Immunogenetics (2021) 73:405–417
https://doi.org/10.1007/s00251-021-01217-6

ORIGINAL ARTICLE

Major histocompatibility complex genes and locus organization in the Komodo dragon (*Varanus komodoensis*)

Kent M. Reed[1] · **Robert E. Settlage**[2]

Received: 2 March 2021 / Accepted: 25 April 2021 / Published online: 12 May 2021

Abstract

We performed a meta-analysis of the newly assembled Komodo dragon (*Varanus komodoensis*) genome to characterize the major histocompatibility complex (MHC) of the species. The MHC gene clusters of the Komodo dragon are gene dense, complex, and contain counterparts of many genes of the human MHC. Our analysis identified 20 contigs encompassing ~6.9 Mbp of sequence with 223 annotated genes of which many are predicted orthologs to the genes of the human MHC. These MHC contigs range in size from 13.2 kb to 21.5 Mbp, contain an average of one gene per 30 kb, and are thought to occur on at least two chromosomes. Eight contigs, each > 100 kb, could be aligned to the human MHC based on gene content, and these represent gene clusters found in each of the recognized mammalian MHC subregions. The MHC of the Komodo dragon shares organizational features of other non-mammalian taxa. Multiple class Iα and class IIβ genes are indicated, with linkage between classical class I and immunoproteasome genes and between framework class I genes and genes associated with the mammalian class III subregion. These findings are supported in both Komodo genome assemblies and provide new insight into the MHC organization of these unique squamate reptiles.

Keywords Komodo dragon · MHC · Genome · Evolution

Introduction

The basic features of a cell-mediated adaptive immune system appear to have been established over 500 million years ago in the immediate common vertebrate ancestor (Sutoh and Kasahara 2021). A key component in adaptive immune response is the major histocompatibility complex (MHC), a genomic locus found in all jawed vertebrates (gnathostomes) that is key in self/non-self recognition and pathogen defense. The MHC locus was originally identified by Peter Gorer and George Snell through tissue graft rejection experiments and is perhaps the most extensively studied vertebrate gene locus (Klein 1986). Initial sequence descriptions of the human MHC regions were based on both genetic mapping and the sequencing and assembly of various human haplotypes (Trowsdale et al. 1991; Campbell and Trowsdale 1993). The first sequence-based gene map was completed in 1999 (The MHC Sequencing Consortium 1999), and an integrated gene map of the extended human MHC was constructed by integrating all data publicly available at that time (Horton et al. 2004). The classical MHC (~3.6 Mb) includes three functional subregions: the class I (*HLA-F* to *MICB,* 1.8 Mb), class III (*PPIAP9* to *BTNL2*, 0.7 Mb), and class II (*HLA-DRA* to *HCG24*, ~1.0 Mb). The flanking extended class I (*SCGN* to *ZFP57)* and extended class II (*COL11A2* to *KIFC1*) subregions are ~4.0 Mb and 0.25 Mb, respectively (Horton et al. 2004; Shina et al. 2009). Thus, the extended human MHC, or HLA "super-locus," spans ~7.8 Mb on Chr6 (based on assembly GRCh38.p13). Paralogous MHC regions thought to be the result of ancient genome-wide duplication are located on chromosomes 1, 9, and 19 (Kasahara et al. 1996, 1997; Katsanis et al. 1996).

Detailed investigations of mice and humans have found the MHC to contain several classes of genes responsible for antigen presentation to the host immune system. The classical loci of the MHC are traditionally used to divide the mammalian MHC into the three functional regions. The class I region contains the classical class I genes that are

✉ Kent M. Reed
kmreed@umn.edu

[1] Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, St. Paul, MN 55108, USA

[2] Advanced Research Computing, Virginia Tech University, Blacksburg, VA 24061, USA

expressed on all nucleated cells. Class I molecules primarily present peptides derived endogenously from cytosolic proteins, but also present peptides generated from exogenous proteins to cytotoxic T cells eliciting non-self-response. Self-MHC class I molecules are also recognized by inhibitory natural killer cell receptors as part of NK cell tolerance. The class II region contains the classical class II genes primarily expressed on antigen-presenting cells (B-cells, basophils, dendritic cells, and macrophages) and primarily present exogenously derived peptides (invader discovery). The class III region contains genes with functional roles in innate immunity and inflammation as well as non-immune-related loci (Milner and Campbell 2001).

Increasingly, MHC assemblies are now accessible through whole genome sequencing projects. Comparison of evolutionary diverse taxa identifies lineage-specific rearrangements and provides evidence supporting hypotheses on the origin of MHC gene clusters and paralogous loci. Analysis of model organisms including zebrafish (Sambrook et al. 2005), *Xenopus* (Ohta et al. 2006), and the chicken (Kaufman et al. 1999) provided insights into how the orientation of MHC regions and their accompanying genes have diverged during vertebrate evolution. For example, a set of non-classical class I genes is uniquely present in non-placental mammals (marsupials and monotremes) having been lost from the eutherian (placental mammal) lineage (Papenfuss et al. 2015).

Based on comparative phylogenetic analyses of fish, amphibians, reptiles (including birds), and mammals, the ancestral form of the MHC, common in non-mammalian vertebrates and in some marsupials, includes classical class I loci linked to components of the antigen processing machinery (Sambrook et al. 2002, 2005; Belov et al. 2006; Jaratlerdsiri et al. 2014; Kaufman 2018). Linkage of classical class I and II genes likely did not appear until the amphibian lineage (Ohta et al. 2006), although linkage of the class Iα genes with class IIα and IIβ genes is found in one family of shark (Ohta et al. 2000), and a class I-like locus has been linked to class II loci in trout (Dijkstra et al. 2007). In the chicken, a highly organized and compact MHC (*B*-locus) was described as a "minimum essential MHC" containing a relatively small number of genes necessary for adaptive immunity (Kaufman et al. 1999). Linkage of class I-like loci to class III genes in birds, fish and monotremes clearly indicates that this arrangement has an ancient origin (Sambrook et al. 2002, 2005). In humans, the class I genes are linked to "framework" loci (Amadou 1999) and the class I and class II regions are separated from each other by the class III region. This I-III-II structure is characteristic of placental mammals and likely evolved less than 200 million years ago as a result of an inversion that separated the class Iα gene from the antigen processing machinery (Kumar and Hedges 1998; Kaufman 2018).

The four orders of non-avian reptiles include over 11,000 named species (ReptileDB) and are divided into two major clades: the archosaurs (Testudines (turtles) and Crocodilia (alligators, crocodiles, and gharials), and the Squamata (snakes and lizards)) and the monogeneric Sphenodontia (tuataras). Although subtle functional differences exist, reptiles appear to have the major components of the immune systems of mammals. For example, reptiles rely heavily on an efficient and diverse innate immune response, while displaying a slower and less robust adaptive immune response (Zimmerman et al. 2010; Zimmerman 2018, 2020). Comparative studies of non-avian reptiles including the green anole (*Anolis carolinensis*, Alfoldi et al. 2011) and species of Crocodylia (Jaratlerdsiri et al. 2014) have provided important contrasts to the avian MHC modeled extensively on galliformes. The MHC of the green anole (squamata) is reported as large and complex (~300 genes) with associated class I and II genes and a class III region closely linked to a framework class I region (Alfoldi et al. 2011). The saltwater crocodile has class I genes co-located with processing genes (transporter associated with antigen processing, *TAP*) consistent with the MHC of birds, but with linkage of class I and framework genes (e.g., *TRIM39*) resembling the eutherian MHC (Jaratlerdsiri et al. 2014). Sequencing and chromosome mapping in (*Sphenodon punctatus*) characterized the tuatara MHC as having high repeat content and low gene density (Miller et al. 2015). Antigen processing or framework genes were not found on the MHC gene-containing clones in this species.

Analysis of taxa from diverse groups highlights the dynamic nature of the MHC and provides important insight into MHC structure in the Reptilia. The Komodo dragon (*Varanus komodoensis*) is a member of the lizard family Varanidae (monitor lizards), an ancient clade that originated an estimated 40 million years ago. It is the largest living species of lizard, growing to a maximum length of ~3 m and weighing up to ~70 kg (Ciofi 1999). Two assemblies of the Komodo dragon genome have recently been published. The first (Lind et al. 2019, the Gladstone Institute assembly, hereafter referred to as GS) focused on identifying genomic adaptations in the cardiovascular and chemosensory systems. The second (van Hoek et al. 2019, the Virginia Tech assembly, hereafter referred to as VT) studied aspects of innate immunity genes for antimicrobial host-defense peptides (defensins and cathelicidins). We utilized these genome assemblies to identify components of the Komodo dragon MHC. Comparison of gene clusters like the MHC in this evolutionarily distinct group, provide clues to the origin and the evolution of the squamate reptiles.

Methods and materials

Ortholog identification and genome comparisons

We used an iterative best-hit methodology to manually identify annotated orthologs for genes of the human MHC in *V. komodoensis*. The annotated gene file from NCBI

was used as an information source for ortholog search based on identical gene symbols. Because the extended human class I region (130 + genes) includes members of four gene families (histone, butrophylin (BTN), olfactory receptor (OR), and zinc finger proteins (ZFP)), this subregion was excluded from our gene search as identification of orthologs is problematic with the potential for paralogous sequence matches. The identified Komodo genome contigs were individually inspected after aligning the predicted gene transcripts (open reading frames, ORFs) with the contig sequences using Sequencher (Gene Codes Corp). Genes annotated as hypothetical proteins were further investigated using Basic Local Alignment Search Tool (BLAST) searches for orthology.

To further refine the comparisons between the two Komodo dragon genome assemblies additional BLAST comparisons were performed. To normalize the comparison, predicted coding sequences were extracted from both assemblies. These were then used as subject sequences vs Uniprot (The UniProt Consortium 2021) via blastX and cross genome via blastN. Independently assembled Komodo dragon plasma transcripts (Bishop et al. 2017) were used as subjects in a separate blastN comparison of the assembled transcripts vs the VT and GS genome assemblies.

Visualization

Oxford grids

All vs all comparisons (LASTZ, Harris 2007) were used to align contigs > 10 kbp from the VT and GS assemblies (VT: total Len 1,504,696,282, num of seq 819, min 10,004, max 101,035,691, avg 1,837,235.997558, N50 24,139,216; GS total Len 1,507,945,839, num of seq 1411, min 10,002, max 138,280,312, avg 1,068,707.18568391, N50 23,831,982) of the Komodo dragon genome. Pairwise plots were produced using the R software package. GS contigs of interest were filtered to include those that contained MHC annotations. Manual inspection of the grids identified contigs from the two assemblies that shared long regions of sequence similarity and those with notable rearrangements.

Hive plots

Hive plots (HiveR, Hanson 2020) were created to better compare the three genomes: human chr6, VT, and GS contigs containing MHC genes. This was done in two ways. First, by using gene coordinates given by the reference annotations. Second, by adding additional alignment information obtained from all-vs-all blastN alignment of sliding 10 kbp segments of the appropriate contigs from the two Komodo assemblies. Specifically, SJPD01000006.1, SJPD01000114.1, SJPD01000117.1, SJPD01000119.1, SJPD01000140.1, SJPD01000167.1, SJPD01000168.1, SJPD01000211.1, SJPD01000215.1, SJPD01000224.1 from the GS assembly and VEXN01000097.1, VEXN01000312.1, VEXN01001601.1, VEXN01008528.1, VEXN01011230.1, VEXN01018551.1, VEXN01019096.1, VEXN01022374.1, VEXN01030207.1 from the VT assembly. Blast hits were filtered using the following criteria: bitscore > 1000, percent id > 90%, and hit length > 2500 bp. Sequences were arranged by increasing radius. For the GS assembly, the plot order is: SJPD01000006.1, SJPD01000167.1, SJPD01000140.1, SJPD01000117.1, SJPD01000119.1, SJPD01000114.1, SJPD01000168.1, SJPD01000215.1, SJPD01000224.1. For the VT assembly, the plot order is: VEXN01008528.1, VEXN01011230.1, VEXN01001601.1, VEXN01030207.1, VEXN01019096.1, VEXN01022374.1, VEXN01000097.1. The following contigs were plotted with reversed coordinates to give better visual concordance between the assemblies: SJPD01000140.1 SJPD01000117.1 SJPD01000119.1 VEXN01011230.1 VEXN01008528.1. Further refinement could include reversing both VEXN01008528.1 and SJPD01000167.1 and shuffling GS contig at position 2 (SJPD01000167.1) to after current contig 4 (SJPD01000117.1).

Results and discussion

Identification and characterization of MHC gene clusters in the Komodo dragon genome

Our investigation was initiated using the VT genome assembly (VKom_1.0, van Hoek et al. 2019). The annotation pipeline of van Hoek et al. predicted protein-coding genes with MAKER2 software (Holt and Yandell 2011) utilizing assembled RNAseq data of the Komodo dragon (Bishop et al. 2017) and protein sequences of the anole lizard (*A. carolinensis*, version AnoCar2.0) and python (*Python bivittatus*, version bivittatus-5.0.2). These were then integrated with prediction methods Blastx, SNAP (Korf 2004) and Augustus (Stanke and Waack 2003). During the writing of this manuscript (August 2020) an annotation of the GS assembly was accessioned at NCBI (reference genome, ASM479886v1). This annotation also used a MAKER pipeline with assembled RNAseq transcripts, protein homology, and de novo predictions as evidence.

Survey of the GS assembly identified 20 contigs containing predicted gene orthologs corresponding to the gene set of the human MHC and 2 contigs containing transcripts annotated as the MHC class I-related gene protein (*MR1*, Table 1). MHC contigs ranged in size from 13.2 kb (SJPD01000805.1) to 21.5 Mbp (SJPD01000006.1) with

an average of 349 kb. The number of annotated genes per contig ranged from 1 to 49 corresponding to an average of 1 gene per 30 kb. Eight contigs > 100 kb (SJPD01000006.1, SJPD01000114.1, SJPD01000117.1, SJPD01000119.1, SJPD01000140.1, SJPD01000168.1, SJPD01000215.1, and SJPD01000224.1) could be aligned to the human MHC based on gene content (Fig. 1). The remaining four notably contained predicted transcripts of class II histocompatibility antigens (SJPD01000211.1, SJPD01000217.1, and SJPD01000223.1) or members of the MHC-associated TRIM gene family (SJPD01000167.1). A list of all putative MHC gene transcripts contained within these contigs is included in Supplementary Table S1.

Approximately 0.8 Mbp (3′ end) of the largest contig (SJPD01000006.1) contained MHC-related genes. This contig included predicted genes that aligned with the 5′ end of the human classical class I subregion (Fig. 1). These framework genes include a group of tripartite motif (TRIM) protein genes (*TRIM7*, *15*, *17*, *27*, and *39*), E3 ubiquitin-protein ligase PPP1R11 (*PPP1R11*), RING finger protein 39 (*RNF39*), and DNA-directed RNA polymerase I subunit RPA12 (*ZNRD1*). Five transcripts in the contig are annotated as hypothetical proteins and two transcripts (*VESPs*) have similarity to unique venom proteins in reptiles (Vespryn/Ohanin) that induce hypolocomotion and hyperalgesia in mice (Pung et al. 2005).

Annotated transcripts in this contig also included genes not part of the classical human MHC subregions (Supplementary Table S1). In humans, *BTN2A1* (butyrophilin subfamily 2 member A1) is located upstream in the extended class I subregion. Butrophylin genes (*BTN)* are also present in the chicken MHC-B locus (Shiina et al. 2007). Annotated gene orthologs in this contig not occurring on human Chr6 include receptor of activated protein C kinase 1 (*RACK1*) located on Chr5, and zinc finger protein with KRAB and SCAN domains 2 (*ZKSCAN2*) located on Chr16 and L-amino-acid oxidase [*OXLA*, *LAAO*, *LAO*], also known as interleukin 4 induced 1 (*IL4I1*), located on human Chr19. Interestingly, *OXLA* is found linked to *BTN2* within the chicken MHC-*B* locus (Shiina et al. 2007) and is linked to *RACK1* within the green anole genome (AnoCar 2.0). The 3′ end of this contig (SJPD01000006.1) contains 2 orfs (-strand) predicted with the NCBI ORF finder with BLAST hits to the framework gene *TRIM39*.

A second scaffold with orthologous loci to the human classical class I subregion is SJPD01000140.1. This rather

Table 1 Komodo dragon genome contigs in the GS assembly containing annotated transcripts corresponding to orthologs of human MHC-region genes

Contig	Length (bp)	Chr	Scaffold	Predicted loci	Percent repeats	Human MHC subregion/genes
SJPD01000006.1	884,375*	1	6_4	23	8.8	Classical class I
SJPD01000114.1	1,332,393	1	102	49	13.8	Classical class III
SJPD01000117.1	1,308,088	1	105	39	10.2	Classical class I, classical class III
SJPD01000119.1	1,187,035	1	107	38	11.9	Classical class III
SJPD01000140.1	590,929	1	128	13	7.5	Classical class I
SJPD01000167.1	262,987	1	155	10	8.9	TRIM genes
SJPD01000168.1	262,698	Un	156	9	8.9	Classical class II, extended class II
SJPD01000211.1	166,416	1	199	7	17.9	Class II histocompatibility antigens
SJPD01000215.1	134,949	13	103	4	9.5	Classical class II
SJPD01000217.1	140,475	1	205	8	12.8	Class II histocompatibility antigen
SJPD01000223.1	150,566	1	211	5	16.4	Class II histocompatibility antigen
SJPD01000224.1	123,899	Un	212	4	12.9	Extended class II
SJPD01000248.1	99,590	1	240	4	7.1	Class I histocompatibility antigens, MR1
SJPD01000309.1	80,374	1	301	1	12.5	Class II histocompatibility antigen
SJPD01000351.1	65,419	1	343	2	19.1	Class II histocompatibility antigen
SJPD01000353.1	68,976	1	345	1	12.6	Class I histocompatibility antigen
SJPD01000422.1	55,258	Un	414	2	5.1	Class I histocompatibility antigen
SJPD01000575.1	40,460	Un	567	1	12.5	Class I histocompatibility antigen
SJPD01000805.1	13,170	Un	797	1	4.6	Class II histocompatibility antigen
SJPD01000984.1	17,623	Un	1,033	2	9.0	Class I histocompatibility antigen
MHC related						
SJPD01000156.1	339,576	1	144	6	12.9	MHC class I-related gene protein (MR1)
SJPD01000455.1	33,428	Un	447	1	13.9	MHC class I-related gene protein (MR1)

*Includes 884,375 bp of 21,484,375 bp total contig length

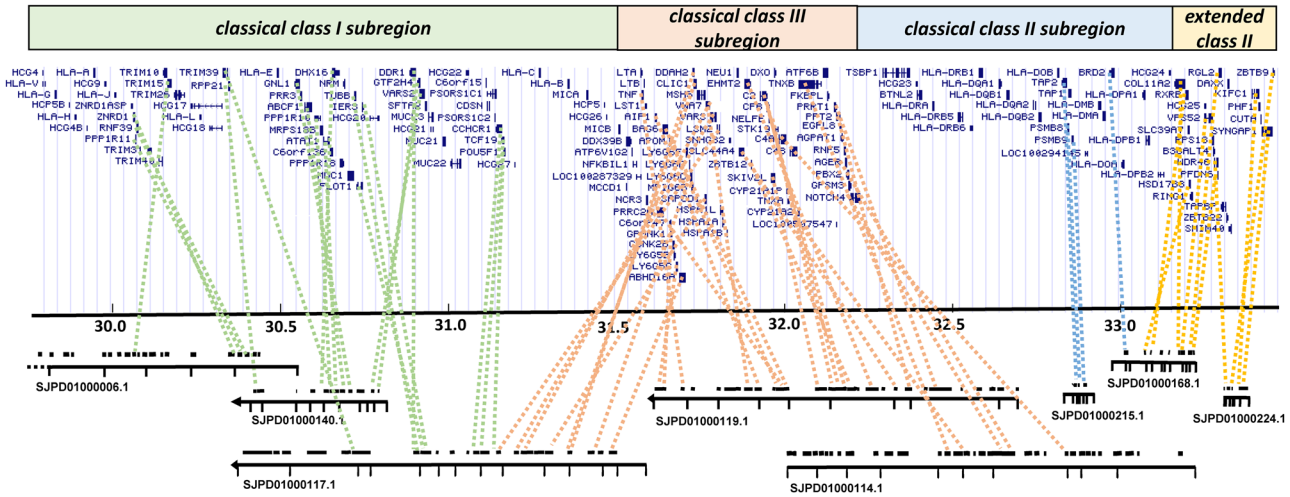

Fig. 1 Idiogram of sequence contigs of the Komodo dragon (Gladstone, GS assembly) aligned to the human MHC. Dashed lines connect predicted gene orthologs. GS contigs are drawn to scale

compact 590 kb contig contains 12 annotated genes and one hypothetical protein (Supplementary Table S1, Supplementary Fig. 1). With one exception (*TPRN*, taperin, human Chr19), each of the named genes within this contig have a human MHC ortholog. Order of the genes within this block is highly conserved, but inverted relative to the human. Within the GS assembly, *DHX16* (pre-mRNA-splicing factor ATP-dependent RNA helicase DHX16) is annotated as two transcripts and the exons for alpha-tubulin N-acetyltransferase 1 (*ATAT1*) are incorrectly annotated as it spans the neighboring gene, *PPP1R10*.

The contig SJPD01000117.1 contains clusters of orthologous MHC genes found in the human classical class I and classical class II subregions (Fig. 1). Notably, these include three C-type lectin domain family 2 genes. C-type lectins are a protein superfamily of proteins that can act as important signaling molecules in the innate immune response (Fujita et al. 2004; Bermejo-Jambrina et al. 2018; Brown et al. 2018) and are a component of the chicken MHC (Miller and Taylor 2016). In this regard, the Komodo lectin-like sequences have closest BLAST similarity to the *BLEC1* gene of the chicken (~ 36% protein identity). The MHC orthologs of the SJPD01000117.1 contig are interspersed by genes, such as *TRAF1*, *NPDC1* found on human Chr9 and *CLEC2B*, *CLEC2D*, and *KLRF1* found on human Chr12. KAF7236218.1, a gene with similarity to *ABHD16A* (abhydrolase domain containing 16A, phospholipase), occurs at the start of this contig (class I subregion) and KAF7236205.1 with similarity to *TRIM39* (class I subregion) occurs towards the opposite end. The contig also includes a *BTN1a1* gene (human extended class I subregion) and an ortholog for *SIRT5* that occurs on human Chr6 but outside the extended MHC boundary (13,570,123–13,619,252 bp). Only a single transcript within the contig is annotated as a hypothetical protein.

Linkage of the gene clusters within SJPD01000117.1 is consistent with the class I/III linkage reported in the green anole (Jaratlerdsiri et al. 2014). Of the MHC gene clusters, one includes *IER3*, *FLOT1*, *TBB7*, and *DDR1* transcripts (Supplementary Table S1). These genes in the human MHC are in close proximity to genes in the classical class I subregion as discussed previously. A second cluster includes *LSM2*, *POU5F3* (likely a *POU5F1* ortholog), *TCF19*, and *CCHCR1*. These genes, with the exception of *LSM2,* are also in the classical class I subregion. *LSM2* lies within the classical class III subregion as do genes in the third cluster (*LY6G6E*, *CLIC1*, *SAPCD2* (likely a *SAPCD1* ortholog), *VWA7*, and *VARS*).

In addition to SJPD01000117.1, two contigs (SJPD01000119.1, SJPD01000114.1) contain orthologs to genes of the human class III subregion (Fig. 1). The first contains 19 annotated MHC orthologs, plus transcripts annotated as non-MHC genes found on human Chr1 (*ZSCAN20*), Chr5 (*TRIM7*), and Chr9 (*CRAT*, *RGS3*, and *TRAF1*) (Supplementary Table S1). The 5′ end of this contig contains two genes (*BTN2A2* and *TRIM27*) located on human Chr6 upstream in the extended class I subregion.

Similar to SJPD01000119.1, the second contig (SJPD01000114.1) contains a suite of annotated class II subregion genes. These include *TNXB*, *CYP21*, *SEC22B-b*, *GPSM3*, *EGFL8*, *AGPAT1*, *RNF5*, *PBX2*, and *C4* (Supplementary Table S1). Transcripts annotated as non-MHC genes include *ZFP2*, *CENPA*, *ZSCAN2*, the *ZNF* genes (*ZNF16*, *79*, and *84*) and *ZBED9*, a gene found in the human extended class I subregion. Also included in this contig are 25 hypothetical protein genes. The majority of these flank

the MHC-associated gene blocks and have BLAST similarity to zinc finger proteins (ZFPs). Four hypothetical proteins, clustered between *CENPA* and *SEC22B-b*, have similarity to *NOTCH4* which is found syntenic within this gene cluster in the human MHC (Fig. 1). BLAST analysis of the hypothetical protein sequences did not find an ortholog to the immunoproteasome gene (proteasome 20S subunit beta 10, *PSMB10*) in the hypothesized ancestral location of this gene adjacent to *C4* (Ohta et al. 2006).

Three contigs (SJPD01000215.1, SJPD01000168.1, and SJPD01000224.1) contain orthologs to genes of the human classical class II and extended class II subregions (Fig. 1, Table 1). The smallest contig SJPD01000215.1 (135 kb) contains transcripts annotated as *ABCB9* (ATP-binding cassette sub-family B member 9), the inducible immunoproteasome genes *PSMB8* and *PSMB9* (Proteasome subunit beta type-8 and type-9) and *TAP2* (antigen peptide transporter 2) and aligns with a syntenic block within the human classical class II subregion. Closer examination of the SJPD01000215.1 sequence identified a likely unannotated *TAP1* gene between *PSMB8* and *PSMB9*. Flanking these central core genes were unannotated orfs (- strand) predicted using the NCBI ORF finder. At the 5′ end of the contig were 2 orfs with BLAST hits to craniofacial development protein 2 (*CFDP2*) and at the 3′ end, 2 orfs (-strand) were identified with BLAST hits to class I histocompatibility antigens. Linkage of the immunoproteasome genes to class I genes is a common feature of non-mammalian MHC loci (Ohta et al. 2006).

The largest contig in the class II region (SJPD01000168.1, 262.7 kb) contains 9 annotated genes including *BRD2*, *HSD17B8*, *RXRBA*, collagen alpha chain transcripts (*COL11A*, *COL5A*), *RGL2* and *PFDN6* (Supplementary Table S1). The predicted transcript of the single hypothetical protein had BLAST similarity to *VPS52* (vacuolar protein sorting-associated protein 52). With the exception of *BRD2* these are genes of the extended class II subregion. A third small contig (SJPD01000224.1, 123.9 kb) contains 4 annotated genes; *WDR46*, *ZBTB22*, *CTK2*, and *SYNGAP1*. The annotation of *CTK2* is problematic and BLAST search of this predicted transcript indicates similarity to kinesin-like protein (*KIFC1*). The orthologous gene lies adjacent to *SYNGAP1* at the border of the extended class II region in the human MHC.

Classical class I and II gene clusters

MHC class I genes

The ability to counter a wide variety of pathogens is in part attributable to diversity in classical class I and class II MHC genes that typically display high allelic polymorphism and sequence diversity. Classical MHC I molecules present antigenic peptide ligands on infected cells to cytotoxic ($CD8^+$) T cells or exogenous proteins through cross-presentation. Components of T cell subsets ($CD8^+$ and $CD4^+$) have been identified in reptiles (reviewed in Zimmerman 2020). MHC class I molecules consist of α and β_2-microglobulin (*B2M*) peptide chains and vertebrate genomes typically possess multiple class I genes (both classical and non-classical).

Our search of the GS assembly identified 6 annotated class I genes or partial coding regions on 5 small contigs (SJPD01000248.1, SJPD01000353.1, SJPD01000422.1, SJPD01000575.1, and SJPD01000984.1, Table 1). These contigs ranged in size from 17.6 to 99.6 kb and contained an average of 2 annotated genes. Included in the class I genes were three transcripts annotated as class I histocompatibility antigen, F10 alpha chain (*HA1F*), 2 designated as RLA class I histocompatibility antigen, alpha chain 11/11 (*HA1A*), and one as H-2 class I histocompatibility antigen, K-K alpha chain (*H2-K1*) (Supplementary Table S1).

The *B2M* gene is not annotated in the GS assembly. However, BLAST search of the genome using the *B2M* mRNA sequence of *Crocodylus porosus* identified sequence similarity to *B2M* in contig SJPD01000079.1 (SLA01 scaffold67). This match corresponded to a sequence annotated as *TRIM69* and was located adjacent to *SORD* and *TERB2*, syntenic with *B2M* on human Chr15. Location of *B2M* outside of the MHC in the Komodo dragon is consistent with that of most vertebrates (Kaufman 2018).

MHC class II genes

The classical class II genes are integral players in the adaptive immune response in that they function to present exogenous proteins to CD4 + T cells. The class II molecule is a heterodimer consisting of an alpha and a beta chain, each encoded by separate genes composed of 5 exons. Exon 1 encodes the leader peptide, exons 2 and 3 encode extracellular domains, exon 4 encodes the transmembrane domain and exon 5 the cytoplasmic tail. Determining the number of class II genes from whole genome sequences can be difficult due to highly conserved gene segments and the presence of multiple alleles. As such, determination of the number of genes is often only accomplished through large insert clone sequencing. Our search of the GS assembly identified 12 transcripts annotated as class II beta genes (or partial CIIB coding regions) on 6 assembled contigs (SJPD01000211.1, SJPD01000217.1, SJPD01000223.1, SJPD01000309.1, SJPD01000351.1, and SJPD01000805.1, Table 1). Alignment of partial class IIβ gene transcripts (Fig. 2) showed highly variable exon 2 sequences supporting presence of multiple genes/alleles.

The number of class I and class IIβ genes appear to be highly variable in non-mammals with substantial duplication in some taxa. Teleost fish possess three major groups

of class II genes (Dijkstra et al. 2007, 2013). The MHC of *Xenopus* contains a single class Iα gene and 3 class IIβ genes (Kobari et al. 1995; Ohta et al. 2019). In contrast, the green anole (*A. carolinensis*) appears to only have a single class IIβ gene (Alfoldi et al. 2011). Multiple class I and class II genes are present passerine (eg. zebra finch, Balakrishnan et al. 2010) and galliform birds (e.g., chicken, Kaufman et al. 1999). Nine class I and 6 class II genes were reported for the saltwater crocodile (Jaratlerdsiri et al. 2014) and multiple cIIβ genes are present in alligators (St John et al. 2012). Miller et al. (2015) found a total of 7 class I sequences and 11 class IIβ sequences in the Tuatara, a rhynchocephalian reptile and Glaberman et al. (2009) identified 8 class IIβ sequences assignable to five locus groups in the Galápagos marine iguana (*Amblyrhynchus cristatus*). Thus, identification of multiple class I and class II genes in the Komodo dragon is consistent with those of other reptilian groups.

Other contigs

Three additional contigs were identified in the GS assembly that contained potential MHC-associated genes (Table 1). The first (SJPD01000156.1) is 339 kb contig that contains transcripts annotated as zinc finger protein with KRAB and SCAN domains 7 (*ZKSCAN7*), major histocompatibility complex class I-related gene protein (*MR1*), and vomeronasal type-2 receptor 26 (*Vmn2R26*) (Supplementary Table S1). *ZKSCAN* and *MR1* are not part of the human MHC (Chr 3 and Chr 1, respectively). In addition, there are three loci annotated as hypothetical proteins that have BLAST similarity to *Vmn2R26;* a mouse gene with no human ortholog. Presence of multiple *Vmn2R-like* transcripts may reflect the expansion of type 2 vomeronasal receptors in the Komodo dragon and several other squamate reptiles (Lind et al. 2019).

The *MR1* transcript in SJPD01000156.1 is one of 6 transcripts annotated as *MR1* identified in our queries of the GS assembly (Supplementary Table S1). A second contig (SJPD01000455.1, 33.4 Kb) contains a single annotated *MR1* transcript (Supplementary Table S1). In humans, *MR1* is located outside the MHC on Chr1 and encodes a non-classical MHC class I antigen-presenting molecule that presents metabolites of microbial vitamin B to mucosal-associated invariant T-cells (Kjer-Nielsen et al. 2012). Genes closely related to class I genes

Fig. 2 Alignment of deduced amino acid sequences for the peptide binding regions (exon 2) and exon 3 of the Komodo dragon MHC class IIβ loci

are found outside the MHC in other non-mammals (Flajnik et al. 1993; Briles et al. 1993). Orthologs of *MR1* have not been identified in non-mammalian vertebrates (Kaufman 2018) and annotation of these in the GS assembly is dubious. BLAST searches (blastP) of the NCBI nr database found significant hits of the Komodo dragon sequences (> 50% identity) to predicted amino acid sequences of class I antigen genes found in other reptiles suggesting they perhaps represent non-classical class I genes.

Like *MR1*, *CD1* is related to MHC class I and class II molecules but is structurally more closely related to class I and present lipid antigens to T cells. Considered a third family of antigen-presenting molecules, CD1 molecules were found in the genomes of the green anole and members of Crocodylia suggesting a common presence in reptiles (Yang et al. 2015). In these species, *CD1* genes are either found linked to the MHC or to an MHC paralogous locus. A *CD1* ortholog was not identified in the Komodo dragon MHC contigs and queries (nucleotide and protein) of both genome assemblies failed to identify sequences with significant similarity to *CD1* genes of chicken, *Xenopus*, *Anolis*, or crocodilians.

The final contig identified in our study (SJPD01000167.1) contains transcripts of four genes annotated as *TRIM10*, *TRIM27*, *TRIM39*, and *LY6G6C*. Orthologs of each of these are found on human Chr6 and all, except *TRIM27* (extended class I), are within the human MHC classical class I or class III subregions. Also present in the contig are *FBXL15* (human Chr10) and *SAA1* (Chr11).

Comparison of genome assemblies and organization of the MHC

Components of the Komodo dragon MHC were identified in both of the recently published genome assemblies (GS and VT). The VT assembly at 1.6 Gb is slightly longer than the GS assembly (1.51 Gb) (Table 2), but the GS assembly has greater sequence depth (144 x vs 45 x, respectively). The number of scaffolds and contigs also differ between the two assemblies; however, this is in part due to the use of different length cutoffs for contigs included in the final assemblies (> 10 kb in GS and > 1 kb in VT). The GS assembly is the NCBI annotated genome reference.

In general, the MHC gene clusters identified in the GS assembly were present in the VT assembly providing support for conservation of these syntenic blocks as opposed to assembly artifacts. Within the VT assembly, 18 contigs were identified with predicted gene orthologs corresponding to the gene set of the human MHC (Supplementary Table S2). These ranged in size from 6 kb (VEXN01032169.1) to 2.7 Mbp (VEXN01011230.1) with an average of 770 kb. The number of annotated genes per contig ranged from 2 to 69 with an average of 17.3 genes per contig. Seven contigs (VEXN01000097.1, VEXN01001601.1, VEXN01008528.1, VEXN01011230.1, VEXN01019096.1, and VEXN01030207.1) could be aligned to the human MHC based on gene content (Supplementary Table S2, Supplementary Fig. 2).

Table 2 Assembly statistics for the genome of the Komodo dragon. Included for comparison are the assemblies compiled by the Gladstone Institute (GS, ASM479886v1) and Virginia Tech (VT, VCOM_VKom_1.0). Lengths are in nucleotides

	GS	VT
Total sequence length	1,507,945,839	1,606,418,949
Total ungapped length	1,493,307,389	1,599,505,109
Number of scaffolds	1,411	84,067
Scaffold N50	23,831,982	23,221,552
Scaffold L50	17	17
Number of contigs	20,696	154,978
Contig N50	189,463	37,135
Contig L50	2,367	12,348
Coverage	144x	45x
Protein sequences	18,293	17,213*

*Protein coding sequences based on the annotation of van Hoek et al. (2019)

Comparison of the VT contigs with the homologous GS sequences found general support for the contig assembles between the genome builds. The Oxford Grid is a useful approach to examining conserved synteny between species or in this case separate genome assemblies (Edwards 1991). Oxford Grids for the Komodo dragon show that complete and near linear alignment of the GS contigs within five of the longer VT contigs with only minor sequence inversions (Fig. 3). Although gene contents are very similar, assembly of two GS contigs (SJPD01000168.1 and SJPD01000224.1) are considerably different than their VT counterparts (VEXN01019096.1 and VEXN01022374.1, respectively). To investigate this further, we created non-overlapping 10kbp fragments of these contigs, aligned the two fragment sets using blastN and then filtered the hits to those at > 90% identity. In this analysis, VEXN01022374.1 (158,440 bp) vs SJPD01000224.1 (123,899 bp) gave 94,820 bp in 34 fragments > 1 k or 71,879 bp in 13 fragments > 2.5 k. VEXN01019096.1 (384,414 bp) vs SJPD01000168.1 (262,698 bp) gave 219,380 bp in 92 fragments > 1 k (or 149,877 bp in 34 fragments > 2.5 k). These results and the alignments (Fig. 3) show the contig pairs are highly fragmented compared to each other with many instances of inversions and translocations.

Our analyses rely on the quality of the annotation of the GS genome build and key in this process is gene prediction and homology identification. We performed independent comparisons of the MHC-associated GS gene set with the Universal Protein Resource (UniProt) database via BLAST which generally confirmed our identification of MHC

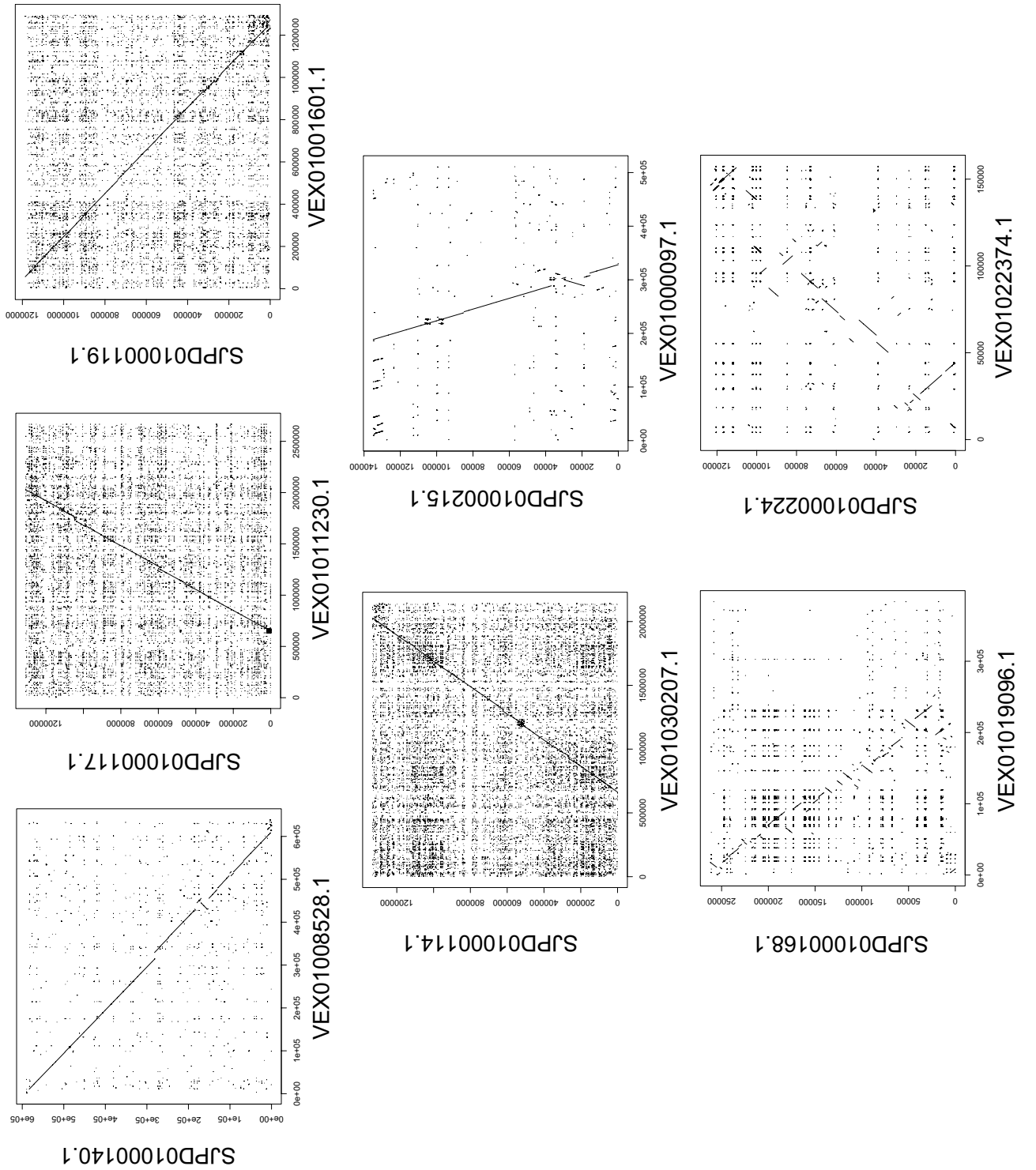

Fig. 3 Sequence Oxford Grid dot plots for alignment of MHC contigs of the Gladstone (GS) assembly (vertical) as depicted in Fig. 1 to those of the Virginia Tech (VT) assembly (horizontal)

orthologs in the Komodo dragon genome. Of the 229 genes included in Supplementary Table S1, only 22 lacked UniProt similarity matches. This is not unexpected given the taxonomic placement of *Varanus* and the mammalian bias of the database.

Having an independent second genome assembly (VT) also provided the opportunity to test for congruence. Although similar in approach, annotation of the two genome assemblies used slightly different pipelines. Comparison of the annotated gene lists for the two Komodo genome

assemblies (GS vs VT BLAST analysis) found significant matches for the majority of genes included in the MHC-associated contigs. Of the genes included in Supplementary Table S1, 207 had significant BLAST similarity to at least one VT annotated gene and gene designations were generally shared for the MHC orthologs. Of the 22 without significant matches, 8 are annotated as hypothetical proteins. Some genes are clearly miss-annotated, perhaps as a result of assembly anomalies. For example, KAF7236117.1 is annotated as replicase polyprotein 1a. This protein is a multifunctional viral protein involved in the transcription and replication of viral RNAs. BLAST search with the KAF7236117.1 amino acid sequence (*Varanus* excluded) found significant similarity to hypothetical proteins in other reptilian species. Three loci that did not have significant VT matches are clustered at the end of the contig SJPD01000167.1 (Supplementary Table S1). KAF7235319.1 is annotated as spindle pole body protein pcp1. This protein is a component of the fission yeast (*Schizosaccharomyces pombe*) spindle pole body that binds calmodulin (Ohta et al. 2012). KAF7235313.1 is annotated as serum amyloid A protein (*SAA1*) which is a highly conserved acute-phase protein. However, BLAST search found similarity to lymphocyte antigen 6 complex locus protein G6c-like (*LY6G6C*) that is adjacent to this locus (KAF7235314.1) in contig SJPD01000167.1. Finally, KAF7235318.1 is annotated as *RER4* (Protein RETICULATA-RELATED 4, chloroplastic). The RER proteins are plant-specific components of the envelope membranes of chloroplasts (Pérez-Pérez et al. 2013). Our BLAST search suggests that this locus is also a fragment of *LY6G6C*. Expression of these three loci was not observed in our examination of de novo-assembled transcripts from the Komodo leukocyte-enriched RNAseq data of Bishop et al. (2017) (data not shown).

Hive plots are visualization tools most commonly used to depict relationships within networks. We used hive plots to highlight our comparison of the Komodo dragon MHC clusters with that of the human MHC and also contrasts the two genome builds. Three-way alignment between human chr6, Gladstone (GS), and Virginia Tech (VT) Komodo dragon assemblies (Fig. 4) demonstrates the syntenic gene clusters identified between the two species and the high degree of concordance between the two genome assemblies. Evident in this plot are positional rearrangements resulting either from translocations, misidentification of orthologs or assembly differences. The latter is most notable between the GS and VT assemblies for the class I subregion (yellow lines in Fig. 4). Because the contigs assigned to chr1 are not ordered in the GS or VT assemblies, we used reversed coordinates of contigs on the GS arm where the blastN hits aligned to a single VT contig and concordance could be visually improved. We also reversed coordinates of both GS and VT contigs where appropriate to give better concordance with the human chr6. The alignment of the fragment between the two assemblies suggests the GS assembly could be improved by creating a super scaffold by combining SJPD01000006.1, SJPD01000167.1, SJPD01000140.1, and SJPD01000117.1 (Fig. 4). This approach significantly improved the concordance while also confirming the presence of assembly differences in this region highlighted in the Oxford grids (Fig. 3).

The Komodo dragon has a very typical reptilian karyotype consisting of 2n = 40 chromosomes with 8 pairs

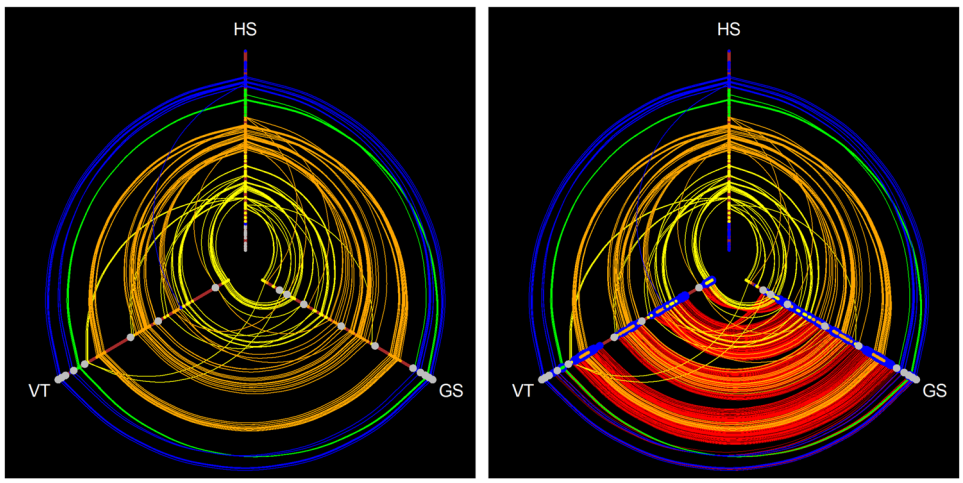

Fig. 4 Hive plots of three way alignment between human chr 6, Gladstone (GS) and Virginia Tech (VT) Komodo dragon assemblies. On the left MHC genes are highlighted as follows: yellow = class I subregion, orange = class III subregion, green = class II subregion while flanking genes are in blue. Grey nodes are contig ends. On the right, the plot is augmented using pairwise links of blastN hits corresponding to 10 k segments from the corresponding GS/VT contigs shown as red lines. These are filtered by bitscore > 1000, percent id > 90%, and hit length > 2500 bp. Contig order and direction is as given in the text

of macrochromosomes, 12 pairs of microchromosomes and a Z/W sex chromosome system (Pokorna et al. 2016). Although it is difficult to draw definitive conclusions about the overall MHC structure, gene content of the GS contigs supports class I/III linkage as observed in SJPD01000117.1. Based on the chromosome assignments of scaffolds in the GS assembly (Lind et al. 2019), the MHC contigs we identified reside on at least 2 chromosomes (1 and 13). The majority of contigs ($n = 14$) are assigned to Chr1, a single contig (SJPD01000215.1, scaffold 103) is assigned to Chr13, with 7 contigs unassigned. Physical chromosomal assignment via FISH and the relative order of these contigs within the genome have not been experimentally determined. As such, chromosome relationships among the MHC gene clusters with orthology to the human MHC subregions (I, II, and III) are unresolved in the Komodo dragon.

The lack of contigs in the GS assembly containing both classical class I and class II genes in combination with MHC framework genes also makes it difficult to summarize the functional organization of the Komodo dragon MHC. In the non-mammalian vertebrates studied to date, the class Iα gene(s) occur within a series of antigen processing genes (Kaufman et al. 1999; Ohta et al. 2006; Jaratlerdsiri et al. 2014). As suggested by Ohta et al. (2006), this arrangement precludes designation of a class I region (as seen in humans) where the antigen processing (*TAP*) genes for example are found within the class II region. Strong linkage disequilibrium is seen between *TAP*, *PSMB*, and class Iα genes in the teleosts medaka and zebrafish (Tsukamoto et al. 2009; McConnell et al. 2016). This gene relationship is extreme in the chicken where the TAP genes are flanked by class I genes and are virtually inseparable by recombination (Kaufman et al. 1999). This functional clustering also appears to be present in the Komodo dragon. As discussed above, the GS contig assigned to Chr13 (SJPD01000215.1) contains transcripts of immunoproteasome and antigen processing genes (*ABCB9*, *PSMB8*, *PSMB9*, *TAP2*, and likely *TAP1*) and unannotated putative class I ORFs. This gene cluster in the VT assembly is positioned on a 515 Kb contig (VEXN01000097.1, scaffold ScpDV4C_95) that is flanked by multiple class I transcripts, consistent with the functional organization observed in other non-mammalian vertebrates (Kaufman et al. 1999; Ohta et al. 2006; Jaratlerdsiri et al. 2014).

Repetitive elements, especially transposable elements, are important in genome evolution as they can facilitate large-scale changes (Jurka et al. 2007) including those within the MHC (Kulski et al. 1997; Reed et al. 2011). In the Komodo dragon genome, repetitive elements are estimated to account for 32% of the assembled sequence (Lind et al. 2019). The majority of identified repeats were transposable elements (LINE2, L3/CR1, 13% of genome) or unclassified (11%). We used RepeatMasker (v4.0.9) to screen for repeats within the MHC-associated contigs of the GS assembly. The percent sequence denoted as repeats within these contigs ranged from 5.1 to 19.1% with an average of 11.2% (Table 1). On average 76% of this repeat sequence corresponded to retroelements. This finding indicates the MHC-associated regions have on average less repetitive DNA than the genome as a whole. Understanding the potential role of repetitive DNA in MHC evolution within the species necessitates further study.

Conclusions

The recent completion of two whole genome assemblies of the Komodo dragon (*V. komododensis*) allowed for the analysis of scaffolds and contigs containing gene clusters corresponding to the MHC subregions in the human. We found the assembled genome to include 20 MHC-related contigs encompassing ~6.9 Mbp of sequence with 223 annotated genes/orfs plus 2 contigs with transcripts that may represent non-classical class Iα genes. The annotated MHC genes include loci involved in antigen processing and presentation, complement, inflammation, immune regulation as well as genes with non-immune functions and hypothetical proteins of uncertain orthology. The evolutionarily ancient varanid reptiles, including Komodo dragons have evolved robust innate immune systems (Lind et al. 2019). The organization of the Komodo dragon MHC resembles that of other non-mammalian taxa. Our analysis of MHC gene clusters finds a gene dense and complex region(s) that contain counterparts of the human MHC and provides insight into the MHC of these unique squamate reptiles.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s00251-021-01217-6.

Author contribution Both authors contributed to all phases of this work and read and approved the final manuscript.

Funding No funding was received for conducting this study.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Alfoldi J, Di Palma F, Grabherr M et al (2011) The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 477:587–591

Amadou C (1999) Evolution of the Mhc class I region: the framework hypothesis. Immunogenetics 49:362–367

Balakrishnan CN, Ekblom R, Völker M, et al (2010) Gene duplication and fragmentation in the zebra finch major histocompatibility complex. BMC Biol 8:29

Belov K, Deakin JE, Papenfuss AT et al (2006) Reconstructing an ancestral mammalian immune supercomplex from a marsupial Major Histocompatibility Complex. PLoS Biol 4:317–328
Bermejo-Jambrina M, Eder J, Helgers LC et al (2018) C-type lectin receptors in antiviral immunity and viral escape. Front Immunol 9:590
Bishop BM, Juba ML, Russo PS et al (2017) Discovery of novel antimicrobial peptides from *Varanus komodoensis* (Komodo dragon) by large-scale analyses and de-novo-assisted sequencing using electron-transfer dissociation mass spectrometry. J Proteome Res 16:1470–1482
Briles WE, Goto RM, Auffray C, Miller MM (1993) A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. Immunogenetics 37:408–414
Brown GD, Willment JA, Whitehead L (2018) C-type lectins in immunity and homeostasis. Nat Rev Immunol 18:374–389
Campbell RD, Trowsdale J (1993) Map of the human MHC. Immunol Today 14:349–352
Ciofi C (1999) The Komodo dragon. Sci Am 280:84–91
Dijkstra J, Katagiri T, Hosomichi K et al (2007) A third broad lineage of major histocompatibility complex (MHC) class I in teleost fish; MHC class II linkage and processed genes. Immunogenetics 59:305–321
Dijkstra JM, Grimholt U, Leong J, Koop BF, Hashimoto K (2013) Comprehensive analysis of MHC class II genes in teleost fish genomes reveals dispensability of the peptide-loading DM system in a large part of vertebrates. BMC Evol Biol 13:260
Edwards JH (1991) The Oxford grid. Ann Hum Genet 55:17–31
Flajnik MF, Kasahara M, Shum BP, Salter-Cid L, Taylor E, Du Pasquier L (1993) A novel type of class I gene organization in vertebrates: a large family of non-MHC-linked class I genes is expressed at the RNA level in the amphibian *Xenopus*. Embo J 12:4385–4396
Fujita T, Matsushita M, Endo Y (2004) The lectin-complement pathway–its role in innate immunity and evolution. Immunol Rev 198:185–202
Glaberman S, Moreno MA, Caccone A (2009) Characterization and evolution of MHC class II B genes in Galapagos marine iguanas (*Amblyrhynchus cristatus*). Dev Comp Immunol 33:939–947
Harris RS (2007) Improved pairwise alignment of genomic DNA. Ph.D. Thesis, The Pennsylvania State University
Hanson BA (2020) HiveR: 2D and 3D Hive Plots for R. R package version 0.3.63
Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491
Horton R, Wilming L, Rand V et al (2004) Gene map of the extended human MHC. Nat Rev Genet 5:889–899
Jaratlerdsiri W, Deakin J, Godinez RM et al (2014) Comparative genome analyses reveal distinct structure in the saltwater crocodile MHC. PLoS One 9:e114631
Jurka J, Kapitonov VV, Kohany O, Jurka MV (2007) Repetitive sequences in complex genomes: structure and evolution. Annu Rev Genomics Hum Genet 8:241–259
Kasahara M, Hayashi M, Tanaka K et al (1996) Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. Proc Natl Acad Sci USA 93:9096–9101
Kasahara M, Nakaya J, Satta Y, Takahata N (1997) Chromosomal duplication and the emergence of the adaptive immune system. Trends Genet 13:90–92
Katsanis N, Fitzgibbon J, Fisher EM (1996) Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. Genomics 35:101–108
Kaufman J (2018) Unfinished business: Evolution of the MHC and the adaptive immune system of jawed vertebrates. Annu Rev Immunol 36:383–409
Kaufman J, Milne S, Göbel T et al (1999) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401:923–925
Kobari F, Sato K, Shum BP et al (1995) Exon-intron organization of *Xenopus* MHC class II β chain genes. Immunogenetics 42:376–385
Kjer-Nielsen L, Patel O, Corbett AJ et al (2012) MR1 presents microbial vitamin B metabolites to MAIT cells. Nature 491:717–723
Klein J (1986) Seeds of time: Fifty years ago Peter A. Gorer discovered the H-2 complex. Immunogenetics 24:331–338
Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5:59
Kulski J, Gaudieri S, Bellgard M et al (1997) The evolution of MHC diversity by segmental duplication and transposition of retroelements. J Mol Evol 45:599–609
Kumar S, Hedges B (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920
Lind AL, Lai YYY, Mostovoy Y et al (2019) Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. Nat Ecol Evol 3:1241–1252
McConnell SC, Hernandez KM, Wcisel DJ et al (2016) Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. Proc Natl Acad Sci USA 113:E5014-5023
Miller HC, O'Meally D, Ezaz T, Amemiya C et al (2015) Major histocompatibility complex genes map to two chromosomes in an evolutionarily ancient reptile, the Tuatara *Sphenodon punctatus*. G3 (Bethesda) 5:1439–1451
Miller MM, Taylor RL Jr (2016) Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. Poult Sci 95:375–392
Milner CM, Campbell RD (2001) Genetic organization of the human MHC class III region. Front Biosci 6:D914-926
Ohta M, Sato M, Yamamoto M (2012) Spindle pole body components are reorganized during fission yeast meiosis. Mol Biol Cell 23:1799–1811. https://doi.org/10.1091/mbc.E11-11-0951
Ohta Y, Goetz W, Hossain MZ, Nonaka M, Flajnik MF (2006) Ancestral organization of the MHC revealed in the amphibian *Xenopus*. J Immunol 176:3674–3685
Ohta Y, Kasahara M, O'Connor TD, Flajnik MF (2019) Inferring the "Primordial Immune Complex": origins of MHC class I and antigen receptors revealed by comparative genomics. J Immunol 203:1882–1896
Ohta Y, Okamura K, McKinney EC, Bartl S, Hashimoto K, Flajnik MF (2000) Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. Proc Natl Acad Sci USA 97:4712–4717
Papenfuss AT, Feng ZP, Krasnec K et al (2015) Marsupials and monotremes possess a novel family of MHC class I genes that is lost from the eutherian lineage. BMC Genomics 16:535
Pérez-Pérez JM, Esteve-Bruna D, González-Bayón R et al (2013) Functional redundancy and divergence within the *Arabidopsis* RETICULATA-RELATED gene family. Plant Physiol 162:589–603. https://doi.org/10.1104/pp.113.217323
Pokorna MJ, Altmanova M, Rovatsos M et al (2016) First description of the karyotype and sex chromosomes in the Komodo dragon (*Varanus komodoensis*). Cytogenet Genome Res 148:284–291
Pung YF, Wong PT, Kumar PP et al (2005) Ohanin, a novel protein from king cobra venom, induces hypolocomotion and hyperalgesia in mice. J Biol Chem 280:13137–13147

Reed KM, Bauer MM, Monson MS et al (2011) Defining the turkey MHC: identification of expressed class I- and class IIB-like genes independent of the MHC-B. Immunogenetics 63:753–771

Sambrook JG, Figueroa F, Beck S (2005) A genome-wide survey of major histocompatibility complex (MHC) genes and their paralogues in zebrafish. BMC Genomics 6:152

Sambrook JG, Russell R, Umrania Y et al (2002) *Fugu* orthologues of human major histocompatibility complex genes: a genome survey. Immunogenetics 54:367–380

Shiina T, Briles WE, Goto RM et al (2007) Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. J Immunol 178:7162–7172

Shiina T, Hosomichi K, Inoko H et al (2009) The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet 54:15–39

St John JA, Braun EL, Isberg SR et al (2012) Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. Genome Biol 13:415

Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19(Suppl 2):ii215–225

Sutoh Y, Kasahara M (2021) The immune system of jawless vertebrates: insights into the prototype of the adaptive immune system. Immunogenetics 73:5–16

Tsukamoto K, Sakaizumi M, Hata M et al (2009) Dichotomous haplotypic lineages of the immunoproteasome subunit genes, PSMB8 and PSMB10, in the MHC class I region of a Teleost Medaka, *Oryzias latipes*. Mol Biol Evol 26:769–781

The MHC Sequencing Consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. Nature 401:921–923

Trowsdale J, Ragoussis J, Campbell RD (1991) Map of the human MHC. Immunol Today 12:443–446

The UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49:D1

van Hoek ML, Prickett MD, Settlage RE et al (2019) The Komodo dragon (*Varanus komodoensis*) genome and identification of innate immunity genes and clusters. BMC Genomics 20:684

Yang Z, Wang C, Wang T et al (2015) Analysis of the reptile CD1 genes: evolutionary implications. Immunogenetics 67:337–346

Zimmerman LM (2018) Reptilia: Humoral Immunity in Reptiles. *In*: Cooper E. (eds) Advances in Comparative Immunology. Springer, Cham. https://doi.org/10.1007/978-3-319-76768-0_20

Zimmerman LM (2020) The reptilian perspective on vertebrate immunity: 10 years of progress. J Exp Biol 223:jeb214171

Zimmerman LM, Vogel LA, Bowden RM (2010) Understanding the vertebrate immune system: insights from the reptilian perspective. J Exp Biol 213:661–671