PRIMER NOTE

Sequencing primers and SNPs for rapidly evolving reproductive loci in endangered ibex and their kin (Bovidae, *Capra* spp.)

STEVE JORDAN,*† CHRISTIAN MIQUEL,† PIERRE TABERLET† and GORDON LUIKART†‡

**Department of Biology, Bucknell University, Lewisburg, PA 17837, USA, †Laboratoire d'Ecologie Alpine, Génomique des Populations et Biodiversité, CNRS UMR 5553, Université Joseph Fourier, BP 53, F-38041 Grenoble, Cedex 9, France*

Abstract

Rapidly evolving genes (e.g. candidate selected loci) are of increasing interest to molecular ecologists and conservation geneticists. Here, we report primers for five regions from three independent nuclear reproductive genes that reliably generate polymorphic sequences across the widespread wild goats of the *Capra ibex* species group and likely many other species of bovids. From three to nine single-nucleotide polymorphisms (SNPs) were identified in each gene region among *C. ibex* subspecies. Average numbers of SNPs per 1000 bp across all five gene regions was 15.0, with a high of 25.3 in the *ZP3* exons 3 and 4 sequence and a low of 6.1 in the *TNP1* sequence.

Keywords: candidate genes, genomics, molecular adaptation, phylogenomics, phylogeny, selection, SNPs

Received 2 September 2005; revision received 30 January 2006; accepted 7 February 2006

The term 'ibex' refers to a group of morphologically similar, wild goat-like montane ungulates of the genus *Capra* that have been historically found from Spain to Mongolia, dipping into Pakistan, the Middle East and Ethiopia. Some ibex taxa (e.g. from northeast Africa) are endangered. Much debate has taken place over the taxonomic status of these widely separated populations of *Capra 'ibex'*, with arguments made both for lumping and splitting (reviewed in Manceau *et al*. 1999). Ibex taxonomy is based mainly on horn morphology in males, which is likely subject to selection and of limited taxonomic value (Manceau *et al*. 1999). For the time being, *Capra ibex* is considered to be one species, with several subspecies (Shackleton 1997). DNA sequence and single-nucleotide polymorphism (SNP) data are needed to help resolve species status among ibex and related wild ungulate species, of which nearly 70% are threatened with extinction (Shackleton 1997). SNPs are also needed within species for studies of population genetics and molecular adaptation.

Rapidly evolving loci are especially useful in molecular ecology as they are likely to be under selection (which can increase polymorphism within and especially among populations and taxa). Reproduction-related loci are interesting as they represent candidate 'speciation' genes, perhaps contributing to the development of reproductive isolation via reduced gamete–gamete recognition (Swanson & Vacquier 2002). These loci are also candidate-selected genes for polygamous species that might have sperm competition (e.g. ungulates). Candidate-selected loci are interesting as they can help elucidate the genetic basis of adaptation, fitness and sexual selection.

We have developed primer pairs to amplify and sequence five nuclear gene regions suitable for exploration of both the phylogeny of *Capra* and tests of selection and molecular adaptation. These regions, including parts of five exons and three introns of the *zona pellucida 2* (*ZP2*) gene, three exons and two introns of the *zona pellucida 3* (*ZP3*) gene and two exons and one intron of the *transition protein 1* (*TNP1*) gene, are among the fastest evolving mammalian proteins (Swanson & Vacquier 2002; Torgerson *et al*. 2002). The *ZP2* and *ZP3* genes code for glycoproteins that compose (in

Correspondence: Steve Jordan, Fax: 570 577-3537; E-mail: sdjordan@bucknell.edu

‡Present address: Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA, and CIBIO, Centro de Investigação em Biodiversidade e Recursos, Genéticos Campus Agrário de Vairão Universidade do Porto, 4485-601 Vairão, Portugal.

© 2006 The Authors
Journal Compilation © 2006 Blackwell Publishing Ltd

part) the zona pellucida (ZP), and are involved in the recognition and binding of sperm by the ZP (Swanson & Vacquier 2002). *TNP1* is a protein that replaces histones in mammalian spermatids during spermatogenesis (Yu *et al*. 2000). While the chromosomal location of *ZP2* is not known, *ZP3* and *TNP1* are respectively found on cattle chromosomes 25 and 2 (www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9913). These genes are all on different chromosomes in humans. We targeted exons in order to increase chances of identifying nucleotides and gene regions under selection. Exons allow for robust tests for selection signatures, such as an excess of nonsynonymous mutations.

Primer sequences and attributes are summarized in Table 1. All primers are intron spanning and anchored in exons, which are indicated by numbers in the primer names. For example, the *ZP2* exons 11 and 13 primer pair spans two introns. Fragments amplified by the two *ZP3* primer pairs overlap by 122 bp within exon 4. All of the primers were designed using the ICCARE website (Muller *et al*. 2004) in conjunction with the Primer3 website (Rozen & Skaletsky 2000). The ICCARE website allows users to compare ESTs of a research organism (in this case the cow, *Bos taurus*) with genomic sequence of *Homo sapiens*. An alignment was generated between these two genomes for each locus in question, and we then sought to situate the primer in a region that was conserved between the two species.

These are not degenerate primers. In instances where the cow and human sequences disagreed, we used the *Bos* sequence as the primer. In order to maximize the chance of primer attachment to template DNA, forward primers were generally designed with the 3′ base as a second codon position, and reverse primers were designed with the 3′ base as a first codon position. This allowed for two nucleotides on the critical 3′ end of each primer before encountering a third codon position, which is typically less conserved and thus could lead to polymerase chain reaction (PCR) failure. We aimed for annealing temperatures of 60 °C. Primers were initially screened on divergent ungulate species (*B. taurus*, *Capra* spp., *Ovis canadensis*, *Rupicapra rupicapra*). Primer pairs that generated successful PCR products and clean, polymorphic sequences were used in the final analysis on the six ibex taxa.

We used these primers to PCR amplify and sequence five gene regions in 24 individual wild ibex, representing six species or subspecies (Table 2). Analysed ibex include individuals from remote populations that were difficult to sample. We used the Amplit*Taq* gold DNA polymerase (ABI), and thus began each PCR with 10 min at 95 °C. We then ran 35 cycles consisting of 95 °C for 30 s, the appropriate annealing temperature (Table 1) for 60 s, and 72 °C for 60 s. Sequencing of both DNA strands was carried out by Genome Express (Meylan) using an ABI 3730 sequencer

Table 1 Summary information for five primer pairs designed to PCR amplify and sequence portions of three reproductive genes in *Capra ibex*, including the number of SNPs per 1000 bp. See text for details

Locus	Protein product function	Exons	Primer name	Tms	PCR annealing temp. (°C)	Approx. real PCR length	Sequence	SNPs per 1000 bp
Zona pellucida 2 (ZP2)	Sperm receptor on oocyte	8–9	ZP2-X8F	59.2	58	350	CCATCTCTACATGGTGCCTCT	13.7
			ZP2-X9R	58.2			TTGTTTTGAGGAGAGTTTTGCT	
Zona pellucida 2		11–13	ZP2-X11F	58.7	58	600	GGGTTTATGGATGTCGAGGT	17.4
			ZP2-X13R	59.6			GTAGGTTTGCAGGGTCAAGG	
Zona pellucida 3 (ZP3)	Sperm receptor on oocyte	3–4	ZP3A-X3F	60.1	58	340	TGCCATTCAGGACCACAGT	25.3
			ZP3A-X4R	60.0			GGAAGTCCACGATGGTGTG	
Zona pellucida 3		4–5	ZP3A-X4F	60.2	58	350	GAGAAGATGACGCCCACCT	12.5
			ZP3A-X5RA	59.5			CATTAGCAAAACGGAACACATC	
Transition protein 1 (TNP1)	Histone replacement in spermatogenesis	1–2	TNP1-X1F	59.4	59	500	ACCAGCCGCAAATTAAAGAG	6.1
			TNP1-X2R	60.32			GCCTCATTTTGTCATGATTGG	

© 2006 The Authors
Journal Compilation © 2006 Blackwell Publishing Ltd

Table 2 Distribution of SNPs across three genes (five regions) sequenced from 24 *Capra* individuals from eight countries. The consensus base is noted below each column, and variants from that base, as well as their frequency, are noted within the column. Variants marked with nc for the codon position are found in introns S = synonymous, NS = nonsynonymous. Question marks indicate that the sequence quality was not high enough for base calling. IUPAC ambiguity codes are used to indicate polymorphisms (Y, C or T; W, A or T; R, A or G; K, G or T; M, A or C) S = synonymous, NS = nonsynonymous

			Locus site												
			ZP2.X8–9 (293 bp)				ZP2.X11–13 (518 bp)								
Species/subspecies	Country	*n*	73	74	111	260	4	41	161	178	365	370	424	451	473
Capra ibex ibex	Italy	2													
Capra ibex pyrenaica	Spain	1													
Capra ibex nubiana	Israel	4								2:G, 2:R					
Capra ibex nubiana	Saudi Arabia	5								1:R					
Capra ibex sibirica	Mongolia	6	2:Y			3:T, 1:Y	3:T, 1:Y	3:A, 1:R	6:T				3:G, 1:K	3:C, 1:Y	3:A, 1:R
Capra ibex sibirica	Pakistan	1							1:T						
Capra caucasica	Russia	3		1:A			1:Y, 1:?	1:R, 1:?	1:Y, 1:?			1:R, 1:?	1:K, 1:?	1:Y, 1:?	1:R, 1:?
Capra cylindricornis	Russia	2		2:W	2:Y		2:Y	2:R	2:Y		2:R	2:R	2:K	2:Y	2:R
	Codon position		nc	nc	nc	First	Third	First	nc	nc	nc	nc	Third	Third	First
	Cons. base		C	T	T	C	C	G	C	A	G	G	T	T	G
	Syn. or nonsyn.		Intron	Intron	Intron	NS	S	NS	Intron	Intron	Intron	Intron	S	S	NS

			Locus site													
			ZP3.X3–4 (277 bp)							ZP3.X4–5 (319 bp)				TNP1.X1–2 (491 bp)		
Species/subspecies	Country	*n*	32	134	148	176	226	238	271	21	71	83	116	140	215	373
Capra ibex ibex	Italy	2														
Capra ibex pyrenaica	Spain	1														
Capra ibex nubiana	Israel	4					2:Y		2:A, 2:R		2:Y		2:A, 2:R		1:W, 1:A	
Capra ibex nubiana	Saudi Arabia	5							5:A				5:A			
Capra ibex sibirica	Mongolia	6		6:G	6:G			3:Y				3:Y		1:A, 1:R, 4:?	2:A, 4:?	1:Y, 4:?
Capra ibex sibirica	Pakistan	1		1:G	1:G											
Capra caucasica	Russia	3	1:M	1:R		1:Y				1:Y						
Capra cylindricornis	Russia	2	2:M	2:R	1:K	2:Y				2:Y						
	Codon position		First	First	Third	First	Third	Third	Third	First	Third	Third	Third	nc	nc	nc
	Cons. base		C	A	T	C	C	C	G	C	C	C	G	G	T	C
	Syn. or nonsyn.		NS	NS	S	NS	S	S	S	NS	S	S	S	Intron	Intron	Intron

© 2006 The Authors
Journal Compilation © 2006 Blackwell Publishing Ltd

(Applied Biosystems) and standard conditions. Sequences are available from GenBank (Accession nos DQ357834–DQ357941).

SNPs were considered valid if they were observed in at least two individual animals (e.g. two heterozygotes, as in Table 2 for the ZP2.X8-9 gene, nucleotide site 73). Thus, our ascertainment procedure excluded potential SNPs observed as singletons across the entire set of 24 individuals. For example, because we wanted to be conservative in our identification of polymorphisms, variable sites that occurred in only one animal out of the 24 were not included in the list of polymorphisms, even if those sites were apparently homozygous (data not shown). Sites were considered heterozygous (indicated with IUPAC ambiguity codes in Table 2) if the lower peak height was greater than 50% of the higher peak height.

We found a total of 23 variable positions across the five loci sequenced (Table 2). All polymorphic sites identified in the *ZP3* exons 4 and 5 fragment were found in the exon 4 portion and were redundant with the other *ZP3* primer pair. Average SNPs per 1000 bp across all five loci was 15.0, with a high of 25.3 in the *ZP3* exons 3 and 4 sequence and a low of 6.1 in the *TNP1* sequence (Table 1). Ten of the polymorphisms occurred in introns, six were nonsynonymous and seven were synonymous.

Major goals of developing these primers are to verify species status, make phylogenies, test for local adaptation, identify speciation genes and identify SNPs for use in population genetic studies (e.g. of microgeographic population structure, dispersal and gene flow). The primers described here promise to do all of the above within wild goat (*Capra*) species and many related species of bovids, such as cattle relatives (*Bos* spp.) that are approximately 20 million years divergent from *Capra*.

Acknowledgements

This work was supported by a US National Science Foundation International Research Postdoctoral Fellowship granted to SJ (grant 0107373). Thanks to P. Weinberg, Y. Zagdsuren, A. Virk and K. Scribner for samples.

References

Manceau V, Després L, Bouvet J, Taberlet P (1999) Systematics of the genus *Capra* inferred from mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution*, **13**, 504–510.

Muller C, Denis M, Gentzbittel L, Faraut T (2004) The Iccare web server: an attempt to merge sequence and mapping information for plant and animal species. *Nucleic Acids Research*, **32**, W429–W434.

Rozen S, Skaletsky HJ (2000) PRIMER 3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386. Humana Press, Totowa, NJ, USA.

Shackleton DM (1997) *Wild Sheep and Goats and Their Relatives: Status Survey and Conservation Action Plan for Caprinae*. IUCN, Gland, Switzerland.

Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. *Nature Reviews Genetics*, **3**, 137–144.

Torgerson DG, Kulathinal RJ, Singh RS (2002) Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Molecular Biology and Evolution*, **19**, 1973–1980.

Yu YE, Zhang Y, Unni E *et al.* (2000) Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. *Proceedings of the National Academy of Sciences, USA*, **97**, 4683–4688.

© 2006 The Authors
Journal Compilation © 2006 Blackwell Publishing Ltd