Discover
Natural
Fibres
Initiative

THE ICAC RECORDER

International
Cotton
Advisory
Committee

Technical
Information Section

VOL. XXXI No. 1
MARCH 2013

Update on Cotton Production Research

Contents

Introduction

The first article in this issue is about the work done in the USA on many variables of pest pressure, including estimated arthropod-related yield losses, percentage of affected area, direct control costs and insecticide sprays. Yield losses attributable to arthropods, individual pests and the cost of direct control are measured every year in a nation-wide survey under a project funded by The Cotton Foundation. The same sources have been used since 1979 to collect the same information, thus providing an excellent database on long-term changes and shifts in the criteria mentioned above. A review of the last 20 years of data show that the boll weevil and bud/bollworms in 1993/94 were responsible for 84% of all yield losses, which were estimated to amount to 6.9% of potential yields. Over the last two decades, the intensity of arthropod damage has changed significantly in the USA. The boll weevil eradication program, the planting of biotech cotton and other programs have lowered yield losses to less than 5% since 2001/02 and only 2.06% in 2012/13. The whitefly population was already on the decline in 1993/94. Now, thrips are causing more loss in yield than bollworms. There is a consensus in the cotton community in the US that Lygus bugs are on the increase, but the data do not seem to support that contention. The article provides a good view of the arthropod situation and changes in the pest pattern.

The primary objective of insect resistant biotech cotton is to reduce the use of insecticides. Refuge requirements were recommended and vigorously implemented in most countries, but not in all. Five percent unsprayed and 20% sprayed of the conventional area was a common recommendation from technology developers, but that recommendation is changing. Resistance to Cry proteins is also emerging. New technologies are coming. Dow Agrosciences is conducting trials with Wide-strike (Cry1Ac + Cry1F); Bayer has initiated trials with twin-link (Cry1Ab + Cry2Ae), and a company in India is working with Cry1Ac + Cry1EC. Barring the development of a single gene as effective as two or three stacked genes, any new technology that is eventually commercialized will have a stacked gene effect. The fact remains that when Cry1Ac was introduced, the target insects had not been exposed directly to Cry proteins. At the time of the commercialization of Bollgard II, a number of lepidopterans had already grown accustomed to Cry1Ac. The introduction of Bollgard III will have to deal with populations that have generations of experience in surviving the Cry proteins. The probability of developing cross-resistance among Cry proteins is greater than with a vegetative insecticidal protein combination. There is a need to shield the newer technologies for a longer period of time. The second article titled, 'Insecticide Use and Toxin Resistance in Biotech Cotton,' is about these issues.

Dr. Andrew Paterson, ICAC Researcher of the Year 2012, authored the third article on 'The Post-Genomic Era for Cotton.' The article explores general features of the cotton genome and fundamental messages learned from the sequences, along with new capabilities that the sequence provides to research and development. In particular, the genome sequence provides a means of coalescing many diverse data types, some of which still need to be created for cotton, to gain new understanding from otherwise disparate data. The genome sequence is not an ending but a beginning – specifically, a beginning of a new era of research and development using powerful new tools and approaches to identify and manipulate cotton genes of economic importance. The benefits of genome sequencing are improved quality, productivity, and stability; reduced input needs that improve sustainability and environmental stewardship; and value-added features tailored to human needs rather than natural adaptation. The conversion of 'sequence' to 'knowledge,' is a challenge for cotton researchers.

Nomenclature Used in the Planting Seed Industry

In the next few months, the Technical Information Section will undertake a study of the nomenclature of planting seed. Seed production and distribution is spread across organizations/departments/ministries within the governments and private sectors of various countries. Breeders develop varieties and are responsible for producing pure seed, which is used for multiplication purposes. Consequently, the step 1, 2 or 3

The ICAC RECORDER (ISSN 1022-6303) is published four times a year by the Secretariat of the International Cotton Advisory Committee, 1629 K Street, NW, Suite 702, Washington, DC 20006-1636, USA. Editor: M. Rafiq Chaudhry <rafiq@icac.org>. Subscription rate: $205.00 hard copy, $170.00 electronic version.

multiplied seed reaches the farmer's field. There is a different nomenclature used by different countries for each category of planting seed, though they mean the same quality of seed. The objective of this study is to collect information on seed categories, seed certification standards and seed producing and distribution systems. The information will be used to propose uniform categories of planting seed such that each category means the same quality standard in all countries.

Cost of Production of Cotton

The Technical Information Section of ICAC will be collecting data from countries on the cost of production of cotton in the next few months. The report will be published before the ICAC Plenary Meeting in Colombia in September 2013.

Yield Losses and Arthropod Management in the US: A Twenty-Year Review

Cotton yields are dependent on many variables, but the two with the greatest year-to-year variation are weather, particularly rain under rainfed conditions, and insect damage. Researchers have yet to develop varieties that require less water for normal growth and hence are minimally affected by year-to-year rainfall variations. Insect research, on the other hand, has taken tremendous strides forward, particularly since the development of insecticides: first, to introduce and promote the use of insecticides, and later, to reduce their use. The leading research targets are: insect biology, the stage at which the insect is most susceptible to control, the sort of chemicals to be used, and how long to use them. In most cases the cotton research community unreservedly followed the insect trends, thereby increasing emphasis on the most dangerous pests within the complex and on the most feasible ways to control them. Control measures have changed in almost all countries as sucking insects and bud/bollworms emerged as the most damaging pests. Jassids and whiteflies were at the top of the list of the most damaging insects in many countries, but later, bollworms took the lead until the introduction of biotech cotton in the mid 1990s. Now once again the trend is, or seems to be, changing, particularly in the countries that are adopting biotechnology. Pest pressure can be tracked/monitored using many indicators, such as: number of hotspots per unit area, percentage area affected, number of insect catches per trap (though not reliable for field infestation and not applicable to all pests), area requiring sprays, among others. Entomologists in the US have been collecting data since 1979 on many variables of pest pressure, including estimated arthropod-related yield losses, percentage of affected area, direct control costs and insecticide sprays. This article centers on the estimated yield losses attributable to arthropods, individual pests and the cost of direct control measures in the last 20 years.

Sources of Data

Data on yield losses and direct management costs for all years were collected from state coordinators and taken from surveys by county agents in cotton producing states, extension specialists, private consultants and research entomologists. The same sources were used every year. According to the sources, who have compiled the data for years, the data have been averaged out over a total reporting unit. Accordingly, 8% damage on 25 hectares out of 100 hectares means 2% damage to 100 hectares.

Sources of Funding

The Cotton Foundation, a membership organization, provided funding for this work. The Foundation members include banks, seed companies, chemical and equipment manufacturers, publishers and others with the objective of fostering a healthy U.S. cotton industry.

Arthropods

Arthropods are technically defined as invertebrate animals of the phylum Arthropoda, including insects, arachnids, and crustaceans that are characterized by jointed limbs and cuticles.

Direct Management Costs

Direct management costs include: 'at planting insecticide costs,'(an estimate of the cost of systemic insecticides applied at planting to control thrips and other pests of seedling cotton); 'Bt cotton costs,'(an estimate of the technology fee); 'eradication costs '(which include the maintenance fee in the states that have eradicated the weevil and that have concluded other eradication projects); and 'scouting costs,' in addition to the traditional `foliar insecticide costs.'

Major Pests

In 1993/94, the five major pests that caused the greatest losses in yields in the US were: boll weevil (*Anthonomus grandis)*, bud/bollworms, and Lygus bug *(Lygus lineolaris* and *Lygus Hesperus* (in the west)*,* thrips *(Thrips* spp.) and beet armyworm (*Spodoptera exigua).* The boll weevil alone caused almost 2% of all yield losses (defined as actual yields compared with yields that would otherwise have been attained), while the budworms and bollworms together caused only 1.6% of losses. The boll weevil and bud/bollworms were responsible for 84% of all yield losses, which were estimated to amount to 6.9% in 1993/94. Over the last two decades, the intensity of arthropod damage has changed as described below.

Boll Weevil

In the last two decades, the boll weevil topped the list of most damaging insects only three times: in the crop years of 1993/94, 1997/98 and 2000/01. The yield damage caused by the boll weevil has been declining steadily. For the seven years since 2006/07, no boll weevil damage to yield has been reported in the United States. The boll weevil eradication program that was launched in the early 1970s has succeeded in eliminating the boll weevil from most of the production areas of the US. Monitoring of the pest through pheromone traps continues in most production areas at a minimum cost, and active eradication is still under way in one production region of Texas, the Rio Grande Valley.

Bud/bollworms

Bud/bollworms caused the greatest yield losses and headed the *most damaging* list 12 times. Yield loss was as high as 4.2% in 1995/96. The level of yield losses caused by bud/bollworms has varied, but in general, the trend is downward. In view of the fact that most of the area is planted to stacked-gene insect-resistant biotech cotton, bud/bollworms, pink bollworm and other caterpillars are not expected to gain strength.

Lygus

Yield losses caused by Lygus have been varying from year to year, as is the case with most other arthropods. In the last 20 years, Lygus was the most damaging pest on only two occasions, both during the past five years. There is general consensus in the cotton community in the US that Lygus bugs are on the increase on cotton in the country, but the figures in the table next page would not seem to support that contention. It seems that Lygus management has definitely become a greater priority as a result of a series of factors, particularly widespread use of insect growth regulators for whitefly control, biotech cotton for lepidopteran control, and increased numbers of crops serving as preferred host for the bug. In particular, the availability of pest-specific pesticides, i.e. insect growth regulators and insect-resistant biotech cotton, has resulted in a significant reduction in the number of applications of insecticides that are active against Lygus. Reduction in insecticide use opened a window that allowed Lygus to cause considerable damage. Many researchers have come to a similar conclusion. According to Taillon *et al* (2012), the tarnished plant bug (*Lygus lineolaris)* has become the most destructive pest in cotton since the eradication of the boll weevil and the commercialization of biotech cotton. Prior to 1995, plant bugs were controlled with insecticides targeting other insect pests such as the tobacco budworm/cotton bollworm and boll weevil. According to Taillon *et al* (2012), Arkansas growers treated 92% of the cotton area in 2010.

Thrips

Several species of thrips are known to infest cotton in the US. The species complex and their abundance vary widely across the cotton producing states. The thrips population surpasses threshold levels, but the attack is generally limited to the early stages of crop growth, so the impact on yield remains at a minimum. Given the minimal damage done by bud/bollworms, the absence of yield loss caused by boll weevil and generally lower losses in yields due to arthropods in 2009/10, thrips was able to make it to the top of the list of yield loss culprits.

Beet Armyworm

The beet armyworm has never headed the list of yield loss producers in the last 20 years. In recent years, there is almost no loss at all in yield due to the beet armyworm, with most of the credit attributed to the use of biotech cotton varieties.

Other Pests

The whitefly (*Bemisia tabaci)* population had been drastically reduced by 1993/94. Whitefly infestations are mostly limited to the western states. Many factors, including: chemical adjustments, biological control and various IPM measures helped reduce the whitefly population to negligible levels. The other arthropods that continue to cause losses in yield are aphids (*Aphis gossypii)*, spider mites (*Tetranychus* spp.) and various species of stink bugs. Losses due to aphids have been consistent over the years, except for 1999/00 when non-traditional arthropods caused the greatest yield losses and fleahoppers topped the list at 2.4%, followed by aphids at 2.2%. Losses in yields due to bollworms were less than 50% of the loss caused individually by fleahoppers or aphids. Generally, spider mites have a greater impact on fiber quality than on yield. Spider mites continue to be a pest of significance on cotton but they have never achieved alarming proportions. Sporadic infestations within fields and across states generally do not require insecticide sprays targeting mite control. Various species of stinkbugs are gaining importance and in recent years insecticide sprays have been needed on larger areas.

Many other insect pests affect cotton yields in the US. The pink bollworm (*Pectinophora gossypiella)* is one of the pests that gained importance in Arizona, New Mexico, western Texas and California. In 1995, it was believed that there was greater potential for good economic management of the pink bollworm in the southwestern desert than at any previous time. Area-wide management based on pheromone and sterile moth technologies was found to be effective when applied on a field- by-field basis. Cultural control measures that helped to kill larvae were also found successful in complementing long-term control measures. Cry genes, which were expected to make their way to the commercial stage, were viewed as a strong integrated tactic for the eradication of the pink bollworm (Staten, 1995), a major pest causing the greatest yield losses in China, India and Pakistan in the early 1990s. Staten's (1995) views were born out and the pink bollworm has had no negative impact on yield in the US for the last 8-10 years. Biotech cotton played a major role in controlling the pink bollworm. Around 20 insect pests have been monitored to assess their individual impact on cotton yields in the US. Six or seven other insects were also monitored intermittently. The situation varies from one state to another, as some insects

Table 1: Cotton Yield Losses in the USA

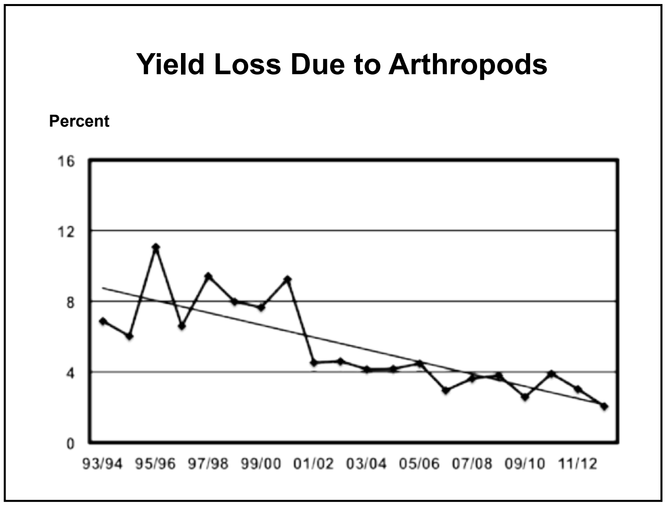
Crop Year	Top Ranking Pest and Damage		No. of Insecticide Applications to Biotech Cotton Against Caterpillars	Bud/bollworms	Lygus	Thrips	Boll Weevil	Aphids	Spider Mites	Stink Bugs	Additional Remarks
	Pest	% Loss in Yield									
93/94	Bool weevil	1.88	No applicable	1.58	0.88	0.88	1.88	0.30	0.34	0.02	Pink bollworm caused only 0.02% loss in yield.
94/95	Bud/bollworms	1.96	No applicable	1.96	1.16	0.16	1.75	0.30	0.29	0.09	Pink bollworm caused only 0.16% loss in yield.
95/96	Bud/bollworms	3.97	Not reported	4.16	1.02	0.34	1.66	1.09	0.61	0.17	
96/97	Bud/bollworms	2.55	Not reported	2.55	0.68	0.40	1.86	0.48	0.11	0.14	
97/98	Bool weevil	4.02	Not reported	2.14	0.91	0.47	4.02	0.75	0.14	0.23	Boll weevil was even absent in 9 states.
98/99	Bud/bollworms	2.71	Not reported	2.71	1.04	0.35	2.30	0.33	0.24	0.15	
99/00	Fleahopper	2.36	0.29	1.04	0.92	0.27	1.20	2.17	0.12	0.37	The only time fleahoppers caused more damage than any other arthropod.
00/01	Bool weevil	2.87	0.33	1.41	0.52	0.59	2.87	0.44	0.22	0.53	
01/02	Bud/bollworms	1.27	0.40	1.27	0.98	0.80	0.10	0.28	0.03	0.77	
02/03	Bud/bollworms	2.34	0.52	2.34	0.72	0.45	0.18	0.12	0.03	0.45	
03/04	Bud/bollworms	1.42	0.55	1.42	0.90	0.26	0.08	0.09	0.12	0.74	Pink bollworm had minor damage on yield.
04/05	Bud/bollworms	1.40	0.47	1.40	1.06	0.56	0.07	0.06	0.08	0.59	Fourth year of less that 5% loss in yield due to arthropods.
05/06	Bud/bollworms	1.54	0.54	1.54	0.90	0.43	0.05	0.13	0.35	0.64	
06/07	Bud/bollworms	0.88	0.59	0.88	0.67	0.38	0.01	0.18	0.20	0.42	Minor loses due to pink bollworm (0.007%).
07/08	Bud/bollworms	0.91	0.50	0.91	0.68	0.58	0.01	0.32	0.24	0.27	Minor loses due to pink bollworm (0.001%).
08/09	Lygus	1.00	0.75	0.76	1.00	0.52	0.00	0.20	0.16	0.75	Almost no losses due to pink bollworm.
09/10	Thrips	0.71	0.75	0.49	0.61	0.71	0.00	0.03	0.10	0.37	No losses due to pink bollworm. Minimum overall loss in yield in 20 years.
10/11	Bud/bollworms	1.19	1.06	1.19	0.68	0.20	0.00	0.29	0.20	0.72	No losses due to pink bollworm.
11/12	Lygus	1.03	0.82	0.38	1.03	0.70	0.01	0.32	0.17	0.51	No losses due to pink bollworm.
12/13											

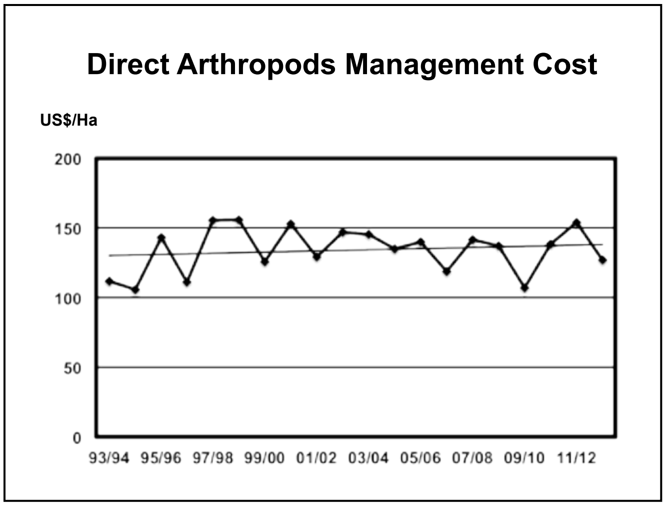
find more suitable conditions in which to flourish with one set of growing conditions than with others, but in general, none of the other pests, including those that were not discussed above, required any insecticide sprays. The boll weevil monitoring and the pink bollworm control strategies cost US$8.9/ha and US$0.82/ha, respectively, at the national level.

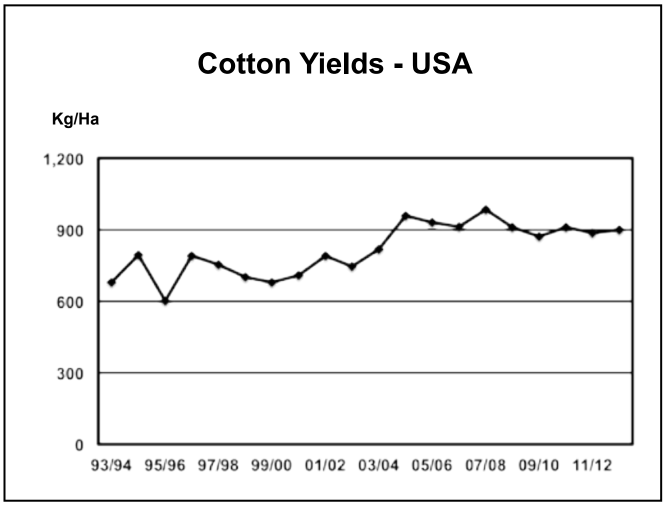
Yield Losses over the Years and Direct Management Costs

Data for the last 20 years show that in 1993/94, the average national yield loss was 6.9%. The average yield in the US was 679 kg/ha, which means that 47 kilograms of lint/ha were lost due to the insect pest complex. The Cotlook A Index for 1993/94 was US$1.56/kg of lint, putting the loss at US$73/ha. In 1994/95, losses declined to 6% due to a generalized reduction of pest pressure. In the next season, 1995/96, losses rebounded to 11.1%, the highest loss rate in the preceding 20 years, which brought the average national yield down to 602 kg/ha. The calculated yield loss in 1995/96 was 69 kg lint/ha or US$130/ha, this, in spite of the US$143/ha spent on direct management costs. Yield loss due to arthropods has been on the decline since 1995/96. Yield losses have been down to less than 5% since 2001/02 and only 2.06% in 2012/13. The season average for the Cotlook A Index for 2012/13 is not available yet, but based on a 3% loss in yield in 2011/12, losses in terms of USD come to only US$56/ha, this in spite of an almost 30% above-average long-term A Index price.

The chart below clearly shows a declining trend in losses due to arthropod damage. Annual insect-related loss figures are also reported in USD spent on direct management of arthropods. The cost of arthropod management showed a slightly increasing trend. On the average over the last 20 years, about US$135/ha were spent to control arthropods. Cotton yields increased and showed a positive trend but did not prove to have a strong negative correlation with the decrease in arthropod-related losses. The weak correlation coefficient of -0.80 indicates that increases in yields have also come from developments/improvements in other agronomic practices. The role of biotech cotton in this connection is discussed separately.







Arthropod-related Yield Reductions by State

Cotton is produced in 17 states in the US. The 20-year average reduction in yield due to arthropods for the 17 states is estimated at 5.6%. However, of all the cotton producing states, cotton yields in Alabama were the ones most affected by arthropods. The 20-year average loss in yield was 10.2%, followed by 7.7% in Oklahoma, and a 7.5% reduction in Tennessee. Bud and bollworms caused most of the yield damage in Alabama. Almost half of the yield loss was due to bud and bollworms in 1995/96. The damage was severe throughout the state, but central Alabama was most severely affected. The 20-year average data for yield reduction also showed that Kansas and Virginia were least affected by arthropods. The declining trend in losses due to arthropods is evident in all states.

Impact of Biotech Cotton on Losses caused by Arthropods

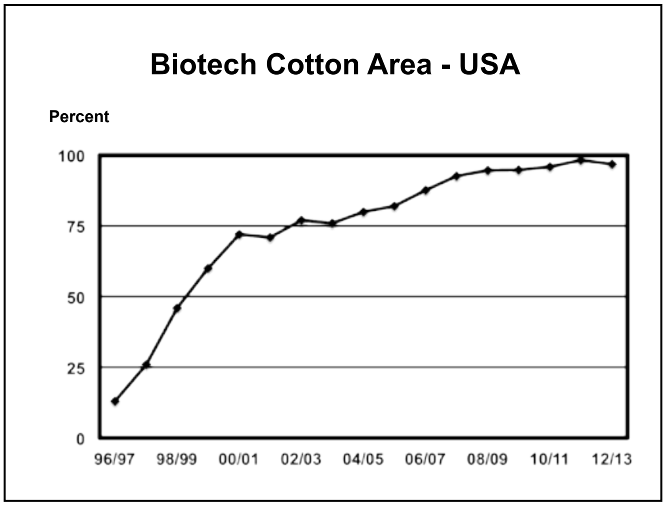
Biotech cotton is an evident success story in the US. Seven years after its commercial launch in 1996/97, biotech cotton was being planted on more than three quarters of the country's cotton area. At present, almost all cotton planted in the US is biotech varieties, except for organic cotton and a little conventional cotton.

Biotech cotton played a major role in changing the extent of losses due to insects and the pest complex. Comments here, as above, are limited exclusively to insect-resistant biotech cotton. Some of the more important facts about its impacts are discussed here.

- Biotech cotton drastically lowered losses due to arthropods without any significant impact in terms of increased control costs. A slight increase in direct control costs could be easily annulled by the obvious upward trend in yields as shown below. However, it remains to be seen whether the recent stagnation in yields will mobilize additional expenditure on direct control costs.
- Season-to-season variations in losses due to bud/bollworms persisted even prior to the commercialization of biotech cotton. Biotech cotton that assures blanket

Table 2: Percent Yield Reduction Due to Arthropods by State

State	1993/94	1994/95	1995/96	1996/97	1997/98	1998/99	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10	2110/11	2011/12	2012/13	Average
Alabama	12.77	14.74	49.22	5.21	9.61	10.78	8.46	4.01	9.19	16.14	6.86	4.05	7.20	7.84	7.20	8.09	6.15	3.74	2.60		10.20
Arizona	4.41	10.58	8.67	10.00	6.36	8.74	4.09	3.18	7.44	4.84	5.75	5.12	8.69	2.32	1.54	2.96	3.31	3.32	3.66		5.53
Arkansas	5.70	7.36	5.72	9.29	6.60	8.57	3.91	8.79	2.33	8.48	5.77	4.85	4.00	3.97	3.45	4.95	3.67	5.71	6.42		5.77
California	4.74	2.83	11.29	3.06	6.79	5.42	2.15	5.10	1.82	1.08	2.55	2.02	6.61	3.83	3.19	2.87	1.93	2.87	3.97		3.90
Florida	5.73	4.28	8.43	3.66	8.05	3.49	2.16	1.85	5.31	4.69	5.25	5.29	3.60	1.36	8.69	8.20	7.95	7.98	5.50		5.34
Georgia	2.37	2.49	3.80	2.79	4.05	4.25	3.09	3.46	3.17	4.22	4.17	2.26	4.21	3.56	3.43	2.25	3.03	2.32	1.87		3.20
Kansas								1.28	0.00	1.15	5.52	3.07	0.41	0.30	1.46	4.90	3.20	4.90	4.90		2.59
Louisiana	9.93	9.01	9.11	8.06	7.42	6.38	3.23	2.86	4.80	5.09	4.53	3.46	3.00	4.81	4.25	6.55	4.22	6.61	7.12		5.81
Mississippi	16.51	12.09	13.24	7.08	6.74	8.35	3.93	4.53	7.85	6.84	5.86	3.79	4.47	2.71	2.63	2.34	5.94	8.94	7.06		6.89
Missouri	5.33	2.79	4.99	3.41	6.91	10.69	4.01	7.74	3.93	7.81	10.27	5.85	0.86	0.91	5.13	0.46	6.11	2.27	1.14		4.77
New Mexico	6.81	4.52	10.09	12.87	10.67	8.45	6.08	8.42	0.97	2.51	1.18	3.43	3.13	2.00	0.34	1.06	0.19	1.40	0.02		4.43
North Carolina	3.13	10.69	4.65	6.23	9.35	4.44	6.40	7.97	8.12	1.55	4.03	10.64	6.79	5.09	2.83	3.95	3.33	2.95	2.95		5.53
Oklahoma	8.35	5.62	13.17	11.78	12.91	8.71	9.35	7.78	7.07	9.40	11.72	10.84	13.29	1.57	1.35	3.18	7.76	3.05	0.22		7.74
South Carolina	14.26	4.68	6.67	9.57	6.98	4.88	4.41	7.00	6.98	8.52	8.09	6.13	8.00	5.94	5.94	5.02	4.45	5.02	2.78		6.60
Tennessee	17.40	6.63	15.05	4.91	7.16	14.70	9.99	9.27	3.28	6.70	5.63	3.26	3.16	6.29	4.06	6.24	4.39	5.71	9.40		7.54
Texas	4.50	2.90	15.58	7.77	15.53	11.50	13.65	16.20	3.63	2.09	1.90	4.02	3.55	1.11	3.62	0.49	1.14	4.30	1.95		6.08
Virginia	0.00	0.50	3.01	2.20	0.89	0.01	0.59	2.05	1.07	3.33	5.33	1.19	5.58	2.58	2.00	2.00	2.00	1.00	2.00		1.96
USA	6.88	6.03	11.08	6.61	9.42	7.98	7.66	9.26	4.53	4.61	4.16	4.18	4.47	2.96	3.62	3.8	2.58	3.91	3.03		5.62



protection against caterpillars without any variation in the pest population did not provide any stability/consistency in loses due to caterpillars.

- Biotech cotton, augmented by other area-wide management control strategies, including the sterile moth release technology, played a significant role in controlling the pink bollworm.
- Biotech cotton had no direct impact on boll weevil elimination in the US, but does deserve partial credit for reducing the pink bollworm to economically feasible levels.
- Some secondary pests, including Lygus, stinkbugs and some regional pests, may gain importance in the next few years.

References

Anonymous. Lygus in Cotton: Identification, Biology & Management. The University of Arizona, Tucson, AZ, available at http://ag.arizona.edu/crop/cotton/insects/lygus/lygus1.pdf

Staten, R. 1995. Pink Bollworm Management: Prospects for the Future. *Proceedings Beltwide Cotton Conferences,* 1995, pp 153-156. National Cotton Council of America, Memphis, TN, USA.

Tailon, N.M., G.M. Lorenz III, W.A. Plummer, B.C. Thrash, J.W. Fortner, C.K. Colwell and G. Wilson. Efficacy of Insecticides for Control of Plant Bugs in Arkansas. Summaries of Arkansas Cotton Research 2011, Edited by Derrick M. Oosterhuis, University of Arkansas, research Series 602, November 2012.

Williams, Michael. Cotton Insect Loss Estimates annual reports from 1994-2013. *Proceedings Beltwide Cotton Conferences* from 1994-2013. National Cotton Council of America, Memphis, TN, USA.

Insecticide Use and Toxin Resistance in Biotech Cotton

The primary objective of commercializing insect-resistant biotech cotton was to reduce insecticide use. Australia, Burkina Faso, China, Colombia, India and Pakistan had more serious insecticide resistance problems than the US or Mexico, countries where the high cost of production was a greater concern. Argentina, Brazil and South Africa adopted biotech cotton more as a new technology for improving productivity. Tackling the resistance problem and containing ever-rising insecticide costs were the primary objectives of adopting insect-resistant biotech cotton, other objectives, such as raising yields, were secondary. However, the insecticide use issue was the one that proved to be pivotal and truly common to all countries that have commercialized biotech cotton. Minimal use of insecticides is desirable, but increases in the population of other non-target pests, which in turn may damage not only cotton but also other crops grown as part of the cotton production system, are not acceptable. This article is limited to insect-resistant biotech cotton and focuses on recent information on insecticide use and insect resistance to the toxin in biotech cotton.

Pesticide Use and Losses

Pesticide use is controversial, but the fact is that pesticides are generally profitable in agriculture, and that is why they are still used. The cost/benefit ratio was undoubtedly higher when pesticides were first adopted in the 1940s, 1950s and 1960s and has slowly declined since. The experience of working with pesticides has also shown that their use does not always decrease crop losses. For example, despite a more than 10-fold increase in insecticide use (organochlorines, organophosphates and carbamates) in the US between 1945 and 2000, total crop losses from insect damage nearly doubled, from 7% to 13% (Pimentel, 2005). Increases in cotton losses due to the insects developing resistance to insecticides were reported in every country where the resistance problem arose. Most of the benefits of insecticide use are measured in terms of direct losses and net crop returns. An extensive article on direct losses of cotton due to arthropods in the US precedes this article. Benefit assessments often fail to include the indirect environmental and economic costs associated with the recommended application of insecticides, in terms of quantity and quality. No matter how precisely insecticides are applied, there will always be consequences, which may become apparent in just a few years or take decades to show up.

It is estimated that the cost of pesticides used in US agriculture today comes to about US$5 billion. These figures are not recent, and the extensive use of herbicides may have more than cancelled the benefits of the reduction of insecticide use on insect-resistant biotech corn, cotton and soybeans. Of course, the direct income generated by the investment of those five billion dollars is much greater than this figure. However, the damage incurred as a result of using insecticides is not assessed in terms of the impact on all related fields. Pimentel (2005) reported (table below) that total losses in terms of direct, social and environmental costs are much higher than the amounts measured in terms of direct pesticide cost and associated overheads. The table below shows that pest-related losses may go down and direct income may surpass the huge direct expenditure on pesticides, but when one takes into account the losses/impacts resulting from pesticide use, the cumulative consequences are actually double the value of the losses as compared to the direct cost of pesticides.

Total Estimated Environmental and Social Costs from Pesticides in the USA

Costs	US$ (Millions)
Public health impacts	1,140
Domestic animals deaths and contaminations	30
Loss of natural enemies	520
Cost of pesticide resistance	1,500
Honeybee and pollination losses	334
Crop losses	1,391
Fishery losses	100
Bird losses	2,160
Groundwater contamination	2,000
Government regulations to prevent damage	470
Total:	9,645

In the US, the data relating to insecticide sprays on biotech cotton showed that over recent years there has been a slight increase in the number of insecticide sprays against target pests. However, the increase is nominal; it may be due to a number of reasons and may not be related to Cry proteins at all. It is true that, unlike the costs stemming from insecticide control, which in the past kept increasing on par with losses due to insect damage, the use of insect-resistant biotech cotton is not increasing the damage due to insects. The Bt toxin technology is working and will go on working for as long as the resistance problem can be delayed, and ultimately, prevented. Similarly, the negative effects, including the social and environmental impacts, are minimal. On the other hand, the positive impacts, as reflected in the table above, while they *are* reported, they are usually not quantified and often underestimated.

In China, the American bollworm, *H. armigera* (known as the cotton bollworm) had developed serious resistance to insecticides by the late 1980s. By 1990/91, it had become almost impossible to control the bollworm with insecticides. The well-known bollworm outbreak on cotton in 1992/93, which reduced cotton yields by 24% from those in 1991/92, convinced the government of the need to recommend systemic insecticide use on wheat to control first-generation bollworms. Higher survival rates in the first generation increased the chances that the following generations would show up on cotton in greater numbers. Just as with the need to spray wheat due to the widespread bollworm attack on cotton and the development of resistance by the bollworm, Chinese researchers have reported that the benefits of biotech insect control may even spill over to other crops that have not been genetically transformed but are grown on the same farms within the cotton production system. Wu *et al*. (2012) found that broad-spectrum insecticides kill harmful arthropods, but they also destroy many natural enemies that provide a form of biological pest control. A decrease in the use of insecticide sprays linked to insect-resistant biotech cotton could enhance bio-control services. Wu *et al*. (2012) showed that, on the basis of data from 1990 to 2010 in six provinces of northern China, there was a marked increase in the abundance of three types of arthropod predators (ladybirds, lacewings and spiders) and a decreased population of aphids attributable to the widespread adoption of biotech cotton and a reduced number of insecticide sprays on cotton.

They also found evidence that those predators that do spill over from biotech cotton fields onto neighboring crops like maize, peanut and soybean growing along with cotton, might provide additional bio-control.

This paper does not go into the matter of mirids, which may have increased on cotton, at least in the US. In the past, when insecticides were sprayed to control bud- and bollworms, they also controlled mirids at extremely low population levels. Non-use or minimal use of insecticides when mirids are present in the fields provides a space for mirid populations to multiply on cotton.

Since 2007/08, India has planted more surface area to biotech cotton than any other country in the world, exceeding ten million hectares in each of the last three seasons. More area under insect-resistant biotech cotton means reduced use of insecticides. In India, most of the insecticides were used against various kinds of bollworms. Since the commercialization of

Quantity of Insecticides Used on Cotton in India

Year	Against Bollworms (In Metric Tons)	Against Sucking Insects (In Metric Tons)
1995	5,748	2,965
1996	5,920	3,743
1997	6,973	3,621
1998	7,930	3,857
1999	7,522	4,487
2000	6,647	3,716
2001	9,410	3,312
2002	4,470	2,110
2003	6,599	2,909
2004	6,454	2,735
2005	2,923	2,688
2006	1,874	2,374
2007	1,201	3,805
2008	652	3,877
2009	500	5,816
2010	249	7,270
2011	222	6,372

Source: Compiled by Kranthi & Reddy (2012), CICR, In *Bt Cotton: Q & A* biotech cotton, the quantity of insecticides used has drastically declined to less than 4% of the total applied in 2003/04. The data also show that the quantity of insecticides used against sucking insects has almost doubled in the same period. The increase in insecticide use to control sucking insects may be attributed to higher pest pressure, but also to higher emphasis on sucking insects in order to increase yields.

The Philosophy Behind the Refuge Crop

The development of resistance to insecticides by pests that attack conventional cotton taught growers very useful lessons that were first applied when the time came to revisit insecticide use recommendations; those same lessons later alerted biotechnology developers to take a cautious approach to biotech applications. Lethal doses (LD50 and LD 90) had gone up in many countries, thus leading to excessive use of insecticides and reduced returns on investments. Consequently, in the more severely affected countries, resistance management programs were launched, to which the pesticide industry made a considerable contribution. The pesticide industry, which had promoted extensive use of insecticides, fine-tuned their message to include spray regimes and to avoid not only lower but also higher doses of insecticides. Countries and regions that followed those recommendations were able to delay and/or prevent the development of resistance, in particular by the American bollworm, which had emerged as the most notorious pest in terms of resistance development. Producers recognized the danger that if bollworms can develop resistance to hard chemicals such as insecticides, they might also be capable of developing resistance to the Bt toxins, and this consideration became an intrinsic part of the recommendations that accompanied the commercialization of biotech cotton. No one ever doubted that bollworms might eventually develop resistance to Cry1Ac; the debate centered on how long the resistance to the Bt toxin might be delayed. Researchers spent copious amounts of time and resources investigating the mechanisms by which pests developed resistance, and their efforts helped to identify measures that might contribute to delaying or preventing it.

Understanding the genetics of insecticide resistance in a field population was considered to be a prerequisite for resistance management programs based on hard science. Identifying the resistance-conferring genes, estimating their frequency in field populations, understanding the mode of inheritance of resistance-causing genes, and the factors influencing the frequency of resistance alleles over time were of prime importance (Kranthi, 2005). Such studies led to the idea of recommending the use of refuge populations when introducing biotech cotton. The aim of the refuge crop was to generate significant numbers of susceptible bollworm moths that had not been exposed to any of the Cry proteins. The recommendation was that a refuge crop should be grown within a specified proximity to the insect-resistant biotech cotton. The underlying principle was that moths produced in the refuge crop would automatically disperse to form part of the local mating population, where they would mate with resistant as well as susceptible populations produced on biotech and conventional (refuge) cotton. Hybridization of the resistant and susceptible populations would naturally disperse resistance alleles and keep them thinning in the population until all the population became susceptible. Continuity in the thinning of the resistance alleles was required, thus making it necessary to grow refuge crops year after year.

Refuge requirements were recommended and vigorously implemented in most countries, but not in all. Five percent unsprayed and 20% sprayed of the conventional area was a common recommendation from technology developers. Despite the difficulty of implementing the refuge recommendations in small-scale farming systems, the strategy worked very well for years. In a country like India, where the farming system may allow for the cultivation of a single hectare of cotton, or even less, it was mandatory for farmers to buy a proportional quantity of conventional planting seed along with biotech seed. Refuge recommendations were not followed strictly in some farms and regions, but since biotech cotton was adopted slowly, hybridization of the bollworm populations continued naturally for years. Fortunately, resistance to both the Bt proteins was found to be recessive, so if a resistant moth having ‘rr’ genes mated with a susceptible moth from the refuge having an ‘ss’ composition, the offspring they produced from the hybrid population, ‘rs’, would also be susceptible to the Cry protein and would be killed by the Bt toxin (Ceeney *et al.*, 2013).

The success of the refuge strategy encouraged countries to relax refuge requirements. Production of a conventional alternate host crop at the same time that biotech cotton was grown was cited as one of the strongest arguments in favor of a reduced emphasis on the refuge crop. Conventional cotton, or unsprayed cotton, was originally the only option for farmers to produce natural feeding populations. Later, however, even biotech crops other than cotton that hosted the target bollworms were included in the refuge requirements. In some cases, even eliminating conventional or unsprayed cotton was recommended based on the principle that other non-biotech crops were producing sufficient quantities of susceptible moths. In the US, soybeans, peanuts, weedy hosts, corn and sorghum are some of the common crops sharing the same target pests. In Australia, sorghum and corn used to be refuge options, but that practice was discontinued when the frequency of resistance to the Cry2Ab (found in Bollgard II) was seen to be increasing in populations of *H. punctigera*. Fortunately, the species *H. punctigera* has not damaged cotton in any other country that has commercialized biotech cotton (except Australia). The requirement of continuous monitoring of the resistance level and the frequency of resistance alleles, which was, and continues to be a key factor in insecticide resistance monitoring programs, is necessary and must be followed.

In conventional cottons, new insecticide products were being introduced on a regular basis so that insecticide doses could be manipulated or different insecticides could be mixed in various proportions, but no such options are available to growers of biotech cotton. The only additional feature available, albeit a very powerful one, is gene stacking, which was not an option in the case of insecticides. Growers were, of course, interested in the convenience and the time saved with natural refuges, along with the increased income associated with not having to plant conventional or unsprayed cotton. Then, the introduction of stacked gene biotech cotton provided growers with a double defense against the American bollworm: it strengthened their confidence in the technology, on the one hand, while encouraging relaxation of refuge requirements on the other.

Toxin Expression

Many studies carried out in a broad range of countries have proven that toxin expression in Bollgard I and Bollgard II cottons is influenced by genotypes, plant health, plant organ, age of plant tissues and growing conditions. Growing conditions include nitrogen levels, which affect plant stress levels. Water supply (rainfed or irrigated) and unfavorable temperatures may also affect stress levels. An extensive article on toxin expression in biotech cotton was published in the June 2008 issue of the *ICAC RECORDER*. Toxin expression is highest in leaves, followed by squares, bolls and flowers. Cotton leaves were found to have fewer toxins per gram weight than leaves from other biotech crops. In cotton, the expression level is highest at around 85-100 days after planting, after which the amount of toxin starts to decline. Therefore, biotech cotton controls bollworms most effectively around the time when the toxin level is highest, particularly in flower buds and bolls. The decline in toxin quantity does not mean that it loses its ability to provide adequate protection against the target pests. It does mean, however, that it is imperative to recognize a clear distinction between lower levels of toxin and higher levels of resistance developed by the target insects.

Resistance to Vip3A

In the last 17 years, since the commercial introduction of insect-resistant biotech cotton in the form of Bollgard I, Bollgard II, WideStrike™, Event 1 (India) and Guokang (in China), only crystal (Cry) proteins have been commercialized. Several studies have shown, in laboratory populations as well as in field populations of various Lepidopteran species that they are capable of developing resistance to Cry toxins (Gulzar *et al*., 2012 and Mahon *et al*., 2012). It was also found that several target pests had developed gene polymorphism for resistance to Cry proteins even before populations were exposed to biotech crops. Field level resistance to at least four different target pests has been reported. The resistance problem would have been worse if refuge requirements had not been advocated from the beginning and if the Cry2Ab gene had not been stacked in cotton. But in the long run, stacking of one gene is not enough to eliminate the resistance problem. The group of Lepidoptera, which is the target in insect-resistant biotech cotton, is susceptible to Cry1 toxins (e.g., Cry1Ac, Cry1Ab, Cry1F) and Cry2 toxins (e.g., Cry2Ab, Cry2Aa, Cry2Ae), several of which have been exploited in existing transgenic crops, including cotton. Within the Cry1 class, insects that are resistant to one toxin are often, but not invariably, cross-resistant to others. Less is known about cross resistance within the Cry2 class, but Cry2Ab resistant *H. armigera* has been found to be resistant to Cry2Aa and Cry2Ab. Cry2Aa and Cry2Ab-resistant *H. armigera* and *H. punctigera* are also resistant to Cry2Ae. Therefore, it is likely for most systems, that when resistance to a toxin in the Cry1 or Cry2 class emerges, plant breeders are left with limited alternative Cry toxins to exploit (Mahon *et al*., 2012).

For the time being, there are no concerns about toxin strength decreasing below threshold levels as Cry1Ac and Cry2Ab back up each other. It has been established that Cry2Ab is affected less by genotype makeup, plant organ and the age of plant tissues than other genes. So, it stands to reason that because the target insects are less exposed to lower doses of Cry2Ab, it would be more probable for them to take longer to develop resistance to Cry2Ab.

There is an ongoing need to strengthen defenses against target pests that not only vary their modes of attack on cotton, but also have various levels and mechanisms for developing resistance to toxins. One such strategy is the use of novel vegetative insecticidal proteins (Vips), which have been abundantly researched for over a decade. Vips are also derived from *Bacillus thuringiensis,* but they are genetically distinct from Cry proteins. The rationale behind this strategy is the expression of two insecticidal proteins that differ in their modes of action, more particularly, in their recognition of specific binding sites in the target insect. Studies have also shown that Vip3A proteins were as toxic as (or even more toxic than) Cry1 proteins. According to Sena *et al* (2009), Vip3A and Cry1A proteins have independent binding sites in *H. virescens, H. zea,* and *S. frugiperda* and probably also in other lepidopteran species. Considering that binding to specific receptors is a key step in the mode of action of *Bt* insecticidal proteins, the odds of finding high levels of cross-resistance between Cry1A or Cry1F proteins and Vip3A proteins in the species described above are extremely low. Binding assays identified specific independent binding sites for Cry1 and Vip3A proteins. Cry1Ab and Cry1Fa competed for the same binding sites, whereas Vip3Aa competed for the binding sites of Vip3Af. Gulzar *et al.* (2012) have even reported that Vip3A resistance was also associated with reduced egg viability and mating success and a lower intrinsic rate of population increase at temperatures below 20°C and above 30°C, i.e., the optimal laboratory culture temperature. The fitness costs associated with resistance are usually greater under stressful conditions. Such data can help predict the impact of fitness costs on the rate of development of resistance in the field and in the design of resistance management strategies that more fully exploit fitness costs. (Gulzar *et al*., 2012).

Ten years have passed since 2003/04 when Bollgard I was approved in India and Bollgard II was approved for commercial use in Australia and the United States. Since then, researchers have been looking for new genes with modes of action different from those of the Cry genes. Monsanto's third generation biotech cotton (which may be traded as 'Genuity Bollgard III' or simply as Bollgard III) will incorporate Vip3A in addition to Cry1Ac and Cry2Ab. Dow AgroSciences is striving to improve its Widestrike® by stacking Vip3A with existing Cry 1Ac and Cry 1F toxins. According to Mahon *et al.* (2012), Bayer Cropscience plans to license a biotech cotton containing Cry1b and Vip3A that will be traded as Vipcot. It is expected that Bollgard III will be released before Vip3A is also commercialized in other combinations. As far as the resistance problem is concerned, maintaining a high level of resistance to Cry2Ab has an enormous value for the success of Vip toxins in Bollgard III or in any other combination.

Back to Conventional Cotton

A number of papers dealing with conventional cotton were presented at the US 2013 Beltwide Cotton Conferences in January 2013, and some of them bore titles like 'Producing Conventional Cotton Again,' 'Conventional vs. Transgenic: The Showdown,' 'Economics of Conventional Cotton,' 'Conventional Cotton: A Growers Perspective,' as well as some others, including a workshop on conventional cotton. Similar papers, albeit not as provocative, were also presented at the 2012 Beltwide Cotton Conferences. Low lepidopteran pest pressure, attributed to insect-resistant biotech cotton, makes it feasible to produce conventional cotton at costs that are lower than what they were at the time when biotech cotton was introduced, or even comparable to those of biotech cotton (lower pest pressure, but high technology fee). However, that situation is exceptional and cannot be considered a valid reason to revert to conventional cotton, which would bring back the same problems faced in the past. The results of those discussions would seem to indicate that conventional non-biotech cotton can be grown and economically protected against insect damage if affordable and effective insecticides are available. This postulate, however, cannot justify the option of making any such shifts because management of conventional non-biotech cotton may require additional investments in scouting and a greater number of insecticide sprays, and would be feasible only if most of the area surrounding conventional fields is planted to biotech cotton.

Summary

New technologies will continue to be developed in the fight against the insects that consistently cause huge losses to farmers. Those developments would improve the economic viability of cotton and would have the least possible impact on the environment. Dow Agrosciences is conducting trials with Wide-strike (Cry1Ac + Cry1F); Bayer has initiated trials with twin-link (Cry1Ab + Cry2Ae), and a company in India is working with Cry1Ac + Cry1EC. Barring the development of a single gene as effective as two or three stacked genes, any new technology that is eventually commercialized will have a stacked gene effect. The fact remains that when Cry1Ac was introduced, the target insects had not been exposed directly to Cry proteins. At the time of the commercialization of Bollgard II, a number of lepidopterans had already grown accustomed to Cry1Ac. The introduction of Bollgard III will have to deal with populations that have generations of experience in surviving the Cry proteins. The probability of developing cross-resistance among Cry proteins is greater than with a vegetative insecticidal protein combination. So the purpose of sharing the information is to welcome new technological developments for expanded use of toxin proteins and to preemptively devise resistance management practices that would shield the newer technologies for a longer period of time.

References

Ceeney, Sally, Geoff Baker, Mary Whitehouse, Peter Gregg, Colin Tann, Tracey Level, Sharon Downes and Lewis Wilson. 2013. Refuge Crops – Investing in the Cotton's Future. *The Australian Cottongrower*, Vol., 33, No. 7, 2013.

Downes, Sharon, Rod Mahon, Tom Walsh, Bill James and Sally Ceney. 2013. Vip3A resistance genes: A threat to Bollgard III. *The Australian Cottongrower*, Vol., 33, No. 7, 2013.

Gulzar, Asim, Brian Pickett, Ali H. Sayyed and Denis J. Wright. 2012. Effect of Temperature on the Fitness of a Vip3A Resistant Population of *Heliothis virescens* (Lepidoptera: Noctuidae), *Journal of Economic Entomology* 105(3):964-970.

Kranthi, K.R. 2005. *Insecticide Resistance – Monitoring, Mechanisms and Management Manual*. An output of the Common Fund for Commodities project 'Sustainable control of the cotton bollworm, *Helicoverpa armigera* in small scale production systems (CFC/ICAC 14),' Central Institute for Cotton Research, PB No. 2, Shankarnagar, PO Nagpur 440010, India.

Kranthi, K. R. 2012. *Bt Cotton Q & A*. Indian Society for Cotton Improvement (ISCI), c/o Central Institute for Research on Cotton Technology, Adenwala Road, Matunga, Mumbai 400 019, India.

Mahon, Rod J., Sharon J. Downes and Bill James. 2012. Vip3A Resistance Alleles Exist at High Levels in Australian Targets before Release of Cotton Expressing this Toxin. PLoS One. 2012; 7(6): e39192, Published online at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3382218/

Pimentel, David. 2005. Environmental and economic costs of the application of pesticides primarily in the United States. *Environment, Development and Sustainability* (2005) 7: 229–252, also available at http://www.beyondpesticides.org/documents/pimentel.pesticides.2005update.pdf.

Sena, Janete A. D., Carmen Sara Hernández-Rodríguez and Juan Ferré. 2009. Interaction of *Bacillus thuringiensis* Cry1 and Vip3A Proteins with *Spodoptera frugiperda* Midgut Binding Sites. *Appl Environ Microbiol*. 2009, April: 75(7), 2236–2237.

Wu, Yanhui, Kongming Wu, Yuying Jiang, Yuyuan Guo and Nicolas Desneux. 2-12. Widespread adoption of Bt cotton and insecticide decrease promotes bio control services. *Nature* 487, 362–365 .

The Post-Genomic Era for Cotton

Andrew H. Paterson, ICAC Researcher of the Year 2012
Regents Professor and Head, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, USA

Background and Rationale

The scientific infrastructure in support of cotton research and improvement took a 'giant leap forward' with the release of the first 'gold-standard' cotton reference genome sequence on 5 January 2012. Later in 2012, two independent publications (Paterson *et al.* 2012; Wang *et al.* 2012) provided initial descriptions of the basic genome of cotton, with one of these also revealing new insights into the genes and processes that have permitted the tetraploid species *Gossypium hirsutum* ('Upland' cotton) and *G. barbadense* (Egyptian, Sea Island, and Pima cotton) to largely supplant the diploids *G. herbaceum* and *G. arboreum* as the providers of the world's leading natural fiber (Paterson *et al.* 2012).

This paper explores some general features of the cotton genome and fundamental messages learned from the sequences, along with new capabilities that the sequence provides to research and development. In particular, the genome sequence provides a means of coalescing many diverse data types, some of which still need to be created for cotton, to gain new understanding from otherwise disparate data. The cotton science infrastructure will quickly grow beyond this 'reference' genome sequence to include much of the diversity among species and genome types in the *Gossypium* genus – and that enriched information has enormous implications for improving the yield and quality of cotton and the sustainability and profitability of its production.

Finally, and perhaps most importantly, the genome sequence is not an ending but a beginning – specifically, a beginning of a new era of research and development using powerful new tools and approaches to identify and manipulate cotton genes of economic importance. While the potential benefits of this era are tangible and large, realizing this potential will require a host of additional enabling tools, technologies, and resources to be developed and creatively deployed, necessitating a new higher level of investment – but offering a new higher level of return on investment.

Which Cotton Genome Should be Sequenced First?

As is widely known, *G. hirsutum* and *G. barbadense*, and the other (wild) tetraploid cotton species, originated from interspecific hybridization between an A-genome African diploid species resembling *G. herbaceum* and a D-genome American diploid species (Skovsted 1934; Beasley 1940) resembling *G. raimondii* or *G. gossypioides* (Gerstel 1958; Phillips 1963). A- and D-genome groups are estimated to have diverged from a common ancestor 5-10 million years ago (MYA), then were reunited about 1-2 MYA (Wendel and Cronn 2003) via polyploidization in an A-genome cytoplasm (Wendel 1989; Small and Wendel 1999) following trans-oceanic dispersal to the New World of an A-genome propagule.

Capitalizing on more than a decade of prior research and preparation, in 2005, the worldwide cotton community prioritized the putative D genome progenitor, *G. raimondii* as the first *Gossypium* genotype to be fully sequenced. From first principles, it was preferred to first sequence a homozygous diploid expected to have only two nearly-identical copies of most genes, rather than a tetraploid which would have much more DNA including four copies of most genes comprised of two pairs that were just different enough from one another to be confusing. Although it does not itself produce spinnable fibers, ironically the *G. raimondii*-derived portion of the tetraploid cotton genome (the D_t 'subgenome') accounts for a somewhat larger share of genetic variation in fiber characteristics than the 'A_t' subgenome derived from an ancestor that does produce spinnable fibers (Jiang *et al.* 1998; Rong *et al.* 2007). *Gossypium raimondii* had the important advantages of having only half as much DNA, and in particular much less repetitive 'junk' DNA than the A genome progenitor. A rich history of genetic mapping and molecular analysis had shown *G. raimondii* to have virtually all genes present in the A genome or tetraploid cottons, and that the genes were largely in the same arrangement in the respective genomes. In partial summary, it was clear that information from *G. raimondii* would 'translate' well to cottons of economic importance, while its reduced size and complexity would reduce the cost and time associated with its sequencing and result in an improved outcome.

General Features of Cotton Revealed by the Genome Sequence

(Paterson *et al*. 2012)

Despite having the least-repetitive DNA of the eight *Gossypium* genome types, *G. raimondii* was nonetheless 61% derived from 'transposable elements', often thought of as 'junk DNA' (Paterson *et al.* 2012). One particular class that accounts for the largest share of many flowering plant genomes, long-terminal-repeat retrotransposons (LTRs), likewise account for about 53% of the *G. raimondii* genome.

To identify the genes of *G. raimondii*, computational approaches to recognize common features of genes such as 'start' and 'stop' sites were applied in conjunction with massively parallel sequencing of gene-encoded messenger RNA, to reveal 37,505 genes and 77,267 protein-coding transcripts (some genes encoding multiple transcripts). Remarkably, genes comprise only 44.9 Mb (6%) of the

G. raimondii genome and are largely located in distal chromosomal regions.

One surprise from the genome sequence was that shortly after its divergence from an ancestor shared with cacao (*Theobroma cacao*) at least 60 million years ago, the cotton lineage experienced an abrupt 5–6-fold ploidy increase. It was already well known that flowering plants had experienced polyploidy more frequently than other taxa – indeed, the common ancestor of most if not all eudicot (broad-leaf) plants experienced a genome triplication about 125 million years ago (Paterson *et al.* 2010). However, this was the first (and to date the only) discovery of such a large ploidy increase in such a short time.

The abrupt 5–6-fold ploidy increase together with the additional polyploidy that formed the common ancestor of *G. hirsutum* and *G. barbadense* and the wild tetraploid cottons, rendered cotton among the most complex of flowering plant genomes, only known to be matched by members of the *Brassica* genus. However, in modern cottons, this complex history of genome duplications is reflected in different ways for different genes and gene functional groups. For example, paleopolyploidy increased the complexity of a Malvaceae-specific clade of Myb family transcription factors, perhaps contributing to the differentiation of epidermal cells into fibers rather than the mucilages of other Malvaceae such as cacao. However, cottons pest- and disease-resistance genes experienced rapid turnover and evolved largely after the 5–6-fold ploidy increase.

Another surprise has been the extent to which the two 'subgenomes' of tetraploid cotton have exchanged information with one another since being joined in a common nucleus by polyploidy. Indeed, the vast majority of mutations that differentiate tetraploid cotton from its diploid progenitors involved non-reciprocal DNA exchanges between the A_t and D_t subgenomes, with random mutations contributing little. Curiously, these exchanges have been asymmetric, with more than twice as many D_t-genome alleles 'copied' on the A_t genome than the reciprocal. A tantalizing hypothesis is that the nascent polyploid may have gained fitness from D-genome alleles native to its New World habitat – however this offers no intuitive explanation for the evolution of the superior fibers of polyploids relative to A-genome diploids. Further investigation is in progress.

The Genome Sequence as a Means of Coalescing Diverse Data Types into New Understanding

Specific DNA sequences of 16 or more nucleotides in length are generally specific to single locations in higher eukaryotic genomes, and a host of biological information has been attached to 'sequence-tagged sites' that are generally substantially longer than this. For example, hundreds of 'quantitative trait loci' responsible for variation in economically important traits have been associated with DNA markers that have been sequenced and genetically mapped (e.g. (Rong *et al.* 2007)). Recently, massively parallel sequencing of short nucleic acid molecules has become an effective means of quantitating expression levels of vast numbers of genes under diverse conditions.

The contiguity and specificity afforded by a reference genome sequence provides a powerful means to coalesce diverse data types. By aligning different sets of DNA markers to the reference genome sequence, it is routine to align and compare different QTL mapping studies to identify 'QTL hotspots", regions of the genome that contain QTLs affecting multiple fiber traits more frequently than can be accounted for by chance (Rong *et al.* 2007). Likewise, voluminous gene expression data permits one to map sequence tags to the genome to identify concentrations of genes exhibiting coordinated changes in expression of functionally diverse genes under parallel sets of conditions.

Intersections among diverse data types that are revealed by using the reference sequence may suggest relationships of functional importance. For example, among 48 genes for which expression is up-regulated in domesticated *G. hirsutum* fibers at 10 days post-anthesis, 20 (a 10-fold enrichment relative to random genes) are within QTL hotspot D_t09.2 affecting length, uniformity, and short fiber content. Thirteen (a 15-fold enrichment) are in homoeologous hotspot A_t09 affecting fiber elongation and fineness. Of 45 genes down-regulated in domesticated *G. barbadense* at 20 DPA, 16 (35.6%) map to D_t09.2, and 8 (17.7%) to A_t09. In 79% of cultivated *G. barbadense*, this A_t region (then called chr. 5) has been unconsciously introgressed by plant breeders with *G. hirsutum* DNA, suggesting an important contribution to productivity of *G. barbadense* cultivars(Wang *et al.* 1995). Without the genome sequence to discern that these diverse data types each reveal non-random patterns that are concentrated in the same small region of the genome, they would merely represent interesting independent observations. Having discerned their relationships, we are much closer to identifying the causal gene(s).

A particularly powerful application of the genome sequence is to align the genes and chromosomes of one organism to those of another – for example, alignment to the botanical model *Arabidopsis thaliana* holds particularly great potential for increasing knowledge of cotton gene functions, albeit by analogy (Rong *et al.* 2005). For example, research into the genetic control of cotton fiber development may benefit from rich progress in understanding the growth and development of hair-bearing epidermal cells (trichomes) in *Arabidopsis*. Indeed, *Gossypium* and *Arabidopsis* are thought to have shared common ancestry about 83-86 million years ago (Benton 1993), and cotton may be the best crop outside of the Brassicales in which to employ 'translational genomics' from *Arabidopsis*.

Capturing the Spectrum of Diversity in the *Gossypium* Genus

The genus *Gossypium* occurs naturally throughout tropical and subtropical regions of the world, with at least 45 diploid species ($2n = 26$) that fall into genomic groups A, B, C, D, E, F, G, or K. The A-genome clade, also including B, E, and F genome types distinguished from one another based on pairing behavior, chromosome sizes, and relative fertility in interspecific hybrids (Beasley 1942) occur naturally in Africa and Asia, while the D-genome clade occurs in America. A third diploid clade exists in Australia, including C, G, and K genome types.

The diversity present in the two cultivated species, based on a subset of the diversity present in only two of the eight genome types, provides only a small 'sliver' of the naturally-occurring 'solutions' (adaptations) that *Gossypium* species have devised to survive and flourish in the face of often harsh and always fluctuating conditions. Indeed, while the importance of exotic germplasm is widely understood in terms of providing 'obvious' traits such as resistance to new disease strains, rich knowledge of other crops has shown beyond doubt that many alleles from exotic germplasm have 'cryptic' benefits that only become obvious when placed in elite backgrounds. Thus, a high priority is to clothe the reference genome with knowledge of the spectrum of extant diversity in each gene and indeed, each nucleotide. Only with such data can the intrinsic genetic potential of the genus be truly understood the intrinsic genetic potential of the genus, and craft improvement strategies can be crafted that optimally integrate full utilization of this potential with the need for 'extrinsic' (transgenic) solutions.

Much of the additional information needed to characterize the spectrum of extant *Gossypium* diversity will come not from 'gold-standard' reference sequences, which are costly and time-consuming to assemble rigorously, but from new 'resequencing' technologies that have relatively high per-nucleotide error rates but which can be mitigated by sequencing each nucleotide many times to arrive at a consensus that is often correct (Shendure and Aiden 2012).

The first such 'draft sequence' was conducted in *G. raimondii* itself (Wang *et al.* 2012), and comparison to the gold-standard sequence (Paterson *et al.* 2012) is illustrative (Table 1). The draft sequence is highly fragmented – the gold-standard sequence comprising nearly 80% fewer 'scaffolds' (genomic regions that could be assembled into single tracts of sequence at appropriate quality control standards), that were an average of ~8x longer (18.8 versus 2.3 Mb). The longest such scaffold approached the length of an entire chromosome arm in the reference sequence (52.1 Mb), being only about 26% of this in the draft. Virtually all (98.3% of) scaffolds in the reference sequence contained a sufficient number of DNA-based genetic markers to be aligned and oriented to genetically-defined chromosomes from the rich history of prior research in cotton genetics, versus only about half (52.4%) for the draft sequence. Estimates of the number of cotton genes based on the reference and draft assemblies were similar, indicating an important strength of draft sequencing – to quickly and economically capture the subtle differences in 'spelling' (sequences) of genes in different cotton genotypes.

Table 1: Parameters of Different *G. raimondii* Genome Assemblies

	Draft	Reference
Scaffold number	4715	1084
N50 (Megabases)	2.3	18.8
Longest scaffold	12.8	52.1
Anchored and oriented % genome	52.40%	98.30%

An important early application of draft sequencing has been to reveal clues into the early steps in the evolution of spinnable fibers (Paterson *et al.* 2012). From unremarkable hairs found on all *Gossypium* seeds, 'spinnable' fibers, i.e. with ribbon-like structure which allows spinning into yarn, evolved in the A-genome following divergence from the B, E, and F genomes ~5-10 MYA (Senchina *et al.* 2003). To clarify the evolution of spinnable fibers, we sequenced the *G. herbaceum* A and *G. longicalyx* F genomes, which respectively differ from *G. raimondii* by 2,145,177 single nucleotide variations (SNVs) and 477,309 indels; and 3,732,370 SNVs and 630,292 indels (Paterson *et al.* 2012). Across entire genes, 36 *G. herbaceum - G. raimondii* and 11 *G. herbaceum - G. longicalyx* ortholog pairs show evidence of diversifying selection. A striking example is Gorai.009G035800, a germin-like protein that is differentially expressed between normal and naked-seed cotton mutants during fiber expansion (Kim and Triplett 2004) and between wild and elite *G. barbadense* at 10 days post-anthesis (Paterson *et al.* 2012). We also identified 'striking mutations' of *G. herbaceum* genes since their divergence from *G. longicalyx* and *G. raimondii* (hence correlated with fiber evolution) including 1,090 non-synonymous mutations in 959 genes comprising the most severe 1% of functional impacts inferred using a modified entropy function (Reva *et al.* 2011); 3,525 frameshifts (3,021 genes); 1077 (987) premature stops; 527 (513) splice site mutations; 102 (102) initiation alterations; and 95 (94) extended reading frames. These striking mutations are enriched ($p=2.6x10^{-18}$) within fiber-related 'quantitative trait locus' (QTL) hotspots in A_tD_t tetraploid cottons (Rong *et al.* 2007), suggesting that post-allopolyploidy elaboration of fiber development (Jiang *et al.* 1998) involved recursive changes in A_t and new changes in D_t genes.

In partial summary, an important next step beyond the 'gold-standard' reference sequence will be to catalog the spectrum of diversity among *Gossypium* species, toward cataloguing of the true genetic potential of the genus to provide intrinsic low-cost genetic solutions to challenges that affect the yield, quality of cotton and economic and environmental sustainability of its production. The draft sequence of *G. raimondii* (Wang *et al.* 2012) overestimates the degree to which many such subsequent draft sequences might be assembled, as it benefitted from additional measures that are frequently not economical (and for example were not done in *G. herbaceum* or *G. longicalyx*). However, as information about the basal *Gossypium* genome accumulates, additional sequences will need to reveal smaller and smaller changes, for example single nucleotides in specific genes, and the need for high assembly quality will decline. Indeed, as we begin to sequence elite germplasm and learn about patterns of association ('linkage disequilibrium') of alleles at different loci along the chromosome, we will quickly reach a point such that sequencing of only a small subset of loci is a sufficient proxy to impute the probable genotype across the entire genome. Such technology, already in place in leading crops such as maize (Buckler *et al.* 2009; McMullen *et al.* 2009; Tian *et al.* 2011), is expected to be important in the application of genomic tools to mainstream crop improvement.

How do we Identify the Genes of Economic Importance?

"The greatest challenge facing the cotton community is the conversion of sequence to knowledge ...." (Paterson 2007)

With the genome and a much improved understanding of cotton's evolutionary history in hand, and a catalog of the spectrum of cotton's natural diversity imminent, how will we convert these new resources into low-cost genetic solutions to challenges that affect the yield, quality of cotton and economic and environmental sustainability of its production?

It was quickly identified that much cotton sequence is repetitive "junk DNA" –this cannot be dismissed as unimportant, but is relatively low in unique information content. While much of the repetitive DNA is thought to be 'junk DNA' that continues to exist because of its ability to multiply rapidly (Doolittle and Sapienza 1980), some proximally-repeated elements serve essential functions (centromeres), or encode products needed in large quantities (rDNA). Moreover, there is growing evidence of roles of repetitive DNA in the regulation of gene expression (Myers *et al.* 2011), and even some highly-repetitive regions of a genome contain occasional genes (Nagaki *et al.* 2004). Therefore, while the repetitive fraction of the genome will be a relatively low priority for functional analysis, it cannot be summarily dismissed.

Some cotton sequence will quickly be converted to information based on similarity to known sequences (from *Arabidopsis* in particular). As noted above, the relatively close relationship of cotton and *Arabidopsis*, and potential importance of using functional genomic information and tools from *Arabidopsis* to aid in dissecting economically-important pathways in cotton make this system an excellent case study for exploring comparisons of gene order among divergent taxonomic families.

However, to understand and manipulate the features that make cotton unique will require a host of new enabling tools, technologies, and resources; in particular targeting genes and regulatory features that are substantially different from those of other organisms. Because the basic gene set for flowering plants has largely been revealed (Paterson *et al.* 2010) by the many genomes now sequenced, a natural priority in cotton functional genomics will be to characterize genes that are related to its unique features. There are few if any other examples of seedborne epidermal plant cells that reach 1-2" or more in length and are nearly pure cellulose. How will we recognize the genes that confer these features, and how will we determine how they work?

Rapid gene evolution may be due to a lack of structural or functional constraints, or to strong positive selection for functional divergence. Established statistical approaches allow one to distinguish clearly between these possibilities (Yang 1997; Nielsen and Yang 1998; Yang 1998; Yang *et al.* 2000a). For example, rapidly evolving genes in Drosophila, mammals, and several other species are vital to reproductive success, cell-cell recognition, and cellular response to pathogens (e.g., (Yang *et al.* 2000b; Swanson *et al.* 2001a; Swanson *et al.* 2001b)). Examples of such cotton genes have been noted above, by identifying genes experiencing extensive change in the Gossypium A-genome following divergence from the F and D genomes (Paterson *et al*. 2012).

However, recognition of genes that have evolved rapidly does not by itself reveal their functions. More generally, following two episodes of polyploidy, many cotton genes may now have different (or at least partly different) functions than *Arabidopsis* genes with similar sequences. There is every reason to anticipate that the functions of some genes have been subdivided [*subfunctionalized* – (Lynch and Force 2000)] between duplicated *Gossypium* copies, while other duplicated copies may have evolved completely new functions (neofunctionalization) that do not exist in *Arabidopsis* or other outgroups. Indeed, several genomes other than *Arabidopsis* are potentially more informative to cotton in terms of understanding gene evolution and function – because *Arabidopsis* itself has experienced two genome duplications since its divergence from cotton (Bowers *et al.* 2003). The genomes of grape (Lin *et al.* 2011), papaya, and cacao have each remained unduplicated since their divergence from cotton – but have received far less attention to understanding gene functions, and accordingly offer far less information to cotton at present. Nonetheless, the cotton community should remain attuned to new information about these genomes as it may 'translate' especially well to cotton.

In partial summary, to understand and manipulate the features that make cotton unique will require new enabling tools, technologies, and resources. A few particularly high priorities among these are likely to include (in random order):

1) Large-scale expression profiling of the full set of cotton genes (indeed, preferably the entire genome) across a comprehensive sampling of *Gossypium* species, tissues,

organs and developmental states, to permit deductions about gene function based on coordinated expression patterns. Such information is rapidly accumulating thanks to the ability of next-generation sequencing technologies (Shendure and Aiden 2012) to economically and quickly capture information about messenger RNA, as well as DNA.

2) Large-scale sampling of patterns of between-species divergence and within-species diversity of the full set of cotton genes (indeed, preferably entire genomes), as detailed above providing the means to distinguish among genes that show evolutionary patterns such as:

   - Divergence to novel function in a particular clade (for example, the A-genome diploids), followed by purifying selection within that clade suggesting that the new function is under strong selection;
   - Divergence to new function in a clade, with continuing positive selection within the clade such as might be expected in the ongoing 'arms war' between plants and their pests;
   - Conservative evolution across otherwise divergent clades, suggesting that the ancestral function is broadly adaptive and under purifying selection.

3) Comprehensive mutant resources. Strategies for *Gossypium* functional genomics need to anticipate that many genes may be implicated in crop improvement by association genetics approaches that would benefit from functional validation. Comprehensive mutant populations, using established techniques (McCallum *et al.* 2000; Till *et al.* 2003; Slade *et al.* 2005; Comai and Henikoff 2006; Tsai *et al.* 2011) that are likely to become still faster and less costly using next-generation sequencing technologies, can provide a means by which functional analysis of *Gossypium* genes can be carefully-targeted to complement and supplement more extensive resources for *Arabidopsis* and other botanical models. This approach will provide for both the study of genes/gene families that are less tractable in other plants, and also for targeting functional analyses to specific genes implicated in key cotton traits by association genetics or other approaches. Such resources are ideally needed for each of the two cultivated tetraploid species (to permit study of duplicated gene fates during all-important adaptation to the polyploid state) and each of the diploid genome types, with priority placed on the A and D genome progenitors of the tetraploid.

4) Well-characterized populations of diverse genotypes that are carefully selected to broadly and deeply sample allelic variation within particular gene pools. Such 'diversity panels' (Morris *et al.* 2013) comprised of a few hundred individuals, including careful phenotyping of these individuals, offer the means to utilize historical accumulations of recombination events to associate relatively abundant alleles with phenotypes, providing more precise 'mapping' than can generally be accomplished using conventional QTL mapping (Paterson *et al.* 1988).

5) New genetic populations of two types:

   - 'Tiling paths' of NIILs that collectively cover the genome of a target genotype, toward genome wide application (Eshed and Zamir 1995) of the 'substitution mapping' strategy (Paterson *et al.* 1990) providing for fine-scale (1-3 cM) dissection of complex variation into individual components. This approach reveals both predictable alleles and 'cryptic' variation (Gibson and Dworkin 2004) not expected based on the parental phenotypes but often of practical value (Eshed and Zamir 1995; Tanksley *et al.* 1996; Fulton *et al.* 1997; Bernacchi *et al.* 1998a; Bernacchi *et al.* 1998b; Bernacchi *et al.* 1998c; Fridman *et al.* 2004; Chee *et al.* 2005a; Chee *et al.* 2005b; Draye *et al.* 2005; Schauer *et al.* 2006). The precision afforded by the NIILs provides a foundation for establishing causality between phenotypes and specific mutations.
   - Nested association mapping populations, that combine the ability to search much diverse germplasm for novel variation with the ability to precisely map the novel variation, guiding the breeder to the specific recombinants needed to separate desirable from undesirable alleles/effects (Yu *et al.* 2005; Yu *et al.* 2008).

Synthesis

In closing, the potential benefits of the post-genomic era in cotton are real and large – improved quality, productivity, and stability; reduced input needs that improve sustainability and environmental stewardship; and value-added features tailored to human needs rather than natural adaptation. The 8 divergent genomes in the *Gossypium* (cotton) *genus* enjoy a broad spectrum of morphological and physiological diversity that has permitted species within the genus to adapt to a wide range of ecosystems in warmer, arid regions of the world. Virtually all of this diversity is conferred by genes that are not yet identified, and the vast majority is found in taxa that are presently beyond the reach of mainstream breeding programs. Identification of genes native to *Gossypium* that confer desirable adaptations or traits, together with their rapid and specific transfer to elite genotypes, may provide a means to harness this variability in a manner that is minimally subject to public concerns.

The greatest challenge facing the cotton community is the conversion of 'sequence' to 'knowledge,' a challenge that will require investment, creativity, investment, energy, investment, coordination, investment, patience, and investment. The sequence(s) are laying bare the secrets of the genetic potential of the *Gossypium* genus, if we are clever enough to find appropriate ways to recognize them. In the 'simple' botanical model *Arabidopsis thaliana*, publication of its sequence in

2000 (Initiative 2000) was followed shortly by the inception of the *Arabidopsis* 2010 project by the US National Science Foundation, and similar projects in other countries, with the goal of determine the function of each of the (~30,000) *Arabidopsis* genes by the year 2010. To date, the *Arabidopsis* 2010 project alone has invested more than $200 million toward this goal (www.nsf.gov/bio/pubs/awards/2010awards.htm), with additional investments made in other countries, and by private firms. While the cotton genome will derive much benefit from Arabidopsis (detailed above), the greater complexity of cotton will require a similar level of investment in its unique genes and features in order to fully realize the potential benefits of its sequencing.

While some ongoing investments in cotton genomics may be in intellectual property of potential commercial value that are appropriately made in the private sector, *many will be in pre-competitive enabling tools that might most efficiently be produced in the public domain or by public-private consortia*. In an industrial crop such as cotton, public-private consortia are particularly attractive, engaging core competencies of public researchers as a 'virtual research and development network' that offers new opportunities for small and medium-sized businesses while also enhancing opportunity for large businesses, by providing new tools, information, and young scientists with the expertise to put these resources to work. Many of the challenges, particularly regarding the spectrum of adaptations that permit cotton to adapt to such a wide range of ecosystems, may best be met by international collaborations.

Acknowledgements

The author thanks many colleagues and collaborators with whom he has partnered in research mentioned herein, and alphabetically Dr. Peng Chee (University of Georgia), Dr. Don Jones (Cotton Incorporated), and Dr. Robert Wright (Texas Tech University) for helpful comments on this manuscript.

References

Beasley, J. O., 1940 The origin of American tetraploid *Gossypium* species. *Amer Naturalist* **74:** 285-286.

Beasley, J. O., 1942 Meiotic chromosome behavior in species hybrids, haploids, and polyploids of *Gossypium. Genetics* **27:** 25-54.

Benton, M. J., 1993 *The Fossil Record* 2. Chapman and Hall, New York.

Bernacchi, D., T. Beck-Bunn, D. Emmatty, Y. Eshed, S. Inai *et al.*, 1998a Advanced back-cross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from Lycopersicon hirsutum and L-pimpinellifolium (vol 97, pg 170, 1998). T*heoretical and Applied Genetics* **97:** 1191-1196.

Bernacchi, D., T. Beck-Bunn, Y. Eshed, S. Inai, J. Lopez *et al.*, 1998b Advanced backcross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from Lycopersicon hirsutum and L-pimpinellifolium. *Theoretical and Applied Genetics* **97:** 170-180.

Bernacchi, D., T. Beck-Bunn, Y. Eshed, J. Lopez, V. Petiard *et al.*, 1998c Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from Lycopersicon hirsutum. *Theoretical and Applied Genetics* **97:** 381-397.

Bowers, J. E., B. A. Chapman, J. Rong and A. H. Paterson, 2003 Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422:** 433-438.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, and P. J. Brown *et al.*, 2009 The Genetic Architecture of Maize Flowering Time. *Science* **325:** 714-718.

Chee, P., X. Draye, C. X. Jiang, L. Decanini, T. A. Delmonte *et al.*, 2005a Molecular dissection of interspecific variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: I. Fiber elongation. *Theoretical and Applied Genetics* **111:** 757-763.

Chee, P. W., X. Draye, C. X. Jiang, L. Decanini, T. A. Delmonte *et al.*, 2005b Molecular dissection of phenotypic variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: III. Fiber length. *Theoretical and Applied Genetics* **111:** 772-781.

Comai, L., and S. Henikoff, 2006 TILLING: practical single-nucleotide mutation discovery. *Plant Journal* **45:** 684-694.

Doolittle, W., and C. Sapienza, 1980 Selfish genes, the phenotype paradigm, and genome evolution. *Nature* **284:** 601-603.

Draye, X., P. Chee, C. X. Jiang, L. Decanini, T. A. Delmonte *et al.*, 2005 Molecular dissection of interspecific variation between *Gossypium hirsutum* and *G-barbadense* (cotton) by a backcross-self approach: II. Fiber fineness. *Theoretical and Applied Genetics* **111:** 764-771.

Eshed, Y., and D. Zamir, 1995 An Introgression Line Population of Lycopersicon Pennellii in the Cultivated Tomato Enables the Identification and Fine Mapping of Yield-Associated OTL. *Genetics* **141:** 1147-1162.

Fridman, E., F. Carrari, Y. S. Liu, A. R. Fernie and D. Zamir, 2004 Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* **305:** 1786-1789.

Fulton, T. M., T. BeckBunn, D. Emmatty, Y. Eshed, J. Lopez *et al.*, 1997 QTL analysis of an advanced backcross of Lycopersicon peruvianum to the cultivated tomato and comparisons with QTLs found in other wild species. *Theoretical and Applied Genetics* **95:** 881-894.

Gerstel, D. U., 1958 Chromosomal translocations in interspecific hybrids of the genus *Gossypium*. *Evolution* **7:** 234-244.

Gibson, G., and I. Dworkin, 2004 Uncovering cryptic genetic variation. *Nature Reviews Genetics* **5:** 681-U611.

Initiative, T. A. G., 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408:** 796-815.

Jiang, C. X., R. J. Wright, K. M. El-Zik and A. H. Paterson, 1998 Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). Proceedings of the National Academy of Sciences of the United States of America **95:** 4419-4424.

Kim, H. J., and B. A. Triplett, 2004 Cotton fiber germin-like protein. I. Molecular cloning and gene expression. *Planta* **218:** 516-524.

Lin, L., H. Tang, R. O. Compton, C. Lemke, L. K. Rainville *et al.*, 2011 Comparative analysis of *Gossypium* and Vitis genomes indicates genome duplication specific to the *Gossypium* lineage. *Genomics* **97:** 313-320.

Lynch, M., and A. Force, 2000 The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154:** 459-473.

McCallum, C. M., L. Comai, E. A. Greene and S. Henikoff, 2000 Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiology* **123:** 439-442.

McMullen, M. D., S. Kresovich, H. S. Villeda, P. Bradbury, H. H. Li *et al.*, 2009 Genetic Properties of the Maize Nested Association Mapping Population. *Science* **325:** 737-740.

Morris, G. P., P. Ramu, S. P. Deshpande, C. T. Hash, T. Shah *et al.*, 2013 Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proceedings of the National Academy of Sciences of the United States of America **110:** 453-458.

Myers, R. M., J. Stamatoyannopoulos, M. Snyder, I. Dunham, R. C. Hardison *et al.*, 2011 A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **9:** e1001046.

Nagaki, K., Z. K. Cheng, S. Ouyang, P. B. Talbert, M. Kim *et al.*, 2004 Sequencing of a rice centromere uncovers active genes. *Nature Genetics* **36:** 138-145.

Nielsen, R., and Z. Yang, 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148:** 929-936.

Paterson, A. H., 2007 Sequencing the cotton genomes, pp. in *World Cotton Research Conference*. International Cotton Advisory Committee, Lubbock TX.

Paterson, A. H., J. W. Deverna, B. Lanini and S. D. Tanksley, 1990 Fine Mapping of Quantitative Trait Loci Using Selected Overlapping Recombinant Chromosomes, in an Interspecies Cross of Tomato. *Genetics* **124:** 735-742.

Paterson, A. H., M. Freeling, H. Tang and X. Wang, 2010 Insights from the Comparison of Plant Genome Sequences. *Annual Review of Plant Biology* **61:** 349-372.

Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln *et al.*, 1988 Resolution of Quantitative Traits Into Mendelian Factors By Using a Complete Linkage Map of Restriction Fragment Length Polymorphisms. *Nature* **335:** 721-726.

Paterson, A. H., J. Wendel, F., H. Gundlach, H. Guo, J. Jenkins *et al.*, 2012 Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492:** 423-427.

Phillips, L. L., 1963 The cytogenetics of *Gossypium* and the origin of New World cottons. *Evolution* **17:** 460-469.

Reva, B., Y. Antipin and C. Sander, 2011 Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* **39:** e118.

Rong, J., J. E. Bowers, S. R. Schulze, V. N. Waghmare, C. J. Rogers *et al.*, 2005 Comparative genomics of *Gossypium* and *Arabidopsis*: Unraveling the consequences of both ancient and recent polyploidy. *Genome Research* **15:** 1198-1210.

Rong, J., E. A. Feltus, V. N. Waghmare, G. J. Pierce, P. W. Chee *et al.*, 2007 Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. *Genetics* **176:** 2577-2588.

Schauer, N., Y. Semel, U. Roessner, A. Gur, I. Balbo *et al.*, 2006 Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature Biotechnology* **24:** 447-454.

Senchina, D. S., I. Alvarez, R. C. Cronn, B. Liu, J. K. Rong *et al.*, 2003 Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Molecular Biology and Evolution* **20:** 633-643.

Shendure, J., and E. L. Aiden, 2012 The expanding scope of DNA sequencing. *Nature Biotechnology* **30:** 1084-1094.

Skovsted, A., 1934 Cytological studies in cotton. II. Two interspecific hybrids between Asiatic and New World cottons. *J. Genet* **28:** 407-424.

Slade, A. J., S. I. Fuerstenberg, D. Loeffler, M. N. Steine and D. Facciotti, 2005 A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nature Biotechnology* **23:** 75-81.

Small, R. L., and J. F. Wendel, 1999 The mitochondrial genome of allotetraploid cotton (*Gossypium* L.). *Journal of Heredity* **90:** 251-253.

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner and C. F. Aquadro, 2001a Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proceedings of the National Academy of Sciences of the United States of America **98:** 7375-7379.

Swanson, W. J., Z. H. Zhang, M. F. Wolfner and C. F. Aquadro, 2001b Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proceedings of the National Academy of Sciences of the United States of America **98:** 2509-2514.

Tanksley, S. D., S. Grandillo, T. M. Fulton, D. Zamir, Y. Eshed *et al.*, 1996 Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L-pimpinellifolium*. *Theoretical and Applied Genetics* **92:** 213-224.

Tian, F., P. J. Bradbury, P. J. Brown, H. Hung, Q. Sun *et al.*, 2011 Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics* **43:** 159-162.

Till, B. J., S. H. Reynolds, E. A. Greene, C. A. Codomo, L. C. Enns *et al.*, 2003 Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Researc*h **13:** 524-530.

Tsai, H., T. Howell, R. Nitcher, V. Missirian, B. Watson *et al.*, 2011 Discovery of Rare Mutations in Populations: TILLING by Sequencing. *Plant Physiology* **156:** 1257-1268.

Wang, G. L., J. M. Dong and A. H. Paterson, 1995 The Distribution of *Gossypium-hirsutum* Chromatin in *Gossypium-barbadense* Germ Plasm - Molecular Analysis of Introgressive Plant-Breeding. *Theoretical and Applied Genetics* **91:** 1153-1161.

Wang, K., Z. Wang, F. Li, W. Ye, J. Wang *et al.*, 2012 The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics* **44:** 1098-1103.

Wendel, J. F., 1989 New World Tetraploid Cottons Contain Old-World Cytoplasm. Proceedings of the National Academy of Sciences of the United States of America **86:** 4132-4136.

Wendel, J. F., and R. C. Cronn, 2003 Polyploidy and the evolutionary history of cotton, pp. 139-186 in *Advances in Agronomy, Vol 78*.

Yang, Z., 1998 Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol*. **15:** 568-573.

Yang, Z., R. Nielsen, N. Goldman and A. Krabbe Pedersen, 2000a Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**.

YANG, Z. H., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13:** 555-556.

YANG, Z. H., R. NIELSEN, N. GOLDMAN and A. M. K. PEDERSEN, 2000b Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155:** 431-449.

YU, J., G. PRESSOIR, W. BRIGGS, I. V. BI, M. YAMASAKI *et al.*, 2005 A Unified Mixed-Model Method for Association Mapping that Accounts for Multiple Levels of Relatedness. *Nature Genetics* **38:** 203-208.

YU, J. M., J. B. HOLLAND, M. D. MCMULLEN and E. S. BUCKLER, 2008 Genetic design and statistical power of nested association mapping in maize. *Genetics* **178:** 539-551.