ELSEVIER

Gene 228 (1999) 33–42

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Characterization of a species-specific repetitive DNA from a highly endangered wild animal, *Rhinoceros unicornis*, and assessment of genetic polymorphism by microsatellite associated sequence amplification (MASA)

Sher Ali [a,*], Md. Asim Azfer [a,b], Anu Bashamboo [a], Pradeep Kumar Mathur [c], Pradeep Kumar Malik [c], Vinod Behari Mathur [c], Atanu Kumar Raha [d], Shehnaz Ansari [a]

[a] *Molecular Genetics Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110 067, India*
[b] *Department of Zoology, Aligarh Muslim University, Aligarh 202 002, India*
[c] *Wildlife Institute of India, Chandrabani, Dehradun 248 001, India*
[d] *West Bengal Forest Department, Bikas Bhavan, North Block, Salt Lake City, Calcutta 700 001, India*

Received 28 October 1998; received in revised form 28 December 1998; accepted 28 December 1998; Received by J. Wild

Abstract

We have cloned and sequenced a 906 bp *Eco*RI repeat DNA fraction from *Rhinoceros unicornis* genome. The contig pSS(R)2 is AT rich with 340 A (37.53%), 187 C (20.64%), 173 G (19.09%) and 206 T (22.74%). The sequence contains MALT box, NF–E1, Poly-A signal, lariat consensus sequences, TATA box, translational initiation sequences and several stop codons. Translation of the contig showed seven different types of protein motifs, among which, EGF-like domain cysteine pattern signatures and Bowman–Birk serine protease inhibitor family signatures were prominent. The presence of eukaryotic transcriptional elements, protein signatures and analysis of subset sequences in the 5′ region from 1 to 165 nt indicating coding potential (test code value = 0.97) suggest possible regulatory and/or functional role(s) of these sequences in the rhino genome. Translation of the complementary strand from 906 to 706 nt and 190 to 2 nt showed proteins of more than 7 kDa rich in non-polar residues. This suggests that pSS(R)2 is either a part of, or adjacent to, a functional gene. The contig contains mostly non-consecutive simple repeat units from 2 to 17 nt with varying frequencies, of which four base motifs were found to be predominant. Zoo-blot hybridization revealed that pSS(R)2 sequences are unique to *R. unicornis* genome because they do not cross-hybridize, even with the genomic DNA of South African black rhino *Diceros bicornis*. Southern blot analysis of *R. unicornis* genomic DNA with pSS(R)2 and other synthetic oligo probes revealed a high level of genetic homogeneity, which was also substantiated by microsatellite associated sequence amplification (MASA). Owing to its uniqueness, the pSS(R)2 probe has a potential application in the area of conservation biology for unequivocal identification of horn or other body tissues of *R. unicornis*. The evolutionary aspect of this repeat fraction in the context of comparative genome analysis is discussed. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Endangered species; Genome analysis; Oligonucleotides; One-horned rhino genome; Species-specific probe

1. Introduction

The great Indian one-horned rhino, *Rhinoceros unicornis*, regarded as a World heritage belonging to Schedule-I Animals of the Indian Wildlife (Protection Act, 1972), is a highly endangered species. Once widely distributed over the large tracts of tall, wet grassland and riverine forests along the foothills of Himalaya, rhinos are now confined to a few scattered protected areas (PAs) in the states of Assam, West Bengal and

* Corresponding author. Tel.: +91-11-610-3799/618-9622; fax: +91-11-616-2125; e-mail: sheralib5@hotmail.com; sherali@nii.res.in.

Abbreviations: AP-PCR, arbitrarily primed polymerase chain reaction; GNP, Gorumara National Park; JWS, Jaldapara Wildlife Sanctuary; MASA, microsatellite associated sequence amplification; mtDNA, mitochondrial DNA; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; PA, protected area; PCR, polymerase chain reaction; PIC, polymorphic information content; RAPD, random amplification of polymorphic DNA.

0378-1119/99/$ – see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0378-1119(99)00015-3

Uttar Pradesh in India. The surviving populations in West Bengal have become isolated, since no forest corridor exists between the two PAs, namely Jaldapara Wildlife Sanctuary (JWS) and Gorumara National Park (GNP). The absence of any likelihood of infusion of newer genetic material into the already confined gene pool, insurmountable biotic pressure, ever-growing habitat encroachment and the continued menace of poaching have become causes of concern for the surviving populations.

R. unicornis has 82 chromosomes, where the male is heterogametic with XY chromosomes and the female is homogametic with two XX chromosomes (Wurster and Benirschke, 1968) of which one X chromosome is Lyonized and forms Barr-body. Allozyme studies in African and Asian rhinoceroses (Swart et al., 1994) and mitochondrial DNA (mtDNA) analysis of black rhinoceros *Diceros bicornis* (Ashley et al., 1990) have revealed a low level of intra-species genetic variation. The mtDNA sequences of the *R. unicornis* were used for establishing phylogenetic relationships among related mammals (Xu et al., 1996). Further, on the basis of such analysis, divergence time of approx. 27 million years before present between *Ceratotherium simum* and *R. unicornis* has been reported (Xu and Arnason, 1997). Earlier studies on the genetic variation of rhinoceroses were confined to the analysis of protein loci and mtDNA. In the absence of any information on the polymorphic DNA in the rhino or its genome organization, we studied one-horned rhino by employing established protocols to uncover genetic polymorphism within this species (John and Ali, 1997; Afroze et al., 1998; Mattapallil and Ali, 1998). We cloned and sequenced a repetitive fragment from the *R. unicornis* genome and assessed its polymorphic information content. For ascertaining genetic diversity by Southern blot analysis, several cloned and synthetic oligonucleotide probes were used. In addition, microsatellite associated sequence amplification (MASA) was conducted. The assessment of genetic diversity within and between the rhino populations would facilitate effective conservation and better management of this endangered species, besides providing information on its genome organization.

2. Materials and methods

2.1. *Collection of blood samples from* R. unicornis

Blood samples from six wild rhinos (four males and two females) from JWS and one (male) from the National Zoological Park (NZP) New Delhi, were obtained with prior permission of the Union Government of India and State Government of West Bengal. In the wild, rhinos were immobilized with a tranquilizer shot using Immobilon (1.5–1.8 ml per 1800–2200 kg body weight). From the ear vein of the immobilized animals, about 10 ml blood was drawn into a Vacutainer® (Becton Dickinson) containing 200 μl of 0.5 M EDTA as an anticoagulant. The animals were then given 1.5–2.5 ml of Revivon antidote and carefully monitored until they resumed normal activities in their secure habitat. From the captive rhino of Assam origin at NZP, about 0.5 ml blood was drawn into a Vacutainer® as described above.

2.2. *DNA isolation*

For DNA isolation, blood or tissue samples from different sources including *R. unicornis* were processed. DNA was extracted from peripheral blood of humans *Homo sapiens*, cattle *Bos indicus*, buffalo *Bubalus bubalis*, sheep *Ovis aries*, goat *Capra hircus*, rat *Rattus norvegicus*, mouse *Mus musculus*, catfish *Heteropneustes fossilis*, rabbit *Oryctolagus cuniculus*, bird *Columba livia*, pig *Sus scrofa* and camel *Camelus dromedarius* following standard protocols (John and Ali, 1997). Whole body, excluding the alimentary canal, of the house cricket *Acheta domesticus* was used for DNA isolation. Kangaroo *Tammar wallaby* DNA was a kind gift from Dr D.W. Cooper of Macquarie University, Sydney, Australia and South African black rhino *D. bicornis* DNA was kindly provided by Dr Colleen O'Ryan, Department of Chemical Pathology, University of Cape Town Medical School, Cape Town, South Africa. DNA from *R. unicornis* was used for restriction survey, molecular cloning, sequencing, slot blot, Southern blot and DNA profiling analyses. DNA from both species of rhinos was used for microsatellite associated sequence amplification (MASA) and the same, with other species, was used for zoo-blot hybridization.

2.3. *Cloning of* Eco*RI repeat fraction from the* R. unicornis *genome*

During a restriction survey of rhino genome with a battery of enzymes (*Eco*RI, *Bam*HI, *Hae*III, *Hin*dIII, *Dra*I, *Mbo*I, *Rsa*I and *Hin*fI), an *Eco*RI repeat fraction of about 0.9 kb was identified. For cloning of this fragment, pGEM-7 (+) vector and NM554f′ *Escherichia coli* competent cells were used following standard procedures (Sambrook et al., 1989).

2.4. *Characterization and sequencing of recombinant clone(s)*

Recombinant clones were characterized by direct PCR screening of the colonies using M13 forward and reverse universal primers (Sambrook et al., 1989) followed by Southern hybridization with a gel-purified *Eco*RI fragment of *R. unicornis* DNA. Of the several recombinant clones, one pSS(R)2 was fully sequenced

using an automated sequencer (ABI-PRISM, 377 version 2.1.1, Perkin-Elmer, USA) following the manufacturer's standard protocol and sequences were deposited in GenBank (accession No. Y-13565).

2.5. Slot-blot hybridization

Approximately 200 ng of heat-denatured genomic DNA from different sources including *R. unicornis* and *D. bicornis* were slot blotted onto the nylon membrane (Pharmacia LKB, Sweden) and UV cross-linked. Blot hybridization with pSS(R)2 and other cloned and oligo probes (Table 1) and autoradiographies were done following standard procedures (John and Ali, 1997).

2.6. Southern-blot hybridization

For Southern-blot hybridization, 2–3 μg of *R. unicornis* genomic DNA was subjected to independent restriction digestion with *Eco*RI, *Bam*HI, *Hae*III, *Hin*dIII, *Dra*I, *Rsa*I, *Mbo*I and *Hin*fI enzymes (*NEB*, USA) following the supplier's specifications. Electrophoretic separation of the digested DNA, its transfer onto the membrane, hybridization with cloned and synthetic oligo probes (Table 1) and autoradiographies were conducted as described earlier (John and Ali, 1997).

2.7. Microsatellite associated sequence amplification (MASA)

Random amplification of polymorphic DNA (RAPD) (Welsh and McClelland, 1990), also referred to as arbitrarily primed polymerase chain reaction (AP-PCR) (Williams et al., 1990), was conducted with arbitrary oligo primers to detect the sequence variations in a species. Analogous to RAPD analysis, we performed MASA in rhinos, using a set of oligo primers OAT15.2, OAT18.2 (Epplen, 1988) and O33.15 (Jeffreys et al., 1985) harboring microsatellite core sequences (Table 1) to uncover genetic heterogeneity. DNA probes carrying these sequences have been reported to detect genetic polymorphism in a number of vertebrate species (Epplen, 1988). For MASA reaction, annealing temperatures were empirically optimized; 58°C was found to be appropriate for OAT15.2 and OAT18.2 primers, and 60°C for the primer O33.15. Reactions were carried out in a volume of 25 μl containing approx. 25 ng template DNA, 20 pmol of primers each, 0.25 units of *Taq* DNA polymerase (Bangalore Genei, India), 2.5 mM $MgCl_2$, 200 μM of each dNTP, 50 mM KCl, 20 mM Tris–HCl (pH 8.3), 0.1% Triton X-100 and an equal volume of mineral oil on a Thermal Cycler (Perkin Elmer, Cetus). The reaction mixture was first heat denatured at 96°C for 2 min, and amplified for 35 cycles comprising subsequent steps of denaturation at 94°C for 1 min, annealing at the temperature(s) mentioned above for 1 min and extension at 72°C for 1 min. On completion of the cycles, the amplified products were further incubated at 72°C for 5 min. Approximately 15 μl of the amplified products were resolved on 2% agarose gel in 1×TAE buffer, stained with ethidium bromide and photographed under UV.

Table 1
Probes used in Southern blot and MASA analyses of the rhino genomes

Serial no.	Probe name[a]	Probe characteristics
1	pSS(R)2	(cloned) 906 bp
2	pSS9	(cloned) 824 bp
3	pSS12	(cloned) 1378 bp
4	pDS5	(cloned) 1369 bp
5	OAT15.2	5′-GACAGACAGACAGAC-3′
6	OAT18.2	5′-CAGACAGACAGACAGACA$_4$-3′
7	OAT36	5′-(GACA)$_9$-3′
8	O33.15	5′-CACCTCTCCACCTGCC-3′

[a] Numbers 1–4 represent genome-derived cloned probes and 5–9 are short synthetic oligos.

2.8. Construction of phylogenetic tree based on MASA with OAT15.2 primer

Genomic DNA samples from different species, together with *R. unicornis* samples, were used for random amplification of microsatellite associated sequences. The size of the fully resolved bands on the gel and overall band similarities between different species were ascertained on an optically enhanced densitometer using Diversity One Version 1.6 software package (PDI, New York) at a resolution of 169 μm and tolerance value of 0.75%. Based on the average percentage of matched bands amongst different species, the phylogenetic tree was generated automatically using the above-mentioned software.

2.9. Statistical analysis

The cladogram was based on the statistical analysis for which conventional estimates of probabilities of identity were followed (Jeffreys et al., 1985: John and Ali, 1997). For the rhino-derived probe pSSR(2), we relied on fully resolved bands in the range of 0.5–20 kb after zooming the autoradiogram on the above-mentioned optically enhanced densitometer. Digests from enzymes found not to be informative were excluded from such analysis. Phylogenetic analysis of the data was carried out using the Sequential, Agglomerative, Hierarchical and Nested (SAHN) clustering method in the NTSYS-pc (version 1.70) program (Rohlf and Slice, 1992). Each resolved band was scored as binary codes, for the presence '1' or absence '0' of a specific DNA fragment in the operational taxonomic units (OTUs).

From the data matrix, a similarity matrix (using the Jaccard coefficient) and subsequently a phenogram via the Unweighted Pair Group average Method with Arithmetic averages (UPGMA) were generated. The genetic distances between a pair of animals were calculated from the mean allelic frequencies at the microsatellite repeat locus pSS(R)2 (see Table 1) employing the bias-corrected procedure for Nei's standard distance (Nei, 1978).

2.10. DNA sequences analyses

The pSS(R)2 sequence was searched for protein and nucleic-acid motifs using the Gene Runner Program (Hastings Software, USA). The relative occurrence of short tandem repeat motifs and test-code values for overall coding potentials of the pSS(R)2 contig and its subset sequences were ascertained using DNAsis software (LKB/Pharmacia, Sweden).

3. Results

The details of the different probes used independently for Southern-blot and MASA analysis are given in Table 1. Of the several probes cloned, pSS(R)2 (accession No. Y-13565) was derived from the *R. unicornis* genome and pSS9 (accession No. Y-11529) and pSS12 (accession No. Y-12073) were both from the caprine genome. The other cloned probe pDS5 (accession No.Y-07658) was of bubaline origin. The origin and rationale of oligo probes O33.15 (Jeffreys et al., 1985; Ali and Wallace, 1988), OAT15.2 and OAT18.2 (John and Ali, 1997), and OAT36 (Ali et al., 1993) have been reported earlier.

3.1. Zoo-blot hybridization

The probes listed in Table 1 were used for zoo-blot analysis to ascertain their presence in the *R. unicornis* and other genomes prior to using them for Southern blot and MASA analyses. The bubaline- (pDS5) and goat- (pSS9 and pSS12) derived cloned probes did not cross-hybridize with *R. unicornis* DNA and were not used for Southern blot analyses. Similarly, cloned probe pSS(R)2 derived from *R. unicornis* did not cross-hybridize with any of the non-rhino or South African black rhino *D. bicornis* genomes (Fig. 1).

3.2. Southern blot hybridization of R. unicornis *genomic DNA with pSS(R) 2 probe*

A total of eight different enzymes chosen empirically were used independently for digestion of *R. unicornis* genomic DNA and hybridized with labeled probe pSS(R)2. The *Eco*RI-digested samples revealed eight to

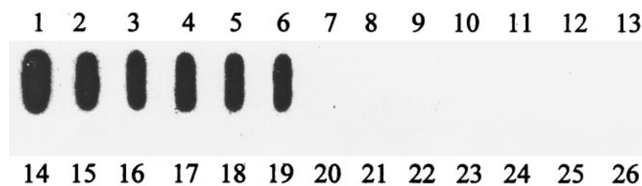


Fig. 1. Evolutionary uniqueness of the pSS(R)2 clone from *R. unicornis* based on slot blot hybridization with DNA samples from different organisms. Spots represent DNA samples: 1–6, *R. unicornis*; 7, human; 8 and 9, goat; 10, sheep; 11, buffalo; 12, cattle; 13, pig; 14, kangaroo; 15, rabbit; 16, rat; 17, mouse; 18, catfish; 19, camel; 20, bird; 21, house cricket; 22, *E. coli*; 23 and 24, *D. bicornis*; 25 and 26 control ($2 \times$ SSC). Note the exclusive signals in the *R. unicornis* sample and their absence in all the other samples.

ten discernible bands with varying signal intensities in the range of 0.3–10 kb, with a few uncommon bands showing a low level of polymorphism. As expected, an isomorphic band at about 900 bp was most prominent (not shown). The other enzyme digests such as *Bam*HI, *Hae*III, *Hin*dIII and *Hin*fI showed either multilocus monomorphic hybridization pattern or smeary signals and no polymorphism was detected (not shown). With *Dra*I enzyme, eight discernible monomorphic bands in the range of 470 bp to 3 kb were detected, of which four bands in the lower molecular weight regions were strong (Fig. 2a). Of the four clearly discernible bands common to all the samples, a 1 kb band was stronger than the remaining three of about 1.2 kb, 860 and 470 bp. The 860 bp was shared by all the samples, but showed strong signals in only three samples (Fig. 2a, lanes 4, 5 and 6). The *Mbo*I-digested samples revealed a total of seven monomorphic bands, of which four (500, 440, 300 and 250 bp) were clearly discernible. Of all the bands detected, that of 300 bp was most prominent, while three of about 1.2 kb, 730 and 670 bp were weak (Fig. 2b). *Rsa*I-digested samples uncovered a number of monomorphic bands in the range of 5 kb to 500 bp, of which the 1.20 kb band was most prominent. The next most prominent one was of about 2.35 kb (Fig. 2c). Other monomorphic bands in higher and lower molecular weight regions were faint and less discernible.

3.3. Southern hybridization of R. unicornis *genomic DNA with synthetic oligo probes*

Several oligo probes (Table 1) in combination with eight different enzymes were screened for detecting fragment length polymorphism, of which OAT36 in combination with *Bam*HI, *Eco*RI, *Hin*dIII, *Hae*III and *Hin*fI detected discernible monomorphic bands in the range of about 400 bp to 9 kb with similar signal intensities. Other enzyme digests showed smeary signals (data not shown).

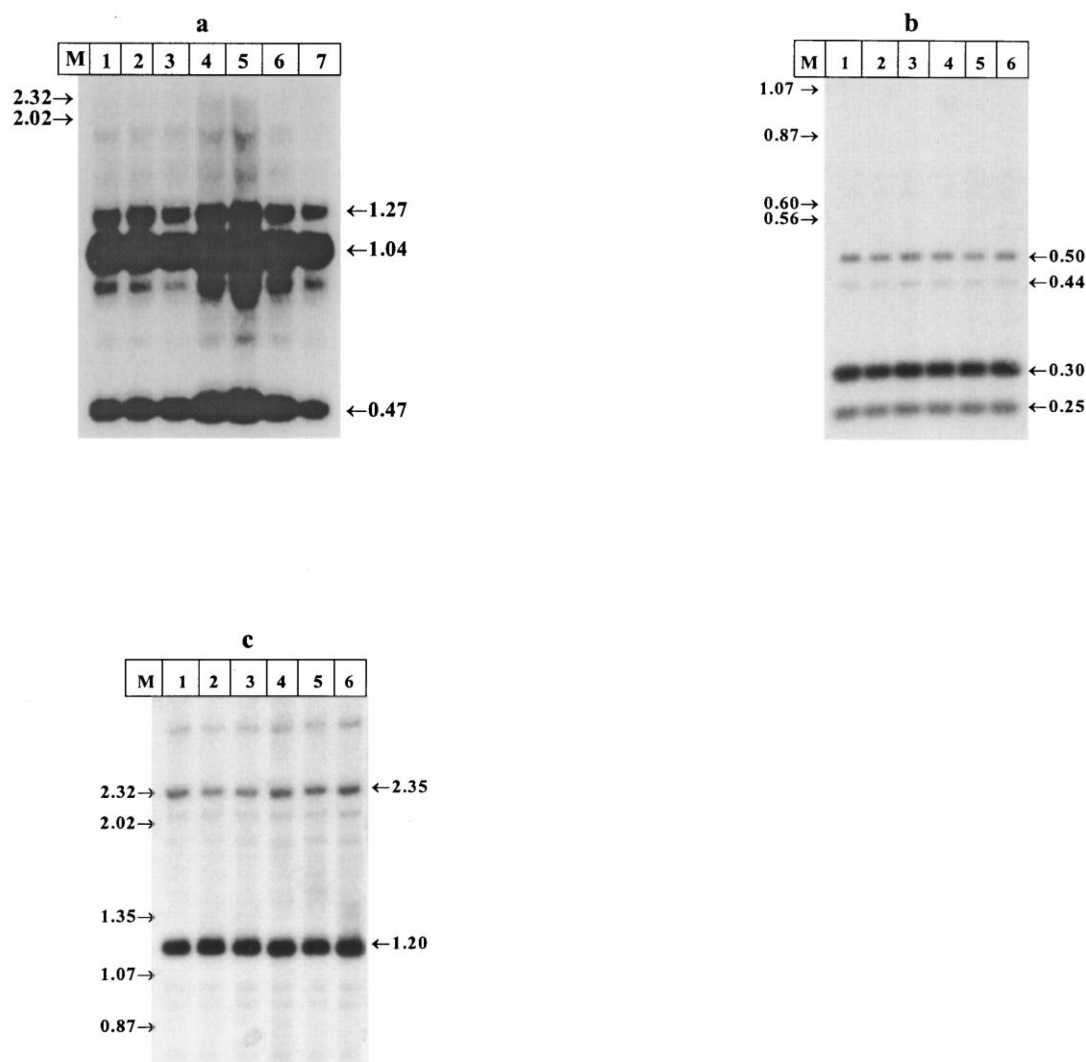


Fig. 2. Genetic homogeneity within the *R. unicornis* species from JWS uncovered by pSSR(2) probe and *Hin*dIII (a), *Mbo*I (b) and *Rsa*I (c) enzyme combinations. Lanes: 1–6, samples from JWS; 7, samples from NZP [Assam origin, present in (a) only]. Note several isomorphic bands with varying degrees of signal intensity in the range of 250 bp to 5 kb and very prominent isomorphic bands of about 1 kb in (a), 300 bp in (b) and 1.2 kb in (c). Molecular size marker λ *Hin*dIII and or ϕx 174/*Hae*III DNA (given in kb) were included in the gel.

3.4. MASA profiling of rhino genomes

Three primers (OAT15.2, OAT18.2 and O33.15) were used for MASA analysis of rhinos genomic DNA (Table 1). Unlike RAPD, MASA enables amplification of micro/minisatellite associated sequences adjacent to the primer annealing sites. With OAT15.2 oligo primer, several common isomorphic bands in both rhino species (*R. unicornis* and *D. bicornis*) were detected, of which four bands of about 1.4 kb, 900, 750 and 670 bp were prominent (Fig. 3a). In *R. unicornis*, a band of about 1.91 kb showed copy number variation (Fig. 3a, lanes 2–7) which was absent in *D. bicornis* samples (Fig. 3a, lanes 8–9). Similarly, two prominent isomorphic bands of about 600 and 800 bp specific to *R. unicornis* and several isomorphic bands specific to *D. bicornis* were detected.

With OAT18.2 primer, the overall number of isomorphic bands detected was more than that with OAT15.2 primer. Of the several isomorphic bands of varying sizes, those of 1.4 kb, 980, 870, 740 and 670 bp were prominent (Fig. 3b). A 630 bp band prominent in *R. unicornis* was absent in *D. bicornis* samples. In *D. bicornis*, of the several isomorphic bands, a 2.4 kb band was prominent. MASA analysis of rhinos genomic DNA with O33.15 primer revealed a large number of isomorphic bands, of which 1070, 1000, 930, 750 and 720 bp were prominent (Fig. 3c). Five isomorphic bands of about 2.06, 1.89, 1.43 kb and 840 and 690 bp were very distinct in *R. unicornis*, but absent in *D. bicornis* (Fig. 3c, lanes 2–7). In *D. bicornis*, three common isomorphic bands of about 1.98, 1.57 kb and 990 bp were detected (Fig. 3c, lanes 8–9). Although all the above primers uncovered species-specific pattern, none of these detected polymorphic

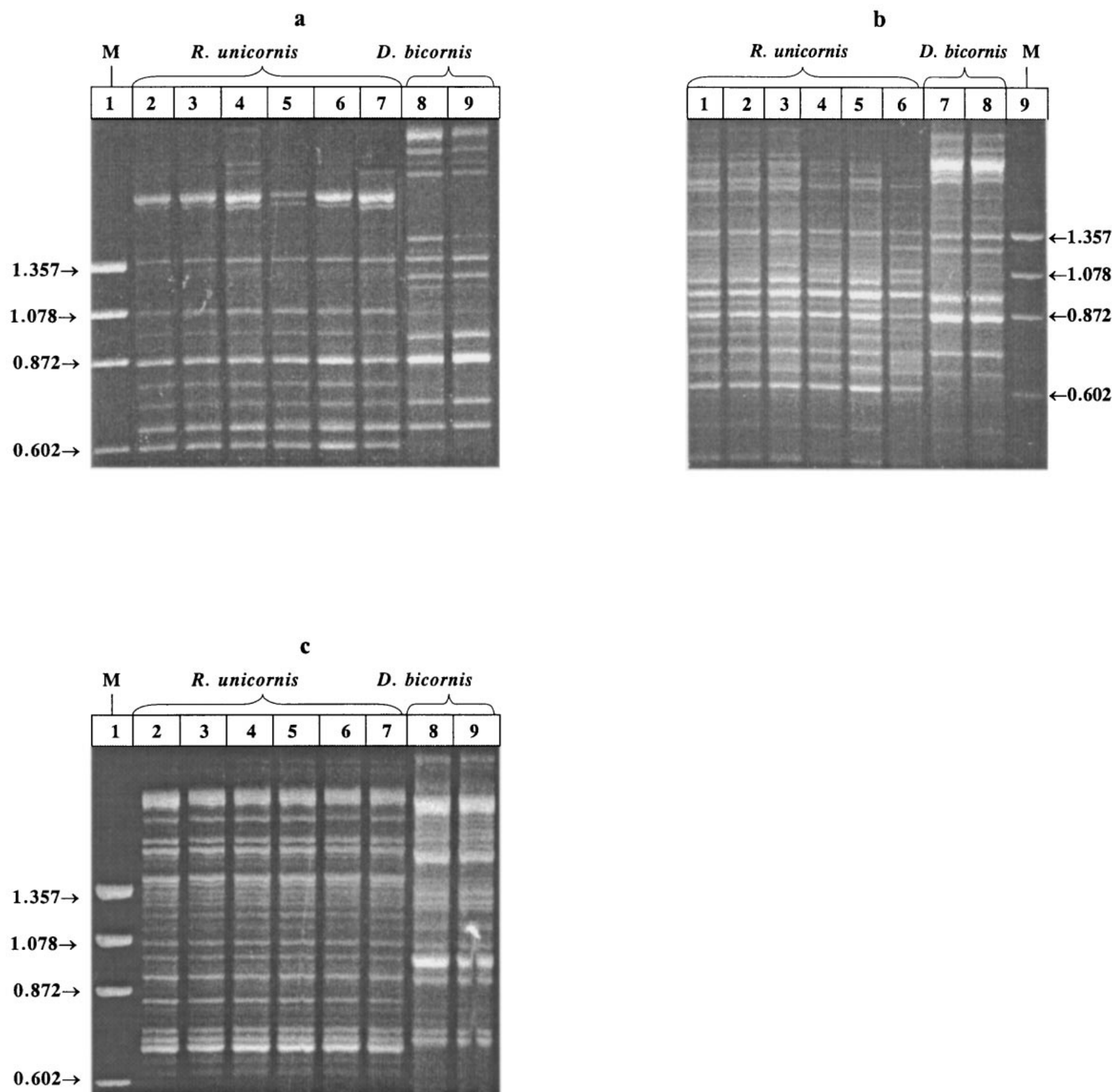


Fig. 3. MASA of rhino genomic DNA with oligo primers: (a) OAT15.2, (b) OAT18.2 and (c) O33.15 resolved on 2% agarose gel. Lanes represent six samples of *R. unicornis* from JWS and two from South African black rhino *D. bicornis*. Marker ϕx 174/*Hae*III digest is given in kb.

bands in any of the rhino species. The OAT15.2 revealed relatively reduced number of bands giving rise to an easily identifiable MASA profile between the two species of rhinos.

3.5. Jaccard's similarity coefficient

Our combined approach of RFLP and MASA for studying the genetic variation in *R. unicornis* revealed a high level of homogeneity. The limited number of variable bands detected by *Eco*RI-digested genomic DNA hybridized with pSS(R)2 clone were used for statistical calculation to obtain Jaccard's similarity coefficient (Fig. 4a). All the males (A, B, C, G, F) clustered together, whereas two females formed a separate unit. It therefore appears that females segregating together are genetically closer. Sample 'F' within the male cluster was separate as compared with the others (data not shown). Both male and female samples showed about 86% similarity reflecting their close genetic affinities with respect to pSS(R)2 repeat sequences.

3.6. Phylogenetic relationship of rhino with other species based on MASA analysis

The oligo primer OAT15.2 used for MASA showed genome-specific amplicons with DNA samples from rhino and other vertebrate species (data not shown). As mentioned earlier, based on the average percentage of matched bands using Diversity One® Version 1.6 software (PDI, New York, USA), a phylogenetic tree was constructed to ascertain the genetic affinities of rhino with other vertebrate species with respect to OAT15.2 repeat loci (Fig. 4b). Rhino samples from Jaldapara

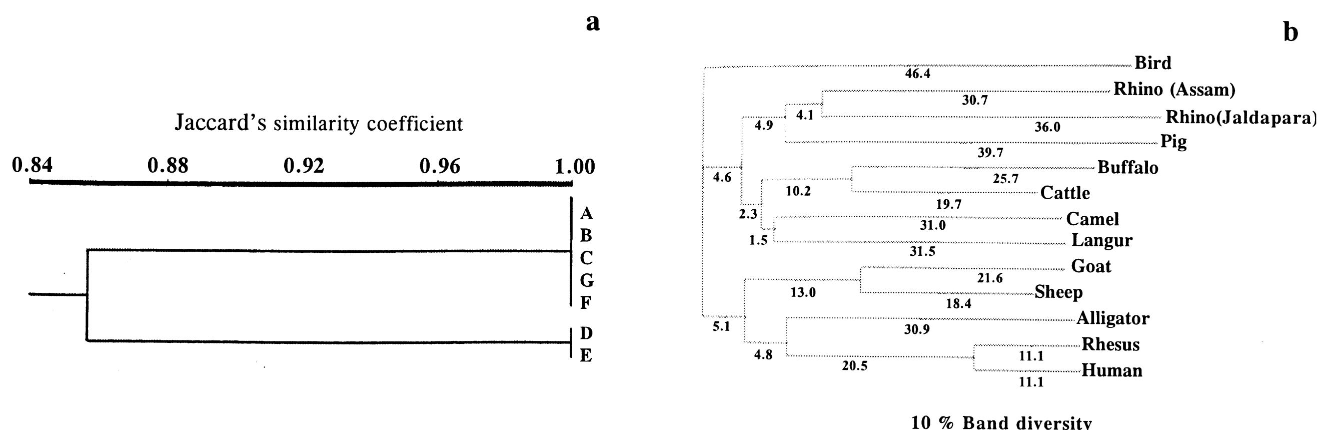


Fig. 4. Genetic distance amongst rhino samples based on Jaccard's similarity coefficient and their phylogenetic relationship. (a) Jaccard's similarity coefficient is based on DNA band profiles of seven rhino samples (autoradiogram not shown) using SHAN clustering method of UPGMA averages. Samples A, B, C, D, E, and G are from Jaldapara WLS and 'F' from NZP (Assam origin). Samples D and E are females and the others are males. Note that sample 'F' forms a separate unit within the male cluster signifying its separate origin. (b) Phylogenetic tree based on average percentage of matched bands detected by MASA reaction using oligo probe 0AT15.2 (figure not shown). The band size was estimated on an optically enhanced automated densitometer using Diversity One software (DPI, New York). Two rhinos (one each from JWS and NZP) are placed in the same cluster compared with other animals.

and Assam form one cluster but still these two maintained separate units, whereas the other vertebrate species, depending upon their genetic affinities, segregated accordingly. The phylogenetic tree in the present study is specific to OAT15.2 repeat loci and may not be construed for final positioning of a clad in the context of molecular systematics. Amplicons generated by MASA using other oligo primers were not used for construction of a phylogenetic tree, because primers OAT15.2 and OAT18.2 both carry similar sequence repeat motif whereas primer O33.15 revealed numerous bands (not shown).

3.7. Sequence analyses of the pSSR(2) contig

Several eukaryotic transcriptional elements, such as MALT Box, NF–E1, poly-A signal, lariat consensus sequence, TATA box, translational initiation sequence and stop codons, were detected in the pSSR(2) contig. Search of protein motifs revealed two protein kinase phosphorylation sites, five *N*-glycosylation sites, 11 casein-kinase II phosphorylation sites, 45 prokaryotic membrane lipoprotein lipid attachment sites and 176 *N*-myristoylation sites, in addition to two EGF-like domain cysteine pattern signatures and five Bowman–Birk serine protease inhibitors family signatures. In addition, different types of repeat motifs with varying frequencies were detected (Fig. 5a, b). Of the several repeat motifs the dimeric repeats were of 16 different types, of which the most common was purine 'AA', occurring 136 times, and the least common was CG, present 11 times (Fig. 5a). The tetramers were present for a maximum of 174 different types, of which AAAT were detected 19 times (Fig. 5b). The least frequent repeat motifs amongst tetramers were present twice, which were of 33 different types. A nt (5′-AACAAGTTT-3′) repeat, present non-consecutively, had three copies within the pSSR(2) array. Repeat motifs of 10–17 nt were present twice, whereas 18 nt and beyond were absent. The pSSR(2) contig is AT-rich with 340 A (37.53%), 187 C (20.64%), 173 G (19.09%) and 206 T (22.74%).

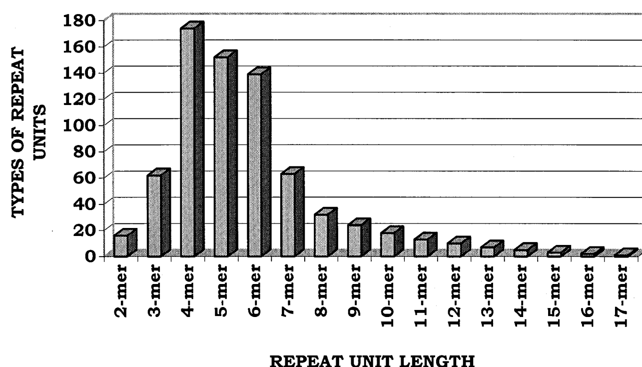
3.8. Coding potentials of pSSR(2) contig

Computer analysis using DNAsis software (LKB/Pharmacia) showed that a subset of the contig pSSR(2) from 166 to 906 nt is a non-coding region (test code value = 0.41). However, subset sequences in the 5′ regions from 1 to 165 nt showed TC values in the range of 0.96–0.97 representing coding sequences as per the algorithm of protein coding prediction values. Blast search revealed varying degrees of sequence homology with human chromosome 4, 6 and 11 besides other functionally unidentified sequences from *Schizosaccharomyces pombe* and *Caenorhabditis elegans* (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast).

4. Discussion

4.1. Sequence variations in R. unicornis genome

The present study was aimed at assessing the intraspecies sequence variation using rhino-derived cloned pSS(R)2 and several synthetic oligo probes. This probe detected a low level of polymorphism in combination with *Eco*RI enzyme, whereas other enzymes were not informative in detecting any polymorphism. Similarly, synthetic oligo probe OAT36 in combination with the



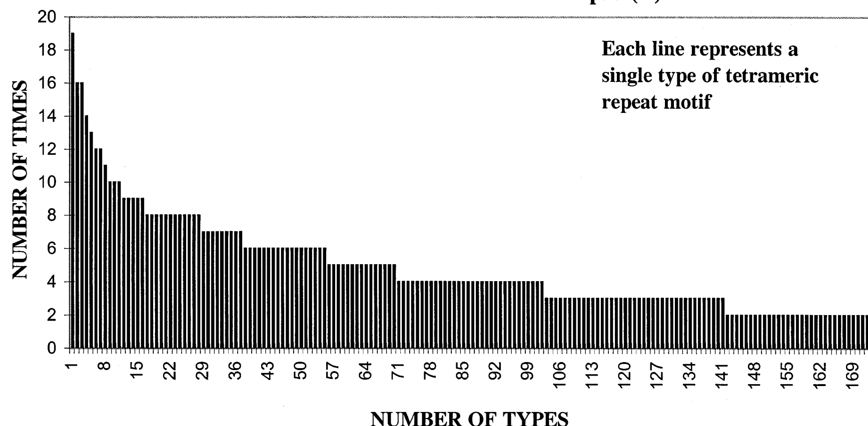


Fig. 5. Relative occurrence of different types of repeat motifs in the *R. unicornis* contig pSSR(2) ranging from 2 to 17 nt (a) and only tetrameric ones (b) (see text for details).

above-mentioned enzymes gave rise mostly to a monomorphic band pattern. Selection, genetic drift, and factors such as migration and population mating structure are known to influence allele distribution (Oakey and Tyler-Smith, 1990). However, our consideration of the effects of these causative factors on allelic distribution is reduced, because *R. unicornis* samples used in the present study originated from a single confined population. The reduced level of polymorphism detected in pSS(R)2 locus is a reflection of genetic homogeneity in the rhino and this could simply be the result of the restricted sampling in our study. This may also be attributed to population mating structure and not to any other kind of selection process. This is corroborated by MASA profiles based on the oligo primers known to detect genetic variability. The RAPD provides several advantages over conventional techniques to uncover genetic polymorphism. However, the target site selection by RAPD primer(s) resulting in multiple amplicons is unable to score for specific recombination ‘hot spots’ due to the random nature of the primers. Mini- or microsatellite sequences have been implicated with the high rate of recombinatorial activities leading to sequence polymorphism in the genome (Royle et al., 1988). We therefore designed a strategy to specifically amplify such sequences by employing microsatellite-specific 15–18-mer primers. In MASA, the amplicons overwhelmingly correspond to microsatellite associated sequences covering a larger pool of genetic variabilities (Azfer et al., 1999). The low level of polymorphism detected by pSS(R)2/*Eco*RI combination and the presence of several isomorphic bands in the rhinos detectable by MASA are in accordance with each other. Thus, MASA profiling substantiates the absence of overall allelic heterozygosity evident from Southern-blot analysis. The reduced level of polymorphism may be attributed to the small number of rhino samples used in the present study. This aspect can be categorically resolved only by analysing more samples, which might not be feasible due to logistic constraints.

4.2. Genetic similarities among R. unicornis

Among all the samples of *R. unicornis*, the overall similarity with respect to pSS(R)2 locus was found to be 86%. All the animals from a single geographical habitat segregated together owing to their genetic affinities (Fig. 4a). Jaccard's similarity coefficient showed that in the cluster, a single male ('F') from a different location forms a separate unit. The apparent genetic similarity among the six samples and relatively limited variation between these six and the single male is not surprising, because all six samples originated from one (Jaldapara) habitat in West Bengal, whereas male 'F' from the National Zoological Park was of Assam origin. This is a significant observation indicating at least some level of genetic variation between the two isolated populations. In the light of this observation, the animals from these two populations may be considered for reshuffling to achieve infusion of unrelated genetic material in the two confined gene pools. The low level of genetic diversity (Sulaiman and Hasnain, 1996; Ali et al., 1999) is likely to cause inbreeding depression.

On the population front, loss of a single male that has already sired several offspring will be less damaging compared with the one that has not. Thus, total protection of the rhino populations by strict control on poaching would allow them to maintain a natural level of heterozygosity while propagating their germplasm in an unconstrained manner. Notwithstanding present results, conclusive evidence may be obtained only if similar analyses were conducted on a larger number of samples from other areas to gain an insight into intra-species genetic variations. This would also enable the delineation of rhinos from different Indian populations into a single clad or possible sub-species.

4.3. Coding potentials of pSS(R)2 contig

Logistic constraints did not permit us to use fresh tissues from rhino for Northern-blot analysis or RT–PCR-mediated expression studies. Thus, we opted for an indirect approach to address this issue. In this context, search of nucleic-acid motifs within the pSS(R)2 contig showed the presence of two NF–E1 sites and several eukaryotic transcriptional elements. Translation of all the three frames showed EGF-like domain cysteine pattern signatures and Bowman–Birk serine protease inhibitor family signatures amongst seven different protein motifs. The presence of EGF-like domain cysteine pattern signature may be of significance, because this is a conserved structure in a large number of unrelated proteins such as transforming growth factor alpha, amphiregulin, coagulation factor VI, IX, X and XII, Tenscin and complement components C_6, C_7, C_8, etc. (Davis, 1990). However, a common feature is that these are membrane-bound secretory proteins of extracellular domains. Since the search of nucleic acid also showed NF–E1 motifs which are important in the context of heavy-chain components, the presence of the EGF cysteine pattern is supported because immunoglobulin are present both in secreted (sIg) and membrane bound (mIg) forms. The Nf1 protein binds to the E1 motif that is believed to be a transcription activating site in the μ domain of the heavy chain enhancer (Birnstiel, et al., 1985; Moore and Sharp, 1985). The μ_m-C-terminal portion is encoded by two exons. Untranslated nt including AATAAA sequences towards the 5′ regions of the poly-adenylation site follow the second μ_m and all the μ_s exons. Thus, a possible involvement of pSS(R)2 repeat array or its subset sequence with the regions coding C_H domains for μ_m and μ_s chains may not be ruled out. Despite the lack of direct evidence of functional significance of eukaryotic transcriptional elements in pSS(R)2 array, the presence of several protein attachment sites and signatures suggest that these sequences may have important role(s).

5. Conclusions

The contig analysis suggests that pSS(R)2 is either a part of or adjacent to a functional gene and its putative involvement in regulation of the expression of neighboring gene(s) is still possible. This is indirectly supported by the coding potential (TC = 0.97) of the subset sequences in the 5′ region from 1 to 165 bp and its complementary strand. The MASA analysis indicates a high level of genetic homogeneity in the *R. unicornis* genome. Notwithstanding the regulatory role or transcriptional status of pSS(R)2, this unique clone may be used for ascertaining the origin of tissue, blood, horn, hoof or skin samples from this species, facilitating the steps taken for conservation of a critical biodiversity.

Acknowledgements

Part of this work was supported by grant No. SP/SO/D-51/93 from the Department of Science and Technology, New Delhi to S. Ali and a core grant from the Department of Biotechnology, Government of India to the National Institute of Immunology, New Delhi. We express our sincere thanks to S.K. Mukherjee, S.K. Basu, Arin Ghosh, V.B. Sawarkar and Subroto Mukherjee who took personal interest to ensure the success of this project. D.M. Singh and N.P. Selvam of the National Zoological Park, New Delhi, assisted in collection of blood sample from the captive rhino of Assam origin. Our special thanks are due to Colleen O'Ryan, Department of Chemical Pathology, University of Cape Town Medical School, Cape Town, South Africa for a kind gift of the *D. bicornis* DNA samples,

Seyed E. Hasnain and Dinkar Sahal for their many valuable suggestions on this paper and Khem Singh Negi for technical assistance.

References

Afroze, D., Misra, M.A., Sulaiman, I.M., Sinha, S., Sarkar, C., Mahapatra, A.K., Hasnain, S.E., 1998. Genetic alteration in brain tumors identified by RAPD analysis. Gene 206, 45–48.

Ali, S., Ansari, S., Ehtesham, N.Z., Azfer, M.A., Homkar, U., Gopal, R., Hasnain, S.E., 1999. Analysis of the evolutionarily conserved repeat motifs in the genome of the highly endangered central Indian Swamp deer *Cervus duvauceli branderi*. Gene 223, 361–367.

Ali, S., Gauri, V., Bala, S., 1993. Detection of genome specific monomorphic loci in *Bos taurus* and *Bubalus bubalis* with oligodeoxyribonucleotide probes. Anim. Genet. 24, 199–202.

Ali, S., Wallace, R.B., 1988. Intrinsic polymorphism of variable number tandem repeat loci in the human genome. Nucleic Acids Res. 16, 8487–8496.

Ashley, M.V., Melnick, D.J., Western, D., 1990. Conservation genetics of the black rhinoceros *Diceros bicornis*, I: evidence from the mtDNA of three populations. Conserv. Biol. 1, 71–77.

Azfer, M.A., Bashamboo, A., Ahmed, N., Ali, S., 1999. Random amplification of polymorphic DNA with conserved sequences reveals genome specific monomorphic amplicons: implications in clad identification. J. Biosci. 24, in press

Birnstiel, M.L., Busslinger, M., Strub, K., 1985. Transcription termination and 3′ processing: the end is in site!. Cell 41, 349–359.

Davis, C.G., 1990. The many faces of epidermal growth factor repeats. New Biol. 2, 410–419.

Epplen, J.T., 1988. On simple repetitive GATA/GACA sequences in animal genomes: a critical appraisal. J. Hered. 79, 409–417.

Jeffreys, A.J., Wilson, V., Thein, S.L., 1985. Hypervariable minisatellite regions in human DNA. Nature 314, 67–73.

John, M.V., Ali, S., 1997. Synthetic DNA-based genetic markers reveal intra- and inter- species DNA sequence variability in the *Bubalus bubalis* and related genome. DNA Cell Biol. 16, 369–378.

Mattapallil, M.J., Ali, S., 1998. Parentage assessment of an IVF calf from *Bubalus bubalis* by DNA fingerprinting. Gene 206, 209–214.

Moore, C.L., Sharp, P.A., 1985. Accurate cleavage and polyadenylation of exogenous RNA substrate. Cell 41, 845–855.

Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89, 583–590.

Oakey, R., Tyler-Smith, C., 1990. Y chromosomal DNA haplotyping suggests that most European and Asian men are descended from one of two males. Genomics 7, 325–330.

Rohlf, F.J., Slice, D.E., 1992. Numerical taxonomy and multivariate analysis system: version 1.70. In: Rohlf, F.J. (Ed.), Applied Biostatistics. Stony Brook, New York, pp. 1-1–5-5.

Royle, N.J., Clarkson, R.E., Wong, Z., Jeffrey, A.J., 1988. Clustering of hypervariable minisatellite in the proterminal regions of human autosomes. Genomics 3, 352–360.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. 2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sulaiman, I.M., Hasnain, S.E., 1996. RAPD markers reveal genetic homogeneity within endangered Himalayan species *M. paniculata* and *M. simplicifolia*. Theoret. Appl. Genet. 93, 91–96.

Swart, M.K.J., Ferguson, J.W.H., DuToit, R., Flamand, J.R.B., 1994. Substantial genetic variation in Southern African black rhinoceros *Diceros bicornis*. J. Hered. 85, 261–266.

Welsh, J., McClelland, M., 1990. Fingerprinting genome using PCR with arbitrary primers. Nucleic Acids Res. 18, 7213–7218.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V., 1990. DNA polymorphisms amplified by arbitrary primers are useful genetic markers. Nucleic Acids Res. 18, 6531–6535.

Wurster, D.H., Benirschke, K., 1968. The chromosomes of the Great Indian Rhinoceros (*Rhinoceros unicornis* L.). Experientia 24, 5–11.

Xu, Xiufeng, Janke, A., Arnason, U., 1996. The complete mtDNA sequence of the greater Indian rhinoceros, *Rhinoceros unicornis*, and the phylogenetic relationship among Carnivora, Perissodactyla, and Artiodactyla (+Cetacea). Mol. Biol. Evol. 13, 1167–1173.

Xu, Xiufeng, Arnason, U., 1997. The complete mtDNA sequence of the white rhinoceros, *Ceratotherium simum*, and comparison with the mtDNA sequence of the Indian rhinoceros, *Rhinoceros unicornis*. Mol. Phylogenet. Evol. 7, 189–194.