

## History and Development of Genetics Research in India: Three case studies

DP Kasbekar\*

(Received 12 October 2014; revised 18 May 2016)

### Abstract

The paper discusses three major contributions in genetics research from India; Genetic and cytogenetic studies on sugarcane undertaken at the Sugarcane Breeding Institute in Coimbatore during 1912- 1938 by C. A. Barber, T. S. Venkatraman, and E. K. Janaki Ammal, the publication of D. D. Kosambi's mapping function (1944), and the discovery of the Bombay blood group in KEM Hospital, Mumbai by Y.M. Bhende and co-workers in 1952. Most current research on these topics is now done outside India, possibly because India never developed the critical mass of researchers to sustain the early leads of these pioneers.

**Key words:** Consanguinity, DNA profiling, Mapping function, Sugarcane interspecific and intergeneric hybrids, Sugarcane cytogenetics

### 1. INTRODUCTION

Professionally, I am a *Neurospora* geneticist, not a historian of Genetics. Therefore, an article on the history and development of genetics research in India by me carries less authoritative heft than (I imagine) do my articles in *Neurospora* genetics. Nevertheless, I took up the challenge to write this paper for three reasons; (1) I could bring in a different viewpoint than a professional historian, (2) it might goad others more competent than me to study this otherwise neglected topic, and (3) it gave me the opportunity to share the thrill of reading a 95 year old research paper published from Coimbatore (see Barber, 1920 in supplementary material 1).

This article touches upon three major early research contributions in genetics from India: (1) interspecific and intergeneric hybridization in sugarcane; (2) D. D. Kosambi's map function; and (3) the discovery of the Bombay blood group. Given that the locus of current research in all these areas has indisputably shifted out of India, the

question why India has not been able to sustain its early leads into the present time has more than mere passing interest.

### 2. SUGARCANE: EARLY GENETIC AND CYTOGENETIC STUDIES

Sugarcane (*Saccharum officinarum*) was domesticated about 3000 years ago in Papua New Guinea. Domesticated cane is derived from the wild species *Saccharum robustum* (basic chromosome number  $x = 10$ ,  $2n = 80$ ), with some admixture from the related grasses *Erianthus arundinaceus* and *Miscanthus sinensis*. Domestication involved the selection of sweeter clones for chewing, and these clones then introgressed "en passant" with related varieties and species and still sweeter hybrids were further selected and dispersed. The domesticated high-sugar containing *S. officinarum* varieties were designated as the "noble" canes by later Dutch colonizers. Another wild species, *Saccharum spontaneum* ( $x = 8$ ,  $2n = 40-128$ ), is widespread

\* Haldane Chair, Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad 500 001 Email: kas@cdfd.org.in

in South and South-East Asia, and in ancient times it too was independently domesticated in India and led to the development of technology to extract sugarcane juice and make jaggery and sugar. The Indian canes, however, had much lower sugar content than the “noble” canes of Indonesia and Papua New Guinea.

Although sugarcane varieties are vegetatively propagated, sugarcane can flower naturally in some regions. However, the same varieties might not flower or may do so only rarely in other regions. Sugarcane seed is called fuzz, and is used only in breeding programs. Flowering behavior is affected by many factors including day length, temperature, and moisture. In equatorial regions sugarcane can flower throughout the year, but the flowers are sterile and preclude breeding activities. Flowering period becomes restricted as one moves north or south from the equator, but the fertility of the flowers increases. Fertility levels are maximal around latitudes 11° N and S, and these latitudes are ideal for sugarcane breeding. Coimbatore is one such place, and has additional advantages for flowering and seed set with regard to rainfall, temperature, humidity and winds. In 1912, the Sugarcane Breeding Institute (SBI) was established in Coimbatore, with C. A. Barber as its first “Head of Station & Government Sugarcane Expert” and it undertook the breeding of improved sugarcane varieties. The SBI website ([www.sugarcane.res.in](http://www.sugarcane.res.in)) “knowledge bank” page is very informative on the early research on sugarcane.

Barber, assisted by T. S. Venkatraman, performed the first interspecies crosses using the noble *Saccharum officinarum* canes as females and the hardier, less sweet, Indian *S. spontaneum* as males. The hybrid progeny had the high sugar content of the former and the increased hardiness to abiotic and biotic stress of the latter. Crossing of *S. spontaneum* males to *S. officinarum* females is called as “nobilization” of *S. spontaneum*. Barber and Venkatraman’s interspecies crosses

worked brilliantly because 1n pollen of *S. spontaneum* fertilized 2n eggs of *S. officinarum* and produced n + 2n zygotes that retained the full chromosome complement of *S. officinarum*. However, the zygotes from the reciprocal cross, were n + n type, as they were in the self-crosses within the parental types. Backcross of the hybrid male to *S. officinarum* female again showed n + 2n transmission, although in further backcrosses the production of 2n eggs depended on the specific *S. officinarum* clone used. The n + 2n transmission allowed the *S. officinarum* 2n genome to be retained and allowed the elimination of unwanted *S. spontaneum* chromosomes and retention of only the desirable ones (the n-bearing segregants) (Bhat and Gill, 1985).

The mechanistic basis of 2n transmission through the *S. officinarum* egg is still not understood. Various hypotheses put forth to explain 2n transmission include: (1) formation of unreduced egg cells (although this was rendered less likely by the observation of segregation of maternal characters among the hybrid progeny), (2) chromosome doubling via endo-duplication after either the first or second meiotic division, (3) post-meiotic fusion of meiotic products, (4) post-meiotic endomitosis in the egg cell, and (5) differential survival of n + 2n versus n + n zygotes. It is now suspected that fertilization by *S. spontaneum* pollen might induce chromosome doubling in the egg cell to produce 2n gametes. Attempts to exploit the secondary gene-pool by nobilizing *Erianthus arundinaceus* by crossing it as the male to *S. officinarum* female were made, but now the *S. officinarum* female was found to transmit only via n eggs instead of 2n. Moreover the resulting intergeneric hybrids are sterile, which precluded further crosses. To overcome these bottlenecks, contemporary research in SBI is making crosses between *E. arundinaceus* and *S. spontaneum*, to explore the use the resulting intergeneric hybrids as males in the expectation that they might trigger 2n gametes formation in the *S. officinarum* used as the female.

The Coimbatore breeds (“Co-canes”) transformed worldwide sugarcane production. Co 205, the first interspecies commercial variety was released in 1918, and in the first year itself it increased sugarcane production in Punjab by 50%. By 1927, Co 205 was being planted in Cuba and Florida. In 1928 Co 312 was released and for the next 30 years it was the most popular variety in sub-tropical India, until even superior varieties were bred.

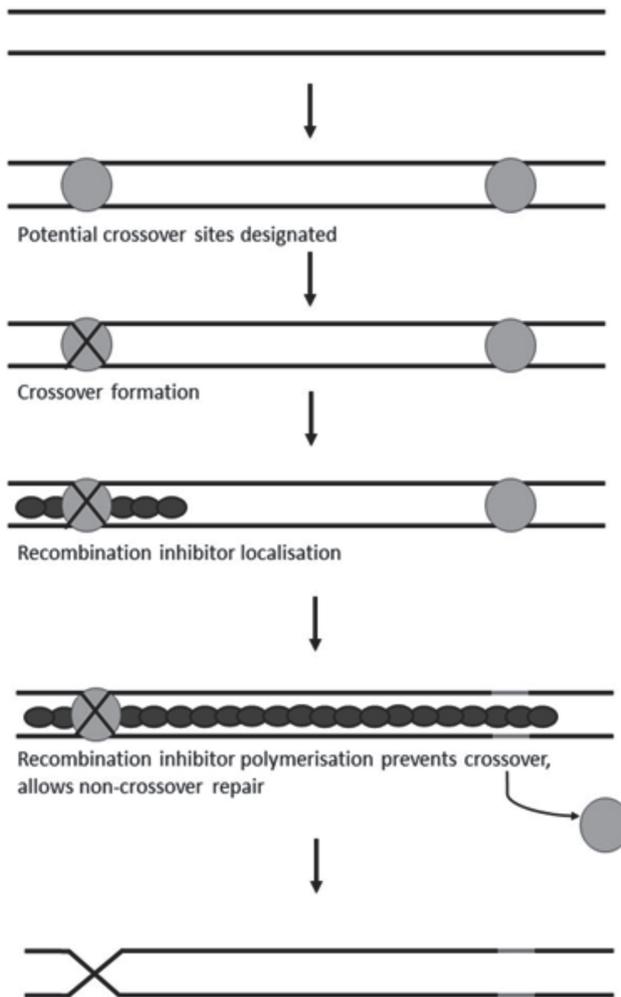
Indian scientists also did first-rate cytogenetics studies in the 1930s. To give just two examples, K. Ramiah and colleagues had identified the second instance of triploidy in rice (Ramiah *et al.*, 1933), only a year after the first instance was reported by Nakamori and colleagues from Japan. E. K. Janaki Ammal (1939) found that the species *S. spontaneum* is a polyploid series with  $2n = 48, 56, 64, 72,$  and  $80$  in India, and  $2n = 80, 96, 112$  in further India and South-East Asia (Burma through Indonesia). Based on this finding she inferred that the species originated in India and that its basic chromosome number is 8. Further, presence of odd multiples of 8 (56, 72) was expected to lead to univalents or multivalents in meiosis, but instead she found 28 and 36 bivalents. This led Janaki Ammal to suggest that these clones arose from hybridization between ancestral  $x = 10$  and  $x = 6$  forms (eg.,  $28 = 3 \times 6 + 10$ ;  $36 = 3 \times 10 + 6$ ).

### 3. KOSAMBI’S MAPPING FUNCTION.

The first paragraph from D. D. Kosambi’s 1944 paper (written when he taught at Fergusson College, Poona) succinctly defines the problem he wanted to address, and so I reproduce it here in its entirety. “Suppose three consecutive loci  $a, b, c$  of the same linkage group to have the recombination fractions (percentage divided by 100)  $(a, b) = y_1, (b, c) = y_2, (a, c) = y_{12}$ . Then it is known that for small values of  $y_1$  and  $y_2, y_{12} = y_1 + y_2$  approximately. For slightly larger values, we have a better approximation given by  $y_{12} = y_1 +$

$y_2 - y_1 y_2$ ; for still larger values, the approximation has again to be replaced by  $y_{12} = y_1 + y_2 - 2y_1 y_2$ . It is desired to obtain one single formula that will cover the entire range 0-1/2 of  $y$ -values in a reasonably satisfactory manner. This must also correspond to a single-valued, monotonically increasing, continuous function  $x$  of  $y$  in such a way that the corresponding identity becomes  $x_{12} = x_1 + x_2$ . The variable  $x$  will then be called the map distance corresponding to the given  $y$ .” Vinod (2011) has done a fine job describing Kosambi’s method to estimate map distances from recombination values for the calculus-challenged reader, and compared it with the methods of Morgan and Bridges (1916) and Haldane (1919).

A genetic map plots the relative positions of different genes on a chromosome. The distance between two genes is proportional to the fraction of progeny that show a crossover between them. However, if the distance is sufficiently large then the progeny derived from an even numbers of crossovers would be added to the non-crossover type, and map distance would be underestimated. On the other hand, in many eucaryotes, the occurrence of one crossover suppresses a second crossover from happening nearby (“nearby” can be of the order of mega-basepairs). This is called crossover interference. The mechanistic basis of crossover interference is still not understood, indeed, Hillers (2004) suggests that interference might be only an incidental consequence of some other aspect of meiosis. [Interestingly, Siddiqi and Putrament (1963), reported a brilliant study showing polarized negative interference in the *pab1* region of *Aspergillus nidulans*.] Crossover interference can be ignored if the genetic interval is sufficiently large. We can consider three levels of interference; complete, incomplete, and no interference at all. Complete interference does not allow double crossovers, and is essentially compatible with the way map distances were derived by Morgan and Bridges, Haldane’s mapping function was suitable when there was no



**Fig. 1.** A mechanism proposed for crossover interference. Formation of a crossover (X) leads to localisation of a recombination inhibitor (stippled ovals) at the crossover and its polymerisation along the vicinity, preventing other potential sites (right circle) nearby from forming crossovers, but allowing non-crossover repair. Figure adapted from Figure 4 of Youds J. L. and Boulton, S. J. (2011).

interference, and Kosambi's mapping function is best suited under conditions of incomplete interference (Vinod, 2011). Briefly, Haldane had recommended  $d = \frac{1}{2} \ln(1-2r)$ , whereas Kosambi suggested  $d = \frac{1}{4} \ln \left[ \frac{1+2r}{1-2r} \right]$ , where  $d$  is the map distance, and  $r$  is the recombination fraction.

Although in organisms with sequenced genomes the genome sequence has now supplanted genetic maps, in organisms whose genomes have not yet been sequenced genetic

maps remain important. Sugarcane is one such organism. Its 10 Gb genome is not easy to sequence with current technology. Al-Janabi *et al.* (1993), who developed a genetic linkage map for *S. spontaneum*, had used Kosambi's mapping function to convert recombination percentages into map units. They also cited Janaki Ammal's 1939 Current Science paper and another 1936 paper by her in the Indian Journal of Agricultural Sciences. That Kosambi and Janaki Ammal were cited 50 or more years post-publication, gives us a measure of the long shadows cast by the two giants. Indeed, I would go as far as to compare D. D. Kosambi with the great Cambridge mathematician G. H. Hardy, of Hardy-Weinberg Equilibrium fame. Both mathematicians made a single foray apiece into Genetics, and each managed to galvanize the subject in his own way. The cytogenetics of grasses by Janaki Ammal, brings to mind comparison with Barbara McClintock, the iconic maize geneticist.

#### 4. BOMBAY BLOOD GROUP

"Kahaani", the 2012 Bollywood thriller (starring the talented Vidya Balan in the lead role), probably introduced the Bombay blood group to more Indians than any article on the history of Indian genetics research can hope to match. The film's main antagonist had this rare blood group, and the protagonists track down the doctor who sourced the blood required for his transfusion. Bombay blood type was discovered in KEM Hospital, Mumbai by Bhende *et al.* (1952). Two patients (X was a railway accident victim, and Y had suffered a stab wound) were seen within a span of a few weeks. Both needed blood transfusion and both were apparently of blood group O. Surprisingly, their serum agglutinated the red blood cells of all the common ABO blood groups, thus showing that it contained anti-A, anti-B, and anti-H (i.e. anti-O) antibodies. Therefore the patients were transfused with plasma and not blood. The researchers then made a systematic

examination of blood from 160 donors and found a first time male donor, Z, as the third case of Bombay type blood. Only the red blood cells of these three men themselves was not agglutinated by their own sera, or by anti-A, anti-B, or anti-H sera. Serum of all three men contained anti-A, anti-B, and anti-H antibody. Absorption of their serum by group O red cells removed the anti-H, leaving the anti-A and anti-B unchanged, whereas absorption of the serum by red cells of Z left all the antibodies unchanged. The antibody that agglutinated O cells was inhibited by saliva from “secretors” regardless of their blood-group, but not by saliva of non-secretors (see below for an explanation). The degree of inhibition was comparable with the degree of inhibition of an eel anti-H serum caused by the same saliva. Thus the three men appeared to possess a new allelomorph of the ABO locus, and they could receive blood transfusion of only Bombay type blood. The parents of patient X were first cousins, although the parents of the other two patients were not related to each other or to patient X. No such blood had been noted in Europe or the US although millions of appropriate tests must have been made. Ludwik Hirszfild (1952), however, recalled seeing a patient during the Second World War (1941), in the kitchen of his provisional home in the Warsaw ghetto. The patient exhibited unagglutinable erythrocytes and strong anti-O agglutinins in her serum. Hirszfild published his findings after the war (1946) in the Polish Medical Journal, although by then Roza Amsel, who had assisted him these studies, was murdered in 1943 “during the extermination of all Jews in Warsaw” (Hirszfild 1952).

The ABO blood group was the first genetic polymorphism found in humans. We now know the ABO blood group antigens are sugars. The H locus on chromosome 19 encodes a fucosyltransferase that catalyses the last step in the production of H antigen on the red blood cells, which is the common precursor of all the ABO

blood types. The homologous Se (secretor) locus also is on chromosome 19, about 35 kb away from the H locus, and encodes another fucosyltransferase that produces a soluble form of the H antigen found in saliva and other body fluids. Secretors have the *Se/Se* or *Se/se* genotype, whereas non-secretors are *se/se* individuals. This explains why the anti-O antibody in the serum of the Bombay patients was inhibited by saliva from secretors but not of non-secretors. The H and Se loci are also referred to as FUT1 and FUT2 to underscore the fact they are paralogous loci, that arose from the duplication of a common ancestral gene. The ABO locus on chromosome 9 encodes a glycosyltransferase that can catalyse addition of another sugar moiety to the H antigen. The A allele encodes a glycosyltransferase that attaches N-acetylgalactosamine to the D-galactose end of the H antigen to produce A antigen, and the resulting blood group is called A. The B allele encodes a glycosyltransferase that attaches D-galactose to the D-galactose end of the H antigen to produce B antigen, and the resulting blood group is called B. If both enzymes are present, as in an AB individual, then both A and B antigens are produced and the blood group is called AB. If both enzymes are absent, the antigen remains as H and the blood group is called O. An individual of Bombay blood type is homozygous for the non-functional h allele of the H locus and fails to make H antigen. Therefore A, B, AB and O which are all different manifestations of H are all absent from the red cells. The Bombay Blood Group is termed as ‘Oh’, which means absence of ‘H’. Previous to this discovery, the detection of O was on the basis of absence of both A and B antigens. The authors hoped that their discovery would encourage transfusion services to test the serum of all blood donors against group O red cells.

Recent studies have shown the A and B blood groups to be a trans-species polymorphism that remained under balancing selection for tens of millions of years (Ségurel *et al.*, 2012).

### 5. CONSANGUINITY AND ENDOGAMY AMONG INDIANS

Discovery of the Bombay blood type reflects the high level of endogamy and consanguinity in the Indian sub-continent. Consanguinity is mating between two related individuals. The closest consanguineous matings are between first-degree relatives, such as parent-child and brother-sister. In most cultures such matings (called “incest”) are taboo. An example of father-daughter incest uncovered in the Centre for DNA Fingerprinting and Diagnostics is shown in Supplementary Material 2A. One can easily work out a “coefficient of inbreeding (F)” for a child born out of any consanguineous mating. The coefficient of inbreeding is the probability that an allele is homozygous because both the father and mother carry the same allele and pass it on to their child. F also identifies the fraction of the genome that is homozygous by descent. In a child from a first-degree (incestuous) mating  $F = 0.25$ ; in a child born out of an uncle-niece mating  $F = 0.125$ ; and in a child born from a first cousin mating  $F = 0.0625$ . Since the fraction of the genome homozygous due to consanguinity is F, the genome fraction that cannot be ascribed to consanguinity has to be  $(1 - F)$ . If a recessive disease-causing allele is present in the population with frequency  $q$ , then in an individual with the disease the probability that the allele is homozygous because of consanguinity is  $Fq$ , and the probability it is homozygous independent of consanguinity is  $(1 - F)q^2$ . Thus, for any affected individual, the probability that consanguinity is the underlying cause is  $\alpha = Fq/[Fq + (1 - F)q^2]$ . If F is large relative to  $q$ ,  $\alpha$  approaches 1. In other words, homozygosity of a very rare recessive genetic condition is almost always due to consanguinity.

Endogamy is the practice of marrying within a specific group. If the group is small, then endogamy increases the chance of consanguinity. Endogamy is common among migrants in new countries who attempt to make roots and resist

complete integration (Darr and Modell, 1988). Sanghvi (1966) reported the coefficient of inbreeding in rural Andhra Pradesh was as high as 0.048.

In DNA profiling services undertaken at the Centre for DNA Fingerprinting and Diagnostics we routinely see effects of endogamy and consanguinity in India. DNA Fingerprinting uses PCR to amplify different “short tandem repeat” (STR) loci on the different chromosomes. STRs are typically repeats of 3-5 base pairs and the different “alleles” of a locus differ in the number of repeats. Primers specific to single copy sequences flanking the STRs are used to PCR amplify the two (paternally- and maternally-derived) alleles of each locus in an individual, and allele size is determined via capillary electrophoresis. Since differences in repeat numbers at the STR loci apparently do not affect phenotype, and the number of repeat units can vary extensively because of replication slippage, it is surmised that unrelated individuals almost certainly have different numbers of repeat units in their alleles. In other words, the STRs are polymorphic, and each allele is shared typically by 5-20% of the population. DNA profiling examines multiple STR loci in the same experiment in a multiplex reaction that simultaneously tests multiple loci (typically 17), thus the allele pattern is quite specific for an individual. Only monozygotic twins are expected to share the same DNA profile. Testing more STR loci increases the discrimination of the test. Given that the loci are so polymorphic, an individual from a randomly mating population is expected to be homozygous at most in only a few loci. We found one individual who displayed homozygosity at nine of the 17 loci examined (see Supplementary Material 2B). Degraded DNA or very small amounts of DNA can cause allelic dropouts, leading one to mistakenly think a heterozygote is a homozygote. However, we are confident that the observed homozygosity is not due to allelic

dropout, and instead reflects the effect of endogamy in the population.

## 6. CONCLUSIONS

The buzz of current research in sugarcane genetics, linkage map construction, and blood group research has indisputably shifted out of India (eg. see Aitken *et al.*, 2014; Ségurel *et al.*, 2012), and leads us back to the question posed in the introduction: why has India not sustained its early leads in these areas into the present time? The answer in my opinion is quite prosaic – India just did not have the critical mass of researchers to take the baton from the pioneers. We hope the recent sprouting of many new universities and research institutes will enable scientific research in India to attain “criticality” in the near future and enable the best research to become self-sustaining.

## ACKNOWLEDGEMENTS

I thank S. P. R. Prasad (CDFD) for providing the DNA profiles presented as Supplementary material 2A and 2B, Parmit Singh (NIH) for help in obtaining several articles, Vidyand Nanjundiah (CHG) and Mandar V. Deshmukh (CCMB) for suggestions to improve the article, and D. Balasubramanian (LVPEI) for encouraging me to take up the challenge to write it in the first place. I am supported by the Haldane Chair in the Centre for DNA Fingerprinting and Diagnostics.

**Supplementary material 1:** Barber, C.A. 1920. The origin of the sugar cane. *The International Sugar Journal* 22: 249-251. [Click here to open](#)

**Supplementary material 2A:** DNA profiles establish that the source of exhibit B (baby of victim) is the daughter born from an incestuous mating of the source of exhibit C (alleged father) with his daughter, source of exhibit A (victim). Each “exhibit” is a blood sample from that individual that was subjected to DNA profiling

using standard protocols. At each locus, two alleles are amplified, one paternally-derived and the other maternally-derived. [Click here to open](#)

**Supplementary material 2B:** DNA profiles of the source of exhibit B (baby girl) shows her to have the same allele at nine of 15 loci examined. The homozygous alleles are highlighted in each profile. The girl’s mother was homozygous at three loci and the alleged father was homozygous at five. The alleged father, however, is not the real father, because the alleles of the baby marked in red could not have come from him. This was a case of disputed paternity. In the profiles shown in Supplementary material 2A, individuals A, B, C, and D were homozygous at, respectively, three, four, four, and four loci. [Click here to open](#)

## BIBLIOGRAPHY

- Aitken, K. S.; McNeil, M. D.; Berkman, P. J.; Hermann, S.; Kilian, A.; Bundock, P. C.; and Li, J. Comparative mapping in the Poaceae family reveals translocations in the complex polyploidy genome of sugarcane. *BMC Plant Biol.* 14(2014):190.
- Al-Janabi, S. M.; Honeycutt, R. J.; McClelland, M. and Sobral, B. W. S. A genetic linkage map of *Saccharum spontaneum* L. ‘SES 208’. *Genetics* 134(1993):1294-1260.
- Barber, C.A. The origin of the sugar cane. *The International Sugar Journal*, 22 (1920): 249-251.
- Bhat, S. R. and Gill, S. S. The implications of 2n egg gametes in nobilization and breeding of sugarcane. *Euphytica*, 34 (1985):377-384.
- Bhende, Y. M.; Deshpande, C. K.; Bhatia, H. M.; Ranger, R.; Race, R. R.; Morgan, W. T. J. and Watkins, W. T. A “new” blood-group character related to the ABO system. *Lancet* 1(May 3 issue) (1952): 903-904
- Darr, A., and Modell, B. The frequency of consanguineous marriage among British Pakistanis. *J. Med. Genet.* 25(1988): 186-190
- Haldane, J. B. S. The combination of linkage values, and the calculation of distance between linked factors. *J. Genet.* 8(1919):299-309.
- Hillers, K. J. Crossover interference. *Curr. Biol.* 14(2004): R1036-R1037.

- Hirszfeld, L A "new" blood-group character related to the ABO system. *Lancet* 1(Oct 25 issue). (1952):826-827
- Janaki Ammal E. K. Triplo-polyploidy in *Saccharum spontaneum*. *Current Science* 8(1939):74-77.
- Janaki Ammal E. K. Chromosome numbers in Sugarcane x Bamboo hybrids. *Nature* 141(1938): 925.
- Kosambi, D. D. The estimation of map distance from recombination values, *Ann. Eugenics* 12(1944):172-175.
- Morgan, T. H. and Bridges, C. B. *Sex linked inheritance in Drosophila*. Carnegie Institution of Washington, 1916 p.88.
- Ramiah, K.; Parthasarathi, N.; Ramanujam, S. A triploid plant in rice (*Oryza sativa*) *Current Science* 2(1933):170-171
- Sanghvi L. D. Inbreeding in India. *Eugenics Quarterly* 13(1966):291-301.
- Siddiqi, O. H. and Putrament, A. Polarized negative interference in the *pab1* region of *Aspergillus nidulans*. *Genet. Res.* 4(1963): 12-20.
- Ségurel, L.; Thompson, E. E.; Flutre, T.; Lovstad, J.; Venkat, A.; Margulis, S. W.; Moyse, J.; Ross, S.; Gamble, K.; Sella, G.; Ober, C. and Przeworski, M. The ABO blood group is a trans-species polymorphism in primates *Proc. Natl. Acad. Sci. USA* 109(2012): 18493-18498.
- Vinod, K. K. Kosambi and the genetic mapping function. *Resonance* 16(2011):540-550.
- Youds J. L. and Boulton, S. J. The choice in meiosis – defining the factors that influence crossover or non-crossover formation. *J. Cell Sci.* 124(2011):501-513.