

Microsatellites in the silkworm, *Bombyx mori*: Abundance, polymorphism, and strain characterization

K. Damodar Reddy, E.G. Abraham, and J. Nagaraju

Abstract: We have isolated and characterized microsatellites (simple sequence repeat (SSR) loci) from the silkworm genome. The screening of a partial genomic library by the conventional hybridization method led to the isolation of 28 microsatellites harbouring clones. The abundance of (CA)_n repeats in the silkworm genome was akin to those reported in the other organisms such as honey bee, pig, and human, but the (CT)_n repeat motif is less common compared to bumble bee and honey bee genomes. Detailed analysis of 13 diverse silkworm strains with a representative of 15 microsatellite loci revealed a number of alleles ranging from 3 to 17 with heterozygosity values of 0.66–0.90. Along with strain-specific microsatellite markers, diapause and non-diapause strain-specific alleles were also identified. The repeat length did not show any relationship with the degree of polymorphism in the present study. The co-dominant inheritance of microsatellite markers was demonstrated in F₁ offspring. A list of primer sequences that tag each locus is provided. The availability of microsatellite markers can be expected to enhance the power and resolution of genome analysis in silkworm.

Key words: microsatellites, simple sequence repeats, polymorphisms, silkworm strains, *Bombyx mori*.

Résumé : Les auteurs ont isolé et caractérisé des microsatellites/SSR chez le vers à soie. Le criblage d'une banque génomique partielle par hybridation conventionnelle a permis d'isoler 28 clones contenant un microsatellite. L'abondance des microsatellites de type (CA)_n dans le génome du vers à soie était semblable à ce qui a été rapporté précédemment chez d'autres espèces telles que l'abeille, le porc et l'humain tandis que les séquences (CT)_n étaient moins abondantes que chez le bourdon et l'abeille. Une analyse détaillée de 13 souches de vers à soie avec 15 microsatellites représentatifs a révélé entre 3 et 17 allèles et des valeurs d'hétérozygotie variant entre 0,66 et 0,90. En plus de marqueurs spécifiques de certaines souches, des allèles spécifiques aux souches à diapause et sans diapause ont été identifiés. Le nombre de répétitions du motif de base n'était pas lié au degré de polymorphisme dans cette étude. La codominance des microsatellites a été démontrée dans des progénitures F₁. Une liste des amorces spécifiques à chaque locus est fournie. La disponibilité de marqueurs microsatellites permettra vraisemblablement d'accroître la puissance et la résolution des analyses génomiques chez le vers à soie.

Mots clés : microsatellites, SSR, polymorphismes, souches du vers à soie, *Bombyx mori*.

[Traduit par la Rédaction]

Introduction

Simple sequence repeats (SSRs), known as microsatellites, are tandemly repeated DNA sequence motifs (usually 2–5 bp long) that occur at multiple sites of eukaryotic genomes. The repeats are abundant and highly polymorphic markers in human and other mammalian genomes, as well as in plant genomes (Tautz 1989; Weber and May 1989; Broun and Tanksley 1996). The important feature of this class of

repetitive DNA is hypervariability, mainly expressed as variation in the copy number of tandem repeats at a particular locus. Due to their hypervariability, relative ease of scoring by polymerase chain reaction (PCR), codominant nature, and high reproducibility, they are now considered to be among the most powerful genetic markers. It is typical to observe loci with more than 10 alleles and heterozygosities above 0.60 even in relatively small samples. As a result, microsatellites have become the markers of choice in a wide variety of applications: high resolution mapping of human (Dib et al. 1996), mouse (Dietrich et al. 1996), dog (Ostrander et al. 1993), chicken (Crooijmans et al. 1996), and many plant species (Akkaya et al. 1995), as well as studies of kinship (Morin et al. 1994), population genetic structure (Bruford and Wyne 1993), conservation genetics (Gotelli et al. 1994), and forensics (Jeffreys et al. 1991).

In insects, very few studies on microsatellites have been carried out. In *Drosophila*, the presence of (CT)_n, (GT)_n and (GTC)_n and their relative abundance has been studied (Pardue et al. 1987; Tautz and Renz 1984) and the length

Corresponding Editor: K. Kasha.

Received August 28, 1998. Accepted March 4, 1999.

K.D. Reddy, E.G. Abraham, and J. Nagaraju.¹ Seribiotech Research Laboratory, Central Silk Board, No. 8, West of Chord Road, Bangalore 560 086, India.

¹Author to whom all correspondence should be addressed at the Centre for DNA Fingerprinting and Diagnostics, CCMB Campus, Uppal Road, Hyderabad 500 007 India (e-mail: jnagaraju@mailcity.com).

polymorphism of a (CAG)_n microsatellite located in the *Notch* gene of *D. melanogaster* has been reported (Tautz 1989). In *Anagasta kuehniella* (formerly *Ephistia kuehniella*) the (GATA)_n tracts have been shown to be hypervariable (Traut 1987). In two hymenopteran species, dinucleotide microsatellite repeats (CT)_n and (GT)_n have been found to occur at every 15 kb and 34 kb in honey bee, and 40 kb and 500 kb in bumble bee genomes, respectively (Estoup et al. 1993). The intense linkage map of the malarial vector mosquito, *Anopheles gambiae* has been prepared using microsatellite markers (Zheng et al. 1996).

The combined strengths of a lepidopteran molecular model and an important economic insect make the silkworm, *Bombyx mori*, an ideal candidate to initiate intensive genome analysis using potential new experimental strategies now being developed in other systems. The silkworm harbours more than 400 hereditary mutants, many of which have been put on a classical linkage map (Doira 1992). More than 3000 silkworm strains representing a repertoire of genetic differences for various complex traits are being maintained in Europe and Asia (Nagaraju and Singh 1997). As of now, silkworm strains are described based on their morphological characteristics that are highly variable and environment dependent. These limitations call for use of reliable DNA-based techniques for varietal characterization. The polymorphic DNA markers as revealed by different DNA marker assays could be used to construct high-density mapping of silkworm genome and genetic fingerprinting of diverse silkworm varieties. Moving in this direction, a few studies have been carried out using RAPDs (random amplified polymorphic DNAs, Nagaraju and Nagaraju 1995; Promboon et al. 1995; Yasukochi 1998), RFLPs (restriction fragment length polymorphisms, Shi et al. 1995), Bkm probe (Nagaraju et al. 1995), inter-SSR (Reddy et al. 1999), and M₁₃ probe (Tretjk et al. 1992). In order to take advantage of the high informative microsatellite markers to augment the marker resources for the genetic analysis of silkworm, in the present study we isolated, characterised, and analysed 28 microsatellite loci.

Materials and methods

Silkworm strains

Six diapausing (HU₂₀₄, KA, NB₁, NB₁₈, and NB_{4D2}) and seven non-diapausing (*C. nichi*, *Moria*, *Nistari*, *Pure Mysore*, *Daizo*, *Gungnong* and *Sarupat*) *Bombyx mori* strains were used in the present study.

The non-diapausing strains are mostly of tropical origin and are characterized by rapid breeding, lower body weight, shorter silk fibre of inferior quality, and resistance to pathogens such as *Bombyx* nuclear polyhedrosis virus (BmNPV). They undergo 5–6 life cycles/year (polyvoltines). Among the seven non-diapausing strains used in the present study, *Pure Mysore* (Karnataka state, India), is unique in that it takes more than 28 days to complete its larval life as compared to the 20–22 days observed for almost all of the non-diapausing strains. It spins small, spindle shaped, and light greenish-yellow coloured cocoons. *Moria* and *Sarupat* strains (Assam, India), spin small, flossy, and creamish-coloured cocoons of spindle shape. *Daizo*, of Chinese origin, makes small, flossy, short, spindle-shaped, dark greenish-yellow cocoons and is characterized by unstable voltinism, i.e., it tends to lay both diapausing and non-diapausing eggs depending upon environmental factors (such as temperature), prevailing during silkworm rearing. *Nistari* (West Bengal, India) spins small, spindle-shaped, golden-yellow-coloured

cocoons. *C. nichi*, of Chinese origin is said to be of the diapausing type when it was brought to India from Japan almost 80 years ago, and has become more of a non-diapausing type with associated characters, as a result of continuous breeding under tropical conditions. It spins small, white and peanut-shaped cocoons.

The diapausing strains, conversely, are of temperate origin, undergo egg diapause, attain higher body weight, secrete longer silk fibre of superior quality, and show susceptibility to different pathogens, particularly to BmNPV and to high levels of heat, humidity, and inadequate sanitary conditions during silkworm rearing. They complete one (univoltine) or two (bivoltine) life cycles in a year. KA, NB₁, and NB₇ spin white-coloured and oval-shaped cocoons (Chinese type), and NB₁₈, NB_{4D2}, and HU₂₀₄ make white-coloured, peanut-shaped (Japanese type) cocoons. All the strains, except HU₂₀₄ which is derived from a Chinese hybrid, are bred either from single- or double-cross Japanese hybrids. NB₁₈ and NB_{4D2} are sister lines bred from a Japanese double-cross hybrid. The characteristics of the strains are as described by Nagaraja and Nagaraju (1995).

Construction of partial genomic library

Genomic DNA from an indigenous silkworm strain, *Nistari*, was extracted according to Nagaraja and Nagaraju (1995). Silkworm DNA was digested to completion simultaneously with *Hae*III, *Eco*RV, *Rsa*I, and *Sau*3A1 (Amersham) according to the manufacturer's instructions. The pBluescript SK(+) vector DNA was digested with *Bam*H1 and *Hind*II and ligated to insert DNA and transferred into competent *Escherichia coli* DH5α cells using standard procedures (Sambrook et al. 1989). A total of 3513 recombinant clones were picked up by blue–white colour selection from the library.

The DNA from the recombinant clones was transferred to nylon membranes (Hybond N, Amersham) by the dot-blot method. Membranes were soaked in 2× SSC for 5 min, then air-dried and used for hybridization.

Library screening for SSR sequences

Oligonucleotides (GT)₁₀ and (CT)₁₀ end-labeled with λ-[³²P]-dATP were used as probes. Membranes were pre-hybridized at 65°C for 3 h in pre-hybridization solution [0.05% sodium pyrophosphate, 0.5% SDS, 6× SSC, 5× Denhardt's solution and *E. coli* DNA (10 μg/mL)], and then hybridized in the same volume of hybridization solution containing the probe and sonicated, denatured *E. coli* DNA (10 μg/mL) for 24 h at 50°C. The blots were washed twice with 0.1% SDS, 6× SSC and 0.05% sodium pyrophosphate at room temperature and then with the same, fresh solution at 40°C for 30 min. The washed, hybridized, and dried blots were applied to autoradiographic film for 3 h to overnight at –80°C and positive clones were selected. The clones which gave signals by the time point of 3 h exposure were taken as high-intensity clones. The same filters were reused for hybridization after stripping the probe. Stripping was carried out by incubating the blots at 45°C for 30 min in 0.4 M NaOH, and 15 min in 0.1× SSC, 0.1% SDS and 0.2 M Tris–Cl, pH 7.5.

Sequencing of positive clones

Purified plasmid DNA from clones selected with the (GT)₁₀ and (CT)₁₀ was transferred into *E. coli* JM109 cells and plated on LB-plates and grown overnight at 37°C. Single colony overnight culture was used for double-strand DNA extraction. Single-strand DNA was prepared according to Sambrook et al. (1989) and sequenced using a Sequenase v. 2.0 DNA sequencing kit (USB) with T₇ primer. Double-strand DNA was sequenced using a AmpliCycle sequencing kit (Perkin–Elmer) and incorporation of α-³²P dCTP.

Table 1. Number of genomic hybridization signals detected by (CA)₁₀ and (GA)₁₀.

Nature of repeat motif or DNA probe	Microsatellites			Frequency in genomic library
	Positive signals	Average distance ^a	No. present ^b	
(CA) ₁₀ /(GT) ₁₀	57	49 kbp	10 816	2 × 10 ⁻² kbp
(GA) ₁₀ /(CT) ₁₀	27	104 kbp	5096	0.96 × 10 ⁻² kbp

^aAverage distance between two microsatellites.

^bNo. of microsatellites present in the *Bombyx mori* genome, obtained by dividing the size of the genome (i.e. 530 000 kb), by the average distance separating 2 loci.

Table 2. Average distance in kb between (GT)_n and between (CT)_n microsatellites in the silkworm, other insects, and vertebrates.

SSR	Silkworm	Honey bee	Bumble-bee	Salmon	Pig	Rat	Human
(GT) _n	49	34	500	12	47	15	30
(CT) _n	104	15	40	ND	ND	50	113

Note: ND, not determined.

Primer design, PCR, and electrophoresis

Primers (17–25 bp) for unique sequences flanking the microsatellite loci were designed, synthesised, and were purified. PCR was performed using a Perkin–Elmer 480 thermal cycler. Typical PCR reactions (20 µL) were performed in 10× PCR buffer (500 mM KCl, 100 mM Tris–Cl and 1.0% Triton X–100), 1 mM of dATP, dGTP, and dTTP each and 0.25 mM dCTP, 4 µM of each primer, 2 µCi of α³²P–dCTP (1000 Ci/mmol; 1Ci = 37 GBq), 1 U of Taq DNA polymerase (Perkin–Elmer, Cetus) and 20 ng of template DNA. Different concentrations of Mg Cl₂ (1–4 mM) were used depending on the locus amplified. PCR cycles for microsatellite loci included a 2 min at 94°C initial denaturation step, followed by 30 cycles of 94°C (30 s), annealing temperatures 42°C–55°C (30 s), and 72°C (45 s). The last elongation step was extended to 10 min at 72°C. Aliquots of amplified DNA from individual PCR reactions were mixed with formamide stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol in the ratio of 3:2 and denatured at 75°C for 2 min, chilled on ice, then loaded on to 6% acrylamide – 8 M urea sequencing gel. An M13 sequence ladder was used on each gel as a size marker. After electrophoresis, the gels were fixed for 2× 20 min with 10% glacial acetic acid. The gels were dried at room temperature, and applied to autoradiographic film for 4–12 h.

Statistical analysis

For band-sharing analysis, each strain was scored for presence (1) or absence (0) of the allelic products. Similarity index matrices were generated based on the number of shared alleles; similarity index (S) = $2n_{ab}/n_a + n_b$, where n_a and n_b represent the total number of bands present in lanes a and b, respectively, and n_{ab} is the number of bands which are shared by both lanes (Nei and Li 1979). For the phylogenetic study, distance (D) was calculated as $1 - S$ and the D values were used to construct a phylogenetic tree using the UPGMA (unweighted pair group method with arithmetic means) analysis program in PHYLIP v. 3.572 (Felsenstein 1993).

The heterozygosity values, which are generally used to describe the informativeness of a genetic marker, were calculated using the algorithm: Heterozygosity = $1 - \sum p_i^2$, where p_i^2 is the frequency of i^{th} allele in the 13 silkworm strains examined.

Results and discussion

Frequency of microsatellites in the silkworm genome

To estimate the frequency of microsatellite repeats in the silkworm genome, 3513 clones from a partial genomic li-

brary of *B. mori* were hybridized with the dinucleotide repeats (GT)₁₀ and (CT)₁₀. Based on moderate to strong signals, the GT probe hybridized to 1.6% and the CT probe to 0.77% of the clones (Table 1). For each microsatellite tested, the average distance between two loci was obtained by dividing the length of the silkworm DNA screened by the number of positive signals. For 3513 clones screened, the corresponding length of silkworm DNA was 3513×0.8 kb (average insert size), equaling 2810 kbp. For each microsatellite, this value was divided by the number of signals observed. Given an average 0.8 kb insert size in the clones based on insert size verification in the recombinant clones in the partial genomic library, the frequency of each type of repeat in the genome was calculated as $N/(0.8 \times T)$, where N = number of positive clones and T = total number of clones. The total fraction of the partial genome analysed (2810 kb) for the presence of microsatellite repeats account for 0.53% of the *Bombyx* genome of 530 Mb. The frequency of occurrence of dinucleotide repeats in the genome was estimated to be, on an average, one (GT)_n every 49 kb, and one (CT)_n microsatellite every 104 kb. The haploid genome size of silkworm is 530 Mb (Gage 1974), which would account for as many as 10 000 and 5000 (GT)_n and (CT)_n microsatellites, respectively. Although this is an estimate based on an analysis of only 0.53% equivalent of the total genome, it indicates that the silkworm genome contains abundant amounts of the two types of dinucleotide repeats. These results show that the dinucleotide repeat motifs occur in the silkworm genome at a frequency comparable with other animals (Table 2). For example, the (GT)_n motifs occur at about every 49 kb compared to at about every 34 kb, 47 kb, 15 kb, and 30 kb in honey bee, pig, rat, and human respectively (Estoup et al.1993; Hamada et al.1982; Wintero et al. 1992; Beckmann and Weber 1992). The (CT)_n motif is less common in the silkworm, occurring at about every 104 kb. This is in contrast to bumble bee, honey bee, and a few plant species in which the (CT)_n motif predominates (Estoup et al. 1993; Guilford et al. 1997), and is more akin to the human genome. The predominant occurrence of (GT)_n motifs in the animal genome is explained by the abundant presence of methylated C residues at CpG couplets which may favour the prevalence of TG repeats over GA repeats (Tautz et al.

Table 3. Microsatellite loci, repeat pattern, forward and reverse flanking primer sequences, number of alleles, allelic range, heterozygosity values and annealing temperature (T) in *Bombyx mori*.

Locus symbol	Repeat motif length	Primer sequence	No. alleles	Allele size range (bp)	heterozygosity	T (°C)	MgCl ₂ (mM)
Sat 158	(CA) ₉ (CA) ₄	5'-ctccagacaaccattagat-3' 5'-cagcatccatccttattat-3'	11*	190–246	0.76	49	1
Sat 211	(GT) ₁₀	5'-gatcgaactacgcaattacg-3' 5'-cagcatccatccttattat -3'	6*	6–146	0.89	49	2
Sat 256	(CA) ₅ (TA) ₇	5'-attgctcggacaggagtgt-3' 5'-tgtgaggacagggcgtta -3'	7*	111–194	0.82	53	3
Sat 346	(GA) ₉	5'-gaagacagagcgaagtga-3' 5'-atggattcctgctgtagat-3'	6*	139–189	0.83	53	3
Sat 892	(GT) ₁₀	5'-caataatgcttacgagttaa-3' 5'-tatcgtagttcctgactt -3'	3*	175–187	0.66	47	3
Sat951	(GA) ₂₃	5'-attgtaaccgattgagaga-3' 5'-atcgcacaataagttcact-3'	5*	107–122	0.68	48	1
Sat962	(GT) ₁₀	5'-cataataaataataataat-3' 5'-tgtagtggatagtcagtat -3'	6*	108–114	0.78	42	3
Sat1013	(GT) ₉	5'-aacagatgctgaggactggt-3' 5'-tgccattcacaatacaaat-3'	5*	135–162	0.80	50	1
Sat1411	(GT) ₈ (GT) ₅	5'- gaatgttctgctggtgg-3' 5'-taatgttttatactttattatag -3'	8*	109–162	0.68	45	3
Sat1423	(CA) ₁₁	5'-cttcgatcaccgcttctc-3' 5'-cgctacgaaataaccattatctgaca-3'	9*	130–176	0.82	55	2
Sat1893	(CA) ₁₀	5'-aatgcagaatcgttaatttt-3' 5'-ttgaccacagacaataagc -3'	7*	98–158	0.85	45	2
Sat2550	(GA) ₁₂	5'-ggtccttgaactcggat-3' 5'-cagagacctgcccgttgcgtt-3'	9*	121–168	0.87	53	2
Sat2604	(CT) ₉	5'-gctcgcattatgcaatcctc-3' 5'-cgtcattgccttcattcagttc-3'	8*	143–186	0.86	53	2
Sat2763	(GT) ₂₂ (GT) ₇	5'-acgcgctctacaaaataccatta-3' 5'-gatcaccggttctgctctcg-3'	17*	105–179	0.90	53	2
Sat3513	(GT) ₅ (TGCG) ₆ (TA) ₂	5'-cgcaattctgtattagataa-3' 5'-taaaggattattcttattcg-3'	6*	134–223	0.84	46	3
Sat101	(GT) ₉	5'-caaaagttgtgacaataat-3' 5'-agaatacagagaagaagtgc-3'	2**	ND	–	45	3
Sat405	(GT) ₁₄	5'-cgtgtgaactccaagaggtc-3' 5'-cacagtcgaaccccggtatg-3'	5**	ND	–	46	3
Sat706	(GA) ₁₂	5'-tcacatattattacaagac-3' 5'-gttgactgtttgtttta-3'	3**	ND	–	42	2
Sat814	(CA) ₇	5'-tattgaggtgcgatacttg-3' 5'-cgtcggactcttgaatgga-3'	2**	ND	–	47	2
Sat1523	(CA) ₁₀	5'-tcccacatttatagagaat-3' 5'-ttgaccgagactgataaac-3'	2**	ND	–	45	4
Sat1697	(GT) ₈	5'-atacttatgaacctgatagg-3' 5'-ttacaaactaccatttat-3'	2**	ND	–	40	4
Sat2442	(GT) ₇ (CA) ₆	5'-tgtgtccctttttacg-3' 5'-ctgataacataacaaggtgt-3'	5**	ND	–	45	1
Sat2856	(CT) ₁₇ (CT) ₈	5'-tttgggtgcaataaaacat-3' 5'-agcgacacttctgctatcgt-3'	5**	ND	–	49	4
Sat3215	(GT) ₅ (GT) ₄	5'-agaaccagtcgaagtaac-3' 5'-actaatcaactcgtgacagt-3'	2**	ND	–	49	2
Sat3426	(GT) ₈	5'-ttctgcttgattcgtg-3' 5'-cgtgaggtctcggtagcc-3'	2**	ND	–	53	4
Sat3430	(GT) ₅ (GT) ₂	5'-catcaggtggacaagaatca-3' 5'-atgtatcgcaactcatca-3'	4**	ND	–	47	3
Sat2746	(GT) ₁₃	5'-ctcattgaccacagattga-3' 5'-ttaactaaatgataaaacc-3'	NA	–	–	42	2
Sat2991	(GA) ₆ (GA) ₉	5'-ggtcacgatgcagtaca-3' 5'-gcctggagccataactactt-3'	NA	–	–	48	2

Note: (*), based on 13 silkworm strains; (**), based on Nistari and NB₁ strains; NA, not amplified; ND, not determined.

1986; Wright 1994). Since methylated C can undergo deamination, in a CG repeat this would probably result in a transition of C to T and thereby increase the abundance of TG couplets.

Microsatellite sequences

The sequences of 28 (GT)_n and (CT)_n containing clones were determined. Following the classification of dinucleotide repeats by Weber (1990), eighteen microsatellites were perfect (no interruption in the run of repeats), eight were imperfect (interruption in the run of repeats), and two were compound (a mixture of repeats of two or more motifs) repeats. Proportions of each category are summarized in Table 3. We found perfect, imperfect and compound microsatellites in the 28 loci to be present at 62%, 29%, and 7% respectively (Table 4). This is in close agreement with 72% perfect, 28% imperfect, and 11% compound repeats in human (Weber 1990) but differs from those of sheep (Buchanan et al. 1993). It is unclear whether these differences could be due to biased estimation introduced by the method of isolation of microsatellite loci or could reflect the actual difference between the silkworm and mammalian genomes.

Among 21 (GT)_n repeats analysed, 13 clones contained perfect (GT)_n repeats ranging in repeat number from 8 to 14, with a mean of 10 repeats per clone. Six clones contained imperfect (GT)_n repeats. Two clones contained compound repeats, one with (CA)₅(TA)₇ repeats and the other with (GT)₅(GCGT)₆(GT)₄(AT)₂. Out of seven (CT)_n repeats, five were perfect repeats ranging in number from 9 to 23 with an average of 13 repeats and two were imperfect (CT)_n repeats (Table 3).

We developed primer pairs for the flanking sequences of the 28 microsatellite loci to amplify the repeats. A summary of primers and PCR amplification conditions is presented in Table 3. In order to determine the levels of polymorphism of microsatellite loci, genomic DNA from 13 diverse silkworm strains was used as templates for PCR amplification, using 17 pairs of primers specific to either (GT)_n or (CT)_n repeats. Each DNA sample was pooled from 10 moths of each of the strains. Out of 17, 15 loci revealed discrete amplified products in different silkworm strains. Polymorphisms were detected for all the fifteen loci analysed in the 13 silkworm strains. An example of typical SSR profile in silkworm strains for locus *sat211* is shown in Fig. 1a.

Allele frequencies at microsatellite loci

For the 15 microsatellite loci, 11 (GT)_n and 4 (CT)_n; a total of 113 alleles were detected in 13 diverse silkworm strains. The number of alleles scored at each locus in 13 strains varied from as few as 3 (*sat892*) to as many as 17 (*sat2763*, Table 3). The analysis of 11 (GT)_n loci revealed a total of 85 alleles, ranging from 3 to 17, with an average of 7.7 alleles per locus. The four (CT)_n loci revealed a total of 28 alleles ranging from 5–9, with an average of 7 alleles per locus. A representative locus for each class of perfect (*sat211*), imperfect (*sat2763*) and compound (*sat3513*) loci is described below:

sat211

Based on the analysis of the size of amplified bands, we detected a total of 6 different alleles. The longest product

Table 4. Percentage of different categories of (GT)_n and (CT)_n microsatellites in silkworm, honey bee (Estoup et al. 1993) and human (Weber 1990). Number in parentheses is the number of (GT)_n or (CT)_n microsatellites analysed.

Type of motif	silkworm		honey bee		human
	GT(21)	CT(7)	GT(23)	CT(52)	GT(57)
perfect	62	71	48	46	61
imperfect	29	29	22	31	27
compound	9	0	30	23	2

was 146 bp and the shortest one was 96 bp. This locus revealed alleles specific to diapausing and non-diapausing strains. Ninety-six-bp and One hundred thirty-four-bp alleles were specific to diapausing strains and 99-bp and 146-bp alleles were specific to non-diapausing strains (Fig. 1a). However, the allelic status needs to be further clarified through segregation analysis of F₂ or backcross population. Such diapause and non-diapause strain-specific markers are useful in silkworm genetic analysis and breeding programs since these two groups of silkworm strains that belong to temperate and tropical zones differ in a large number of qualitative and quantitative traits. The possible linkage of such specific markers with the traits of interest could be established using appropriate populations for focused and rapid practical breeding programs. Besides, in the mapping programme, the mapping population, such as F₂, is derived from a cross of two divergent parents that belong to the diapausing and non-diapausing groups, respectively. The high degree of polymorphism between two such parental strains provides greater marker resources for mapping. In fact, the ongoing mapping programme uses the F₂ population derived from a cross of a diapausing strain, C₁₀₈ and P₅₀ (a variant of Daizo) (Shi et al. 1995; Promboon et al. 1995). The latter has unstable voltinism depending on the environmental conditions (see Materials and methods) but has all the qualitative and quantitative characters of non-diapausing strains.

sat3513

Analysis of allelic length variations of this locus revealed 6 alleles ranging between 134 and 223 bp in Moria, Nistari, Pure Mysore, Daizo, and Sarupat strains. Such a large gradient of variability cannot be explained by variation in the length of the microsatellite repeats, and perhaps occurs due to small insertions and deletions preferentially extended to one end of the repeat (Jeffreys et al. 1991; Kostilla et al. 1995). Further cloning and analysis of the alleles of different sizes are needed to verify such observations (Fig. 1b).

sat2763

This microsatellite was found to be highly polymorphic, showing as many as 17 alleles, ranging from 105–179 bp on a panel of 13 strains (Fig. 1c). The alleles are distributed between three classes; (i) 105–126 bp; (ii) 138–150 bp; and (iii) 158–179 bp. The clear allelic separation was seen only in the first two classes and in the last class, most of the alleles were associated with the stuttering bands.

Fig. 1. (a) Polymorphisms detected by microsatellite loci in 13 silkworm strains; locus *sat211*. Arrows indicate diapause and non-diapause strain-specific alleles.

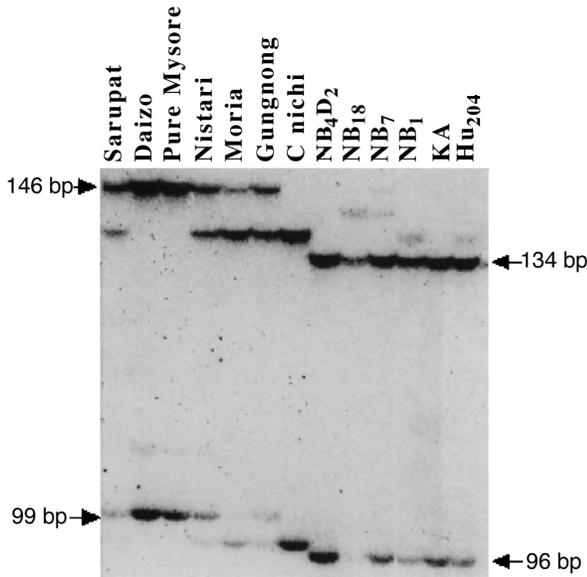


Fig. 1. (c) Polymorphisms detected by microsatellite loci in 13 silkworm strains; locus *sat2763*. Arrows indicate diapause and non-diapause strain-specific alleles.

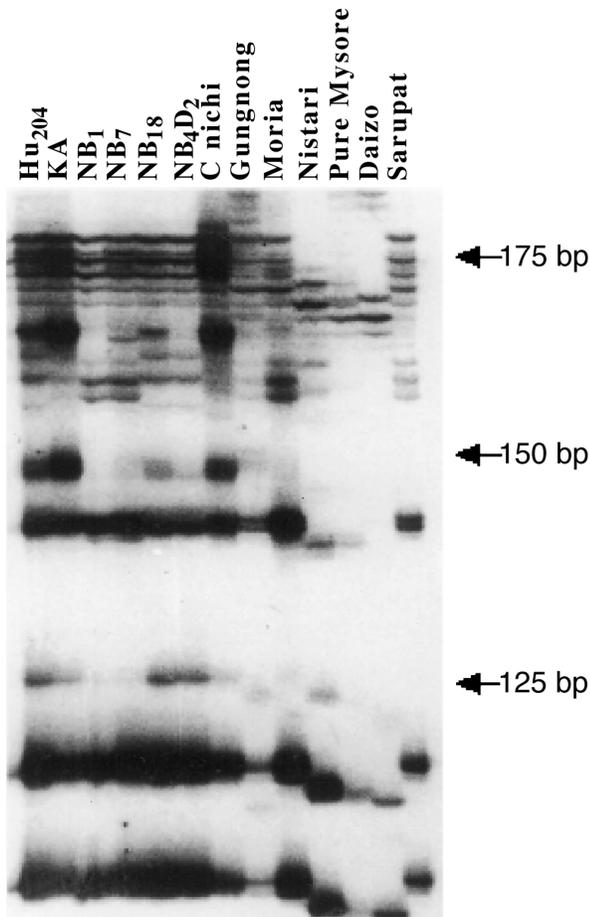
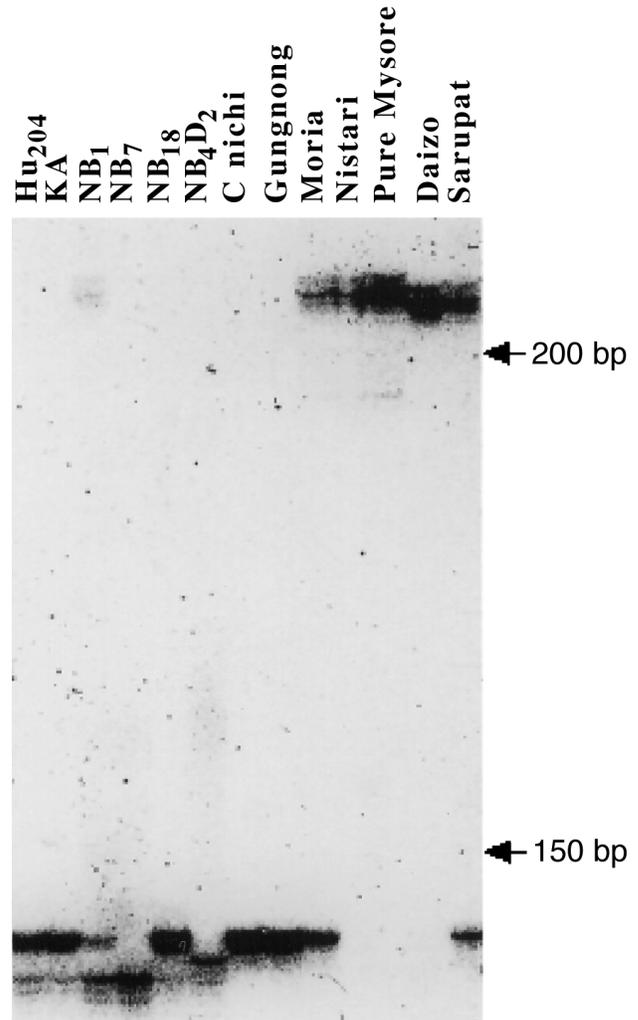


Fig. 1. (b) Polymorphisms detected by microsatellite loci in 13 silkworm strains; locus *sat3513*. Arrows indicate diapause and non-diapause strain-specific alleles.



Heterozygosity values

To assess the level of polymorphism, 13 silkworm strains were analysed with the 15 primer pairs. The average heterozygosity value for the 15 microsatellite loci was 0.79, ranging from 0.66–0.90. The most polymorphic locus, *sat2763*, showed the highest heterozygosity value of 0.90, and the locus *sat892* showed the lowest heterozygosity value of 0.66 (Table 3). Average heterozygosity values were 0.80 for perfect, 0.78 for imperfect, and 0.76 for compound repeats. The silkworm varieties chosen for the present study reveal marked differences for such diverse characters as geographic origin, voltinism, silk fibre length, body weight, larval span, biochemical traits, fecundity, etc., hence, it is likely that we have examined a reasonable representation of the alleles in silkworms and that the heterozygosity values given are realistic estimates of these markers. However, further analysis on segregation of each of the microsatellite loci would clear up the questions of artefacts or multiple alleles or multiple loci, enabling the use of microsatellites for genetic analysis of silkworm.

The relationship between the repeat length and degree of polymorphism was also examined. Although the number of loci available for each class for such an observation was low, the trend showed that there is no clear relationship between total repeat length and the degree of polymorphism in the present study. For example locus *sat951* (GA)₂₃, the largest perfect repeat in this study, has a heterozygosity value of 0.68 with 5 alleles, and locus *sat2604*, with the smallest (GA)₉ perfect repeat, has a heterozygosity of 0.86 with 8 alleles. The loci that are similar in repeat length also differ in their number of alleles and degree of polymorphism. The loci *sat211*, *sat892*, *sat962*, and *sat1893*, which carry similar repeat lengths of (GT)₁₀ showed 6, 3, 6, and 7 alleles, respectively (Table 3). These results contrast to the earlier studies in human and other species where longer repeats have been reported to generate higher degree of polymorphism (Weber 1990; Wu and Tanksley 1993; Fisher et al. 1998), and are consistent with the results obtained for teleost fishes and marine mammals (Slettan et al. 1993; Wright 1994).

Allelic variations and their codominant inheritance

The allelic variation within strain was examined using the locus *sat211* on Nistari and NB₁ individuals. The amplification products originating from individuals of the same strain revealed clear similarity patterns and strain-specific products. The NB₁ strain produced the diagnostic amplification products of 134 and 96 bp while Nistari revealed specific products of 99 bp and 138 bp in four individuals, 146 and 102 bp in two individuals and 146 and 138 bp in one individual (Fig. 2a), which also has alleles of 102 and 98 bp, that, considering their segregation in F₁ offspring (Fig. 2b), are probably originating from the duplicated locus and are not amplification artefacts.

The locus *sat211* was also tested for the co-dominant inheritance of alleles in the F₁ hybrid offspring of the parental strains, Nistari and NB₁. Products of 146- and 138-bp alleles of Nistari, segregated in the F₁ offspring (Fig. 2b) indicated the heterozygous nature of both the alleles while a homozygous NB₁ specific allele was present in all the offspring, as expected.

Genetic distances analysis

The dendrogram generated by UPGMA analysis based on 15 microsatellite loci resolved the 13 silkworm strains into two groups; one comprising the diapausing strains and the other the non-diapausing strains (Fig. 3). Again, the non-diapausing strains are clearly resolved into two subgroups, one comprising Gungnong, Moria, and Sarupat, and the other Nistari, Pure Mysore, and Daizo strains. It is clear from the dendrogram that the strains *C. nichi* and Daizo are the most divergent. The closest nearest neighbor strains were found to be KA and NB₁, NB₁₈ and NB_{4D2}, Pure Mysore and Nistari, and Moria and Sarupat. These results are in close agreement with the previously reported RAPD (Nagaraja and Nagaraju 1995), a Bkm-2(8) probe (containing 66 copies of GATA repeats interspersed with TA repeats, Nagaraju et al. 1995) and ISSR (Reddy et al. 1999) based dendrograms. The microsatellite-based analysis reflected the geographical, pedigree, and morphological relations of the strains used in the study. For example, Moria and Sarupat,

Fig. 2. (a) Allelic variations and their codominant inheritance detected by microsatellite locus *sat211*. Individuals of Nistari and NB₁ strains. Arrows indicate NB₁ and Nistari strain-specific alleles. The squares and circles represent males and females respectively.

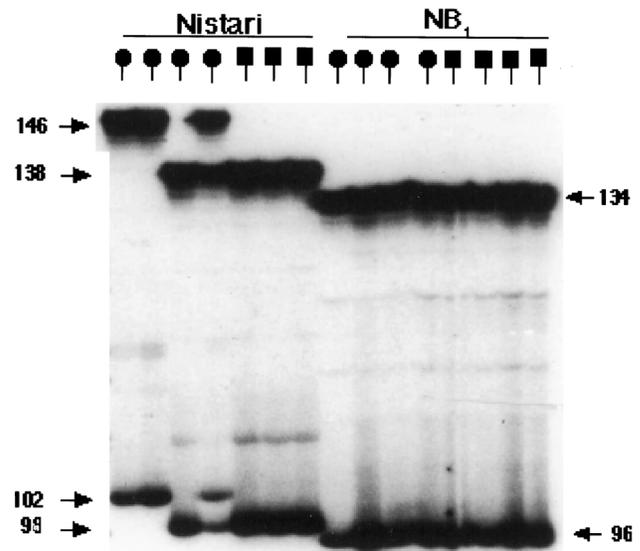
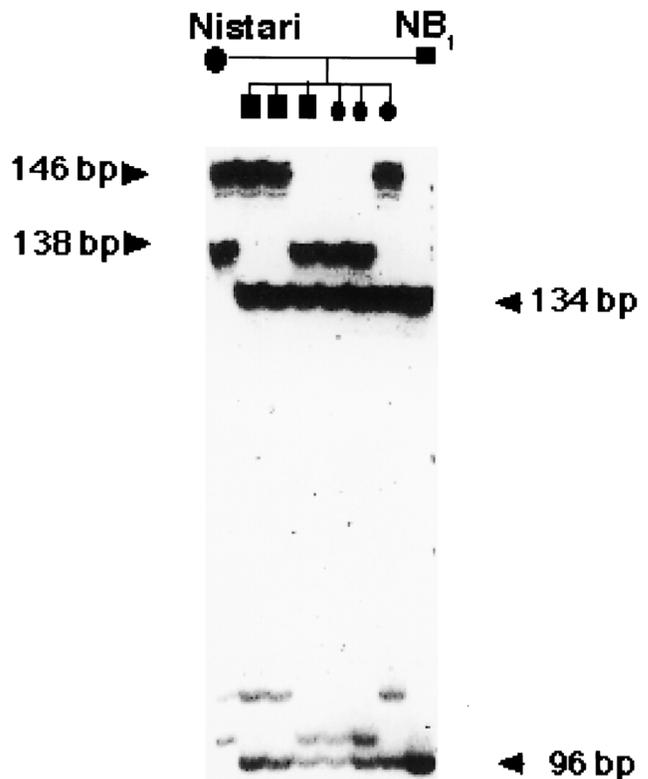
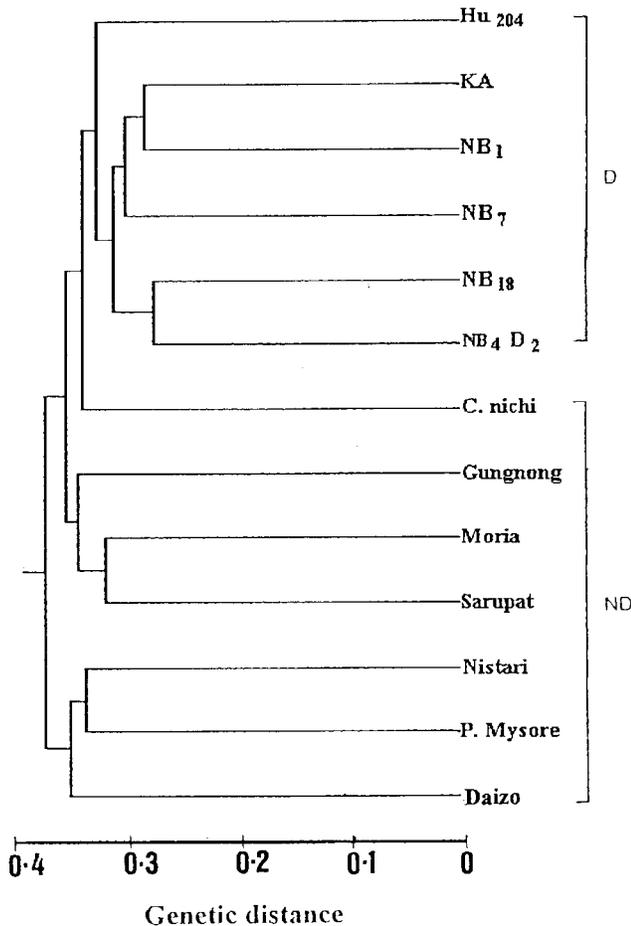


Fig. 2. (b) Allelic variations and their codominant inheritance detected by microsatellite locus *sat211*. Codominant inheritance of microsatellite alleles in the offspring of Nistari and NB₁ silkworm strains. Arrows, see above.



which share common geographical habitat, i.e., the Assam state of India, the two sister inbred lines, NB_{4D2} and NB₁₈, which are derived from the same Japanese double-cross hy-

Fig. 3. UPGMA-derived dendrogram illustrating relationships among silkworm strains determined by microsatellite length polymorphisms. D, diapausing strains. ND, non-diapausing strains.



brid, Nistari and Pure Mysore, which share either a common ancestor or the origin of Pure Mysore from Nistari, and NB₁, NB₇, and KA, which spin Chinese type cocoons are grouped in the different, respective sub-groups. All these sub groupings are in close agreement with the groupings obtained using RAPDs, Bkm multilocus probe, and ISSRs.

The present study is a step towards harnessing highly informative microsatellite loci that provide a ubiquitous marker system for characterization of silkworm genome. An additional use of microsatellite markers in silkworm is their possible application as markers for the related silk-secreting wild silk moths, which belong to the family Saturniidae. We have evidence to show that a majority of the *Bombyx* microsatellite primers effectively amplify the DNA of the wild silk moth species (Nagaraju and Reddy, in preparation). Such a study would facilitate the genetic analysis of the wild silk moths, some of which are endangered and the genetics for which is little known.

Acknowledgements

This work was supported, in part, by a grant (no. 1859/95) from the Department of Biotechnology, Govt. of India to J.

Nagaraju. The financial support of the Department of Biotechnology and Central Silk Board is greatly appreciated.

References

- Akkaya, M.S., Shoemaker, R.C., Specht, J.E., Bhagwat, A.A., and Cregan, P.B. 1995. Integration of simple sequence repeat DNA markers into a soybean linkage map. *Crop Sci.* **35**: 1439–1445.
- Broun, P., and Tanksley, S.D., 1996. Characterization and genetic mapping of simple sequences in the tomato genome. *Mol. Genet.* **250**: 29–49.
- Buchanan, F.C., Littlejohn, R.P., Galloway, S.M., and Crawford, A.M. 1993. Microsatellites and associated repetitive elements in the sheep genome. *Mamm. Genome*, **4**: 258–264.
- Bruford, M.W., and Wyne, R.K. 1993. Microsatellites and their application to population genetic studies. *Curr. Opin. Genet. Dev.* **3**: 939–943.
- Crooijmans, R.P.M.A., Vanoers, P.A.M., Strijk, J.A., Vanderpoel, J.J., and Groenen, M.A.M. 1996. Preliminary linkage map of the chicken (*Gallus domesticus*) genome based on microsatellite markers—77 new markers mapped. *Poult. Sci.* **75**: 746–754.
- Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J., and Weissbach, J. 1996. A comprehensive genetic map of the human genome based on 5264 microsatellites. *Nature*, **380**: 152–154.
- Dietrich, W.F., Miller, J., Steen, R., Merchant, M.A., Damronboles, D., Husain, Z., Dredge, R., Daly, M.J., Ingalls, K.A., O'Connor, T.J., Evans, C.A., DeAngelis, M.M., Levinson, D.M., Kruglyak, L., Goodman, N., Copeland, N.G., Jenkins, N.A., Hawkins, T.L., Stein, L., Page, D.C., and Lander, E.S. 1996. A comprehensive genetic map of the mouse genome. *Nature*, **380**: 149–152.
- Doira, H. 1992. Genetical stocks and mutations of *Bombyx mori*: Important genetic resources. Institute of Genetic Resources, Kyushu University, Fukoka, Japan.
- Estoup, A., Solignac, M., Harry, M., and Cornuet, J.M., 1993. Characterization of (GT)_n and (CT)_n microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. *Nucleic Acids Res.* **21**: 1427–1431.
- Felsenstein, J. 1993. PHYLIP 3.5, available from joe@genetics.Washington.edu. University of Washington.
- Fisher, P.J., Richardson, T.E., and Gardner, R.C. 1998. Characteristics of single- and multi-copy microsatellites from *Pinus radiata*. *Theor. Appl. Genet.* **96**: 969–979.
- Gage, L.P. 1974. Polyploidization of the silk gland of *Bombyx mori*. *J. Mol. Biol.* **86**: 97–108.
- Gotelli, D., Sillero-Zubiri, C., Applebaum, G.D., Roy, M.S., Girman, D.J., Garcia-Moreno, J., Ostrander, E.A., and Wayne, R.K. 1994. Molecular genetics of the most endangered canid: The Ethiopian wolf *Canis simensis*. *Mol. Ecol.* **3**: 301–312.
- Guilford, P., Prakash, S., Zhu, J.M., Rikkerink, E., Gardiner, S., Bassett, H., and Forster, R. 1997. Microsatellites in *Malus × domestica* (apple): Abundance, polymorphism and cultivar identification. *Theor. Appl. Genet.* **94**: 249–254.
- Hamada, H., Petrino, M.G., and Kakunaga, T. 1982. A novel repeated element with Z-DNA forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. U.S.A.* **79**: 6465–6469.
- Jeffreys, A.J., Allen, A.J., Hagelberg, E., and Sonnberg, A. 1991. Identification of the skeletal remains of Josef Mengele by DNA analysis. *Nature*, **352**: 427–429.
- Kostilla, S., Sirkaa-Liisa, V., Vakkari, P., and Pulkkinen, P. 1995. Microsatellite sequences in a conifer, *Pinus sylvestris*. *Genome*, **38**: 1244–1248.

- Morin, P.A., Moore, J.J., Chakraborty, R., Jin, L., Goodall, J., and Woodruff, D.S. 1994a. Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science*, **265**: 1193–1201.
- Nagaraja, G.M., and Nagaraju, J. 1995. Genome fingerprinting in silkworm, *Bombyx mori* using random arbitrary primers. *Electrophoresis*, **16**: 1633–1638.
- Nagaraju, J., and Singh, L. 1997. Assessment of genetic diversity by DNA profiling and its significance in silkworm *Bombyx mori*. *Electrophoresis*, **18**: 1676–1681.
- Nagaraju, J., Sharma, A., Sethuraman, B.N., Rao, G.V., and Singh, L. 1995. DNA fingerprinting in silkworm *Bombyx mori* using banded krait minor satellite DNA-derived probe. *Electrophoresis*, **16**: 1639–1642.
- Nei, M., and Li, W. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **76**: 5269–5273.
- Ostrander, E.A., Sprague, J.F., and Rine, J., Jr. 1993. Identification and characterization of dinucleotide repeat (CA)_n markers for genetic mapping in dog. *Genomics*, **16**: 207–213.
- Pardue, M.L., Lowenhaupt, K., Rich, A., and Nordheim, A. 1987. (dC-dA)_n-(dG-dT)_n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. *EMBO J.* **6**: 1781–1789.
- Promboon, A., Shimada, T., Fuziwara, H., and Kobayashi, M. 1995. Linkage map of random amplified polymorphic DNA (RAPDs) in the silkworm, *Bombyx mori*. *Genet. Res. Cambridge*, **66**: 1–7.
- Reddy, K.D., Nagaraju, J., and Abraham, E.G. 1999. Genetic characterization of silkworm *Bombyx mori* by Simple Sequence Repeat (SSR)-anchored PCR. *Heredity*, **83**: In press.
- Sambrook, J., Fritsch, E.F., and Maniatis, T.C. 1989. *Molecular cloning: A laboratory manual* (2nd ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Shi, J., Heckel, D.G., and Goldsmith, M.R. 1995. A genetic linkage map for the domesticated silkworm, *Bombyx mori*, based on restriction fragment length polymorphisms. *Genet. Res. Cambridge*, **66**: 109–126.
- Slettan, A., Olsaker, I., and Lic, O. 1993. Isolation and characterization of (GT)_n repetitive sequences from Atlantic salmon, *Salmo salar* L. *Animal Genet.* **24**: 195–197.
- Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* **17**: 6463–6471.
- Tautz, D., and Renz, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acid Res.* **12**: 4127–4138.
- Tautz, T., Trick, M., and Dover, G.A. 1986. Cryptic simplicity in DNA is a major source of DNA variation. *Nature (London)*, **332**: 652–658.
- Traut, W. 1987. Hypervariable Bkm DNA loci in moth *Ephestia kuehniella*: Does transposition cause restriction fragment length polymorphism? *Genetics*, **115**: 493–498.
- Tretjk, A.P., Ryskov, A.P., Sevastyanova, G.A., Fillipovich, Y.B., and Strunnikov, V.A. 1992. DNA fingerprints of *Bombyx mori* L. Testing of genotypic variability of parthenogenetic strains. *FEBS Lett.* **303**: 258–260.
- Weber, J.J. 1990. Informativeness of human (dC-dA)_n-(dG-dT)_n polymorphisms. *Genomics*, **7**: 524–530.
- Weber, J.L., and May, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**: 388–396.
- Wintero, A.K., Fredholm, M., and Thomsen, P.B. 1992. Variable (dG-dT)_n-(dC-dA)_n sequences in the porcine genome. *Genomics*, **12**: 281–288.
- Wright, J.M. 1994. Mutation at VNTRs: Are minisatellites the evolutionary progeny of microsatellites? *Genome*, **37**: 345–347.
- Wu, K.S., and Tanksley, S.D. 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. *Mol. Gen. Genet.* **241**: 225–235.
- Yasukochi, Y., 1998. A dense genetic map of the silkworm, *Bombyx mori*, covering all chromosomes based on 1018 molecular markers. *Genetics*, **150**: 1513–1525.
- Zheng, L., Benedict, M.Q., Cornel, A.J., Collins, H.F., and Kafatos, F.C. 1996. An integrated genetic map of the human malaria vector mosquito *Anopheles gambiae*. *Genetics*, **143**: 941–952.