Molecular Ecology (2003) **12**, 283–291

Y chromosome conserved anchored tagged sequences (YCATS) for the analysis of mammalian male-specific DNA

LINDA HELLBORG and HANS ELLEGREN
Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

Abstract

Y chromosome haplotyping based on microsatellites or single nucleotide polymorphisms has recently proven to be a powerful approach for evolutionary studies of human populations, and also holds great promise for the studies of wild species. However, the use of the approach is hampered in most natural populations by the lack of Y chromosome markers and sequence information. Here, we report the large-scale development of Y chromosome conserved anchor tagged sequence (YCATS) markers in mammals by a polymerase chain reaction screening approach. Exonic primers flanking 48 different introns of Y-linked genes were developed based on human and mouse sequences, and screened on a set of 20 different mammals. On average about 10 introns were amplified for each species and a total of 100 kb of Y chromosome sequence were obtained. Intron size in humans was a reasonable predictor of intron size in other mammals ($r^2 = 0.45$) and there was a negative correlation between human fragment size and amplification success. We discuss a number of factors affecting the possibility of developing conserved Y chromosome markers, including fast evolution of Y chromosome sequences due to male-biased mutation and adaptive evolution of male-specific genes, dynamic evolution of the Y chromosome due to being a nonrecombining unit, and homology with X chromosome sequences.

Keywords: cross-species amplification, PCR, sex chromosomes, SNP, Y chromosome

Received 12 April 2002; revision received 9 September 2002; accepted 9 September 2002

Introduction

DNA-based approaches for the study of relationships between individuals, populations and species form the essence of molecular ecology and have also become important for related subject areas such as conservation genetics and systematics. The molecular ecology community has benefited enormously from technical and conceptual advances made initially in connection with efforts to map and subsequently to sequence the genomes of human and model organisms. For instance, subsequent to the introduction of microsatellite genotyping as a key tool for human genetic mapping (Litt & Luty 1989; Tautz 1989; Weber & May 1989), the isolation and application of microsatellites in natural populations soon occurred (Ellegren 1991; Schlotterer *et al*. 1991). Other techniques originally developed for the purpose of genetic mapping have rapidly been adopted by molecular ecologists as well [e.g. restriction fragment length polymorphisms (RFLP), random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphisms (AFLP)].

It is probably a good bet that single nucleotide polymorphism (SNP) markers will become a key tool in the next generation of molecular ecology research, just as microsatellites and other types of markers have been instrumental during the past 5–10 years. SNPs are co-dominant biallelic markers at individual nucleotide positions and can be found in both coding and noncoding sequences. Although they obviously do not display the hypervariability characteristic of microsatellites, they represent an almost inexhaustible source of markers spread over the genome. For example, there is an estimated density of one SNP every 1300 base pairs (bp) of two human genomes and more than one million SNPs have so far been identified in the human genome project (Altshuler *et al*. 2000;

Correspondence: Hans Ellegren. Fax: + 46 18 4716310; E-mail: Hans.Ellegren@ebc.uu.se

International SNP Consortium 2001). Equally importantly, SNPs are amenable to large-scale, automated genotyping with techniques based on, for example, oligonucleotide hybridization (Fan *et al*. 2000), primer extension (Pastinen *et al*. 2000) and pyrosequencing (Alderborn *et al*. 2000).

One of the most prominent recent applications of SNPs in human evolutionary studies is based on the use of markers from the Y chromosome (Stumpf & Goldstein 2001). Because it is mainly a nonrecombining chromosome, analysis of Y chromosome SNPs allows the construction of specific haplotypes defining patrilines. Using this approach, a number of studies have recently attempted to dissect the evolutionary history of modern humans (Hill *et al*. 2000; Semino *et al*. 2000; Underhill *et al*. 2000; Bamshad *et al*. 2001; Bosch *et al*. 2001; Capelli *et al*. 2001; Kayser *et al*. 2001; Ke *et al*. 2001; Quintana-Murci *et al*. 2001). In general, Y chromosome haplotyping is complementary to data from markers based on mitochondrial DNA (defining matrilines) or autosomal sequences (biparentally inherited) as it allows specifically paternal relationships to be studied. With the access to Y chromosome haplotyping, it is therefore possible to compare directly the relative contribution of fathers in shaping the genetic signatures of populations.

While it seems likely that Y chromosome markers will become valuable to molecular ecological research, their identification, and therefore also their application, is not without difficulties. First, cloning of Y chromosome sequences may not be a trivial task. For instance, the Y chromosome contains few genes (Tilford *et al*. 2001; making cDNA library screening tedious) and chromosome-specific libraries (derived from flow sorting or chromosome scraping) are rarely available for wild species. Second, nucleotide diversity in the sex-limited chromosome may be lower than in the rest of the genome (Filatov *et al*. 2000; International SNP Consortium 2001; Montell *et al*. 2001) because the absence of recombination makes it sensitive to selective sweeps and background selection (Rice 1987; Charlesworth *et al*. 1993). While the latter should basically be a matter of how much sequence is available for screening, the critical problem may be to initially access (amplify) Y chromosome sequences of a less well-characterized genome.

O'Brien *et al*. (1993) and Lyons *et al*. (1997) introduced the concept of large-scale use of conserved anchored tagged sequences (CATS) for targeting sequences in uncharacterized genomes (see also Jiang *et al*. 1998). Based on sequence data available for two or more organisms, the approach relies on the design of primers in exonic regions subject to selective constraints. Such primers are likely to amplify across a range of species. Moreover, if the primers are placed in exon sequences flanking an intron, they may provide access to noncoding sequences which can be screened for genetic variability. Nucleotide diversity is generally much higher in noncoding DNA than in coding DNA.

CATS have been widely used for comparative mapping in mammals (e.g. Lyons *et al*. 1999; Priat *et al*. 1999; Lahbib-Mansais *et al*. 2000; Pinton *et al*. 2000) and to some extent, and typically on a much smaller scale, for the characterization of sequences in the genomes of wild species (Slade *et al*. 1993; Friesen *et al*. 1999; Bierne *et al*. 2000; Saetre *et al*. 2001). Here we apply the CATS approach for the first large-scale targeting of Y chromosome sequences, which we term 'YCATS', in a broad range of mammalian species.

Materials and methods

Our strategy for developing intronic YCATS involved a two-step process. First, human genomic Y chromosome gene sequences were aligned with the homologous mRNA sequences from at least one other mammalian species (generally mouse) and in some cases from two or more species (horse, pig and different primates). This allowed confirmation of exon–intron organization and identification of conserved exonic regions or sequences. Second, as many Y chromosome genes have independently evolving gametologues on the X chromosome, a reminiscence of an ancestral autosomal state of sex chromosomes (Lahn & Page 1999), we added the corresponding human and mouse X chromosome sequences to the alignment. From this we could identify exonic nucleotide positions conserved between available Y chromosome sequences but different from the corresponding positions in X chromosome sequences (see example in Fig. 1). Such conserved (between Y chromosome sequences) but unique (relative to the corresponding X chromosome sequences) Y chromosome sites either represented synonymous or nonsynonymous substitutions, or short insertion/deletions (indels). In general, primers were then designed to have the 3′ ultimate base(s) at these sites. Primer sequences are given in Table 1. Forty-eight introns were analysed and screened in a diverse range of 20 mammalian species (see Table 2).

Twenty-five nanograms DNA were used in 20 μL polymerase chain reaction (PCR) reactions containing 1× Gold PCR buffer, 2.5 mM $MgCl_2$, 5 μM of each primer, 2 μM dNTP and 0.25 U Ampli*Taq* Gold polymerase (Perkin Elmer). For optimization two PCR profiles were run for each primer pair. Both profiles included one cycle of 95 °C for 10 min, followed by 20 cycles of 95 °C for 30 s and a touchdown from either 65 °C to 55 °C or from 55 °C to 45 °C for 1 min decreasing by 0.5 °C/cycle, and then 72 °C for 1.50 min. This was followed by 20 cycles of 95 °C for 30 s, 55 °C or 45 °C for 1 min and 72 °C for 1.50 min, and a final extension step of 72 °C for 10 min. PCR products were separated on 2.0% agarose gels, stained with ethidium bromide and visualized under ultraviolet light. At least one male and one female DNA sample were used for each

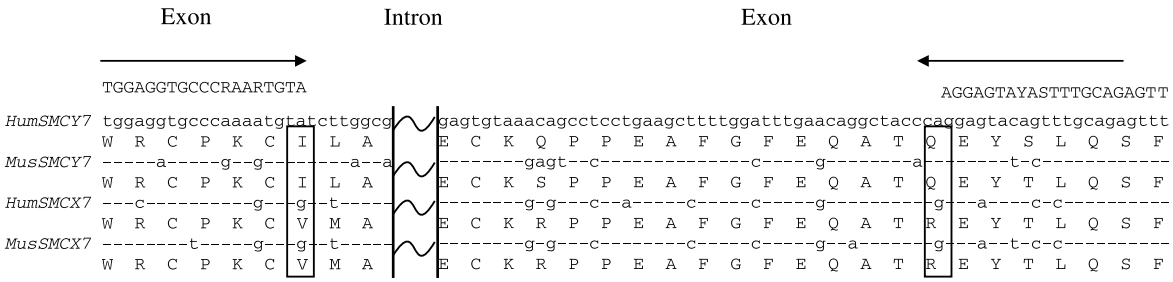

Fig. 1 An example of the strategy used for primer design. For each gene both the DNA and the amino acid sequence is given. Identical nucleotide positions are denoted by '–'. Arrows indicate primer position (with the primer sequence given below). Boxes highlight exonic nucleotide positions (triplets) conserved between available Y chromosome sequences but different from the corresponding positions in available X chromosome sequences.

Table 1 PCR Primer sequences* of YCATS markers

Locus	Forward primer	Reverse primer
DBY1	AGCAGTTTTGGRTCTCGWGA	CCAACGACTATGWCCACT
DBY2	CACTTAAGGAACARAGAARCAT	GCTTTCCTCTAGAATCWC
DBY3	ACTATCGACAGAGYAGTGGTT	TCCATAACCATCAYTATTGTAG
DBY4	TGATGGTATTGGYRRTCGTGA	CGGTTGCCTCTACTGGTATA
DBY5	CTTTGAGAAATATGATGATATA	CGAGTAAGTTCAATGTTC
DBY6	TCCTACTCCAGTGCAAAAA	ATCTGACTCAGWATGGGT
DBY7	GGTCCAGGAGARGCTTTGAA	CAGCCAATTCTCTTGTTGGG
DBY8	CCCCAACAAGAGAATTGGCT	CAGCACCACCATAKACTACA
DBY9	CTAGAGTTCGTCCTTGTGTA	AATCCCTATTCCAGCATCCT
DBY10	CACCAAAGGGSGTTCGT	CATCCAAAAAGTCACGGAGCA
DBY11	TCTTTTTGGCTGTAGGCAGAGTA	TGGTGAAGGGCCTCCTCTCG
DBY12	GGAGGCCCTTCACCAGTTT	CATGTTTCACATTTGAAATGTCT
DBY13	CAGGACGTGTAGGAAACC	GTTCATAAGCCATARTTTCC
DBY14	CAAGAAGTGCCTTCTTGGTTG	GGCTCCAAATCCTCCACTG
RPS4Y1	CCGAAGCATTGGATGTCTT	AGTCTATTCCTRAGGAAG
RPS4Y3	GATATGTATGCAACGBTTCATC	ACGGCCCTTGGTGTCATASAC
RPS4Y4	GAACAGATTCTCTTCCGTCG	GCTGCAACACGCTTTAAGTG
SMCY1	TGACAATTTCARRTTTACTCCT	CTGTAAAGGTCCAAGATCTT
SMCY2	ATTCCCAATGTGGAGMGGAA	CATAGCCACCTTCCTCMAT
SMCY3	ATTTACCCTTATGAAATRTTT	TCAAATGGGTGWGTGTYACAT
SMCY4	TCCAAGTTCAGCWGCTAYAGT	TCCTCCTCTGTRGGYTCT
SMCY5	GGTCCCAAGATGATRGGCT	AGTTGGGGGGCATRTSAC
SMCY6	CTACGGAAGAATCACAGCAG	ATTTCAGGAAGSGGTGGYAA
SMCY7	TGGAGGTGCCCRAARTGTA	AACTCTGCAAASTRTACTCCT
SMCY8	TTGGATCTCAATCTAGCAGT	CGCTCAGCYTCCGTGAYG
SMCY9	ATGCTACAWAARCTGAAGA	TGCTCTTAGCTCTTCAAA
SMCY11	CTGCCCTGYRCCATGCAT	TCCACCTGTTSMAGRACAT
SMCY12	TGGGAGGAAAAGGCYCATT	CTCACGTATTATRRCTTCCAA
SMCY14	TGCTGGGGCTGTCTGCA	CTTCTCCTTCTGTTCCCCT
SMCY16	TGGCAGCTGCTGCAGGCTG	TTCTACTCAGAWCCACCTTYTG
SMCY17	AATGATCTGTGCAAGTGCTC	GTCAAATGACTCAGCYCGAAT
UBE1Y1	GCCCAGAGACAGAACACG	CAACGGCGGTCCACGTATAA
UBE1Y6	CCCCTGCAGACCKRCAT	AAGGCCAAGTTGATRAARCT
UBE1Y7	TGGCTCTGCCTTTSTTYAGC	AGGTTGTATGCCYTGYACA
UTY5	TTGGTTTGGTCTAYTTCTAC	GGTCAACATAAAGGACRTCT
UTY7	CAAGTAAAAGCAACTGTATTG	TGTCTCCTACTAGATCCATAT
UTY11	CATCAATTTTGTAYMAATCCAAAA	TGGTAGAGAAAAGTCCAAGA
ZFY2	CAGTGGACCAGCAAGATA	GATCCATTTCTGAYTCTGCA
ZFY4	AAGAAGATAAGTTTACAT	CTGTTCAGCTGTCTCAT

*For degenerate primers Y = C/T, M = A/C, R = A/G, S = C/G, B = C/G/T and N = A/C/G/T.

Table 2 Estimated size (in bp) of Y-specific amplification products obtained in a range of mammalian species

	Species*																			
Locus	Hosa**	Patr	Lyly	Calu	Gugu	Urar	Myda	Eqca	Rata	Caca	Susc	Loaf	Grgr	Caur	Hagr	Mile	Cafi	Ereu	Leeu	Miag	Soar
DBY1	927	1000†									1000‡								650‡		1500‡
DBY2	2549	3000‡																			
DBY3	637			300†	300§	350§		500†			400†				400†						700†
DBY4	281	250†	250†						300†	300†			600†		300†			300†		550†	
DBY5	776	700†		500†	1100†	500†					700†			900†							
DBY6	845	800†		800†				800†			800†						900†				
DBY7			350‡	350‡	500‡		400‡	350‡	350‡	330‡	350‡	300‡	400‡	600‡	750†	700‡	250‡			550‡	500‡
DBY8	185	200‡	200†	200‡	200‡	200‡	200‡	200†	200‡	200‡	200‡	200‡	200‡	200‡	200‡	200‡	200‡	200‡		200‡	200‡
DBY9	590									450‡			300†	500‡	650‡	750‡	680‡			800†	
DBY10	191	200†			300†												200†			200†	
DBY11	475																400†		400‡	600†	
DBY12	163		200‡			200‡							200†		200†						
DBY13	406	400†					450†										400†				
DBY14	441	450†					500‡			350†							120†	500‡		500‡	500†
RPS4Y1	1944	2000‡																			
RPS4Y3	1523	2000‡															600†				
RPS4Y4	620	600‡		700†										150†							
SMCY1							2000†	1000†													
SMCY2	305	300‡						300‡													
SMCY3	1725	3000†	900†				600†	1000‡			1500†				800†						
SMCY4				550†															2000†		
SMCY5	231	200‡		700†		200†		200‡			600‡		200†		250†						
SMCY6				600†				900‡													
SMCY7	512	500‡	700‡	500†	500†	500‡		400‡			450‡		500‡	500†	450‡	500†	450‡			600‡	
SMCY8	382	400†											100†				400†			350†	
SMCY9	230	200†																	500†		
SMCY11	630	650‡					550‡	600§					350‡							1000†	
SMCY12													150†	500‡	450†	500‡					
SMCY14		250†																		150†	
SMCY16	244						450†						500†							250†	
SMCY17	171		200†	200†		200†	200†	200†				230†		230†		230†	200‡			200†	
UBE1Y1	314¶										450†									150†	450‡
UBE1Y6	288¶			300†	250†			200‡	300‡	300‡	200†		300‡		300‡	300†		300‡		280‡	
UBE1Y7	217¶												500†		650†	600†					
UTY5	3695								600†												
UTY7	2923	3000†																			
UTY11	900	900‡	700†	400†			550†	600†		750†			550†	650†	650†	650†				800†	1000†
ZFY2							270†														
ZFY4	200	200†																			200†

*Hosa = Human, Patr = Chimpanzee, Lyly = Lynx, Calu = Wolf, Gugu = Wolverine, Urar = Brown bear, Myda = Daubenton's bat, Eqca = Horse, Rata = Reindeer, Caca = Roe deer, Susc = Pig, Loaf = African elephant, Grgr = Risso's Dolphin, Caur = Northern fur seal, Hagr = Grey seal, Mile = Southern elephant seal, Cafi = Beaver, Ereu = Hedgehog, Leeu = Brown hare, Miag = Field Vole and Soar = European Shrew.
†Touchdown with an annealing temperature gradient from 55 °C to 45 °C over 20 cycles.
‡Touchdown with an annealing temperature gradient from 60 °C to 50 °C over 20 cycles.
§Touchdown with an annealing temperature gradient from 65 °C to 55 °C over 20 cycles.
¶Since *UBE1Y* is not present in human, the sizes of the fragments are taken from mouse.
**Human sizes are obtained from sequence data.

primer pair and an amplification was considered Y-specific when a single fragment was obtained in male but not in female amplification. The majority of fragments were sequenced to confirm Y chromosome origin. Those products were purified with a Qia-Quick PCR purification kit (Qiagen) and sequenced using BigDye™ Terminator Cycle Sequencing chemistry (Perkin Elmer), following the manufacturer's protocols. The sequences were recorded with an ABI377 semi-automated sequencing instrument (Perkin-Elmer).

SNP screening was performed using single-strand conformation polymorphism (SSCP) analysis of lynx and field vole for fragments shorter than 400 bp. Longer fragments were sequenced directly. PCR fragments were separated in native polyacrylamide gels, electrophoresed at 1 W and visualized with silver staining. Different electromorphs identified with SSCP were confirmed with DNA sequencing according to above.

Results

Overall amplification success

We focused on Y chromosome introns of sizes amenable to PCR (in most cases less than 1000 bp) and where suitable primer sites could be identified in flanking exons. Amplification of 48 Y-linked introns from six different genes was attempted in a range of 20 species representing 10 different mammalian orders. Thirty-nine introns could be amplified in one or more species with a mean of 9.9 ± 4.8 introns per species (Table 2; in the following analyses only these 39 introns are considered). This corresponds to an overall success rate of 26% and translates into amplification of approximately 100 kilobases (kb) of mammalian Y chromosome DNA. On average, 5.1 ± 4.2 kb of sequence were obtained per species (range 0.7–21.2).

Failure of Y-specific amplification was either in the form of no or nonspecific amplification (85%), or due to the fact that a similarly sized product was obtained in amplification of both male and female DNA (15%). For the latter, possible explanations include that the X chromosome gametologue had been amplified from males and females, despite the measures taken at primer design. Alternatively, the appropriate Y chromosome sequence may indeed have been specifically amplified from males, but in the absence of competing template, the primers may have amplified the X gametologue in females. This assumes the introns of X and Y gametologues be indistinguishable in size under the conditions here employed. Similarly, both gametologues may have been amplified in males. We did not investigate these instances further but they could represent additional targets for access to the Y chromosome sequence.

We sequenced at least part of the PCR product for 29 of the 39 introns that showed amplification in males only; a total of approximately 58 kb of sequence data were derived in this way. While male specificity in amplification is in itself a very strong indication that the Y chromosome sequence has been amplified, alignment with human genomic X and Y sequences consistently confirmed that the products obtained were indeed of Y chromosome origin (data not shown). We therefore conclude that truly male-specific sequences were targeted. Moreover, as the genes selected for use in this study represent single-copy genes (Lahn *et al*. 2001), it seems reasonable to assume that true orthologues were amplified. Several other genes on the mammalian Y chromosome exist in multiple copies (Lahn & Page 1997).

Not surprisingly, amplification success varied between species. The highest ratio was obtained for chimpanzee (59%), which constituted the closest relative to either of the two species generally used for primer design. In one sense this ratio may seem low given the close relationship between human and chimpanzee. However, the success ratio for human DNA was only 74% and the results obtained for other species should thus be seen in relation to this figure. The second highest success ratio was observed for field vole, which is compatible with the use of mouse sequences for primer design.

To make a preliminary test of how well these primers work in mammals other than those used in the main screening, we selected five markers (*SMCY7*, *SMCY17*, *DBY3*, *DBY5* and *DBY7*) giving Y-specific amplification in wolves and applied them to four other canid species: coyote, bush dog, Blanford's fox and North American grey fox. All five primer pairs gave Y-specific amplification products in the other canids, with similar fragment sizes to those obtained in wolves. This indicates that the information provided in Table 2 could be used as a guideline for the possibility of amplification of Y chromosome sequences in close relatives to the species included in our screening.

Fragment size and amplification success

Figure 2 illustrates the relationship between fragment size in human and PCR product size obtained in amplification from other mammalian species. Although there is a strong positive relationship ($r^2 = 0.45$, $P = 0.0001$), there is considerable variation in fragment length for orthologous loci. It seems reasonable to assume that this variation is mainly due to variation in intron size; this was also confirmed for those loci that were sequenced. There was a negative correlation between human fragment size and amplification success ($r^2 = 0.20$, $P = 0.006$). Moreover, as the data in Fig. 2 are better explained by a second-order than a linear model ($r^2 = 0.56$ vs. $r^2 = 0.45$), there is probably a bias among long introns to amplify in species where they are comparably short.

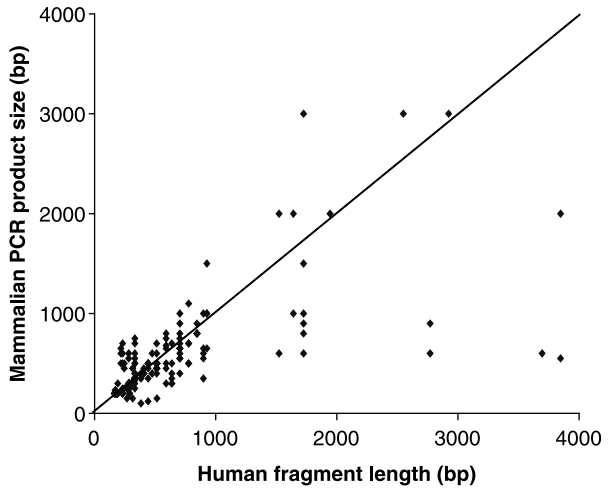

Fig. 2 The relationship between human fragment length (bp) and the PCR product size obtained in amplification of the orthologous intron in other mammals.

Screening for Y chromosome SNPs

In a pilot study we applied the new Y chromosome sequences to screen for SNPs in two mammalian species, the field vole and the lynx, using SSCP. This indicated that Y chromosome variability may differ significantly between species. Seven segregating sites were uncovered in 5 kb of Y chromosome sequence among 18 unrelated male Scandinavian field voles. In contrast, no variation was revealed among 53 unrelated male Scandinavian lynxes in 2.3 kb of Y chromosome sequence (data not shown). Levels of variability at autosomal microsatellites are low in Scandinavian lynxes (Hellborg *et al*. 2002) so it may be that the SNP frequency seen in lynx Y is unusually low, rather than that seen in field voles being unusually high. In any case, this shows the feasibility of the approach for revealing Y chromosome variability based on YCATS.

Discussion

The intrinsic evolutionary properties of the mammalian Y chromosome, including the patterns of mutation and selection and the process of sex chromosome evolution, pose problems for the development of conserved Y-linked genetic markers. In the following we shall discuss these properties in relation to our data. We shall also discuss strategies that can be employed to facilitate the development of new Y-linked markers.

Male-biased mutation rates

Being a male-specific chromosome and given that males tend to mutate more frequently than females (Hurst & Ellegren 1998), Y is exposed to more mutations than the rest of the genome. Synonymous substitution rates on Y thus generally exceed the rates in autosomal or X-linked genes. As a consequence, and everything else being equal, we shall expect divergence at silent exonic PCR primer sites on Y to be more pronounced than in other genes. For this reason conserved Y primers may be particularly hard to develop.

It is difficult to address quantitatively the effect of elevated mutation rates on PCR performance, mainly because there exists no clear relationship between performance and the degree and character of primer sequence divergence. The fact that we never achieved more than 60% amplification success in individual species, and generally less than 25%, may suggest that cross-species amplification was significantly hampered by rapid evolution of primer sequences. However, this is probably a premature conclusion for several reasons. First, amplification success in human material was only 74% despite using the human sequence for primer design. This low intraspecific success ratio is perhaps surprising, and is unaccounted for, but may partly relate to any of the factors discussed below. In any case, this indicates general problems in amplifying Y chromosome sequences, at least in our study. Second, we have no information on the amplification success of autosomal markers, derived in the same way, on the mammalian species used in this study. Hence we cannot say if Y chromosome amplification success was particularly low.

A male-biased mutation rate will affect all loci on the mammalian Y chromosome and should in this sense thus not influence the choice of targets for marker development. However, recent observations have suggested that the local mutation rate varies significantly between different genomic regions (Lercher *et al*. 2001; Smith *et al*. 2002). Future development of conserved genetic markers should preferably target genes or gene regions with low synonymous substitution rates. Obviously, this would apply to marker development from all chromosomes, not only the Y chromosome.

Positive selection on Y chromosome sequences

The obvious rationale of using exonic sites as conserved anchored tags for primer design is that purifying (negative) selection typically puts coding sequences under constraint. However, adaptive evolution (positive selection) will occasionally drive coding sequences toward new functions, leading to elevated rates of nonsynonymous substitution. Under these conditions it may be more difficult to identify conserved primer binding sites. A hallmark of positive selection in DNA sequences is that the rate of nonsynonymous substitution (K_a) exceeds the rate of synonymous substitution (K_s) (whilst under neutrality

$K_a/K_s = 1$, and under purifying selection $K_a/K_s < 1$). A number of genes have been found associated with K_a/K_s ratios above one and among these genes involved in male reproduction are highly overrepresented (Wyckoff *et al*. 2000; Yang & Bielawski 2000). Many Y-linked genes are indeed involved in male reproduction and we suggest that they may be less useful for the development of conserved male-specific markers. As the number of genes on mammalian Y is limited (< 30; Lahn *et al*. 2001), targets for marker development therefore need to be carefully selected; in this study we did not include genes involved in male reproduction.

Dynamic evolution of the Y chromosome

While the mammalian X chromosome is highly conserved in terms of size and gene content across mammals (Ohno 1973), Y is very dynamic in these respects. Being a non-recombining unit (with the exception of the pseudo-autosomal region; PAR), Y is subject to the constraints inherently associated with clonal transmission (i.e. an inability to segregate deleterious mutations at the nucleotide or chromosomal level). Most genes on Y that once were shared with X have degraded and eventually become lost. For this reason genes that are Y-linked in some mammalian lineages may not necessarily be so in others. Moreover, there are several examples of recent acquisition of genes to the mammalian Y chromosome by transposition or retroposition (Lahn *et al*. 2001). Generally, such genes are involved in male-specific processes (spermatogenesis).

The dynamic nature of the mammalian Y chromosome adds to the care that should be taken in the design of conserved markers. In our case we cannot exclude the possibility that the failure to amplify some genes in some species was due to these genes being lost from Y. Comparative mapping information will be important for assessing such lineage-specific gene losses from the Y chromosome.

X–Y chromosome gametologues

As indicated above, the Y chromosome differs from all other mammalian chromosomes in that a majority of the genes have a related gametologue on another chromosome, namely, the X chromosome. This peculiar type of chromosomal homology is likely to represent a reminiscence of an ancestral autosomal state of X and Y, prior to their differentiation. It is commonly believed that sex chromosomes evolve by suppression of recombination around a sex-determining locus, triggered by linkage to sexually antagonistic alleles or by events of inversion, followed by gradual decay and contraction of the sex-limited chromosome. Recent observations suggest that human sex chromosome differentiation can be defined by at least four evolutionary strata on the X chromosome, each stratum representing a chromosomal segment with a specific divergence time relative to Y (Lahn & Page 1999). This would indicate that sex chromosome evolution was punctuated by at least four events suppressing recombination in stepwise fashion, from the distal long arm to the distal short arm. Human X–Y gametologues within the respective strata share similar K_s values, which is indicative of common divergence times, but they may show different K_a values. Estimates of divergence times are 130–350 million years ago (Ma) for stratum 4 (distal on Xq), 130–170 Ma for stratum 3, 80–130 Ma for stratum 2 and 30–50 Ma for stratum 1 (distal on Xp, next to the pseudo-autosomal region; Lahn & Page 1999).

The process of mammalian sex chromosome evolution has several implications for the development of Y chromosome markers. First, the presence of a related gametologue on X implies a risk for cross-amplification. Everything else being equal, the relative risk for spurious amplification should decrease with increasing X–Y divergence, suggesting that genes from the older strata might be better suited in this respect. Second, for a Y-linked gene to possibly share character states across all mammals, X–Y divergence must obviously predate the split of extant mammalian lineages. As human stratum 1 evolved subsequent to the major mammalian radiation, genes in this stratum are either still recombining in other mammalian lineages or have been subject to independent divergence. In any case, sequence information from Y-linked genes corresponding to stratum 1 should be of little use for the development of conserved Y-specific markers across all mammals.

Acknowledgements

We would like to thank Anna-Carin Andersson, Peter Arctander, Ulla Gustafsson, Maarit Jaarola, Lori Lawson, Gabriella Lindgren, Anna Lindvall, Jennifer Seddon, Göran Spong and Carl-Gustaf Thulin for providing samples and Frank Hailer for help in the laboratory. H.E. is a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg foundation. This study was supported by the Swedish Research Council.

References

Alderborn A, Kristofferson A, Hammerling U (2000) Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Research*, **10**, 1249–1258.

Altshuler D, Pollara VJ, Cowles CR *et al*. (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, **407**, 513–516.

Bamshad M, Kivisild T, Watkins WS *et al*. (2001) Genetic evidence on the origins of indian caste populations. *Genome Research*, **11**, 994–1004.

Bierne N, Lehnert A, Bedier E, Bonhomme F, Moore SS (2000) Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. *Molecular Ecology*, **9**, 233–235.

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *American Journal of Human Genetics*, **68**, 1019–1029.

Capelli C, Wilson JF, Richards M *et al.* (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *American Journal of Human Genetics*, **68**, 432–443.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.

Ellegren H (1991) DNA typing of museum birds. *Nature*, **354**, 113.

Fan JB, Chen X, Halushka MK *et al.* (2000) Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Research*, **10**, 853–860.

Filatov DA, Monéger F, Negrutiu I, Charlesworth D (2000) Low variability in a Y-linked plant gene and its implications for Y chromosome evolution. *Nature*, **404**, 388–390.

Friesen VL, Congdon BC, Kidd MG, Birt TP (1999) Polymerase chain reaction (PCR) primers for the amplification of five nuclear introns in vertebrates. *Molecular Ecology*, **8**, 2147–2149.

Hellborg L, Walker CW, Ruesness EK *et al.* (2002) Genetic differentiation and low levels of variation in northern European lynx (*Lynx lynx*) populations revealed by microsatellites and mitochondrial DNA analysis. *Conservation Genetics*, **3**, 97–111.

Hill EW, Jobling MA, Bradley DG (2000) Y chromosome variation and Irish origins. *Nature*, **404**, 351–352.

Hurst LD, Ellegren H (1998) Sex biases in the mutation rate. *Trends in Genetics*, **14**, 446–452.

International SNP Consortium (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, **409**, 928–933.

Jiang Z, Priat C, Galibert F (1998) Traced orthologous amplified sequence tags (TOASTs) and mammalian comparative maps. *Mammalian Genome*, **9**, 577–587.

Kayser M, Brauer S, Weiss G, Schiefenhovelm W, Underhill PA, Stoneking M (2001) Independent histories of human Y chromosomes from Melanesia and Australia. *American Journal of Human Genetics*, **68**, 173–190.

Ke Y, Su B, Song X *et al.* (2001) African origin of modern humans in East Asia: a tale of 12 000 Y chromosomes. *Science*, **292**, 1151–1153.

Lahbib-Mansais Y, Leroux S, Milan D *et al.* (2000) Comparative mapping between humans and pigs: localization of 58 anchorage markers (TOASTs) by use of porcine somatic cell and radiation hybrid panels. *Mammalian Genome*, **11**, 1098–1106.

Lahn BT, Page DC (1997) Functional coherence of the human Y chromosome. *Science*, **278**, 675–680.

Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. *Science*, **286**, 964–967.

Lahn BT, Pearson NM, Jegalian K (2001) The human Y chromosome, in the light of evolution. *Nature Reviews Genetics*, **2**, 207–216.

Lercher MJ, Williams EJB, Hurst LD (2001) Local similarity in evolutionary rates extends over whole chromosomes in human-rodent and mouse-rat comparisons: implications for understanding the mechanistic basis of the male mutation bias. *Molecular Biology and Evolution*, **18**, 2032–2039.

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*, **44**, 397–401.

Lyons LA, Laughlin TF, Copeland NG *et al.* (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics*, **15**, 47–56.

Lyons LA, Kehler JS, O'Brien SJ (1999) Development of comparative anchor tagged sequences (CATS) for canine genome mapping. *Journal of Heredity*, **90**, 15–26.

Montell H, Fridolfsson AK, Ellegren H (2001) Contrasting levels of nucleotide diversity on the avian Z and W sex chromosomes. *Molecular Biology and Evolution*, **18**, 2010–2016.

O'Brien SJ, Womack JE, Lyons LA, Moore KJ, Jenkins NA, Copeland NG (1993) Anchored reference loci for comparative genome mapping in mammals. *Nature Genetics*, **3**, 103–112.

Ohno S (1973) Ancient linkage groups and frozen accidents. *Nature*, **244**, 259–262.

Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvanen AC (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Research*, **10**, 1031–1042.

Pinton P, Schibler L, Cribiu E, Gellin J, Yerle M (2000) Localization of 113 anchor loci in pigs: improvement of the comparative map for humans, pigs, and goats. *Mammalian Genome*, **11**, 306–315.

Priat C, Jiang ZHH, Renier C, Andre C, Gailbert F (1999) Characterization of 463 Type I markers suitable for dog genome mapping. *Mammalian Genome*, **10**, 803–813.

Quintana-Murci L, Krausz C, Zerjal T *et al.* (2001) Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *American Journal of Human Genetics*, **68**, 537–542.

Rice WR (1987) Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics*, **116**, 161–167.

Saetre GP, Borge T, Lindell J *et al.* (2001) Speciation, introgressive hybridization and nonlinear rate of molecular evolution in flycatchers. *Molecular Ecology*, **10**, 737–749.

Schlotterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. *Nature*, **354**, 63–65.

Semino O, Passarino G, Oefner PJ *et al.* (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science*, **290**, 1155–1159.

Slade RW, Moritz C, Heidman A, Hale PT (1993) Rapid assessment of single-copy nuclear-DNA variation in diverse species. *Molecular Ecology*, **2**, 359–373.

Smith NGC, Webster MW, Ellegren H (2002) Deterministic mutation rate variation in the human genome. *Genome Research*, **12**, 1350–1356.

Stumpf MP, Goldstein DB (2001) Genealogical and evolutionary inference with the human Y chromosome. *Science*, **291**, 1738–1742.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic markers. *Nucleic Acids Research*, **12**, 4127–4138.

Tilford CA, Kuroda-Kawaguchi T, Skaletsky H *et al.* (2001) A physical map of the human Y chromosome. *Nature*, **409**, 943–945.

Underhill PA, Shen P, Lin AA *et al.* (2000) Y chromosome sequence variation and the history of human populations. *Nature Genetics*, **26**, 358–361.

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed with the polymerase chain reaction. *American Journal of Human Genetics*, **44**, 388–396.

Wyckoff GJ, Wang W, Wu CI (2000) Rapid evolution of male reproductive genes in the descent of man. *Nature*, **403**, 304–309.

Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution*, **15**, 496–503.

Linda Hellborg is a PhD student at the Evolutionary Biology Centre of Uppsala University in a group led by Prof. Hans Ellegren, studying carnivore conservation genetics and the use of sex-linked markers.
