Chapter

Cotton as a Model for Polyploidy and Fiber Development Study

Venera S. Kamburova, Ilkhom B. Salakhutdinov, Shukhrat E. Shermatov, Zabardast T. Buriev and Ibrokhim Y. Abdurakhmonov

Abstract

Cotton is one of the most important crops in the world. The *Gossypium* genus is represented by 50 species, divided into two levels of ploidy: diploid (2n = 26) and tetraploid (2n = 52). This diversity of *Gossypium* species provides an ideal model for studying the evolution and domestication of polyploids. In this regard, studies of the origin and evolution of polyploid cotton species are crucial for understanding the ways and mechanisms of gene and genome evolution. In addition, studies of polyploidization of the cotton genome will allow to more accurately determine the localization of QTLs that determine fiber quality. In addition, due to the fact that cotton fibers are single trichomes originating from epidermal cells, they are one of the most favorable model systems for studying the molecular mechanisms of regulation of cell and cell wall elongation, as well as cellulose biosynthesis.

Keywords: cotton, polyploidy, genome evolution, cotton fiber, cell elongation

1. Introduction

Currently, the cotton (*Gossypium* L.) is one of the most important textile crops in the world, producing natural and quality fiber. For example, in 2017/18, the cotton world production and use were estimated at 25.1 million tons [1, 2]. As predicted, world cotton production will grow and reaching 26.1 million tons in 2026 [3].

The *Gossypium* genus is represented by more than 50 species, divided after ploidy into two groups: diploid (2n = 2x = 26) and tetraploid (2n = 4x = 52) [1, 4]. Moreover, 45 of species are diploid, and five remained species are tetraploid [4]. Among them, the diploid species such as *G. arboretum* L., *G. herbaceum* L. and tetraploid *G. hirsutum* L. and *G. barbadense* L. are cultivated only [4, 5]. Consequently, this kind of diversity of *Gossypium* species is a suitable model for studying the evolution, domestication and polyploidy, also to study of ploidy effect on the most important agronomic traits of cotton (e.g. fiber quality), as well as the expression and inheritance of corresponding genes of interest [6].

Similar to most plants, the evolution of cotton was characterized by repeating cycles of whole genome duplication [1, 6, 7]. At the same time, a parallel level of cytogenetic and genomic diversity emerged during the global widespread of the cotton, that finally led to the appearance of eight groups of diploid (n = 13) species (groups A-G and K of genomes) [1, 6]. It should be noted that despite the existence

of different types of polyploidy [1, 6], the most common type is allopolyploidy, when two differentiated genomes, usually of various species, are combined in one cell nucleus as a result of hybridization [1, 6].

Thus, allopolyploid duplication of the genome leads to numerous of molecular genetic interactions, interlocus concerted evolution, difference of genomic evolution rates, interlocus transfer of genetic material, and possibly to changes in gene expression [1, 6]. In addition, allopolyploidy may have stimulated the morphological, ecological and physiological adaptation of cotton through natural selection based on a higher level of variability such as a result of duplication of the gene set [1, 6].

For the same reasons, the genome duplication may have given new opportunity for cotton improvement by directional selection [7, 8]. Another important aspect of allopolyploidy is that not every allopolyploid has to strictly correspond to concept of the simple summation of the ancestral diploid genomes. In some cases, the fusion of two different genomes is accompanied by significant genomic reorganization and non-Mendelian genetic inheritance as result [7, 9].

Consider to the mentioned above, we would attempt to analyze the consequences of evolution of polyploids, including on genomic, epigenomic and phenotypic levels in this chapter.

2. Evolution of *Gossypium* genus

According to molecular genetic data, the history of cotton evolution has amounted about 10–15 million years, after the *Gossypium* diverged from other *Gossypieae* [6, 10, 11]. In the same time, the evolution of eight groups of diploid species (genomic groups A-G and K) also occurred by the cotton widespread, that led to the arising of parallel level of cytogenetic and genomic diversity [1, 6, 11]. It should be noted that molecular genetic and cytogenetic studies show that the species lineages on genealogical tree of the genus coincide with genomic groups A-G, K, and AD and geographic origin [11, 12].

The evolution studies of the *Gossypium* have shown that the origination of tetraploid species proceeded by polyploidization of A- (African) and D-genomes (American) diploid species [1, 6, 11]. Alloploidization of these two genomes occurred about 1.5–2 million years ago, resulting in five different genomes: *G. darwinii*, *G. tomentosum*, *G. mustelinum*, *G. hirsutum* and *G. barbadense*, where the last two belong to cultivated species [13]. It was also proved that during the alloploidization process the *G. arboreum* and *G. herbaceum* were as receptors of A-genome and should be a predecessors, because all existing polyploid species contain the cytoplasm of the A genome. At the same time, the D-genome donor was appear *G. raimondii* [11].

After occurrence of the predecessor of allotetraploid species, at the initial stage of divergence led to the origination of two evolutionary lines of cotton with AD genomes: the first includes *G. mustelinum* (AD4 genome), the second one – all other species (AD1 – AD3 and AD5 genomes). In other words, the follow-up divergence of the second evolutionary line of AD genomes led to the emergence of recent allotetraploid cotton species such as *G. hirsutum* (AD1 genome), *G. barbadense* (AD2 genome), *G. tomentosum* (AD3 genome), and *G. darwinii* (AD5 genome) [11, 12].

One of an important evolutionary events for *Gossypium* was appear the domestication of four wild species. This selection was based on the length and quality of cotton fiber, which is anatomically specialized unicellular trichomes located on the surface of the epidermis of seeds [10, 11]. This sequential process led to the domestication of four species of cotton: two American – *G. hirsutum* and *G. barbadense* and two Afro-Asian – *G. arboretum* and *G. herbaceum* [11].

Followed phylogenetic studies have shown the trait of prolonged elongation of trichomes has appeared first time in the A/F-genomes. Possibly, it was the reason to domestication of *G. arboretum* and *G. herbaceum* (A-genome). Unlike A-genome a number of species with D-genome (*G. thurberi*, *G. trilobum*, *G. davidsonii* and *G. klotzschianum*, and three species of the *Cauducibracteolata* subsection) lack of clearly visible fibers [11, 12]. This suggests that the traits of prolongation of trichomes were probably inherited by the allotetraploid (AD-genome) from the A-genome [11].

Moreover, the domestication of cotton species led to a change not only in the length of the fiber, but also in the chemical composition of its: the fiber of wild species besides cellulose contains suberin, while in cultivated species it is cellulose only [11].

Summarizing the information mentioned above, it should be noted that the *Gossypium* diverged from other *Gossypieae* in the Pleistocene period eventualy. This genus has evolved in two ways: divergence at diploid species(genomic groups A-G and K) and allopolyploidization of A- and D-genomes, followed by arising of tetraploid species (AD1 - AD5-genomes). Besides this, the domestication of these species and artificial selection based on fiber quality have also greate influenced on evolution of cultivated cotton.

3. Mechanisms of polyploidy

Polyploidization of eukaryotic genomes is an important evolutionary event that had a significant effect on the evolution of plants, including cotton [14–16]. Polyploids are divided into two large groups: autopolyploids and allopolyploids [17–20]. The difference between these two groups basically lies in the hybridization type: intraspecific hybridization occurs in autopolyploids, while allopolyploids arise by the combination of processes such as interspecies hybridization and duplication of chromosomes [17, 20].

In turn, there are two types of allopolyploids: true and segmental allopolyploids. True allopolyploids emerged due to hybridization of distantly related species, but segmental allopolyploids through hybridization of closely related species with partially different genomes [20]. In this case, segmental allopolyploids can be considered as an intermediate type between true allopolyploids and autopolyploids [20].

In autopolyploids, the presence of more than two homologous chromosomes in the genome may lead to formation of multivalents during meiosis. It contributes to the polysomic type of inheritance of traits. Whereas, in true allopolyploids bivalents are formed, that leads to disomic inheritance of traits. At the same process, in segmental allopolyploids monovalent, bivalent and/or multivalent chromosome pairing is observed during meiosis [20].

The second mechanism is the fusion of unreduced gametes – the basic factor of the natural emergence of polyploidy. In this case, the fusion of unreduced gametes may lead to unilateral- (fusion with a typically reduced gamete) or bilateral polypolydization (fusion with another unreduced gamete) [20].

The formation of unreduced gametes can occur due to errors during meiosis. In this case, errors during meiosis I (first division restitution – FDR) can be a consequence of a fail to chromosome pairing in prophase I (synaptene/pachytene) or separation of homologous chromosomes in anaphase I [20]. At the same time, errors during meiosis II (second division restitution - SDR) occur in anaphase II due to the fail to separation and segregation of sister chromatids [20]. Both of FDR and SDR lead to a chromosome set doubling in gametes, resulted in dyads or triads formation [21].

Depending on the meiotic restitution mechanism, a polyploidization consequences will differ. Thus, after FDR, the heterozygosity level of unreduced gametes will be similar to the original gametes, while SDR leads to a decrease in the level of heterozygosity of unreduced gametes [20]. The heterozygosity level of a resulting polyploids will be of decisive importance both in the struggle for survival as well as by artificial selection.

Polyploidy had a significant effect on the evolution process and formation of species by increasing phenotypic variability, heterosis, and mutation resistance. On the other hand, in terms of evolution, allopolyploidization (interspecific hybridization) is more preferable due to the pronounced effect of heterosis, that manifest in increasing of biomass, growth and its rate, fertility and resistance of occured hybrids to stress [22]. Thus, in tetraploid cultivated cotton species (*G. hirsutum* and *G. barbadense*) the quality and yield of fiber are much higher than cultivated diploids (*G. arboretum* and *G. herbaceum*) [23].

Resuming the above, polyploidization is rather widespread phenomenon in plant evolution (the number of polyploid species is approximately ¼ of the total number of vascular plant species) [24]. At the same time, the polyploidy occurrence brings an evolutionary "benefit" to a species, increasing its chances in the struggle for survival.

4. Genomic consequences of polyploidization

The allopolyploidization process of cotton genome could not be considered as the simple sum of the A- and D-genomes. It has been shown that genome duplication leads to various molecular genetic interactions e.g.: interlocus consistent evolution, different rates of genomes evolution, interlocus transfer of genetic material and changes in gene expression [1, 6, 17].

Additionally, according to the latest molecular data tetraploid cotton species are at least paleo-octaploids, and diploid species are paleo-tetraploids. Due to this fact cotton may be a good model system for studying consequences of genome polyploidization [6, 9, 25].

In connection with the above, let us review the changes that occurred after polyploidization of the cotton genome.

4.1 Genome stability

Despite the fact that diploid Gossypium species have the same chromosome basic number (n = 13), the DNA length in different species widely varies from ~900 Mb in D-genomes to ~2500 Mb in K-genomes [1, 6, 17]. Moreover, the analysis of bivalents formation in the metaphase of meiosis also suggest that diploid cotton species are actually paleopolyploid organisms [6]. A number of studies have also shown that the ancestor of *Gossypium* went off through cycles of polyploidization, followed by the loss of a part of homologous genes and diploidization [6, 26, 27].

In this respect it should be noted that allopolyploidization of cotton has not only characterized by rearrangements at the chromosome level [1, 6]. This assumption was confirmed by both classical cytogenetic and molecular genetic data [1, 6]. Thus, cytogenetic data show that chromosomes of A- and D-genome less form bivalents after crossing of allotetraploids compared to diploid species hybrids [1, 6]. For example, hybrids of allotetraploids form less than one bivalent per cell in the meiotic metaphase, while hybrids of present diploids of A- as well as D-genome form, on average, 5.8 and 7.8 bivalents [1, 6].

Additionally, the analysis of the order and syntheny of genes in the A- and D-genomes as well as allopolyploid genomes (A versus At and D versus Dt) showed a low level of structural chromosome rearrangements with a retention of collinear linkage groups [28]. Along with this, AFLP analysis of nine artificial allotetraploid and allohexaploid cotton species showed a significant additivity of genetic loci [1, 6].

Summarizing the facts, it can be assumed that the cotton genome stabilization after polyploidization led to such reorganization of the original genomes that they were no longer able to homeological pairing [1, 6].

Thus, it can be concluding that the cotton genome is quite stable and genome stabilization is not achieved through structural rearrangements unlike some other plant models with polyploid genome.

4.2 Mobile elements in genome

As mentioned above, the genome size of different cotton species differs significantly even the same basic number of chromosomes [1, 6, 29]. This may be conditioned with a number mobile genetic elements (MGE) in the *Gossypium* genome [6]. Wu et al. (2017) have shown that the *Gossypium* genome contains a large number of MGE, particularly a long terminal repeat (LTR) retrotransposons in compare to *Theobroma cacao* (L.) and *A. thaliana* (L.) Heynh [30].

Moreover, the analysis of the genomes of *G. raimondii*, *G. arboreum*, and *G. hirsutum* showed that the greatest number of MGE, especially LTR-retrotransposons is observed in A- and AD-genome [6, 12, 31, 32]. However, the frequency of occurrence of *Copia* LTR retrotransposons is higher in *G. raimondii* (D5 genome) – the smallest genome size (885 Mb). At the same time, the occurrence frequency of the *Gypsy* LTR retroelements is higher in species with a large genome size [6, 32–34]. Additionally, it was established that the wide distribution of *GORGE3* (*Gossypium* retrotransposable *gypsy*-like element) in A- and AD-genome was the reason for their upsizing [31, 32, 35, 36].

It has been also found that besides the genome resizing in various cotton species, MGEs have also affected on the expression of genes responsible for fiber development [30, 32]. Thus, in D-subgenome was observed the insertion of the *Copia* LTR retrotransposon into promoter region of the gene encoding the transcription factor *GhMYB25*. This well consists with the facts of hyperexpression of the D-genome homeolog in *G. hirsutum* [32]. Similarly, the insertion of the LINE retrotransposon into promoter of ethylene response factor (*GhERF*) gene in D-subgenome increases the expression level of the D-homeologue in compare to its A-copy [32].

It has been also suggested that the silencing of CICR (Chinese Institute of Cotton Research) LTR elements had an appreciable effect on the formation of allotetraploid cotton species, because the occurrence frequency of these MGEs is significant in the A-subgenomes, and practically not occur in the D-subgenomes [37].

Summarize this, presence of mobile elements in a genome, their polymorphism and occurrence frequency, probably had the significant influence on the cotton evolution. In addition, MGE are involved in regulation of activity of genes responsible for fiber quality.

4.3 Asymmetric evolution of the genome

Hereof the *Gossypium* has both diploid and tetraploid genome, it makes cotton an ideal model to study of the homeologous genes evolution and their expression after polyploidization.

As mentioned above, the extended trichomes elongation trait was probably inherited by the allotetraploid AD-genomes from the A-genome [11]. Further evolution of domesticated tetraploids (*G. hirsutum* and *G. barbadense*) was done under the influence of artificial selection directed on improving fiber quality. Its led to the asymmetric evolution of the A- and D-subgenomes. According Li et al. (2015) in *G. hirsutum* the mutation frequency and formation rate of single nucleotide polymorphisms (SNPs) within intergenic collinear regions of the Dt-subgenome were significantly higher than in the At-genome [31]. Meanwhile, established Ks values for pairs of collinear genes in the At- and Dt-subgenomes were less than in the corresponding diploid A- and D-genomes. It was also shown reducing of dN/dS ratio in Dt/D pair in comparison with *T. cacao* and similar indicators for At/A [31].

In addition, scientists have found a greater extension of total rearrangements in At-subgenome (372.6 Mb) compared to Dt-subgenome (82.6 Mb) by comparative study of interchromosomal rearrangements and SNP frequency in *G. hirsutum* and *G. barbadense* [38]. It was also shown that SNP frequency is increased in the At-subgenome in both *G. hirsutum* and *G. barbadense* by comparing the Dt-subgenome (5.95 per thousand nucleotides in At-subgenome versus of 5.81 in the Dt-subgenome) [38].

These data also show that allotetraploid genomes due to genetic redundancy are being under less pressure from stabilizing selection, and directed selection by fiber quality has a greater effect on the At-subgenome [31, 38].

The asymmetry of these subgenomes is also appeared by the mutation types occurring in allotetraploid genomes of *G. hirsutum* and *G. barbadense*. Thus, it was found that duplications in the At-subgenome were more conserved than in the Dt-subgenome of *G. hirsutum*. At the same time, there are more conservative deletions in Dt-subgenome compared to the At-subgenome of *G. barbadense* [39]. These data indicate that artificial selection during cotton domestication furthered the fixation of duplications in the At-subgenome in *G. hirsutum*, and deletions in the Dt-subgenome of *G. barbadense*. It may have contributed to the development of a higher fiber quality in Pima cotton that distinguishes the species from others [39].

Differences in subgenomes are also manifested by different occurrence of frequency and activity of MGE. Two independent research groups have found that MGE number in At-subgenome exceeded the same parameter in Dt-subgenome [31, 40]. At the same time, the frequency of LTR-*Gypsy* occurrence in the At-subgenome was significantly higher than in the Dt-subgenome [31, 40]. Li et al. (2015) have also found that subgenomes differ not only in the MGEs number within them, but also by transcriptional activity and location [31]. Thus, it was shown that the transcription level of both LTR-*Copia* and LTR-*Gypsy* was increased in the Dt-subgenome compared with the At-subgenome [31]. However, LTR-*Copia* were more active and more frequently located near the coding genes when compared to LTR-*Gypsy* [16].

The asymmetry is also manifested in the unequal expression of At- or Dt-homeologs, which regulate fiber development in cotton [31, 41–43]. The expression level of homeologs of some transcription factors (eg, *MYB*) was significantly increased in the At-subgenome [31]. And the comprehensive proteomic analysis of the fiber of allopolyploid species (*G. hirsutum* and *G. barbadense*) have shown that A-patterns of expression prevailed in *G. hirsutum* over ones in *G. barbadense* at different stages of fiber development. Thus, the expression level changed the direction of dominance from D-genome to A-genome [42].

Moreover, the results obtained using the RNA-seq technology on *G. hirsutum* have shown a shift on the level of homeologs expression towards the A-subgenome in allotetraploid cotton [44]. This shift of gene expression can be explained by the deactivation of homeologs in non-dominant D-subgenome due to negative

regulators (miRNA and transcriptional repressors) [6, 44]. It was also established that genes in A-subgenome may be responsible for the fiber development by regulation of fatty acids biosynthesis/metabolism and microtubules growing process. While the genes in D-subgenome may be involved to the transcription regulation and stress response [44].

Thus, the analysis of the available data allows to speak about the asymmetric evolution of allopolyploid cotton subgenomes with a shift in dominance towards A-subgenome.

5. Effects of polyploidy on fiber development

The fiber is one of the key point for domestication of four *Gossypium* species: two diploid *G. arboretum* and *G. herbaceum* (A-genome), as well as two tetraploid species *G. hirsutum* and *G. barbadense* (AD-genome) [11]. In the meantime, the domestication process of tetraploid species was independent, that have been confirmed both the sequencing data and significant differences in cotton fiber at the proteome level [42, 43, 45].

Cotton fiber is basically elongated single cell of seed epidermis (trichome) with a clear gradation of development stages: fiber initiation, elongation, secondary biosynthesis of the cell walls and maturation [33, 46, 47]. It first appeared among ancestral diploid cotton with A-genome after divergence with F-genome [1, 6, 48]. Allotetraploid species (AD genomes) have significantly higher fiber quality, that can be explained by the nucleotypic effect after allopolyploidization of A- and D-genome [48, 49].

Polyploidization has also led to increase of the number of nuclear genes associated with fiber development [47]. E.g., a number of studies have shown the content of Malvaceae specific genes of *MIXTA* family, encoding *MYB* transcription factors and regulating fiber development is significantly higher in allotetraploid species [50, 51]. Additionally, stabilization of the natural and artificial selection contributed a changes at the expression level of fiber development genes. It has been achieved either by epigenetic modifications (DNA methylation, miRNA and siRNA biogenesis) or by histones modification, among other factors [48, 52].

The fiber development in cotton is a complex process ensured by the coordinated action of many genes involvong to biosynthesis of polysaccharides, lipids and phytohormones, pro- and antioxidant system, calcium homeostasis, as well as transcription factor genes (*MYB, C2H2, bHLH, WRKY* and *HD-ZIP*) [40, 53–55]. At the same time, in tetraploid species, the expression and co-expression of genes at different stages of fiber development is different: some genes are expressed at the stage of fiber initiation, others - at the stages of fiber elongation and secondary cell walls biosynthesis [53, 54]. It has been shown that genes in the Dt-subgenome are predominantly expressed at the stage of fiber initiation, very important parameter to the fiber yield [1, 33].

The difference of gene expression level between *G. hirsutum* and *G. barbadense* was also established using whole genomes alighnment of both species. It was shown that a longer fiber of *G. barbadense* may be a result of more continuous activity of genes encoding sucrose transporter (*GbTSTl*), Na^+/H^+-antiporter (*GbNHXl*), aluminum-activated malate transporter (*GbALMT16*), vacuolar-localized vacuolar invertase (*GbVIN1*) and plasmodesmata (*PD*) [8].

It was also found that the fiber development in tetraploid is specified by gene expression in both At- and Dt-subgenome [1, 40, 48, 55]. Despite the fact that major genes for fiber quality were introduced into allopolyploids from A-genome, the genes in Dt-subgenome also take a significant effect on the fiber development in

tetraploid cotton [48]. For example, several researchers on the base of an integrated genetic and physical map of fiber development genes supposed that a transcription factors regulating the expression of fiber genes in At-subgenome are transcribed in Dt-subgenome [1, 56].

Along with this, another research group has identified 811 positively selected genes (PSG) in *G. hirsutum*, 591 of them were associated with fiber development [40, 55]. Along with this, another research group has identified 811 positively selected genes (PSG) in *G. hirsutum*, 591 of them were associated with fiber development [40, 55]. Moreover, 58% of these PSGs were localized in At-subgenome, and 42% of PSGs were identified in the Dt-subgenome only. Moreover, it has been shown that PSGs in At-subgenome are associated with beta-D-glucan biosynthesis, regulation of signal transduction, as well as carbohydrates and sucrose biosynthesis. While, PSGs in Dt-subgenome determine the stress responses, which, as is known, reflect on fiber development [40, 55, 57].

All of these results were confirmed by studies of functional enrichment of proteins differentially expressed in cotton fiber [42]. The results of the study of proteome in *G. hirsutum* and *G. barbadense* have shown that the dominant expression pattern of *G. hirsutum* was more similar to A-genome (*G. arboretum*), while dominant expression pattern of *G. barbadense* was different dependent on fiber development stage, and switched from Dt- subgenome to At-subgenome [42]. In this case, the dominant patterns of At-subgenome produced the enzymes involved to biosynthesis of alcohols, monosaccharides and hexoses, while the patterns of Dt-subgenome produced proteins involved in various stress responses [42]. These results allowed to suggest that similarity in fiber appearance of these two species arose during evolution but through different pathways at the proteomic level [42].

The results obtained by genome sequencing of tetraploid *G. hirsutum* and diploid *G. arboretum* and *G. raimondii* have shown that difference of gene expression between *G. hirsutum* and *G. raimondii* was significantly higher than between *G. hirsutum* and *G. arboretum* [44]. It has been also demonstrated a shift of the expression level towards the At-subgenome, explained by the authors as an activation/deactivation of Dt-homeologs by negative regulators such as miRNA and transcription repressors. Deactivation of Dt homeologues was confirmed by a reduced number of nonfunctional genes in the Dt-subgenome [44]. The other authors have shown that the Dt-subgenome dominant pattern of *G. hirsutum* is associated with stress responses (genes encoding phosphatidylinositol phosphate kinase PIPK, PIP (internal plasma membrane protein), calmodulin (CaM), ethylene receptors and ethylene response factors (ERF), ABA receptors (PYR/PYL), protein kinase SnRK and protein kinase PP2C [8].

Thus, all of these data show that hybridization of A- and D-genome in allopolyploids had a significant effect on the fiber development in cotton due to both nucleotypic effect as well as changes and differentiation at the expression level of homeologuesof in At- and Dt-subgenome. Obviously, At-genes are associated with the fiber development, while Dt-genes regulate the activity of At-genes towards to fiber quality and determine the adaptive capabilities of allotetraploid cotton to adverse environment conditions [8, 42, 44].

6. Differential evolution of subgenomes

Following the fusion of two genomes into a single nucleus due to allopolyploidy, it is expected that some genes will acquire mutations and become pseudogenes,

while others may diverge and acquire new functions [17–19]. However, it can be expected that these and other phenomena affecting the genes molecular evolution, will be equally distributed in the two allopoliploid genomes. This leads to a useful null hypothesis, that is, the evolutionary rates of nucleotide substitutions will be equivalent for duplicated homeologists [17–19]. This leads to the null hypothesis, according to which the evolutionary rates of nucleotide substitutions will be equivalent for duplicated homeologs [17–19]. Inference expectation is that both gene copies accumulate intraspecific diversity at equivalent rates. However, this is not always true, for example, when there is strong directional selection per gene copy [17–19]. However, in the presence of strong stabilizing selection per gene copy, this condition got broken [17–19].

Despite this, this model can be useful in study the mechanisms underlying differential evolutionary rates or different levels of diversity. Thus, if one of the homeolog becomes pseudogenized, while the others remain under the pressure of purifying selection, an increase in nucleotide diversity can be expected at a higher rate in the first locus than in the last one [15, 19]. Finding duplicated genes in the same nucleus simplifies the problem of isolating potentially important genomic forces from population-level factors that can influence diversity patterns, such as the selection system or effective population size [15, 19]. Since population factors are neutral in regards to the two homeologs, the observed differences in diversity are almost certainly associated with genetic or genomic processes [15, 17–19].

Gossypium allopolyploids is a suitable model for these studies, especially when the two genomes are largely collinear but genome size differ in twice [1, 14, 15, 58–60]. The assumption of unequal speeds evolution in A- and D-genomes was confirmed by the observations that synthetic A- and D-genomic hybrids may be formed only when the A genome is used as a recipient [6]. This phenomenon is confirmed by divergent indicators. Thus, the study of the levels of RFLP polymorphisms found in allopolyploid cotton has shown that the number of polymorphisms in the Dt-subgenome was greater than in the At-subgenome [1, 14, 15, 58–60]. Similarly, two independent phylogenetic analyzes allowed to find out that D-genomic sequences in allopolyploids have longer phylogenetic branches and higher evolutionary rates in comparison to their homeologous A-genomic sequences [1, 14, 58]. Moreover, localization of quantitative traits loci indicates higher rates of evolution in the D-subgenome [1, 14, 58–60].

In addition, a direct test of the null hypothesis of the nucleotide substitution rates equivalence for homeologous genes is provided by measuring of the levels of nucleotide diversity [1, 17–19]. If evolutionary forces are equal for duplicated genes, mutations must accumulate randomly towards the homeolog. Therefore, the number of detected alleles should be approximately equal for two gene copies in the study of allelic polymorphism [1, 17–19, 58]. This approach was used by researchers in the study of the nucleotide sequences of the alcohol dehydrogenase gene (*AdhA*) in *G. hirsutum* and *G. barbadense* [61]. In both allopolyploid species the estimates of nucleotide diversity were twice as high for the Dt-homeolog of *AdhA* gene [60]. Similar data were obtained in the study of other gene of alcohol dehydrogenase (*AdhC*) [62].

Thus, these data allowed to suggest the existence of the increasing rate of Dt-subgenome evolution of the allopolyploid *Gossypium*. In addition, the evolutionary forces affecting *Gossypium* subgenomes can be fundamentally different. At the same time, it should be noted that the molecular mechanisms underlying the differential evolution of subgenomes remain unclear. However, it is logical to assume that they are associated with a double difference in genomic size.

7. Conclusion and future prospect

Summarizing the aforementioned, due to *Gossypium* diversity including both diploid (2n = 2x = 26) and tetraploid (2n = 4x = 52) species, cotton may be an ideal model for studying the evolution of allopolyploids, as well as the influence of ploidy for the most important agronomic traits – cotton fiber quality [1, 6, 33, 55]. In addition, the presence four cultivated species (diploid - *G. arboretum* and *G. herbaceum* and tetraploid - *G. hirsutum* and *G. barbadense*) allow to use this plant as a model for studying the effect of artificial selection in domestication process to shift of the homeologous expression level in tetraploid towards one of the subgenome [1, 6, 33, 55]. Moreover, because of cotton fiber is a single and easily isolated cell with a clear gradated of developmental stages, it is a good model to study of fiber development mechanisms [47].

This chapter presents the results of research on the evolution of *Gossypium*, mechanisms of polyploidization, genomic consequences of polyploidy, including the role of mobile genetic elements and asymmetric expression of homeologues, as well as the polyploidy effect on fiber quality traits. These data clarify the evolution history of this genus and mechanisms that regulate the formation and elongation of fiber.

Despite the volume of the obtained data, there are many unsolved issues in cotton genomics. Thus, the study the subgenome asymmetry using LTR-elements will help to clarify the evolution of *Gossypium* genomes and their divergence in time. Analysis of MGE polymorphisms may help identify genes involving to development of cotton fiber.

In addition, the issues of sub- and neofunctionalization of duplicated genes remain unclear, as well as the mechanism and relationship of epigenetic regulation in asymmetric expression of homeologous genes.

Continuation of comparative transcriptome and proteomic studies will also make it possible to more accurately differentiate the of natural and artificial selection influence on cultivated cotton species. At the same time, these studies can be a good basis for a more complete characterization of the metabolic pathways underlying the fiber formation and development.

Such research as genotyping and more accurate assembly of reference genomes, pan-genomic approaches (sequencing of gene pool in a populations), big data analysis, genome editing, de-novo domestication and genomic selection, combined with the available data, will allow for more efficient development of new cotton varieties with the desired properties as well as developing of personalized farming technologies for this crop.

DOI: http://dx.doi.org/10.5772/intechopen.99568

Author details

Venera S. Kamburova*, Ilkhom B. Salakhutdinov, Shukhrat E. Shermatov, Zabardast T. Buriev and Ibrokhim Y. Abdurakhmonov
Center of Genomics and Bioinformatics, Academy of Science of Republic of Uzbekistan, Tashkent, Uzbekistan

*Address all correspondence to: venera.k75@gmail.com

References

[1] Wendel JF, Flagel LE, Adams KL. Jeans, Genes, and Genomes: Cotton as a Model for Studying Polyploidy. In: Soltis P, Soltis D, editors. Polyploidy and Genome Evolution. Berlin: Springer; 2012. p. 181-207. DOI: 10.1007/978-3-642-31442-1_10

[2] Khan MA, Wahid A, Ahmad M, Tahir MT. World Cotton Production and Consumption: An Overview. In: Ahmad S, Hasanuzzaman M, editors. Cotton Production and Uses. Singapore: Springer Nature Singapore Pte Ltd; 2020. p. 1-7. DOI: 10.1007/978-981-15-1472-2_1

[3] Cotton [Internet]. Available from: http://www.fao.org/3/BT093e/BT093e.pdf

[4] Emani C. Transgenic Cotton for Agronomical Useful Traits. In: Ramawat K, Ahuja M, editors. Fiber Plants. Sustainable Development and Biodiversity. Vol. 13. Cham: Springer; 2016. p. 201-216. DOI: 10.1007/978-3-319-44570-0_10

[5] Sun Y, Zhang X, Huang C, Guo X, Nie Y. Somatic embryogenesis and plant regeneration from different wild diploid cotton (*Gossypium*) species. Plant Cell Rep. 2006;25:289-296. DOI: 10.1007/s00299-005-0085-2

[6] Strygina K, Khlestkina E, Podolnaya L. Cotton genome evolution and features of its structural and functional organization. Bio. Comm. 2020;65(1):15-27. DOI: 10.21638/spbu03.2020.102

[7] McGrath CL, Lynch M. Evolutionary Significance of Whole-Genome Duplication. In: Soltis P, Soltis D, editors. Polyploidy and Genome Evolution. Berlin: Springer; 2012. p. 1-20. DOI: 10.1007/978-3-642-31442-1_1

[8] Hu Y, Chen J, Fang L, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat Genet. 2019;51:739-748. DOI: 10.1038/s41588-019-0371-5

[9] Meng F, Pan Y, Wang J, Yu J, et al. Cotton Duplicated Genes Produced by Polyploidy Show Significantly Elevated and Unbalanced Evolutionary Rates, Overwhelmingly Perturbing Gene Tree Topology. Front. Genet. 2020;11:239. DOI: 10.3389/fgene.2020.00239

[10] Brubaker CL, Bourland FM, Wendel JF. The origin and domestication of cotton. In Smith CW, Cothren JT, editors. Cotton: Origin, History, Technology and Production. New York: Wiley; 1999. p. 3-31.

[11] Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM. Evolution and Natural History of the Cotton Genus. In: Paterson AH, editor. Genetics and Genomics of Cotton. Plant Genetics and Genomics: Crops and Models. Vol. 3. New York: Springer; 2009. p. 3-22. DOI: 10.1007/978-0-387-70810-2_1

[12] Kim, H. J. Fiber Biology. In: Fang DD, Percy RG, editors. Cotton, 2nd ed., Agron. Monogr. 57. Madison: ASA, CSSA, and SSSA; 2015. p. 97-128. DOI: 10.2134/agronmonogr57.2013.0022

[13] Abdurakhmonov IY, Buriev ZT, Shermatov SS, Abdullaev AA, Urmonov K, Kushanov F, et al. Genetic diversity in *Gossypium* genus. In Galiskan M, editor. Genetic diversity in Plants. Rijeka: InTech Press; 2012. p. 331-338. DOI: 10.5772/35384

[14] Wendel JF, Grover CE. Taxonomy and Evolution of the Cotton Genus, *Gossypium*. In: Fang DD, Percy RG, editors. Cotton, 2nd ed., Agron. Monogr. 57. Madison: ASA, CSSA,

and SSSA; 2015. p. 25-44. DOI: 10.2134/agronmonogr57.2013.0020

[15] Madlung A. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity. 2013;110:99-104. DOI: 10.1038/hdy.2012.79

[16] Chen ZJ, Sreedasyam A, Ando A, et al. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. Nat Genet. 2020;52:525-533. DOI: 10.1038/s41588-020-0614-5

[17] Soltis PS. Hybridization in Plants. In Levin SA, editor. Encyclopedia of Biodiversity (Second Edition). Oxford: Academic Press; 2013. p. 166-176. DOI: 10.1016/b978-0-12-384719-5.00202-1

[18] Counterman BA. Hybrid Speciation. In Kliman RM, editor. Encyclopedia of Evolutionary Biology. Oxford: Academic Press; 2016. p. 242-248. DOI: 10.1016/b978-0-12-800049-6.00072-x

[19] Blackman BK. Speciation Genes. In Kliman RM, editor. Encyclopedia of Evolutionary Biology. Oxford: Academic Press; 2016. p. 166-175. DOI: 10.1016/b978-0-12-800049-6.00066-4

[20] Sattler MC, Carvalho CR, Clarindo WR. The polyploidy and its key role in plant breeding. Planta. 2016;243:281-296. DOI: 10.1007/s00425-015-2450-x

[21] Ramanna MS, Jacobsen E. Relevance of sexual polyploidization for crop improvement: a review. Euphytica. 2003;133:3-8. DOI: 10.1023/A:1025600824483

[22] Chen ZJ. Molecular mechanisms of polyploidy and hybrid vigor. Trends Plant Sci. 2010;15:57-72. DOI: 10.1016/j.tplants.2009.12.003

[23] Renny-Byfield S, Wendel JF. Doubling down on genomes: polyploidy and crop plants. Am J Bot. 2014;101:1-15. DOI: 10.3732/ajb.1400119

[24] Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. On the relative abundance of autopolyploids and allopolyploids. New Phytol. 2016;210(2):391-398. DOI: 10.1111/nph.13698

[25] Renny-Byfeld S, Gong L, Gallagher JP, Wendel JF. Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. Molecular Biology and Evolution. 2015;32(4):1063-1071. DOI: 10.1093/molbev/msv001

[26] Renny-Byfeld S, Gallagher JP, Grover CE, Szadkowski E, et al. Ancient gene duplicates in *Gossypium* (cotton) exhibit near-complete expression divergence. Genome Biology and Evolution. 2014;6(3):559-571. DOI: 10.1093/gbe/evu037

[27] Birchler J.A. Genetic Consequences of Polyploidy in Plants. In: Soltis P, Soltis D, editors. Polyploidy and Genome Evolution. Berlin: Springer; 2012. p. 21-32. DOI: 10.1007/978-3-642-31442-1_2

[28] Zhang T, Endrizzi JE. Cytology and Cytogenetics. In: Fang DD, Percy RG, editors. Cotton, 2nd ed., Agron. Monogr. 57. Madison: ASA, CSSA, and SSSA; 2015. p. 129-154. DOI: 10.2134/agronmonogr57.2013.0023

[29] Wang M, Yuan D, Zhang X. Genome Sequencing. In: Fang DD, Percy RG, editors. Cotton, 2nd ed., Agron. Monogr. 57. Madison: ASA, CSSA, and SSSA; 2015. p. 289-302. DOI: 10.2134/agronmonogr57.2013.0028

[30] Wu Z, Yang Y, Huang G, Lin J, Xia Y, Zhu Y. Cotton functional genomics reveals global insight into genome evolution and fiber development. Journal of Genetics and Genomics 2017;44(11):511-518. DOI: 10.1016/j.jgg.2017.09.009

[31] Li F, Fan G, Lu C, Xiao G, et al. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nature Biotechnology. 2015;33(5):524-530. DOI: 10.1038/nbt.3208

[32] Wang K, Huang G, Zhu Y. Transposable elements play an important role during cotton genome evolution and fiber cell development. Science China Life Sciences. 2016;59(2):112-121. DOI: 10.1007/s11427-015-4928-y

[33] Pan Y, Meng F, Wang X. Sequencing Multiple Cotton Genomes Reveals Complex Structures and Lays Foundation for Breeding. Front. Plant Sci. 2020;11:560096. DOI: 10.3389/fpls.2020.560096

[34] Orozco-Arias S, Isaza G, Guyot R. Retrotransposons in Plant Genomes: Structure, Identification, and Classification through Bioinformatics and Machine Learning. Int. J. Mol. Sci. 2019;20:3837. DOI: 10.3390/ijms20153837

[35] Palmer SA, Clapham AJ, Rose P, Freitas FO, et al. Archaeogenomic Evidence of Punctuated Genome Evolution in *Gossypium*. Mol. Biol. Evol. 2012;29(8):2031-2038. DOI: 10.1093/molbev/mss070

[36] Wang M, Li J, Wang P, Liu F, et al. Comparative Genome Analyses Highlight Transposon-Mediated Genome Expansion and the Evolutionary Architecture of 3D Genomic Folding in Cotton. Molecular Biology and Evolution. DOI: 10.1093/molbev/msab128

[37] Lu H, Cui X, Liu Z, Liu Y, et al. Discovery and annotation of a novel transposable element family in *Gossypium*. BMC Plant Biology. 2018;18(1):307. DOI: 10.1186/s12870-018-1519-7

[38] Wang M, Tu L, Yuan D, Zhu D, et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat. Genet.* 2019;51:224-229. DOI: 10.1038/s41588-018-0282-x

[39] Page JT, Liechty ZS, Alexander RH, Clemons K, et al. DNA sequence evolution and rare homoeologous conversion in tetraploid cotton. PLoS Genet. 2016;12:e1006012. DOI: 10.1371/journal.pgen.1006012

[40] Zhang T, Hu Y, Jiang W, Fang L, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 2015;33:531-537. DOI: 10.1038/nbt.3207

[41] Yoo MJ, Wendel JF. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLoS Genet.* 2014;10:e1004073. DOI: 10.1371/journal.pgen.1004073

[42] Hu G, Koh J, Yoo MJ, Chen S, Wendel JF. Gene-Expression Novelty in Allopolyploid Cotton: A Proteomic Perspective. Genetics 2015;200:91-104. DOI: 10.1534/genetics.115.174367

[43] Hovav R, Faigenboim-Doron A, Kadmon N, Hu G, et al. A transcriptome profile for developing seed of polyploid cotton. The Plant Genome. 2015;8:eplantgenome2014.08.0041. DOI: 10.3835/plantgenome2014.08.0041

[44] Peng Z, Cheng H, Sun G, et al. Expression patterns and functional divergence of homologous genes accompanied by polyploidization in cotton (*Gossypium hirsutum* L.). Sci. China Life Sci. 2020;63:1565-1579. DOI: 10.1007/s11427-019-1618-7

[45] Fang L, Gong H, Hu Y, Liu C, et al. Genomic insights into divergence and dual domestication of cultivated

allotetraploid cottons. Genome Biology. 2017;18:33. DOI: 10.1186/s13059-017-1167-5

[46] Panchy N, Lehti-Shiu MD, Shiu SH. Evolution of gene duplication in plants. Plant Physiology. 2016;171:2294-2316. DOI: 10.1104/pp.16.00523

[47] Haigler CH, Betancur L, Stiff MR, Tuttle JR. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. Front Plant Sci. 2012;3:104. DOI: 10.3389/fpls.2012.00104

[48] Yang Z, Qanmber G, Wang Z, Yang Z, Li F. Gossypium Genomics: Trends, Scope, and Utilization for Cotton Improvement. Trends Plant Sci. 2020;25:488-500. DOI: 10.1016/j.tplants.2019.12.011

[49] Snodgrass SJ, Jareczek J, Wendel JF. An examination of nucleotypic effects in diploid and polyploid cotton. AoB Plants. 2017;9:plw082. DOI: 10.1093/aobpla/plw082

[50] Brockington SF, Alvarez-Fernandez R, Landis JB, Alcorn K, et al. Evolutionary Analysis of the MIXTA Gene Family Highlights Potential Targets for the Study of Cellular Differentiation. Molecular Biology and Evolution. 2013;30:526-540. DOI: 10.1093/molbev/mss260

[51] Wu H, Tian Y, Wan Q, Fang L, et al. Genetics and evolution of MIXTA genes regulating cotton lint fiber development. New Phytol. 2018;217:883-895. DOI: 10.1111/nph.14844

[52] Zheng D, Ye W, Song Q, Han F, Zhang T, Chen ZJ. Histone Modifications Define Expression Bias of Homoeologous Genomes in Allotetraploid Cotton. Plant Physiol. 2016;172:1760-1771. DOI: 10.1104/pp.16.01210

[53] Gallagher JP, Grover CE, Hu G, Jareczek JJ, Wendel JF. Conservation and Divergence in Duplicated Fiber Coexpression Networks Accompanying Domestication of the Polyploid *Gossypium hirsutum* L. G3: Genes, Genomes, Genetics. 2020;10:2879-2892. DOI: 10.1534/g3.120.401362

[54] Ashraf J, Zuo D, Wang Q, Malik W, et al. Recent insights into cotton functional genomics: progress and future perspectives. Plant Biotechnology Journal. 2018;6:699-713. DOI: 10.1111/pbi.12856

[55] Fang L, Guan X, Zhang T. Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.). The Crop Journal. 2017;5:159-165. DOI: 10.1016/j.cj.2016.07.001

[56] Xu Z, Yu JZ, Cho J, Yu J, Kohel RJ, Percy RG. Polyploidization Altered Gene Functions in Cotton (*Gossypium* spp.). PLoS ONE. 2010;5(12):e14351. DOI: 10.1371/journal.pone.0014351

[57] Zhu G, Li W, Wang G, Li L, Si Q, Cai C, Guo W. Genetic Basis of Fiber Improvement and Decreased Stress Tolerance in Cultivated Versus Semi-Domesticated Upland Cotton. Front. Plant Sci. 2019;10:1572. DOI: 10.3389/fpls.2019.01572

[58] Wendel JF, Cronn RC. Polyploidy and the evolutionary history of cotton. Adv Agron. 2003;78:139-186.

[59] Page JT, Huynh MD, Liechty ZS, Grupp K, et al. Insights into the Evolution of Cotton Diploids and Polyploids from Whole-Genome Re-sequencing. G3 (Bethesda). 2013;3(10):1809-18. DOI: 10.1534/g3.113.007229

[60] Buriev ZT, Saha S, Shermatov SE, Jenkins JN, et al. Molecular evolution of the clustered MIC-3 multigene family of

Gossypium species. Theor Appl Genet. 2011;123:1359-1373. DOI: 10.1007/s00122-011-1672-y

[61] Small RL, Wendel JF. Copy number lability and evolutionary dynamics of the Adh gene family in diploid and tetraploid cotton (*Gossypium*). Genetics. 2000;155:1913-1926.

[62] Small RL, Wendel JF. Differential evolutionary dynamics of duplicated paralogous Adh loci in allotetraploid cotton (*Gossypium*). Mol. Biol. Evol. 2002;19:597-607. DOI: 10.1093/oxfordjournals.molbev.a004119