Investigation of HLA-DQB in *Megaptera novaeangliae*

Female mating preferences and probabilistic paternity analysis

Master Thesis

Laetitia G. Schmid at Stockholm University

Under supervision of Per J. Palsbøll and Martine Bérubé

2009/10

Previous page: A humpback whale with her calf. Free wallpaper downloaded from animal.discovery.com the 1st of December 2009.

PREFACE

The following 78 pages tell the detailed story of my Master's thesis in the evolutionary genetics group of marine mammals at Stockholm University under the supervision of Per Palsbøll and Martine Bérubé. Fascinated by the major histocomopatibility complex (MHC), I succeeded in convincing Per to supervise a study about female mate choice based on a presumed to be highly polymorphic locus in the MHC of humpback whales, *Megaptera novaeangliae*.

The Master's thesis functions mainly as the completion of an academic study. It calls for the application of the most substantial acquired knowledge throughout the study to conduct an own, independent work under the supervision of experienced insiders. In my case I learned a lot by being forced to eventually produce after years of consuming. New qualities such as independence, self-confidence, stamina and frustration tolerance became suddenly important. Well, my scientific hunger is still constantly growing, and I am more excited than ever before. The following pages represent the main production during my time in Per's and Martine's group although I was from time to time obsessed by some migration with isolation considerations. I start with a long introduction leading the freshwater reader into the area of immunology, explaining the function of the MHC and the role of female mate choice in this context; as well as describing the saltwater study species. An extensive section listing all experimental procedures and analyses of the produced data follows. Everything is listed including attempts that failed. The same goes for the section listing all the results. In the discussion I tried to make the connections and explain my findings in humpback whales in the context of the general theory about female mate choice and its statements of grounds. Failed attempts are also discussed. After a short conclusion I included a section, called "supplementary information", containing all the substantial data and results necessary to follow the main thesis, such as sampling IDs, genotypes, population genetic test statistics and additional results concerning female mate choice. The penultimate section, termed "appendix", shows the script written in R, photographs of the PCR amplifications attempting to design locus-specific primers and the output of the software PHASE used to statistically infer alleles. This section is not necessary to understand the main thesis. As usual, the thesis closes with the references.

I would like to thank Per and Martine for supervising me during the last year. Martine especially for helping me develop hands on doing precise lab work and care for the environment in the lab, and Per for answering all my questions at any time of the day and guiding me towards studying the MHC in humpback whales. I felt very comfortable in the evolutionary genetics group. This was definitely because of Jean-Luc who was always there for me whenever needed and who taught me to have more confidence, as well as Morten who made me happy by taunting whenever possible. The group seemed to be perfect after the inclusion of Mimmy, the best one of all. I also want to thank Donny for his support and patience. We succeeded. And it is getting better and better. At this point I want to mention Mathias, the second most important man in my life for everything. He makes the difference. Thank you Benjamin for the support with R and being my most successful friend. Then I want to thank my family, especially my mother for giving me the strength needed, the Ökvist family for supporting me here in Stockholm and all my friends here in Stockholm, especially Estelle, Dries and Ruben. Black and Brown. And Lina. And last but not least my friends at home, especially Chrige and Donny's parents in Houston. Thank you iChat and Messenger. And thank you Spotify! And thank you Ziggy.

Stockholm, 5th of May 2010, Laetitia G. Schmid

TABLE OF CONTENTS

INTRODUCTION

Typically, selectively neutral markers have been used for conservation genetic studies of wild populations [1]. It is possible to answer questions about demographic parameters and population structure, based solely on presumed to be neutral genetic data with tiny little pieces of tissue from a representative sample of the study population [2]. Study goals include tracing back lineages, revealing pedigrees, defining evolutionary significant units and/or management units. I am very interested in applying coalescence theory to the study of population structures and histories based on neutral markers, i.e. the neutral genetic structure of humpback whales in the Northern Atlantic Ocean. However, evolutionary relevant and adaptive processes within and among populations can only be investigated with coding genes, such as the major histocompatibility complex. Here the study goals include the identification of the genetic architecture of local adaptive traits and adaptive population divergence. Eventually the knowledge of local adaptation will turn out to be crucial for defining management units. Currently, endeavors such as the 1000 genomes project [3], powered by next-generation sequencing techniques, allow us to tackle new challenges and allow for the extension to new amounts of data to solve old problems. For example, using whole genome association studies, it is now possible to identify genes linked with almost any genetic trait or to scan the entire genome for regions under selection. In this study, I investigated female mate choice based on the major histocompatibility complex in a population of *Megaptera novaeangliae*.

Major Histocompatibility Complex

The major histocompatibility gene complex (MHC) plays an important role in the immune system, autoimmunity and reproductive success of all vertebrates except for the jawless vertebrates [4]. The MHC consists of a group of closely linked genes, which encode cell-surface glycoproteins that bind antigens derived from pathogens or parasites and present them to T-lymphocytes, which trigger the appropriate immune response.

There are two classes that can be distinguished [5, 6]. Class I genes code for glycoproteins that bind intracellular peptides mainly derived from viral proteins or cancer infected cells. Class II genes monitor the extracellular environment and present peptides derived from parasites (e.g. bacteria, nematodes or cestodes).

MHC class II molecules are made up of two polypeptide chains. Each polypeptide chain consists of an intracellular domain, a transmembrane domain, and two extracellular domains (alpha and beta) [4, 7, 8]. MHC class II molecules bind only peptides with amino acids at key positions that fit to their own binding pocket which is built up of amino acids with specific properties (polarity, acidity, electric charge) in the extracellular domain [4, 8, 9]. The investigated exon in this study codes for one of the extracellular domains: the class II HLA-DQ beta (HLA-DQB) domain. Each extracellular-domain molecule can bind to a limited range of antigens (figure 1). The maintenance and the renewal of variation in the antigen binding sites is an important genetic component in the cascade leading to an appropriate immune response since this part of the MHC acts to prevent attacks by viruses, bacteria and other parasites [4]. The more different alleles there are, the greater are the chances to detect and react against various pathogens [10, 11]. On the other hand there is an upper limit to the number of different MHC molecules [12]. All T-lymphocytes that can recognize self-peptides bound to a new MHC variant must be removed in order to maintain self-tolerance. With an increasing number of MHC molecules more T-lymphocytes must be produced by the thymus. The latter has an upper limit above which it cannot produce more diverse T-lymphocytes. A

resulting loss of the T-lymphocytes repertoire is disadvantageous for the immune system because a hypothetical individual with all the MHC molecules required to present the necessary peptides would not have any T-lymphocytes left to respond to them [13-16]. So no individual can respond to all pathogen-peptides in its environment.

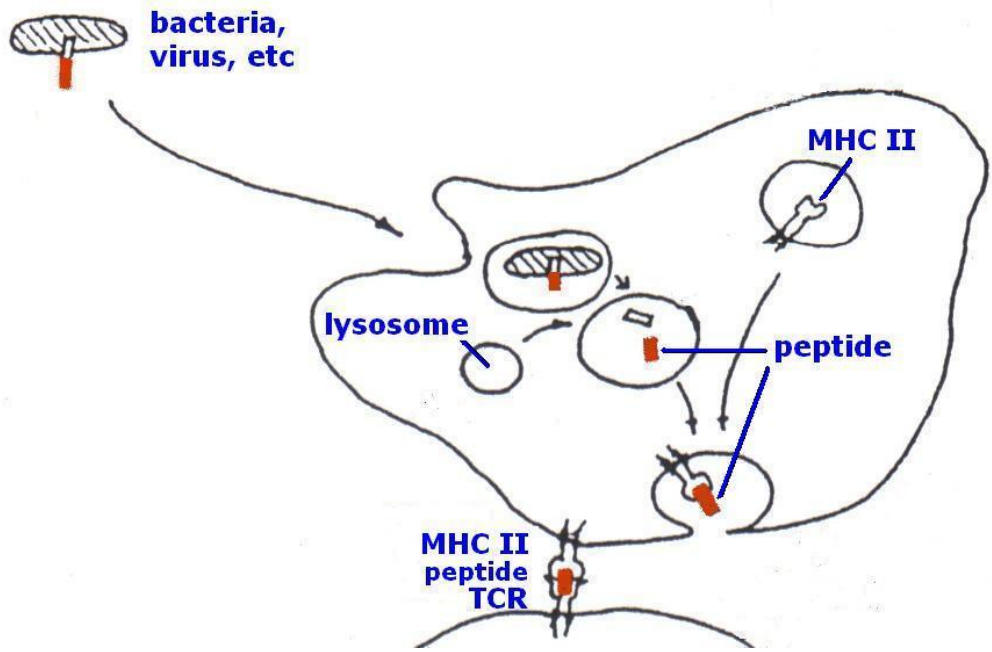


Figure 1: Schematic procession of an antigen by an antigen-presenting cell. The extracellular antigen (red) is taken up by the cell and split up into peptides in the lysosome. One of the peptides builds a MHC class II complex together with a MHC dimer produced in the endoplasmatic reticulum. This complex will be anchored on the cell wall and present the antigen to a cytotoxic t-cell (TCR), which triggers the appropriate immune response.

MHC sequence variants have been shown to correlate with important biological traits, such as susceptibility to infectious and autoimmune diseases, kin recognition, cooperation, pregnancy outcome and mating preferences [1]. Most of our knowledge about this supergene complex has been derived from studies in humans or model species under experimental, laboratory conditions [4, 17-22]. The MHC is one of the most polymorphic gene complexes in the genome of vertebrates.

Three main selection mechanisms driving the high diversity of the MHC are summarized in Spurgin and Richardson 2010 [23]. The *heterozygote advantage* hypothesis states that individuals heterozygous at MHC loci are able to respond to a greater range or pathogen peptides than homozygote individuals [10, 11].

The *frequency dependent selection* proposes that rare alleles have an advantage. There is strong selection on pathogens to overcome the resistance of the most common host MHC alleles. The result is a cyclical, co-evolutionary arms race in which pathogens and MHC alleles fluctuate in frequency and therefore maintain diversity in a dynamic process ("Red Queen hypothesis" or "moving-target hypothesis") [24-27].

The *fluctuating selection* hypothesis assumes that spatial and temporal heterogeneity in type and abundance of pathogens can maintain diversity at the MHC [28-31]. Selection is in this case directional and not cyclical and the pathogen fluctuations are determined externally, rather than by co-evolution of host and pathogen.

It is important to keep in mind that the three main selection mechanisms are by no means mutually exclusive and may operate in concert. It is almost impossible to isolate and quantify single selective forces [23]; and we might end up with an interaction of signals of micro-recombination, parasite-mediated selection and mating preferences, the last one enhancing the effect of different selection mechanisms [32, 33].

MHC-disassortative mating represents an additional selective force resulting in indirect genetic benefits [34]. Mating choice is either for "good genes" or for genetic compatibility. In the former, individuals prefer mates that display increased vigor, such as better body condition or costly secondary sexual characteristics. These individuals may be more disease resistant. In the latter, individuals prefer mates with a complementary genotype to their own to maximize diversity in their offspring. Preferring MHC dissimilar mates produces more heterozygous offspring, which might not be more resistant per se but profit from a higher probability to carry the needed allele. Particular allele combinations in heterozygote human individuals were beneficial against HIV [35] and hepatitis [36]. Odors play an important role for vertebrates in mating preferences [37]. It has been shown that MHC metabolites are involved in odor production [38-41], and females are especially sensitive to these odors during their recipient phase [42-44]. Female mate choice depending on the MHC has been found in several different species: i.e. mice and rats [26, 45-52], lizards [53], fish [54-56] and birds [57-60]. Besides Wedekind's pioneer study in humans from Switzerland, it has been applied to other human populations: other European populations [61-63], Hutterites [64] and a South-Amerindian population [65]. There is obviously a publication bias in showing female mate choice for dissimilar alleles at the MHC, I could only find four studies which show no such patterns [63, 65-67].

Humpback Whales

In this study I investigated female mate choice based on the locus HLA-DQB exon 2 in humpback whales (figure 2). Humpback whales (*Megaptera novaeangliae*, Borowski, 1781) inhabit all major oceans except for the Arctic Ocean [68]. They belong to the baleen whales and are one of the bigger rorqual species. Humpback whales are considered "vulnerable" to extinction by the World Conservation Union and are an Appendix I (endangered) species under the Convention on International Trade in Endangered Species (CITES) [69]. Their distribution is characterized by summer occupancy of mid- to high-latitude feeding areas and winter occupancy of low-latitude breeding areas where mating and calving takes place. Fidelity to a specific feeding area has shown to be very strong, and calves born in low latitudes are guided back by their mother to her feeding range during the period of maternal care [70-73].

In the North Atlantic Ocean there are four main summer feeding aggregations: the Gulf of Maine, eastern Canada, western Greenland and Iceland & Norway [74, 75]. The North Atlantic humpback whale population was estimated to number around 11'000 individuals in 1992-1993 (95% CI 10,290 – 13,390) [76]. They were commercially exploited from the 17th until the 20th century, and there is no reliable estimate of the population size prior to exploitation [69]. During winter these discrete feeding populations move to shared low-latitude breeding grounds along the Atlantic margins of the Antilles, from Cuba to northern Venezuela. The largest modern breeding aggregations occur at the Greater Antilles in the West Indies, where photo-identification research has confirmed the presence of all primary feeding areas in the North Atlantic ocean. Photo identification also revealed humpback whale annual movements of over 8000 km distances in the North Atlantic ocean [77]. This species

migrates farther than any other mammal. Despite their wide-ranging migrations the individual humpback whales from the different feeding aggregations exhibit very high rates of annual return to the same feeding aggregation [72, 74, 77]. The mean sighting dates for whales in the West Indies that had fed in the Gulf of Maine were significantly earlier than those for animals that had fed in Greenland or Iceland & Norway. Males were generally seen earlier on the breeding ground than were females. These results imply that, although there is an annual mixing of whales from the different feeding areas in the breeding ground, there are reduced mating opportunities due to the staggered timing of migration.

Figure 2: A humpback whale and her calf [78].

Humpback Whales in the Gulf of Maine

The population in this study is summer-feeding in the Gulf of Maine (GoM, figure 3) on the northeastern coast of North America from Cape Cod up to southwest Nova Scotia [69]. In American waters the species is protected under the U. S. Endangered Species Act and Marine Mammal Protection Acts, but farther up north in Canadian waters the North Atlantic populations are not considered to be of any management concern. The size of the GoM population is estimated to lie in the high hundreds [79]. Individuals are observed mainly between April and October. One of the main factors influencing the humpback whales' distribution is the distribution of prey in the feeding area. Humpback whales are generalists, feeding on different schooling fish species and euphausiids. The GoM humpback whales feed mostly on sand lance, *Ammodytes* species, Atlantic herring, *Clupea harengus* and euphausiid species. The average age at first birth for females is 8.78 years (s = 2.33) in the Gulf of Maine feeding population [69]. This is also the time when survival in females peaks. Breeders exhibit lower survival than non-breeders, and the cost for breeders persists into adulthood. The average mature GoM female gives birth to a single calf every two to three years [68, 72]. The calves are born in the breeding area between January and March and are dependent on their mother until autumn of their first year. The Gulf of Maine exhibits high rates of boat traffic. More than half of all individuals of the GoM population have experienced non-lethal entanglements in fishing gear and an average of 9.7 dead individuals are recovered each year [69].

Individual humpback whales can be identified from their natural markings such as ventral pigmentation of the flukes, the shape and size of the dorsal fin [80] or with genetic tagging [75]. Whale watching vessels in the Gulf of Maine operate on a near-daily basis in the southwest GoM from mid-April through October, 1979 until the present. During the Years of the North Atlantic Humpback Whale (YoNAH) project, the composition of GoM whales in the West Indies (WI) breeding ground (figure 3) has been estimated using mark-recapture studies to investigate population size and population structure across the North Atlantic Ocean from January through March 1992-1993 [74-77]. WI data was primarily obtained at Silver Bank, a breeding aggregation located approximately 70 miles north of the Dominican Republic.

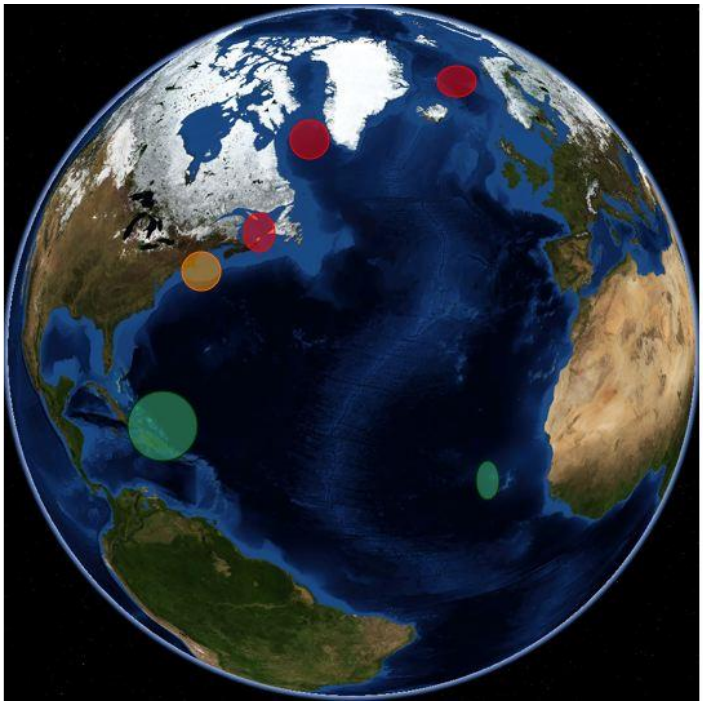


Figure 3: Global view on the North Atlantic Ocean. The four main humpback whale summer feeding aggregations are shown as shaded orange and red circles, orange: Gulf of Maine (study population), three red from west to east: eastern Canada, western Greenland, and Iceland & Norway. The breeding areas are shaded in green: bigger area in the west: West Indies, smaller area in the east: Cape Verde [77].

Thesis Overview

In this thesis I investigated female mate choice for dissimilar alleles relative to her own using an adaptive gene complex trying to reveal information about selective forces acting on humpback whales feeding in the Gulf of Maine. The gene under investigation (major histocompatibility complex, MHC) is a supergene complex with important functions in the immune system, present in all mammals. The MHC class II genes code for peptide-binding proteins on antigen-presenting cells. MHC-sequence variants have been shown to correlate with important biological traits, such as susceptibility to infectious and autoimmune diseases, kin-recognition, cooperation, pregnancy outcome and mating preferences. The majority of studies investigating female mate choice depending on the MHC have shown preferences for males with dissimilar alleles at the HLA-DQB locus (complementarity). The common explanation for these findings is that the more different alleles there are, the greater are the chances to detect and react against various pathogens. Female mate choice depending on the MHC has been investigated in a wide array of terrestrial vertebrates and fish [18, 34, 53, 55, 57, 81-85], but not yet in cetaceans. I expected to find a replication of the common paradigm of female mate choice for dissimilar alleles in humpback whales. Two reasons make humpback whales an excellent system to study female mate choice for complementarity in marine mammals: 1) Males, overrepresented relative to females in the breeding ground, have no means to force females to mate. It is the females that allow a particular male to mate with her after having spent some time with different males. 2) Females are sampled together with their calves in the feeding ground, after having successfully traveled back from the breeding ground. Only the healthy and fitter females make their way back with a calf. This provides the opportunity to assign calves to the right mother and infer the paternal contribution in the calf. The locus HLA-DQB exon 2, investigated in this study, could not be separated from another co-amplified locus. I sequenced directly the simultaneous amplification of two very similar genes of the MHC in 27 mother-calf pairs sampled during the feeding season in the Gulf of Maine and a random sample of 43 males sampled during the 1992 breeding season in the West Indies. The genetic similarity of the mother and her calf was quantified and compared to the expected distribution of genetic similarity assuming random mating between the sampled mothers and the random sample of males. There was a trend for females producing calves genetically similar to their own HLA-DQB genotype, either due to mate choice or post zygotic selection at the HLA-DQB locus.

MATERIAL & METHODS

Genetic Material

All skin biopsies were collected from free-ranging humpback whales (n = 97, 43 males and 27 mother-calf pairs, sample IDs in table S3, S4 and S5 supplementary information) using standard biopsy techniques [86, 87] and stored in a saturated salt solution with 25% dimethyl sulfoxide [88] at -20 degrees Celsius (°C). The epidermal samples of 43 males for the probabilistic paternity simulation were collected as a part of the Years of the North Atlantic Humpback Whale project (YoNAH) during 1992 and 1993 in the West Indies breeding area [77]. The epidermal samples of the mother-calf pairs were obtained from 54 whales between 1992 and 2005 during the feeding season in the Gulf of Maine.

DNA was extracted using a Qiagen DNeasy tissue kit (Qiagen, Inc.) following the manufacturer's protocol. Sex determinations were conducted as described by Bérubé and Palsbøll [89, 90]. Blank extractions and blank PCR reactions were included as controls to account for possible contaminations.

HLA-DQB exon 2 in *Megaptera novaeangliae* in the North Atlantic Ocean

A 172 base pair (bp) long fragment (excluding primer sequences) was polymerase chain reaction (PCR) amplified using the universal primer pair DQB1 (reverse; [91, 92]) and DQB2 (forward; [91, 92]). The optimal thermocycling profile consisted of an initial denaturation step at 94 °C for 2 min, followed by 33 cycles of 94 °C denaturation for 15 s, 54 °C annealing for 15 s and 72 °C extension for 15 s, with a final 10 min extension soak at 72 °C to allow for complete extension of the PCR product. The reactions started with 0.75 ul (10 uM) forward primer, 0.75 ul (10 uM) reverse primer, 1.5 ul PCR buffer (including 2.5 mM $MgCl_2$), 6 ul dGATC-mix (0.5 uM per nucleotide), 0.08 ul Taq™ DNA polymerase, 4.92 ul ddH_2O and 1 ul (10 ng/ul) genomic DNA for a final reaction volume of 15 ul per individual. Blank PCR reactions were included as controls to account for possible contaminations.

The PCR products, stained with ethidium bromide, were checked on a 2% agarose gel (SIGMA-ALDRICH) in 10 mM TBE (Tris-borate/EDTA), pH of 7-8.5, run under 175 Volt and then purified by a Shrimp Alkaline Phosphatase and Exonuclease I (SAP-Exo) protocol. SAP-Exo cleaning per 5 ul PCR amplification: 2 ul SAP (Promega), 0.5 ul Exo (New England Biolabs), 0.2 ul Buffer and 1.3 ul ddH_2O, incubation at 37 °C for 40 min.

Purified PCR products were sequenced directly in both directions on an ABI 3130 genetic analyzer (Applied Biosystems) using BigDye terminator chemistry (PerkinElmer). The BigDye® Terminator v3.1 sequencing standard kit (Applied Biosystems) was used following the manufacturer's instructions using a POP-7™ polymer and the standard run module. The cycle sequencing reactions consisted of 0.5 ul 1/4 Terminator Ready Reaction Mix, Big Dye version 3.1, 5 ul PCR template, 0.17 ul primer (at 10 mM), 2.83 ul ddH_2O and 1.5 ul of the 5X ABI sequence buffer per reaction. The thermocycling protocol was 25 cycles of denaturation at 96 °C for 10 s, annealing at 50 °C for 5 s and extension at 60 °C for 240 sec.

The purification protocol of the cycle sequencing reaction contained the following substances per individual: 1.5 ul (3 M) sodium acetate, 31.25 ul 95% ethanol and 7.25 ul ddH_2O in the first step and 250 ul 70% ethanol in the second step.

The sequences that had been amplified with the universal DQB primer pair showed highest similarity with HLA-DQB exon 2 in several cetacean species, cow and human after a nucleotide sequence search (BLAST) on the National Center for Biotechnology Information (NCBI) homepage. It will be subsequently called HLA-DQB exon 2. Several triple peaks suggested a duplication of the locus of interest. The simplest way to separate simultaneously amplified loci is to design more specific primers in the region of both ends of the sequence or in the flanking intron sequences.

Allele-specific Primers I (before cloning) to separate two loci

Heimeier *et al*. had amplified HLA-DQB exons 2 and 3 successfully in the Hector's dolphin, *Cephalorhynchus hectori*, in 2009 [93]. In a first step I amplified a sequence in *M. novaeangliae* with the primers designed for HLA-DQB exon 3 by Heimeier *et al.* (DQBex3F and DQBex3R2, table S1 supplementary information) under the same conditions applied to exon 2.

As a second step I tried to amplify the flanking introns of the sequence HLA-DQB exon 2. For intron 2 I used Heimeier's reverse primer situated in the end of HLA-DQB exon 3 (DBQex3R2, figure S1 supplementary information, table S1 supplementary information) and I also designed several other reverse primers situated in my monomorphic exon 3 sequences of *M. novaeangliae* (DQBex3R1, DQBEXON3R3, DQB_Intron2R, DQB_Intron2_R2 and DQB_Intron2_R3, figure S1 supplementary information, table S1 supplementary information). As forward primers I designed two primers in a short conserved region within HLA-DQB exon 2 of *M. novaeangliae* (DQB_Intron2F and DQB_Intron2_F2, figure S1 supplementary information, table S1 supplementary information) and one primer in the beginning of the HLA-DQB intron 2 sequence of *C. hectori* published by Heimeier *et al*. 2009 (DBQ_Intron2_F3, figure S1 supplementary information, table S1 supplementary information). I also tried to amplify intron 1 with a highly conserved primer used in the study by Heimeier *et al.* 2009 to amplify HLA-DQB intron 1 in *C. hectori* as forward primer and DQB1 as the corresponding reverse primer. This universal forward primer, situated in the end of exon 1, was originally designed for cows and pigs [94].

The conditions for the PCR amplifications for intron 1 and 2 were the following: one cycle of 94 °C for 2 min in the beginning, followed by 32 cycles of a denaturing step at 94 °C for 60 sec, an annealing step of 54 °C for 60 sec and an extension step of 72 °C for 240 sec. In the end there was a final extension soak of 72 °C for 10 min (this program takes more than four hours). I also applied a gradient PCR program with the same conditions except for the annealing temperature which rose from 52 °C to 62 °C. This program was repeated with several individuals and all the different primer combinations, including DQB1 and DQB2. The reactions started with 0.5 ul (10 uM) forward primer, 0.5 ul (10 uM) reverse primer, 1 ul PCR buffer (including 2.5 mM $MgCl_2$), 4 ul dGATC-mix (0.5 µM per nucleotide), 0.08 ul Taq™ DNA polymerase, 2.92 ul ddH_2O and 1 ul (10 ng/ul) genomic DNA for a final reaction volume of 10 ul per individual. Blank PCR reactions were included as controls to account for possible contaminations. The PCR products, stained with ethidium bromide, were checked on a 2% agarose gel (SIGMA-ALDRICH) in 10 mM TBE (Tris-borate/EDTA), pH of 7-8.5, run under 175 Volt.

Cloning

Because I was not able yet to sequence any allele-specific flanking intron sequences of HLA-DQB exon 2 in the study species, I decided to clone some individuals for HLA-DQB exon 2. The resulting alleles should have served to design allele-specific primers to amplify allele-specific flanking intron sequences (amplified with these allele-specific primers within DQB exon 2 and a corresponding primer situated in exon 1, or a corresponding primer situated in exon 3, respectively). The allele-specific flanking intron sequences should themselves have served to design allele- and later locus-specific primers for the duplicated HLA-DQB exon 2 sequences (separation of the two simultaneously amplified loci).

It was also necessary to clone more individuals to assess the scoring of four alleles at two simultaneously amplified loci in the directly sequenced samples.

Individuals GM030005, WI920001 and WI920009 were PCR amplified three times before cloning and then either cloned separately several times (WI920009) or the PCR amplifications were mixed and cloned once (GM030005, WI920001). WI920015 was PCR amplified once and cloned once (table S2 supplementary information). The amplification conditions for HLA-DQB exon 2 for cloning differed slightly from the standard conditions. I adjusted the PCR conditions according to Becker and Lenz (2008) to minimize the formation of sequence chimeras during PCR amplification and mosaic sequences during cloning [95]. These adjustments include: (1) at least two independent amplification reactions per individual (except for individual WI920015), (2) a reduction of the PCR cycle number to 25 in the first round whereby an aliquot of this amplification was taken as DNA template added to fresh PCR educts for a second amplification of 10 steps and (3) a longer elongation time of 60 sec per cycle is taken.

I used a TOPO TA cloning kit (invitrogen). The sequences were cloned into a pCR®2.1-TOPO cloning vector and then transformed into One Shot ®TOP10 chemically competent cells. Briefly, the kit makes use of the fact that Taq™ DNA polymerase adds a single deoxyadenosine (A) to the 3’-end of the PCR product during standard PCR amplification [96, 97]. TOPO TA cloning ligates the PCR product into the linearized pCR®2.1-TOPO cloning vector with the help of the enzyme topoisomerase I, which is covalently bound to the vector. The vector itself has single, overhanging 3’-deoxythymidine (T) residues. The method relies on the ability of adenine (A) and thymine (T) on different DNA fragments to hybridize and, in the presence of a ligase, become ligated together. When the free 5'-ends of the PCR product strands attack the topoisomerase/3'-end of each vector strand, the strands are covalently linked by the topoisomerase.

Positive clones were selected by PCR amplification screening using the provided primers M13 forward (5’-GTA AAA CGA CGA CCA G-3’) and M13 reverse (5’-CAG GAA ACA GCT ATG AC-3’) located within the vector. The protocol was the same as used before for directly sequencing HLA-DQB exon 2 with the annealing temperature lowered to 50 °C. Between 42 and 60 positive clones per individual were sequenced in both directions on an ABI 3130 sequencer using BigDye terminator chemistry, as described above.

Allele-specific Primers II (after cloning) to separate the two loci

With the sequences of the five resulting alleles of individuals GM030005 and WI920009, I designed seven more allele-specific primers (AL1_403_F, AL2_403_F, AL1_716A_F, AL1_716B_F, DQBHM01, DQBLC01 and DQBLC02, figure S1 supplementary information, table S1 supplementary information). The amplification conditions for the allele-specific PCR reactions after cloning were the same as above: 94 °C denaturing temperature for 60 sec, 54 °C annealing temperature for 60 sec, 72 °C extension temperature for 240 sec for 32 cycles. The PCR amplification was initiated by 94 °C for 2 min and finished with an extension step of 72 °C for 10 min. Depending on the primer pairs, the annealing temperature and the number of cycles could vary (table A1 appendix). I also applied a gradient PCR program with the same conditions except for the annealing temperature which rose from 52 °C to 62 °C. The reactions started with 0.5 ul (10 uM) forward primer, 0.5 ul (10 uM) reverse primer, 1 ul PCR buffer (including 2.5 mM $MgCl_2$), 4 ul dGATC-mix (0.5 µM per nucleotide), 0.08 ul Taq™ DNA polymerase, 2.92 ul ddH_2O and 1 ul (10 ng/ul) genomic DNA for a final reaction volume of 10 ul per individual.

I tested all possible primer combinations to amplify HLA-DQB- intron 1 or intron 2 at different annealing temperatures and number of cycles. The PCR products, stained with ethidium bromide, were checked on a 2% agarose gel (SIGMA-ALDRICH) in 10 mM TBE (Tris-borate/EDTA), pH of 7-8.5, run under 175 Volt. Blank PCR reactions were included as controls to account for possible contaminations.

If primer combinations at a specific temperature resulted in a few discrete bands, I excised the band of the expected length from the gel. For the gel excisions the buffer of the gel electrophoresis was changed after each reaction to avoid cross contamination. For the extractions I used the QIAEX II Gel Extraction Kit (Qiagen, Inc.) and followed the manufacturer's instructions. Briefly, the kit purifies DNA from any agarose gel in TBE buffer (Tris-borate/EDTA, 10 mM, pH of 7-8.5). Silica particles absorb DNA molecules in the presence of a high salt concentration. All non-nucleic acid impurities, such as agarose, proteins, salts and ethidium bromide are removed during washing steps with differently concentrated alcohol dilutions. DNA is finally re-suspended in TBE (Tris-borate/EDTA, 10 mM, pH of 7-8.5) or sterilized water.

Excised and purified amplifications were sequenced directly in both directions on an ABI 3130 sequencer using BigDye terminator chemistry, as described above.

Direct Sequencing

The 172 bp long fragment (excluding primer sequences) from the locus HLA-DQB exon 2 was PCR amplified, in a total of 97 *M. novaeangliae* individuals from the North Atlantic Ocean, using the universal primer pair DQB1 (reverse; [91, 92]) and DQB2 (forward; [91, 92]). The thermocycling profile was the same that amplified successfully from beginning on: an initial denaturation step at 94 °C for 2 min, followed by 33 cycles of 94 °C denaturation for 15 s, 54 °C annealing for 15 s and 72 °C extension for 15 s, with a final 10 min extension at 72 °C to allow for complete extension of the PCR product. The reactions started with 0.75 ul (10 uM) forward primer, 0.75 ul (10 uM) reverse primer, 1.5 ul PCR buffer (including 2.5 mM $MgCl_2$), 6 ul dGATC-mix (0.5 uM per nucleotide), 0.08 ul Taq™ DNA polymerase, 4.92 ul ddH_2O and 1 ul (10 ng/ul) genomic DNA for a final reaction

volume of 15 ul per individual. Blank PCR reactions were included as controls to account for possible contaminations.

Every individual was amplified and sequenced at least four times. At least three replicates are needed because during the PCR amplification one locus may stochastically be amplified in a higher frequency relative to the other. The PCR products, stained with ethidium bromide, were checked on a 2% agarose gel (SIGMA-ALDRICH), pH of 7-8.5, in 10 mM TBE (Tris-borate/EDTA) run under 175 Volt and purified by a Shrimp Alkaline Phosphatase and Exonuclease I (SAP-Exo) protocol, as described above. Purified PCR products were sequenced directly after purification in both directions on an ABI 3130 sequencer using BigDye terminator chemistry.

Data Analysis

The software Geneious [98] was used to call the bases at each nucleotide position. I used default settings except for the resolution, which was constantly set to 200% and the chromatogram ranged from a minimum of 100% to a maximum of 140%. Three to five replicates of each individual were scored. At each polymorphic position four bases (A,C,G,T) had to be scored. As an additional control, another person scored 20 individuals. Consistently scored polymorphic positions were used for a single nucleotide polymorphism (SNP) analysis of 27 mother-calf pairs and 43 random males. When both loci are treated as a super-locus, it is possible to sequence both simultaneously resulting in four alleles (figure 4). The comparison of mother-calf pairs with four alleles per position (multi-SNPs) allows one to infer the similarity between mothers and their calves. When this is compared to the similarity of randomly generated calves and their mothers, the similarity between mothers and fathers can be indirectly described.

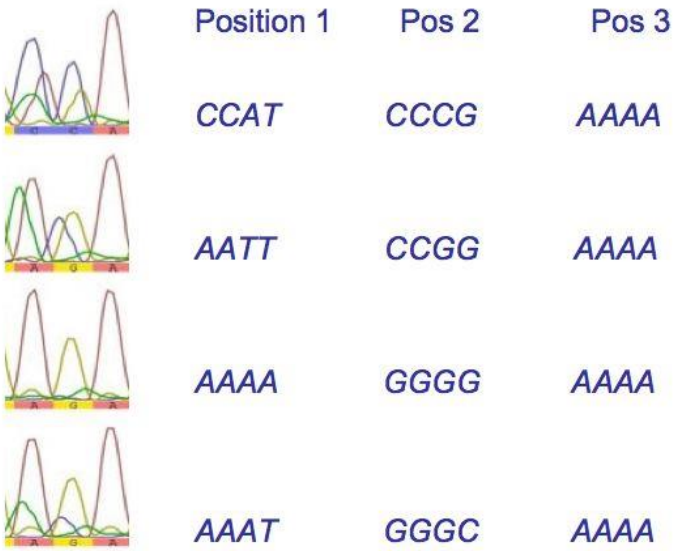


Figure 4: Translation of the electropherograms from the ABI sequencer to letters of the four bases (A = Adenosine, C = Cytosine, G = Guanine and T = Tyrosine) for three example positions in four different individuals.

A script was written in R, a programming language and software environment for statistical computing and graphics [99] (the script is given in the appendix). The function of the script was (a) to count the average of all shared bases between mothers and their calves at every polymorphic position and thereby giving a similarity score *SS* (equation [i]) for each mother-calf pair (mother-calf pairs and their sample IDs are given in table S5 supplementary information).

$$SS = \frac{1}{n}\sum_{i=1}^{n} |M_i \cap C_i| \qquad \text{[i]}$$

where *SS* = similarity score, *n* = number of polymorphic positions (multi-SNPs), $|M_i \cap C_i|$ = number of elements in the intersection of the mothers' bases and the calves' bases at position *i*.

In a second step (b) the script was used to simulate random female mate choice in the population, using a random sample of males and creating multiple iterations of calves. With these randomly generated calves the script was used to calculate the density distribution of the similarity scores between the simulated calves and their mothers (c), representing the null-hypothesis of random mate choice. Steps a, b and c were conducted in an initial analysis of 16 consistently scored multi-SNPs in a single locus (HLA-DQB exon 2). A second estimation (steps a, b and c) was done with seven non-synonymous and highly polymorphic positions out of the 16 (table S5 supplementary information) which are involved in antigen-binding according to the aligned homologous sequences from cows [100].

The probability of the observed similarity score for each mother-calf pair given random mating was assessed using a Mann-Whitney U test. The same test was used to test if the distributions of similarity scores for each position in the observed mother-calf pairs were different from the random distribution.

The average *SS* over all 27 females was calculated in a simulation where females were mated with the most dissimilar male in the sample relative to her own multi-SNP genotype to see if it was theoretically possible to produce a significantly low *SS* with the data available.

To investigate if females preferably mate with more related males than expected (i.e. if there is inbreeding), steps a, b and c were also analyzed with a different marker system: five microsatellite loci. This test is devised to assess mate choice at the genome-wide level. The observed similarity of mothers and their calves for the microsatellite loci (relatedness) was calculated as the fraction of shared alleles per locus between two individuals ($M_{xz'}$ [101]). The null hypotheses of no deviations from Hardy-Weinberg proportions and genotypic linkage disequilibria for these five microsatellite loci were tested using the software GENEPOP 4.0 [102, 103]. Specifically, deviations from Hardy-Weinberg proportions were assessed using the Markov chain approach of Guo and Thompson [104], and the one-sided test of heterozygote deficiency was assessed by an estimate of F_{is} [105]. The presence of genotypic linkage disequilibrium was tested using a Markov chain method as described by Raymond and Rousset (1995 [106]). At a first step all tests were performed with 10,000 dememorizations, 1000 batches and 5000 iterations. For p-values close to significance at the 95% level, when the standard error was large or the number of switches (the number of times the sample configuration changes in the MC run) was low, the number of batches was increased to 10,000. The global significance of p-values was assessed by applying the sequential Bonferroni correction to adjust for the effect of multiple testing [107].

Using the nine HLA-DQB exon 2 alleles obtained by cloning of four individual humpback whales in this study (figure 5) and 23 alleles from humpback whales sampled mostly in the Pacific Ocean [71] and published on GenBank, the actual four alleles in every directly sequenced individual were inferred using the software PHASE [108]. The default input parameters were used (number of iterations = 1000, thinning interval = 1 and burn-in = 100). For individuals harboring more than two alleles of a different kind, one or two alleles were subtracted by eye using the nine cloned alleles prior to the phasing analysis. PHASE can only handle input files with two alleles per sequence.

To assess the functionality of the sequences and if selection is acting on the investigated locus, two different tests for selection were performed using the software DnaSP [109]: Tajima's *D* [110] as well as Fu and Li's *D** and *F** [111]. These tests were applied to (i) an alignment of the nine alleles obtained by cloning, (ii) an alignment of the nine cloned alleles and all remaining alleles from GenBank and (iii & iv) an alignment of (i) and (ii) respectively, using only the positions involved in antigen-binding [100].

RESULTS

HLA-DQB exon 2 in *Megaptera novaeangliae* in the North Atlantic Ocean

The polymerase chain reaction (PCR) amplification and sequencing of a 172 bp long fragment with the universal primers DQB1 (reverse, table S1 supplementary information; [91, 92]) and DQB2 (forward, table S1 supplementary information; [91, 92]) turned out to be successful in a sub-sample of *M. novaeangliae* feeding annually in the Gulf of Maine. The resulting sequences showed several double and two triple peaks. Therefore I decided to design allele-specific primers to separate simultaneously amplified loci in this species. The sequence showed highest similarity with HLA-DQB exon 2 in several cetacean species after a nucleotide sequence search (BLAST) on the National Center for Biotechnology Information (NCBI) homepage.

Allele-specific Primers I (before cloning)

In a first step I amplified a sequence in *M. novaeangliae* with the primers designed for HLA-DQB exon 3 in *C. hectori* by Heimeier *et al.* [93] (DQBex3F and DQBex3R2, table S1 supplementary information) under the same conditions applied to exon 2. Exon 3 turned out to be monomorphic in the study species (figure S2 supplementary information). No stop codons were identified.

All primer combinations used resulted either in no amplification or in the amplification of several sequences of different length. Even the gradient amplification resulted in several sequences of different length up to a threshold 60 °C at which amplification failed altogether.

Cloning

Four individuals with different patterns when sequenced directly (1 x 3:1, 1 x 2:2, 1x 2:1:1 and 1x 1:1:1:1) were cloned. Between 42 und 60 clones per individual were sequenced afterwards, resulting in nine different alleles occurring in more than one copy (table 1, figure 5), four of them have been described before in humpback whales [112]. The frequency of alleles found by cloning confirmed the scoring results done by direct sequencing (table 1). The cloning of every individual resulted in some sequences with only one or two copies that were different from the others by one or two base pairs and must have been created by PCR- or cloning-artifacts. They were therefore termed “false singletons”. Individual WI920009 was PCR amplified and cloned three times and resulted in one false singleton, whereas individual WI920015 was PCR amplified once and cloned once and resulted in six false singletons. The other individuals lied in between (table S2 supplementary information). Singletons did not contain any stop codons and looked like variants of functional sequences. The clones of individual GM030005 and WI920009 served to design more allele-specific primers (AL1_403_F, AL2_403_F, AL1_716A_F, AL1_716B_F, DQBHM01, DQBLC01 and DQBLC02, figure S1 supplementary information, table S1 supplementary information).

Cloning worked most efficiently when I purified the PCR product with the QIAquick PCR purification kit (Qiagen, Inc.) following the manufacturer’s protocol and cloned them directly without diluting to the desired 3:1 insert to vector ratio. Except for a few white, false positive colonies, all colonies were blue. I assume that the insertion was in-frame and the

enzyme beta-galactosidase retained functional. PCR purification did not have a major effect on cloning efficiency; but without purifying, cloning resulted in a substantial amount of primer insertions that were difficult to detect by gel electrophoresis screening.

Table 1: Results from the cloning of four humpback whale individuals; one female sampled in the Gulf of Maine and three males sampled in the West Indies (Sample IDs starting with GM or WI respectively). The allele numbers correspond to the Sample ID and are simply given as numbering of different alleles found in an individual. The allele names are given according to the suggestion of Klein *et al*. (1990) [113]. Between 42 and 60 clones were sequenced in all individuals.

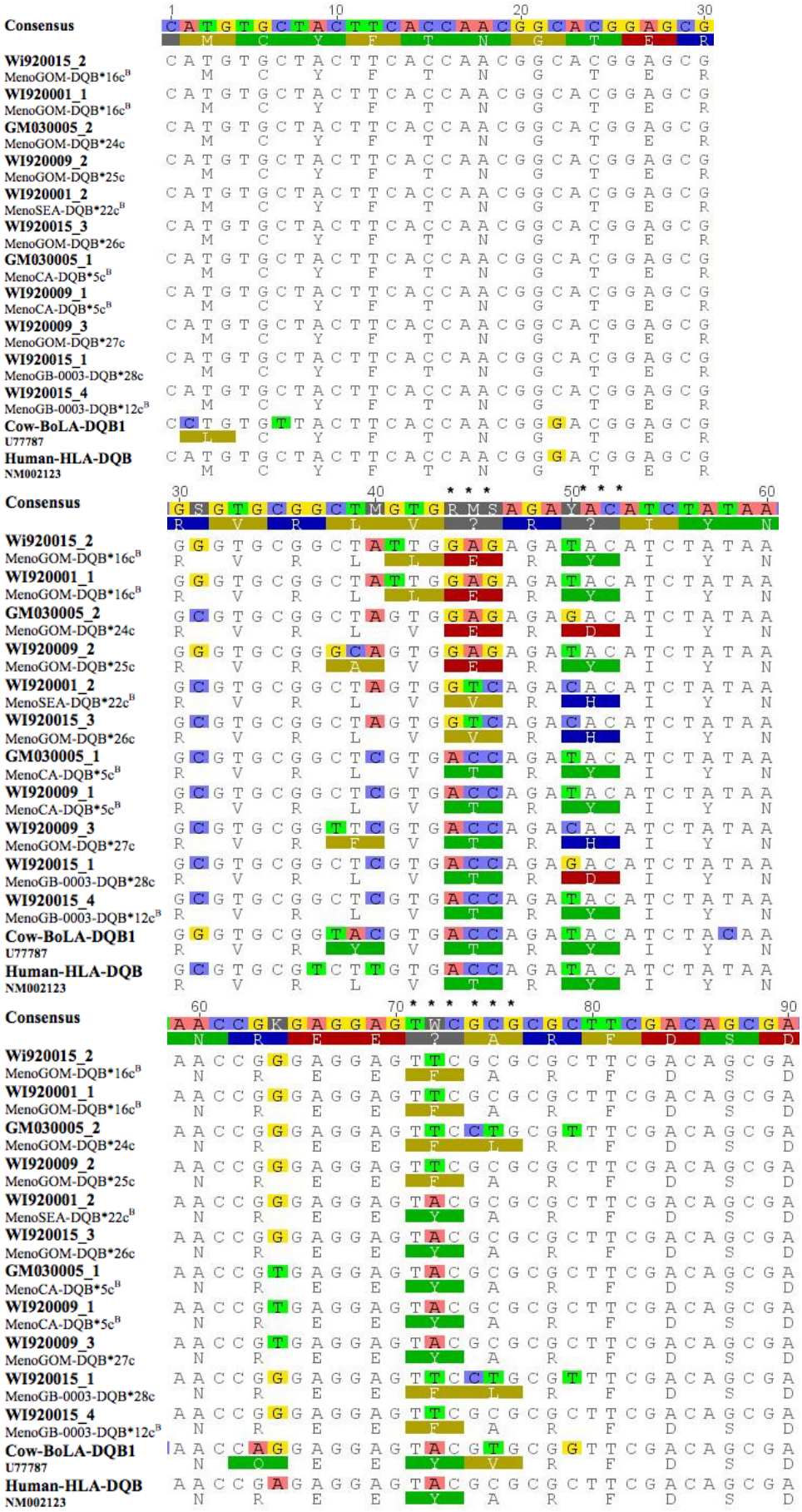
Sample ID	Allele number	Allele name (Klein 1990)	Number of clones
GM030005	GM030005_1	MenoCA-DQB*5c[B]	37/60
GM030005	GM030005_2	MenoGOM-DQB*24c	20/60
WI920001	WI920001_1	MenoGOM-DQB*16c[B]	22/42
WI920001	WI920001_2	MenoSEA-DQB*22c[B]	18/42
WI920009	WI920009_1	MenoCA-DQB*5c[B]	28/60
WI920009	WI920009_2	MenoGOM-DQB*25c	18/60
WI920009	WI920009_3	MenoGOM-DQB*27c	13/60
WI920015	WI920015_1	MenoGOM-DQB*28c	26/53
WI920015	WI920015_2	MenoGOM-DQB*16c[B]	11/53
WI920015	WI920015_3	MenoGOM-DQB*26c	5/53
WI920015	WI920015_4	MenoGB-0003-DQB*12c[B]	5/53

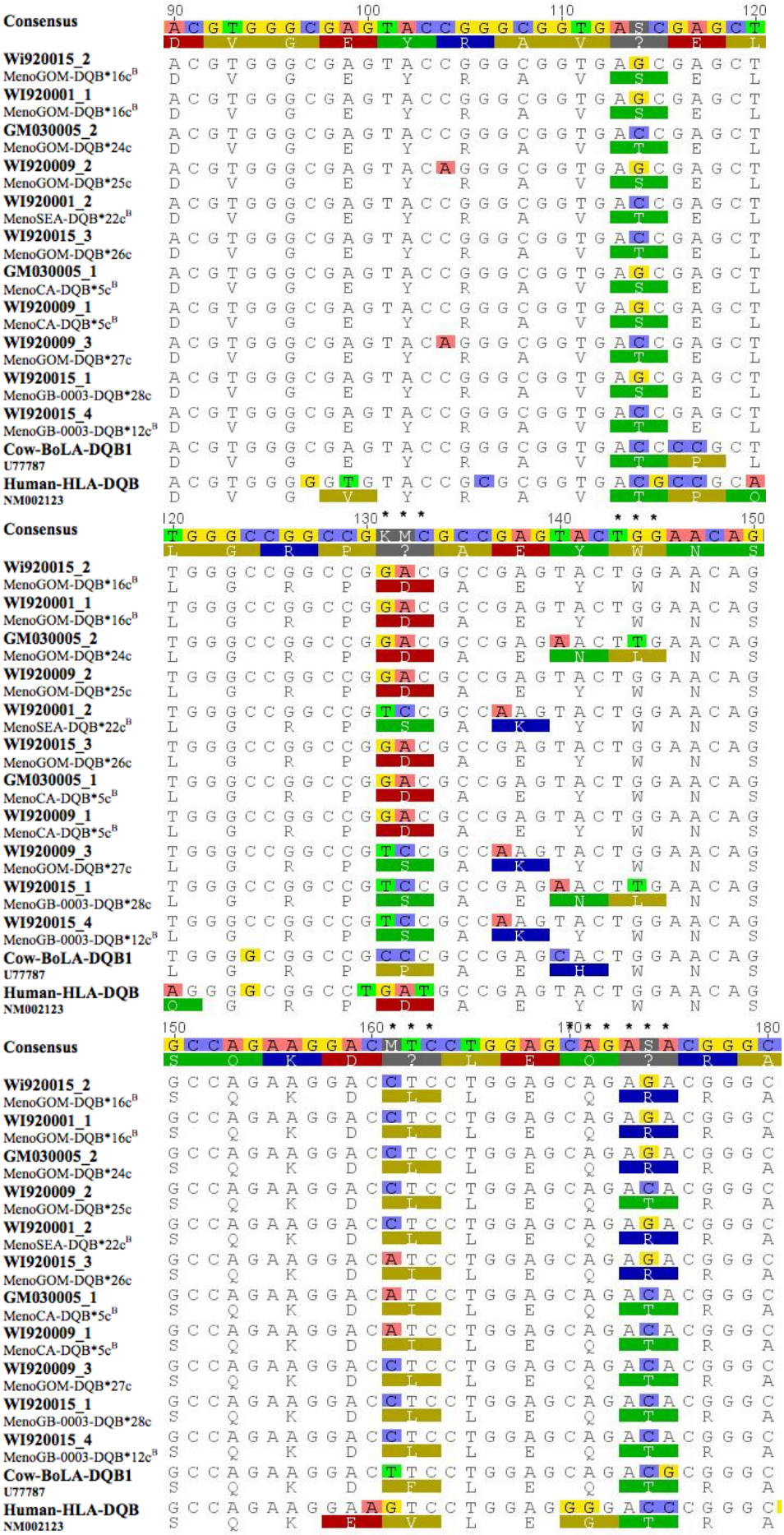
[B] Alleles MenoCA-DQB*5c, MenoGOM-DQB*16c, MenoSEA-DQB*22c and MenoGB-0003-DQB*12 have been described in (Baker *et al.* 2006).

Table 2: Results of two different tests for selection acting on a nucleotide sequence including p-values. Tajima's *D* [110] and Fu's and Li's test *D** and *F** [111] were calculated using all DQB exon 2 alleles from this study and all available DQB exon 2 alleles available on GenBank. The tests were calculated using first an alignment of all sequences, followed by an alignment of all nucleotide positions involved in antigen binding of all alleles available. The tests were also calculated only for an alignment of all nine sequences in this study, followed by an alignment of all nucleotide positions involved in antigen binding of these nine sequences. The nucleotide diversity for all different kinds of alignments is also given.

	Tajima's *D*	P-Value	Fu's test *D**	P-Value	Fu's test *F**	P-Value	Nucl. div.
All alleles	2.03	~0.05	1.33	~0.05	1.45	~0.05	0.058
Antigen-binding	2.41	<0.05 *	1.61	<0.02 **	1.74	<0.05 *	0.198
9 alleles	1.29	n.s.	1.28	n.s.	1.31	n.s.	0.054
Antigen-binding	1.89	~0.05	1.51	<0.02 **	1.59	<0.05 *	0.21

***** significant, $p < 0.05$, ****** significant, $p<0.02$


Consensus
Wi920015_2
MenoGOM-DQB*16c^B
WI920001_1
MenoGOM-DQB*16c^B
GM030005_2
MenoGOM-DQB*24c
WI920009_2
MenoGOM-DQB*25c
WI920001_2
MenoSEA-DQB*22c^B
WI920015_3
MenoGOM-DQB*26c
GM030005_1
MenoCA-DQB*5c^B
WI920009_1
MenoCA-DQB*5c^B
WI920009_3
MenoGOM-DQB*27c
WI920015_1
MenoGB-0003-DQB*28c
WI920015_4
MenoGB-0003-DQB*12c^B
Cow-BoLA-DQB1
U77787
Human-HLA-DQB
NM002123


Consensus
Wi920015_2
MenoGOM-DQB*16c^B
WI920001_1
MenoGOM-DQB*16c^B
GM030005_2
MenoGOM-DQB*24c
WI920009_2
MenoGOM-DQB*25c
WI920001_2
MenoSEA-DQB*22c^B
WI920015_3
MenoGOM-DQB*26c
GM030005_1
MenoCA-DQB*5c^B
WI920009_1
MenoCA-DQB*5c^B
WI920009_3
MenoGOM-DQB*27c
WI920015_1
MenoGB-0003-DQB*28c
WI920015_4
MenoGB-0003-DQB*12c^B
Cow-BoLA-DQB1
U77787
Human-HLA-DQB
NM002123

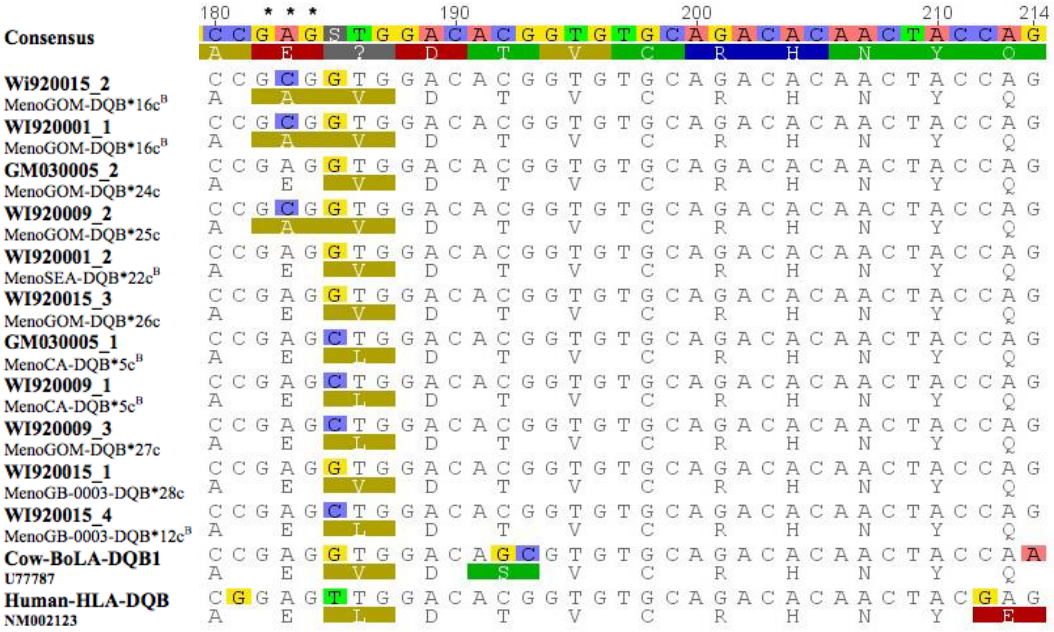


Figure 5: Alignment of the alleles of HLA-DQB exon 2 for the four cloned individuals, corresponding to table 1. Position 1 in the alignment of all my cloned alleles and the homologous gene in cow, *Bos taurus* (GENBANK accession number U77787) and human, *Homo sapiens* (GENBANK accession number NM002123) corresponds to the nucleotide position 40 in the human and cow HLA-DQB exon 2 [8, 92, 114]. Positions thought to be involved in peptide binding are marked with a star on top (*) and follow Brown *et al.* (1984) [100]. For the consensus sequence as well as for all individuals the nucleotide and corresponding amino acid sequences are given.

Allele-specific Primers II (after cloning)

Here again, primer combinations used resulted either in no amplification or in the amplification of several sequences of different length. Even the gradient amplification resulted in several sequences of different length up to a threshold 60 °C and from then on amplification failed (PCR reactions and conditions I consider most useful are listed in the appendix, table A1. They can be used for further experimentation and attempts to isolate single alleles). The attempt to sequence individual sequences directly, even after gel excision, still resulted in double and triple peaks.

Direct Sequencing

Ninety-seven individual humpback whales were sequenced between three and five times for both HLA-DQB exon 2- similar loci. Sixteen polymorphic positions were consistently scored and used for further analyses. Fourteen were dimorphic and two showed triple peaks. Fourteen positions resulted in non-synonymous amino acid changes and 10 positions are presumably involved in peptide binding [100]. In order to confirm my calling of the bases, another person scored 20 individuals; and the scores were identical to mine. Scoring of the chromatograms in Geneious resulted in 62 different multi-SNP genotypes, whereas 47 were single individual multi-SNP genotypes (table S3 and table S4 supplementary information). There were three groups of more than three identical multi-SNP genotypes.

Data Analysis

Sixteen polymorphic positions of the two simultaneously amplified loci were used for further data analysis (table S3 and S4 supplementary information). The script in R is given in the appendix. When all 16 positions were used for steps a, b and c of the script as described in the material and methods section, there were nine females that did not mate within the interval of random mating which lies between 2.5% and 97.5% of the simulated mating distribution (figure 7). One female had a significantly different calf, and eight females had significantly similar calves. The 16 SNPs include both synonymous and non-synonymous substitutions. Some of the positions showed higher variability than others. Since all 16 positions were treated as independent loci, the number of permutations for the whole sequence is very high. The script needed to run for at least 100,000 iterations to allow for the occurrence of all possible multi-SNP genotype combinations. Some rare combinations produced with the available females and males had a very low frequency; this makes an extreme genotype relative to the others more extreme. Therefore a second analysis was done using only seven polymorphic positions that are involved in antigen-binding. P44, P45, P50, P72, P131, P161 and P183 are the seven non-synonymous sites used for this analysis (table S5 supplementary information). Here the extreme genotypes were resumed with a smaller number of iterations, and the number of permutations was lower. Nevertheless this analysis included only the most functional positions in the sequence and was thereby much more informative (figures 6 and 7). The permutation produced on one hand too many possible sequences, but on the other hand they also accounted for the possible genotypes of non-sampled whales. Now 10 females were significantly different from the random distribution, one was significantly different from its calf and the other nine were significantly similar to their calves. It is noteworthy for both analyses that even females harboring genotypes that could be explained with only one or two alleles produced similar calves and did not go for different male genotyes (“A” in S5

supplementary information). Overall the average *SS* in all samples was not significantly different from the random distribution (figure 6). When females were mated with the most dissimilar males available, *SS* became significantly low (red line figure 6).

The probabilities of the observed similarity score for each mother-calf pair given random mating assessed using a Mann-Whitney U test are given for significant deviations in figure 7. The test was only applied to the simulations using seven positions to account for their higher resolution (involvement in antigen-binding; P44, P45, P50, P72, P131, P161 and P183). The same test was used to test if the distributions of similarity scores for each position in the observed mother-calf pairs were different from the random distribution (figure S3). There was no distribution that deviated significantly from the random simulation for each of the 16 positions.

To test if females preferably mate with more related males than expected; steps a, b and c were also analyzed with a different marker system, five microsatellite loci (figures 6 and 7). The results show a qualitative contrast to the MHC locus revealing information about inbreeding. Except for female 8, there was no sign of similarity or dissimilarity in the neutral markers of the females significant for the other marker system. The close to significant p-value for the microsatellites in female 8 was not significant anymore after applying Bonferroni correction. A summary of all three simulations is also given in table S5 supplementary information.

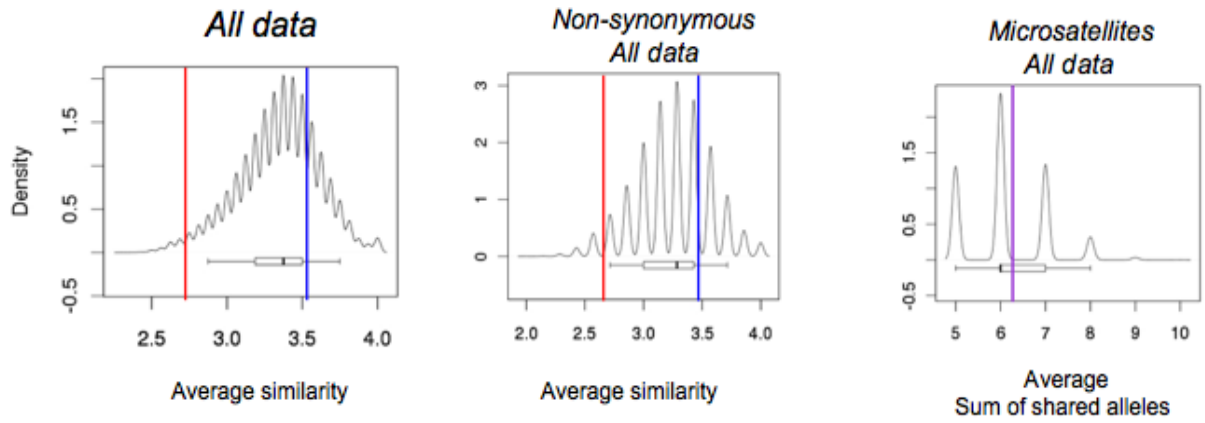
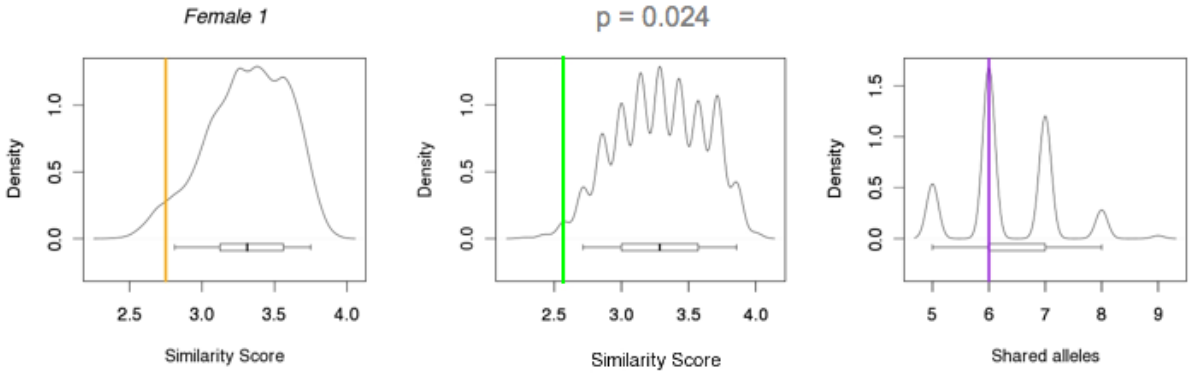
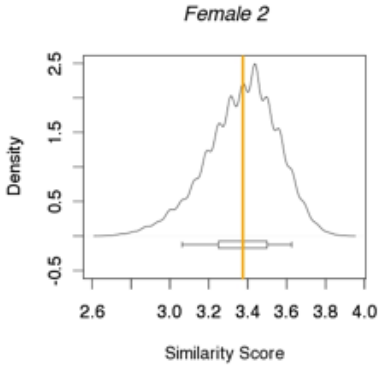


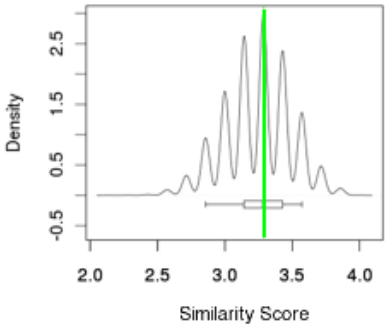
Figure 6: Expected density distribution and observed values for three simulations for the 27 females altogether. From the left to the right: the first graph shows the expected density distribution of similarity scores *SS* averaged over all females when mated randomly with the 43 available males, producing a calf with them and calculating the distribution of the thereby produced *SS* between a female and the produced calves using 16 polymorphic positions. In the middle, the same is repeated with seven polymorphic positions resulting in non-synonymous amino acid changes. On the right, the density distribution of the same analysis using five microsatellite loci shows the number of shared alleles instead of the average similarity scores. Blue line = observed *SS*, red line = *SS* for all females when forced to mate with the most dissimilar male in the sample, purple line = average of the sum of shared alleles for the microsatellites for all mother-calf pairs.

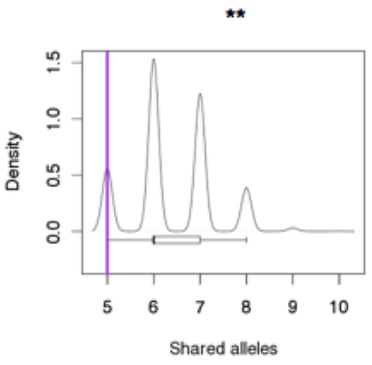

Female 1
Density
Similarity Score

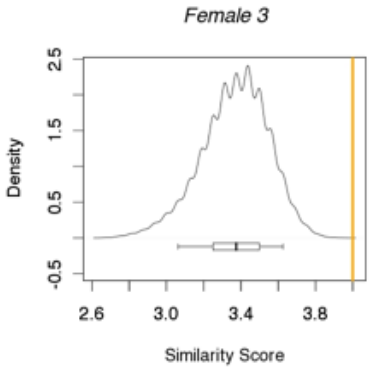
p = 0.024
Density
Similarity Score

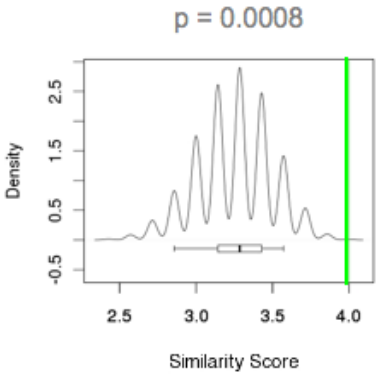
Density
Shared alleles


Female 2
Density
Similarity Score

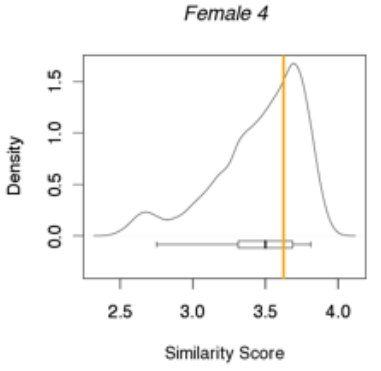

Density
Similarity Score

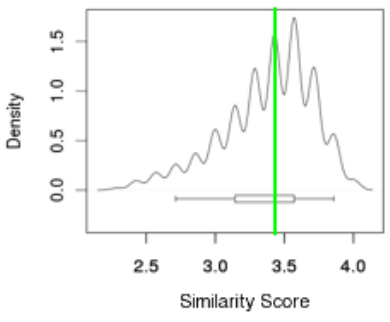

**
Density
Shared alleles

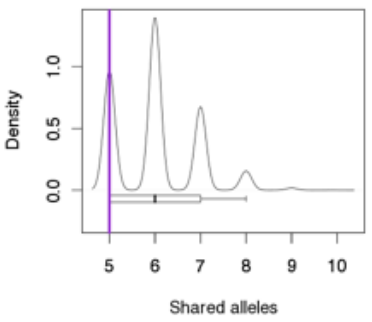

Female 3
Density
Similarity Score

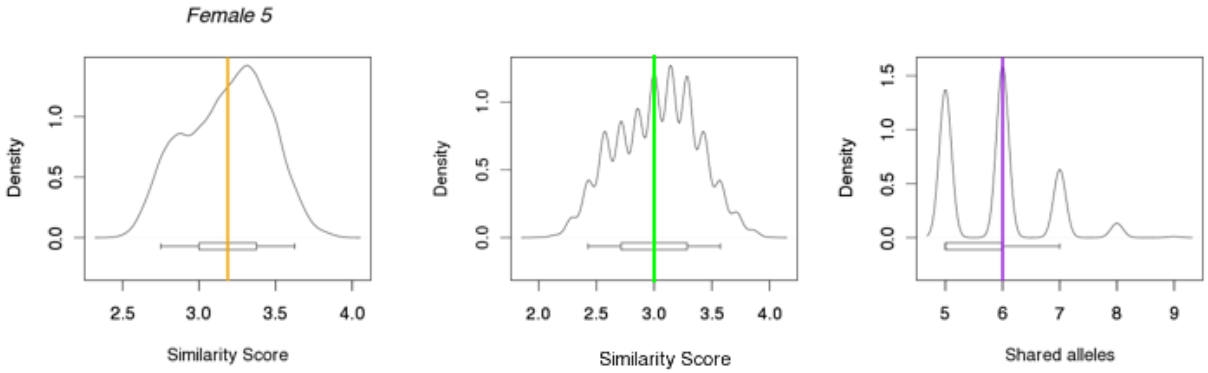

p = 0.0008
Density
Similarity Score

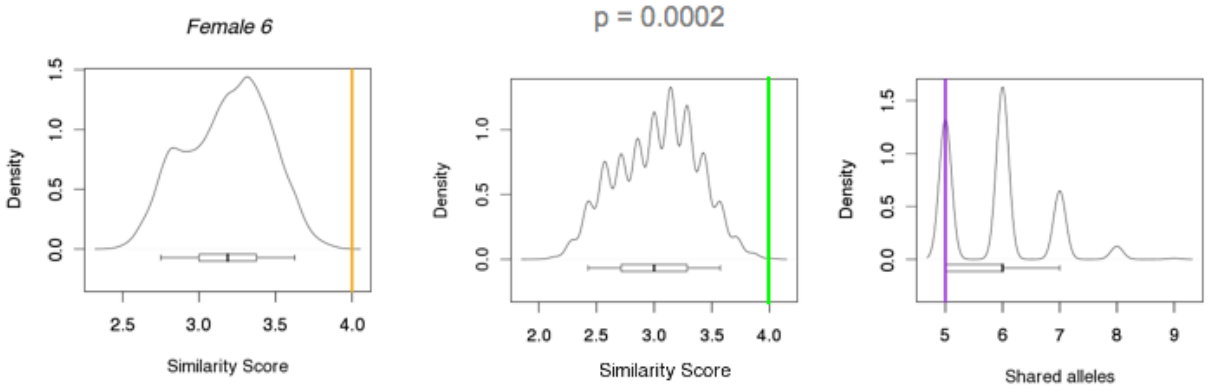
NA

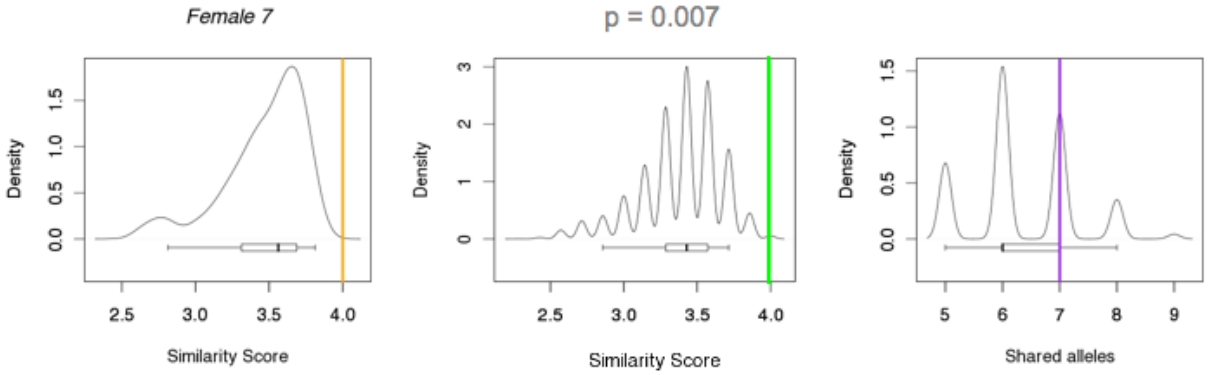

Female 4
Density
Similarity Score

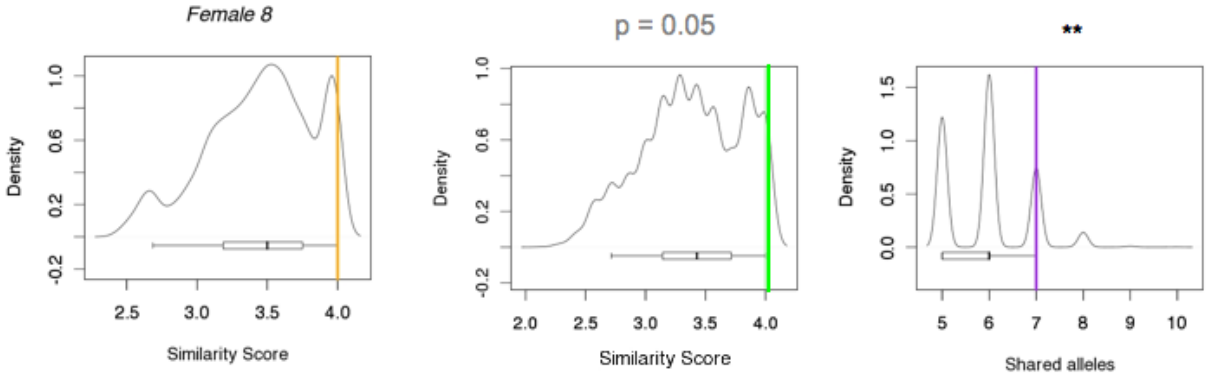

Density
Similarity Score

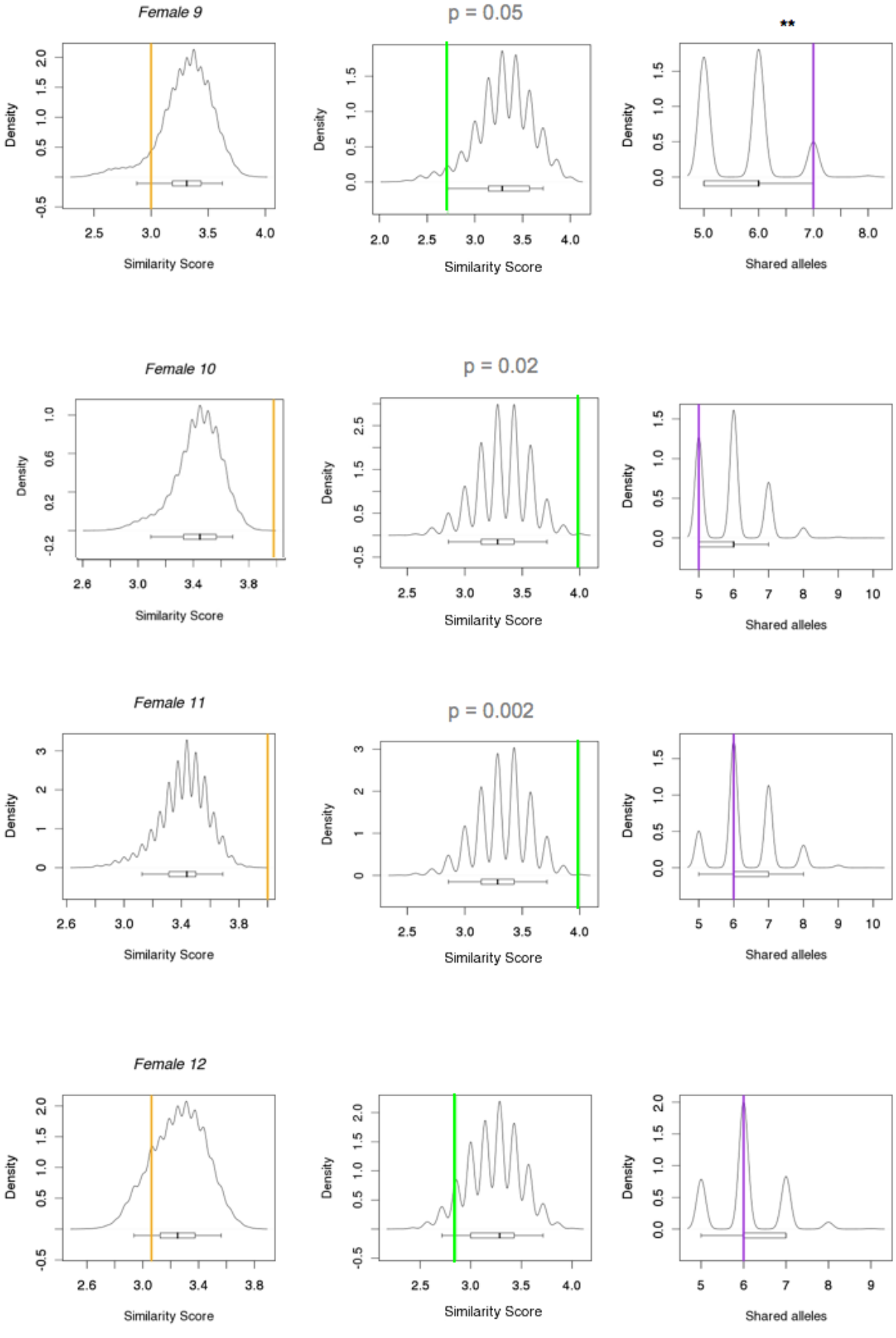

Density
Shared alleles

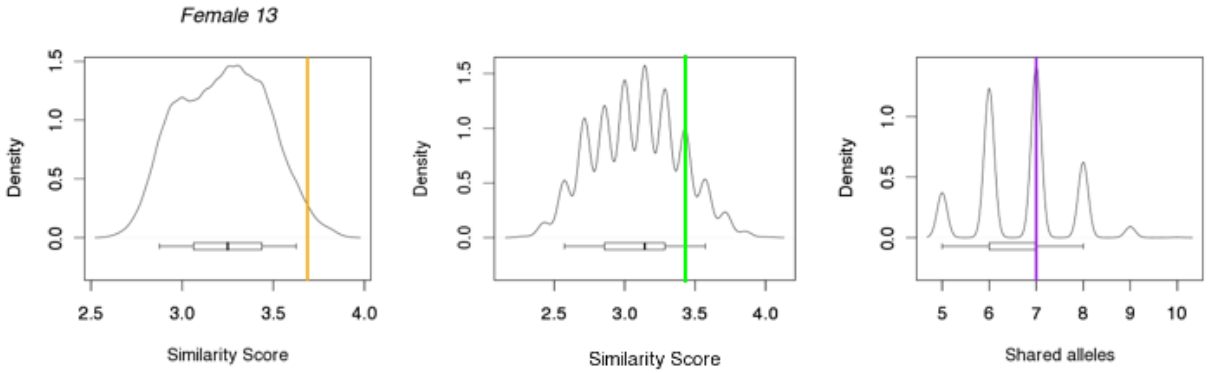

Female 5
Density
Similarity Score
Density
Similarity Score
Density
Shared alleles

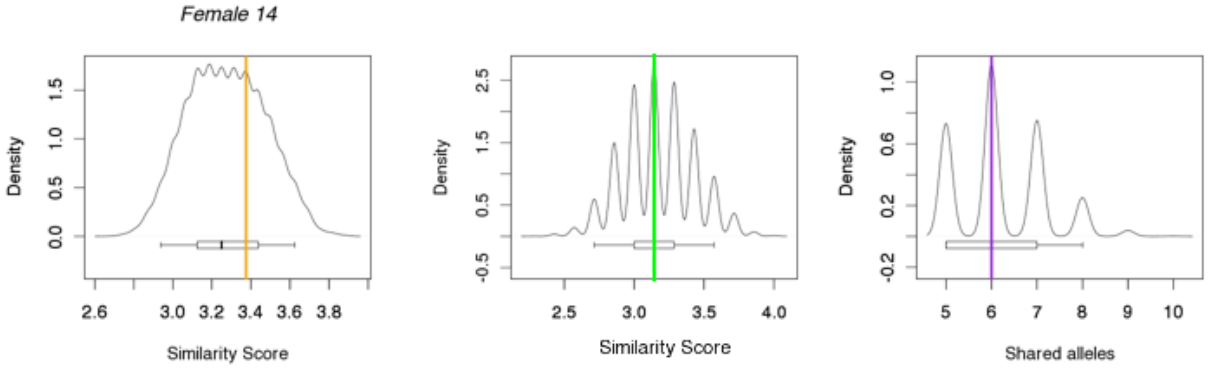

Female 6
p = 0.0002
Density
Similarity Score
Density
Similarity Score
Density
Shared alleles

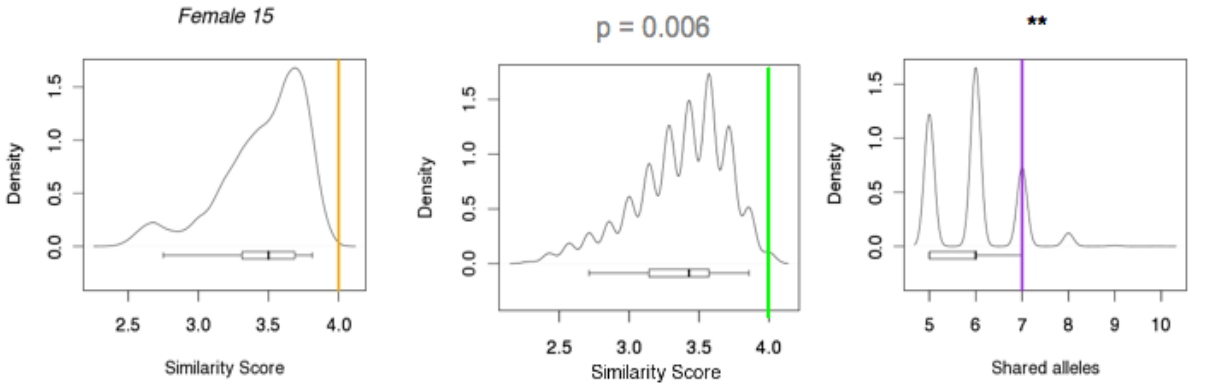

Female 7
p = 0.007
Density
Similarity Score
Density
Similarity Score
Density
Shared alleles

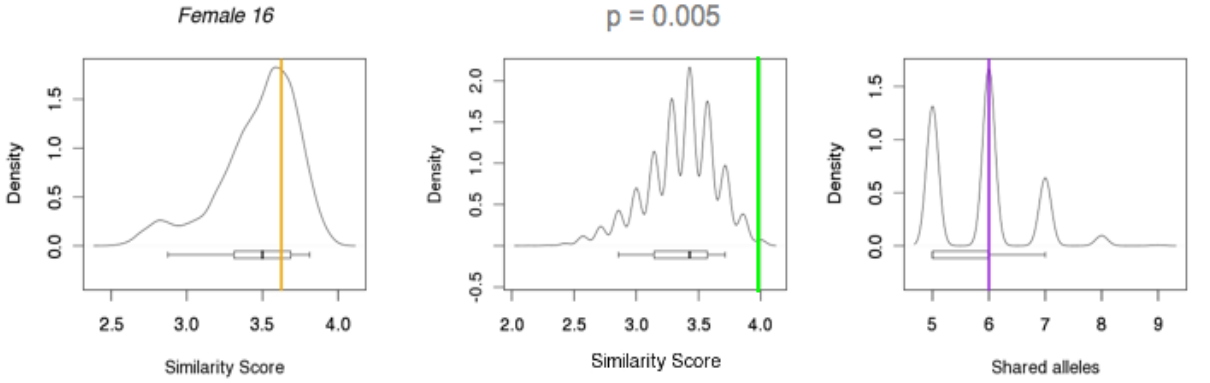

Female 8
p = 0.05
**
Density
Similarity Score
Density
Similarity Score
Density
Shared alleles

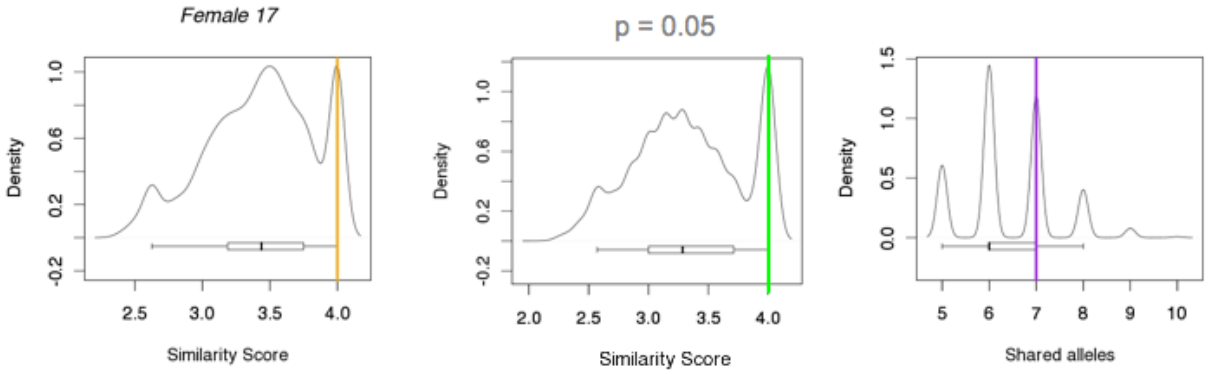

Female 9
p = 0.05
**
Female 10
p = 0.02
Female 11
p = 0.002
Female 12
Density
Similarity Score
Shared alleles

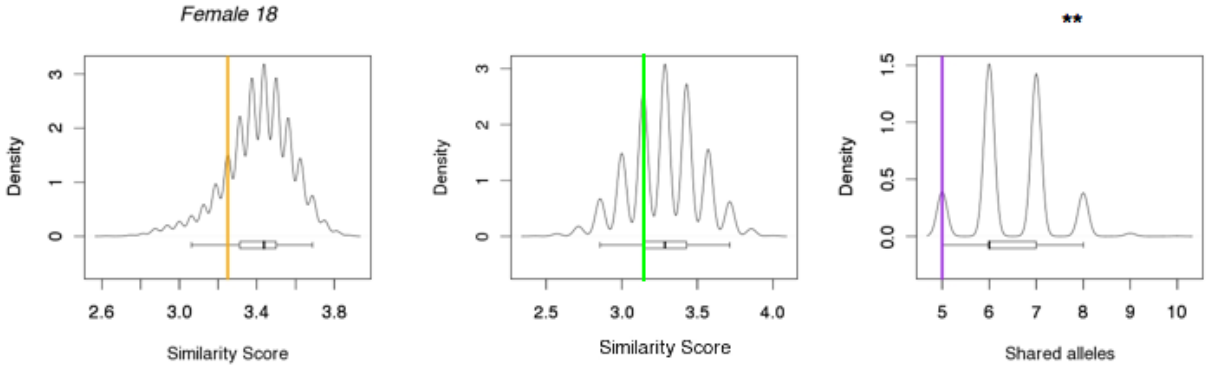

Female 13
Density
Similarity Score
Density
Similarity Score
Density
Shared alleles

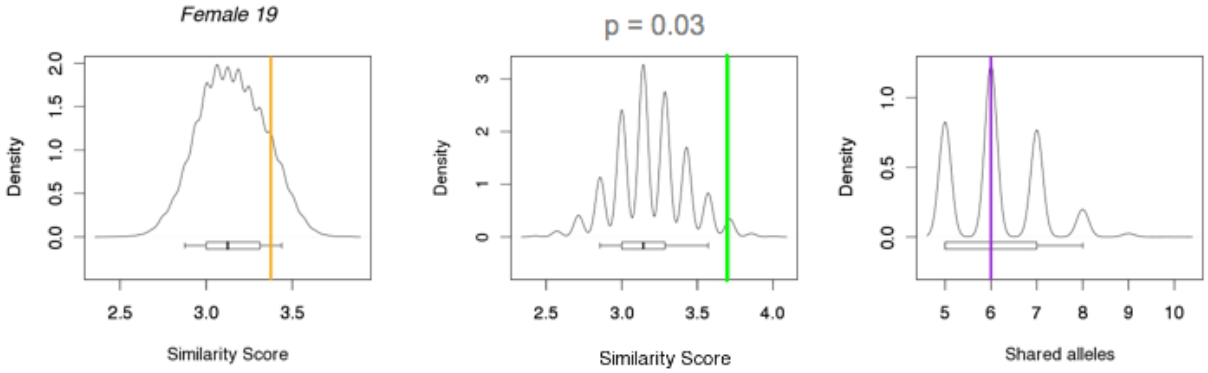

Female 14
Density
Similarity Score
Density
Similarity Score
Density
Shared alleles

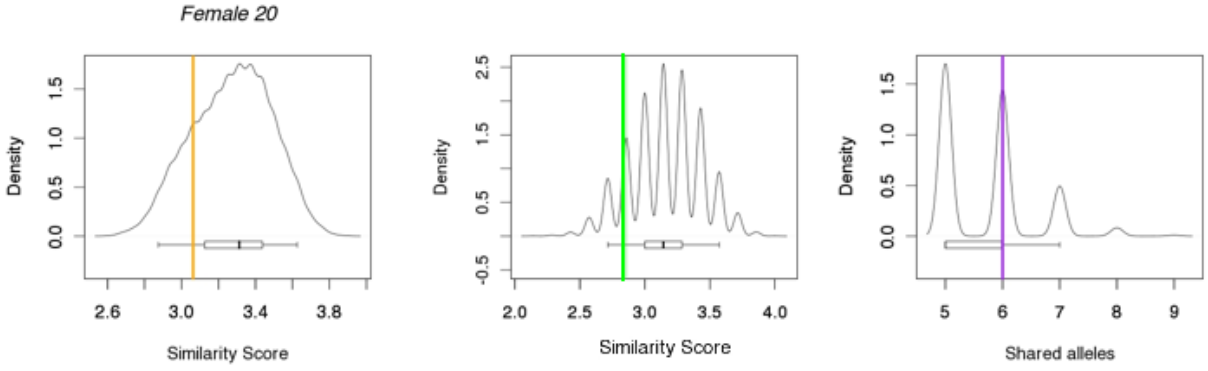

Female 15
Similarity Score
p = 0.006
Density
Similarity Score
**
Density
Shared alleles


Female 16
Density
Similarity Score
p = 0.005
Density
Similarity Score
Density
Shared alleles

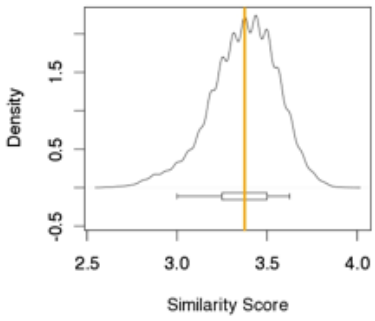

Female 17
Density
Similarity Score
p = 0.05
Density
Similarity Score
Density
Shared alleles

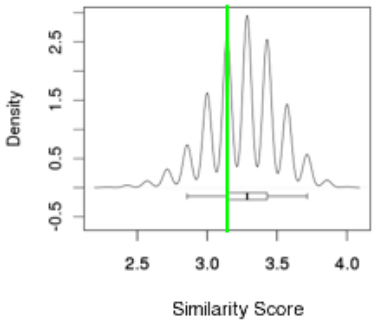

Female 18
Density
Similarity Score
Density
Similarity Score
**
Density
Shared alleles

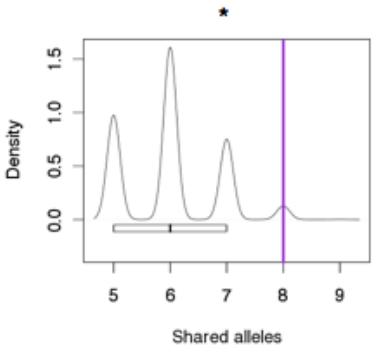

Female 19
Density
Similarity Score
p = 0.03
Density
Similarity Score
Density
Shared alleles


Female 20
Density
Similarity Score
Density
Similarity Score
Density
Shared alleles

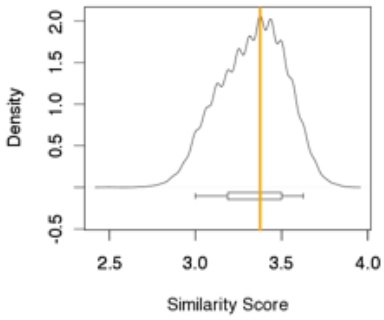
Female 21

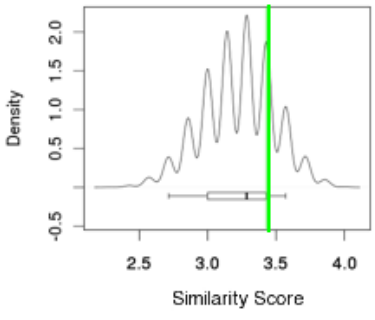


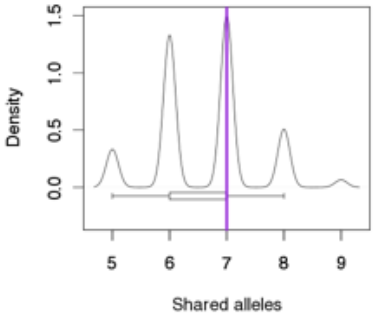




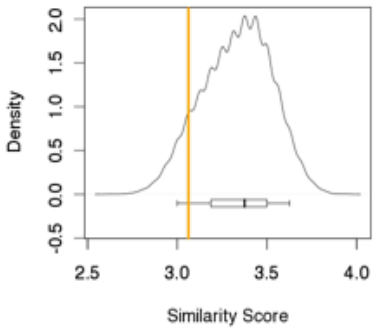
Female 22

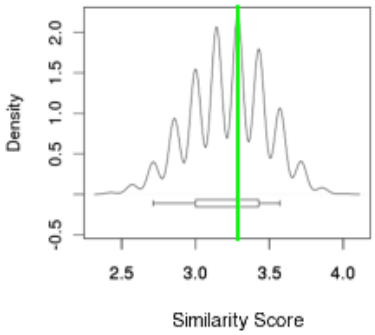


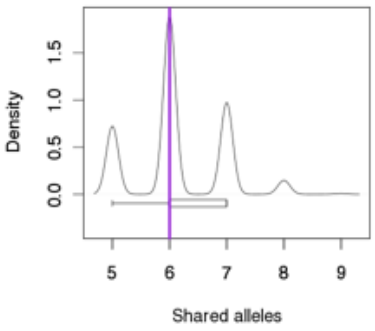




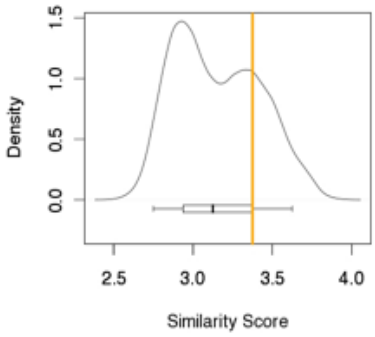
Female 23

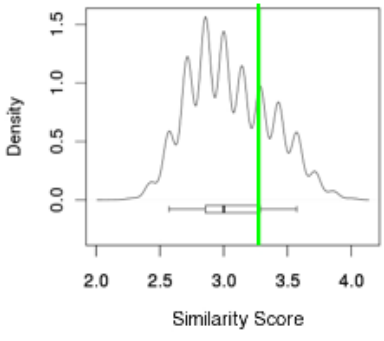


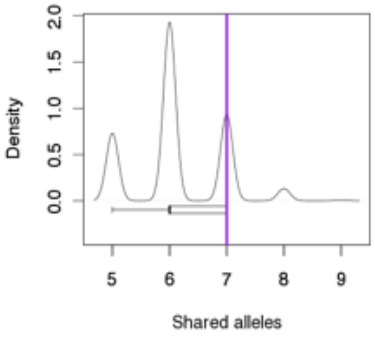




Female 24







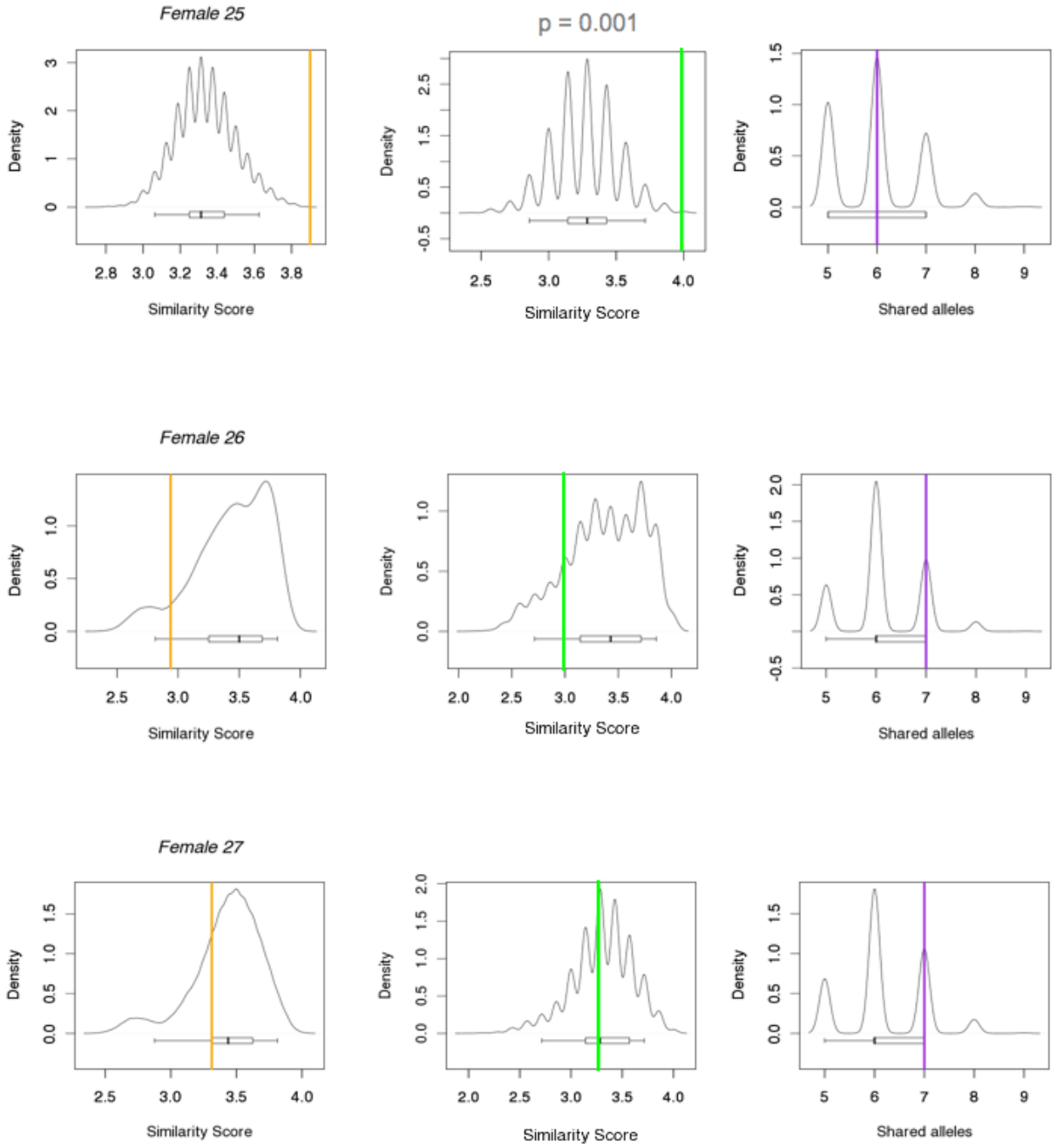


Figure 7: Expected density distribution and observed values for three simulations for every female. From the left to the right: the first graph shows the expected density distribution of similarity scores *SS* for a particular female when mated randomly with the 43 available males, producing a calf with them and calculating the distribution of the thereby produced *SS* between her and the produced calves, using 16 polymorphic positions. In the middle the same is repeated with seven polymorphic positions resulting in non-synonymous amino acid changes. On the right is the density distribution of the same analysis using five microsatellite loci showing the number of shared alleles instead of the average similarity scores. The orange, green and purple lines depict the observed scores of the analyses of 16 polymorphic positions, seven non-synonymous positions and microsatellite analysis, respectively for every observed female and her calf. Significant p-values are given for the probability of observing a given similarity score under random mating. * = significantly different value from simulated distribution, ** = p-value at 0.05, NA = microsatellite data not available.

There did not appear to be any pervasive effects of inbreeding, population subdivision, or null alleles in the data set of microsatellites (table S6 to table S9 supplementary information). When only the mothers from the Gulf of Maine were treated as a single group (Mothers GoM and analyzed), two loci were in linkage: TAA031 and GATA098. This is most probably an effect of sampling. In the other groups there was no linkage observed. In a sample of over 3000 humpback whales from the North Atlantic Ocean there were no deviations from Hardy Weinberg equilibrium and no signs of linkage disequilibrium [74-77]. The allele frequencies between mothers sampled in the Gulf of Maine and random males sampled in the West Indies deviated slightly from each other. The differentiation was highest at the locus GATA053 (table S9).

Using the nine HLA-DQB exon 2 alleles obtained by cloning and 23 alleles from humpback whales sampled mostly in the Pacific Ocean [71] and published on GenBank, the actual four alleles in every directly sequenced individual were inferred using the software PHASE [108]. For individuals harboring more than two alleles of a different kind one or two alleles were subtracted by eye using the nine cloned alleles prior to the analysis with phase. This analysis resulted in 63 different alleles in the whole sample and 43 from whales in this study (table A2 appendix). This high variation made the procedure of forcing every individual to harbor at least one of the nine alleles resolved from cloning much too constraining. Some mother-calf pairs did not share two out of four bases per position anymore. The 63 alleles were not used for further analyses.

Tajima's *D* test is based on the differences between the number of segregating sites and the average number of nucleotide differences [110], and it tests the hypothesis that all mutations are selectively neutral. Fu and Li's tests also test the hypothesis that all mutations are selectively neutral [115]. The *D** test statistic is based on the differences between the number of singletons (mutations appearing only once among the sequences) and the total number of mutations [111], whereas the *F** statistic is based on the differences between the number of singletons and the average number of nucleotide differences between pairs of sequences. The two test approaches were applied to (i) an alignment of the nine alleles obtained by cloning, (ii) an alignment of the nine cloned alleles and all remaining alleles from GenBank and (iii & iv) an alignment of (i) and (ii) respectively, using only the positions involved in antigen-binding [100] (table 2). Alignments (i) and (ii) were close to significance, whereas (iii) and (iv) were significant for all tests. The sequences appeared to be functional and selectively important. A neighbor-joining tree of all alleles showed that there was no clustering of them depending on geography (North Atlantic Ocean vs. Pacific Ocean, results not shown here).

DISCUSSION

The goal of this study was to investigate if female humpback whales, *Megaptera novaeangliae*, preferably mate with males harboring different alleles to their own at the HLA-DQB exon 2 locus. By comparing observed similarity between mothers and their calves to the similarity of randomly generated calves and their mothers, mate choice with respect to the MHC was tested. Thereby the similarity between mothers and fathers could be indirectly assessed and the result of pre- and/or post-copulatory selective mechanisms was investigated.

Two simultaneously amplified Loci

Different species seem to exhibit different rates of gene conversion and recombination at the MHC and its region, leading to multiple simultaneously amplified MHC genes which are closely linked [12]. This makes it difficult to isolate and amplify single loci [116-118]. In this study I did not succeed in amplifying only a single HLA-DQB exon 2-like diploid sequence but ended up with the co-amplification of two loci. All nucleotide sequences obtained from cloned PCR products as well as obtained by direct sequencing appeared to be fully functional devoid of stop codons. In total 74.9% of the substitutions were non-synonymous for all alleles available, and 75.8% for the 9 cloned alleles in this study. It has been argued earlier that the HLA-DQB gene in humpback whales is duplicated [112]. Curiously, the functionally related HLA-DRB gene in humpback whales showed insertions and deletions when amplified with a universal primer pair designed for different cetacean species. The HLA-DRB locus, one of the most polymorphic loci in other vertebrates [7, 94, 119-121], might have lost its functionality in humpback whales after a duplicated version of the HLA-DQB gene had taken over its function. An alternative explanation could be that the universal DQB primers amplify both genes, HLA-DQB and HLA-DRB, in humpback whales; and both genes are highly linked. It could also be that the HLA-DQB locus itself is duplicated and both copies remained functional. All three explanations would explain why there could be a 3:1 ratio of different alleles in the cloned individual GM030005, indicating a homozygote locus and a heterozygote locus sharing one allele with each other. I sequenced 60 clones for this individual, resulting in 37 alleles of one kind and 20 alleles of another kind. This is closer to a 3:1 ratio than to a 2:2 ratio but it is not obvious.

In order to investigate the function of a gene it is necessary to know the sequence of the haplotypes. To quantify the similarity between genotypes, the knowledge of the phased sequences would allow a more precise estimate. The amino acids, as a proxy for the functionality, could be compared among different genotypes; and functional domains could be described [122]. In this thesis I could not infer the haplotypes of all individuals and had to treat all polymorphic positions as independent from each other in the analysis simulating random mate choice. As shown in figure S3 (supplementary information), the single positions alone did not deviate from a random distribution. It must have been the linkage of some of the polymorphic positions to a sequence that were most informative and could produce significant results. (figures 6 and 7).

There are several different ways to separate the two loci with molecular methods. The target of research, i.e. the extreme polymorphisms, lead to an increased rate of artifact (singleton) formation *in vitro* [123]. The most straightforward solution would be to use next-generation sequencing technology. Tagged primers with an individual barcode of a few nucleotides offer the opportunity to genotype in parallel hundreds of individuals at several

loci [123, 124]. The large number of sequence reads can then be used to identify artifacts by frequency distribution thresholds, working with complex multi-allelic templates and large sample sizes [118]. Babik *et al.* genotyped 96 bank voles using degenerate primers. This study yielded 98 sequences per individual representing a mean of 5 loci. By running a fraction of duplicates, they quantified an intrinsic threshold of 3% whereby alleles occurring less than that per individual are considered artifacts. For the time being, this method is too expensive for a one year Master thesis.

Other methods include SSCP (Single Strand Conformation Polymorphisms) which is based on the different conformational properties of single-stranded nucleotide sequences during gel electrophoresis [125]. SSCP was used to separate the two simultaneously amplified HLA-DQB exon 2-like loci in humpback whales from the Pacific Ocean (C. S. Baker, *pers. com*). A system needs to be established empirically for non-model species. This did not work out after two years of attempts in the Pacific Ocean humpback whale. A similar method is DGGE (Denaturing Gradient Gel Electrophoresis) that relies on a conformational change of a sequence caused by a chemical gradient across the gel [126]. As before for SSCP, this system needs to be established prior to analysis. A further related analysis is RSCA (Reference Strand-Mediated Conformational Analysis) where allelic variants are separated from each other as heteroduplexes containing a fluorescent-labeled reference sequence (FLR) [127]. By using several FLRs, alleles may be assigned unique migration patterns on a gel. All three methods reveal only information about the genotype of an individual. To obtain sequence information, the bands need to be excised from the gel, purified and sequenced. Another method worth mentioning is simply cloning. The last four methods mentioned above are all expensive, labor-intensive, subject to PCR artifact and always likely to underestimate true variability. For this study, I decided to design locus-specific primers in a first step, clone a subset of individuals in a second step and infer the rest of the sequences with a statistical method in a third step. This is a trade-off of cost and work under the time constraints of a Master's thesis (maximum 60 credits translating into one year). I sequenced between 42 and 60 clones per individual and detected thereby 12 singletons (table S2 supplementary information) although I followed the suggestions to reduce artifacts [95]. The quantification of the artifact formation in bigger scale projects is therefore a must. For future investigations of the HLA-DQB locus in humpback whales, I suggest using next-generation sequencing directly. If financial constraints make this impossible, I would extend the current work using inverse PCR [128-131] to obtain the nucleotide sequence of the flanking introns thereby increasing the length of nucleotide sequence from which allele-specific primers may be designed.

During this study, I ended up amplifying two loci simultaneously and decided to treat the data as a tetraploid locus. The comparison of mother-calf pairs with four possible states (A, C, G, and T) at each SNP permitted an estimation of the degree of genetic similarity between each mother and her calf. When this was compared to the similarity of randomly generated calves and their mothers, I could indirectly describe the similarity between mothers and fathers. There were at least four sequencing replicates for every individual. At least three replicates were needed because during the PCR amplification one locus may stochastically be amplified in a higher frequency relative to the other. I found 62 different combinations (multi-SNP genotypes) in a sample of 97 individuals. The minimum number of alleles required to produce 62 different individuals with four alleles amplified simultaneously is five (for three alleles amplified simultaneously it would be seven alleles and for two alleles it would be 11). Since most of the individuals had two to three bases at the same position, the minimum is seven alleles. The phasing analysis revealed 43 alleles in this sample, divided by two to divide into the two loci; this is what has been found in most wild vertebrate populations of this sample size [57, 67, 94, 117, 132-136].

The method applied in this study is prone to produce noise during the simulation caused by the huge increase of possible combinations treating all multi-SNPs as unlinked. Using only the seven most polymorphic, non-synonymous positions reduced the noise revealing that indeed many of the calves were more similar to their mother than expected by chance.

Female Mate Choice

Why is there female mate choice? To answer this question we have to go back and ask why there is sexual reproduction at all.

To maintain sexual reproduction, the recombination of genes through selective mate choice must achieve at least a twofold genetic benefit in each generation [12]. During the process of sexual selection, half of the genome is passed on to the offspring through Mendelian inheritance to combine it with the half of the genome from a different individual. Hypotheses put forward to explain the advantage of sexual reproduction include the "Muller's ratchet" [137] where asexual lines will accumulate deleterious mutations until the genome goes extinct due to high genetic load. However, this mechanism is too weak and too slow to explain that producing some mutation-free offspring would compensate for the twofold efficiency advantage of asexual reproduction. Kondrashov tried to rescue Muller's ratchet and added two important assumptions [138]. The mutation rate must be greater, and multiple mutations must have an increasingly damaging effect (synergistic epistasis). Most experiments so far have failed to demonstrate Kondrahov's hypothesis [12]. In the context of parasites, a "Red Queen hypothesis" might explain how sex is favored if selective mate choice produces a twofold benefit in each generation [25, 27, 139, 140]. Infectious diseases occur in a vast amount of different species. Caused by a changing parasitic environment, new combinations of genes for resistance are required in every generation to cope with the currently dominating parasites. The individual conducting mate choice may complement his or her own set of alleles with a more or less diverse set of partner alleles to reach an optimal number of different MHC alleles for the offspring [85, 141]. The optimal complement from the partner has to include those alleles that provide resistance against current parasites in the population. In some species this can be revealed by the expression of costly secondary sexual characters [59] or generally by the condition of the possible partner. Without any information, preferring MHC dissimilar mates might be a best-of-bad-job-rule [51, 52, 142]. MHC ligand peptides may provide olfactory clues aiding mate choice [41, 45, 52, 143]. This would allow a context-dependent decision making on condition-dependent traits. Mate choice provides a mechanism to reinforce all three possible selection mechanisms. As the Red Queen said to Alice in Wonderland: "It takes all the running you can do, to keep in the same place." [144]. The Red Queen hypothesis can be tested in populations with both asexual and sexual reproduction. A correlation has been shown in the lizard *Heteronotia binoei* [145].

Two requirements have to be fulfilled to justify sexual reproduction as a hypothesis to overcome asexual reproduction. First, there should be variation of genes conferring resistance in the population. The choosing partner has to be able to find other genes than those that he or she dropped. Second, the choosing partner has to know her or his own resistance genes; and accordingly he or she must also be able to recognize the partner's genes. This can function through an olfactory system. As mentioned earlier, the MHC is one of the most polymorphic if not the most polymorphic locus in vertebrates. Many species have been shown to be able to recognize kin and preferable mating partners based on the MHC; for example, mice and rats [26, 45-52], lizards [53], fish [54-56] and birds [57-60], as well as humans [61-63]. In rodents there are specific receptors to respond differentially to the nine-amino-acid peptide ligands

that bind to MHC proteins (i.e. the vomeronasal organ) [143]. MHC peptide ligands are released when MHC proteins are broken down, and they are filtered into urine (as waste products). Female sticklebacks prefer males with an appropriate number of alleles [85], and a male's MHC diversity is assessed by self-referent matching of MHC peptide ligands detected in the water in which the male has been swimming.

Besides the odor, the condition of the potential partner is also a signal in the context of parasites. From a distance the choosy partner selects by smell the partner offering the right combination of MHC alleles and, after a closer look at condition and behavior, a healthy partner will be chosen.

Female Mate Choice in the Humpback Whale

The humpback whale mating system is polygynous with relatively little skew in male reproductive success [146] and with attributes similar to a lek [147]. The males have no means to force a female to mate with them. They were observed performing different behaviors during the same season such as singing alone or being part of a competitive group [69]. Presumably they attract females by singing first and during a further step fight with other males for access in the competitive group that courts her, i.e. they rotate in strategy. This behavior offers the possibility to the females to evaluate the male's quality through honest signals. Females take the risk of traveling down to the breeding ground every time for mating. Because of their investment in parental care, one would expect females to be the choosier sex [148-150]. The prerequisites for female mate choice are met in the following way:

(1) There is variation of genes conferring resistance in the population of humpback whales in this study. Both simultaneously amplified loci seem to be functional and the females can choose from between nine to 43 different alleles. Cloning and sequencing of the pooled sample of amplifications with the two universal DQB primers in five individuals in another study of humpback whales revealed 23 alleles [112]. The variation to choose from is highest in the biggest breeding ground (West Indies) where the male to female ratio lies around 2.5 [77]. By traveling to the breeding ground where the males gather, the females increase the collection of genes to choose from.

(2) Females can get into the ovulative phase caused by the presence of a male in the breeding ground (J. Jacobsen, *pers. obs.*). It is not fully understood how the humpback whales perceive their odor environment through smell. I searched for a vomeronasal organ-like sequence, which is present in most mammal families as well as bird and fish families [49, 143, 151], in the only whole genome available of a cetacean species, *Tursiops truncatus*. There was no agreement but the sequence coverage is very low, and the whole genome is not finished yet. The lek-like breeding ground offers the opportunity for males to show their quality through singing [152-155] and competitive courtship behavior in the male competitive groups [156].

Although the simulations run in this study produced too many permutations by treating all polymorphic positions independently, nine out of 27 females turned out to produce significantly similar calves. This represents one third of the whole dataset. Only one female produced a significantly dissimilar calf. Random mate choice based on the MHC HLA-DQB locus can therefore not be assumed. There is no trend of female mate choice for dissimilar alleles at this locus as it has been seen in so many other studies. In mice it has been suggested that the MHC could act as an odor source to discriminate between related and non-related

mates. Here this does not apply. Females tended to mate with more similar males. Pre-zygotically, this could represent a choice for good genes. The confounding factor of inbreeding can be ruled out as an alternative explanation since I controlled for it, running the analysis for five microsatellite loci. The signal observed in this study needs to be investigated further using concrete allele sequence information. A trend of female mate choice in this direction would call for new interpretations of female mate choice based on the MHC. One way of interpretation is that female humpback whales in the breeding ground prefer to mate with HLA-DQB- similar males from the pool of all males in an attempt to provide their offspring with the optimal genes in the context of their fidelity to the same feeding ground. This hypothesis could be tested simply by comparing HLA-DQB- allele frequencies among feeding grounds.

Alternatively, the signal could be produced by post-copulatory mechanisms, such as post-zygotic selection. A similar mechanism has been shown in sticklebacks where only offspring with intermediate variation of MHC class II genes survived more often, although females mated with a wider range of dissimilar mates [157, 158].

CONCLUSION

The universal primer pair to amplify the HLA-DQB exon 2 locus in cetaceans amplifies two highly similar loci in humpback whales, *Megaptera novaeangliae*. The design of locus-specific primers was not successful. Not even the cloning of several individuals and the sequencing of their alleles solved the problem. Nevertheless, the cloning revealed that individuals could be sequenced directly for both loci simultaneously and scored at every polymorphic position (multi-SNP) independently for four instead of two alleles. The cloned alleles showed signs of functionality and selection. The rate of non-synonymous changes was greater than synonymous changes, and most polymorphisms were found in the codons involved in antigen-binding.

Sixteen consistently scored multi-SNPs of the two genes in the major histocompatibility complex (MHC) led to the discovery of eight out of 27 females that produced significantly similar calves. When only the positions involved in antigen-binding were investigated this trend became even clearer, including an additional female. Female mate choice for similar alleles has not been shown yet in any system with the exception of inbred species. Here I controlled for inbreeding using neutral microsatellite loci. Two possible hypotheses to explain this finding are pre-zygotic female mate choice for good genes, in this case similar genes dependent on the feeding ground and/or post-zygotic selective mechanisms favoring the development of the genetically similar embryo.

SUPPLEMENTARY INFORMATION

Table S1: Primers used in this study. The numbering corresponds to figure S1. Several of these primers are listed in the appendix for the allele-specific PCR reactions table A1.

	Name	5'-3' Sequence	Target	Designer
1	AL1_403_F	CGG GCA GTG GAG AGA TAC AT	Forward	LGS *
2	AL1_716A_F	TAT AAC CGG GAG GAG TTC CTG CGT	Forward	LGS
3	AL1_716B_F	GGA GGA GTT CCT GCG TT	Forward	LGS
4	AL2_403_F	TGC GGT TCG TGA CCA GAC	Forward	LGS
5	DQB1	CTG GTA GTT GTG TCT GCA CAC	Reverse	Murray (1995)
6	DQB2	CAT GTG CTA CTT CAC CAA CGG	Forward	Murray (1995)
7	DQBex3F	GAA CCT ACA GTG ACC ATC TC	Forward	Heimeier (2009)
8	DQBex3R1	GAG ATG GTC ACT GTA GGT TCC ACT	Reverse	LGS
9	DQBex3R2	TCT GGA GGC TGG AGT GCT CC	Reverse	Heimeier (2009)
10	DQB_Intron2F	CGG TGT GCA GAC ACA ACT A	Forward	LGS
11	DQB_Intron2R	AAG CAT CTG GAA GGT CCA GT	Reverse	LGS
12	DQB_Intron2_F2	GAG GTG GAC ACG GTG TGC AGA CAC AAC TAC	Forward	LGS
13	DQB_Universal	ATG TYT GGG ATG GTG SCT CTG	Forward	Heimeier (2009)
14	DQBEXON3R3	AGC ATC ACA AGC ATC TGG AAG GTC CAG T	Reverse	LGS
15	DQBHM01	TGC GGC TCG TGA CCA GAT	Forward	LGS
16	DQBLC01_F	ACT KGA ACA GCC AGA AGG ACC	Forward	LGS
17	DQBLC02_F	ACT GGA ACA GCC AGA AGG ACA	Forward	LGS
18	DQB_Intron2_F3	GAG CAG AGA CGG GCC GAG GTG GAC A	Forward	LGS
19	DQB_Intron2_R2	CTA ATA AGA GGG GTG GAC ACA ACG CCA GCT GT	Reverse	LGS
20	DQB_Intron2_R3	CGG CGG AGA TGG TCA CTG TAG GTT CCA CT	Reverse	LGS

***** LGS = Laetitia G. Schmid

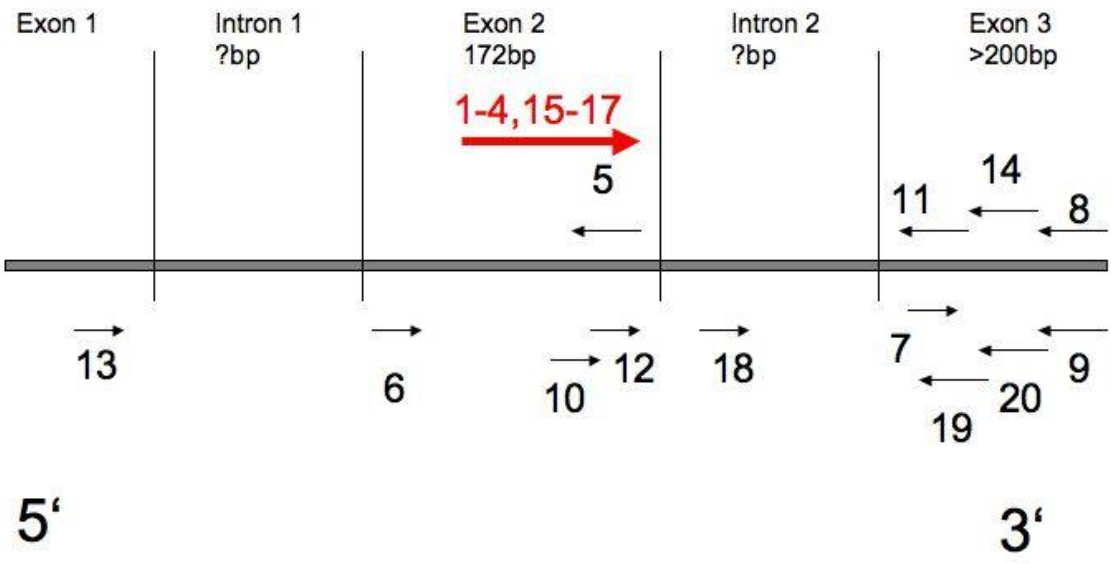
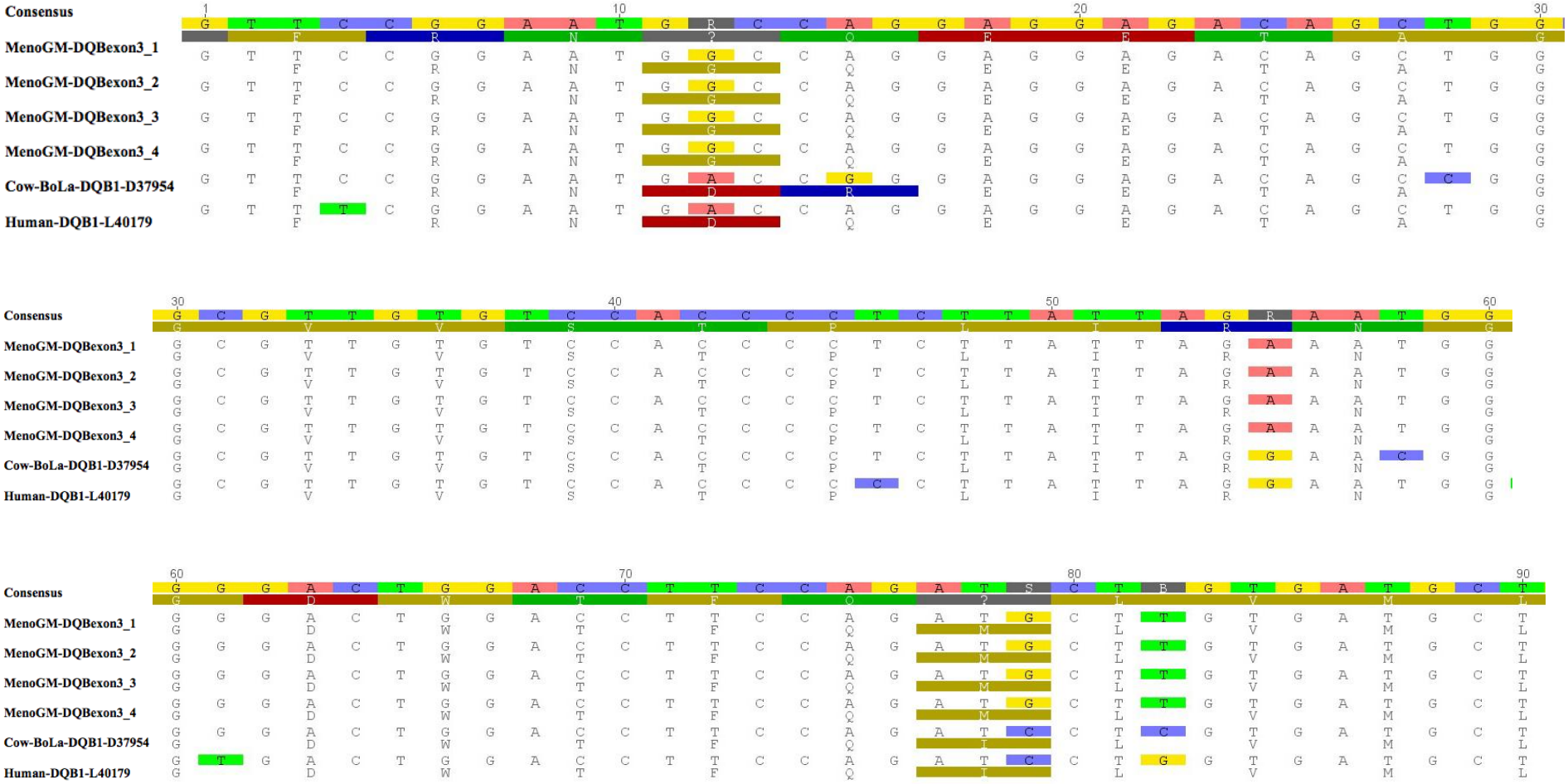


Figure S1: Schematic representation of the DQB gene including exon 1 to 3 in 5’-3’ direction. The arrows represent primers and their orientation. The corresponding names to numbers belonging to each arrow are given in table S1.

Table S2: Sample IDs of the four individuals cloned for HLA-DQB exon 2 in this study. The numbers of PCR amplification per individual, numbers of cloning per individual and numbers of singletons (cloned sequences appearing not more than twice in 42 to 60 clones each) are given.

Sample ID	n PCR	n Cloning	n Singletons
GM030005	3	1	3
WI920001	3	1	2
WI920009	3	3	1
WI920015	1	1	6


Consensus
MenoGM-DQBexon3_1
MenoGM-DQBexon3_2
MenoGM-DQBexon3_3
MenoGM-DQBexon3_4
Cow-BoLa-DQB1-D37954
Human-DQB1-L40179

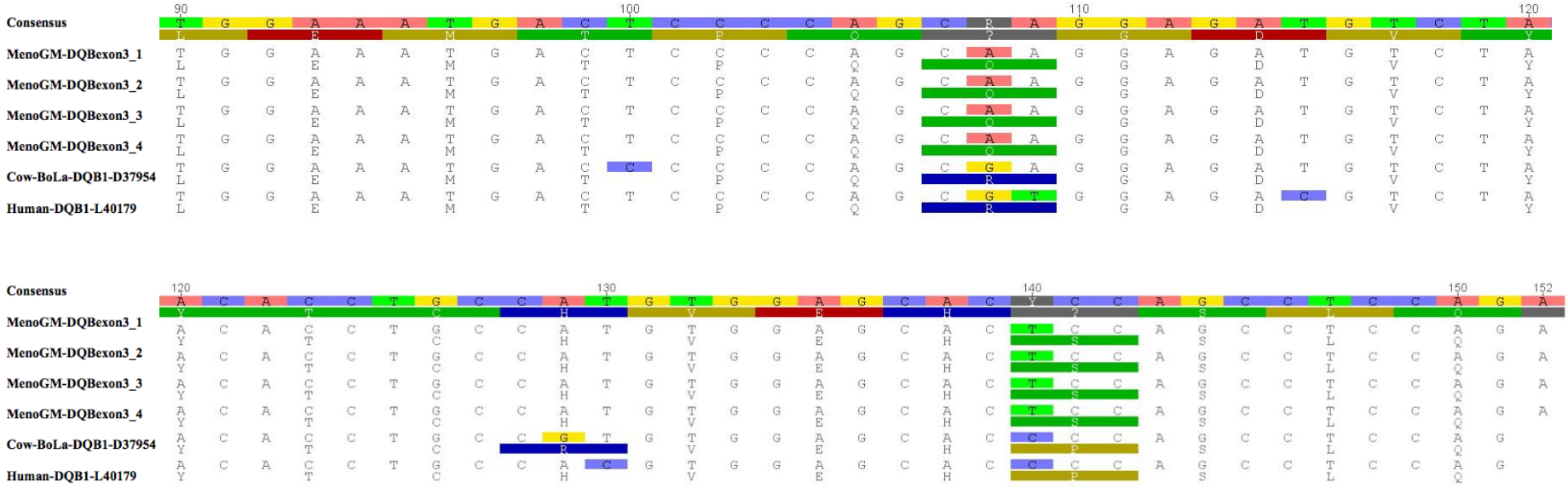


Figure S2: Alignment of four of the monomorphic exon 3 sequences in humpback whales summer-feeding in the Gulf of Maine (MenoGM-DQBexon3_1 to _4). The outgroups taken for this alignment are a HLA-DQB exon 3 sequence of the cow (Cow-BoLaDQB1-D37954) and one of human (Human-DQB1-L40179), downloaded from GenBank accession numbers D37954 and L40179, respectively. For the consensus sequence as well as for all individuals the nucleotide and corresponding amino acid alignment is given. Position 1 in the alignment corresponds to position 3 of human and cow sequences.

Table S3: Multi-SNP genotypes for all male humpback whales sampled in the West Indies. All four observed alleles are given for 16 polymorphic positions, used in this study. Identical genotypes are shaded with the same color. Font colors red and blue, as well as light and dark blue shadings are genotypes that also occured in the Gulf of Maine whales. The rest are multi-SNP genotypes only occurring in one individual overall. The numbering of the 16 positions corresponds to the alignment in figure 5.

SampleID	P38	P39	P40	P41	P44	P45	P46	P50	P64	P72	P131	P132	P161	P174	P183	P185
WI920425	CCCT	TTTT	CCCC	GGGG	AAAA	CCCC	CCCC	CTTT	TTTT	AAAA	GGGT	AAAC	AAAC	CCCC	AAAA	CCCC
WI920408	CCCC	TTTT	CCCC	GGGG	AAAA	CCCC	CCCC	TTTT	TTTT	AAAA	GGGG	AAAA	AAAA	CCCC	AAAA	CCCC
WI920009	CCGT	CTTT	ACCC	GGGG	AAAG	ACCC	CCCG	CTTT	GTTT	AAAT	GGGT	AAAC	AACC	CCCC	AAAC	CCCG
WI920352	CCCC	TTTT	AACC	GGGG	AAGG	AACC	CCGG	GGTT	GGGT	ATTT	GGTT	AACC	ACCC	CCCG	AAAA	CCGG
WI920004	CCCC	TTTT	AACC	GGTT	AAGG	AACC	CCGG	TTTT	GGGG	AATT	GGGG	AAAA	AACC	CCGG	AACC	CCGG
WI920353	CCCC	TTTT	AAAC	GGTT	AAGG	AACC	CCGG	TTTT	GGTT	AATT	GGGG	AAAA	AACC	CCGG	AACC	CCGG
WI920410	CCCC	TTTT	AACC	GGGG	AAGG	ACCT	CCCG	CGTT	GGTT	AATT	AGGT	AAAC	AACC	CCGG	AAAC	CCGG
WI920007	CCCC	TTTT	AACC	GGGG	AAGG	ACCT	CCCG	CTTT	GGTT	AATT	GGGT	AAAC	AACC	CCCG	AAAC	CCGG
WI920015	CCCC	TTTT	AACC	GGGT	AAGG	ACCT	CCCG	CGTT	GGGG	ATTT	GGTT	AACC	ACCC	CCGG	AAAC	CGGG
WI920426	CCCC	TTTT	AACC	GGGG	AAGG	CCTT	CCCC	CCTT	GGTT	AAAA	GGTT	AACC	AACC	CCGG	AAAA	CCGG
WI920433	CCCC	TTTT	AACC	GGGG	AAGG	CCTT	CCCC	CCTT	GGTT	AATT	GGGG	AAAA	AACC	CCGG	AAAA	CCGG
WI920370	CCCC	TTTT	AAAC	GGGT	AGGG	AAAC	CGGG	GTTT	GGGT	ATTT	GGGG	AAAA	ACCC	CGGG	AAAC	CGGG
WI920437	CCCC	TTTT	AAAC	GGGT	AGGG	ACTT	CCCG	CCTT	GGGT	AATT	GGTT	AACC	ACCC	CCGG	AAAC	CGGG
WI920027	CCCC	TTTT	AAAA	GTTT	GGGG	AAAA	GGGG	TTTG	GGGG	TTTT	AGGG	AAAT	CCCC	GGGG	ACCC	GGGG
WI920415	CCCC	TTTT	AAAC	GGGT	AGGG	AACT	CCGG	CGTT	GGGT	AATT	AGGT	AACT	ACCC	CGGG	AAAC	CGGG
WI920023	CCCC	TTTT	AAAC	GGGT	AGGG	AACT	CCGG	CGTT	GGGT	AATT	GGTT	AACC	ACCC	CCGG	AAAC	CGGG
WI920360	CCCC	TTTT	AAAC	GGTT	AGGG	AACT	CCGG	CTTT	GGGG	ATTT	GGTT	AACC	CCCC	CCGG	AAAC	CGGG
WI920017	CCCC	TTTT	AAAC	GGTT	AGGG	AACT	CCGG	CTTT	GGGT	AATT	GGGT	AAAC	ACCC	CCGG	AACC	CCGG
WI920028	CCCC	TTTT	AAAC	GGGT	AGGG	ACTT	CCCG	CCTT	GGGT	AATT	GGGT	AAAC	ACCC	CCGG	AAAC	CGGG
WI920361	CCCC	TTTT	AAAA	GGTT	GGGG	AAAT	CGGG	CGTT	GGGG	ATTT	GGGG	AAAA	CCCC	GGGG	AACC	GGGG
WI920367	CCGG	CCTT	AAAA	GGTT	GGGG	AAAT	CGGG	CTTT	GGGG	ATTT	GGGT	AAAC	CCCC	CCGG	ACCC	GGGG
WI920366	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGGG	AAAA	CCCC	GGGG	AACC	GGGG
WI920365	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CGTT	GGGG	ATTT	GGGG	AAAA	CCCC	GGGG	AACC	GGGG
WI920362	CCCC	TTTT	AAAC	GGTT	AGGG	AAAC	CGGG	GTTT	GGGT	ATTT	GGGG	AAAA	ACCC	CGGG	AAAC	CGGG
WI920441	CCCC	TTTT	AAAC	GGTT	AGGG	AAAC	CGGG	GTTT	GGGT	ATTT	GGGG	AAAA	ACCC	CGGG	AAAC	CGGG
WI920022	CCCC	TTTT	AAAC	GGGT	AGGG	ACTT	CCCG	CCTT	GGGT	AATT	GGTT	AACC	ACCC	CCGG	AAAC	CGGG
WI920356	CCCC	TTTT	AAAC	GGGT	AGGG	ACTT	CCCG	CCTT	GGGT	AATT	GGTT	AACC	ACCC	CCGG	AAAC	CGGG
WI920355	CCGG	CCTT	AAAA	GGTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	CCGG	CCCC	GGGG
WI920423	CCGG	CCTT	AAAA	GGTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	CCGG	CCCC	GGGG

WI920021	CCCC	TTTT	AAAA	GTTT	GGGG	AAAT	CGGG	CTTT	GGGG	ATTT	GGGG	AAAA	CCCC	GGGG	ACCC	GGGG
WI920359	CCCC	TTTT	AAAA	GTTT	GGGG	AAAT	CGGG	CTTT	GGGG	ATTT	GGGG	AAAA	CCCC	GGGG	ACCC	GGGG
WI920024	CCCC	TTTT	AAAA	TTTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	GGGG	CCCC	GGGG
WI920369	CCCC	TTTT	AAAA	TTTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	GGGG	CCCC	GGGG
WI920416	CCCC	TTTT	AAAA	TTTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	GGGG	CCCC	GGGG
WI920427	CCCC	TTTT	AAAA	TTTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	GGGG	CCCC	GGGG
WI920428	CCCC	TTTT	AAAA	TTTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	GGGG	CCCC	GGGG
WI920431	CCCC	TTTT	AAAA	TTTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	GGGG	CCCC	GGGG
WI920001	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGTT	AACC	CCCC	GGGG	AACC	GGGG
WI920010	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGTT	AACC	CCCC	GGGG	AACC	GGGG
WI920349	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGTT	AACC	CCCC	GGGG	AACC	GGGG
WI920363	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGTT	AACC	CCCC	GGGG	AACC	GGGG
WI920417	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGTT	AACC	CCCC	GGGG	AACC	GGGG
WI920430	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGTT	AACC	CCCC	GGGG	AACC	GGGG

Table S4: Multi-SNP genotypes for all mothers and calves of humpback whales sampled in the Gulf of Maine. All four observed alleles are given for 16 polymorphic positions used in this study. Identical genotypes are shaded with the same color. Font colors red and blue, as well as light and dark blue shadings are genotypes that also occured in the Gulf of Maine whales. Purple and grey shading should help grouping the individuals into groups of identical genotypes. The rest are multi-SNP genotypes only occurring in one individual overall. Sample IDs in italics depict identical mother-calf pairs. The numbering of the 16 positions corresponds to the alignment in figure 5.

SampleID	P38	P39	P40	P41	P44	P45	P46	P50	P64	P72	P131	P132	P161	P174	P183	P185
GM030005	CCCC	TTTT	ACCC	GGGG	AAAG	ACCC	CCCG	GTTT	GTTT	AAAT	GGGG	AAAA	AAAC	CCCG	AAAA	CCCG
GM930026	CCCC	TTTT	AACC	GGGG	AAGG	AACC	CCGG	GGTT	GGTT	AATT	GGGG	AAAA	AACC	CCCG	AAAA	CCGG
GM030217	CCCC	TTTT	AACC	GGGG	AAGG	AACC	CCGG	GGTT	GGTT	AATT	GGGG	AAAA	AACC	CCGG	AAAA	CCGG
GM940040	CCCC	TTTT	AACC	GGTT	AAGG	AACC	CCGG	GGTT	GGGG	TTTT	GGTT	AACC	CCCC	CCGG	AACC	CCGG
GM930028	CCGG	CCTT	AAAC	GGGG	AAGG	AACC	CCGG	GGTT	GGGT	AATT	GGGG	AAAA	ACCC	CCGG	AACC	CCGG
GM990125	CCCC	TTTT	AAAC	GGTT	AAGG	AACC	CCGG	GGTT	GGTT	AATT	GGGG	AAAA	AACC	CCGG	AACC	CCGG
GM970025	CCCC	TTTT	AACC	GGTT	AAGG	AACC	CCGG	TTTT	GGGT	ATTT	GGGG	AAAA	ACCC	CCGG	AACC	CCGG
GM980051	CCGG	CCTT	AACC	GGGG	AAGG	AACC	CCGG	TTTT	GGTT	AAAA	GGGG	AAAA	AACC	CCCC	AACC	CCGG
GM970008	CCCC	TTTT	AACC	GGGG	AAGG	ACCT	CCCG	CGTT	GGTT	AATT	GGGG	AAAA	AACC	CCGG	AAAA	CCGG
GM970009	CCCC	TTTT	AACC	GGGG	AAGG	CCTT	CCCC	CCTT	GGTT	AAAA	GGGG	AAAA	AACC	CCGG	AAAA	CCGG
GM010056	CCCC	TTTT	AACC	GGGG	AAGG	CCTT	CCCC	CCTT	GGTT	AATT	GGTT	AACC	AACC	CCGG	AAAA	CCGG
GM990026	CCGG	CCTT	AAAA	GGGG	AGGG	AAAC	CGGG	GGTT	GGGT	ATTT	GGGT	AAAC	ACCC	CCGG	AAAC	CGGG
GM980029	CCGG	CCTT	AAAA	GGTT	AGGG	AAAC	CGGG	CTTT	GGGT	ATTT	GGGT	AAAC	CCCC	CCGG	ACCC	CGGG
GM920089	CCCC	TTTT	AAAA	GGGT	AGGG	AACT	CCGG	CGTT	GGGT	AATT	GGGT	AAAC	ACCC	CGGG	AAAC	CGGG
GM050090	CCCC	TTTT	AAAC	GGTT	AGGG	AACT	CCGG	CTTT	GGGG	ATTT	GGTT	AACC	ACCC	CCGG	AACC	CCGG
GM930109	CCCC	TTTT	AAAC	GGGG	AGGG	ACTT	CCCG	CGTT	GGTT	AATT	GGTT	AACC	AACC	CCGG	AAAA	CCGG
GM990065	CCCC	TTTT	AAAC	GGGT	AGGG	ACTT	CCCG	CCTT	GGGT	AAAT	GGTT	AACC	ACCC	CGGG	AAAC	CGGG
GM920110	CCCC	TTTT	AAAC	GGGT	AGGG	ACTT	CCCG	CCTT	GGGT	AATT	GGGT	AAAC	ACCC	CCGG	AAAC	CGGG
GM980007	CCGG	CTTT	AAAC	GGGG	AGGG	ACTT	CCCG	CCTT	GGGT	AATT	GGTT	AACC	ACCC	CCGG	AAAC	CGGG
GM940039	CCCC	TTTT	AAAC	GGGT	AGGG	ACTT	CCCG	CGTT	GGGG	ATTT	GGTT	AACC	CCCC	CCGG	AAAC	CGGG
GM990027	CCCC	CTTT	AAAA	GGTT	GGGG	AAAA	GGGG	GTTT	GGGG	TTTT	GGGG	AAAA	CCCC	CGGG	AACC	GGGG
GM920222	CCGG	CCTT	AAAA	GGTT	GGGG	AAAA	GGGG	TTTT	GGGG	AATT	GGGG	AAAA	AACC	CCGG	CCCC	GGGG
GM970026	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGGT	AAAC	CCCC	GGGG	AACC	GGGG
GM930029	CCGG	CCTT	AAAA	GGGT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGTT	AACC	CCCC	CCGG	AACC	GGGG
GM990104	CCGG	CCTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGGG	AAAA	CCCC	CCGG	AACC	GGGG
GM990105	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGGG	AAAA	CCCC	GGGG	AACC	GGGG
GM980049	CCCC	TTTT	AACC	GGTT	AAGG	AACC	CCGG	TTTT	GGTT	AATT	GGGG	AAAA	AACC	CCGG	AACC	CCGG
GM050008	CCCC	TTTT	AACC	GGTT	AAGG	AACC	CCGG	TTTT	GGTT	AATT	GGGG	AAAA	AACC	CCGG	AACC	CCGG

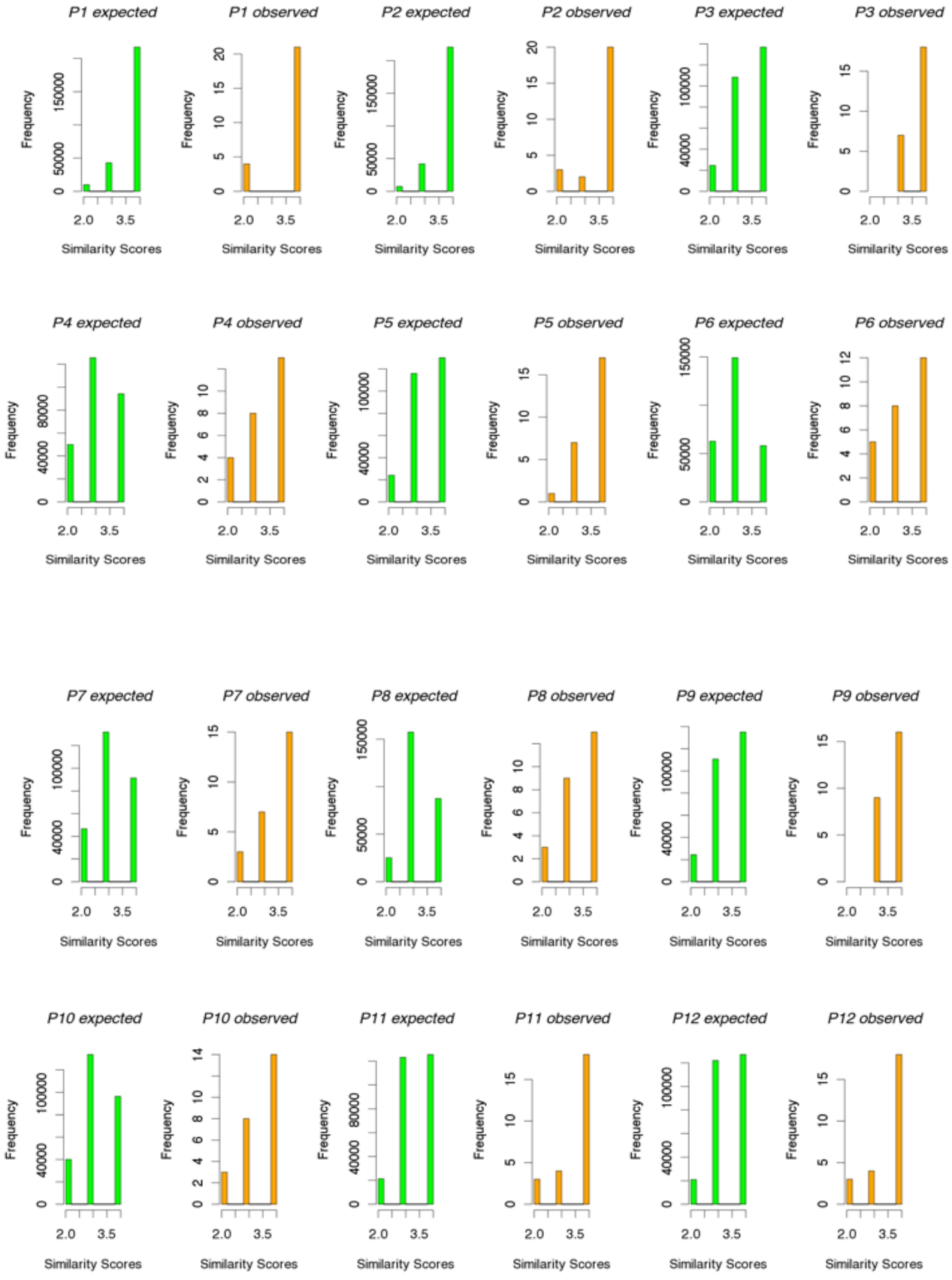
GM050009	CCCC	TTTT	AACC	GGTT	AAGG	AACC	CCGG	TTTT	GGTT	AATT	GGGG	AAAA	AACC	CCGG	AACC	CCGG
GM920224	CCCC	TTTT	AACC	GGGG	AAGG	CCTT	CCCC	CCTT	GGTT	AAAA	GGTT	AACC	AACC	CCGG	AAAA	CCGG
GM970041	CCCC	TTTT	AACC	GGGG	AAGG	CCTT	CCCC	CCTT	GGTT	AAAA	GGTT	AACC	AACC	CCGG	AAAA	CCGG
GM920224	CCCC	TTTT	AACC	GGGG	AAGG	CCTT	CCCC	CCTT	GGTT	AAAA	GGTT	AACC	AACC	CCGG	AAAA	CCGG
GM960006	CCCC	TTTT	AAAC	GGGT	AGGG	AAAC	CGGG	GTTT	GGGT	ATTT	GGGG	AAAA	ACCC	CCGG	AAAC	CGGG
GM960028	CCCC	TTTT	AAAC	GGGT	AGGG	AAAC	CGGG	GTTT	GGGT	ATTT	GGGG	AAAA	ACCC	CCGG	AAAC	CGGG
GM990111	CCCC	TTTT	AAAC	GGGT	AGGG	AAAC	CGGG	GTTT	GGGT	ATTT	GGGG	AAAA	ACCC	CCGG	AAAC	CGGG
GM940030	CCCC	TTTT	AAAC	GGGT	AGGG	AAAC	CGGG	GTTT	GGGT	ATTT	GGGG	AAAA	ACCC	CGGG	AAAC	CGGG
GM940031	CCCC	TTTT	AAAC	GGGT	AGGG	AAAC	CGGG	GTTT	GGGT	ATTT	GGGG	AAAA	ACCC	CGGG	AAAC	CGGG
GM930040	CCCC	TTTT	AAAA	GGGT	AGGG	AACT	CCGG	CGTT	GGGT	AATT	GGTT	AACC	ACCC	CCGG	AAAC	CGGG
GM920207	CCCC	TTTT	AAAA	GGGT	AGGG	AACT	CCGG	CGTT	GGGT	AATT	GGTT	AACC	ACCC	CCGG	AAAC	CGGG
GM930040	CCCC	TTTT	AAAA	GGGT	AGGG	AACT	CCGG	CGTT	GGGT	AATT	GGTT	AACC	ACCC	CCGG	AAAC	CGGG
GM050099	CCCC	TTTT	AAAC	GGGG	AGGG	ACTT	CCCG	CGTT	GGGT	AATT	GGTT	AACC	ACCC	CCGG	AAAC	CGGG
GM050099	CCCC	TTTT	AAAC	GGGG	AGGG	ACTT	CCCG	CGTT	GGGT	AATT	GGTT	AACC	ACCC	CCGG	AAAC	CGGG
GM970052	CCCC	TTTT	AAAA	TTTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	GGGG	ACCC	GGGG
GM920096	CCCC	TTTT	AAAA	TTTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	GGGG	ACCC	GGGG
GM920033	CCCC	TTTT	AAAA	GTTT	GGGG	AAAT	CGGG	CTTT	GGGG	ATTT	GGGT	AAAC	CCCC	GGGG	AAAC	GGGG
GM920034	CCCC	TTTT	AAAA	GTTT	GGGG	AAAT	CGGG	CTTT	GGGG	ATTT	GGGT	AAAC	CCCC	GGGG	AAAC	GGGG
GM960023	CCCC	TTTT	AAAA	GTTT	GGGG	AAAT	CGGG	CTTT	GGGG	ATTT	GGGT	AAAC	CCCC	GGGG	ACCC	GGGG
GM980001	CCCC	TTTT	AAAA	GTTT	GGGG	AAAT	CGGG	CTTT	GGGG	ATTT	GGGT	AAAC	CCCC	GGGG	ACCC	GGGG
GM920031	CCCC	TTTT	AAAA	GTTT	GGGG	AAAT	CGGG	CTTT	GGGG	ATTT	GGGT	AAAC	CCCC	GGGG	ACCC	GGGG
GM930035	CCCC	TTTT	AAAA	TTTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	GGGG	CCCC	GGGG
GM920202	CCCC	TTTT	AAAA	TTTT	GGGG	AAAA	GGGG	TTTT	GGGG	TTTT	GGGG	AAAA	CCCC	GGGG	CCCC	GGGG
GM960022	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGTT	AACC	CCCC	GGGG	AACC	GGGG
GM920092	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGTT	AACC	CCCC	GGGG	AACC	GGGG
GM990066	CCCC	TTTT	AAAA	GGTT	GGGG	AATT	CCGG	CCTT	GGGG	AATT	GGTT	AACC	CCCC	GGGG	AACC	GGGG

Next two pages:

Table S5: 27 mother-calf pairs used in this study. Sample IDs, status and pair number for the 27 mother-calf pairs are given. P44, P45, P50, P72, P131, P161 and P183 are the seven non-synonymous sites used for the simulation "nonsyn". "A" depicts the minimum number of alleles per individual. The simulations are "all" = 16 polymorphic positions, "nonsyn" = seven non-synonymous positons out of the 16 and "usat" = five neutral microsatellites. Trends are marked with different shading colors, red = different, green = similar. Shading of the sample IDs and genotypes correspond to table S3 and S4 supplementary information.

SampleID	Status	Pair	P44	P45	P50	P72	P131	P161	P183	A	all	nonsyn	Usat
GM960022	calf	1	GGGG	AATT	CCTT	AATT	GGTT	CCCC	AACC	2	sign. Different	sign. Different	Random
GM920222	mother	1	GGGG	AAAA	TTTT	AATT	GGGG	AACC	CCCC	2	sign. Different	sign. Different	Random
GM920092	calf	2	GGGG	AATT	CCTT	AATT	GGTT	CCCC	AACC	2	random	random	random/different
GM930040	mother	2	AGGG	AACT	CGTT	AATT	GGTT	ACCC	AAAC	3	random	random	random/different
GM920207	*calf*	*3*	*AGGG*	*AACT*	*CGTT*	*AATT*	*GGTT*	*ACCC*	*AAAC*	*3*	*sign. Similar*	*sign. Similar*	*NA*
GM930040	*mother*	*3*	*AGGG*	*AACT*	*CGTT*	*AATT*	*GGTT*	*ACCC*	*AAAC*	*3*	*sign. Similar*	*sign. Similar*	*NA*
GM970026	calf	4	GGGG	AATT	CCTT	AATT	GGGT	CCCC	AACC	3	random	random	Random
GM960023	mother	4	GGGG	AAAT	CTTT	ATTT	GGGT	CCCC	ACCC	2	random	random	Random
GM920110	calf	5	AGGG	ACTT	CCTT	AATT	GGGT	ACCC	AAAC	3	random	random	Random
GM920224	mother	5	AAGG	CCTT	CCTT	AAAA	GGTT	AACC	AAAA	2	random	random	Random
GM970041	*calf*	*6*	*AAGG*	*CCTT*	*CCTT*	*AAAA*	*GGTT*	*AACC*	*AAAA*	*2*	*sign. Similar*	*sign. Similar*	*Random*
GM920224	*mother*	*6*	*AAGG*	*CCTT*	*CCTT*	*AAAA*	*GGTT*	*AACC*	*AAAA*	*2*	*sign. Similar*	*sign. Similar*	*Random*
GM920033	*calf*	*7*	*GGGG*	*AAAT*	*CTTT*	*ATTT*	*GGGT*	*CCCC*	*AAAC*	*2*	*sign. Similar*	*sign. Similar*	*Random*
GM920034	*mother*	*7*	*GGGG*	*AAAT*	*CTTT*	*ATTT*	*GGGT*	*CCCC*	*AAAC*	*2*	*sign. Similar*	*sign. Similar*	*Random*
GM970052	*calf*	*8*	*GGGG*	*AAAA*	*TTTT*	*TTTT*	*GGGG*	*CCCC*	*ACCC*	*2*	*similar*	*similar*	*Similar*
GM920096	*mother*	*8*	*GGGG*	*AAAA*	*TTTT*	*TTTT*	*GGGG*	*CCCC*	*ACCC*	*2*	*similar*	*similar*	*similar*
GM930028	calf	9	AAGG	AACC	GGTT	AATT	GGGG	ACCC	AACC	3	different	different	similar
GM930029	mother	9	GGGG	AATT	CCTT	AATT	GGTT	CCCC	AACC	3	different	different	similar
GM940030	*calf*	*10*	*AGGG*	*AAAC*	*GTTT*	*ATTT*	*GGGG*	*ACCC*	*AAAC*	*2*	*sign. Similar*	*sign. Similar*	*random*
GM940031	*mother*	*10*	*AGGG*	*AAAC*	*GTTT*	*ATTT*	*GGGG*	*ACCC*	*AAAC*	*2*	*sign. Similar*	*sign. Similar*	*random*
GM960006	*calf*	*11*	*AGGG*	*AAAC*	*GTTT*	*ATTT*	*GGGG*	*ACCC*	*AAAC*	*3*	*sign. Similar*	*sign. Similar*	*random*
GM960028	*mother*	*11*	*AGGG*	*AAAC*	*GTTT*	*ATTT*	*GGGG*	*ACCC*	*AAAC*	*3*	*sign. Similar*	*sign. Similar*	*random*
GM980029	calf	12	AGGG	AAAC	CTTT	ATTT	GGGT	CCCC	ACCC	3	random	random	random
GM980007	mother	12	AGGG	ACTT	CCTT	AATT	GGTT	ACCC	AAAC	3	random	random	random
GM970008	calf	13	AAGG	ACCT	CGTT	AATT	GGGG	AACC	AAAA	3	sign. Similar	random	random
GM970009	mother	13	AAGG	CCTT	CCTT	AAAA	GGGG	AACC	AAAA	2	sign. Similar	random	random
GM970025	calf	14	AAGG	AACC	TTTT	ATTT	GGGG	ACCC	AACC	3	random	random	random
GM930026	mother	14	AAGG	AACC	GGTT	AATT	GGGG	AACC	AAAA	2	random	random	random
GM980001	*calf*	*15*	*GGGG*	*AAAT*	*CTTT*	*ATTT*	*GGGT*	*CCCC*	*ACCC*	*2*	*sign. Similar*	*sign. Similar*	*similar*

GM920031	*mother*	*15*	*GGGG*	*AAAT*	*CTTT*	*ATTT*	*GGGT*	*CCCC*	*ACCC*	*2*	*sign. Similar*	*sign. Similar*	*similar*
GM990104	calf	16	GGGG	AATT	CCTT	AATT	GGGG	CCCC	AACC	2	random	sign. Similar	random
GM990105	mother	16	GGGG	AATT	CCTT	AATT	GGGG	CCCC	AACC	2	random	sign. Similar	random
GM930035	*calf*	*17*	*GGGG*	*AAAA*	*TTTT*	*TTTT*	*GGGG*	*CCCC*	*CCCC*	*1*	*similar*	*similar*	*random*
GM920202	*mother*	*17*	*GGGG*	*AAAA*	*TTTT*	*TTTT*	*GGGG*	*CCCC*	*CCCC*	*1*	*similar*	*similar*	*random*
GM990125	calf	18	AAGG	AACC	GGTT	AATT	GGGG	AACC	AACC	3	random	random	different
GM920089	mother	18	AGGG	AACT	CGTT	AATT	GGGT	ACCC	AAAC	3	random	random	different
GM980049	calf	19	AAGG	AACC	TTTT	AATT	GGGG	AACC	AACC	2	random	sign. Similar	random
GM980051	mother	19	AAGG	AACC	TTTT	AAAA	GGGG	AACC	AACC	2	random	sign. Similar	random
GM990111	calf	20	AGGG	AAAC	GTTT	ATTT	GGGG	ACCC	AAAC	2	random	random	random
GM930109	mother	20	AGGG	ACTT	CGTT	AATT	GGTT	AACC	AAAA	3	random	random	random
GM940040	calf	21	AAGG	AACC	GGTT	TTTT	GGTT	CCCC	AACC	2	random	random	sign. Similar
GM940039	mother	21	AGGG	ACTT	CGTT	ATTT	GGTT	CCCC	AAAC	3	random	random	sign. Similar
GM050090	calf	22	AGGG	AACT	CTTT	ATTT	GGTT	ACCC	AACC	3	random	random	random
GM050099	mother	22	AGGG	ACTT	CGTT	AATT	GGTT	ACCC	AAAC	3	random	random	random
GM010056	calf	23	AAGG	CCTT	CCTT	AATT	GGTT	AACC	AAAA	2	random	random	random
GM050099	mother	23	AGGG	ACTT	CGTT	AATT	GGTT	ACCC	AAAC	3	random	random	random
GM030217	calf	24	AAGG	AACC	GGTT	AATT	GGGG	AACC	AAAA	2	random	random	random
GM030005	mother	24	AAAG	ACCC	GTTT	AAAT	GGGG	AAAC	AAAA	3	random	random	random
GM050008	*calf*	*25*	*AAGG*	*AACC*	*TTTT*	*AATT*	*GGGG*	*AACC*	*AACC*	*2*	*sign. Similar*	*sign. Similar*	*random*
GM050009	*mother*	*25*	*AAGG*	*AACC*	*TTTT*	*AATT*	*GGGG*	*AACC*	*AACC*	*2*	*sign. Similar*	*sign. Similar*	*random*
GM990026	calf	26	AGGG	AAAC	GGTT	ATTT	GGGT	ACCC	AAAC	3	random	random	random
GM990027	mother	26	GGGG	AAAA	GTTT	TTTT	GGGG	CCCC	AACC	3	random	random	random
GM990065	calf	27	AGGG	ACTT	CCTT	AAAT	GGTT	ACCC	AAAC	3	random	random	random
GM990066	mother	27	GGGG	AATT	CCTT	AATT	GGTT	CCCC	AACC	3	random	random	random


P1 expected
P1 observed
P2 expected
P2 observed
P3 expected
P3 observed
P4 expected
P4 observed
P5 expected
P5 observed
P6 expected
P6 observed
P7 expected
P7 observed
P8 expected
P8 observed
P9 expected
P9 observed
P10 expected
P10 observed
P11 expected
P11 observed
P12 expected
P12 observed
Frequency
Similarity Scores
2.0
3.5

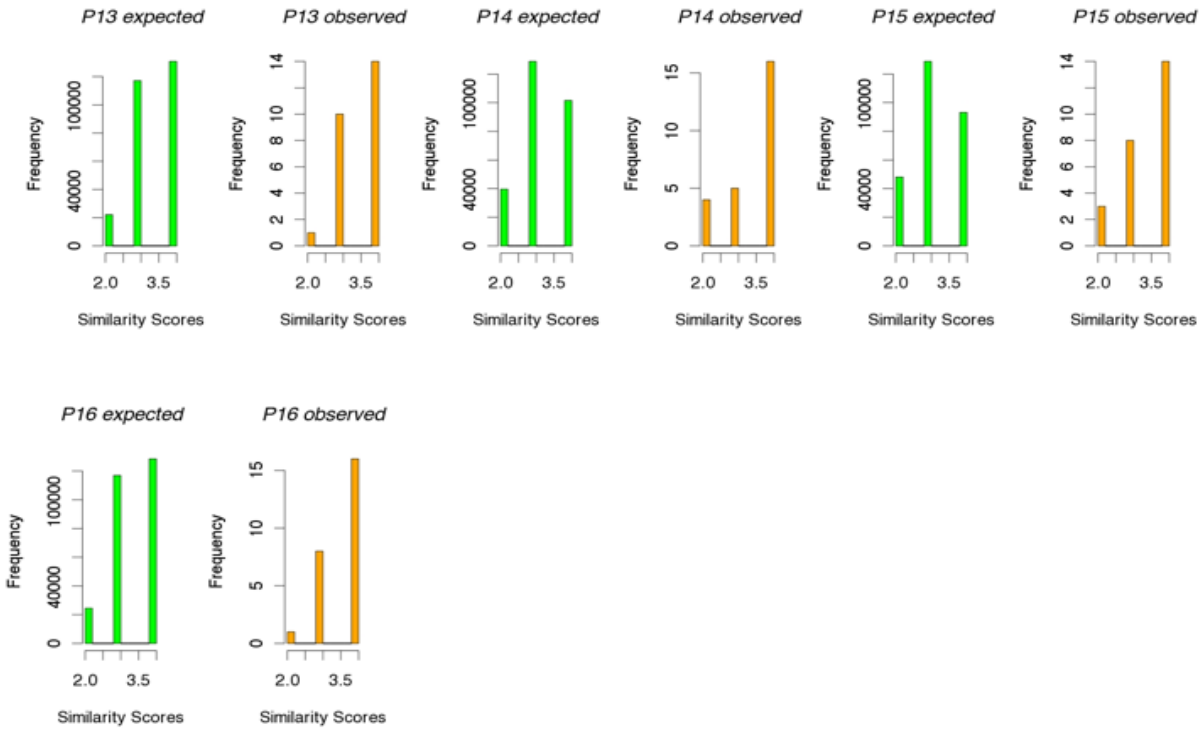


Figure S3: Comparison of the distribution of similarity scores for all 16 polymorphic positions used in this study. Expected distributions obtained with the simulation using the 27 mothers and 43 random males in this study are given in green bars and observed distributions in the 27 corresponding mother-calf pairs are given in orange bars. None of the comparisons were significantly different from each other.

Table S6: P-Values and their standard errors for the tests for deviations from Hardy-Weinberg equilibrium using the software GENEPOP [102, 103]. The tests were done for all five microsatellite loci each (GATA028, TAA031, GATA053, GATA098 and GATA417). The whales were subdivided into four different groupings: Males WI = 43 random samples of male whales sampled in the West Indies, Mothers GoM = 27 samples of female humpback whales in the Gulf of Maine having had a calf, Calves = 27 calves of the Mothers GoM and GoM all = Males Wi and Mothers GoM combined. Specifically, deviations from Hardy-Weinberg proportions were assessed using the Markov chain approach of Guo and Thompson = P-Value (1992 [104]), and the one-sided test of heterozygote deficiency was assessed by an estimate of F_{is} = P-one-sided [105]. Steps = the number of times the sample configuration changes in the MC run. P-values close to significance or significant were Bonferroni-corrected for multiple testing and the adjusted p-vales are given = P-adjusted.

Locus	P-Value	S.E.	P-adjusted	P-one-sided	S.E.	Steps
GATA028						
Males WI	0.111	0.015	1.000	0.102	0.011	4029
Mothers GoM	0.793	0.022	1.000	0.333	0.022	2773
Calves	0.469	0.020	1.000	0.888	0.014	6592
GoM all	0.368	0.028	1.000	0.102	0.012	7091
TAA031						
Males WI	0.945	0.009	1.000	0.394	0.025	7068
Mothers GoM	0.929	0.007	1.000	0.733	0.017	9424
Calves	0.920	0.007	1.000	0.702	0.018	8733
GoM all	0.929	0.012	1.000	0.610	0.027	8057
GATA053						
Males WI	0.357	0.016	1.000	0.531	0.022	13631
Mothers GoM	0.234	0.014	1.000	0.312	0.015	11940
Calves	0.816	0.012	1.000	0.869	0.011	9521
GoM all	0.045	0.007	0.224	0.181	0.014	22887
GATA098						
Males WI	0.282	0.021	1.000	0.861	0.012	11544
Mothers GoM	0.447	0.020	1.000	0.349	0.020	5854
Calves	0.644	0.032	1.000	0.363	0.028	2166
GoM all	0.385	0.029	1.000	0.665	0.028	5688
GATA417						
Males WI	0.861	0.021	1.000	0.904	0.018	3841
Mothers GoM	0.129	0.012	1.000	0.967	0.006	8689
Calves	0.435	0.022	1.000	0.326	0.023	8099
GoM all	0.505	0.029	1.000	0.970	0.008	6692

Next page:

Table S7: Test for linkage disequilibrium of all possible pairs of five microsatellite loci (GATA028, TAA031, GATA053, GATA098 and GATA417). The presence of genotypic linkage disequilibrium was tested using a Markov chain method as described by Raymond and Rousset (1995 [106]) using the software GENEPOP [102, 103]. At a first step all tests were performed with 10'000 dememorizations, 1000 batches and 5000 iterations. The whales were subdivided into four different groupings: Males WI = 43 random samples of male whales sampled in the West Indies, Mothers GoM = 27 samples of female humpback whales in the Gulf of Maine having had a calf, Calves = 27 calves of the Mothers GoM and GoM all = Males Wi and Mothers GoM combined. For p-values close to significance at the 95% level, when the standard error was large or the number of switches (the number of times the sample configuration changes in the MC run) was low, the number of batches was increased to 10'000. Significance of p-values was assessed by applying the sequential Bonferroni correction to adjust for the effect of multiple testing [107] = P-adjusted. Steps = the number of times the sample configuration changes in the MC run.

Population	Locus 1	Locus 2	P-Value	S.E.	Steps	P-adjusted
GoM all	GATA028	TAA031	0.395	0.003	1235272	1.000
GoM all	GATA028	GATA053	0.851	0.002	1307908	1.000
GoM all	TAA031	GATA053	0.020	0.001	422013	0.201
GoM all	GATA028	GATA098	0.327	0.002	2164609	1.000
GoM all	TAA031	GATA098	0.005	0.000	924012	0.050
GoM all	GATA053	GATA098	0.058	0.001	998016	0.585
GoM all	GATA028	GATA417	0.862	0.002	1058445	1.000
GoM all	TAA031	GATA417	0.006	0.001	334993	0.058
GoM all	GATA053	GATA417	0.180	0.003	340359	1.000
GoM all	GATA098	GATA417	0.147	0.002	773965	1.000
Males WI	GATA028	TAA031	0.549	0.013	12217	1.000
Males WI	GATA028	GATA053	0.853	0.008	15877	1.000
Males WI	TAA031	GATA053	0.026	0.005	5714	1.000
Males WI	GATA028	GATA098	0.248	0.009	26598	1.000
Males WI	TAA031	GATA098	0.403	0.013	12146	1.000
Males WI	GATA053	GATA098	0.215	0.010	16535	1.000
Males WI	GATA028	GATA417	0.426	0.014	10456	1.000
Males WI	TAA031	GATA417	0.204	0.012	3406	1.000
Males WI	GATA053	GATA417	0.761	0.013	4789	1.000
Males WI	GATA098	GATA417	0.253	0.012	10247	1.000
Mothers GoM	GATA028	TAA031	0.563	0.012	14799	1.000
Mothers GoM	GATA028	GATA053	0.640	0.011	17014	1.000
Mothers GoM	TAA031	GATA053	0.066	0.007	8906	1.000
Mothers GoM	GATA028	GATA098	0.475	0.010	25317	1.000
Mothers GoM	TAA031	GATA098	0.002	0.001	15505	0.040
Mothers GoM	GATA053	GATA098	0.167	0.009	17868	1.000
Mothers GoM	GATA028	GATA417	0.784	0.009	17408	1.000
Mothers GoM	TAA031	GATA417	0.005	0.002	7941	0.106
Mothers GoM	GATA053	GATA417	0.135	0.009	9724	1.000
Mothers GoM	GATA098	GATA417	0.069	0.006	16994	1.000
Calves only	GATA028	TAA031	0.853	0.021	2372	1.000
Calves only	GATA028	GATA053	0.579	0.032	2179	1.000
Calves only	TAA031	GATA053	0.234	0.036	1032	1.000
Calves only	GATA028	GATA098	0.334	0.031	2703	1.000
Calves only	TAA031	GATA098	0.096	0.022	1491	1.000
Calves only	GATA053	GATA098	0.278	0.036	1375	1.000
Calves only	GATA028	GATA417	0.794	0.028	1814	1.000
Calves only	TAA031	GATA417	0.423	0.043	738	1.000
Calves only	GATA053	GATA417	1.000	0.000	792	1.000
Calves only	GATA098	GATA417	1.000	0.000	1109	1.000

Table S8: Allele frequency differences between the 27 samples of humpback whale mothers from the Gulf of Maine and the 47 samples of random males from the West Indies for all five microsatellite loci.

GATA028

Population	Alleles							
	148	152	155	156	159	179	183	191
Males WI	0.686	0.035	0.012	0.116	0.058	0.023	0.012	0.058
Mothers GoM	0.673	0.038	0.019	0.096	0.096	0.019	0.019	0.038

TAA031

Population	Alleles									
	102	105	106	108	109	111	114	118	880	960
Males WI	0.093	0.267	0.012	0.198	0	0.023	0.035	0.023	0.058	0.058
Mothers GoM	0.058	0.212	0	0.212	0.019	0.058	0	0	0.154	0.058

	970	990
Males WI	0.023	0.209
Mothers GoM	0.038	0.192

GATA053

Population	Alleles									
	176	180	184	188	192	196	200	204	208	212
Males WI	0	0.198	0.105	0.012	0.023	0.267	0.14	0.081	0.14	0.035
Mothers GoM	0.22	0.16	0	0.06	0.28	0.08	0.06	0.12	0.02	0

GATA098

Population	Alleles									
	107	109	111	115	117	119	123	125	127	131
Males WI	0.093	0	0.023	0.07	0	0.116	0.047	0	0.116	0.023
Mothers GoM	0.12	0.08	0	0	0.02	0.16	0.02	0.02	0.08	0

	910
Males WI	0.512
Mothers GoM	0.5

GATA417

Population	Alleles									
	195	199	203	206	207	210	211	214	218	222
Males WI	0.14	0.058	0.023	0.012	0.07	0.256	0.012	0.093	0.023	0.209
Mothers GoM	0.135	0.154	0.058	0	0.058	0.173	0.038	0.154	0.019	0.154

	226	230	234	281
Males WI	0.035	0.035	0.012	0.023
Mothers GoM	0.058	0	0	0

Table S9: Population differentiation of the samples from the Gulf of Maine and the samples from the West Indies used in this study. The pairwise F_{ST} values were calculated as one locus estimates following standard ANOVA as in Weir and Cockerham (1984) [105].

GATA028	TAA031	GATA053	GATA098	GATA417	overall
0.015	0.005	0.088	0.002	0.001	0.018
p = 0.003	p = 0.432	p <0.001	p = 0.090	p = 0.475	

APPENDIX

A Script to simulate the null-hypothesis of random mating

```
# convert the observed alleles to vectors
alleleToVector<-function(inputAllele){
   return(unlist(strsplit(inputAllele,"")))
}

# convert the vectors back to alleles
vectorToAllele<-function(inputVector){
   return(paste(inputVector,collapse=""))
}

# create a calf with a mother and a random male
createCalfAllele<-function(dadAllele, momAllele){

   dad<-alleleToVector(dadAllele)
   mom<-alleleToVector(momAllele)

   if (length(dad) != length(mom))
      return(FALSE) # something wrong with input...

   l<-length(dad)
   while (length(dad)>l/2){
      rdm<-ceiling(runif(1,0,length(dad)))
      dad<-dad[-rdm]
}

   while (length(mom)>l/2){
      rdm<-ceiling(runif(1,0,length(mom)))
      mom<-mom[-rdm]
}

   calf<-c(dad,mom)
   return(vectorToAllele(calf))
}

# calculate the similarity between mothers and their calves
lettermatch <- function(mother, calf) {
  tb <- merge(as.data.frame(table(strsplit(mother, ""))),
            as.data.frame(table(strsplit(calf, ""))), by="Var1")
  sum(apply(tb[-1], 1, min))
}


# read in the data
data_mom<-read.delim("females_in.txt", stringsAsFactors=FALSE)
data_mom<-data_mom[data_mom$Status=="mother",]

data_papi<-read.delim("males_in.txt", stringsAsFactors=FALSE)
```

```
# simulation

set.seed(100111)

noIterations<-10000
noAllels<-16
means01<-data.frame()
allCalfs<-list()
allData<-list()

for (iteration in 1:noIterations){
output_calf<-matrix(ncol=18, nrow=nrow(data_mom))

for (j in 1:nrow(data_mom)){
   momAllele<-data_mom[j,4:19]
   rdmDad<-ceiling(runif(1,0,length(data_papi[,1])))
   dadAllele<-data_papi[rdmDad,4:19]
   calf<-c()

   for (k in 1:16){
      calf[k]<-createCalfAllele(momAllele[,k],dadAllele[,k])
}

   calf<-c(data_mom$SampleID[j],data_papi$SampleID[rdmDad],calf)
   output_calf[j,]<-calf
}
      output_calf<-as.data.frame(output_calf)
      names(output_calf)<-c("MomID","DadID",names(data_mom[4:19]))
      #loop over all MomIDs that made calfs (should be all of them, but make it save)
      dataOfThisIteration<-data.frame()
      for (id in 1:nrow(output_calf)){
        #get correct Mom for each ID
        momToConsider<- data_mom[id,4:19]

        #get correct Calf for each ID
        calfToConsider<-output_calf[id,3:18]

      comparison<-c()
        for (i in 1:length(momToConsider)){
          comparison<-c(comparison,lettermatch(as.character(momToConsider[i]),
as.character(unlist(calfToConsider[i]))) )
}
       dataOfThisIteration<-rbind(dataOfThisIteration,comparison)
       names(dataOfThisIteration)<-names(data_mom[4:19])
}

allData<-c(allData,list(dataOfThisIteration))
allCalfs<-c(allCalfs,list(output_calf))
# save a temporary output file with the current data being produced
if (i %% 10) write.table(allData,"temporary_allData.txt")
}
```

```
# plot the density of the simulated mother-calf similarities
par(mfrow=c(1,1))

density.raw<-c()
for (i in 1:noIterations){
density.raw<-c(density.raw,rowMeans(allData[[i]]))}
plot(density(density.raw,to=4))

bxp.custom<-function(file2,at,color,...){
 b<-boxplot(file2, plot=F)
 b$stats[, 1] = quantile(file2, prob = c(0.05, 0.25, 0.5, 0.75, 0.95),na.rm=T)
 b$out<-file2[file2<b$stats[1,1] | file2>b$stats[5,1]]
 b$group<-rep(1,length(b$out))
bxp(b,add=T,horizontal=T,at=at,outline=F,col="red",col.sub="red",border=color,...)
}

plotDensitiesWithBoxPlot<-function(file1,...){
 d1<-density(file1,na.rm=T)

 ymax<-max(d1$y, to=1)
 ymin<--1*0.2*ymax

# first way to compare stuff
 plot(d1,xlab="Similarity Score",ylim=c(ymin,ymax),main="MHC female mate choice")
 bxp.custom(file1, at=ymin+0.15*ymax,  color="black", pars = list(boxwex
=0.08*ymax,col.sub="red",staplewex=0.8),lty=1)
}

# and save it (i.e. as a jpg-picture)
jpeg("analysis_allData.jpg",quality=100,pointsize=20)
plotDensitiesWithBoxPlot(density.raw)
abline(v=3.528, col = "blue", lwd=4)
dev.off()

# Spring 2010 LGS
```

Allele- and locus-specific PCR reactions (trial and error)

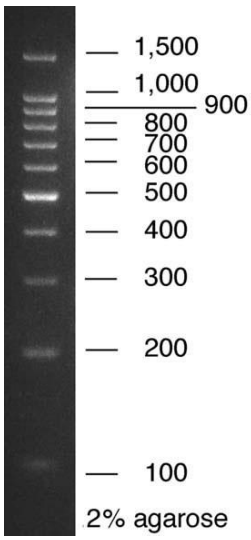
This is a summary of the allele- and locus-specific PCR reactions I would consider working with if I tried in a second attempt to separate the two HLA-DQB-like loci. All the following reactions contain a forward primer within HLA-DQB exon 2 and a reverse primer within HLA-DQB exon 3 from Gulf of Maine humpback sequences. The idea was to sequence allele- or locus-specific HLA-DQB intron 2 sequences that should have served to design reverse primers to eventually amplify the two duplicated loci separately together with the universal DQB exon 2 forward primer "DQB2". All forward primers below were designed using cloned alleles from the individuals GM030005 and WI920009. I used three different DNA ladders.

Table A1: PCR conditions used in the attempts to amplify single alleles in the region of HLA-DQB exon 2. Names of forward and reverse primers are given (for the exact primer sequences see tabe S1 supplementary information), as well as band sizes appearing in the amplification reactions for which PCR IDs are given. Also given are the annealing temperatures of the specific amplification reactions and the numbers of cycles used.

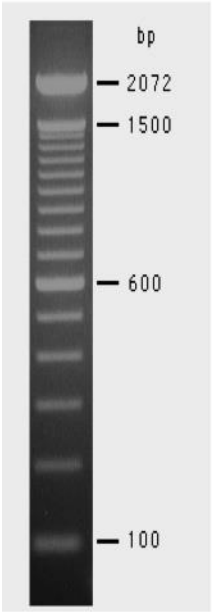
Forward Primer	Reverse Primer	Band size (bp)	PCR ID	Annealing (T)	Cycles (n)
DQBLC01_F (i)	DBQ_INTRON2R	500	4823/4764	58	35
DQBLC01_F (i)	DBQ_INTRON2R	800	4823/4764	58	35
DQBLC01_F (i)	DBQ_INTRON2R	900	4823/4764	58	35
DQBLC02_F (ii)	DBQEX3R2	600	4829/4812	65	35
DQBLC02_F (ii)	DQB_INTRON2_R3	700 + several	4779	60	35
AL1_716A_F (iii)	DBQEX3R2	900	4824/4808/4775	60	36
AL1_716B_F (iv)	DBQEX3R2	900	4809/4776/4825	60	37
DQBHM01 (v)	DBQ_INTRON2R	800	4820	59	33
AL1_403_F (vi)	DBQ_INTRON2R	700	4766	56	36
AL2_403_F (vii)	DQB_INTRON2_R3	600(+/-)	4806	61	35
AL2_403_F (vii)	DQB_INTRON2R	600 (500/700)	4828/4810	58	35

3 different DNA ladders:

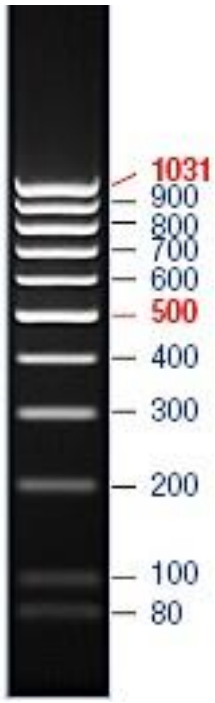
1) Promega's "DNA Ladder Molecular Weight Marker" a 100 base pair (bp) DNA ladder, catalogue number g2101
URL: http://www.promega.com/catalog/catalogproducts.aspx?categoryname=productleaf_179&ckt=1



2) Invitrogen's 100 bp DNA ladder, catalogue number 15628-050
URL:
http://tools.invitrogen.com/content/sfs/manuals/15628050.pdf

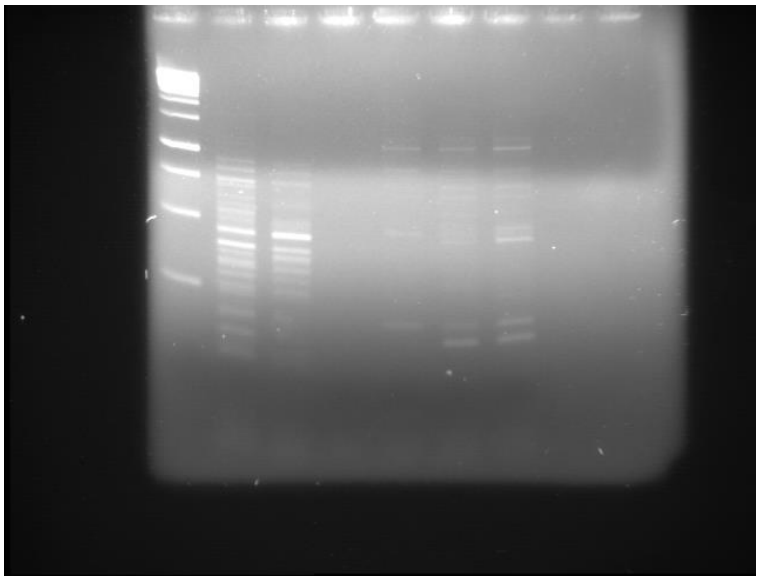


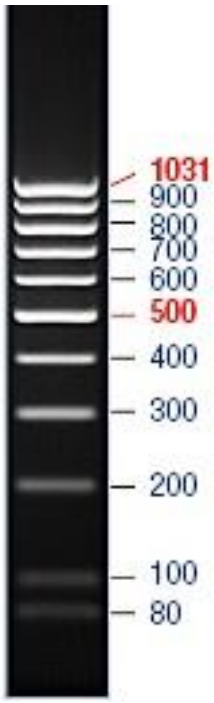
3) Fermentas' "MassRuler™ Low Range DNA Ladder, ready-to-use, 80-1031 bp"
URL:
http://www.fermentas.com/templates/files/tiny_mce/family_images/sm038_fam.jpg



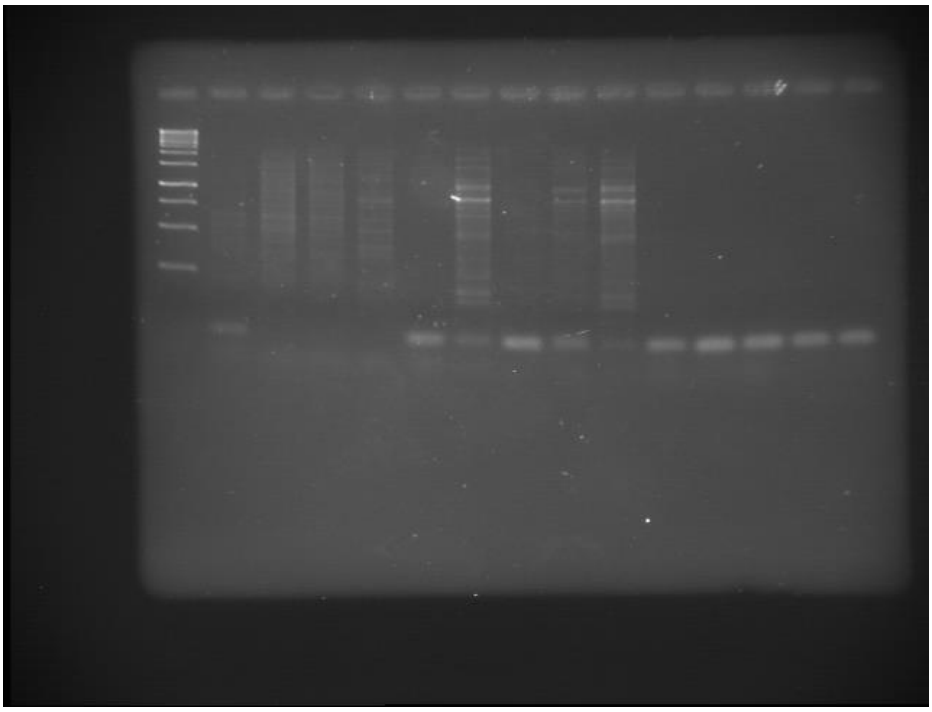
i. Locus-specific 1

Forward primer: DQBLC01_F,
Reverse primer: DBQ_INTRON2R,
PCR IDs 4823 and 4764





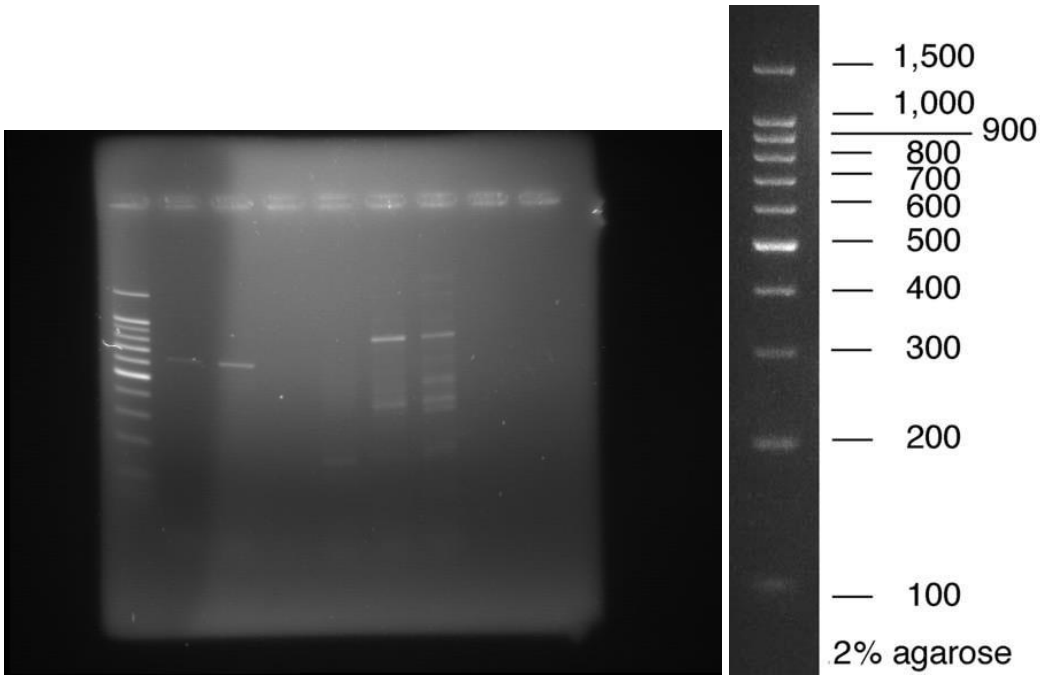
(PCR ID 4823: 5th, 6th and 7th lane, DNA ladder representing lane 1)



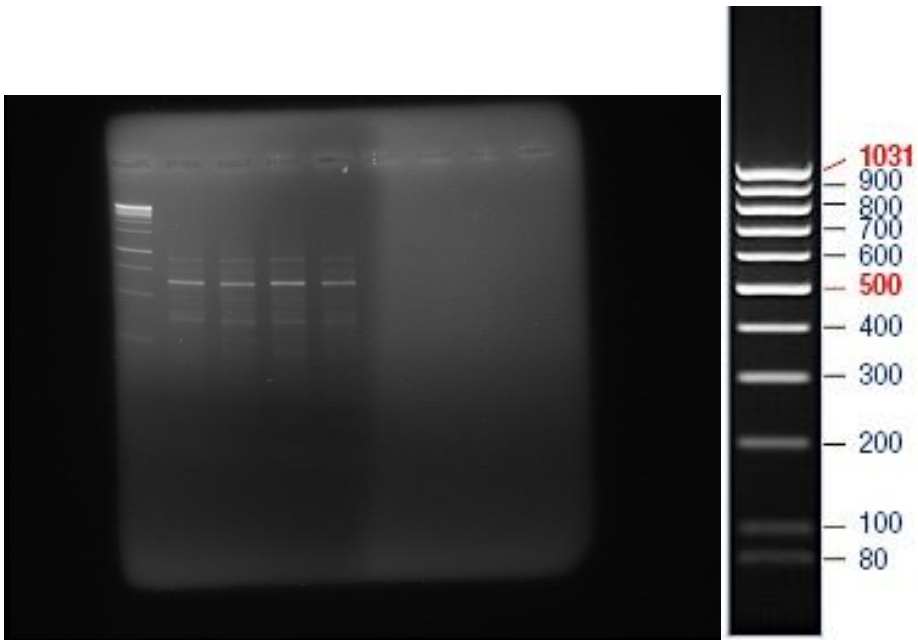
(PCR ID 4764)

ii. Locus-specific 2

Forward primer: DQBLC02_F,
Reverse primer: DBQEX3R2,
PCR IDs 4829 and 4812

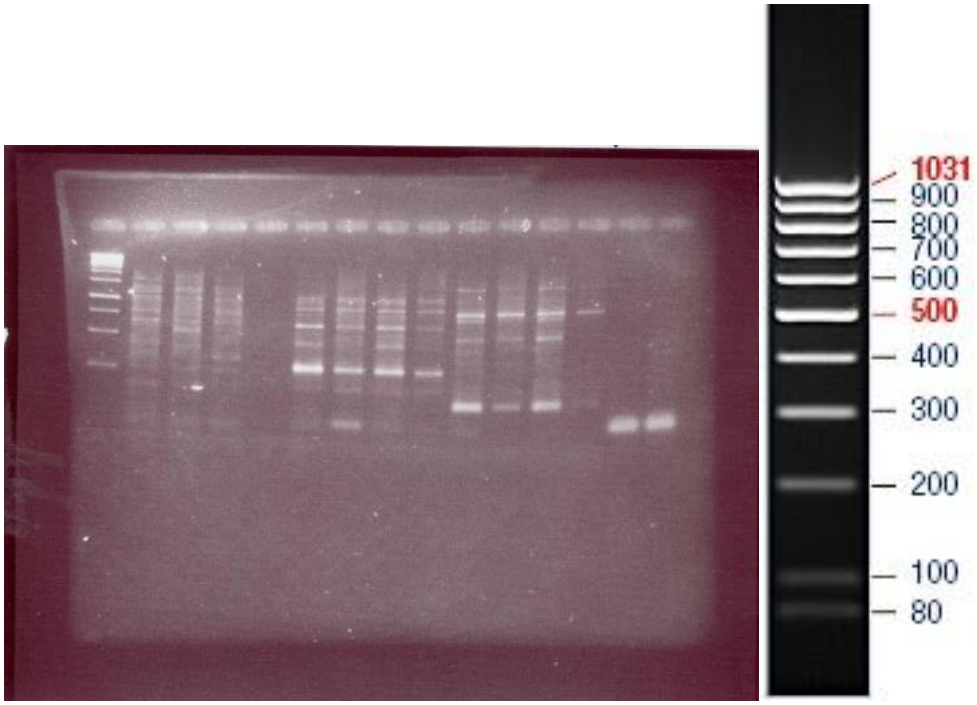


(PCR ID 4829: 2nd to 5th lane, DNA ladder representing lane 1)



(PCR ID 4812)

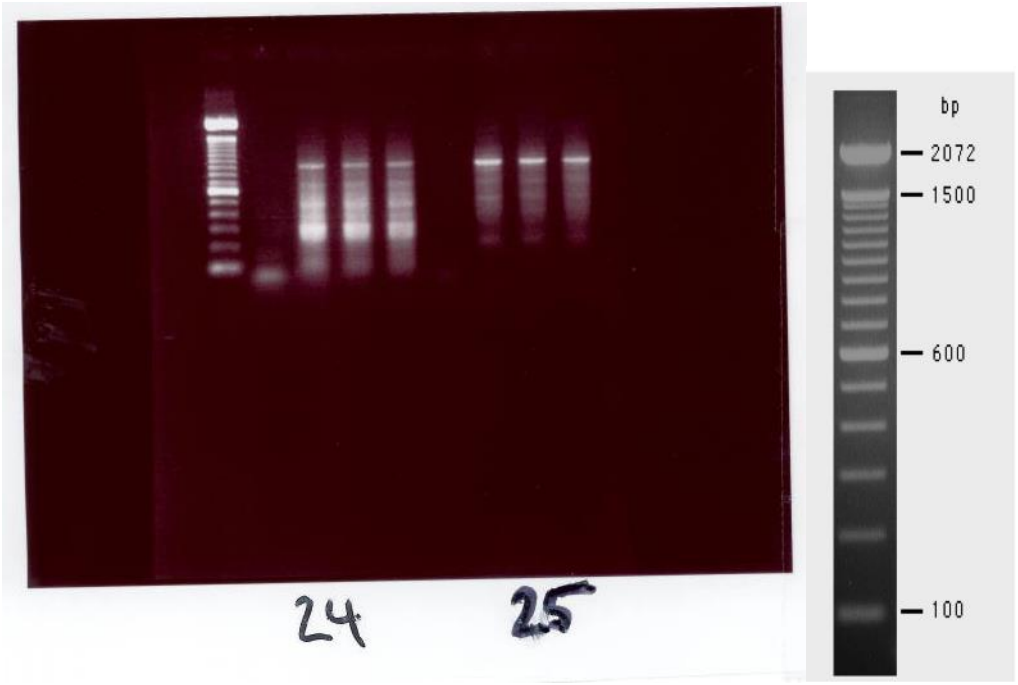
Forward primer: DQBLC02_F,
Reverse primer: DQB_INTRON2_R3,
PCR ID 4779



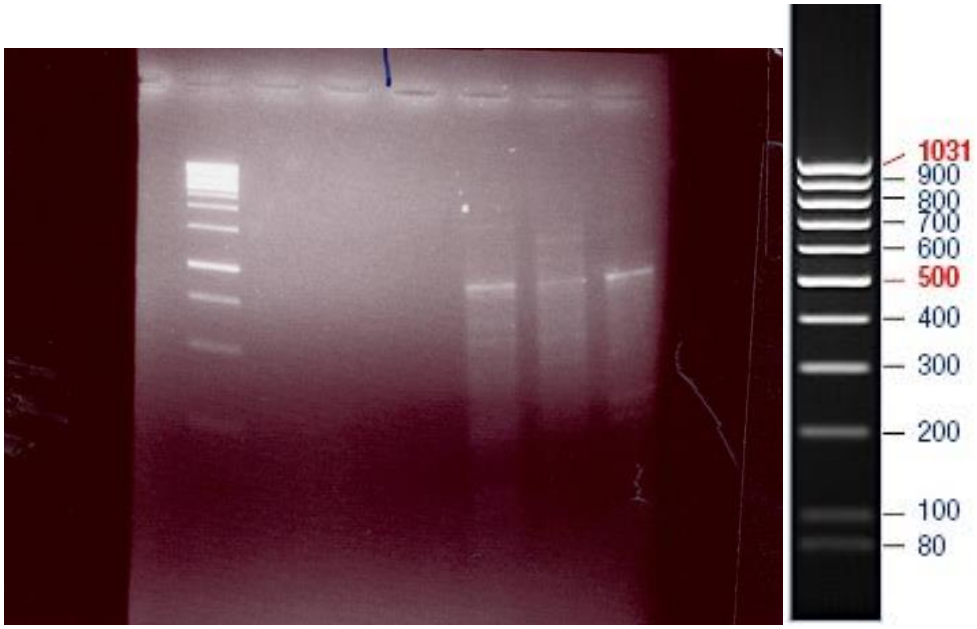
(PCR ID 4779: 10th to 13th lane, DNA ladder representing lane 1)

iii. Allele 1 of individual GM030005-specific

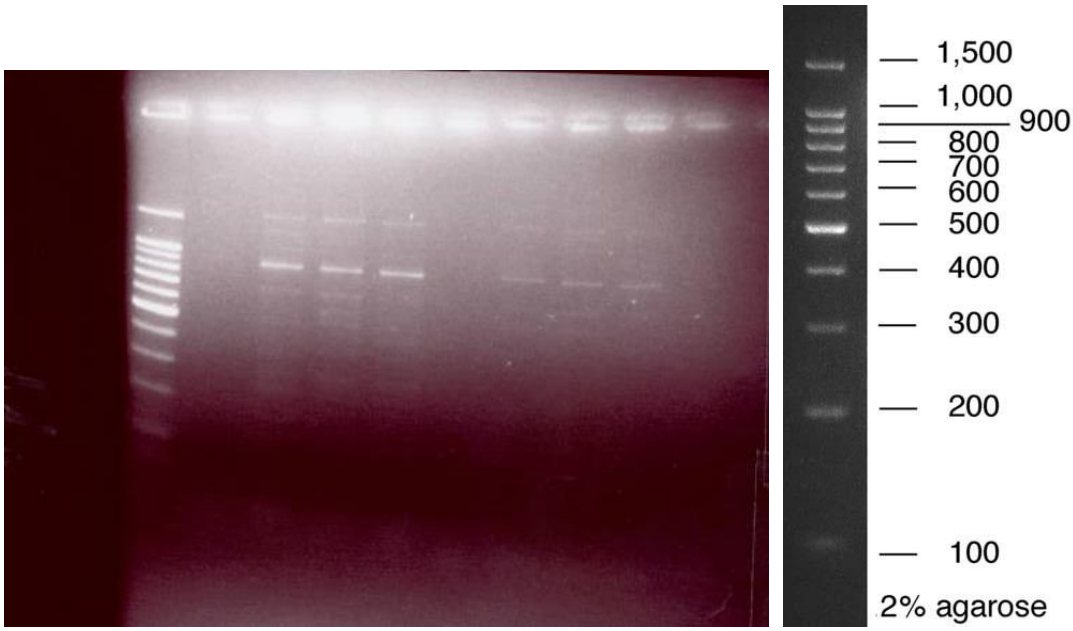
Forward primer: AL1_716A_F,
Reverse primer: DBQEX3R2,
PCR IDs 4824, 4808 and 4775



(PCR ID 4824: 2nd to 5th and 7th to 9th lane, DNA ladder representing lane 1)



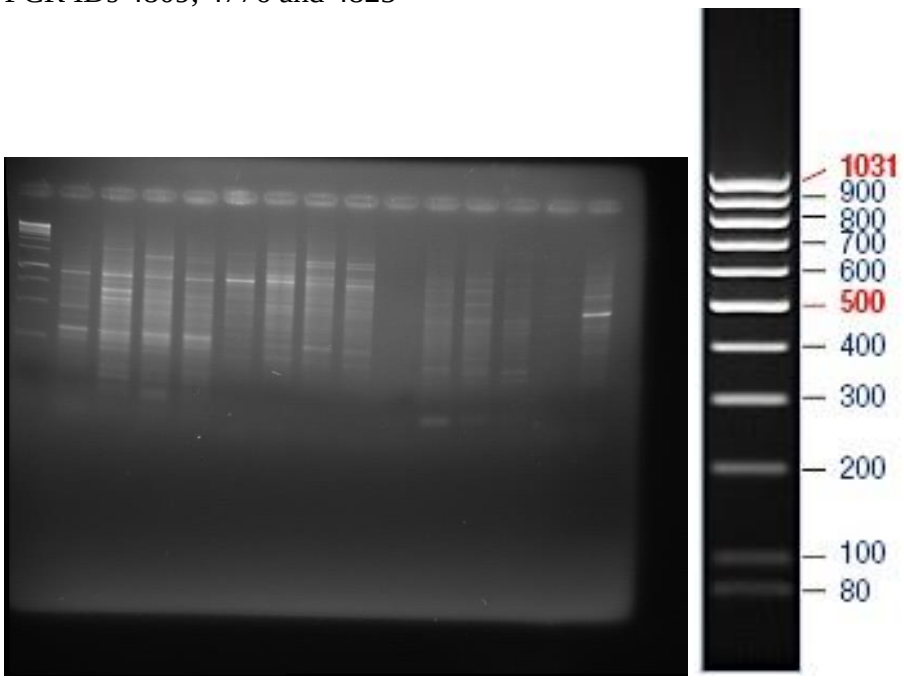
(PCR ID 4808: 4th to 7th lane, DNA ladder representing lane 1)



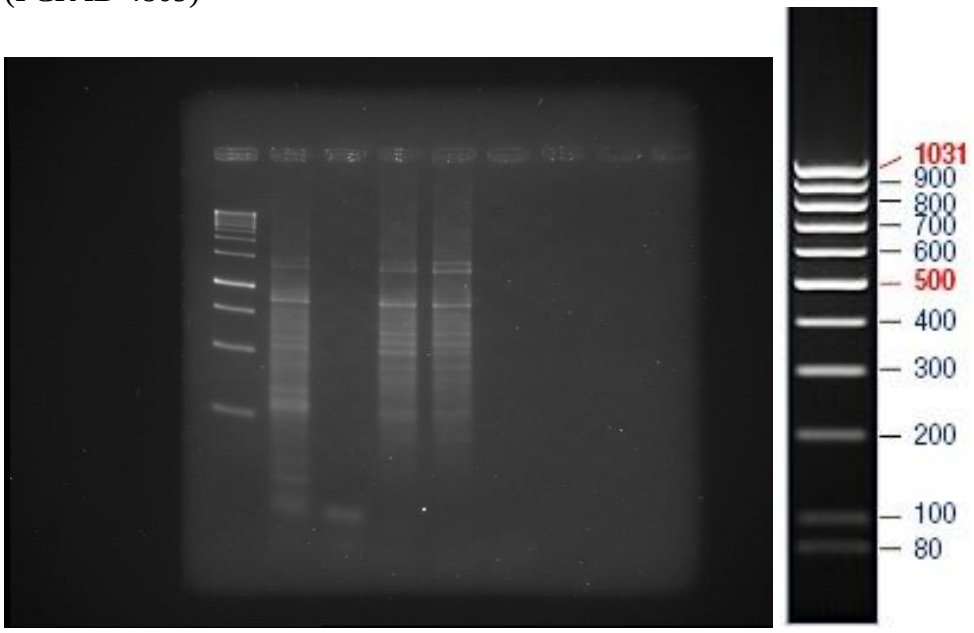
(PCR ID 4775)

iv. Allele 2 of individual GM030005-specific

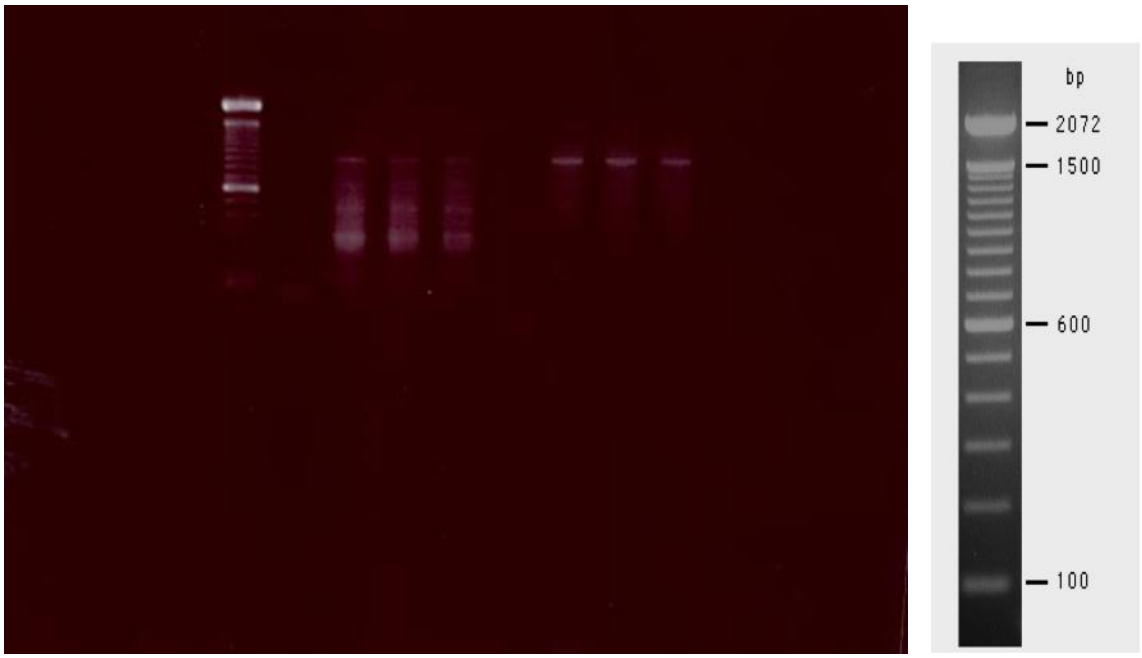
Forward primer: AL1_716B_F,
Reverse primer: DBQEX3R2,
PCR IDs 4809, 4776 and 4825



(PCR ID 4809)



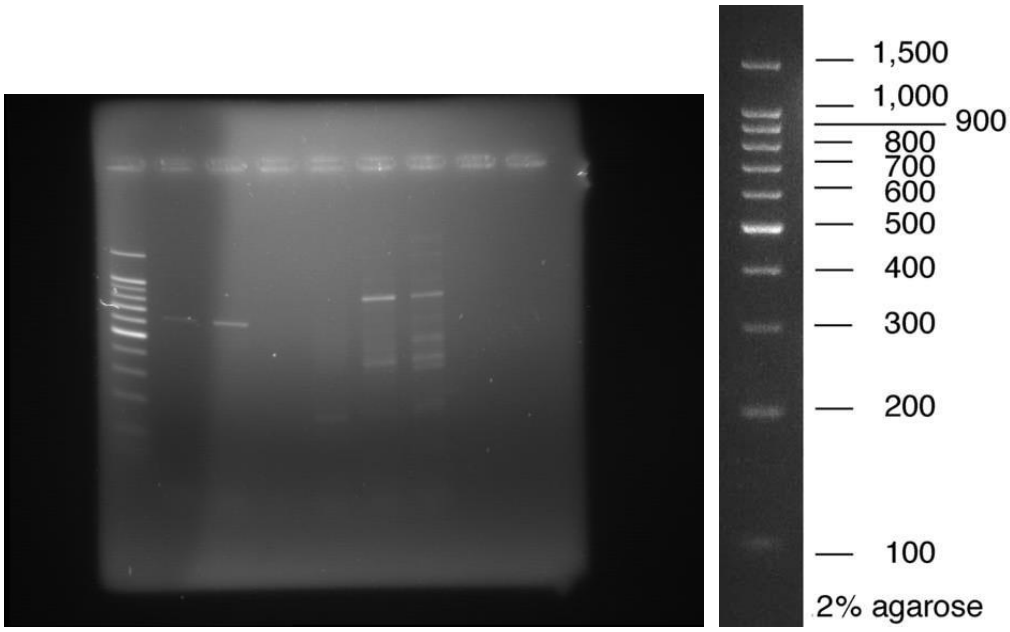
(PCR ID 4776)



(PCR ID 4825: 7th to 9th lane, DNA ladder representing lane 1)

v. Allele 3 of individual GM030005-specific

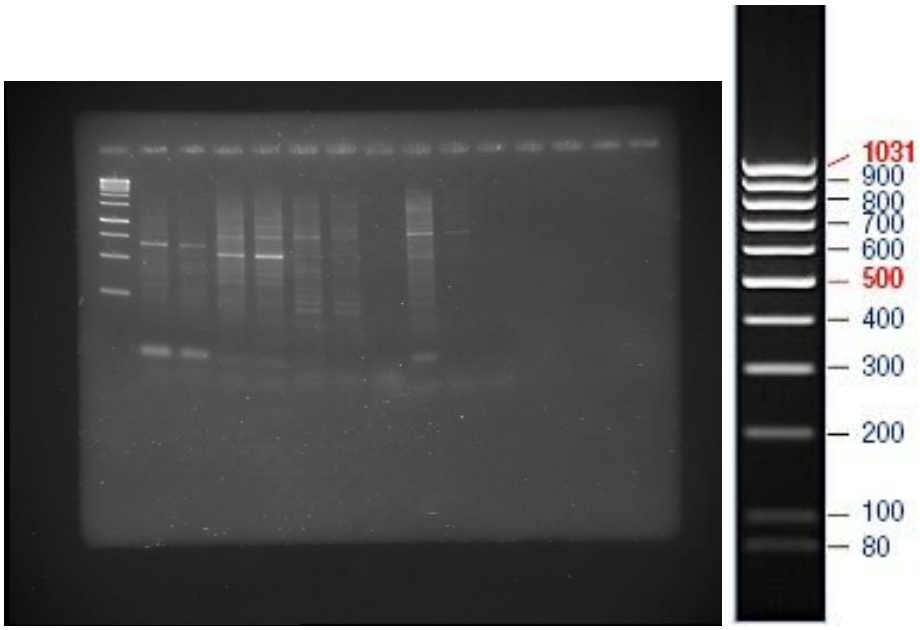
Forward primer: DQBHM01,
Reverse primer: DBQ_INTRON2R,
PCR ID 4820



(PCR ID 4820: 5th to 7th lane, DNA ladder representing lane 1)

vi. Allele 1 of individual WI920009-specific

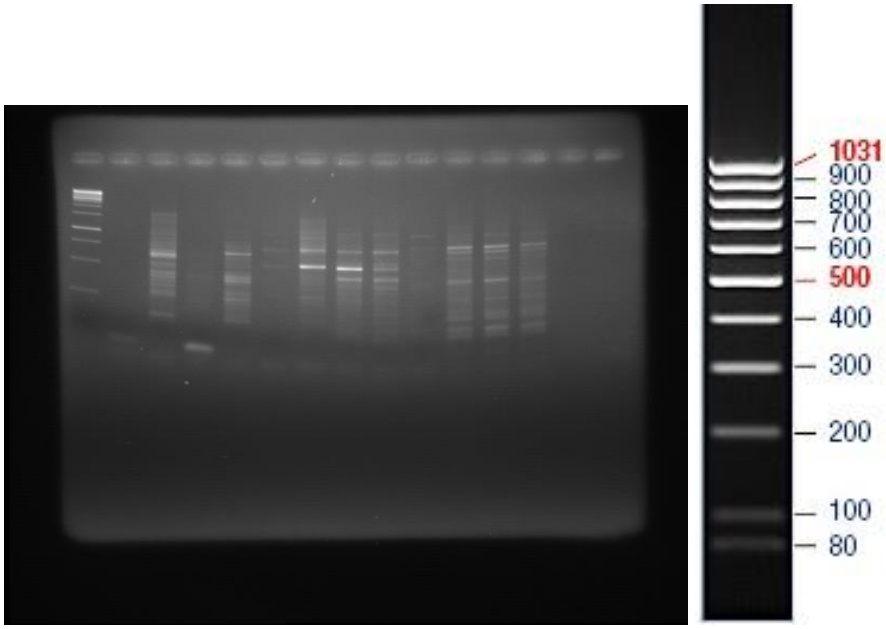
Forward primer: AL1_403_F,
Reverse primer: DBQ_INTRON2R,
PCR ID 4766



(PCR ID 4766: 2nd and 3rd line, DNA ladder representing lane 1)

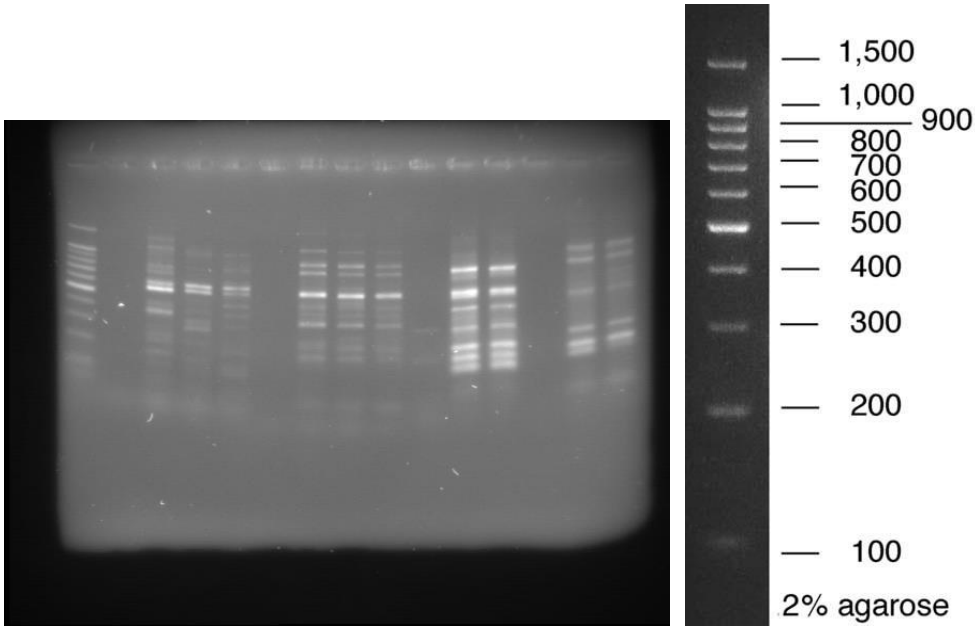
vii. Allele 2 of individual WI920009-specific

Forward primer: AL2_403_F,
Reverse primer: DQB_INTRON2_R3,
PCR ID 4806

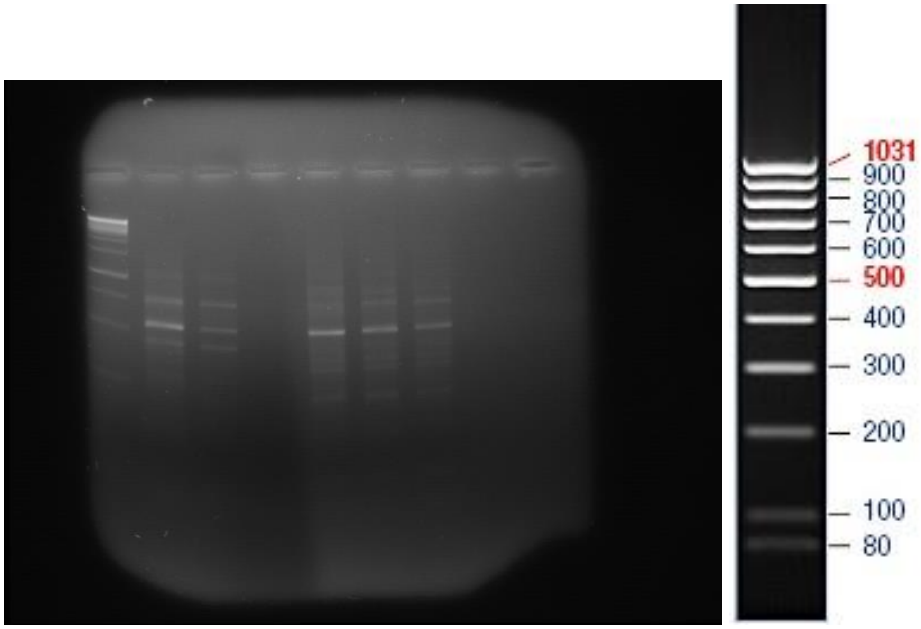


(PCR ID 4806)

Forward primer: AL2_403_F,
Reverse primer: DQB_INTRON2R,
PCR ID 4828 and 4810



(PCR ID 4828)



(PCR ID 4810)

OUTPUT PHASE

All available HLA-DQB exon 2 sequences of humpback whales from this study and downloaded from GenBank were used for this analysis. PHASE is a software that infers statstically the two alleles of an unphased heterzygous sequence [108]. The sequences were reduced to an alignment of only the 16 polymorphic sites. All individuals were forced to harbor not more than two different sequences prior to the phasing by subtracting two alleles (two of the four letters at every position) using the nine cloned alleles in this study.

Number of sequences: 262 Number of sequences used: 262
Selected region: 1-16 Number of sites: 16
Total number of sites (excluding sites with gaps / missing data): 16
Sites with alignment gaps: not considered
Number of variable sites: 16

========== Haplotype Distribution ==========
Number of Haplotypes, h: 63

Hap_1: 32 [MenoGOM-DQB*05c-1 MenoGOM-DQB*05c-2 MenoGOM-DQB*05c-1 MenoGOM-DQB*05c-2 MenoSEA-DQB*19c-1 MenoSEA-DQB*19c-2 WI920425-1 WI920425-2 WI920408-1 WI920408-2 WI920009-1 WI920426-2 WI920370-2 WI920362-2 WI920441-2 GM030005-1 GM030005-2 GM930026-2 GM030217-2 GM930028-1 GM930109-2 GM980049-2 GM050008-2 GM050009-2 GM920224-2 GM970041-2 GM920224-2 GM960006-2 GM960028-2 GM990111-2 GM940030-2 GM940031-2]
Hap_2: 5 [MenoGOM-DQB*24c-1 MenoGOM-DQB*24c-2 GM030217-1 GM970008-1 GM990026-2]
Hap_3: 55 [MenoGOM-DQB*16c-1 MenoGOM-DQB*16c-2 MenoGOM-DQB*16c-1 MenoGOM-DQB*16c-2 MenoCAx-DQB*08c-1 MenoCAx-DQB*08c-2 WI920353-2 WI920015-2 WI920027-2 WI920361-2 WI920355-1 WI920423-1 WI920021-2 WI920359-2 WI920024-1 WI920024-2 WI920369-1 WI920369-2 WI920416-1 WI920416-2 WI920427-1 WI920427-2 WI920428-1 WI920428-2 WI920431-1 WI920431-2 WI920001-2 WI920010-2 WI920349-2 WI920363-2 WI920417-2 WI920430-2 GM970025-1 GM980029-1 GM920089-2 GM990065-1 GM970026-2 GM980049-1 GM050008-1 GM050009-1 GM970052-2 GM920096-2 GM960023-1 GM960023-2 GM980001-1 GM980001-2 GM920031-1 GM920031-2 GM930035-1 GM930035-2 GM920202-1 GM920202-2 GM960022-2 GM920092-2 GM990066-2]
Hap_4: 21 [MenoGOM-DQB*22c-1 MenoGOM-DQB*22c-2 MenoGBx-DQB*11c-1 MenoGBx-DQB*11c-2 MenoCAx-DQB*02c-1 MenoCAx-DQB*02c-2 WI920426-1 WI920415-1 WI920001-1 WI920010-1 WI920349-1 WI920363-1 WI920417-1 WI920430-1 GM930029-1 GM920224-1 GM970041-1 GM920224-1 GM960022-1 GM920092-1 GM990066-1]
Hap_5: 10 [MenoGOM-DQB*25c-1 MenoGOM-DQB*25c-2 MenoSEA-DQB*18c-1 MenoSEA-DQB*18c-2 MenoCAx-DQB*03c-1 MenoCAx-DQB*03c-2 WI920009-2 WI920367-2 WI920355-2 WI920423-2]
Hap_6: 2 [MenoGOM-DQB*27c-1 MenoGOM-DQB*27c-2]
Hap_7: 2 [MenoGOM-DQB*28c-1 MenoGOM-DQB*28c-2]
Hap_8: 3 [MenoGOM-DQB*26c-1 MenoGOM-DQB*26c-2 WI920015-1]
Hap_9: 7 [MenoGOM-DQB*12c-1 MenoGOM-DQB*12c-2 MenoGBx-DQB*12c-1 MenoGBx-DQB*12c-2 GM940039-2 GM050099-2 GM050099-2]
Hap_10: 2 [MenoSEA-DQB*23c-1 MenoSEA-DQB*23c-2]
Hap_11: 2 [MenoSEA-DQB*22c-1 MenoSEA-DQB*22c-2]
Hap_12: 2 [MenoSEA-DQB*21c-1 MenoSEA-DQB*21c-2]
Hap_13: 4 [MenoSEA-DQB*20c-1 MenoSEA-DQB*20c-2 GM980051-1 GM980051-2]
Hap_14: 3 [MenoGOM-DQB*17c-1 MenoGOM-DQB*17c-2 GM920222-2]
Hap_15: 6 [MenoGOM-DQB*16c-1 MenoGOM-DQB*16c-2 WI920433-2 WI920028-2 GM970008-2 GM920110-2]
Hap_16: 2 [MenoGOM-DQB*15c-1 MenoGOM-DQB*15c-2]
Hap_17: 3 [MenoGBx-DQB*14c-1 MenoGBx-DQB*14c-2 GM970025-2]
Hap_18: 2 [MenoCAx-DQB*13c-1 MenoCAx-DQB*13c-2]
Hap_19: 3 [MenoCAx-DQB*10c-1 MenoCAx-DQB*10c-2 WI920352-1]
Hap_20: 2 [MenoCAx-DQB*09c-1 MenoCAx-DQB*09c-2]
Hap_21: 10 [MenoCAx-DQB*07c-1 MenoCAx-DQB*07c-2 WI920362-1 WI920441-1 GM970052-1 GM920096-1 GM920033-1 GM920033-2 GM920034-1 GM920034-2]
Hap_22: 2 [MenoCAx-DQB*06c-1 MenoCAx-DQB*06c-2]
Hap_23: 2 [MenoCAx-DQB*05c-1 MenoCAx-DQB*05c-2]

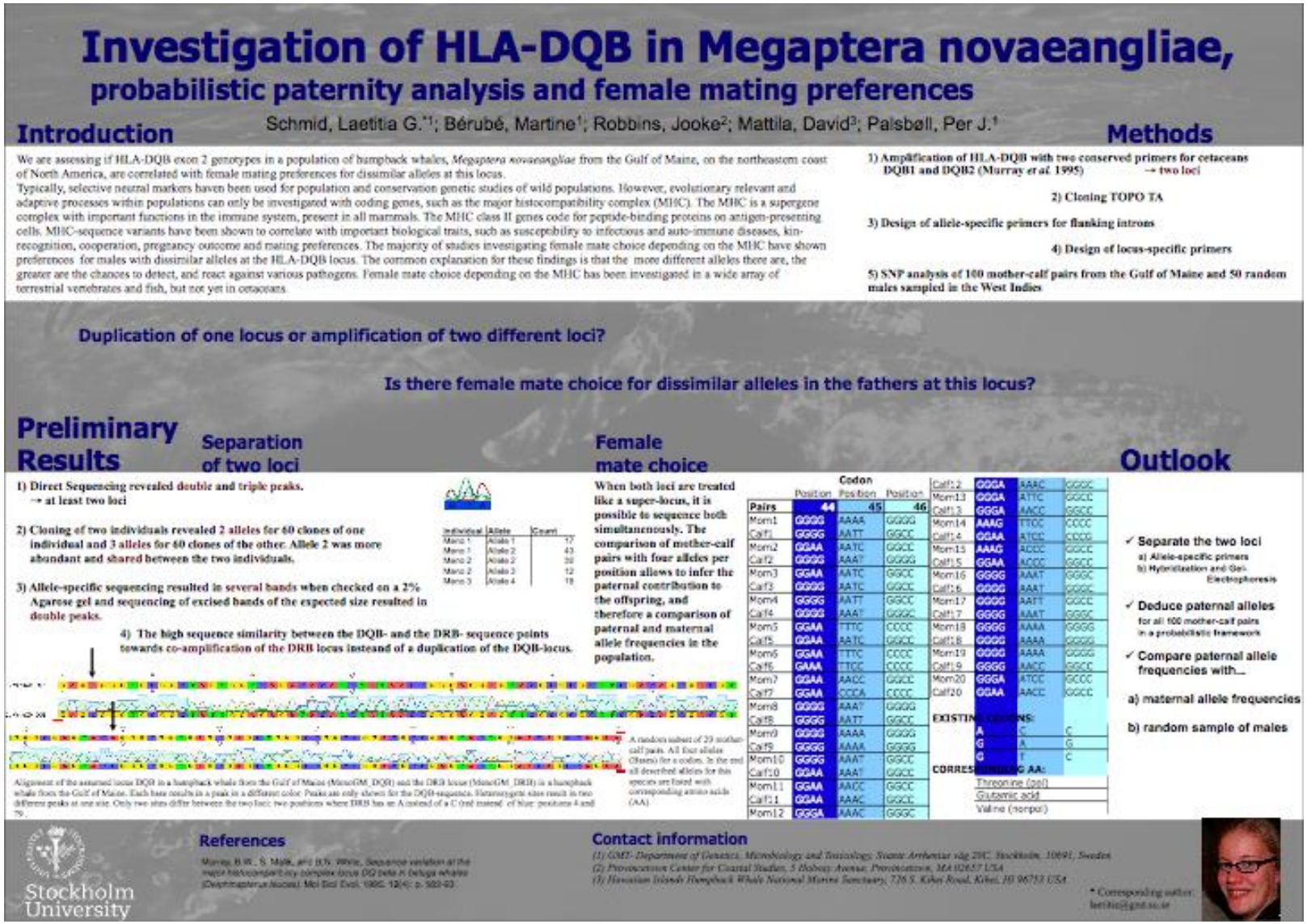
Hap_24: 2 [MenoCAx-DQB*04c-1 MenoCAx-DQB*04c-2]
Hap_25: 3 [MenoCAx-DQB*01c-1 MenoCAx-DQB*01c-2 GM050090-2]
Hap_26: 5 [WI920352-2 WI920370-1 GM990027-1 GM940030-1 GM940031-1]
Hap_27: 2 [WI920004-1 WI920004-2]
Hap_28: 1 [WI920353-1]
Hap_29: 1 [WI920410-1]
Hap_30: 5 [WI920410-2 WI920437-2 WI920022-2 WI920356-2 GM010056-2]
Hap_31: 1 [WI920007-1]
Hap_32: 1 [WI920007-2]
Hap_33: 3 [WI920433-1 WI920366-1 WI920366-2]
Hap_34: 9 [WI920437-1 WI920023-2 WI920028-1 WI920022-1 WI920356-1 GM920110-1 GM930040-2 GM920207-2 GM930040-2]
Hap_35: 1 [WI920027-1]
Hap_36: 1 [WI920415-2]
Hap_37: 1 [WI920023-1]
Hap_38: 4 [WI920360-1 GM960006-1 GM960028-1 GM990111-1]
Hap_39: 2 [WI920360-2 GM940040-2]
Hap_40: 2 [WI920017-1 GM050090-1]
Hap_41: 2 [WI920017-2 GM970009-2]
Hap_42: 8 [WI920361-1 WI920365-1 WI920021-1 WI920359-1 GM970009-1 GM970026-1 GM990105-1 GM990105-2]
Hap_43: 1 [WI920367-1]
Hap_44: 1 [WI920365-2]
Hap_45: 1 [GM930026-1]
Hap_46: 1 [GM940040-1]
Hap_47: 1 [GM930028-2]
Hap_48: 1 [GM990125-1]
Hap_49: 1 [GM990125-2]
Hap_50: 1 [GM010056-1]
Hap_51: 1 [GM990026-1]
Hap_52: 1 [GM980029-2]
Hap_53: 1 [GM920089-1]
Hap_54: 1 [GM930109-1]
Hap_55: 1 [GM990065-2]
Hap_56: 1 [GM980007-1]
Hap_57: 1 [GM980007-2]
Hap_58: 1 [GM940039-1]
Hap_59: 2 [GM990027-2 GM930029-2]
Hap_60: 1 [GM920222-1]
Hap_61: 2 [GM990104-1 GM990104-2]
Hap_62: 3 [GM930040-1 GM920207-1 GM930040-1]
Hap_63: 2 [GM050099-1 GM050099-1]

Table A2: 63 alleles (Hap = haplotypes) revealed by the analysis using the software PHASE [108]. One example individual harboring the allele and the allele sequence consisting of 16 polymorphic positions (for the positions see figure 5, table S3 and table S4) are given.

Hap	SampleID	Sequence
1	WI920370-2	CTCGACCTTAGAACAC
2	MenoGOM-DQB*24c-1	CTAGGAGGGTGACGAG
3	MenoGOM-DQB*16c-1	CTATGAGTGTGACGCG
4	MenoGOM-DQB*22c-1	CTAGGTCCGATCCGAG
5	MenoGOM-DQB*25c-1	GCAGGAGTGTGACCCG
6	MenoGOM-DQB*27c-1	TTCGACCCTATCCCAC
7	MenoGOM-DQB*28c-1	CTCGACCGGTTCCCAG
8	MenoGOM-DQB*26c-1	CTAGGTCCGAGAAGAG
9	MenoGOM-DQB*12c-1	CTCGACCTGTTCCCAC
10	MenoSEA-DQB*23c-1	CTAGGTCCGTTCCCAC
11	MenoSEA-DQB*22c-1	CTAGGAGGGTATCGCG

12	MenoSEA-DQB*21c-1	CTCGACCTTTGAACCG
13	MenoSEA-DQB*20c-1	GCAGGAGTGAGACCCG
14	MenoGOM-DQB*17c-1	GCAGGAGTGAGAACCG
15	WI920433-2	CTCGACCTTTGACCAC
16	MenoGOM-DQB*15c-1	CTCGACCTGAGAACCG
17	MenoGBx-DQB*14c-1	CTCGACCTGTGACCAC
18	MenoCAx-DQB*13c-1	CTAGGTCCGAGAACAC
19	MenoCAx-DQB*10c-1	CTAGGAGGTAGAACAC
20	MenoCAx-DQB*09c-1	CTAGGAGGGTATCGAG
21	MenoCAx-DQB*07c-1	CTATGAGTGTGACGAG
22	MenoCAx-DQB*06c-1	GCAGGAGTTAGAACAC
23	MenoCAx-DQB*05c-1	CTAGGTCCGAGAACCG
24	MenoCAx-DQB*04c-1	CTGGGTCCGATCACCG
25	MenoCAx-DQB*01c-1	CTCGACCTGTGAACAC
26	WI920437-2	CTCGACCTTTTCCCAC
27	WI920004-2	CTCGACCTGAGAACAC
28	WI920353-1	CTAGACCTTAGAACAC
29	WI920410-1	CTAGGAGGGTAACGCG
30	WI920410-2	CTCGACCTTTTCCCAC
31	WI920007-1	CTAGGAGAGTGACGAG
32	WI920007-2	CTAGGTCTTAGAACCG
33	WI920433-1	CTAGGTCCGTGACGAG
34	WI920437-1	CTATGAGTGTGACCCG
35	WI920027-1	CTATGAGTGTATCGCG
36	WI920415-2	CTCGACCTTAATACAC
37	WI920023-1	CTAGGAGGTATCCGAG
38	WI920360-1	CTAGGAGTGTGACCAG
39	WI920360-2	CTCTACCTGTTCCCAC
40	WI920017-1	CTATGAGTGTTCCCCC
41	WI920017-2	CTCGACCTTAGACCAC
42	WI920361-1	CTAGGTCCGAGACGAG
43	WI920367-1	CTATGTCCGATCCGAG
44	WI920365-1	CTAGGTCCGAGACGAG
45	GM930026-1	CTAGGAGGGTGACCAG
46	GM940040-1	CTCGACCGGTTCCCCC
47	GM930028-2	GCAGACCGGAGACGCC
48	GM990125-1	CTATACCGGAGACGCC
49	GM990125-2	CTATGAGTTTGAACCG
50	GM010056-1	CTAGGTCCGTTCCGAG
51	GM990026-2	CTAGGAGGGTGACGAG
52	GM980029-2	GCAGACCCTATCCCAC
53	GM920089-1	CTAGACCTTATCCCAC
54	GM930109-1	CTAGGTCTTTTCACAC
55	GM990065-2	CTCGACCTTATCCCAC
56	GM980007-1	CTCGACCTTTGAACAC
57	GM980007-2	GTAGGTCCGATCCGAG
58	GM940039-1	CTATGTCTGTGACCCG
59	GM990027-2	CCATGAGTGTGACCCG
60	GM920222-1	CTATGAGTGAGAAGCG
61	GM990104-1	GCAGGTCCGAGACCAG
62	GM930040-1	CTAGACCTTATCACAC
63	GM050099-1	CTAGGTCTTATCCCCG

Poster presented at the 16th Biennial Conference on the Biology of Marine Mammals in Québec City, October 2009.



REFERENCES

1. Sommer, S., *The importance of immune gene variability (MHC) in evolutionary ecology and conservation.* Front Zool, 2005. **2**: p. 16.
2. Avise, J.C., *Perspective: conservation genetics enters the genomics era.* Conservation Genetics, 2010. **11**: p. 665-669.
3. ; Available from: http://www.1000genomes.org/page.php.
4. Klein, J., *The natural history of the major histocompatibility complex*. 1986, New York: Wiley.
5. Dengjel, J., et al., *Autophagy promotes MHC class II presentation of peptides from intracellular source proteins.* Proc Natl Acad Sci U S A, 2005. **102**(22): p. 7922-7.
6. Ohta, T., *On the pattern of polymorphisms at major histocompatibility complex loci.* J Mol Evol, 1998. **46**(6): p. 633-8.
7. Groenen, M.A.M., et al., *The nucleotide sequence of bovine MHC class II DQB and DRB genes.* Immunogenetics, 1990. **31**: p. 37-44.
8. Brown, J.H., et al., *Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1*. Nature, 1993. **364**(6432): p. 33-9.
9. Chaves, F.A., et al., *Peptide-binding motifs for the I-A(d) MHC class II molecule: Alternate pH-dependent binding behavior.* Biochemistry, 2006. **45**(20): p. 6426-6433.
10. Doherty, P.C. and R.M. Zinkernagel, *Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex.* Nature, 1975. **256**(5512): p. 50-2.
11. Hughes, A.L. and M. Nei, *Patterns of nucleotide substitutions at major histocompatibility complex class I loci reveals overdominant selection.* . Nature, 1988. **335**(167-170).
12. Milinski, M., *The Major Histocompatibility Complex, Sexual Selection and Mate Choice.* Annu Rev Ecol Evol Syst, 2006(37): p. 159-86.
13. De Boer, R.J. and A.S. Perelson, *How diverse should the immune system be?* Proc R Soc Lond B, 1993. **252**: p. 171-75.
14. Borghans, J.A., A.J. Noest, and R.J. De Boer, *Thymic selection does not limit the individual MHC diversity*. Eur J Immunol, 2003. **33**(12): p. 3353-8.
15. De Boer, R.J., et al., *Heterozygote advantage fails to explain the high degree of polymorphisms of the MHC.* Immunogenetics, 2004. **55**: p. 725-31.
16. Nowak, M.A., K. Tarczy-Hornoch, and J.M. Austyn, *The optimal number of major histocompatibility complex molecules in an individual.* Proc Natl Acad Sci U S A, 1992. **89**(22): p. 10896-9.
17. Hedrick, P.W., *Pathogen resistance and genetic variation at MHC loci.* Evolution, 2002. **56**(10): p. 1902-8.
18. Wedekind, C., et al., *Non-random fertilization in mice correlates with the MHC and something else*. Heredity, 1996. **77**: p. 400-409.
19. Hedrick, P.W., *Evolutionary genetics to the major histocompatibility complex.* The American Naturalist, 1994. **143**(6): p. 945-964.
20. Hughes, A.L. and M. Nei, *Maintenance of MHC polymorphism.* Nature, 1992. **355**(6359): p. 402-3.
21. Wedekind, C. and D. Penn, *MHC genes, body odours, and odour preferences.* Nephrology Dialysis Transplantation, 2000. **15**(9): p. 1269-1271.
22. Hughes, A.L. and M. Nei, *Models of host-parasite interaction and MHC polymorphism.* Genetics, 1992. **132**(3): p. 863-4.
23. Spurgin, L.G. and D.S. Richardson, *How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings.* Proc Biol Sci. **277**(1684): p. 979-88.
24. Slade, R.W. and H.I. McCallum, *Overdominant vs. frequency-dependent selection at MHC loci.* Genetics, 1992. **132**(3): p. 861-4.
25. Ladle, R.J., *Parasites and sex: catching the Red Queen.* Trends Ecol Evol, 1992. **7**: p. 405-8.
26. Penn, D.J. and W.K. Potts, *The evolution of mating preferences and major histocompatibility complex genes.* The American Naturalist, 1999. **153**: p. 145-164.
27. Hamilton, W.D., R. Axelrod, and R. Tanese, *Sexual reproduction as an adaptation to resist parasites (a review)*. Proc Natl Acad Sci U S A, 1990. **87**(9): p. 3566-73.

28. Hill, A.V., et al., *Common west African HLA antigens are associated with protection from severe malaria.* Nature, 1991. **352**(6336): p. 595-600.
29. Ekblom, R., M. Grahn, and J. Hoglund, *Patterns of polymorphism in the MHC class II of a non-passerine bird, the great snipe (Gallinago media).* Immunogenetics, 2003. **54**(10): p. 734-41.
30. Ekblom, R., et al., *Spatial pattern of MHC class II variation in the great snipe (Gallinago media).* Mol Ecol, 2007. **16**(7): p. 1439-51.
31. Ekblom, R., et al., *Balancing selection, sexual selection and geographic structure in MHC genes of Great Snipe.* Genetica, 2008.
32. Hamilton, W.D., *The genetical evolution of social behaviour. I.* J Theor Biol, 1964. **7**(1): p. 1-16.
33. Edwards, S.V. and P.W. Hedrick, *Evolution and ecology of MHC molecules: from genomics to sexual selection.* Trends Ecol Evol, 1998. **13**(8): p. 305-311.
34. Miller, H.C., et al., *Influence of major histocompatibility complex genotype on mating success in a free-ranging reptile population.* Proc Biol Sci, 2009. **276**(1662): p. 1695-704.
35. Carrington, M., G.M. Nelson, and M.P. Martin, *HLA and HIV-1: heterozygote advantage and B*35-Cw04 diadvantage.* Science, 1999. **283**(5408): p. 1748-5172.
36. Thursz, M.R., et al., *Heterozygote advantage for HLA class-II type in hepatitis B virus infection.* Nat Genet, 1997. **17**(1): p. 11-2.
37. Penn, D. and W.K. Potts, *How do major histocompatibility complex genes influence odor and mating preferences?* Proc R Soc Lond B, 1998. **264**: p. 1471-1479.
38. Singer, A.G., *Chemistry of odor types in mice: Fractionation and bioassay.* J Chem Ecol, 1993. **19**: p. 569-579.
39. Yeager, M., M. Carrington, and A.L. Hughes, *Class I and class II MHC bind self peptide sets that are strikingly different in their evolutionary characteristics.* Immunogenetics, 2000. **51**(1): p. 8-15.
40. Howard, J.C., *H-2 and mating preferences.* Nature, 1977. **266**: p. 406-408.
41. Pearse-Pratt, R., et al., *Evolutionary and immunological implications of the role of the MHC in olfactory signalling*, in *Chemical Signals in Vertebrates*, R.L. Doty and D. Müller-Schwarze, Editors. 1992, VI. Plenum Press: New York. p. 167-175.
42. McClintock, M.K., *Menstrual synchorony and suppression.* Nature, 1971. **229**(5282): p. 244-5.
43. Monti-Bloch, L., et al., *The human vomernasal system.* Psychoneuroendocrinology, 1994. **19**: p. 673-686.
44. Stern, K. and M.K. McClintock, *Regulation of ovulation by human pheromones.* Nature, 1998. **392**(6672): p. 177-9.
45. Singh, P.B., R.E. Brown, and B. Roser, *MHC antigens in urine as olfactory recognition cues* Nature, 1987. **327**: p. 161-164.
46. Novotny, M.V., et al., *Chemical identification of MHC-influenced volatile compounds in mouse urine. I: Quantitative Proportions of Major Chemosignals.* J Chem Ecol, 2007. **33**(2): p. 417-34.
47. Brown, R.E., B. Roser, and P.B. Singh, *Class I and class II regions of the major histocompatibility complex both contribute to individual odors in congenic inbred strains of rats.* Behav Genet, 1989. **19**: p. 659-674.
48. Lanyon, C.V., et al., *Murine scent mark microbial communities are genetically determined.* FEMS Microbiol Ecol, 2007. **59**(3): p. 576-83.
49. Leinders-Zufall, T., et al., *MHC class I peptides as chemosensory signals in the vomeronasal organ.* Science, 2004. **306**: p. 1033-1037.
50. Singer, A.G., G.K. Beauchamp, and K. Yamazaki, *Volatile signals of the major histocompatibility complex in male mouse urine.* Proc Natl Acad Sci U S A, 1997. **94**(6): p. 2210-4.
51. Potts, W.K., C.J. Manning, and E.K. Wakeland, *Mating patterns in seminatural populations of mice influenced by MHC genotype.* Nature, 1991. **352**(6336): p. 619-21.
52. Yamazaki, K., et al., *Control of mating preferences in mice by genes in the major histocompatibility complex.* J Exp Med, 1976. **144**(5): p. 1324-35.

53. Olsson, M., et al., *Major histocompatibility complex and mate choice in sand lizards.* Proc Biol Sci, 2003. **270 Suppl 2**: p. S254-6.
54. Olsen, K.H., et al., *MHC and kin discrimination in juvenile Arctic charr, Salvelinus alpinus (L.).* Anim Behav, 1998. **56**(2): p. 319-327.
55. Landry, C., et al., *'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (Salmo salar).* Proc Biol Sci, 2001. **268**(1473): p. 1279-85.
56. Aeschlimann, P.B., et al., *Female sticklebacks Gasterosteus aculaetus use self-reference to optimize MHC allele number during mate selection.* Behav Ecol Sociobiol, 2003. **54**: p. 119-126.
57. Bonneaud, C., et al., *Complex Mhc-based mate choice in a wild passerine.* Proc Biol Sci, 2006. **273**(1590): p. 1111-6.
58. Richardson, D.S., et al., *MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler.* Proc Biol Sci, 2005. **272**(1564): p. 759-67.
59. Hale, M.L., et al., *Is the peacock's train an honest signal of genetic quality at the major histocompatibility complex?* J Evol Biol, 2009. **22**(6): p. 1284-94.
60. Gillingham, M.A., et al., *Cryptic preference for MHC-dissimilar females in male red junglefowl, Gallus gallus.* Proc Biol Sci, 2009. **276**(1659): p. 1083-92.
61. Egid, K. and J.L. Brown, *The major histocompatibility complex and female mating preferences in mice.* Anim Behav, 1989. **38**: p. 4186-4188.
62. Wedekind, C. and S. Furi, *Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity?* Proceedings of the Royal Society of London Series B-Biological Sciences, 1997. **264**(1387): p. 1471-1479.
63. Jacob, S., et al., *Paternally inherited HLA alleles are associated with women's choice of male odor.* Nat Genet, 2002. **30**(2): p. 175-9.
64. Ober, C., et al., *HLA and mate choice in humans.* Am J Hum Genet, 1997. **61**(3): p. 497-504.
65. Hedrick, P.W. and F.L. Black, *HLA and mate selection: no evidence in South Amerindians.* Am J Hum Genet, 1997. **61**(3): p. 505-11.
66. Paterson, S., *No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population.* Proc R Soc Lond B, 1997. **264**(1813-1819).
67. Westerdahl, H., *No evidence for an MHC-based female mating preference in great reed warblers.* Molecular Ecology, 2004. **13**: p. 2465-2470.
68. Clapham, P.J. and J.G. Mead, *Megaptera Novaeangliae.* Mammalian Species, 1990. **604**: p. 1-9.
69. Robbins, J., *Structure and dynamics of the Gulf of Maine humpback whale population*, in *School of Biology*. 2007, University of St. Andrews: St. Andrews. p. 1-168.
70. Martin, A.R., et al., *Migration of humpback whales between the Carribean and Iceland.* Journal of Mammalogy, 1984. **65**(2): p. 330-333.
71. Baker, C.S., et al., *Migratory movement and population structure of humpback whales Megaptera novaeangliae in the central and eastern North Pacific.* Marine Ecology Progress Series, 1986. **31**: p. 105-119.
72. Clapham, P.J. and C.A. Mayo, *Reproduction and recruitment of individually indentified humpback whales, Megaptera novaeangliae, observed in Massachusetts Bay, 1979-1985.* Canadian Journal of Zoology, 1987. **65**: p. 2853-2863.
73. Clapham, P.J. and C.A. Mayo, *Reproduction of humpback whales (Megaptera novaeangliae) observed in the Gulf of Maine.*, in *Reports of the International Whaling Comission (special issue)*. 1990. p. 171-175.
74. Stevick, P.T., et al., *Segretation of migration by feeding ground origin in North Atlantic humpback whales (Megaptera novaeangliae).* J Zool Lon, 2003. **259**: p. 231-237.
75. Palsboll, P.J., et al., *Genetic tagging of humpback whales.* Nature, 1997. **388**(6644): p. 767-9.
76. Stevick, P.T., et al., *Population spatial structuring on the feeding grounds in North Atlantic humpback whales.* Journal of Zoology, 2003. **270**: p. 244-255.
77. Smith, T.D., et al., *An ocean-basin-wide mark-recapture study of the North Atlantic humback whale (Megaptera novaeangliae).* Marine Mammal Science, 2006. **15**(1): p. 1-32.
78. *Humpback Mother and Calf.* 2009; Available from: http://animal.discovery.com/features/free-wallpaper/humpback-whale-wallpaper.html.

79. Clapham, P.J., et al., *Abundance and demographic parameters of humpback whales from the Gulf of Maine, and stock definition relative to the Scotian Shelf.* Journal of Cetacean Research and Management, 2003. **5**: p. 13-22.
80. Katona, S.K. and H.P. Whitehead, *Identifying humpback whales using their natural markings.* Polar Record, 1981. **20**: p. 439-444.
81. Wedekind, C., et al., *Mhc-Dependent Mate Preferences in Humans.* Proceedings of the Royal Society of London Series B-Biological Sciences, 1995. **260**(1359): p. 245-249.
82. Edwards, S.V., et al., *Toward an evolutionary genomics of the avian Mhc.* Immunol Rev, 1999. **167**: p. 119-32.
83. Milinski, M. and C. Wedekind, *Evidence for MHC-correlated perfume preferences in humans.* Behavioral Ecology, 2001. **12**(2): p. 140-149.
84. Zelano, B. and S.V. Edwards, *An MHC component to kin recognition and mate choice in birds: predictions, progress, and prospects.* Am Nat, 2002. **160 Suppl 6**: p. S225-37.
85. Milinski, M., et al., *Mate choice decisions of stickleback females predictably modified by MHC peptide ligands.* Proc Natl Acad Sci U S A, 2005. **102**(12): p. 4414-8.
86. Lambertsen, R.H., *A biopsy system for large whales and its use for cytogenetics.* Journal of Mammalogy, 1987. **68**(2): p. 443-445.
87. Palsboll, P.J., F. Larsen, and E. Sigurd Hansen, *Sampling of skin biopsies from free-ranging large cetaceans in West Greenland:Development of new biopsy tips and bolt designs.* Rep. Int. Whaling Commiss. Spec. Iss., 1991(13): p. 71-79.
88. Amos, W. and A.R. Hoelzel, *Long-term preservation of whale skin for DNA analysis*. Genetic ecology of whales and dolphins. Rep. int. Whal. Comm. (spec. iss. 13), ed. A.R. Hoelzel. 1991.
89. Palsboll, P.J., et al., *Determination of gender in cetaceans by the polymerase chain reaction.* Canadian Journal of Zoology, 1992. **70**(11): p. 2166-2170.
90. Berube, M. and P. Palsboll, *Identification of sex in cetaceans by multiplexing with three ZFX and ZFY specific primers.* Mol Ecol, 1996. **5**(2): p. 283-7.
91. Tsuji, K., M. Aizawa, and T. Sasazuki, *HLA 1991, Proceedings of the XI International Histocompatibility Workshop and Conference*. Oxford Scientific Publications. Vol. 1 and 2. 1991, Oxford: Oxford University Press.
92. Murray, B.W., S. Malik, and B.N. White, *Sequence variation at the major histocompatibility complex locus DQ beta in beluga whales (Delphinapterus leucas).* Mol Biol Evol, 1995. **12**(4): p. 582-93.
93. Heimeier, D., et al., *Confirmed expression of MHC class I and class II genes in the New Zealand endemic Hector's dolphin (Cephalorhynchus hectori).* Marine Mammal Science, 2009. **25**(1): p. 68-90.
94. Kanai, T.H., et al., *Allelic diversity at class II DRB1 and DQB loci of the pig MHC (SLA).* Immunogenetics, 1999. **50**(5-6): p. 295-300.
95. Lenz, T.L. and S. Becker, *Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci--implications for evolutionary analysis.* Gene, 2008. **427**(1-2): p. 117-23.
96. Campbell N, A., *Biology*. 4th ed. 1996, Menlo Park, CA: The Benjamin/Cummings Publishing Company. 379-380.
97. Hoover C. *Invitrogen Web Catalog*. 1998 [cited 2009 Aug 18]; Available from: http://www.invitrogen.com/pdf_manuals/topota_man.pdf.
98. Drummond, A.J., et al., *Geneious v4.8*. 2009.
99. R Development and Core Team, *R: A Language and Environment for Statistical Computing*. 2009: Vienna, Austria.
100. Brown, G., et al., *T-cell proliferation and expression of MHC class II antigens.* Scand J Immunol, 1984. **19**(4): p. 373-7.
101. Blouin, M.S., et al., *Use of microsatellite loci to classify individuals by relatedness.* Mol Ecol, 1996. **5**(3): p. 393-401.
102. Raymond, M. and F. Rousset, *GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism.* J. Heredity, 1995. **86**: p. 248-249.
103. Rousset, F., *Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux.* Mol. Ecol. Resources, 2008. **8**: p. 103-106.

104. Guo, S.W. and E.A. Thompson, *Performing the exact test of Hardy-Weinberg proportion for multiple alleles.* Biometrics, 1992. **48**: p. 361-372.
105. Weir, B.S. and C.C. Cockerham, *Estimating F-statistics for the analysis of population structure*. Evolution, 1984. **38**: p. 1358-1370.
106. Raymond, M. and F. Rousset, *An exact test for population differentiation.* Evolution, 1995. **49**: p. 1280-1283.
107. Rice, W.R., *Analyzing tables of statistical tests*. Evolution, 1989. **43**: p. 223-225.
108. Stephens, M., N. Smith, and P. Donnelly, *A new statistical method for haplotype reconstruction from population data.* American Journal of Human Genetics, 2001. **68**: p. 978-989.
109. Rozas, J., et al., *DnaSP, DNA polymorphism analyses by the coalescent and other methods.* Bioinformatics, 2003. **19**: p. 2496-2497.
110. Tajima, F., *Statistical method for testing the neutral mutation hypothesis by DNA polymorphisms.* Genetics, 1989. **123**: p. 585-595.
111. Fu, Y.-X. and W.-H. Li, *Statistical test of neutrality of mutations.* Genetics, 1993. **133**: p. 693-709.
112. Baker, C.S., et al., *Diversity and duplication of DQB and DRB-like genes of the MHC in baleen whales (suborder: Mysticeti)*. Immunogenetics, 2006. **58**(4): p. 283-96.
113. Klein, J., et al., *Nomenclature for the major histocompatibility complexes of different species: a proposal.* Immunogenetics, 1990. **31**(4): p. 217-9.
114. Ellis, S.A. and K.T. Ballingall, *Cattle MHC: evolution in action?* Immunol Rev, 1999. **167**: p. 159-68.
115. Kimura, M., *The neutral theory of Molecular Evolution*. 1983, Cambridge, Massachusetts: Cambridge University Press.
116. Binz, T., et al., *SSCP analysis of Mhc class IIB genes in the threespine stickleback.* Journal of Fish Biology, 2001. **58**(3): p. 887-890.
117. Richardson, D.S. and H. Westerdahl, *MHC diversity in two Acrocephalus species: the outbred Great reed warbler and the inbred Seychelles warbler.* Mol Ecol, 2003. **12**(12): p. 3523-9.
118. Babik, W., et al., *New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC systems.* Mol Ecol Res, 2009(9): p. 713-719.
119. Edwards, S.V., et al., *Ancestral polymorphism of Mhc class II genes in mice: implications for balancing selection and the mammalian molecular clock.* Genetics, 1997. **146**(2): p. 655-68.
120. Murray, B.W. and B.N. White, *Sequence variation at the major histocompatibility complex DRB loci in beluga (Delphinapterus leucas) and narwhal (Monodon monoceros)*. Immunogenetics, 1998. **48**(4): p. 242-52.
121. Musolf, K., Y. Meyer-Lucht, and S. Sommer, *Evolution of MHC-DRB class II polymorphism in the genus Apodemus and a comparison of DRB sequences within the family Muridae (Mammalia: Rodentia)*. Immunogenetics, 2004. **56**(6): p. 420-6.
122. Bondinas, G.P., A.K. Moustakas, and G.K. Papadopoulos, *The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function.* Immunogenetics, 2007. **59**(7): p. 539-53.
123. Wegner, K.M., *Massive parallel MHC genotyping: titanium that shines.* Mol Ecol, 2009. **18**(9): p. 1818-20.
124. Binladen, J., M.T.P. Gilbert, and J.P. Bollback, *The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing.* PLoS ONE, 2007. **2**: p. 197.
125. Sunnucks, P., et al., *SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology.* Mol Ecol, 2000. **9**(11): p. 1699-710.
126. Langefors, A., et al., *Screening of Mhc variation in Atlantic salmon (Salmo salar): a comparison of restriction fragment length polymorphism (RFLP), denaturing gradient gel electrophoresis (DGGE) and sequencing.* Mol Ecol, 2000. **9**(2): p. 215-9.
127. Drake, G.J.C., et al., *The use of reference strand-mediated conformational analysis for the study of cheetah (Acinonyx jubatus) feline leucocyte antigen class II DRB polymorphisms.* Molecular Ecology, 2004. **13**: p. 221-229.

128. Offringa, R. and F. van der Lee, *Isolation and characterization of plant genomic DNA sequences via (inverse) PCR amplification.* Methods Mol Biol, 1995. **49**: p. 181-95.
129. Pang, K.M. and D.A. Knecht, *Partial inverse PCR: a technique for cloning flanking sequences.* Biotechniques, 1997. **22**(6): p. 1046-8.
130. Li, Z.X., et al., *Allele-specific, inverse-PCR amplification for genotyping MN blood group.* Biotechniques, 1998. **25**(3): p. 358-60, 362.
131. Raponi, M., I.W. Dawes, and G.M. Arndt, *Characterization of flanking sequences using long inverse PCR.* Biotechniques, 2000. **28**(5): p. 838-40, 842, 844.
132. Sunna Sigurdardottir, et al., *Gene duplications and sequence polymorphism of bovine class II DQB genes.* Immunogenetics, 1992. **35**: p. 205-213.
133. Hoelzel, A.R., J.C. Stephens, and S.J. O'Brien, *Molecular genetic diversity and evolution at the MHC DQB locus in four species of pinnipeds.* Mol Biol Evol, 1999. **16**(5): p. 611-8.
134. Hambuch, T.M. and E.A. Lacey, *Enhanced selection for MHC diversity in social tuco-tucos.* Evolution, 2002. **56**(4): p. 841-845.
135. Hayashi, K., et al., *Sequence variation of the DQB allele in the cetacean MHC.* Mammal study, 2003. **28**: p. 89-96.
136. Nigenda-Morales, S., et al., *MHC DQB-I Polymorphism in the Gulf of California Fin Whale (Balaenoptera physalus) Population.* Journal of Heredity, 2007. **WEISS ICH NÖD**: p. WEISS ICH NÖD.
137. Muller, H.J., *The Relation of Recombination to Mutational Advance.* Mutat Res, 1964. **106**: p. 2-9.
138. Kondrashov, A.S., *Deleterious mutations and the evolution of sexual reproduction.* Nature, 1988. **336**(6198): p. 435-40.
139. Bell, G., *The Masterpiece of Nature*. 1982, London: Croom Helm.
140. VanValen, L., *A new evolutionary law.* Evol. Theory, 1973. **3**: p. 1-30.
141. Reusch, T.B., et al., *Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism.* Nature, 2001. **414**(6861): p. 300-2.
142. Potts, W.K., C.J. Manning, and E.K. Wakeland, *The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test.* Philos Trans R Soc Lond B Biol Sci, 1994. **346**(1317): p. 369-78.
143. Hurst, J.L., *Female recognition and assessment of males through scent.* Behav Brain Res, 2009. **200**(2): p. 295-303.
144. Caroll, L., *Through the Looking Glass*. 1872, London: MacMillan.
145. Moritz, C., et al., *Parasite loads in parthenogenetic and sexual lizards (Heteronotia binoei): support for the Red Queen hypothesis.* Proc R Soc Lond B, 1991. **244**: p. 145-49.
146. Cerchio, S., et al., *Paternity in humpback whales, Megaptera novaeangliae: assessing polygyny and skew in male reproductive success.* Animal Behaviour, 2005. **70**: p. 267-277.
147. Clapham, P.J., *The social and reproductive biology of humpback whales: an ecological perspective.* Mammal Review, 1996. **26**: p. 27-49.
148. Burley, N., *Parental investment, mate choice, and mate quality.* Proc Natl Acad Sci U S A, 1977. **74**(8): p. 3476-9.
149. Wade, M.J. and S.M. Shuster, *The evolution of parental care in the context of sexual selection: a critical reassessment of parental investment theory.* Am Nat, 2002. **160**(3): p. 285-92.
150. Kokko, H. and M.D. Jennions, *Parental investment, sexual selection and sex ratios.* J Evol Biol, 2008. **21**(4): p. 919-48.
151. Bertmar, G., *Evolution of the vomeronasal organs in vertebrates.* Evolution, 1981. **35**(2): p. 359-366.
152. Darling, J.D. and M. Bérubé, *Interactions of singing humpback whales with other males.* Marine Mammal Science, 2001. **17**(3): p. 570-584.
153. Mercado, E., 3rd, L.M. Herman, and A.A. Pack, *Song copying by humpback whales: themes and variations.* Anim Cogn, 2005. **8**(2): p. 93-102.
154. Darling, J.D., M.E. Jones, and C.P. Nicklin, *Humpback whale songs: Do they organize males during the breeding season?* Behaviour, 2006. **143**: p. 1051-1101.
155. Au, W.W., et al., *Acoustic properties of humpback whale songs.* J Acoust Soc Am, 2006. **120**(2): p. 1103-10.

156. Clapham, P.J., et al., *Composition and dynamics of humpback whale competitive groups in the West Indies.* Behaviour, 1992. **122**: p. 3-4.
157. Wegner, K.M., et al., *Parasite selection for immunogenetic optimality.* Science, 2003. **301**(5638): p. 1343.
158. Eizaguirre, C., *The role of the pleiotropic genes of the major histocompatibility complex in evolution: The example of the three-spined stickleback*, in *MPI Evolutionary Ecology*. 2008, Christian-Albrechts-Universität: Kiel. p. 211.