Received: 19 May 2023 | Revised: 1 December 2023 | Accepted: 22 December 2023

DOI: 10.1111/1755-0998.13955

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

A pan-cetacean MHC amplicon sequencing panel developed and evaluated in combination with genome assemblies

Dorothea Heimeier[1] | Ellen C. Garland[2] | Franca Eichenberger[2] | Claire Garrigue[3,4] | Adriana Vella[5] | C. Scott Baker[6] | Emma L. Carroll[1]

[1]School of Biological Sciences, University of Auckland—Waipapa Taumata Rau, Auckland, New Zealand

[2]Sea Mammal Research Unit, School of Biology, University of St. Andrews, Fife, UK

[3]UMR ENTROPIE, (IRD, Université de La Réunion, Université de la Nouvelle-Calédonie, IFREMER, CNRS, Laboratoire d'Excellence—CORAIL), Nouméa, New Caledonia

[4]Opération Cétacés, Nouméa, New Caledonia

[5]Conservation Biology Research Group, Department of Biology, University of Malta, Msida, Malta

[6]Marine Mammal Institute, Hatfield Marine Science Center, Oregon State University, Corvallis, Oregon, USA

Correspondence
Dorothea Heimeier and Emma L. Carroll, School of Biological Sciences, University of Auckland – Waipapa Taumata Rau, 3A Symonds Street, Auckland 1010, New Zealand.
Email: d.heimeier@auckland.ac.nz and e.carroll@auckland.ac.nz

Ellen C. Garland, Sea Mammal Research Unit, School of Biology, University of St. Andrews, St. Andrews, Fife KY16 8LB, UK.
Email: ecg5@st-andrews.ac.uk

Funding information
Royal Society, Grant/Award Number: RGF\EA\180213, RGF\R1\181014, UF160081 and URF\R\221020; Rutherford Discovery Fellowship from the Royal Society of New Zealand Te Apārangi

Handling Editor: Joanna Kelley

Abstract

The major histocompatibility complex (MHC) is a highly polymorphic gene family that is crucial in immunity, and its diversity can be effectively used as a fitness marker for populations. Despite this, MHC remains poorly characterised in non-model species (e.g., cetaceans: whales, dolphins and porpoises) as high gene copy number variation, especially in the fast-evolving class I region, makes analyses of genomic sequences difficult. To date, only small sections of class I and IIa genes have been used to assess functional diversity in cetacean populations. Here, we undertook a systematic characterisation of the MHC class I and IIa regions in available cetacean genomes. We extracted full-length gene sequences to design pan-cetacean primers that amplified the complete exon 2 from MHC class I and IIa genes in one combined sequencing panel. We validated this panel in 19 cetacean species and described 354 alleles for both classes. Furthermore, we identified likely assembly artefacts for many MHC class I assemblies based on the presence of class I genes in the amplicon data compared to missing genes from genomes. Finally, we investigated MHC diversity using the panel in 25 humpback and 30 southern right whales, including four paternity trios for humpback whales. This revealed copy-number variable class I haplotypes in humpback whales, which is likely a common phenomenon across cetaceans. These MHC alleles will form the basis for a cetacean branch of the Immuno-Polymorphism Database (IPD-MHC), a curated resource intended to aid in the systematic compilation of MHC alleles across several species, to support conservation initiatives.

KEYWORDS
cetacean, humpback whale, major histocompatibility complex, MHC evolution, MHC organisation, southern right whale

1 | INTRODUCTION

The major histocompatibility complex (MHC) is a genetic region which encodes highly polymorphic cell surface molecules that are on the front line of pathogen detection (Klein, 1986). Antigen-presenting MHC molecules from class I and class II families bind either self- or pathogen-derived peptides within the cell that are then expressed on the cell surface. These cell surface peptides are then presented to T-cell and/or Killer cell receptors which initiate an appropriate immune response when a non-self peptide is presented (Rock et al., 2016). Class I and II molecules are similar but differ slightly in structure and function. The class I molecule is a heterodimer consisting of a transmembrane alpha chain non-covalently linked to β2-microglobulin, whereas class II molecules are heterodimers composed of α and β chains that are encoded by two separate MHC class II genes (DRA (α) and DRB (β); DQA (α) and DQB (β)) (Yeager & Hughes, 1999). Class I molecules predominantly (but by no means exclusively) bind peptides that originate from proteins in the cytoplasm and nucleus, including from replicating viruses and a few cytoplasmic bacteria. In comparison, class II molecules bind peptides from proteins located in intracellular vesicles, many of which originate from the extracellular surroundings (and thus can include many different kinds of pathogens and other antigens). Binding occurs when a pathogen protein fits in the peptide-binding groove or region of a MHC molecule, encoded by exons 2 and 3 in class I and exon 2 in class II. It is in this region where most of the polymorphic variation of an MHC gene is located and enables the molecule to bind a large number of antigens. Binding of the peptide–MHC molecule complex by a T cell receptor is followed by signalling that initiates an immune response.

Due to its role in the immune system, MHC diversity can be used as a marker or proxy for fitness to assess the ability of natural populations and species to respond to pathogens (Piertney & Oliver, 2006; Sommer, 2005). The MHC can also be used to investigate mate choice (Kamiya et al., 2014; Leclaire et al., 2019). Hundreds of studies have investigated MHC diversity and cover a large range of taxa, including birds (reviewed by O'Connor et al., 2019), mammals (e.g., Castro-Prieto et al., 2011; Gigliotti et al., 2021; Huang et al., 2019), fish (reviewed by Yamaguchi & Dijkstra, 2019) and reptiles (e.g., Pearson et al., 2017). Generally, higher MHC diversity enables a population to present a more diverse range of antigens, and hence, on an individual level, is related to higher fitness (Worley et al., 2010) and higher mating success (Kamiya et al., 2014). Hundreds, and in some species (i.e., humans), thousands, of alleles have been described for MHC genes (Robinson et al., 2017). This high number of alleles in a population is maintained by balancing selection (Radwan et al., 2020; Spurgin & Richardson, 2010). While many model species have had their MHC diversity explored, understanding MHC diversity in non-model species is essential as emerging pathogens are a threat to species in modified and impacted environments (Avila et al., 2018; Schmeller et al., 2020).

The MHC has been found in all jawed vertebrates investigated thus far and in placental mammals MHC is roughly organised into gene clusters of similar functions, including class I, class II and class III (Figure 1) (Kaufman, 2018; Kelley et al., 2005; Kumánovics et al., 2003). Some rearrangements have occurred that are lineage-specific, such as a large-scale inversion that separated class II into class IIa and IIb in cetartiodactyla (Ruan et al., 2016; Skow et al., 1996), which was first reported in bovine (*Bos taurus*) (Andersson et al., 1988). In eutherian mammals, class III genes are located between class II and class I. The latter two are bound by conserved framework genes (Abduriyim et al., 2019; Belov et al., 2006; although see Krasnec et al., 2015). The framework hypothesis (Amadou, 1999) propose that the expansion of class I genes—at least in mammals—only occurs between certain insertion points (i.e., the framework) with a varying degree of expansion between species. Within this region, there are three duplicated blocks (α, κ and β) in class I, each enclosed by a set of framework genes, which are highly conserved in gene content and order among mammals (Abduriyim et al., 2019). Within these framework genes, the class I genes have expanded and diversified between species, following

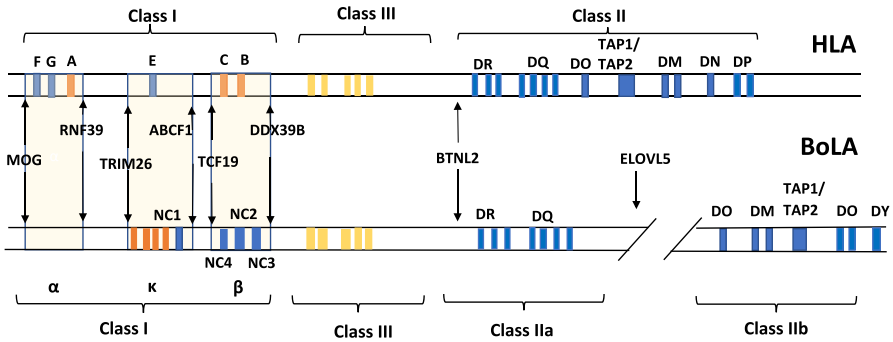

FIGURE 1 A schematic that shows the comparative organisation of the genomic MHC region of human (HLA) and bovine (BoLA). The α duplication block is between the MOG (myelin oligodendrocyte glycoprotein) and RNF39 (RING finger protein 39) genes, the κ duplication block between TRIM26 (tripartite motif containing 26) and ABCF1 (ATP-binding cassette subfamily F member 1), and the β duplication block between TCF19 (transcription factor 19) and DXX39B (DExD-box helicase 39B). BTNL2 (butyrophilin like 2) and ELOVL5 (ELOVL fatty acid elongase 5) encompass the class IIa region Orange = 'classical class I' MHC genes (Halenius et al., 2015), blue = 'non-classical' MHC genes, yellow = class III genes. Schematic not drawn to scale and not all genes are shown for simplicity.

the birth-and-death model of evolution where some loci become duplicated while others become non-functional (Nei et al., 1997). Polymorphism and evolution of the MHC gene family are believed to be shaped by the pathogen landscape an organism is exposed to, although the exact mechanisms are not yet fully understood (Bentkowski & Radwan, 2019; Manczinger et al., 2019; Prugnolle et al., 2005). As the study of MHC is being extended to a growing number of species, it is becoming more evident that the fixed class I gene number in humans is an exception and variable gene content haplotypes are more widespread than previously thought. In some taxa, such as bovine species, the number of class II and class I genes that form a haplotype can vary (Schwartz & Hammond, 2015), and therefore, investigating MHC diversity can be difficult in the absence of extensive a priori knowledge of genomic organisation and copy number.

There is a distinct lack of knowledge of MHC genomic organisation and diversity in the infraorder Cetacea, comprising whales, dolphins and porpoises. Recently, cetacean MHC class II organisation was characterised (Alves de Sá et al., 2019; Ruan et al., 2016; Zhang et al., 2019), but no comparable work has been done on MHC class I organisation, except for a single class I assembly from the Yangtze finless porpoise (*Neophocaena asiaeorientalis*) (Ruan et al., 2016). This lack of information is likely due to a combination of limited genomic resources and the increased difficulty of assembling class I due to its plasticity compared to class II (Westerdahl et al., 2022). Recently, there has been a rapid increase in genetic and genomic information in non-model species, including cetaceans (Cammen et al., 2016). Initiatives such as DNA Zoo (https://www.dnazoo.org) and the vertebrate genome project (https://vertebrategenomesproject.org) (Rhie et al., 2021) are leading this knowledge extension, with the former recently publishing their 250th genome from mostly non-model, endangered species (Dudchenko et al., 2017).

In cetaceans, MHC diversity of small segments of class II exon 2 (less than 200 bp of mostly DQB and DRB1) has been characterised with Sanger sequencing (e.g., Arbanasić et al., 2014; Heimeier et al., 2018; Villanueva-Noriega et al., 2013; Yang et al., 2012) while only a handful of studies have investigated class I diversity (e.g., Flores-Ramírez et al., 2000; Gillett et al., 2014; Xu et al., 2007). Amplicon sequencing is a well-used tool to genotype MHC diversity in individuals for model species. For example, it has been used for medical purposes in humans where MHC alleles are already known (Shortreed et al., 2020) and has been applied to non-human primates where genotyping can be done by matching amplicons to a reference database. Recent advances in next-generation sequencing offer a method to genotype MHC rapidly and accurately in non-model species (Razali et al., 2017; Rekdal et al., 2018; Stutz & Bolnick, 2014), despite the remaining challenges in identifying alleles.

Here, we use existing genome assemblies (i.e., NCBI and DNA Zoo) to explore MHC gene polymorphism and MHC gene content in cetaceans. We identified MHC framework genes and extracted MHC class I and class IIa regions from genome assemblies of 27 cetacean species. Using these genome resources, we extracted a curated set of full-length MHC genes to develop universal, pan-cetacean primers for the complete exon 2 for class IIa (DQA, DQB, DRA, DRB) and class I genes. Exon 2 was chosen instead of exon 3 because in bovine, the closest related model species, exon 2 in class I alleles is more polymorphic than exon 3 and alleles are determined by exon 2 variable sites (Heimeier, unpublished). We then tested these markers in 19 species by creating a multi-locus panel and linking amplicon data to expected sequences based on species-specific genome assemblies which allowed verification of gene content of MHC genes in these assemblies. This study provides a basis for the design and analysis of cetacean MHC diversity for future studies, particularly for evaluating MHC class I diversity.

2 | MATERIALS AND METHODS

2.1 | Comparative MHC organisation in cetaceans

2.1.1 | Cetacean genome assemblies from NCBI and DNA Zoo and MHC region extraction

We identified cetacean genome assemblies from NCBI and DNA Zoo (that were available as of May 2021, $n = 33$ usable assemblies from 27 species, Table S1). For each genome assembly, framework genes for MHC class I and IIa were mapped using the default Geneious 10.0.9 (Biomatters Ltd., NZ) mapping algorithm (see Figure 1 for framework genes). The reference framework gene sequences and MHC class I and IIa sequences were sourced from the fully annotated NCBI genome assemblies from the bottlenose dolphin (*Tursiops truncatus*) for comparisons with toothed whales, and the blue whale (*Balaenoptera musculus*) for baleen whales (hereafter called 'reference sequences'). In those assemblies for which we could identify framework genes, the region between those genes was extracted. MHC genes were confirmed within the extracted region by mapping the following class I and IIa reference genes against it: (1) full-length sequences; or (2) full coding sequences (exons only) (accession numbers included in Table S1). Secondary confirmation was made by aligning the extracted MHC class I and IIa region sequences against the reference sequences with Mauve genome aligner (Darling et al., 2004) using the progressive aligner algorithm and default settings to identify large-scale region rearrangements and inversions.

When MHC genes were found in the extracted regions, the full-length sequence was annotated according to the reference. Full-length annotated gene sequences were then extracted and checked using Geneious by aligning them to the predicted coding sequence of the appropriate reference. Functional genes were presumed and annotated on the extracted region when no stop codons or frameshift mutations were found; otherwise, genes were annotated as pseudogenes. This dataset of confirmed extracted MHC genes or pseudogenes from the genome assemblies from each species formed our 'curated dataset' for further analysis,

which was undertaken on a gene level (class II: DRA, DRB, DQA, DQB; and class I).

2.1.2 | Comparative gene phylogeny and arrangement

The curated, aligned dataset for each gene was used to build a phylogeny and examine gene arrangement across cetaceans. For the phylogeny, we downloaded the full-length bovine sequences for MHC class I and IIa genes (accession numbers included in phylogenetic tree) as an outgroup because the MHC is well understood in bovine, and it is a closely related taxa to cetaceans. The sequences were aligned against the curated dataset for its respective gene with MAFFT (Katoh et al., 2005) with default settings and allowing gaps. The alignment was exported, and a Neighbour-Joining (NJ) tree with Tamura-Nei parameters was built in MEGA 6 (Tamura et al., 2013). The method was chosen because of high similarity of sequences and faster performance compared to maximum likelihood. The newly annotated region between framework genes (i.e., DRA, DRB, DQA, DQB, class I in the order they were found in the original assembly) for each genome assembly was displayed by anchoring from the 5′ framework gene.

2.2 | Design and validation of pan-cetacean MHC amplicon sequencing approach

2.2.1 | Primer design

Using the curated dataset, suitable primers were designed with Primer3 (Untergasser et al., 2012) in conserved regions that amplified a fragment around 400 bp that included the complete exon 2 for each gene: DRA, DRB, DQA, DQB and class I. The class I primer pair in our study was designed to amplify the multiple class I genes in silico found in both the blue whale and bottlenose dolphin genome assemblies. The final primer sequences for DQA, DQB, DRA, DRB-a and class I can be found in Table S2, along with details on PCR reaction mix and conditions (Supporting Information Methods).

2.2.2 | Selection of validation dataset and library preparation

Species were selected to test the pan-cetacean MHC genotyping panel to (1) represent the broader cetacean phylogeny (McGowen et al., 2020) and (2) represent the curated dataset where possible. A total of 19 cetacean species, of which 17 species were chosen from the New Zealand Cetacean Tissue Archive (NZCeTA) housed at the University of Auckland Waipapa Taumata Rau (Thompson et al., 2013) (see Table S3 for details on all samples). Tissue samples from strandings in New Zealand were taken by the Department of Conservation New Zealand and sent to the NZCeTA with approval from mana whenua (Māori indigenous groups). Biopsy samples from New Zealand cetaceans include two Hector's dolphins (*Chephalorhyncus hectori*) (Hamner et al., 2017) and two bottlenose dolphins (*T. truncatus*) (Tezanos-Pinto et al., 2009). Further biopsies include two rough-toothed dolphins (*Steno bredanensis*) and two Blainville beaked whales (*Mesoplodon densirostris*) from French-Polynesia (Albertson et al., 2017; Oremus et al., 2012).

To extend this proof-of-concept study to cetacean populations, multiple samples from two further species, the humpback whale (*Megaptera novaeangliae*) and the southern right whale (*Eubalaena australis*) were included in the study. Thirty southern right whale biopsy samples were collected from New Zealand in 2020 (Carroll et al., 2022). Twenty-five humpback whale biopsy samples were included from the New Caledonian breeding ground in the South Pacific (Derville et al., 2019; Garrigue et al., 2001). Four complete humpback whale paternity trios (comprised of offspring, mother and candidate father; Eichenberger et al., 2022) were part of this sample set to investigate Mendelian inheritance patterns.

DNA extractions were performed with either standard phenol–chloroform extraction and ethanol precipitation methods (Sambrook et al., 1989), as modified for small tissue samples (Baker et al., 1994) or by DNAeasy kit (Qiagen). Genomic integrity was checked on a 0.8% agarose gel and concentration of DNA was standardised to 50 ng/μL measured by nanodrop and verified by agarose gel before PCR. Class IIa genes DRA, DRB-a, DQA, DQB and class I were amplified as outlined in Table S2 and PCR products were run out on a 1.5% agarose gel to check amplification success and fragment size. The sizes of inserts without primers ranged from 318 to 405 bp and in total spanned the complete exon 2 sequence of each of the genes. A total of 5 μL of each amplicon was pooled for each of the 96 individuals and purified with 25 μL of Ampure beads according to the manufacturer's protocol with slight modifications. Specifically, an additional ethanol wash step was performed, and the PCR product was eluted with 32 μL Ultra-pure water. The concentration of the elute was measured by Qubit and samples were diluted to 5 ng/μL.

Individual samples were indexed by Auckland Genomics (University of Auckland) with Nextera indexes supplied by IDT (San Diego, USA). The Illumina metabarcoding protocol for indexing (16S Metagenomic Sequencing Library Preparation, Illumina) was followed using Platinum SuperFi Taq (Thermofisher) for the amplification. Library quality and concentration of a random subset of 12 amplicons were checked with a High Sensitivity DNA Assay on the Bioanalyzer 2100 (Agilent). The library was sequenced on an Illumina MiSeq platform (NanoSeq) by Auckland Genomics using 2 × 250 cycle paired-end kit. To increase read number, the whole library was subjected to a second NanoSeq run using the same protocol. Additionally, a subset of samples which included 25 humpback whale samples and 18 samples from other cetacean species were sequenced again with adjusted conditions (28 PCR cycles) for amplicons DRB-a, DQB and class I, and sequenced as part of a full MiSeq run (repeat samples marked in Figure S1). Genotyping was performed using reads from all available sequencing runs per sample, following technical replicate and reproducibility analyses (see below).

2.2.3 | MHC genotyping pipeline

We have adapted the MHC genotyping pipeline used in AMPLISAS (Alvaro et al., 2016) for use in Geneious. Step-by-step analyses are listed below, but important differences are that reads for each individual were processed separately, and mapping to a target gene instead of sequence demultiplexing by primer was used to separate reads from different amplicons. The pipeline's threshold between dominant (likely to be 'true' alleles) and artefact sequences for class I was empirically decided for each sample based on the degree of change (DoC) curve (see below) (as described by Lighten et al., 2014). The pipeline is summarised in Figure 2. First, Illumina sequencing data that was provided as demultiplexed reads, already trimmed of sequencing adapter and index sequences, was imported into Geneious. Initial quality control (QC) involved trimming both paired end (PE) reads to remove bases with >0.01 error probability. PE reads were then merged with BBMerge using default settings (Bushnell et al., 2017). After merging, each read was mapped to individual full-length MHC class I and class IIa genes extracted from the NCBI reference from either the blue whale or the bottlenose dolphin, for baleen whales and toothed whales, respectively. Mapping proceeded with the inbuilt Geneious mapper, with a minimum overlap identity of 85% (90% for DRB-a) and allowing gaps.

After mapping to the respective gene reference, each gene was analysed separately, and the primer sequences were then trimmed from both ends of the mapped reads with Geneious. For each sample, reads were de novo assembled to form clusters of highly similar reads with 100% identity and a minimum of 200 bp overlap. Clusters were sorted in descending order of number of reads per sample (read depth) and a consensus sequence was created for each top cluster. Top clusters were selected until the subsequent cluster had a significant drop-off in read depth determined by a degree of change (DoC) in the cumulative percentage of reads (vertical blue line: Figure S3). For class IIa, a minimum of 10 reads per cluster was deemed sufficient to be included in the analysis, since those identified allele sequences could be confirmed with previously published alleles (see Class IIa genes, Table S8). For class I, if no DoC break could be established, only clusters with more than 500 reads were included. Also, samples with less than 500 reads for the top cluster were excluded from further analysis. The name of each consensus sequence, representing a potential allele per sample, retained the sample ID, cluster number and read depth of the cluster from which the consensus sequence originated. Consensus sequences for each

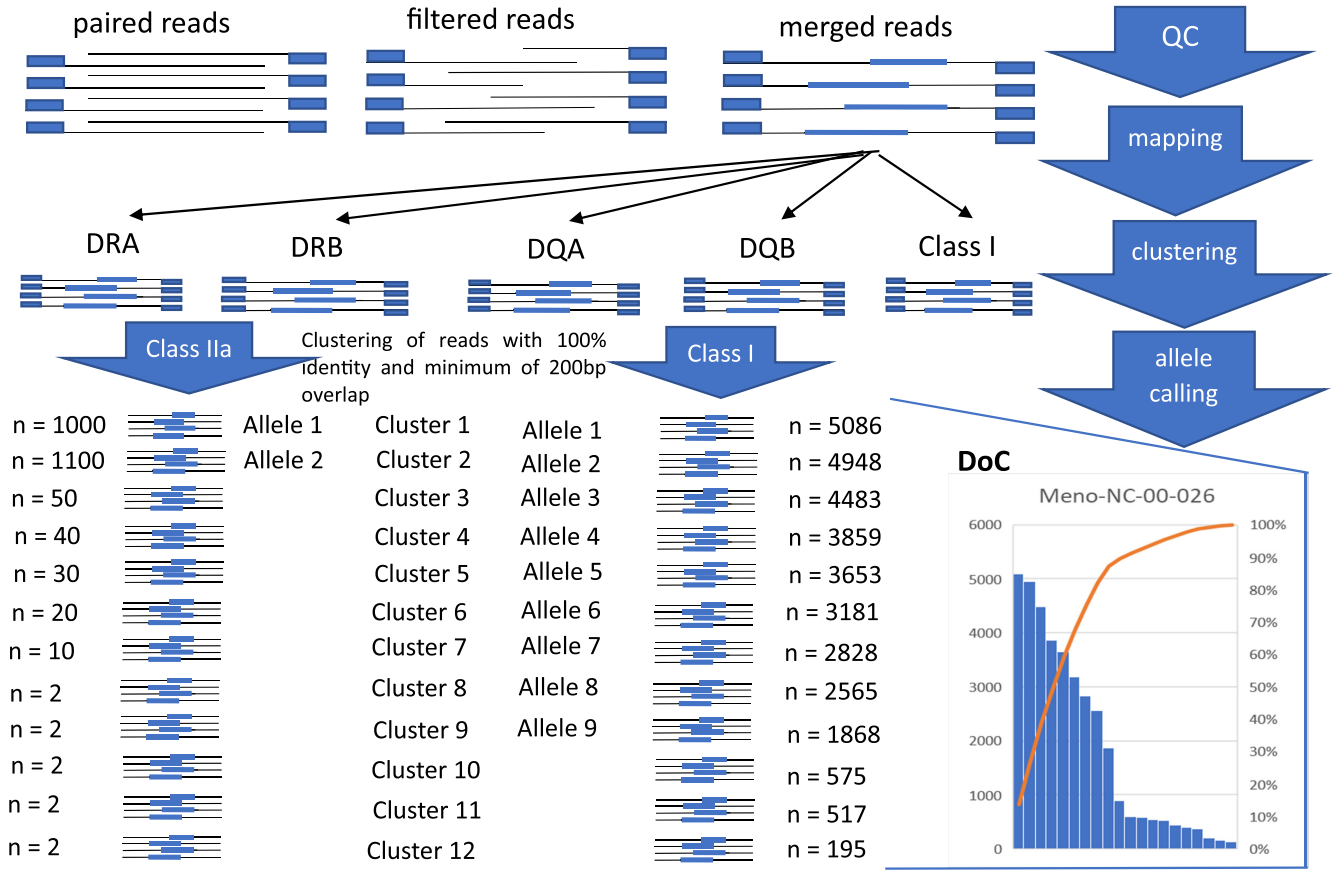

FIGURE 2 Schematic of workflow from quality control (QC), mapping to each gene and clustering reads to allele assignment. The read numbers for clusters are examples typical for class II and class I. The degree of change (DoC) graph shows typical cluster read numbers for class I and the drop-off in cluster number as a bend in the cumulative percentage curve.

gene, across all individuals, were aligned in MAFFT. Potential alleles were reconciled per species to give a list of unique alleles per gene (termed the 'curated allele list', Table S8).

Functionality was assumed if predicted CDS was in reading frame with no stop codons. Alleles were given a name that consisted of the four-letter species abbreviation followed by the gene name, double asterisks (**) and then consecutive numbers in no particular order.

2.2.4 | Replication and reproducibility analyses

To have confidence in the allele designations, we assessed the reproducibility of the genotyping pipeline with one technical plate containing 94 samples, eight of which were technical sample replicates (same genomic DNA). Six of the technical sample replicates were amplified with different PCR cycle numbers. A further 42 samples were subject to a third technical replicate that also served to increase read numbers for these samples.

The inclusion of paternity trios further allowed us to check the validity of allele identification by assessing the Mendelian inheritance of MHC genes. This is especially valuable for MHC genes with more than one locus (e.g., class I genes). Haplotypes were manually inferred for the four humpback whale trios included in this study based on allele occurrence.

2.2.5 | Comparison with published data

On GenBank, we used the search terms of the individual gene names and 'cetacea' and downloaded all available results (as of August 2022). For each gene, sequences with exon 2 were retained and aligned with MAFFT with the full-length and amplicon allele list and trimmed to the amplicon length. The alignment also included the homologous bovine MHC genes. The alignment was imported into MEGA and a Neighbour-Joining tree using Tamura-Nei parameters was created.

We were unable to determine genomic origin (κ or β block) from the class I exon 2 amplicon using a phylogenetic approach. Instead, we mapped the putative alleles for each species back to the phylogenetically closest genome assembly that contained all κ, β and middle-class I loci. Mapping was done in Geneious for all genes at both, 95% and 92% identity.

3 | RESULTS

3.1 | Comparative MHC organisation in cetaceans

From 28 available and usable cetacean genome assemblies, 26 species provided class I and 21 provided class IIa regions, enclosed by framework genes. The genomes represented good coverage across the phylogenetic lineage of cetaceans: 20 toothed whales across six

TABLE 1 Overview and summary table of species used in this study and diversity at MHC loci in each.

Species information						MHC class II results				MHC class I results						
						Total (max) allele number				Total (max)	κ-1			Short κ	middle	β
	Family	Common name	Scientific name	Code	Assembly	*n*	DRA	DRB-a	DQA	DQB		a	b	c			
Parv.																	
Baleen whale	Eschrichtiidae	Grey whale	*Eschrichtius robustus*	*Esro*	DNA Zoo												
	Balaenopteridae	Blue whale	*Balaenoptera musculus*	*Bamu*	NCBI												
	Balaenopteridae	Eden's whale	*Balaenoptera edeni*	*Baed*		2	2 (2)	2 (2)	1 (1)	2 (2)	6 (6)				[1][b]		[1][b]
	Balaenopteridae	Humpback whale	*Megaptera novaeangliae*	*Meno*	DNA Zoo	26	1 (1)	12 (2)	1 (1)	14 (2)	33 (9)						
	Balaenopteridaea	Minke whale	*Balaenoptera acutorostrata*	*Baac*	NCBI	2	2 (2)	1 (1)	0 (0)	0 (0)							
	Balaenopteridaea	Pygmy blue whale	*Balaenoptera musculus brevicauda*	*Bamu*		1	1 (1)	1 (1)	2 (2)	2 (2)	7 (7)						
	Balaenopteridaea	Rice's whale	*Balaenoptera ricei*	*Bari*	DNA Zoo												
	Balaenidae	North Atlantic right whale	*Eubalaena glacialis*	*Eugl*	DNA Zoo												
	Balaenidae	Southern right whale	*Eubalaena australis*	*Euau*	DNA Zoo	30	2 (1)	6 (2)	2 (2)	15 (2)	32 (6)				2		

TABLE 1 (Continued)

Species information							MHC class II results				MHC class I results				
							Total (max) allele number				Total (max)	κ-1	Short κ	middle	β
Toothed whales	Physeteridae	Sperm whale	*Physeter macrocephalus*	*Phca*	DNA Zoo										
	Kogiidae	Dwarf sperm whale	*Kogia sima*	*Kosi*		1	1 (1)	1 (1)	2 (2)	2 (2)	2 (2)				
	Kogiidae	Pygmy sperm whale	*Kogia breviceps*	*Kobr*		1	1 (2)	0 (0)	2 (2)	0 (0)	2 (2)				
	Ziphiidae	Blainville's beaked whale	*Mesoplodon densirostris*	*Mede*	DNA Zoo	2	2 (2)	2 (2)	2 (2)	4 (2)	7 (4)				
	Ziphiidae	Cuvier's beaked whale	*Ziphius cavirostris*	Zica		2	1 (1)	3 (2)	2 (2)	2 (2)	9 (5)				
	Ziphiidae	Gervais' beaked whale	*Mesoplodon europaeus*	*Meeu*		1	1 (1)	2 (2)	2 (2)	0 (0)	5 (5)				
	Ziphiidae	Grey's beaked whale	*Mesoplodon grayi*	*Megr*		2	1 (1)	3 (2)	2 (2)	3 (2)	4 (4)				
	Ziphiidae	Stejneger's beaked whale	*Mesoplodon stejnegeri*	*Mest*	DNA Zoo										
	Lipotidae	Baji	*Lipotes vexillifer*	*Live*	NCBI										
	Monodontidae	Beluga	*Delphinapterus leucas*	*Dele*	DNAZoo										
	Monodontidae	Narwhal	*Monodon monoceros*	*Momo*	NCBI										
	Phocoenidae	Harbour porpoise	*Phocoena phocoena*	*Phoh*	DNAZoo										
	Phocoenidae	Vaquita	*Phocoena sinus*	*Phsi*	NCBI										
	Phocoenidae	Yangtze finless porpoise	*Neophocaena asiaeorientalis*	*Neas*	NCBI										
	Delphinidae	Atlantic spotted dolphin	*Stenella frontalis*	*Stfr*	DNA Zoo										
	Delphinidae	Atlantic white-sided dolphin	*Lagenorhynchus acutus*	*Laac*	DNA Zoo										
	Delphinidae	Bottlenose dolphin	*Tursiops truncatus*	*Tutr*	DNA Zoo										
	Delphinidae	Bottlenose dolphin	*Tursiops truncatus*	*Tutr*	NCBI	3	3 (2)	3 (2)	3 (2)	4 (2)	19 (9)				
	Delphinidae	Commerson's dolphin	*Cephalorhynchus commersonii*	*Ceco*	DNA Zoo										
	Delphinidae	Dusky dolphin	*Lagenorhynchus obscurus*	*Laob*		2	1 (1)	1 (1)	2 (2)	1 (1)	11 (7)				
	Delphinidae	Hector's dolphin	*Cephalorhynchus hectori*	*Cehec*		2	2 (2)	1 (1)	2 (2)	2 (2)	3 (3)				
	Delphinidae	Indo-Pacific bottlenose dolphin	*Tursiops aduncus*	*Tuad*	DNA Zoo										
	Delphinidae	Indo-Pacific humpbacked-dolphin	*Sousa chinensis*	*Soch*	NCBI										
	Delphinidae	Killer whale	*Orcinus orca*	*Oror*	DNA Zoo										
	Delphinidae	Long-finned pilot whale	*Globicephala melas*	*Glme*	DNA Zoo	2	2 (2)	3 (2)	1 (1)	1 (1)	10 (8)				
	Delphinidae	Pacific white-sided dolphin	*Sagmatias obliquidens*	*Saob*	DNA Zoo										
	Delphinidae	Pantropical spotted dolphin	*Stenella attenuata*	*Stat*	DNA Zoo										
	Delphinidae	Risso's dolphin	*Grampus griseus*	*Grgr*	DNA Zoo	2	1 (1)	3 (2)	2 (2)	4 (2)	10 (7)				
	Delphinidae	Rough-toothed dolphin	*Steno bredanensis*	*Stbr*		2	1 (1)	3 (2)	1 (1)	3 (2)	14 (7)				
	Delphinidae	Striped dolphin	*Stenella coeruleoalba*	*Stco*		1	1 (1)	0 (0)	1 (1)	1 (1)	5 (5)				
	Delphinidae	Short-beaked common dolphin	*Delphinus delphis*	*Dede*		1	2 (2)	2 (2)	2 (2)	2 (2)	7 (7)				
	Total[a]					85	28	49	32	62	186				

Note: Shown is each the parvorder (Parv.), family, common name, scientific name, and four-letter codes used in phylogenetic trees and for allele names for each species. Where a genome assembly is available for a species, we show the origin (assembly) and where a species was used in the amplicon panel, we show the sample size (n). For class IIa genes we present the amplicon panel results showing total number of alleles across all samples per species and gene for exon 2 and the maximum number of alleles found per individual in parentheses next to it (max). For class I, we present the amplicon panel results showing total number of alleles across all samples per species (Total) and the maximum number of alleles found per individual in parentheses next to it across all genes (max). These were not able to be localised to class I gene region in most cases, except where the sequence matched the reference genome. Shading: blue indicates presence of gene in assembly; grey indicates the region (dark grey) or gene (light grey) were not in the assembly; no colour in this section indicates no available reference genome. Note that *DRBp* and DRB-b were not listed as they were not amplified.

[a]Includes alleles shared between species.

[b]Matched the genome sequence of Rice's whale.

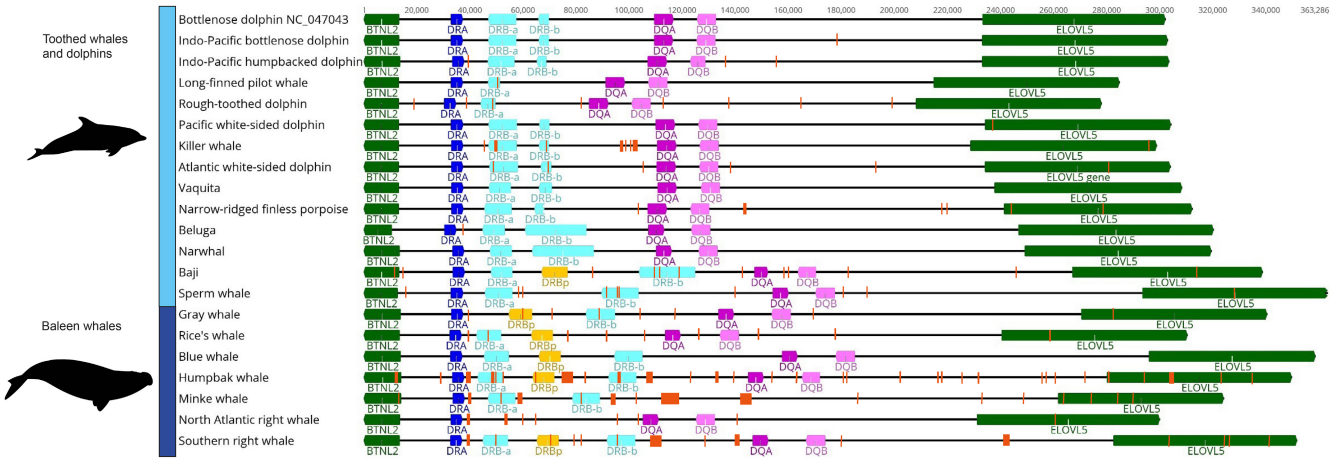

FIGURE 3 MHC class IIa regions (not aligned) with annotated MHC genes and framework genes (green) for each for which a class IIa region assembly was available. Species name can be found on the left, length of the region is indicated in kilobases (kb); see Table S1 for further details including accession numbers. Toothed whales shaded in light blue and baleen whales in dark blue. Assembly gaps are marked by a red line, not drawn to scale and indicative only.

of 10 extant families, and seven baleen whales across three of four extant families (Table 1).

3.1.1 | Class IIa

A total of 21 genome assemblies contained class IIa framework genes encompassing the class IIa genes DRA, DRB, DQA, DQB (Table 1, Figure 3; Table S1 for accession numbers). The MHC class IIa region was similar in length across toothed whales at about 300 kb (including the framework genes BTNL2 and ELOVL5), whereas the region was about 10–50 kb longer in baleen whales (except for Rice's whale, *Balaenoptera ricei*, and the North Atlantic right whale, *Eubalaena glacialis*). The MHC class IIa region was well conserved across the cetacean genome assemblies investigated, containing one DRA, one DQA and one DQB gene each, except for the common minke whale (*B. acutorostrata*) assembly which was missing the DQA-DQB genes and coincided with an assembly gap, as noted previously (Alves de Sá et al., 2019). We observed variation in gene copy number in DRB with between one and three loci per species. Most Delphinidae had two DRB copies (DRB-a and DRB-b, named in order of appearance). The exceptions were the rough-toothed dolphin and the long-finned pilot whale (*Globicephala melas*), which had shortened class IIa regions with an incomplete DRB-a gene and no DRB-b gene; while primers did not bind to these genome assemblies in silico they did amplify and produce DRB-a exon 2 sequences (see below).

Most baleen whales had DRB-a and -b, and an additional DRB pseudogene (DRB-p; as was annotated in the blue whale genome assembly; yellow in Figure 3). This pseudogene was also found in the Baji (*Lipotes vexillifer*). The full-length gene DRA, DQA and DQB phylogenies formed well-supported clades separating cetacean parvorders (Table 1, Figure 4a,c,d). The DRB phylogeny showed distinct clades with high bootstrap support for the two gene loci (DRB-a and DRB-b; Figure 4b), while pseudogenes DRB-p formed their own clade. Within the gene-specific clades, baleen whales and toothed whales were broadly separated.

3.1.2 | Class I

A total of 26 genome assemblies from 25 species contained at least one of the two outer framework genes of the class I region (Figure 5) and were examined further. Eight of these 26 assemblies did not contain any class I genes (Figure 5, red labels). One bottlenose dolphin assembly (HiC; GCF_001922835.1) had no class I genes, while in contrast, another conspecific assembly (NC_047043) had five class I genes. It is of note that the assemblies differ extensively in completeness, as can be observed based on the assembly gaps (numerous vs none, respectively). There was also considerable variation in length of the class I region (between 450 and 800 kb) that did not seem specific to particular taxa.

In genome assemblies with identified class I genes, there was variation in gene copy number (Figure 5). In the case of the κ class I genes, the bottlenose dolphin had up to three copies, the blue whale and Yangtze finless porpoise two and all others only one. In the case of the class I gene from β block, all assemblies that contained a gene only contained one copy. We also found one additional class I gene outside the κ and β blocks in eight assemblies (labelled here as 'middle class I'; Figure 5). As above, the presence or absence of class I genes did not appear to be specific to particular taxa.

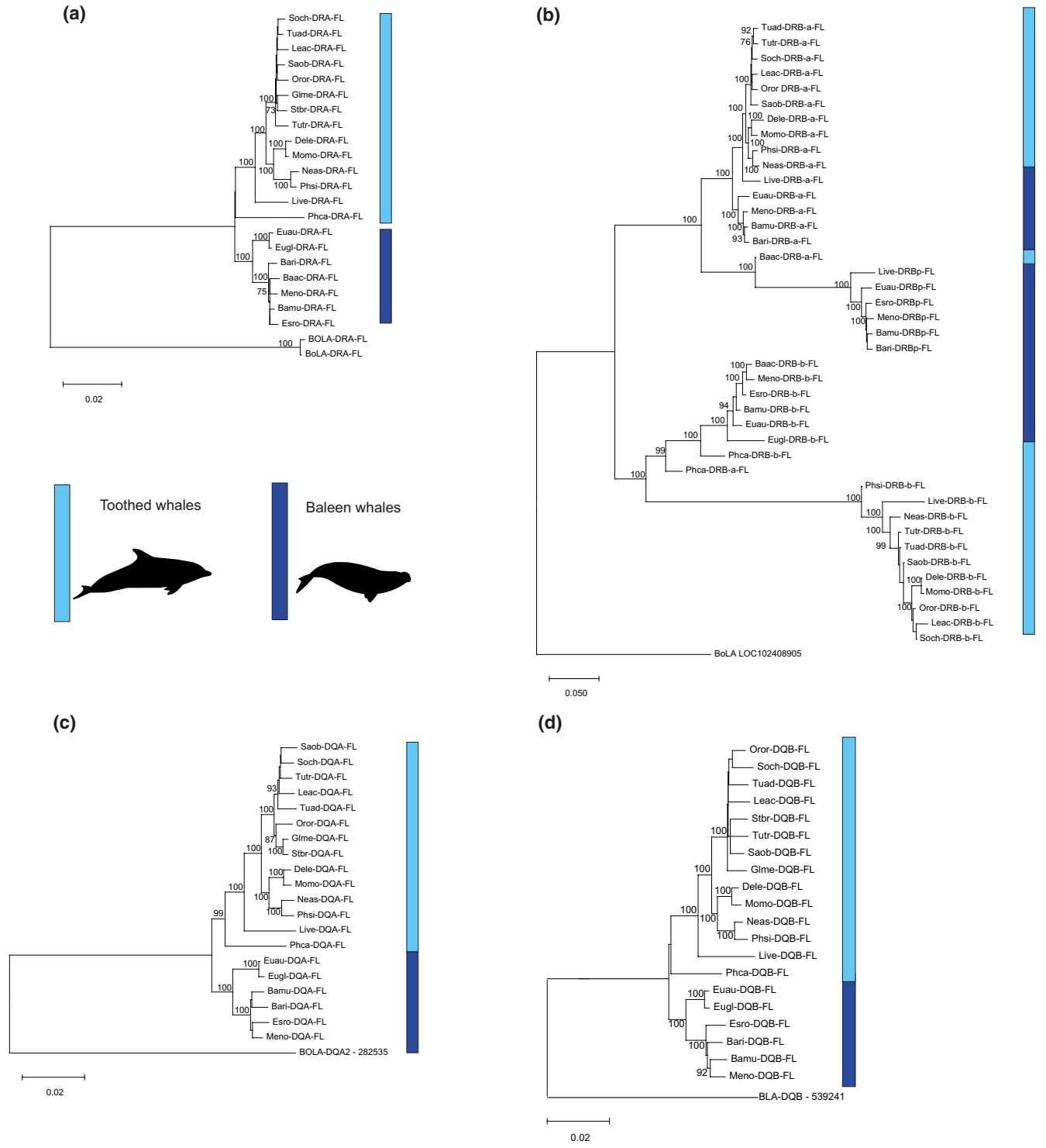

FIGURE 4 Neighbour-Joining phylogenetic trees for full-length (FL) genes for (a) DRA, (b) DRB, (c) DQA and (d) DQB. Bootstrap support over 80% is shown. The evolutionary distances were computed using the Tamura-Nei method and are in the units of the number of base substitutions per site. Light blue = toothed whales and dolphins; dark blue = baleen whales.

A total of 41 full-length class I gene sequences were extracted from the assemblies and formed well-supported clades correlating to their class I block as well as their parvorder (Figure 6). Interestingly, the κ class I-b (s) from the blue whale clustered strongly with the other baleen whale κ class I sequences, but the blue whale κ class I-a and bottlenose dolphin κ class I-a, -b and -c clustered with the odontocete κ class I sequences.

3.2 | Validation of pan-cetacean MHC amplicon sequencing approach

3.2.1 | Library sequencing success

We simultaneously genotyped exon 2 of five MHC genes in 94 individual samples from 19 different cetacean species using amplicon

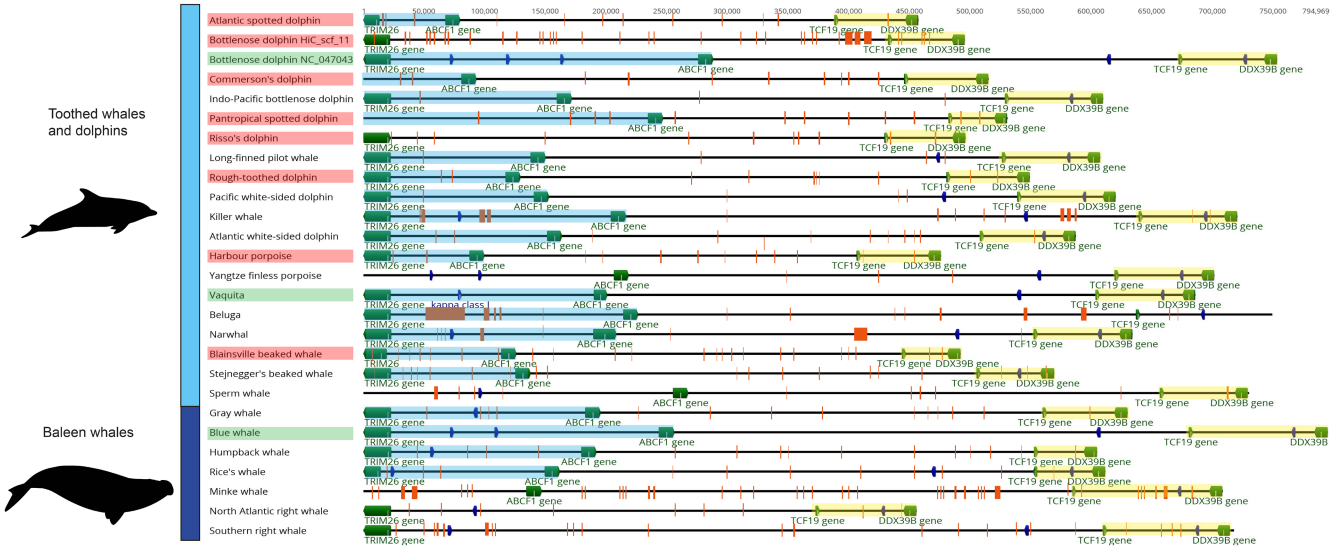

FIGURE 5 MHC class I region with duplication blocks from κ to β (not aligned) with annotated class I genes (blue) and framework genes (green) for each assembly for which a class I region was available. Species name can be found on the left, length of the region is indicated in kilobases (kb). Assemblies with class I regions that contain no class I genes are shaded in red, kappa block shaded blue and beta block shaded yellow. Toothed whales and dolphins shaded in light blue and baleen whales in dark blue. Assembly gaps are marked by a red line, not drawn to scale and indicative only. Assemblies without gaps are shaded in green. See Table S1 for further details including accession numbers.

sequencing on the Illumina MiSeq platform. Across all combined sequencing runs and combined replicate samples, we generated a total of 1,844,808 QC reads with an average of 21,704 reads per individual sample (SD = 15,963; max = 65,020, min = 700) (Table S4). Only one sample failed with less than 200 reads, and after accounting for technical sample replicates, we had data from 85 unique individuals. Reproducibility was 100%, as all replicate samples resolved to the same alleles after analysis.

Reads were successfully separated into amplicons by mapping to a class IIa or class I gene with no or only a very small number of unmapped reads (2%; Table S4). The majority of reads (62%) mapped to class I (Table S4), which was likely to be the result of one or more of the following: (1) higher concentration of the PCR amplicon for class I after 30 cycles compared to 25 cycles for class IIa; (2) preferential sequencing of smaller class I amplicons (also observed in DRA/DQA vs DRB/DQB) and (3) pooling in non-equimolar ratios of amplicons.

3.2.2 | Genotyping pipeline and allele summary

Class IIa genes

For all class IIa genes, between 85% and 90% of mapped reads grouped into clusters (Table S4), with the majority of reads per individual clustered into the top one or two clusters (Figure S1). Heterozygous individuals had a similar read depth of their top two clusters (alleles). A total of 171 alleles were found across all class IIa genes from 85 individual cetaceans including some alleles (DRA and DQA only) shared by two or more species; read depth information is in Table S5.

We found the highest number of alleles for DQB ($n = 62$) and a slightly smaller number for DRB-a ($n = 49$). The number of DRA ($n = 28$) and DQA ($n = 32$) alleles was about half that of DQB, and their nucleotide sequence was also more similar to each other than DRB-a and DQB alleles to each other (Table 1 and Table S5). Two DRA and two DQA alleles were shared across six and three species, respectively (Tables S7 and S8). Allele sharing was only found within families. Low diversity for DRA and DQA alleles was also found within the two studied populations, with both DRA and DQA alleles being monomorphic in 25 humpback whales and DRA in 30 southern right whales. All DRA, DQA and DQB alleles were in reading frame and assumed to be functional, as was DRB-a in all toothed whales. Predicting functionality from DRB-a alleles in baleen whales was ambiguous. When we used the same reading frame as for toothed whales, all DRB-a alleles were predicted to be non-functional with multiple stop codons. An alternative reading transferred from BoLA-DRB2 (E-S beta chain isoform X2), predicted all baleen whales DRB-a alleles to be functional, except for all southern right whale alleles.

Class I genes

For class I, an average of 80% of mapped reads grouped into clusters (Table S4) with between two and eleven clusters as top clusters per individual (mean = 6.2 + −2.2) (Figures S2 and S3). A total of 183 alleles were found across all class I genes from 75 individual cetaceans, with three shared alleles between Delphinidae species (Table 1, Table S8). Between two and nine potential alleles were found per individual across all cetacean species investigated (Table 1). The read depth of alleles differed substantially within and between individuals from an average of 3340 for the top cluster to an average of 860 reads for the lowest cluster.

For baleen whales (except for humpback whales and southern right whales), only two individuals (the pygmy blue whale (*Balaenoptera musculus brevicauda*) and Eden's whale (*Balaenoptera*

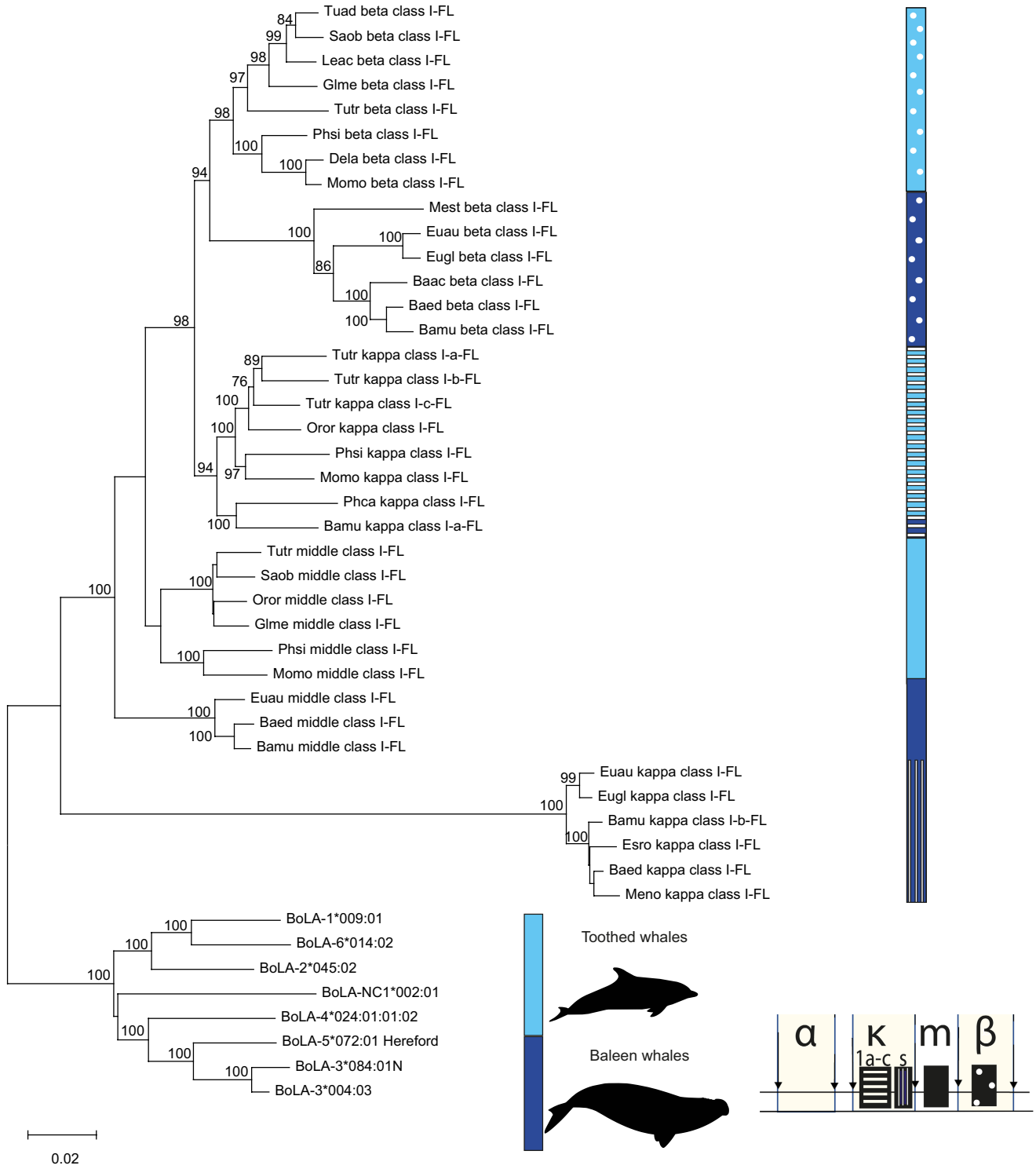

FIGURE 6 Neighbour-Joining phylogenetic tree with cetacean full-length (FL) sequences for class I genes from genomic assemblies rooted with bovine (BoLA) gene sequences. Bootstrap support over 80% is shown. The evolutionary distances were computed using the Tamura-Nei method and are in the units of the number of base substitutions per site. Gene accession numbers can be found in Table S1. Light blue indicates sequences from toothed whales, and dark blue from baleen whales. Patterns refer to location of genes in the MHC class I region (see bottom right schematic); dotted = beta; plain = middle; horizontal stripes = kappa and vertical stripes = kappa (short class I gene (s), present only in baleen whales).

edeni)) had enough reads for class I analysis (Table 1 and Figure S3). For toothed whales, 105 alleles were identified from 22 individuals. A total of 33 alleles were found across the 25 humpback whales with an average of eight alleles per individual at class I. In humpback whales, all alleles, except for five (termed singletons) were found in at least two individuals. Of the remaining 28 alleles, two were found in each of the 25 humpback whale samples. Further, five alleles were shorter in length than all other alleles (318 and 330 vs. 341 bp). Thirty southern right whales had a total of 32 alleles with an average of five alleles per individual. Thirteen alleles were singletons. Like the humpback whale, of the remaining 19 alleles, we found four alleles that were present in the majority of southern right whales. Two alleles were shorter in length than the other alleles, as found in the humpback whale. All alleles, except for four (Euau_class I_N**03:01 and 02; Meno_class I_N** 24:01 and 02) were in reading frame and therefore presumed to be functional.

3.2.3 | Comparison of identified alleles to published data (assemblies and GenBank entries)

Class IIa genes

All class IIa genes (DRA, DRB-a, DQA and DQB) were successfully amplified from each species in our study with a few exceptions (Table S7). DRB-a was not amplified in the PCR for striped dolphin (*S. coeruleoalba*) and pygmy sperm whale (*Kogia breviceps*). DQA and DQB did not amplify for the minke whale. Alleles for all examined class IIa genes were limited to two alleles per individual. This is consistent with the finding of only one locus for each gene in all cetacean genome assemblies.

Across all 171 class IIa alleles, we found a total of 56 alleles in our curated allele list that had been previously described on GenBank and/or were present in the genome assemblies (see Table S8 for a complete list). When these were included in an amplicon-length phylogeny (Figure S5a), only DRA maintained the clustering of cetacean families as found in the full-length phylogeny (Figure 4a). In the DRB amplicon-length phylogeny (Figure S5b), DRB-a sequences from baleen whales still clustered together, but the genomic origin (DRB-a or DRB-b) cannot be established from the phylogenetic tree of the amplicon (including only exon 2). In the DQA amplicon-length phylogeny (Figure S5c) each of the cetacean families is still clustered together. In the DQB amplicon-length phylogeny (Figure S5d) all clustering of cetacean families, even between baleen and toothed whales was lost.

Class I genes

Class I sequences were successfully amplified from all species in our amplicon panel study, except for the minke whale, due to low read numbers. This seems to be in contrast to some assemblies in which no class I gene was found to be present (i.e. Risso's dolphin, Rough-toothed dolphin and Blainsville's beaked whale) and could indicate towards collapsed assemblies or other assembly issues. In silico, the class I primer pair annealed to all three class I genes (κ, β and middle); each class I gene clustered together in the full-length gene phylogeny (Figure 6). This structure however was not resolved in the amplicon phylogeny (not shown), except for the mysticete κ class I sequences (κ class I-b (s)) which were still clustering together with high bootstrap support. Amplicon sequences from this locus were shorter and different enough in sequence to be allocated to their gene origin in the genome (Figure 6). Since our aim was to validate the gene content of genome assemblies with amplicons, it was desirable to place amplicons within their gene of origin. So, instead of generating a phylogenetic tree, we mapped the potential alleles at 95% and 92% identity from a species to the closest available genomic assembly that had a complete set of class I genes to predict its most likely genomic origin (Table 1, Table S6). We used the blue whale, Rice's whale and southern right whale for baleen whales; the bottlenose dolphin for dolphins and at the time of analysis no genome was available for use as reference for dwarf and pygmy sperm whale and the beaked whales.

Across all 183 class I alleles identified, we found five previously described alleles from GenBank and seven alleles on genome assemblies. The shorter class I alleles (318 bp) from baleen whales all mapped to κ-class Ib genes of the blue whale, Rice's whale and southern right whale assembly. Three alleles were identical to genomic genes: two Eden's whale class I alleles matched to κ-class Ib and β class I genes of Rice's whale assembly; a southern right whale allele to a southern right whale κ-class Ib gene. Furthermore, more than 14 alleles were only 1–3 bp different to a gene in an assembly (light green shading in Table S6).

3.3 | Validation of alleles and haplotype inference through Mendelian inheritance patterns

The inclusion of four paternity trios (mother–putative father–offspring) from humpback whales allowed us to verify our allele assignment by determining haplotypes based on Mendelian inheritance patterns. From allele occurrence in paternity trios (Figure 7), we assigned class IIa and class I haplotypes for those paternity trios. We did not include genes with low variability. The offspring in paternity trio C had a very low read count for class I, so we matched all reads from this individual to the class I allele list and could identify two additional alleles at a low read count of 40 and 47. Class I haplotypes in all trios contained between one and three presumed class I alleles and the same allele was found on different haplotypes (i.e.: I_N**03:01, I_N**17, I_N**19 and I_N**20). A total of 14 class IIa haplotypes and 12 class I haplotypes were identified in those 12 humpback whales. The pattern of Mendelian inheritance was upheld in three paternity trios, but the offspring in trio D did not match one of the haplotypes from the candidate father from class IIa or class I. However, the paternity trio D was only assigned at an 80% confidence level in Cervus 3.0.7 (Kalinowski et al., 2007) and further showed one mismatch at a microsatellite locus compared to all other paternity trios that were assigned at the highest confidence level (95%) with no mismatches (Eichenberger et al., 2022). The paternity assignment of trio D could, therefore, be a false paternity assignment (type I error). This would explain the unmatching haplotypes

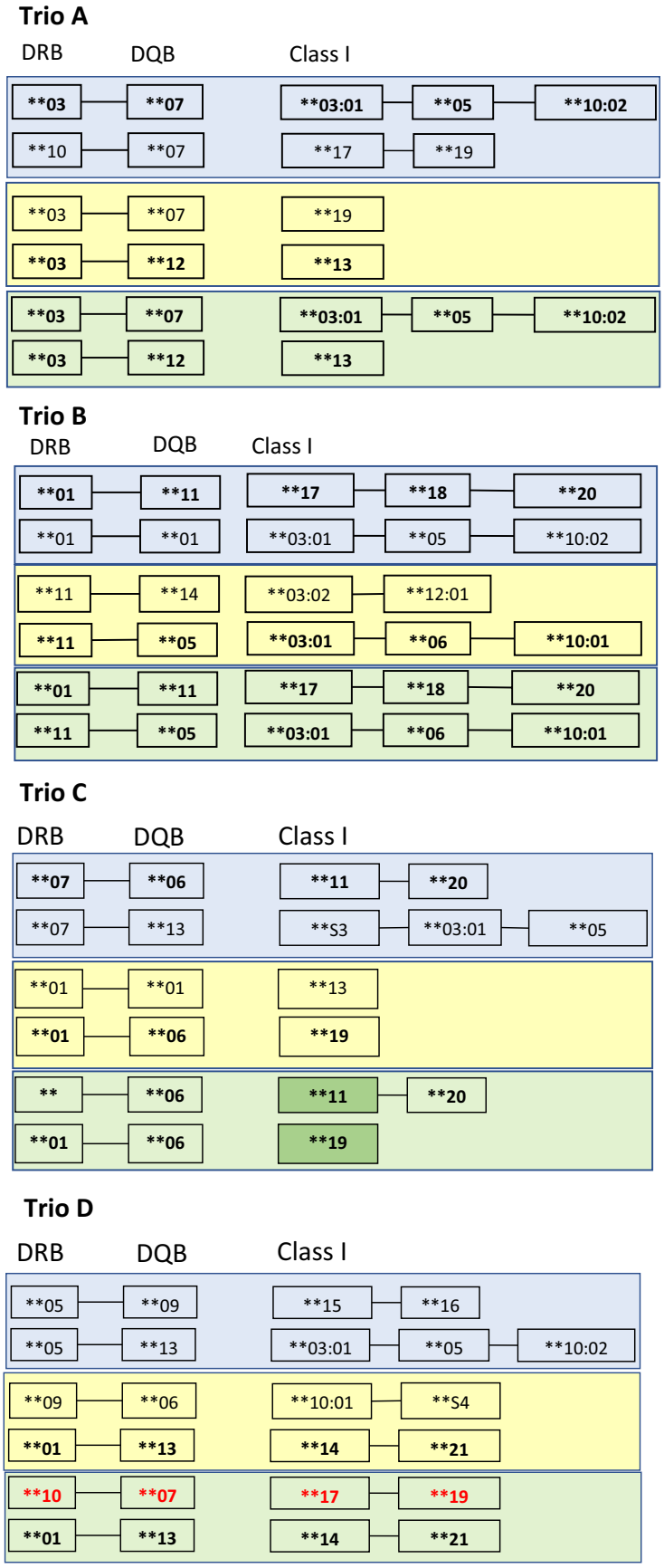

FIGURE 7 Class IIa (DRB-a, DQB) and class I haplotypes for four presumed paternity trios from humpback whales (Eichenberger et al., 2022). Allele designation starts with ** and the number given to the allele. Blue boxes contain the father's haplotypes, yellow boxes the mother's and green boxes the offspring's haplotype. The haplotype in bold has been passed on to the offspring. Red indicates a mismatch in haplotype in the offspring, suggesting the paternity assignment may be incorrect. Dark green shaded boxes mean that the alleles were only found after matching all reads of this individual to the curated allele list—or in other words, those alleles were only present in the allele clusters after the DoC cut-off.

from class IIa and class I. The pattern of Mendelian inheritance was upheld in all known mother–offspring pairs of all paternity trios. The class I haplotypes we found in the trios were found six times in six other humpback whales. We were not able to establish more potential class I haplotypes with confidence due to varying gene content with shared alleles between haplotypes.

4 | DISCUSSION

Here we presented a pan-cetacean MHC genotyping panel that provided consistent results across species. We compared the results from our proof-of-concept study to the gene content of available genome assemblies for cetaceans. This provides an exciting basis for future work in assessing the past, current and future functional diversity of cetacean species. This is important from an evolutionary standpoint, but also for conservation management; as of December 2020, 25% of cetaceans are considered threatened and a further 10% data deficient (https://iwc.int/management-and-conservation/cetaceans-and-extinction). Information on this functional marker will provide insights into the capacity of these species to respond to emerging diseases (e.g.: Coker et al., 2023).

4.1 | MHC organisation in cetaceans in comparison with amplicon data

4.1.1 | MHC class IIa

Here we have built on recent work characterising the cetacean MHC class IIa region (Alves de Sá et al., 2019) and extended it to an assessment of the class I region with a larger dataset spanning almost 80% of cetacean families. In agreement with previous work (Alves de Sá et al., 2019), our comparative genomic analyses found that the class IIa region is highly conserved, with most cetaceans having one DRA and a single DQA and DQB gene. This was further confirmed by our amplicon sequencing, which found two alleles at DRA, DQA and DQB across almost all the cetacean species examined. These three loci also showed Mendelian inheritance in the humpback whale trios investigated. An exception was the minke whale which did not amplify for either DQA or DQB; genes which are also missing in the genome assembly. However, the missing DQ pair also coincided with an assembly gap so further research is needed to confirm the missing DQ genes in minke whale.

Our finding of a single DQB gene is at odds with some previous studies that reported a duplicated DQB gene in some baleen whale species (Baker et al., 2006; Moreno-Santillán et al., 2016) and the Baji (Yang et al., 2005). Here (see Figure S4), we show that the original DQB primer pair that was used in those studies (DQB1 and DQB2; first used in a cetacean by Murray et al. (1995)) amplifies a 172 bp fragment of exon 2 of DQB, but also binds to one or more of the DRB in some species where the 5′ end of DQB-2 primer differed at three nucleotides (Figure S4a,b for alignments). These amplicons are of the same length and are similar in sequence to the expected DQB sequence and cannot be differentiated by locus or species based on phylogenetic methods (Baker et al., 2006). This suggests that the previous finding of duplicated DQB loci was more likely the result of amplification bias rather than a true duplication event. For example, a previously documented DQB allele (GenBank: LiveDQB*8_AY177286) from the Baji has been found to be identical to the DRB-b locus in the Baji genome (NW_006786873). This highlights how increasing genomic resources can help with study design and identify and correct errors that have arisen from previously limited genetic data.

The DRB gene was the only class IIa gene that was duplicated in all cetaceans, with a DRB pseudogene found in baleen whales and the Baji (confirming previous work by Alves de Sá et al., 2019). Full-length phylogenetic trees of DRB clearly distinguished sequences by genome location, supporting a duplication of DRB-a that resulted in DRB-b before the split of baleen and toothed whales 25 million years ago (McGowen et al., 2020).

We investigated only one DRB gene across cetaceans, as our primer pair was designed to amplify DRB-a only. Amplicon sequencing confirmed a maximum of two alleles per species at this locus across 11 cetacean species examined. A comparison of genome assembly and amplicon data highlighted a discrepancy for the long-finned pilot whale and rough-toothed dolphin genome assemblies. In these species, the DRB primer pair did not amplify the region in silico but did produce DRB amplicons and putative alleles. On closer examination, the only DRB gene present in both genomes is an amalgamation of the first half of DRB-b (including exon 1) and the second half of DRB-a (after exon 2 including exons 3 and 4) and coincides with an assembly gap in both genomes. This is pointing towards assembly issues, likely not representative of the true genomic sequence.

4.1.2 | MHC class I

MHC class I organisation is much more complex and variable compared to class IIa, likely related to its faster evolutionary rate and higher gene copy number (Minias & Remisiewicz, 2021). We concentrated our analysis on the region between framework genes DDX39B and TRIM26 as these enclose the β and κ blocks in which an expansion of class I genes occurred in Laurasiatheria, a superorder of placental mammals to which cetaceans belong (Abduriyim et al., 2019). Gene copy number, length and content of class I region varied within the genome assemblies of toothed whales and dolphins and baleen whales, and even within their families. However, there is a clear disparity between our amplicon data, seemingly confirming class I genes in some species, but missing in the genome assembly in the expected location (i.e. Humpback whale, Blainville's beaked whale, Risso's dolphin and rough-toothed dolphin; Table 1). These findings suggest caution is required by researchers when using these assemblies, as there seems to be some assembly issues in this gene region. Future studies are needed to validate gene content and the true MHC class I organisation for each of those species.

In humpback whales and southern right whales, the same or very similar alleles can be found in all individuals in the middle class I and β block class I genes. Similarly, bovines have three class I genes located in the β block (Birch et al., 2008; Ellis & Hammond, 2014). The class I gene situated between duplication blocks ('middle class I') is not common. However, the consistent presence of this gene in 10 assemblies

across all cetaceans seems to suggest that the position of the gene in the genome is correct, as well as the finding that MHC class I genes located outside those blocks have been previously reported in sheep (Siva Subramaniam et al., 2015) and a class I pseudogene in a similar position found for cattle (Birch et al., 2008). Further expression studies will be needed to validate this designation.

Our analysis of humpback whale paternity trios supported variable gene content of class I genes on haplotypes. In the four investigated trios, we found haplotypes with one to three class I genes. This is very similar to bovine for which haplotypes have been described that contain between one and four class I genes (Codner et al., 2012; Hammond et al., 2012). The κ block is where ruminants class I genes have duplicated and expanded (Siva Subramaniam et al., 2015) and this is also where we find multiple class I genes in cetaceans. However, this was only observed in three assemblies, the bottlenose dolphin, the blue whale and the Yangtze finless porpoise, which coincidentally are the three most curated assemblies with no assembly gaps: the first two from NCBI, the latter assembled from BAC clones. Also, amplicons were observed from all investigated species with a maximum of nine alleles per individual. This further points towards assembly difficulties for most of the cetacean class I regions. In conclusion, we cautiously predict that middle and β class I genes are present in all cetaceans, and the κ region can contain between one and three class I genes. In addition, we predict that there is a shorter class I gene (shorter at exon 2) found only in baleen whales. The individuals we used in our panels were not the same ones that were used for genome assemblies so that we could have amplified alleles that are present in one individual but not in another that was used for genome assembly. This is always a concern when using reference genomes that present one haplotype only. We tried to be conservative in our conclusions and utilising amplicons and genomes across all cetaceans, a gene that is present in most of the species is likely to be present in all individuals of a species as well.

4.2 | Future research directions and limitations

Here we presented a pan-cetacean MHC genotyping panel that provided consistent results across species. The utility of such a panel is likely to increase as analyses of MHC and other immune system genes become more common due to the expanding incidences of emerging diseases in cetacean populations, which are expected to increase under rapid climate change (Kebke et al., 2022; van Bressem et al., 2009). As well as an indicator of population health and investigating the relationship between the immune system and microbiome (e.g., Fleischer et al., 2022), MHC is also used as a proxy for functional genomic diversity (e.g., Manlik et al., 2019; Slade & Mccallum, 1992) and is investigated across species for its role in mate choice (e.g., Santos et al., 2018; Schwensow et al., 2007).

The quality and comprehensiveness of genomes continue to improve with the advent of long-read sequencing technologies, with direct relevance to conservation (e.g., Whibley, 2021). Our comparative approach to assess MHC class I and IIa organisation in cetaceans by regions that are defined by the framework genes in combination with a genotype amplicon panel is a powerful approach for highlighting which assemblies are likely suffering from assembly artefacts, in addition to traditional metrics like N50 and BUSCO scores (Jauhal & Newcomb, 2021). However, most genome assemblies shown here need careful future curation for the class I MHC region.

Our analysis of humpback whale paternity trios highlighted the role of kin relationships to infer phasing and identify haplotypes. Future work could build on our amplicon panel using long-read sequencing, as has been suggested in other non-model organisms (e.g., O'Connor et al., 2019). Future full-length resolution of class I alleles, which include exon 3, could split alleles based on only exon 2 found in this study. However, this seems unlikely, given that bovine MHC class I alleles are more variable at exon 2 than exon 3 (Heimeier, unpublished; and also see https://www.ebi.ac.uk/ipd/mhc/group/BoLA/) as well as that class I alleles showed Mendelian inheritance patterns in our humpback whale paternity trios. This panel development, along with the comparative analysis that found some previously unrecognised redundancy of cetacean MHC genes within the published literature and on GenBank, has prompted us to expand IPD-MHC to cetacean species (Maccari et al., 2017, 2020). This initiative will allow the standardisation of MHC nomenclature and to undertake an MHC 'inventory' for all cetacean species to reduce redundancy.

4.3 | Conclusions

It has been suggested that the focus of conservation genetics should shift from the assessment of neutral genetic diversity to functional genetic diversity (e.g., Teixeira & Huber, 2021). MHC has remained one of the most important and commonly studied functional genomic markers, due to its importance in disease resistance and mate choice (e.g., Kamiya et al., 2014). Here, we show that the genomics revolution is producing sufficient resources to allow for the design, amplification and validation of an MHC panel that works across a non-model species infraorder, Cetacea.

Here, we designed a genomically informed approach to develop a pan-cetacean MHC panel that we subsequently successfully applied to 19 species from six families of whales and dolphins. The comparison of genome assemblies and the amplicon panel results highlighted potential scope for improvement in available genomes for this often complex and repetitive region. To support improved standardisation of MHC nomenclature in cetaceans going forward, we have extended a free online resource MHC-IDP (Maccari et al., 2017, 2020) for cetaceans, which we hope will benefit the cetacean research community and beyond.

AUTHOR CONTRIBUTIONS

Conceptualisation: Dorothea Heimeier, Ellen C. Garland, Emma L. Carroll; Funding: Ellen C. Garland, Emma L. Carroll; Sample collection: Emma L. Carroll, Claire Garrigue, Adriana Vella, Scott Baker; Sample selection: Franca Eichenberger and Dorothea Heimeier;

MHC Methodology: Dorothea Heimeier and Emma L. Carroll; Bioinformatics: Dorothea Heimeier; Paternity analysis: Franca Eichenberger; Writing–original draft: Dorothea Heimeier, Emma L. Carroll, Ellen C. Garland; All authors critically revised the manuscript and approved the final manuscript.

ACKNOWLEDGEMENTS

This study was funded by a Royal Society Research Grants for Research Fellows (RGF\R1\181014) to E.C.G. E.C.G. is funded by a Royal Society University Research Fellowship (UF160081 & URF\R\221020). F.E. is supported by a University of St Andrews School of Biology Ph.D. scholarship and a Royal Society Research Fellows Enhancement Award (RGF\EA\180213 to E.C.G). E.L.C. is funded by a Rutherford Discovery Fellowship from the Royal Society of New Zealand Te Apārangi. Surveys of humpback whales in New Caledonia were made possible by contributions from Fondation d'Entreprise Total and Total Pacifique, the Provinces Sud, North and Isles and Inco S.A. We thank Dominique Boillon, Claire Bonneville, Solène Derville, Magaly Chambellant, Rémi Dodemont, Jacqui Greaves, Veronique Pérard and all the volunteers who helped in the field. Thank you to Michael Poole for contributing genetic samples to this study. We thank the Auckland Genomics Unit at Auckland University for sequencing services. This study was approved by the University of St Andrews School of Biology Ethics Committee (ref: SEC2018004). Unpublished genome assemblies and sequencing data are used with permission from the DNA Zoo Consortium (dnazoo.org). Acknowledgement information per species is provided in full in the Supplementary Information (Acknowledgements and references for genome assemblies). Samples from southern right whales were collected under New Zealand Marine Mammal Protection Act Permit 84845-MAR and Marine Reserve Act Permit 87513-MAR following University of Auckland Animal Ethics approved protocol 002072 to ELC. This work was supported by funding from Live Ocean, Lou and Iris Fisher Trust, Royal Society Rutherford Discovery Fellowship and University of Auckland Science Faculty Research Development Fund grants to ELC and in-kind support from the Cawthron Institute, Antarctic New Zealand, Australian Antarctic Division and British Antarctic Survey. We thank Captain Steve Kafka, Sandra Carrod and crew of the Evohe Jim Dilley, Tori Muir and Johan Domeij and our research team Simon Childerhouse, Leena Riekkola, Rochelle Constantine, Ros Cole, Esther Stuck, Bill Morris and Richie Robinson for their help in the field. We thank the Kaitiaki Roopū o Murihiku for discussions around and support of the right whale project. Samples from the New Zealand Cetacean Tissue archive were collected by the New Zealand Department of Conservation—Te Papa Atawhai in consultation with mana whenua (Indigenous groups) of Aotearoa New Zealand and curated by ELC and Rochelle Constantine as part of the New Zealand Cetacean Tissue Archive at the University of Auckland—Waipapa Taumata Rau. Open access publishing facilitated by The University of Auckland, as part of the Wiley - The University of Auckland agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

All predicted to be functional MHC class II alleles are available on GenBank (Accession numbers OR901500-OR901654). MHC class I alleles and class IIa (non-functional DRB-a) alleles are available on Dryad. Also, all raw reads have been submitted to Dryad https://doi.org/10.5061/dryad.wh70rxwvb.

BENEFIT SHARING STATEMENT

The authors of this work recognise the rights of Indigenous peoples to make decisions about the future use of information, biological collections, data and digital sequence information that derives from associated lands, waters and territories. To support the practice of proper and appropriate acknowledgement into the future of these rights, we request that those seeking to reuse data from the New Zealand Cetacean Tissue Archive (see Table S3) contact EC (e.carroll@auckland.ac.nz) ahead of use and publication.

ORCID

Dorothea Heimeier https://orcid.org/0000-0001-8731-0842
Ellen C. Garland https://orcid.org/0000-0002-8240-1267
Franca Eichenberger https://orcid.org/0000-0002-3345-4792
Claire Garrigue https://orcid.org/0000-0002-8117-3370
Adriana Vella https://orcid.org/0000-0003-2246-5354
C. Scott Baker https://orcid.org/0000-0002-6276-3244
Emma L. Carroll https://orcid.org/0000-0003-3193-7288

REFERENCES

Abduriyim, S., Zou, D. H., & Zhao, H. (2019). Origin and evolution of the major histocompatibility complex class I region in eutherian mammals. *Ecology and Evolution*, *9*(13), 7861–7874. https://doi.org/10.1002/ece3.5373

Albertson, G. R., Baird, R. W., Oremus, M., Poole, M. M., Martien, K. K., & Baker, C. S. (2017). Staying close to home? Genetic differentiation of rough-toothed dolphins near oceanic islands in the central Pacific Ocean. *Conservation Genetics*, *18*(1), 33–51. https://doi.org/10.1007/s10592-016-0880-z

Alvaro, S., Herdegen, M., Migalska, M., & Radwan, J. (2016). AMPLISAS: a web server fpr multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources*, *16*, 498–510.

Alves de Sá, A. L., Breaux, B., Burlamaqui, T. C. T., Deiss, T. C., Sena, L., Criscitiello, M. F., & Cruz Schneider, M. P. (2019). The marine mammal class II major histocompatibility complex organization. *Frontiers in Immunology*, *10*, 1–14. https://doi.org/10.3389/fimmu.2019.00696

Amadou, C. (1999). Evolution of the Mhc class I region: the framework hypothesis. *Immunogenetics*, *49*(4), 362–367. https://doi.org/10.1007/s002510050507

Andersson, L., Lundén, A., Sigurdardottir, S., Davies, C., & Rask, L. (1988). Linkage relationships in the bovine MHC region. High recombination frequency between class II subregions. *Immunogenetics*, *27*(4), 273–280. https://doi.org/10.1007/BF00376122

Arbanasić, H., Đuras, M., Podnar, M., Gomerčić, T., Ćurković, S., & Galov, A. (2014). Major histocompatibility complex class II variation in bottlenose dolphin from Adriatic Sea: Inferences about the extent of balancing selection. *Marine Biology*, *161*(10), 2407–2422. https://doi.org/10.1007/s00227-014-2515-6

Avila, I. C., Kaschner, K., & Dormann, C. F. (2018). Current global risks to marine mammals: Taking stock of the threats. *Biological Conservation*, *221*, 44–58. https://doi.org/10.1016/j.biocon.2018.02.021

Baker, C. S., Slade, R. W., Bannister, J. L., Abernethy, R. B., Weinrich, M. T., Lien, J., Urban, J., Corkeron, P., Calmabokidis, J., Vasquez, O., & Palumbi, S. R. (1994). Hierarchical structure of mitochondrial DNA gene flow among humpback whales, *Megaptera novaeangliae*, world-wide. *Molecular Ecology*, *3*, 313–327.

Baker, C. S., Vant, M. D., Dalebout, M. L., Lento, G. M., O'Brien, S. J., & Yuhki, N. (2006). Diversity and duplication of DQB and DRB-like genes of the MHC in baleen whales (suborder: Mysticeti). *Immunogenetics*, *58*(4), 283–296. https://doi.org/10.1007/s00251-006-0080-y

Belov, K., Deakin, J. E., Papenfuss, A. T., Baker, M. L., Melman, S. D., Siddle, H. v., Gouin, N., Goode, D. L., Sargeant, T. J., Robinson, M. D., Wakefield, M. J., Mahony, S., Cross, J. G. R., Benos, P. V., Samollow, P. B., Speed, T. P., Marshall Graves, J. A., & Miller, R. D. (2006). Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. *PLoS Biology*, *4*(3), 317–328. https://doi.org/10.1371/journal.pbio.0040046

Bentkowski, P., & Radwan, J. (2019). Evolution of major histocompatibility complex gene copy number. *PLoS Computational Biology*, *15*, 1–15.

Birch, J., Codner, G., Guzman, E., & Ellis, S. A. (2008). Genomic location and characterisation of nonclassical MHC class I genes in cattle. *Immunogenetics*, *60*(5), 267–273. https://doi.org/10.1007/s00251-008-0294-2

Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge—Accurate paired shotgun read merging via overlap. *PLoS ONE*, *12*(10), 1–15. https://doi.org/10.1371/journal.pone.0185056

Cammen, K. M., Andrews, K. R., Carroll, E. L., Foote, A. D., Humble, E., Khudyakov, J. I., Louis, M., McGowen, M. R., Olsen, M. T., & van Cise, A. M. (2016). Genomic methods take the plunge: Recent advances in high-throughput sequencing of marine mammals. *Journal of Heredity*, *107*(6), 481–495. https://doi.org/10.1093/jhered/esw044

Carroll, E. L., Riekkola, L., Andrews-Goff, V., Baker, C. S., Constantine, R., Cole, R., Goetz, K., Harcourt, R., Lundquist, D., Meyer, C., Ogle, M., O'Rorke, R., Patenaude, N., Russ, R., Stuck, E., van der Reis, A. L., Zerbini, A. N., & Childerhouse, S. (2022). New Zealand southern right whale (Eubalaena australis; Tohorā nō Aotearoa) behavioural phenology, demographic composition, and habitat use in port Ross, Auckland Islands over three decades: 1998–2021. *Polar Biology*, *45*(8), 1441–1458. https://doi.org/10.1007/s00300-022-03076-7

Castro-Prieto, A., Wachter, B., & Sommer, S. (2011). Cheetah paradigm revisited: MHC diversity in the world's largest free-ranging population. *Molecular Biology and Evolution*, *28*(4), 1455–1468. https://doi.org/10.1093/molbev/msq330

Codner, G. F., Birch, J., Hammond, J. A., & Ellis, S. A. (2012). Constraints on haplotype structure and variable gene frequencies suggest a functional hierarchy within cattle MHC class i. *Immunogenetics*, *64*(6), 435–445. https://doi.org/10.1007/s00251-012-0612-6

Coker, O. M., Osaiyuwu, O. H., & Oladiran, A. (2023). Major Histocompatibility Complex (MHC) Diversity and its implications for human and wildlife health and Conservation—A review. https://doi.org/10.46325/gabj.v7i2.318

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, *14*, 1394–1403. https://doi.org/10.1101/gr.2289704.tion

Derville, S., Torres, L. G., Dodémont, R., Perard, V., & Garrigue, C. (2019). From land and sea, long-term data reveal persistent humpback whale (Megaptera novaeangliae) breeding habitat in New Caledonia. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *29*, 1697–1711. https://doi.org/10.1002/aqc.3127

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). De novo assembly of the Aedes aegypti genome using hi-C yields chromosome-length scaffolds. *Science*, *356*(6333), 92–95. https://doi.org/10.1126/science.aal3327

Eichenberger, F., Carroll, E. L., Garrigue, C., Rendell, L., Steel, D., Bonneville, C., Jarman, S., & Garland, E. C. (2022). Variation in male reproductive success in a singing cetacean. *24th Biennial Conference on the Biology of Marine Mammals*, *1-5 August*., 197–198.

Ellis, S. A., & Hammond, J. A. (2014). The functional significance of cattle major histocompatibility complex class i genetic diversity. *Annual Review of Animal Biosciences*, *2*, 285–306. https://doi.org/10.1146/annurev-animal-022513-114234

Fleischer, R., Schmid, D. W., Wasimuddin, Brändel, S. D., Rasche, A., Corman, V. M., Drosten, C., Tschapka, M., & Sommer, S. (2022). Interaction between MHC diversity and constitution, gut microbiota and astrovirus infections in a neotropical bat. *Molecular Ecology*, *31*(12), 3342–3359. https://doi.org/10.1111/mec.16491

Flores-Ramírez, S., Urban-Ramirez, J., & Miller, R. D. (2000). Major histocompatibility complex class I loci from the gray whale (*Eschrichtius robustus*). *The Journal of Heredity*, *91*(4), 279–282.

Garrigue, C., Greaves, J., & Chambellant, M. (2001). Characteristics of the new Caledonian humpback whale population. *Memoirs of the Queensland Museum*, *47*(2), 539–546.

Gigliotti, A., Bowen, W. D., Hammill, M., Puryear, W., Runstadler, J., Wenzel, F., & Cammen, K. M. (2021). Sequence diversity and differences at the highly duplicated MHC-I gene reflect viral susceptibility in sympatric pinniped species.

Gillett, R. M., Murray, B. W., & White, B. N. (2014). Characterization of class I- and class II-like major histocompatibility complex loci in pedigrees of north atlantic right whales. *Journal of Heredity*, *105*(2), 188–202. https://doi.org/10.1093/jhered/est095

Halenius, A., Gerke, C., & Hengel, H. (2015). Classical and non-classical MHC I molecule manipulation by human cytomegalovirus: So many targets—but how many arrows in the quiver? *Cellular and Molecular Immunology*, *12*(2), 139–153. https://doi.org/10.1038/cmi.2014.105

Hammond, J. A., Marsh, S. G. E., Robinson, J. T., Davies, C. J., Stear, M. J., & Ellis, S. A. (2012). Cattle MHC nomenclature: Is it possible to assign sequences to discrete class I genes? *Immunogenetics*, *64*(6), 475–480. https://doi.org/10.1007/s00251-012-0611-7

Hamner, R. M., Constantine, R., Mattlin, R., Waples, R., & Baker, C. S. (2017). Genotype-based estimates of local abundance and effective population size for Hector's dolphins. *Biological Conservation*, *211*, 150–160. https://doi.org/10.1016/j.biocon.2017.02.044

Heimeier, D., Alexander, A., Hamner, R. M., Pichler, F. B., & Baker, C. S. (2018). The influence of selection on MHC DQA and DQB haplotypes in the endemic New Zealand hector's and Maui dolphins. *Journal of Heredity*, *109*(7), 744–756. https://doi.org/10.1093/jhered/esy050

Huang, S., Huang, X., Li, S., Zhu, M., & Zhuo, M. (2019). MHC class I allele diversity in cynomolgus macaques of Vietnamese origin. *PeerJ*, *2019*(11), 1–22. https://doi.org/10.7717/peerj.7941

Jauhal, A. A., & Newcomb, R. (2021). Assessing genome assembly quality prior to downstream analysis: N50 versus BUSCO. *Molecular Ecology Resources*, *21*(6), 1416–1421. https://doi.org/10.1111/1755-0998.13364

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, *16*, 1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Kamiya, T., O'Dwyer, K., Westerdahl, H., Senior, A., & Nakagawa, S. (2014). A quantitative review of MHC-based mating preference: The role of diversity and dissimilarity. *Molecular Ecology*, *23*(21), 5151–5163. https://doi.org/10.1111/mec.12934

Katoh, K., Kuma, K. I., Toh, H., & Miyata, T. (2005). MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, *33*(2), 511–518. https://doi.org/10.1093/nar/gki198

Kaufman, J. (2018). Unfinished business: Evolution of the MHC and the adaptive immune system of jawed vertebrates. *Annual Review of Immunology*, *36*, 383–409. https://doi.org/10.1146/annurev-immunol-051116-052450

Kebke, A., Samarra, F., & Derous, D. (2022). Climate change and cetacean health: Impacts and future directions. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *377*(1854), 20210249. https://doi.org/10.1098/rstb.2021.0249

Kelley, J., Walter, L., & Trowsdale, J. (2005). Comparative genomics of major histocompatibility complexes. *Immunogenetics*, *56*(10), 683–695. https://doi.org/10.1007/s00251-004-0717-7

Klein, J. (1986). *Natural history of the major histocompatibility complex*. (J. Wiley, Ed.).

Krasnec, K. V., Sharp, A. R., Williams, T. L., & Miller, R. D. (2015). The opossum MHC genomic region revisited. *Immunogenetics*, *67*(4), 259–264. https://doi.org/10.1007/s00251-015-0826-5

Kumánovics, A., Takada, T., & Lindahl, K. F. (2003). Genomic organization of the mammalian MHC. *Annual Review of Immunology*, *21*(1), 629–657. https://doi.org/10.1146/annurev.immunol.21.090501.080116

Leclaire, S., Strandh, M., Dell'Ariccia, G., Gabirot, M., Westerdahl, H., & Bonadonna, F. (2019). Plumage microbiota covaries with the major histocompatibility complex in blue petrels. *Molecular Ecology*, *28*(4), 833–846. https://doi.org/10.1111/mec.14993

Lighten, J., Van Oosterhout, C., & Bentzen, P. (2014). Critical review of NGS analyses for de novo genotyping multigene families. *Molecular Ecology*, *23*(16), 3957–3972. https://doi.org/10.1111/mec.12843

Maccari, G., Robinson, J., Hammond, J. A., & Marsh, S. G. E. (2020). The IPD Project: a centralised resource for the study of polymorphism in genes of the immune system. 49–55.

Maccari, G., Robinson, J. T., Ballingall, K. T., Guethlein, L. A., Grimholt, U., Kaufman, J., Ho, C. S., De Groot, N. G., Flicek, P., Bontrop, R. E., Hammond, J. A., & Marsh, S. G. E. (2017). IPD-MHC 2.0: An improved inter-species database for the study of the major histocompatibility complex. *Nucleic Acids Research*, *45*(D1), D860–D864. https://doi.org/10.1093/nar/gkw1050

Manczinger, M., Boross, G., Kemény, L., Müller, V., Lenz, T. L., Papp, B., & Pál, C. (2019). Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLoS Biology*, *17*(1), 1–21. https://doi.org/10.1371/journal.pbio.3000131

Manlik, O., Krützen, M., Kopps, A. M., Mann, J., Bejder, L., Allen, S. J., Frère, C., Connor, R. C., & Sherwin, W. B. (2019). Is MHC diversity a better marker for conservation than neutral genetic diversity? A case study of two contrasting dolphin populations. *Ecology and Evolution*, *9*(12), 6986–6998. https://doi.org/10.1002/ece3.5265

McGowen, M. R., Tsagkogeorga, G., Álvarez-Carretero, S., dos Reis, M., Struebig, M., Deaville, R., Jepson, P. D., Jarman, S., Polanowski, A., Morin, P. A., & Rossiter, S. J. (2020). Phylogenomic resolution of the cetacean tree of life using target sequence capture. *Systematic Biology*, *69*(3), 479–501. https://doi.org/10.1093/sysbio/syz068

Minias, P., & Remisiewicz, M. (2021). Distinct evolutionary trajectories of MHC class I and class II genes in old world Fi Nches and buntings. *Heredity (Edinb)*, *126*, 974–990. https://doi.org/10.1038/s41437-021-00427-8

Moreno-Santillán, D. D., Lacey, E. A., Gendron, D., & Ortega, J. (2016). Genetic variation at exon 2 of the MHC class II DQB locus in blue whale (*Balaenoptera musculus*) from the Gulf of California. *PLoS ONE*, *11*(1), 1–15. https://doi.org/10.1371/journal.pone.0141296

Murray, B. W., Malik, S., & White, B. N. (1995). Sequence variation at the major histocompatibility complex locus DQB in beluga whales. *Molecular Biology and Evolution*, *12*(4), 582–593.

Nei, M., Gu, X., & Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences*, *94*(15), 7799–7806. https://doi.org/10.1073/pnas.94.15.7799

O'Connor, E. A., Westerdahl, H., Burri, R., & Edwards, S. V. (2019). Avian mhc evolution in the era of genomics: Phase 1.0. *Cells*, *8*(10), 1–21. https://doi.org/10.3390/cells8101152

Oremus, M., Poole, M. M., Albertson, G. R., & Baker, C. S. (2012). Pelagic or insular? Genetic differentiation of rough-toothed dolphins in the Society Islands, French Polynesia. *Journal of Experimental Marine Biology and Ecology*, *432–433*, 37–46. https://doi.org/10.1016/j.jembe.2012.06.027

Pearson, S. K., Bull, C. M., & Gardner, M. G. (2017). Egernia stokesii (gidgee skink) MHC I positively selected sites lack concordance with HLA peptide binding regions. *Immunogenetics*, *69*(1), 49–61. https://doi.org/10.1007/s00251-016-0947-5

Piertney, S. B., & Oliver, M. K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, *96*(1), 7–21. https://doi.org/10.1038/sj.hdy.6800724

Prugnolle, F., Manica, A., Charpentier, M., Guégan, J. F., Guernier, V., & Balloux, F. (2005). Pathogen-driven selection and worldwide HLA class I diversity. *Current Biology: CB*, *15*(11), 1022–1027. https://doi.org/10.1016/j.cub.2005.04.050

Radwan, J., Babik, W., Kaufman, J., Lenz, T. L., & Winternitz, J. (2020). Advances in the evolutionary understanding of MHC polymorphism. *Trends in Genetics*, *36*(4), 298–311. https://doi.org/10.1016/j.tig.2020.01.008

Razali, H., Connor, E. O., Drews, A., Burke, T., & Westerdahl, H. (2017). A quantitative and qualitative comparison of illumina MiSeq and 454 amplicon sequencing for genotyping the highly polymorphic major histocompatibility complex (MHC) in a non-model species. *BMC Research Notes*, *10*, 346. https://doi.org/10.1186/s13104-017-2654-1

Rekdal, S. L., Anmarkrud, J. A., Johnsen, A., & Lifjeld, J. T. (2018). Genotyping strategy matters when analyzing hypervariable major histocompatibility complex-experience from a passerine bird. *Ecology and Evolution*, *8*(3), 1680–1692. https://doi.org/10.1002/ece3.3757

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., … Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, *592*(7856), 737–746. https://doi.org/10.1038/s41586-021-03451-0

Robinson, J. T., Guethlein, L. A., Cereb, N., Yang, S. Y., Norman, P. J., Marsh, S. G. E., & Parham, P. (2017). Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles. *PLoS Genetics*, *13*(6), 1–28. https://doi.org/10.1371/journal.pgen.1006862

Rock, K. L., Reits, E., & Neefjes, J. (2016). Present yourself! By MHC class I and MHC class II molecules. *Trends in Immunology*, *37*(11), 724–737. https://doi.org/10.1016/j.it.2016.08.010

Ruan, R., Ruan, J., Wan, X. L., Zheng, Y., Chen, M. M., Zheng, J. S., & Wang, D. (2016). Organization and characteristics of the major histocompatibility complex class II region in the Yangtze finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*). *Scientific Reports*, *6*, 1–11. https://doi.org/10.1038/srep22471

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual*. Cold SpringHarbor Laboratory Press.

Santos, P. S. C., Mezger, M., Kolar, M., Michler, F. U., & Sommer, S. (2018). The best smellers make the best choosers: Mate choice is affected by female chemosensory receptor gene diversity in a mammal. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1893), 20182426. https://doi.org/10.1098/rspb.2018.2426

Schmeller, D. S., Courchamp, F., & Killeen, G. (2020). Biodiversity loss, emerging pathogens and human health risks. *Biodiversity and Conservation*, *29*(11–12), 3095–3102. https://doi.org/10.1007/s10531-020-02021-6

Schwartz, J. C., & Hammond, J. A. (2015). The assembly and characterisation of two structurally distinct cattle MHC class I haplotypes point to the mechanisms driving diversity. *Immunogenetics*, *67*(9), 539–544. https://doi.org/10.1007/s00251-015-0859-9

Schwensow, N., Eberle, M., & Sommer, S. (2007). MHC-associated mate choice in a wild promiscuous primate. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 555–564. https://doi.org/10.1098/rspb.2007.1433

Shortreed, C. G., Wiseman, R. W., Karl, J. A., Bussan, H. E., Baker, D. A., Prall, T. M., Haj, A. K., Moreno, G. K., Penedo, M. C. T., & O'Connor, D. H. (2020). Characterization of 100 extended major histocompatibility complex haplotypes in Indonesian cynomolgus macaques. *Immunogenetics*, *72*(4), 225–239. https://doi.org/10.1007/s00251-020-01159-5

Siva Subramaniam, N., Morgan, E. F., Wetherall, J. D., Stear, M. J., & Groth, D. M. (2015). A comprehensive mapping of the structure and gene organisation in the sheep MHC class I region. *BMC Genomics*, *16*(1), 1–17. https://doi.org/10.1186/s12864-015-1992-4

Skow, L. C., Snaples, S. N., Davis, S. K., Taylor, J. F., Huang, B., & Gallagher, D. H. (1996). Localization of bovine lymphocyte antigen (BoLA) DYA and class I loci to different regions of chromosome 23. *Mammalian Genome*, *7*(5), 388–389. https://doi.org/10.1007/s003359900112

Slade, R. W., & Mccallum, H. I. (1992). Overdominant vs. frequency-dependent selection at MHC loci. *Genetics*, *132*, 861–862.

Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*, *2*, 16. https://doi.org/10.1186/1742-9994-2-16

Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings. Biological Sciences/the Royal Society*, *277*(1684), 979–988. https://doi.org/10.1098/rspb.2009.2084

Stutz, W. E., & Bolnick, D. I. (2014). Stepwise threshold clustering: A new method for genotyping MHC loci using next-generation sequencing technology. *PLoS ONE*, *9*(7), 25–27. https://doi.org/10.1371/journal.pone.0100587

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729. https://doi.org/10.1093/molbev/mst197

Teixeira, J. C., & Huber, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(10), 1–10. https://doi.org/10.1073/pnas.2015096118

Tezanos-Pinto, G., Baker, C. S., Russell, K., Martien, K., Baird, R. W., Hutt, A., Stone, G., Mignucci-Giannoni, A. A., Caballero, S., Endo, T., Lavery, S., Oremus, M., Olavarría, C., & Garrigue, C. (2009). A worldwide perspective on the population structure and genetic diversity of bottlenose dolphins (*Tursiops truncatus*) in New Zealand. *Journal of Heredity*, *100*(1), 11–24. https://doi.org/10.1093/jhered/esn039

Thompson, K. F., Millar, C. D., Scott Baker, C., Dalebout, M., Steel, D., van Helden, A. L., & Constantine, R. (2013). A novel conservation approach provides insights into the management of rare cetaceans. *Biological Conservation*, *157*, 331–340. https://doi.org/10.1016/j.biocon.2012.07.017

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), 1–12. https://doi.org/10.1093/nar/gks596

van Bressem, M. F., Raga, J. A., di Guardo, G., Jepson, P. D., Duignan, P. J., Siebert, U., Barrett, T., de Oliveira Santos, M. C., Moreno, I. B., Siciliano, S., Aguilar, A., & van Waerebeek, K. (2009). Emerging infectious diseases in cetaceans worldwide and the possible role of environmental stressors. *Diseases of Aquatic Organisms*, *86*(2), 143–157. https://doi.org/10.3354/dao02101

Villanueva-Noriega, M. J., Baker, C. S., & Medrano-González, L. (2013). Evolution of the MHC-DQB exon 2 in marine and terrestrial mammals. *Immunogenetics*, *65*(1), 47–61. https://doi.org/10.1007/s00251-012-0647-8

Westerdahl, H., Mellinger, S., Sigeman, H., Kutschera, V. E., Proux-Wéra, E., Lundberg, M., Weissensteiner, M., Churcher, A., Bunikis, I., Hansson, B., Wolf, J. B. W., & Strandh, M. (2022). The genomic architecture of the passerine MHC region: High repeat content and contrasting evolutionary histories of single copy and tandemly duplicated MHC genes. *Molecular Ecology Resources*, *22*, 2379–2395. https://doi.org/10.1111/1755-0998.13614

Whibley, A. (2021). Genome insights give cause for optimism in the ongoing battle to save the vaquita. *Molecular Ecology Resources*, *21*(4), 1005–1007. https://doi.org/10.1111/1755-0998.13345

Worley, K., Collet, J., Spurgin, L. G., Cornwallis, C., Pizzari, T., & Richardson, D. S. (2010). MHC heterozygosity and survival in red junglefowl. *Molecular Ecology*, *19*(15), 3064–3075. https://doi.org/10.1111/j.1365-294X.2010.04724.x

Xu, S., Sun, P., Zhou, K., & Yang, G. (2007). Sequence variability at three MHC loci of finless porpoises (*Neophocaena phocaenoides*). *Immunogenetics*, *59*(7), 581–592. https://doi.org/10.1007/s00251-007-0223-9

Yamaguchi, T., & Dijkstra, J. M. (2019). Major histocompatibility complex (Mhc) genes and disease resistance in fish. *Cells*, *8*(4), 1–31. https://doi.org/10.3390/cells8040378

Yang, G., Yan, J., Zhou, K., & Wei, F. (2005). Sequence variation and gene duplication at MHC DQB loci of baiji (*Lipotes vexillifer*), a Chinese river dolphin. *The Journal of Heredity*, *96*(4), 310–317. https://doi.org/10.1093/jhered/esi055

Yang, W., Chou, L., & Hu, J. (2012). Sequence analysis of MHC class II genes in cetaceans. In B. Abdel-Salam (Ed.), *Histocompatibility* (pp. 117–132). InTech.

Yeager, M., & Hughes, A. L. (1999). Evolution of the mammalian MHC: Natural selection, recombination, and convergent evolution. *Immunological Reviews*, *167*, 45–58.

Zhang, Z., Sun, X., Chen, M., Li, L., Ren, W., Xu, S., & Yang, G. (2019). Genomic organization and phylogeny of MHC class II loci in cetaceans. *Journal of Heredity*, *110*(3), 332–339. https://doi.org/10.1093/jhered/esz005

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Heimeier, D., Garland, E. C., Eichenberger, F., Garrigue, C., Vella, A., Baker, C. S., & Carroll, E. L. (2024). A pan-cetacean MHC amplicon sequencing panel developed and evaluated in combination with genome assemblies. *Molecular Ecology Resources*, *00*, e13955. https://doi.org/10.1111/1755-0998.13955