Investigation of gammaherpesvirus-induced lymphoproliferative disorders through sequencing of the B cell receptor repertoire

Baiwir Jérôme[1], Xiao Xue[1], Lété Céline[1], Javaux J.[1], Pathammavong M.[1], Sandor R.[1], Leemans S.[1], Machiels Bénédicte[1,*], Gillet Laurent[1,*]

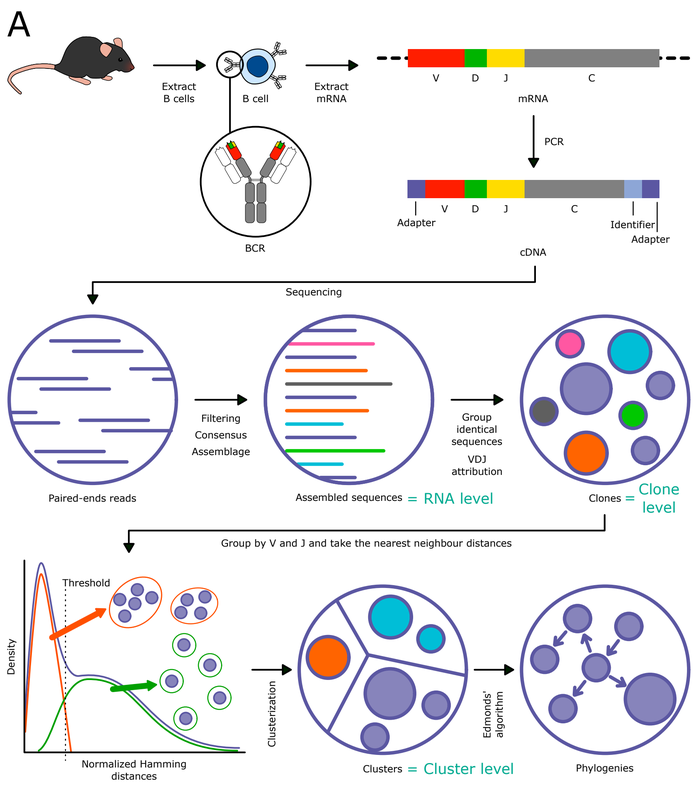
1.Immunology-vaccinology, Department of infectious and parasitic diseases, Veterinary public health, FARAH, University of Liège, Belgium

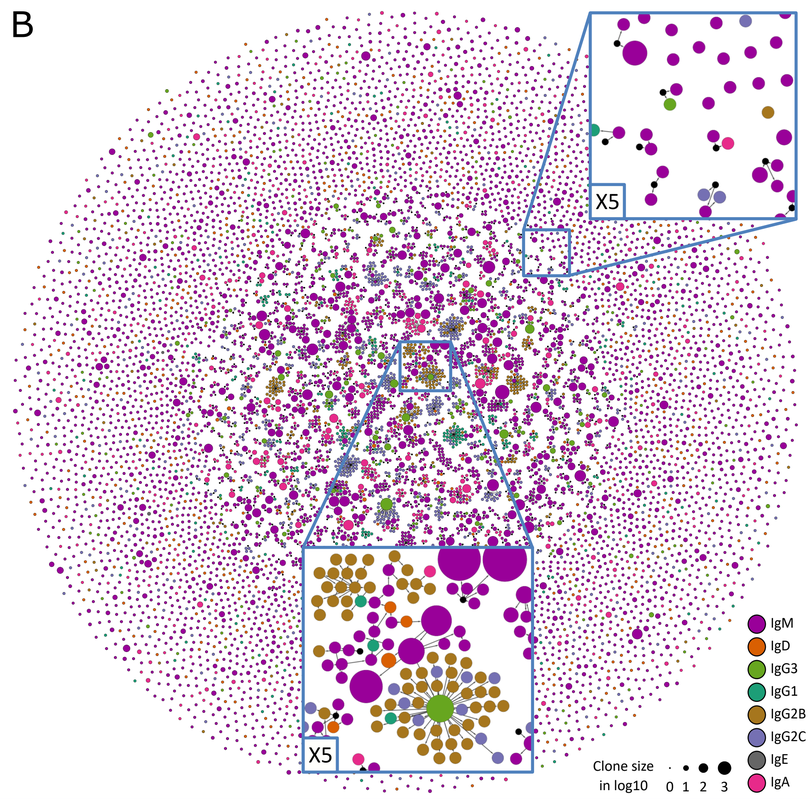
*** Co-senior authors**

Corresponding author : jerome.baiwir@uliege.be

Gammaherpesviruses (γHVs) are ubiquitous viruses that have co-evolved with their hosts. Although these infections remain asymptomatic in most of the individuals, they can cause cancers, mainly lymphoproliferative disorders, in immunocompromised people. In grafted people under immunosuppressive treatment, post-transplant lymphoproliferative disorders are a real risk of death. After primary infection, most of γHVs undergo latent expansion in germinal center (GC) B cells and persists in memory cells. Here, we characterized the impact of the Murid Herpesvirus 4 (MuHV-4), a mouse γHV, on the antibody repertoire of its host.

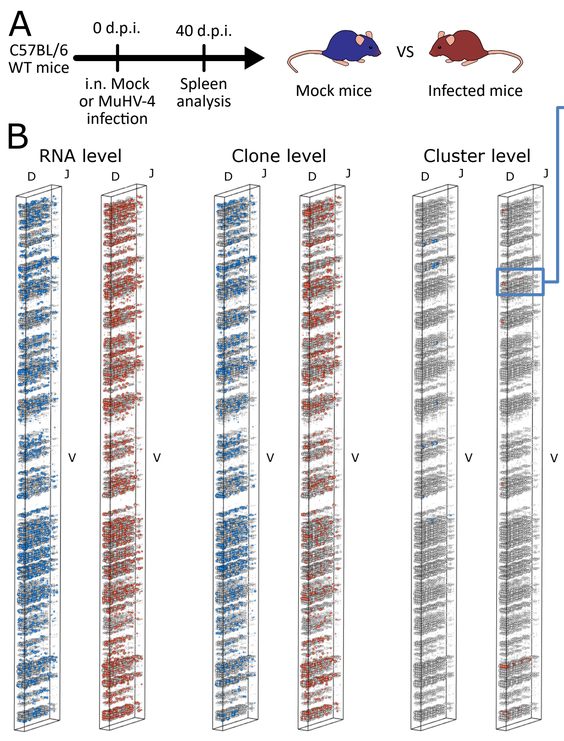
1. Pre-processing implemented to analyse the BCR repertoires of mice

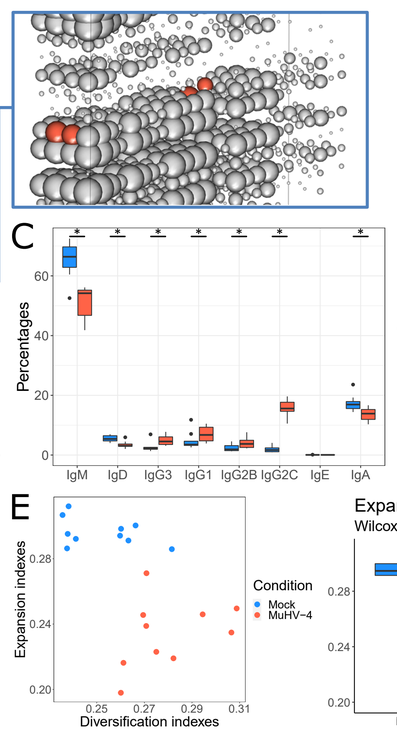




A. Pre-processing of the BCR sequences. B cells are isolated from mice following diverse experimental designs. An artistic representation of a BCR is shown in a circle. V, D, J, and C domains from the heavy chains are represented in red, green, yellow, and grey, respectively. The light chains are represented in white. The mRNAs of the B cells are then extracted, among which the mRNAs coding for the BCRs. These mRNAs are then amplified by PCR, during which adapter sequences for sequencing, and identifiers are inserted. The obtained sequences are then sequenced with 250 bp paired-ended MiSeq (Illumina). The paired-end reads are then filtered based on the quality of their sequencing, keeping reads with a phred >= 30. Reads harbouring the same identifiers are used to build consensus sequences, then the full sequences are assembled based on the paired reads. At this step, we recovered the individual sequences of mRNAs of BCRs, this is referred to the "RNA level". Identical sequences harbouring distinct identifiers have their information merged in order to obtain the clones. More identical sequences means bigger clones. The sequences of the clones are then aligned to the IMGT database containing the V, D, and J sequences in order to identify the combinations beard by the clones. This step is referred to the "clone level". Clones are then grouped by identical V, J, and CDR3 length. Inside these groups, we look at the sequences of the CDR3 of the clones. The smallest normalized Hamming distances separating two clones (nearest neighbour) is computed for each clone. All these nearest neighbour distances can be visualised as a bimodal distribution (blue line). The two underlying modes consist of a mode of clones that have sequences really closed to each other (in orange) that will be part of a same cluster, and another mode of clones that have sequences too different to be clustered with any other clone (in green). An optimal threshold can be calculated in order to best cleave these two modes. A second step (not shown here) is realised by comparing the full sequences of the clones, and not based on the CDR3 only, between them in order to better refine the clusters. At this step, we clustered the clones together based on their similarities, this is referred to the "cluster level". Applying the Edmonds' algorithm to each cluster gives then a model of relation between the clones of the clusters. The BCR repertoires can then be analysed. **B.** Visualisation of the BCR repertoire of a MuHV-4 infected mouse having undergone the pipeline in A. Each circle represents a distinct clone. The size of the circles is proportional to the number of identical sequences retrieved for each clone. Arrows connect clones with sequences derived from each other. The different colors of the circles correspond to the isotype of the sequences. Boxes represent zoomed portions of the BCR repertoire. This visualisation is a sample of the sequenced BCR repertoire. It is limited to 10,000 clones using an algorithm derived from the CC algorithm described by Bashford-Rogers et al, Nature 2019, vol 574 that preserves the initial architecture of the BCR repertoire.

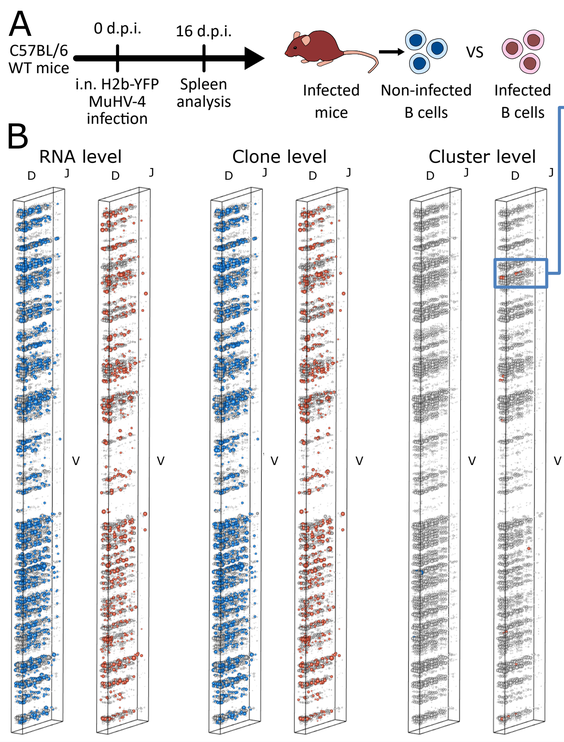
2. Comparison of the BCR repertoires of MuHV-4 infected and non-infected mice

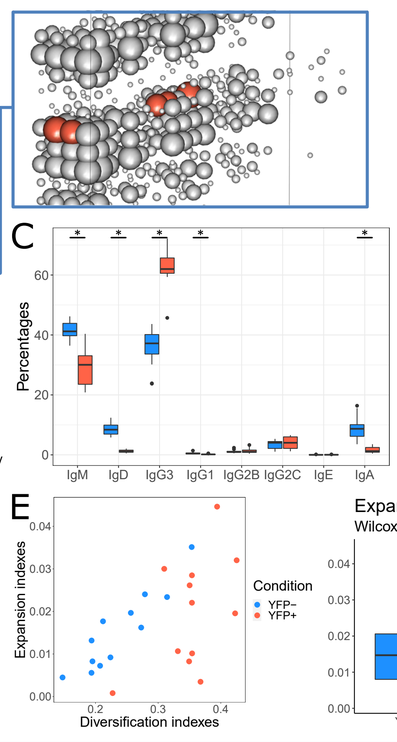




A. Experimental design. **B.** Distinct VDJ recombination frequencies. Each VDJ recombination is represented as a sphere. The position of the spheres correspond to the V, D, and J regions used, and the size of the spheres is proportional to the number of times that combination is retrieved. Red and blue spheres correspond to combinations more often used in MuHV-4 and Mock infected mice respectively (FDR < 0.05). Grey spheres are non significant. A zoom is displayed in the blue box. **C.** Percentages of isotypes in Mock and MuHV-4 infected mice 40 d.p.i. n = 10 in each group. Stars represent a FDR < 0.05 after Wilcoxon rank test. **D.** Isotype switching. The sizes of the circles are proportional to the percentage of clones with a particular isotype. The thickness of the arrows are proportional to the percentage of switches. Red and blue stars represent respectively significantly (FDR < 0.05) higher and lower numbers of isotype switches in B lymphocytes from MuHV-4 infected mice in comparison to their Mock infected counterparts. **E.** Expansion indexes and diversification indexes as defined by Bashford-Rogers et al, Nature 2019, vol 574. The expansion index is a measure of disparity of the size of the clones. The diversification index is a measure of disparity of the size of the clusters. B cells from MuHV-4 infected mice show a lower disparity in the size of their clones and a higher disparity in the size of their cluster in comparison to the B cells of non-infected mice. **F.** Density plot of the selection scores of IGH sequences of B cells from Mock and MuHV-4 infected mice. A higher score implies a higher frequency of replacement mutations.

3. Comparison of the BCR repertoires of infected and non-infected B cells of MuHV-4 infected mice





A. Experimental design. **B.** Distinct VDJ recombination frequencies. Each VDJ recombination is represented as a sphere. The position of the spheres correspond to the V, D, and J regions used, and the size of the spheres is proportional to the number of times that combination is retrieved. Red and blue spheres correspond to combinations more often used in infected and non-infected B cells respectively (FDR < 0.05). Grey spheres are non significant. A zoom is displayed in the blue box. **C.** Percentages of isotypes in infected and non-infected B cells 16 d.p.i. n = 12 in each group. Stars represent a FDR < 0.05 after Wilcoxon rank test. **D.** Isotype switching. The sizes of the circles are proportional to the number of sequences of a particular isotype. The thickness of the arrows are proportional to the number of switches. Red and blue stars represent respectively significantly (FDR < 0.05) higher and lower numbers of isotype switches in infected B cells in comparison to non-infected B cells. **E.** Expansion indexes and diversification indexes as defined by Bashford-Rogers et al, Nature 2019, vol 574. The expansion index is a measure of disparity of the size of the clones. The diversification index is a measure of disparity of the size of the clusters. Infected B cells show a higher disparity in the size of their cluster in comparison to non-infected B cells of infected mice. **F.** Density plot of the selection scores of IGH sequences of infected and non-infected B cells. A higher score implies a higher frequency of replacement mutations.

These results highlight the profound effect of γHV infection on the immune repertoire of their host. It suggests that MuHV-4 infection is not random and establishes preferentially in some B cells and that γHV infection influences the proliferation profile of B cells both acting directly on B cells or via the micro-environment. In the future, identifying the common determinants of these infectable B cell subsets and how γHVs hijack the normal cell cycle could help us to better understand γHVs lifecycle and the lymphoproliferative disorders that they induce. More generally, it could help us to better understand how our environment and especially some infections agents shape our immune responses.

We thank L. Dams and C. Espert for their excellent technical and secretary assistance.