2022

MHC Diversity Analysis of Spheniscus demersus for in situ and ex situ Populations

Antonieta van den Berg Monsalve
vandenbergmonsalve1@otterbein.edu

MHC Diversity Analysis of *Spheniscus demersus* for in situ and ex situ Populations

Otterbein University
Department of Biology and Earth Science
Westerville, Ohio 43081
Antonieta van den Berg Monsalve

Submitted in partial fulfillment of the requirements for
Graduation with Honors

Simon K. Lawrance Ph. D.
Project Advisor

Advisor's Signature

Halard Lescinsky Ph. D.
Second Reader

Second Reader's Signature

John Tansey Ph. D.
Honors Representative

Honors Representative's Signature

Acknowledgments

This research and thesis could not have been completed without the support and guidance of Dr. Simon Lawrance. I'm grateful for his expertise as I was able to cultivate my experience as a scientist and contribute to a contemporary conservation project that has been almost ten years in the making. I would like to extend my gratitude to Dr. Mike Cranfield and Mystic Aquarium whose penguin sample contributions made this research possible. Many thanks to my predecessors Athena Schalk, Caitlin E. Castagno, Courtney Kast, and Kelly M. Huth for paving the way for fellow undergraduate researchers as myself. The research for this thesis was made possible through Otterbein's Student Research Fund, whose support financed much of the lab work necessary for our project. Special thanks to my second reader, Dr. Halard Lescinsky, and my Honors Representative, Dr. John Tansey, for their time and support. Finally, I want to thank Erin Ulrich for supplying the Tris acetate EDTA and other necessary reagents for gel electrophoresis.

Abstract

The South African Penguin has been listed as endangered by the International Union for Conservation of Nature. Major histocompatibility complex (MHC) alleles provide valuable statistics because of their variation in terms of vertebrate genomes and are pertinent to immune and reproductive health. A greater number of MHC alleles correlates with survivability of a population because the genes control the immune and reproductive systems. Legacy research by Otterbein alumni and Dr. Simon Lawrance studied MHC samples from wild, in situ, and captive, ex situ, penguin populations. By comparing these populations' major histocompatibility complexes through biostatistical analysis, contributions to conservation of the species both inside and out of the wild can be made. One goal of this research was to compare the MHC of in situ and ex situ penguins through analysis of previously obtained allele sequences and identification of unreported alleles. Another goal was to expand the scope of the previous work done by colleagues through applying new primers to old samples. This primer set encompasses the entirety of Exon 2 which contains the coding sequences for the MHC class II beta chain. Out of six prior MHC class II samples, three unreported alleles were analyzed as potential new alleles. The first polymerase chain reaction product of penguin blood using the new primers, Lpen.hum1F and Lpen.hum2R, was sequenced and its results show a promising level of polymorphism across the length of the sequence.

Table of Contents

Introduction

As a species' population declines, so does its genetic variation. This loss of genetic diversity can trigger health problems due to inbreeding and weakened immune systems. Human interference and habitat loss have contributed greatly to the population decline of South African Penguins. Egg harvests, oil spills, and food scarcity has driven the population of in situ penguins from a bustling half million to a meager 50,000 [11]. According to the International Union for Conservation of Nature, South African penguins, *Spheniscus demersus*, are currently on the endangered species list [11]. With the low population size of *Spheniscus demersus* in the wild and in captivity their genetic variation as a species is at risk. An important step towards conservation is determining if the wild population's genetic diversity is represented within the captive population. One conservation method to keep the penguin populations healthy is to use the ex situ populations as a genetic bank for the species survival, but this is only viable if the population is diverse enough to survive future adversity.

The current goal of the Lawrance Lab is to assess the severity of the genetic loss, using the major histocompatibility complex (MHC) class II genes to compare the wild (in situ) populations and captive (ex situ) populations. Major histocompatibility complex alleles correlate with survivability of a population because the genes control the immune and reproductive systems. [26]. Critical contributions by Otterbein University alumni Athena Schalk, Caitlin E. Castagno, Courtney Kast, Kelly M. Huth, and research advisor Dr. Simon Lawrance provided the information and data necessary to continue the laboratory's legacy research [5, 10, 13, 25]. The samples covering the in situ population of *Spheniscus demersus* hail from Dassen Island, Dyer Island, and Robben Island off the coast of South Africa, and were supplied by Dr. Cranfield [25].

Mystic Aquarium has also collaborated in the past to provide samples of their ex situ penguins [5]. The samples analyzed in this research were provided by Mystic Aquarium and Dr. Cranfield.

The major histocompatibility complex is a highly variable genome component and a focal research point in many conservation studies [3,19]. The MHC gives rise to the immune system by protecting the organism from pathogens and toxins [12]. MHC proteins exhibit extraordinary polymorphism [27]. An alpha chain and beta chain form the two halves of MHC class II proteins, which are generally expressed on the surface of B lymphocytes, dendritic cells, and macrophages [1]. MHC class II proteins strictly bind exogenous proteins, such as bacterial or microbial proteins that were taken into the cell [27].

Genetic diversity and pathogen resistance are two important points in the conservation of this species. Ex situ populations are especially susceptible to aspergillosis and malaria [5]. Aspergillosis is caused by a fungus and is characterized by respiratory tract disease which can lead to death [7]. Cases of avian malaria, caused by exposure to *Plasmodium* by a mosquito vector, are also responsible for penguin illness and deaths [4]. While aquariums are equipped with their own veterinarians and ways to reduce incidences of disease, the immune system of the penguins plays a large role in recovery from illnesses.

In various studies it was concluded that MHC class II genes are pertinent to the assessment of genetic health of the penguins, and provide useful tools for conservation management [15,28]. Kikkawa et al. studied the polymorphism of the MHC class II genes in four different species of penguins belonging to the genus *Spheniscus*. The team's methodology included: penguin sample collection, DNA extraction, PCR, sequencing, and sequence data analysis [15]. Nucleotide sequences were compared between penguins to identify deletions within introns and allele percentage frequencies were calculated [15]. The results from Kikkawa

et al. showcase a similar process to that of Bollmer et al. where the focus was on the MHC of Galápagos penguins [3]. Both studies used computer software to assemble and edit the gathered sequences in order to identify nucleotide anomalies between penguins.

By comparing these populations' MHC class II genes through biostatistical analysis, contributions to conservation of the species both inside and out of the wild are possible. Two possible problems that could arise, in light of the species' current situation. Captive populations might lack significant genetic variation, or both ex situ and in situ populations have lost significant genetic diversity and the lack of variance would be unsustainable for future generations of South African penguins. The research in this thesis aims to provide a starting point for greater insight into conservation efforts by analyzing the MHC of *Spheniscus demersus* in regards to the *DRB1*-like region of the MHC class II molecule and beyond.

Methods

Bioinformatic Techniques:

Previously completed wet lab work by the Lawrance lab found six different alleles in the overall *Spheniscus demersus* population by using polymerase chain reaction (PCR) and denaturing gradient gel electrophoresis (DGGE) [10]. These samples were PCR 182, PCR 172, PCR 181, PCR 171, PCR 173, and PCR 111. After gathering sequences from prior research, six nucleotide sequences were chosen for analysis: PCR 182, PCR 172, PCR 181, PCR 171, PCR 173, and PCR 111. These sequences were chosen based on their incidence, as to include only sequences that were dissimilar. The blood samples used for PCR 172, PCR 181, PCR 171, and PCR 173 were provided by Mystic Aquarium in 2012 [13]. The blood sample used for PCR 172 was collected from a male penguin named Gold. PCR 181 and PCR 173 were both carried out

with blood sampled from a female penguin named Green. Another male penguin named Chocolate provided the blood sample for PCR 171. Dr. Mike Cranfield sent in the blood sample for PCR 111 in 2008. It was collected from a penguin in Dyer Island, off the coast of South Africa [25]. The origin of the penguin for PCR 182 could not be found.

NCBI BLAST was used to search the nucleotide database in order to confirm the identity of the alleles [20]. Then the nucleotide sequences were translated using the ExPASy translation tool and the correct frame was chosen by searching the BLAST protein database [6,2]. The amino acid sequences of these alleles were then aligned to each other using the National Center for Biotechnology Information's Basic Local Alignment Search Tool in order to identify the amino acid substitutions between all six sequences [2].

Once confirmed, the sequences were aligned to the DRB region of a human major histocompatibility class II protein structure, 1j8h, provided by Hennecke and Wiley using BLAST [8,2]. The six sample sequences were superimposed onto Lefranc's IMGT/Collier-de-Perles result for 1j8h using image editing software Paint.NET [17].

Polymerase Chain Reaction:

The first round of polymerase chain reaction (PCR) included four samples containing avian DNA labeled: WB (White Bird), BB (Blue Bird), PB (Penguin Blood), CL (Chicken Liver), and a fifth sample as a negative control with dH_2O. PCR was completed using primers pen1 and pen4, and served to amplify the DRB like region within Exon 2 of the MHC class II gene [14]. The sequences for the primers are 5' AACGGCACCGAGCGGGTGAGGT 3' for pen 1 and 5' CCCGTAGTTGTGTTGGCAG 3' for pen4, which had been used in previous Lawrance Lab research [5, 10, 13, 14, 25]. A stock solution of the pen1 and pen 4 primers was made by

combining 10 µl of pen1 primer, 10 µl of pen4 primer, and 80 µl of dH_2O. Then 1 µl of the primer stock solution was combined with 12.5 µl of Buffer D, 10 µl of dH_2O, 0.5 µl of Failsafe enzyme, and 1 µl of template DNA into a PCR microtube. Thirty-five cycles of PCR were performed in a thermocycler after a first hold at 95℃ for 10 minutes. Each cycle ran at 95℃ for thirty seconds, 55℃ for thirty seconds, and 72℃ for thirty seconds, and after the thirty-fifth cycle the sample was held for a second time at 72℃ for thirty minutes. The resulting samples were labeled WB1, BB2, PB3, CL4, and 05.

Using the prior four samples WB, BB, PB, and CL as well as a control with dH_2O, the second round of PCR served to amplify the region of the MHC class II gene which encompasses all of Exon 2. It was completed using primers Lpen.hum1F and Lpen.hum2R. The sequences for the primers are 5' GTCAGGATCACACCGCCACTC 3' for Lpen.hum1F and 5' ACACGCTCTCCCCTCCTGTG 3' for Lpen.hum2R [14]. A stock solution of the Lpen.hum1F and Lpen.hum2R primers was made by combining 10 µl of Lpen.hum1F primer, 10 µl of Lpen.hum2R primer, and 80 µl of dH_2O. Then 1 µl of the primer stock solution was combined with 12.5 µl of Buffer D, 10 µl of dH_2O, 0.5 µl of Failsafe enzyme, and 1 µl of DNA template into a PCR microtube. Thirty-five cycles of PCR were performed in a thermocycler after a first hold at 95°C for 10 minutes. Each cycle ran at 95℃ for thirty seconds, 55℃ for thirty seconds, and 72℃ for thirty seconds, and after the thirty-fifth cycle the sample was held for a second time at 72°C for thirty minutes. The PCR product samples were labeled WB11, BB12, PB13, CL14, and 015.

Gel Electrophoresis:

In order to confirm the difference in size between both PCR reactions, gel electrophoresis was performed. A Tris acetate EDTA buffer was prepared using 250 mL of dH_2O, 5ml of TAE 50x stock solution, and 1μL of ethidium bromide. Then, 50 mL of the TAE buffer were combined with .5 g of agarose to make an 1% agarose gel. The agarose solution was heated in a microwave for 1 minute and 33 seconds. The agarose gel solution was allowed to cool to 55°C before being poured into a gel tray with a comb and allowed to set. After placing the set gel in the buffer tank, the remaining 200 mL of the TAE buffer were poured into the buffer tank. In the first well, 5 μL of a 100 base pair ladder were loaded in. For each following well, 1 μL of 6x loading dye was added to 5 μL of sample and mixed before 5 μL of the combined solution were pipetted into each well. For the ten well gel, five samples of the second PCR reaction, which used the Lpen.hum1F and Lpen.hum2R primers, were loaded starting in the second lane of the gel. Four samples of the first PCR reaction using the pen1 and pen4 primers were loaded starting in the seventh lane of the gel. Finally, the samples were electrophoresed for 35 minutes at 100 volts. After 35 minutes of gel electrophoresis, the gel was removed and placed on a transilluminator for photo documentation.

DNA Sequencing:

The PB13 product resulting from the second PCR reaction was prepared for sequencing. Based on the gel electrophoresis results, 5 μL of the PB13 PCR product were combined with 5 μL of dH_2O into a thermocycler tube. After 1 μL of shrimp alkaline phosphatase and 0.5 μL of exonuclease I were added, the tube was incubated in the thermocycler for 15 minutes at 37℃ and 15 minutes at 80℃. Once the two holds were complete, 5 μL of the treated PB13 solution were aliquoted with 5 μL of dH_2O into two separate 1.5 mL microcentrifuge tubes: LPF and LPR.

Into a tube labeled LPF, 5 µL of the forward primer, Lpen.hum1F, were added. For tube LPR 5 µL of the reverse primer Lpen.hum2R were added. These steps were repeated with PB3 from the first PCR reaction, but primers pen1 and pen4 were added at the end instead Lpen.hum1F and Lpen.hum2R. The tubes were assigned barcodes and sent to Eurofins for sequencing.

Results

Nucleotide Sequences and Confirmation

The gathered sequences for PCR 182, PCR 172, PCR 181, PCR 171, PCR 173, and PCR 111 are shown below in Table 1.

Table 1: Six sequences from PCR 182, PCR 172, PCR 181, PCR 171, PCR 173, PCR 111, and their sizes.

Sample:	Length:	Nucleotide Sequence (5' to 3'):
PCR 182	157 bp	ATGTGGAGAGGTACATCCACAACCGGCAGGAGTTACTGCACTTCGACAGCGACGTGG GGTACTATGTGGCCGACACCCCCCTGGGTGAGCCTGATGCCAAGTACTGGAACAGCC AGACGGACATACTGGAGGATGAACGGGCTGCGGTGGACACGTA
PCR 172	157 bp	TCGTGCACAGGGACATCTACAACCGGCAGCAGGACGTGCACTTCGACAGCGACGTG GGGTACTATGTGGCCGACACCCCCCTGGGTGAGCCTGATGCCAAGTACTGGAACAG CCAGACGGACTTACTGGAGCAGAGACGGGCTGCGGTGGACACGTA
PCR 181	157 bp	ATGTGGTGAGGGACATCTACAACCGGCAGCAGGACGTGCACTTCGACAGCGACGTG GGGCAGTTTGTGGCCGACACCCCCCTGGGTGAGCCTGATGCCAAGTACTGGAACAG CCAGACGGACTTCCTGGAGCAGAAACGGGCTGCGGTGGACACGTT
PCR 171	157 bp	ATGTGGTGAGGGCCATCTACAACCGGCAGCAGTACGTGCACTTCGACAGCGACGTG GGGCAGTTTGTGGCCGACACCCCCCTGGGTGAGCCTGATGCCAAGTACTGGAACAG CCAGACGGACATACTGGAGCAGAAACGGGCTGAGGTGGACACGGT
PCR 173	157 bp	ATGTGGTGAGGGCCATCTACAACCGGCAGCAGTACGTGCACTTCGACAGCGACGTG GGGCAGTTTGTGGCCGACACCCCCCTGGGTGAGCCTTCTGCCAAGTACTGGAACAG CCAGACGGACATACTGGAGCAGAAACGGGCTGAGGTGGACACGGT
PCR 111	157 bp	ATGTGGTGAGGTACATCTACAACCGGCAGCAGTACGTGCACTTCGACAGCGACGTGG GGCAGTTTGTGGCCGACACCCCCCTGGGTGAGCCTTCTGCCAAGTACTGGAACAGC CAGACGGACATACTGGAGCAGAAACGGGCTGAGGTGGACACGGT

These sequences were used to search for reported alleles from other studies using NCBI BLAST. Matches for *Spheniscus demersus* were chosen based on percent identity; any other

penguin species were ignored. PCR 182 matched to AB301478.1 in NCBI's GenBank, which corresponds to cape001, with a 100% identity and an E value of 2e-74 [21]. PCR 172 matched to AB301944.1 in NCBI's GenBank, which corresponds to cape002, with a 100% identity and an E value of 2e-74 [22]. PCR 181 matched to AB301945.1 in NCBI's GenBank, which corresponds to cape003, with a 99% identity and an E value of 6e-70 [23]. PCR 171 matched to AB301946.1 in NCBI's GenBank, which corresponds to cape004, with a 100% identity and an E value of 2e-74 [24]. PCR 173 also matched to AB301946.1 in NCBI's GenBank, which corresponds to cape004, with a 99% identity and an E value of 4e-71 [24]. PCR 111 matched to AB301946.1 as well, with a 97% identity and an E value of 1e-67 [24]. The alignments are shown below in Figures 1, 2, and 3.

The nucleotide sequences which were aligned to reported DRB-like sequences for *Spheniscus demersus* with a percent identity of a hundred will be referred to as the allele with which they matched. Nucleotide sequences that scored percent identities lower than a hundred were identified according to their matches and either a letter in addition to the reported allele, for PCR 181 and PCR 173, or one number greater in the case of PCR 111. The following conclusions are summarized in Table 2. PCR 182 was confirmed to be cape001, PCR 172 was confirmed to be cape002, and PCR 171 was confirmed to be cape004. Unconfirmed sequences PCR 181, PCR 173, and PCR 111 were cataloged as alleles cape003b, cape004b, and cape005 respectively.

Table 2: Sample names along with their corresponding allele matches, expected values, percent identities, and their new allele ID.

Sample:	Matched Allele ID:	Expected Value:	Percent Identity:	New Allele ID:
PCR 182	cape001	2e-74	100%	cape001
PCR 172	cape002	2e-74	100%	cape002
PCR 181	cape003	6e-70	99%	cape003b
PCR 171	cape004	2e-74	100%	cape004
PCR 173	cape004	4e-71	99%	cape004b
PCR 111	cape004	1e-67	97%	cape005

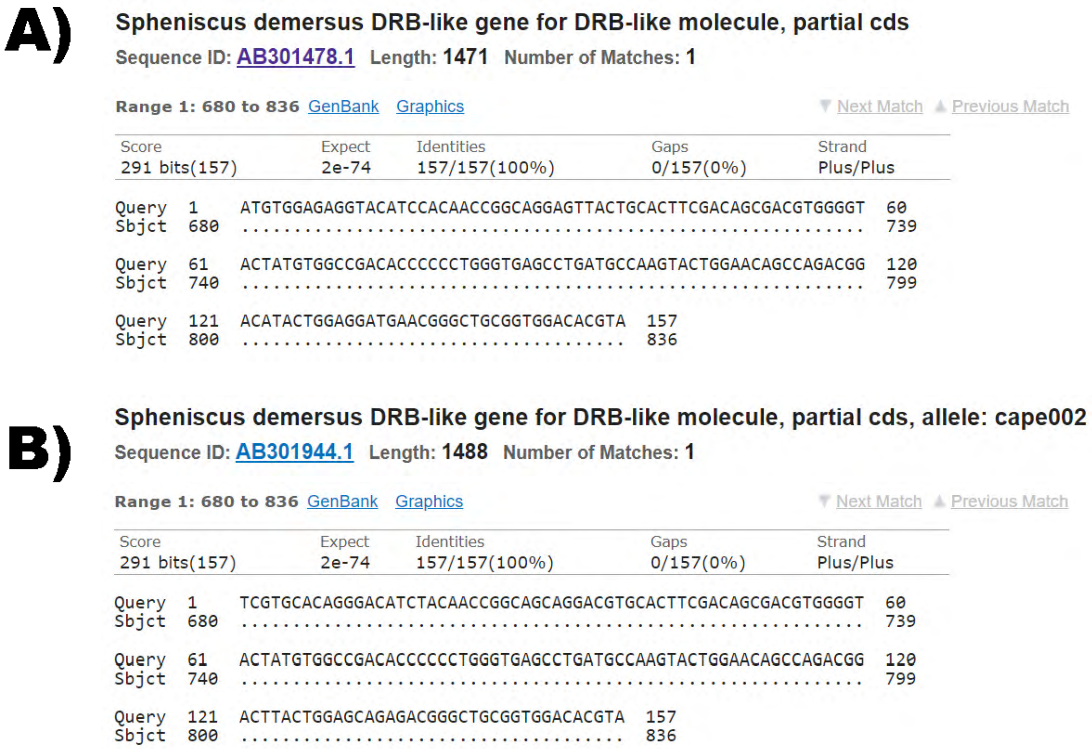

Figure 1: A) PCR 182 matched to AB301478.1 which was reported as cape001. B) PCR 172 matched to AB301944.1 which was reported as cape002.

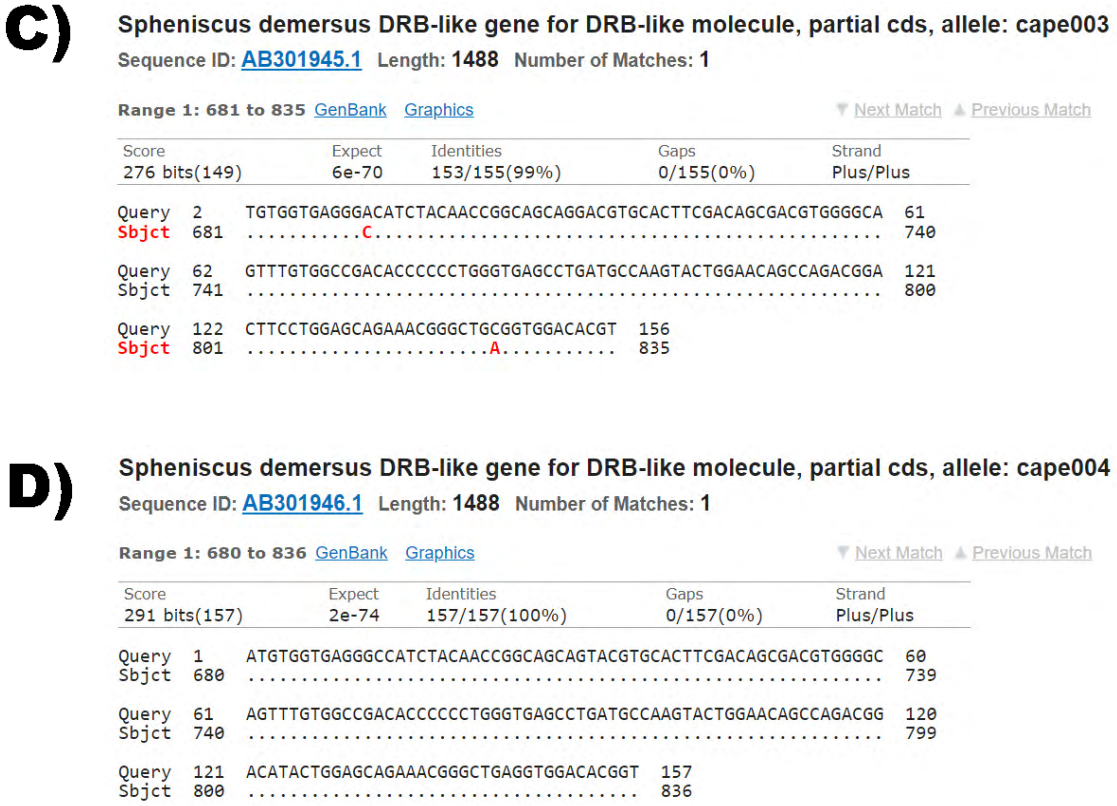

Figure 2: C) PCR 181 matched to AB301945.1 which was reported as cape003. D) PCR 171 matched to AB301946.1 which was reported as cape004.

E)

Spheniscus demersus DRB-like gene for DRB-like molecule, partial cds, allele: cape004

Sequence ID: AB301946.1 Length: 1488 Number of Matches: 1

Range 1: 680 to 836 GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
279 bits(151)	4e-71	155/157(99%)	0/157(0%)	Plus/Plus

```
Query  1    ATGTGGTGAGGGCCATCTACAACCGGCAGCAGTACGTGCACTTCGACAGCGACGTGGGGC  60
Sbjct  680  ............................................................  739

Query  61   AGTTTGTGGCCGACACCCCCCTGGGTGAGCCTTCTGCCAAGTACTGGAACAGCCAGACGG  120
Sbjct  740  ..............................GA..........................  799

Query  121  ACATACTGGAGCAGAAACGGGCTGAGGTGGACACGGT  157
Sbjct  800  .....................................  836
```

F)

Spheniscus demersus DRB-like gene for DRB-like molecule, partial cds, allele: cape004

Sequence ID: AB301946.1 Length: 1488 Number of Matches: 1

Range 1: 680 to 836 GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
268 bits(145)	1e-67	153/157(97%)	0/157(0%)	Plus/Plus

```
Query  1    ATGTGGTGAGGTACATCTACAACCGGCAGCAGTACGTGCACTTCGACAGCGACGTGGGGC  60
Sbjct  680  ..........GC..............................................  739

Query  61   AGTTTGTGGCCGACACCCCCCTGGGTGAGCCTTCTGCCAAGTACTGGAACAGCCAGACGG  120
Sbjct  740  ..............................GA..........................  799

Query  121  ACATACTGGAGCAGAAACGGGCTGAGGTGGACACGGT  157
Sbjct  800  .....................................  836
```

Figure 3: E) PCR 173 also matched to AB301946.1 which was reported as cape004. F) PCR 111 matched to AB301946.1 which was reported as cape004.

Amino Acid Translation and Alignment

The nucleotide sequences from PCR 182, PCR 172, PCR 181, PCR 171, PCR 173, and PCR 111, were translated into amino acid sequences in 5'to 3' Frame 3 using ExPASy [6]. The translated amino acid sequences are shown below in Table 3.

Table 3: Length and amino acid sequence for the six sequences pertaining to the MHC class II DRB-like gene. New allele identifiers were assigned to the sequences with less than a 100% identity. PCR 181 is cape003b, PCR 173 is cape004b, and PCR 111 is cape005.

Sample:	Allele ID:	Length:	Amino Acid Sequence (5' to 3' Frame 3):
PCR 182	cape001	51 residues	VERYIHNRQELLHFDSDVGYYVADTPLGEPDAKYWNSQTDILEDE RAAVDT
PCR 172	cape002	51 residues	VHRDIYNRQQDVHFDSDVGYYVADTPLGEPDAKYWNSQTDLLEQ RRAAVDT
PCR 181	cape003b	51 residues	VVRDIYNRQQDVHFDSDVGQFVADTPLGEPDAKYWNSQTDFLE QKRAAVDT
PCR 171	cape004	51 residues	VVRAIYNRQQYVHFDSDVGQFVADTPLGEPDAKYWNSQTDILEQ KRAEVDT
PCR 173	cape004b	51 residues	VVRAIYNRQQYVHFDSDVGQFVADTPLGEPSAKYWNSQTDILEQ KRAEVDT
PCR 111	cape005	51 residues	VVRYIYNRQQYVHFDSDVGQFVADTPLGEPSAKYWNSQTDILEQ KRAEVDT

NCBI BLAST

BLAST was used to align the amino acid sequences for PCR 172, PCR 181, PCR 171, PCR 173, and PCR 111 to the amino acid sequence of PCR 182 . The resulting alignments are shown below in Figure 4.

Figure 4: Amino acid sequence for PCR 182 aligned to the sequences of PCR 172, PCR 181, PCR 171, PCR 173, and PCR 111.

Superimposed IMGT Colliers de Perles

The sequence amino acid sequences for PCR 182, PCR 172, PCR 181, PCR 171, PCR 173, and PCR 111 were aligned to the amino acid sequence for the beta chain of 1j8h. Visualizations of the six amino acid sequences were made based on Lefranc's 1j8h IMGT Colliers de Perles and the six penguin amino acid sequences [17]. They are shown below in Figures 5, 6, 7, 8, 9 and 10.

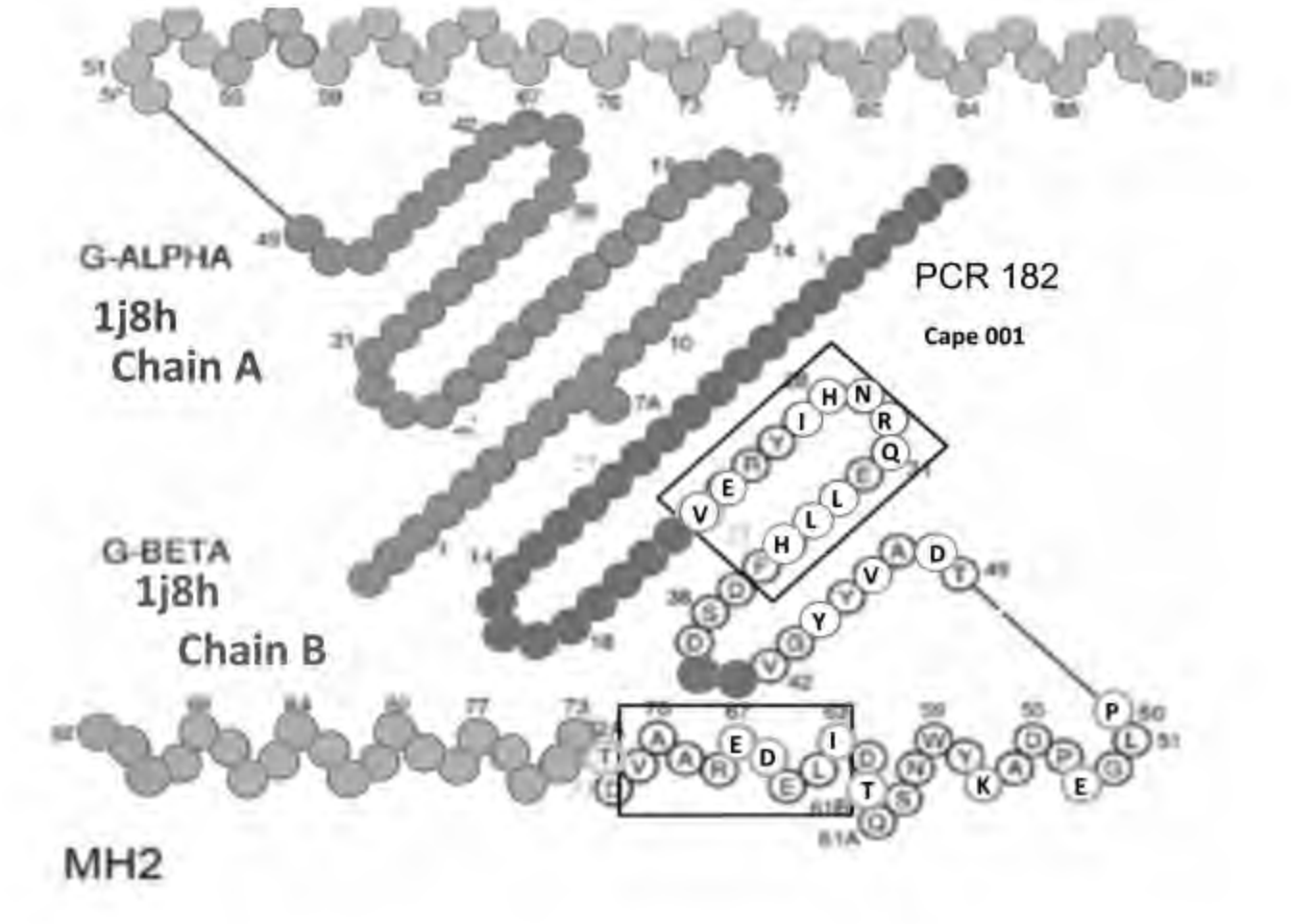

Figure 5: The amino acid sequence for PCR 182 was superimposed onto the 1j8h Beta Chain, in order to approximate the coverage of the DRB-like region of the allele.

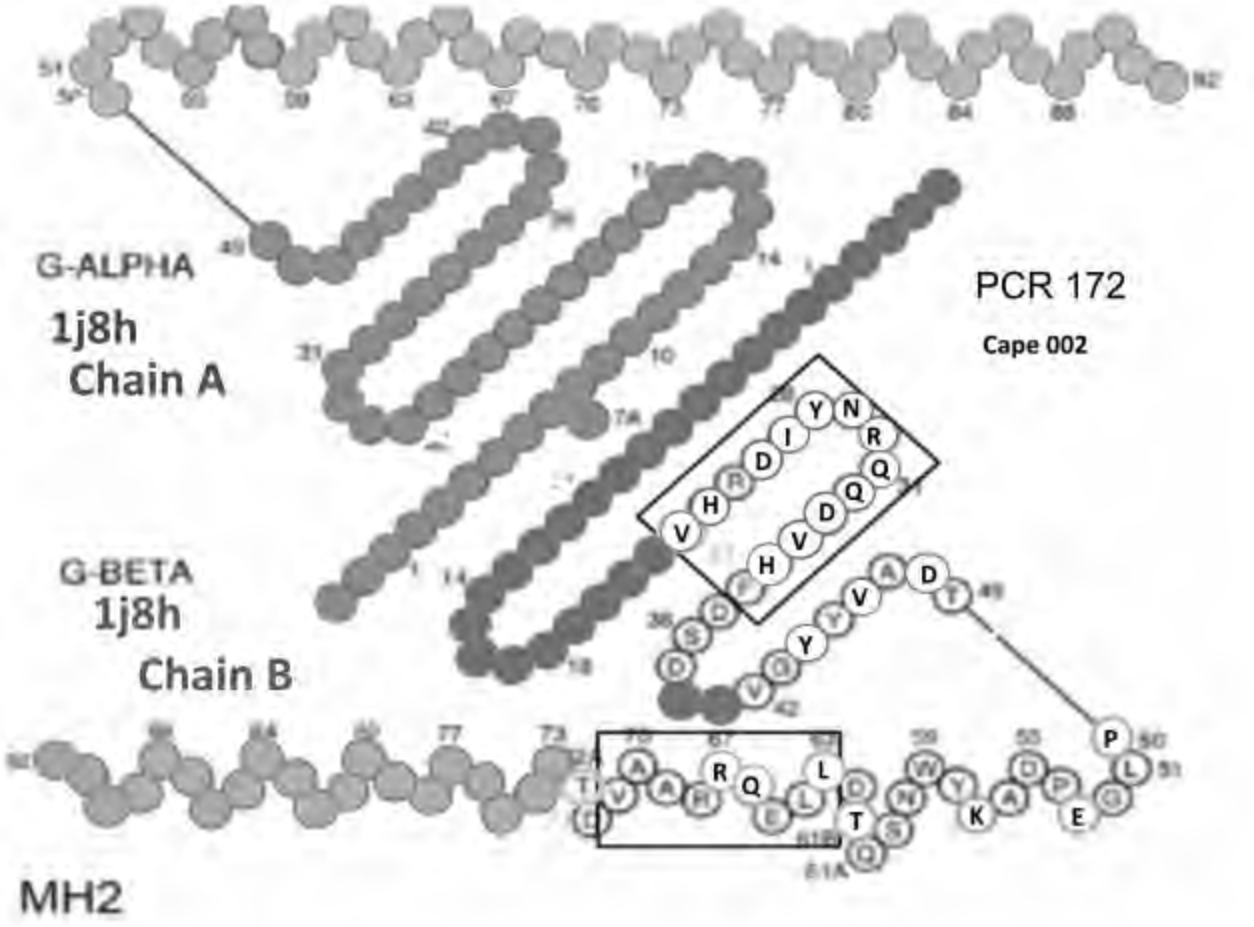

Figure 6: The amino acid sequence for PCR 172 was superimposed onto the 1j8h Beta Chain, in order to approximate the coverage of the DRB-like region of the allele.

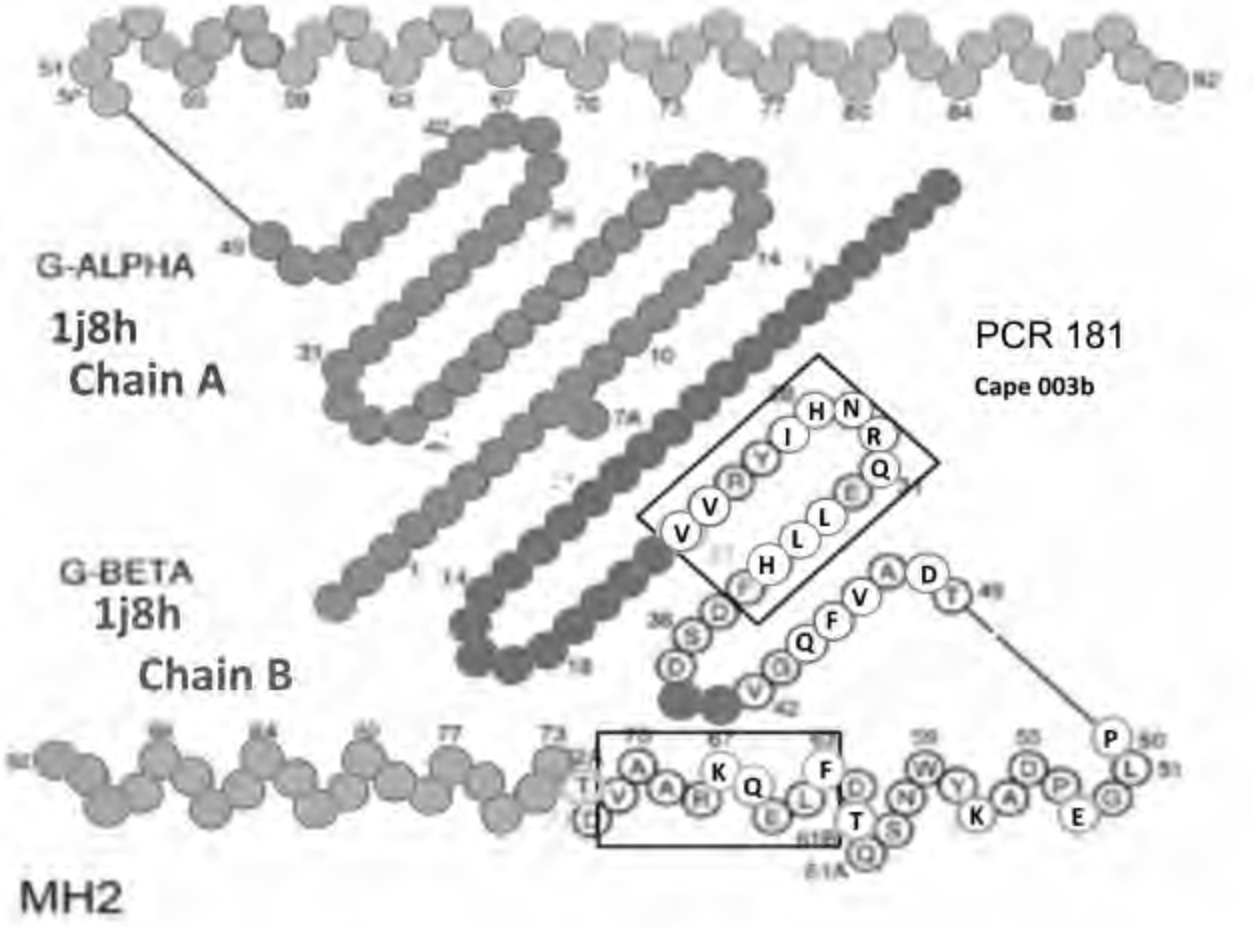

Figure 7: The amino acid sequence for PCR 181 was superimposed onto the 1j8h Beta Chain, in order to approximate the coverage of the DRB-like region of the allele.

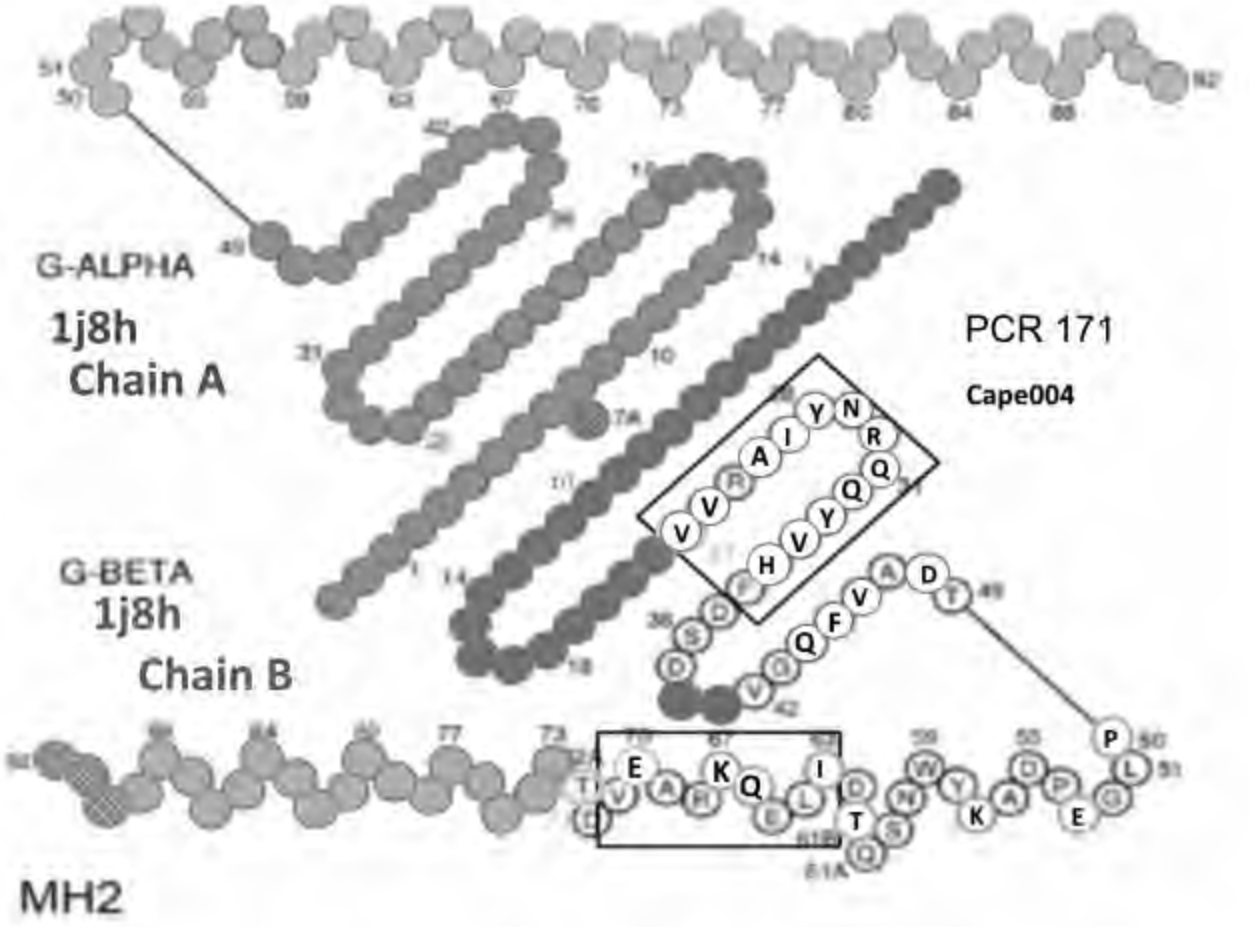

Figure 8: The amino acid sequence for PCR 171 was superimposed onto the 1j8h Beta Chain, in order to approximate the coverage of the DRB-like region of the allele.

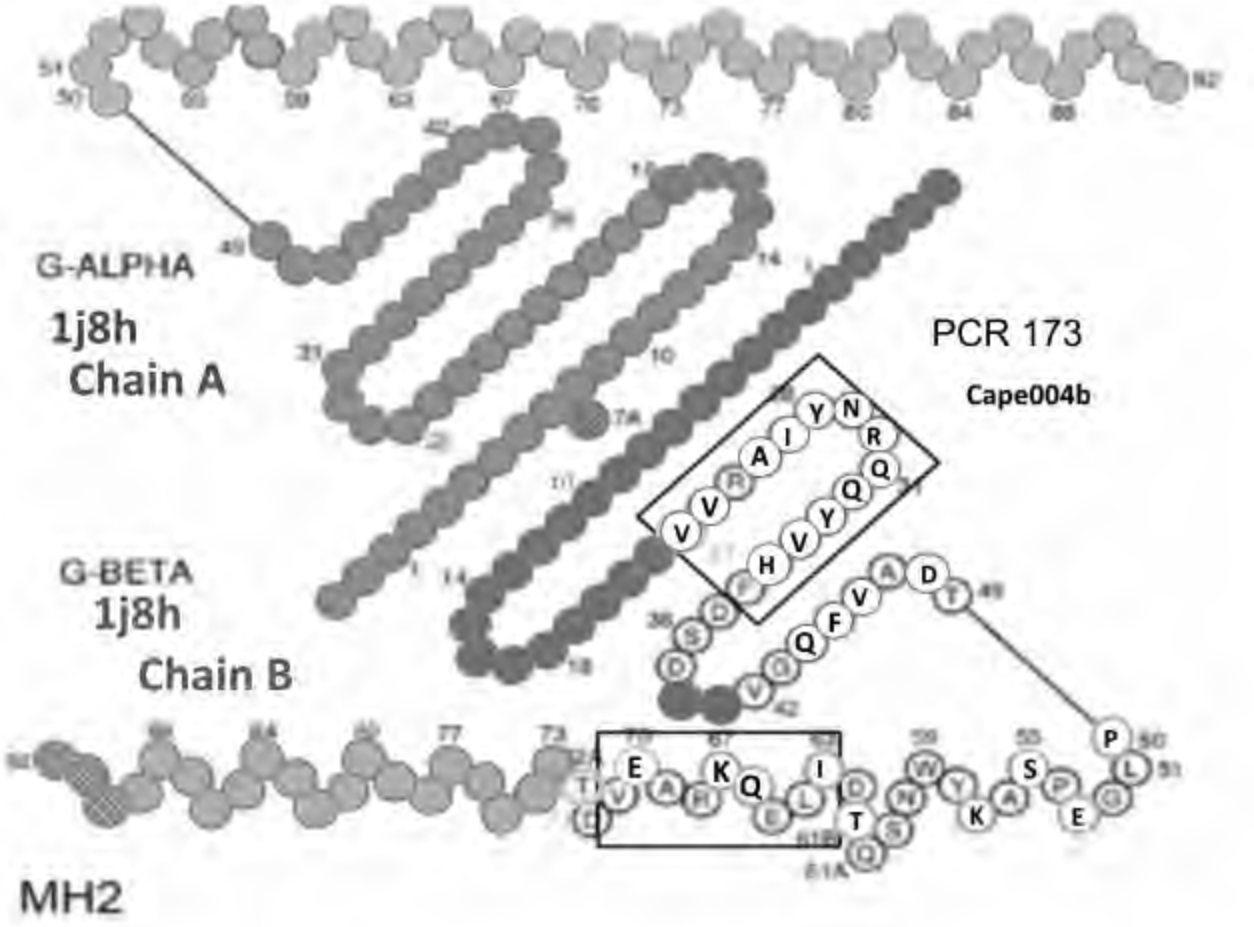

Figure 9: The amino acid sequence for PCR 173 was superimposed onto the 1j8h Beta Chain, in order to approximate the coverage of the DRB-like region of the allele.

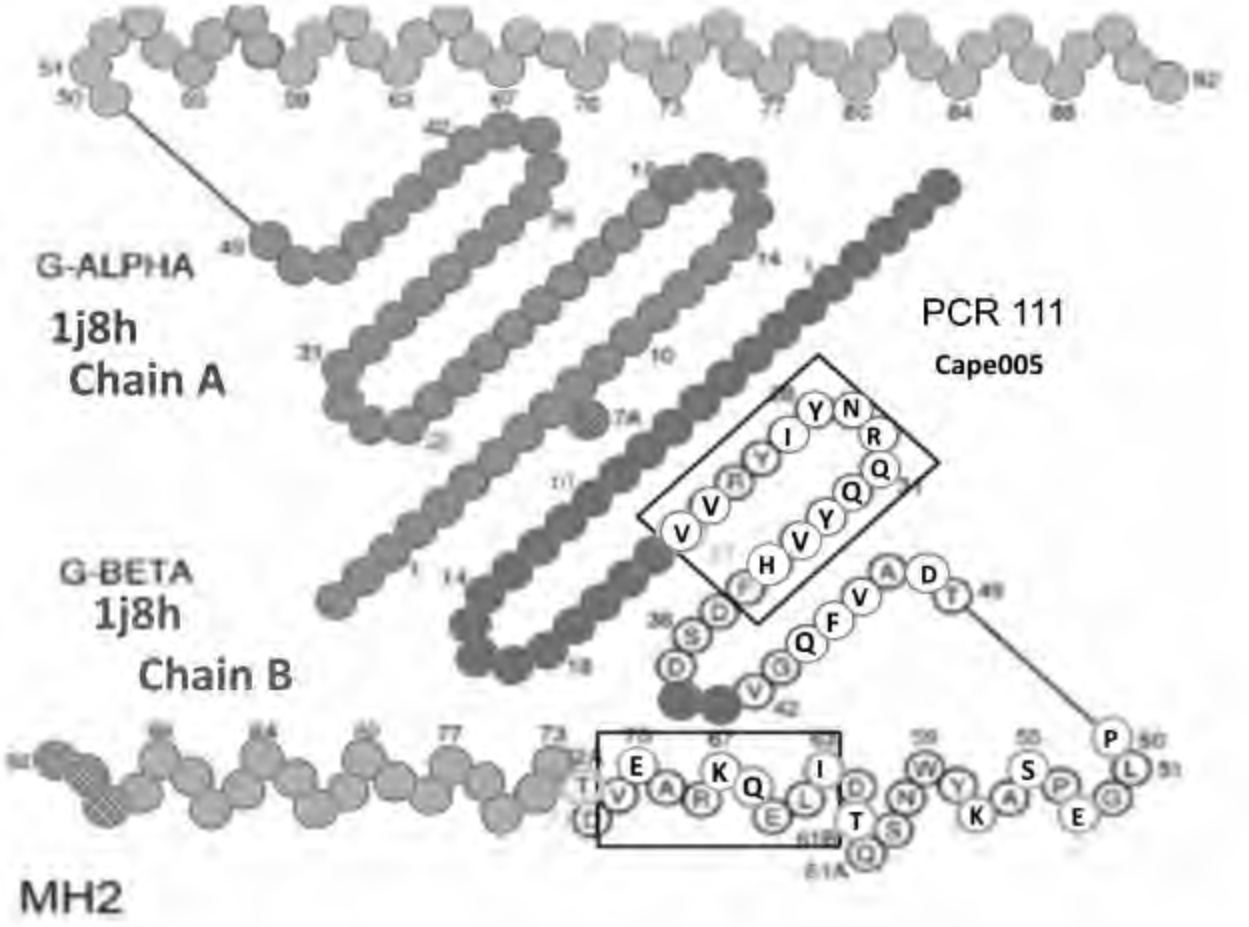

Figure 10: The amino acid sequence for PCR 111 was superimposed onto the 1j8h Beta Chain, in order to approximate the coverage of the DRB-like region of the allele.

Gel Electrophoresis

Shown below in Figure 11 is the gel containing the products from the second round of PCR: WB11, BB12, PB13, CL14, and 015 in lanes two through six. Four products from the first PCR reaction, WB1, BB2, PB3, and CL4 were loaded into lanes seven through ten of the gel. Table 4, found below, contains the lane number and the contents of each well including the primers.

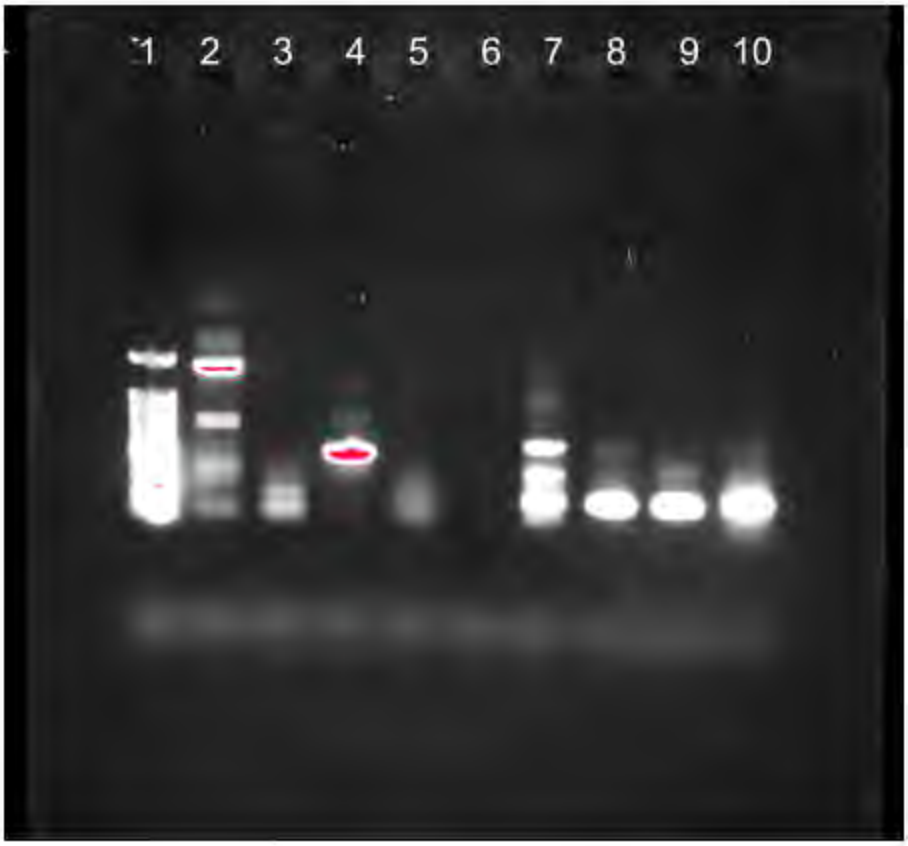

Figure 11: Lane 1 contains the 100 base pair ladder from Promega. Lanes 2 through 6 contain five samples from the second PCR reaction that used primers Lpen.hum1F and Lpen.hum2R. Lanes 7 through 10 contained the first four samples from the first PCR reaction which used pen1 and pen4 as the primers.

Table 4: Lane numbers for the gel and their respective samples. The primers corresponding to each sample are also included.

	Lane 1	**Lane 2**	**Lane 3**	**Lane 4**	**Lane 5**	**Lane 6**	**Lane 7**	**Lane 8**	**Lane 9**	**Lane 10**
Sample:	100 bp	WB11	BB12	PB13	CL14	dH_2O	WB1	BB2	PB3	CL4
Primers:	None	Lpen.hum1F/2R	Lpen.hum1F/2R	Lpen.hum1F/2R	Lpen.hum1F/2R	Lpen.hum1F/2R	pen1/4	pen1/4	pen1/4	pen1/4

DNA Electropherogram

Sample PB13 sequence results using the Lpen.hum1F and Lpen.hum2R primers were reported to be 452 base pairs long. The electropherogram appears to have various places along the sequence where double peaks are visible. The electropherogram for the Lpen.hum1F sense strand is shown below in Figure 12.

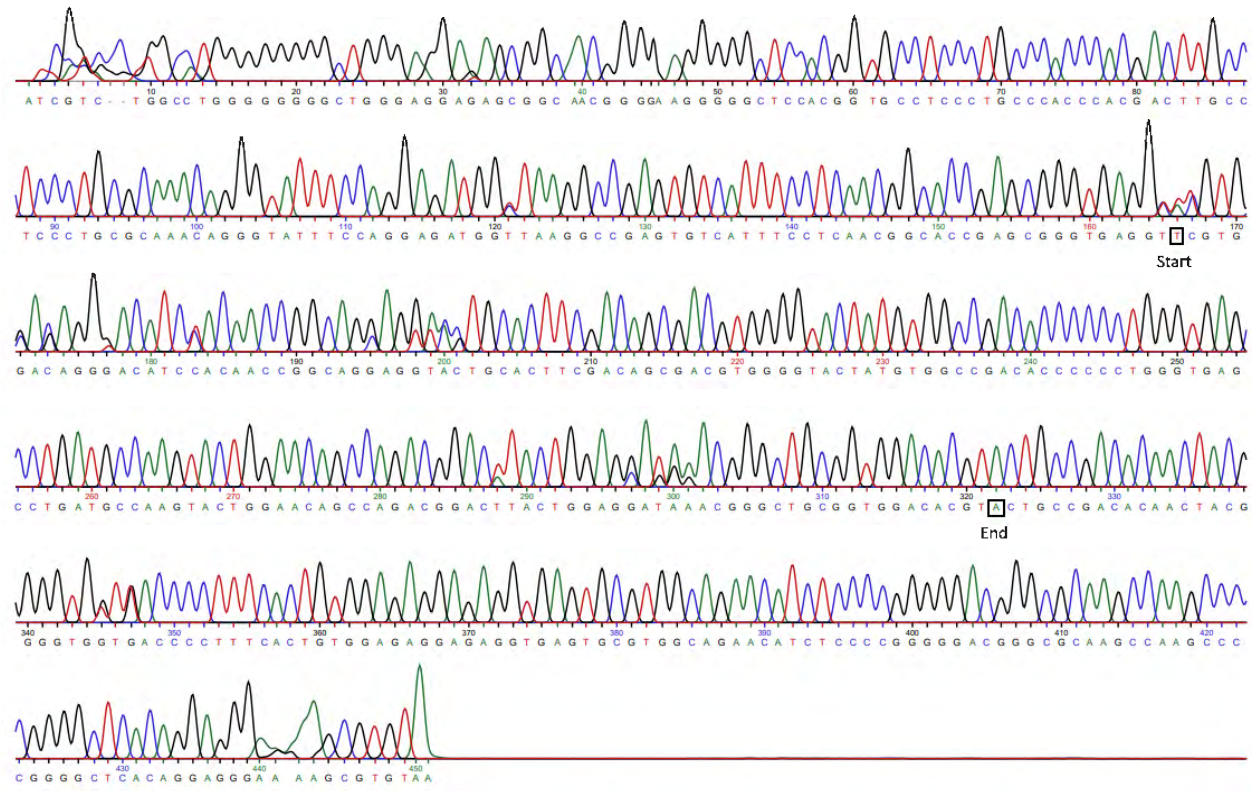

Figure 12: Electropherogram for sample PB using Lpen.hum1F primer, sequencing done by Eurofins. The DRB-like region, spanning 157 base pairs, is marked with the black Start and End boxes. The nucleotides are color coded. Adenine is shown in green, thymine in red, guanine in black, and cytosine in blue.

Discussion

The BLAST search using the nucleotide sequences for PCR 182, PCR 172, PCR 181, PCR 171, PCR 173, and PCR 111 resulted in the confirmation of several alleles. PCR 182, PCR 172, and PCR 171 were confirmed to be alleles cape001, cape002, and cape004 respectively, as shown in Figures 1 and 2. The three other samples had various single nucleotide polymorphisms (SNPs) within their matches, which were depicted in Figures 2 and 3. In order to better represent these polymorphisms, PCR 181, PCR 173, and PCR 111 were assigned allele identifiers cape003b, cape004b, and cape005 respectively. On top of the SNPs, the amino acid translation for the three latter samples each varied by at least one residue. Figure 4 shows the alignments between the amino acid sequence for PCR 182 and the remaining five sequences. Each amino acid sequence was unique. Therefore, it is probable that the SNPs between the unconfirmed or unreported alleles lead to nonsynonymous mutations in the corresponding amino acid sequences. In a small sample size of six, three possibly new alleles were discovered. Whether or not the single residues of difference between the sequences contributes a significant change in immune fitness will have to be elucidated in another study.

As visualized in Figures 5 through 10, the DRB-like region coded by the 157 base pair sequence only covers about one third of the beta chain of the MHC class II molecule. This is a rudimentary analysis, since despite the MHC class II molecules having highly conserved regions between species, the MHC class II molecule shown belongs to *Homo sapiens* not *Spheniscus demersus*. Structures for a penguin MHC class II protein could not be located in the Protein Data Bank [9]. Regardless, the DRB-like region spans an inner portion of the beta chain. In order to elucidate preceding and following amino acid residues on the beta chain not included in the DRB-like region, primers Lpen.hum1F and Lpen.hum2R reported by Kikkawa et al. were used

to amplify the entire beta chain of the MHC class II molecules (Kik 2005). The gel electrophoresis shown in Figure 11 confirms the difference in size between the PCR products for PB3 and PB13. PB3 was amplified using pen1 and pen4, and has a reported size of 198 bp [14]. PB13 was amplified using Lpen.hum1F and Lpen.hum2R whose product size is reported to be 476 bp [14].

According to the electropherogram shown in Figure 12, the sequence for PB13 contains polymorphisms within the DRB-like region and outside of it. These are demonstrated by the double peaks at some locations of the nucleotide sequence, signifying that PB13 contains two alleles, and is in other words heterozygous. Further testing by using denaturing gradient gel electrophoresis will facilitate the separation of these alleles and their consequential identification [16]. Once the alleles sequence results are received a consensus alignment from the results of the forward primer and reverse primer reads will be processed in order to yield a consensus sense strand sequence [18]. Expansion into the rest of the beta chains for each penguin sample could reveal a greater part of the alleles and the beta chain of MHC class II proteins for *Spheniscus demersus*.

At this point in time, the research was only able to analyze seven penguin blood samples in total. Dozens of samples remain from prior research, and even more semi purified blood samples from in situ penguins are ready to be processed and analyzed. With some samples being over a decade old, tests will have to be run in order to assess the quality of the available samples. This can be done by repeating the sequencing protocol for samples PCR 182, PCR 172, PCR 181, PCR 171, PCR 173, PCR 111 and comparing it to the sequences reported in this thesis. It would also be in the interest of this research to obtain recent samples from previous affiliates.

This way the study could possibly reflect over a decade of selective pressure on the penguin populations of South Africa and aquariums.

References

1. Abbas, Abul K, and Andrew H. Lichtman. Basic Immunology: Functions and Disorders of the Immune System. Philadelphia, PA: Elsevier Saunders, 2016.
2. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410.
3. Bollmer, Jennifer L., Vargas, F. Hernán, Farker, Patricia G. “Low MHC Variation in the Endangered Galápagos Penguin (Spheniscus Mendiculus).” *Immunogenetics*. vol. 59. no. 7. July 2007. pp. 593–602. Print
4. Botes, Annelise, Thiart, Hanlie, Parsons, Nola J., & Bellstedt, Dirk U.. (2017). Conservation implications of avian malaria exposure for African penguins during rehabilitation. South African Journal of Science, 113(7-8), 1-8. https://dx.doi.org/10.17159/sajs.2017/20160244
5. Castagno, Caitlin. "Major Histocompatibility Complex Variability and Infectious Disease in the Mystic Aquarium ex situ Population of African Penguins (Spheniscus demersus)" (2014). Distinction Thesis Projects.
6. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A.ExPASy: the proteomics server for in-depth protein knowledge and analysisNucleic Acids Res. 31:3784-3788(2003).
7. Gaire, “Avian Aspergillosis” Index of Bird Diseases Avian Web n.d. Web March 2022 www.beautyofbirds.com/aspergillosis.html
8. Hennecke J, Wiley DC. Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. J Exp Med. 2002 Mar 4;195(5):571-81. doi: 10.1084/jem.20011194. PMID: 11877480; PMCID: PMC2193773
9. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank Nucleic Acids Research, 28: 235-242.
10. Huth, Kelly, "MAJOR HISTOCOMPATIBILITY COMPLEX GENETIC DIVERSITY IN THE MYSTIC AQUARIUM EX SITU POPULATION OF AFRICAN PENGUINS (SPHENISCUS DEMERSUS)" (2013). Honors Thesis Projects.

11. IUCN. 2018. The IUCN Red List of Threatened Species. Version 2018-2. Available at: www.iucnredlist.org.
12. Janeway, Charles A., Travers, Paul, Walport, Mark. “Immunobiology: The Immune System in Health and Disease”. New York: Garland Publishing. 1999. pp.116-118. Print
13. Kast, Courtney, "Major Histocompatibility Complex Diversity in the Mystic Aquarium African Penguin (Spheniscus demersus)Population" (2013). Distinction Thesis Projects.
14. Kikkawa, E.F., Tsuda, T.T., Naruse, T.K. et al. “Analysis of the sequence variations in the Mhc DRB1-like gene of the endangered Humboldt penguin (Spheniscus humboldti).” Immunogenetics 57, 99–107 (2005). https://doi.org/10.1007/s00251-005-0774-6
15. Kikkawa EF, Tsuda TT, Sumiyama D, Naruse TK, Fukuda Michio, Kurita M, Wilson RP, LeMaho Y, Miller GD, Tsuda M, et al. “Trans-species polymorphisms of the mhc class II DRB-like gene in banded penguins (genus spheniscus).” *Immunogenetics* Vol. 61. May 2009. pp. 341-352. Print
16. Knapp, L.A., Lehmann. E., Hennes, L., et al. “High-resolution HLA-DRB typing using denaturing gradient gel electrophoresis and direct sequencing.” *Tissue Antigens* vol. 50. 1997. pp. 170-177. Print
17. Lefranc MP. Immunoglobulin and T Cell Receptor Genes: IMGT® and the Birth and Rise of Immunoinformatics. Frontiers in Immunology. 2014;5(5-6):505-28. doi:10.3389
18. *MacVector 9.5 User Guide*. MacVector, Inc, 2007.
19. Manlik, O, Krützen, M, Kopps, AM, et al. “Is MHC diversity a better marker for conservation than neutral genetic diversity? A case study of two contrasting dolphin populations.” Ecol Evol. 2019; 9: 6986– 6998. https://doi.org/10.1002/ece3.5265
20. *National Center for Biotechnology Information*, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/.
21. Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – . Accession AB301478.1, Spheniscus demersus DRB-like gene for DRB-like molecule, partial cds, DNA; [cited 2022 March 17]. Available from: https://www.ncbi.nlm.nih.gov/nucleotide/AB301478.1
22. Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – . Accession AB301944.1, Spheniscus

demersusDRB-like gene for DRB-like molecule, partial cds, allele: cape002, DNA; [cited 2022 March 17]. Available from: https://www.ncbi.nlm.nih.gov/nucleotide/AB301944.1

23. Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – . Accession AB301945.1, Spheniscus demersusDRB-like gene for DRB-like molecule, partial cds, allele: cape003, DNA; [cited 2022 March 17]. Available from: https://www.ncbi.nlm.nih.gov/nucleotide/AB301945.1
24. Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – . Accession AB301946.1, Spheniscus demersusDRB-like gene for DRB-like molecule, partial cds, allele: cape004, DNA; [cited 2022 March 17]. Available from: https://www.ncbi.nlm.nih.gov/nucleotide/AB301946.1
25. Schalk, Athena, "Diversity of the Major Histocompatibility Complex in African Penguins (*Spheniscus demersus*) *in situ*" (2020). Honors Thesis Projects.
26. Sommer, Simone. “The importance of immune gene variability (MHC) in evolutionary ecology and conservation.” Frontiers in zoology vol. 2 16. 20 Oct. 2005, doi:10.1186/1742-9994-2-16
27. Strachan, Tom, Read, Andrew. *Human Molecular Genetics* CRC Press, 2020, pp. 391
28. Ujvari, Beata, and Katherine Belov. “Major Histocompatibility Complex (MHC) markers in conservation biology.” International journal of molecular sciences vol. 12,8 (2011): 5168-86. doi:10.3390/ijms12085168C102861978