TITLE: **International Genetic, Cytogenetic and Germplasm Resources for Cotton Genomics and Genetic Improvement**

DISCIPLINE: **Genomics and Biotechnology**

AUTHORS:

D.M. Stelly (Corresponding author)
Texas A&M University
Department of Soil & Crop Sciences
370 Olsen Boulevard
College Station, TX 77843-2474
Phone: 979-845-2745
Fax: 979-845-0456 (Attn. David Stelly)
Email: stelly@tamu.edu

J.-M. Lacape
CIRAD
Avenue Agropolis
TA A-96/03
Montpellier, 34398
France

D.E.G.A. Dessauw
CIRAD
TA B-10/02
Avenue Agropolis
Montpellier, 34398
France

R. Kohel
USDA-ARS
2765 F & B Road
College Station, TX 77845

G. Mergeai
Faculté des Sciences Agronomiques de Gembloux,
Passage des Déportés,
2-5030,Gembloux,
Belgium

M. Sanamyan
National University of Uzbekistan
Department of Biology
Laboratory of Cotton Genetics
Vuzgorodok, Tashkent
Tashkent, 700174
Uzbekistan

I.Y. Abdurakhmonov
Institute of Genetics and Plant Experimental Biology
Yuqori-Yuz, Kibray region
Tashkent, 702151
Uzbekistan

T. Zhang
Nanjing Agriculture University
National Key Laboratory of Crop Genetics and Germplasm
Enhancement and Cotton Research Institute
Nanjing, 210095
China

K. Wang
Cotton Research Institute
Chinese Academy of Agricultural Sciences
Avenue Huanghe
Anyang, 455000
China

B. Zhou
Nanjing Agriculture University
National Key Laboratory of Crop Genetics and Germplasm
Enhancement and Cotton Research Institute
Nanjing, 210095
China

S. Saha
USDA-ARS
Genetics and Precision Agriculture Research Unit
810 Hwy 12 East
Mississippi State, MS 39762

J. Frelichowski
USDA-ARS
Crop Germplasm Research Unit
2881 F&B Road
College Station, TX 77845

ABBREVIATIONS: ICGI, International Cotton Genome Initiative; RIL, recombinant inbred line; USDA, United States Department of Agriculture; CIRAD, Centre de coopération Internationale en Recherche Agronomique pour le Développement; MAS, marker assisted selection

International Genetic, Cytogenetic and Germplasm Resources for Cotton Genomics and Genetic Improvement

ABSTRACT

A general understanding of the nature, uses and need of cotton germplasm collections is important to maintaining support for them and their utilization. Such support is vital to addressing the future challenges and needs of cotton, which will be many and varied. The solutions to many of these challenges will be genetic, and involve the scientific utilization of cotton genetic and germplasm resources. In this presentation, we will look at a number of key sites from around the world, to illustrate the types of genetic, cytogenetic and germplasm collections that exist. The taxa that constitute cotton germplasm can be categorized into primary, secondary and tertiary "gene pools" that progressively connote both the relative difficulty of introgression and their likelihood of harboring new genetic diversity. We will exemplify past and ongoing usage. Patterns of introgression reflect the gene-pool classifications. The International Cotton Genome Initiative (ICGI) is facilitating globalization of scientific interactions that affect these collections, their access, and their contributions to cotton science and genetic improvement – these will benefit cotton worldwide. Lastly, we will look at what the near future may bring in the way of new paradigms for utilization of genetic stocks, cytogenetic stocks, germplasm and molecular markers in cotton science and genetic improvement.

KEY WORDS:

cotton, germplasm, breeding, improvement, diversity

"International Genetic, Cytogenetic and Germplasm Resources for Cotton Genomics and Genetic Improvement". As illustrated by many of the presentations at the WCRC, the scientific and practical inter-relationships between germplasm, genomics and genetic improvement are numerous, and often bi-directionally informative. Accordingly, this talk is titled "*International Genetic, Cytogenetic and Germplasm Resources for Cotton Genomics and Genetic Improvement"*. Moreover, it should be noted that the maintenance, development, and uses of cotton germplasm collections are deeply rooted in the International Cotton Genome Initiative (ICGI), in which they constitute one of the five existing Work Groups (see http://icgi.tamu.edu/).

Main topics. It is important for cotton, as for all economically important plants, that broad recognition exist for the need of germplasm resources, and their crucial roles in genetic improvement. It is also important that everyone, not just curators, geneticists and breeders, be wary of the biological fragility of germplasm collections, so that our institutions and governments are vigilant and consistent in their support for germplasm collections. Adequate appreciation of these values can perhaps be best gained through some comprehension of the absolute need for genetic diversity in genetic improvement efforts, an awareness of the amount of diversity in the cotton germplasm collections, challenges in using the germplasm, and the need for genetic improvement as part of a successful future for cotton, its producers and related industries.

What is "Germplasm"? Let us start with the concept of "germ plasm", or as it is often written now "germplasm". Etymologically, "Germ" refers to something that initiates development or serves as an origin, and "-plasm" refers to something that is formative or a formed material. But, more importantly, the concept of "germplasm" traces back to the latter part of the 19th century, when the nature of reproduction and inheritance were hotly debated

topics, and the subject of intense research fueled by advances in optics and cytological stains. One of the great advances was the "Germplasm Theory" of August Weismann (ca. 1880), which soundly countered the "Theory of Acquired Characteristics" advocated by Lysenko. In simple terms, Weismann said that the "*soma*" or body was totally separate from the "*germ*", and only the latter gave rise to the next generation, through to equal contributions from male and female. In effect, he said that modifications of the *soma* had no effect on the *germ*, and thus there is no transmission of acquired characteristics. In alluding to "germplasm" today, we extend the concept of "germ" to the genus level, including all species in the genus *Gossypium*.

Genetic fitness as a "Landscape". To appreciate genetic diversity and the importance of germplasm as a source of genetic diversity for genetic improvement, it is worthwhile to conceptualize genetic fitness as a geographic "landscape", with different species occupying the mountain tops, i.e., "adaptive peaks", *sensu* Sewall Wright. The "adaptive peak" of fitness (each species) is conferred *collectively* by the set of genes, in that environment, at least. The analogy implies that [1] genes work as teams and traits or performance could be very similar, even if genes are different; and [2] it may be quite difficult to move a gene from one peak to another, because the level of fitness in the intervening generations will be very low, i.e., across a fitness "valley" where the generations will be potentially weak or inviable. The analogy implies, too, that once transferred, an alien gene's functional ramifications *might* be quite different than expected, i.e., due to interactions with a new set of genes. Perhaps more important, however, is that while it may in some cases be possible to assess a potential source of alien genes according to its performance in its original genetic background, the effects of some genes will change once they are put into a new genetic background. This phenomenon, *epistasis*, is well known, and more common in "wide crosses" between less related types or species. The implications of

epistasis, however, seem to be greatly under-appreciated if not ignored when reviewers of proposals for germplasm introgression demand a priori evidence of a valuable gene in the source germplasm. The "Catch-22" is that introgression of such genes involved in such epistasis would be requisite to their detection. This might be particularly important for multigenic traits that have been under selection pressure in the cultivated forms. Analogous difficulties extend to prediction of "transgressive segregation", i.e., the occurrence of traits more extreme than found in either parent. Thus, while *a priori* evidence of value is always nice, one must challenge its pre-existence as a stipulation for investments in introgression, because empirically we can predict that beneficial genetic variation exists in different species. Large amounts of recent data from cotton and tomato strongly support this concept, e.g., Saranaga et al. (2001); Gur and Zamir (2004); Lacape et al. (2005); Jenkins et al. (2006); Saha et al. (2006); and He et al. (2007).

Germplasm or gene "pools". We know that genetic diversity is essential to genetic progress from breeding, but too much of good thing can be a hindrance. The facility and predictability of results from alien germplasm transfer is highest when it is closely related to the target. Thus, Harlan and DeWet (1971) defined "gene pools" as a convenient way of categorizing potential sources of germplasm for a given target, e.g., an agronomic crop. Briefly, germplasm in the "Primary Gene Pool" will hybridize and breed readily with the target species, i.e., using conventional techniques. For the "Secondary Gene Pool" however, the processes are more difficult and special methods are required to hybridize and breed. Species in the "Tertiary Gene Pool" are problematic, and are generally very difficult or impossible to hybridize or introgress into the target; and if so, extreme measures are required.

Categorization of *Gossypium* into gene pools. The concept of "Gene Pools" extends meaningfully to *Gossypium*. For Upland and Pima types of cottons, the primary gene pool

contains all of the five “AD” species. Among these, hybrids and introgressed materials are produced readily. This grouping reflects differences in chromosome number per genome, where “genome” refers to a complete set of chromosomes. Upland and Pima cottons, like all “(AD)” species have 26 chromosomes per haploid genome, as in pollen, whereas all other *Gossypium* species have just 13 chromosomes per genome. The “AD” species diverged during evolution from an ancestor formed by uniting genomes very similar to extant A and D genomes, so even today strong genetic similarities exist in the AD and the A and D genomes. Nevertheless, the difference in chromosome numbers leads to extensive sterility in hybrids, so that exceptional efforts are required to introgress germplasm from species with “A” or “D” genomes, as well as “B” and “F” genomes. Introgression is even more difficult from species containing the other genomes, i.e., C, E, G, and K, which thus constitute the “Tertiary Gene Pool”. In preparing for sexual reproduction, the meiotic interactions required between homologous chromosomes for successful reproduction are steeply reduced between them and the AD-genome chromosomes. These lead to imbalanced chromosome numbers during sexual reproduction, a lot of sterility and very low rates of genetic exchange. Extreme methods are required to address these and other difficulties from these species into any of the AD species. Thus, we find that the “gene pools” relate to the genomic groups defined according to chromosome number and ability of their chromosomes pair and recombine during meiosis.

Geographic distribution of cotton germplasm. When genomic groups are visualized according to their world distributions, their distributional correlation with geography is readily apparent. Thus, the gene pools also correlate well with geographic distributions, as well as phylogenetic relatedness, e.g., as deduced by DNA sequence comparisons.

The *Gossypium* mountain range. Let us consider the situation more specifically, e.g., the *Gossypium* mountain range. Each species with a 13-chromosome haploid genome contains and estimated 30,000 genes, working as a team to achieve high levels of fitness and local adaptation. There remains some variation among individuals in each population, though the amount can differ widely. By definition, however, there are major differences between species, especially across genomic groups. If only 10% differ, ~3000 new alleles would potentially be available for transfer.

Elite types – a peak in the *Gossypium* mountain range. We can extend this *Gossypium* Mountain Range analogy to our target of interest, e.g., the cultivars of *G. hirsutum* cottons. These are a very select group within the species, and the vast majority of *G. hirsutum* genotypes differ in numerous ways from the elite individuals. For our purposes, we can envision elite cultivars and breeding lines as at the very top of the *G. hirsutum* peak.

Gene transfer – going from one peak to another. What do we anticipate genetically during alien gene transfer? What sorts of genetic results do we expect? The patterns of expectation conform well to this model, too. Each species is genetically unique, in part by chance fixation during evolution, but also because it arose by way of selection for fitness its "environment". Given 30,000+ genes per genome, extant species are expected to differ at many genes, different forms of which are alluded to as "alleles". Alien gene transfer, if successful, can result in many new genes becoming available for breeding. Consider gene transfer from *G. tomentosum* (Hawaiian) and *G. mustelinum* (Brazilian) to *G. hirsutum*, where the former is more closely related to Upland cotton. Moving genes between species entails a difficult process, because "regions" of extremely low genetic fitness must be traversed. Natural selection can directly affect many genes, and lead to their loss or reduction in frequency, as well as many other

genes that are “linked” to them, i.e., near them on the same chromosome. Albino seedlings that occur during *G. mustelinum* introgression visually exemplify this phenomenon, which has far-reaching impact on breeding. Interspecific germplasm transfer is a genetically complicated process that requires many recombination events and multiple generations to disassociate desirable and undesirable alleles. As in the model, the more genetically distant the alien gene source, the deeper the valley, and the more difficult the transfer. On the other hand, the potential gains, genetically and economically, would also be higher. Thus, gene transfer from a more distant species, e.g. *G. mustelinum*, would be expected from more difficult, less genomically comprehensive, and more time-consuming than from a more closely related species, e.g. *G. tomentosum*, but also potentially more valuable.

Germplasm -- a key for cotton improvements. Essentially all naturally existing traits of cotton are subject to genetic effects and thus amenable genetic improvement through germplasm introgression and breeding. These include but are not limited to the following classes of traits: morphology, structure and habit, biochemical constitution, stress resistances and tolerances to biotic and abiotic challenges, plant fertility and reproduction traits, seed oil content and profiles, and fiber quality and yield. In some cases, simply inherited plant and resistance traits offer quantum, high-value advantages. In most cases, however, the advances are incremental, because most traits of greatest importance, such as fiber yield and quality are subject to the effects of many loci and their many interactions with each other and the environment, as well as environmental effects, per se.

Germplasm collection components. The above indicate that there are very practical reasons to have comprehensive germplasm collections that include all of the germplasm pools. For each of the major crops, like cotton, there is also a need for long-term funding of projects

that introgress germplasm from various species and discover valuable effects. Additional concerns are that wild cotton populations have diminished greatly already and continue to diminish at an alarming rate. Each of the major cotton germplasm collections includes thousands of accessions. These can be separated into three main categories, [1] naturally occurring types, [2] new off-types and mutants, and [3] artificially constructed or engineered individual, lines or populations. The largest numbers of accessions, by far, fall into the first and third categories. The first includes species, racestocks, biotypes, and so on, which are the sources of most new variability. The third includes germplasm from research and breeding efforts, including obsolete cultivars from conventional breeding, chromosome substitution and recombinant inbred populations. The second includes spontaneous and induced mutations, i.e., another source of variability.

Examples of national collections. World cotton germplasm accessions total about 50,000, about half or more of the accessions are held in the collections of Uzbekistan (17K), China (8K), the USA (9K), and France (3K) (Bernard and Hau, 2007).

Genus coverage / USDA & CIRAD collections. Natural and man-made lines from breeding and mapping predominate the collections, and only relatively small numbers of accessions involve species outside of the four cultivated species *G. hirsutum, G. barbadense, G. arboreum* and *G. herbaceum*. This is illustrated in a side-by-side comparison of USA (USDA) and French (CIRAD) collections (Table 1).

Factors affecting ease of use. The process of germplasm introgression into cotton is influenced by characteristics of the germplasm donor, relative to the target, and by certain aspects of a targeted trait. Juvenility and/or photoperiodicity of an alien germplasm source can significantly delay the initiation of introgression, and impede overall progress. In interspecific

and wider crosses, the effectiveness can be increasingly marginalized by extensive “breakdown”. Differences in ploidy level and the meiotic (dis)similarity of genomic group(s) are common impediments. Genetic incompatibilities can also affect introgression, such as the complementary hybrid lethals in the D3-genomes of *G. davidsonii* and *G. klotzchianum*. It is noteworthy that a trait is more likely to be targeted for improvement by introgression given prior evidence of its importance, the existence of natural variation, heritability, simple inheritance, and evaluation methods that are quick, inexpensive and effective. Linkages between genes conferring desirable and undesirable traits can greatly influence the usefulness of introgression products, so prior or concomitant genetic dissection and linkage mapping are often bundled with introgression efforts. The availability of good markers for marker-assisted selection makes possible the introgression of beneficial genes for traits that are difficult or expensive to evaluate, or plagued by linkage to an undesirable gene.

Traits involving introgression from primary and secondary gene pools. As might be expected, the Primary Gene Pool has been used most extensively as for germplasm introgression, followed by the Secondary Gene Pool. Germplasm-based introgression has been used to increase diversity for fiber yield and quality traits, as well as morphological, fertility and pathogen resistance. Examples include the following:

- Blight resistance (*Xanthomonas*, i.e., *B* genes)
- Glanding; nectaries; leaf shapes…; pubescence
- Male sterility (genic, CMS)
- Fiber yield, qualities, color

Usage patterns - primary gene pool. A number of different approaches have been used with varying degrees of effectiveness to introgress primary gene pool germplasm. They include

F2, F2-derived inbreds, RILs, backcross-inbreds and isolines, chromosome substitution, complex populations with and without male sterility and, soon, chromosome-specific RILs.

Primary gene pool -- complementarity among patterns of usage. The methods offer complementarity by varying widely in the time and labor required, and their susceptibility to constraints imposed by genetic and biological problems that about in wide crosses. While we still have much to learn about these problems and their interactions with various breeding schemes, it is implicit that some progress can be made with virtually any breeding method, and also that there are marked differences among the breeding schemes. When severe, these hindrances constitute barriers, and alternative methods must be deployed. In contemporary introgression projects, molecular markers are often used to create linkage maps and/or associate map loci with the genes significantly affecting trait governance.

Germplasm introgression from secondary and tertiary gene pools. In cotton, targets for improvement by means of germplasm introgression have typically included lint yield, fiber quality, manipulation of the cytoplasmic-genic male fertility, pathogen, pest or abiotic stress resistance. Three examples of recent and ongoing usage to address rapidly escalating pathogen problems include the Burewala Strain of Cotton Virus in Pakistan (Naveed and Zahid, 2007), reniform nematodes in Belgium and the USA (Robinson et al., 2006), and fusarium in Australia (McFadden et al., 2004). The prospect of manipulating the patterns of glanding through introduction of the so-called glanded-plant glandless-seed trait from Australian species (Altman et al. 1987; Brubaker et al., 1999; Guy Mergeai, 2006), with the ultimate aim of enhancing value of cotton seed as feed without undermining resistance to pests.

Patterns of usage -- 2^o and 3^o pools. The main patterns of germplasm introgression differ markedly for secondary and tertiary germplasm (e.g., see Brubaker et al., 1999; Mergeai,

2006; Robinson et al., 2006), yet both entail manipulations of ploidy levels (chromosome number) and genomic composition, which are often difficult and time-consuming. For the secondary gene pool, utilization usually entails synthesis of balanced or quasi-balanced ~52-chromosome synthetics (e.g., A_hAD_hD, A_hFD_hD, A_hBD_hD), followed by further hybridization and various amounts of backcrossing and/or inbreeding. These typically lead to significant levels of homologous recombination in some regions of secondary gene pool genomes, if not most. Chromosome addition lines have been derived in some instances, e.g., some African species. For tertiary germplasm, the more common route has been synthesis of hexaploids, followed by backcrossing to derive monosomic addition lines. Recombination, if any occurs, is quite limited, especially with Australian genomes, and its distribution remains to be characterized.

Solving challenges and answering opportunities for cotton on global scale. The future of cotton research holds a number of anticipatable challenges and opportunities, and probably many that are not so obvious now. Clearly, cotton must compete well against man-made fibers for marketshare. New resistance genes must be identified and deployed to offset ever-evolving pathogens and pests, as well as to reduce costs and hazards of production associated with chemical control methods. They might also be used to reduce reliance on expensive GMO methods, and instead of GMOs in countries or cultivation regimes where GMOs are not allowed, and/or for resistance traits not yet available through GMOs. Cotton production systems must become increasingly sustainable, ecologically, environmentally "friendly" and healthy for the producer and farm worker. Water usage must be reduced, as water is almost certain to continue becoming less available and more costly. Moreover, with increased population and urbanization, the urban populations may impose relatively severe requirements

for changes that affect water usage and quality. There are opportunities to enhance economic yields by expanding the value of cottonseed, e.g., as a source of higher quality vegetable oil, better feed and as a bio-fuel or bio-diesel. Genetic modifications will be key to addressing the above, and so our genetic bank, the world's cotton germplasm collections, are certain to be a major contributor.

The future. As we look to the future, it is important that we maintain focus on [1] maintaining and improving the collections, [2] sharing them and thus keeping them well "backed up", [3] characterizing them, [4] studying and improving the germplasm transfer methods, and [5] actually transferring germplasm and "pre-breeding" to put new genetic variation into genetic backgrounds that are fit for use by applied breeding programs. Increasingly, high-throughput markers will be used not only to better characterize and maintain the collections but also to conduct interspecific introgression of germplasm. In addition to enabling genetic dissection of complex traits, the high-throughput markers offer the prospect of detailing the ramifications of genetic and cytogenetic hurdles to introgression, as well as the definition of the localization and biological basis of those barriers. Once the genes are introgressed, the markers will increasingly be used in concert with phenotypic breeding, i.e., for MAS that will facilitate simultaneous manipulation of multiple genes for cotton improvement.

REFERENCES

Altman, D.W., D.M. Stelly, and R.J. Kohel. 1987. Introgression of the glanded-plant and glandless-seed trait from *Gossypium sturtianum* Willis into cultivated Upland cotton using ovule culture. Crop Sci. 27:880-884.

Brubaker, C.L., A.H.D. Brown, J.McD. Stewart, M.J. Kilby, and J.P. Grace. 1999. Production of fertile hybrid germplasm with diploid Australian *Gossypium* species for cotton improvement. Euphytica 108: 199–213.

Dessauw D., and B. Hau. 2007. Inventory and history of the CIRAD cotton (*Gossypium* spp.) germplasm collection. Plant Genetic Resources Newsl. 147: 52-58.

Gur, A., and D. Zamir. 2004. Unused natural variation can lift yield barriers in plant breeding. Publ. Libr. Sci. (Biol.) 2(10): e245.

Jenkins, J.N., J. Wu, J.C. McCarty, S. Saha, O. Gutiérrez, R. Hayes, and D.M. Stelly. 2006. Genetic effects of thirteen *Gossypium barbadense* L. chromosome substitution lines in topcrosses with Upland cotton cultivars: I. Yield and yield components. Crop Sci 46: 1169-1178.

Harlan, J.R., and J.M.J. de Wet. 1971. Toward a rational classification of cultivated plants. Taxon 20: 509-517.

He, D-H, Z-X. Lin, X-L. Zhang, Y-C. Nie, X-P. Guo, Y-X. Zhang, and W. Li. 2007. QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* x *Gossypium barbadense*. Euphytica 153: 181-197.

Koto, E. 1983. Tentative d'utilisation de l'espece sauvage diploıde *Gossypium longicalyx* pour l'amelioration de l'espece cultive´e tetraploıde *G. hirsutum* L. par la methode des lignees d'addition et de substitution. Ph.D. thesis. Universite Pierre et Marie Curie. Paris. 91 pp.

Lacape, J-M., and T.B. Nguyen. 2005. Mapping quantitative trait loci associated with leaf and stem pubescence in cotton. J. Hered. 2005:96(4): 441–444.

McFadden, H., D. Beasley, and C.L. Brubaker. 2004. Assessment of *Gossypium sturtianum* and *G. australe* as potential sources of Fusarium wilt resistance to cotton. Euphytica 138: 61–72.

Mergeai, G. 2006. Introgressions interspécifiques chez le cotonnier. Cahiers Agric. 15: 135-143.

Naveed, M., and I.A. Zahid. 2007. Resistance Found for Burewala Strain Cotton Virus BSCV. Whitefly Newsl. Jan. 2007.

Robinson, A.F., A.A. Bell, D. Stelly, N. Dighe, and M. Menz. 2006. Progress report on introgression of reniform nematode resistance from *G. longicalyx* into Upland cotton. Proc. Beltwide Cotton Proc. Res. Conf., San Antonio, TX. 3-6 Jan. 2006. Natl. Cotton Counc. Amer., Memphis, TN.

Saha, S., J.N. Jenkins, J. Wu, J.C. McCarty, O.A. Gutierrez, R.G. Percy, R.G. Cantrell, and D.M. Stelly. 2006. Effects of chromosome specific introgression in Upland cotton on fiber and agronomic traits. Genetics 172: 1927-1938.

Saranga Y., M. Menz, C.X. Jiang, R.J. Wright, D. Yakir, and A.H. Paterson. 2001. Genomic dissection of genotype x environment interactions conferring adaptation of cotton to arid conditions. Genome Res. 11: 1988–1995.

Table 1. Summary of USA and French cotton germplasm collections, by gene pool, species, and genome (ID).

Pool	Species	ID	US-07	FR-05	Pool	Species	ID	US-07	FR-05
	G. hirsutum	(AD)1	5243	2173		*G. stocksii*	E1	4	3
	G. barbadense	(AD)2	1600	483		*G. somalense*	E2	3	3
1	*G. tomentosum*	(AD)3	28	2		*G. areysianum*	E3	2	1
	G. mustelinum	(AD)4	22	1		*G. incanum*	E4	2	2
	G. darwinii	(AD)5	142	95		*G. benadirense*	E		
	G. herbaceum	A1	223	50		*G. bricchettii*	E		
	G. arboreum	A2	2490	69		*G. vollesenii*	E		
	G. anomalum	B1	4	20		*G. sturtianum*	C1	10	3
	G. triphyllum	B2	3			*G. nandewarense*	C1N	6	
	G. capitis-viridis	B3	1	3		*G. robinsonii*	C2	11	1
	G. longicalyx	F	4	2		*G. australe*	G	29	2
	G. thurberi	D1	37	6		*G. nelsonii*	G	7	1
2	*G. armourianum*	D2-1	10	3	3	*G. bickii*	G1	6	3
	G. harknessii	D2-2	20	3		*G. anapoides*	K1		
	G. davidsonii	D3-d	31	4		*G. populifolium*	K2	22	
	G. klotzchianum	D3-k	66	45		*G. cunninghamii*	K3	7	
	G. aridum	D4	31	2		*G. pulchellum*	K4	1	
	G. raimondii	D5	56	1		*G. pilosum*	K5	5	
	G. gossypioides	D6	8	2		*G. costulatum*	K6	6	
	G. lobatum	D7	7	2		*G. enthyle*	K7		
	G. trilobum	D8	11	3		*G. exiguum*	K8		
	G. laxum	D9	9	1		*G. londonderriense*	K9		
	G. turneri	D10	8	1		*G. marchantii*	K10		
	G. schwendimanii	D11	3	1		*G. nobile*	K11		
						G. rotundifolium	K12		
						NI			79
						TOTAL		10057	3070