DARWINIANA 44(1): 7-17. 2006 ISSN 0011-6793

ARTÍCULO INVITADO

BEYOND TAXONOMY: PROSPECTS FOR UNDERSTANDING MORPHOLOGICAL DIVERSITY IN THE GRASSES (POACEAE)

Elizabeth A. Kellogg

Department of Biology, University of Missouri-St. Louis, One University Boulevard, St. Louis, Missouri 63121, USA; tkellogg@umsl.edu

Abstract. Kellogg, E.A. 2006. Beyond taxonomy: prospects for understanding morphological diversity in the grasses (Poaceae). *Darwiniana* 44(1): 7-17.

The grass family (Poaceae) is unusually well known taxonomically, with over 10,000 described species. These are characterized and classified according to a variety of morphological characters, particular those of the inflorescence and the spikelet but also a wealth of micromorphological and molecular characters. Much data on the family is accessible electronically. Available phylogenies have identified 12 major lineages that are classified as subfamilies; within these subfamilies, especially for subfamily Panicoideae, the phylogeny is becoming increasingly clear. Developmental studies focus on the activity and fate of meristems throughout the plant, and illustrate how adult morphology arises. Analyses of quantitative trait *loci* (QTL) verify that some taxonomic characters are controlled by distinct genes, thus supporting their use in taxonomy. Studies of gene expression identify several genes, including *LEAFY HULL STERILE1, RAMOSA1,* and *TEOSINTE BRANCHED1* that may have contributed to diversification of the grasses.

Keywords. Poaceae, phylogeny, taxonomy, development, quantitative genetics, gene expression.

Resumen. Kellogg, E. A. 2006. Más allá de la taxonomía: perspectivas para entender la diversidad morfológica en la familia de las Gramíneas (Poaceae). *Darwiniana* 44(1): 7-17.

La familia de las Gramíneas (Poaceae) es bien conocida desde el punto de vista taxonómico, con más de 10.000 especies descriptas, caracterizadas y clasificadas de acuerdo con una variedad de caracteres morfológicos, en particular de la inflorescencia y de la espiguilla, como también por una gran cantidad de caracteres micromorfológicos y moleculares. Información abundante sobre esta familia es accesible electrónicamente. Las filogenias disponibles han identificado 12 linajes principales clasificados como subfamilias, dentro de cada una de ellas, y en particular dentro de la subfamilia Panicoideae, la filogenia se está revelando con creciente claridad. Los estudios de desarrollo se enfocan en la actividad y el destino de los meristemas de toda la planta, e ilustran cómo se origina la morfología adulta. Los análisis de *loci* de caracteres cuantitativos (QTL) verifican que algunos caracteres taxonómicos son controlados por genes diferentes, apoyando el uso de estos caracteres en la taxonomía. Los estudios de expresión génica identifican varios genes, entre ellos *LEAFY HULL STERILE1, RAMOSA1,* y *TEOSINTE BRANCHED1*, que pueden haber contribuido a la diversificación de las gramíneas.

Palabras clave. Poaceae, filogenia, taxonomía, desarrollo, genética cuantitativa, expresión de genes.

INTRODUCTION

For centuries, humans have classified the organisms around them using visible characteristics. This process gradually became formalized into what is now called alpha taxonomy, in which organisms are assigned to groups and the groups are given names. Naming organisms is of course central to all communication among humans. Names are the entities that allow us to make sense of the world. Contemplate briefly the possibility of referring to your friends and family members by numbers, or by strings of words, rather than names. One might imagine saying, “Good morn-

Original recibido el 31 de marzo de 2006; aceptado el 15 de mayo de 2006

ing, person in the office down the hall!" Or, "How are things, 1247?" Alpha taxonomy solves the basic problem of communication by providing the words from which the language of biology can be developed.

In the process of assigning names to organisms, the alpha taxonomist also describes morphological diversity. In plant taxonomy, the characters used initially are those that can be observed with the naked eye or with a low-powered lens. Initially, these morphological characters are interpreted as reflecting relationships among organisms; characteristics that are shared by two or more groups of specimens are hypothesized to be homologous. The hypothesis of homology is then tested by addition of micromorphological and molecular characters and construction of a phylogeny.

In some plant families, it is now possible to test the hypothesis of homology in an even more precise way by examining the genetic basis of the characters. One such family is the grass family (Poaceae). In this brief paper, I will outline some recent work in the family, and highlight the possibility of developing a real understanding of morphological diversity with current genetic and genomic tools.

ALPHA TAXONOMY

The alpha taxonomy of the grass family is remarkably well understood, thanks to years of hard work by literally hundreds of illustrious taxonomists. Revisions of large and complex genera appear frequently (e.g. Saarela et al., 2003; Spangler, 2003; Molina & De Agrasar, 2004; Finot et al., 2005; Zuloaga & Morrone, 2005, to name just a few in the last 36 months). Although new species are continually being discovered, these can usually be assigned to known genera or at least tribes. Because of the excellent taxonomic base, this is now an ideal time to be undertaking monographic work in grasses. It is possible to assemble a broad set of specimens from almost any genus or tribe, and to know that the sample approaches completeness. This means that generalizations are possible and that they are unlikely to be overturned by discovery of a vast number of radically different specimens. The literature on grasses is also well cataloged, especially for the New World (Soreng et al., 2000 onwards).

This is not to say that all taxonomic questions are answered. There remain many groups in which names are inconsistently applied, where the same species goes by different names in different parts of the world, where typification is unclear, and where regional floristic treatments provide the only keys to species. Many sizable genera or generic groups need worldwide taxonomic treatments, if only so that species can be delimited more precisely. Some parts of the world remain seriously undercollected, and it is notoriously difficult to obtain new material from some countries (e.g. India, Brazil).

The grasses also have been a model for development of electronic resources, which are too numerous to list here; a set of relevant links is provided by Soreng et al. (2000 onwards). Grass Genera of the World (Watson & Dallwitz, 1992 onwards) was developed as a prototype for an electronic taxonomic resource, and was one of the model databases for which the Descriptive Language for Taxonomy (DELTA) was created (Dallwitz et al., 1999 onwards). That database now includes all genera (ca. 800), and can be searched by an interactive key (software downloadable from the web). The generic database has also been extended to the species level (Clayton et al., 2002 onwards), so that descriptions of 10,800 species of grasses are now available electronically. This number is about 800 more than the number suggested by Clayton and Renvoize two decades ago (Clayton & Renvoize, 1986), representing an average of 40 new species per year. Other exemplary electronic resources include AusGrass (Sharp & Simon, 2002) and Grasses of Iowa (http://www.eeob.iastate.edu/research/iowagrasses/index.html). With these and other examples, it is easy to see how taxonomic data on the grasses can be used on the web.

The taxonomic needs in the grasses are thus the same as those in many other taxa - continental-scale or world-wide revisions of genera and tribes, and electronic access to identification materials. Nonetheless, compared to other families of a similar size (e.g., Leguminosae, Rubiaceae), the extent and quality of alpha taxonomy in the grasses is remarkable.

PHYLOGENETICS

Phylogenetic analysis has become closely linked to alpha taxonomy, becoming central to production of a classification that reflects shared history. At least as important, phylogenies help understand and interpret morphological variation. Although many morphological phylogenies have been produced, molecular data are increasingly used for phylogenies and also for delimitation of species.

In the grasses, a molecular phylogeny of the family was generated by the Grass Phylogeny Working Group (GPWG, 2001), based on data from four chloroplast loci, two nuclear loci, and chloroplast restriction sites. Most of the relationships in that phylogeny have been supported by subsequent studies (e.g. Malcomber & Kellogg, 2004). In particular, *Streptochaeta* plus *Anomochloa* are sister to the rest of the family, the next diverging group is Pharoideae, next is Puelioideae, and then follows the major radiation of the family. Panicoideae, Arundinoideae, Centothecoideae, Chloridoideae, Aristidoideae, and Danthonioideae form a well-supported clade (the PACCAD clade), although no macromorphological character is uniquely shared among them. The relationships among this set of subfamilies are not certain, although most data support a sister relationship between Panicoideae and Centothecoideae. Sister to the PACCAD clade in the GPWG analysis is a clade comprised of Bambusoideae, Ehrhartoideae, and Pooideae. This clade is not strongly supported in most analyses, and other relationships among the three subfamilies cannot be ruled out. In particular, Pooideae might be sister to the PACCAD clade, although this remains unsupported.

Molecular phylogenetic data are also available for all subfamilies, most tribes, and many of the ca. 800 genera. The most comprehensive molecular phylogenetic studies published to date have been in the subfamily Panicoideae, in which 59 of ca. 200 genera have been sampled (Giussani et al., 2002; Aliscioni et al., 2003). The enormous genus *Panicum* has been confirmed as polyphyletic. This was not a surprise, but many species had been left in *Panicum* because there was no other obvious location for them.

Together with monographic work, molecular phylogenetic data have permitted a revised and stable classification for much of the family. Twelve subfamilies are now recognized, with the possible addition of a 13th subfamily to accommodate previously unplaced species of the PACCAD clade (Sanchez-Ken et al., unpublished). Several tribes (e.g. Danthonieae, Aristideae) are now identical in circumscription to their respective subfamilies and are therefore redundant. Major outstanding tribal-level questions include delimitation of Poeae from Aveneae (Pooideae) and Chlorideae from Eragrostideae (Chloridoideae). The divisions between the members of each of these pairs have been unclear even since early morphological phylogenetic work (Kellogg & Watson, 1993). Available molecular data do not support monophyly of any of these tribes as conventionally understood (Van den Borre & Watson, 1997; Soreng & Davis, 2000). Possible solutions may include sinking Aveneae into Poeae and Eragrostideae into Chlorideae, although recognition of multiple smaller tribes might also be consistent with phylogenies. At the moment, the limitation is indequate sampling of genes and taxa in published phylogenies, combined with weak statistical support for many nodes. Additional phylogenetic data for the Chloridoideae and Pooideae are forthcoming in the proceedings of the most recent monocot conference (Columbus et al., in press), and will provide additional useful information.

Generic delimitation and relationships among genera are still very much a work in progress. Again, more is known about Panicoideae than other subfamilies, although data are accumulating on species related to *Avena* (Grebenstein et al., 1998), *Festuca* (Torrecilla et al., 2003), *Poa* (Gillespie & Soreng, 2005), *Bouteloua* (Columbus et al., 1998), and *Eleusine* (Neves et al., 2005), to name only a few. Generic relationships within Triticeae have always been and remain problematic, although molecular data have clearly identified the source of the problems (Blattner, 2004; Helfgott & Mason-Gamer 2004; Mason-Gamer, 2005).

DEVELOPMENTAL MORPHOLOGY

Good monographic and phylogenetic studies permit - perhaps even demand - re-examination of

phenotypes. Characters thought to indicate relationship are now seen as parallelism (e.g. single-flowered spikelets in Pooideae, highly branched inflorescences in Panicoideae) or convergence. Mapping adult characters on phylogenetic trees points to complex patterns of homoplasy. Homoplasious characters are non-homologous (where the word "homology" means "synapomorphy"(Patterson, 1982)).

A powerful use of developmental data is in assessing similarity between structures. Adult structures may appear similar even though they are developmentally distinct, and conversely structures that originate in similar ways may become distinct late in development. Developmental studies of structures that arise in parallel can determine whether the structures are developmentally the same or different, and thus whether taxonomists have correctly or incorrectly given them the same name. Here I mention several characters that vary among grasses, and sketch some of the taxonomic insights that have come from developmental studies.

The terminology applied to inflorescences in the grasses traditionally equates a spikelet with a flower. Inflorescences with spikelets attached directly to the inflorescence axis are called spikes, those with spikelets on short pedicels, racemes, and everything else is termed a panicle. This terminology is misleading in its simplicity, and in particular in the inclusion of so many disparate structures under the term "panicle". This problem has been observed by several workers (e.g. Camara-Hernandez & Rua, 1991; Vegetti & Weberling, 1996), who realized that more precise descriptions of morphology would be necessary. Vegetti and his colleagues have produced detailed descriptions of the adult morphology of a number of species of grasses, including various Andropogoneae (del Pilar Schneider & Vegetti, 1992; Vegetti, 1992) and Oryzeae (Vegetti, 2000). These papers highlight the real complexity of the grass inflorescence, and focus on such characteristics as terminal spikelets and reiterated branching patterns. The resulting typology of grass inflorescences is a helpful step beyond the misleading simplicity of the spike-raceme-panicle terms.

Plant form is generated ultimately by the activity of meristems. To make appropriate comparisons among plants and to determine which structures are comparable requires investigation of meristems, their activity and their products. Recently Rua and Reinheimer (unpublished) have reviewed the typology literature, and have suggested that developmental data present a real opportunity for improving understanding of inflorescences and for analyzing their development in a comparative manner.

Developmental data have helped clarify the distribution of such characters as phyllotaxis, numbers of orders of branching, and number of branches at each order (reviewed by Malcomber et al., 2006). For example, the phyllotaxis of the inflorescence in the grasses is ancestrally spiral. The leaves are produced in a distichous phyllotaxis, but when the meristem is transformed to a reproductive meristem, its first lateral structures (branches) are produced in a spiral; this shift has been investigated extensively in maize. Spiral primary branches are also produced in *Joinvillea*, *Ecdeiocolea* (grass outgroups), *Streptochaeta*, *Oryza*, and most panicoids and chloridoids. In the Pooideae in contrast, the distichous phyllotaxy of the leaves continues into the inflorescence; the distichy of the primary inflorescence branches thus appears to be a synapomorphy for Pooideae (Evans, 1940). This means that "panicles" in Pooideae are architecturally different from "panicles" in Panicoideae, having arisen at their highly branched condition from different starting points. Similar persistent distichy is observed in *Urochloa* (Reinheimer et al., 2005), some bamboos (E. A. Kellogg, unpublished observations), some Andropogoneae (LeRoux & Kellogg, 1999), and *Chionochloa macra*, the one member of Danthonioideae that has been studied (Martin et al., 1993).

Another such character is spikelet pairing. Data from development and genetics show that the spikelet pair is actually a small branch complex; one branch produces a single lateral meristem and then terminates in a spikelet, as does the lateral meristem. Thus within the spikelet pair, one of the spikelets is terminal and the other is lateral on the short branch. Because the short branch elongates very little, the two spikelets appear to develop from a common meristem. This interpretation is supported by genetic studies (e.g., Vollbrecht et al., 2005; Bortiri et al., 2006), by developmental observations (e.g. Orr et al., 2002), and

by exploration of gene expression. For example, the gene *Barrenstalk1*, which is expressed in the axil of any incipient branch is expressed in the axil of the sessile spikelet in maize, suggesting that the sessile spikelet is indeed a lateral branch (Gallavotti et al., 2004). Spikelet pairing is generally cited as synapomorphic for Andropogoneae, but developmental studies show that it must be a deeper synapomorphy, because it appears in other Panicoideae such as *Ixophorus* (Kellogg et al., 2004) and *Paspalum* (Kellogg, 2000).

The most striking characteristic of the grasses is the spikelet, a short shoot subtended by two bracts (glumes) and containing one or more flowers. The spikelet meristem initiates the two glumes in succession, followed by one or more bracts known as lemmas. In the axils of the lemmas, floral meristems form and produce, successively, an adaxial structure (palea), lodicules, stamens and a gynoecium. The lodicules are in the position of an inner whorl of tepals, and share genetic similarities with inner perianth structures in other plants (Whipple et al., 2004; Whipple & Schmidt, 2006). Thus, although their function is clearly mechanical in the grasses, all evidence points to their being a highly modified inner perianth. The evolutionary history of the palea is more controversial; although it is in the position of an outer perianth whorl, it is adaxial and in many grasses looks very much like a prophyll. Likewise the lemma, although clearly abaxial and bract-like, shares some similarities with an outer perianth whorl. The entire organ complex - lemma, palea, lodicules, stamens and gynoecium - is thus called a floret to indicate its uncertain homologies with a flower.

Although initiation of glumes and lemmas is strictly acropetal as far as is known, differentiation of the floral meristems may proceed from bottom to top (acropetal) or from top to bottom (basipetal) (Malcomber & Kellogg, 2004; Reinheimer et al., 2006). Species with the former pattern have multiple florets per spikelet, whereas the latter pattern appears in species with only two or three florets per spikelet. Taxonomically, basipetal maturation characterizes Panicoideae and Ehrhartoideae, whereas acropetal maturation is common in Pooideae and Chloridoideae. Maturation pattern correlates with expression of some genes (see below).

Other aspects of grass morphology still need to be investigated developmentally. For example, awns form on the lemma of many species, and in some species also on the glumes. In evolutionary time, awns are gained and lost multiple times, and their morphology is variable. It seems likely that their development is also variable, but they have never been investigated in a comparative context.

Parts of the spikelet and floret are described as "absent" in some species, but this is based solely on examination of adult morphology. For example, reduction in stamen number is common, and may correlate with obligate inbreeding. It is unclear whether all three stamens initiate in such species, or whether stamens are suppressed from the start. Also, different species may achieve stamen reduction in different ways.

In many grasses - often segregated into their own genera - adult morphology is highly modified. Such genera include *Sesleria*, which has "bracts" in the lower part of the inflorescence that are generally interpreted as the glumes of sterile spikelets (Clayton & Renvoize, 1986), *Coix*, in which a highly modified leaf sheath forms a hard bead-like case around the pistillate spikelet (Jacques-Félix, 1961), and *Lygeum*, in which the parts of the spikelet are largely indistinguishable at maturity. None of these has been studied developmentally to determine exactly when development diverges from normal.

In summary, developmental data have helped to clarify the nature of multiple taxonomic characters. The term "panicle" is really too imprecise for description of inflorescences and should probably be discarded. Phyllotaxis of the primary inflorescence branches does vary and could be given more attention. The order of development of flowers within a spikelet may be more fundamental than the actual number of flowers, and thus would repay more investigation in more taxa. The mechanism of reduction of parts could be quite different among unrelated species. And the many species with particularly odd morphology could be understood better if examined throughout their growth.

GENOME SEQUENCING, MAPPING, AND QTL ANALYSIS

The grasses have incomparable genomic

resources. Whole genome sequences are available for *Oryza sativa* ssp. *japonica* and ssp. *indica* (Goff et al., 2002; Yu et al., 2002), sorghum will be completed soon after this review appears (http://www.jgi.doe.gov/sequencing/cspseqplans2006.html), and maize will undoubtedly be completed soon after that (http://www.nsf.gov/news/news_summ.jsp?cntn_id=104608&org=NSF). Collections of expressed sequence tags (ESTs) are available for *Avena sativa*, *Brachypodium distachyon*, *Festuca arundinacea*, *Hordeum vulgare*, *Panicum virgatum*, *Saccharum officinarum*, *Secale cereale*, *Sorghum propinquum*, *Triticum aestivum*, and *Triticum monococcum* (http://www.plantgdb.org/prj/ESTCluster/progress.php).

Genome maps are also available for many grasses. These maps are constructed by crossing two plants and then self-pollinating their F_1 hybrid to produce a population of F_2 individuals. The parental plants are generally chosen because they differ in some aspect of morphology or biochemistry of interest. Known pieces of DNA ("markers") are screened in the parents either by use of restriction enzymes or by sequencing; any piece of DNA that is polymorphic between the parents can be located on the genetic map. Each F_2 plant is assayed for each marker, and is scored according to whether it is homozygous for the alleles from parent 1, homozygous for alleles from parent 2, or heterozygous. These data are then analyzed statistically to determine which markers are genetically linked, and the groups of linked markers assembled into a genetic map, with an ordered list of markers on each chromosome.

A startling result from early mapping studies was that molecular markers in multiple species of grasses were arranged in approximately the same order in the genomes (Gale & Devos, 1998), demonstrating that the genomes of the grasses are all largely collinear despite the considerable morphological and physiological differences among the plants. Suddenly geneticists working on rice, maize, and wheat discovered that they were working on organisms that were more similar than they had ever thought possible. With this discovery came the realization that systematic and phylogenetic data were necessary for understanding genetics. The discovery of collinear genomes thus unified cereal genetics with taxonomy and phylogenetics. (A similar phenomenon has occurred simultaneously in Solanaceae, in which a major international collaboration has originated out of the confluence of genomics and systematics.)

Once a genetic map has been constructed, it can be used for multiple purposes, including the very powerful approach of quantitative trait analysis (Lynch & Walsh, 1998). In this method, the same F_2 individuals that were used to create the map can be scored for any phenotypic trait of interest; in agronomy this is often yield, but can be such characters as oil content of the seed, or plant height, or number of inflorescence branches. The phenotype of each plant is then correlated with its genotype at each marker. If phenotypic variation correlates with genotypic variation at a particular locus, then there is evidence that near that marker is a gene that controls the phenotype.

QTL analysis can be used to determine the number of genes underlying a particular phenotype. For example, Doebley and Stec (1991, 1993) were able to demonstrate that the dramatic morphological differences between maize and its wild ancestor teosinte were controlled by only a small number of major loci, thus confirming a hypothesis first proposed by Beadle (1939, 1980). Similarly, the differences between foxtail millet and its wild progenitor green millet are due to about 14 loci affecting branching in the inflorescence and in the vegetative parts of the plant (Doust et al., 2004, 2005).

QTL analysis can also be used to determine whether particular characters are controlled by the same or by different loci. For example, Doust et al. (2004) showed that tillering and axillary branching in *Setaria* were controlled by distinct loci, even though both would be classed morphologically as axillary branches. Doust et al. also found that branching in *Setaria* was only slightly affected by the locus *TEOSINTE BRANCHED1* (*TB1*), even though this is a major locus leading to the reduction in branching in domesticated maize (Wang et al. 1999). In contrast, branching in *Setaria* is strongly influenced by a locus that contains the gene *BARRENSTALK1* (*BA1*) (Doust & Kellogg, 2006). Interestingly, *TB1* controls branch outgrowth, whereas *BA1* controls branch initiation. In cultivated maize, axillary branches and tillers are initiated but fail to grow out, con-

sistent with high expression of *TB1*. In *Setaria*, axillary branches fail to initiate at all.

The challenge in QTL analysis is making the mapping population and the genetic map in the first place. Currently a limitation is creating an F_1 hybrid between disparate parental plants, and then producing F_2 or backcross offspring. Once it is made, however, the power for determining numbers of genes is considerable.

After QTL have been identified it is possible to exploit the collinearity of the grass genomes. Markers in one species can be placed on a map in another species. For example, Doust et al. (2005) have recently identified QTL in *Setaria* that affect the numbers of orders of branching in the inflorescence. The QTL region is bracketed by a pair of markers that originated in rice. Doust et al. then looked in rice at the region between those two markers to retrieve the list of genes that fall into that region. The list is quite long (ca. 400 genes), which seems daunting, but is almost 100 times fewer genes than in the whole genome, and thus represents narrowing of the field of possibilities by two orders of magnitude. Within that list of 400 genes are a number whose function is known, including *TB1*. Recent work with *TB1* suggests that it might be a regulator of inflorescence branching in foxtail millet, even though it is not in maize (A.N. Doust and E.A. Kellogg, unpublished data).

In summary, genome maps provide a tool for investigating the numbers and locations of genes that underlie taxonomic characters. Aspects of inflorescence and vegetative branching have been mapped and shown to be under the control of multiple genes. For example, basal branching (tillering) and axillary branching are genetically distinct, and thus the genetic results support their use as independent taxonomic characters.

GENE EXPRESSION

Once candidate genes have been identified, via comparison with a model system and/or via QTL mapping, there remains the difficult problem of determining their function. To do this rigorously requires extensive experimental work on biochemistry, reverse genetics, and transformation in one or more models. A first step, however, is to examine expression of the gene to determine in which tissues and cells it is transcribed.

It is important here to distinguish between biochemical function and developmental role, both of which are sometimes described as "gene function". Biochemical function refers to the specific protein-protein or protein-nucleic acid interactions of the gene product - which residues contact which other residues and what the substrate is if the gene encodes an enzyme. Developmental role refers to the tissue and organ development that is affected by the gene. A gene can acquire a novel developmental role by changes in where or when it is expressed, even though its biochemical function remains conserved. Changes in expression pattern are thought to be caused by changes in promoter or other regulatory sequences in the gene itself, so-called cis-regulatory changes, rather than changes in the regulator, which would be trans-regulatory (Doebley & Lukens, 1998). Studies of expression pattern provide little insight into biochemical function, but can suggest change or conservation of developmental role. Thus such studies serve as a correlational test of hypotheses of function, and also as a heuristic device to develop novel hypotheses.

Expression of many developmentally important genes has now been studied in some detail, some in multiple grasses (reviewed by Malcomber et al., 2006). Genes such as *KNOTTED1*, *BARREN STALK*, *RAMOSA1*, and the B- and C-class MADS box genes have similar expression patterns in all grass species investigated to date (Gallavotti et al., 2004; Whipple et al., 2004; Bortiri et al., 2006; Whipple & Schmidt, 2006; Doust & Kellogg, unpublished). This implies conservation of biochemical function and developmental role for these genes among the grasses. More interesting from an evolutionary point of view are the genes whose expression pattern varies among the grasses. These are good candidates for genes that control the inflorescence diversity seen by taxonomists. In this category are the FRUITFULL-like and *SEPALLATA*-like MADS box genes, plus the branching gene *RAMOSA1*.

The *FRUITFULL* (*FUL*)-like proteins have sometimes erroneously been called A-class proteins, based on their general similarity to *APETALA1* (*AP1*), which affect sepal and petal identity in Arabidopsis. It is now clear, however, that *AP1*

and its function in perianth identity are restricted to eudicots (Litt & Irish 2003). Most other angiosperms -including the grasses- have genes more similar to *FUL*, which in Arabidopsis affects carpel and ovule identity. *FUL*-like Proteins have a C-terminus that is clearly distinct from that of *AP1*.

Grasses have three *FUL*-like genes, two of which were produced by the whole genome duplication at the origin of the grasses (Preston and Kellogg, submitted). In the grasses, *FUL*-like genes are expressed in a broad range of tissues, including the spikelet, but also in some species, in the leaves. The expression pattern is variable, and this may correlate with variable function. One of the *FUL* loci (called *FUL1* by Preston & Kellogg, *VRN1* by other authors) controls vernalization response in wheat, barley, and ryegrass (*Lolium*) (Yan et al., 2003; Yan et al., 2004; Fu et al., 2005; Jensen et al., 2005). In winter wheat, which requires a cold treatment to flower, *FUL1* is upregulated in response to cold. When *FUL1* function is removed, winter wheat fails to flower. It will be of considerable interest to know whether *FUL1* is generally involved in cold response in the grasses, or whether this phenomenon occurs only in Pooideae.

Grasses also have five sets of genes related to the *SEPALLATA* genes of *Arabidopsis* (Malcomber and Kellogg, 2005). In *Arabidopsis*, these genes interact with other MADS-box transcription factors to specify identity of floral organs. In the grasses, only one of the genes, *LEAFY HULL STERILE1*, has been characterized in any detail (Jeon et al. 2000). When the gene is mutated in rice, the spikelet is disrupted and the floral organs become defective. Conversely, when the gene is overexpressed, the two sterile lemmas (sometimes called "glumes" in the taxonomic literature) become similar to the upper lemma and palea in morphology (Prasad et al., 2001). Data from rice thus show that the gene is involved in spikelet architecture and in lemma and palea identity. Studies in multiple species of grasses show that *LHS1* is expressed in lemmas and paleas in all species investigated to date (Malcomber & Kellogg, 2004; Reinheimer et al., 2006).

Based on the mutant phenotype in rice and on the expression pattern of comparable genes in maize, Cacharrón et al. (1999) suggested that *LHS1* genes were required to specify the upper floret of the spikelet. Malcomber and Kellogg (2004) and Reinheimer et al. (2006) tested this hypothesis by examining expression of *LHS1* in multiple species of grasses. They found that expression pattern correlated with the pattern of spikelet development. Species with two- or three-flowered spikelets and top-down maturation of the florets expressed *LHS1* only in the upper floret. However, in species with more than three florets and bottom-up maturation, *LHS1* was expressed in all florets, albeit transiently in some species. Thus gene expression correlates with developmental pattern.

RAMOSA1 is another protein that may be responsible for variation in inflorescence form in the grasses. In the maize inflorescence, it regulates the development of spikelet pairs (Vollbrecht et al., 2005). When the gene is mutated, the short branches that would normally form spikelet pairs instead elongate and produce several unpaired spikelets. The gene is expressed early in inflorescence development, about the time that spikelet pairs form. In *Miscanthus sinensis* and *Sorghum bicolor*, two other species that also produce spikelet pairs, *RA1* is also expressed during spikelet pair development. In contrast, rice does not produce spikelet pairs, and the gene appears to be entirely absent from the rice genome. Work is now ongoing to determine the phylogenetic distribution of the *RA1* gene and whether it originated at the same time as spikelet pairing.

The inflorescence is important in grass evolution and classification, but other striking characters have repaid detailed study. The most notable of these is the C_4 photosynthetic pathway, which was first discovered in sugarcane (Hatch, 1999). Internal anatomy of the leaves correlates with activity of C_4 enzymes (Hattersley & Watson, 1975), and with $\partial^{13}C$ ratios (Brown 1977). The grasses include several different C_4 subtypes, which differ in their internal anatomy (Hattersley & Watson, 1992). The development and biochemistry of the various subtypes have been studied extensively. Comparative studies of protein localization show that the major photosynthetic enzymes - phosphoenol pyruvate carboxylase and ribulose 1,5 bisphosphate carboxylase/oxygenase- are distributed in the mesophyll and bundle sheath, respectively, of plants with all of the C_4

subtypes. Other enzymes, however, are localized differently depending on the taxonomic group (Sinha & Kellogg, 1996).

In summary, gene expression studies are just beginning to provide data that can illuminate similarities and differences among species. Vernalization is a physiological phenomenon that affects habitat and ecology of many grasses, not just the crops, so an understanding of the variability of its regulation will be important for understanding natural systems. The genetic control of spikelet morphology is barely understood, but *LHS1* expression suggests that the different developmental patterns are highly regulated at the genetic level. Inflorescence branching and photosynthetic pathway are characters that exhibit considerable parallelism among grasses. Understanding their genetic basis is important to understanding what causes parallelism and convergence.

CONCLUSIONS

I have tried to show in this brief essay that a) the grasses are diverse and remarkably well-known taxonomically; b) developmental descriptions for an increasing number of taxa provide the basis for precise comparisons of phenotypes; c) genetic mapping can identify potential genes underlying phenotypic variation; d) gene expression studies are beginning to identify genes that may control the morphological diversity. Thus the diversity seen in the field and in the herbarium by the alpha taxonomist can be re-described at the developmental and genetic level.

Good taxonomy relies entirely on the careful and critical evaluation of taxonomic characters-specifically, morphology. Phylogenetic data are powerful precisely because they help taxonomists understand morphology. By knowing which species are closely related to which other ones we can test hypotheses of morphological synapomorphies (homology) and homoplasies. Developmental, genetic, and genomic data permit an even deeper analysis of characters, their similarities and their differences. The grasses thus provide an example of the power of 21st century biology for helping to understand the basis of morphological diversity.

ACKNOWLEDGEMENTS

I would like to thank Fernando Zuloaga and Raúl Pozner for the opportunity to write this review. Work described herein was supported in part by the National Science Foundation.

BIBLIOGRAPHY

Aliscioni, S. A., L. M. Giussani, F. O. Zuloaga & E. A. Kellogg. 2003. A molecular phylogeny of *Panicum* (Poaceae: Paniceae): tests of monophyly and phylogenetic placement within the Panicoideae. *Amer. J. Bot.* 90: 796-821.

Beadle, G. W. 1939. Teosinte and the origin of maize. *J. Heredity* 30: 245-247.

Beadle, G. W. 1980. The ancestry of corn. *Sci. Amer*. 242: 112-119.

Blattner, F. R. 2004. Phylogenetic analysis of *Hordeum* (Poaceae) as inferred by nuclear rDNA ITS sequences. *Molec. Phylog. Evol.* 33: 289-299.

Bortiri, E., G. Chuck, E. Vollbrecht, T. Rocheford, R. Martienssen & S. Hake. 2006. *ramosa2* encodes a LATERAL ORGAN BOUNDARY domain protein that determines the fate of stem cells in branch meristems of maize. *Pl. Cell* 18: 574-585.

Brown, W. V. 1977. The Kranz syndrome and its subtypes in grass systematics. *Mem. Torrey Bot. Club* 23: 1-97.

Cacharrón, J., H. Saedler & G. Theissen. 1999. Expression of MADS box genes *ZMM8* and *ZMM14* during inflorescence development of *Zea mays* discriminates between the upper and the lower floret of each spikelet. *Developm. Genes Evol.* 209: 411-420.

Camara-Hernandez, J. & G. H. Rua. 1991. The synflorescence of Poaceae. *Beitr. Biol. Pflanzen* 66: 297-311.

Clayton, W. D., K. T. Harman, & H. Williamson. 2002 onwards. World grass species: descriptions, identification and information retrieval. Royal Botanic Gardens, Kew.

Clayton, W. D. & S. A. Renvoize. 1986. Genera graminum. Her Majesty's Stationery Office, London.

Columbus, J. T., E. A. Friar, C. W. Hamilton, J. M. Porter, L. M. Prince, & M. G. Simpson, editors. In press. Monocots: comparative biology and evolution. Rancho Santa Ana Botanic Garden, Claremont, California, USA.

Columbus, J. T, M. S. Kinney, R. Pant, & M. E. Sigueiros-Delgado. 1998. Cladistic parsimony analysis of internal transcribed spacer region (nrDNA) sequences of *Bouteloua* and relatives (Gramineae: Chloridoideae). *Aliso* 17: 99-130.

Dallwitz, M. J., T. A. Paine, & E.J. Zurcher. 1999 onwards. User's guide to the DELTA Editor. http://delta-intkey.com.

del Pilar Schneider, M., & A. C. Vegetti. 1992. The synflorescence in species of Sorghinae (Andropogoneae-Poaceae). *Beitr. Biol. Pflanzen.* 67: 225-239.

Doebley, J., & L. Lukens. 1998. Transcriptional regulators and the evolution of plant form. *Pl. Cell* 10: 1075-1082.

Doebley, J. & A. Stec. 1991. Genetic analysis of the morphological differences between maize and teosinte. *Genetics* 129: 285-295.

Doebley, J. & A. Stec. 1993. Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. *Genetics* 134: 559-570.

Doust, A.N., K.M. Devos, M.D. Gadberry, M.D. Gale, & E.A. Kellogg. 2004. Genetic control of branching in foxtail millet. *Proc. Natl. Acad. Sci. USA* 101: 9045-9050.

Doust, A. N., K. M. Devos, M. D. Gadberry, M. D. Gale, & E. A. Kellogg. 2005. The genetic basis for inflorescence variation between foxtail and green millet (Poaceae). *Genetics* 169: 1659-1672.

Doust, A. N. & E. A. Kellogg. 2006. Genotype-environment interactions for branching in the weed green millet (*Setaria viridis*) and the crop foxtail millet (*S. italica*) (Poaceae). *Molec. Ecol*. 16: 1335-1349.

Evans, M. W. 1940. Developmental morphology of the growing point of the shoot and the inflorescence in grasses. *J. Agric. Res*. 61: 481-520.

Finot, V. L., P. M. Peterson, F. O. Zuloaga, R. J. Soreng & O. Matthei. 2005. A revision of *Trisetum* (Poaceae: Pooideae: Aveninae) in South America. *Ann. Missouri Bot. Gard.* 92: 533-568.

Fu, D., P. Szucs, L. Yan, M. Helguera, J. S. Skinner, J. von Zitzewitz, P. M. Hayes & J. Dubcovsky. 2005. Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. *Molec. Genet. Genomics* 273: 54-65.

Gale, M. D. & K. M. Devos. 1998. Plant comparative genomics after 10 years. *Science* 282: 656-659.

Gallavotti, A., Q. Zhao, J. Kyozuka, R. B. Meeley, M. K. Ritter, J. F. Doebley, M. E. Pe & R. J. Schmidt. 2004. The role of *barren stalk1* in the architecture of maize. *Nature* 432: 630-635.

Gillespie, L. J. & R. J. Soreng. 2005. A phylogenetic analysis of the bluegrass genus *Poa* based on cpDNA restriction site data. *Syst. Bot.* 30: 84-105.

Giussani, L. M., J. H. Cota, F. O. Zuloaga & E. A. Kellogg. 2002. A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C_4 photosynthesis. *Amer. J. Bot*. 88: 1993-2012.

Goff, S. A. et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296: 92-100.

Grass Phylogeny Working Group. 2001. Phylogeny and subfamilial classification of the Poaceae. *Ann. Missouri Bot. Gard.* 88: 373-457.

Grebenstein, B., M. Röser, W. Sauer & V. Hemleben. 1998. Molecular phylogenetic relationships in Aveneae (Poaceae) species and other grasses as inferred from ITS1 and ITS2 rDNA sequences. *Pl. Syst. Evol.* 213: 233-250.

Hatch, M. D. 1999. C_4 photosynthesis: a historical overview. Pages 17-46 in R.F. Sage and R.K. Monson (eds.), C_4 Plant Biology. Academic Press, San Diego, California, USA.

Hattersley, P. W. & L. Watson. 1975. Anatomical parameters for predicting photosynthetic pathways of grass leaves: the 'maximum lateral cell count' and the 'maximum cells distant count'. *Phytomorphology* 25: 325-333.

Hattersley, P. W. & L. Watson. 1992. Diversification of photosynthesis. Pages 38-116 in G.P. Chapman (ed.), *Grass evolution and domestication.* Cambridge University Press, Cambridge, UK.

Helfgott, D. M. & R. J. Mason-Gamer. 2004. The evolution of North American *Elymus* (Triticeae, Poaceae) allotetraploids: evidence from phosphoenolpyruvate carboxylase gene sequences. *Syst. Bot.* 29: 850-861.

Jacques-Félix, H. 1961. Observations sur la variabilité morphologique de *Coix lacryma-jobi. J. Agric. Trop. Bot. Appl*. 8: 44-56.

Jensen, L. B., J. R. Andersen, U. Frei, Y. Xing, C. Taylor, P. B. Holm & T. Lubberstedt. 2005. QTL mapping of vernalization response in perennial ryegrass (*Lolium perenne* L.) reveals co-location with an orthologue of wheat *VRN1*. *Theor. Appl. Genet.* 110: 527-536.

Jeon, J. S., S. Jang, S. Lee, J. Nam, C. Kim, S. H. Lee, Y. Y. Chung, S. R. Kim, Y. H. Lee, Y. G. Cho & G. An. 2000. *Leafy hull sterile1* is a homeotic mutation in a rice MADS box gene affecting rice flower development. *Pl. Cell* 12: 871-884.

Kellogg, E. A. 2000. Molecular and morphological evolution in Andropogoneae. Pages 149-158 in S.W.L. Jacobs and J. E. Everett (eds.) *Grasses: Systematics and evolution.* CSIRO, Melbourne.

Kellogg, E. A., K. M. Hiser, & A. N. Doust. 2004. Taxonomy, phylogeny, and inflorescence development of the genus *Ixophorus* (Panicoideae: Poaceae). *Int. J. Plant Sci.* 165: 1089-1105.

Kellogg, E. A. & L. Watson. 1993. Phylogenetic studies of a large data set. I. Bambusoideae, Andropogonodae, and Pooideae. *Bot. Rev.* 59: 273-343.

LeRoux, L. G. & E. A. Kellogg. 1999. Floral development and the formation of unisexual spikelets in the Andropogoneae (Poaceae). *Amer. J. Bot.* 86: 354-366.

Litt, A. & V. F. Irish. 2003. Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* 165: 821-833.

Lynch, M. & B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland, Massachusetts.

Malcomber, S. T. & E. A. Kellogg. 2004. Heterogeneous expression patterns of the *SEPALLATA*-like gene *LEAFY HULL STERILE1* (*LHS1*) in grasses (Poaceae) suggets separate roles in meristem determinacy, palea/lemma identity, and flower sexuality. *Pl. Cell* 16: 1692-1706.

Malcomber, S. T. & E. A. Kellogg. 2005. *SEPALLATA* gene diversification: brave new whorls. *Trends.Plant Sci.* 10: 427-435.

Malcomber, S. T., J. C. Preston, R. Reinheimer, J.Kossuth & E A. Kellogg. 2006. Developmental gene evolution and the origin of grass inflorescence diversity. *Advances. Bot. Res.* in press.

Martin, M., P. E. Jameson, A. F. Mark, E. C. Yeung & R. P. Pharis. 1993. Early panicle development in *Chionochloa macra* plants induced to flower by 2,2 dimethyl gibberellin A_4 or long days. *New Zealand. J. Bot*. 31: 193-201.

Mason-Gamer, R. J. 2005. The ß-amylase genes of grasses and a phylogenetic analysis of the Triticeae. *Amer. J. Bot.* 92: 1045-1058.

Molina, A. M. & Z. E. R. De Agrasar. 2004. Taxonomic revision of the species of the genus *Chloris* (Poaceae: Chloridoideae) in South America. *Candollea* 59: 347-427.

Neves, S. S., G. Swire-Clark, K. W. Hilu & W. V. Baird. 2005. Phylogeny of *Eleusine* (Poaceae: Chloridoideae) based on nuclear ITS and plastid *trnT-trnF* sequences. *Molec. Phylog. Evol.* 35: 395-419.

Orr, A. R., K. Mullen, D. Klaahsen & M. D. Sundberg. 2002. Inflorescence development in a high-altitude annual Mexican teosinte (Poaceae). *Amer. J. Bot.* 89: 1730-1740.

Patterson, C. 1982. Morphological characters and homology. Pages 21-74, in K.A. Joysey & A.E. Friday (ed.), Problems of phylogeny reconstruction. Academic Press, London.

Prasad, K., P. Sriram, C. S. Kumar, K. Kushlappa & U.

Vijayraghavan. 2001. Ectopic expression of rice *OsMADS1* reveals a role in specifying the lemma and palea, grass floral organs analogous to sepals. *Developm. Genes Evol*. 211: 281-290.

Reinheimer, R., S. T. Malcomber & E. A. Kellogg. 2006. Evidence for distinct roles of the *SEPALLATA* gene *LEAFY HULL STERILE*1 in *Eleusine indica* and *Megathyrsus maximus* (Poaceae). *Evol. Developm.* 8: 293-303.

Reinheimer, R., R. Pozner & A. C. Vegetti. 2005. Inflorescence, spikelet, and floral development in *Panicum maximum* and *Urochloa plantaginea* (Poaceae). *Amer. J. Bot.* 92: 565-575.

Rua, G. H. & R. Reinheimer. Homology of inflorescences revisited. Unpublished MS.

Saarela, J. M., P. M. Peterson, R. J. Soreng & R. E. Chapman. 2003. A taxonomic revision of the eastern North American and eastern Asian disjunct genus *Brachyelytrum* (Poaceae): evidence from morphology, phytogeography and AFLPs. *Syst. Bot.* 28: 674-692.

Sharp, D. & B. K. Simon. 2002. AusGrass: Grasses of Australia. CSIRO/ABRS.

Sinha, N. R. & E. A. Kellogg. 1996. Parallelism and diversity in multiple origins of C_4 photosynthesis in the grass family. *Amer. J. Bot.* 83: 1458-1470.

Soreng, R. J., G. Davidse, P. M. Peterson, F. O. Zuloaga, E. J. Judziewicz, T. S. Filgueiras & O. Morrone. 2000 onwards. Catalogue of New World grasses (Poaceae). http://mobot.mobot.org/W3T/Search/nwgc.html

Soreng, R. J. & J. I. Davis. 2000. Phylogenetic structure in Poaceae subfamily Pooideae as inferred from molecular and morphological characters: misclassification versus reticulation, pp. 61-74, in S.W.L. Jacobs and J. Everett (eds.), *Grasses: systematics and evolution.* CSIRO, Melbourne, Australia.

Spangler, R. E. 2003. Taxonomy of *Sarga*, *Sorghum* and *Vacoparis* (Poaceae: Andropogoneae). *Austral. Syst. Bot.* 16: 279-299.

Torrecilla, P., J. A. L. Rodriguez, D. Stancik & P. Catalan. 2003. Systematics of *Festuca* L. *sects*. *Eskia* Willk., *Pseudatropis* Kriv., *Amphigenes* (Janka) Tzvel., *Pseudoscariosa* Kriv. and *Scariosae* Hack. based on analysis of morphological characters and DNA sequences. *Pl. Syst. Evol.* 239:113-139.

Van den Borre, A. & L. Watson. 1997. On the classification of the Chloridoideae (Poaceae). *Aust. Syst. Bot.* 10: 491-531.

Vegetti, A.C. 1992. Contribution to the study of the synflorescence in *Themeda* Forssk. (Andropogoneae-Poaceae). *Beitr. Biol. Pflanzen* 67: 251-258.

Vegetti, A. C. 2000. Typology of synflorescences in Oryzeae (Poaceae). *Phyton* 40: 71-88.

Vegetti, A. C. & F. Weberling. 1996. The structure of the paracladial zone in Poaceae. *Taxon* 45: 453-460.

Vollbrecht, E., P. S. Springer, L. Goh, E .S. Buckler IV & R. Martienssen. 2005. Architecture of floral branch systems in maize and related grasses. *Nature* 436: 1119-1126.

Wang, R.-L., A. Stec, J. Hey, L. Lukens & J. Doebley. 1999. The limits of selection during maize domestication. *Nature* 398: 236-239.

Whipple, C. J., P. Ciceri, C. M. Padilla, B. A. Ambrose, S. L. Bandong & R. J. Schmidt. 2004. Conservation of B-class floral homeotic gene function between maize and Arabidopsis. *Development* 131: 6083-6091.

Whipple, C. J. & R. J. Schmidt. 2006. Genetics of grass flower development. *Advances Bot. Res*.: In press.

Yan, L., A. Loukoianov, A. Blechl, G. Tranquilli, W. Ramakrishna, P. SanMiguel, J. L. Bennetzen, V. Echenique & J. Dubcovsky. 2004. The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* 303: 1640-1644.

Yan, L., A. Loukoianov, G. Tranquilli, M. Helguera, T. Fahima, & J. Dubcovsky. 2003. Positional cloning of the wheat vernalization gene *VRN1*. *Proc. Natl. Acad. Sci. USA* 100: 6263-6268.

Yu, J., et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L., ssp. *indica*). *Science* 296: 79-92.

Zuloaga, F., & O. Morrone. 2005. Revision de las especies de *Paspalum* para America del sur Austral. Missouri Botanical Garden Press, St. Louis, Missouri, USA.