778	**0.836144**	0.07	0.1	0.009	**0.875201**	0.05	0.08	0.01125	**0.905256**

	F3H white	F3H red	χ2	p-value	F3'H white	F3'H red	χ2	p-value	F3'5'H white	F3'5'H red	χ2	p-value	UDP white	UDP red	χ2	p-value
1	0.08	0.07	0.001428571		0.03	0.03	0		0.08	0.05	0.018		0.03	0.09	0.04	
1	0.01	0.07	0.051428571		0.03	0.02	0.005		0.01	0.04	0.0225		0.06	0.06	0	
1	0.05	0.09	0.017777778	**0.7904146**	0.05	0.03	0.013333333	**0.8922951**	0.08	0.08	0	**0.8405061**	0.05	0.1	0.025	**0.798761**
2	0.47	0.63	0.040634921		0.51	0.48	0.001875		0.41	0.63	0.076825397		0.45	0.44	0.000227273	
2	0.45	0.6	0.0375		0.66	0.68	0.000588235		0.66	0.43	0.123023256		0.62	0.63	0.00015873	
2	0.62	0.51	0.02372549	**0.7496084**	0.53	0.42	0.028809524	**0.8596332**	0.54	0.5	0.0032	**0.6522712**	0.44	0.4	0.004	**0.9471972**
3	1.38	1.488	0.00783871		1.224	1.428	0.029142857		1.272	1.464	0.025180328		1.332	1.26	0.004114286	
3	1.344	1.296	0.001777778		1.524	1.416	0.008237288		1.404	1.428	0.000403361		1.284	1.248	0.001038462	
3	1.476	1.212	0.05750495	**0.795575**	1.56	1.332	0.039027027	**0.7822267**	1.248	1.2	0.00192	**0.868281**	1.308	1.464	0.016622951	**0.8826854**
4	1.332	1.44	0.0081		1.32	1.296	0.000444444		1.272	1.248	0.000461538		1.44	1.284	0.018953271	
4	1.38	1.248	0.013961538		1.284	1.296	0.000111111		1.356	1.428	0.003630252		1.368	1.236	0.014097087	
4	1.236	1.212	0.000475248	**0.8806679**	1.248	1.284	0.001009346	**0.9684448**	1.248	1.368	0.010526316	**0.903766**	1.272	1.368	0.006736842	**0.8418972**
5	0.79	0.8	0.000125		0.73	0.8	0.006125		0.8	0.79	0.000126582		0.73	0.61	0.023606557	
5	0.73	0.65	0.009846154		0.71	0.77	0.004675325		0.6	0.7	0.014285714		0.6	0.63	0.001428571	
5	0.71	0.68	0.001323529	**0.915363**	0.63	0.79	0.032405063	**0.835339**	0.79	0.72	0.006805556	**0.884187**	0.71	0.68	0.001323529	**0.8710275**
6	0.65	0.49	0.052244898		0.61	0.41	0.097560976		0.66	0.68	0.000588235		0.7	0.41	0.205121951	
6	0.68	0.47	0.093829787		0.44	0.41	0.002195122		0.61	0.61	0		0.61	0.42	0.085952381	
6	0.51	0.43	0.014883721	**0.6882757**	0.54	0.57	0.001578947	**0.7502334**	0.59	0.47	0.030638298	**0.8597359**	0.43	0.41	0.00097561	**0.5889099**
7	0.21	0.38	0.076052632		0.3	0.24	0.015		0.38	0.3	0.021333333		0.28	0.37	0.021891892	
7	0.24	0.31	0.015806452		0.28	0.25	0.0036		0.4	0.24	0.106666667		0.33	0.37	0.004324324	
7	0.27	0.28	0.000357143	**0.7613784**	0.4	0.31	0.026129032	**0.8325031**	0.34	0.36	0.001111111	**0.7193555**	0.22	0.3	0.021333333	**0.8273835**
8	0.1	0.08	0.005		0.05	0.09	0.017777778		0.02	0.08	0.045		0.02	0.1	0.064	
8	0.06	0.06	0		0.07	0.04	0.0225		0.09	0.1	0.001		0.07	0.03	0.053333333	
8	0.01	0.07	0.051428571	**0.8122325**	0.09	0.02	0.245	**0.5932627**	0.04	0.01	0.09	**0.7122904**	0.08	0.08	0	**0.7319449**