

Reduced Polymorphism in the Chimpanzee Semen Coagulating Protein, Semenogelin I

Sarah B. Kingan, Marc Tatar, David M. Rand

Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA

Received: 6 January 2003 / Accepted: 28 February 2003

Abstract. The semen of many primate species coagulates into a mating plug believed to prevent the sperm of subsequent mating events from accessing the ova. The texture of the coagulum varies among species: from a semisoft mass in humans to a firm plug in chimpanzees. In humans, a component of the coagulum, semenogelin I, also inhibits sperm motility. We tested the hypothesis that polymorphism and divergence at semenogelin I differ among hominoid species with different mating systems. Sequence data for the semenogelin I locus were obtained from 12 humans, 10 chimpanzees, 7 gorillas, and 1 bonobo. Mitochondrial D-loop data were collected from a subset of individuals to assess levels of variation at an unlinked locus. HKA tests using D-loop sequence data revealed a significant reduction of polymorphism at semenogelin I in chimpanzees, consistent with predictions of a selective sweep at this locus. This result was supported by independent HKA tests using polymorphism data from a putatively neutral locus from the literature. Humans show a similar trend toward reduced polymorphism, although HKA tests were only marginally significant. Gorilla sequence data show evidence of functional loss at the semenogelin I locus, indicated by stop codons within the putative open reading frame as well as high levels of polymorphism. Elevated K_a/K_s ratios within the *Pan-Homo* clade suggest a history of positive selec-

tion at semenogelin I. Our results suggest that there is a positive relationship between the intensity of sperm competition in a species and the strength of positive Darwinian selection on the seminal protein semenogelin I.

Key words: Sperm competition — Mating system — Polyandry — Apes — Humans — Hominoids — Semenogelin — Prostate-specific antigen — Positive selection — Semen

Introduction

Sperm competition is a potent evolutionary force mediating male–male competition (Parker 1970; Smith 1984; Birkhead and Moller 1998). This component of sexual selection has been studied in many taxa and is found to be “nearly ubiquitous” in the animal kingdom (Smith 1998). Data collected over the past 20 years indicate a role for sperm competition in polyandrous primate taxa. Harcourt et al. (1981) found an association between mating system and relative testis size, where primate genera with multimale mating systems have significantly heavier testes relative to their body weight than do polygynous or monogamous genera. More recently, Anderson and Dixson (2002) found that at both the genus and the species level, primates with multimale mating systems have significantly larger spermatozoa midpiece volumes than do primates with single-male mating systems. This relationship may indicate selection for increased mitochondrial loading and

Correspondence to: Sarah B. Kingan, Genomic Analysis and Technology Core, Biosciences West, Room 246B, University of Arizona, Tucson, AZ 85721, USA; email: Sarah_Kingan@alumni.brown.edu

greater sperm motility in species with sperm competition (Anderson and Dixson 2002).

Many reproductive traits such as genital morphology (Eberhard 1985) and proteins involved in sperm competition and gamete recognition (Swanson and Vacquier 2002) can evolve rapidly through positive Darwinian selection. Notable among these rapidly evolving reproductive proteins are several *Drosophila* accessory gland proteins (review by Chapman 2001; see also Tsaur et al. 1998, 2001; Tsaur and Wu 1997; Aguadé 1999; Begun et al. 2000). Among marine invertebrates, the abalone sperm protein lysin (Hellberg and Vacquier 1999; Yang et al. 2000) and sea urchin sperm protein bindin (Metz and Palumbi 1996; Biermann 1998) evolve through positive selection. Swanson et al. (2001b) demonstrated positive selection on mammalian zona pellucida egg coat proteins as well as oviductal glycoprotein. More recently, Torgerson et al. (2002) demonstrated that sperm-specific proteins have higher rates of nonsynonymous divergence between humans and mice than do proteins expressed in other tissues. In primates, genes affecting sperm morphology, protamine 1 and 2, appear to evolve under positive selection (Wyckoff et al. 2000).

The semen of many primate species coagulates into a mating plug upon ejaculation (Roussel and Austin 1967). The human plug is gelatinous in texture, but a more firm plug forms in chimpanzees (Dixson and Mundy 1994) and rhesus monkeys (van Pelt and Keyser 1970). The human coagulum derives from the seminal vesicles and is inseminated after the liquid portion, which contains prostate and Cowper's gland secretions as well as concentrated spermatozoa (Marson et al. 1989; Tauber et al. 1975; van Pelt and Keyser 1970). The human semen coagulum is composed of semenogelin I and II (SgI and SgII) [Peter et al. 1998], which are encoded by neighboring genes in the q12–q13.2 region of chromosome 20 (NCBI Locus Link). In addition to its coagulating properties, human SgI inhibits sperm motility (Robert and Gagnon 1999) as well as sperm capacitation (de Lamirande et al. 2001). Following ejaculation *in vitro*, both human semenogelins are degraded by prostate-specific antigen (PSA), a protease that is synthesized in the prostate and present in the liquid portion of the ejaculate (Lilja 1985). As PSA cleaves SgI, the coagulum liquefies and any spermatozoa trapped in the plug are released and become motile (Robert et al. 1997).

Figure 1 schematically models the spatial distribution within the vagina of ejaculates from sequential matings by two males. Temporal separation of the liquid and coagulating components may stratify the respective ejaculates of the primary and secondary mating and prevent their mixing except at the interface. In this model, the semenogelin coagulum of the first male forms a physical barrier to the sperm of the

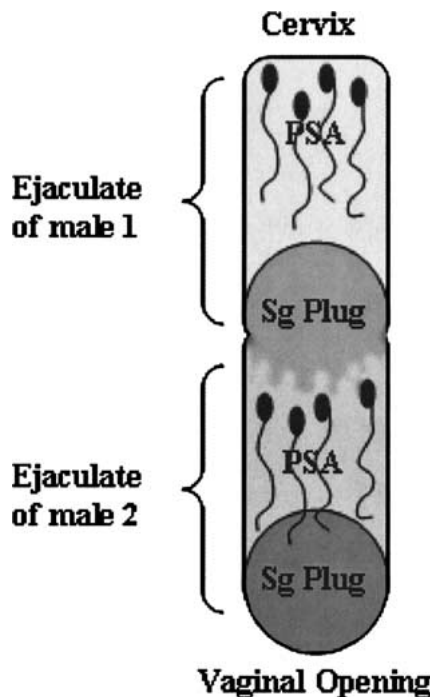


Fig. 1. Schematic of vagina containing ejaculates from two sequential matings. The liquid portion (containing concentrated spermatozoa and PSA) is inseminated first and appears above the semenogelin plug in both ejaculates. The plug of male 1 may block the sperm of male 2 from accessing the ovum, while the PSA of male 2 may degrade the first male's plug.

second male as well as inhibiting sperm capacitation and movement when spermatozoa of the second male come into contact with the coagulum. PSA in part protects the self-sperm of a male from his own semenogelin plug and, when in a second position mating, degrades the plug of the previous male.

Both semenogelin genes are members of the REST family (rapidly evolving substrates for transglutaminase), which encode the coagulating proteins of rats, SVS-II, and guinea pigs, SVP-1 (Lundwall and Lazure 1995). Members of the REST gene family are characterized by three exons encoding the signal peptide, secreted protein, and 3' untranslated region, respectively. While the first and third exons are highly conserved, the second exon is rapidly evolving among mammalian taxa. Consequently, among rats and guinea pigs, the mature proteins encoded by members of the REST family bear little resemblance to one another (Lundwall and Lazure 1995).

In this report, we test hypotheses of neutral evolution at the seminal vesicle protein semenogelin I in four hominoid species. A primary question is whether the degree of polyandry in the mating systems of each species determines the strength of positive selection on this seminal protein presumed to directly mediate sperm competition. We aim to determine whether evidence for positive selection is strong in species with highly polyandrous behavior. Signatures of positive

selection include elevated K_a/K_s ratios resulting from adaptive evolution, elevated divergence between species, and, in the case of a selective sweep, reduced polymorphism.

To contrast primate mating systems, we evaluated sequence patterns in semenogelin I across chimpanzees, gorillas, bonobos and humans. Sperm competition is expected to be intense in both common and pygmy chimpanzees, where females routinely mate with several males during a single reproductive cycle (Dixson 1998). Gorilla females in contrast, live in harems with several other females and nearly exclusively mate with the single alpha male (Dixson 1998); for gorillas we expect little to no sperm competition. Human mating practices range from monogamy, to polygyny, to polyandry (Levinson and Ember 1996) and levels of sperm competition are likely to be intermediary.

We found reduced levels of polymorphism in the common chimpanzee semenogelin I gene, consistent with a selective sweep model of molecular evolution. Human polymorphism at semenogelin I also was reduced, although our data yielded only marginal significance. Gorilla sequences exhibited high levels of polymorphism, including a stop codon at intermediate frequency, suggestive of functional loss. Our analyses of nonsynonymous and synonymous variation suggest a history of adaptive evolution in the *Pan-Homo* clade compared with *Gorilla*. In sum, among hominoids the degree of polyandry appears to be associated with measures of positive selection on the seminal coagulating protein semenogelin I.

Materials and Methods

DNA Samples

Twelve human samples, including five of African descent, were collected at Brown University as buccal swabs. Informed consent was received in writing and all procedures were approved by the Brown University Internal Review Board. Ten unrelated chimpanzee (*Pan troglodytes*) DNA samples were donated by the Southwest National Primate Research Center. Accurate subspecies information is not available for these individuals, although pedigree records and our own D-loop sequence data indicate that the sample was composed of *P. t. troglodytes* and *P. t. verus* individuals as well as several hybrids. Seven unrelated western lowland gorilla DNA samples (*Gorilla gorilla gorilla*) and 1 bonobo (*Pan paniscus*) sample were provided by the Zoological Society of San Diego.

PCR Amplification

DNA amplifications were performed in 25- μ L PCR reactions containing 0.5 units of Taq polymerase in storage buffer B with 1 \times magnesium-free buffer (Promega), 200 mM of each dNTP, 10–20 ng of DNA, and 1–3 mM MgCl₂ (see Table 1 for primer sequences and MgCl₂ concentrations). The MJ Research Programmable Thermal Controller and PTC-100 were used, with the generic conditions of 95°C for 5 min, 30–40 cycles at 95°C for 30 s, λ °C

Table 1. Primer sequences and PCR conditions for amplification of SgI and the mitochondrial D-loop

SgI ^a	Upper primer	Sequence (5'–3')	Lower primer	Sequence (5'–3')	Species	MgCl ₂ (mM)	Anneal. temp. (°C)
	U734	CTTTGACTTCCCTGCTCCTC	L1282	CTCTGCCTTCTTGTTATTATG	Human	2.0	54
	U1141	GGCAGTTTTTCTATTCA	L1686	GTTGAAGTTTGCTCTGG	Chimp	1.0	54
	U1625	ATGTGGTTGAAGTGAGAG	L2344	CGAATGGTAGGTTTATGT	Gorilla	3.0	52
	U2319	TAACAACGACCGAAACC	L3055	CAAAGAGCGAGGGAAGC	Human	2.0	50
	U2319I	GACTTTTGTGGGATTGA	L3055I	GAGAGGGAGGACTGATAG	Chimp	2.0	50
	L15926 ¹	TCAAAGCTTACACCAGTCTTGTA	H16498 ¹	CCTGAAGTAGGAACCAGATG	Gorilla	1.5	48
D-Loop ^b	D-88 ²	CTCTGTTCTTTCATGGGGAAGC	D-441 ²	CGGGATAATTGATTTTCACGGAGG	Human	2.5	56
					Gorilla	2.5	56
					Chimp	2.0	62
						Sequencing primers	

^a Numbers for SgI primers refer to bases in published human semenogelin I gene: GenBank accession number M81650 (Lilja et al. 1989).

^b D-loop primer sequences and conditions previously published: ¹Di Rienzo and Wilson (1991); ²Morin et al. (1994).

(see Table 1) for 30 s, and 72°C for 1.5 min, followed by 72°C for 5 min. Reactions were visualized on an agarose gel. Most primer sets amplified all four species, with the exception of U1625–L2344, which only amplified humans. For some gorilla samples a specific version of primer L1282 was used, which had Y degenerate sites (C + T) at positions 17 and 20.

A second locus, the mitochondrial control region, or D-loop, was sequenced in a subset of the four species (eight humans, eight chimps, six gorillas, and one bonobo) to provide polymorphism data from an unlinked locus. Human and gorilla primers and conditions were published previously (Di Rienzo and Wilson 1991). Chimpanzee primers are from Morin et al. (1994). PCR conditions follow the same generic program as for semenogelin I reactions. See Table 1 for primer sequences, MgCl₂ concentrations, and annealing temperatures.

Cloning

Forward and reverse sequence reads of diploid PCR product were obtained from each individual. Heterozygous polymorphisms were confirmed by cloning individual alleles to obtain haploid sequences. We cloned PCR products using the TOPO TA Cloning Kit from Invitrogen. From positive bacterial colonies, we performed a boil prep and then a PCR using M13 primers per the manufacturer's instructions. Both alleles were inferred based on the forward and reverse sequence of one clone. The presence in the population of both alleles was confirmed for most polymorphism by individuals homozygous for each allele or a second heterozygous individual. Only two gorilla polymorphisms were inferred based on the diploid sequence of a single heterozygous individual. We constructed "artificial" full-length haplotypes based on free recombination of the four cloned haplotype sequences. The full-length gametic phase would not alter the interpretation of our results, as our analysis was based on the number of variable sites rather than linkage.

Sequencing

PCR products were cleaned using Qiagen's QIAquick PCR Purification kit. Twenty-microliter cycle sequencing reactions were run using ABI's Big Dye Terminator Cycle Sequencing Ready Reaction Kit with 10–20 ng of PCR product and 64 nmol of primer. Samples were run on an ABI 377 sequencer. Forward and reverse reads were obtained for all sequences except three human and six gorilla D-loop samples, where a string of Cs prohibited sequencing the entire amplicon. For these samples, two forward and two reverse reads were obtained for the regions flanking the poly(C) string.

Sequence Editing and Analysis

Sequencher for Macintosh v. 3.0 (Gene Codes Corporation 1995) and Bioedit (Hall 1999) were used to edit and align our data against a published human sequence (GenBank accession number M81650 [Lilja et al. 1989]). Levels of polymorphism and divergence were quantified using DnaSP (Rozas and Rozas 1999). The poly(C) region of the D-loop was excluded from our analysis. We calculated Tajima's *D* (TD), an analysis of the allele frequency spectrum, using DnaSP (Rozas and Rozas 1999). This statistic compares the two neutral estimators of nucleotide diversity, π (the average pairwise difference) and θ_w (diversity based on the number of segregating sites). A neutral locus in a panmictic, infinitely large population should have a TD of zero. A positive TD indicates balancing selection, population structure, or population contraction, whereas a negative value indicates positive directional selection or population expansion. Two neutrality tests were performed. The McDonald–Kreitman (1991) test (MK test) compares the ratios of synonymous

to nonsynonymous differences within and between species, testing the neutral assumption of equal ratios. These tests were performed by hand using a 2×2 *G* test for heterogeneity. The Hudson–Kreitman–Aguadé (1987) test (HKA test) compares polymorphism and divergence at two or more loci, testing the assumption that the ratio of polymorphism to divergence should be equal across all neutral loci. We used semenogelin I and mitochondrial D-loop sequences from the same individuals to compare patterns of DNA variation at two unlinked loci. The model was run using the freeware HKA (Hey 2001). Due to concerns over mutational saturation and interspecific sequence alignment at the D-loop, additional HKA tests were run using SgI and HOXB6 (Deinard and Kidd 1999). HOXB6 is a putatively neutral locus in a region of moderate recombination (Nachman 2001) for which polymorphism data from all four species were available and similar population sampling strategies were employed. We used the maximum likelihood software, PAML (Yang 1997) to analyze our data for lineage-specific elevation of K_a/K_s ratios as well as selection on particular nucleotide positions with the program Codeml. For the lineage-specific analysis we used the model M2, which allows different K_a/K_s ratios for each branch, with the NSSites parameter set to 0 (one K_a/K_s for all sequence sites). We compared the likelihood of this model to the null: M0 (one K_a/K_s for the entire tree) with NSSites = 0. For the site-specific analysis we ran M0 with NSSites = 2 (variable K_a/K_s across sequence sites, allowing for selection), compared to the null model of M0 with NSSites equal to 0 (one K_a/K_s for all sites) and 1 (variable K_a/K_s across sites, neutral evolution).

Results

Region Sequenced

In nine humans 2190 bases were sequenced; in the remaining three humans, 1561 bases were sequenced. In all 10 chimpanzees 1585 bases were sequenced. In one bonobo, 958 bases were sequenced. In three gorillas 1629 bases were sequenced, in three other gorillas 912 bases were sequenced, and in the remaining gorilla 563 bases were sequenced. See Fig. 2 for data in regions sampled. The repetitive nature of SgI (Robert and Gagnon 1999) made sequencing regions of the second exon difficult in several samples, and a region of exon 2 comprising 629 bases was not successfully amplified or sequenced in chimps, bonobos, gorillas, or three humans. Sequences have been submitted to GenBank under the accession numbers AY174423–AY174458.

Polymorphism and Divergence

Polymorphism in gorillas was substantially higher than in either chimpanzees or humans (see Table 2). Gorillas had a total of 20 polymorphic sites: 10 replacement, 5 synonymous, and 5 intronic. The human sample contained five polymorphisms: one replacement and four intronic. Chimps had one polymorphic site: an intronic singleton. Additionally, the gorilla sample contained one polymorphic insertion–deletion in intron 2 at a frequency of 0.33 (based on diploid sequence) and a second insertion–deletion in intron 2

Table 2. Polymorphism at semenogelin I^a

	Number of sites				π	θ	TD	AA poly.	K_a/K_s
	Total	Rep.	Syn.	Intron					
Human	5	1	0	4	0.00096	0.00086	0.349	0.234	NA
Chimp	1	0	0	1	0.00006	0.00018	-1.164	0	NA
Gorilla	20	10	5	5	0.00781	0.00617	0.865	1.610	0.72

^a Rep., replacement; syn., synonymous. π is the average number of nucleotide differences per site between two sequences (Nei 1987); θ is based on the number of segregating sites in the sample (Nei 1987). TD, Tajima's D; AA poly., percentage amino acid

polymorphism. K_a/K_s is the ratio of the number of nonsynonymous changes per site to the number of synonymous changes per site. Data for AA poly. and K_a/K_s are the average for all pairwise intraspecific comparisons.

Table 3. Divergence at semenogelin I^a

	Number of sites				K_a (ave.)	K_s (ave.)	K_a/K_s (ave.)	AA div. (ave.)
	Total	Rep.	Syn.	Intron				
Human–chimp	12.09	3.63	1.00	7.46	0.71 (0.80)	0.70 (1.11)	1.01 (0.84)	1.65 (1.34)
Human–bonobo	7.75	4.63	1.00	2.12	0.90	0.70	1.28	2.10
Human–gorilla	19.70	7.70	3.33	8.67	1.59 (0.93)	2.52 (1.48)	0.66 (0.82)	3.70 (1.58)
Chimp–bonobo	1.00	1.00	0	0	0.19	0	NA	0.45
Chimp–gorilla	18.50	6.00	3.33	9.17	1.29 (0.90)	2.53 (1.64)	0.56 (0.69)	2.74 (1.65)
Gorilla–bonobo	14.16	7.33	3.33	3.50	1.46	2.36	0.68	3.11

^a Number of sites is the average of all pairwise comparisons; rep., replacement; syn., synonymous. K_a the number of nonsynonymous changes per site; K_s is the number of synonymous changes per site; AA div., average amino acid divergence for all pairwise compar-

isons. K_a , K_s , and AA div. are reported as percentages. Values in parentheses are the averages across many genes from Chen and Li (2001); not all comparisons were available.

fixed relative to the other species. Many gorilla polymorphisms occurred at a high frequency (see Fig. 2): Only 2 of the 20 polymorphic sites had allele frequencies lower than 0.10, and 6 of the 20 sites had allele frequencies lower than 0.20. Surprisingly, we found a high-frequency (0.50) stop codon 15 bases into exon 2 in gorillas. Two individuals were homozygous for the stop codon, and two were heterozygous. This stop codon was observed in an independent study (Jensen-Seaman, personal communication).

We measured divergence as the average of all pairwise K_a and K_s values, the number of nonsynonymous and synonymous differences per site, respectively (see Table 3). Gorillas were the most divergent species and were similarly divergent from humans, chimpanzees, and bonobos. K_a values ranged from 1.29% for chimp–gorilla to 1.59% for human–gorilla. K_s values ranged from 2.36% for bonobo–gorilla to 2.53% for chimp–gorilla. At synonymous sites, humans were equally divergent from both *Pan* species ($K_s = 0.70\%$) but were more divergent from bonobos than chimps at nonsynonymous sites ($K_a = 0.90\%$ and $K_a = 0.71\%$, respectively). Divergence between the two chimpanzee species was slight: only a single replacement site ($K_a = 0.19\%$, $K_s = 0$).

In terms of percentage amino acid divergence, gorillas were again the most divergent hominoid. The highest amino acid divergence in our data set was between humans and gorillas (3.70%). Amino acid divergence between gorillas and bonobos was higher than between gorillas and chimps (3.11% and 2.74%, respectively). Humans were more divergent from bonobos than chimps (2.10 and 1.65%, respectively) and chimp–bonobo divergence was the lowest (0.45%).

Synonymous and Nonsynonymous Variation

Neither intraspecific nor interspecific K_a/K_s ratios showed a significant departure from the neutral prediction of 1.0 (Kumar et al. 1993; see Tables 2 and 3). Intraspecific K_a/K_s ratios could not be calculated for humans or chimpanzees because humans had no synonymous polymorphism and chimps had no polymorphism in coding regions. The intraspecific gorilla K_a/K_s was 0.72, slightly higher than the interspecific values comparing gorillas with humans (0.66), chimps (0.56), and bonobos (0.68). Interestingly, human–chimp (1.01) and human–bonobo (1.28) K_a/K_s values were higher than the values obtained when either species was compared with gorillas, although these values did not show significant

Table 4. Data for HKA tests^a

	S_{A1}	S_{A2}	S_{B1}	S_{B2}	D_A	D_B	p
SgI–D-loop							
H–C	5 (5.38)	1 (9.54)	26 (25.62)	74 (65.46)	12.09 (3.17)	59.61 (68.53)	<0.0005*
H–G	5 (14.18)	20 (14.14)	26 (16.82)	18 (23.86)	19.70 (16.39)	88.46 (91.77)	NS
H–B	5 (9.62)	NA	26 (21.38)	NA	7.75 (3.13)	67.50 (72.12)	<0.05
C–G	1 (18.62)	20 (9.53)	74 (56.38)	18 (28.47)	18.60 (11.44)	91.79 (98.95)	<0.005*
C–B	1 (1.68)	NA	74 (73.32)	NA	1.00 (0.32)	47.38 (48.06)	NS
G–B	20 (19.37)	NA	18 (18.63)	NA	14.16 (14.79)	105.17 (104.54)	NS
SgI–HOXB6							
H–C	5 (6.03)	1 (4.48)	4 (2.97)	8 (4.52)	12.09 (7.58)	6.13 (10.64)	<0.05
H–G	5 (7.30)	20 (17.80)	4 (1.70)	6 (8.20)	19.70 (19.60)	13.00 (13.1)	NS
H–B	5 (6.78)	NA	4 (2.22)	NA	7.75 (5.97)	7.80 (9.58)	NS
C–G	1 (5.59)	20 (16.38)	8 (3.41)	6 (9.62)	18.50 (17.53)	13.88 (14.84)	<0.05
C–B	1 (1.63)	NA	8 (7.37)	NA	1.00 (0.37)	3.38 (4.00)	NS
G–B	20 (18.67)	NA	6 (7.33)	NA	14.16 (15.49)	15.67 (14.34)	NS

^a Two sets of HKA tests were performed comparing data at semenogelin I (SgI) to the mitochondrial D-loop and to the HOXB6 intergenic region. The two loci for each set of tests are referred to as A and B; the two species, as 1 and 2. S is the number of segregating sites; D is the number of divergent sites. S_{A1} is the number of segregating sites at locus A for species 1. For example, the first entry in the table shows the number of segregating sites at SgI in

humans. Observed data are followed by expected values in parentheses. p is the p value for each test. NS, nonsignificant at the $p < 0.05$ level. Tests that remain significant after correcting for multiple tests are indicated by an asterisk, after the p value. Data that show significant deviations from expectations are in bold face. H, human; C, chimp; G, gorilla; B, bonobo.

departures from neutrality. Chimp–bonobo K_a/K_s ratios could not be calculated due to a lack of synonymous divergence.

Neutrality Tests

HKA tests comparing nucleotide variation of the same samples at SgI and the mitochondrial D-loop consistently showed reduced chimp polymorphism at semenogelin I (see Table 4). When chimps were compared with humans and gorillas significant departures from neutrality at semenogelin I were observed. These tests remained significant after correcting for multiple tests. For both tests, chimp polymorphism was lower than expected. Human–chimp and chimp–gorilla divergence at SgI were higher than expected. Also, gorilla polymorphism was higher than expected compared with that of chimps. Chimp–bonobo tests did not show a significant departure from neutrality. This was likely due to the low number of segregating sites in chimpanzees, lack of polymorphism data in bonobos, and slight divergence between the two species. The human–bonobo test at these two loci was individually significant, although it was not significant after correcting for multiple tests. The test showed higher than expected divergence between humans and bonobos at SgI and reduced human polymorphism at SgI. For all significant tests, deviation at the D-loop was slight compared to deviation at the semenogelin I locus.

A second set of tests was performed comparing patterns of variation at semenogelin I in our population sample to data from the HOXB6 locus (Dei-

nard and Kidd 1999; see Table 4). Similar patterns were observed in both sets of tests: Lower than expected chimp polymorphism and elevated divergence between chimp and human at SgI were consistently found across both sets of tests. However, neither elevated gorilla polymorphism nor elevated chimp–gorilla divergence at SgI was observed when the locus was compared to HOXB6. While the two tests comparing chimps to humans and gorillas were individually significant, after correcting for multiple tests they were not significant.

Neither MK tests (McDonald and Kreitman 1991) nor Tajima's D (TD) showed any significant departure from neutrality within any species at SgI (MK results not shown; see Table 2 for TD values). The chimpanzee TD value was negative (–1.164), whereas both the human and the gorilla TDs were slightly positive (0.349 and 0.865, respectively). Lineage-specific elevation of K_a/K_s ratios and positively selected sites were evaluated by maximum likelihood with PAML (Yang 1997). All tests were nonsignificant relative to models with one K_a/K_s ratio for the entire tree or all sites. The power of this test, however, was low due to the limited number of variable sites.

Discussion

Our survey of polymorphism and divergence at semenogelin I in several hominoids revealed a reduction of polymorphism in chimpanzees consistent with a selective sweep model of molecular evolution. There was a nonsignificant reduction of variation in humans, while gorilla sequences appeared to be evolving neutrally. This pattern is consistent with stronger selection

on semenogelin I in species with higher degrees of polyandry, however, data from more species would provide further support for this intriguing relationship.

Our first set of HKA tests (see Table 4) compared polymorphism to divergence at SgI and the mitochondrial D-loop in the same individuals. Tests revealed significantly reduced chimp polymorphism, elevated chimp–human divergence, elevated chimp–gorilla divergence, and elevated gorilla polymorphism at SgI. Chimp–human and chimp–gorilla tests remained significant after correcting for multiple tests. Our findings of reduced chimpanzee polymorphism and elevated human–chimp divergence at SgI were supported by the results of additional HKA tests which used previously published data from HOXB6 (Deinard and Kidd 1999) rather than the D-loop. We used this data set because HOXB6 is an autosomal locus in a region of moderate recombination and the chimpanzee sampling strategy of the authors was similar to ours, comprising different subspecies (Deinard, personal communication). We sought to avoid comparing our results to a locus in a low recombinational regime to avoid the effects of background selection and hitchhiking on patterns of nucleotide variation (Nachman 2001). Because both data sets pooled across chimp subspecies, we could minimize the biasing effects of population structure on the HKA tests.

We chose to confirm the results of our first set of HKA tests with a second putatively neutral locus for two reasons. First, deviations from neutrality have been observed in mitochondrial loci (Hey 1997; Nachman et al. 1996) indicating that the D-loop is not an ideal neutral locus for comparison in selection studies. In addition, mutational saturation and difficulties with alignment of gorilla sequences to those of the other three species called into question the divergence values we found for D-loop. All HKA tests that compared chimps to humans and gorillas showed significant departures from neutral expectation in the direction of reduced chimp polymorphism at SgI. (However, the only tests that remained significant after correcting for multiple tests were the chimp–human and chimp–gorilla HKA tests comparing SgI to the D-loop.) Therefore, our result of reduced chimp polymorphism and elevated chimp–human divergence at SgI is robust across two putatively neutral loci. In addition, when levels of polymorphism at semenogelin I are compared to those at two other putatively neutral loci (Xq13.3 and 16s rRNA [Kaessmann et al. 1999; Noda et al. 2001]), in regimes of low or no recombination where nucleotide variability could be reduced by selection on linked sites, the reduction of chimpanzee polymorphism at SgI is still apparent. The consistent rejection of neutrality for tests with chimp as well as the high number of individually significant tests (42%) suggests that

the occurrence of significant HKA tests is not due to the number of tests conducted.

The D-loop sequence data we collected allowed us to reject pedigree or sampling effects as the cause of reduced chimp polymorphism at SgI. The chimp sample showed the highest levels of polymorphism in the D-loop of the three species surveyed (74 sites among 354 bases, $\pi = 0.077$, $\theta = 0.084$). Our sample of chimps comprises purebred and hybrid individuals from at least two subspecies (*Pan troglodytes troglodytes* and *P. t. verus*) as determined by information provided by the Southwest National Primate Research Center and analysis of the D-loop sequence (data not shown). In light of the high diversity at the mitochondrial D-loop, the presence of only one segregating site at semenogelin I in the chimpanzee sample is all the more remarkable.

Analyses of nonsynonymous and synonymous nucleotide variation suggest that divergent evolutionary forces may be acting across the *Gorilla* and *Pan–Homo* clades. K_a/K_s ratios across the sample of alleles in the *Pan–Homo* clade are higher than intraspecific and interspecific values for gorilla. In addition, we find that both human–chimp and human–bonobo K_a/K_s estimates for SgI are higher than estimates based on surveys of primate nuclear genes (Chen and Li 2001; Chen et al. 2001; Ohta 1995). Chen et al. (2001) described chimp–human divergence and reported an average K_a and K_s across 88 genes which give a ratio of 0.48. For the subset of 15 duplicate genes, such as SgI and SgII, Chen et al. (2001) estimate a K_a and K_s that give a ratio of 0.56. We estimate the human–chimp K_a/K_s at SgI to be 1.01 ± 0.14 (mean \pm SD), which is higher than 97.5% of the values for all the genes reported by Chen et al. (2001). The relatively high values of SgI K_a/K_s in the *Pan–Homo* clade, rather than their absolute magnitude, suggest that this locus has differentially evolved among taxa of this clade. The substitution patterns are consistent with positive selection in a background of purifying selection, as seen in several mammalian reproductive proteins (Swanson et al. 2001b). However, the power to test this hypothesis formally based on K_a/K_s is limited because of the low number of variable sites among our samples. Sequence data from more divergent species are needed.

Though not significant, the trend observed in TD values also indicates divergent evolutionary forces in each taxon. The chimpanzee TD is negative (-1.164), whereas both the human and the gorilla TD are slightly positive (0.349 and 0.865, respectively). The negative chimp TD is consistent with recovery from selective sweep, although the power of the test, based on only one singleton, is very low. The human TD is close to zero so does not indicate either strong demographic or selective effects on the locus and is rather inconclusive. The gorilla TD, the largest value

in our data set, reflects the high levels of intermediate-frequency variants and could be suggestive of population contraction of the Western lowland gorilla subspecies (*Gorilla gorilla gorilla*). However, this scenario is not consistent with the TD value for the D-loop, which is nearly zero (TD = -0.013).

Amino acid and nucleotide divergence (K_a and K_s) between human and chimp at semenogelin I show no clear trend compared to the averages across many genes reported by Chen and Li (2001) for several hominoid species. Amino acid percentage divergence is slightly higher at SgI, while both synonymous (K_s) and nonsynonymous divergence (K_a) at SgI are lower. To understand better the amino acid variation present in humans and chimpanzees, we categorized the nature of the amino acid substitutions in the *Pan-Homo* clade according to the classification developed by Li et al. (1984). (We did not perform this analysis for gorillas because of the questionable function of the gene in this species.) Based on Grantham's (1974) physiochemical distance between amino acid pairs, Li et al. classified amino acid substitutions as conservative, moderately conservative, moderately radical, and radical. In the human lineage, the fixed mutation at site 554 (Glu→Gln) was conservative, the polymorphism at site 500 (Thr→Ser) was moderately conservative, and the fixed mutation at site 372 (Ser→Phe) was radical. (The mutation at site 388 was synonymous.) Bonobos differ from all species at site 537 (His→Pro), a moderately conservative change. A radical change occurred in the lineage leading to both chimpanzees at site 677 (Gly→Trp). The amino acid changes observed in the human and chimp lineages are not biased toward conservative changes as expected under neutral evolution and the presence of two radical changes is consistent with selection for new forms of semenogelin I in this clade.

Gorillas present a distinctly different pattern of nucleotide variation at semenogelin I relative to chimps and humans. In fact, many segregating alleles of gorilla SgI appear to be nonfunctional as evidenced by a stop codon in the second exon. The locus shows high polymorphism (20 sites in all, including the stop codon at 50% frequency), high amino acid and nucleotide divergence from other species, and no departure from neutrality based on the HKA tests. All measures of divergence from both chimp and human (K_a , K_s , and percentage amino acid divergence) are higher than 97.5% of the values reported by Chen and Li (2001) for 37 nuclear genes. However, while loss of function is often signified by elevated K_a/K_s ratios characteristic of relaxation of selection, both intraspecific and interspecific K_a/K_s ratios for gorillas are consistent with the average values reported by Chen and Li (2001). Potentially, polymorphism in the remaining functional alleles is limited in gorillas due to a structural constraint on

the protein, as may occur in the *Pan-Homo* clade where the absolute magnitude of K_a/K_s is modest despite elevation relative to gorilla.

The polymorphic stop codon in gorillas appears to be relatively young. If we assume that the stop codon is neutral and is halfway to fixation by drift (which would take $4N_e$ generations, where N_e is the effective population size), we estimate the allele to be only $(1/2)(4)(20,000)(15) = 600,000$ years old, assuming a generation time of 15 years (Chen and Li 2001) and an effective population size of 20,000 (Jensen-Seaman et al. 2001). The loss of function is probably an allele effect that arose after the gorilla lineage split from that leading to humans and chimpanzees 6.3 to 8.5 million years ago (Chen and Li 2001). If selection was relaxed early on in the lineage leading to gorillas, we would expect to see a much higher level of nucleotide diversity, similar to the level of a pseudogene. The average Jukes-Cantor distance for SgI between humans and gorillas is 1.31% and that between chimps and gorillas is 1.19%; both estimates are lower than all the values reported by Chen and Li (2001) for seven pseudogenes, which range from 1.49 to 3.62% between humans and gorillas and 1.84 to 4.27% between chimps and gorillas. This result implies a relatively recent loss of function or the retention of some function despite the stop codon. It is possible that polyandrous mating behavior was prevalent in the common ancestor of humans, chimpanzees, and gorillas and that polygyny recently arose in the lineage leading to gorillas.

We have demonstrated that chimpanzees, the most polyandrous hominoid, show a significant departure from neutral evolution at a protein believed to play an important role in sperm competition, semenogelin I. In gorillas, a harem polygynous ape presumed to have little or no sperm competition, the gene appears to be subjected to less functional constraint, indicating that it may play a limited role in sperm competition. Predicting the expected intensity of sperm competition in ancestral *Homo* is controversial. Based on regression of testis weight to mating system, Harcourt et al. (1981) estimated the intensity of sperm competition in humans to be intermediate to that of chimps and gorillas. Here we find patterns of nucleotide variability at SgI in humans to resemble more closely the patterns seen in chimps than in gorillas. Overall, across these hominoid species we detect a trend between the degree of polyandry and the strength of positive selection on the seminal coagulating protein, semenogelin I. Further sequencing of the SgI locus, in addition to semenogelin II, which encodes the other coagulating protein of semen, as well as the protease of the semenogelins, prostate specific antigen (PSA), may help elucidate the effects of sperm competition on the evolution of reproductive proteins in primates. Sequence data from more

divergent taxa, including the socially monogamous gibbon (*Hylobates* spp.), as well as a polygynous and polyandrous baboon species (*Papio* spp.), would further increase our statistical power to test patterns of selection based on polymorphism and divergence.

Acknowledgments. S.B.K. would like to thank Oliver Ryder and Leona Chemnick at the Center for Reproduction of Endangered Species/Zoological Society of San Diego as well as Karen Rice and Mary Sparks at the Southwest National Primate Research Center for providing valuable DNA samples. Michael Palmer, Lea Sheldahl, Faye Lemieux, and Raymond Qwan provided support in conducting this research. Michael Hammer, Michael Nachman, Tasha Altheide, Jason Wilder, and two anonymous reviewers provided helpful comments on the manuscript. This research was funded by the Royce Fellowship for Undergraduate Research at Brown University. Additional support was provided by a grant to M.T. from the NIH (AG16632) and grants to D.M.R. from the NSF (DEB9981497 and DEB0108500).

References

- Aguadé M (1999) Positive selection drives the evolution of the Acp29AB accessory gland protein in *Drosophila*. *Genetics* 152:543–551
- Anderson MJ, Dixon AF (2002) Sperm competition: Motility and the midpiece in primates. *Nature* 416:496
- Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG (2000) Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics* 156:1879–1888
- Biermann CH (1998) The molecular evolution of sperm bindin in six species of sea urchins (*Echinoidea: Strongylocentrotidae*). *Mol Biol Evol* 15:1761–1771
- Birkhead TR, Moller AP (eds) (1998) Sperm competition and sexual selection. Academic Press, San Diego
- Chapman T (2001) Seminal fluid-mediated fitness traits in *Drosophila*. *Heredity* 87:511–521
- Chen F-C, Li W-H (2001) Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet* 68:444–456
- Chen FC, Vallender EJ, Wang H, Tzeng CS, Li WH (2001) Genomic divergence between human and chimpanzee estimated from large-scale alignments of genomic sequences. *J Hered* 92:481–489
- Deinard A, Kidd K (1999) Evolution of a HOXB6 intergenic region within the great apes and humans. *J Hum Evol* 36:687–703
- de Lamirande E, Yoshida K, Yoshiike TM, Iwamoto T, Gagnon C (2001) Semenogelin, the main protein of semen coagulum, inhibits human sperm capacitation by interfering with the superoxide anion generated during this process. *J Androl* 22:672–679
- Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci USA* 88:1597–1601
- Dixon AF (1998) Primate sexuality: Comparative studies of the prosimians, monkeys, apes, and human beings. Oxford University Press, Oxford, pp 22–50
- Dixon AF, Mundy NI (1994) Sexual behavior, sexual swelling, and penile evolution in chimpanzees (*Pan troglodytes*). *Arch Sex Behav* 23:267–280
- Eberhard WG (1985) Sexual selection and animal genitalia. Harvard University Press, Cambridge, MA, pp 2–3
- Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science* 185:862–864
- Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
- Harcourt AH, Harvey PH, Larson SG, Short RV (1981) Testis weight, body weight and breeding system in primates. *Nature* 293:55–57
- Hellberg ME, Vacquier VD (1999) Rapid evolution of fertilization selectivity and lysine cDNA sequences in teguline gastropods. *Mol Biol Evol* 16:839–848
- Hey J (1997) Mitochondrial and nuclear genes present conflicting portraits of human origins. *Mol Biol Evol* 14:166–172
- Hey J (2001) HKA: A computer program for tests of natural selection (<http://lifesci.rutgers.edu/~heylab/>)
- Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159
- Jensen-Seaman MI, Deinard AS, Kidd KK (2001) Modern African ape populations as genetic and demographic models of the last common ancestors of humans, chimpanzees, and gorillas. *J Hered* 92:475–480
- Kaessmann H, Wiebe V, Paabo S (1999) Extensive nuclear DNA sequence diversity among chimpanzees. *Science* 286:1159–1162
- Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetic analysis, version 1.01. The Pennsylvania State University, University Park, pp 24–26
- Levinson D, Ember M (eds) (1996) Encyclopedia of cultural anthropology. Henry Holt and Company, New York
- Li WH, Wu CI, Luo CC (1984) Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J Mol Evol* 21:58–71
- Lilja H (1985) A kallikrein-like serine protease in prostatic fluid cleaves the predominant seminal vesicle protein. *J Clin Invest* 76:1899–1903
- Lilja H, Abrahamsson P-A, Lundwall A (1989) Semenogelin, the predominant protein in human semen. *J Biol Chem* 264:1894–1900
- Lundwall A, Lazure C (1995) A novel gene family encoding proteins with highly differing structure because of a rapidly evolving exon. *FEBS Lett* 374:53–56
- Marson J, Gervais D, Meuris S, Cooper RW, Jouannet P (1989) Influence of ejaculation frequency on semen characteristics in chimpanzees (*Pan troglodytes*). *J Reprod Fertil* 85:43–50
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654
- Metz EC, Palumbi SR (1996) Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol Biol Evol* 13:397–406
- Morin PA, Moore JJ, Chakraborty R, Jin L, Goodall J, Woodruff DS (1994) Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* 265:1193–1201
- Nachman MW (2001) Single nucleotide polymorphism and recombination rate in humans. *Trends Genet* 17:481–485
- Nachman MW, Brown WM, Stoneking M, Aquadro CF (1996) Nonneutral mitochondrial DNA variation in humans and chimpanzees. *Genetics* 142:953–963
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
- Noda R, Kim CG, Takenaka O, Ferrell RE, Tanoue T, Hayasaka I, Ueda S, Ishida T, Saitou N (2001) Mitochondrial 16S rRNA sequence diversity of hominoids. *J Hered* 92:490–496
- Ohta T (1995) Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J Mol Evol* 40:56–63

- Parker GA (1970) Sperm competition and its evolutionary consequences in the insects. *Biol Rev* 45:525–567
- Peter A, Lilja H, Lundwall A, Malm J (1998) Semenogelin I and semenogelin II, the major gel-forming proteins in human semen, are substrates for transglutaminase. *Eur J Biochem* 252:216–221
- Robert M, Gagnon C (1999) Semenogelin I: A coagulum forming, multifunctional seminal vesicle protein. *Cell Mol Life Sci* 55:944–960
- Robert M, Gibbs BF, Jacobson E, Gagnon C (1997) Characterization of prostate-specific antigen proteolytic activity on its major physiological substrate, the sperm motility inhibitor precursor/semenogelin I. *Biochemistry (Wash)* 36:3811–3819
- Roussel JD, Austin CR (1967) Enzymatic liquefaction of primate semen. *Int J Fertil* 12:288–290
- Rozas J, Rozas R (1999) DnaSP version 3: An integrated program for molecular populations genetics and molecular evolution analysis. *Bioinformatics* 15:174–175
- Smith RL (ed) (1984) Sperm competition and the evolution of animal mating systems. Academic Press, London
- Smith RL (1998) Foreword. In: Birkhead TR, Moller AP (eds) Sperm competition and sexual selection. Academic Press, San Diego
- Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. *Nat Rev Genet* 3:137–144
- Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF (2001a) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc Natl Acad Sci USA* 98:7375–7379
- Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001b) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci USA* 98:2509–2514
- Tauber PF, Zaneveld LJ, Propping D, Schumacher GF (1975) Components of human split ejaculates. I. *J Reprod Fertil* 43:249–267
- Torgerson DG, Kulathinal RJ, Singh RS (2002) Mammalian sperm proteins are rapidly evolving: Evidence of positive selection in functionally diverse genes. *Mol Biol Evol* 19:1973–1980
- Tsaur S-C, Wu C-I (1997) Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of *Drosophila*. *Mol Biol Evol* 14:544–549
- Tsaur S-C, Ting C-T, Wu C-I (1998) Positive selection driving the evolution of a gene of male reproduction, Acp26Aa, of *Drosophila*: II. Divergence versus polymorphism. *Mol Biol Evol* 15:1040–1046
- Tsaur S-C, Ting C-T, Wu C-I (2001) Sex in *Drosophila mauritiana*: A very high level of amino acid polymorphism in a male reproductive protein gene, Acp26Aa. *Mol Biol Evol* 18:22–26
- van Pelt LF, Keyser PE (1970) Observations on semen collection and quality in macaques. *Lab Anim Care* 20:726–733
- Wyckoff GJ, Wang W, Wu C-I (2000) Rapid evolution of male reproductive genes in the descent of man. *Nature* 403:304–309
- Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum Likelihood. *CABIOS* 13:555–556
- Yang Z, Swanson WJ, Vacquier VD (2000) Maximum-likelihood analysis of molecular adaptation in abalone sperm lysis reveals variable selective pressures among lineages and sites. *Mol Biol Evol* 17:1446–1455