Short Technical Reports

Universal Linker and Ligation Procedures for Construction of Genomic DNA Libraries Enriched for Microsatellites

BioTechniques 27:500-507 (September 1999)

ABSTRACT

Microsatellite loci are highly informative genetic markers useful for population genetic studies, linkage mapping and parentage determination. Methods to identify novel microsatellite loci commonly use subtractive hybridization to enrich small-insert genomic libraries for repeat sequences. A critical step in enrichment is attachment of an oligonucleotide linker to genomic DNA fragments so that repeat-containing sequences can be recovered by PCR for cloning. Current linkers and ligation methods rely on single restriction enzymes to size-fraction genomic DNA and generate complementary ends. These restriction enzyme/linker combinations are often species-specific, give poor recovery of repeat-enriched DNA and yield library inserts that are not a broad sample of the genome. We have developed a blunt-end linker, named SNX for its restriction sites, that allows the use of combinations of restriction enzymes to digest the majority of genomic DNA into the 200–1000-bp range. SNX is attached to genomic DNA with a simultaneous ligation/restriction reaction that is highly efficient and improves recovery of sequences after subtractive hybridization. SNX can be used for microsatellite enrichment in any species, since ligation is independent of the restriction enzymes used to size-fraction genomic DNA. These methods improve current repeat-enrichment strategies, resulting in representative small-insert libraries with a very high proportion of positive clones.

INTRODUCTION

Tandemly repeated DNA sequences of 1–6 bp are referred to as simple sequence repeats (SSRs), sequence tagged sites (STS) or microsatellite loci. Microsatellite loci often exhibit many alleles per locus and high levels of heterozygosity and have become a widely used marker type for parentage analysis and population genetic studies (5,9). Although microsatellite loci are useful genetic markers, their development in species without substantial DNA sequence data requires efficient methods to select suitable microsatellite markers from among many candidate repeat sequences. Selectively enriching genomic DNA libraries for repeat-sequences is commonly used to increase the number of cloned fragments that contain repeats compared to random-insert libraries. Repeat-enrichment strategies generally rely on subtractive hybridization, where repeat-oligonucleotides are used to capture single-stranded genomic DNA fragments containing complementary sequences. Only those genomic DNA fragments containing repeats are used for cloning, resulting in libraries enriched for microsatellite sequences. Numerous studies have presented methods of subtractive hybridization to enrich genomic libraries (1,3,4,7,8) (see References 3 and 7 for a schematic representation of these procedures).

Ligation of a double-stranded linker to restriction-digested genomic DNA is usually required before subtractive hybridization. After linker attachment, genomic DNA of unknown sequence can be enriched for repeats and then made double-stranded and amplified by polymerase chain reaction (PCR) using the linker sequence as a priming site. We have found that the design of the linker itself and the ligation methods used to attach it to genomic DNA greatly influence the quality and enrichment rate of resulting genomic DNA libraries. Without high-efficiency ligation of linkers, a large proportion of genomic DNA fragments cannot be recovered for cloning, regardless of the efficiency of subtractive hybridization.

The design of a linker sequence is critical because it must ligate to genomic DNA, serve as a PCR primer and contain the necessary restriction sites for ligation into cloning vectors. Microsatellites are most often cloned with small-insert libraries, where the majority of inserts can be sequenced entirely using primers that flank the cloning site (e.g., T7 and T3). This requires that libraries be constructed with genomic DNA fragments ≤1000 bp, a process called insert size-selection often accomplished by cutting DNA fragments of the desired size range from an agarose gel (1,4). Currently used linkers usually provide an overhang for ligation to genomic DNA that requires the use of a single restriction enzyme to size-select genomic DNA (e.g., Reference 3). This is a limitation since, depending on the species, the restriction enzyme that creates a complementary overhang can yield a small proportion of the genome as fragments ≤1000 bp. Single restriction enzyme size-selection can effectively discard potential microsatellite loci in fragments >1000 bp and result in libraries that are limited samples of a genome (1,6).

We have developed a new linker and ligation strategy that greatly simplifies the construction of repeat-enriched, small-insert genomic libraries. The SNX linker (Figure 1) contains two restriction sites for ligation into vector, and an *Xmn*I site that is created if two linker monomers ligate to form a dimer. SNX is blunt-end ligated to restricted genomic DNA using a simultaneous ligation-*Xmn*I restriction reaction. The ligation conditions are designed to compensate for the inefficiency of blunt-end ligation and the tendency of linkers to ligate to themselves instead of to genomic DNA due to their higher molar end concentration. Blunt-end ligation of SNX to genomic DNA (after treatment of genomic DNA with single-strand specific nuclease) permits use of any restriction enzyme or combination of enzymes that yield the majority of genomic DNA fragments ≤1000 bp without size selection and associated loss of DNA. Similarly, a simultaneous ligation-restriction reaction was used to ligate repeat-enriched genomic DNA into vector, preventing dimerization of inserts. Using the SNX linker and established subtractive hybridization methods, a single genomic library can be constructed rapidly that is highly enriched for microsatellites and sampled from a large proportion of genome.

MATERIALS AND METHODS

Genomic DNA was extracted from a tree species, four penguin species and

Short Technical Reports

Table 1. Plant, Bird and Primate Species, Source and Method of Extraction of Genomic DNA and Restriction Enzyme Combinations Used for Library Construction

Species	Tissue Type	Extraction Method	Restriction Enzymes	Enrichment Sequences
Corythophora alta	frozen leaves	DNeasy® Plant Mini-Prep[b]	*Alu*I, *Hae*II, *Rsa*I, *Nhe*I	GT, CT, CCG, CTC, CAG
Spheniscus spp.	blood[a]	QIAamp Blood kit[b]	*Rsa*I, *Nhe*I	AAG, AAT, CAT, CAC, GGA GT
Lagothrix lagotricha[c]			*Hae*III, *Rsa*I, *Nhe*I	CAG, CT, GT
Ateles geoffroyi				
Alouatta spp.				

[a]Preserved in lysis buffer (0.1 M Tris, 0.1 M NaCl, 10 mM EDTA, 1.0% SDS).
[b]Purchased from Qiagen.
[c]Extracted DNA provided by Center for Reproduction of Endangered Species, San Diego, CA, USA.

several primates (Table 1). Various combinations of 4- and 6-bp recognition restriction enzymes were tested to determine combinations that resulted in a majority of DNA fragments within the 200–1000-bp size range for each species. All enzyme combinations included *Nhe*I because any genomic fragments containing this site would be cut during ligation into vector (Figure 1; and as described below). *Xmn*I, which prevents dimerization of the SNX linker during ligation to genomic DNA, cuts infrequently in these genomes, so *Hha*I or *Hae*II were not included in digestions (Figure 1). Genomic DNA was digested at 37°C overnight (Table 2A) and then heated at 65°C for 20 min to inactivate restriction enzymes. Five units of mung bean nuclease (MBN) were added directly to digested genomic DNA and incubated at 30°C for 30 min to remove single-stranded overhangs and leave all genomic fragments with blunt ends. Genomic digests were purified using QIAquick™ PCR Purification Columns (Qiagen, Valencia, CA, USA) and eluted in 50 µL of EB buffer provided with the columns. Genomic DNA ends were then dephosphorylated by adding 6 µL NEB buffer 2 (New England Biolabs, Beverly, MA, USA), 3 µL sterile water and 1 uL (10 U) calf intestinal phosphatase (CIP) to the 50 µL of digested, MBN-treated genomic DNA with incubation at 37°C for 2 h. CIP was removed by again purifying genomic DNA with a QIAquick PCR column and eluting the DNA in 30 µL of EB buffer. All restriction enzymes, modifying enzymes and buffers were obtained from New England Biolabs.

Digested, MBN-treated and dephosphorylated ("processed") genomic DNA was ligated to double-stranded SNX linkers using high-concentration DNA ligase to achieve efficient blunt-end ligation. *Xmn*I was also included in this ligation to maintain a high concentration of SNX linkers as monomers available to ligate to genomic DNA (Figure 2A). This simultaneous restriction/ligation reaction was possible because both *Xmn*I and ligase are fully active in NEB Buffer 2. Table 2B gives ligation reaction components and volumes. Two microliters of each ligation was used as template in PCRs (Table 2C, substituting 33.7 µL of water), which were visualized on 2.0% agarose/TBE gels containing 1× GelStar® stain (FMC BioProducts, Rockland, ME, USA). This PCR is the best measure of ligation success, since it demonstrates if genomic fragments can be amplified after enrichment (Figure 2B). Multiple ligation reactions were constructed to supply SNX linker-attached genomic DNA for multiple subtractive hybridizations.

Ligations were used without further

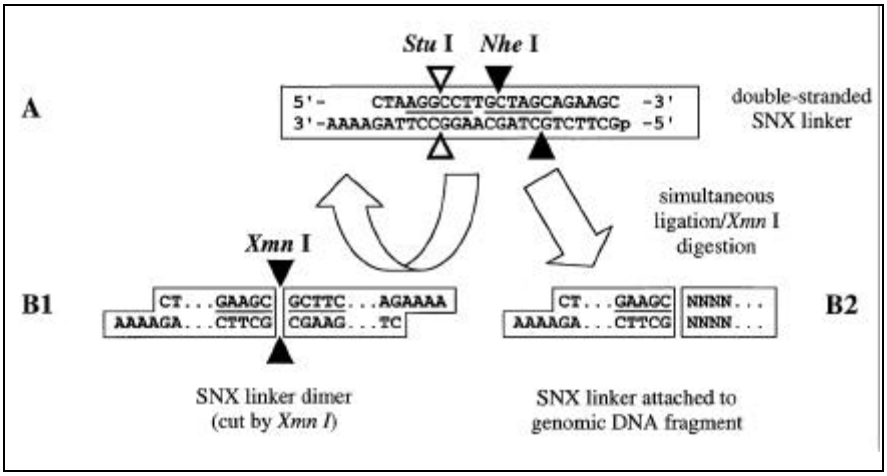

Figure 1. The "SNX" linker and its activity. The SNX linker is named for the two restriction sites it contains (*Stu*I and *Nhe*I) and the restriction site it forms when it dimerizes (*Xmn*I). The *Stu*I and *Nhe*I sites provide blunt and four-base overhang cloning sites, respectively (A). During the simultaneous ligation/restriction reaction, when two linkers ligate together, they form an *Xmn*I site (B1). Including *Xmn*I in the ligation reaction will cut these linker dimers apart, maintaining a large proportion of linkers as monomers available to ligate to genomic DNA (B2). Linkers remain attached to genomic DNA fragments created by any restriction enzyme other than *Xmn*I. *Hha*I and *Hae*II both cut within *Xmn*I sites and can be added to genomic digests if *Xmn*I cuts genomic DNA frequently. The SNX oligonucleotide sequences are: SNX forward 5′-CTAAGGCCTTGCTAGCAGAAGC-3′ and SNX reverse 5′-pGCTTCT-GCTAGCAAGGCCTTAGAAAA-3′, where p indicates phosphorylation.

Table 2. Reaction Components, Volumes and Conditions for Restriction Digests, Ligations and Amplifications

A. Restriction Digest of Genomic DNA[a]		B. Ligation of SNX Linker to Processed Genomic DNA[b]	
Genomic DNA (ca. 200 ng/µL)	81.0 µL	Double-stranded SNX linker (5 µM)	11.7 µL
NEB Buffer #2	10.0	NEB Buffer #2	3.0
BSA (NEB 100×)	1.0	BSA (NEB 100×)	0.3
*Hae*III (20 U)	2.0	Processed genomic DNA	10.0
*Rsa*I (20 U)	2.0	ATP (10 mM)	3.0
*Alu*I (16 U)	2.0	*Xmn*I (20 U)	1.0
*Nhe*I (10 U)	2.0	Ligase (high concentration, 1000 U)	1.0
Total volume:	100.0	Total volume:	30.0
C. PCR for Enriched Genomic DNA[c]		**D. Ligation of Enriched Genomic DNA into Vector[b]**	
Sterile, distilled water	25.7 µL	Amplified, enriched genomic DNA	12.8 µL
Repeat-enriched genomic DNA	10.0	NEB Buffer #2	2.0
10× Thermopol buffer (20 mM $MgSO_4$)	5.0	BSA (NEB 100×)	0.2
dNTP mixture (8 mM; 2 mM each nucleotide)	5.0	ATP (10 mM)	3.0
SNX linker (forward only, 10 µM)	4.0	*Xba*I cut pBluescript SK(+) plasmid (100 ng/µL)	10.0
$Vent_R$® (exo-) DNA Polymerase[d]	0.3	*Nhe*I (5U)	1.0
		Ligase (low concentration, 400 U)	1.0
Total volume:	50.0	Total volume:	30.0

[a]Reaction used for *C. alta*; DNA volume, enzymes and buffer may vary.

[b]Ligations were incubated overnight in a thermal cycler programmed for cycles of 16°C for 30 min and 37°C for 10 min.

[c]Temperature profile: 96°C for 5 min followed by 40 cycles of 96°C for 45 s, 60°C for 1 min, 72°C for 1 min in a Model PTC-200™ (MJ Research, Watertown, MA, USA). Include double-stranded SNX linker, processed genomic DNA and water as controls.

[d]From New England Biolabs.

All enzymes, reagents and buffers were obtained from New England Biolabs, except ATP, which was purchased from Sigma (St. Louis, MO, USA).

purification for subtractive hybridization with 30 bp, 5′ biotinylated oligonucleotides (Operon Technologies, Almeda, CA, USA) as the hybridization target and streptavidin-coated magnetic beads (Dynabeads®; Dynal, Lake Success, NY, USA) as the substrate to capture genomic DNA-oligonucleotide hybrids. This procedure has been described in detail elsewhere (3,8). Briefly, subtractive hybridizations were carried out independently for each oligonucleotide repeat sequence (see Table 1). Six microliters of SNX linker-attached genomic DNA and 2 µL of a repeat-oligonucleotide (1 µM) were added to 50 µL of 2× hybridization solution [12× standard saline citrate (SSC), 0.1% sodium dodecyl sulfate (SDS)] and 42 µL water. These hybridizations were heated at 95°C for 15 min to denature double-stranded DNA and then incubated at hybridization temperature overnight. Hybridization temperatures were 65°C for most repeat-sequences; however, it was 45°C for A/T-rich repeats such as AAT. Hybridization reactions were mixed with 300 µg of streptavidin beads (prerinsed four times in 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 M NaCl). Tubes were incubated for several hours in a shaker-bath at 43°C to allow beads to bind genomic DNA-oligonucleotide hybrids. Beads were washed once in 2× SSC, 0.1% SDS at room temperature, twice in 1× SSC, 0.1% SDS at 45°C and twice in 1× SSC, 0.1% SDS at 60°C. A/T-rich repeats were washed at a maximum of 45°C. All washes were 200 µL for 5 min in a heat block set at the appropriate temperature. Bound genomic DNA was eluted by adding 60 µL of preheated T-dot-E (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) and incubating for 10 min at 95°C.

Repeat-enriched DNA was made double-stranded and amplified in a PCR using the forward SNX linker as a primer (Table 2C). Amplification was verified by running a 5-µL aliquot on 2.0% agarose/TBE gels containing 1× GelStar stain. Four-base overhangs on amplified, enriched genomic DNA fragments were prepared by adding 1 µL (5 U) *Nhe*I and 0.46 µL bovine serum albumen (BSA) directly to the remaining PCR and incubating overnight at 37°C. This digest was purified with a QIAquick PCR column, and the DNA was eluted in 30 µL of EB buffer. Table 2D lists the reaction to ligate this enriched genomic DNA into a pBlue-

script® II SK(+) vector (Stratagene, La Jolla, CA, USA) linearized with *Xba*I and dephosphorylated. Ligations of enriched genomic DNA into vector included *Nhe*I [which does not cut pBluescript SK(+)] to assure that multiple inserts could not ligate together to create chimeric clones. Plasmid ligations were transformed into competent *E. coli* (XL1-Blue MRF′; Stratagene) and plated at low density on LB containing 75 mg/L ampicillin and 100 µL per plate of isopropyl-β-D-thiogalactopyranoside (IPTG)/5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) solution (1× SOC, 5.6 mg/mL IPTG, 5.6 mg/mL X-gal; Reference 10).

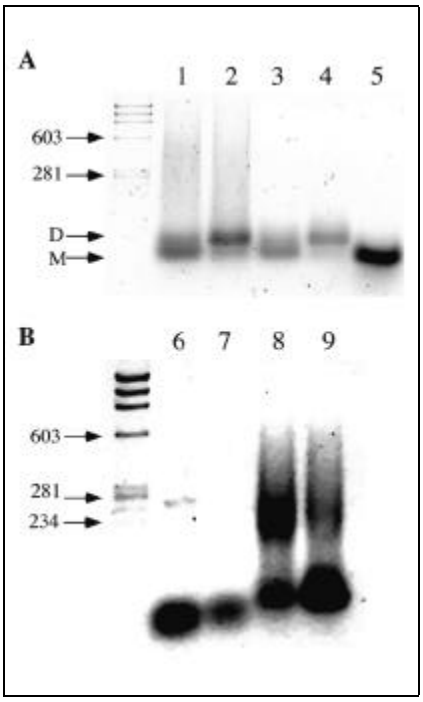

Figure 2. Negative images of SNX linker ligation (A) and PCR amplification of ligations (B). Ligation of SNX to genomic DNA with (A1) or without (A2) *Xmn*I. Control ligation reactions using double-stranded SNX only with *Xmn*I (A3) and without *Xmn*I (A4). Lane A5 contains unmodified monomeric double-stranded SNX linker. Arrows at left without numbers indicate approximate size of monomeric (M) and dimeric (D) SNX linker. Existing linkers can produce PCR artifacts, as seen near 281 bp in a negative-control PCR (lane B6 shows the SAUL linker of References 1 and 4). Double-stranded SNX linker does not produce artifacts during PCR (B7). More genomic DNA is recovered by PCR when *Xmn*I is included in ligation reactions (B8) than ligation without *Xmn*I (B9). The templates used for the PCRs shown in B8 and B9 are shown in lanes A1 and A2, respectively. Molecular weights in bp are indicated by arrows at the left.

Colonies were lifted off LB plates with nylon membranes (MAGNA Lift; Micron Separations Inc. [MSI], Westboro, MA, USA) using standard methods (10). Several steps were taken to prevent high background observed with nonradioactive detection of colony lifts. Filters were air-dried, and the DNA was immobilized with UV illumination. Then filters were washed for 1 h in 5 mL of proteinase K buffer (100 mM Tris-HCl, 50 mM EDTA, 0.5% SDS; Reference 10) containing 25 µL of proteinase K (10 mg/mL) to remove cellular debris. Filters were prehybridized for several hours and hybridized overnight in 0.25 M Na_2HPO_4, pH 7.2, 7% SDS, 1 mM Na-EDTA, pH 8.0, 1% BSA and fraction 5 (11). The hybridization solution contained all bi-

otinylated oligonucleotides (0.2 μM of each) used for subtractive hybridization. Hybridized filters were washed once in 2× SSC, 0.1% SDS at room temperature, twice in 2× SSC, 0.1% SDS at 45°C and twice in 1× SSC, 0.1% SDS at 65°C. Filters with colonies bearing A/T-rich repeats were only washed once at 65°C. Probe DNA bound to filters was detected with the Phototope®-Star Chemiluminescent Detection Kit (New England Biolabs) following the "maximum volume" protocol for colony-lift filters.

Positive colonies were picked with sterile pipet tips, washed into 100 μL of T-dot-E, and cells were lysed by incubation at 100°C for 10 min to release plasmids. Inserts were PCR-amplified with T7 and T3 primers and sequenced on a Model 373XL or 377XL Fluorescent Sequencer (PE Biosystems, Foster City, CA, USA). Primers flanking microsatellite regions were designed and tested with Amplify version 1.2 (2) and then empirically optimized to produce single-band PCR products in agarose gels. PCR products with incorporated FdCTPs were sized on a fluorescent sequencer with GENESCAN™ (PE Biosystems). The full protocol is available at **http://www.georgetown.edu/faculty/hamiltm1**/ or from the corresponding author.

RESULTS AND DISCUSSION

Digestion of genomic DNA with several enzymes yields a large proportion of fragments within the 200–1000-bp size range that can be used directly for library construction. More SNX linker is available as monomers for attachment to genomic DNA when *Xmn*I is included in ligations (Figure 2A). This results in more efficient attachment of SNX linker to genomic DNA and yields more genomic DNA that can be recovered by PCR (Figure 2B). Although the percentage of hybridization-positive clones was difficult to quantify accurately due to the large numbers of colonies, plates exhibited approximately 20% to 95% hybridization-positive clones depending on the enrichment repeat. This resulted in literally thousands of clones that could be sequenced for each species.

In total, several hundred positive clones were sequenced and almost all contained a discrete microsatellite sequence. In *Corythophora alta*, only three redundant clones were observed of 120 sequenced. Six primer sets for *C. alta* that have been screened to date all produced highly polymorphic loci with between 7 and 17 alleles per locus. For GT repeats in *Spheniscus spp.*, 65 clones were sequenced, with 12 being redundant. Eight primer sets have been tested thus far, yielding five polymor-

phic loci. For primates, of 77 clones sequenced to date, five were redundant and all but two contained microsatellite loci. Screening of primer sets is currently in progress.

The ligation methods reported here simplify the cloning process and yield more efficient recovery of microsatellite-enriched sequences. The blunt-end SNX linker can be utilized with any genome and does not depend on the cutting frequency of a specific restriction endonuclease. The ability to digest a large proportion of genomic DNA into the size range that can be sequenced entirely provides several advantages. A single genomic library prepared with these methods should yield microsatellite loci representing the majority of the genome and obviates the need to construct multiple libraries, each based on a different restriction endonuclease/linker combination. These methods might aid microsatellite marker development for linkage mapping studies, where markers are needed from all regions of a genome. They should also facilitate microsatellite marker development for population genetic studies, which depend on randomly located genetic markers to infer subdivision and phylogenetic relationships.

REFERENCES

1. **Armour, J.A. L., R. Neumann, S. Gobert and A.J. Jeffreys.** 1994. Isolation of human simple repeat loci by hybridization selection. Hum. Mol. Genet. *3*:599-605.
2. **Engels, W.R.** 1993. Contributing software to the Internet: the Amplify program. Trends Biochem. Sci. *18*:448-450.
3. **Fischer, D. and K. Bachmann.** 1998. Microsatellite enrichment in organisms with large genomes (*Allium cepa L.*). BioTechniques *24*:796-802.
4. **Fleischer, R.C. and S. Loew.** 1995. Construction and screening of microsatellite-enriched genomic libraries, p. 459-468. *In* J. Ferraris and S. Palumbi (Eds.), Molecular Zoology: Advances, Strategies and Protocols. Wiley-Liss, New York.
5. **Goldstein, D.B. and D.D. Pollock.** 1997. Launching microsatellites: a review of the mutation processes and methods of phylogenetic inference. J. Hered. *88*:335-342.
6. **Hamilton, M.B. and R.C. Fleischer.** Cloned microsatellite repeats differ between 4-base restriction endonucleases. J. Hered. (In press.)
7. **Kandpal, R.P., G. Kandpal and S.M. Weissman.** 1994. Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. Proc. Natl. Acad. Sci. USA *91*:88-92.
8. **Kijas, J.M.H., J.C.S. Fowler, C.A. Garbett and M.R. Thomas.** 1994. Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. BioTechniques *16*:656-662.
9. **McDonald, D.B. and W.K. Potts.** 1997. DNA Microsatellites as genetic markers at several scales, p. 29-48. *In* D.P. Mindell (Ed.), Avian Molecular Evolution and Systematics. Academic Press, San Diego.
10. **Sambrook, J., E.F. Fritsch and T. Maniatis.** 1989. Molecular Cloning: A Laboratory Manual. CSH Laboratory Press, Cold Spring Harbor, N.Y.
11. **Westneat, D.F., W.A. Noon, H.K. Reeve and C.F. Aquadro.** 1988. Improved hybridization conditions for DNA 'fingerprints' probed with M13. Nucleic Acids Res. *16*:4161.

We thank T. Burg, A. Colwell, D. Fonseca, C. McIntosh and L. Shapiro for insightful questions and comments. E.L.P. thanks D. Boersma, G. Miller, M. Schwartz and the Baltimore Zoo for penguin samples. A.D. thanks the Center for Reproduction of Endangered Species for DNA samples. This work was supported by the Smithsonian Institution, the National Zoological Park, a National Science Foundation (NSF) Research Training Grant to the University of Maryland and Smithsonian Institution, an NSF Graduate Fellowship to E.L.P., the Biological Dynamics of Forest Fragments Project and the National Geographic Society.

Address correspondence to Dr. Matthew B. Hamilton, Georgetown University, Department of Biology, Reiss Sciences Bldg. 406, Washington, DC 20057-1229, USA. Internet: hamiltmb@gusun. georgetown.edu

Received 15 March 1999; accepted 24 May 1999.

Matthew B. Hamilton, Elaine L. Pincus[1,2], Anthony Di Fiore[1,2] and Robert C. Fleischer[2]
Georgetown University
Washington, DC
[1]University of Maryland
College Park, MD
[2]Smithsonian Institution
National Zoological Park
Washington, DC, USA