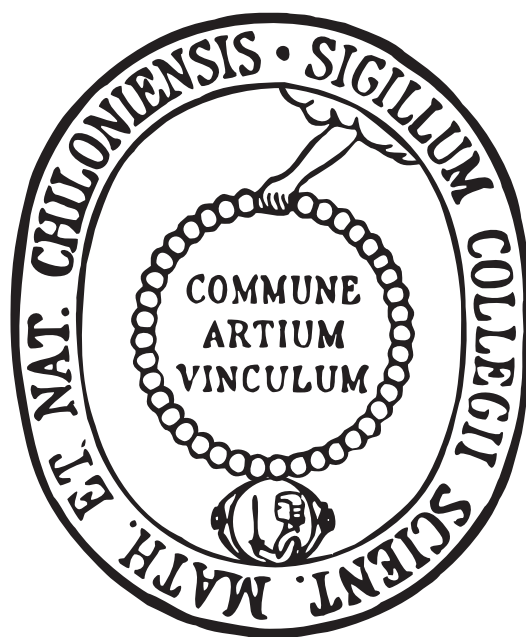


Pyrrolizidine alkaloids and homospermidine synthases in grasses (Poaceae)



Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel
vorgelegt von

Anne-Maria Wesseling

Kiel, 2017

Erster Gutachter: Prof. Dr. Dietrich Ober

Zweite Gutachterin: Prof. Dr. Eva H. Stukenbrock

Tag der mündlichen Prüfung: 27.02.2017

Zum Druck genehmigt: 27.02.2017

Thesis abstract

The specialized metabolism of plants (also called secondary metabolism) and the evolution of proteins are two fascinating and active fields of biological research which overlap in the subject of the present doctoral thesis: pyrrolizidine alkaloid biosynthesis. On the one hand, pyrrolizidine alkaloids (PAs) represent a typical group of plant specialized metabolites. These compounds are produced only by specific flowering plant taxa and are important for the plant's interactions with its environment. In the case of PAs, the compounds are mainly known as chemical defense molecules protecting the plant from herbivory. On the other hand, this specialized biosynthetic pathway provides interesting subjects for the study of protein evolution. The production of PAs in plants relies on the enzymatic activity of homospermidine synthases (HSS) which evolved from a gene duplicate of a deoxyhypusine synthase (DHS), a ubiquitous and essential enzyme of primary metabolism. The evolutionary history of the HSS is especially fascinating as the gene evolved several times independently in different flowering plant taxa.

The range of flowering plants species that produces PAs is wide and includes several grass species (family Poaceae). While there is extensive literature on HSS biochemistry and evolution as well as on natural occurrences and the structural diversity of PAs in general, only very little of it deals with PAs produced by the members of the grass family. The goal of the research presented here was to gain new insights into the specifics of this trait in grasses.

A special emphasis was put on investigating the evolution of the HSS and the paralogous DHS within the grass family. DHS as well as HSS coding sequences were identified from a variety of species. Subsequent phylogenetic analyses revealed that in the course of Poaceae evolution the *dhs* gene experienced several duplication events two of which gave rise to two distinct *hss* gene clusters specific for different subfamilies within the Poaceae. The molecular evolution is further reconstructed using in-depth selection analysis methods in combination with *in vitro* functional characterization of extant and, especially, ancestral enzymes. Using these techniques revealed how gene duplicates were retained in the respective genome either through neo- or subfunctionalization.

Insights into the prevalence of PA biosynthesis within the grass family were gained using liquid chromatography-mass spectrometry (LC-MS). A selection of grass species from the Pooideae subfamily were grown under controlled conditions and examined for PA occurrences. The results showed that the presence of plant-produced PAs is actually limited within grasses in several ways. First, the only PA structures detected were thesinine and glycosides thereof. Second, in contrast to the previously confirmed widespread occurrence of *hss* genes only very few grass species, all belonging to the *Lolium-Festuca* species complex, actually produce PAs.

Taking all results together we now have a much better understanding of grass PA biosynthesis. While our knowledge on the evolution of the grass HSS and DHS as well as PA biosynthesis in general was greatly advanced there are still many aspects which invite further investigation in future.

Kurzdarstellung

Der spezialisierte Stoffwechsel von Pflanzen (auch bekannt als Sekundärmetabolismus) und die Evolution von Proteinen sind zwei faszinierende Forschungsgebiete der Biologie, die sich im Thema der vorliegenden Doktorarbeit, der Pyrrolizidinalkaloid-Biosynthese, wiederfinden. Einerseits handelt es sich bei der Molekülgruppe der Pyrrolizidinalkaloide (PAs) um ein typisches Beispiel für spezialisierte Metabolite, da diese nur in spezifischen Angiospermen-Taxa zu finden sind. Zudem sind diese Verbindung am Zusammenspiel zwischen der Pflanze und ihrer Umwelt beteiligt. So sind die PAs vor allem bekannt für ihre Funktion in der chemischen Abwehr von Herbivoren. Andererseits liefern die spezialisierten Stoffwechselwege auch interessante Beispiele für die Proteinevolutions-Forschung. In Pflanzen katalysiert eine Homospermidinsynthese (HSS) den ersten Schritt der PA-Biosynthese. Das entsprechende HSS-Gen ist aus einem Genduplikat der Desoxyhypusinsynthese (DHS) entstanden. Dabei ist besonders interessant, dass das HSS-Gen im Laufe der Angiospermen-Evolution mehrere Male aus unterschiedlichen Duplikationsereignissen heraus, in verschiedenen Pflanzengruppen unabhängig voneinander entstanden ist.

Die Biosynthese von PAs findet man in den unterschiedlichsten Angiospermen-Arten, und so auch in einigen Grasarten (Familie Poaceae). Allerdings ist über gras-spezifische HSS sowie über die Verbreitung und Vielfalt der PA-Strukturen in Gräsern, im Gegensatz zu anderen Pflanzengruppen, sehr wenig bekannt. Die Forschungsergebnisse, die in dieser Arbeit vorgestellt werden, sollen daher die Wissenslücke, die in diesem Bereich existiert, füllen.

Ein Schwerpunkt der Untersuchungen lag auf der Evolution der HSS und DHS in der Gras-Familie. Hierfür wurden die codierenden Sequenzen der entsprechenden Gene aus verschiedenen Gras-Arten identifiziert und phylogenetisch untersucht. Dabei hat sich gezeigt, dass innerhalb der Gräser die DHS mehrere Duplikationen durchlaufen ist und dass dabei zweimal unabhängig voneinander in unterschiedlichen Gras-Untergruppen eine HSS entstanden ist. Die Evolutionsgeschichte dieser Gene wurde weiter unter zur Hilfenahme von Selektionsanalysen weiter aufgeklärt und mit Ergebnissen aus *in vitro* funktionellen Analysen von rezenten und insbesondere von ancestralen Proteinen ergänzt. Die Ergebnisse geben uns grundsätzliche Einblicke über die Evolution von Genduplikaten und wie diese zum Beispiel durch Mechanismen wie die Neo- oder Subfunktionalisierung im Genom fixiert werden können.

Um zu ermitteln wie weit die Fähigkeit zur PA-Biosynthese innerhalb der Gräser verbreitet ist, wurden Pflanzen aus der Gras-Unterfamilie Pooideae unter kontrollierten Bedingungen angezogen und dann mittels Flüssigchromatographie gekoppelt mit Massenspektrometrie (LC-MS, von englisch *liquid chromatography-mass spectrometry*)

auf das Vorkommen von PAs untersucht. Diese Untersuchungen haben ergeben, dass, im Gegensatz zu den innerhalb der Gräser weitverbreitete HSS-Genen, die PA-Biosynthese tatsächlich in nur sehr wenigen Gras-Arten vorzufinden ist. Alle identifizierten PA-produzierende Arten gehören dabei zum *Festuca-Lolium*-Artkomplex. Zudem ist die Vielfalt der PA-Strukturen in den untersuchten Arten nur auf die Verbindung Thesinin und verschiedene Thesinine-Glykoside beschränkt.

Alle Ergebnisse zusammengenommen haben wir jetzt viel bessere Kenntnisse von der PA-Biosynthese in Gräsern und dazu viele interessante Anhaltspunkte für die weitere Erforschung dieser Eigenschaft in Gräsern.

Contents

General introduction.....	1
Protein evolution and gene duplications.....	1
Specialized metabolism in plants: Pyrrolizidine alkaloids	7
Thesis objective and outline.....	15
Chapter 1: Complex molecular evolution of the paralogous <i>dhs</i> and <i>hss</i> genes in grasses (Poaceae).....	17
Introduction	17
Results.....	20
Discussion.....	29
Materials and Methods	34
Acknowledgments	38
Supplementary material	39
Chapter 2: Characterization of pyrrolizidine alkaloid occurrence in the grass subfamily Pooideae (Poaceae).....	65
Abstract	65
Introduction	66
Results.....	67
Discussion.....	75
Experimental.....	77
Acknowledgments	80
Supplementary material	81
Chapter 3: Further analyses on grass HSS and DHS	85
Background.....	85
Results.....	85
Discussion and outlook.....	92
Materials and Methods	96
Chapter 4: Further LC-MS based investigations on PA biosynthesis in grasses	99
Background.....	99
Results and Discussion	99
Material and methods	103
Conclusions and perspectives	105
Contributions	107
Danksagung & Acknowledgements.....	109
Bibliography.....	111
Erklärung.....	123

General introduction

Protein evolution and gene duplications

Proteins are complex biomolecules and essential components of every living organism. They perform a diverse set of functions, for example, serving as structural elements, acting as transporters, or regulating metabolic processes. In accordance with the multitude of biological processes, we find a great diversity of proteins in nature (see e.g. the UniProt database; UniProt Consortium, 2015).

In the study of proteins one of the questions that arises is how new genes and thereby new proteins evolve. Extensive research has identified **the duplication of genes** as the evolutionary process that has contributed most to the expansion of the genetic repertoire. Gene duplication refers to any molecular process during which DNA regions containing one or more genes are replicated and then integrated into the genome resulting one organism having two copies of the same gene. Through the generation of new genetic material gene duplications provide opportunities for evolution to develop new and specialized functions, allowing organisms to better adapt to changing environments (Zhang, 2003). Gene duplications can lay the groundwork for the evolution of species-specific functions, might drive speciation and have even been suggested to have played a major role in the origin and extensive diversification of angiosperms (Zhang, 2003; De Bodt et al., 2005).

Genes (and their protein products) that are related due to their common origin by gene duplication are **paralogous** to each other and form a gene family. The prevalence of gene duplications in all domains of life is evident by the great number, size and ubiquity of gene families that are found in all domains of life (Zhang, 2003). Using genomic sequence data Lynch and Conery (2000) calculated a gene duplication rate of one duplication per gene per hundred million years in eukaryotes, which is comparable to the nuclear substitution rate.

Four different molecular processes are considered to be responsible for the generation of the majority of gene duplicates (Hahn, 2009). Unequal crossing-over between sister

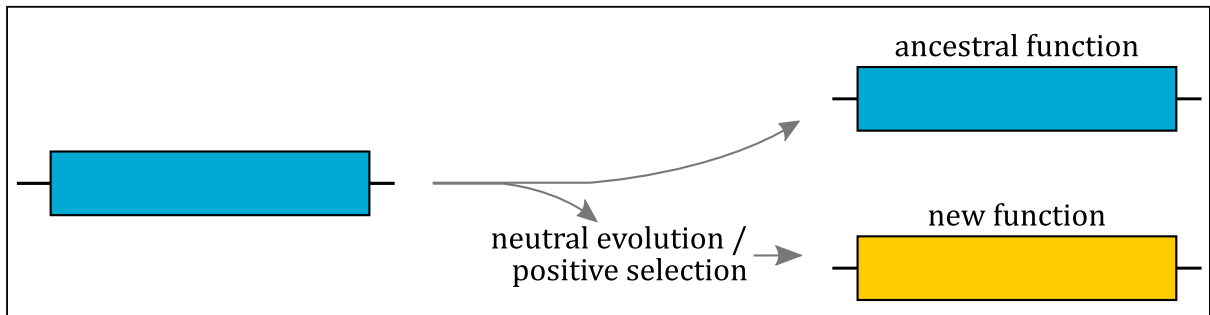
chromatids can result in a tandem arrangement of gene duplicates on one of the chromatids and gene loss on the other, a process referred to as **tandem duplication** (e.g. Shoja and Zhang, 2006). During **retrotransposition** an mRNA molecule is reverse transcribed to DNA which is reinserted at a random position in the genome while during **duplicative transposition** replicated DNA sequences are transposed to other regions of the genome without involving an RNA intermediate (Hahn, 2009). In the fourth and most extreme cases an entire genome is duplicated. These **whole genome duplications** (WGDs) are either the result of aberrant cell division during which chromosomes fail to separate (resulting in autopolyploidy) or of an hybridization event which combines the genomes of two closely related species (allopolyploidy) (Prince and Pickett, 2002). The contribution of each mechanism to the overall generation of gene duplicates has been shown to vary greatly between genomes (Hahn, 2009). For example, while WGDs are rare in animals they occur frequently in the plant kingdom (Prince and Pickett, 2002).

Evolution after gene duplication

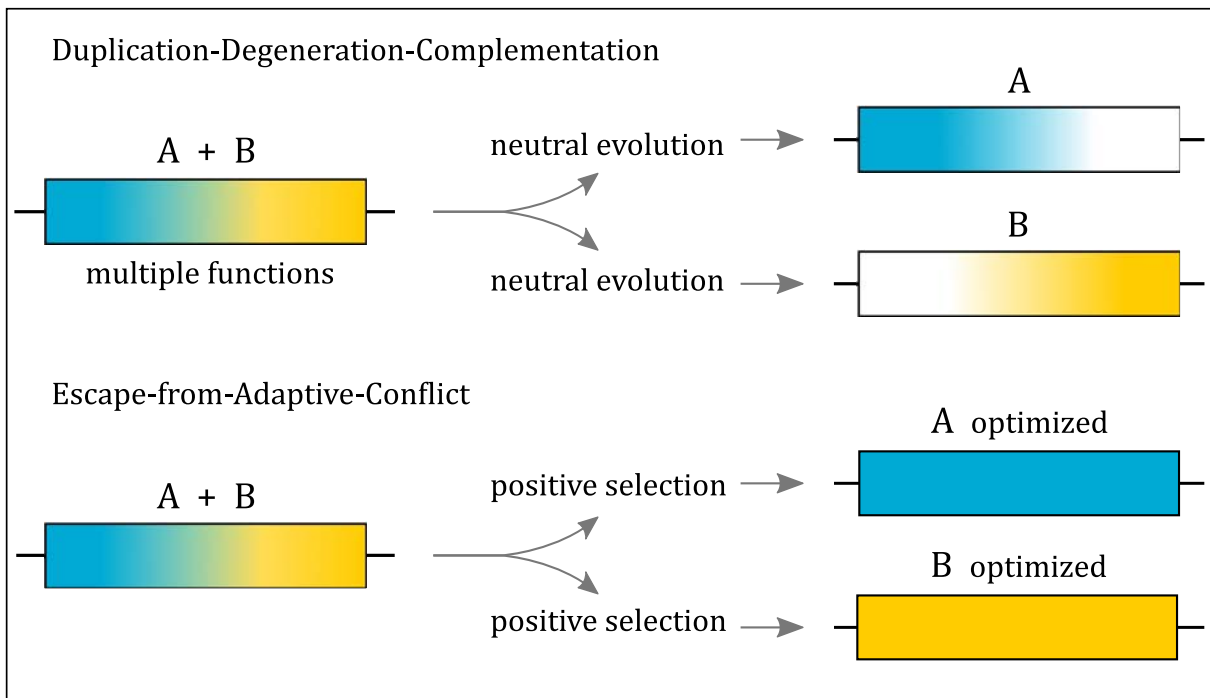
The duplication of a gene constitutes only the very first step in the emergence of new proteins. Gene duplicates mostly start out identical and only over time will diverge in their nucleotide sequence and therefore their function. Over many decades various potential evolutionary fates of gene duplicates have been explored and described in gene duplication models (for a historical review see Taylor and Raes, 2004). Discussion on this subject has been extensive and in their review Innan and Kondrashov counted a total of ten different gene duplication models in the literature (Innan and Kondrashov, 2010).

The first question after a gene duplication event is whether both gene copies will be retained in the genome or one of the copies is lost again. Different assumptions about the benefits of the duplication of genes themselves are made in the various models which consequently make different predictions about the probability of both duplicates being fixed. For example, in some models it is presumed that having two copies of the same gene is neither beneficial nor deleterious and therefore fixation or gene loss would be the result of neutral processes such as genetic drift (see e.g. Ohno, 1970; Innan and Kondrashov, 2010). In various models a generally beneficial increase of gene dosage has

Neofunctionalization



Subfunctionalization



Nonfunctionalization

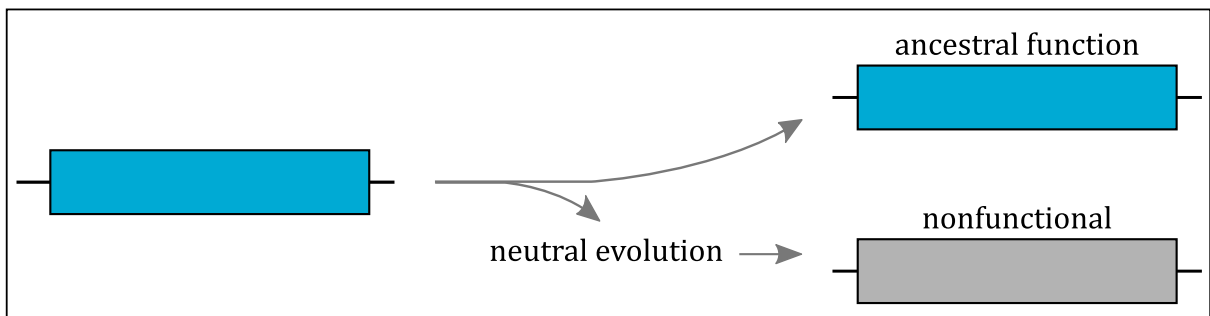


Fig. 1: Overview of the most prominent gene duplication models

Depicted are the different courses that functional divergence might take after a gene duplication event. Boxes represent genes; colors (blue, yellow) and letters (A, B) indicate differing gene functions; neutral evolution and positive selection refer to the selective forces which act on coding regions of the gene duplicates.

been predicted which would lead to the gene duplicates to be positively selected for leading to rapid fixation of both gene copies (Kondrashov et al., 2002; Kondrashov and Kondrashov, 2006)

In contrast to these, the **dosage-balance hypothesis** predicts that gene duplication events that disturb the balance (stoichiometry) of complex protein interaction networks (e.g. single-gene duplications) are generally harmful and therefore selected against (Papp et al., 2003). Indeed, there is a bias in gene loss and retention depending on whether a gene's function involves extensive protein-protein interactions (Papp et al., 2003; Li et al., 2006; Conant and Wolfe, 2008). In support of the dosage-balance hypothesis, it was shown that the functional bias is reversed when the whole genome is duplicated and gene duplicates which were lost after small-scale duplication events were now more likely to be retained (Maere et al., 2005; Conant and Wolfe, 2008).

A different and much discussed aspect in gene duplicate evolution is in which “direction” gene functions evolve and which evolutionary forces drive sequence evolution and thereby functional divergence (see box 1 for more information on selection with regards to coding sequences). In the following, prominent models dealing with the question of functional divergence are presented in accordance with the classification of Conant and Wolfe (2008). A schematic overview is given in figure 1.

The **neofunctionalization** model postulates that one of the gene copies will acquire a function that was not present before the duplication event while the other copy conserves the function of the ancestral gene. The model was first proposed by Ohno in 1970 (Conant and Wolfe, 2008). This specific outcome is possible if the second gene copy is redundant and therefore under no functional constraints. The copied gene may accumulate mutations which affect its function but which are ultimately neutral because of the redundancy (Ohno, 1970; Conant and Wolfe, 2008). Should a new function emerge and prove to be beneficial, positive selection would further drive the optimization of the gene through more beneficial mutations (Zhang et al., 1998; Conant and Wolfe, 2008). However, considering that out of all possible mutations most will cause the loss of function, **nonfunctionalization** of one of the duplicates is the most likely outcome under relaxed selection.

Another term consistently found in the literature dealing with gene duplications is **subfunctionalization**. Conant and Wolfe (2008) use this as an umbrella term for two

Box 1: Coding sequences, mutations and selection

When discussing point mutations in the coding region of genes it is important to differentiate between **synonymous and nonsynonymous substitutions**. Within the universal genetic code up to six different nucleotide triplets (so-called codons) code for one of 20 amino acids. Therefore, not every nucleotide substitution will result in a change of amino acids of a protein and is therefore synonymous. In the context of protein evolution only those mutations are relevant that cause a change in a protein's amino acid sequence and could therefore potentially change protein functionality.

The ratio of synonymous substitution rate (d_S) to nonsynonymous substitution rate (d_N) is known as ω and can inform us on the type that has affected a coding sequence (see e.g. Nielsen and Yang, 1998).

$$\frac{\text{rate of nonsynonymous substitutions}}{\text{rate of synonymous substitutions}} = \frac{d_N}{d_S} = \omega$$

d_S = number of synonymous substitutions per synonymous site

d_N = number of nonsynonymous substitution per nonsynonymous site

In most cases, nonsynonymous substitutions will be selected against as most mutations are deleterious, an evolutionary process also called **purifying selection**. This will usually result in a synonymous substitution rate that is higher than the nonsynonymous substitution rate ($\omega < 1$). In the case of no functional constraints acting on a protein-coding gene it may **evolve neutrally** and accumulates both synonymous and nonsynonymous substitutions at the same rate ($\omega = 1$). As a third option a protein-coding gene may experience directional or **positive selection** meaning that specific beneficial amino acid changes accumulate, resulting in **adaptive evolution**. In this case the nonsynonymous substitution rate can be higher than the synonymous substitution rate resulting in ω larger than 1.

Codon substitution models have been developed to describe nucleotide sequence evolution based on parameters such as d_N and d_S . These substitution models are implemented in programs used to determine which type of selection has acted on a specific coding sequence. Since their first conception (Goldman and Yang, 1994; Muse and Gaut, 1994), codon substitution models have been modified to allow for **variation of ω between different sites, lineages, or a combination of both** to increase the power of selection analyses which allow evolutionary biologists detailed insight into the evolution of genes (e.g. Weadick and Chang, 2012; Yang et al., 2005; Zhang et al., 2005).

different models addressing the evolution of multifunctional genes: the **duplication-degeneration-complementation** (DDC) and the **escape-from-adaptive-conflict** (EAC) model. The first, DDC, states that both gene copies of a multifunctional protein might experience mutations. And while some protein functions are lost (degeneration) the mutations are ultimately neutral because the lost activity is compensated for by the other gene copy (complementation). Joint protein function and activity would then match that of the ancestral protein (He and Zhang, 2005). The EAC model, on the other hand, assumes that several functions are combined within an ancestral protein and that each function cannot be optimized without compromising the other (conflict). Gene duplication then provides the opportunity for optimization of a subset of functions (escape) through the accumulation of adaptive mutations (Innan and Kondrashov, 2010; Stoltzfus, 1999).

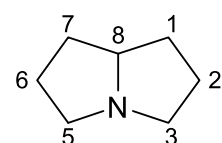
In addition to the models described above, it was also suggested that in some cases gene duplicate evolution follows a combination of gene duplications models (e.g. He and Zhang, 2005; Rastogi and Liberles, 2005). To determine which gene duplication model a specific set of duplicates has followed, evolutionary biologists use phylogenetic **selection analyses** which are based on codon substitution models (see box 1 for an overview and Murrell et al., 2012; Zhang et al., 1998 for exemplary studies). For a complete picture of the evolutionary processes paralogous protein pairs under investigation are functionally characterized in combination studying selection pressure (e.g. Kaltenegger et al., 2013).

Specialized metabolism in plants: Pyrrolizidine alkaloids

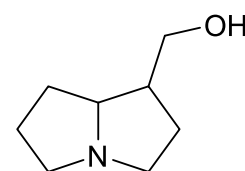
Plant specialized metabolites (also referred to as secondary metabolites) are organic, low molecular weight compounds which by definition are only found in specific plant lineages while missing from others (Fraenkel, 1959; Pichersky et al., 2006). They represent adaptations to the plant's ecological niche by interacting with biotic and abiotic factors in the environment, for example, as pollinator attractants, pathogen or herbivore defense molecules (Fraenkel, 1959; Pichersky et al., 2006; Pichersky and Lewinsohn, 2011). Pyrrolizidine alkaloids (PAs) are an exemplary class of specialized metabolites as they are found only in specific plant lineages: PA-containing species have been reported, for example, for the daisy, borage, morning glory and legume families (Asteraceae, Boraginaceae, Convolvulaceae and Fabaceae) but have also been found in monocots, specifically, in some orchid and grass species (Orchidaceae and Poaceae) (Crow, 1962; Hartmann and Witte, 1995; Stegelmeier et al., 1999; Koulman et al., 2008).

All PAs have in common that their structure is based on a pyrrolizidine, a *N*-containing double ring system (fig. 2) (Hartmann and Witte, 1995). Plant-produced PA structures are generally derived from 1-hydroxymethylpyrrolizidine (fig. 2), the so-called necine base, and occur as ester alkaloids where the necine base is esterified with necic acids forming, for example, monoesters, open-chain or macrocyclic diesters (fig. 2) (Hartmann and Witte, 1995). Pairing off the various necine bases with different necic acids gives rise to a great variety of PA structures with more than 350 different naturally occurring PAs (not counting their *N*-oxide variants) having been identified to date (Stegelmeier et al., 1999; Roeder, 2000).

The function PAs are mainly known for is herbivore deterrence by intoxication (e.g. Dam et al., 1995; Macel et al., 2005). After ingestion by (non-adapted) herbivores PAs that



pyrrolizidine



1-hydroxymethylpyrrolizidine

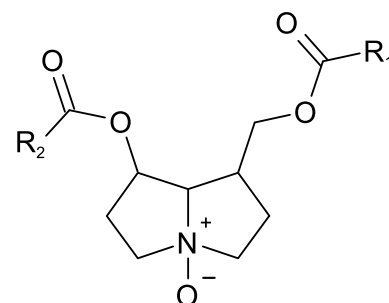
pyrrolizidine alkaloid *N*-oxide

Fig. 2: basic PA structures

R_1 -, R_2 -COOH = necic acids

contain an 1,2-unsaturated necine base are converted by cytochromes P450 (CYPs) to pyrrolic metabolites which are genotoxic and, in mammals, also hepatotoxic (reviewed in Chen et al., 2010; Li et al., 2011; Macel, 2011). The 1,2-saturated PAs which are found, for example, in *Phalaenopsis* orchids would be considered non-toxic. However, specific accumulation of these PAs in the reproductive tissues of the orchids still suggests that they are part of the plant's chemical defense mechanism against herbivores despite their lack of toxicity (Mattocks and White, 1971; Frölich et al., 2006).

As is typical for specialized metabolites, PAs are synthesized in plants from precursor molecules obtained from primary metabolism. As could be shown by a series of tracer experiments PA biosynthesis largely depends on amino acid metabolites (Hartmann and Witte, 1995). The majority of studies have focused on necine base biosynthesis and have cumulatively demonstrated the polyamine homospermidine is produced which in turn is converted via an iminium ion intermediate into a necine base (see fig. 3 and Hartmann and Witte, 1995). Investigations into which enzymes are involved in the biosynthetic pathway have led to the identification of the homospermidine synthase (HSS), which catalyzes the first pathway-specific step in PA biosynthesis (Ober and Hartmann, 1999a). Research to identify the other enzymes involved in the biosynthetic pathway is in progress (see e.g. Sievert et al., 2015) and until today the HSS remains the only characterized enzyme of PA biosynthesis.

Homospermidine synthases

Since their discovery and being the only known protein component of PA biosynthesis, HSS (EC 2.5.1.45) have been under intense study. The enzyme catalyzes the NAD⁺-dependent transfer of an aminobutyl moiety from spermidine to putrescine thereby producing diaminopropane and the necine base precursor molecule homospermidine (fig. 3) (Ober and Hartmann, 1999a; Ober et al., 2003a).

One aspect of the HSS, which has been thoroughly investigated, is its evolutionary origin. With the identification of the first HSS, the striking similarities between this protein and the deoxyhypusine synthase (DHS) became apparent (Ober and Hartmann, 1999a). Besides occurring as homotetramers and some biochemical properties both enzymes also perform the same reaction with the only difference being that the DHS transfers the aminobutyl moiety from spermidine onto a specific, protein-bound lysine residue on the eukaryotic initiation factor 5A (eIF5A) instead of putrescine (Lee et al., 1999; Ober and

Hartmann, 1999a; Ober et al., 2003a). High sequence identity of the protein sequences (e.g. 79% between the *Senecio vernalis* HSS and DHS) confirmed that the enzymes are paralogous to each other meaning they share a common evolutionary origin (Ober and Hartmann, 1999a). Considering that DHS function is essential for the viability of all eukaryotes and that at least one DHS-coding gene can be found in every eukaryotic and archaea genome (Park, 2006) it is apparent that the HSS has evolved from a gene duplicate which originally encoded a DHS. More detailed analyses of the enzymatic activities of both enzymes have further shown that in the course of evolution the HSS has simply lost the ability to bind the protein substrate thereby retaining only the HSS function which is considered a “side activity” of the DHS (Ober et al., 2003a).

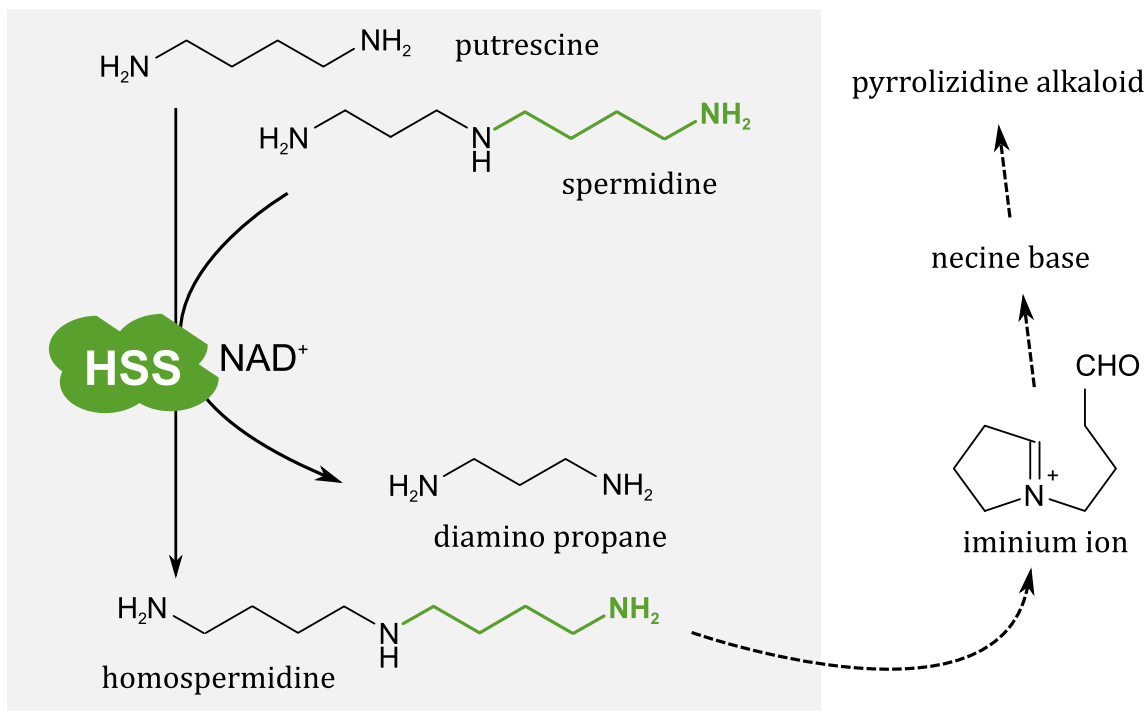


Fig. 3: First pathway-specific reaction of PA biosynthesis catalyzed by the HSS

The first sequence for a HSS was identified from *Senecio vernalis* (Asteraceae) and since then HSS-coding sequences have been identified from a variety of PA-producing species from diverse flowering plant lineages (Ober and Hartmann, 1999a; Reimann et al., 2004; Kaltenecker et al., 2013; Irmer et al., 2015). Where investigated, the HSS often displays tissue- and cell-specific expression, which is again a characteristic feature of plant specialized metabolism (Moll et al., 2002; Anke et al., 2004, 2008; Hartmann, 2007; Niemüller et al., 2012). The expression patterns vary between the various species, for

example, being restricted to leaf epidermis in *Heliotropium indicum* (Heliotropiaceae), to the endodermis cells in roots in *Senecio* species or being located in the root cortex in *Eupatorium cannabinum* (both Asteraceae), suggesting a polyphyletic origin of PA biosynthesis (Anke et al., 2004). Phylogenetic analyses of the collected DHS and HSS sequence data have revealed that the HSS has indeed evolved several times independently from different DHS gene duplication events in the different flowering plant lineages. At present, we know of six distinct *hss* gene lineages: two within the Asteraceae, one each in the Boraginales, Convolvulaceae and Fabaceae in addition to a duplication event which took place in the course of monocot evolution and has given rise to an HSS lineage specific to orchids (Orchidaceae) (Reimann et al., 2004; Kaltenecker et al., 2013; Irmer et al., 2015).

Most of our insight into biochemical characteristics of the HSS came from experiments conducted with *S. vernalis* HSS, the first PA-biosynthetic enzyme identified. Molecular evolution of HSS, however, has been studied in most detail in morning glories (Convolvulaceae, Kaltenecker et al., 2013). Using phylogenetic analyses based on codon substitution models it was determined how selection shaped HSS evolution after the gene duplication event. The study showed how, at first, the DHS gene copy that would later become the HSS-encoding gene experienced only purifying selection. This suggests that directly after the gene duplication event no functional divergence took place. Later on, the lineage underwent neutral evolution before finally experiencing positive selection. Progress of functional divergence suggested by the selection analyses could be confirmed by the variance in HSS activity observed for the various Convolvulaceae HSS. Altogether the study could show that HSS evolution, at least in this specific lineage, is so complex that it shows signatures consistent with more than one of the common gene duplication models (Kaltenecker et al., 2013).

While there are many studies which have dealt with PA-production in plants there are still many questions that remain open concerning the evolution of PA biosynthesis in general and HSS evolution specifically. For example, the molecular mechanism through which the DHS genes have been duplicated for any of the known HSS lineages is unknown mainly for lack of genomic sequence information; none of (known) PA-producing plant species have been included in any of the collaborative genomic sequencing efforts (Phytozome v11.0, Goodstein et al., 2012). Furthermore, the HSS (and DHS) genes have not been identified in many of the less prominent PA-producing plant

taxa like the grass family (Poaceae) and it is therefore possible that there are even more HSS lineages which might have evolved independently within the flowering plants.

Occurrence of pyrrolizidine alkaloids in grasses

As the grass family (Poaceae) also contains numerous crop plants the alkaloid content of members of this taxon is of special interest and has been under investigation for many decades (see e.g. Crow, 1962, Schardl et al., 2007, and references therein). Since then, several instances of PA occurrence have been described for various grass species.

The first PAs to have ever been isolated and identified from grasses belonged to the class of loline alkaloids (see Schardl et al., 2007 for a historical overview). However, these PAs are derived from 1-aminopyrrolizidine structures, which contain an ether bridge between the C2 and C7 (fig. 4). Comprehensive research has confirmed that these PAs are actually not products of plant specialized metabolism but are produced by fungal endophytes via a biosynthetic pathway that does not include homospermidine as an intermediate (Schardl et al., 2007). Recorded occurrences of these PAs are limited to members of the cool-season grasses subfamily (Pooideae, see Schardl et al., 2007; for details on grass systematics

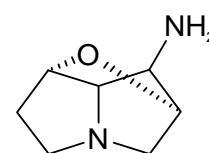
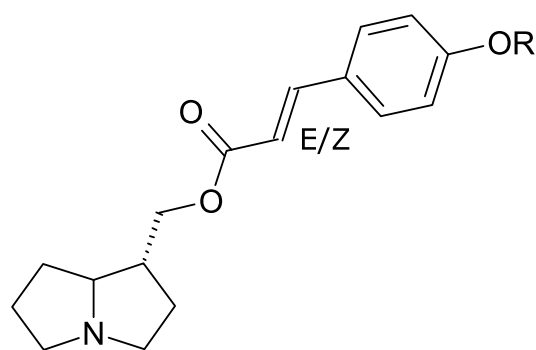


Fig. 4: Norloline



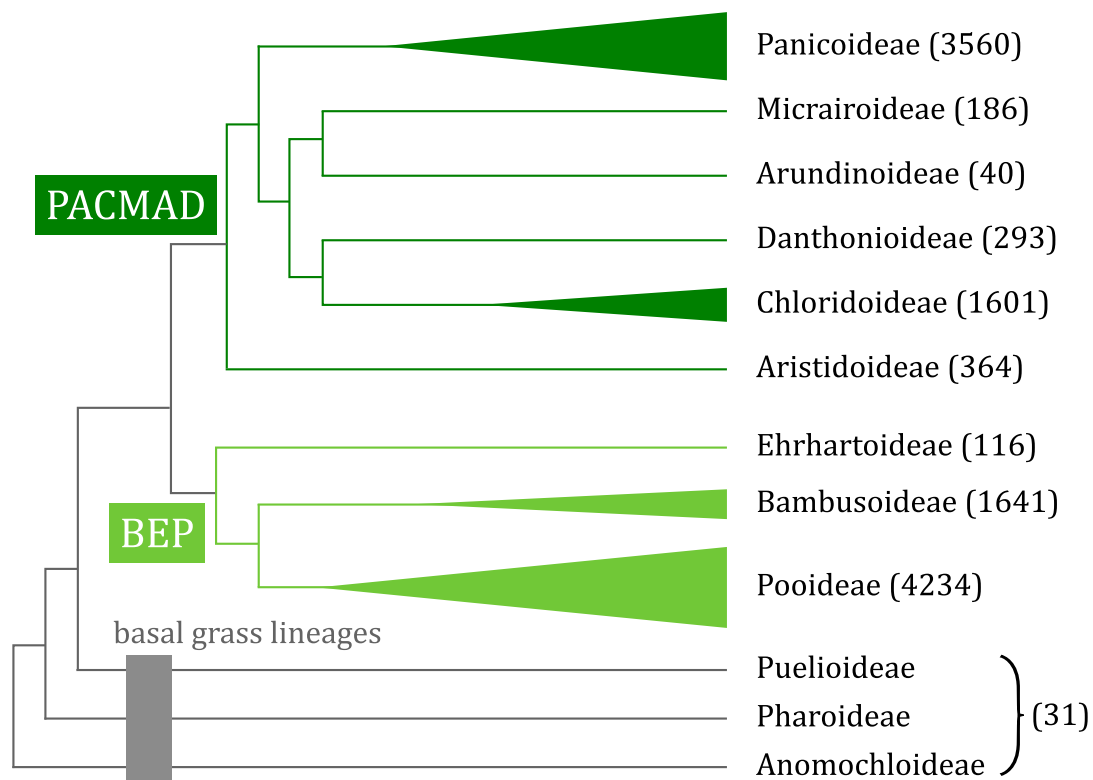
R = H thesinine
 R = Rha thesinine-rhamnoside
 R = Rha-Gly thesinine-rhamnoside-glycoside

Fig. 5: Thesinine-conjugates

refer to box 2). A single report from 2008 describes the presence of PAs that are not structurally related to lolines from perennial rye-grass and tall fescue (*Lolium perenne* and *Festuca arundinacea*, both belong to the Pooideae) independent of an infection by a fungal endophyte (Koulman et al., 2008). The structure of the PA was determined by NMR-spectroscopy to be *E*- and *Z*-stereoisomers of thesinine which was conjugated with mono- or disaccharides (fig. 5) (Koulman et al., 2008). The necine base of this PA is saturated and is therefore thought to be not toxic. Which functions these

Box 2: The grass family

Poaceae (also called “true grasses”) constitute a family of monocotyledonous plants within the angiosperms. The first grass plant appeared in the Late Cretaceous around 83 Ma ago, possibly even earlier (Janssen and Bremer, 2004; Prasad et al., 2011). Since then the Poaceae family has undergone considerable taxonomic radiation resulting in approximately 12000 extant species (Soreng et al., 2015). These grasses can be assigned to one of twelve subfamilies, which in turn either belong to the PACMAD or BEP subclades or the basal grass lineages (see below).



Phylogenetic relationship between Poaceae subfamilies

Topology and species numbers in brackets are based on Soreng et al., 2015

Exemplary grass species:

Panicoideae: maize (*Zea mays*), sugarcane (*Saccharum officinale*)

Chloridoideae: pampas grass (*Cortaderia*)

Pooideae: wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), ryegrasses (*Lolium*)

Ehrhartoideae: rice (*Oryza sativa*)

Bambusoideae: bamboos (various species)

glycosylated PAs might have in grasses is unknown and it was hypothesized that these PAs are not involved in herbivore defense as the two Poaceae species are widely used pasture grass, which are also fed to livestock (Koulman et al., 2008).

In 1962, Crow reported the isolation of the alkaloid thelepogine from *Thelepogon elegans* (Panicoideae) and determination of the crystal structure revealed that this C₂₀-molecule also contained a pyrrolizidine ring system (CD, fig. 6) (Fridrichsons and Mathieson, 1960). This PA is unusual in that it is not an ester alkaloid. Because of the presence of two other rings (AB, fig. 6) the structure is reminiscent of the diterpene

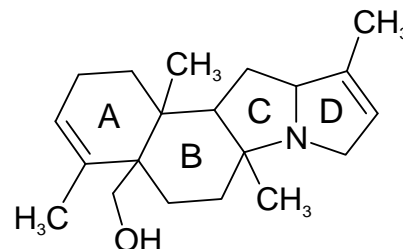


Fig. 6: Thelepogine

manool and a terpenoid precursor was suggested to be part of this PAs biosynthetic pathway (Fridrichsons and Mathieson, 1960).

One more study has indicated that PAs might be present in a grass species, this time from the Danthonioideae subfamily. In 1986, Wassel and colleagues extracted from *Schismus barbatus* an alkaloid which shared physical properties as well as characteristic mass spectrometry fragments with senecionine (Wassel et al., 1987). This specific PA is well known from various *Senecio* species and notorious for its toxicity (see e.g. Langel et al., 2011 or Xiong et al., 2011).

Excluding the extensive research that has been conducted on lolines none of the reports on PA occurrences were followed up by further studies so that we do not know anything about the biosynthesis, occurrence patterns or function of plant-produced PAs in this taxon.



Thesis objective and outline

PA biosynthesis is a fascinating trait exhibited by a variety of plant species belonging to different flowering plant taxa and, in many aspects, has been studied in great detail. However, as has been illustrated in this introduction, the production of PAs by grasses has not been investigated to any depth in regards to the general occurrence of PAs in this plant family, the spectrum of PA structures that are produced or the protein factors (and their evolution) involved in the biosynthetic pathway. The **objective** of my research efforts described in this doctoral thesis was therefore to rectify this oversight and specifically investigate PA biosynthesis in grasses.

The outcome of my research is presented in four chapters. Chapters 1 and 2 were written in the form of articles and are to be published in scientific journals. The study constituting **Chapter 1** investigates gene duplicate evolution by example of grass DHS and HSS molecular evolution. For that purpose, *hss* and *dhs* coding sequences were identified and their phylogenetic relationships determined. In addition to extensive phylogenetic analyses, *in vitro* experiments were performed first to determine specific enzyme activities and then to reconstruct the progression of functional divergence between the homologs. The investigations described in **Chapter 2** were performed in order to expand our basic knowledge on how widespread PA biosynthesis is within the grass family. Focusing on the Pooideae subfamily, a targeted metabolomics approach was used to screen grass extracts for the presence of any PA molecules. **Chapters 3 and 4** contain a miscellany of analyses which touch on various aspects of grass PA biosynthesis: On the one hand, the question of which amino acid residues in the protein sequences of grass HSS and DHS could potentially play an essential role in enzyme specificity is explored (Chapter 3). On the other hand, the results of an experiment are detailed which investigates whether the next reaction step in the PA biosynthetic pathway after the synthesis of homospermidine is catalyzed by a diamine oxidase (DAO) as it has been suggested for other PA producing plants (Chapter 4). Finally, the scientific part of the thesis closes with a conclusion meanwhile putting into perspective future research on HSS evolution and PA biosynthesis in grasses.

Chapter 1 and 2 contain co-authored material and the according author contributions are listed at the end of this thesis.



Chapter 1:

Complex molecular evolution of the paralogous *dhs* and *hss* genes in grasses (Poaceae)

Anne-Maria Wesseling, Elisabeth Kaltenecker and Dietrich Ober
Botanisches Institut, CAU Kiel, Am Botanischen Garten 1-9, 24118 Kiel, Germany

Author contributions are detailed at the end of this thesis.

Introduction

How new genes with novel functions emerge is one of the fundamental questions explored in the field of molecular evolution. Early on, gene duplication and subsequent functional divergence had been identified as a mechanism which majorly contributes to the evolution of new genes (see Taylor and Raes, 2004). As a result of extensive research and discussion several gene models have been developed dealing with the evolutionary fates of gene duplicates (reviewed e.g. in Conant and Wolfe, 2008; Innan and Kondrashov, 2010). Out of all models the most prominent are the neofunctionalization and subfunctionalization models which make differing predictions about the emergence and maintenance of new and ancestral functions in duplicated gene copies and about which kind of selection pressures drive functional divergence (Ohno, 1970; Force et al., 1999; Stoltzfus, 1999). There is an ongoing debate over whether a certain mode of gene duplication evolution is more prevalent than others (Hahn, 2009) and various genome-wide studies have been performed in an attempt to answer this question (e.g. Kondrashov et al., 2002; He and Zhang, 2005; Kondrashov and Kondrashov, 2006). Retaining a duplicate in the genome may also involve a combination of the suggested models and mechanisms and knowing the relative contribution of each is of great interest (Hahn, 2009; Panchy et al., 2016). Detailed studies of selected gene duplicates and the reconstruction of their individual history complement the genome-wide analyses in addressing this question.

Specialized (or secondary) metabolism is a characteristic feature of plants and it is thought that pathway evolution has greatly benefited from gene duplications. The term “specialized metabolite” describes a low molecular, organic compound that is specific to only few plant lineages. And while these metabolites are not part of the essential primary metabolic pathways shared by all plants, they still present important adaptations to the plant’s sessile lifestyle and respective ecological niche, for example, by protecting the plant from herbivory (Fraenkel, 1959; Pichersky et al., 2006). In consequence of the diversity in molecules, a diverse set of specialized enzymes are needed. There are numerous examples of gene duplication events having provided the genetic material for the evolution of new proteins for various specialized metabolic pathways (reviewed e.g. by Ober, 2005; Moghe and Last, 2015).

One example of such a gene duplicate having been recruited for a pathway in specialized metabolism is homospermidine synthase (HSS). This enzyme catalyzes the production of homospermidine which constitutes the first pathway-specific step in the biosynthesis of pyrrolizidine alkaloids (PA), a class of specialized metabolites found in a variety of flowering plant lineages and which is mostly known to function as chemical defense molecule against insect herbivores (Dam et al., 1995; Macel, 2011). The gene coding for the HSS has evolved from a gene duplicate originally coding for a deoxyhypusine synthase (DHS) (Ober and Hartmann, 1999a; Reimann et al., 2004). The DHS catalyzes the post-translational modification of the eukaryotic initiation factor 5A (eIF5A), a function that is essential in eukaryotes and consequently all eukaryotes carry at least one copy of the *dhs* gene (Park et al., 1997; Dever et al., 2014).

The paralogous relationship between the *hss* and *dhs* is apparent from their high sequence identity in addition to the fact that both enzymes share various biochemical characteristics. For example, both enzymes occur as homotetramers (Kang et al., 1995; Lee et al., 1999; Ober and Hartmann, 1999a). Furthermore, HSS and DHS display the same reaction mechanism insofar as they catalyze the transfer of an amino-butyl moiety from a spermidine molecule onto an acceptor molecule (Ober and Hartmann, 1999a). The reactions differ in the acceptor that is modified in the reaction. In case of the DHS a specific lysine residue of a protein (eIF5A) acts as the acceptor, while the HSS modifies a putrescine molecule. However, DHS displays some promiscuity in substrate selection as the enzyme is also able to perform the HSS reaction, which is thought to be the source of trace amounts of homospermidine found in some (non-PA-producing) plants (Ober and

Hartmann, 1999b; Ober et al., 2003b). Extensive phylogenetic analyses have shown HSS evolution to be a case of molecular convergence in that the *hss* gene has evolved at least six times from independent *dhs* gene duplications in different PA-producing plant lineages making it an excellent model to study and especially compare the evolutionary fates of gene duplicates (Reimann et al., 2004; Kaltenecker et al., 2013; Irmer et al., 2015).

PAs have been extracted from and studied in a variety of flowering plant species ranging from eudicots (e.g. Boraginaceae, Asteraceae) to monocots (Orchidaceae) (Hartmann and Witte, 1995; Stegelmeier et al., 1999; Ober and Kaltenecker, 2009). Most recently, the Pooideae (a subfamily of the grass family, Poaceae) was added to the list of taxa which contain PA-producing species (Koulman et al., 2008). Glycosylated PA molecules were isolated and identified from the grass species *Lolium perenne* and *Festuca arundinacea*. It is, however, not clear which function these specific compounds serve in grass plants (Koulman et al., 2008). In addition, the according biosynthetic pathway was completely unexplored in grasses.

We took this as an opportunity to investigate the evolution of the *dhs* and its paralogs in grasses in which context we also identified grass-specific *hss* coding sequences. For this we compiled *dhs* homolog coding sequences from across the grass family and performed phylogenetic analyses. These revealed *dhs* evolution in grasses to be very complex involving many gene duplications. Combining the phylogenetic results with functional characterization of selected DHS homologs we were able to identify two *hss* gene clusters originating from two independent *dhs* duplication events. We applied a series of codon substitution model-based selection analyses to identify when and which kind of selection accompanied the functional shift from a DHS- to an HSS-encoding gene. The present study not only takes fundamental steps towards the elucidation of PA biosynthesis in grasses but the detailed analyses also give us fascinating insights into the diversity of evolutionary mechanisms for gene duplicate retention and functional divergence.

Results

dhs homologs in the Poaceae

We identified partial and full-length coding sequences for *dhs* homologs from a variety of grass species by PCR amplification from cDNA. In addition, we assembled *dhs* coding sequences from transcriptome data available through the Sequence Read Archive (SRA, Leinonen et al., 2011). Finally, we searched online databases for suitable sequence accessions to further complement our dataset. In this way we obtained more than 80 distinct sequences from a total of 46 grass species from eight different grass subfamilies. Paralog numbers per grass species vary from one to nine (see supplementary table 1 for a complete list). However, those species with the highest numbers of *dhs* homologs (*Triticum aestivum* with nine genes and *Festuca arundinacea* with at least four genes) are hexaploid. Therefore, in the analyzed grasses the number of *dhs* homologs actually varies from one to three per genome and subgenome of the polyploid species, respectively.

We used the accumulated sequence information to reconstruct gene evolution using maximum likelihood (ML) and Bayesian estimation and using the *dhs* of *Joinvillea ascendens* (Joinvilleaceae, Poales) as an outgroup. Independent of the reconstruction method used, the topologies of the sequence trees are to a large extent consistent (see fig. 1 for the ML tree and supplementary fig. 1 for the Bayesian inference). In the following, we will therefore only consider the ML tree.

The topology of the tree largely agrees with that of the currently accepted grass species phylogeny in that sequences from species of the same grass subfamily mostly form monophyletic clusters (Soreng et al., 2015) the only exception being the sequences which belong to the Panicoideae subfamily. However, this will be further discussed in the next results section. Moreover, the phylogenetic relationships between these subfamily clusters match that of the grass phylogeny with one exception: the Pooideae do not cluster with the Bambusoideae and Ehrhartoideae to form the so-called BEP clade (Bambusoideae-Ehrhartoideae-Pooideae). However, placing these subfamilies within Poaceae phylogeny has proven to be notoriously difficult and in this respect our results concerning the BEP relationships are consistent with the literature (see e.g. Ma et al., 2012 and Wu and Ge, 2012).

Repeated *dhs* gene duplications in Poaceae

Reconstructing *dhs* homolog evolution has revealed that in the course of Poaceae evolution the *dhs* gene was repeatedly duplicated (fig. 1). While several gene duplications that we identified affect only one or few closely related species in our tree and have therefore probably been relatively recent events (e.g. the paralogs of *Oryza*, Ehrhartoideae or *Panicum virgatum*, Panicoideae) there are duplications which have split the tree into big and very distinct clusters of *dhs* paralogs. The AB duplication event occurred very early in Poaceae evolution, most likely before the divergence of the PACMAD subfamilies. This is the reason why the Panicoideae sequences do not form a single monophyletic cluster. Most PACMAD lineages have only retained one of the gene copies in their genomes (cluster A) resulting in two gene clusters of very unequal in size. It seems like gene duplicates of cluster B have only been retained in few Panicoideae species from the Andropogoneae tribe. The imbalance in number of genes between these two clusters is further increased by another duplication within cluster A which gave rise to clusters A1 and A2 (fig. 1). During evolution of the Pooideae subfamily another independent duplication event has occurred producing gene duplicate clusters exclusive to this subfamily (C and D, fig. 1). Similar to clusters A and B, the two Pooideae clusters are unequal in that only genes from cluster C were retained in all Pooideae species.

Gene duplicates hold the potential for functional divergence and, in the case of a *dhs* gene duplication, opens up the opportunity for HSS evolution as part of PA biosynthesis (Ober and Kaltenecker, 2009; Kaltenecker et al., 2013). However, there are no known sequence motifs by which DHS and HSS protein sequences can be universally distinguished and therefore enzyme's substrate preference needed to be determined experimentally. We selected a total of eight coding sequences to represent different sequence clusters (asterisks, fig. 1) which were heterologously expressed in *E. coli*. Purified proteins were then subjected to two different enzyme activity assays to determine DHS and HSS activity, respectively. The results confirm that clusters A and C both contain DHS-encoding sequences, with no detectable difference between the representatives of clusters A1 and A2 (supplementary table 2). The clusters A and C contain sequences of those grass species in which the *dhs* was shown to occur as a

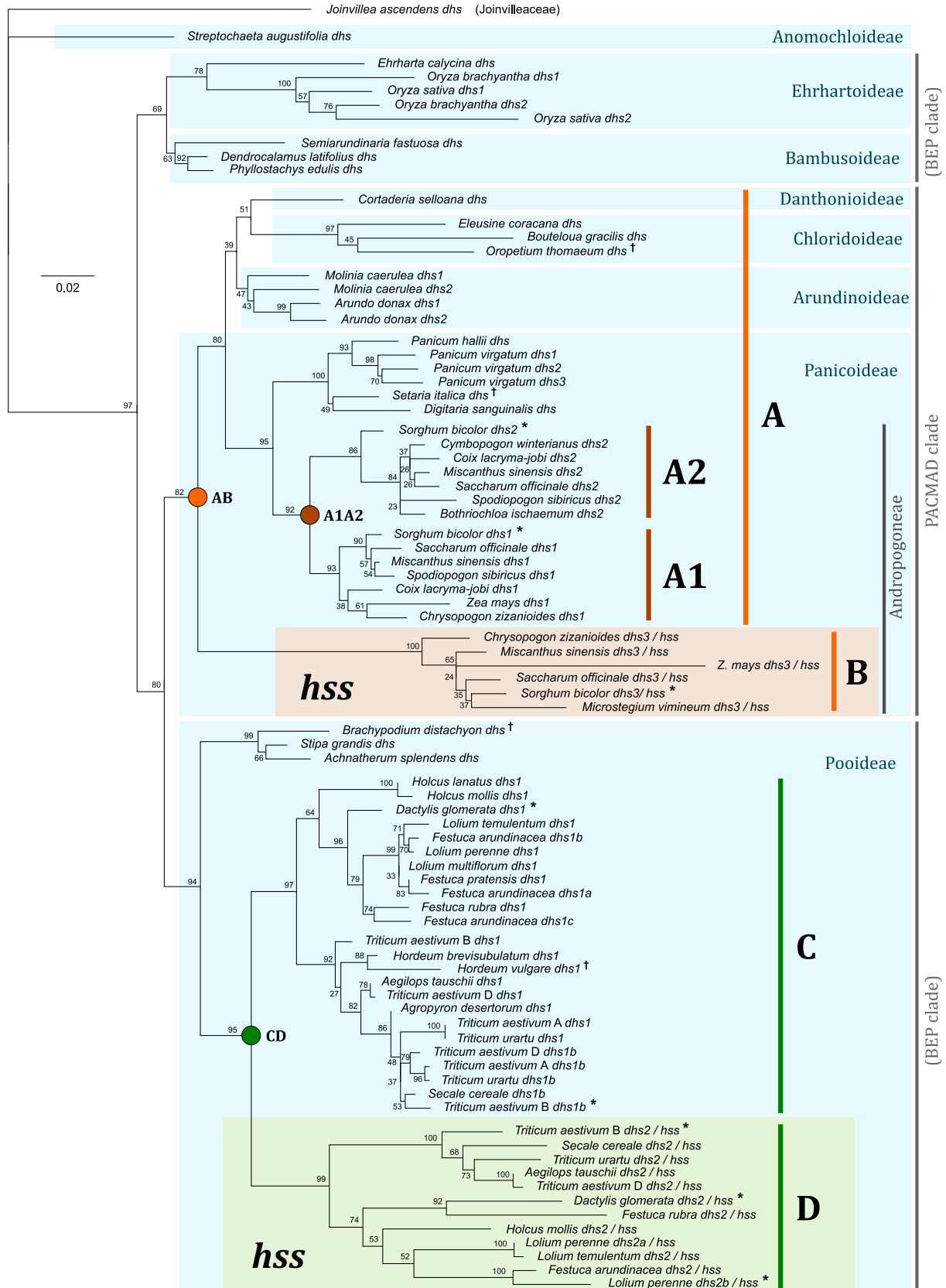


Fig. 1: Maximum likelihood gene tree of *dhs* homologs from various Poaceae species. Blue boxes indicate Poaceae subfamily affiliation. Node labels present bootstrap support values (500 replicates), asterisks indicate which coding sequences were tested in enzyme activity assays, obelisks indicate single-copy *dhs* genes, and colored circles mark selected gene duplication events. *Triticum aestivum* A, B and D indicate the three subgenomes of this species.

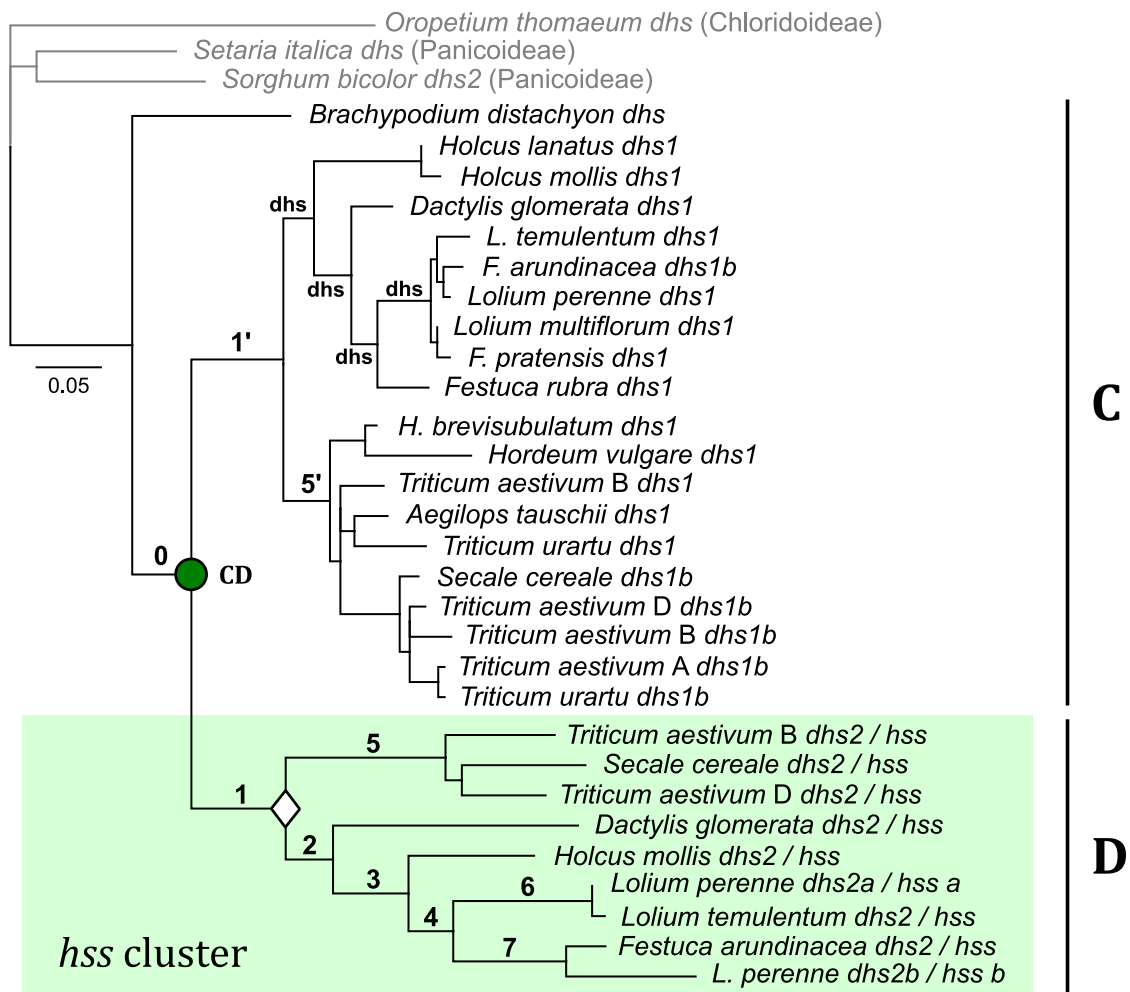
single-copy gene based on whole-genome sequence data (e.g. *Hordeum vulgare* and *Setaria italica*, see also obelisks in fig. 1 or supplementary table 1). Furthermore, the enzyme assays have revealed that enzymes encoded by sequences in clusters B and D have lost their eIF5A-modifying capacity while HSS activity was not affected (supplementary table 2). We therefore label these clusters “*hss* clusters” while distinguishing between Panicoideae *hss* and Pooideae *hss* (clusters B and D, fig. 1, respectively).

Selection during functional divergence in the Pooideae *hss* cluster

To further investigate the evolutionary trajectory which the Pooideae *hss* genes have followed after the gene duplication event (green circle, fig. 2), we conducted a series of sequence analyses aimed at detecting episodic changes in selection pressure. To this end, we employed a variety of selection analyses which estimate changes in synonymous and nonsynonymous substitution rates (d_S and d_N , respectively). The ratio of d_N to d_S , also known as ω , is used as an indicator of whether purifying ($\omega < 1$), neutral evolution ($\omega = 1$), or positive selection ($\omega > 1$) has acted on the coding sequence of a gene.

To estimate ω for internal branches of cluster D (fig. 2) we first applied a so-called 2-ratios-branch model, which partitions the branches in a sequence tree into foreground branches, which according to the model may be under positive selection, and background branches, on which positive selection is not allowed (Yang, 1998). For this and all following analyses described in this section, we used a dataset of select Pooideae *dhs* and *hss* coding sequences (fig. 2, for details see materials and methods section). Estimating ω for the internal branches suggested an increase in the nonsynonymous mutation rate only for the phase directly following the gene duplication event (branch 1, fig. 2). However, a likelihood ratio test (LRT) showed that the result was not statistically different from neutral evolution (i.e. 2-ratios model with $\omega = 1$).

To further explore this we used the same data to estimate ω values in the context of branch-site models, which take into account ω variation among sites and accordingly divide sites (codon triplets) into different categories instead of averaging ω over the whole sequence (Zhang et al., 2005). Analyses using branch-site models are considered to have more power to detect positive selection compared to the simpler branch model



	2-ratios-branch model			branch-site model		
	ω_f	ω_b	p (LRT)	ω	p (LRT1)	p (LRT2)
branch 1	∞	0.08	0.226	143	0.147	0.00015
branch 2	0.12	0.09	-	1	-	-
branch 3	0.12	0.09	-	1	-	-
branch 4	0.13	0.09	-	1	-	-
branch 5	0.05	0.09	-	1	-	-
branch 6	0.05	0.09	-	∞	0.057	0.016
branch 7	0.40	0.09	-	1.3	0.891	0.210

	aBSREL		branches	RELAX	
	ω	p		K	p
branch 1	∞	0.087	1 vs 0	0	1.8×10^{-5}
branch 2	0.127	1	1' vs 0	0.24	0.034
branch 3	0.077	1	1 vs 1'	0	7.6×10^{-3}
branch 4	0.150	1	2,3,4 vs dhs	0.95	0.864
branch 5	0.047	1	5 vs 5'	1.07	0.852
branch 6	0.41	1	6 vs 4	2.01	0.168
branch 7	0.089	0.762	7 vs 4	2.83	0.225
	16.4		6 vs 7	2.98	0.030

Fig. 2: caption on next page

Fig. 2 (p. 24): Overview of selection analyses performed on Pooideae *hss* lineage. Branch lengths present the number of nucleotide substitutions per codon (1-branch model). ω_f – ω of foreground branch, ω_b – ω of background branches. LRT – likelihood ratio test based on 2-ratios-branch models, LRT1 – branch-site test for positive selection, LRT2 – branch-site test for both positive selection and relaxed functional constraints. For adaptive branch-site REL (aBSREL) tests the p values were corrected for multiple testing. Green circle and white diamond mark nodes which were chosen for ancestral sequence reconstruction.

based tests. It is more likely that only few sites experience positive selection, while the majority of sites on a coding sequence remain under purifying or relaxed selection pressure (Zhang et al., 2005). Results of the branch-site calculations for branch 1 again indicate that a portion of sites have evolved under the influence of positive selection (fig. 2). Testing the branch-site model estimation results against a null hypothesis in another LRT (here called LRT1; Zhang et al., 2005) did not confirm that the branch-site model with $\omega > 1$ fit significantly better than the null hypothesis, $\omega = 1$, for any of the branches (fig. 2). A second test (here LRT2; Yang et al., 2005; Zhang et al., 2005) allows testing for episodic positive selection as well as relaxed constraints on a lineage of interest. According to this test, the results for branch 1 are highly significant, indicating that the duplicated gene likely experienced at least relaxed functional constraints if not positive selection (fig. 2).

In addition to the branch and branch-site model calculations we performed the adaptive branch-site random effects model likelihood test (aBSREL) to find signatures of episodic positive selection (Smith et al., 2015). The test estimates the appropriate number of site categories and the according ω for each branch individually and thereby avoids loss of power through violation of model assumptions (Pond et al., 2011; Smith et al., 2015). Applying this test confirmed the results of the branch and branch-site model by estimating increased ω values for some branches including branch 1, however, statistical support was not significant (fig. 2).

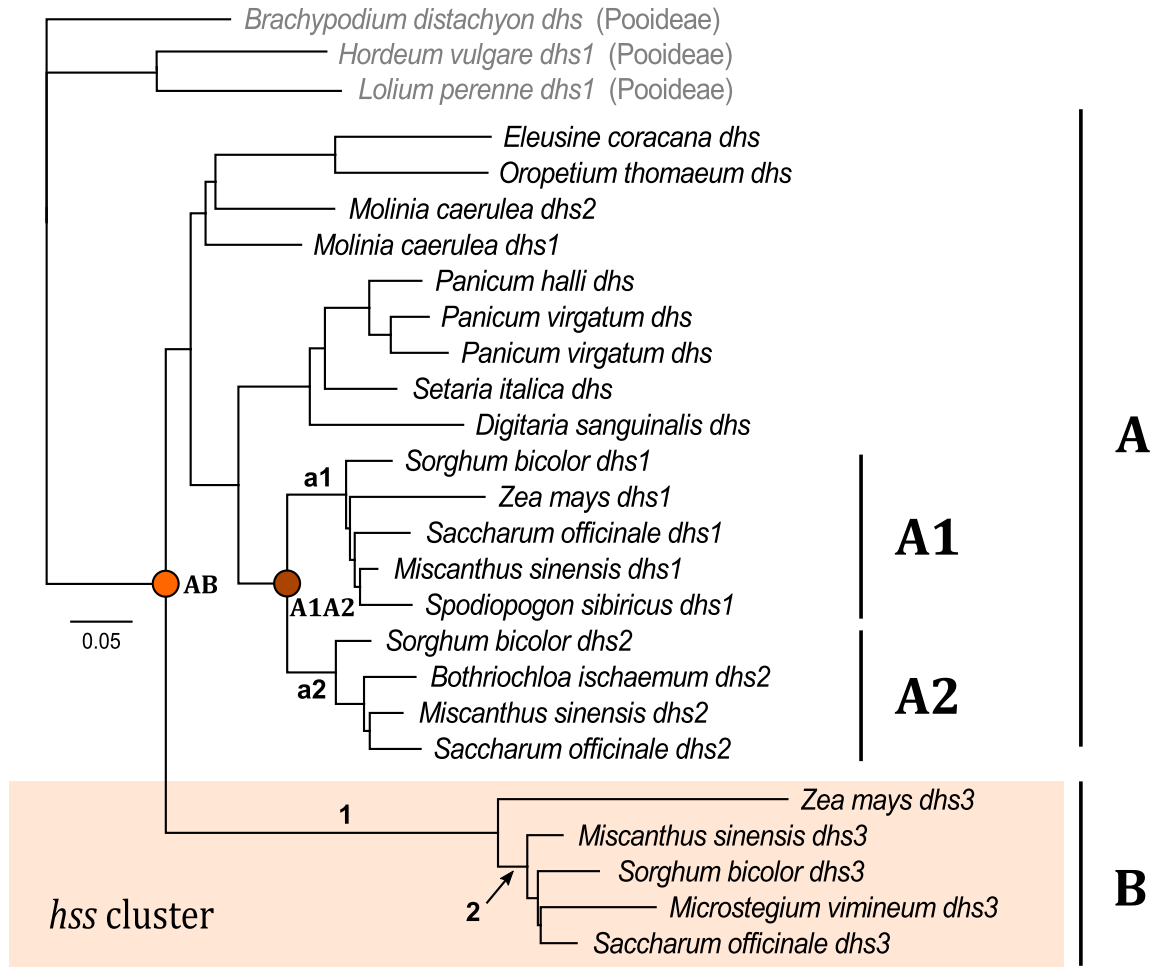
Next, we checked for relaxation of selective constraints using RELAX, a test that was specifically designed to distinguish between the relaxation of selection pressure and its intensification, which includes both positive and purifying selection (Wertheim et al., 2015). By implementing another variation of branch-site models, RELAX calculates a so-called relaxation coefficient K which relates the parameters for the site categories

between two sets of branches (Wertheim et al., 2015). That way, $K < 1$ indicates selection relaxation and conversely $K > 1$ signifies that on the tested lineage an intensification of selection pressure has taken place. Testing branch 1 against the branch directly preceding the CD duplication event (branch 0, fig. 2) we found significant relaxation of selection ($K = 0$). Testing branch 1' of the *dhs* cluster against branch 0 also indicated relaxation of selection, however, less pronounced ($K = 0.24$), and comparing branch 1 of the *hss* cluster against the parallel evolving branch 1' showed that selection was significantly more relaxed on the *hss* lineage. In contrast to this, testing the succeeding set of *hss* branches against their *dhs* counterparts (2,3,4 vs *dhs*, 5 vs 5', fig. 2) found neither strong intensification nor relaxation of selection pressure intensity. The K values (circa 1) suggest that selection pressure was about the same.

To evaluate the significance of the temporal variations in selection intensity it is necessary to establish if and how enzyme function changed during that particular phase of evolution. We therefore reconstructed the ancestral DHS enzyme that was present at the time of the gene duplication event as well as the enzyme duplicate which had evolved during an initial phase of changed selection pressure (fig. 2, green circle and white diamond, respectively). The two reconstructed protein sequences differ at 13 sites spread along the entire length of the proteins (supplementary fig. 2). The ancestral proteins were heterologously expressed, purified, and subjected to HSS and DHS activity assays. The results show that the mutations, which the ancestral *dhs* gene had accumulated in the early stages of its evolution, cumulatively caused a reduction in specific DHS activity (supplementary table 3). On the other hand, specific HSS activity of the enzyme increased during this phase of HSS evolution (supplementary table 3).

Selection after gene duplications during Panicoideae evolution

We also examined *dhs* paralog lineages which originated in the PACMAD clade in more detail again focusing on the molecular evolution of the gene using a variety of selection analyses (cf. previous results section). We tested the parallel evolving branches resulting from the A1A2 duplication event (branches a1 and a2, fig. 3). In addition, the internal branches of the Panicoideae *hss* cluster were examined (branches 1 and 2, fig. 3). Using 2-ratios-branch models we estimated ω values of the respective branches and found ω



	2-ratios-branch model				branch-site model		
	ω_f	ω_b	p (LRT)	p (LRTb)	ω	p (LRT1)	p (LRT2)
branch 1	0.22	0.10	-	0.002	2.0	0.468	0.003
branch 2	0.29	0.11	-	0.315	1	-	-
branch a1	0.10	0.11	-	0.943	1	-	-
branch a2	0.00	0.11	-	0.021	1	-	-

Fig. 3: Overview of selection analyses performed on Panicoideae *hss* lineage. Branch lengths present the number of nucleotide substitutions per codon (1-branch model). ω_f – ω of foreground branch, ω_b – ω of background branches. LRT – likelihood ratio test based on 2-ratios-branch models, LRT1 – branch-site test for positive selection, LRT2 – branch-site test for both positive selection and relaxed functional constraints.

values that were smaller than 1 but which were on the other hand elevated in comparison to the background ω , in the case of branches 1 and 2 of the *hss* cluster branches (fig. 3). Since none of the ω values were larger than 1 the LRT1 was not performed. Instead, the branch model results were compared in a LRT against branch model results where the foreground ω was set at 0.1 which represents the calculated background ω (LRTb, fig. 3). This verified our suspicion that the increase in ω for branch 1 in comparison to the background is indeed significant. In contrast to this, the ω values and LRTb results for branches a1 and a2 suggest that both of the lineages after the A1A2 gene duplication event experienced purifying selection similar to or stronger than that of the average background selection pressure of this sequence tree (fig. 3). Branch-site model estimations only resulted in $\omega > 1$ for sites on branch 1 of the *hss* cluster. Likelihood ratio tests comparing the results of the branch-site calculations against those of the different null hypotheses (LRT1 and LRT2, fig. 3) show that either relaxed or positive selection has acted on branch 1 ($p \ll 0.05$).

Discussion

We have identified *dhs* homolog sequences from various grasses and used these for phylogenetic analyses. The *dhs* gene in this plant family has experienced a complex evolutionary history. Based on their *in vitro* biochemical activity we were able to identify two *hss* clusters which have evolved from two independent gene duplication events specific to two different grass subfamilies, the Pooideae and Panicoideae (fig. 1). As the Pooideae cluster contains the species *Festuca arundinacea* and *Lolium perenne* that have been described to produce PAs the involvement of these *hss* in PA biosynthesis is most likely. Including the two distinct grass *hss* lineages, a total of eight independent *dhs* gene duplications were confirmed in flowering plant *hss* evolution, with the two distinct *hss* clusters in the Poaceae being the second example of this gene having evolved twice within one plant family (described only once before for the two PA-producing lineages within the Asteraceae, see Anke et al., 2004; Reimann et al., 2004).

The results of the various tests and experiments that were conducted in this study allow for various interpretations on the exact nature of selection acting on the various *hss* lineages. Within the Pooideae *hss* cluster, for the phase directly following the gene duplication event, we found greatly elevated ω values and also significant relaxation of selection pressure (fig. 2). The nonsynonymous substitutions that have occurred under the changed selection pressure have caused the early functional divergence of the gene, which we could experimentally verify by showing an incipient loss of DHS activity and increase in HSS activity in the ancestral enzymes (fig. 2 and supplementary table 3). It is possible that both, positive selection and neutral evolution, have acted during this early phase, may it be consecutively or on different sites simultaneously. Following this phase of relaxed selection pressure, purifying selection in the *hss* cluster reaches levels similar to that of the respective branches in the *dhs* cluster. The fact that purifying selection acted again on both lineages could indicate that both, the *dhs* and *hss* genes, served a biological function which might have initiated the conservation of these genes in the grass ancestors.

Concurrent with *hss* evolution in the Pooideae, we also found either positive or relaxed selection following the clade-specific gene duplication event resulting in the Panicoideae *hss* lineage (fig. 3). Analyzing sequence evolution of Panicoideae *hss* we are confronted with the problem that, due to extensive gene loss (indicated by whole-genome sequence

data) branch 1 of this cluster is comparatively long (dS value of 0.253 in comparison to 0.077 for branch 1 in the Pooideae *hss* cluster). During this time span represented by this branch (more than 17 MA, see Estep et al., 2014, Cotton et al., 2015) the gene has most likely undergone various stages of gene duplicate evolution and experienced different kinds of selection. While it is limited what we can conclude about the evolutionary history of the Panicoideae *hss* based on selection analysis results this second *hss* cluster should prove to be insightful when *hss* sequence evolution is further investigated as it allows us to directly compare functional and sequence evolution through functional analyses. Two independent *hss* lineages stem from closely related taxa, with similar genetic backgrounds and where epistatic effects might be comparable [cf. Chapter 3].

The Panicoideae *dhs* have provided us with another interesting aspect of *dhs* evolution. DHS-encoding genes have been found to be single-copy genes in most eukaryotes or, where several paralogs have been identified, only one was found to function as a DHS (Álvarez et al., 2008; Chawla et al., 2010; Reimann et al., 2004; Sasaki et al., 1996; Wang et al., 2003). However, two of the *dhs* paralogs identified from various Panicoideae species have experienced only purifying selection after the gene duplication and have been found to code for a functional DHS in enzyme activity assays (clusters A1 and A2, fig. 1 and fig. 3). The two functional *dhs* copies therefore present an exception to the “*dhs* is single-copy gene” rule. It is noteworthy that the *S. bicolor* genome in addition to the two *dhs* copies also contains two gene copies coding for the DHS protein substrate eIF5A (86.3% protein sequence identity, supplementary fig. 3). The question arises whether these *dhs* gene copies were able to avoid being purged from the genome through subfunctionalization of DHS function. Enzyme activity assays have given no indications that both enzymes greatly diverge in their biochemical function. Expressional divergence could be a mode of subfunctionalization which would not be detectable by the analyses at hand. This would be consistent with purifying selection on coding sequences which was detected in this specific case of *dhs* evolution.

Poaceae *hss* evolution in context with gene duplication models

How does the Pooideae *hss* evolution fit with the established models of gene duplication evolution? Looking at protein function in terms of enzyme specificity *dhs* paralog evolution could be considered a mixture of sub- and neofunctionalization. On the one

hand, both functions are already present in the ancestral *dhs* gene (deduced from *dhs* activities of various plants including grasses; Ober and Hartmann, 1999b). On the other hand, consistent with neofunctionalization which predicts that one of the gene duplicates will retain the ancestral gene function while the other acquires a novel function, DHS activity is conserved in the extant *dhs* genes after duplication, and in the second gene copy HSS activity with simultaneous loss of the DHS activity presents a new function. However, based on the biological function, which is the decisive factor when it comes to natural selection, *dhs/hss* evolution in the Pooideae is most likely a case of neofunctionalization. The model, as it was first proposed by Ohno (Ohno, 1970), postulates after duplication the complete redundancy of the second gene copy, which is therefore free from any functional constraints and accumulates substitutions under no selective influences (i.e. neutral evolution). In most cases this would lead to the non-functionalization of the gene. However, in a few cases this could result in the acquisition of a new beneficial function. The results of the selection analyses best match the neofunctionalization theory. There are indications that (at least for a while) the gene evolved neutrally (RELAX results, fig. 2). The elimination of the DHS function through inhibition of the protein-protein interaction between the enzyme and its protein substrate can probably be achieved by just a few of many possible substitutions which at the same time do not inhibit the HSS “side activity” and therefore could have easily occurred during neutral evolution. Additionally, few adaptive substitutions could have occurred also under positive selection.

Another hypothesis on how genes behave after duplications has been formulated in the paralog interference theory, which specifically discusses how obligate homooligomeric proteins evolve after their duplication (Kaltenegger and Ober, 2015). It was predicted that both proteins will interact with each other as they are identical in the time directly following the duplication event. Therefore any diverging mutation in one of the copies will ultimately affect the functionality of either gene. This is also the case for the DHS, especially as the enzyme’s active site is located in the protein-protein interface of two subunits (Liao et al., 1998). Both genes will therefore experience purifying selection to avoid a decrease in fitness until paralog interference can be escaped for example through expressional separation of the duplicates (Kaltenegger and Ober, 2015). The evolution of the *hss* gene within the Convolvulaceae was interpreted to be an example for such an evolutionary scenario (Kaltenegger et al., 2013; Kaltenegger and Ober,

2015). The Convolvulaceae-specific *dhs* gene duplication is followed by a period of purifying selection before the gene duplicate of the *hss* cluster experience relaxed selection and finally positive selection (Kaltenegger et al., 2013). The *hss* evolution in the Pooideae, however, seems to have taken a different route although restrictions affecting Convolvulaceae *dhs* and *hss* evolution should theoretically also apply to the Pooideae genes. Our analyses could not produce any evidence that a phase of purifying selection directly succeeded the Pooideae-specific gene duplication event. After further inspection of the according genomic sequence scaffolds we found that two of the *dhs* paralogs from *T. aestivum*, *dhs1* and *dhs2/hss*, are located on the same chromosome in all *T. aestivum* subgenomes as well as in some close relatives (supplementary fig. 4 and supplementary table 1). The close proximity of the gene locations on the chromosomes suggest that these copies might have been created through a tandem duplication process. One possible explanation is that through this small-scale duplication event genomic regions responsible for transcription regulation were eliminated from one of the gene copies and that this way expressional separation occurred simultaneously with the gene duplication event. Therefore the gene copies might have skipped the predicted initial phase of purifying selection and instead directly continued with the processes leading to neofunctionalization.

Pooideae *hss* evolution in context with PA biosynthesis in grasses

As illustrated above, we have identified a gene lineage of an enzyme involved in PA-biosynthesis in *L. perenne* and *F. arundinacea*. PA-production by a grass plant was so far described only for few closely related Pooideae species: *Lolium* species that are outbreeding are represented in our sequence dataset by the *Lolium perenne* and *L. multiflorum* genes, and *F. arundinacea* (Cheng et al., 2015; Koulman et al., 2008; Wesseling et al., unpublished results [Chapter 2]). However, we found that *hss* genes are wide-spread among Pooideae species in general. Several species outside of the *Lolium-Festuca* complex and in which no PAs were detected, e.g. *T. aestivum* and *Dactylis glomerata* (Wesseling et al., unpublished results [Chapter 2]), have been shown to contain (and transcribe) biochemically functional *hss*. The analyses presented here demonstrate that the shift in enzyme specificity from DHS to HSS was initiated early in Pooideae evolution before *T. aestivum* and other species branched off from the lineage leading to the PA-containing *Lolium-Festuca* species complex.

There are several possible scenarios which would be consistent with the observed enzyme activities and selection pressure changes (fig. 1 and 2, supplementary table 2 and 3). On the one hand, the ability of grasses to produce PAs might have appeared at the same time as the HSS-specific activity of the enzyme and later on some Pooideae lineages experienced secondary loss of this ability. In this case, *hss* genes have been retained in genomes without having a function in PA biosynthesis. On the other hand, it is possible that the neutrally evolved, “pre-adapted” *hss* gene was recruited only later during *Lolium-Festuca* species radiation for the task of PA-production. A third possibility combines both scenarios. First, the gene duplicate was recruited for a specific biological function (unknown but not excluding PA-biosynthesis) and then re-recruited for PA-production in the various *Lolium* and *Festuca* species. Our results cannot exclude the possibility that some branches in the *hss* cluster (e.g. *L. perenne* paralogs *hss a* and *hss b*) have gone through a second round of selection pressure changes (fig. 2) however clearing this up will require a much closer look at *hss* evolution in this specific taxon. Future analyses would have to include a higher phylogenetic resolution of *hss* evolution and should result in deeper insights on protein evolution and specifically *hss* evolution in direct correlation with the evolution grass PA biosynthesis.

Materials and Methods

Identifying and cloning of coding sequences of *dhs* homologs

Sequence data which provided the basis for all subsequent analyses were collected from miscellaneous sources: on the one hand, we used a variety of PCR techniques to amplify putative *dhs* and *hss* coding sequences from selected grass species and on the other hand we mined various online sequence data bases for relevant sequence information.

For PCR amplification total RNA was isolated from plant samples using the RNeasy Plant Mini Kit (Qiagen) and reverse transcribed to cDNA using the RevertAid Premium Reverse Transcriptase (ThermoFisher Scientific) following the manufacturer's instructions. Most coding sequences were initially amplified in 3'RACE-PCRs (rapid amplification of cDNA ends - polymerase chain reactions) following the protocol of Scotto-Lavino et al. (2007). For this we used degenerate oligonucleotides designed to match conserved regions of Poaceae and Poales *dhs* sequences as forward primers (P01 - P04) and Q_T, Q_o, and Q_i as reverse primers (sequences of all primers are listed in supplementary table 4). Full-length sequence information of select *dhs* homologs was obtained through 5'RACE-PCRs. For this, a polyC tail using a terminal deoxynucleotidyl transferase to cDNA samples which had been reverse transcribed using a gene-specific reverse primer (P05). PCRs were then performed with AAP or AUAP as forward primers and a variety of reverse primers (P06, P07). All RACE-PCRs were performed using GoTaq DNA Polymerase (Promega). Where appropriate (e.g. multiple PCR bands or PCR smear) nested PCRs were performed. Full-length coding sequences were obtained by amplification of cDNA (reverse transcribed with the Q_T-Primer) using primers P08-P53 and a high-fidelity proofreading polymerase (Phusion High-Fidelity Polymerase, ThermoFisher Scientific).

PCR products were purified (NucleoSpin Gel and PCR Clean-up, Macherey-Nagel) and cloned into the pGEM-T-easy vector (Promega) or alternatively into pET23a- or pET22d-based expression vectors (Novagen) which extended the coding sequence by a C-terminal polyhistidine-tag. Either way, constructs were propagated in *E. coli*-Top10 cells and were subsequently prepared for Sanger sequencing (sequencing service of MWG Eurofins Genomics) or at Institute of Clinical Molecular Biology (IKMB) in Kiel.

Sequencing results were annotated and further processed using the Geneious software package (Biomatters).

To complement the sequence data acquired in our lab we searched through various online databases for coding sequences of grass *dhs* homologs using the according BLAST web tools: the GenBank nucleotide collection (nr/nt), GenBank expressed sequence tags division (est), EnsemblGenomes (Kersey et al., 2016), and Phytozome (Goodstein et al., 2012).

To further increase taxonomic sampling we searched the Sequence Read Archive (SRA, (Leinonen et al., 2011)) for appropriate transcriptome data (RNA-seq) of various grasses. Raw data of Illumina runs were downloaded and the reads trimmed (Trimmomatic 0.36, Bolger et al., 2014). We used different tools available from the blast+ package (version 2.2.25, Camacho et al., 2009) to blast the processed reads against a small database of grass *dhs* and *hss* sequences and extracted those reads which produced a significant hit. The specific tools used were blastn (E-value 10) or blastx (E-value 1e-2) as well as fastadb and fastacmd. Extracted reads were then assembled using the Geneious Assembler (Biomatters) using the highest recommended sensitivity while not merging variants with a coverage of over 1. Only unambiguous consensus sequences of contigs covering the 3' end of the coding sequence for a minimum length of 600 bp were added to the dataset.

Supplementary table 1 lists all sequences used in this study with details on how and/or from where the sequences were obtained.

Protein purification and enzyme activity assays

Full-length DHS homologs were overexpressed in *E. coli* BL21(DE3) following the manufacturer's instructions (Novagen). Harvested cells were resuspended in lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol, pH 8.0) and then sonicated. Overexpressed proteins were purified from the lysate by Ni-NTA gravity-flow chromatography (Qiagen). After binding the Ni-NTA resins were washed with a high-salt buffer (50 mM imidazole, 1.5 M NaCl) and the protein was eluted using a high-imidazole buffer (250 mM imidazole, 300 mM NaCl). Purified protein was concentrated (Amicon Ultra-4 Centrifugal Filter Units, Millipore) and rebuffed to an assay buffer (100 mM Glycin-NaOH, 0.1 mM EDTA, 1 mM DTT, pH 9.0). Protein

concentration was calculated from the sample's absorption at 280 nm and their specific extinction coefficient ($M^{-1} cm^{-1}$) predicted by the ProtParam web tool on ExPASy (Gasteiger et al., 2005).

To check enzyme activity we conducted two assays ("DHS" and "HSS") as described previously (Ober and Hartmann, 1999b). In short, purified proteins (3 - 15 μ g) were incubated with radioactively labeled substrates and the amount of radioactive product was quantified at different time points of the reaction (1 - 32 min). The DHS assay uses 82 μ g of heterologously expressed and purified *Z. mays* eIF5A as aminobutyl acceptor (for details on expression and purification refer to Nurhayati et al., 2009).

Sequence tree reconstruction for Poaceae *dhs* homolog evolution

Nucleotide sequences were aligned using the MAFFT algorithm (algorithm = Auto, scoring matrix 200PAM/k=2, gap open penalty = 1.53, offset value = 0.123) (Kato and Standley, 2013). The resulting alignment was trimmed at both ends to match the length of the shortest sequence of the dataset (which was the *J. ascendens dhs* partial coding sequence) resulting in a final, ungapped length of 565 bp. Maximum likelihood (ML) gene trees were inferred employing the GTR+G substitution model with 500 bootstraps by PhyML in Geneious (Guindon and Gascuel, 2003). For Bayesian tree inference we used MrBayes in Geneious using the default settings and the GTR+G substitution model (Huelsenbeck and Ronquist, 2001).

Selection analyses on Pooideae *hss* and Panicoideae *hss*

We compiled selected sequences into two datasets: the first dataset used to analyze the *hss* cluster D contains all available full-length Pooideae *dhs* and *hss* sequences in addition to select full-length *dhs* sequences from other Poaceae subfamilies as outgroup (see fig. 2), while the second dataset to analyze *hss* cluster B includes all full-length PACMAD *dhs* and *hss* sequences and few Pooideae sequences for tree rooting (see fig. 3).

To eliminate the possibility that erroneously aligned non-homologous sites distort the outcome of the selection analyses (Fletcher and Yang, 2010), we used the GUIDANCE2 web tool (Sela et al., 2015) to align coding sequences using the MAFFT algorithm (sequence type = nucleotide, max-iterate = 2, bootstraps = 100). All columns with a confidence score below 0.990 were removed. The alignments were then manually

checked and modified to correct for reading frame shifts. We also screened sequences for recombination using the GARD tool from the web-based datamonkey suite (Delport et al., 2010; Pond et al., 2006). If a predicted recombination breakpoint was confirmed by significant topological incongruence (KH test) the alignment was shortened to only contain non-recombined DNA regions. See supplementary material for GARD output of each dataset and resulting alignments. Furthermore, a premature stop codon in the open reading frame of *T. aestivum dhs2* (chromosome 5D) was replaced by a gap. Based on these alignments ML sequence trees (PhyML, GTR+G) were calculated of which the topologies were used for selection analyses.

A series of branch model and branch-site model calculations implemented in the codeml program of the PAML4.7 package using custom PERL scripts (Yang, 1998, 2007; Zhang et al., 2005). First, a basic 1-ratio model (NSsites = 0, model = 0) was used to estimate branch lengths which were then used as starting values in all subsequent codeml estimations. Branches of the tree were then divided into background branches and a foreground branch which was to be tested for changes in selection pressure. Branch models (NSsites = 2, model = 0) tested in a likelihood ratio test against a null model where ω was fixed at one (LRT, d.f. = 1) (Yang, 1998). Branch-site models (NSsites = 2, model = 2) were calculated using different starting values for ω (1.3, 2, 10 and 100) and the result with the best likelihood was used to test against two different null models in two likelihood ratio tests: a test for positive selection (NSsites = 2, model = 2, fix_omega = 1, omega = 1; LRT1, d.f. = 1) and a test for both positive selection and relaxed constraints (model = 0, NSsites = 1; LRT2, d.f. = 2) (Zhang et al., 2005).

For additional analyses of positive selection we made use of the adaptive branch-site random effects model (aBSREL) implemented in HyPhy 2.2.4 using default settings (Pond et al., 2005; Smith et al., 2015). To specifically test for relaxation of selection pressures instead of positive selection we ran the RELAX test in HyPhy (analysis type "all", version 1.0, Wertheim et al., 2015).

Ancestral sequence reconstruction

Ancestral sequences for nodes of interested were extracted from the codeml output (1-ratio-branch model, see above). For these analyses the full-length ancestral regions not covered by the codeml output were filled manually using a parsimonious approach.

Gaps and amino acids were placed so that the number of character-state changes was minimized. The finished predicted ancestral protein sequence was then codon-optimized and synthesized (GenScript). The open reading frame was subcloned into a vector with a pET23b-backbone. Expression, purification and enzyme assays were performed as described above except that the lysis buffer contained additionally 0.8 mM NAD⁺.

Acknowledgments

We are grateful to Brigitte Schemmerling for her excellent technical assistance and the staff of the Botanical Garden Kiel for help in growing many of the grasses used in this study. We also thank Lars H. Kruse for his helpful hints on processing transcriptome data. Many thanks are due to Jessica Garzke, Julia Jensen-Kroll, Niklas Lauterbach, and Ana Lokmer for assisting in sequence identification and cloning. We thank the Institute of Clinical Molecular Biology in Kiel for providing Sanger sequencing. AMW was supported through a stipend of the International Max Planck Research School (IMPRS) for Evolutionary Biology.

Supplementary material

Supplementary table 1: Overview of all *dhs* homolog sequences used in this study, sorted by taxa

Species	Gene	Accession	Cultivar/Source	Further information*
Joinvilleaceae				
<i>Joinvillea ascendens</i> subsp. <i>glabra</i>	<i>dhs</i>	this study	D. Lorence 9066 (NTBG)	P02 & Q _i (3'RACE)
Poaceae				
Anomocloideae				
<i>Streptochaeta augustifolia</i>	<i>dhs</i>	this study	SRA (SRR3233339)	assembled from RNA-seq data
Arundinoideae				
<i>Arundo donax</i>	<i>dhs1</i>	this study	KIEL	P03 & Q _i (3'RACE)
<i>Arundo donax</i>	<i>dhs2</i>	this study	KIEL	P03 & Q _i (3'RACE)
<i>Molinia caerulea</i>	<i>dhs1</i>	this study	KIEL	P29 & P30 (full-length)
<i>Molinia caerulea</i>	<i>dhs2</i>	this study	KIEL	P29 & P30 (full-length)
Bambusoideae				
<i>Dendrocalamus latifolius</i>	<i>dhs</i>	this study	SRA (SRR1171494)	assembled from RNA-seq data
<i>Phyllostachys edulis</i>	<i>dhs</i>	this study	SRA (SRR2035212)	assembled from RNA-seq data
<i>Semiarundinaria fastuosa</i>	<i>dhs</i>	this study	KIEL	P04 & Q _i (3'RACE)
Chloridoideae				
<i>Eleusine coracana</i>	<i>dhs</i>	this study	KIEL	P02 & Q _i (3'R.), P06 & AUAP (5'R.)
<i>Oropetium thomaeum</i> [†]	<i>dhs</i>	LFJQ0100068	GenBank	
Danthonioideae				
<i>Bouteloua gracilis</i>	<i>dhs</i>	this study	KIEL	P02 & Q _i (3'RACE)
<i>Cortaderia selloana</i>	<i>dhs</i>	this study	KIEL	P04 & Q _i (3'RACE)

Supplementary table 1 (continued)

Species	Gene	Accession	Cultivar/Source	Further information
Ehrhartoideae				
<i>Ehrharta calycina</i>	<i>dhs</i>	this study	GRA 828 (IPK)	P02 & Q _i (3'RACE)
<i>Oryza brachyantha</i>	<i>dhs1</i>	XM_0066661177	GenBank	
<i>Oryza brachyantha</i>	<i>dhs2</i>	XM_006650498	GenBank	
<i>Oryza sativa</i>	<i>dhs1</i>	NM_001057753	GenBank	
<i>Oryza sativa</i>	<i>dhs2</i>	AP005579	GenBank	
Panicoideae				
Andropogoneae				
<i>Bothriochloa ischaemum</i>	<i>dhs1</i>	this study	KIEL	P02 & Q _i (3'R.), P07 & AUAP (5'R.)
<i>Chrysopogon zizanioides</i>	<i>dhs1</i>	this study	SRA (SRR2167610)	assembled from RNA-seq data
<i>Chrysopogon zizanioides</i>	<i>dhs3</i>	this study	SRA (SRR2167610)	assembled from RNA-seq data
<i>Coix lacryma-jobi</i>	<i>dhs1</i>	this study	KIEL	P02 & Q _i (3'RACE)
<i>Coix lacryma-jobi</i>	<i>dhs2</i>	this study	KIEL	P02 & Q _i (3'RACE)
<i>Cymbopogon winterianus</i>	<i>dhs2</i>	this study	SRA (SRR1614278)	assembled from RNA-seq data
<i>Microstegium vimineum</i>	<i>dhs3</i>	YPIC-2010006	Blast4OneKP	
<i>Miscanthus sinensis</i>	<i>dhs1</i>	this study	KIEL	P08 & P09 (full-length)
<i>Miscanthus sinensis</i>	<i>dhs2</i>	this study	KIEL	P10 & P11 (full-length)
<i>Miscanthus sinensis</i>	<i>dhs3</i>	this study	KIEL	P12 & P13 (full-length)
<i>Saccharum officinale</i>	<i>dhs1</i>	this study	KIEL	P14 & P15 (full-length)
<i>Saccharum officinale</i>	<i>dhs2</i>	this study	KIEL	P16 & P17 (full-length)
<i>Saccharum officinale</i>	<i>dhs3</i>	this study	KIEL	P18 & P19 (full-length)

Supplementary table 1 (continued)

Species	Gene	Accession	Cultivar/Source	Further information
Andropogoneae, Panicoideae (continued)				
<i>Sorghum bicolor</i> [‡]	<i>dhs1</i>	this study	KIEL	P20 & P21 (full-length, pET23a) cf. genome accession: XM_0024466442 Chromosome 1
<i>Sorghum bicolor</i> [‡]	<i>dhs2</i>	this study	KIEL	P22 & P11 (full-length, pET23a) cf. genome accession: XM_002444888 Chromosome 7
<i>Sorghum bicolor</i> [‡]	<i>dhs3</i>	this study	KIEL	P23 & P24 (full-length, pET23a) cf. genome accession: XM_002446223 Chromosome 6
<i>Spodiopogon sibiricus</i>	<i>dhs1</i>	this study	KIEL	P02 & Qi (3'RACE)
<i>Spodiopogon sibiricus</i>	<i>dhs2</i>	this study	KIEL	P02 & Qi (3'RACE)
<i>Zea mays</i> [‡]	<i>dhs1</i>	AM411617	Cultivar B73	Nurhayati et al., 2009, (pET3a) genome accession: NM_001155612
<i>Zea mays</i> [‡]	<i>dhs3</i>	this study	Cultivar B73	P25 & P26 (full-length, pET23a) cf. genome accession: NM_001137334
Paniceae				
<i>Digitaria sanguinalis</i>	<i>dhs</i>	this study	KIEL	P27 & P28 (full-length)
<i>Panicum hallii</i>	<i>dhs</i>	Pahal.0052s0086.1	Phytosome	
<i>Panicum virgatum</i>	<i>dhs1</i>	Pavirv00038084m	Phytosome	
<i>Panicum virgatum</i>	<i>dhs2</i>	Pavirv00006392m	Phytosome	
<i>Panicum virgatum</i>	<i>dhs3</i>	Pavirv00048310m	Phytosome	
<i>Setaria italica</i> [†]	<i>dhs</i>	XM_004981776	GenBank	

Supplementary table 1 (continued)

Species	Gene	Accession	Cultivar/Source	Further information
Pooideae				
<i>Achnatherum splendens</i>	<i>dhs</i>	this study	SRA (SRR3089957)	assembled from RNA-seq data
<i>Brachypodium distachyon</i> [†]	<i>dhs</i>	XM_003559868	GenBank	
<i>Stipa grandis</i>	<i>dhs</i>	this study	SRA (SRR1740003)	assembled from RNA-seq data
Poeae				
<i>Dactylis glomerata</i>	<i>dhs1</i>	this study	KIEL	P31 & P32 (full-length, pET23a)
<i>Dactylis glomerata</i>	<i>dhs2</i>	this study	KIEL	P33 & P34 (full-length, pET23a)
<i>Festuca arundinacea</i>	<i>dhs1</i>	GT037976	GenBank	
<i>Festuca arundinacea</i>	<i>dhs1b</i>	this study		P50 & P51 (full-length)
<i>Festuca arundinacea</i>	<i>dhs1c</i>	this study	GR 1333 (IPK)	P01 & Q ₀ (3'RACE)
<i>Festuca arundinacea</i>	<i>dhs2</i>	this study		P52 & P53 (full-length)
<i>Festuca pratensis</i>	<i>dhs1</i>	this study	SRA (SRR3096315)	assembled from RNA-seq data
<i>Festuca rubra</i> subsp. <i>rubra</i>	<i>dhs1c</i>	this study	GR 2197 (IPK)	P36 & P37 (full-length)
<i>Festuca rubra</i> subsp. <i>rubra</i>	<i>dhs2</i>	this study	GR 2197 (IPK)	P02 & Q _i (3'RACE)
<i>Holcus lanatus</i>	<i>dhs1</i>	this study	GR 4224 (IPK)	P31 & P35 (full-length)
<i>Holcus mollis</i>	<i>dhs1</i>	this study	GR 7269 (IPK)	P31 & P35 (full-length)
<i>Holcus mollis</i>	<i>dhs2</i>	this study	GR 7269 (IPK)	P38 & P39 (full-length)
<i>Lolium multiflorum</i>	<i>dhs1</i>	this study	Lema cultivar (NPZ)	P40 & P41 (full-length)
<i>Lolium perenne</i>	<i>dhs1</i>	ViaLactia		P40 & P41 (full-length, pET23a)
<i>Lolium perenne</i>	<i>dhs2a</i>	ViaLactia		P42 & P43 (full-length, pET23a)
<i>Lolium perenne</i>	<i>dhs2b</i>	ViaLactia		P44 & P45 (full-length, pET23a)

Supplementary table 1 (continued)

Species	Gene	Accession	Cultivar/Source	Further information
Poeae, Pooideae (continued)				
<i>Lolium temulentum</i>	<i>dhs1</i>	this study	Ceres cultivar (ABY-Ba)	P40 & P41 (full-length)
<i>Lolium temulentum</i>	<i>dhs2</i>	this study	Ceres cultivar (ABY-Ba)	P42 & P43 (full-length)
Triticeae				
<i>Aegilops tauschii</i> [#]	<i>dhs1</i>	EMT19875 (ASM34733v1)	EnsemblGenomes	
<i>Aegilops tauschii</i> [#]	<i>dhs2</i>	EMT19873 (ASM34733v1)	EnsemblGenomes	
<i>Agropyron desertorum</i>	<i>dhs1</i>	this study	SRA (SRR3087737)	assembled from RNA-seq data
<i>Hordeum brevisubulatum</i>	<i>dhs1</i>	JN107536	GenBank	
<i>Hordeum vulgare</i> [†]	<i>dhs1</i>	AK248438	GenBank	MLOC_51599 (ASM32608v1)
<i>Secale cereale</i>	<i>dhs1b</i>	this study	SRA (ERR1254238)	assembled from RNA-seq data
<i>Secale cereale</i>	<i>dhs2</i>	this study	SRA (ERR1254238)	assembled from RNA-seq data
<i>Triticum urartu</i> [#]	<i>dhs1</i>	TRIUR3_02692	EnsemblGenomes	
<i>Triticum urartu</i> [#]	<i>dhs1b</i>	TRIUR3_10402	EnsemblGenomes	
<i>Triticum urartu</i> [#]	<i>dhs2</i>	TRIUR3_02694	EnsemblGenomes	
<i>Triticum aestivum</i> [#]	subgenome A <i>dhs1</i>	TRIAE_CS42_5AL_TGACv1_373990_AA1186800	EnsemblGenomes	found on chromosome arm 5AL (TGACv1_scaffold_373990_5AL)
<i>Triticum aestivum</i> [#]	subgenome B <i>dhs1</i>	TRIAE_CS42_5BL_TGACv1_404663_AA1307680	EnsemblGenomes	found on chromosome arm 5BL (TGACv1_scaffold_404663_5BL)
<i>Triticum aestivum</i> [#]	subgenome D <i>dhs1</i>	TRIAE_CS42_5DL_TGACv1_433196_AA1405200	EnsemblGenomes	found on chromosome arm 5DL (TGACv1_scaffold_433196_5DL)

Supplementary table 1 (continued)

Species	Gene	Accession	Cultivar/Source	Further information
Triticeae, Pooideae (continued)				
<i>Triticum aestivum</i> †	subgenome A	<i>dhs1b</i> this study	TRI 12922 (IPK)	P46 & P47 (full-length, pET21d) found on chromosome arm 3AS (TGACv1_scaffold_211731_3AS: 18156-30895)
<i>Triticum aestivum</i> †	subgenome B	<i>dhs1b</i> TRAES3BF010400570CFD	EnsemblGenomes	found on chromosome 3B TGACv1_scaffold_222693_3B
<i>Triticum aestivum</i> †	subgenome D	<i>dhs1b</i> this study	TRI 12922 (IPK)	P46 & P47 (full-length), (also found on chromosome arm 3DS, IWGSC1.0: IWGSC_CSS_3DS_scaff_2491766, -2376710, -231871, -2305013) includes premature stop codon
<i>Triticum aestivum</i> †	subgenome A	<i>dhs2</i> TRIAE_CS42_5AL_TGACv1_373990_AA1186800	EnsemblGenomes	found on chromosome arm 5AL (TGACv1_scaffold_373990_5AL)
<i>Triticum aestivum</i> †	subgenome B	<i>dhs2</i> this study	TRI 12922 (IPK)	P48 & P49 (full-length, pET21d) found on chromosome arm 5BL TGACv1_scaffold_404663_5BL: 106636-110594)
<i>Triticum aestivum</i> †	subgenome D	<i>dhs2</i> TRIAE_CS42_5DL_TGACv1_435223_AA1447460	EnsemblGenomes	found on chromosome arm 5DL (TGACv1_scaffold_435223_5DL)

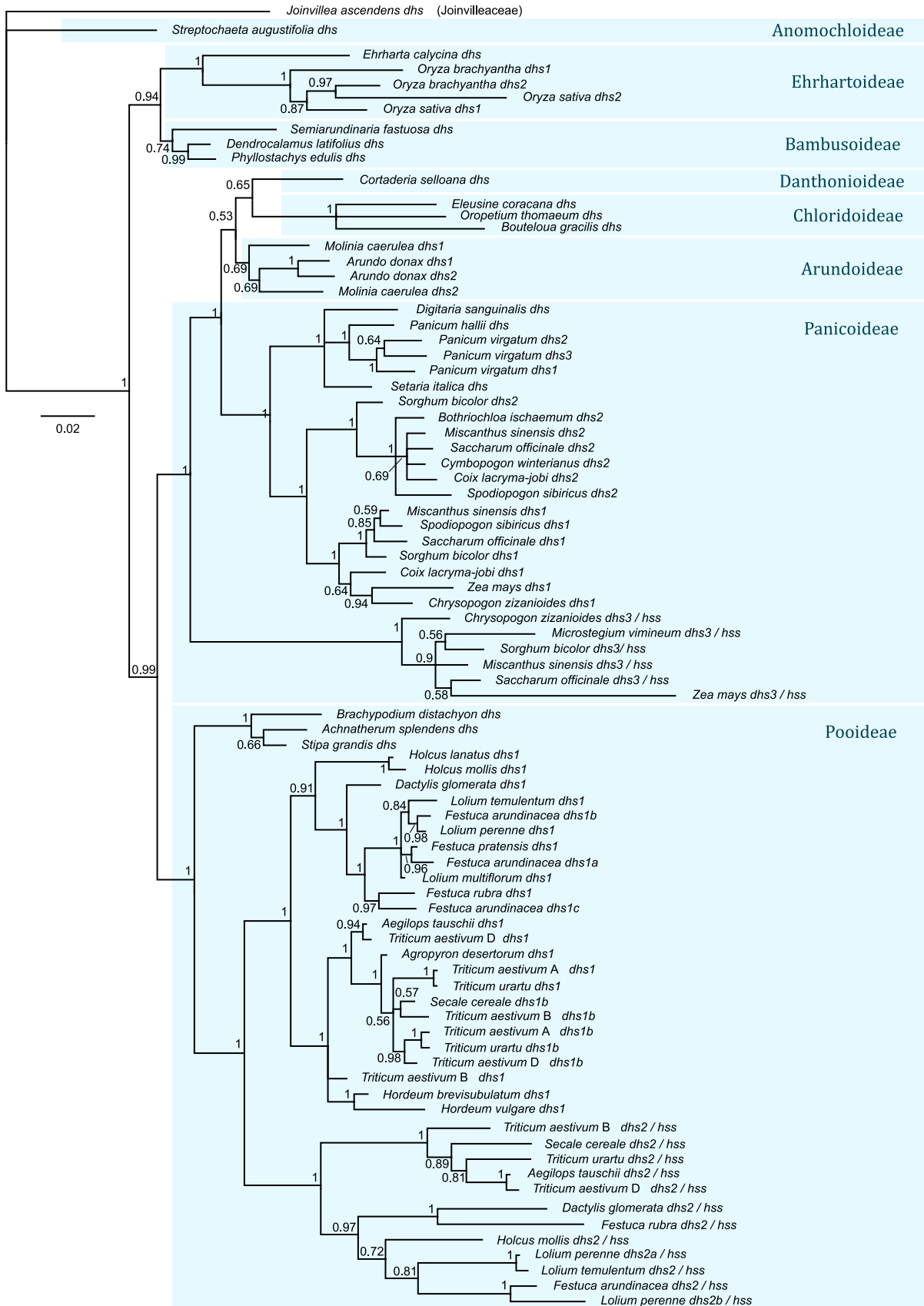
* Primer sequences in supplementary table 4

† Grass species with sequenced genome which contain only single copy of the *dhs* gene

‡ Grass species with sequenced genome which contain several copies of *dhs* homologous genes

Primer sequences (P01-P53) can be found in supplementary table 4

IPK - Genebank Gatersleben of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany; NPZ - Norddeutsche Pflanzenzucht Hans-Georg Lembke KG, Holtsee, Germany; ABY-Ba - Genebank IBERS, Aberystwyth University, Wales; Kiel - Botanischer Garten der Christian-Albrechts-Universität zu Kiel, Germany; NTBG - National Tropical Botanical Garden, Hawaii, USA



Supplementary fig. 1: Bayesian gene tree of *dhs* homolog coding sequences obtained various Poaceae species. Blue boxes indicate Poaceae subfamily affiliation. Node labels present posterior probabilities. Subgenome locations of *T. aestivum* coding sequences is indicated (A, B or D).

Supplementary table 2: Specific DHS and HSS enzyme activities of select Poaceae DHS homologs sorted by gene cluster (see fig. 1)

	Gene	DHS activity	HSS activity
Panicoideae			
A1	<i>Sorghum bicolor dhs1</i>	39.1 pkat/mg	368.3 pkat/mg
A2	<i>Sorghum bicolor dhs2</i>	37.0 pkat/mg	471.7 pkat/mg
B	<i>Sorghum bicolor dhs3</i>	1.5 pkat/mg	487.8 pkat/mg
Pooideae			
C	<i>Dactylis glomerata dhs1</i>	46.0 pkat/mg	205.6 pkat/mg
	<i>Triticum aestivum A dhs1b</i>	17.0 pkat/mg	350.5 pkat/mg
D	<i>Dactylis glomerata dhs2</i>	4.1 pkat/mg	907.1 pkat/mg
	<i>Triticum aestivum B dhs2</i>	0.6 pkat/mg	246.5 pkat/mg
	<i>Lolium perenne dhs2b</i>	0.2 pkat/mg	88.5 pkat/mg

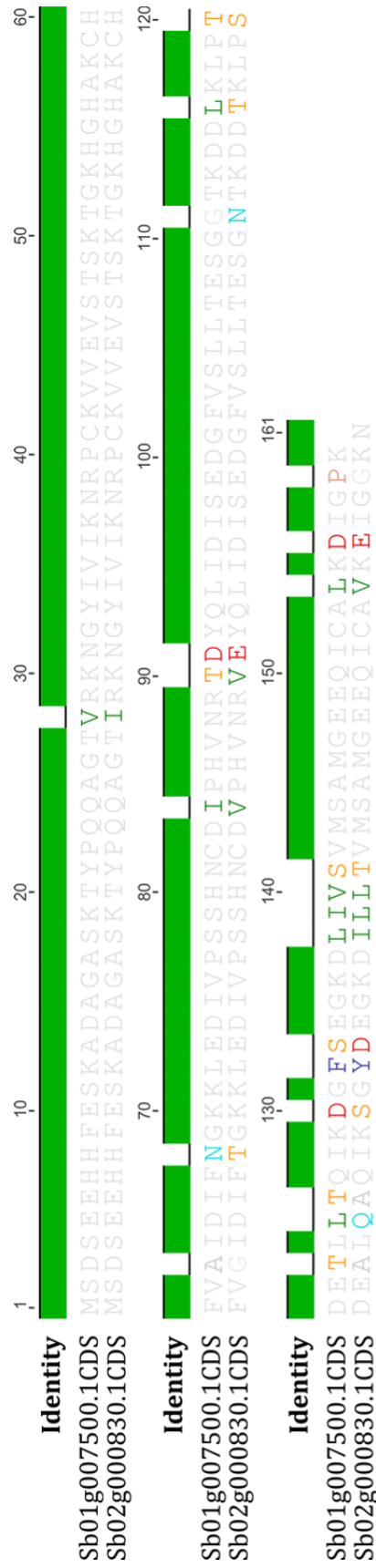
Supplementary table 3: Specific DHS and HSS enzyme activities of an ancestral DHS (protein 1) and HSS (protein 2) specific to the Pooideae *hss* cluster (green circle and white diamond in fig. 2, respectively)

	DHS activity	HSS activity
ancestral protein 1	14.9 pkat/mg	90.8 pkat/mg
ancestral protein 2	3.2 pkat/mg	124.0 pkat/mg

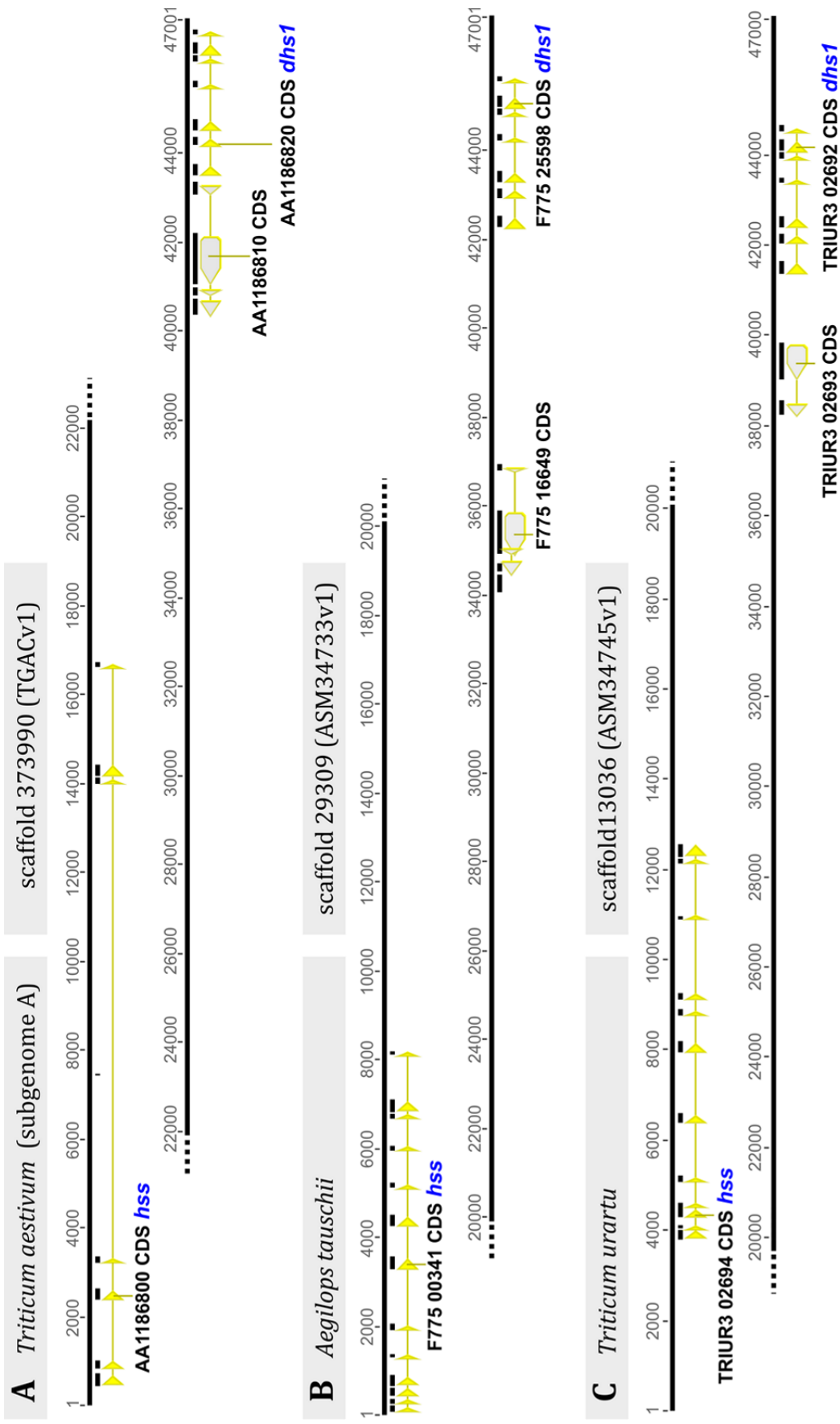
	1	10	20																										
ancestral protein 1	M	A	G	E	G	S	S	G	G	G	R	N	T	D	A	L	E	G	V	R	A	I	V	L	K	P	S	E	S
ancestral protein 2	M	A	G	E	G	S	S	G	G	G	R	N	T	D	A	L	E	G	V	R	A	I	L	L	K	P	S	E	S
	30	40	50																										
ancestral protein 1	L	D	E	S	R	F	T	R	I	A	G	A	D	F	N	D	A	G	L	G	L	A	G	L	L	G	S	L	A
ancestral protein 2	L	D	E	S	R	S	T	R	I	A	G	A	D	F	N	D	A	G	L	G	L	A	G	L	L	G	S	L	A
	60	70	80																										
ancestral protein 1	S	T	G	F	Q	A	S	N	L	G	D	A	I	D	V	V	N	Q	M	L	D	W	R	L	S	H	E	K	P
ancestral protein 2	S	T	G	F	Q	A	S	H	L	G	D	A	I	D	V	V	N	Q	M	L	D	W	R	L	S	H	E	K	P
	90	100	110																										
ancestral protein 1	S	E	D	C	D	E	A	E	L	D	P	K	Y	R	E	S	V	K	C	K	I	F	L	G	F	T	S	N	L
ancestral protein 2	S	E	D	C	D	E	A	E	L	D	P	K	Y	R	E	S	V	K	C	K	I	F	L	G	F	T	S	N	L
	120	130	140																										
ancestral protein 1	V	S	S	G	I	R	D	V	I	R	F	L	V	Q	H	H	M	V	D	V	I	V	T	T	A	G	G	I	E
ancestral protein 2	V	S	S	G	I	R	D	V	I	R	F	L	V	Q	H	H	M	V	D	V	I	V	T	T	A	G	G	I	E
	150	160	170																										
ancestral protein 1	E	D	L	I	K	C	L	A	P	T	Y	R	G	D	F	S	L	P	G	A	L	L	R	S	K	G	L	N	R
ancestral protein 2	E	D	L	I	K	C	L	A	P	T	Y	R	G	D	F	S	L	P	G	A	L	L	R	S	K	G	L	N	R
	180	190	200																										
ancestral protein 1	I	G	N	L	L	V	P	N	D	N	Y	C	K	F	E	N	W	I	M	P	L	F	D	Q	M	L	L	E	Q
ancestral protein 2	I	G	N	L	L	L	P	N	D	N	Y	C	K	F	E	K	W	I	M	P	L	L	D	Q	M	L	L	E	Q
	210	220	230																										
ancestral protein 1	S	T	E	N	V	W	T	P	S	K	V	I	A	R	L	G	K	E	I	N	D	E	S	S	Y	L	Y	W	A
ancestral protein 2	S	T	K	N	V	W	T	P	S	K	V	I	A	R	L	G	K	E	I	N	D	E	S	S	Y	L	Y	W	A
	240	250	260																										
ancestral protein 1	Y	K	N	N	I	P	V	Y	C	P	A	L	T	D	G	S	L	G	D	M	L	F	C	H	A	V	R	N	P
ancestral protein 2	H	K	N	N	I	P	V	Y	C	P	A	L	T	D	G	S	I	G	D	M	L	F	C	H	A	V	H	N	P
	270	280	290																										
ancestral protein 1	G	L	I	I	D	I	V	Q	D	I	R	L	M	N	G	E	A	I	H	A	T	P	R	K	T	G	V	I	I
ancestral protein 2	G	L	I	I	D	I	V	Q	D	V	R	L	M	N	G	E	A	I	H	A	T	P	R	K	T	G	V	I	I
	300	310																											
ancestral protein 1	L	G	G	G	L	P	K	H	H	I	C	N	A	N	M	F	R	N	G	A	D	Y	A	V	Y	I	N	T	A
ancestral protein 2	L	G	G	G	L	P	K	H	H	I	C	N	A	N	M	L	R	N	G	A	D	Y	A	V	Y	I	N	T	A
	320	330	340																										
ancestral protein 1	Q	E	F	D	G	S	D	S	G	A	Q	P	D	E	A	V	S	W	G	K	I	K	G	S	A	K	P	V	K
ancestral protein 2	Q	E	F	D	G	S	D	S	G	A	H	P	D	E	A	V	S	W	G	K	I	K	G	S	A	K	P	V	K
	350	360	370	372																									
ancestral protein 1	V	H	C	D	A	T	I	A	F	P	L	L	V	A	A	T	F	A	R	R	F	H	G	A					
ancestral protein 2	V	H	C	D	A	T	I	A	F	P	L	L	V	A	A	T	F	A	R	R	F	H	G	A					

 nonpolar/hydrophobic	 polar	 acidic polar	 basic polar
---	--	---	--

Supplementary fig. 2: Alignment of reconstructed ancestral protein sequences



Supplementary fig. 3: Alignment of *Sorghum bicolor* eIF5A protein sequences



Supplementary fig. 4: Co-localization of *hss* and *dhs1* genes on the same chromosome *T. aestivum* and close relatives
Based on genomic sequence data, coding regions (CDS) of genes are marked

Supplementary table 4: Primers and nucleotide sequences used in this study.
Suppl. table 1 contains information for which amplification each primer was used for.

Name	Sequence (5' to 3')	Reference
Q_T	CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGC TTTTTTTTTTTTTTTTTTTT	1
Q_i	GAGGACTCGAGCTCAAGC	1
Q_o	CCAGTGAGCAGAGTGACG	1
AUAP	GGCCACGCGTCGACTAGTAC	2
AAP	GGCCACGCGTCGACTAGTAC (G) ₁₆	2
P01	CCGGCGCCGACTTCAAC	this study
P02	TTGYAARTTTGARAANTGGAT	this study
P03	GACGCCATCGACGTCGTCAAACAGA	this study
P04	GGARGAYCTCATHAAATGCC	this study
P05	CATGATCCAGTTCTCAAACCTTG	this study
P06	CTTTGCCAAGACGAGCTATCAC	this study
P07	ACCTCTGTAAGTTGGTGCAAGG	this study
P08	ATAATCATATGGCCGGCGCGGGA	this study
P09	ATAATGTCGACTTTGGAGCTGTGAACCTTACGTGCA	this study
P10	ATAATCATATGGCCGGCGCGGGAGG	this study
P11	ATATAGTCGACCTTGGAGCCATGAACCTTGCGTGCAAA	this study
P12	ATAGCATTAAATATGGGTGCCGCGCCG	this study
P13	ATAATCTCGAGGTTTGTGATTTAGTGCCGTGAACCTTG	this study
P14	GGCTCCGGACTCCAGAGATG	this study
P15	CCGAAAGGGATTCAGTTTGTTC	this study
P16	ATAATCATATGGCCGGCGCGGCA	this study
P17	ATAATGTCGACTTTGGAGCCGTGAACCTTGC	this study
P18	ATAATCATATGGGTGCCGCGCCCGGA	this study
P19	ATAATCTCGAGTCAGTTTGTGATTTAGTGCCATGAACCTTGT	this study
P20	ATATACATATGGCCGGCGCGGGAGGCAGCGG	this study
P21	ATATAGTCGACTTTGGAGCTGTGAACCTTACGTGCAAAAT	this study
P22	ATATACATATGGCCGGCGCGGGAGGCAGCGG	this study
P23	ATATACATATGGGTGCCGCGCCGGGAGGAGG	this study
P24	ATATACTCGAGGTTTGTGATTTAGTGCCATGAACCTT	this study
P25	ATATACATATGGGTGCCGAGTTGGGAGAAGC	this study
P26	ATATCTCGAGGTTTGTGATTTAGTGCCGTGAACCTT	this study
P27	AGCTCCAGACTCCAGAGATGG	this study
P28	CAGCTGGATCCATAGCTCACTG	this study
P29	AAGGCTTCACCGGGGATG	this study
P30	GTCAAGTAAAGCTCGTGCTCC	this study

Supplementary table 4 (continued)

Name	Sequence (5' to 3')	Reference
P31	ATAATATTAATTCCAGCCTCCAGGGATTCG	this study
P32	ATAATCTCGAGGTTGGTTGAATTTGTGCCATG	this study
P33	ATAATATTAATAGCAGCTCGGACTCCAGAGATG	this study
P34	GAGGTCCTTAGTTTCGTCATCAG	this study
P35	CGCCWGAAATTCAGTWGGTT	this study
P36	AGTGATTCGGCGGAGATG	this study
P37	CTCCGAAAATTCAGTTGGTCTG	this study
P38	GTACGGTGAAGCTCGAGTTCTC	this study
P39	CGCCACCTGAAATTCAGTAGG	this study
P40	ATATACATATGGCCGGCGAGGGAAGCAGCG	this study
P41	ATATACTCGAGGTTGGTTCGGATTTGTGCCATGAAACC	this study
P42	ATATACATATGGCCGGCGATGGAGGCGGCGA	this study
P43	ATATACTCGAGATTTGCGCCATGAAACCTACGTGCAAA	this study
P44	ATATACATATGGCTGCCAATGGAGACGTCAACG	this study
P45	ATATACTCGAGTTCCAGGCCATGAAATTTACGGGC AAA	this study
P46	ATTAACCATGGGAGGTGACGGGATGAG	this study
P47	ATATTCTCGAGATTGGTTGAATTTGCACTATGAG	this study
P48	TCGCCGGCCATGGCCGGCGAGG	this study
P49	AATAACTCGAGATTCACACCATGAAACGTACGTG	this study
P50	TATATCATATGGCCGGCGAGGGAAGCAGC	this study
P51	TATATCTCGAGGTTGGTTCGGATTTGTGCCATG	this study
P52	TATATCATATGACTGCCAATGGAGACGGC	this study
P53	TATATCTCGAGGCCAGGCCATGAAACTTACG	this study

Y = C or T, R = A or G, H = A, C, or T, N = any base (following IUPAC nomenclature)

1 - Scotto-Lavino, E., Du, G., and Frohman, M.A. (2007). *Nat. Protoc.* 1, 2742–2745.

2 - Ober, D., and Hartmann, T. (1999b). *J. Biol. Chem.* 274, 32040–32047.

Recombination analysis of “Pooideae” full-length sequence alignment

GARD output on datamonkey website (Pond et al., 2006)

Recombination report

Number of breakpoints	AICc	Δ AIC	Segments
0	14045.9		1 (total length of 1074 bp)
1	13969.7	76.2049	2, with breakpoint at 378 bp

KH testing report

Location of breakpoint	LHS p-value	RHS p-value	Significance
378	0.00020	0.00040	*** (p-value = 0.01)

Conclusion

Evidence of **1** breakpoint(s) was found using AICc; an improvement of 76.2049 points was achieved over the model without recombination. Of those, **1** breakpoint(s) also indicated significant topological incongruence using the KH test.

GARD recombination analysis was run using the HKY85 model.

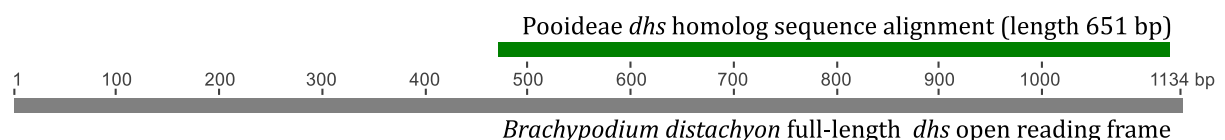
KH = Kishino Hasegawa

LHS p-value = Adjusted KH p-value that the partition to the left of the of this BP has a topology different from the one inferred from the partition on the right

RHS p-value = Adjusted KH p-value that the partition to the right of the of this BP has a topology different from the one inferred from the partition on the left

Pond, S.L.K., Posada, D., Gravenor, M.B., Woelk, C.H., and Frost, S.D.W. (2006). *Bioinformatics* 22, 3096–3098.

Final “Pooideae” alignment length, *B. distachyon dhs* ORF as reference:



“Pooideae” sequence alignment used for selection analyses (in FASTA format)

```
>Oropetium thomaemum dhs
AGAGGTGATTTTCTTTACCTGGAGCACTCTGCGGTCAAAGGGACTGAACCGGATAGGAAATTTGTTGGTGCCGAATGA
TAACTATTGCAAGTTTGAGAACTGGATCATGCCACTCTTTGACAAAGATGTTGCTGGAAACAATCTCTGAGAACGTTTGGGA
CACCATCAAAGGTGATAGCTCGTCTTGGCAAAGGAGATAAACGATGAAAAGTTCTTACCTTTATTGGGCATACAAGGACAAT
ATCCCTGTCTACTGTCCAGCGCTGACCGATGGATCACTTGGAGACATGTTGTTTTGTCATGCAGTCTGCAATCCTGGTCT
TATAATTGATATTGTACAAGATATACGATTGATGAATGGGGAAGCCATCCACGCAACCCCAAGGAAGACAGGGATCATAG
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCCAATATGTTTTCGCAACGGTGCAGATTATGCAGTATACATT
AACACAGCTCAAGAAATTTGATGGTAGTGACTCTGGAGCACAGCCTGACGAAAGCAGTTTTCATGGGAAAGATCAAGGGTTC
AGCAAACCTGTGAAGGTGCATTGCGATGCAAGTATTGCTTTCCCTCTCTTGTGCTGCAACATTTGCGCCTAAGGTTT
ACGGCTACAAA
>Sorghum bicolor dhs2
AGAGGTGATTTTCTTTACCTGGAGCACTACTGCGGTCAAAGGGACTGAACCGGATAGGAAATCTGTTGGTGCCGAATGA
TAACTATTGCAAGTTTGAGAACTGGATCATGCCACTCTTTGACCAGATGCTATTGCAACAATCTACTGAGAATGTTTGGGA
CACCATCAAAGGTGATAGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTATCTTTACTGGGCATACAAGAAATAAC
ATACCTGTATACTGCCCTGCATTGACTGATGGATCACTCGGAGACATGCTGTTCTGTCAATGCAGTCCGTAATCCTGGCCT
TATTATTGACATTGTACAAGATATACGGCTGATGAATGGGGAAGCCATCCATGCAACACCAAGGAAGACAGGGATTATAG
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCCAATATGTTTTCGCAATGGTGCAGATTATGCTGTTTTATATT
AACACAGCTCAAGAGTTTGTATGGTAGTGACTCTGGAGCACAGCCTGATGAAGCAGTCTCATGGGCAAGATCAAGGGTTC
AGCAAACCTGTAAAGGTGCATTGCGATGCAAGCATCGCTTTCCCGCTTCTTGTGCTGCGACATTTGCACGCAAGGTTT
ATGGCTCCAAG
>Setaria italica dhs
AGAGGTGATTTTCTTTACCTGGAGCACTACTGCGGTCAAAGGGTGAACCGGATAGGAAATCTGTTGGTGCCCAATGA
TAACTATTGTAAGTTTGAGAACTGGATCATGCCACTATTGACCAGATGCTACTAGAACAATCTACTGAGAATATTGGA
CACCATCAAAGGTGATAGCTCGTCTTGGCAAAGAAATAAATGACGAAAGCTCCTATCTTTATTGGGCATACAAGAAATAAC
ATTCCTGTATACTGCCAGCATTGACTGATGGATCACTTGGAGACATGCTGTTCTGTCAATGCAGTCCGCAATCCTGGCCT
TATTATTGACATTGTACAAGATATACGGCTGATGAATGGGGAAGCCATCCACGCAAGATCGAGGAAGACAGGGATAATAG
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCTAATATGTTCCGCAATGGAGCAGACTATGCAGTCTATATC
AACACAGCTCAAGAGTTTGTATGGTAGTGACTCTGGAGCACAGCCTGATGAAGCAGTCTCATGGGCAAGATCAAGGGTTC
AGCAAACCTGTAAAGGTGCATTGCGATGCAAGCATCGCTTTCCCGCTTCTTGTGCTGCGACATTTGCACGCAAGGTTT
ATGGCTCCAAG
>Aegilops tauschii dhs1
CGGGTGAATTTTCTTTACCTGGGCACTGCTACGGTCAAAGGGACTTAAACCGGATAGGAAATCTGTTGGTGCCCAATGA
TAACTATTGCAAGTTTGAAAACTGGATCATGCCACTCTTTGACCAGATGCTACAAGAACAATCTACTGAGAATGCTTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTATTGGGCATACAAGAAACAAT
ATTCTGTATACTGCCAGCATTGACTGATGGATCACTCGGAGACATGCTGTTTTGTCATGCAGTTTCGCAATCCTGGTCT
TATTATCGACATTGTACAAGATATAAGGCTGATAAATGGTGAAGCCATTTCATGCAAGCCCAAGGAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTCCGCAATGGTGCAGATTATGCAGTCTATATC
AACACGGCTCAAGAGTTTGTATGGAAGTGACTCAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAAGGTTT
AGCAAAGCTGTAAAGGTTTCACTGTGATGCGACTATCGCTTTCCCGTACTTGTAGCCGCAACATTTGCACGGAGGTTCT
ACAGTGCAAAT
>Triticum urartu dhs1
CGAGGTGAATTTTCTTTACCTGGGCACTGCTACGGTCAAAGGGACTTAAACCGGATAGGAAATCTGTTGGTGCCCAATGA
TAACTATTGCAAGTTTGAAAACTGGATCATGCCACTCTTTGACCAGATGCTACAAGAACAATCTACTGAGAATGTTTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTATTGGGCATACAAGAAACAAT
ATTCTGTATACTGCCAGCATTGACTGATGGATCACTCGGAGACATGCTGTTTTGTCATGCAGTTTCGCAATCCTGGTCT
TATTATCGACATTGTACAAGATATAAGGCTGATAAATGGTGAAGCCATTTCATGCAAGCCCAAGGAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTTTCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAGTTTGTATGGAAGTGACTCAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAAGGTTT
AGCAAAGCTGTAAAGGTTTCACTGTGATGCGACTATCGCTTTCCCGTACTTGTAGCCGCAACATTTGCACGGAGGTTCT
ACAGTGCAAAT
>Triticum aestivum dhs1 chromosome 5B
CGAGGTGAATTTTCTTTACCTGGGCACTGCTACGGTCAAAGGGACTTAAACCGGATAGGAAATCTGTTGGTGCCCAATGA
TAACTATTGCAAGTTTGAAAACTGGATCATGCCACTCTTTGACCAGATGCTACAAGAACAATCTACTGAGAATGTTTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTATTGGGCATACAAGAAACAAT
ATTCTGTATACTGCCAGCATTGACTGATGGATCACTCGGAGACATGCTGTTTTGTCATGCAGTTTCGCAATCCTGGTCT
TATTATCGACATTGTACAAGATATAAGGCTGATAAATGGTGAAGCCATTTCATGCAAGCCCAAGGAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTTTCGCAATGGTGCAGATTATGCAGTCTATATC
AACACGGCTCAAGAGTTTGTATGGAAGTGACTCAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAAGGTTT
AGCAAACCTGTAAAGGTTTCACTGTGATGCGACTATCGCTTTCCCGTACTTGTAGCCGCAACATTTGCACGGAGGTTCT
ATGGTGCAAAT
>Hordeum brevisubulatum dhs1
CGAGGTGAATTTTCTTTACCTGGGCACTGCTACGGTCAAAGGGACTTAAACCGGATAGGAAATCTGTTGGTGCCCAATGA
CAACTACTGCAAGTTTGAAAACTGGATCATGCCACTCTTTGACCAGATGCTACAAGAACAATCTACTGAGAATGTTTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTATTGGGCATACAAGAAACAAT
ATTCTGTATACTGCCAGCATTGACTGATGGATCACTCGGAGACATGCTGTTTTGTCATGCAGTTTCGCAATCCTGGTCT
TATTATTGACATTGTACAAGATATAAGGCTGATAAATGGTGAAGCCATTTCATGCAAGCCCAAGGAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTTTCGCAATGGTGCAGATTATGCAGTCTATATC
AACACGGCTCAAGAGTTTGTATGGAAGTGACTCAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAAGGTTT
AGCAAACCTGTAAAGGTTTCACTGTGATGCGACTATCGCTTTCCCGTACTTGTAGCCGCAACATTTGCACGGAGGTTCT
ATGGTGCAAAT
>Hordeum vulgare dhs1
CGAGGTGAATTTTCTTTACCTGGGCACTGCTGCGGTCAAAGGGACTTAAACCGGATAGGAAATCTGTTGGTGCCCAATGA
CAACTACTGCAAGTTTGAAAACTGGATCATGCCACTCTTTGACCAGATGTTACAAGAACAATCTACTGAGAATGTTTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGCAAAGAAATAAATGACGAAACCTCCTACCTTTATTGGGCATACAAGAAACAAT
ATTCTGTATACTGCCAGCATTGACTGATGGATCACTCGGAGACATGTTGTTTTGTCACGCAAGTTTCGCAATCCTGGTCT
TATTATTGACATTGTACAAGATATAAGGCTGATAAATGGTGAAGCCATTTCATGCAAGCCCAAGGAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTTTCGCAATGGTGCAGATTATGCAGTCTATATC
AACACGGCTCAAGAGTTTGTATGGAAGTGACTCAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAAGGTTT
AGCAAACCTGTAAAGGTTTCACTGTGATGCGACTATCGCTTTCCCGTACTTGTAGCCGCAACATTTGCACGGAGGTTCT
GTAGTGCAAAT
```



```

>Triticum aestivum dhs1b chromosome 3A
CGAGGTGAATTTTCTTTACCTGGGGCACTTCTACGGTCAAAGGACTTAACCGGATAGGAAATCTGTTGGTGCCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCACTCTTTGACCAGATGCTACAAGAACAATCTACTGAGAATGTCTGGA
CACCATCAAAGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTACTGGGCATACAAGAACAAT
ATTCTGTATACCTGCCCAGCATTGACTGATGGATCAGTTGGAGACATGCTGTTTTGTTCATGCAGTTCGCAATCCTGGTCT
TATTATCGACATTGTACAAGATATAAGGCTGATAAATGGTGAAGCCATTTCATGCAAGCCCAAGGAAGACAGGGGCCATAA
TTCTTGGTGGAGGCCTTCCAAAGCATCATATGCAATGCAAAATATGTTTCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAGTTTGTATGGAAGCGACTCAGGAGCACACCCAGATGAAGCAGTTTCATGGGGAAAAGATCAAAGGTTT
AGCAAAGCCTGTAAAGGTTCACTGTGATGCTACTATCGCTTTCCATTACTTGTAGCCGCAACATTTGCACGGAGGTCTC
GTAGTGCAAAT
>Triticum aestivum dhs1b chromosome 3D
CGAGGTGAATTTTCTTTACCTGGGGCACTGCTACGGTCAAAGGACTTAACCGGATAGGAAATCTGTTGGTGCCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCACTCTTTGACCAGATGCTACAAGAACAATCTACTGAGAATGTCTGGA
CACCATCAAAGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTACTGGGCATACAAGAACAAT
ATTCTGTATACCTGCCCAGCATTGACTGATGGATCAGTTGGAGACATGCTGTTTTGTTCATGCAGTTCGCAATCCTGGTCT
TATTATCGACATTGTACAAGATATAAGGCTGATAAATGGTGAAGCCATTTCATGCAAGCCCAAGGAAGACAGGGGCCATAA
TTCTTGGTGGAGGCCTTCCAAAGCATCATATGCAATGCAAAATATGTTTCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAGTTTGTATGGAAGCGACTCAGGAGCACACCCAGATGAAGCAGTTTCATGGGGAAAAGATCAAAGGTTT
AGCAAAGCCTGTAAAGGTTCACTGTGATGCTACTATCGCTTTCCCGTTACTTGTAGCCGCAACATTTGCACGGAGGTCTC
GTAGTGCAAAT
>Triticum urartu dhs1b
CGAGGTGAATTTTCTTTACCTGGGGCACTTCTACGGTCAAAGGACTTAACCGGATAGGAAATCTGTTGGTGCCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCACTCTTTGACCAGATGCTACAAGAACAATCTACTGAGAATGTCTGGA
CACCATCAAAGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTACTGGGCATACAAGAACAAT
ATTCTGTATACCTGCCCAGCATTGACTGATGGATCAGTTGGAGACATGCTGTTTTGTTCATGCAGTTCGCAATCCTGGTCT
TATTATCGACATTGTACAAGATATAAGGCTGATAAATGGTGAAGCCATTTCATGCAAGCCCAAGGAAGACAGGGGCCATAA
TTCTTGGTGGAGGCCTTCCAAAGCATCATATGCAATGCAAAATATGTTTCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAGTTTGTATGGAAGCGACTCAGGAGCACACCCAGATGAAGCAGTTTCATGGGGAAAAGATCAAAGGTTT
AGCAAAGCCTGTAAAGGTTCACTGTGATGCTACTATCGCTTTCCCGTTACTTGTAGCCGCAACATTTGCACGGAGGTCTC
GTAGTGCAAAT
> Triticum aestivum dhs1b chromosome 3B
CGAGGTGAATTTTCTTTACCTGGGGCACTGCTACGGTCAAAGGACTTAACCGGATAGGAAATCTGTTGGTGCCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCACTCTTTGACCAGATGCTACAAGAACAATCTACTGAGAATGTCTGGA
CACCATCAAAGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTACTGGGCATACAAGAACAAT
ATTCTGTATACCTGCCCAGCATTGACTGATGGATCAGTTGGAGACATGCTGTTTTGTTCATGCAGTTCGCAATCCTGGTCT
TATTATCGACATTGTACAAGATATAAGGCTGATAAATGGTGAAGCCATTTCATGCAAGCCCAAGGAAGACAGGGGCCATAA
TTCTTGGTGGAGGCCTTCCAAAGCATCATATGCAATGCAAAATATGTTTCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAGTTTGTATGGAAGCGACTCAGGAGCACACCCAGATGAAGCAGTTTCATGGGGAAAAGATCAAAGGTTT
AGCAAAGCCTGTAAAGGTTCACTGTGATGCTACTATCGCTTTCCCGTTACTTGTAGCTGCAACATTTGCACGGAGGCCTC
ATAGTGCAAAT
>Secale cereale dhs1b
CGAGGTGAATTTTCTTTACCTGGGGCACTGCTACGGTCAAAGGACTTAACCGGATAGGAAATCTGTTGGTGCCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCACTCTTTGACCAGATGCTACAAGAACAATCTACTGAGAATGTCTGGA
CACCATCAAAGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTACTGGGCATACAAGAACAAT
ATTCTGTATACCTGCCCAGCATTGACTGATGGATCAGTTGGAGACATGCTGTTTTGTTCATGCAGTTCGCAATCCTGGTCT
TATTATCGACATTGTACAAGATATAAGGCTGATAAATGGTGAAGCCATTTCATGCAAGCCCAAGGAAGACAGGGGCCATAA
TTCTTGGTGGAGGCCTTCCAAAGCATCATATGCAATGCAAAATATGTTTCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAGTTTGTATGGAAGCGACTCAGGAGCACACCCAGATGAAGCAGTTTCATGGGGAAAAGATCAAAGGTTT
AGCAAAGCCTGTAAAGGTTCACTGTGATGCTACTATCGCTTTCCCGTTACTTGTAGCTGCAACATTTGCACGGAGGTCTC
ATAGTTCAAAT
>Dactylis glomerata dhs1
CGAGGTGAATTTTCTTTACCTGGGGCACTGCTGCGGTCAAAGGACTTAATCGGATAGGAAATCTGTTGGTGCCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCAATCTTTGACCAGATGCTACAAGAACAATCTAGTGAGAATGTCTGGA
CACCATCAAAGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTGTATTGGGCATACAAGAACAAT
ATTCTGTATACCTGCCCAGCATTGACTGATGGATCAGTTGGAGACATGCTGTTTTGTTCATGCAGTTCGCAATCCTGGTCT
TATTATTGACATTGTACATGATATAAGGCTGATAAATGGTGAAGCCATTTCATGCAAGCCCAAGAAAGACTGGGGTCATAA
TTCTTGGTGGAGGCCTTCCAAAGCATCATATATGCAATGCAAAATATGTTTCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAGTTTGTATGGAAGCGACTCAGGGGCACAGCCAGATGAAGCAGTTTCATGGGGAAAAGATCAAAGGTTT
AGCAAACCTGTAAAGTCCATTGTGATGCGACTATTGCTTTCCATTACTCGTTGCCGCAACATTTGCACGTAGGTGTCT
ATGGCACAAAT
>Festuca arundinacea dhs1b
CGAGGCGAATTTTCTTTACCTGGGGCACTGCTGAGGTCAAAGGACTTAATCGGATAGGAAATCTGTTGGTGCCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCAATCTTTCGACAAGATGCTACAAGAACAATCTAGTGAGAATGTCTGGA
CACCATCAAAGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTATTGGGCATACAAGAACAAT
ATTCTGTATACCTGCCCAGCATTGACTGATGGATCAGTTGGAGACATGCTGTTTTGTTCATGCAGTTCGCAATCCTGGTCT
TATTATTGACATTGTACATGATATAAGGCTCATAAATGGTGAAGCCATTTCATGCAAGCCCAAGAAAGACTGGGGTCATAA
TTCTTGGTGGAGGCCTTCCAAAGCATCATATATGCAATGCAAAATATGTTTCGCAATGGTGCAGATTATGCAGTGTACATC
AACACAGCTCAAGAGTTTGTATGGAAGTGAAGTCAAGGAGCACAGCCAGATGAAGCAGTTTCATGGGGAAAAGATCAAAGGTTT
AGCAAACCTGTGAAGTCCATTGTGATGCAACTATTGCTTTCCATTGCTCGTTGCCGCAACATTTGCAGTGTAGGTTTC
ATGGCACAAAT
>Lolium multiflorum dhs1
CGAGGCGAATTTTCTTTACCTGGGGCACTGCTGCGGTCAAAGGACTTAATCGGATAGGAAATCTGTTGGTACCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCAATCTTTCGACAAGATGCTACAAGAACAATCTAGTGAGAATGTCTGGA
CACCATCAAAGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTATTGGGCATACAAGAACAAT
ATTCTGTATACCTGCCCAGCATTGACTGATGGATCAGTTGGAGACATGCTGTTTTGTTCATGCAGTTCGCAATCCTGGTCT
TATTATTGACATTGTACATGATATAAGGCTCATAAATGGTGAAGCCATTTCATGCAAGCCCAAGAAAGACTGGGGTCATAA
TTCTTGGTGGAGGCCTTCCAAAGCATCATATATGCAATGCAAAATATGTTTCGCAATGGTGCAGATTATGCAGTGTACATC
AACACAGCTCAAGAGTTTGTATGGAAGTGAAGTCAAGGAGCACAGCCAGATGAAGCAGTTTCATGGGGAAAAGATCAAAGGTTT
AGCAAACCTGTGAAGTCCATTGTGATGCAACTATTGCTTTCCATTGCTCGTTGCCGCAACATTTGCAGTGTAGGTTTC
ATGGCACAAAT

```

>Festuca pratensis dhs1
CGAGGCGAATTTTCTTTACCTGGGGCACTGCTGCGGTCAAAGGACTTAATCGGATAGGAAATCTGTTGGTACCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCAATCTTCGACAAGATGCTACAAGAACAACTTAGTGAGAATGCTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTATCTTTATTGGGCATACAAGAACAAAT
ATTCTGTATACAGCCAGCATTGACTGATGGATCACTTGGAGACATGCTGTTTTGTCTATGCAGTTCCGCAATCCTGGTCT
TATTATTGACATTGTACATGATATAAGGCTCATAAATGGTGAAGCCATTTCATGCAAGCCCAAGAAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTCCGCAATGGTGCAGATTATGCAGTCTACATC
AACACAGCTCAAGAGTTTCGATGGAAGTGACTCAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAGGGTTC
AGCAAAACCTGTGAAGGTTCCATTGTGATGCAACTATTGCTTTCCATTGCTCGTTGCCGCAACATTTGCCGCTAGGTTTC
ATGGCACAAAT

>Lolium temulentum dhs1
CGAGGCGAATTTTCTTTACCTGGGGCACTGCTGCGGTCAAAGGACTTAATCGGATAGGAAATCTGTTGGTACCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCAATCTTCGACAAGATGCTACAAGAACAACTTAGTGAGAATGCTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTATTGGGCATACAAGAACAAAT
ATTCTGTATACAGCCAGCATTGACTGATGGATCACTTGGAGACATGCTGTTTTGTCTATGCAGTTCCGCAATCCTGGTCT
TATTATAGACATTGTACATGATATAAGGCTCATAAATGGTGAAGCCATTTCATGCAAGCCCAAGAAAGACAGGGGTCATCA
TTCTTGGCGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTCCGCAATGGTGCAGATTATGCAGTCTACATC
AACACAGCTCAAGAGTTTCGATGGAAGTGACTCAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAGGGTTC
AGCAAAACCTGTGAAGGTTCCATTGTGATGCAACTATTGCTTTCCATTGCTCGTTGCCGCAACATTTGCCGCTAGGTTTC
ATGGCACAAAT

>Lolium perenne dhs1
CGAGGCGAATTTTCTTTACCTGGGGCACTGCTGCGGTCAAAGGACTTAATCGGATAGGAAATCTGTTGGTACCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCAATCTTCGACAAGATGCTACAAGAACAACTTAGTGAGAATGCTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGAAAAGAAATAAATGATGAAAAGCTCCTACCTTTATTGGGCATACAAGAACAAAT
ATTCTGTATACAGCCAGCATTGACTGATGGATCACTTGGAGACATGCTGTTTTGTCTATGCAGTTCCGCAATCCTGGTCT
TATTATTGACATTGTACATGATATAAGGCTCATAAATGGTGAAGCCATTTCATGCAAGCCCAAGAAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTCCGCAATGGTGCAGATTATGCAGTGTACATC
AACACAGCTCAAGAGTTTCGATGGAAGTGACTCAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAGGGTTC
AGCAAAACCTGTGAAGGTTCCATTGTGATGCAACTATTGCTTTCCATTGCTCGTTGCCGCAACATTTGCCGCTAGGTTTC
ATGGCACAAAT

>Festuca rubra dhs1
CGAGGCGAATTTTCTTTACCTGGGGCACTGCTGCGGTCAAAGGACTTAATCGGATAGGAAATCTGTTGGTACCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCAATCTTCGACAAGATGCTACAAGAACAACTTAGTGAGAATGCTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGCAAAGAAATAAATGACGAAAAGCTCCTACCTTTATTGGGCATACAAGAACAAAT
ATTCTGTATACAGCCAGCATTGACTGATGGATCACTTGGAGACATGCTGTTTTGTCTATGCAGTTCCGCAATCCTGGTCT
TATTATTGACATTGTACATGATATAAGGCTCATAAATGGTGAAGCCATTTCATGCAAGCCCAAGAAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTCCGCAATGGTGCAGATTATGCTGTCTATATC
AACACAGCTCAAGAGTTTCGATGGAAGTGACTCAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAGGGTTC
AGCAAAACCTGTGAAGGTTCCATTGTGATGCGACTATCGCTTTCCATTACTCGTTGCCGCAACATTTGCACGTAGGTTTC
ATGGCACGAAT

>Holcus lanatus dhs1
CGAGGTGAATTTTCTTTACCTGGGGCACTGCTCCGGTCAAAGGACTCAACCGGATAGGAAATCTGTTGGTACCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCCTCTCTTTGACCAGATGCTACAAGAACAACTTAGTGAGAATGCTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTATTGGGCATACAAGAACAAAT
ATTCTGTATACAGCCAGCATTAACTGATGGATCACTCGGCGATATGCTGTTCTGTCTATGCAGTCCGCAACCCCTGGTCT
TATTATTGACATTGTACACGATATAAGGTTGATAAATGGTGAAGCCATTTCATGCAACCCCAAGAAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTCCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAATTTGATGGAAGTGATTGAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAAGGTTTC
AGCAAAACCTGTAAAGGTTCCATTGTGATGCAACTATCGCTTTCCATTACTGTTGCCGCAACATTTGCACGTAGGATTTC
ATGGCACAAAT

>Holcus mollis dhs1
CGAGGTGAATTTTCTTTACCTGGGGCACTGCTCCGGTCAAAGGACTCAACCGGATAGGAAATCTGTTGGTACCCAATGA
TAACTATTGCAAGTTTGAAAACCTGGATCATGCCCTCTCTTTGACCAGATGCTACAAGAACAACTTAGTGAGAATGCTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGCAAAGAAATAAATGATGAAAAGCTCCTACCTTTATTGGGCATACAAGAACAAAT
ATTCTGTATACAGCCAGCATTAACTGATGGATCACTCGGCGATATGTTGTTCTGTCTATGCAGTCCGCAACCCCTGGTCT
TATTATTGACATTGTACACGATATAAGGTTGATAAATGGTGAAGCCATTTCATGCAACCCCAAGAAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTCCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAATTTGATGGAAGTGATTGAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAAGGTTTC
AGCAAAACCTGTAAAGGTTCCATTGTGATGCAACTATCGCTTTCCATTACTGTTGCCGCAACATTTGCACGTAGGATTTC
ATGGCACAAAT

>Brachypodium distachyon
AGAGGTGATTTTCTTTACCTGGAACTACTGCGGTCAAAGGACTGAACCGGATAGGAAATCTGCTGGTACCCAATGA
TAACTACTGCAAGTTTGAAAACCTGGATCATGCCAATCTTTGACCAGATGCTACTAGAACAACTTAGTGAGAATGTTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGGAAAGAAATAAATGACGAAAATCTCCTACCTTTATTGGGCATACAAGAACAAAT
ATTCTGTATACAGCCAGCATTGACTGATGGATCACTTGGAGACATGCTTTTTCTGTCTATGCAGTCCGCAATCCTGGTCT
TATTATTGACATTGTACACGATATAAGGTTGATAAATGGTGAAGCCATTTCATGCAACCCCAAGAAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAAATATGTTCCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAGTTTCGATGGAAGTGATTGAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAAGGTTTC
AGCAAAACCTGTAAAGGTTCCATTGCGATGCGACTATTGCTTTCCCGTACTTGTGGCGCAACATTTGCACGTAGGTTTC
ATGGTGCGAAT

>Dactylis glomerata dhs2
CGAGGTGATTTTCTTTACCTGGAACTACTGCGGTCAAAGGACTGAACCGTATAGGAAATCTGTTGCTGACCCAACGA
CAACTACTGCAGATTTGAAAATGGATCATGCCAATCTTTGACCAGATGCTACTAGAGCAATCCAGTAAGAATGTTTGGGA
CACCATCAAAGGTGATTGCTCGTCTTGGGAAAGAAATAAATGATGAAAAGCTCCTACCTTTATTGGGCATACAAGAACAAAT
ATTCTGTATACAGCCAGCATTGACTGATGGATCACTTGGAGACATGCTTTTTCTGTCTATGCAGTCCGCAATCCTGGTCT
TATTATTGATATTGTACAAGATATACCGCTGATTAATGGGGAAAGCCATTTCATGCAACCCCAAGGAAAGACAGGGGTCATAA
TTCTTGGTGGAGGCTTCCGAAAGCATCATATATGCAATGCAAAATATGTTCCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAGTTTCGATGGAAGTGATTGAGGAGCACAGCCAGATGAAGCAGTTTCATGGGAAAGATCAAAGGTTTC
AGCAAAACCTGTAAAGGTTCCATTGCGATGCGACTATTGCTTTCCCGTACTTGTGGCGCAACATTTGCACGTAGGTTTC
ATGGTGCGAAT

```

>Festuca arundinacea dhs2
CGAGGTGATTTTTCTCTACCTGGAAAAATGCTGCGCTCAAAGGACTGAACCGGATAGGAAATCTGTTGATGCCAACGA
TAACTATTGCAAGTTTGAAAAATGGATCATGCCACTCCTTGACCAGATGCTACTGGAACAATCTACTAAGAATGTTTGGGA
CACCATCAAAGGTGATTGCTCGCTCGGCAAAGAAATTAATGATGAAAAGCTCCTTCTTACTGGGCACAAAAGAACAAAT
ATCCCTGTATACTGCCAGCATTGACTGATGGATCAATTGGAGACATGTTGTTTTGTCTATGCAGTTCACAATCCTGGTCT
TATTATTGACATTTGTGCAAGATGTAAGGCTGATGGATTGTGAAGCAATTAATGCAACCCCAAGGAAGACAGGGGCTATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAATATGCTCCGCAATGGTACAGATTATGCAGTCTACATC
AACACAGCTCAAGAGTTTGTAGTATTGATGAGGACATCCGGATGAAGCAGTTTCATGGGGAAAGATCAATGGTTC
GGCAAAACCTGTAAAGTCCACTGTGATGCAACCATTGCTTTCCCGTTACTTGTGCGCCGAACATTTGCCCGTAAGTTTC
ATGGCCTGGGC
>Lolium perenne dhs2b
CGAGGTGATTTTTCTCTATCTGGAACAATGCTGCGCTCAAAGGGCTGAACCGGATAGGAAATCTGCTGATGCCAACAA
TAACTATTGCAAGTTTGAAAAATGGATCATGCCACTCCTTGACAAGATGCTACTGGAACAATCTACTAAGAATGTTTGGGA
CACCATCAAAGGTGATTGCTCGCTCGGCAAAGAAATAAATGATGAAAAGCTCCTTCTTACTGGGCACAGAAGAACAAAT
ATCCCTGTATTCTGCCAGCATTGACTGATGGATCAATTGGAGACATGTTGTTTTGCCATGCAGTTAAAAATCCTGGTCT
TATTATTGACATTTGTGCAAGATGTAAGGATGATGGATTGCCAAGCAATTAATGCAACCCCAAGGAAGACAGGGGCTATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAATATGCTCCGCAATGGTACAGATTATGCAGTCTACATC
AACACAGCTCAAGAGTATGATGTTAGTATTGATGAGGACATCCGGATGAAGCAGTTTCATGGGGAAAGATCAATGGTTC
GGCAAAACCTGTAAAGTCCACTGCGATGCAACCATTGCTTTCCCGTTAGTTGTGGCTGCAACATTTGCCCGTAAGTTTC
ATGGCCTGGAA
>Holcus mollis dhs2
CGAGGTGATTTTTCTCTACCTGGAAACAATGCTGCGGTCAAAGGACTGAACCGGATAGGAAATCTGTTGATGCCAACGA
TAACTATTGCAAGTTTGAAAAATGGATCATGCCACTCCTTGACCAGATGCTACTAGAGCAATCTACTAAGAATGTTTGGGA
CACCATCAAAGGTGATTGCTCGCTCGGCAAGAAATAAATGATGAAAAGCTCCTTACTGGGCACAAAAGAACAAAT
ATTCCAGTATTCTGCCAGCATTGACTGATGGATCAATTGGAGACATGTTTATTTGCCATGCAGTTCACAATCCTGGTCT
TATTATTGACATTTGTGCAAGATGTAAGGCTGATGAATGGTGAAGCAATTAATGCAACCCCAAGGAAGACAGGGGCTATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAATATGCTCCGCAATGGTGCAGATTATGCAGTCTATATC
AACACAGCTCAAGAGTTTGTAGTGGCAGTATTCTGAGGACATCCGGATGAAGCGGTTTCATGGGGAAAGATCAATGGTTC
AGCAAAACCTGTAAAGGTCCACTGCGATGCAACCATTGCTTTCCCGTACTTGTGCTGCAACATTTGTCGGTAGGTTTC
ATGGCGTGAGT
>Lolium perenne dhs2a
CGAGGTGATTTTTCTCTACCCGGAACAATGCTGCGCTCAAAGGACTGAACCGGATAGGAAATCTGTTGATGCCAACGA
TAACTATTGCAAGTTTGAAAAATGGATCATGCCACTCCTTGATCAGATGCACCTGGAACAATCTACCAAGAATGTTTGGGA
CACCATCAAAGGTGATTGCTCGCTCGGCAAAGAAATAAATGACGAAAAGCTCCTTACTGGGCACACAAGAACAAAT
ATTCCTGTATACTGCCAGCATTGACTGATGGATCAATCGGAGATATGTTGTTTTGCCATGCAGTTCGCAATCCTGGTCT
TATTATTGACATTTGTGCAAGATGTAAGGCTGATGAATGGTGAAGCAATTAATGCAACCCCAAGGAAGACAGGGGCTATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAATATGCTCCGCAATGGTGCAGATTATGCAGTCTACATC
AACACAGCTCAAGAGTTTGTAGTGGCAGTACTCAGGAGACATCCGGATGAAGCTGTTTCATGGGGAAAGATCAATGGTTC
AGCAAAACCTGTAAAGGTCCATTGTGATGCGACGATCGCTTTCCCTTACTCGTTGCCGCAACATTTGCACGTAGGTTTC
ATGGCGCAAAAT
>Lolium temulentum dhs2
CGAGGTGATTTTTCTCTACCCGGAACAATGCTGCGCTCAAAGGACTGAACCGGATAGGAAATCTGTTGATGCCAACGA
TAACTATTGCAAGTTTGAAAAATGGATCATGCCACTCCTTGATCAGATGCACCTGGAACAATCTACCAAGAATGTTTGGGA
CACCATCAAAGGTGATTGCTCGCTCGGCAAAGAAATAAATGACGAAAAGCTCCTTACTGGGCACACAAGAACAAAT
ATTCCTGTATACTGCCAGCATTGACTGATGGATCAATCGGAGATATGTTGTTTTGCCATGCAGTTCGCAATCCTGGTCT
TATTATTGACATTTGTGCAAGATGTAAGGCTGATGAATGGTGAAGCAATTAATGCAACCCCAAGGAAGACAGGGGCTATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAATATGCTCCGTAATGGTGCAGATTATGCAGTCTACATC
AACACAGCTCAAGAGTTTGTAGTGGCAGTACTCAGGAGACATCCGGATGAAGCTGTTTCATGGGGAAAGATCAATGGTTC
AGCAAAACCTGTAAAGGTTCATTGTGATGCGACGATCGCTTTCCCTTACTCGTTGCCGCAACATTTGCACGTAGGTTTC
ATGGCGCAAAAT
>Secale cereale dhs2
CGAGGTGATTTTTCTCTACCTGGAGCATTGTTGCGGTCCAAAGGACTGAACCGGATGGTAATCTCTTGCTGCCAATGA
TAACTATTGCAAGTTTGAGCAATGGATTATGCCAGTCTTGACCAGATGCTACTAGAACAAATCTACTAAGAATGTTCTGGA
CACCATCAAAGGTGATTGCTCGCTTGGTAAAGAAATAAATGATGAAAAGCTCCTTACTGGGCACACAAGAACAAAT
ATTCCTGTATACTGCCAGCATTGACTGATGGATCAATGGAGACATGTTGTTTGTCTATGCCGTTTACAATCCTGGTCT
TATTATCGACATTTGTGCAAGATGTAAGGCTCATGAATGATGAAGCTATTTCATGCAACCCCAAGGAAGACAGGGGCTATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAATATGCTCCGCAATGGTGCAGATTATGCAGTCTACATC
AACACAGCTCAAGAGTTTGTAGTATTGATGAGGACATCCAGATGAAGCAGTTTCATGGGGAAAGATTAAGGGTTC
GGCAAAACCTGTAAAGGTTCATTGCGATGCGACTATCGCTTTCCCGTTACTTGTGCTGCAACATTTGCACGTAAGTTTC
ATGGTGGCAAT
>Triticum aestivum dhs2 chromosome 5B
CGAGGTGATTTTTCTTTACGAGGAGCATTGTTGCGGTCCAAAGGACTAAACCGGATGGAAATCTCTTGCTGCCAATGA
TAACTATTGCAAGTTTCGAGAAATGGATCATGCCAGTCTTGACCAGATGCTACTAGAACAAATCTACTAAGAATGTTCTGGA
CACCATCAAAGGTGATTGCTCGCTTGGTAAAGAAATAAATGATGAAAAGCTCCTTACTGGGCACACAAGAACAAAT
ATTCCTGTATACTGCCAGCATTGACTGATGGATCAATGGAGACATGTTGTTTTGCCATGCCGTTTACAATCCTGGTCT
TATTATCGACATTTGTGCAAGATGTAAGGCTCATGAATGATGAAGCTATTTCATGCAACCCCAAGGAAGACAGGGGCTATAA
TTCTTGGTGGAGGCTTCCAAAGCATCATATATGCAATGCAAATATGCTCCGCAATGGTGCAGATTATGCCGCTATATC
AACACAGCTCAAGAGTTTGTAGTATTGATGAGGACATCCAGATGAAGCAGTTTCATGGGGAAAGATTAAGGGTTC
AGCAAAACCTGTAAAGGTTCATTGCGATGCGACTATCGCTTTCCCGTTACTTGTGCTGCAACATTTGCACGTAAGTTTC
ATGGTGTGAAT
>Triticum aestivum dhs2 chromosome 5D
CGAGGTGACTTTTTCTTTACCTGGAGCATTGTTGCGGTCAAAGGACTGAACCGGATGGAAATCTCTTGCTGCCAATGA
TAACTATTGCAAGTTTCGAGAAATGGATCATGCCAGTCTTGACCAGATGCTACTAGAACAAATCTACTAAGAATGTTATGGA
CACCATCAAAGGTGATTGCTCGCTTGGTAAAGAAATAAATGATGAAAAGCTCCTTACTGGGCACACAAGAACAAAT
ATTCCTGTATACTGCCAGCATTGACTGATGGATCAATGGGAGACATGTTGTTTTGCCATGCCGTTTACAATCCTGGTCT
TATTATCGACATTTGTGCAAGATGTAAGGCTCATGAATGATGAAGCTATTTCATGCTACCCCAAGGAAGACAGGGGTTATAA
TTCTTGGAGGAGGCTTCCAAAGCATCATATATGCAATGCAAATATGCTCCGCAATGGTGCAGATTATGCCGCTATATC
AACACAGCTCAAGAGTTTGTAGTATTGATGAGGACATCCAGATGAAGCAGTTTCATGGGGAAAGATTAAGGGTTC
AGCAAAACCTGTAAAGGTTCATTGCGATGCGACTATCGCTTTCCCGTTACTTGTGCTGCAACATTTGCACGTAAGTTTC
ATGGTGGCAAT

```

Recombination analysis of PACMAD full-length sequences

GARD output on datamonkey website (Pond et al., 2006)

Recombination report

Number of breakpoints	AICc	Δ AIC	Segments
0	13528		1 (total length of 1080 bp)
1	13467.3	60.6843	2, with breakpoint at 225 bp

KH testing report

Location of breakpoint	LHS p-value	RHS p-value	Significance
225	0.00020	0.06300	* (p-value = 0.1)

KH-insignificant breakpoints arise most frequently due to variation in branch lengths between segments; this could be due to some forms of recombination or to other processes, such as spatial rate variation, heterotachy, etc.

Conclusion

Evidence of **1** breakpoint(s) was found using AICc; an improvement of 60.6843 points was achieved over the model without recombination. Of those, **0** breakpoint(s) also indicated significant topological incongruence using the KH test. Some of the breakpoints may be due to rate variation or heterotachy and not recombination.

GARD recombination analysis was run using the HKY85 model.

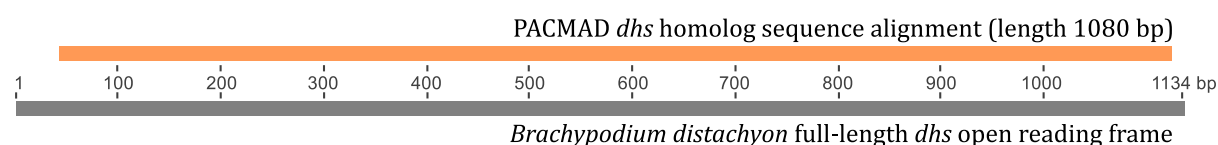
KH = Kishino Hasegawa

LHS p-value = Adjusted KH p-value that the partition to the left of the of this BP has a topology different from the one inferred from the partition on the right

RHS p-value = Adjusted KH p-value that the partition to the right of the of this BP has a topology different from the one inferred from the partition on the left

Pond, S.L.K., Posada, D., Gravenor, M.B., Woelk, C.H., and Frost, S.D.W. (2006). *Bioinformatics* 22, 3096–3098.

Final “PACMAD” alignment length, *B. distachyon dhs* ORF as reference:



“PACMAD” sequence alignment used for selection analyses (in FASTA format)

```

>Bothriochloa ischaemum dhs1
GAGGCGCTGGAGGGCGTGCCTCCATCGTGCTAAAGCCGTCGGAGTCGCTCGACGAGTCGCGGTTCCACGAGGATCGCCGG
CGCAGACTTCAACGACCCGGGCTCGGGCTCGAGGGGCTGCTCGCCTCGCTTGCACACACGGGGTTCCAGGCATCCAACC
TCGGCGACACCATCGACGCTCGTCAACCAGATGTTAGATTGGAGGTTGTCGCATGAGAAGCCAGTGAAGATTGTGATGAA
GCTGAACCTTGACCTAAATATAGAGAATCTGTGAAGTCAAGATATTTCTGGGCTTCACTTCAAACCTCGTGTCTTCTGG
CATCCGGGATATAGTCCGTTTTCTGGCTCAGCATCACATGGTGGATGTTATTGTTACAACCTGCTGGGGGTATAGAGGAGG
ATCTCATTAAGTGCCTTGCACCAACTTACAGAGGTGATTTTTATTACCTGGAGCACTACTGCGGGCTAAAGGACTGAAC
CGGATAGGAAATCTGTTGGTCCGAATGATACTACTGCAAGTTTGGAGAACTGGATCATGCCACTTTTTGACCAGATGCT
ACTGGAACAATCTACTGAGAATGTTGGACACCATCAAAGGTGATAGCTCGTCTTGGTAAAGAAATAAATGATGAGAGCT
CCTATCTTTACTGGGCATACAAGAATAACATACTGTATACTGCCCTGCATTGACTGATGGATCACTTGGAGACATGCTG
TTCTGTGCATGCAGTCCGTAATCCTGGCCTTATTATTGACATTGTACAAGATATACGGCTGATGAATGGGGAAGCCATCCA
TGCAACACCAAGGAAGACAGGGATATAGTTCTTGGTGGAGGCCCTTCAAAGCATCATATATGCAATGCCAATATGTTTTC
GCAATGGTGCAGATTATGCTGTTTATATTAACACAGCTCAAGAGTTTGTAGTGGTACTCTGGAGCACAGCCTGATGAA
GCAGTCTCATGGGCAAGATAAAGGGTTCAGCTAAACCTGTAAGGTGCATTGTGATGCGAGTATCGCTTTCCCGCTTGT
TGTGGCTGCGACATTTGCACGCAAGGTTTCATGGCTCCAA
>Miscanthus sinensis dhs2
GAGGCGTTGGAGGGCGTGCCTCCATCGTGCTAAAGCCGTCGGAGTCGCTCGACGAGTCGCGGTTCCACGAGGATCGCCGG
CGCCGACTTCAACGACCCGGGCTCGGACTCGAGGGGCTGCTCGCCTCGCTTGCACACACGGGGTTCCAGGCATCCAACC
TCGGCGACGCCATCGACGTTGTCAACCAGATGTTAGATTGGAGGTTGTCGCATGAGAAGCCAGTGAAGATTGTGATGAA
GCTGAACCTTGACCTAACTATAGAGAATCTGTGAAGTCAAGATATTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
CATCCGGGATATAATCCGTTTTCTAGCTCAGCATCACATGGTGGATGTTATTGTTACAACCTGCTGGGGGTATAGAGGAGG
ATCTCATTAAGTGCCTTGCACCAACTTACAGAGGTGATTTTTATTACCTGGAGCACTACTGCGGTCAAAGGACTGAAT
CGGATAGGAAATCTGTTGGTCCGAATGATACTATTGCAAGTTTGGAGAACTGGATCATGCCACTTTTTGACCAGATGCT
ACTGGAACAATCTACTGAGAATGTTGGACACCATCAAAGGTGATAGCTCGTCTTGGTAAAGAAATAAATGATGAAAGCT
CCTATCTTTACTGGGCATACAAGAATAACATACTGTATACTGCCCGCATTGACTGATGGATCACTTGGAGACATGCTG
TTCTGTGCATGCAGTCCGTAATCCTGGCCTTATTATTGACATTGTACAAGATATACGGCTGATGAATGGGGAAGCCATCCA
TGCAACACCAAGGAAGACAGGGATATAGTTCTTGGTGGAGGCCCTTCAAAGCATCATATATGCAATGCCAATATGTTTTC
GCAATGGCGCGGATATGCTGTTTATATTAACACAGCTCAAGAGTTTGTAGTGGTACTCTGGAGCACAGCCTGATGAA
GCAGTCTCATGGGCAAGATAAAGGGTTCAGCTAAACCTGTAAGGTGCATTGTGATGCGAGTATCGCTTTCCCGCTTGT
TGTGGCTGCGACATTTGCACGCAAGGTTTCATGGCTCCAA
>Saccharum officinale dhs2
GAGGCGCTAGAGGGCGTGCCTCCATCGTGCTAAAGCCGTCGGAGTCGCTCGAAGAGTCGCGGTTCCACGAGGATCGCCGG
CGCCGACTTCAACGACCCGGGCTCGGGCTCGAGGGGCTGCTGGCCTCGCTTGCACACACGGGGTTCCAGGCATCCAACC
TCGGCGACGCCATCGACGTTGTCAACCAGATGTTAGATTGGAGGTTGTCGCATGAGAAGCCAGTGAAGATTGTGATGAA
GCTGAACCTTGACCCAAAATATAGAGAATCTGTGAAGTCAAGATATTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
CATCCGGGATATAATCCGTTTTCTAGCTCAGCATCACATGGTGGATGTTATTGTTACAACCTGCTGGGGGTATAGAGGAGG
ATCTCATTAAGTGCCTTGCACCAACTTACAGAGGTGATTTTTATTACCTGGAGCACTACTGCGGTCAAAGGACTGAAT
CGGATAGGAAATCTTTGGTCCGAATGGTAACTATTGCAAGTTTGGAGAACTGGATCATGCCACTTTTTGACCAGATGCT
ACTGGAACAATCTACTGAGAATGTTGGACACCATCAAAGGTGATAGCTCGTCTTGGTAAAGAAATAAATGATGAAAGCT
CCTATCTTTACTGGGCATACAAGAATAACATACTGTATACTGCCCGCATTGACTGATGGATCACTTGGAGACATGCTG
TTCTGTGCATGCAGTCCGTAATCCTGGCCTTATTATTGACATTGTACAAGATATACGGCTGATGAATGGGGAAGCCATCCA
TGCAACACCAAGGAAGACAGGGATATAGTTCTTGGTGGAGGCCCTTCAAAGCATCATATATGCAATGCCAATATGTTTTC
GCAATGGCGCGGATATGCTGTTTATATTAACACAGCTCAAGAGTTTGTAGTGGTACTCTGGAGCACAGCCTGATGAA
GCAGTCTCATGGGCAAGATAAAGGGTTCAGCTAAACCTGTAAGGTGCATTGTGATGCGAGTATCGCTTTCCCGCTTGT
TGTGGCTGCGACATTTGCACGTAAGGTTTCACGGCTCCAA
>Sorghum bicolor dhs2
GAGGCGCTGGATGGCGTGCCTCCATCGTGCTAAAGCCGTCGGAGTCGCTCGATGAGTCGCGGTTCCACGAGGATCGCCGG
CGCCGACTTCAACGACCCGGGCTCGGGCTCGAGGGGCTGCTCGCCTCGCTTGCACACACGGGGTTCCAGGCATCCAACC
TCGGCGACGCCATCGACGTTGTCAACCAGATGTTAGATTGGAGGTTGTCGCATGAGAAGCCAGTGAAGATTGTGATGAA
GCTGAACCTTGACCCAAAATATAGAGAATCTGTGAAGTCAAGATATTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
CATCCGGGATATAATCCGTTTTCTAGCTCAGCATCACATGGTGGATGTTATTGTTACAACCTGCTGGGGGTATAGAGGAGG
ATCTCATTAAGTGCCTTGCACCAACTTACAGAGGTGATTTTTATTACCTGGAGCACTACTGCGGTCAAAGGACTGAAC
CGGATAGGAAATCTGTTGGTCCGAATGATACTATTGCAAGTTTGGAGAACTGGATCATGCCACTTTTTGACCAGATGCT
ATTCGAACAATCTACTGAGAATGTTGGACACCGTCAAAGGTGATAGCTCGTCTTGGTAAAGAAATAAATGATGAAAGCT
CCTATCTTTACTGGGCATACAAGAATAACATACTGTATACTGCCCTGCATTGACTGATGGATCACTCGGAGACATGCTG
TTCTGTGCATGCAGTCCGTAATCCTGGCCTTATTATTGACATTGTACAAGATATACGGCTGATGAATGGGGAAGCCATCCA
TGCAACACCAAGGAAGACAGGGATATAGTTCTTGGTGGAGGCCCTTCAAAGCATCATATATGCAATGCCAATATGTTTTC
GCAATGGTGCAGATTATGCTGTTTATATTAACACAGCTCAAGAGTTTGTAGTGGTACTCTGGAGCACAGCCTGATGAA
GCAGTCTCATGGGCAAGATAAAGGGTTCAGCTAAACCTGTAAGGTGCATTGTGATGCGAGTATCGCTTTCCCGCTTGT
TGTGGCTGCGACATTTGCACGCAAGGTTTCATGGCTCCAA
>Miscanthus sinensis dhs1
GAGGCGCTTGGAGGGCGTGCCTCCATCGTGCTAAAGCCGTCGGAGTCGCTCGACGAGTCGCGGTTCCACGAGGATTGCGGG
CGCCGACTTCAACGACCCGGGCTCGGACTCGAGGGGCTGCTCGCCTCGCTTGCAGACACGGGGTTCCAGGCATCCAACC
TCGGTGACGCCATCGACGTTGTCAACCAGATGTTAGATTGGAGGTTGTCGCATGAGAAGCCACTGAGGATTGTGACGAA
GCTGAACCTTGACCTAAATATAGAGAATCTGTGAAGTCAAGATATTTCTGGGCTTCACTTCAAACCTCGTGTCTTCTGG
CATCCGGGATATAATCCGTTTTCTAGCTCAGCATCACATGGTGGATGTTATTGTTACAACCTGCTGGGGGTATAGAGGAGG
ATCTCATTAAGTGCCTTGCACCAACTTACAGAGGTGATTTTTATTACCTGGAGCACTACTGCGGTCAAAGGACTGAAC
CGGATAGGAAATCTGTTGGTCCGAATGATACTATTGCAAGTTTGGAGAACTGGATCATGCCACTTTTTGACCAGATGCT
ACTGGAACAATCTACTGAGAATGTTGGACACCGTCAAAGGTGATAGCTCGTCTTGGTAAAGAAATAAAGGACTGAAC
CCTATCTTTATGGGCATACAAGAATAACATCTGTATACTGCCAGCGTTGACTGATGGATCACTTGGAGACATGCTG
TTCTGTGCATGCAGTCCGTAATCCTGGCCTTATTATTGACATTGTACAAGATATACGGCTGATGAATGGGGAAGCCATCCA
TGCAACACCAAGGAAGACAGGGATATAGTTCTTGGTGGAGGCCCTTCAAAGCATCATATATGCAATGCCAATATGTTTTC
GCAATGGTGCAGATTATGCTGTTTATATTAACACAGCTCAAGAGTTTGTAGTGGTACTCTGGAGCACAGCCTGATGAA
GCAGTCTCATGGGCAAGATAAAGGGTTCAGCTAAACCTGTAAGGTGCATTGTGATGCGAGTATCGCTTTCCCGCTTGT
TGTGGCTGCGACATTTGCACGTAAGGTTTCACAGCTCCAA

```

>Saccharum officinale dhs1
GAGGCACCTTGAGGGCGTGCCTCCATCGTGCTAAAGCCGTCGGAGTCGCTCGACGAGTCGCGGTTCCGCGAGGATCGCCGG
CGCCGACTTCAACGACCCGGGCTCGGACTCGAGGGGCTGCTCGCCTCGCTTGCAGGACCGGGGTTCCAGGCATCCAACC
TCGGCGACGCCATCGACGTCGTAACCCAGATGTTAGATTGGAGGTTGCTCATGAGAAGGCCAGTGAGGATTTGTGATGAA
GCTGAACCTTGACCTAAATATAGAGAATCTGTGAAGTGCAAGATATTTCTGGGCTTCACTTCAAATCTCGTGTCTTCTGG
CATCCGGGATATAATCCGTTTTCTAGCTCAGCATCACATGGTGGATGTTATTGTTACAAGTGTGGCGGTATAGAGGAGG
ATCTCATTAATGCCTTGACCAACCTACAGAGTGATTTTTCTTTACCTGGAGCACTGCTGCGGTCAAAGGACTGAAC
CGGATAGGAAATCTGTTGGTGCCAAATGATAACTACTGCAAGTTGAGAACTGGATCATGCCACTCTTTGACCAGATGCT
ACTGGAGCAATCTACTGAGAATGTTGGACACCGTCAAAGGTGATAGCTCGTCTTGGCAAAGAAATAGATGACAGCT
CTATCTTTTATTGGGCGTACAAGAATAACATTCCTGTATACTGCCAGCTTGGACTGATGGATCACTTGGAGACATGCTG
TTCTGTCTATGCAAGTTCGTAATCTTGGCCTTATTATTGACATTGTGCAAGATATACGACTGATGAATGGGAAGCCATCCA
TGCAACACCAAGGAAGACAGGGATATAGTTCCTTGGTGGAGGCTTCCAAGCATCATATATGCAATGCCAATATGTTCC
GCAATGGTGCAGATTATGCTGTCTATATTAACACGGCCAGAGTTTGTATGGTAGCGATCTTGGAGCACACCCTGATGAA
GCAGTCTCATGGGCAAGATCAAGGGTTGAGCAAAACCTGTAAGGTGCACTGCGATGCAAGCATCGCTTTCCCGCTTCT
TGTGCTGCGACATTTGCACGTAAGGTTACAGCTCCAAA

>Sorghum bicolor dhs1
GAGGCGCTGGACGGCGTGCCTCCATCGTGCTAAAGCCGTCGGAGTCGCTCGACGAGTCGCGGTTACCAGGATCGCCGG
CGCCGACTTCAACGACCCGGGCTCGGACTCGAGGGGCTGCTCGCCTCGCTTGCAGGACCGGGGTTCCAGGCATCCAACC
TCGGTGACGCCATCGACGTCGTAACCCAGATGTTAGATTGGAGGTTATCTCATGAGAAGGCCAGTGAGGATTTGTGATGAC
GCTGAACCTTGACCTAAATATAGAGAATCTGTGAAATGCAAGATATTTCTGGGCTTCACTTCAAACCTCGTGTCTTCTGG
CATCCGGGATATAATCCGTTTTCTAGCTCAGCATCACATGGTGGATGTTATTGTTACAACCTGCTGGGGTATAGAGGAGG
ATCTCATTAATGCCTTGACCAACTTACAGAGTGATTTTTCTTTACCTGGAGCACTGCTGCGGTCAAGAGGACTGAAC
CGGATAGGAAATCTGTTGGTGCCCAACGATAAATATTGCAAGTTTGGAGAACTGGATCATGCCACTCTTTGACCAGATGCT
TCTGGAGCAATCTACTGAGAATGTTGGACACCGTCAAAGGTGATAGCTCGTCTTGGCAAAGAAATAAATGATGAAAGCT
CTATCTTTTATTGGGCATACAAGAATAACATTCCTGTATACTGCCAGCTTGGACTGATGGATCACTTGGAGACATGCTG
TTCTGTCTATGCAAGTTCGTAATCTTGGCCTTATTATTGACATTGTGCAAGATATACGACTGATGAATGGGAAGCCATCCA
TGCAACACCAAGGAAGACAGGGATATAGTTCCTTGGTGGAGGCTTCCAAGCATCATATATGCAATGCCAATATGTTCC
GCAATGGTGCAGATTATGCTGTCTATATTAACACGGCCAGAGTTTGTATGGTAGCGATCTTGGAGCACACCCTGATGAA
GCAGTCTCATGGGCAAGATCAAGGGTTGAGCAAAACCTGTAAGGTGCACTGCGATGCAAGCATCGCTTTCCCGCTTCT
TGTGCTGCGACATTTGCACGTAAGGTTACAGCTCCAAA

>Spodiopogon sibiricus dhs1
GACGCGCTGGAGGGCGTGCCTCCATCGTGCTAAAGCCGTCGGAAATCGCTCGACGAGTCGCGGTTCTCGAGGATCGCCGG
CGCCGACTTCAACGACCCGGGCTCGGACTCGAGGGGCTGCTAGCCTCGCTTGCAGGACCGGGGTTCCAGGCATCTAACC
TCGGTGACGCCATCGACGTCGTAACCCAGATGTTAGATTGGAGGTTGTCGCATGAGAAGGCCAGTGAGGATTTGTGATGAA
GCTGAACCTTGACCTAAATATAGAGAATCTGTGAAGTGCAAGATATTTCTGGGCTTCACTTCAAACCTCGTGTCTTCTGG
CATCCGGGATATAATCCGTTTTCTAGCTCAGCATCACATGGTGGATGTTATTGTTACAAGTGTGGGGTATTGAGGAGG
ATCTCATTAATGCCTTGACCAACTTACAGAGTGATTTTTCTTTACCTGGAGCACTGCTGCGGTCAAAGGACTGAAC
CGGATAGGAAATCTGCTGGTGCCCAATGATAAATATTGCAAGTTTGGAGAACTGGATCATGCCACTCTTTGACCAGATGCT
ACTGGAGCAATCTACTGAGAATGTTGGACACCGTCAAAGGTGATAGCTCGTCTTGGCAAAGAAATAAAGCAGCAAGCT
CTATCTTTTATTGGGCATACAAGAATAACATTCCTGTATACTGCCAGCTTGGACTGATGGATCACTTGGAGACATGCTG
TTCTGTCTATGCAAGTTCGTAATCTTGGCCTTATTATTGACATTGTGCAAGATATACGACTGATGAATGGGAAGCCATCCA
TGCAACACCAAGGAAGACAGGGATATAGTTCCTTGGTGGAGGCTTCCAAGCATCATATATGCAATGCCAATATGTTCC
GCAATGGTGCAGATTATGCTGTCTATATTAACACGGCCAGAGTTTGTATGGTAGCGATCTTGGAGCACACCCTGATGAA
GCAGTCTCATGGGCAAGATCAAGGGTTGAGCAAAACCTGTAAGGTGCACTGCGATGCAAGCATCGCTTTCCCGCTTCT
TGTGCTGCGACATTTGCACGTAAGGTTACAGCTCCAAA

>Zea mays dhs1
GAGGCGCTGGAGGGCGTGCCTCCATCGTGCTAAAGCCGTCGGAGTCGCTCGACGAGTCGCGGTTACCAGGATCGCCGG
CGCCGACTTCAACGACCCGGGCTCGGACTCGAGGGGCTGCTCGCCTCGCTTGCAGGACCGGGGTTCCAGGCATCCAACC
TCGGCGACGCCATCGACGTCGTAACCCAGATGTTAGATTGGAGGTTGTCGCATGAGAAGGCCAGTGAGGATTTGTGATGAA
GCTGAACCTTGACCTAAATATAGAGAATCTGTGAAGTGCAAGATATTTCTGGGCTTCACTTCAAACCTCGTGTCTTCTGG
CATCCGGGATATAATCCGTTTTCTAGCTCAGCATCACATGGTGGATGTTATTGTTACAAGTGTGGGGTATAGAGGAGG
ATCTCATTAATGCCTTGACCAACTTACAGAGGCGATTTTTCTTTACCTGGAGCACTGCTGCGGTCAAAGGACTGAAC
CGGATAGGAAATCTGTTGGTGCCCAATGACAACCTATTGCAAGTTTGGAGAACTGGATCATGCCACTCTTCCAGCAGATGCT
ACAGGACCAATCTACTGAGAATGTTGGACACCGTCAAAGGTGATAGCGCGCTTGGCAAAGAAATAAATGATGAAAGCT
CTATCTTTTATTGGGCATACAAGAATAACATTCCTGTATACTGCCAGCATTGGACTGATGGATCACTTGGAGACATGCTG
TTTTGTCTATGCAAGTTCGTAATCTTGGCCTTATTATTGACATTGTGCAAGATATACGCTGATGAATGGGAAGCCATCCA
TGCAACACCAAGGAAGACAGGGATATAGTTCCTTGGTGGAGGCTTCCGAAAGCATCATATATGCAATGCCAATATGTTCC
GCAATGGTGCAGATTATGCTGTCTATATTAACACGGCTCAAGAGTTTGTATGGTAGCGATCTTGGAGCACACCCTGATGAA
GCAGTCTCATGGGCAAGATTAAGGGTTGAGCAAAACCTGTAAGGTGCACTGCGATGCAAGCATCGCTTTCCCGCTTCT
TGTGCGACCAACATTTGCACGTAAGGTTCAAACTCCAAA

>Digitaria sanguinalis dhs
GAGGCGCTGGAGGGCGTGCCTCCATCGTGCTAAAGCCGTCGGAGTCGCTCGACGAGTCCCCTTTACGAGGATCGCCGG
GCTCGACTTCAACGACCCGGGCTTGGCCTCGAGGGCCTGCTCGGTCGTTTCGCTTCCAGGGGTTCCAGGCGTCCAACC
TTGGCGACGCCATCGACGTCGTAACCCAGATGTTAGATTGGAGGTTGTCGCATGATAAACCAGTGAAGATTCCGATGAT
GCTGAACCTTGACCCCAAGTACAGAGCATCCGTGAAGTGCAAGATATTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
CATCAGGACATAGTCCGATTTCTAGCTCAGCATCGCATGGTGGATGTTATCGTTACAACCTGCTGGGGTATAGAGGAGG
ATCTCATAAATGCCTTGACCAACTTACAGAGTGATTTTTCTTTACCTGGAACATTTGCTGCGGTCAAAGGTCTGAAC
CGAATAGGAAATCTGTTGGTGCCCAATGATAACTACTGCAAGTTTGGAGAACTGGATCATGCCACTATTGACCAGATGCT
ACTAGAACAATCGACCGAGAATGTTGGACACCGTCAAAGGTGATAGCTCGTCTTGGCAAAGAAATAAATGACGAAAGCT
CTATCTTTTATTGGGCATACAAGAACAACATTCCTGTATACTGTCCAGCATTGACTGATGGATCACTTGGAGACATGCTT
TTCTGTCTATGCAAGTTCGTAATCTTGGCCTTATTATTGACATTGTGCAAGATATACGCTGATGAATGGGAAGCCATCCA
TGCAACACCAAGGAAGACAGGGATATAGTTCCTTGGTGGAGGCTTCCGAAAGCATCATATATGCAATGCCAATATGTTCC
GCAATGGTGCAGATTATGCTGTCTATATTAACACGGCTCAAGAGTTTGTATGGTAGCGATCTTGGAGCACACCCTGATGAA
GCAGTCTCATGGGCAAGATTAAGGGTTGAGCAAAACCTGTAAGGTGCACTGCGATGCAAGCATCGCTTTCCCGCTTCT
TGTGCTGCAACATTTGCATGTAAGGTTACGGCTCCAAA

```

>Panicum hallii dhs
GAGGCGCTGGACGGCGTGCCTCCATCGTCTGAAGCCCTCGGAGTCGATCGACGAGTCCCCTTCACGAGGATCGCCGG
CGCCGACTTCAACGACCCCGGGCTCGGGCTCGAGGGCTGCTCGGGTTCGCTCGCGTCCACGGGGTTCAGGGCGTCCAACC
TCGGAGACGCCATCGACGCTCGTCAACCAGATGTTAGATTGGAGGTTGTTGCATGAGAAGCCTGGTGAAGATTGCCGATGAA
GCTGAACCTTGACCCCAATACAGAGAATCTGTGAAGTGAAGATATTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
CATCCGGGATATAATCCGATTTCTAGCTCAGCATCGCATGGTGGATGTTATTGTTACAACCTGCTGGGGTATAGAGGAGG
ATCTCATTAAGTCTTGCACCAACTTATAGAGGTGATTTTCTTTACCTGGAGCACTACTGCGGTCAAAGGATTGAAC
CGGATAGGAAATCTGTTGGTGCCCAATGATAACTATTGCAAGTTTGAAGACTGGATCATGCCACTATTTGACCAGATGCT
ACTAGAACAATCTACCGAGAATGTTGGACACCGTCAAAGGTGATAGCTCGTCTTGGCAAAGAAATAAATGACGAAAGCT
CCTATCTTTATTGGGCATACAAGAATAACATTCTGTATACTGCCAGCATTGACTGATGGATCACTTGGAGACATGCTG
TTCTGTACGCGAGTCCGCAATCTGGGCTTATTATTGACATTGTACAAGATATACGGTTGATGAATGGGGAAGCCATCCA
TGCAACGCCAAGGAAGACAGGGATAAATAGTTCTTGGTGGAGGCCCTTCAAAGCATCATATATGCAATGCTAATATGTTCC
GCAATGGAGCAGACTATGCAGTCTATATTAACACAGCCCAAGAGTTTGTATGGTAGTGATTCAGGAGCACAACTGATGAA
GCAGTTTTCATGGGAAAGATCAAGGGTTACAGGAAACCTGTAAAGGTGCATTGTGATGCGAGTATTCGCTTCCCGCTTCT
GTTTGCAGCAATTTGCACGTAAGGATCACGGATCCAAA

>Panicum virgatum dhs2
GAGGCGCTGGACGGCGTGCCTCCATCGTCTGAAGCCCTCGGAGTCGCTCGACGAGTCCCCTTCACCAGGATCGCCGG
CGCCGACTTCAACGACCCCGGGCTCGGGCTCGAGGGCTGCTCGGGTTCGCTCGCGTCCACGGGGTTCAGGGCGTCCAACC
TCGGCGACGCCATCGACGCTCGTCAACCAGATGTTAGATTGGAGGTTGTCATGAGAAGCCTGGTGAAGATTGTGATGAA
GCTGAACCTTGACCCCAATACAGAGAATCTGTGAAGTGAAGATATTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
CATTCGGGATATAATCCGATTTCTAGCTCAGCATCGCATGGTAGATGTTATTGTTACAACCTGCTGGGGTATAGAGGAGG
ATCTCATTAAGTCTTGCACCAACTTATAGAGGTGATTTTCTTTACCTGGAGCACTACTGCGGTCAAAGGATTGAAC
CGGATAGGAAATCTGTTGGTGCCCAATGATAACTATTGCAAGTTTGAAGACTGGATCATGCCACTATTTGACCAGATGCT
ACTAGAACAATCTACCGAGAATGTTGGACACCGTCAAAGGTGATAGCTCGTCTTGGCAAAGAAATAAAGCAGGAAAGCT
CCTATCTTTATTGGGCATACAAGAATAACATTCTGTGACTGCCATCATTGACTGACGGATCACTTGGAGACATGCTG
TTCTGTACGCGAGTCCGCAATCTGGCCTTATTATTGACATGATACAAGATATACGGTTGATGAATGGGGAAGCCATCCA
TGCAACGCCAAGGAAGACAGGGATAAATAGTTCTTGGTGGAGGCCCTTCAAAGCATCATATATGCAATGCTAATATGTTCC
GCAATGGAGCGGACTATGCAGTCTATATTAACACAGCCCAAGAGTTTGTATGGTAGTGATTCAGGAGCACAGCCTGATGAA
GCAGTTTTCATGGGAAAGATCAAGGGTTACAGGAAACCTGTAAAGGTGCATTGTGATGCGAGTATTCGCTTCCCGCTTCT
TGTGCTGCAACATTTGCACGTAAGTTCACGGATCCAAA

> Panicum virgatum dhs3
GAGGCGCTGGACGGCGTGCCTCCATCGTCTGAAGCCCTCGGAGTCGCTCGACGAGTCCCCTTCACGAGGATCGCCGG
CGCCGACTTCAACGACCCCGGGCTCGGGCTCGACGGCTGCTCGGGTTCGCTCGCGTCCACGGGGTTCAGGGCGTCCAACC
TCGGCGACGCCATCGACGCTCGTCAACCAGATGTTAGATTGGAGGTTGTCGATGAGAAGCCTGGTGAAGATTGTGATGAA
GCTGAACCTTGACCCCAATACAGAGAATCTGTGAAGTGAAGATATTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
CATCCGGGATATAATCCGATTTCTAGCTCAGCATTCGATGGTGGATGTTATTGTTACAACCTGCTGGGGTATAGAGGAGG
ATCTCATTAAGTCTTGCACCAACTTATAGAGGTGATTTTCTTTACCTGGAGCCCTACTGCGGTCAAAGGATTGAAC
CGGATAGGAAATCTTCTGGTGCCCAATGATAACTATTGCAAGTTTGAAGACTGGATCATGCCACTATTTGACCAGATGCT
ACTAGAACAATCTACCGAGAATGTTGGACACCATCAAAGGTGATAGCTCGTCTTGGCAAAGAAATAAAGCAGGAAACT
CCTATCTTTATTGGGCATACAAGAATAACATTCTGTGACTGCCATCATTGACTGATGGATCACTTGGAGACATGCTG
TTCTGTACGCGAGTCCGCAATCTGGCCTTATTATTGACATGATACAAGATATACGGTTGATGAATGGGGAAGCCATCCA
TGCAACGCCAAGGAAGACAGGGATAAATGTTCTTGGTGGAGGCCCTTCAAAGCATCATATATGCAATGCTAATATGTTCC
GCAATGGAGCGGACTATGCAGTCTATATTAACACAGCCCAAGAGTTTGTATGGTAGTGATTCAGGAGCACAGCCTGATGAA
GCAGTTTTCATGGGAAAGATCAAGGGTTACAGGAAACCTGTAAAGGTGCATTGTGATGCGAGTATTCGCTTCCCGCTTCT
TGTGCTGCAACATTTGCACGTAAGTTCACGGATCCAAA

>Setaria italica dhs
GAGGCTCTGGAGGGCGTGCCTCCATCGTCTGAAGCCCTCGGAGTCGCTCGACGAGTCCCCTTCACGAGGATCGCCGG
CGCCGACTTCAACGACCCCGGGCTCGGGCTCGAGGGCTGCTCGGGTTCGCTCGCGTCCACGGGGTTCAGGGCGTCCAACC
TTGGCGACGCCATCGACGCTCGTCAACCAGATGTTAGATTGGAGGTTGTCGATGAGAAGCCTGGCGAAGATTGTGATGAA
GCTGAACCTTGACCCCAATACAGAGAATCTGTGAAGTGAAGATATTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
CATCCGGGATATAATCCGATTTCTAGCTCAGCATTCGATGGTGGATGTTATTGTTACAACCTGCTGGGGTATAGAGGAGG
ATCTCATTAAGTCTTGCACCAACTTACAGAGGTGATTTTCTTTACCTGGAGCACTACTGCGGTCAAAGGGTGAAC
CGGATAGGAAATCTGTTGGTGCCCAATGATAACTATTGTAAGTTTGAAGACTGGATCATGCCACTATTTGACCAGATGCT
ACTAGAACAATCTACTGAGAATATTTGGACACCGTCAAAGGTGATAGCTCGTCTTGGCAAAGAAATAAATGACGAAAGCT
CCTATCTTTATTGGGCATACAAGAATAACATTCTGTATACTGCCAGCATTGACTGATGGATCACTTGGAGACATGCTG
TTCTGTACGCGAGTCCGCAATCTGGCCTTATTATTGACATTGTACAAGATATACGGCTGATGAATGGGGAAGCCATCCA
CGCAGCATCGAGGAAGACAGGGATAAATAGTTCTTGGTGGAGGCCCTTCAAAGCATCATATATGCAATGCTAATATGTTCC
GCAATGGAGCAGACTATGCAGTCTATATCAACACAGCTCAAGAGTTTGTATGGTAGTGATTCAGGAGCACAACTGATGAA
GCAGTTTTCATGGGAAAGATCAAGGGTTACAGGAAACCTGTAAAGGTGCATTGTGATGCGAGTATTCGCTTCCCGCTTCT
TGTGCTGCAACATTTGCACGTAAGTTCACGGATCCAAA

>Molinia caerulea dhs1
GAAGTGTGGAGGGCGTGCCTCCATCGTCTGAAGCCCTCGGAGTCGCTCGACGAGTCCCCTTCACGAGGATCGCCGG
CGCCGACTTCAACGACCCCGGGCTCGGGCTCGAGGGCTGCTCGGGTTCGCTCGCGTCCACGGGGTTCAGGGCGTCCAACC
TCGGCGACGCCATCGACGCTCGTCAACCAGATGTTAGATTGGAGGTTGTCGATGAGAAGCCTGGCGAAGATTGTGATGAA
GCTGAACCTTGACCCCAATATAGAGAATCTGTGAAGTGAAGATATTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
CATCAGGGATATAAATCCGATTTCTGGCTCAGCATCACATGGTGGATGTTATTGTTACAACCTGCTGGGGTATAGAGGAGG
ATCTCATTAAGTCTTGCACCAACTTACAGAGGTGATTTTCTTTACCTGGAGTATTACTGCGGTCAAAGGACTGAAT
CGGATAGGAAATCTGTTGGTGCCCAATGATAACTATTGCAAGTTTGAAGACTGGATCATGCCCTCTTTGACCAGATGCT
GATGGAACAATCTACTGAGAATGTTGGACACCATCAAAGGTGATAGCTCGTCTTGGCAAAGAAATAAAGCAGATGAAAGCT
CCTACCTTTATTGGTCATACAAGAACAACATTCTGTATACTGCCAGCGTTGACTGATGGATCACTTGGAGACATGCTG
TTTGTACACAGTCCGCAATCTGGTCTTATTATTGACATTGTACAAGATATACGGTTGATGAATGGGGAAGCCATCCA
TGAACCCCAAGGAAGACAGGGATCATAGTTCTTGGTGGAGGCCCTTCAAAGCATCATATATGCAATGCTAATATGTTCC
GCAATGGTGCAGATTATGCAGTCTACATTAACACAGCTCAAGAGTTTGTATGGTAGTGATTCAGGAGCACAGCCTGATGAA
GCAGTTTTCATGGGAAAGATCAAGGGTTACAGGAAACCTGTAAAGGTGCATTGTGATGCGAGTATTCGCTTCCCGCTTCT
TGTGCTGCAACATTTTCATGTAAGATTACAGGCTCCAAA

```

>Molinia caerulea dhs2
GATGCGCTGGAGGGCGTGCCTCCATCGTGTGAAGCCCTTCGGAGTTCGCTCGACGAGGCGGGCTCACGAGGATCGCCGG
CGCCGACTTCAACGAAACCGAATCTCGGCCTTGATGGGCTGCTCGGGTTCGCTCGCGTCCACGGGGTTCAGGCCCTCCAGCC
TCGGGGACGCCATCGACTCGTCAAGCCAGATGTTAGATTGGAGGTTGTCGCATGAGAAGCCACACGAGGATTTGTGATGAA
GCTGAACCTGACCCATAATATAGAGAATCTGTGAAGTGCAAGATATTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
CATCCGGGATATAATCCGATTTCTAGCTCAGCATCACATGGTGGATGTTATTGTTACAACCTGCTGGGGGTATAGAGGAGG
ATCTCATTAATGCCTTGACCAACTTACAGAGTGATTTTCTTTACCTGGAGCATTATTGCGGTCAAAGGACTGAAC
CGGATAGGAAATCTGTTGGTGCCCAATGATAACTATTGCAAGTTTGAGAACTGGATCATGCCACTCTTTGACCAGATGCT
ACTGGAACAATCTACTGAGAACGTTTGGACACCATCAAAGGTGATATCTCGTCTTGGCAAAGAAATAAACGATGAAAGCT
CCTACCTTTATTGGGCATACCAGAATAACATTCCTGTATACTGCCAGCGTTGACTGATGGATCACTTGGAGACATGCTG
TTTTGTACAGCCGTACGCAATCTGGTCTTATTATTGACATTGTACAAGATATACGACTGATGAATGGGAAGCCATTCA
TGCAACCCCAAGGAAGACAGGGATCATAGTTCTTGGTGGAGGCTTCCAAGGCATCATATATGCAATGCCAATATGTTTC
GCAATGGCGCAGATTTATGCAGTCTATATTAACACAGCTCAAGAGTTTGATGGTAGTGATCTGGAGCACAGCCTGATGAA
GCAGTTTTCATGGGAAAGATCAAGGGCTCAGCAAACCTGTAAAGGTGCATTGCGATGCAAGTATTGCTTTCCCGCTTCT
TGTGGCCGCAACATTTGCTCGTAAAGGTTACGGCTCCAA
> Eleusine coracana dhs
GACGCGCTGGAGGGCGTGCCTCCATCGTGTGAAGCCGTCGGAGTTCGCTCGAAGAGTTCGCGCTTACGAGGATTGCGCGG
CGCCGACTTCAACGACGCGGCTCGGCCTCGAAGGGCTGCTCGCTTCTGTTGCGTCCACGGGGTTCAGGCCCTCCAATC
TCGGTGACGCCATAGATATCGTCAACCCAGATGTTGGATTGGAGGTTGTCGCATGAGAAGCCACGCGAGGATTTGTGATGAA
GCTGAACCTGACCCATAATATAGAGAATCTGTGAAGTGCAAGATATTTCTGGGCTTCACTTCGAACCTTGTGTCTTCTGG
CATTCCGGATACAAATCAGATTCCTGGCTCAGCATCACATGGTGGATGTTATTGTTACTACAGCTGGAGGTATAGAGGAGG
ATCTCAATAAATGCCTTGTCCAACTTACAGAGTGATTTTCTTTACCTGGAGCACTTCTGCGGTCAAAGGATTTGAAC
CGGATAGGAAATTTGTTGGTGCCAAATGATAACTATTGTAAGTTTGGAAATTTGAGAATTGGATAATGCCCATCTTTGACCAGATGCT
ACTGGAACAATCTTCTGAGAATGTTTGGACGCCATCAAAGGTGATAGCTCGACTTGGCAAGGAGATAAATGACGAAAGTT
CCTACCTTTATTGGGCATACAAGAACAATCCCTGTACTGCCAGCGCTGACTGATGGATCACTTGGAGACATGTTG
TTTTGTCTGACGTCGCAATCTGGTCTTATAATTGATATTGTACAAGATATACGATTTGATGAACGGGGAAGCCATCCA
TGCAACCCCAAGGAAGACCGGGATCATAGTTCTTGGTGGAGGCTTCCAAGGCATCATATATGCAATGCCAATATGTTTC
GCAATGGTGCAGACTACGAGTCTACATTAACACGGCTCAAGAGTTGATGGTAGTGATCTGGAGCACAGCCTGACGAA
GCAGTTTTCATGGGAAAGATCAAGGGTTCAAGAAACCTGTAAAGGTGCATTGATGCAAGTATCGCTTTCCCGCTCCT
TGTGCTGCAACATTTGCGCTAAGGTTTATGGCTCGAA
>Oropetium thomaemum dhs
GACGCTTGGAGGGCGTGCCTCCATCGTGTGAACCCGTCGGAGTTCGCTAGACGAGTTCGCGCTTCCGCGAGGATCGCCGG
CGCCGACTTCAACGACGCGGCTCGGCCTCGATGGGCTGCTCGCTCGCTCGCGTCCACGGGGTTCAGGCCCTCCAAC
TCGGCGACGCCATCGACGCTCGTCAACCCAGATGTTAGATTGGAGATTATCACATGAGAAGCCACGTGAGGATTTGTGATGAA
GCCGAGCTTGACCTAAATATAGAGAATCTGTGAAGTGCAAGATATTTCTGGGCTTCACTTCGAACCTTGTGTCTTCTGG
CATTCCGGATATAATCAAATTTCTGGCTCAGCATCACATGGTGGATGTTATAGTTACTACTGCTGGAGGTATCGAGGAGG
ATCTCATTAATGCCTTGTCCAACTTACAGAGTGATTTTCTTTACCTGGAGCACTCTGCGGTCAAAGGGACTGAAC
CGGATAGGAAATTTGTTGGTGCCGAATGATAACTATTGCAAGTTTGGAAATTTGAGAATTGGATCATGCCAATCTTTGACAAGATGTT
GCTGGAACAATCTTCTGAGAACGTTTGGACACCATCAAAGGTGATAGCTCGTCTGGGCAAGGAGATAAACGATGAAAGTT
CCTACCTTTATTGGGCATACAAGGACAATATCCCTGTACTGTCCAGCGCTGACCGATGGATCACTTGGAGACATGTTG
TTTTGTCTGACGTCGCAATCTGGTCTTATAATTGATATTGTACAAGATATACGATTTGATGAATGGGGAAGCCATCCA
CGCAACCCCAAGGAAGACAGGGATCATAGTTCTTGGTGGAGGCTTCTTAAGCATCATATATGCAATGCCAATATGTTTC
GCAACGGTGCAGATTATGCAGTATACATTAACACAGCTCAAGAAATTTGATGGTAGTGATCTGGAGCACAGCCTGACGAA
GCAGTTTTCATGGGAAAGATCAAGGGTTCAAGAAACCTGTAAAGGTGCATTGATGCAAGTATTGCTTTCCCTCTCCT
TGTGCTGCAACATTTGCGCTAAGGTTTACGGCTCAAA
>Lolium perenne dhs1
GACCCGCTGGAGGGCGTGCCTCCATCGTGTGAAGCCGTCGGAGTTCGCTGGAGGAGTTCGCGGTTACGAGGATCTCCGG
CGCCGACTTCAACGACGCGGCTCGGCCTTGACGGGCTGCTCGGCTCGCTCGCGTCCACGGGGTTCAGGCCCTCCACC
TCGGCGACGCCATCGACTCGTCAATCAGATGTTAGATTGGCGGTTGTCATGAGAAGCCAACTGAAGATTGTGATGAA
GCTGAACCTGACCCGAAATACAGAGAATCTGTCAAGTGCAAGATATTTCTGGTTCACGTCAAACCTTGTCTTCTGG
CATACCGGATGTGATCCGGTTTCTAGTTCAACATCACATGGTTGATGTCATTGTTGCGACTGCTGGGGGTATAGAGGAAG
ACCTCATAAAATGCCTCGCACCTACGATACGAGGCGAATTTTCTTTACCTGGGGCACTGCTGCGGTCAAAGGACTTAA
CGGATAGGAAATCTGTTGGTGCCCAATGATAACTATTGCAAGTTTGGAAACTGGATCATGCCAATCTTTCGACAAGATGCT
ACAAGAACAATCTAGTGAGAATGTCTGGACACCATCAAAGGTGATGCTCGTCTTGGAAAAGAAATAAATGATGAAAGCT
CCTACCTTTATTGGGCATACAAGAACAATATTCCTGTATACTGCCAGCATGACTGATGGATCACTTGGAGACATGCTG
TTTTGTCTGACGTTCCGAATCTGGTCTTATTATTGACATTGTACATGATATAAGGCTCATAAATGGTGAAGCCATTCA
TGCAAGCCCAAGAAAGACAGGGTTCATAATTCTTGGTGGAGGCTTCCAAGGCATCATATATGCAATGCCAATATGTTTC
GCAATGGTGCAGATTATGCAGTGTACATCAACACAGCTCAAGAGTTGATGGAAAGTACTCAGGAGCACAGCCAGATGAA
GCAGTTTTCATGGGAAAGATCAAGGGTTCAAGAAACCTGTAAAGGTGCATTGATGCAACTATTGCTTTCCCTCTCCT
CGTTGCGCCGCAACATTTGCGCTAGGTTTTCATGGCACAAAT
>Hordeum vulgare dhs1
GACCCGCTGGAGGGCGTCCGTCATCGTGTGAAGCCGTCGGAGTTCGCTGGAGGAGTTCGCGGTTACCCAAATCGCCGG
CGCCGACTTCAACGACGCGGCTCGGCCTCGACGGGCTTCTTGGTTCGCTCGCTTCCACGGGGTTCAGGCCCTCAACC
TCGGCGACGCCATCGACTCGTCAATCAGATGATAGATTGGAGGTTGTCGCATGAGAAGCCGAGCGAGGACTGTGATGAA
GCTGAACCTGACCCGAAATACAGAGAATCTGTCAAGTGCAAGATATTTCTTGGTTTACGTCAAACCTTGTCTTCTCCGG
TATACGGGATGTAATACGGTTTCTAGTTCAACATCGCATGGTGGATGTTATTGTTACGACTGCTGGGGGTATAGAGGAAG
ACCTCATAAAATGCCTTGACCAACGTAACGAGGTGAATTTCTTTACCCGGGGCACTGCTGCGGTCAAAGGACTTAAC
CGGATAGGAAATCTGTTGGTGCCCAATGATAACTATTGCAAGTTTGGAAACTGGATCATGCCAATCTTTCGACAAGATGCT
ACAAGAACAATCTACTGAGAATGTCTGGACACCATCAAAGGTGATGCTCGTCTTGGCAAAGAAATAAATGACGAAACCT
CCTACCTTTATTGGGCATACAAGAACAATATTCCTGTATACTGCCAGCATTGACTGATGGATCACTCGGAGACATGTTG
TTTTGTCTGACGTTCCGAATCTGGTCTTATTATTGACATTGTACATGATATAAGGCTCATAAATGGTGAAGCCATTCA
TGCAAGCCCAAGAAAGACAGGGTTCATAATTCTTGGTGGAGGCTTCCAAGGCATCATATATGCAATGCCAATATGTTTC
GCAATGGTGCAGATTATGCAGTGTACATCAACACAGCTCAAGAGTTGATGGAAAGTACTCAGGAGCACAGCCAGATGAA
GCAGTTTTCATGGGAAAGATCAAGGGTTCAAGAAACCTGTAAAGGTGCATTGATGCAACTATTGCTTTCCCTCTCCT
CGTTGCGCCGCAACATTTGCGCTAGGTTTTCATGGCACAAAT

>Brachypodium distachyon dhs

GATGCGCTGGAGGGGGTGCCTGCCATCGTGCTGAAGCCATCTGAATCGCTTGACGAATCGCGGGTGCAGGATCGCCGG
 CGCCGACTTCAACGACGCGCGGCTCGGCTCGGAGGCTGCTCGGGTCTCTTGCTCCACGGGGTCCAGGCGTCCCACC
 TTGGTGACGCCATCGACGCTCGTCAATCAGATGCTAGATTGGAGGTTGCCATGAGAAGCCTAGTGAGGATTGTGATGAA
 GCTGATCTTGACCCTAAATACAGAGAATCAGTGAAGTGAAGATTTTTCTTGGTTTCACTTCAAACCTTGTGTCTTCTGG
 CATACGGGATGTAATTCGGTTTTAGTTCAACATCACATGGTGGATGTTATTGTCACAACTGCTGGGGGTATAGAGGAGG
 ATCTCATTAATGCCTTGACCAACTTACAGAGGTGATTTTTCTTACCTGGAACACTACTGCGGTCAAAGGACTGAAC
 CGGATAGGAAATCTGCTGGTGCCCAATGATAACTACTGCAAGTTTGGAGAACTGGATCATGCCACTCTTTGACCAGATGCT
 ACTAGAACAATCTACTGAGAATGTTGGACACCATCGAAGGTGATTACTCGTCTTGGGAAAGAAATAAATGACGAAACCT
 CCTACCTTTATTGGGCATACAAGAACAATATTCTGTATACTGCCATCATTGACTGATGGATCACTGGAGACATGCTT
 TTCTGTATGCAGTCCGCAATCTGGTTTTATTATTGATATTGTACAAGATATACGGCTGATTAATGGGGAAGCCATTCA
 TGCAACCCCAAGGAAGACAGGCGCTATAATTCTTGGTGGAGGCCCTCCGAAGCATCATATATGCAATGCCAATATGTTCC
 GCAATGGTGCAGATTATGCAGTCTATATCAACACAGCTCAAGAGTTTGTATGGTAGTATTCAGGAGCACAAACCCGATGAA
 GCAGTTTCATGGGAAAGATCAAGGGTTCAGCAAAAACCTGTAAGGTTCCATTGCGATGCGACTATTGCTTTCCCGGTACT
 TGTGGCGCAACATTTGCACGTAGGTTTCATGGTGCGAAT

>Microstegium vimineum dhs3

GACGCCCTGGAAGACGTGCGCTCCATCCTGCTGAAACCGTGGAGCGCCTCGACGACAAGCGCTTCCAGAGAATCGCCGG
 CGCCGACTTCGACGACGCGAGCCTCGGCTCGCGGAGCTGCTCGGCTCGCTCGCGTCCACTGGCTTCCAGGCTCTAACC
 TCGGCGACGCCATCGACGCTCGTCAACCAGATGTTGGAGTGGAGGTTATCCCACGAGAAGCCAAGCGATGATTGTGCTGAA
 GGTGAACCTTGATCTGCATATAGAGAATCCGTCAAGTGAAGATCTTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
 TGTTCGGGAGATTATCCGATTTCTGCTCAGCATGGCATGGTGGATGTCATTGTTACGACTGCTGGGGGTATAGAGGAGG
 ATCTCATCAAATGCCTTGCGCCAACTTACAAAGGTGATTCTCTTACCTGGAGCAATGCTGCGGTCAAAGGACTTAAC
 CGGATAGGAAACCTGTTGGTGCCCAATGATAACTACTGCAAGTTTGGAGAACTGGATCATGCCATCCTTGAACAAGTCT
 GCTGGAACAATCCAAAGGAAATGTTGGACACCATCAAGGTGATTGCTCGTCTTGGCAAGAAATAAATGATGAAAGCT
 CCTACTTTACTGGGCATACAAGAACAACATCCCTGTGTTCTGCCAGCATTGACTGACGGATCAATTGGAGACATGCTA
 TTTTGTCACTCGGTCAACAATCCTGGTCTTATTGTTGATATTGTACAAGATGTAGGGCTGATGAATCGGAAGCCATCCA
 CGCAAGCCCAAGGAAGACAGGGATTATAATTCTTGGTGGAGGCCCTCCAAAGCATCATATATGCAATGCCAATATGATGC
 GGAACGGTGCAGATTATGCAGTGTGTTGTAACACAGCTCAGGAGTATGATGGTAGCGATTCTGGGGCACGGCCTGACGAA
 GCGGTTTCATGGGAAAGATAAAGAGCTCAGCAAAAAGTGTAAAGGTGCATTGCGATGCAACGATTGCTTTTCCGCTTAT
 TGTGGCTGCAACATTTGCACGCAAGGTTTCATGGCACTAAA

>Miscanthus sinensis dhs3

GACGCCCTGGAAGACGTGCGCTCCATCCTGCTGAAACCGTGGAGCGCCTCGACGACAAGCGCTTCCAGAGGATCGCCGG
 CGCCGACTTCGACGACGCGGCGCTCGGCTCGCGGGGCTGCTCGCTCGCTCGCGTCCACGGGGTCCAGGCTCTAACC
 TCGGCGACGCTATCGACGCTCGTCAACCAGATGTTGGACTGGAGGTTATCCCACGAGAAGCCAAGCGATGATTGTGCTGAA
 GGTGAACCTTGATCTGCATATAGAGAATCCGTGAAGTGAAGATCTTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
 TGTTCGGGAGATTATCCGATTTCTGCTCAGCATGGCATGGTGGATGTTATCGTTACAACCTGCTGGGGGTATAGAGGAGG
 ATCTCATCAAATGCCTTGCGCCAACTTACAAAGGTGATTCTCTTACCTGGAGCTCTGCTGCGGTCAAAGGACTTAAC
 CGGATAGGAAACCTGTTGGTGCCCAATGATAACTACTGCAAGTTTGGAGAACTGGATCATGCCAATCCTTGAACAAGTCT
 GCTGGAACAATCCACCGAAATGTTGGACACCATCAAGGTGATTGCTCGTCTTGGCAAGAAATAAATGATGAAAGCT
 CCTACTTTACTGGGCATACAAGAACAACATCCCTGTGTTCTGCCAGCATTGACCAGCGATCAATTGGAGACATGCTA
 TTTTGTCACTCGGTTACAATCCAGGTTCTTATTGTTGACATTGTACAAGATGTACGGCTGATGAATGCGGAAACCATCCA
 CGCAATCCCAAGGAAGACAGGGATTATAATTCTTGGTGGAGGCCCTCCAAAGCATCACATATGCAATGCCAATATGTTGC
 GCAACGGTGCAGATTATGCAGTGTATGTCAACACAGCTCAGGAGTTTGTATGGTAGTATCTGGGGCACGGCCTGACGAA
 GCGGTTTCATGGGAAAGATAAAGAGCTCAGCAAAAACCTGTAAGGTTCCATTGCGATGCGACGATTGCTTTTCTGCTTCT
 TGTGGCTGCAACATTTGCACGCAAGGTTTCACGGCACTAAA

>Saccharum officinale dhs3

GACGCCCTGGAAGACGTGCGCTCCATCCTGCTGAAACCGTGGAGCGCCTCGACGACAAGCACTTCCAGAGGATCGCCGG
 CGCCGACTTCGACGACGCGAGCCTCGGCTCGCGGGGCTGCTCGCTCGCTCGCGTCCACGGGGTCCAGGCTCTAACC
 TCGGCGACGCCATCGACGCTCGTCAACCAGATGTTGGACTGGAGGTTATCCCACGAGAAGCCAAGCGATGATTGTGCTGAA
 GGTGAACCTTGATCTGCATATAGAGAATCCGTGAAGTGAAGATCTTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
 TGTTCGGGAGATTATCCGATTTCTGCTCAGCATGGCATGGTGGATGTTATCGTTACAACCTGCTGGGGGTATAGAGGAGG
 ATCTCATCAAATGCCTTGCGCCAACTTACAAAGGTGATTCTCTTACCTGGAGCGCTGCTGCGGTCAAAGGACTTAAC
 CGGATAGGAAACCTGTTGGTGCCCAATGATAACTACTGCAAGTTTGGAGAACTGGATCATGCCAATCCTTGAACAAGTCT
 GCTGGAACAATCCACCGAAATGTTGGACACCATCAAGGTGATTGCTCGTCTTGGCAAGAAATAAATGATGAAAGCT
 CCTACCTTTACTGGGCATACAAGAACAACATCCCTGTGTTCTGCCAGCATTGACTGACGGCTCAATTGGAGACATGCTA
 TTTTGTCACTCGGTTACAATCCAGGTTCTTATTGTTGACATTGTACAAGATGTACGGCTGATGAATGCGGAAACCATCCA
 CGCAACCCCAAGGAAGACAGGGATCATAATTCTTGGTGGAGGCCCTCCAAAGCATCATATATGCAATGCCAATATGTTGC
 GCAACGGTGCAGATTATGCAGTGTATGTCAACACAGCTCAGGAGTTTGTATGGTAGTATCTGGGGCACGGCCTGACGAA
 GCGGTTTCATGGGAAAGATAAAGAGCTCAGCAAAAACCTGTAAGGTTCCATTGCGATGCAACGATTGCTTTTCCGCTTCT
 TGTGGCTGCAACATTTGCACACAAGGTTTCACGGCACTAAA

>Sorghum bicolor dhs3

GACGCCCTGGAAGACGTGCGCTCCATCCTGCTGAAACCGTGGAGCGCCTCGACGACAAGCGCTTCCAGAGGATCGCCGG
 CGCCGACTTCGACGACGCGAGCCTCGGCTCGCGGGGCTGCTCGCTCGCTCGCGTCCACGGGGTCCAGGCTCTAACC
 TCGGCGACGCCATCGACGCTCGTCAACCAGATGTTGGACTGGAGGTTATCCCACGAGAAGCCAAGCGATGATTGTGCTGAA
 GGTGAACCTTGATCTGCATATAGAGAATCCGTGAAGTGAAGATCTTTCTGGGCTTCACTTCAAACCTTGTGTCTTCTGG
 TGTTCGGGAGATTATCCGATTTCTGCTCAGCATGGCATGGTGGATGTTATTGTTACAACCTGCTGGGGGTATAGAGGAGG
 ATCTCATCAAATGCCTTGACCAACTTACAAAGGTGATTCTCTTACCTGGAGTGTGCTGCGGTCAAAGGACTTAAC
 CGGATAGGAAACCTGTTGGTGCCCAATGATAACTACTGCAAGTTTGGAGAACTGGATCATGCCAATCCTTGAACAAGTCT
 GCTGGAACAATCCACCGAAATGTTGGACACCATCAAGGTGATTGCTCGTCTTGGCAAGAAATAAATGATGAAAGCT
 CCTACTTTACTGGGCATACAAGAACAACATCCCTGTGTTCTGCCAGCATTGACTGACGGCTCAATTGGAGACATGCTA
 TTTTGTCACTCGGTTACAATCCAGGTTCTTATTGTTGACATTGTACAAGATGTACGGCTGATGAATGCGGAAACCATCCA
 CGCAACCCCAAGGAAGACAGGGATTATAATTCTTGGTGGAGGCCCTCCAAAGCATCATATATGCAATGCCAATATGTTGC
 GCAACGGTGCAGATTATGCAGTGTATGTCAACACAGCTCAGGAGTTTGTATGGTAGTATCTGGGGCACGGCCTGACGAA
 GCGGTTTCATGGGAAAGATAAAGAGCTCAGCAAAAACCTGTAAGGTTCCATTGCGATGCAACGATTGCTTTTCCGCTTCT
 TGTGGCTGCAACATTTGCACACAAGGTTTCACGGCACTAAA

>Zea mays dhs3

GACGCGCTGGAAGCGTGCGCTCCATCCTGCTGAAACCGTCGGAGCGCCTCGACGACAAGCGCTTCACGAGGATCGCCGG
CGCCGACTTCGACGACGCCGGCCTCGGCCTCGCCGGGTTGCTCGGCTCGCTCGCGTCCACGGGGTTCCAGGCCCTCCAACC
TCGGCGACGCCATCGACGTCGTCAACCAGATGTTGGACTGGAGGTTGTCCCACGAGAAGCCAAGCGACGACTGCGCTGAA
GCTGAACTCGATCCTGCGTACAGAGAATCCGTCAAGTGCAAGATATTCTGGGGTTCACCTTCGAACCTTGTGTCTCCGG
TGTTCCGGGAGATCATCCGGTTTCTTGTTCAGCATCGCATGGTGGACGTTATTGTTACGACTGCTGGGGGTATAGAGGAGG
ACCTCATCAAATGCCTTGCGCCAACCTACAGAGGCGACTTCACTCTGCCGGAGCGCTGCTGCGGTCCAGAGGACTTAAC
CGGATAGAAACCTGCTGGTGCCCAATGACAACCTACTGCAAGTTCGAGCACTGGATCATGCCGATCCTCGACAAGATGCT
GCTGGAACAGTCCACCCAAAATGTTTGACGCGCGTCCAAGGTGATTGCTCGCCTTGGCAAGGAAATAAATGACGAAAGCT
CCTACCTTTATTGGGCGTACAAGAACAACATCCCTGTGTTCTGCCAGCATTGACTGACGGATCAATTGGGGACATGCTG
TTCTGTCACTCGGTTACAGTCCAGGTCCTGTTGTCGACATTGTACAAGATGTACGGCTGATGAACGGGAAACCATCCA
CGCAAGCCCAAGGAAGACAGGGATATATACTCCTTGGCGGAGGCCTTCCGAAGCATCATATATGCAACGCCAATATGCTGC
GCGACGGAGCTGACTATGCGGTGTATGTCAACACAGCTCAGGAGTTCGATGGCAGTGATTCGGGGCGCGCCTGACGAA
GCGGTTTCATGGGAAAGATCAAGAGCTCAGCGAGAAGTGTGAAGGTGCATTGTGATGCGACGATTGCCTTTCCACTTCT
TGTGGCTGCAACATTTGCACGCAAGGTTACGGCACTAAA

Chapter 2:

Characterization of pyrrolizidine alkaloid occurrence in the grass subfamily Pooideae (Poaceae)

Anne-Maria Wesseling ^a, Tobias J. Demetrowitsch ^b, Karin Schwarz ^b, Dietrich Ober ^a

^a Botanisches Institut, CAU Kiel,
Am Botanischen Garten 1-9, 24118 Kiel, Germany

^b Institut für Humanernährung und Lebensmittelkunde, CAU Kiel
Heinrich-Hecht-Platz 10, 24118 Kiel, Germany

Author contributions are detailed at the end of this thesis.

Abstract

Pyrrolizidine alkaloids (PAs) are a class of specialized metabolites produced by various plants from different lineages of flowering plants including cool-season grasses (Poaceae, subfamily Pooideae). The grasses are only distantly related to the other known PA-producing taxa and it has been suggested that PA biosynthesis evolved independently in this subfamily. In addition, potentially PA-biosynthetic genes have been identified from several different Pooideae species. However, until today the grass species *Lolium perenne* and *Festuca arundinacea* are the only ones which have been tested and confirmed to produce PAs. We therefore grew plants of the Pooideae lineage under controlled conditions and screened them for the occurrence of PAs via an LC-MS based targeted metabolomics approach. The results revealed that PA biosynthesis is limited to a group of very closely related Pooideae species, and displays to a great inter- and intraspecies variability of PA levels as well as a limited diversity of PA structures.

Keywords

Pooideae, LC-QTOF-MS, SCX-SPE, pyrrolizidine alkaloid, thesinine, thesinine-glycoside

Introduction

Pyrrolizidine alkaloids (PAs) are a typical class of plant specialized (or secondary) metabolites produced in various lineages of the flowering plants and their occurrence, structural diversity, and biosynthesis have been studied extensively for several decades (Hartmann and Witte, 1995; Ober and Kaltenecker, 2009; Stegelmeier et al., 1999). In cool-season grasses (Poaceae, subfamily Pooideae) a specific class of PAs, known as lolines, have been described to accumulate in individuals infected with special endophytic fungal symbionts that were shown to synthesize these alkaloids (for an excellent review see Schardl et al., 2007).

In 2008 Koulman and colleagues analyzed grass extracts and described certain PAs as the product of the plant's own metabolism (Koulman et al., 2008). The authors were able to show that specific PA conjugates are present in the pasture grass species *Lolium perenne* and *Festuca arundinacea* irrespective of a fungal infection (Koulman et al., 2008). Structural elucidation revealed these compounds to be the stereoisomers *E*- and *Z*-thesinine in varying states of glycosylation (i.e. aglycones, rhamnosides and rhamnoside-glycosides, see Koulman et al., 2008). PAs are generally known as defensive compounds protecting the plant from herbivores (Macel, 2011). However, 1,2-saturated PAs such as thesinine are considered non-toxic (Hartmann and Witte, 1995). For this reason and the fact that these alkaloids accumulate in highly bred grass cultivars it was concluded that these compounds are not part of the plant's defense mechanism against herbivory (Koulman et al., 2008). The biological function of these glycosylated PAs found in these grasses is therefore unknown.

PA biosynthesis relies on the presence of a homospermidine synthase (HSS), which catalyzes the first pathway-specific step and, thereby, connects essential (or primary) with the specialized metabolism (Ober et al., 2003a). The HSS has its origin in the duplication of a gene encoding the deoxyhypusine synthase (DHS), an enzyme of primary metabolism involved in the post-translational activation of the eukaryotic initiation factor 5A (Park, 2006; Park et al., 2010). A recent study focusing on the molecular evolution of grass-specific HSS and DHS suggested that the ability to produce PAs evolved independently in grasses from other PA-producing taxa (Wesseling and Ober, unpublished results [Chapter 1]). The *hss* genes identified from *L. perenne* and *F. arundinacea* originated from a gene duplication event that took place during Pooideae

species radiation and several other extant Pooideae species have retained functional *hss* genes in their genomes (Wesseling and Ober, unpublished results [Chapter 1]).

Other than the aforementioned pasture grasses, the Pooideae also contain important crop species such as bread wheat (*Triticum aestivum*). As this lineage might have the capacity to produce PAs, we decided to screen further members this taxonomic group, in addition to *L. perenne* and *F. arundinacea*, for the presence of PAs.

We have grown various grass species under controlled conditions and tested them for the presence of PAs through a targeted metabolomics approach using LC-MS. Determining the spread and variations in alkaloid levels in this taxon should give us better general understanding of the subject and provide a basis for subsequent studies on various aspects of PA biosynthesis in grasses.

Results

Approach for the detection of thesinine and its conjugates via LC-QTOF-MS

To detect thesinine and its conjugates crude grass extracts were analyzed by LC-QTOF-MS. The compounds could readily be identified by their respective $[M+H]^+$ ions, namely m/z 288.159 for thesinine (M: $C_{17}H_{21}NO_3$) and m/z 434.217 for thesinine-rhamnoside (M: $C_{23}H_{31}NO_7$). Broadband MSⁿ fragmentation provided further diagnostic ions, namely m/z 142.121 and m/z 124.112, representing the necine base of thesinine and one of its typical fragments, respectively. In accordance with the findings of Koulman and colleagues (2008), who identified *E*- and *Z*-stereoisomers of thesinine-rhamnoside, we observed two distinct m/z 434.2 peaks in the chromatogram with retention times (t_R) of 324 s (defined as compound **1**) and 334 s (compound **2**) (Fig. 1). In our LC-MS setup the aglycone thesinine (**3**) was shown to elute after the thesinine-rhamnosides at 340 s (Suppl. Fig. S1). Stereoisomers for a disaccharide conjugate of thesinine (m/z 549) were only present in traces in our LC-MS analyses (maximal 1000 counts). In every case, these samples also contained the monosaccharidic thesinine-rhamnosides (**1** and **2**), and as detailed analyses of the disaccharide conjugates would therefore not provide any additional information with this regard to the presence or absence of PAs in grasses, they were not further addressed in study. Furthermore, PAs are known to occur as either tertiary amines or *N*-oxides (Hartmann and Witte, 1995; Stegelmeier et al., 1999). The thesinine-conjugates identified by Koulman et al. (2008)

are tertiary amines and checking for according *N*-oxides (calculated m/z : 304.155 and 450.213 for the hypothetical thesinine-*N*-oxide and thesinine-*N*-oxide-rhamnoside, respectively) revealed that these were not present in the extracts.

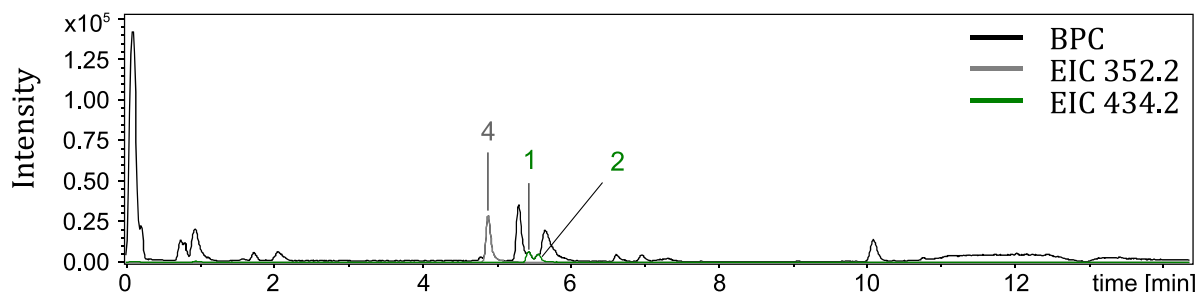


Fig. 1: Base peak chromatogram (BPC) of a *Lolium perenne* extract recorded via LC-QTOF-MS Extracted ion chromatograms (EICs, smoothed) depicting the elution profiles of the thesinine-rhamnosides (m/z 434.218, **1** and **2** and the internal standard retrorsine (m/z 352.173, **4**). For a more detailed view of the elution profiles of thesinine and its conjugates see Suppl. Figure S1.

Occurrence of thesinine and its conjugates is limited to specific Pooideae species

Initial LC-MS analyses of extracts obtained from *Triticum aestivum*, *Dactylis glomerata*, and *Holcus mollis* plants (all Pooideae) suggested a lack of the previously described thesinine-conjugates in these species (results not shown). Therefore, we decided to focus on the *Lolium-Festuca* species complex, the taxon that contains the two Pooideae species *L. perenne* and *F. arundinacea* for which PA biosynthesis has been detected (Cao et al., 2008; Jauhar, 1993a; Catalán et al., 2004). We chose *L. multiflorum* (cultivar Lema), *L. temulentum*, *L. rigidum*, *L. remotum*, *F. pratensis*, and *F. rubra*. Individuals of the known PA-producing species *L. perenne* (cultivar Fennema) and *F. arundinacea* were included as positive controls.

To quantify thesinine-conjugates for a comparison of PA content between samples, we used the PA retrorsine as an internal standard (compound **4**, m/z 352, Suppl. Fig. S1). To confirm stability of LC-MS measurements, we repeatedly measured so-called quality control (QC) samples (a mixture of all plant extracts in equal proportions) for minimal variation in their chromatograms (Demetrowitsch et al., 2015; European Medicines Agency, 2011; Xu et al., 2007). Comparison of peak intensities of five

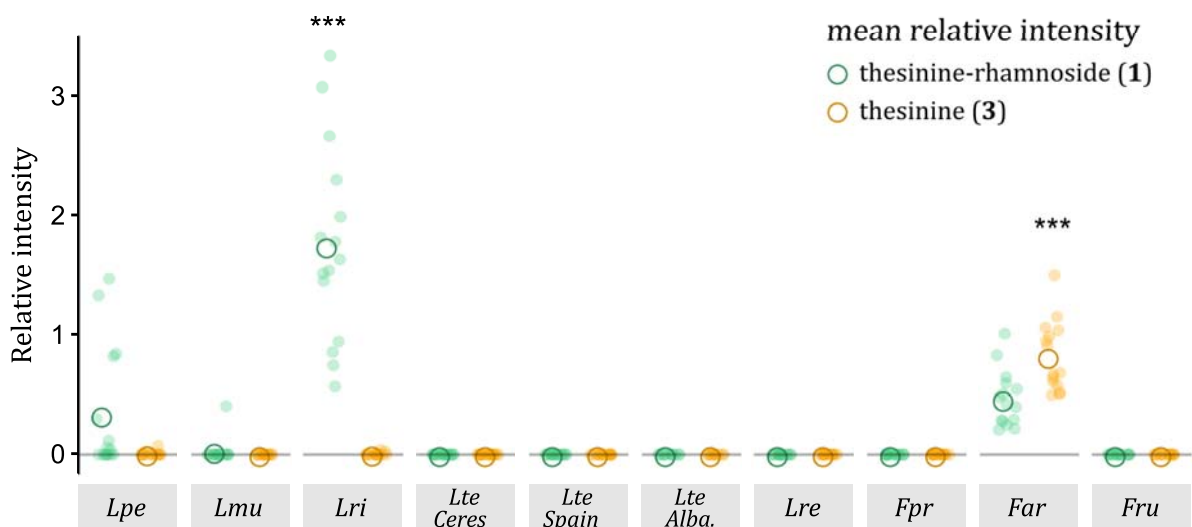


Fig. 2: Thesinine-rhamnoside (1) and thesinine (3) in various *Lolium* and *Festuca* species. Relative intensity represents the peak intensities as determined in LC-MS analyses for **1** (434.2 m/z , 324 s) and **3** (288.1 m/z , 340 s) relative to the internal standard (retrorsine, 0.375 $\mu\text{g}/\text{ml}$). ANOVA of **1**, $p < 0.001$ ***, Tukey post-hoc tests for *Lri*, adjusted $p < 0.001$ ***, ANOVA of **3**, $p < 0.001$ ***, Tukey post-hoc tests for *Far*, adjusted $p < 0.001$ ***. *Lpe* – *L. perenne* (Fennema), *Lmu* – *L. multiflorum* (Lema), *Lri* – *L. rigidum*, *Lte* – *L. temulentum*, *Lre* – *L. remotum*, *Fpr* – *F. pratensis*, *Far* – *F. arundinacea*, *Fru* – *F. rubra*.

randomly chosen m/z signals in addition to our compounds of interest namely **1** to **4**, revealed only little variation between QC measurements with relative standard deviations of peak intensities lower than 15%, except for m/z 437.2 (603 s). The value of 15% has been suggested as the upper limit for standard deviation by EMA (Suppl. Table 1) (European Medicines Agency, 2011). In addition, we performed a principal component analysis (PCA) on the intensities of m/z - t_{R} pairs (the so-called buckets) to further confirm measurement stability. The PCA plot revealed an expected central aggregation of QC samples, and otherwise showed no significant batch bias or other technical artifacts (Suppl. Fig. S2).

For the following quantitative analyses we calculated peak intensities of thesinine-rhamnoside (**1**) and thesinine (**3**) relative to the internal standard (**4**) (referred to as the relative intensity in the present study). Comparison of the average relative intensities calculated for each species showed widely varying levels of **1** and **3** between the different species (Fig. 2). *L. rigidum* showed by far the highest average relative intensities for **1** in comparison with the other tested species (adjusted $p < 0.001$,

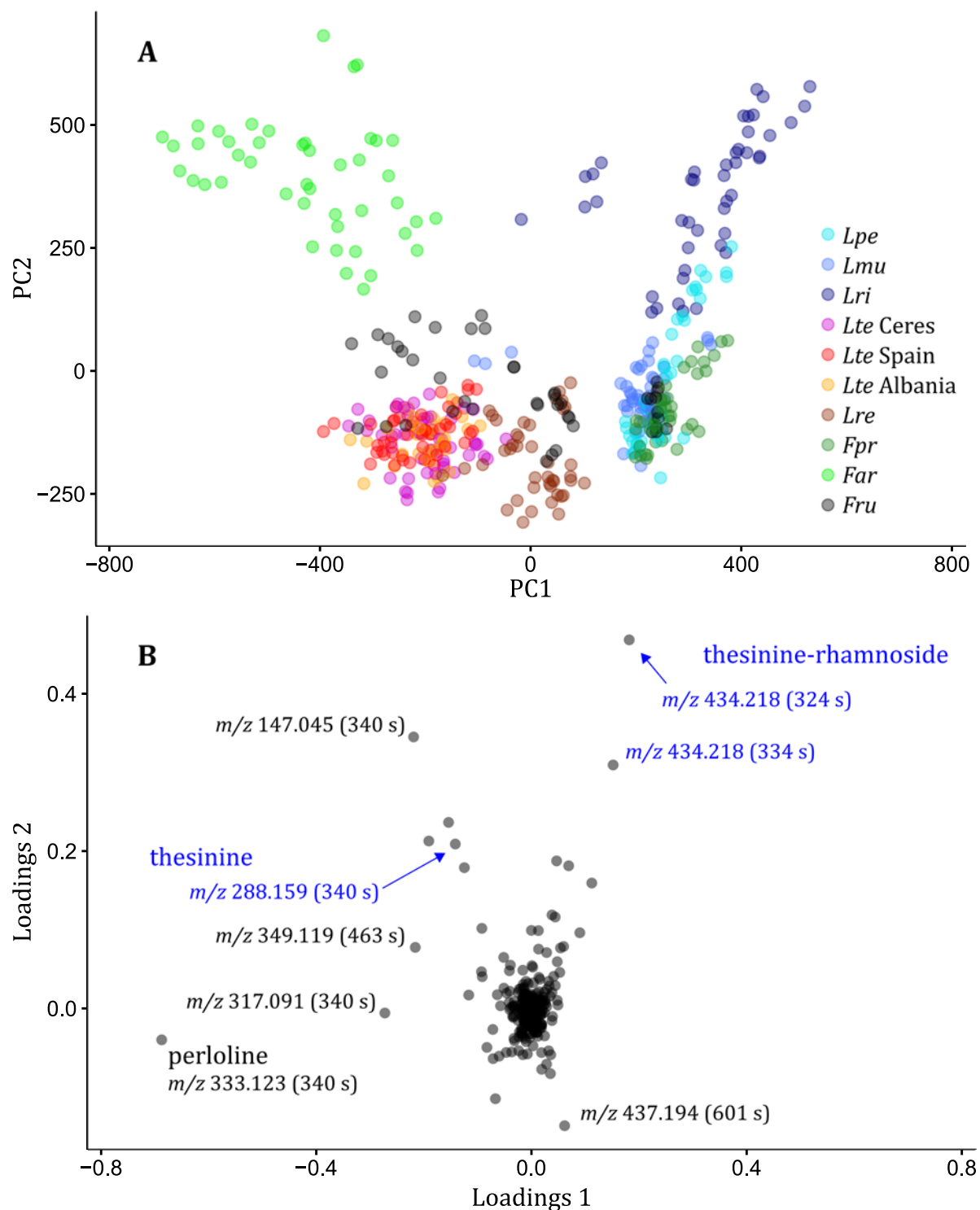


Fig. 3: Principal component analysis of all plant extract LC-MS analyses. A, Scores plot, coloring according to species. B, Loadings plot with notable compounds marked. QC samples measurements, outliers and buckets representing internal standards are excluded. *Lpe* – *L. perenne* (Fennema), *Lmu* – *L. multiflorum* (Lema), *Lri* – *L. rigidum*, *Lte* – *L. temulentum*, *Lre* – *L. remotum*, *Fpr* – *F. pratensis*, *Far* – *F. arundinacea*, *Fru* – *F. rubra*.

ANOVA, Tukey's test). The prominent presence of thesinine-conjugates in this species and in *L. multiflorum* are novel observations. *L. perenne* and *F. arundinacea* displayed intermediate levels of thesinine-rhamnoside, while *L. multiflorum* on average contained the least amount of the alkaloid per biomass (Fig. 2).

The aglycone **3** was only found in significant amounts in *F. arundinacea* (adjusted $p < 0.001$, ANOVA, Tukey's test); lower intensities (up to 2500 counts) were recorded in some samples of the other species. As thesinine was only detected in those *Lolium* samples that also contained thesinine-rhamnoside, it is probably the result of thesinine-conjugate hydrolysis taking place in the extract.

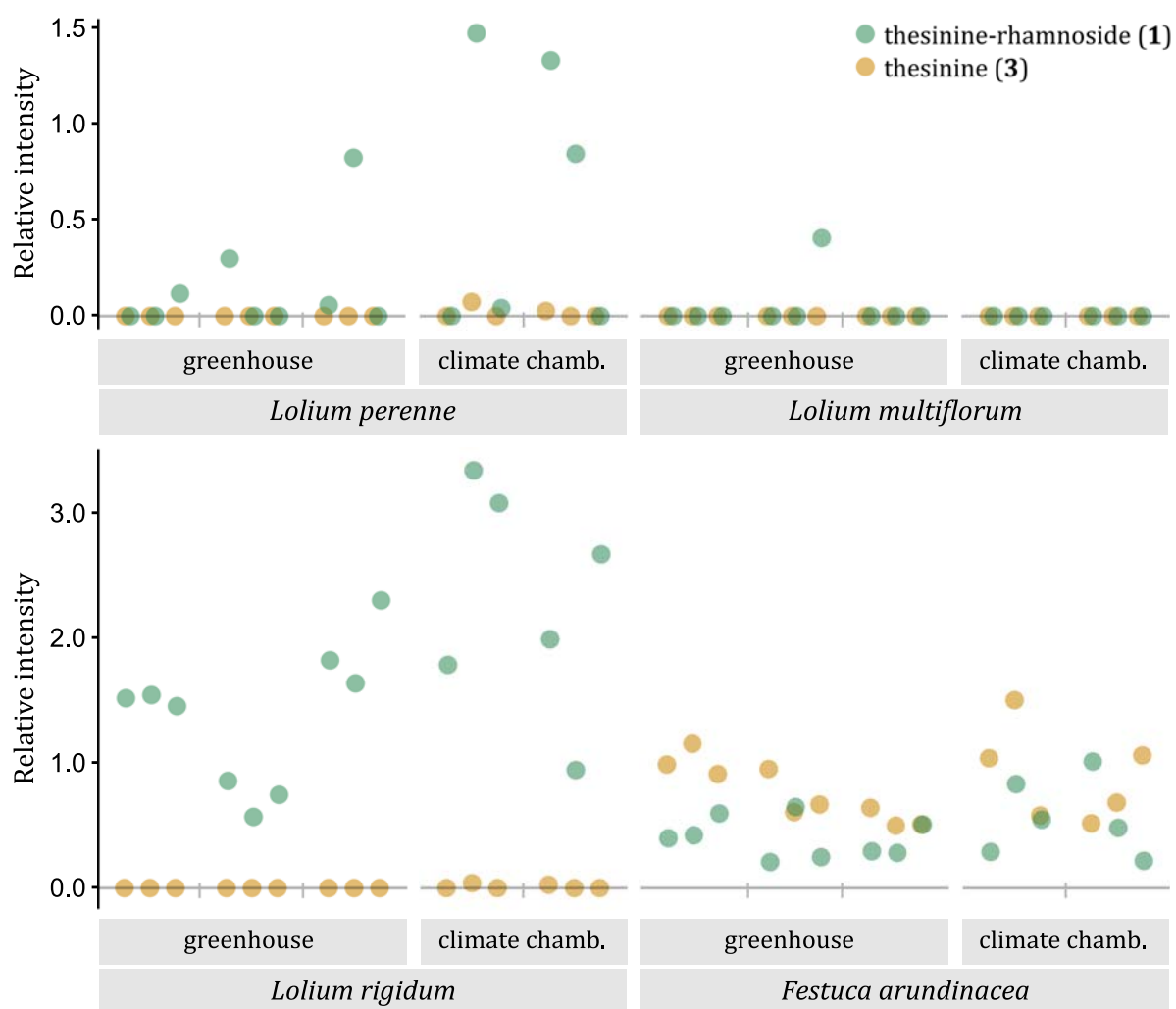


Fig. 4: Occurrence of thesinine-rhamnoside (1) and thesinine (3) in individual samples of *L. perenne*, *L. multiflorum*, *L. rigidum*, and *F. arundinacea*. Intensities extracted from LC-MS data for 1 (m/z 434.2, 324 s, green) and 3 (m/z 288.1, 340 s, orange) displayed relative to the intensity of an internal standard (retrorsine, 0.375 $\mu\text{g}/\text{ml}$). Samples are grouped according to the place the grass plants were grown (greenhouse and climate chamber) and the growth stage at which the plants were sampled: 2, 6 and (in the case of greenhouse-grown plants) 8 weeks.

Another PCA performed exclusively on biological MS data (i.e., without QC sample measurements and excluding internal standards) revealed a strong clustering of samples by species (Fig. 3). The corresponding loadings revealed that the difference between species was mainly attributed to the varying levels of perloline (m/z 333.123, 340 s), a diazaphenanthrene alkaloid known from *L. perenne* and *F. arundinacea* metabolomes (Bush and Jeffreys, 1975; Cao et al., 2008). The thesinine-conjugates **1** and **2** are among the 10 buckets with the highest loadings meaning that their variability greatly contributes to the species-specific differences in metabolic composition.

Variable occurrence patterns of thesinine conjugates between species

Analysis of samples taken from various individuals of the PA containing species for their thesinine-conjugate content revealed an unexpected pattern of intra-species variation with respect to compound **1** and **2** (Fig. 4). Whereas all sampled *L. rigidum* and *F. arundinacea* individuals contained levels of the alkaloid above the detection limit, the occurrence of thesinine-rhamnoside is variable in individuals of *L. perenne* and *L. multiflorum*. The alkaloid was only detected in 8 out of 15 tested *L. perenne* plants and, even more remarkably, only in one (out of 15) *L. multiflorum* plant, which thereby explained the low average relative intensity observed for that species (Fig. 2 and Fig. 4). PA biosynthesis is known to be coupled to plant growth in various species and is localized in specific tissues and cells, resulting in an individual spatio-temporal pattern of PA accumulation (Anke et al., 2004, 2008). By analyzing plants at various growth stages (2, 6, and 8 weeks after germination), we detected increasing amounts of PAs, suggesting a constitutively active PA biosynthesis and accumulation in *L. rigidum* and *F. arundinacea*. Levels of thesinine-rhamnoside **1** and thesinine **3** relative to the biomass did not vary significantly at the diverse growth stages (ANOVA, $p > 0.05$) indicating that, on average, PA levels are kept constant throughout plant development. Whereas for *L. perenne*, thesinine-rhamnoside occurrence is sporadic, we detected no growth stage, for which not at least one plant individual contained thesinine-rhamnoside, showing that this species is capable of PA biosynthesis as early as the three-leaf stage of the grass (two weeks after germination).

The occurrence of thesinine-rhamnoside is not restricted to specific cultivars of *L. perenne* and *L. multiflorum*

The striking sporadic occurrence of the PAs in *L. perenne* and *L. multiflorum* motivated us to test whether this pattern is a specific property of the tested cultivars. We therefore expanded the LC-MS analyses to include a variety of cultivars of both species. For *L. perenne*, we choose the cultivar ‘Chicago’ in addition to ‘Fennema’ (see 2.1), whereas for *L. multiflorum*, we screened the cultivars ‘Lema’ (see 2.1), ‘Pilgrim’, ‘Lipo’, and ‘Fabio’. All cultivars were grown in parallel in the greenhouse and sampled 7 weeks after germination. We used the same LC-QTOF-MS approach as described above, including the use of QCs to show measurement stability (Suppl. Table S1).

Both *L. perenne* cultivars showed, on average, higher relative intensities for **1** and **2** than the *L. multiflorum* cultivars (Fig. 5). Quantification of thesinine-rhamnoside levels in the samples of plant individuals confirmed the sporadic occurrence of the alkaloid in the

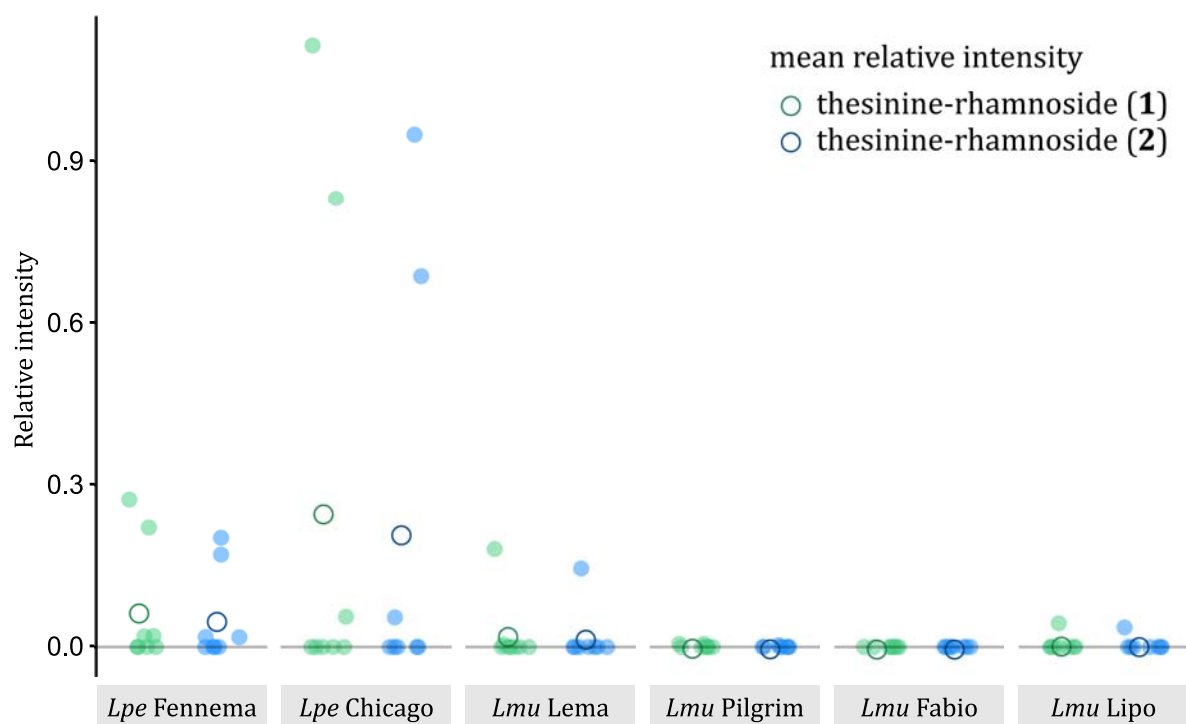


Fig. 5: Relative intensity of thesinine-rhamnosides (1 and 2) in various *L. perenne* and *L. multiflorum* cultivars. Relative intensity is the peak intensity as determined in LC-MS analyses for **1** and **2** (m/z 434.2, 324 s and 334s, respectively) normalized over an internal standard (retrorsine, 0.32 $\mu\text{g/ml}$). *Lpe* – *L. perenne*, *Lmu* – *L. multiflorum*.

two species: about half of the analyzed individuals of *L. perenne* contained thesinine-rhamnosides **1** and **2** (4 plants out of 8 for 'Fennema' and 3 out of 8 for 'Chicago'). For the various *L. multiflorum* cultivars, thesinine-rhamnosides were only detected a total of five times out of the 32 individuals (2 for 'Lema', 2 for 'Pligrim', 0 for 'Fabio', and 1 for 'Lipo').

PA occurrence within grasses is limited to thesinine-conjugates and to a few Pooideae species

In addition to the analyses focusing on the thesinine-conjugates we also screened the Pooideae species for potential accumulation of other PAs. These analyses included not only the *Lolium* and *Festuca* species (and cultivars) that we have described so far, but also the species *D. glomerata*, *H. lanatus* and *T. aestivum*. *Brachypodium distachyon* was included as a species that split off from other Pooideae before the evolution of the Pooideae-specific *hss* gene (Wesseling and Ober, unpublished results [Chapter 1]) and serves as a negative control in this experiment. To enrich potentially occurring PAs and to increase their chance of detection, plant extracts were applied to strong cation exchange solid-phase extraction columns (SCX-SPE) before being subjected to LC-QTOF-MS analysis. The SCX-SPE method was shown to be suitable for the extraction of PAs in their different oxidation states, i.e. tertiary PAs and their respective *N*-oxides (Betteridge et al., 2005; Colegate et al., 2005; Mroczek et al., 2002). We found that the method was also applicable for the extraction of glycosylated thesinines making it a convenient method for extracting and detecting the complete PA repertoire of a plant. To identify PAs, the fragmentation mass spectra of prominent peaks were inspected for typical necine base fragments. Whereas several prominent peaks were apparent in chromatograms of the various plant extracts, no fragments typical for necine bases of PAs were found (except that of the thesinine-conjugates) suggesting that the PA spectrum of grasses tested in this study is limited to the thesinine-conjugates.

Discussion

Despite the agronomical importance of grasses, little is known about PAs that are not the product of endophytic fungi. To shed light on this aspect of PA biosynthesis, we have analyzed the PA occurrence in selected species of cool-season grasses (Pooideae, Poaceae). We performed LC-MS-based screening experiments to test a variety of species from this taxon and found only few incidences of PA production in this taxon (Fig. 1). In our surveys, PA accumulation was limited to four species placed in the so-called broad-leaved fescue lineage of the *Lolium-Festuca* species complex. The three PA-producing *Lolium* species, namely *L. perenne*, *L. multiflorum* and *L. rigidum*, are outbreeders and have been shown to be closely related to each other (Catalán et al., 2004; Cheng et al., 2015). These outbreeding *Lolium* species present a sister group to the inbreeding *Lolium* species which include *L. temulentum* and *L. remotum* and which we have shown here to be PA-free. The PA-producing species *F. arundinacea*, however, presents an exception to this pattern. It is also a broad-leaved fescue but of the subgenus *Schedonorus*, which forms a sister taxon to *Lolium* (Catalán et al., 2004; Cheng et al., 2015). The presence of PAs in *F. arundinacea* is surprising considering their absence in *F. pratensis*. The two species have been shown to be closely related to each other, as one of the *F. arundinacea* subgenomes is derived from an *F. pratensis* ancestor (Humphreys et al., 1995). This unexpected pattern of PA occurrence can be explained either by the independent evolution of PA biosynthesis in *F. arundinacea* or by the secondary loss of it in the inbreeding *Lolium* species. Another possibility is that the plant acquired the ability to produce PAs through introgression from a PA-producing *Lolium* species. A comprehensive record of natural and synthetic hybridizations exists between *Festuca* species of the subgenus *Schedonorus*, to which *F. arundinacea* belongs, and the outbreeding *Lolium* species (Jauhar, 1993b). Indeed, several instances of introgression have already been documented (Humphreys and Thomas, 1993; Humphreys et al., 2003; King et al., 1998; Kopecký et al., 2008). Therefore, the genetic components required for PA biosynthesis might conceivably have reached the *F. arundinacea* genome through this mechanism. Detailed genetic and genomic analyses will be needed to clarify this.

Nevertheless, our observation, that PAs are found in one *Festuca* and several *Lolium* species, which include the crop weed *L. rigidum*, suggests that the ability to produce PAs appeared before the evolution of the extant outbreeding *Lolium* species and is thus not a

result of selective plant breeding processes as has previously been suggested (Koulman et al., 2008). This is also supported by the *dhs/hss* phylogeny which places the origin of the *hss* gene early during Pooideae evolution (Wesseling and Ober, unpublished results [Chapter 1]). In contrast to the limited occurrence of PA within the Pooideae, functional *hss* genes are widespread in the genomes of this taxon. The *hss* gene has been identified, amongst others, for *D. glomerata*, *H. lanatus*, and *T. aestivum* in addition to *L. temulentum* and *F. rubra* (Wesseling and Ober, unpublished results [Chapter 1]), which all tested negative for PAs in this study.

Remarkably, the PA spectrum that we have detected in the various grass species and cultivars is limited to thesinine and its conjugates. PA-producing species of the dicots are often characterized by bouquets of structurally related PAs (Hartmann and Witte, 1995). Despite the structural diversity of known PAs with around 350 different structures having been described in the literature (not including the respective *N*-oxides) (Roeder, 2000), the occurrence of glycosylated PAs is rather unusual. Apart from grasses, they have also been extracted from *Borago* seeds. It is of note that in *Borago* the aglycone of these PAs is also thesinine (Herrmann et al., 2002).

Despite the limited number of PA-producing grass species, we have observed an unexpected inter- and intra-specific variability in PA levels and occurrences. First, the analyzed PA intensities are highly variable (Fig. 1), and the co-occurrence of thesinine and thesinine-rhamnosides are not identical (Fig. 2). Within *F. arundinacea*, for example, the aglycone thesinine is prominently present, in addition to the thesinine-rhamnosides. Second, PAs were only sporadically found in individuals of *L. perenne* and *L. multiflorum*, whereas each tested individual of *L. rigidum* and *F. arundinacea* contained detectable amounts of PAs. As the plants tested in this study were only a few weeks old, we cannot exclude that PA accumulation is more common in later developmental stages. Even so, our data demonstrate the variability between species in that PA production seems to be a constitutive trait in some species, whereas in others, it occurs only sporadically and might possibly be induced.

In conclusion, our data show several unexpected characteristics of grass PA biosynthesis, raising questions concerning the evolutionary history of PA production and also about the function that these alkaloids serve in representatives of this taxon. Further studies are required to shed more light onto PA biosynthesis in grasses.

Experimental

General experimental procedures

Plant extract samples were analyzed by an LC-coupled QTOF-MS set-up. Analytes were separated on a reversed-phase LC column (Nucleodur C18 Gravity column, 100 mm x 2 mm, 1.8 μm , Macherey-Nagel) installed in an Infinity 1260 UHPLC system (Agilent Technologies). A solvent gradient was used at a flow rate of 250 $\mu\text{l min}^{-1}$ (solvent A: H_2O , 0.1% HCOOH , solvent B: CH_3CN , 0.1% HCOOH). The LC gradient began at 0% solvent B starting at 1 min and was ramped to 90% over the next 8 min. Column washing at 90% was maintained for 0.5 min and then ramped down to 0% again over 4.5 min the gradient remained for the next 5 min (total runtime of 15 min).

The measurements were performed with the microTOF-QII mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with an electrospray ionization (ESI) source. The ionization was conducted in the positive ionization mode with a broadband collision-induced dissociation method (bbCID). The source parameters were: dry gas at a temperature of 210 $^{\circ}\text{C}$ with a flow rate of 6 l/min and a nebulizer pressure of 1 bar. LC-MS data were further processed by using the Find Molecular Features (FMF) algorithm implemented in DataAnalysis 4.2 (Bruker Daltonics) by using the following peak finding parameters: $S/N = 3$, correlation coefficient threshold = 0.7, minimum compound length = 7 spectra, smoothing width = 1.

Plant material

For our screening experiments we used a selection of Pooideae species. All species used in this study are summarized in Supplementary Table S2 including information on cultivars and origin. Seeds were placed on wet filter paper until germination. When 2-3 days old, seedlings were transferred to soil-filled pots (Einheitserde Classic, Einheitserde Werkverband e.V.) and then kept either in a greenhouse or climate chamber. For the surveys, the above-ground parts of the grasses were sampled as a whole, flash-frozen in liquid nitrogen, and stored at -20°C until milling and extraction.

Thesinine-conjugate screening in *Lolium-Festuca* species complex

To screen for thesinine-conjugates in members of the *Lolium-Festuca* species complex we grew *L. perenne* 'Fennema', *L. multiflorum* 'Lema', *L. remotum*, *L. rigidum*, and three *L. temulentum* varieties ('Ceres', 'Albania' and 'Spain') plus *F. arundinacea* 'Arola', *F. pratensis*, and *F. rubra* in two lots: in a greenhouse (September/October 2014 in Kiel, Germany) and a climate chamber (16 h light and 21°C, 8 h dark and 18°C). After 2, 6, and 8 weeks, three plants from each lot and of every species/cultivar were sampled as biological replicates. Freeze-dried plants were milled to produce a fine powder, of which 10.0 mg was then extracted with 2.0 ml of 75% MeOH + 0.1% FA.

To be able to assess LC-MS measurement stability we made use of quality control (QC) samples following the protocol of Demetrowitsch et al. (2015). For this, 20 µl of each plant extract was pooled to create a QC sample, which was then always subjected to the same preparation procedure and freeze-thaw cycles samples as the plant extract samples. Samples were diluted 1:4 (25 µl + 75 µl) with MeOH spiked with a pyrrolizidine alkaloid as an internal standard (retrorsine, Sigma-Aldrich) at a final concentration of 0.375 µg/ml. All samples were stored at -80°C until the day of measurement. For LC-MS analyses, the samples were divided into three batches, each containing one of the biological replicates, and each being measured three times (technical replicates) within 43 hours of thawing. Before each batch measurement, the QTOF was cleaned following the protocol of Demetrowitsch et al. (2015), and at the beginning of each measurement cycle, the QC sample was injected twice for column equilibration (measurements were not included in subsequent data analyses). The QC samples were injected again after every 20th measurement followed by a blank sample (MeOH).

All samples were recorded in the bbCID (broadband collision-induced dissociation) mode providing both MS and MS_n spectra at a detection rate of 4 Hz and a collision energy of 10 eV for MS and 20 eV for MS_n results.

After data acquisition, we compared internal standard intensities of all runs and excluded obvious outlier measurements from further analyses. Next, the intensity of nine marker compounds in the QC measurements was compared, and their relative standard deviation was calculated to assess overall measurement stability: *m/z* 434.2, 324 s (**1**), *m/z* 434.2, 334 s (**2**), *m/z* 288.1, 341 s (**3**), *m/z* 352.2, 292 s (**4**), *m/z* 381.1,

56 s, m/z 319.1, 314 s, m/z 163.0, 317 s, m/z 303.0, 364 s, m/z 349.1, 463 s (see Suppl. Table 1).

For relative quantification of the PAs, their peak intensities were extracted by using the Compass PathwayScreener 1.0 (Bruker Daltonics). The automatization parameters for data evaluation were an extracted ion chromatogram width of ± 5 mDa, with the mSigma tolerance being set to 1000 and the area and intensity threshold being set to 100 counts. The minimum peak valley was set to 1 % by a sensitivity level of 99 % and a smoothing width of 0.2. We extracted a complete list of intensities for every found peak and manually checked for their accuracy. For relative quantification, we divided the peak intensities with that of the according internal standard **4**. Finally, the mean was calculated from the technical replicates (referred to, in this study, as the relative intensity). ANOVAs and the Bartlett test were performed by using the R stats package version 3.2.3 (R Core Team, 2013). Normal distribution of residuals was checked visually.

Bucket tables were calculated in ProfileAnalysis 2.1 (Bruker Daltonics) by using the following parameters: mass range of m/z 100 to m/z 750, retention time range 20 s to 720 s, bucket-pair calculation by advanced bucketing with a time range of 5 s and a mass range of 5 mDa. Data were normalized using the quantile algorithm. PCAs were calculated in R by using the functions 'prep' for Pareto scaling (pcaMethods package, Stacklies et al., 2007) and 'prcomp' principal component calculation (R stats package, R Core Team, 2013). The R stats package was also used to calculate MANOVAs.

Thesinine-conjugate screening in *L. perenne* and *L. multiflorum* cultivars

Plants of two *L. perenne* and four *L. multiflorum* cultivars were grown in the greenhouse (June/July 2015) and sampled 7 weeks after germination. Sample preparation, data acquisition, and data analysis largely followed the same protocol as described above. Plant extracts were diluted 1:5 (50 μ l + 200 μ l), and the final concentration of the internal standard retrorsine was 0.32 μ g/ml. All samples and QC samples were analyzed via LC-MS in three batches that were each measured within 24 hours of thawing. For the assessment of measurement stability the following signals were chosen: m/z 381.1, 57 s, m/z 119.1, 124 s, m/z 163.0, 319 s, m/z 333.1, 340 s, m/z 437.2, 603 s in addition to m/z 434.2, 324 s (**1**), m/z 434.2, 334 s (**2**), and the internal standard m/z 352.2, 292 s (**4**).

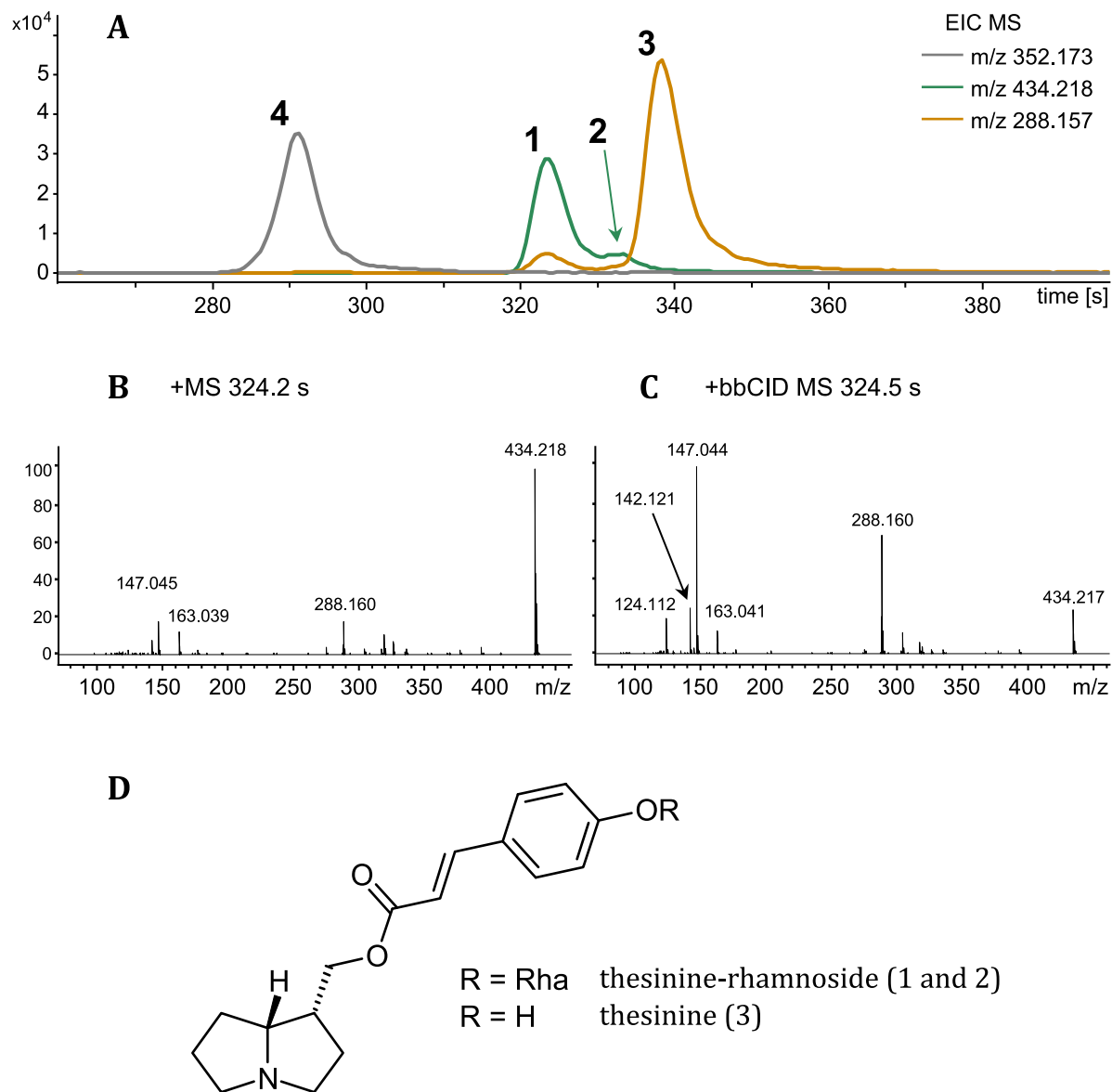
Screening for other pyrrolizidine alkaloids in various Pooideae species

The Pooideae species selected included all the *Lolium* and *Festuca* species and cultivars) plus *D. glomerata*, *H. lanatus*, *T. aestivum*, and *B. distachyon* (see 4.2 and Suppl. Table S2). Powder from various plant parts and growth stages of a specific species were mixed in approximately same proportions and were extracted and prepared based on the protocol of Betteridge et al. (2005). In short, the powdered plant material was extracted with 0.1 N H₂SO₄ (10 ml per 0.1 g freeze-dried material and approx. 2 ml per 0.1 g fresh material). Heliotrine (0.8 µg, Latoxan) and retrorsine (0.8 µg, Sigma-Aldrich) were added as controls to the samples. The extracts were applied to strong cation exchange solid phase extraction (SCX-SPE) columns (Phenomenex) that were treated according to the manufacturer's instructions. Alkaloids were eluted with ammoniated MeOH, which was subsequently evaporated to dryness. The residue was resolved in 2.0 ml MeOH and again diluted 1:20 (10 µl + 190 µl) and analyzed via LC-MS (AutoMSn mode). The LC-MS chromatograms were manually screened for potential necine base fragments: *m/z* 156 , *m/z* 138 , *m/z* 120 for retronecine (and isomers), *m/z* 168 , *m/z* 150 for otonecine, and *m/z* 142, *m/z* 124 for isoretronecanol (and isomers) (Colegate et al., 2005; Zhou et al., 2010).

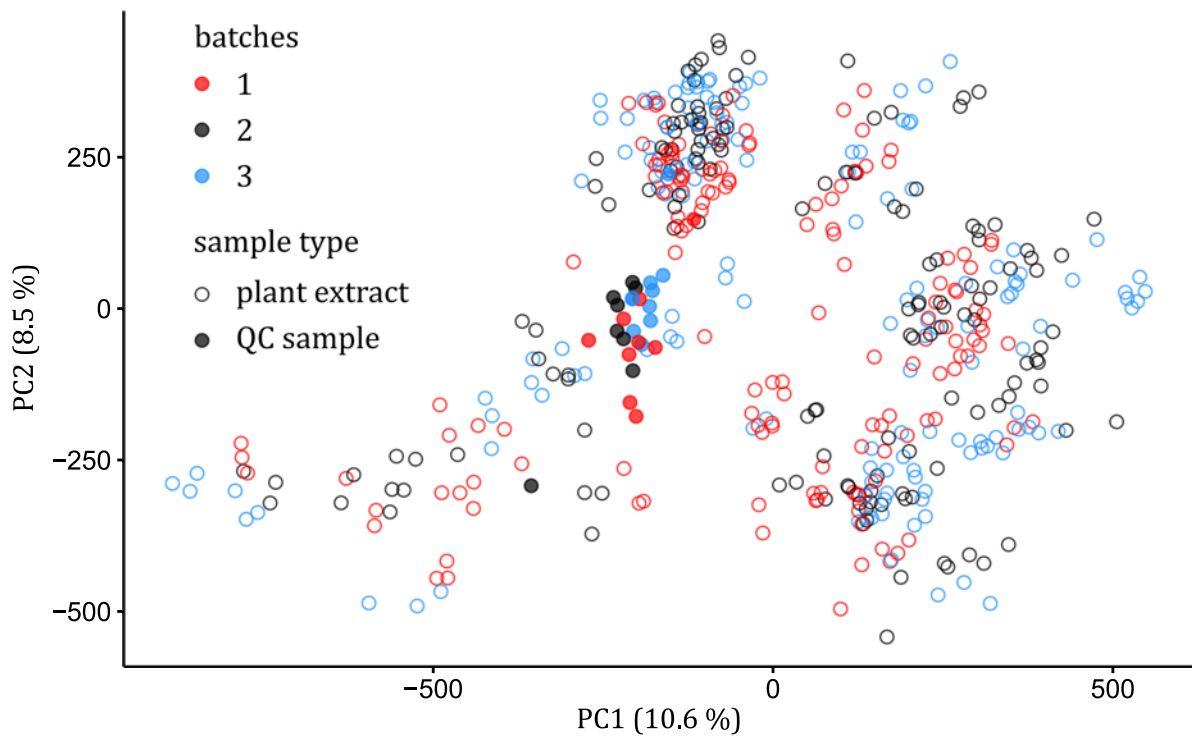
Acknowledgments

We thank the staff of the Botanical Garden Kiel for their help with grass cultivation, Nadja Arens and Hans-Peter Mock for their support during initial LC-MS experiments, Bernhard Ingwersen and Astrid Schirmer at the NPZ-Lembke for helpful hints and seeds of *L. perenne* and *L. multiflorum* cultivars, and Joshka Kaufmann for sharing his R skills.

Supplementary material



Suppl. Fig. S1: Thesinine and thesinine-rhamnosides detected in a *Festuca arundinacea* extract via LC-QTOF-MS analysis. Extracted ion chromatograms (EICs, smoothed, **A**) depicting the elution profiles of the thesinine-rhamnosides (m/z 434.218, **1** and **2**), thesinine (m/z 288.157, **3**), and the internal standard retrorsine (m/z 352.173, **4**). Mass (MS, **B**) and broadband fragmentation spectrum (MSn, **C**) recorded at retention time 323 s (asterisk). **D** Structure of the E/Z-thesinine-conjugate



Suppl. Fig. S2: Principal component analysis of the complete LC-MS dataset of the *Lolium-Festuca* screening. MANOVA of PC1 and PC2 scores reveals that there are no significant differences between batches ($p = 0.31$)

Supplementary Table S1: Stability of marker compounds in quality control (QC) samples throughout the LC-QTOF-MS analysis for the *Lolium-Festuca* screen

Marker compound	Mean intensity	Relative standard deviation	Maximum RT difference
<i>Lolium-Festuca</i> species screen for thesinine and thesinine-rhamnoside			
Thesinine-rhamnoside (1) 434.2 m/z, 324 s	10224	7.0%	3.1 s
Thesinine-rhamnoside (2) 434.2 m/z, 334 s	4699	8.8%	1.6 s
Thesinine (3) 288.1 m/z, 341 s	3966	7.9%	1.3 s
Retrorsine (4) 352.2 m/z, 292 s	36256	11.3%	1.8 s
381.1 m/z, 56 s	23103	7.2%	1.6 s
319.1 m/z, 314 s	7951	13.6%	3.5 s
163.0 m/z, 317 s	19540	5.7%	6.6 s
303.0 m/z, 364 s	11027	11.2 %	5.6 s
349.1 m/z, 463 s	7178	8.2%	3.6 s
<i>Lolium</i> cultivar screen for thesinine-rhamnoside			
Thesinine-rhamnoside (1) 434.2 m/z, 326 s	1316	12.6%	5.5 s
Thesinine-rhamnoside (2) 434.2 m/z, 335 s	1117	8.2%	5.1 s
Retrorsine (4) 352.2 m/z, 294 s	29922	5.9%	5.0 s
381.1 m/z, 57 s	17559	5.3%	2.5 s
119.1 m/z, 124 s	2578	9.5%	12.1 s
163.0 m/z, 319 s	17151	3.1%	6.0 s
333.1 m/z, 340 s	28526	6.6%	7.6 s
437.2 m/z, 603 s	12466	15.6%	3.2 s

Supplementary Table S2: All Pooideae samples with sources used in this study

Species	Additional information	Source [□]
<i>Lolium-Festuca</i> species complex		
<i>Lolium perenne</i>	Fennema †	NPZ-Lembke
<i>Lolium perenne</i>	Chicago †	NPZ-Lembke
<i>Lolium multiflorum</i>	Lema †	NPZ-Lembke
<i>Lolium multiflorum</i>	Pilgrim †	NPZ-Lembke
<i>Lolium multiflorum</i>	Fabio †	NPZ-Lembke
<i>Lolium multiflorum</i>	Lipo †	NPZ-Lembke
<i>Lolium remotum</i>	GR 11839 †	Genebank Gatersleben
<i>Lolium rigidum</i>	GR 11848 †	Genebank Gatersleben
<i>Lolium temulentum</i>	Ceres [‡] , ABY-Ba 13539-1999U [†]	IBERS Aberystwyth
<i>Lolium temulentum</i>	GR 11902 † (Albania [°])	Genebank Gatersleben
<i>Lolium temulentum</i>	GR 11860 † (Spain [°])	Genebank Gatersleben
<i>Festuca arundinacea</i>	GR 11566 †, Arola †	Genebank Gatersleben
<i>Festuca pratensis</i>	GR 9987 †	Genebank Gatersleben
<i>Festuca rubra</i> ssp. <i>rubra</i>	GR 2197 †, Engina †	Genebank Gatersleben
Additional Pooideae species		
<i>Brachypodium distachyon</i>		Botanical Garden Kiel
<i>Dactylis glomerata</i>		Botanical Garden Kiel
<i>Holcus lanatus</i>	GR 4224 [†]	Genebank Gatersleben
<i>Holcus mollis</i>	GR 7269 [†]	Genebank Gatersleben
<i>Triticum aestivum</i> var. <i>lutescens</i>	Florida [‡] , TRI 29469 [†]	Genebank Gatersleben
<i>Triticum aestivum</i> var. <i>lutiflatum</i>	Chinese Spring [‡] , TRI 12922 [†]	Genebank Gatersleben
[□] full names: Genebank Gatersleben - Genebank Gatersleben of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany; NPZ-Lembke - Norddeutsche Pflanzenzucht Hans-Georg Lembke KG, Holtsee, Germany; IBERS Aberystwyth - Genebank IBERS, Aberystwyth University, Wales; Botanical Garden Kiel - Botanischer Garten der Christian-Albrechts-Universität zu Kiel, Germany		
[†] accession number; [‡] cultivar; [°] country of origin		

Chapter 3: Further analyses on grass HSS and DHS

Background

Homospermidine synthase-encoding genes have evolved twice within the grass family from two independent *dhs* gene duplication events (see Chapter 1). These grass *hss* genes therefore belong to either of two gene clusters: the “Pooideae *hss*” and “Panicoideae *hss*”, respectively. Chapter 1 of this thesis encompasses in-depth analyses of the evolutionary history of the grass *dhs* gene duplicates. The presented study showed, amongst other things, that the observed change in enzyme function from a DHS to an HSS was accompanied by temporal changes of selection pressure in both the Pooideae and Panicoideae *hss* lineages. This means that for a short period of time nonsynonymous substitutions had become more permissible or even might have even been positively selected for in order to promote the functional change from *dhs* to *hss*. While the functional divergence could be experimentally verified, it is unclear which of the substitutions were responsible for the change in enzyme function in either grass *hss* lineage.

To identify amino acids sites that are potentially important for the functionality of a DHS or HSS in grasses, an assortment of sequence-based analyses focusing on codon/amino acid sites were performed and the output is presented in the following. These results should serve as a basis for future investigations to experimentally verify which amino acid residues in grass HSS and DHS protein sequences are instrumental for either enzyme’s function.

Results

Nonsynonymous substitution sites of grass HSS and DHS

A first step in the direction to identifying which sites be important for HSS function is to determine which codon sites might have experienced nonsynonymous (i.e. amino acid changing) substitutions in the course of both Pooideae and Panicoideae *hss* evolution. As part of codon substitution model-based calculations performed in codeml (Yang, 2007)

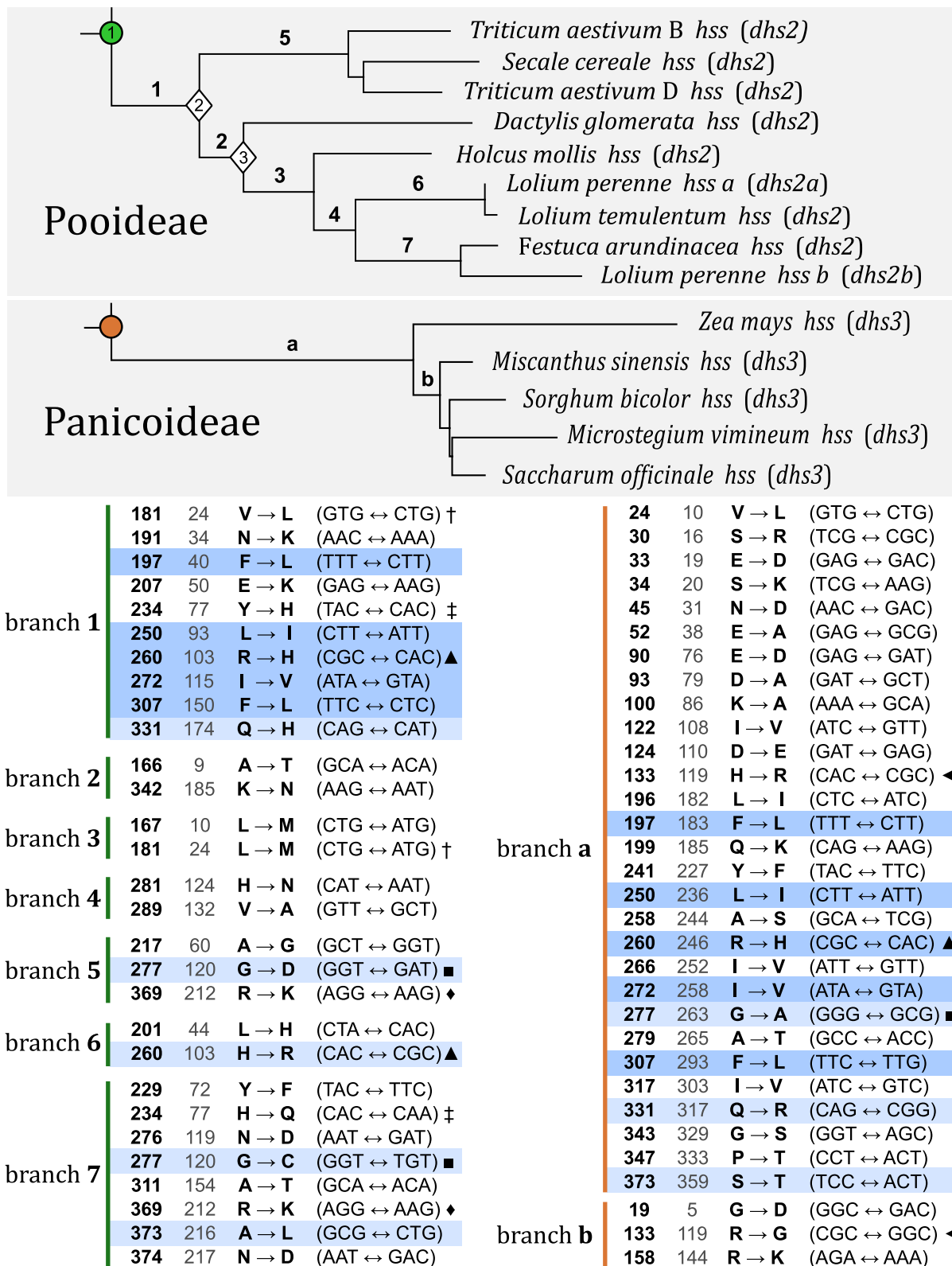


Fig. 1: Nonsynonymous substitutions occurring during grass *hss* evolution

Displayed only for internal branches of Pooideae and Panicoideae *hss* lineages, site position in black/bold according to *Brachypodium distachyon* DHS, site positions in grey corresponds to site position of the specific alignment used for the analyses, blue parallel substitution, light blue marked sites affected by mutations in both lineages. Black symbols denote sites which have experienced more than one substitution in one of the *hss* lineages.

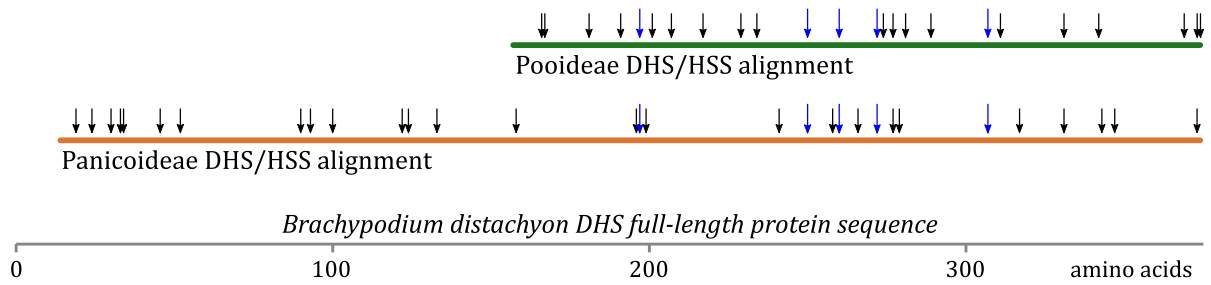


Fig. 2: Distribution of amino acid substitutions of grass HSS evolution

Blue arrows mark parallel substitutions shared between Pooideae and Panicoideae HSS

ancestral coding sequences are reconstructed for every node of a given sequence tree, making it not only possible to infer which site experienced mutations but also when (meaning on which branch) specific substitutions have taken place.

Figure 1 lists all nonsynonymous mutations that can be placed on internal branches of the Pooideae *hss* and Panicoideae *hss* gene clusters (duplication events are marked green and orange, respectively). Since their respective duplication events the genes have experienced a minimum of thirteen nonsynonymous substitutions during Pooideae *hss* evolution over an alignment length of 217 amino acids and no fewer than twenty-five changes on 360 amino acids during Panicoideae *hss* evolution. As only the internal branches were considered the actual number of amino acid changes in each *hss* lineage should be higher and vary between each extant HSS-encoding genes. Looking at the distribution of the predicted nonsynonymous substitution sites reveals no special regions of accumulation in both the Pooideae and Panicoideae *hss* lineages (figure 2).

To determine which sites might have experienced positive selection in the course of functional divergence after the gene duplication event, a Bayes empirical Bayes (BEB) procedure was applied to *dhs* and *hss* sequence data to calculate the posterior probabilities of a codon site having experienced adaptive mutations (Yang et al., 2005). The results for branch 1 of the Pooideae *hss* cluster and branch a of the Panicoideae *hss* cluster (see figure 1) are summarized in table 1. The calculation output indicates that during Pooideae *hss* evolution six specific amino acid sites might have experienced positive selection whereas the results of the BEB for Panicoideae *hss* evolution could only identify one site which might have been under positive selection. However, a

Table 1: BEB inference of amino acid sites under positive selection

branch 1 Pooideae <i>hss</i>				branch a Panicoideae <i>hss</i>			
site		posterior probability*		site		posterior probability*	
191	(34)	N	0.989	100	(86)	K	0.964
197	(40)	F	0.990				
207	(50)	E	0.990				
250	(93)	L	0.994				
272	(115)	I	0.989				
307	(150)	F	0.991				

* Listed are only those sites which displayed a posterior probability > 0.989

complete BEB analysis includes branch-site model based likelihood ratio tests, the so-called branch-site tests for positive selection (Zhang et al., 2005). These had already been performed for these datasets (see fig. 2 and fig. 3 in Chapter 1). And the results could not confirm that positive selection had actually affected the tested branches; only the test which detects positive selection or relaxed functional constraints, without distinguishing either, was significant for either of the *hss* branches (LRT2, fig. 2 and fig. 3 in Chapter 1).

Another approach to identifying amino acid residues, which are potentially important for HSS function, was to identify sites which have experienced substitutions in both the Pooideae and Panicoideae *hss* evolutionary lineages. The comparison has revealed that eight codon sites have experienced nonsynonymous substitutions in *hss* evolution of both grass subfamilies (marked blue, figure 1). It is of note that in the case of Pooideae *hss* evolution the majority of these substitutions took place soon after the gene duplication event (i.e. on branch 1, see figure 1). In addition to this, five out of these six sites have undergone the exact same amino acid change (parallel substitutions) without having necessarily experienced the same nucleotide substitution (darker blue, figure 1). Out of the sites that have experienced a mutation on branch 1 one substitution was reversed back to its original state later on branch 6 (site 260, Pooideae *hss* cluster, fig. 1). This, however, has only affected two of the extant coding sequences included in this sequence dataset, *Lolium perenne dhs2a* and *L. temulentum dhs2*. Furthermore, the

Table 2: Amino acid distribution at select sites across flowering plants

Poaceae	Asteraceae		Boraginales		Convolvulaceae		Monocots ^a		
	HSS evolution	DHS	HSS	DHS	HSS	DHS	HSS	DHS	HSS
Sites which have experienced parallel substitutions in Poaceae <i>hss</i> lineages									
197	F → L	L/F	F	F	L	F	F	L/F	L
250	L → I	L	I/L	L	I	L	I/L	L	I
260	R → H	R	K/S/T	R	R/N	R	N/K/T	R/Q	R
272	I → V	I	V	I	A/V	I	I/V	I	V
277	G → D/C/A	S	N	G/S	N/D	G	D/G	G/S	E
307	F → L	M	M	M	M	M	M	L/M/F	L
331	Q → R/H	R	R	R/C	R	R	R	R/Q	R
373	A → L S → T	E/S/T/-	A/T	S	N/K/S	S	R/F	A/R/G	A
Positively selected sites in Pooideae <i>hss</i> lineage according to BEB results									
191	N → K	D	D	D	D	D	D/N	N/D	D
207	E → K	Q/K	Q/K	E/K	Q/K	Q	Q	E/K	E
Positively selected sites in Panicoideae <i>hss</i> lineage according to BEB results									
100	K → A	A/T	T/-	A	A/L/K	A	A	A/S	A
Number of proteins	4	4	3	3	5	5	4	2	

^a Monocots other than grasses

list contains two sites that have been subjected to substitutions on two branches within the Pooideae *hss* cluster (black diamonds and squares, branches 5 and 7, fig. 1). Out of these two sites, site 369 has undergone parallel substitutions (R → K) while at site 277, which was also mutated during Panicoideae evolution, changes from a glycine (G) into a total of three different amino acids in the branches a, 5, and 7 (see figure 1).

For further assessment of each site in the compiled list of nonsynonymous substitutions, amino acid frequencies at homologous sites in HSS and DHS protein sequences from various Asteraceae, Boraginales and Convolvulaceae species as well as monocot species other than grasses were determined (table 2). At some homologous sites the amino acid patterns closely match the parallel amino acid changes observed in the grass *hss* lineages; others, however, display completely different amino acid substitution patterns.

Table 3: DHS and HSS enzyme activities for mutated *S. bicolor* DHS

<i>S. bicolor</i> DHS2	DHS activity	HSS activity	
S30R	42 pkat/mg	486 pkat/mg	(experiment 1)
S34K	40 pkat/mg	382 pkat/mg	(experiment 1)
I272V	37 pkat/mg	545 pkat/mg	(experiment 2)
S30R S34K I272V	23 pkat/mg	367 pkat/mg	(experiment 2)
wild type *	34 pkat/mg	490 pkat/mg	(experiment 1)
wild type	37 pkat/mg	472 pkat/mg	(experiment 2)
<i>S. bicolor</i> HSS *	1.5 pkat/mg	488 pkat/mg	

* result have already been presented and discussed in chapter 1

For example at site 250 we find a leucine (L) in DHS protein sequences across all plant species and, with some exceptions, isoleucine (I) at the homologous site in HSS protein sequences (in total 9 out of 14 sequences). This substitution matches the leucine to isoleucine substitution (L → I) observed in both grass *hss* lineages at this position. At site 272 all DHS contain an isoleucine (I) while most HSS sequences (12 out of 14) exhibit a valine (V) residue, matching the parallel substitution at this site in grass HSS evolution. Another striking amino acid pattern can be found, for example, at site 197 where the codon triplets either code for leucine (L) or phenylalanine (F) but without any obvious association of either residue to HSS and DHS function. At site 260 arginine (R) residues seem to be conserved in DHS across flowering while for HSS we find variety of different amino acids at this site even within the HSS sequences of one taxon.

One way to test the significance of a specific nonsynonymous substitution to an enzyme's functionality is to artificially introduce the mutation into a wild type (WT) gene and check the effect it has on enzyme activity. For initial testing, three sites that had experienced substitutions in the course of Panicoideae HSS evolution were chosen for site-directed mutagenesis in *Sorghum bicolor* (*Sbi*) DHS2 (see table 3). Three plasmid constructs, each containing an *Sbi*DHS2 open reading frame with a different mutated site, in addition to a fourth construct containing a triple mutant were used for heterologous expression. The proteins were purified and subjected to enzyme activity assays to measure DHS and HSS activity (in two separate experiments, table 3). The

specific enzyme activities of the mutants did not show great deviations from the wild type activity and is nowhere near close to *S. bicolor* HSS activity (cf. Chapter 1).

Additional results from ancestral Pooideae HSS investigations

As part of the study presented in Chapter 1 ancestral DHS and HSS were reconstructed from nodes within the Pooideae *hss* cluster and were then tested in *in vitro* enzyme activity assays to detect functional divergence in the derived proteins. In earlier phases of the study *dhs* and *hss* sequence information for *Triticum aestivum* and its close relatives *T. urartu* and *Aegilops tauschii* was not yet available. At this time the ancestral enzyme at node 3 (fig. 1) was reconstructed based on the then available data (at which time the node was also the first branch point within the Pooideae *hss* cluster following the node representing the gene duplication event). There are only four amino acid differences between the full-length sequence of the ancestral proteins 3 and 2. This however does not include the very *N*-terminal region of the reconstructed proteins.

A

```

ancestral protein 3  1  10  20  30  40  50  60
                    MAGDGGGGDGVGSSGGGRN TDA LEGVRAI L K PS ES LDESRS TRIAGADFN DAG LG LAG LLGS L
ancestral protein 3  70  80  90  100  110  120
                    AS TGFQASH LGDAI DVVNQMLDWR LSH EK PS EDC DEA ELDPKYR ESVKCK I F LGF TSNLVSSG
ancestral protein 3  130 140 150 160 170 180
                    IR DV I R F LVQH MV DV I V TTAGGI EEDLIKCL A P TYR G DFS L PG T LLRSKG LNR I GN LL M PND
ancestral protein 3  190 200 210 220 230 240 250
                    NYCKF EKWIM PLLDQMLLEQS TKNVWTPSKVIAR LGK EIND ESSYLYWAHKNNI PVYCPALTD
ancestral protein 3  260 270 280 290 300 310
                    GSIGDMLFCHAVHNPGLI I DIVQDVR LMNGEAIHA TPRK TGVI I LGGG LPKHHI CNANMLRNG
ancestral protein 3  320 330 340 350 360 370 378
                    ADYAVYINTAQEF DGS DSGAH PDEAVSWGKI GSAK PVKVHCDA TIAF PLLVAA TFARRFHGA
  
```

substitutions: nonpolar/hydrophobic polar

B

```

ancestral protein 3  1  10
                    MAGDGGGGDGVGSSGGG ... ← B. distachyon DHS N-terminus
ancestral protein 1/2  MAGEG-----SSGGG ... ← reconstructed
  
```

Fig. 3: Ancestral Pooideae HSS sequence

A, Reconstructed ancestral HSS from node 3 of the Pooideae *hss* cluster. Amino acid differences to the ancestral reconstruction of the protein at the preceding node are marked with colored boxes. B, Comparison of *N*-terminal region of ancestral reconstruction based on different data. Substitutions are in comparison to ancestral protein 2 (see Chapter 1).

For ancestral proteins 1 and 2 the N-terminus was reconstructed by hand without the help of any software (cf. Materials and Methods of Chapter 1) while for ancestral protein 3 the N-terminal region of *B. distachyon* was copied due to lack of sufficient sequence data at that time. The ancestral protein was heterologously expressed, purified and then its HSS and DHS activity was measured as described in Chapter 1. The results are presented in table 4, which also includes the specific enzyme activities for the ancestral proteins presented in Chapter 1. In comparison with the ancestral proteins 1 and 2 the protein from node 3 displays highest specific HSS activity and intermediate specific DHS activity (lying between the ancestral protein 1 and 2).

Table 4: Specific DHS and HSS enzyme activities of an ancestral DHS and HSS proteins specific to the Pooideae *hss* cluster (see fig. 1)

	DHS activity	HSS activity
ancestral protein 1 *	14.9 pkat/mg	90.8 pkat/mg
ancestral protein 2 *	3.2 pkat/mg	124.0 pkat/mg
ancestral protein 3	9.0 pkat/mg	529.0 pkat/mg

* results have already been presented and discussed in Chapter 1
N-terminal regions of ancestral proteins 1 and 2 are identical

Discussion and outlook

The grass HSS evolved twice from two independent DHS gene duplicates within grasses, once within the Pooideae subfamily and another time before the emergence of the Panicoideae subfamily (see Chapter 1). While the previous study established that a switch in enzyme function did indeed take place, it was not discussed which of the substitutions that occurred in the course of grass HSS evolution might have induced this observable change in function. To explore this issue nonsynonymous (i.e. amino acid changing) substitutions that occurred on internal branches of the two independently evolved grass *hss* clusters were identified. Through a variety of approaches these nonsynonymous substitution sites were then evaluated for their potential importance in HSS (and DHS) functionality.

Positive selection acting on a particular site can be interpreted as a strong indicator that this site played an important role in the enzymes newly evolving function. However, using the BEB approach developed by Yang et al. (2005) it was not possible to unambiguously identify sites at which specific substitutions had been positively selected for in either grass HSS lineage (table 1) the reason being that the associated branch-site test for positive selection did not verify that positive selection had actually affected HSS evolution (Chapter 1; Zhang et al., 2005).

Comparing amino acid substitution sites between the two independent grass *hss* lineages as well as a wider comparison of HSS protein sequences from across the flowering plants has shown more promise than the selection analyses. The results revealed that several homologous sites had actually experienced parallel substitutions (fig. 1 and table 2). There is an ongoing debate over whether molecular convergence (i.e. convergent and parallel amino acid substitutions, see Zhang and Kumar, 1997), is a reliable indicator for adaptive evolution or whether molecular convergence is mainly the result of the small number of possible neutral nonsynonymous substitutions (e.g. Rokas and Carroll, 2008; Goldstein et al., 2015; Zou and Zhang, 2015). The HSS/DHS study system has the advantage that observed amino acid substitutions can easily be correlated with enzymatic function. In addition to this, we know of eight parallel evolved *hss* lineages (see Chapter 1, Reimann et al., 2004, Kaltenecker et al., 2013 and Irmer et al., 2015) which can all be included in experiments to correlate specific substitutions with the functional switch from DHS to HSS. It therefore seems extremely likely that at least a few of the parallel substitutions have played a role in causing the functional switch in the gene. This is strongly supported by the fact that all five parallel substitutions shared by the two grass *hss* lineages occurred in an early phase of Pooideae *hss* evolution (branch 1, fig. 1) during which DHS enzyme activity was reduced and HSS activity increased (see Chapter 1).

Further comparison of amino acid substitution sites/amino acid frequencies across various flowering plant taxa revealed several amino acid sites which could potentially be important for enzyme specificity: sites 250 and 272. At these sites the amino acid frequencies in HSS and DHS are (relatively) consistent across all examined flowering plant taxa. With a few exceptions site 250 in DHS and HSS protein sequences display an isoleucine (I) and a leucine (L) residue, respectively (table 2). At site 272 most flowering plant HSS possess a valine (V) residue where all the examined DHS display an isoleucine

residue (see table 2). Interestingly, the side chains of these amino acids are all nonpolar and it is maybe difficult to imagine how exchanging these amino acids for one another could greatly contribute to changes in enzymatic function. Introducing the according mutation into one of the *S. bicolor* DHS orthologs (DHS2, see table 3) did not induce a noticeable change in enzyme specificity (table 3). However, this I to V substitution at site 272 had already been investigated once before in a previous study conducted by Kaltenecker et al. (2013). In their study they found that this specific site had been under significant positive selection in Convolvulaceae-specific HSS evolution. Substituting the according amino acid at this site in a Convolvulaceae wild type DHS and subjecting this construct to enzyme activity assays confirmed that this substitution indeed has contributed to the switch from a DHS to an HSS at least in this taxon (Kaltenecker et al., 2013).

There are several possible reasons why mutating the *S. bicolor* DHS2 at this position did not yield any positive results other than the possibility that the residue at this site is actually not involved in determining enzyme specificity in this particular protein. For example, changes in enzyme activity in the mutated *S. bicolor* DHS could have been too small to be detectable by the methods employed. Another possible reason is that the I272V mutation might have different effect in different genetic backgrounds, a phenomenon known as epistasis. The substitution might need to be “paired” with another substitution in the same or another gene to take an effect on enzyme activity. Therefore, site 272 should not be disregarded and instead be investigated further by trying out different experimental approaches. The same goes for the L250I substitution which has not been paid attention to in any studies investigating DHS or HSS function. One possible approach would be to test the effects of mutating this amino acid site in an HSS. A better approach would be to test how these substitutions affected those proteins in which they actual occurred. This means reconstructing ancestral proteins, which occurred at the relevant evolutionary time point (e.g. the time of the gene duplication event) and testing the effect different amino acid residues have on this protein construct.

Although the particular sites and amino acid substitutions discussed above show great potential to be of significance for enzyme functionality the other parallel and unique substitutions are not to be disregarded. With the various *dhs* and *hss* lineages having evolved independently for long periods of time every substitution might be subject to

epistatic effects. To unambiguously identify which substitutions were necessary for the evolution of the new function, mutational pathways connecting the (ancestral) DHS and HSS should be explored. Epistatic interactions could potentially allow only a specific succession of substitutions and testing and analyzing the effects of only single substitutions could lead to misinterpretations concerning their function (Anderson et al., 2015; Bridgham et al., 2009; McKeown, 2015; Smith and Rausher, 2011). Based on the data presently available the Pooideae HSS cluster lends itself well to these types of investigations. Phylogenetic resolution is higher here than in the Panicoideae-specific HSS cluster; branch 1 of the Pooideae-specific HSS cluster has experienced 10 amino acid substitutions while branch a of Panicoideae HSS evolution has experienced a total of 17 substitutions over the same sequence length.

Another aspect which has not been specifically investigated here but should be discussed at least briefly is the relevance of the DHS and HSS *N*-terminal region. Even between closely related grass species the *N*-termini of the DHS homologs display great variability in length and sequence so that finding and aligning homologous sites is near impossible for this region. The first 10-15 *N*-terminal amino acids have therefore been disregarded in phylogenetic sequence analyses (here and in Chapter 1). However, analyzing the crystal structure the human DHS has led to the realization that the *N*-terminus contains a ball-and-chain motif essential for catalysis (Umland et al., 2004). This was supported by experiments performed with *N*-terminal deletion mutants which exhibited reduced or even no enzyme activity (Kang and Chung, 1999; Umland et al., 2004). The observations gained from experiments conducted with ancestral grass HSS (and DHS) sequences fit those assumptions. Ancestral protein 3 displays a higher DHS activity than protein 2, which is evolutionary closer to the ancestral DHS (i.e. ancestral protein 1). This could be explained by the fact that the *N*-terminus of ancestral protein 2 was derived from the (presumably well adapted) *B. distachyon* DHS. Therefore, only enzyme activities of ancestral proteins 1 and 2 are suitable for comparisons and can be taken into account for conclusions about HSS evolution. When investigating HSS and DHS biochemistry in future special care should be taken where the *N*-terminus is concerned, especially in ancestral sequence reconstruction experiments.

It is quite clear that there is much left to examine regarding HSS and DHS biochemistry. The results and discussion of this chapter present a framework in which future studies should be conducted.

Materials and Methods

All sequence analyses were performed using the same coding sequence alignments that the selection analyses in chapter 1 are based on. In short, selected full-length *dhs* and *hss* coding sequences of Pooideae and PACMAD (which includes the Panicoideae), respectively, were aligned while being checked for alignment confidence scores as well as potential recombination break points (Pond et al., 2006; Sela et al., 2015, for details see the materials and methods section of chapter 1). To facilitate the comparison of the results of the various analyses, codon/amino acid sites are numbered according to homologous sites on the full-length *Brachypodium distachyon* DHS sequence (accession: XM_003559868).

Inferring substitutions in Panicoideae and Pooideae *dhs/hss* evolution

Coding sequence alignments and the according maximum likelihood tree were used for 1-ratio branch model estimations with integrated ancestral sequence reconstructions (control file: model = 0, NSsites = 0, RateAncestor = 1) (Yang et al., 1995). In the course of these calculations the program gives out a list of codons which have been substituted for each branch of the tree (in “supplemental results file rst”; Yang, 2007).

Identifying sites under positive selection in *hss* evolution

To determine if and which substitutions might have been positively selected for during *hss* evolution in grasses, a Bayes empirical Bayes procedure (BEB) implemented in the codeml branch-site model calculations (model = 2, NSsites = 2) was applied (Yang et al., 2005; Zhang et al., 2005; also see the material and methods section of Chapter 1).

Determining amino acid frequencies at specific sites in angiosperm DHS and HSS

Protein sequences of various plant DHS and HSS (accessions listed in table 4) were aligned using the MAFFT algorithm (default parameters implemented in Geneious, Katoh and Standley, 2013). Amino acid frequencies for specific homologous sites in DHS and HSS for each taxon were counted based on the alignment and using a custom PERL script.

Table 4: List of *dhs* and *hss* sequence accessions from various flowering plant taxa

Asteraceae	
<i>Senecio jacobaea hss1</i> gene	AJ704850
<i>Lactuca sativa dhs</i> gene	AY731231
<i>Eupatorium cannabinum dhs1</i> gene	AJ704841
<i>Eupatorium cannabinum hss1</i> gene	AJ704842
<i>Senecio vernalis dhs1</i> gene	AJ238622
<i>Senecio vernalis hss1</i> gene	AJ238623
<i>Petasites hybridus dhs1</i> gene	AJ704846
<i>Petasites hybridus hss1</i> gene	AJ704847
Boraginales	
<i>Cynoglossum officinale dhs1</i> gene	AJ704839
<i>Cynoglossum officinale hss1</i> gene	AJ704840
<i>Heliotropium indicum hss1</i> gene	AJ704843
<i>Heliotropium indicum dhs1</i> gene	AJ704844
<i>Symphytum officinale hss1</i> gene	AJ704851
<i>Symphytum officinale dhs1</i> gene	AJ704852
Convolvulaceae	
<i>Ipomea neei dhs1</i> gene	HF911504
<i>Ipomea neei hss1</i> gene	HF911505
<i>Ipomea meyeri dhs1</i> gene	HF911519
<i>Ipomea meyeri hss1</i> gene	HF911520
<i>Ipomea alba dhs1</i> gene	HF911522
<i>Ipomea alba hss1</i> gene	HF911523
<i>Ipomoea hederifolia hss1</i> gene	HF911521
<i>Ipomoea purpurea dhs2</i> gene	HF911510
<i>Merremia quinquefolia dhs1</i> gene	HF911517
<i>Merremia quinquefolia hss1</i> gene	HF911518
Monocotyledons	
<i>Elaeis guineensis dhs</i> transcript variant X1	XM_010915973
<i>Phoenix dactylifera dhs</i> transcript variant X2	XM_008811668
<i>Phalaenopsis</i> hybrid cultivar <i>hss1</i> gene	AJ704848
<i>Phalaenopsis amabilis dhs1</i> gene	AM411616
<i>Phalaenopsis amabilis hss1</i> gene	FM201299
<i>Musa acuminata</i> subsp. <i>malaccensis dhs</i> gene	XM_009405857

Site-directed mutagenesis and enzyme activity assays

Mutations were introduced into an *S. bicolor dhs2* expression construct (see chapter 1, here called the wild type) by site-directed mutagenesis-PCR. Mutated plasmids were amplified using a proofreading polymerase (Phusion High-Fidelity or Pfu Polymerase); primer pairs used for each mutant are listed in table 5. After the amplification reaction the samples were treated with the restriction enzyme DpnI (incubation at 37°C for about 1 h) to digest the methylated and unmutated template DNA. For plasmid propagation Top10-*E. coli* cells were transformed with the reaction product and single-clone colonies were picked for plasmid isolation. Mutation success was verified by Sanger sequencing (MWG, Eurofins Genomics). Proteins were expressed and purified before being evaluated in enzyme activity assays performed as described in Chapter 1. In short, overexpressed polyhistidine-tagged proteins were affinity purified and subsequently subjected to “DHS” and “HSS” activity assays to determine at which rate the mutated enzymes (and the wild type) catalyze the modification of *Zea mays* eIF5a or the production of homospermidine, respectively.

Table 5: Primer pairs used for site-directed mutagenesis

Name	Sequence (5' to 3')	Mutation
P1-forward	AAGCCGTCGGAG <u>CGG</u> CTCGATGAGTCGC	S30R
P1-reverse	GCGACTCATCGAG <u>CCG</u> CTCCGACGGCTT	
P2-forward	GAGTCGCTCGATGAGA <u>AG</u> CGGTTACGAGGATCG	S34K
P2-reverse	CGATCCTCGTGAACCG <u>CTT</u> CTCATCGAGCGACTC	
P3-forward	GACATTGTACAAGATG <u>TAC</u> GGCTGATGAATGGGGAAGC	I272V
P3-reverse	GCTTCCCCATT <u>CAT</u> CAGCCGT <u>AC</u> TCTTGTACAATGTC	
P4-forward	AAGCCGTCGGAG <u>CGG</u> CTCGATGAGA <u>AG</u> CGGTTACCG	S30R and S34K
P4-reverse	CGTGAACCG <u>CTT</u> CTCATCGAG <u>CCG</u> CTCCGACGGCTT	

Underlined nucleotides indicate sites for directed mutagenesis

Chapter 4: Further LC-MS based investigations on PA biosynthesis in grasses

Background

The question of which enzymes are involved in plant PA biosynthesis has been investigated for many years and is still underway (Hartmann and Witte, 1995; Sievert et al., 2015). Various enzymes have been proposed as potential candidates, which catalyze the reaction steps following the synthesis of homospermidine. One of the enzyme classes which was suggested to be involved are the diamine oxidases (DAO). This was deduced from detailed inhibitor studies conducted using the DAO inhibitor β -HEH (β -hydroxyethylhydrazine) (Böttcher et al., 1993; Hartmann and Witte, 1995). Since the HSS has been identified from PA-producing Pooideae and it was established that the metabolic pathway evolved independently in this grass subfamily, the next issue to address is whether a DAO is also involved in the biosynthesis of PAs in this taxon.

Results and Discussion

To test which effect DAO inhibition has on the metabolome of a PA-producing grass an experiment was conducted in which tall fescue (*Festuca arundinacea*) was germinated and grown in the presence of the inhibitor molecule β -HEH. As a negative control fescues were grown in parallel without the inhibitor.

As a first observation, the DAO-treated plants display a slower growth rate so that the seedlings were considerably smaller seven days after sowing (1-2 cm seedling length when treated with the inhibitor, 3-4 cm of the controls, fig. 1). *Zea mays* seedlings, which were grown in parallel and under the same conditions as the fescues, displayed similarly stunted growth when treated with β -HEH (fig. 1). It is not very surprising that the inhibitor treatment produced the same effect in these not closely related grass species (*F. arundinacea* and *Zea mays* belong to two different grass subfamilies Pooideae and

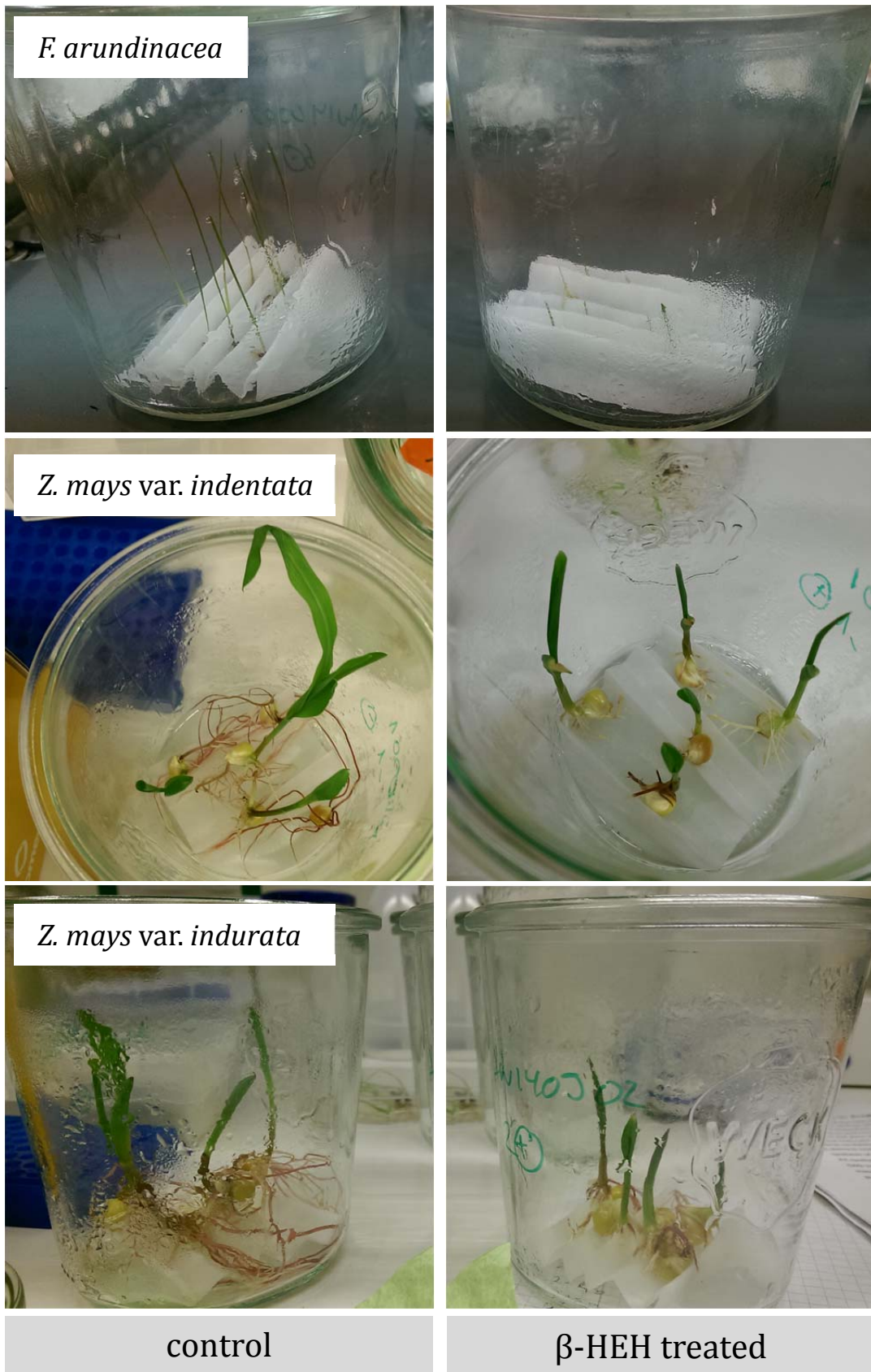


Fig. 1: Germination and growth of various grasses under the influence of β -HEH

Panicoideae, respectively). Amine oxidases are known to be involved in several different plant processes including plant growth and development (reviewed e.g. by Cona et al., 2006). Plants exposed to a general DAO inhibitor such as β -HEH should be affected in any case and the observed effect is not necessarily related to PA biosynthesis.

To directly check which effects the β -HEH treatment has on the tall fescue metabolome, fescues of one treatment were pooled and methanolic extracts were subjected to LC-MS analysis. Since the extraction volume was adjusted according to the plants' dry biomass the base peak chromatograms (BPCs) of the LC-MS analyses are expected very similar in the case that the inhibitor treatment had no effect on the tall fescue's metabolic processes. However, as can be seen in figure 1, several of the prominent peaks of the BPCs differ in intensity or are completely absent in either of the chromatograms. Looking specifically at the levels of the PAs we find on the one hand highly reduced levels of thesinine-rhamnoside but on the other hand an increase in total thesinine counts in comparison to the control measurement (fig. 2 and table 1). While the decrease of thesinine-rhamnoside is expected and fits our expectation of finding reduced PA-levels after DAO-inhibition the results for the aglycone thesinine are in contradiction to this hypothesis. For comparison, the levels of the diazaphenanthrene alkaloid perloline (Cao et al., 2008) also increased slightly in the β -HEH treated sample. In Adding together the peak areas of thesinine and thesinine-rhamnoside for each LC-MS analysis as a proxy for the total PA levels in each sample (provided that both compounds experience similar ionization levels by electrospray ionization) we find an overall slight decrease (table 1).

When interpreting these data previous findings about the variability in PA levels in grasses need to be taken into consideration. As we have learned in Chapter 2 the levels of both thesinine and thesinine-rhamnoside may vary greatly even between individual plants of one species which raises the question whether the observed differences are due to this variability. However, for this experiment around twelve tall fescue seedlings were pooled of each treatment which should compensate, at least partly, the interspecies variability. Nevertheless, there is an obvious need for more experimental work before the role of DAOs in grass PA biosynthesis is fully elucidated. For one, instead of only assessing PA levels after DAO inhibition also levels of homospermidine need to be measured (see e.g. Böttcher et al., 1993). For final confirmation that and

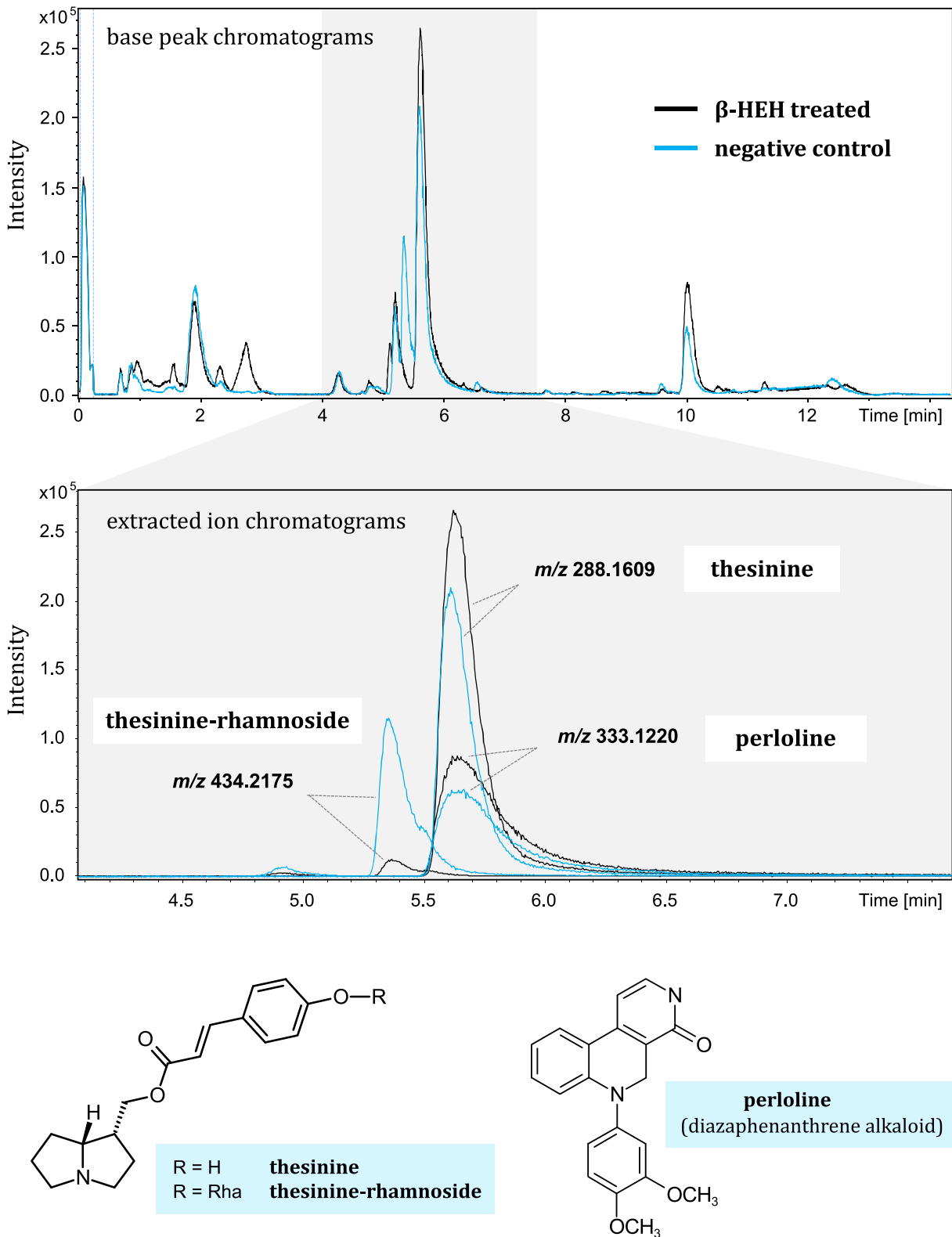


Fig. 2: LC-MS chromatograms of β -HEH treated and control tall fescue seedlings

which (β -HEH-sensitive) DAO is involved in grass PA biosynthesis is to unambiguously identify the according gene. This could be achieved by using reverse genetics approaches such as virus-induced gene silencing or CRISPR-Cas9-mediated gene inactivation. Directly investigating the effect the down regulation of specific DAOs will be more insightful than generally inhibiting DAOs and thereby causing confounding side effects.

Table 1: Peak area in LC-MS chromatograms of various alkaloids

	thesinine-rhamnoside <i>m/z</i> 434.22 <i>t_R</i> 5.3 -5.5 min	thesinine <i>m/z</i> 288.16 <i>t_R</i> 5.5 – 5.9 min	thesinine and conjugates sum of PA peak areas	perloline <i>m/z</i> 333.12 <i>t_R</i> 5.5 -5.8 min
β-HEH treated	86 426	2 618 606	2 705 032	1 204 632
control	822 277	1 991 668	2 813 945	941 166
relative to control	11 %	131 %	96 %	128 %

Material and methods

Grass seeds were placed on wet filter paper for germination in a Weck jar. In the case of the corn (*Zea mays*) the seeds were first disinfected by immersion in 35% H₂O₂ before being incubated for 2-4 hours in sterile Milli-Q H₂O before germination. Seedlings were watered as needed using either 0.1% β -HEH or water as the negative control. Tall fescues were harvested eight days and the corn 14 days after sowing. Fescue samples were pooled; all plant samples were freeze-dried and then powdered. Methanolic extracts were prepared and measured via LC-QTOF-MS (following the protocol described in Chapter 2).

Table 2: Plant species used including source from which seeds were obtained

<i>Festuca arundinacea</i> “Arola”	Genebank Gatersleben, GR 11566
<i>Zea mays</i> var. <i>indentata</i> “ZHR1”	Genebank Gatersleben, ZEA 3639
<i>Zea mays</i> var. <i>indurata</i> “HHS1”	Genebank Gatersleben, ZEA 3645



Conclusions and perspectives

The research presented in this thesis has explored various different aspects concerning the production of pyrrolizidine alkaloids (PAs) by grass plants (Poaceae). At the outset of this work the literature contained little and only anecdotal information on the occurrence of plant-produced PAs and no information at all on any of the putative enzymes involved in PA biosynthesis for this taxon. This is in stark contrast to the extensive literature available on PA biosynthesis in other flowering plant taxa.

We now have an understanding of how widespread this particular trait is within the grass family. Based on the molecular evolution of the *hss* gene it can be deduced that PA biosynthesis evolved within the grass subfamily Pooideae (Chapter 1). Systematic screening of grasses for PAs revealed that the occurrence of the alkaloid is further limited to specific broad-leaved species of the *Lolium-Festuca* species complex within the Pooideae (Chapter 2). Moreover, the LC-MS based screenings showed that the spectrum of PA structures in these grasses is limited to thesinine and its glycosides were the only PA structures detected (Chapter 2). Having this information at hand will be useful when exploring other aspects of PA biosynthesis in grasses. Several key questions still need answering: for example, which biological function these molecules serve in grasses.

Another finding of the *dhs/hss* evolution analysis was that in the course of Poaceae evolution two of the *dhs* gene duplication events gave rise to two independent *hss* gene clusters (Chapter 1). The significance of the “second” *hss* cluster which is comprised of *hss* sequences exclusively from grasses of the Panicoideae subfamily (e.g. maize and sugarcane), none of which are known to contain any kinds of PAs. However, the grass species *Thelepogon elegans*, which was shown to contain the “unusual” PA thelepogine, belongs to the Panicoideae subfamily as well (Crow, 1962; Fridrichsons and Mathieson, 1960). Even more striking, all these grasses belong to the same tribe within the Panicoideae, the Andropogoneae. Despite previous suggestions that thelepogine is probably not derived from homospermidine but might be synthesized via a different biosynthetic pathway the close phylogenetic relationship is too big of coincidence to ignore and begs further investigation.

Furthermore, *dhs/hss* evolution in grasses has proved to be an interesting and insightful study object for protein evolution in general, and the evolution of proteins after gene duplication in particular. The protein's evolutionary history in this plant family is complex, comprising several gene duplication events, not only those which have given risen to the two *hss* clusters (Chapter 1). Amongst other things, the *dhs/hss* gene tree comprises various examples for the neofunctionalization and the subfunctionalization of duplicated genes. The specific evolutionary histories of the various gene duplicates as determined by the detailed selection analyses encourages discussion on how gene function and gene duplication mechanisms can further affect protein evolution.

Finally, while our knowledge on the evolution of the grass HSS homologs and PA biosynthesis in general was greatly advanced there are still many different aspects which invite further investigations and I await future exciting studies.

Contributions

Chapter 1 and 2 contain co-authored material. The study presented in Chapter 1 was designed by me and my supervisor Dietrich Ober. Sequence data were acquired by me and Brigitte Schemmerling. Elisabeth Kaltenecker and I purified proteins. I performed enzyme activity assays and finally analyzed all data. I wrote the manuscript which was then revised by me, Elisabeth Kaltenecker and Dietrich Ober.

The study presented in Chapter 2 was conceived and designed by me under the supervision of Dietrich Ober. I prepared the plants and plant extracts. LC-MS measurements were optimized, planned and conducted in cooperation with Tobias J. Demetrowitsch (Department of Food Technology at the University of Kiel). The data were analyzed by me. I wrote the manuscript with assistance by Tobias J. Demetrowitsch, Dietrich Ober and Karin Schwarz. I would also like to acknowledge the help of a professional editing service in editing Chapter 2.



Danksagung & Acknowledgements

This thesis would never have seen the light of day without the fantastic support and help of many people who I would like to wholeheartedly thank here:

Allen voran danke ich meinem Doktorvater Prof. Dietrich Ober. Danke, dass du mir dieses spannende Projekt anvertraut hast. Mir hat die Bearbeitung dieses tollen Themas und unsere Zusammenarbeit sehr viel Spaß gemacht.

Ich bedanke mich bei Dr. Tobias Demetrowitsch und Prof. Karin Schwarz für die sehr erfolgreiche Kollaboration. Vielen Dank Tobias, dass du dein umfangreiches Metabolomics-Wissen mit mir geteilt hast. danke für deine Unterstützung und Geduld. Ohne diese großartige Kooperation wäre das Chapter 2 dieser Arbeit nie zustande gekommen.

Ein großes Dankeschön geht an Dr. Elisabeth Kaltenegger. Danke für deine Mitarbeit an Chapter 1, deine große Diskussionsbereitschaft und den Ideenaustausch. Dank dir (und deinen fleißigen Helferlein) ist das HSS/DHS-Evolutions-Projekt ja noch zu einem sehr guten Abschluss gekommen.

Vielen Dank an die WissenschaftlerInnen, die mich im Rahmen ihrer Bachelorarbeit, ihres Laborpraktiums, oder Gastwissenschaftleraufenthalts tatkräftig im Labor unterstützt haben. Danke an Julia Jensen-Kroll, Niklas Lauterbach und Dr. Jessica Garzke! Es war mir eine große Freude mit euch zusammenzuarbeiten.

I am deeply grateful to the IMPRS for Evolutionary Biology, first and foremost, for introducing me to some great evolutionary biologists and amazing people. Especially the most awesome of IMPRS cohorts, the students of 2011, have greatly facilitated my transition from a molecular biologist to a molecular evolutionary biologist. Thank you guys! This thesis and I have greatly benefited from our exchange, scientific and otherwise.

Ganz herzlich möchte ich den Mitgliedern der AG Ober, aktuell und ehemalig, für alle Hilfe und jeden Tipp im und ums Labor (oder auch den Computer), den tollen wissenschaftlichen und nicht-wissenschaftlichen Austausch und die unterhaltsamen (und manchmal sehr ausgedehnten) Mittagspausen danken. Thanks & Danke an Brigitte, Inka, Lars, Moritz, Jule, Arun, Margret und Fritz, Franzi, Elisabeth, Doro, Frau Sobbecki, Britta, Franciane, Thomas, und Philipp! Ohne euch wäre nicht einmal die Hälfte meiner Experimente möglich gewesen und ohne euch wäre ich nicht halb so gerne jeden Tag ins Labor gekommen.

Thanks to the IMPRS I was also introduced to and had the amazing opportunity to share this whole PhD experience with the best and my most favorite flatmates in the whole wide world: Anne and Meriem. The last five years with you girls have been a blast. It was great to discuss, laugh, vent, party, and just hang in our kitchen or at the beach (in Kiel or in Portugal) and thereby balance out all the difficulties and stress that came with doing this PhD thing.

Ein großes Dankeschön geht an meine Familie in Münster ohne deren Unterstützung ich es nie von Münster über Heidelberg bis nach Kiel geschafft hätte. Vielen lieben Dank!

I also want to thank Joshka for his constant support, especially in this last and most challenging phase of being a PhD student, *the writing of the thesis*. Your knowledge of statistics, R, and cheese has been an amazing help. Thank you for always being there for me.

In my lifetime I have met a great many people – far too many to count and name individually – who have had a huge and positive impact on my life, as a scientist and as a person. Sending a big thank you from Kiel to all of my friends all over the world!

Bibliography

- Álvarez, I., Costa, A., and Feliner, G.N. (2008). Selecting single-copy nuclear genes for plant phylogenetics: A preliminary analysis for the Senecioneae (Asteraceae). *J. Mol. Evol.* *66*, 276–291.
- Anderson, D.W., McKeown, A.N., and Thornton, J.W. (2015). Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. *eLife* *4*, e07864.
- Anke, S., Niemüller, D., Moll, S., Hänsch, R., and Ober, D. (2004). Polyphyletic origin of pyrrolizidine alkaloids within the Asteraceae. Evidence from differential tissue expression of homospermidine synthase. *Plant Physiol.* *136*, 4037–4047.
- Anke, S., Gondé, D., Kaltenegger, E., Hänsch, R., Theuring, C., and Ober, D. (2008). Pyrrolizidine alkaloid biosynthesis in *Phalaenopsis* orchids: developmental expression of alkaloid-specific homospermidine synthase in root tips and young flower buds. *Plant Physiol.* *148*, 751–760.
- Betteridge, K., Cao, Y., and Colegate, S.M. (2005). Improved method for extraction and LC-MS analysis of pyrrolizidine alkaloids and their *N*-oxides in honey: application to *Echium vulgare* honeys. *J. Agric. Food Chem.* *53*, 1894–1902.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* *30*, 2114–2120.
- Böttcher, F., Adolph, R.-D., and Hartmann, T. (1993). Homospermidine synthase, the first pathway-specific enzyme in pyrrolizidine alkaloid biosynthesis. *Phytochemistry* *32*, 679–689.
- Bridgham, J.T., Ortlund, E.A., and Thornton, J.W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* *461*, 515–519.
- Bush, L.P., and Jeffreys, J.A.D. (1975). Isolation and separation of tall fescue and ryegrass alkaloids. *J. Chromatogr. A* *111*, 165–170.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* *10*, 421.
- Cao, M., Koulman, A., Johnson, L.J., Lane, G.A., and Rasmussen, S. (2008). Advanced data-mining strategies for the analysis of direct-infusion ion trap mass spectrometry data from the association of perennial ryegrass with its endophytic fungus, *Neotyphodium lolii*. *Plant Physiol.* *146*, 1501–1514.
- Catalán, P., Torrecilla, P., Rodríguez, J.Á.L., and Olmstead, R.G. (2004). Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from ITS and *trnL*-F sequences. *Mol. Phylogenet. Evol.* *31*, 517–541.
- Chawla, B., Jhingran, A., Singh, S., Tyagi, N., Park, M.H., Srinivasan, N., Roberts, S.C., and Madhubala, R. (2010). Identification and characterization of a novel deoxyhypusine synthase in *Leishmania donovani*. *J. Biol. Chem.* *285*, 453–463.

-
- Chen, T., Mei, N., and Fu, P.P. (2010). Genotoxicity of pyrrolizidine alkaloids. *J. Appl. Toxicol.* *30*, 183–196.
- Cheng, Y., Ma, X., Zhou, K., Humphreys, M.W., and Zhang, X.Q. (2015). Phylogenetic analysis of *Festuca-Lolium* complex using SRAP markers. *Genet. Resour. Crop Evol.* *63*, 7–18.
- Colegate, S.M., Edgar, J.A., Knill, A.M., and Lee, S.T. (2005). Solid-phase extraction and HPLC-MS profiling of pyrrolizidine alkaloids and their *N*-oxides: a case study of *Echium plantagineum*. *Phytochem. Anal.* *16*, 108–119.
- Cona, A., Rea, G., Angelini, R., Federico, R., and Tavladoraki, P. (2006). Functions of amine oxidases in plant development and defence. *Trends Plant Sci.* *11*, 80–88.
- Conant, G.C., and Wolfe, K.H. (2008). Turning a hobby into a job: How duplicated genes find new functions. *Nat. Rev. Genet.* *9*, 938–950.
- Cotton, J.L., Wysocki, W.P., Clark, L.G., Kelchner, S.A., Pires, J.C., Edger, P.P., Mayfield-Jones, D., and Duvall, M.R. (2015). Resolving deep relationships of PACMAD grasses: a phylogenomic approach. *BMC Plant Biol.* *15*, 178.
- Crow, W. (1962). Alkaloids of the Gramineae: *Thelepogon elegans*. *Aust. J. Chem.* *15*, 159–160.
- Dam, N.M. van, Vuister, L.W.M., Bergshoeff, C., Vos, H. de, and Meijden, E. van D. (1995). The “Raison D’être” of pyrrolizidine alkaloids in *Cynoglossum officinale*: Deterrent effects against generalist herbivores. *J. Chem. Ecol.* *21*, 507–523.
- De Bodt, S., Maere, S., and Van de Peer, Y. (2005). Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* *20*, 591–597.
- Delport, W., Poon, A.F.Y., Frost, S.D.W., and Pond, S.L.K. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* *26*, 2455–2457.
- Demetrowitsch, T.J., Petersen, B., Keppler, J.K., Koch, A., Schreiber, S., Laudes, M., and Schwarz, K. (2015). Validation of a two-step quality control approach for a large-scale human urine metabolomic study conducted in seven experimental batches with LC/QTOF-MS. *Bioanalysis* *7*, 103–112.
- Dever, T.E., Gutierrez, E., and Shin, B.-S. (2014). The hypusine-containing translation factor eIF5A. *Crit. Rev. Biochem. Mol. Biol.* *49*, 413–425.
- Estep, M.C., McKain, M.R., Diaz, D.V., Zhong, J., Hodge, J.G., Hodkinson, T.R., Layton, D.J., Malcomber, S.T., Pasquet, R., and Kellogg, E.A. (2014). Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc. Natl. Acad. Sci.* *111*, 15149–15154.
- European Medicines Agency (2011). Guideline on bioanalytical method validation.
- Fletcher, W., and Yang, Z. (2010). The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Mol. Biol. Evol.* *27*, 2257–2267.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* *151*, 1531–1545.

- Fraenkel, G.S. (1959). The raison d'être of secondary plant substances; these odd chemicals arose as a means of protecting plants from insects and now guide insects to food. *Science* *129*, 1466–1470.
- Fridrichsons, J., and Mathieson, A.M. (1960). The molecular structure and absolute configuration of theleopogine methiodide. *Tetrahedron Lett.* *1*, 18–21.
- Frölich, C., Hartmann, T., and Ober, D. (2006). Tissue distribution and biosynthesis of 1,2-saturated pyrrolizidine alkaloids in *Phalaenopsis* hybrids (Orchidaceae). *Phytochemistry* *67*, 1493–1502.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*, J.M. Walker, ed. (Humana Press), pp. 571–607.
- Goldstein, R.A., Pollard, S.T., Shah, S.D., and Pollock, D.D. (2015). Nonadaptive amino acid convergence rates decrease over time. *Mol. Biol. Evol.* *32*, 1373–1381.
- Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* *40*, D1178–D1186.
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* *52*, 696–704.
- Hahn, M.W. (2009). Distinguishing among evolutionary models for the maintenance of gene duplicates. *J. Hered.* *100*, 605–617.
- Hartmann, T. (2007). From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry* *68*, 2831–2846.
- Hartmann, T., and Witte, L. (1995). Chemistry, biology and chemoeology of the pyrrolizidine alkaloids. *Alkaloids Chem. Biol. Perspect.* *9*, 155–233.
- He, X., and Zhang, J. (2005). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* *169*, 1157–1164.
- Herrmann, M., Joppe, H., and Schmaus, G. (2002). Thesinine-4'-O- β -d-glucoside the first glycosylated plant pyrrolizidine alkaloid from *Borago officinalis*. *Phytochemistry* *60*, 399–402.
- Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinforma. Oxf. Engl.* *17*, 754–755.
- Humphreys, M.W., and Thomas, H. (1993). Improved drought resistance in introgression lines derived from *Lolium multiflorum* \times *Festuca arundinacea* hybrids. *Plant Breed.* *111*, 155–161.
- Humphreys, M., Thomas, H., Morgan, W., Meredith, M., Harper, J., Thomas, H., Zwierzykowski, Z., and Ghesquiere, M. (1995). Discriminating the ancestral progenitors of hexaploid *Festuca arundinacea* using genomic *in situ* hybridization. *Heredity* *75*, 171–174.

-
- Humphreys, M.W., Canter, P.J., and Thomas, H.M. (2003). Advances in introgression technologies for precision breeding within the *Lolium-Festuca* complex. *Ann. Appl. Biol.* *143*, 1–10.
- Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* *11*, 97–108.
- Irmer, S., Podzun, N., Langel, D., Heidemann, F., Kaltenegger, E., Schemmerling, B., Geilfus, C.-M., Zörb, C., and Ober, D. (2015). New aspect of plant-rhizobia interaction: alkaloid biosynthesis in *Crotalaria* depends on nodulation. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 4164–4169.
- Janssen, T., and Bremer, K. (2004). The age of major monocot groups inferred from 800+ *rbcL* sequences. *Bot. J. Linn. Soc.* *146*, 385–398.
- Jauhar, P.P. (1993a). Taxonomic treatments. In *Cytogenetics of the Festuca-Lolium Complex*, (Springer Berlin Heidelberg), pp. 9–27.
- Jauhar, P.P. (1993b). Intergeneric hybridization, genome relationships, and plant improvement. In *Cytogenetics of the Festuca-Lolium Complex*, (Springer Berlin Heidelberg), pp. 149–176.
- Kaltenegger, E., and Ober, D. (2015). Paralogous interference affects the dynamics after gene duplication. *Trends Plant Sci.* *20*, 814–821.
- Kaltenegger, E., Eich, E., and Ober, D. (2013). Evolution of homospermidine synthase in the Convolvulaceae: a story of gene duplication, gene loss, and periods of various selection pressures. *Plant Cell* *25*, 1213–1227.
- Kang, K.R., and Chung, S.I. (1999). Characterization of yeast deoxyhypusine synthase: PKC-dependent phosphorylation *in vitro* and functional domain identification. *Exp. Mol. Med.* *31*, 210–216.
- Kang, K.R., Wolff, E.C., Park, M.H., Folk, J.E., and Chung, S.I. (1995). Identification of YHR068w in *Saccharomyces cerevisiae* chromosome VIII as a gene for deoxyhypusine synthase, expression and characterization of the enzyme. *J. Biol. Chem.* *270*, 18408–18412.
- Katoh, K., and Standley, D.M. (2013). MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* *30*, 772–780.
- Kersey, P.J., Allen, J.E., Armean, I., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C., et al. (2016). Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res.* *44*, D574–D580.
- King, I.P., Morgan, W.G., Armstead, I.P., Harper, J.A., Hayward, M.D., Bollard, A., Nash, J.V., Forster, J.W., and Thomas, H.M. (1998). Introgression mapping in the grasses. I. Introgression of *Festuca pratensis* chromosomes and chromosome segments into *Lolium perenne*. *Heredity* *81*, 462–467.
- Kondrashov, F.A., and Kondrashov, A.S. (2006). Role of selection in fixation of gene duplications. *J. Theor. Biol.* *239*, 141–151.
- Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I., and Koonin, E.V. (2002). Selection in the evolution of gene duplications. *Genome Biol* *3*, 8–1.

- Kopecký, D., Lukaszewski, A.J., and Doležal, J. (2008). Cytogenetics of *Festulolium* (*Festuca* × *Lolium* hybrids). *Cytogenet. Genome Res.* *120*, 370–383.
- Koulman, A., Seeliger, C., Edwards, P.J.B., Fraser, K., Simpson, W., Johnson, L., Cao, M., Rasmussen, S., and Lane, G.A. (2008). *E/Z*-Thesinine-*O*-4'- α -rhamnoside, pyrrolizidine conjugates produced by grasses (Poaceae). *Phytochemistry* *69*, 1927–1932.
- Langel, D., Ober, D., and Pelsler, P.B. (2011). The evolution of pyrrolizidine alkaloid biosynthesis and diversity in the Senecioneae. *Phytochem. Rev.* *10*, 3–74.
- Lee, Y.B., Joe, Y.A., Wolff, E.C., Dimitriadis, E.K., and Park, M.H. (1999). Complex formation between deoxyhypusine synthase and its protein substrate, the eukaryotic translation initiation factor 5A (eIF5A) precursor. *Biochem. J.* *340*, 273–281.
- Leinonen, R., Sugawara, H., and Shumway, M. (2011). The Sequence Read Archive. *Nucleic Acids Res.* *39*, D19–D21.
- Li, L., Huang, Y., Xia, X., and Sun, Z. (2006). Preferential duplication in the sparse part of yeast protein interaction network. *Mol. Biol. Evol.* *23*, 2467–2473.
- Li, N., Xia, Q., Ruan, J., Fu, P., and Lin, G. (2011). Hepatotoxicity and tumorigenicity induced by metabolic activation of pyrrolizidine alkaloids in herbs. *Curr. Drug Metab.* *12*, 823–834.
- Liao, D.-I., Wolff, E.C., Park, M.H., and Davies, D.R. (1998). Crystal structure of the NAD complex of human deoxyhypusine synthase: an enzyme with a ball-and-chain mechanism for blocking the active site. *Structure* *6*, 23–35.
- Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* *290*, 1151–1155.
- Ma, P.-F., Guo, Z.-H., and Li, D.-Z. (2012). Rapid sequencing of the bamboo mitochondrial genome using Illumina technology and parallel episodic evolution of organelle genomes in grasses. *PLoS ONE* *7*, e30297.
- Macel, M. (2011). Attract and deter: a dual role for pyrrolizidine alkaloids in plant-insect interactions. *Phytochem. Rev.* *10*, 75–82.
- Macel, M., Bruinsma, M., Dijkstra, S.M., Ooijendijk, T., Niemeyer, H.M., and Klinkhamer, P.G.L. (2005). Differences in effects of pyrrolizidine alkaloids on five generalist insect herbivore species. *J. Chem. Ecol.* *31*, 1493–1508.
- Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., and Van de Peer, Y. (2005). Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 5454–5459.
- Mattocks, A.R., and White, I.N.H. (1971). Pyrrolic metabolites from non-toxic pyrrolizidine alkaloids. *Nature* *231*, 114–115.
- McKeown, A. (2015). Molecular mechanisms for the evolution of DNA specificity in a transcription factor family. Dissertation, University of Oregon.
- Moghe, G.D., and Last, R.L. (2015). Something old, something new: Conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiol.* *169*, 1512–1523.

-
- Moll, S., Anke, S., Kahmann, U., Hänsch, R., Hartmann, T., and Ober, D. (2002). Cell-specific expression of homospermidine synthase, the entry enzyme of the pyrrolizidine alkaloid pathway in *Senecio vernalis*, in comparison with its ancestor, deoxyhypusine synthase. *Plant Physiol.* *130*, 47–57.
- Mroczek, T., Glowniak, K., and Wlaszczyk, A. (2002). Simultaneous determination of *N*-oxides and free bases of pyrrolizidine alkaloids by cation-exchange solid-phase extraction and ion-pair high-performance liquid chromatography. *J. Chromatogr. A* *949*, 249–262.
- Murrell, B., de Oliveira, T., Seebregts, C., Kosakovsky Pond, S.L., Scheffler, K., and on behalf of the Southern African Treatment and Resistance Network (SATuRN) Consortium (2012). Modeling HIV-1 drug resistance as episodic directional selection. *PLoS Comput. Biol.* *8*, e1002507.
- Nielsen, R., and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* *148*, 929–936.
- Niemüller, D., Reimann, A., and Ober, D. (2012). Distinct cell-specific expression of homospermidine synthase involved in pyrrolizidine alkaloid biosynthesis in three species of the Boraginales. *Plant Physiol.* *159*, 920–929.
- Nurhayati, N., Gondé, D., and Ober, D. (2009). Evolution of pyrrolizidine alkaloids in *Phalaenopsis* orchids and other monocotyledons: Identification of deoxyhypusine synthase, homospermidine synthase and related pseudogenes. *Phytochemistry* *70*, 508–516.
- Ober, D. (2005). Seeing double: gene duplication and diversification in plant secondary metabolism. *Trends Plant Sci.* *10*, 444–449.
- Ober, D., and Hartmann, T. (1999a). Homospermidine synthase, the first pathway-specific enzyme of pyrrolizidine alkaloid biosynthesis, evolved from deoxyhypusine synthase. *Proc. Natl. Acad. Sci.* *96*, 14777–14782.
- Ober, D., and Hartmann, T. (1999b). Deoxyhypusine synthase from tobacco. cDNA isolation, characterization, and bacterial expression of an enzyme with extended substrate specificity. *J. Biol. Chem.* *274*, 32040–32047.
- Ober, D., and Kaltenecker, E. (2009). Pyrrolizidine alkaloid biosynthesis, evolution of a pathway in plant secondary metabolism. *Phytochemistry* *70*, 1687–1695.
- Ober, D., Harms, R., Witte, L., and Hartmann, T. (2003a). Molecular evolution by change of function. Alkaloid-specific homospermidine synthase retained all properties of deoxyhypusine synthase except binding the eIF5A precursor protein. *J. Biol. Chem.* *278*, 12805–12812.
- Ober, D., Gibas, L., Witte, L., and Hartmann, T. (2003b). Evidence for general occurrence of homospermidine in plants and its supposed origin as by-product of deoxyhypusine synthase. *Phytochemistry* *62*, 339–344.
- Ohno, S. (1970). *Evolution by gene duplication* (Springer Science & Business Media).
- Panchy, N., Lehti-Shiu, M., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* *171*, 2294–2316.

- Papp, B., Pál, C., and Hurst, L.D. (2003). Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424, 194–197.
- Park, M.H. (2006). The post-translational synthesis of a polyamine-derived amino acid, hypusine, in the eukaryotic translation initiation factor 5A (eIF5A). *J. Biochem.* 139, 161–169.
- Park, M.H., Lee, Y.B., and Joe, Y.A. (1997). Hypusine is essential for eukaryotic cell proliferation. *Neurosignals* 6, 115–123.
- Park, M.H., Nishimura, K., Zanelli, C.F., and Valentini, S.R. (2010). Functional significance of eIF5A and its hypusine modification in eukaryotes. *Amino Acids* 38, 491–500.
- Pichersky, E., and Lewinsohn, E. (2011). Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.* 62, 549–566.
- Pichersky, E., Noel, J.P., and Dudareva, N. (2006). Biosynthesis of plant volatiles: nature's diversity and ingenuity. *Science* 311, 808–811.
- Pond, S.L.K., Frost, S.D.W., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679.
- Pond, S.L.K., Posada, D., Gravenor, M.B., Woelk, C.H., and Frost, S.D.W. (2006). GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22, 3096–3098.
- Pond, S.L.K., Murrell, B., Fourment, M., Frost, S.D., Delport, W., and Scheffler, K. (2011). A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* msr125.
- Prasad, V., Strömberg, C.A.E., Leaché, A.D., Samant, B., Patnaik, R., Tang, L., Mohabey, D.M., Ge, S., and Sahni, A. (2011). Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nat. Commun.* 2, 480.
- Prince, V.E., and Pickett, F.B. (2002). Splitting pairs: the diverging fates of duplicated genes. *Nat. Rev. Genet.* 3, 827–837.
- R Core Team (2013). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing).
- Rastogi, S., and Liberles, D.A. (2005). Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol. Biol.* 5, 28.
- Reimann, A., Nurhayati, N., Backenköhler, A., and Ober, D. (2004). Repeated evolution of the pyrrolizidine alkaloid-mediated defense system in separate angiosperm lineages. *Plant Cell* 16, 2772–2784.
- Roeder, E. (2000). Medicinal plants in China containing pyrrolizidine alkaloids. *Pharmazie* 55, 711–726.
- Rokas, A., and Carroll, S.B. (2008). Frequent and widespread parallel evolution of protein sequences. *Mol. Biol. Evol.* 25, 1943–1953.
- Sasaki, K., Abid, R., and Miyazaki, M. (1996). Deoxyhypusine synthase gene is essential for cell viability in the yeast *Saccharomyces cerevisiae*. *FEBS Lett.* 384, 151–154.

-
- Schardl, C.L., Grossman, R.B., Nagabhyru, P., Faulkner, J.R., and Mallik, U.P. (2007). Loline alkaloids: Currencies of mutualism. *Phytochemistry* 68, 980–996.
- Scotto-Lavino, E., Du, G., and Frohman, M.A. (2007). 3' End cDNA amplification using classic RACE. *Nat. Protoc.* 1, 2742–2745.
- Sela, I., Ashkenazy, H., Katoh, K., and Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* 43, W7–W14.
- Shoja, V., and Zhang, L. (2006). A roadmap of tandemly arrayed genes in the genomes of human, mouse, and rat. *Mol. Biol. Evol.* 23, 2134–2141.
- Sievert, C., Beuerle, T., Hollmann, J., and Ober, D. (2015). Single cell subtractive transcriptomics for identification of cell-specifically expressed candidate genes of pyrrolizidine alkaloid biosynthesis. *Phytochemistry* 117, 17–24.
- Smith, S.D., and Rausher, M.D. (2011). Gene loss and parallel evolution contribute to species difference in flower color. *Mol. Biol. Evol.* 28, 2799–2810.
- Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K., and Pond, S.L.K. (2015). Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* 32, 1342–1353.
- Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Zuloaga, F.O., Judziewicz, E.J., Filgueiras, T.S., Davis, J.L., and Morrone, O. (2015). A worldwide phylogenetic classification of the Poaceae (Gramineae). *J. Syst. Evol.* 53, 117–137.
- Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J. (2007). pcaMethods - a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* 23, 1164–1167.
- Stegelmeier, B.L., Edgar, J.A., Colegate, S.M., Gardner, D.R., Schoch, T.K., Coulombe, R.A., and Molyneux, R.J. (1999). Pyrrolizidine alkaloid plants, metabolism and toxicity. *J. Nat. Toxins* 8, 95–116.
- Stoltzfus, A. (1999). On the possibility of constructive neutral evolution. *J. Mol. Evol.* 49, 169–181.
- Taylor, J.S., and Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* 38, 615–643.
- Umland, T.C., Wolff, E.C., Park, M.H., and Davies, D.R. (2004). A new crystal structure of deoxyhypusine synthase reveals the configuration of the active enzyme and of an enzyme·NAD·inhibitor ternary complex. *J. Biol. Chem.* 279, 28697–28705.
- UniProt Consortium, T. (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212.
- Wang, T.-W., Lu, L., Zhang, C.-G., Taylor, C., and Thompson, J.E. (2003). Pleiotropic effects of suppressing deoxyhypusine synthase expression in *Arabidopsis thaliana*. *Plant Mol. Biol.* 52, 1223–1235.

- Wassel, G., El-Menshawi, B., Saeed, A., Mahran, G., and Reich, J. (1987). New sources of pyrrolizidine alkaloids: genus *Cordia* (Ehretiaceae) and *Schismus* (Gramineae). *J. Sci. Pharm.* 55, 163–166.
- Weadick, C.J., and Chang, B.S.W. (2012). An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. *Mol. Biol. Evol.* 29, 1297–1300.
- Wertheim, J.O., Murrell, B., Smith, M.D., Pond, S.L.K., and Scheffler, K. (2015). RELAX: Detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* 32, 820–832.
- Wu, Z.-Q., and Ge, S. (2012). The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts. *Mol. Phylogenet. Evol.* 62, 573–578.
- Xiong, A., Yang, L., Ji, L., Wang, Z., Yang, X., Chen, Y., Wang, X., Wang, C., and Wang, Z. (2011). UPLC-MS based metabolomics study on *Senecio scandens* and *S. vulgaris*: an approach for the differentiation of two *Senecio* herbs with similar morphology but different toxicity. *Metabolomics* 8, 614–623.
- Xu, X., Keefer, L.K., Ziegler, R.G., and Veenstra, T.D. (2007). A liquid chromatography–mass spectrometry method for the quantitative analysis of urinary endogenous estrogen metabolites. *Nat. Protoc.* 2, 1350–1355.
- Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15, 568–573.
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
- Yang, Z., Kumar, S., and Nei, M. (1995). A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141, 1641–1650.
- Yang, Z., Wong, W.S.W., and Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118.
- Zhang, J. (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298.
- Zhang, J., and Kumar, S. (1997). Detection of convergent and parallel evolution at the amino acid sequence level. *Mol. Biol. Evol.* 14, 527–536.
- Zhang, J., Rosenberg, H.F., and Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci.* 95, 3708–3713.
- Zhang, J., Nielsen, R., and Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22, 2472–2479.
- Zhou, Y., Li, N., Choi, F.F.-K., Qiao, C.-F., Song, J.-Z., Li, S.-L., Liu, X., Cai, Z.-W., Fu, P.P., Lin, G., et al. (2010). A new approach for simultaneous screening and quantification of toxic pyrrolizidine alkaloids in some potential pyrrolizidine alkaloid-containing plants by using ultra performance liquid chromatography–tandem quadrupole mass spectrometry. *Anal. Chim. Acta* 681, 33–40.

Zou, Z., and Zhang, J. (2015). Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Mol. Biol. Evol.* 32, 2085–2096.





Erklärung

Die vorliegende Doktorarbeit ist - abgesehen von der Beratung durch meinen Betreuer Prof. Dietrich Ober - nach Inhalt und Form meine eigene Arbeit.

Die Arbeit wurde weder ganz noch zum Teil an der CAU Kiel oder an einer anderen Hochschule im Rahmen eines Prüfungsverfahrens vorgelegt.

Die vorliegende Arbeit entstand unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft.

Kiel, 20.12.2016

(Anne-Maria Wesseling)