



CORRESPONDENCE

**REVISED** Major histocompatibility complex (MHC) fragment numbers alone – in Atlantic cod and in general - do not represent functional variability [version 2; referees: 3 approved]

Johannes M. Dijkstra <sup>1</sup>, Unni Grimholt <sup>2</sup>

<sup>1</sup>Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi, 470-1192, Japan

<sup>2</sup>Fish Research Group, Norwegian Veterinary Institute, Oslo, Oslo N-0106, Norway

**v2** First published: 28 Jun 2018, 7:963 (doi: [10.12688/f1000research.15386.1](https://doi.org/10.12688/f1000research.15386.1))  
 Latest published: 06 Sep 2018, 7:963 (doi: [10.12688/f1000research.15386.2](https://doi.org/10.12688/f1000research.15386.2))

**Abstract**

This correspondence concerns a publication by Malmstrøm *et al.* in Nature Genetics in October 2016. Malmstrøm *et al.* made an important contribution to fish phylogeny research by using low-coverage genome sequencing for comparison of 66 teleost (modern bony) fish species, with 64 of those 66 belonging to the species-rich clade Neoteleostei, and with 27 of those 64 belonging to the order Gadiformes. For these 66 species, Malmstrøm *et al.* estimated numbers of genes belonging to the major histocompatibility complex (MHC) class I lineages U and Z and concluded that in teleost fish these combined numbers are positively associated with, and a driving factor of, the rates of establishment of new fish species (speciation rates). They also claimed that functional genes for the MHC class II system molecules MHC IIA, MHC IIB, CD4 and CD74 were lost in early Gadiformes. Our main criticisms are (1) that the authors did not provide sufficient evidence for presence or absence of intact functional MHC class I or MHC class II system genes, (2) that they did not discuss that an MHC subpopulation gene number alone is a very incomplete measure of MHC variance, and (3) that the MHC system is more likely to reduce speciation rates than to enhance them. Furthermore, their use of the Ornstein-Uhlenbeck model is a typical example of overly naïve use of that model system. In short, we conclude that their new model of MHC class I evolution, reflected in their title “Evolution of the immune system influences speciation rates in teleost fish”, is unsubstantiated, and that their “pinpointing” of the functional loss of the MHC class II system and all the important MHC class II system genes to the onset of Gadiformes is preliminary, because they did not sufficiently investigate the species at the clade border.

**Keywords**

fish, MHC, Atlantic cod, evolution, speciation rate

**Open Peer Review**

Referee Status:

	Invited Referees		
	1	2	3
<b>REVISED</b>			
<b>version 2</b>	report	report	report
published			
06 Sep 2018			
<b>version 1</b>			
published	report	report	report
28 Jun 2018			

- Anthony B. Wilson** , Brooklyn College CUNY, USA  
City University of New York, USA
- Brian Dixon** , University of Waterloo, Canada
- Jerzy K. Kulski** , Tokai University School of Medicine, Japan  
The University of Western Australia, Australia

**Discuss this article**

Comments (0)

**Corresponding author:** Johannes M. Dijkstra ([Dijkstra@fujita-hu.ac.jp](mailto:Dijkstra@fujita-hu.ac.jp))

**Author roles:** **Dijkstra JM:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation; **Grimholt U:** Data Curation, Formal Analysis, Investigation, Methodology, Validation

**Competing interests:** No competing interests were disclosed.

**Grant information:** The author(s) declared that no grants were involved in supporting this work.

**Copyright:** © 2018 Dijkstra JM and Grimholt U. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Dijkstra JM and Grimholt U. **Major histocompatibility complex (MHC) fragment numbers alone – in Atlantic cod and in general - do not represent functional variability [version 2; referees: 3 approved]** *F1000Research* 2018, 7:963 (doi: [10.12688/f1000research.15386.2](https://doi.org/10.12688/f1000research.15386.2))

**First published:** 28 Jun 2018, 7:963 (doi: [10.12688/f1000research.15386.1](https://doi.org/10.12688/f1000research.15386.1))

**REVISED Amendments from Version 1**

A description of the changes from version 1 to 2:

Inspired by the comments by Dr. Wilson and Dr. Kulski, we now added three new paragraphs “Detailed discussion of the use of the Ornstein-Uhlenbeck model by Malmstrøm *et al.*”, “Additional criticisms in regard to the modelling by Malmstrøm *et al.*”, and “Conclusion”. These were added because Dr. Wilson asked us to explain more clearly how we think that Malmstrøm *et al.*<sup>1</sup> should have handled their data, and Dr. Kulski raised attention to the modeling weaknesses in the Malmstrøm *et al.* article. To explain the modeling weaknesses we also added [Table S2](#). The other supplementary files were also exchanged, because something had gone wrong in their placement during the discussion with the editors during the first submission so that one file was shown twice.

**See referee reports**

**Correspondence**

In the below, we explain our criticisms of the Malmstrøm *et al.*<sup>1</sup> article as they are summarized in our abstract.

*When was the MHC class II system lost in Gadiformes?* The data as presented by Malmstrøm *et al.*<sup>1</sup> suggest a simultaneous loss of major histocompatibility complex (MHC) IIA, MHC IIB, CD4 and CD74 functions at the evolutionary onset of Gadiformes (see their Figure 2). However, within their datasets for gadiform fishes, sequence reads that represent these genes can readily be detected ([Table S1](#) and [Supplementary File 1](#)). These sequence read numbers are much lower than found for the non-gadiform fish, and they may be contaminations, but that should be appropriately tested. Meanwhile, for several non-gadiform fishes, including for *S. chordatus* which among the investigated fishes is the species closest related to Gadiformes, there are no full-length *MHC IIA*, *MHC IIB*, *CD4* or *CD74* gene sequences in the unitig and scaffold datasets presented by Malmstrøm *et al.*<sup>1</sup> ([Supplementary File 2](#) and [Table S2](#)). We agree with the conclusion by Malmstrøm *et al.*<sup>1</sup> that their data suggest that throughout Gadiformes there is no canonical MHC class II system. However, as for the evolutionary timings of the loss of an intact MHC class II system and of the losses of the individual MHC class II system genes, we find their study technically wanting and preliminary. The combination of (i) not finding intact full-length sequences for all important MHC class II system genes in species closely related to Gadiformes, while (ii) finding reads of these genes in gadiform fishes, prohibits what the authors call “pinpointing the loss of MHC II pathway genes to the common ancestor of Gadiformes”. At least for a few species at either side of the Gadiformes clade border, Malmstrøm *et al.*<sup>1</sup> should have substantiated their claims by addition of specific PCR plus sequencing analyses, which should confirm presence of full-length intact MHC class II genes in the non-gadiform fishes, and their absence in the gadiform fishes.

*Discussion of the MHC class I counting strategy by Malmstrøm *et al.**<sup>1</sup> Whereas our criticisms of the MHC class II system analysis by Malmstrøm *et al.*<sup>1</sup> are about technical issues and the preliminary character of their conclusions, we more

fundamentally disagree with their analyses and discussions of MHC class I. The authors assumed<sup>1</sup>, as postulated by other researchers before them, that there can be a “copy number optimum” of MHC genes affected by a tradeoff between a higher number allowing the presentation of more pathogen antigens while also having a depletion effect on the T cell population. Regardless of the extent to which this mostly theoretical concept is true<sup>2</sup>, the MHC counting strategy by Malmstrøm *et al.*<sup>1</sup> should be deemed incomplete and far too simplistic. For their number determination Malmstrøm *et al.*<sup>1</sup> solely relied on estimation of U plus Z lineage genomic  $\alpha 3$  exon fragment numbers, despite that the typical “birth and death” mode of MHC evolution can produce many pseudogenes<sup>3</sup>. The decision of the authors to only count U plus Z lineage gene fragments was based on their unsubstantiated perception that (neo-)teleost U and Z molecules “predominantly” bind peptide ligands<sup>1</sup>. However, not all teleost U and Z molecules are expected to present peptides<sup>4,5</sup>, for example this is not expected for the majority of U lineage molecules in the neoteleost fish medaka<sup>6</sup> and the non-neoteleost fish rainbow trout<sup>7,8</sup>; how this is in the majority of the species investigated by Malmstrøm *et al.*<sup>1</sup> remains to be determined. Furthermore, it should be realized that MHC class II and non-peptide-binding MHC class I molecules (like maybe teleost molecules of the MHC class I lineages L, P and S<sup>4</sup>) also can contribute to T cell depletion<sup>e.g.9</sup>. Peculiarly, while from their referencing it follows that Malmstrøm *et al.*<sup>1</sup> were aware of an MHC class II impact on T cell depletion, the authors did not look at MHC class II when investigating their optimum MHC number model. A more general shortcoming of the article<sup>1</sup> is the lack of awareness that the direct determiner of the levels of “antigen coverage” and T cell depletion is the variation between the relevant MHC molecules<sup>2</sup>, rather than merely the MHC gene copy number. [Table 1](#) (with detailed explanations in [Supplementary File 3](#)) summarizes that different teleost fish species can have very different levels of variation between MHC molecules<sup>4</sup>, and that despite its many U lineage gene copies the extent of MHC variation in Atlantic cod can be considered as relatively limited. Previously, we showed that salmon, zebrafish and eel share variation in U lineage sequences, dating from probably more than 300 million years ago (MYA), whereas all U lineage variation found within the neoteleost fishes stickleback and Atlantic cod probably was established only after these two species separated around 150 MYA<sup>4</sup>. Without experimental evidence, it cannot simply be assumed that “antigen coverage” and/or T cell depletion are highest in fishes with the highest counts of U plus Z  $\alpha 3$  fragments, while neglecting levels of variance among the intact U and Z molecules and possible presences of other categories of MHC molecules. As a last critical comment we point out that, in stark contrast to the evolution of any other known MHC lineage, most deduced Z lineage molecules are characterized by a putative peptide binding groove which was almost perfectly conserved since >400 MYA<sup>4</sup>; this questions the model by Malmstrøm *et al.*<sup>1</sup> that Z lineage evolution was driven by pathogen antigen variation, and is a further argument against the use of combined U+Z numbers for analysis of MHC evolution.

*Overall discussion of the model by Malmstrøm *et al.**<sup>1</sup> saying that U+Z numbers in teleost fish affect speciation rates and that the half-life for reaching the U+Z optimum number is 23 million

**Table 1. Intra-species major histocompatibility complex (MHC) variation differs among teleost fishes.** This table shows the lowest percentages of amino acid sequence identities between membrane-distal domains ( $\alpha 1 + \alpha 2$  for MHC I,  $\alpha 1$  for MHC IIA,  $\beta 1$  for MHC IIB) of same category MHC molecules found between reported sequences of the same species. In some species no genes for particular categories were found (black boxes), while in other instances only one seemingly intact gene sequence was found (1 sequence) or only pseudogenes were found (pseudogene). A more detailed explanation of this table is provided in [Supplementary File 3](#).

		Species		Neoteleostei				
		Zebrafish	Salmon	Medaka	Fugu	Stickleback	Tilapia	Cod
MHC category		( <i>Danio rerio</i> )	( <i>Salmo salar</i> )	( <i>Oryzias latipes</i> )	( <i>Takifugu rubripes</i> )	( <i>Gasterosteus aculeatus</i> )	( <i>Oreochromis niloticus</i> )	( <i>Gadus morhua</i> )
MHC class I	U classical	40%	47%	52%	75%	76%	52%	58%
	U all	27%	38%	32%	29%	62%	27%	58%
	Z	70%	84%	84%	1 sequence	1 sequence	78%	89%
	L	27%	51%				1 sequence	
	P		pseudogene		85%			1 sequence
	S		99%					
MHC class II	DA IIA	34%	84%	48%	72%	64%	39%	
	DA IIB	34%	76%	56%	76%	57%	46%	
	DB IIA	23%	20%	20%		1 sequence	21%	
	DB IIB	31%	25%	26%		1 sequence	22%	
	DE IIA		99%					
	DE IIB		99%					

years. Malmstrøm *et al.*<sup>1</sup> postulated their multiple-regime Ornstein-Uhlenbeck model with very slow progress towards optimum MHC numbers because it was the best fitting model among the few models that they tested. However, an even better fitting model would be that in each species the respective optimal U and Z gene organizations were achieved. Further criticism is that their calculation methods for optimum U+Z numbers and half-life periods incorporated calculations of U+Z gene multiplication speeds, which suffered from the fact that (like for their other considerations) Malmstrøm *et al.* considered all U and Z genes as identical mathematical units<sup>1</sup>. For such speed calculations U and Z genes should have been studied separately, and it also should have been realized that whereas from some U or Z genes multiple new copies were generated, others were lost in accordance with the “MHC gene birth and death” model<sup>3</sup>. Lastly, even if, regardless of the discussable calculations for speeds and optimum numbers, there is a positive association in neoteleost fish between speciation rates and U+Z  $\alpha 3$  fragment numbers (see their Figure 3), then still their model which considers MHC genes as “speciation genes that promote rapid diversification”<sup>1</sup> would be implausible in regard to cause and effect. Namely, in most species, there is a strong evolutionary pressure to maintain old allelic variation within MHC genes (trans-species polymorphism<sup>3,4,10</sup>), which, if anything, is likely to slow down speciation rates because it increases the required size of the founder population<sup>10</sup>. If old allelic or haplotype variation can't be maintained because of rapid speciation through small founder populations, it can be speculated that a species might benefit from an enhanced capacity for the creation of new MHC

allelic and/or haplotype variation by duplications/deletions and recombination<sup>11</sup> between a high number of linked MHC gene copies. However, in that scenario it wouldn't be the MHC organization which drives the speciation rate, as suggested by Malmstrøm *et al.*<sup>1</sup>, but the other way around.

*Detailed discussion of the use of the Ornstein-Uhlenbeck model by Malmstrøm et al.*<sup>1</sup>. The manner in which Malmstrøm *et al.*<sup>1</sup> used the multiple-regime Ornstein-Uhlenbeck (OU) model concerns model fitting rather than model testing. The few restrictions within this system, because multiple regimes are allowed, and the lack of testing of the conclusions with independent datasets, makes the OU-based conclusions by Malmstrøm *et al.*<sup>1</sup> statistically wanting. Cooper *et al.* (2016)<sup>12</sup> listed limitations of, and recommendations for, meaningful use of the OU modeling system. The Malmstrøm *et al.*<sup>1</sup> study did not comply with those recommendations, namely:

1. Cooper *et al.*<sup>12</sup> found that OU model error rates are unacceptably high when tree size is small (< 200 species tips). Malmstrøm *et al.*<sup>1</sup> only used 66 species tips for their OU calculations of U+Z  $\alpha 3$  fragment optima.
2. Cooper *et al.*<sup>12</sup> concluded that the system can be strongly affected by measurement errors. Comparison between different studies questions the reliability of the estimation of 80 U lineage  $\alpha 3$  fragment per haploid cod genome by Malmstrøm *et al.*<sup>1</sup>. Namely, with several of the same authors, an earlier study concluded approximately 100 U lineage  $\alpha 3$  fragments per haploid

genome based on qPCR analysis (Star *et al.*, 2011<sup>13</sup>), and in a later study<sup>14</sup>, upon improved analysis of Atlantic cod genomic scaffold sequences by Tørresen *et al.* (2018<sup>14</sup>), a similar set of authors could only detect 13 different sequences with  $\alpha1+\alpha2+\alpha3$  fragment combinations, 13 with only  $\alpha1$ , 7 with only  $\alpha2$ , 16 with only  $\alpha3$ , and 4 with  $\alpha2+\alpha3$  fragments. Based on those various findings, Tørresen *et al.*<sup>14</sup> stated that they found 53 copies of U lineage genes (sum of all hits) and that those represented 76% of the number previously estimated by Malmstrøm *et al.*<sup>1</sup>, which according to Tørresen *et al.*<sup>14</sup> was 70. This appears to be an incorrect statement, because, according to their Supplementary Table 3<sup>1</sup>, Malmstrøm *et al.*<sup>1</sup> concluded a copy number of approximately 80.1 (and not 70)  $\alpha3$  fragments per haploid genome, while Tørresen *et al.*<sup>1</sup> only found 20 different sequences with  $\alpha3$  fragments per diploid genome which theoretically could be derived from only 10 loci. Incorporation of measurement errors into the OU modeling by Malmstrøm *et al.*<sup>1</sup> are not only suggested by comparison of various studies on cod MHC class I numbers, but also by their<sup>1</sup> *Hox* gene copy analysis. Namely, they<sup>1</sup> tried to validate the copy number estimation procedure by estimating the “relatively conserved number” of *Hox* gene copies for all investigated species. However, whereas their<sup>1</sup> estimation of 50.3 *Hox* copies average among the investigated fish species may approximate the biological reality, without a reasonable biological explanation the finding of variation in *Hox* copy numbers between 30 and 99 in different species (see their Supplementary Figure 3 and Supplementary Table 3<sup>1</sup>) suggests measurement errors inherent in their analytical system. The fact that Malmstrøm *et al.*<sup>1</sup> included a *Hox* control revealed their awareness that their Illumina read libraries might not equally represent all genomic regions, and their *Hox* analysis results should not have relieved them of those worries.

3. Cooper *et al.*<sup>12</sup> concluded that OU calculations are sensitive to intraspecific variation. MHC class I gene copy numbers can differ widely between individuals (haplotype variation), for example, it was estimated that the U lineage gene copy number in the neoteleost fish tilapia can differ between 11 and 17 per haplotype<sup>15</sup>. Also, concerning Atlantic cod haplotype variation, the presence/absence of U lineage genes was found<sup>16</sup>. The availability of copy number variation within species makes it highly unlikely that the half-life for reaching optimum numbers of U+Z  $\alpha3$  fragments is 23 million years in teleost fish, since there is already an existing copy number variation from which can be selected. A period of 23 million years typically includes multiple speciation events (see Table S2).

4. Cooper *et al.*<sup>12</sup> recommended that OU-based conclusions should be tested by comparison with other analyses. Table S2 shows that the Malmstrøm *et al.*<sup>1</sup> conclusion that species richness is associated with elevated U+Z  $\alpha3$  numbers may agree with their division into clades (“regimes”) used to fit their model, but that such a conclusion does not hold if other branchings in the neoteleost phylogenetic tree are compared (Table S2). For example, Percomorphaceae excluding Ophidiiformes have more species and a higher average number of U+Z  $\alpha3$  numbers than Ophidiiformes, which agrees with the model, but at the same time Carangiformes appear to have many more species

but considerably lower U+Z  $\alpha3$  numbers than Abantiformes (Table S2), which does not agree with the model. Table S2 shows that there is not a general tendency for the branches with more species to have higher U+Z  $\alpha3$  numbers. The calculated half-life of 23 million years for reaching optimum copy number also should be considered as a reason to abandon the model; namely, the long period would imply that hardly any of the investigated species have the optimum MHC number, whereas the previously discussed haplotype variations between individuals of the same species reveal that the system is quite plastic. After all, the type of gene copy number variation investigated here is generally not about gene duplicates that acquired entirely new functions, but more commonly about duplicated genes that acquired only slightly modified functions (for example presenting a slightly different set of peptides), and genes and pseudogenes which may predominantly serve as a reservoir for increased genomic recombination, or which may have no function at all. For example, the turnover of closely related MHC class I genes during primate evolution has produced a human MHC genomic region with only 3 classical class I genes *HLA-A*, *-B*, and *-C*, 3 nonclassical genes *HLA-E*, *-F* and *-G* that also encode molecules with peptide binding ability, and numerous pseudogenes including *HLA-H*, *-J*, *-K*, *-L*, and *-V*, and additional orphan sequence fragments that together account for 6 MHC class I coding genes and 13 pseudogenes<sup>17,18</sup>. In evolution, by recombination events, increase or decrease in numbers of tandemly organized similar genes can go very rapidly (e.g. 19). A half-life of 23 million years also would have trouble explaining how the Atlantic cod has 80 U lineage  $\alpha3$  fragments, whereas the fish *Theragra chalcogramma*, which separated from Atlantic cod only ~3 million years ago, only has 31; this observed difference implies very rapid large changes in copy numbers.

*Additional criticisms in regard to the modelling by Malmstrøm et al.*<sup>1</sup>.

There are several more important criticisms that can be made about the modelling by Malmstrøm *et al.*<sup>1</sup>.

(A) Malmstrøm *et al.*<sup>1</sup> claimed that their study is about teleosts, but basically (with 64 of 66 investigated species) they only investigated within one clade of teleosts, namely the neoteleosts, and neglected information obtained for fish from other large clades of teleosts.

(B) By focusing solely on the  $\alpha3$  exons, Malmstrøm *et al.*<sup>1</sup> investigated the part of MHC class I genes which is the least informative for function.

(C) In teleosts, although probably the least pronounced in neoteleosts, selection on U lineage variation happens at two levels<sup>4,20</sup>. In fish like cyprinids and salmonids, extreme evolutionary pressure appears to maintain ancient allelic sequence variation from hundreds of millions of years ago, which extends far beyond the peptide binding groove. Simultaneously, there is also the “normal” balancing/diversifying selection on peptide binding groove variation that has been well described for mammals. It is hard to take any model on teleost MHC class I



evolution seriously that does not recognize these two different levels of selection.

(D) Questions also can be raised concerning the reliability of the BAMM method used by Malmstrøm *et al.*<sup>1</sup> for analysis of speciation rates<sup>21</sup>. By using this method, Malmstrøm *et al.*<sup>1</sup> found higher speciation rates in species-rich clades than in species-poor clades, which seems to be a logical conclusion. However, when Rabosky, the author of the BAMM method<sup>22</sup>, applied the method to fish, a “negative relationship between species richness and mean speciation rate” was found, which is counterintuitive and curiously not explained in the respective article<sup>23</sup>.

(E) Malmstrøm *et al.*<sup>1</sup> used a bivalent BiSSE model for determining statistical significance of a positive association between elevated speciation rates and high U+Z  $\alpha 3$  numbers. It seems that they speculated that the concluded significance of the finding derived from accelerated speciation in fish lineages where U+Z  $\alpha 3$  numbers were higher than a threshold of about 20–25 copies, although that would not explain the overall distribution seen in their Figure 3b. As a potential biological explanation for their proposed model, they<sup>1</sup> suggested that the effect of T cell depletion on hybrid fitness becomes more pronounced in the 20–25 copy range and that this might affect mate choice in species with copy numbers above this threshold, promoting inbreeding and reinforcement. However, their speculation can not explain how some species can have many more than 20–25 copies, for example the >100 copies observed in *Muraenolepis marmoratus*, and why such high copy numbers are not associated with even higher speciation rates (see their Figure 3b).

(F) Malmstrøm *et al.*<sup>1</sup> concluded “an optimum of 6.8 MHC I copies for basal branches of the phylogeny, which is in concordance with the hypothesized MHC I repertoire of early gnathostomes”. However, previously estimated U+Z numbers in zebrafish, Mexican tetra (cavefish) and Atlantic salmon were 14, 31 and 14, respectively<sup>4</sup>, and U+Z  $\alpha 3$  copies that Malmstrøm *et al.*<sup>1</sup> estimated for *Borostomias antarcticus* were 20, suggesting that the U+Z number was considerably higher than 6.8 at the start of the neoteleost lineage. This should have been acknowledged by the authors.

In regard to the modelling methods used by Malmstrøm *et al.*<sup>1</sup>, we also recommend readers to read the extensive critical comments by reviewer Dr. Jerzy Kulski, which he placed under the first version of our manuscript. In those comments he discusses the models used by Malmstrøm *et al.*<sup>1</sup>.

## Conclusion

Malmstrøm *et al.*<sup>1</sup> used low-coverage genome sequencing for comparison of 66 mostly neoteleost fish, and so helped with elucidating their phylogeny. They found that intact MHC class II system genes may be completely absent in Gadiformes, and believed that related non-gadiform fishes have intact MHC class II system genes. However, their genomic databases were incomplete and in the case of many Gadiformes spiked with reads from MHC class II system genes that may or may not be contaminations, so that final conclusions require some additional analysis of at least a few species at the gadiform/non-gadiform clades border. We suggest that they need to perform a number of PCR and sequencing experiments to clarify this matter. When comparing class I and class II situations in their investigated neoteleosts, Malmstrøm *et al.*<sup>1</sup> also found that their earlier theory, which was that the absence of an MHC class II system might explain the high number of MHC class I genes in Atlantic cod<sup>13</sup>, could not be corroborated. Instead, solely based on estimations of U+Z  $\alpha 3$  fragment numbers, they<sup>1</sup> proposed a new theory on MHC class I evolution which they referred to in their manuscript title. We hope to have shown sufficiently that their conclusions on MHC class I evolution were unsubstantiated, that estimation of U+Z  $\alpha 3$  fragment numbers is not a proper way to analyze MHC functions or MHC evolution, and that, apart from not investigating logical units that are better suited for their methods of modeling, also the number estimations and modelling systems used by Malmstrøm *et al.*<sup>1</sup> were flawed and/or non-trustworthy. Before any meaningful discussion can be started about the evolution of MHC class I genes in neoteleosts, a much higher level of information about sequences and genomic positions is necessary.

## Data availability

The datasets analyzed in this study originate from Malmstrøm *et al.* (2016)<sup>1</sup> and are publicly available in the NCBI bioproject <https://www.ncbi.nlm.nih.gov/bioproject/PRJEB12469> and in the DRYAD repository <https://datadryad.org/resource/doi:10.5061/dryad.326r8>. Details are explained in [Supplementary File 1](#), [Supplementary File 2](#) and [Supplementary File 3](#).

## Grant information

The author(s) declared that no grants were involved in supporting this work.

## Acknowledgement

We appreciate the work of the reviewers Dr. Brian Dixon, Dr. Jerzy Kulski and Dr. Anthony Wilson, who helped us to improve our manuscript.

## Supplementary material

Supplementary Table S1: Examples of sequence reads of major histocompatibility complex (MHC) class II system genes found in single read archive (SRA) datasets published by Malmstrøm *et al.* for Gadiformes and closely related fishes.

[Click here to access the data.](#)

Supplementary Table 2: Species numbers in teleost clades, and estimated U and Z a3 copy numbers in teleost species, investigated by Malmstrøm and co-workers.

[Click here to access the data.](#)

Supplementary File 1: List of sequence reads in SRA datasets of Gadiformes published by Malmstrøm *et al.* that match with major histocompatibility complex (MHC) class II system genes.

[Click here to access the data.](#)

Supplementary File 2: Investigation of unitigs with (partial) major histocompatibility complex (MHC) class II system genes which are listed by Malmstrøm *et al.* in their Table S7 for the non-gadiform fishes *S. chordatus*, *C. roseus*, *Z. faber*, *T. subterraneus*, *P. transmontana*, and *P. japonica*.

[Click here to access the data.](#)

Supplementary File 3: Detailed explanation of Table 1.

[Click here to access the data.](#)

## References

- Malmstrøm M, Matschiner M, Tørresen OK, *et al.*: **Evolution of the immune system influences speciation rates in teleost fishes.** *Nat Genet.* 2016; **48**(10): 1204–10.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Borghans J, Keşmir C, de Boer RJ: **MHC diversity in Individuals and Populations.** In: Flower D, Timmis J, editors. *In Silico Immunology.* Springer, New York NY; 2007; 177–195.  
[Publisher Full Text](#)
- Nei M, Rooney AP: **Concerted and birth-and-death evolution of multigene families.** *Annu Rev Genet.* 2005; **39**: 121–52.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grimholt U, Tsukamoto K, Azuma T, *et al.*: **A comprehensive analysis of teleost MHC class I sequences.** *BMC Evol Biol.* 2015; **15**: 32.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Malmstrøm M, Jentoft S, Gregers TF, *et al.*: **Unraveling the evolution of the Atlantic cod's (*Gadus morhua* L.) alternative immune strategy.** *PLoS One.* 2013; **8**(9): e74004.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nonaka MI, Aizawa K, Mitani H, *et al.*: **Retained orthologous relationships of the MHC Class I genes during euteleost evolution.** *Mol Biol Evol.* 2011; **28**(11): 3099–112.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dijkstra JM, Kiryu I, Yoshiura Y, *et al.*: **Polymorphism of two very similar MHC class Ib loci in rainbow trout (*Oncorhynchus mykiss*).** *Immunogenetics.* 2006; **58**(2–3): 152–67.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Miller KM, Li S, Ming TJ, *et al.*: **The salmonid MHC class I: more ancient loci uncovered.** *Immunogenetics.* 2006; **58**(7): 571–89.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Schümann J, Pittoni P, Tonti E, *et al.*: **Targeted expression of human CD1d in transgenic mice reveals independent roles for thymocytes and thymic APCs in positive and negative selection of Valpha14i NKT cells.** *J Immunol.* 2005; **175**(11): 7303–10.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Klein J, Sato A, Nikolaidis N: **MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics.** *Annu Rev Genet.* 2007; **41**: 281–304.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Doxiadis GG, de Groot N, Otting N, *et al.*: **Haplotype diversity generated by ancient recombination-like events in the MHC of Indian rhesus macaques.** *Immunogenetics.* 2013; **65**(8): 569–84.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cooper N, Thomas GH, Venditti C, *et al.*: **A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies.** *Biol J Linn Soc Lond.* 2016; **118**(1): 64–77.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Star B, Nederbragt AJ, Jentoft S, *et al.*: **The genome sequence of Atlantic cod reveals a unique immune system.** *Nature.* 2011; **477**(7363): 207–10.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tørresen OK, Briec MSO, Solbakken MH, *et al.*: **Genomic architecture of haddock (*Melanogrammus aeglefinus*) shows expansions of innate immune genes and short tandem repeats.** *BMC Genomics.* 2018; **19**(1): 240.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Murray BW, Nilsson P, Zaleska-Rutczynska Z, *et al.*: **Linkage Relationships and Haplotype Variation of the Major Histocompatibility Complex Class I A Genes in the Cichlid Fish *Oreochromis niloticus*.** *Mar Biotechnol (NY).* 2000; **2**(5): 437–448.  
[PubMed Abstract](#)
- Miller KM, Kaukinen KH, Schulze AD: **Expansion and contraction of major histocompatibility complex genes: a teleostean example.** *Immunogenetics.* 2002; **53**(10–11): 941–63.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Shiina T, Blancher A, Inoko H, *et al.*: **Comparative genomics of the human, macaque and mouse major histocompatibility complex.** *Immunology.* 2017; **150**(2): 127–38.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Olivieri DN, Gambon-Deza F: **Primate MHC class I from Genomes.** *bioRxiv.* 2018; 266064.  
[Publisher Full Text](#)
- Norman PJ, Abi-Rached L, Gendzekhadze K, *et al.*: **Meiotic recombination generates rich diversity in NK cell receptor genes, alleles, and haplotypes.** *Genome Res.* 2009; **19**(5): 757–69.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shum BP, Guethlein L, Flodin LR, *et al.*: **Modes of salmonid MHC class I and II evolution differ from the primate paradigm.** *J Immunol.* 2001; **166**(5): 3297–308.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Moore BR, Höhna S, May MR, *et al.*: **Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures.** *Proc Natl Acad Sci U S A.* 2016; **113**(34): 9569–74.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rabosky DL: **Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees.** *PLoS One.* 2014; **9**(2): e89543.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rabosky DL, Chang J, Title PO, *et al.*: **An inverse latitudinal gradient in speciation rate for marine fishes.** *Nature.* 2018; **559**(7714): 392–395.  
[PubMed Abstract](#) | [Publisher Full Text](#)

# Open Peer Review

Current Referee Status:   

---

## Version 2

Referee Report 17 September 2018

doi:[10.5256/f1000research.17714.r38018](https://doi.org/10.5256/f1000research.17714.r38018)

 **Anthony B. Wilson**  <sup>1,2</sup>

<sup>1</sup> Department of Biology, Brooklyn College CUNY, New York, NY, USA

<sup>2</sup> The Graduate Center, City University of New York, New York, NY, USA

No further amendments are required. As highlighted by Dr. Kulski in his review, I'm hopeful that this debate will prompt the authors and others to tackle this fascinating evolutionary question!

**Competing Interests:** No competing interests were disclosed.

**Referee Expertise:** Adaptive Immunity, Evolutionary Biology, Speciation

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Referee Report 13 September 2018

doi:[10.5256/f1000research.17714.r38019](https://doi.org/10.5256/f1000research.17714.r38019)

 **Brian Dixon** 

Department of Biology, University of Waterloo, Waterloo, ON, Canada

The additional analysis and explanation of the basis for their comments strengthens this article.

**Competing Interests:** No competing interests were disclosed.

**Referee Expertise:** Fish Immunology, MHC genes, antigen presentation

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Referee Report 07 September 2018

doi:[10.5256/f1000research.17714.r38020](https://doi.org/10.5256/f1000research.17714.r38020)





**Jerzy K. Kulski**  <sup>1,2</sup>

<sup>1</sup> Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, Isehara, Japan

<sup>2</sup> School of Psychiatry and Clinical Neurosciences, The University of Western Australia , 35 Stirling Highway Crawley, WA 6009, Australia

No further comments were required.

**Competing Interests:** No competing interests were disclosed.

**Referee Expertise:** MHC genes, retrotransposons, evolutionary genomics and biology, population diversity.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

**Version 1**

Referee Report 06 August 2018

doi:[10.5256/f1000research.16766.r36540](https://doi.org/10.5256/f1000research.16766.r36540)



**Jerzy K. Kulski**  <sup>1,2</sup>

<sup>1</sup> Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, Isehara, Japan

<sup>2</sup> School of Psychiatry and Clinical Neurosciences, The University of Western Australia , 35 Stirling Highway Crawley, WA 6009, Australia

The correspondence by Dijkstra & Grimholt<sup>1</sup> provides critical concerns about a publication by Malmstrøm et al in Nature Genetics in October 2016<sup>2</sup>, concluding that their new model of MHC class I evolution, reflected in their title “Evolution of the immune system influences speciation rates in teleost fish”, is unsubstantiated. I concur with their three main criticisms “(1) that the authors did not provide sufficient evidence for presence or absence of intact functional MHC class I or MHC class II system genes, (2) that they did not discuss that an MHC subpopulation gene number alone is a very incomplete measure of MHC variance, and (3) that the MHC system is more likely to reduce speciation rates than to enhance them.”

All three critical points are well founded and stand-alone without too much need for further support. However, I have added the following 14-point commentary for Dijkstra & Grimholt<sup>1</sup>, Malmstrøm et al (2016)<sup>2</sup> and others to consider and elaborate on if they would like to because they are important concerns in the field of MHC genomics, biological function and evolution.

1. According to Dijkstra & Grimholt<sup>1</sup> “the MHC counting strategy by Malmstrøm et al.<sup>2</sup> should be deemed incomplete and far too simplistic. For their number determination Malmstrøm et al.<sup>2</sup> solely relied on estimation of U plus Z lineage genomic  $\alpha$ 3 exon fragment numbers, despite that the typical “birth and death” mode of MHC evolution can produce many pseudogenes.” I agree that this is a major problem

with the Malmstrøm et al (2016)<sup>2</sup> paper, one that also omits the other categories of the MHC I such as the L, S and P lineages that might contribute to a much large number of MHC I-like genes. In this regard, it seems that Malmstrøm et al.<sup>2</sup> have taken only the genomic exon fragment numbers of the MHC I U and Z lineages to represent the entire immune system of their title.

2. Dijkstra & Grimholt<sup>1</sup> are also right to point out that low coverage sequencing by next generation sequencing (NGS) can result in the artifactual loss of genes that in turn can lead to misleading or incorrect conclusions when counting for gene copy numbers or looking for gene losses. Malmstrøm et al (2016)<sup>2</sup> sequencing coverage was 9 to 39x and they recovered only about 75% of the conserved eukaryotic genes. Therefore, in this situation, the *de novo* low coverage sequencing data should not have been used as evidence for the absence of genes from the genome without providing a properly organized high coverage map of genomic assemblies to show where the sequences are missing in the genomes. The reviewers and editors of the Malmstrøm et al (2016)<sup>2</sup> paper should have been aware of this basic problem of using low coverage NGS, particularly with respect to looking for a few needles in a haystack. For a better understanding of the advantages and disadvantages of MHC genotyping and haplotyping by NGS see the review by Shiina et al (2016)<sup>3</sup>.

3. Malmstrøm et al (2016)<sup>2</sup> reported that there was no overall correlation between the combined MHC copy numbers and the HOX gene copy numbers that were used as a control. As already pointed out by Dijkstra & Grimholt<sup>1</sup>, the U and Z genes should have been analysed separately. Nevertheless, it would have been interesting to see how these duplicated HOX development genes, which also have been implicated in driving speciation, compared with the properly classified duplicated MHC class I adaptive immune genes (separate Z, X, L, S and P lineages) at the classical and non-classical level in Fig 3 during speciation rate simulations<sup>2</sup>. In addition, it appears from Fig S3 that Malmstrøm et al (2016)<sup>2</sup> might have missed an inverse relationship between MHC & HOX for the MHC copy numbers up to 25 and direct correlation between MHC & HOX for the MHC copy number from 25 to 50.

4. It seems absurd to count up only the short sequences of a3 fragments from a low coverage sequence library and extrapolate the numbers counted in Fig 2b to reconstruct an artificial model for duplicated MHC gene copies influencing speciation or evolution without first knowing their categories (classical versus non-classical presentations), functions, overall structure and coding ability, transcriptional activity and genomic locations. Malmstrøm et al (2016)<sup>2</sup> provided no properly organized genomic assemblies or genomic gene maps and no information about genomic distribution of the MHC I or II sequenced fragments or the duplication mechanisms involved. If they had done so, they might have added important information to better assess and place the threshold MHC I copy numbers and gene distributions into some sort of genomic perspective<sup>4</sup>. More reliable models for the evolution of MHC class I genomic duplications might be achieved by providing duplication gene maps and the phylogenetic relationships of the duplicated gene sequences showing likely duplication mechanisms, where and how these genes are located relative to each other, and how the genomic structures have changed in a comparative analysis between species. See Kulski et al. (2004)<sup>5</sup> for an example of one such duplication model. Mapping with phylogeny is a more informative approach than just constructing phylogenetic trees using one or more single exonic sequences from a limited number of each species and then claiming that the changes in copy numbers influence the speciation rates of almost the entire number of extant fish species. Perhaps, the Malmstrøm et al (2016)<sup>2</sup> low coverage sequence libraries could still be used effectively to reconstruct full-length gene structures and targeted genomic regions that harbour multiple copies of the MH(C) genes in a comparative analysis.

## 5. The multiple-optima Ornstein Uhlenbeck (OU) model<sup>6</sup>.

(a) According to Malmstrøm et al (2016)<sup>2</sup> the multiple-optima OU model vastly outperformed alternatives such as Brownian motion, white noise, single-peaked OU and early-burst models. This finding corroborated their hypothesis that MHC I copy number evolution is characterized by selection toward intermediate optima, resulting from a tradeoff between detection and elimination of pathogens. Presumably, the authors preferred the OU multiple-optima model to the other models because it supported rather than falsified their hypothesis. Of course, this highly artificial computing model did not detect a tradeoff between detection and elimination of pathogens, this would be the authors' own biological hypothesis and bias. The OU model intrinsically sets optima (biases) according to its built-in algorithm<sup>6</sup>, and this is one of the main objections to the use of this prediction model. The OU model artificially generates bias because its purpose is to find the trait optimum that stabilizes selection<sup>6</sup>. The misguided conclusion by Malmstrøm et al (2016)<sup>2</sup> in using the OU model is that the trait optimum influences speciation.

(b) Interestingly, Malmstrøm et al (2016)<sup>2</sup> did not directly test the opposing hypothesis that the MHC I copy number evolution is not characterized by selection toward intermediate optima, and does not result from a tradeoff between detection and elimination of pathogens. Possibly, their best control in this regard was the simple Brownian model that did not work as well for them as the multiple-optima OU model that has extra parameters such as the addition of an overall optimum trait value to which all lineages are attracted. Other evolutionists often prefer the Brownian model for the reason that it is a simple, neutral model without the added bias of creating optimum trait value.

(c) The OU multiple-optima model is not a fool-proof algorithm, and a number of evolutionists believe that it can be an unreliable or misleading model. According to Cooper et al (2016)<sup>7</sup>, although widely used, the properties of the OU model, and other direct extensions of the Brownian model, are poorly understood leading to the potential for inappropriate use and misinterpretation of results. In particular, Cooper et al (2016)<sup>7</sup> used computer simulation studies to demonstrate that the single peaked OU model error rates are unacceptably high when tree size is small (< 200 species tips), when likelihood ratio tests or Akaike information criterion (AICc) are used to select the best model, and when measurement errors are introduced into the data. They also showed that when the alpha parameter of trait evolution was extremely small (<1) in the OU model it was indistinguishable from Brownian motion, and as the alpha value became larger it favoured OU prediction models, until the larger values of alpha were indistinguishable from white noise and it was therefore independent of phylogeny. The alpha values for the Malmstrøm et al (2016)<sup>2</sup> model selection analysis were markedly less than one (Supplementary Table 13), suggesting that they could have accepted the Brownian model over the OU model as the better model fit.

## 6. The BiSSE threshold model.

Malmstrøm et al (2016)<sup>2</sup> carried out binary state speciation and extinction (BiSSE) analysis to estimate differences in diversification rate between lineages with high and low MHC I copy numbers. They found that diversification rates based on correlation estimates differed most when the threshold was placed between 20 and 25 copies (Fig. 3c). With a threshold in this range, the model with two separate speciation rates for lineages with high and low copy numbers was statistically better supported than a model with a single speciation rate parameter. On this basis, they concluded that, 'These results suggest that the influence of MHC I genes on speciation rates is stronger in species that have already evolved at least 20 copies.' In comparison, the number of MHC I gene copy numbers in humans (excluding haplotype differences) is approximately 18 genes; 6 classical and non-classical MHC I genes, 5 CD1 genes, and 5 PHFZ genes (MICA, MICB, MR1, HFE, Zn-A2-GP, etc). Thus, in comparison to some fish species, humans are diversifying along very nicely as a 'diversified' species approaching the 'magical'

threshold of between 20 and 25 MHC I copies.

It is noteworthy that Maddison et al (2007)<sup>8</sup> highlighted the following assumptions that need to be taken into account when using their BiSSE package. For the BiSSE model analysis none of the characters associated with speciation rates can be said to be causing or influencing evolution, even if Maddison et al (2007)<sup>8</sup> write, 'the correct conclusion given a significant result using our method is that the character examined or a codistributed character *appear to be controlling diversification rates.*' At best, the binary character state is an association, at worse a misleading one. Maddison et al (2007)<sup>8</sup> provided the following cautions and assumptions about the likelihood of the BiSSE (binary-state speciation and extinction) model:

- (a) the transitions happen instantaneously over the time scales considered (i.e., ignore periods of time during which a species is polymorphic).
- (b) these events are independent of one another; in particular, the character state change does not, in and of itself, cause speciation (or vice versa).
- (c) an accurate rooted phylogenetic tree with branch lengths is known (the "inferred tree") and the character state is known for each of the terminal taxa.
- (d) the tree is assumed complete: all extant species in the group have been found and included.
- (e) all terminal taxa are contemporaneous, and the tree is ultrametric (i.e., the total root-to-tip distance is the same for all tips).

I'm not convinced that Malmstrøm et al (2016)<sup>2</sup> considered or accepted these constraining assumptions when using BiSSE modeling.

#### 7. Speciation and diversification rates.

(a) In the study by Santini et al (2009)<sup>9</sup>, the speciation rates within the Percomorpha clade were calculated to be at least ten times greater than in the Gadiformes order. Yet, according to Malmstrøm et al (2016)<sup>2</sup>, there were fewer than 20 copies of the U genes for each of the 5 species in the Perciformes clade, compared to more than 20 copies and up to 100 copies of the U genes in 16 of 30 species in the Gadiformes order (Fig 2b, Malmstrøm et al 2016<sup>2</sup>). The use of only 5 species in the Perciformes order is a gross underestimate of the ten thousand or more species found in that order<sup>9</sup>. Moreover, in the Gadiformes order, there were closely related species (n=16) with > 20 U genes and different groups of closely related species (n=14) with < 20 U genes. Again, the number of species that were sequenced in the Gadiformes order is grossly under-represented. The Gadiformes comprises 10 families and more than 600 species<sup>9</sup>, whereas Malmstrøm et al (2016)<sup>2</sup> sequenced only 27, i.e.,  $27 \times 100/600 = 2,700/600 = 4.7\%$  analysis of extant sequence, a percentage that is simply not good enough to support their extravagant conclusions. Is < 5% of the 600 species really representative of the Gadiformes. Malmstrøm et al (2016)<sup>2</sup> have to be more temperate with their conclusions using such a small representative sample. There are clear inconsistencies with MHC I a3 fragment copy numbers in the Gadiformes order. The MHC I a3 frag copy numbers are low (<20) for Moridae and *M. occidentalis* in Macrourinae, and for Phycinae, Lotinae, and three species in Gadinae. Five species of Gadinae have between 20 and 40 copies. On the other hand, Bregmacerotidae, Merlucciidae, Melanonidae, Muraenolepidodae, and Trachyrincinae have MHC I a3 fragment copy numbers between 50 and 100. The threshold levels (20 to 25) are all over the place. Moreover, the lineage, genomic block duplication and hitchhiking (linkage) effects on MHC gene duplications (8 to >100 copies) in the Gadiformes have not been taken into account in the analysis of speciation rates (Fig 3, Malmstrøm et al 2016<sup>2</sup>), and, therefore, make the entire analysis unreliable.

(b) “Diversification rate analyses were calculated on the basis of the time-calibrated phylogeny and counts of species richness in each of the 37 mutually exclusive clades of teleost fishes”<sup>2</sup>; mostly from the Gadiformes order. The MHC I speciation model of Malmstrøm et al (2016)<sup>2</sup> appears contradictory for the Perciformes (10,033 species) that have speciation rates 18 times greater than Gadiformes (555 species)<sup>9</sup> and yet the MHC I a3 fragment copy numbers are at least two times lower in Perciformes than Gadiformes (Fig 2b, Malmstrøm et al 2016<sup>2</sup>). Also, the Anabantiformes have 252 species – a species rate 40 times lower than Gadiformes<sup>9</sup> and yet their MHC I a3 fragment copy numbers are at least two times higher than in Perciformes.

(c) Considering that there are more than 29,000 species of teleost fishes<sup>9</sup>, a highly limited analysis by Malmstrøm et al (2016)<sup>2</sup> of using a sample group of less than 0.2% of the available extant species cannot be considered to be statistically, taxonomically or biologically significant or sufficiently reliable to conclude that, “Evolution of the immune system influences speciation rates in teleost fish”<sup>2</sup>. What does a species half-life of 25 million years mean in the context of 29,000 species of teleost fishes? If the multiple-regime OU model is wrong, highly biased or misinterpreted then does it validate or support the overall hypothesis of Malmstrøm et al (2016)<sup>2</sup>? Also, what does an optimal trait actually mean in the context of 29,000 species? If, a suboptimal number of MHC I copies are detrimental to a species, then how have divergent species managed to survive for so long with a half-life of 25 million years of adaptation? Also, if, as Malmstrøm et al (2016)<sup>2</sup> say, ‘Such gene family expansions may promote biological diversification by introducing new raw genetic material, potentially resulting in sub- or neofunctionalization and thus novel immunological pathways.’, then which of the non-optimal (greater than or less than the threshold of 20 -25 copies) MHC I genes are detrimental to the species? In this regard, there must be a gradation of functionally good and bad MHC I genes as their copy numbers approach the threshold (optimally good) and then deteriorate beyond it. Is this assumption of a MHC I copy number functional trait value as a quantitative marker of speciation at all testable?

8. In their discussion, Malmstrøm et al (2016)<sup>2</sup> referenced the hypothesis of T cell depletion and hybrid fitness by Eizaguirre et al (2009)<sup>10</sup> and concluded that, “Our analyses identify this threshold at 20–25 MHC I copies, suggesting that the effect of T cell depletion on hybrid fitness becomes more pronounced in this range and that this might affect mate choice in species with copy numbers above this threshold, promoting inbreeding and reinforcement.” Eizaguirre et al (2009)<sup>10</sup> suggested that, “Super-optimal individual MHC diversity should be a common disadvantage for species hybrids in vertebrates, resulting in elevated parasite loads.” In this regard, if high copy numbers of the MHC class I genes leads to hybridization and loss of the immune system as inferred by Eizaguirre et al (2009)<sup>10</sup>, then this more than likely would lead to extinction of populations and species. Extinction would be the most extreme and bizarre form of immune system influence on speciation rates. Furthermore, it is extremely speculative for Malmstrøm et al (2016)<sup>2</sup> to say that high copy numbers of the MHC class I genes with copy numbers above the threshold of 20 to 25 copies, promotes inbreeding and reinforcement, because, in fact, there is no such evidence for it. A more reasonable hypothesis is that high copy numbers of linked MHC class I genes, such as in rhesus macaque, or the mouse<sup>11</sup>, or the cod<sup>12</sup>, might benefit the species to better adapt to microbial inhabitants in a greater variety of geographical environments, although the evidence for this is tenuous as well. Despite ongoing debates, the selective advantage of MHC diversity in host-pathogen coevolution might not be easily resolved (at the macroevolution level) because of the constant number of insults by large numbers of pathogens in the life-time of an individual organism, population or a species and the arms race or Red Queen effect. Studies on extant species will always discover an example of a pathogen associated with a polymorphic MHC gene that might favour selective advantage for host-pathogen coevolution, whereas the pathogen that caused the extinction of a species



is rarely or never found. To conclude that the immune system (that is, different copy numbers of the class I MHC genes<sup>2</sup>) influences speciation rates, it would have to be shown that the immunity gene products can commonly create reproductive barriers or genetic incompatibilities among populations that permit the maintenance of the genetic and phenotypic distinctive-ness of these populations in geographical proximity<sup>13</sup>; and this was not done<sup>2</sup>.

9. Malmstrøm et al (2016)<sup>2</sup> did not provide any reliable evidence to support their speculation that evolution of the immune system influences speciation rates in teleost fish or that increasing MHC I diversity facilitates speciation<sup>7</sup>. Instead, Malmstrøm et al (2016)<sup>2</sup> used their limited data and analyses using speculative models to jump to highly unsupported conclusions and quickly position the cart before the horse. Dijkstra & Grimholt<sup>1</sup> pointed out that the Malmstrøm et al (2016)<sup>2</sup> title “Evolution of the immune system influences speciation rates in teleost fish”, is unsubstantiated, and that their hypothesis seems to be “the wrong way around”. It should have been, “Speciation (rate) influences the evolution of the immune system in teleost fish.” Or, “Speciation rates are associated with diversity of MHC class I genes in teleost fishes”, which perhaps is too obvious and underwhelming. This is not simply the chicken or the egg causality dilemma; in fact, the change in title is better supported by the literature and the established theories of MHC genomic evolution in vertebrates<sup>4</sup>. However, because it is less “sexy” and controversial than the original title, it might not have been so readily published.

10. A large number and variety of genome-wide gene duplications have been associated with speciation<sup>13</sup>, that is, genomic gene duplications are not limited to only class I MHC genes. If MHC I gene duplications effect or affect speciation, how do the other hundreds of gene duplications contribute to speciation rates? Also, do sequence variants or mutations in non-duplicated genes have any influence on speciation rates? It seems absurd to pick on only one group of gene duplications (e.g., MHC class I genes<sup>2</sup>) as those that are responsible for speciation and ignore all the others as an inconvenience. For example, a relatively recent comparative genomic study revealed how genomes change with speciation in an examination of genomes from five cichlid fish species, an ancestral lineage from the Nile, and four species from the East Africa lakes, Tanganyika, Malawi, and Victoria<sup>14</sup>. Compared to the ancestral Nile lineage, the East African cichlid genomes had many alterations in regulatory elements, accelerated evolution of protein-coding elements in genes for pigmentation, an excess of gene duplications, and other distinct features that affect gene expression associated with transposable element insertions and novel microRNA. Each species also contains a reservoir of mutations different from the other species<sup>14</sup>. Much of the diversity between the cichlid fish species evolved in a nonparallel manner often rapidly due to sexual selection and genetic conflicts between males and females or between different regions of the genome at a regulatory level<sup>14</sup> rather than by the slower and weaker forces of classical natural selection<sup>13</sup>. If sexual selection and genetic conflict at the genomic regulatory level are the prime movers of speciation rate, it is difficult to conclude that the variable diversity of a few MHC gene copies are responsible for speciation as well as for the many other associated genomic changes associated with speciation.

11. Malmstrøm et al (2016)<sup>2</sup> informed us in the introduction section of their publication that “Our results highlight the plasticity of the vertebrate adaptive immune system and support the role of MHC genes as ‘speciation genes’, promoting rapid diversification in teleost fishes.” MHC class I gene copy number variability occurs across many different species, families, orders and domains. Because there is such enormous variability in MHC class I gene copy number for hundreds<sup>4</sup> or possibly even thousands of different chordate species, it is not possible to conclude meaningfully that the expansion of MHC class I genes provides an undefined advantage of one species over another. For example, the great apes (humans, chimpanzees, gorillas and orangutans) have about six functional MHC class I genes, whereas

the old and new world monkeys often have up to 15 or more<sup>4,11</sup>. Is this evidence that the MHC class I genes influence the rate of speciation in primates? And if so, what does that really mean in the whole scheme of things? How do the species with low copy numbers of MHC class I genes survive so well over millions of years without the presence of another 90 to 100 copies of MHC class I genes? This question is often neglected, and yet it is important for a better understanding of the function and evolution of MHC genes between and within the vertebrate species.

#### 12. Taxonomic and lineage markers.

Mutations, indels and duplications drive diversity and evolution. However, most mutated genes within species and their families do not create or influence speciation rates in the sense that Malmström et al (2016)<sup>2</sup> use the term, 'speciation genes'. In comparative genomics and their sequence relationships between different species, most genomic sequences range between newly derived genes and the ultraconserved or the essential core coding and noncoding genes with varying amounts of sequence differences. Some genic and nongenic sequences such as the MHC genes and retrotransposons are highly polymorphic and therefore are useful taxonomic markers at the individual, population, species and broader lineage levels. The MHC gene sequences clearly are one of these useful taxonomic or lineage markers along with olfactory receptors, immunoglobulins, globins, HOX, TOLLS, KIRs, mitochondrial DNA, ribosomal RNA sequences and thousands of others that can be used comparatively in the phylogeny to undertake an examination of the accuracy and reliability of current taxonomical rankings and sister lineages. However, because many thousands of coding and noncoding genes (or sequences) are variants (polymorphic) or vary in copy numbers, we do not immediately or easily imply that all or some of them are responsible for speciation without providing further concrete evidence. This kind of extrapolation without the burden of proof is absurd and wrong. Similarly, to say that the polymorphisms demonstrate natural selection as if natural selection was a biological or molecular mechanism is meaningless without showing experimentally how these polymorphisms benefit or disadvantage the organism over all the other different polymorphisms.

13. On the basis of either *a priori* or *a posteriori* reasoning, the immune system obviously affects the wellbeing of individuals and populations, but whether it can be extrapolated to speciation events and speciation rates remains highly dubious and most probably unlikely. It seems too farfetched to blame MHC class I genes with high copy numbers over threshold levels of promoting inbreeding and reinforcement<sup>2</sup> because this in turn could create hybrid inviability or sterility resulting in postzygotic isolation. Although the population conditions in many models of rapid speciation do favour inbreeding and/or hybridization<sup>13</sup>, none of the teleost species tested by Malmström et al (2016)<sup>2</sup> were shown to be either inbreeding or in postzygotic isolation. The factors responsible for either prezygotic or postzygotic isolation are likely to be independent of the adaptive immune system, although zealots might argue otherwise. Hybridization between diverging lineages in post-zygotic reproductive isolation can trigger genome instability. For most animals without an adaptive immune system and for plants without a MHC, speciation depends on the shrinkage, expansion and equilibrium (e.g., aneuploidization and dysploidy) of the genome and the containment and functionality of all the essential genomic information to develop an optimal balance between stability and plasticity within the organism in order to first survive and then propagate and expand itself as a new species<sup>13</sup>. In those rare and 'traumatic' transitional situations, there is no need for particular 'speciation' genes such as variable copies of the class I MHC genes to influence speciation. The rarely observed transition from population 'trauma' to a new speciation event depends on an array of totally different factors for creating postzygotic isolation events including interbreeding between semi-isolated populations and an elaboration on the existence of stress-induced changes in chromosomal and ploidy integrity both in hostile and non-hostile environments.

14. Finally, Malmstrom et al (2016)<sup>2</sup> admirably sequenced 66 teleost species by a next generation sequencing method and identified an array of MHC I and MHC II exonic fragments for phylogenetic and speciation analysis using the multiple-regime OU model to predict the optimal MHC I copy number as an evolutionary trait optimum affecting speciation. However, the conclusions of the paper by Malmstrom et al (2016)<sup>2</sup> especially for the MHC I gene copy numbers are unreliable because they are based on far too many assumptions, speculations, contradictions, incomplete or missing data and unproven predictive models with little or no empirical evidence in support. Nevertheless, their simple, but controversial hypothesis is published, and now it is up to them and others to test its validity and "consider plausible alternative hypotheses in a firm hypothesis-testing framework in which alternative hypotheses make clear [and sensible] predictions of emerging patterns that can be unambiguously associated with particular models."<sup>7</sup>

## References

1. Dijkstra J, Grimholt U: Major histocompatibility complex (MHC) fragment numbers alone – in Atlantic cod and in general - do not represent functional variability. *F1000Research*. 2018; **7**. [Publisher Full Text](#)
2. Malmstrøm M, Matschiner M, Tørresen OK, Star B, Snipen LG, Hansen TF, Baalsrud HT, Nederbragt AJ, Hanel R, Salzburger W, Stenseth NC, Jakobsen KS, Jentoft S: Evolution of the immune system influences speciation rates in teleost fishes. *Nat Genet*. **48** (10): 1204-10 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Shiina T, Suzuki S, Kulski J: MHC Genotyping in Human and Nonhuman Species by PCRbased Next-Generation Sequencing. 2016. [Publisher Full Text](#)
4. Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H: Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunol Rev*. 2002; **190**: 95-122 [PubMed Abstract](#)
5. Kulski JK, Anzai T, Shiina T, Inoko H: Rhesus macaque class I duplicon structures, organization, and evolution within the alpha block of the major histocompatibility complex. *Mol Biol Evol*. 2004; **21** (11): 2079-91 [PubMed Abstract](#) | [Publisher Full Text](#)
6. Hansen TF: STABILIZING SELECTION AND THE COMPARATIVE ANALYSIS OF ADAPTATION. *Evolution*. 1997; **51** (5): 1341-1351 [PubMed Abstract](#) | [Publisher Full Text](#)
7. Cooper N, Thomas GH, Venditti C, Meade A, Freckleton RP: A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biol J Linn Soc Lond*. 2016; **118** (1): 64-77 [PubMed Abstract](#) | [Publisher Full Text](#)
8. Maddison WP, Midford PE, Otto SP: Estimating a binary character's effect on speciation and extinction. *Syst Biol*. 2007; **56** (5): 701-10 [PubMed Abstract](#) | [Publisher Full Text](#)
9. Santini F, Harmon LJ, Carnevale G, Alfaro ME: Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC Evol Biol*. 2009; **9**: 194 [PubMed Abstract](#) | [Publisher Full Text](#)
10. Eizaguirre C, Lenz TL, Traulsen A, Milinski M: Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. *Ecol Lett*. 2009; **12** (1): 5-12 [PubMed Abstract](#) | [Publisher Full Text](#)
11. Shiina T, Blancher A, Inoko H, Kulski JK: Comparative genomics of the human, macaque and mouse major histocompatibility complex. *Immunology*. 2017; **150** (2): 127-138 [PubMed Abstract](#) | [Publisher Full Text](#)
12. Parham P: How the codfish changed its immune system. *Nature Genetics*. 2016; **48** (10): 1103-1104 [Publisher Full Text](#)
13. Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre GP, Bank C, Brännström A, Brelford A, Clarkson CS, Eroukhanoff F, Feder JL, Fischer MC, Foote AD, Franchini P, Jiggins CD, Jones FC, Lindholm AK, Lucek K, Maan ME, Marques DA, Martin SH, Matthews B, Meier JI, Möst M, Nachman MW, Nonaka E, Rennison DJ, Schwarzer J, Watson ET, Westram AM,

Widmer A: Genomics and the origin of species. *Nat Rev Genet.* 2014; **15** (3): 176-92 [PubMed Abstract](#) | [Publisher Full Text](#)

14. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, Turner-Maier J, Johnson J, Alcazar R, Noh HJ, Russell P, Aken B, Alföldi J, Amemiya C, Azzouzi N, Baroiller JF, Barloy-Hubler F, Berlin A, Bloomquist R, Carleton KL, Conte MA, D'Cotta H, Eshel O, Gaffney L, Galibert F, Gante HF, Gnerre S, Greuter L, Guyon R, Haddad NS, Haerty W, Harris RM, Hofmann HA, Hourlier T, Hulata G, Jaffe DB, Lara M, Lee AP, MacCallum I, Mwaiko S, Nikaido M, Nishihara H, Ozouf-Costaz C, Penman DJ, Przybylski D, Rakotomanga M, Renn SCP, Ribeiro FJ, Ron M, Salzburger W, Sanchez-Pulido L, Santos ME, Searle S, Sharpe T, Swofford R, Tan FJ, Williams L, Young S, Yin S, Okada N, Kocher TD, Miska EA, Lander ES, Venkatesh B, Fernald RD, Meyer A, Ponting CP, Strelman JT, Lindblad-Toh K, Seehausen O, Di Palma F: The genomic substrate for adaptive radiation in African cichlid fish. *Nature.* 2014; **513** (7518): 375-381 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the rationale for commenting on the previous publication clearly described?**

Yes

**Are any opinions stated well-argued, clear and cogent?**

Yes

**Are arguments sufficiently supported by evidence from the published literature or by new data and results?**

Yes

**Is the conclusion balanced and justified on the basis of the presented arguments?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Referee Expertise:** MHC genes, retrotransposons, evolutionary genomics and biology, population diversity.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 03 Sep 2018

**Johannes M. Dijkstra**, Fujita Health University, Japan

Dear Dr. Jerzy Kulski,

Thank you for your review and support of our article. We appreciate that an expert such as you is willing to join the public debate so that erroneous/unsubstantiated messages like the ones presented by Malmström *et al.* can not take hold in our field.

Your comments are very extensive and valuable, and we now refer the readers to them.

Our special fields of scientific expertise are MHC genes and molecules, and in our first manuscript version we only concentrated on those. If there is no sufficient unity among the units used for mathematical modeling, that modeling, or the functional explanation of the resulting model, can

never make sense. However, we now realized that for a large audience these MHC-specific issues might not be so clear, and that it is better to also address the questionable modeling methods used by Malmstrøm *et al.* Therefore, we now have added two paragraphs dedicated to this questionable modeling, titled “*Detailed discussion of the use of the Ornstein-Uhlenbeck model by Malmstrøm et al.*” and “*Additional criticisms in regard to the modelling by Malmstrøm et al.*”<sup>1</sup>. We realize that we use different language in regard to these topics than a theoretical biologist would use, but we hope that nevertheless we address the issues clearly and correctly.

Sincerely,

also on behalf of Dr. Unni Grimholt

Hans (J.M.) Dijkstra

**Competing Interests:** No competing interests were disclosed.

Referee Report 27 July 2018

doi:10.5256/f1000research.16766.r36541



**Brian Dixon** 

Department of Biology, University of Waterloo, Waterloo, ON, Canada

The critique of Malmstrøm *et al.* presented here makes some very valid points that are well supported by the literature.

It has long been true in fish MHC research that the fact that a gene has not been reported to be present in a particular species does not mean that it is not present. Modern genomics techniques have presented better proof for this assertion, with the lack of an MHCII/CD4 pathway in gadids being the most prominent example, but even modern genomics techniques are not iron clad 100% proof and should be checked very carefully before definitive statements are made. Thus the comments about verifying the presence or absence of specific genes in the numerous species by other means is valid.

Additionally, the treatment of all U and Z genes as identical units while ignoring the allelic diversity of each gene within those classes is indeed a serious flaw in the reasoning of Malmstrøm *et al.* There is significant variability in diversity U gene families which will have differing effects on T cell selection that simply counting gene numbers will not address.

Dijkstra and Grimholt's critique should be carefully read and addressed.

**Is the rationale for commenting on the previous publication clearly described?**

Yes

**Are any opinions stated well-argued, clear and cogent?**

Yes



**Are arguments sufficiently supported by evidence from the published literature or by new data and results?**

Yes

**Is the conclusion balanced and justified on the basis of the presented arguments?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Referee Expertise:** Fish Immunology, MHC genes, antigen presentation

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 03 Sep 2018

**Johannes M. Dijkstra**, Fujita Health University, Japan

Dear Dr. Brian Dixon,

Thank you for your review and support of our article. We appreciate that an expert such as you is willing to join the public debate so that erroneous/unsubstantiated messages like the ones presented by Malmstrøm *et al.* can not take hold in our field.

Sincerely,

also on behalf of Dr. Unni Grimholt

Hans (J.M.) Dijkstra

**Competing Interests:** No competing interests were disclosed.

Referee Report 19 July 2018

doi:10.5256/f1000research.16766.r35833



**Anthony B. Wilson**  1,2

<sup>1</sup> Department of Biology, Brooklyn College CUNY, New York, NY, USA

<sup>2</sup> The Graduate Center, City University of New York, New York, NY, USA

Dijkstra & Grimholt present a critical analysis of Malmstrom *et al.*'s 2016 *Nature Genetics* article<sup>1</sup>, which investigated the evolution of MHC I and II loci in gadiform fishes using a low coverage genomic screen of 66 species, inferring a link between adaptive immune evolution and speciation rates in this group. Dijkstra & Grimholt's criticisms are wide ranging – I deal with each of their major areas of concern below:

**I. MHC Class II loss in gadiform fishes.** The authors highlight two serious flaws in the Malmstrom analysis, demonstrating that the original dataset contains sequence reads of MHC II and associated loci

in several species that were overlooked in the original analysis. Equally importantly, datasets for several of the outgroup taxa lack these genes, raising questions concerning the reliability of the underlying data. Malmstrom *et al.*'s genomic screen is understandably low coverage given the taxonomic breadth of their survey, but I agree with Dijkstra & Grimholt that based on the existing evidence, one cannot confidently infer the timing of MHC II gene loss in this group.

**II. MHC I allele counting strategy.** Dijkstra & Grimholt challenge the allele counting strategy used by Malmstrom *et al.*, particularly their focus on U & Z loci (teleost fish have at least 5 different MHC I lineages<sup>2</sup>), based on their assumption that these loci are chiefly involved in binding peptide ligands. While I agree that grouping U & Z loci together simplifies their known functional complexity (I was rather confused by this approach myself when reading the original paper), here I feel that Dijkstra & Grimholt could be more constructive in their criticism. At present, it's not entirely clear what type of analysis they would feel would be most suitable. I would also suggest providing slightly more context on the study system to assist readers who may be unfamiliar with the original work.

While Dijkstra & Grimholt have elsewhere provided compelling evidence that Z loci may have a very different function, it's not clear whether they're suggesting that Malmstrom *et al.* should have focused solely on U loci, or whether it would have been more appropriate to include all MHC lineages in their analyses. Either way, I would have liked to see whether analyzing the data in the manner preferred by the authors would impact the conclusions of the original article.

I agree that experimental evidence would be necessary to conclusively demonstrate a link between allelic diversity and function, but given the taxonomic breadth of Malmstrom *et al.*'s study, surely they wouldn't expect experimental evidence for all species included in the original study - How much experimental evidence would they deem sufficient? At present, it's not clear whether they're simply suggesting that Malmstrom *et al.* should have been more circumspect in their conclusions, or whether they feel that the results of the analysis are entirely unreliable. Clarification of this point is essential.

**III. Testing the relationship between MHC allelic diversity and speciation rates in gadiform fishes.** The authors raise concerns on the modelling approach used by Malmstrom *et al.*, including their combined analysis of U and Z loci (see above), and their lack of a biologically realistic model of gene evolution, incorporating MHC gene gain and loss<sup>3</sup> - I agree with these criticisms. I do, however, take some issue with their contention that Malmstrom *et al.*'s hypothesis is wholly invalid. While there is indeed strong evidence of trans-species MHC polymorphism in some well-studied vertebrate lineages, this does not invalidate an experimental test of an alternative hypothesis. If Dijkstra & Grimholt feel that Malmstrom *et al.* have their hypothesis "the wrong way around", are there any data/analyses that could convince them otherwise?

## References

1. Malmstrøm M, Matschiner M, Tørresen OK, Star B, Snipen LG, Hansen TF, Baalsrud HT, Nederbragt AJ, Hanel R, Salzburger W, Stenseth NC, Jakobsen KS, Jentoft S: Evolution of the immune system influences speciation rates in teleost fishes. *Nat Genet.* **48** (10): 1204-10 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Grimholt U, Tsukamoto K, Azuma T, Leong J, Koop BF, Dijkstra JM: A comprehensive analysis of teleost MHC class I sequences. *BMC Evol Biol.* 2015; **15**: 32 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Nei M, Rooney AP: Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet.* 2005; **39**: 121-52 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the rationale for commenting on the previous publication clearly described?**

Yes

**Are any opinions stated well-argued, clear and cogent?**

Yes

**Are arguments sufficiently supported by evidence from the published literature or by new data and results?**

Partly

**Is the conclusion balanced and justified on the basis of the presented arguments?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Referee Expertise:** Adaptive Immunity, Evolutionary Biology, Speciation

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 03 Sep 2018

**Johannes M. Dijkstra**, Fujita Health University, Japan

Dear Dr. Anthony B. Wilson,

Thank you for your review and support of our article. We appreciate that an expert such as you is willing to join the public debate so that erroneous/unsubstantiated messages like the ones presented by Malmstrøm *et al.* can not take hold in our field.

You suggest us to provide lay readers with more background information. Following your suggestion, we have tried to do so, by writing a tentative introduction section, but decided not to use it as we found that the necessary size and discussions would distort the article too much. This article is a correspondence, a discussion about another paper, and we feel that this discussion character should be clear throughout the article. Furthermore, F1000Research proscribes that "Correspondence articles are short, peer reviewed comments", and, as it is, the article is already quite lengthy.

You ask us to be more constructive in our criticism towards Malmstrøm *et al.*, and to explain to them what they should have been doing instead. First of all, we would have liked them to make their claims on MHC class II system genes solid, because that seems within close experimental reach, and we believe that they should have concentrated their article on that topic. In regard to MHC class I, we would have liked them to either study those genes intensively, or to have refrained from any modeling, and especially to have refrained from highlighting a resulting model in the title. In the supplementary files 1 and 2 we already made some detailed comments about which experiments would be necessary to make the MHC class II system claims more solid and within acceptable standards. Now we also added a conclusion section to the main text which explains, in general terms, what we would like Malmstrøm *et al.* to do or not to do as follows:

**"Conclusion**

“Malmstrøm *et al.*<sup>1</sup> used low-coverage genome sequencing for comparison of 66 mostly neoteleost fish, and so helped with elucidating their phylogeny. They found that intact MHC class II system genes may be completely absent in Gadiformes, and believed that related non-gadiform fishes have intact MHC class II system genes. However, their genomic databases were incomplete and in the case of many Gadiformes spiked with reads from MHC class II system genes that may or may not be contaminations, so that final conclusions require some additional analysis of at least a few species at the gadiform/non-gadiform clades border. We suggest that they need to perform a number of PCR and sequencing experiments to clarify this matter. When comparing class I and class II situations in their investigated neoteleosts, Malmstrøm *et al.*<sup>1</sup> also found that their earlier theory, which was that the absence of an MHC class II system might explain the high number of MHC class I genes in Atlantic cod<sup>13</sup>, could not be corroborated. Instead, solely based on estimations of U+Z a3 fragment numbers, they<sup>1</sup> proposed a new theory on MHC class I evolution which they referred to in their manuscript title. We hope to have shown sufficiently that their conclusions on MHC class I evolution were unsubstantiated, that estimation of U+Z a3 fragment numbers is not a proper way to analyze MHC functions or MHC evolution, and that, apart from not investigating logical units that are better suited for their methods of modeling, also the number estimations and modelling systems used by Malmstrøm *et al.*<sup>1</sup> were flawed and/or non-trustworthy. Before any meaningful discussion can be started about the evolution of MHC class I genes in neoteleosts, a much higher level of information about sequences and genomic positions is necessary.”

As to your question how we would have addressed the issue of MHC numbers/variation and thymic T cell depletion. The answer is that we probably wouldn't try. Meaningful modeling would require a better understanding of how positive and negative T cell selections in the thymus, involving multiple different MHC molecules, contribute both quantitatively and qualitatively to the T cell pool. Regardless of the thymic selection model, we might be interested in MHC class I gene numbers, but then we would first want to separate the question into functional MHC subclasses. For example, it may be interesting to see whether there is some evolutionary pattern in the number of classical type polymorphic MHC class I genes (a question which Malmstrøm *et al.* erroneously seem to think that they were addressing), in the number of genes of nonclassical MHC class I families, or the number of MHC class I pseudogenes. Possibly, selection for increased diversification speeds may select for unstable haplotypes with many tandemly organized genes and pseudogenes that can function as a recombination reservoir. It would be interesting to see whether there are differences in numbers of MHC class I pseudogenes between those fish species that more stably maintained ancient variation and those that more rapidly acquired new variation. Basically, we would try to get good information on all investigated genes (and not just count a3 fragments), and then would try to answer questions one at the time and directly connected to the data (and not try to make an unsubstantiated overarching model). Without good data, we simply wouldn't start modeling.

I hope our above answer, and our extended criticism in the article main text of the models used by Malmstrøm *et al.*, is also sufficient as a response to the issues that you raised in paragraph III of your review.

Sincerely,

also on behalf of Dr. Unni Grimholt

Hans (J.M.) Dijkstra

**Competing Interests:** No competing interests were disclosed.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

F1000Research