

**METHODOLOGY ARTICLE**

**Open Access**

# Evolution of MHC class I genes in the endangered loggerhead sea turtle (*Caretta caretta*) revealed by 454 amplicon sequencing

Victor A Stiebens<sup>1</sup>, Sonia E Merino<sup>2</sup>, Frédéric J J Chain<sup>3</sup> and Christophe Eizaguirre<sup>1,3\*</sup>

## Abstract

**Background:** In evolutionary and conservation biology, parasitism is often highlighted as a major selective pressure. To fight against parasites and pathogens, genetic diversity of the immune genes of the major histocompatibility complex (MHC) are particularly important. However, the extensive degree of polymorphism observed in these genes makes it difficult to conduct thorough population screenings.

**Methods:** We utilized a genotyping protocol that uses 454 amplicon sequencing to characterize the MHC class I in the endangered loggerhead sea turtle (*Caretta caretta*) and to investigate their evolution at multiple relevant levels of organization.

**Results:** MHC class I genes revealed signatures of trans-species polymorphism across several reptile species. In the studied loggerhead turtle individuals, it results in the maintenance of two ancient allelic lineages. We also found that individuals carrying an intermediate number of MHC class I alleles are larger than those with either a low or high number of alleles.

**Conclusions:** Multiple modes of evolution seem to maintain MHC diversity in the loggerhead turtles, with relatively high polymorphism for an endangered species.

**Keywords:** Major histocompatibility complex, Loggerhead sea turtle, Trans-species polymorphism, Reptiles, Intermediate diversity

## Background

All organisms are confronted with diseases, which can be particularly threatening to endangered species that show reduced genetic diversities [1]. In vertebrates, growing evidence suggests that genetic diversity is especially important at the level of the major histocompatibility complex (MHC, [2-4]). Since the primary function of MHC molecules is to present parasite-derived peptides to T-lymphocytes, it has been argued that parasites and pathogens are major selective pressures acting on the evolution of MHC genes [1,2,5,6]. There are two main types of MHC molecules, class I and class II. Both classes of molecules function as shuttles that transport

peptides from the cytoplasm and display them on the cell surface. MHC class I molecules in particular are expressed by nearly all cell types and present peptides that are derived from proteins degraded by the proteasome [7].

The MHC polymorphism is especially high in the region that encodes for the peptide-binding domain. The residues of the  $\alpha 1$  and  $\alpha 2$  domains of the MHC class I molecules form the peptide-binding region. Antigenic peptides are anchored at specific residues called antigen binding sites, which are commonly found to be evolving under positive selection in natural populations (e.g. [8]).

The polymorphism present at the MHC genes has regularly been investigated at multiple levels of organization. Firstly, a very particular feature of MHC genes is the existence of trans-species polymorphism (TSP) which has been observed in various taxa (e.g. [9-12]). TSP can occur either through allelic lineages being maintained over long periods

\* Correspondence: ceizaguirre@geomar.de

<sup>1</sup>Department of Evolutionary Ecology of Marine Fishes, GEOMAR|Helmholtz Center for Ocean Research, Kiel 24105, Germany

<sup>3</sup>Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Ploen 24306, Germany

Full list of author information is available at the end of the article

of time across speciation events [13,14] or through convergent evolution presumably due to similar parasite pressures [15,16]. Secondly, genetic diversity at MHC loci has been used to measure the immunological fitness of wild populations [1]. Although a direct link between pathogen-mediated population decline and low MHC variation has been difficult to demonstrate in natural populations [17], several studies have reported decreased pathogen resistance among MHC homozygotes (reviewed in [5]). Thirdly, at the individual level, MHC diversity has been associated with numerous fitness traits such as secondary sexual ornamentations [18,19], parasitism [20,21], and life time reproductive success [22]. Although patterns are not clear, several studies have found fitness advantages in individuals carrying either an intermediate number of MHC alleles [20,22-24] or a maximum number of alleles (Heterozygote advantage – [21,25,26]).

Despite a tremendous research effort to understand the evolution of MHC genes and their relevance for conservation biology, surprisingly few studies have focused on the group of non-avian reptiles. The best-characterized MHC example in this taxa is that of the Tuatara in which the second exon of the MHC class I is comprised of two sets of duplicated alleles in most individuals [26,27].

In this study, we used 454 deep amplicon sequencing to investigate the variation of the MHC class I alpha-1 heavy chain in a population of the loggerhead sea turtle (*Caretta caretta*) nesting at the Cape Verde archipelago. Next generation sequencing offers new tools to characterize extreme variation within and between individuals. The use of individually barcoded primers during amplification allows the sequencing of PCR products derived from hundreds of individuals in a single 454 experiment, even for dense gene complexes [28-30]. The read length of 454 sequencers also permits coverage of the entire polymorphic exons of the MHC.

The Cape Verde population of loggerhead turtles is the second largest in the Atlantic [31]. Recently, Monzon-Arguello *et al.* [31] revealed the significant genetic divergence between the Cape Verde rookery and other Atlantic and Mediterranean rookeries. Furthermore, Stiebens *et al.* [32] showed strong signs of philopatry at the island level, suggesting a complex structure of the rookery with independent colonies. Additionally, in Cape Verde, the fungus *Fusarium solani* was found to be the cause of infections in turtle eggs that accounted for over 80% of mortality in a challenged experiment [33], supporting the need to characterize immune relevant genes.

In this study, after characterizing the MHC class I  $\alpha$  genes in the loggerhead turtle, we investigate different modes of evolution at different levels of organization from species to individuals.

## Results

### Phylogeny of MHC genes in reptiles

To investigate the phylogenetic coherence between neutral and adaptive markers, we built two phylogenetic trees of reptiles using i) mtDNA control region and ii) MHC class I  $\alpha$  genes. The trees suggest different evolutionary scenarios (Figure 1A & B). On the one hand, the mtDNA control region clearly separates reptile species where each node is supported by high bootstrap values. In contrast, the MHC class I phylogeny is much weaker and mainly separates the outgroup and the *Sphenodon* MHC sequences. Interestingly, the loggerhead turtle shows MHC alleles that display closer allelic relationships between species than within species - suggesting trans-species polymorphism over a large range of reptile species and/or a duplication event prior to speciation.

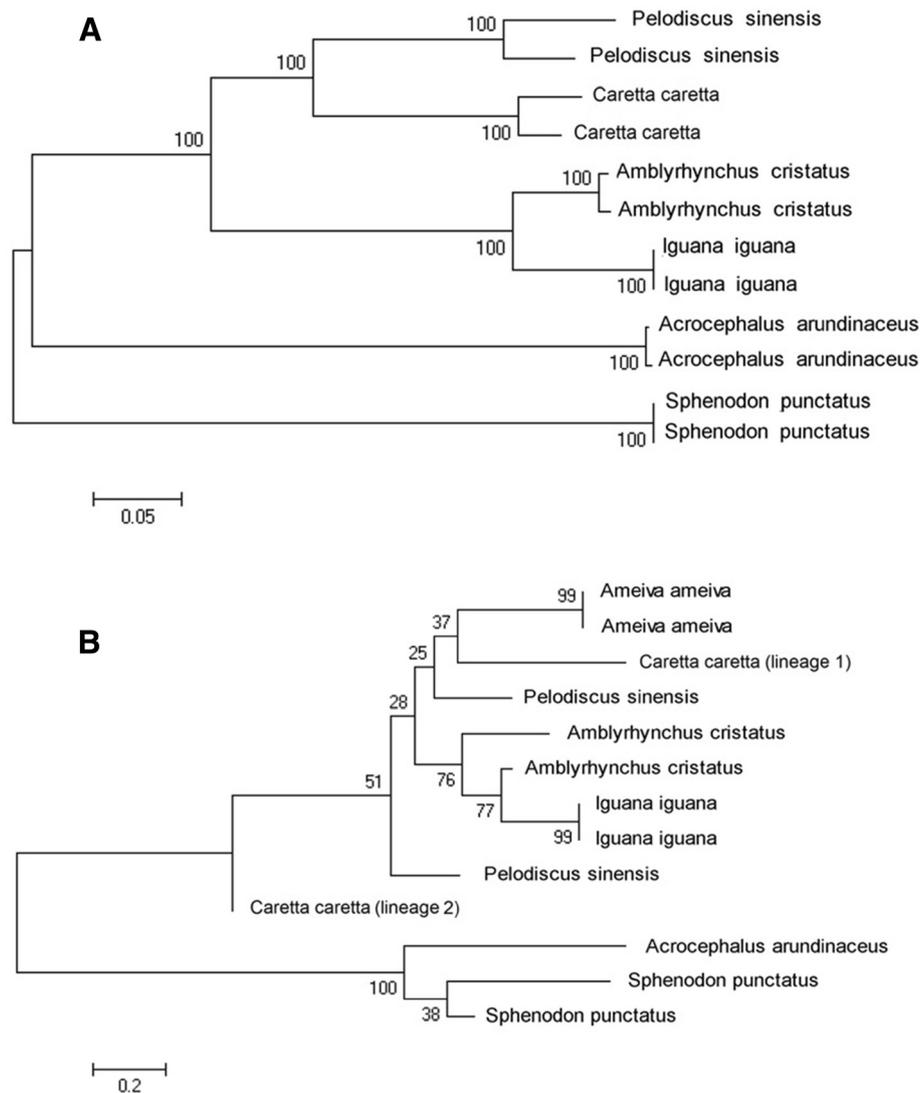
### Phylogeny of MHC in the Cape Verde rookery

The phylogenies within the loggerhead turtle population from Cape Verde based on mtDNA and MHC class I alleles were also discordant. For the mtDNA, we found two strong clusters arising from the presence of an extremely divergent haplotype (CC2 in ACCSTR, <http://accstr.ufl.edu/resources/mtdna-sequences/>) that differs from the other haplotypes from a maximum of 35 point mutations (Additional file 1: Figure S1). As expected from the reptile phylogeny, the MHC neighbor-joining tree identified two main lineages supported by high bootstrap values (Figure 2), which suggests at least one duplication event and/or the maintenance of old allelic lineages. No particular link could be identified between the two phylogenetic trees.

### MHC allelic pool

For the 40 turtles sequenced in this study, we obtained approximately 4100 usable 454 reads. After data filtering (see Methods), 34 different alleles were detected with coverage depths varying between 54 and 106 reads per allele (accession numbers: KF021627-KF021666). Allele abundances within the population varied from 0.025 to 0.275 (Figure 3). We found 12 singleton alleles (i.e. found in one individual), but all alleles were present in both independent PCR reactions.

Out of the of 216 basepair (bp) sequence, bp differences ranged from 1 to 69 with a median of 18 (mean = 34.42  $\pm$  9.89 bp), and from 1 to 32 amino acid changes (median of 11, mean = 16.64  $\pm$  5.65, Additional file 1 document 2). As would be expected under parasite-mediated balancing selection [34], MHC genes in turtles show strong signs of positive selection:  $Z = 1.983$ ,  $p = 0.025$ . Likelihood ratio tests also suggest that several codon sites in the MHC class I gene are evolving under positive selection (Table 1, Figure 4).



**Figure 1 Neighbor-joining trees.** (A) based on the mtDNA control region of five reptile and one bird species. (B) based on the MHC class I based on the same species (Node values (in %) are obtained from 1000 bootstraps). Although (A) shows species clustered together, (B) demonstrates trans-species polymorphism of the MHC class I gene in reptiles.

None of the alleles that appeared in more than one individual were in linkage disequilibrium with one other ( $p = 0.649$ ).

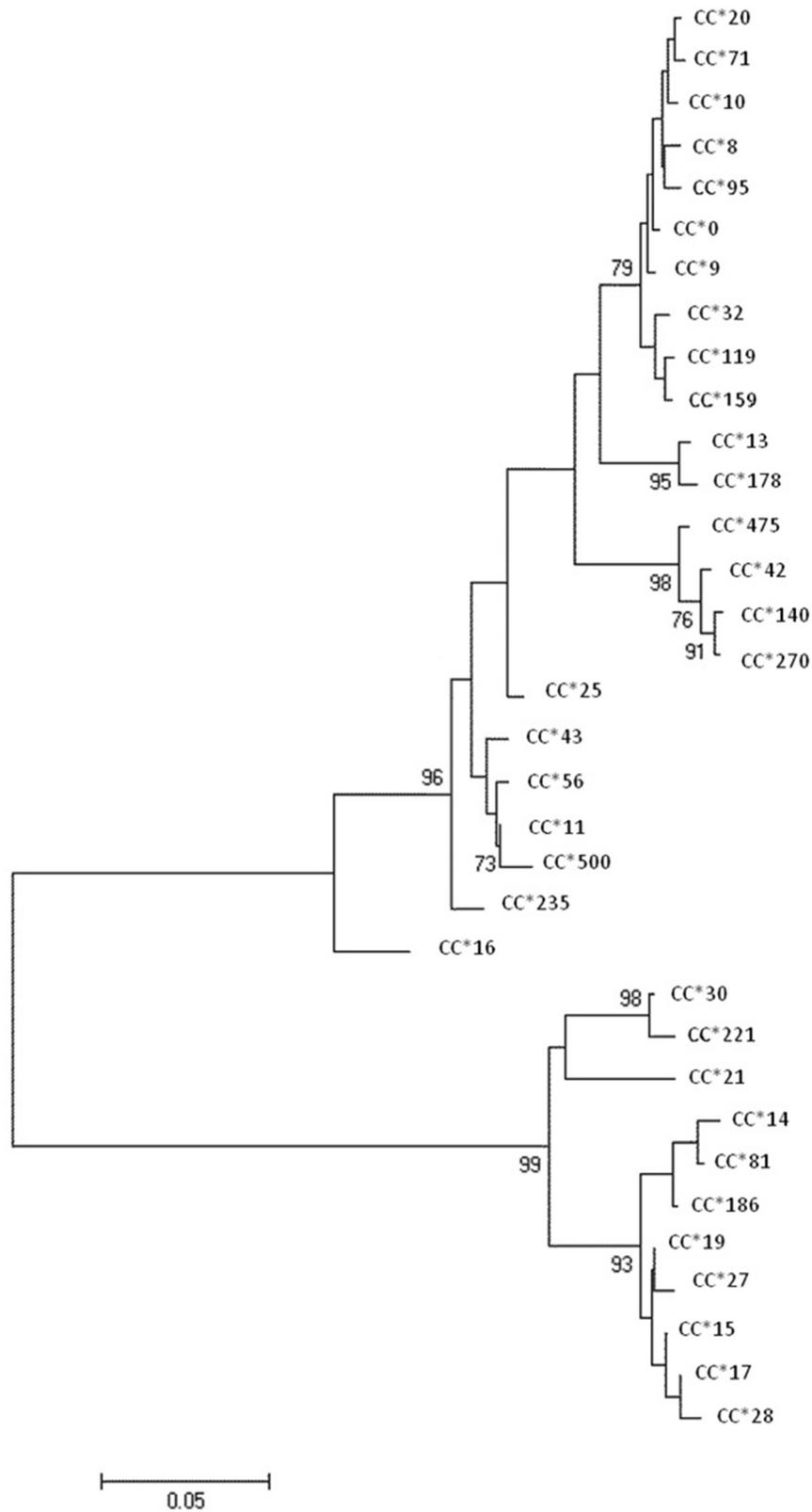
The Hudson four-gamete test [36] implemented in DnaSP [37] detected eight recombination events (RM). These values indicate the minimum number of recombination events in the history of the samples.

Within the 34 alleles found, GENECONV analyses detected six fragments significantly involved in gene conversion events. In addition, the numbers of pairwise internal fragments exceeded the random-assumption of 5% (here, 15.9%) suggesting the occurrence of gene conversion in turtle MHC class I genes.

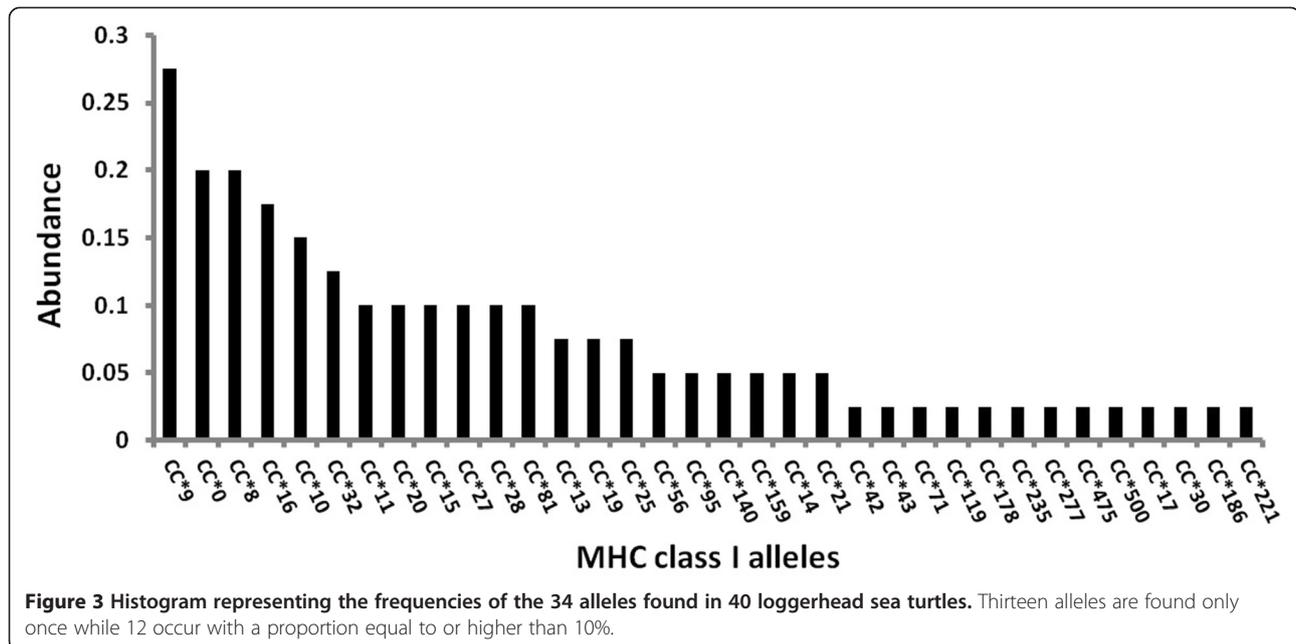
#### Individual MHC allele variation

Individual diversity ranged from one to four alleles (median = 2), indicating the presence of up to four MHC class I loci in this loggerhead population. Out of 7 individuals for which cloning was also performed, 6 genotypes were identical between cloning and 454 sequencing. For the remaining individual, one allele was missing in the cloning approach, but increasing the number of sequenced clones a-posteriori revealed the presence of this allele in this individual (Additional file 1 document 3).

Identifying fitness proxies in sea turtles is difficult but numerous studies have found that larger turtles have higher clutch size [38,39]. Thus, we used individual size



**Figure 2 Neighbor-joining tree based on 1000 bootstraps for all 34 MHC alleles detected.** Three main lineages supported by high bootstrap values are found suggesting gene duplication and/or maintenance of old allelic lineages. Node values are given in percentages, only values higher than 70% are represented.



(by using the residuals of the correlation between the curved carapace length and the curved carapace width) as an estimate of turtle body condition. We found that turtles with intermediate MHC diversity were larger than turtles with either higher or lower number of MHC alleles (Quadratic term, Estimate = -0.194, St.Err. = 0.082, t-value = -2.38, p = 0.023, Figure 5) – suggesting an evolutionary advantage to intermediate MHC diversity.

### Discussion

In this work, we characterized the allelic diversity for genes of the major histocompatibility complex in the endangered loggerhead sea turtle (IUCN 2007). Loggerhead turtles are confronted with multiple direct and indirect anthropogenic threats menacing their genetic diversity – a crucial component of population viability [1]. The MHC genes are not only good estimators of genetic diversity but also play important roles in the onset of the adaptive immune system [5,6]. Here, we used high-throughput genotyping to assess MHC adaptive genetic diversity. Despite the numerous advantages of using next generation sequencing, 454 amplicon sequencing is particularly prone to sequencing errors such as homopolymers [40,41]

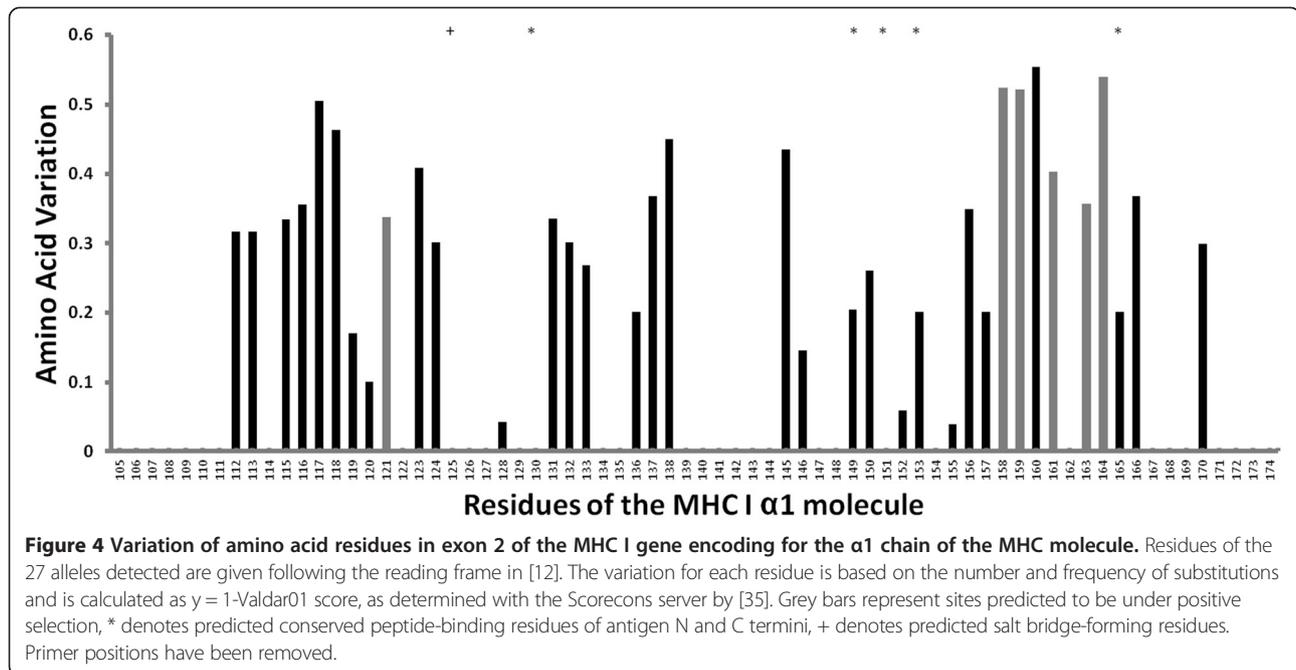
resulting in an increased frequency of indels [29] or to increased number of sequenced chimeras [34]. Nonetheless, the consequences of such effects can be diminished by combining precautionary PCR preparation (reconditioning steps), independent replicate reactions (using differently labeled primers), accurate primer design [42,43] and sufficient depth of sequencing coverage [34]. Following all those recommendations we were able to address the evolutionary history of MHC class I genes in the endangered loggerhead turtles at multiple evolutionary levels.

Firstly, at a large taxonomic range, we found clear species clustering for the mtDNA control region. Even though the mode of inheritance and the evolutionary rates of both mtDNA and MHC markers are different, contrary to the neutrally evolving mtDNA marker, the MHC genes showed a closer relationship between species than within loggerhead turtle alleles suggesting the existence of TSP within the reptile taxa. TSP corresponds to the maintenance of allelic lineages that are passed on during speciation events to each of the newly formed species [44]. TSP has been reported in related iguana species [12], and our results suggest that TSP spans an even larger taxonomic range, which may arise from the slow evolutionary rate of

**Table 1** Table summarizing codon-based tests for positive selection

Likelihood models	Statistical test	P-value	Estimate for $\omega > 1$	Proportion of sites $\omega > 1$	Codon sites $\omega > 1$
M1a vs M2a	18.212	0.00011	3.41529	0.20074	121,158,159
M7 vs M8	23.862	6.58E-06	3.02509	0.26839	121,158,159,161,163,164
M8a vs M8	15.922	3.30E-05			

$\omega = (dN / dS)$  where dN represents the number of non synonymous substitutions and dS the number of synonymous substitutions. Numbers of codon sites refer to position in the reading frame.



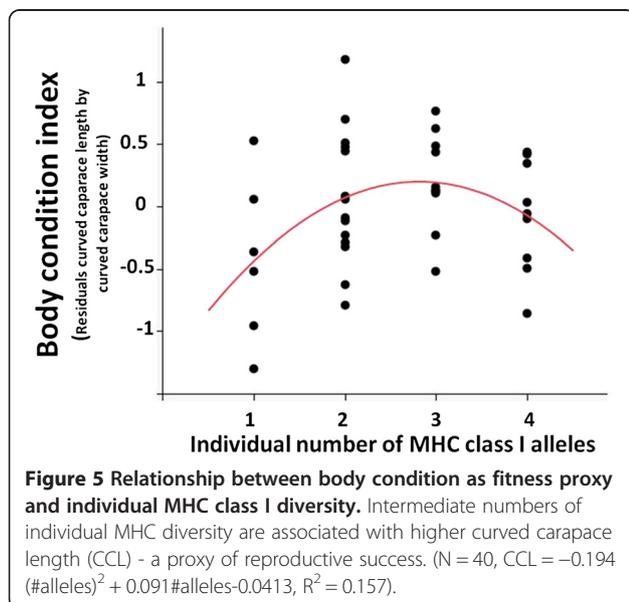
the basal class of reptiles [45] and/ or via long-term balancing selection.

Given the observed signature of TSP, it was therefore not surprising to find that the sequenced MHC alleles in the loggerhead turtle population from Cape Verde clustered into 2 groups supported by high bootstrap values. Genotypes with such diverse MHC alleles are expected to bind more dissimilar antigens that could then favor their maintenance on an evolutionary time scale [44]. Interestingly, several hypotheses have been proposed to explain the maintenance of MHC polymorphism, but, given the function of these genes, parasite-mediated balancing

selection is the most likely driving force (reviewed in [5,6]) as recently shown experimentally [46]. The exceptional allelic diversity usually observed in natural populations, both in terms of the number of specific alleles as well as in terms of amino acid diversity, provides the potential to adapt to a given parasite spectrum.

In the sequenced turtles, we found 34 different alleles suggesting that the MHC class I diversity in the endangered loggerhead turtle is not particularly low compared to other endangered species such as the Namibian cheetah [47] or the European Bison [48]. From a conservation perspective, the fact that numerous individuals carry a unique allelic repertoire indicates the importance of preserving this diversity. Furthermore, our results show that turtles possess up to 4 different MHC alleles, suggesting at least one event of duplication. Since the number of functional MHC loci in the genome represents the bottleneck for adaptation to parasites and pathogens, it might be selectively advantageous to retain duplications at these loci [49]. On an evolutionary time scale, the number of loci within a species is not fixed and may vary over time in a birth-and-death process of gene duplications and deletions [13,50].

It is also worth noting that we found evidence for MHC class I amino acid sites evolving under positive selection. This further supports the view of balancing selection also acting on MHC evolution in turtles. With our dataset, we not only tackled the puzzling evolutionary question of the maintenance of MHC polymorphism but also showed that gene conversion and recombination between copies exist - both playing a role in the generation of high allelic polymorphisms [51,52].



Recombination between loci may explain the occurrence of sequence variants that are particularly divergent, which may then provide particular advantage against parasitic attack. Since many classical MHC genes occur as clusters of functionally intact, duplicated genes, interlocus recombination through unequal crossing-over may also generate sequence polymorphism [53].

Finally, with our dataset we were also able to investigate the relationship between individual MHC diversity and a fitness relevant trait. Identifying relevant fitness traits is complex in marine turtles as reproductive success cannot be followed over generations. Numerous studies, nonetheless, have found that larger turtles achieve a higher clutch size e.g. [38,39]. Here, our results suggest that individuals with an intermediate MHC diversity were larger than those with either high or low diversity. Several studies have reported a relationship between individual MHC diversity and fitness traits, supporting either an advantage for an intermediate diversity [20,24] or for increased heterozygosity [21,25,54]. An intermediate diversity is thought to be due to a combined action of parasite-mediated selection and an excessively strong negative T-cell selection that takes place under high individual MHC diversity [55]. In the case of the loggerhead turtles, up to four MHC alleles seems rather low to trigger increased costs of negative T-cell selection. However the best estimates obtained from mathematical models suggest that such costs can exist with an individual number of expressed MHC molecules in the range of 3 to 25, when combining both MHC class I and class II [55]. This can then apply to the loggerhead turtles. Besides the tropical python, this is the second report of higher individual fitness measure with intermediate MHC class I diversity in reptiles. This correlation may stem from either an advantage of individuals with intermediate MHC diversity being able to better fight off parasites and therefore allocate more energy to growth, or from non-random mortality with regards to MHC. This would result in larger individuals, with intermediate MHC diversity, being older. Both hypotheses are not mutually exclusive but at this stage cannot be disentangled. Another possible explanation is that our data reflect an advantage to heterozygote individuals over homozygotes which would also be predicted by the heterozygote advantage theory (reviewed in [5]). In either case of an optimal diversity or an advantage to heterozygotes, our results suggest an associated cost of homozygosity, a major concern for endangered species such as the loggerhead sea turtle.

## Conclusions

The MHC class I data presented here can serve as an important launching point for studies of conservation genetics, particularly with regard to disease resistance/susceptibility in the loggerhead turtle and other endangered

species. Over the last two decades, the MHC has emerged as a valuable complex of genes for evaluating the relative influence of natural selection versus drift and migration on the levels of genetic variation in populations. This is important when considering that selection and adaptation may have its greatest effect on functionally important genes, including genes affecting resistance to pathogens. Evidence for natural selection of the MHC in the loggerhead turtle adds additional insights into the evolution of this gene complex in a phylogenetically basal lineage and demonstrates the potential importance of MHC in the sustainability of an endangered population.

## Methods

### Sampling

Tissue samples from 40 nesting loggerhead sea turtles were collected between July and September 2010 on the island of Sal, Cape Verde. A 3 mm sample was taken from the superficial part of the non-keratinized skin of the flippers using a single-use disposable scalpel immediately after egg deposition. Samples were individually preserved in ethanol until DNA extraction.

### DNA extraction

All tissues were washed in Milli-Q water for 1 minute and were air dried for 15 minutes. DNA extraction was performed using the DNeasy<sup>®</sup> 96 Blood & Tissue Kit (QIAGEN, Hilden, Germany). All steps followed the manufacturer's protocol with the exception of the elution, which was conducted in two steps of 100 µl, re-using the first elution.

### mtDNA sequencing

In order to compare the MHC based phylogeny with a phylogeny obtained from a neutral marker, we amplified 723 bp of the mtDNA control region for all individuals using LCN15382 and H950 primers [56]. After amplification and cleaning of PCR product using EXoSap, sequences were loaded into an ABI 3730 Genetic Analyzer (Applied Biosystems, Darmstadt, Germany). For more details, see Stiebens et al. [32]. Four different haplotypes were found: CcA1.3, CcA17.1, CcA17.2 and CcA2.1 following the Archie Carr Center for Sea Turtle Research nomenclature (<http://accstr.ufl.edu/resources/mtdna-sequences/>).

### MHC primer design

In order to design primers to characterize the highly polymorphic MHC class I exon 2, GeneBank was searched for MHC sequences of related species to the loggerhead turtle. Reptile and avian MHC class I sequences were aligned using BioEdit version 7.0.5.3 [57] and consisted of sequences from reptiles *Malaclemys terrapin* (Genebank accession numbers: GQ495891.1), *Pelodiscus sinensis*

(AB185243.1 and AB022885.1), *Sphenodon punctatus* (FJ457094.1, FJ457093.1), and a bird species *Gallus gallus* (AY123227.1). Within this alignment, conserved regions in the exon 2 were selected to design several primer pairs. The exon 2 was chosen because it encodes for a part of the peptide-binding groove involved in parasite recognition. After various PCR tests for the best primer combination, Cc-MHC-I-F (5'-GATGTATGGGTGTGATCTCCGGG-3') and Cc-MHC-I-R (5'-TTCACCTCGATGCAGGTCDNCTCCAGGT-3') showed consistent amplification of multiple MHC class I sequences across several cloning procedures. Although, the Cc-MHC-I-R primer shows polymorphism from the 16<sup>th</sup> to 18<sup>th</sup> base pair, no better primers could be designed.

#### MHC amplification, cloning, and sequencing

To reduce the risk of PCR artifacts, two independent 20  $\mu$ l PCR reactions were prepared. Each "replicate" consisted of 2  $\mu$ l 10 $\times$  Dreamtaq<sup>®</sup> Buffer, 1  $\mu$ l dNTP's (10 mM), 2  $\mu$ l of each primer (5pmol/ $\mu$ l), 0.2  $\mu$ l Taq Polymerase (Dreamtaq) and 2  $\mu$ l template DNA [ $\sim$ 20  $\mu$ g/ $\mu$ l]. Thermal profile started with an initial denaturing step at 95°C for 3 minutes, followed by 30 cycles of 30 seconds at 94°C, 30 seconds at 66°C and 1 minute at 72°C. The final elongation was set for 5 min at 72°C. The volumes of both reactions were then pooled, of which 30  $\mu$ l was loaded in an agarose gel (1.5%, 5 h at 45 V). This procedure was recommended by [43] and [58] in order to reduce PCR artifacts. Bands of the expected size ( $\sim$ 220 bp) were excised.

Gel purification followed manufacturer's protocol for the NucleoSpin Extract II Kit (Macherey-Nagel, Düren, Germany). PCR amplicons were cloned with the Qiagen<sup>®</sup> PCR cloning Kit (Qiagen, Hilden, Germany). The manufacturer's ligation protocol was followed, except that the ligation-reaction-mixture consisted of 1  $\mu$ l pDrive Cloning Vector, of 5  $\mu$ l Ligation Master Mix and of 4  $\mu$ l PCR products. The transformation protocol was modified as follows: 5  $\mu$ l of the ligation-reaction mixture were mixed with 25  $\mu$ l competent cells. Reactions were then heated for 40 seconds at 42°C. Later, 150  $\mu$ l SOC medium were added and to allow recombinant growth for Kanamycin selection, the reaction mixture was first incubated for 30 minutes at 37°C (slightly shaken) and then plated on a Kan<sup>®</sup> IptgX-Gal plate. Plasmids were extracted with the Invisorb<sup>®</sup> Spin Plasmid Mini Two Extraction Kit (Invitex, Berlin, Germany) as described in Kit's provided protocol, with a final elution step of 50  $\mu$ l. Cycle sequencing took place in 10  $\mu$ l PCR reactions consisting of 1  $\mu$ l Big Dye<sup>®</sup> Buffer, 1  $\mu$ l Big Dye<sup>®</sup> Terminator, 1  $\mu$ l of the universal M13 Forward primer, 3  $\mu$ l of HPLC water and 4  $\mu$ l of extracted plasmid template. The thermal cycling protocol had a first step for 1 minute at 96°C, then 26 cycles at 96°C for 10 seconds and 50°C for 5 seconds. The elongation final step was set at 60°C for 4 minutes. DNA was precipitated and re-

diluted in HiDi before being loaded on an ABI 3130 Genetic Analyzer (Applied Biosystems, Darmstadt, Germany). After comparisons of the different sequences obtained with the different primer pairs, the best combination (i.e. the one providing most sequences) was used for high throughput sequencing on a next generation sequencing platform.

#### Barcoded 454 sequencing of MHC genes

The 454 next generation sequencing platform using a barcoded deep amplicon approach [29,30] was chosen because of the long sequence reads and large coverage to help determine high intra and inter individual variability. To this end, DNA concentrations were standardized to 10 ng/ $\mu$ l in order to maximize the likelihood of equal coverage of all samples. As previously described, two independent PCR reactions were performed. For each replicate, the protocol was split into two steps. In the first step, PCR conditions were kept as described above, but the number of PCR cycles was reduced to 25. The first PCR products were used as a template for another 10 PCR cycles. The reconditioning procedure coupled with independent PCR reactions reduces the final proportion of artifacts [42], a major problem with new sequencing technologies. The reconditioning step used 454 sequencing adaptors (Forward side TitaA CCATCTCATCCCTGC GTGTCTCCGACTCAG; Reverse side TitaB CCTATCCC CTGTGTGCCTTGGCAGTCTCAG, GATC, Constance, Germany), followed by a 10 nucleotide individual tag (MID, Roche) and the newly developed MHC class I primer pair. The MID tags were designed such that the random accumulation of up to two polymerase errors in the MID would still lead to the correct individual identification. For a given individual, replicated PCRs had the same forward MID tags but different reverse MID tags which allowed us to track the product of each PCR reaction all along the amplification and sequencing.

After amplification, amplicons were cleaned using the Qiagen PCR Purification Kit (Qiagen, Hilden, Germany). The cleaned products were run on gels, to verify the presence of the expected bands. From all cleaned samples, DNA concentration was re-measured and all samples were pooled so that each PCR reaction contributed to an equal amount of 100 ng/sample. To remove potential unspecific amplicons, the final pool was loaded on a 1.5% agarose gel (14 h at 30 V). Bands of  $\sim$ 340 bp were cut out and products were extracted as described above.

#### Individual MHC genotyping

MHC alleles were called and assigned to each individual using Perl scripts. Reads were screened for the forward and reverse sequencing primers, allowing one nucleotide mismatch or indel (insertion/deletion) in case of sequencing errors and otherwise discarded. Remaining reads were then

assigned to individuals based on MID tags, again allowing for one nucleotide mismatch or indel. Reads were then trimmed (removing the primer and MID sequence) and aligned using BioEdit, resulting in a set of putative allele variants for each individual. To cull out less reliable sequence variants, alleles were retained only if they met the following criteria per individual: (1) if they appeared in both independent PCR preparations (both MID tags) and (2) if their frequency (in terms of proportion of reads) was above 10% of the most frequently occurring allele within that individual. The remaining variants, although they might stem from different loci, are referred to as “alleles” and make up our final allele dataset.

Errors occurring during the 454 sequencing include substitutions and small indels [29,30], and these were expected to occur randomly across the sequence. From our MID tags, the frequency of errors resulting in base substitutions was low. Therefore, the probability of multiple, identical substitution errors is estimated to be low [30]. Single-base indels occurring in homopolymer tracts were relatively common and were non-randomly distributed along the sequence. However, such variants were removed with our method because of their low frequency of occurrence within an individual and across independent replicate PCR reactions.

#### Data analyses

Under positive selection, a relative excess of non-synonymous over synonymous substitutions is expected [59]. We calculated the relative rates of synonymous ( $d_S$ ) and non-synonymous ( $d_N$ ) substitutions following the method of Nei and Gojobory [60] with the Jukes-Cantor [61] correction for multiple substitutions implemented in MEGA 4 [62]. The rate ratio  $d_N/d_S$  was tested for significant deviation from one using a Z-test.

MEGA 4 was also used to build a neighbor-joining tree with 1000 bootstraps for all MHC alleles found in the sampled turtles. Two additional neighbor-joining trees were simulated: one based on the control region of the mitochondrial genome (mtDNA) of 6 reptile species and one based on the MHC class I of 5 reptile species.

Maximum likelihood site models implemented in the CODEML program from PAML version 4.4 [63] were used to test for evidence of positive selection and to identify branch-specific positively selected codon sites [ $\omega > 1$ , where  $\omega = (d_N/d_S)$ ]. The maximum likelihood procedures evaluate heterogeneous rate ratios ( $\omega$ ) among sites by applying different models of codon evolution. Three likelihood-ratio tests of positive selection were performed comparing the models M1a (nearly neutral) vs M2a (positive selection), M7 ( $\beta$ ) vs M8 ( $\beta + \omega$ ), and M8a ( $\beta + \omega = 1$ ) vs M8 [64]. In these likelihood-ratio tests, two nested models are compared: a model based on the null hypothesis of no positive selection, and a model that allows some sites to evolve

under positive selection. The null model M1a assumes two site classes in the molecule with  $0 < \omega_0 < 1$  and  $\omega_1 = 1$  in proportions  $p_0$  and  $p_1 = 1 - p_0$ . The alternative model M2a incorporates another class of sites with  $\omega_2 > 1$  and the proportion  $p_2$  estimated from the data. The null model M7 assumes a beta distribution for  $\omega$ , not allowing positive selection ( $0 < \omega < 1$ ). The alternative model M8 has additional classes of sites that allow some codons to evolve under positive selection ( $\omega > 1$ , [62]). A third null model M8a differs from model M8 in that its additional class of sites are evolving neutrally ( $\omega = 1$ ). In the models M2a and M8, positively selected sites are inferred from posterior probabilities calculated by the Bayes empirical Bayes method [65]. Because MHC alleles are so variable and often represent ancient lineages (TSP), we thought the evaluation of dN and dS appropriate despite the comparison within a species.

We used the ScoreCons online server [35] to determine variation for amino acid residues in the exon 2 of the loggerhead turtles. The software MultiLocus 1.22 [66] was used to estimate linkage disequilibrium between detected alleles using 10000 randomizations.

The minimum number of recombinant events (RM) was calculated after Hudson and Kaplan [four-gamete method, McVean *et al.* [36] using the software DnaSP.

The program GENECONV version 1.81 was used to detect sequence fragments that were likely to have been subjected to gene conversions. GENECONV detects pairs of sequences that share unusually long stretches of similarity given their overall polymorphism [67]. We used global and pairwise permutation tests (10,000 replicates) to assess significance.

Although fitness is difficult to estimate in loggerhead turtles, studies have shown that larger females have a higher clutch size, linking turtle morphometrics to high fecundity [38,39]. As a fitness proxy we used the curved carapace length corrected (residuals of correlation) for curved carapace width, as equivalent to body condition. Residuals for this correlation were then tested against individual number of MHC alleles (linear and quadratic terms) following [20]. Curved carapace length and curved carapace width were measured for all turtles immediately after egg deposition.

#### Additional file

**Additional file 1: Figure S1.** Neighbor-joining tree of the mtDNA control region. All sequences have been deposited on Archie Carr Centre for Sea Turtle Research (<http://accstr.ufl.edu/resources/mtdna-sequences/>).  
**Document 2** Amino acid alignment of loggerhead turtle MHC class I alleles. Dots indicate identity with the loggerhead Cc\*0 sequence.  
**Document 3** Table summarizing the genotyping of 7 turtles using two different methods: cloning/sequencing vs. 454 sequencing. Allele identities are given together with the number of clones picked and sequenced for each individual. Row in bold shows a discrepancy between cloning and 454 sequencing. § indicates a posteriori screen.

### Competing interest

The authors declare no competing interests.

### Authors' contributions

CE designed the study. VS, SEM, and CE participated in sample collection. VS and CE performed the statistical analyses. FC wrote the bioinformatic scripts. VS, FC and CE drafted the manuscript. All authors read and approved the final manuscript.

### Acknowledgments

We would like to thank SOS Tartarugas Cabo Verde, particularly Jacquie Cozens, J. Kutz, H. Taylor, S. M. Correia, A. Nascimento da Luz from INDP provided tremendous assistance in the field as well. We also thank A. Hasselmeier for assistance in the lab. CE is supported by Leibniz Institute Competitive funds, the German Science Foundation grants (DFG, El 841/4-1 and El 841/6-1), and a National Geographic Grant (GEFNE69-13).

### Author details

<sup>1</sup>Department of Evolutionary Ecology of Marine Fishes, GEOMAR|Helmholtz Center for Ocean Research, Kiel 24105, Germany. <sup>2</sup>National Institute for Development and Fishery (INDP), Mindelo, Cape Verde. <sup>3</sup>Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Ploen 24306, Germany.

Received: 6 January 2013 Accepted: 17 April 2013

Published: 30 April 2013

### References

- Sommer S: Major histocompatibility complex and mate choice in a monogamous rodent. *Behav Ecol Sociobiol* 2005, **58**(2):181–189.
- Apanius V, Penn D, Slev PR, Ruff LR, Potts WK: The nature of selection on the major histocompatibility complex. *Crit Rev Immunol* 1997, **17**(2):179–224.
- Siddle HV, Kreiss A, Eldridge MDB, Noonan E, Clarke CJ, Pyecroft S, Woods GM, Belov K: Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proc Natl Acad Sci U S A* 2007, **104**(41):16221–16226.
- Ellison A, Allainguillaume J, Girdwood S, Pachebat J, Peat KM, Wright P, Consuegra S: Maintaining functional major histocompatibility complex diversity under inbreeding: the case of a selfing vertebrate. *Proc Biol Sci* 2012, **279**:5004–5013.
- Piertney SB, Oliver MK: The evolutionary ecology of the major histocompatibility complex. *Heredity* 2005, **96**(1):7–21.
- Milinski M: The major histocompatibility complex, sexual selection, and mate choice. *Annu Rev Ecol Syst* 2006, **37**(1):159–186.
- Janeway CA, Travers P, Walport M, Sclomchik MJ: *Immunobiology: the immune system in health and disease*. 6th edition. New York: Garland Science Publishing; 2005.
- Miller HC, Allendorf F, Daugherty CH: Genetic diversity and differentiation at MHC genes in island populations of tuatara (*Sphenodon* spp.). *Mol Ecol* 2010, **19**(18):3894–3908.
- Edwards SV, Chesnut K, Satta Y, Wakeland EK: Ancestral polymorphism of MHC class II genes in mice: Implications for balancing selection and the mammalian molecular clock. *Genetics* 1997, **146**(2):655–668.
- Seddon JM, Ellegren H: MHC class II genes in European wolves: a comparison with dogs. *Immunogenetics* 2002, **54**(7):490–500.
- Graser R, OhUigin C, Vincek V, Meyer A, Klein J: Trans-species polymorphism of class II MHC loci in danio fishes. *Immunogenetics* 1996, **44**(1):36–48.
- Glaberman S, Caccone A: Species-specific evolution of class I MHC genes in iguanas (order: squamata; subfamily: iguaninae). *Immunogenetics* 2008, **60**(7):371.
- Klein J, Satta Y, Ohuigin C, Takahata N: The molecular descent of the major histocompatibility complex. *Annu Rev Immunol* 1993, **11**:269–295.
- Figuerola F, Gunther E, Klein J: MHC polymorphism predating speciation. *Nature* 1988, **335**(6187):265–267.
- Christin PA, Weinreich DM, Besnard G: Causes and evolutionary significance of genetic convergence. *Trends Genet* 2010, **46**:400–405.
- Lenz TL, Eizaguirre C, Kalbe M, Milinski M: Evaluating patterns of convergent evolution and trans-species polymorphism at MHC immunogenes in two sympatric stickleback species. *Evolution* 2013. In press, doi: 10.1111/evo.12124.
- Gutierrez-Espeleta GA, Hedrick PW, Kalinowski ST, Garrigan D, Boyce WM: Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation? *Heredity* 2001, **86**:439–450.
- Buchholz R, Jones Dukes MD, Hecht S, Findley AM: Investigating the turkey's 'snood' as a morphological marker of heritable disease resistance. *J Anim Breed Genet* 2004, **121**(3):176–185.
- Jäger I, Eizaguirre C, Griffiths SW, Kalbe M, Krobbach CK, Reusch TBH, Schaschl H, Milinski M: Individual MHC class I and MHC class IIB diversities are associated with male and female reproductive traits in the three-spined stickleback. *J Evol Biol* 2007, **20**(5):2005–2015.
- Wegner KM, Kalbe M, Kurtz J, Reusch TBH, Milinski M: Parasite selection for immunogenetic optimality. *Science* 2003, **301**(5638):1343.
- Oliver MK, Telfer S, Piertney SB: Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). *Proc Biol Sci* 2009, **276**(1659):1119–1128.
- Kalbe M, Eizaguirre C, Dankert I, Reusch TBH, Sommerfeld RD, Wegner KM, Milinski M: Lifetime reproductive success is maximized with optimal MHC diversity. *Proc R Soc Lond B Biol Sci* 2009, **276**:925–934.
- Bonneaud C, Perez-Tris J, Federici P, Chastel O, Sorci G: Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution* 2006, **60**(2):383–389.
- Madsen T, Ujvari B: MHC class I variation associates with parasite resistance and longevity in tropical pythons. *J Evol Biol* 2006, **19**(6):1973–1978.
- Penn DJ, Damjanovich K, Potts WK: MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc Natl Acad Sci U S A* 2002, **99**(17):11260–11264.
- Miller HC, Belov K, Daugherty CH: MHC class I genes in the tuatara (*Sphenodon* spp.): evolution of the MHC in an ancient reptilian order. *Mol Biol Evol* 2006, **23**(5):949–956.
- Miller HC, Miller KA, Daugherty CH: Reduced MHC variation in a threatened tuatara species. *Anim Conserv* 2008, **11**(3):206–214.
- Binladen J, Gilbert M, Bollback J, Panitz F, Bendixen C: The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2007, **2**:e197.
- Babik W, Taberlet P, Ejsmond MJ, Radwan J: New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Mol Ecol Resour* 2009, **9**(3):713–719.
- Galan M, Guivier E, Caraux G, Charbonnel N, Cosson JF: A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics* 2010, **11**:15.
- Monzon-Arguello C, Rico C, Naro-Maciel E, Varo-Cruz N, Lopez P, Marco A, Lopez-Jurado LF: Population structure and conservation implications for the loggerhead sea turtle of the Cape Verde Islands. *Conserv Genet* 2010, **11**(5):1871–1884.
- Stiebens VA, Merino SE, Roder C, Chain FJJ, Lee PLM, Eizaguirre C: Living on the edge: how philopatry maintains adaptive potential. *Proc R Soc B* 2013. in press.
- Sarmiento-Ramirez JM, Abella E, Martin MP, Telleria MT, Lopez-Jurado LF, Marco A, Dieguez-Urbeondo J: *Fusarium solani* is responsible for mass mortalities in nests of loggerhead sea turtle, *Caretta caretta*, in Boavista, Cape Verde. *FEMS Microbiol Lett* 2010, **312**(2):192–200.
- Babik W: Methods for MHC genotyping in non-model vertebrates. *Mol Ecol Resour* 2010, **10**(2):237–251.
- Valdar WSJ: Scoring residue conservation. *Proteins* 2002, **48**(2):227–241.
- Hudson RR, Kaplan NL: Statistical properties of the number of recombination events in the history of a sample of DNA-sequences. *Genetics* 1985, **111**(1):147–164.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R: DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 2003, **19**(18):2496–2497.
- Hays GC, Speakman JR: Clutch size for Mediterranean loggerhead turtles (*Caretta caretta*). *J Zool* 1992, **226**(2):321–327.
- Broderick AC, Glen F, Godley BJ, Hays GC: Variation in reproductive output of marine turtles. *J Exp Mar Biol Ecol* 2003, **288**(1):95–109.
- Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Foltis KM, Soltis DE: Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol* 2006, **6**:17.
- Brockman W, Alvarez P, Young S, Garber M, Giannoukos G, Lee WL, Russ C, Lander ES, Nusbaum C, Jaffe DB: Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res* 2008, **18**(5):763–770.

42. Lenz TL, Becker S: Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci – Implications for evolutionary analysis. *Gene* 2008, **427**:17–123.
43. Lenz TL, Eizaguirre C, Becker S, Reusch TBH: RSCA genotyping of MHC for high-throughput evolutionary studies in the model organism three-spined stickleback *Gasterosteus aculeatus*. *BMC Evol Biol* 2009, **9**:57.
44. Lenz TL: Computational prediction of MHC II-antigen binding supports divergence allele advantage and explains trans-species polymorphism. *Evolution* 2011, **65**(8):2380–2390.
45. Bowen BW, Kamezaki N, Limpus CJ, Hughes GR, Meylan AB, Avise JC: Global phylogeography of the loggerhead turtle (*Caretta caretta*) as indicated by mitochondrial-DNA haplotypes. *Evolution* 1994, **48**(6):1820–1828.
46. Eizaguirre C, Lenz TL, Kalbe M, Milinski M: Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nat Commun* 2012, **3**:621.
47. Castro-Prieto A, Wachter B, Sommer S: Cheetah paradigm revisited: MHC diversity in the world's largest free-ranging population. *Mol Biol Evol* 2011, **28**:1455–1468.
48. Babik W, Kawalko A, Wojcik JM, Radwan J: Low major histocompatibility complex class I (MHC I) variation in the European Bison (*Bison bonasus*). *J Hered* 2012, **103**:349–359.
49. Eizaguirre C, Lenz TL: Major Histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. *J Fish Biol* 2010, **77**:2023–2047.
50. Nei M, Gu X, Sitnikova T: Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci U S A* 1997, **94**(15):7799–7806.
51. Hughes AL, Yeager M: Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet* 1998, **32**:415–435.
52. Parham P, Ohta T: Population biology of antigen presentation by MHC class I molecules. *Science* 1996, **272**(5258):67–74.
53. Ohta T: Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. *Immunol Rev* 1999, **167**:319–325.
54. Lenz TL, Wells K, Pfeiffer M, Sommer S: Diverse MHC IIB allele repertoire increases parasite resistance and body condition in the long-tailed giant rat (*leopoldamys sabanus*). *BMC Evol Biol* 2009, **9**:13.
55. Woelfing B, Traulsen A, Milinski M, Boehm T: Does intra-individual MHC diversity keep a golden mean? *Phil Trans Biol Sci* 2009, **364**(1513):117–128.
56. Abreu-Grobois FA, Horrocks J, Formia A, Dutton P, Lereux R, Velez-Zuazo X, Soares L, Meylan P, Brown D: New mtDNA control region primers which work for a variety of marine turtle species may increase the resolution capacity of mixed stock analyses, Book of Abstracts. 26th Annual Symposium on Sea Turtle Biology and Conservation. Athens, Greece: International Sea Turtle Society; 2006:179.
57. Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp* 1999, **41**:95–98.
58. Kanagawa T: Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng* 2003, **96**:317–323.
59. Hughes AL, Nei M: Pattern of nucleotide substitution at major histocompatibility complex class-I loci reveals overdominant selection. *Nature* 1988, **335**(6186):167–170.
60. Nei M, Gojobory T: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986, **3**:418–426.
61. Jukes TH, Cantor CR: Evolution of protein molecules. In *Mammalian protein metabolism*, Volume 3. Edited by Munroe HN. New York: Academic Press; 1969.
62. Tamura K, Dudley J, Nei M, Kumar S: MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007, **24**(8):1596–1599.
63. Yang Z: PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 1997, **13**:555–556.
64. Yang Z, Nielsen R, Goldman N, Krabbe Pedersen A-M: Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000, **155**:431–449. 101.
65. Yang Z, Wong WSW, Nielsen R: Empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 2005, **22**(4):1107–1118. 1.
66. Agapow PM, Burt A: Indices of multilocus linkage disequilibrium. *Mol Ecol Notes* 2001, **1**(1–2):101–102.
67. McVean G, Awadalla P, Fearnhead P: A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 2002, **160**(3):1231–1241.

doi:10.1186/1471-2148-13-95

**Cite this article as:** Stiebens et al.: Evolution of MHC class I genes in the endangered loggerhead sea turtle (*Caretta caretta*) revealed by 454 amplicon sequencing. *BMC Evolutionary Biology* 2013 **13**:95.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

