Molecular Assessment of Sex Chromosome Polymorphisms in the Bowhead Whale

John W. Bickham[1], Heather K. Downing[2], John C. Patton[1], John C. George[3], and Robert S. Suydam[3].

[1]Department of Forestry and Natural Resources, and Center for the Environment, Purdue University, West Lafayette, IN 47907; [2]Department of Biology, Purdue University, West Lafayette, IN 47907; [3]North Slope Borough, Department of Wildlife Management, Barrow, AK 99723, USA

Abstract

The unique inheritance patterns of the mammalian X and Y chromosomes provide the opportunity to expand our understanding of heretofore virtually unknown components of population biology. The inheritance patterns, expected neutral mutation rates, and effective population sizes of the X and Y chromosomes differ from those of both the maternally inherited mitochondrial DNA and the bi-parentally inherited autosomes which are the typical elements used in population genetics. To better comprehend the population biology of the X and Y elements in bowhead whales, we sequenced 21,750 nucleotides for USP9Y and 11,150 nucleotides for USP9X. We observed a single nucleotide substitution in USP9Y and 8 variable sites in USP9X; 6 point mutations and two variable microsatellite satellite repeats. Variation in the X chromosome is of a level comparable to that expected from theoretical mutation rates for this element. However, much less variation than expected was observed in the Y chromosome based on theoretical mutation rates and from previous studies on human Y chromosome variation. Our data suggest that bowheads have experienced a Y-chromosome selective "sweep" in the recent evolutionary past which contrasts markedly with a previously presented estimate of 1.2 million years for the time to most recent common ancestor for mtDNA. This paper describes a system for the analysis of X and Y chromosome variation applicable to the rest of the great whales.

Introduction

Recent interest in genetics of the bowhead whale (*Balaena mysticetus*) has stemmed from the endangered status of populations that were decimated by commercial whaling, the need to successfully manage aboriginal hunts in Alaska and Russia, and aspects of its unique biology including its longevity; it is the longest lived mammal with estimates up to 200 years for some individuals. The understanding and preservation of the genetic diversity of species and populations are key metrics in conservation biology. Since the decline of genetic diversity is well known to be associated with increased probability of extinction, accurate measures of genetic diversity and their temporal trends are useful data for population managers. By far the most commonly used genetic marker for population genetic studies of cetaceans is the mitochondrial DNA (mtDNA) control region. Since the mtDNA is maternally inherited, data from this marker is informative only for the female component of population genetics. Roman and Palumbi (2003) used this marker to estimate long-term female effective population sizes (N_{ef}) and from N_{ef} they calculated total population sizes for pre-whaling populations of North Atlantic fin, humpback, and minke whales. Compared to present day population estimates, historical population sizes for fin and humpback whales was 6 and 20 times current population estimates, respectively. The implication of this for the IWC is related to the RMP requirement that populations of cetaceans should not be harvested when below 54% of carrying capacity. If the historic population estimates based on mtDNA are accurate, then the North Atlantic populations of the three whale species studied by Roman and Palumbi (2003) likely would not qualify for harvest. However, caution must be used in the interpretation of population genetic estimators such as effective population size because many assumptions that go into the calculations are in themselves highly uncertain.

Therefore, estimates should be based not upon one locus but on multiple loci including ones linked to the different mammalian inheritance patterns; mtDNA (strict maternal inheritance), bi-parentally inherited autosomal loci, X-linked loci (two copies in females and one in males), and Y-linked loci with strict paternal inheritance. For bowhead whales population genetics data exist for mtDNA (LeDuc et al., 2008), autosomal microsatellites (Givens et al., 2010), and autosomal Single Nucleotide Polymorphisms (SNPs) (Morin et al., unpublished). This paper presents the results of sequence analyses of paralogous X- and Y-chromosomal loci useful in determining levels of sequence diversity characteristic of these two distinct inheritance systems.

The Ubiquitin Specific Peptidase-9 gene (USP9) was selected for analysis to determine levels of variability in bowhead whales. This gene functions to bind and modify ubiquitin. Ubiquitin is a regulatory protein found in all eukaryotic cells and involved in cellular control (Hochstrasser, 2009). This protein is involved in a number of physiological processes, including conjugating with other proteins to mark them for degradation by the ubiquitin-proteasome system. USP9 codes for an enzyme which removes ubiquitin from attached proteins by hydrolyzing the isopeptide bonds. Due to a translocation from the X to the Y, this is a paralogous gene pair in mammals; USP9X is found on the X chromosome, and USP9Y is found on the Y chromosome. Following translocation to the non-recombining region of the Y, USP9Y has evolved independently and gained its own function. This paralogy allows for a basis of comparison between the sex chromosomes. Because the functions of the two genes are similar, differences in mutation rates and evolutionary patterns are likely due to differences inherent to their different linkage groups.

There is obvious value in direct comparisons of population genetic or evolutionary studies between maternally inherited mtDNA and paternally inherited Y-chromosomal genes. Nevertheless, there are no population genetic studies of Y-linked genes for any cetacean species and in part this is due to the difficulty of discovering polymorphic markers for the Y chromosome. This problem has only been adequately addressed in humans where the phylogeography of males has been well studied but based on an extensive effort to sequence multiple genes. In particular, Shen et al. (2000) surveyed approximately 41,000 nucleotides per individual from three genes in a sample of 70 human Y-chromosomes. They observed 51 polymorphic sites in their study, or around 1 variant per 900 base pairs. This is an indication of how many base pairs need to be surveyed to find an equivalent amount of variation, assuming that variation in bowheads is comparable to that in humans. Examination of Figure 1 in Semino et al. (2010) illustrates the number and frequency of Y-chromosome haplotypes in local human populations. Given those numbers, it is clear that we would expect to observe multiple haplotypes with a sample size of 10 males.

The overview of male evolutionary history as determined by the population genetics of the human Y chromosome has helped set the framework for this study. Shen et al. (2000) have shown that polymorphic genes on the Y-chromosome are evolutionarily neutral, with few recurrent substitutions, and represent a useful system for the reconstruction of evolutionary history.

Methods

Spleen samples were obtained from 14 male and 4 female bowhead whales taken during subsistence hunts from 2008-2010 at Barrow, Alaska. To extract DNA, tissue samples were incubated at 55°C in a Longmire's solution with proteinase K for 24 hrs. Digested tissue

samples were then extracted with phenol:chloroform and precipitated essentially as described in Sambrook et al. (2001).

After extraction, samples were amplified for different regions along the USP9 gene using primers from adjacent exons for intron spanning. The PCR protocol was essentially that of Cronin et al. (1996), as modified by proper annealing temperatures of individual primer pairs. Appendix 1 lists the primers and temperatures used for amplification of the USP9 gene. Amplification mixture consisted of 5 µl of each primer, 10X dNTPs, 10X BSA, and10X salt solution, respectively, 1.5 units Taq polymerase, and 25 µl Sigma water, for a total reaction volume of 50 µl per sample. PCR reactions were carried out on a BioRad MyCycler or an ABI GeneAmp 2700 thermocycler. The cycle-sequence protocol was: 4 min at 94°C, followed by 36 cycles of a 30 sec denaturation step at 94°C, a 30 sec annealing step of 50-65°C (depending on specific needs of the primer pair), and a 2.5-3.5 min extension step at 70°C, depending on the expected size of the products. For a few primer pairs, a touchdown procedure was used, where after three cycles, the initial annealing temperature was reduced by three degrees for the subsequent three cycles, and then returned to the previous annealing temperature for the duration of the amplification reaction. PCR primers were mainly developed from GenBank data using the cow sequence, which represents the most closely related genomic sequence to whales (primer sequences available from the authors). Intron spanning PCR was employed in this study to amplify the more highly variable introns of the USP9 gene (Fig. 1). Amplified DNA products were separated by gel electrophoresis in a 0.8% agarose medium and stained with ethidium bromide; bands were visualized on a BioRad GelDox XR imaging system.

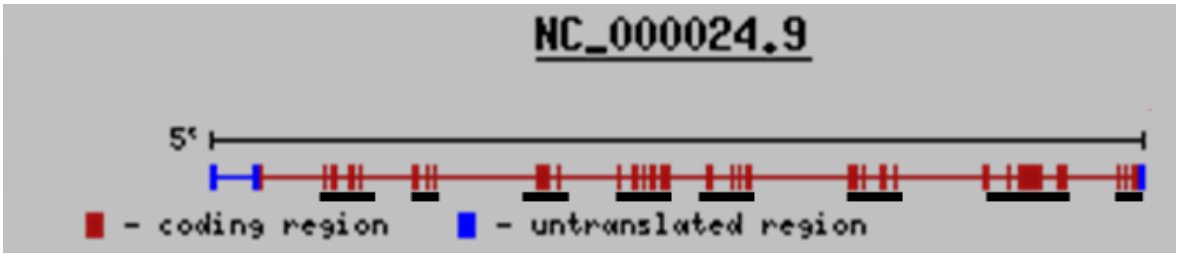

Figure 1.—Illustration of the structure of the USP9 gene from GenBank. Islands of exons that were targeted are underlined in black. PCR primers were developed from cow exon sequences and designed to span the intervening introns.

PCR products were either excised from the gel and purified with a Qiagen™ Band Excision Clean-up Kit or directly purified via Qiagen™ PCR Clean-up kit, depending on the complexity of the generated bands. An appropriate amount of purified DNA was then dried down and an ABI BigDye v 3.1 sequencing reaction was performed using the recommended protocol. Sequencing was performed on an ABI 3730 automated sequencer at the Purdue Genomics Core Facility. The sequences obtained were then blasted against known sequences in GenBank to confirm that they were indeed USP9 and determine their X or Y-chromosome specificity. X and Y specific primers used in further amplifications or sequencing were developed from these preliminary sequences as a perfect match to the *Balaena mysticetus* sequence; primer sequences are available from the authors. Sequences of amplification products as well as those taken from GenBank were aligned and assembled using Sequencher™ 4.7.

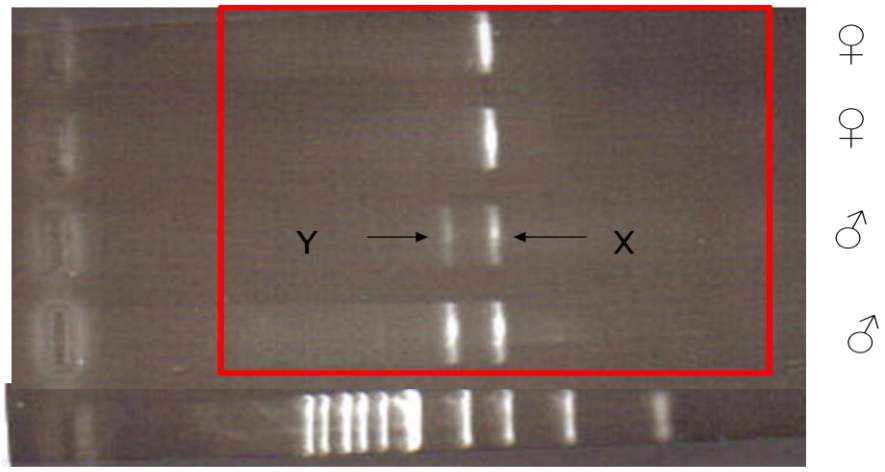

Figure 2.—The results of amplification of a fragment of the USP9Y and USP9X genes in the region between exons 14 and 15. Different banding patterns can be observed in males and females, with the common band being from USP9X and the band exclusively seen in male samples being from the USP9Y gene.

Results

The method of exon-primed, intron crossing (EPIC) amplification (Palumbi and Baker, 1994) provided successful amplifications of nine regions of the Y and eight regions of the X-chromosome. Figure 1 illustrates the underlying structure of the USP9 gene. "Islands" of clustered exons are found throughout the gene, and are underlined in black in the figure. The exons are flanked on each side by relatively small introns which were the targets of the sequencing efforts. The exons provided conserved sequence for the primers and allowed for the successful amplification of intervening introns by intron spanning. Clusters of these exon islands are flanked by very large introns, which are too long to allow successful amplification. The products obtained from individual islands are shown in Tables 1 and 2. Figure 2 shows the products of amplification of the fragment between exons 14 and 15 run out on a 0.8% agarose gel. The difference in banding patterns on the gel of amplified USP9X and USP9Y regions for males and females are due to differing sizes of the introns between the two paralogous genes. The common band between the four individuals is from USP9X. The extra band of the two male samples is a region from USP9Y. Band excision and gel DNA extraction was used to separate and sequence the X and Y-chromosome fragments of the amplifications.

A total of more than 32,000 nucleotides have been sequenced for USP9Y and USP9X for varying numbers of individuals. As can be seen in Table 1, this includes approximately 21,750 nucleotides sequenced for nine regions of USP9Y. Not all individuals have been sequenced for all regions of the two genes. For the Y-chromosome we sequenced a total of 138,350 bases which yields 6.36 X coverage of the 9 regions studied. A single variable site was found in intron 37. This variant is the result of an A/T transversion. Table 2 reports the results of sequence analysis of 11,150 base pairs for USP9X. For the X-chromosome we sequenced a total of 94,150 bases which yields 8.44 X coverage of the 8 regions studied. Eight variable sites were found; two transitions in intron 18 (A/G,C/T), one transition in intron 25 (C/T), one transition in intron 37 (A/G), a highly complex microsatellite region in intron 43, and in intron 45 there were three transitions (A/G, A/G, C/T) and a variable tetrameric microsatellite repeat (TATG). For the complex microsatellite region in intron 43,

we have sequenced 11 males from which 6 alleles were identified (Fig. 3). Because males have only one X chromosome, these sequences represent a haplotype, thus allowing unambiguous identification of the variants on the X-chromosome of each male. This is advantageous, as interpreting the complex patterns in a female heterozygote could be difficult.

```
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTC
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTC
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTC
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGC::::::::::::::::::::::::
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGC::::::::::::::::::::::::
GCAGGTGAGGGTCTCTCTCCCTCTCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGCTCTCTCTTGC::::::::::::::::::::::::

TCTTGC::TCTCTCTCTCTCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTC::::::TATATATATATATGTGTATATATATAC
TCTTGC::TCTCTCTCTCTCGCGCGCGCGC::::TCTCTCTCTCTCTCTCTCTC::TATATATATATATATATGTGTATATATATAC
TCTTGC::TCTCTCTCTCTCGCGCGCGCGC::::::TCTCTCTCTCTCTCTCTCTATATATATATATATATATGTGTATATATATAC
::::::TCTCTCTCTCTCT::CGCGCGCGC::::::TCTCTCTCTCTCTCTCTCTC::TATATATATATATATATGTGTATATATATAC
::::::TCTCTCTCTCTCT::CGCGCGCGC::::TCTCTCTCTCTCTCTCTCTC::::::TATATATATATATGTGTATATATATAC
::::::TCTCTCTCTCTCT::CGCGCGCGC::::::::::TCTCTCTCTCTCTCTCTATATATATATATATATATGTGTATATATATAC

ACACACACATATAATTTTGTAGAAATCT
ACACACACATATAATTTTGTAGAAATCT
ACACACACATATAATTTTGTAGAAATCT
ACACACACATATAATTTTGTAGAAATCT
ACACACACATATAATTTTGTAGAAATCT
ACACACACATATAATTTTGTAGAAATCT
```

Figure 3.—Nucleotide sequence of a segment of intron 43 of USP9X of 6 bowhead whales showing allelic variation in a complex microsatellite.

Table 1.—The number of sequenced base pairs for 2-12 bowhead whales in 9 regions of the USP9Y gene. "EX" denotes an exon.

Region	Bases Sequenced	Variable sites
EX 8- EX 11	**2,400**	
EX 14- EX 15	**1,800**	
EX 17- EX 20	**3,600**	
EX 23- EX 24	**550**	
EX 25- EX 26	**950**	
EX 28- EX 31	**2,800**	
EX 34- EX 39	**7,200**	**A/T transversion**

EX 40- EX 42	**950**	
EX 43- EX 46	**1,500**	
TOTAL:	**21,750 nucleotides**	**1 variable site**

Table 2.—The number of sequenced base pairs for 2-12 individuals in the USP9X gene. "EX" denotes an exon, "in" denotes an Intron.

Region	Bases Sequenced	Variable sites
EX 14- EX 15	**1300**	
EX 18- EX 19	**1000**	**A/G, C/T transitions**
EX 25- EX 26	**1050**	**C/T transition**
EX 28 -EX 31	**2100**	
EX 36- EX 38	**2150**	**A/G transition**
EX 40-in41	**1000**	
EX 43- EX 45	**1250**	**Complex Microsatellite**
EX 45- EX 46	**1300**	**G/A, G/A, C/T transitions; 1 variable tetrameric microsatellite**
TOTAL:	**11,150 nucleotides**	**9 variable sites**

Discussion

The objective of this study is to develop a system for sequence analysis of X- and Y-linked genes to include a sufficiently large number of nucleotides for meaningful documentation of the paternal and X-chromosomal evolutionary histories of the bowhead whale. We report here the results of a sequence analysis for approximately 21,000 bp of the bowhead Y and 11,000 bp of the bowhead X. These 32,000 nucleotides are located in several 'island' regions of exons, with spans of introns of generally less than 3 kilobase pairs (kbp) in length, allowing the use of a standard PCR protocol to amplify them.

On the Y chromosome, 21,750 base pairs of the USP9 gene were sequenced with 6.36 X coverage. From these sequences, we observed a single variable site in a single whale (11 individuals were sequenced for the variant position). This was a point mutation of a single nucleotide in the sequence of intron 37 (Table 1). This compares with our survey of 11,150 nucleotides on the USP9X gene with 8.44 X coverage where a total of 9 variable sites were discovered including seven point mutations and two variable microsatellites. For the complex microsatellite in intron 43 we sequenced 11 males and observed 6 different alleles. Because the X in males is haploid it is possible to obtain unambiguous sequence data for each variant of the complex repeat.

The neutral mutation rate of the Y-chromosome is expected to be greater than the autosomes which in turn are expected to be greater than the X-chromosome. This is because 100% of the Y-chromosomes pass through the male, 50% of the autosomes pass through the male, and 33.3% of the X-chromosomes pass through the male. Higher mutation rates are predicted for the Y-linked genes because of the greater number of germ cell divisions in spermatogenesis than in oogenesis (Miyata et al., 1987). Thus, the higher the percentage of chromosomes in a population that have passed through the male, the higher the relative mutation rate. Because spermatogenesis produces so many more gametes than oogenesis, mutations are much more likely to occur in the male lineage. This relationship of Y>autosomes>X has been confirmed empirically (Miyata et al., 1987). One of the most thorough studies to have compared X and Y-chromosome sequence evolution (Slattery and O'Brien, 1988) showed Y-chromosomal sequences evolved 2X faster than X-chromosomal sequences in a paralagous gene pair, Zfx and Zfy. But in population genetics, variation within a species is due to both mutation rate and effective population size N_e. Because N_e for males is much smaller than females, due to the high variance in reproductive success for males, new Y-chromosome variants come to fixation very quickly. Thus, the number of variants within a population will be reflected in the time since the last Y-chromosome sweep or the Time to Most Recent Common Ancestor (TMRCA).

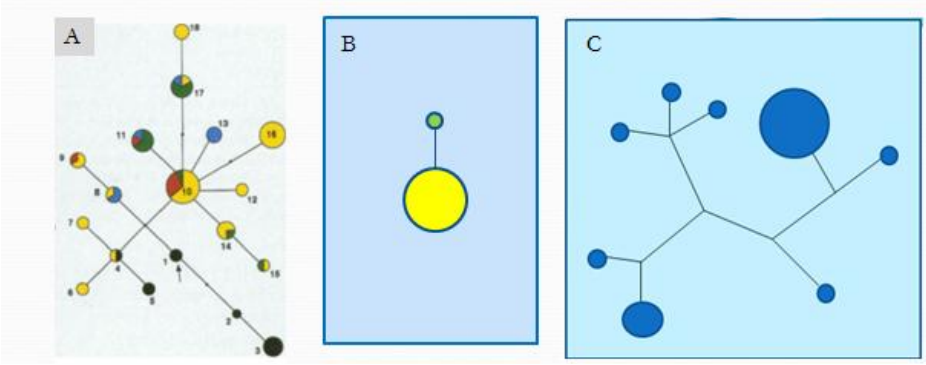

Figure 4.—Haplotype networks for: A) Shen et al. (2000) study on 16,000 nucleotides of the human SMCY gene demonstrating the amount of diversity in *Homo sapiens sapiens*, B) 21,750 nucleotides analyzed for the bowhead USP9Y gene demonstrating the amount of diversity observed thus far in the study, and C) 1,140 nucleotides analyzed for the mitochondrial gene, cytochrome B in bowhead whales which compares diversity of the maternally inherited mtDNA to the paternally inherited Y chromosome.

In comparison to the study by Shen et al. (2000), it is clear that bowhead whales have less variation in the Y-chromosome than is seen in humans. Shen et al. (2000) observed approximately 1 variant per 900 base pairs in the SMCY gene on the human Y chromosome. In bowheads thus far we have observed only 1 variant per 21,750 base pairs; a single point mutation was found in one individual. For comparison, we randomly chose 12 bowheads which were sequenced for the 1,140 nucleotides of the cytochrome B gene. Genes on the mtDNA coding regions evolve about 10 times more rapidly than is expected in the Y, thus 1,000 nucleotides on the mtDNA should show variation comparable to 10,000 nucleotides on the Y chromosome. Thus using 1,140 nucleotides from the cytochrome B gene should be an underestimate of the comparable amount of variation of 21,750 base pairs of the intron regions of the USP9Y gene. Yet there is still a clearly greater amount of diversity in cytochrome B than seen in the USP9Y gene. Figure 4 shows the phylogenetic networks of

A) 16,000 nucleotides of the SMCY gene for humans (Shen et. al., 2000), B) 21,750 nucleotides of the USP9Y gene in this study, and C) 1,140 nucleotides of the cytochrome B gene for bowheads. The variation demonstrated in the randomly sequenced 12 individuals for cytochrome B shows variation comparable to that seen by Semino his study of the Y chromosome haplotypes in European males. These studies are not comparable on the basis of sample size. The sample size for the human study is 1,000 individuals. Our sample size from the Y-chromosome averages 6.36 individuals over the 21,750 base pairs studied.

Conclusions

This study describes methods to analyze X and Y-chromosomal genetic loci for population studies of bowhead whales. The conserved PCR primers will allow the development of equivalent analyses in other cetacean species. In addition, the data show that X-chromosome haplotype diversity is high, or at least within expected levels, in bowheads but Y-chromosomal diversity is very low. This is most likely due to a recent selective sweep and indicates that male reproductive success likely is highly variable. This would be expected if sperm competition plays a role in bowhead reproductive strategy as has been suggested and it is consistent with the observation of "super males" with extremely large testes.

Acknowledgments

We thank the Alaska Eskimo Whaling Commission (AEWC) and the Barrow Whaling Captains' Association for their confidence, guidance and support of our research. We gratefully acknowledge funding provided by the North Slope Borough Department of Wildlife Management and National Oceanic and Atmospheric Administration (through the AEWC).

Literature Cited

Cronin, MA, Bodkin J, Ballachey B, Estes J, Patton JC (1996) Mitochondrial DNA variation among subspecies and populations of sea otters (*Enhydra lutris*). *Journal of Mammalogy*, 77, 546-577.

Hochstrasser, Mark (2009). Origin and function of ubiquitin-like proteins. *Nature* 458, 422–429.

Miyata T, Hayashida H, Kumar K, Mitsuyasu K, Yasunaga T (1987) Male-driven molecular evolution: A model and nucleo tide sequence analysis. Cold Spring Harbor Symp. Quant. Biol. 52, 863–867.

Palumbi SR, Baker SC (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution* 11:426-435.

Sambrook J, MacCallum P, Russel DW (2001) Molecular cloning: A Laboratory Manual Coldspring Harbor Press, 3rd addition.

Semino et al. (2010) The Genetic Legacy of Paleolithic *Homo sapiens sapiens* in Extant Europeans: A Y chromosome Perpsective. *Science*, 290: 1155-1159.

Shen, et al. (2000) Population genetic implications from sequene variation in four Y chromosome genes. *PNAS*, 97, 7354-7359.

Slattery JP, O'Brien SJ (1998) Patterns of *Y* and *IX* chromosome DNA sequence divergence during the Felidae radiation. *Genetics* 148**,** 1245–1255.

Thompson R, Pritchard JK, She P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *PNAS*, 97, 7360-7365.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner, PJ, Cavalli-Sforza LL (2000) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of Human Genetics*, 65: 43-62.