

Genetic polymorphisms in mitochondrial DNA hypervariable regions I, II and III of the Malaysian population.

Lian Lay Hoong^{1*} and Koh Chong Lek²

¹Department of Molecular Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.

²National Institute of Education, Nanyang Technological University, Nanyang Walk, Singapore 637616.

Received 20 June 2005 / Accepted 14 October 2005

Abstract. Nucleotide sequence variabilities in the three hypervariable (HV-) regions, HV1, HV2, and HV3, of the noncoding control region of human mitochondrial DNA (mtDNA) from a portion of the Malaysian population were elucidated with the use of polymerase chain reaction (PCR) appropriate gel detection and fluorescence based sequencing methodologies. Genomic DNA were extracted from 195 randomly selected individuals (15 samples from each of the 13 ethnicities ranging from Malay, Chinese, Indian, Punjabi, indigenous Sarawakian, indigenous Sabahan, and Orang Asli) and nucleotide sequence variabilities in HV1, HV2, and HV3 regions, were determined. These were located in the control region that contains sequences responsible for transcription and replication control of the mtDNA. The control region is located between the tRNA genes that encoded proline and phenylalanine, respectively. We found that the noncoding segments of the control region were polymorphic in the Malaysian samples. The polymorphisms within the control region exhibited a significant degree of diversity, thus enabling the three HV-regions of the control region of mtDNA to be used as additional markers in individual identification in forensic investigations, supplementing nuclear DNA markers. The results generated could also further complement anthropology and population studies in Asia.

Keywords. hypervariable regions, Malaysia, mtDNA, sequence polymorphism

INTRODUCTION

Human mitochondrial DNA (mtDNA) is a double-stranded circular molecule present in 1,000 to 10,000 copies per cell. The complete nucleotide sequence of the 16,569 base-pair (bp) molecule was determined in 1981 (Anderson et al., 1981).

The mitochondrial genome can be divided into two sections: a large coding region, which is responsible for the production of various biological molecules involved in the process of energy production in the cell, and a smaller 1.2 kilobase pair fragment, called the control region. It is found to be highly polymorphic and harbors three hypervariable regions (HV), HV1, HV2 and HV3. The HV-regions have been used extensively in practical forensic investigations, because mtDNA is stable during long storage, owing to various factors: its high copy number and circular form makes it less susceptible to exonuclease degradation (Parson et al., 1998). Because mtDNA is inherited through the mother, as long as an individual shares maternal descent with a candidate sample source, it can be used to verify identity. However, recent studies have also shown that there is new evidence for the occasional inheritance of paternal mitochondrial DNA (Kraytsberg et al.,

2004). MtDNA also evolves five to ten times faster than chromosomal DNA, and this relatively higher mutation rate gives rise to more polymorphic sites. Studies have also shown that the substitution rate in the control region is about ten times higher than that in the remainder genome, hence the presence of the HV regions (Parsons et al., 1997). MtDNA can be detected successfully in various biological samples, including blood, blood-stains, bone, buccal cells, faeces, hair, nails, skin, semen stains, teeth, and urine (Rousselet and Mangin, 1998).

MATERIALS AND METHODS

Population sample collection. Blood and throat washings were randomly obtained from 195 healthy unrelated volunteer

*Author for Correspondence.

Mailing address: Department of Molecular Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia. Tel: 603 7967 5740; Fax: 603 7967 4957; Email: lhlian@um.edu.my

Table 1. Oligonucleotide primers used in the PCRs. The Universal M13 and -21M13 sequences are in the italic form (ABI PRISM, 1998).

Primer name (Forward; F, Reverse; R)	Primer sequence (5' → 3') numbered according to Anderson <i>et al.</i> (1981)	Region amplified	Fragment size (bp)
HVI-F	<i>TGT AAA ACG ACG GCC AGT</i> - 16420-TGA TTT CAC GGA GGA TGG TG-16401	HV1(15975 to 16420)	484
HV1-R	<i>CAG GAA ACA GCT ATG ACC</i> - 15975-CTC CAC CAT TAG CAC CCA AAG-15995		
HV2-F	<i>CAG GAA ACA GCT ATG ACC</i> - 00008-GGT CTA TCA CCC TAT TAA CCA C-00029	HV2(00008 to 00429)	421
HV2-R	<i>TGT AAA ACG ACG GCC AGT</i> - 00429-CTG TTA AAA GTG CAT ACC GCC-00409		
HV3-F	<i>CAG GAA ACA GCT ATG ACC</i> - 00362-CAA AGA ACC CTA ACA CCA GC-00381	HV3(00362 to 00599)	255
HV3-R	<i>TGT AAA ACG ACG GCC AGT</i> - 00599-TTG AGG AGG TAA GCT ACA TA-00580		

donors, recruited from 13 ethnic groups comprising Malay, Chinese, Indian, Punjabi, Iban, Bidayuh, Kadazan-Dusun, Rungus, Bajau, Murut, Orang Sungai, Bisaya and Orang Asli. Each group was represented by 15 subjects.

DNA extraction and amplification. Genomic DNA was extracted from blood and throat washings following a standard phenol-chloroform method (Sambrook *et al.*, 1989). Amplification of HV1, HV2 and HV3 regions was carried out using three sets of primers encompassing the three HV-regions, respectively (Tables 1 and 2). PCR products were gel eluted and purified for sequencing.

DNA sequencing. Automated DNA sequencing was carried out using dye-terminator chemistries, on an automated DNA sequencer (ABI PRISM 377). The universal M13 forward and reverse primers were used accordingly for the sequencing of the direct PCR method for HV1, HV2, and HV3 following manufacturer's recommendations.

Analysis of data. The sequences were aligned and edited between positions 15994 to 16375 (for HV1), 00063 to 00350 (for HV2) and 00428 to 00584 (for HV3). According to the TWGDAM (Technical Working Groups for DNA Analysis Methods), a group of forensic research laboratories in the United States that sets standards for DNA technology, the minimum sequence that will be accepted for the mtDNA database for HV1 is from positions 16024 to 16365, and for HV2, from positions 00073 to 00340. In this study, we have reaffirmed the existence of HV3 (from positions 00438 to 00574) as reported by Lutz and his co-workers. (Lutz *et al.*, 1998).

Table 2. Parameters of PCR conditions (35 cycles) for amplifying HV-regions of mtDNA.

Cycling condition	Temperature (°C)	Duration
Initial denaturation	94	5 minutes
Denaturation	94	45 seconds
Annealing	66	1 minute
Extension	72	1 minute
Final extension step	72	5 minutes

RESULTS AND DISCUSSION

Data of nucleotide polymorphisms for HV1, HV2 and HV3 are presented in Figures 1, 2 and 3, respectively. We found that in HV1, HV2 and HV3, nucleotide polymorphisms were evident at expected positions. These results are summarised in Figures 1, 2 and 3. The novel mutations are those that are typed in bold face fonts.

A total of 164 polymorphic sites were observed, 108 in HV1, 41 in HV2, and 15 in HV3 (Figures 1, 2 and 3). Owing to the fact that some of these variable sites had more than one type of polymorphism, we report a total of 182 polymorphisms, with 118 in HV1, 46 in HV2, and 18 in HV3. Over 70% of the individuals analysed exhibited between 9 to 12 polymorphisms in the control region. Nucleotide substitutions (78.61%) were the most common, compared to insertions (16.48%) and deletions (4.93%). The most common transition scored was A → G, while A → C substitution was the most frequent transversion. The most common insertion encountered was an additional C residue while the most frequent deletion involved a dinucleotide repeat of CA.

16024 ↓	TTCTTTTCATG	GGGAAGCAGA	TTTGGGTACC <i>G</i>	ACCCAAGTAT	TGACTCACCC <i>T</i>	50
	ATCAACAACC	GCTATGTATT <i>C CC</i>	TCGTACATTA	CTGCCAGCCA <i>T T T</i>	CCATGAATAT <i>T</i>	100
	TGTACGGTAC <i>C C A</i>	CATAAATACT <i>C C</i>	TGACCACCTG <i>A TT</i>	TAGTACATAA <i>C G</i>	AAACCCAATC <i>CTTTG C</i>	150
	CACATCAAAA <i>T CC</i>	CCCCCTCCCC <i>TTTT C TT</i>	ATGCTTACAA <i>G</i>	GCAAGTACAG <i>C GA</i>	CAATCAACCC <i>GCT GT T</i>	200
	TCAACTATCA <i>C C</i>	CACATCAACT <i>TG TG TC</i>	GCAACTCCAA <i>GTC</i>	AGCCACCCCT <i>TTGATTTT</i>	CACCCACTAG <i>GT GTCG</i>	250
	GATACCAACA <i>AG T</i>	AACCTACCCA <i>T C TTTG</i>	CCCTTAACAG <i>TTTCC</i>	TACATAGTAC <i>C G C</i>	ATAAAGCCAT <i>GCCAT</i>	300
	TTACCGTACA <i>C T</i>	TAGCACATTA <i>G T</i>	CAGTCAAATC <i>G T C T CG</i>	CCTTCTCGTC <i>TTCC T C</i>	CC <i>T</i>	342
				<i>G</i>	↑ 16365	

Figure 1. Variable nucleotide sites of HV1. The reference sequence of Anderson *et al.* (1981) is shown in the lines along with the numbers indicating the total number of bases. Polymorphisms observed in at least one individual are shown in italic below the reference sequence. Novel mutations are in bold.

A total of 172 different lineages were observed in HV1, 91 in HV2, and 26 in HV3. Consequently, 185 unique and 5 shared mtDNA lineages were identified in subjects based on all the three regions in combination. The probability of two randomly selected individuals having identical mtDNA types in a population is 0.54%. The genetic diversity was noted to be the highest in HV1 (0.998) and the lowest in HV3 (0.804). The genetic diversity in HV2 was 0.985. The combined genetic diversity value of these three regions was 0.999. Table 3 presents a summary of the Malaysian data in comparison with other global populations.

Novel polymorphisms were scored in all the three HV-regions (Figures 1, 2 and 3). A total of 18 new polymorphisms, i.e., nucleotide changes that have yet to be reported in the MITOMAP database, were discovered. These polymorphisms were subsequently deposited in the MITOMAP database under appropriate accession numbers (Table 4).

MITOMAP is a database in which data relating to mtDNA sequence variation are organised in a format useful to researchers and clinicians investigating mitochondrial diseases and biodiversity (Kogelnik *et al.*, 1996, 1998). It is very useful to researchers, forensic scientists, population and evolutionary biologists, and anthropologists, in that it allows them to analyse in detail mtDNA variation.

Most of the novel polymorphisms (59%) discovered in the Malaysian population were detected in the indigenous samples from East Malaysia, while the remaining novel polymorphisms were scored in the Malay, Chinese, Indian, and Punjabi groups from Peninsular Malaysia. While a large number of the unreported polymorphisms were unique, the most common polymorphism was the T → C transition at site 16157 (4%).

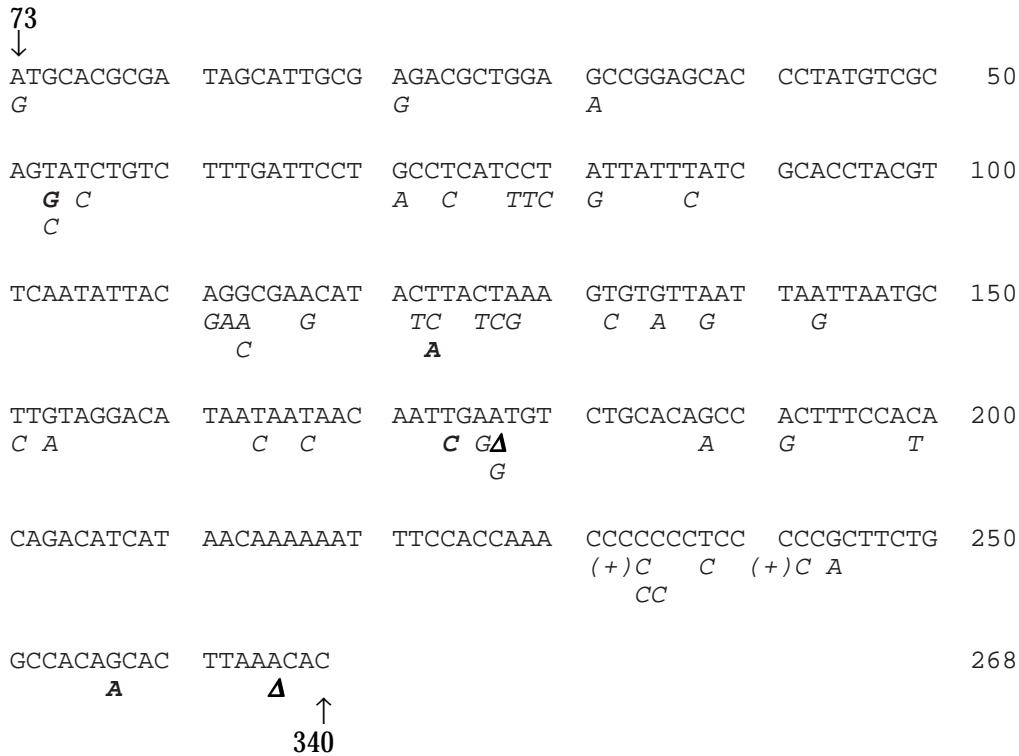


Figure 2. Variable nucleotide sites of HV2. The reference sequence of Anderson *et al.* (1981) is shown in the lines along with the numbers indicating the total number of bases. Polymorphisms observed in at least one individual are shown in italic below the reference sequence. Novel mutations are in bold. (+) indicates the presence an insertion or insertions, while Δ represents a deletion.

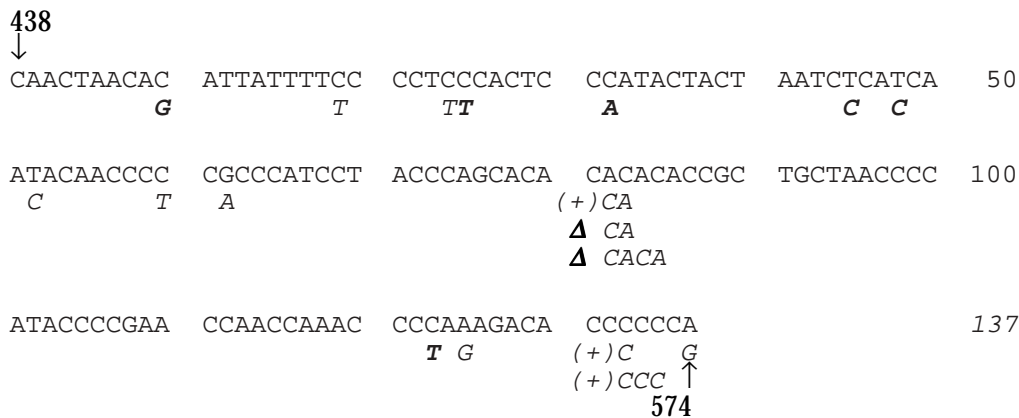


Figure 3. Variable nucleotide sites of HV3. The reference sequence of Anderson *et al.* (1981) is shown in the lines along with the numbers indicating the total number of bases. Polymorphisms observed in at least one individual are shown in italic below the reference sequence. Novel mutations are in bold. (+) indicates the presence of an insertion or insertions, while Δ represents deletions.

Table 3. Comparisons of the characteristics across HV1 and HV2 of the control region in different human population groups.

Population	Indian ^a	Japanese ^b	Malaysian ^c	Korean ^d	African ^e	USCaucasian ^f	Austrian ^g	German ^h	French ⁱ
Sample size	98	162	195	306	111	604	109	200	109
No. of variant sites	83	133	149	197	97	233	124	153	103
A→G	233	392	473	758	323	1112	181	330	194
G→A	66	86	81	164	78	219	39	55	29
T→C	145	405	461	621	382	1007	170	308	158
C→T	117	285	321	563	486	688	122	199	121
% transition	94.85	92.92	92.16	95.08	95.77	97.61	97.34	95.61	98.62
A→T	1	3	2	13	0	2	0	4	1
A→C	23	70	81	33	15	47	9	5	4
G→T	0	0	0	2	18	1	0	0	0
G→C	0	0	3	9	0	6	1	1	1
C→A	0	15	30	28	17	12	2	11	0
C→G	4	1	1	15	6	6	1	19	1
T→A	7	0	5	3	0	0	1	1	0
T→G	0	0	3	6	0	0	0	0	0
% transversion	5.15	7.08	7.84	4.92	4.23	2.39	2.66	4.39	1.38
Insertion	168	282	322	308	140	983	188	291	181
Deletion	0	22	28	39	6	14	7	6	0

Note: % of transitions and transversions were calculated as number of observations divided by total substitution times 100.

^a Mountain *et al.*, 1995; ^{b, c, e, f, i} Budowle *et al.*, 1999; ^c This study; ^d Lee *et al.*, 1997; ^g Parson *et al.*, 1998; ^h Lutz *et al.*, 1998.

CASE STUDY

The protocol used in this study was tested on a forensic case. It involved the discovery of some highly degraded bone samples in Taman Negara, the National Park situated in Pahang. The samples were suspected to belong to a Caucasian female, who was reported missing for some months somewhere in the vicinity. Subsequently, DNA from the bone samples was extracted at the Chemistry Department, Petaling Jaya, Selangor. For comparison, a blood sample was obtained from the putative mother, and mtDNA typing was carried out with DNA extracted from both the blood and bone samples. All three HV-regions were successfully amplified by using the direct PCR protocol. Both the 5'-3' and 3'-5' strands of the mtDNA were sequenced by using the universal and reverse M13 primers, respectively.

The nucleotide sequences derived from the putative mother's blood DNA sample were clear and clean. The nucleotide sequences obtained from the bone DNA samples were comparatively not as good, but both HV1 and HV3 gave consistent sequences of both strands. The data generated from

HV2 sequencing were not so clear, owing to the high background, but were sufficient to distinguish the polymorphisms.

When compared with the reference sequence (Anderson *et al.*, 1981), DNA from the bone and blood samples both had a total of eight nucleotide differences, respectively. The nucleotide polymorphisms in the bone DNA included three from HV1 [16111T, 16223T and 16231C], three from HV2 [199C, 263G and 311 to 315(+)-C], and two in HV3 [489C and 514 to 523(-)CA]. On the other hand, the eight changes detected in the blood DNA consisted of three polymorphisms from HV1, i.e., at positions 16224C, 16311C, and 16362T; four in HV2, i.e., at positions 73G, 263G, 303 to 309(+)-C, and 311 to 315(+)-C; and one from HV3, i.e., 568 to 573(+)-C (Table 5).

Alignment and comparison of the sequences obtained from the bone and blood DNA samples showed that the two sequences were not identical (Table 5). Therefore, it was concluded that the bone sample did not come from a person related to the donor of the blood samples, and hence, did not belong to the suspected victim.

Table 4. List of the novel polymorphisms discovered and their respective accession numbers in the MITOMAP database.

Control region	Base number	Change	Accession no.
HV1	16157	T → C	20000211006
	16157	T → A	20000915003
	16187	C → G	20000224001
	16192	C → A	20000211007
	16259	C → A	20000914004
	16309	A → C	20000914005
	16317	A → C	20000211008
	16318	A → C	20000914006
	16337	C → T	20000915001
	HV2	125	T → G
195		T → A	20000211002
246		T → C	20000211003
329		G → A	20000211005
HV3	447	C → G	20000915002
	463	C → T	20000914001
	469	C → A	20000211009
	482	T → C	20000211001
	485	T → C	20000914002
	560	C → T	20000914003

Table 5. Variable sites scored in the analysis of mtDNA. All nucleotide substitutions observed in HV1, HV2, and HV3 were transitions. In addition, there were three examples of insertions and an observed deletion. Ref stands for the reference sequence (Anderson *et al.* 1981).

Region and Position	Samples		
	Ref	Deceased (Bone Sample)	Putative Mother (Blood Sample)
HV1			
16111	C	T	C
16223	C	T	C
16224	T	T	C
16231	T	C	T
16311	T	T	C
16362	C	C	T
HV2			
73	A	A	G
199	T	C	T
263	A	G	G
from 303 to 309	(C) ₇	(C) ₇	(C) ₈
from 311 to 315	(C) ₃	(C) ₆	(C) ₆
HV3			
489	T	C	T
from 514 to 523	(CA) ₅	(CA) ₄	(CA) ₅
from 568 to 573	(C) ₆	(C) ₆	(C) ₇

HV3 has proven to be useful as an 'auxiliary' polymorphic segment for forensic application. Although the majority of forensic cases only use nucleotide sequences from HV1 and HV2, it should be noted that HV3 has the potential to 'enhance' mtDNA as a marker, more so in forensic investigations (Ivanov *et al.*, 1996; Lutz *et al.*, 1998).

While this is a pilot study, we sincerely hope that the population data generated will enable the loci to be potentially considered as a basic marker in forensic testing and maternity/paternity determination, as well as for further anthropological studies in Malaysia.

ACKNOWLEDGMENT

This study was supported by (1) a medical research grant from the Dr. Ranjeet Bhagwan Singh Medical Research Fund awarded to Lian LH; and (2) a PJP (Vot F 0251/98) grant from the Faculty of Science, University of Malaya, awarded to Lian LH. The authors would like to thank Ms Lim Moo Eng for her invaluable technical assistance.

REFERENCES

- Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290: 457-465.
- Budowle, B., Wilson, M.R., DiZinno, J.A., Stauffer, C., Fasano, M.A., Holland, M.M. and Monson, K.L. 1999. Mitochondrial DNA regions HVI and HVII population data. *Forensic Science International* 103: 23-35.
- Ivanov, P.L., Wadhams, M.J., Roby, R.K., Holland, M.M., Weedn, V.W. and Parsons, T.J. 1996. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nature Genetics* 12: 417-420.
- Kogelnik, A.M., Lott, M.T., Brown, M.D., Navathe, S.B. and Wallace, D.C. 1996. MITOMAP: A human mitochondrial genome database. *Nucleic Acids Research* 24: 177-179.
- Kogelnik, A.M., Lott, M.T., Brown, M.D., Navathe, S.B., and Wallace, D.C. 1998. MITOMAP: A human mitochondrial genome database - 1998 update. *Nucleic Acids Research* 26: 112-115.

- Kraytsberg, Y., Schwartz, M., Brown, T.A., Ebralidse, K., Kunz, W.S., Clayton, D.A., Vissing, J., and Khrapko, K. 2004. Recombination of Human Mitochondrial DNA. *Science* 304: 981.
- Lee, S.D., Shin, C.H., Kim, K.B., Lee, Y.S. and Lee, J.B. 1997. Sequence variation of mitochondrial DNA control region in Koreans. *Forensic Science International* 87: 99-116.
- Lutz, S., Weisser, H-J., Heizmann, J. and Pollak, S. 1998. Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *International Journal of Legal Medicine* 111: 67-77.
- Mountain, J.L., Hebert, J.M., Bhattacharyya, S., Underhill, P.A., Ottolenghi, C., Gadgil, M. and Cavalli-Sforza, L.L. 1995. Demographic history of India and mtDNA-sequence diversity. *American Journal of Human Genetics* 56: 979-992.
- Parson, W., Parsons, T.J., Scheithauer, R. and Holland, M.M. 1998. Population data for 101 Austrian Caucasian mitochondrial DNA D-loop sequences: Application of mtDNA sequence analysis to a forensic case. *International Journal of Legal Medicine* 111: 124-132.
- Parsons, T.J., Muniec, D.S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M.R., Berry, D.L., Holland, K.A., Weedn, V.W., Gill, P., and Holland, M.M. 1997. A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics* 15: 363-367.
- Rousselet, F. and Mangin, P. 1998. Mitochondrial DNA polymorphisms: a study of 50 French Caucasian individuals and application to forensic casework. *International Journal of Legal Medicine* 111: 292-298.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. Molecular Cloning: A laboratory manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- ABI PRISM Mitochondrial DNA sequencing. 1998. Technical Booklet, Perkin-Elmer.