European Journal of Biology and Biotechnology
www.ejbio.org

Mapping on Sea-Star MHC Genes in Invertebrates

Michel Leclerc and Loïc Baerlocher

ABSTRACT

MHC genes have been discovered in Echinodermata (Invertebrates containing 5 classes).2 classes (Ophuirids, Crinoïds) out of 5 possess HLA E, HLA B (Class I), HLADRB1, HLADQB1 (Class II). By the use of Mapping we identified 2 other MHC genes (Class II) in another Echinodermata: the sea star Asterias rubens (Asterids).

Keywords: Mapping; Invertebrates, MHC genes; sea star.

Published Online: April 06, 2021

ISSN: 2684-5199

DOI:10.24018/ejbio.2021.2.2.170

Michel Leclerc *
Immunology of Invertebrates, France.
(e-mail: mleclerc45@gmail.com)
Loïc Baerlocher
Fasteris, Plan les Ouattes, Switzerland.

Corresponding Author

I. Introduction

The Ophuirid (Ophiocomina nigra), the Crinoïd (Antedon bifida) possess MHC genes (Classes I, II) [1]. We attempt, in the precedent work, to discover in the Sea star Asterias rubens genome new MHC genes, using Mapping.

II. Materials and Methods

This section describes how the reads obtained by sequencing are mapped against the reference genome and describes the result of the mapping.

A. Software

	BWA
Description	Burrows-Wheeler Alignment Tool
Version	0.7.5a
Source	http://bio-bwa.sourceforge.net/
Citation (2)	Li H. and Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler Transform. Bioinformatics,
	Samtools
Description	Toolbox for manipulation of SAM/BAM files
Version	1.2
Source	http://www.htslib.org/

B. Alignment Method

The alignment is done using the mapping software bwa-0.7.5a [3]. We use BWA with a maximum of 2 mismatches in the first 32 bases of the sequences, and a maximum of n mismatches in total. The table below summarizes the number of mismatches allowed according to the length of the reads:

Read length	Max num. of mismatch
17	2
38	3
64	4
93	5
124	6
157	7
190	8
225	9

Reads mapping to several positions on the reference sequence with the same mapping quality are attributed at random to one of the positions with a mapping quality of 0. When an input read has N's in their nucleotide sequence, BWA replaces the Ns by a random nucleotide.

In the case of pair-end sequencing, both reads of each pair should align on the genome with a position delta equal to the insert size, and they should map on opposite strands. As a consequence, if one of the reads is mapping on the genome but not the other, BWA tries to 'force' the alignment in the same area of the genome using a mapping method more tolerant to insertion and deletion.

For reads smaller than 70 bases, the mapping is performed using the tool BWA-ALN. For reads equal or larger than 70 bases, the tool BWA-MEM is used. BWA-MEM allows to split the reads and align independently the read segments. In such cases, the longest mapped segment of a read is designed as “primary alignment”, while any other mapped segments of the read are considered as “secondary alignments”.

The statistics in the report are computed from the aligned segments.

C. Reference Sequences

This section provides details about the reference sequences used in the alignments.

The follow 11 genes (MHC genes) have been retrieved

from NCBI and have been used for the mapping.

Gene ID	Locus	Length
ID-3105	NC_000006.12:29942532-29945870	3’339
ID-3106	NC_000006.12:c31357179-31353875	3’305
ID-3107	NC_000006.12:c31272092-31268749	3’344
ID-3115	NC_000006.12:33075990-33089696	13’707
ID-3117	NC_000006.12:32637406-32654846	17’441
ID-3118	NC_000006.12:32741391-32747198	5’808
ID-3119	NC_000006.12:c32666657-32659467	7’191
ID-3122	NC_000006.12:32439887-32445046	5’160
ID-3123	NC_000006.12:c32589848-32578775	11’074
ID-3133	NC_000006.12:30489508-30494194	4’687
ID-3135	NC_000006.12:29826474-29831130	4’657

III. Results

A. Mapping

The result of the mapping of the reads on the reference sequences is summarized in the table(s) below.

Description of the columns:

a) 'read set': this column identifies the set of reads that have been mapped on the reference (the name is made of the identifier of the lane where the reads have been sequenced and the code of the library).

b) '#reads': number of reads in the library.

c) '#mapped reads': number of primary alignments (for a read with multiple segments aligned, only the larger one is taken into account).

d) '%mapped reads': '#mapped reads' / '#reads' × 100.

e) '%multiple mappings': proportion of mapped segments matching several positions on the reference sequences (more exactly, this value is the number of mapped segments having a mapping quality < 4 divided by the number of mapped segments).

f) '%mis': proportion of sequenced bases having a mismatch with the reference sequence.

In the case of pair-end sequencing, the table(s) also report(s):

a) '%proper': the proportion of properly mapped read-pairs, i.e., reads mapping onto the reference with its mate mapping within a distance of 500bp on the opposite strand.

b) '%single': the proportion of read-pairs where only one read of the pair is mapped onto the reference.

c) '%both: the proportion of read-pairs where both pairs are mapped onto the reference.

d) '#HQlink': the number of linking pairs having a mapping quality higher than 5, i.e., the number of pairs having both reads mapping to different reference sequences with a mapping quality higher than 5 to avoid the ones falling into repeat regions of the genome.

read set	#reads (R1 +R2)	#mapped reads (R1 + R2)	%mapped reads	#secondary mappings	%multiple mappings	%mis	%proper	%single	%both	#HQlink
L008_GZK-2	127'897'314	23'020	0.02%	501	39.93%	0.84%	0.00%	0.01%	0.00%	76

TABLE I: Mapping Results for References_11-genes Reads Mapped per Gene

Gene ID	Mapped reads	Comment
ID-3105	0	
ID-3106	0	
ID-3107	0	
ID-3115	180	
ID-3117	7’789	One strong peak of about 30 bp
ID-3118	35	
ID-3119	0	
ID-3122	40	
ID-3123	11’814	One strong peak of about 30 bp
ID-3133	24	
ID-3135	0	

We have added 2 pictures taken from IGV display to represent the mapping on the gene of interest for the two highest counts genes (Fig. 1, 2).

IV. Conclusion

Mapping allows us to envisage the presence of new MHC genes in Echinodermata genomes and, particularly sea star Asterias rubens genome.

Further studies are necessary to determine these MHC genes. But in the present time, it seems that HLADQA1 gene (from MHC Class II) exists in Asterias rubens genome (Asterids) when compared to other MHC genes from Table I as previously said in Results and shown in Fig. 1.

HLADRB1 gene (Class II) which is also described in the present data, had been already demonstrated in Crinoïds and Ophuirids [1].

It confirms the originality of our work. through the evolution of immune genes in the animal kingdom.

European Journal of Biology and Biotechnology
www.ejbio.org

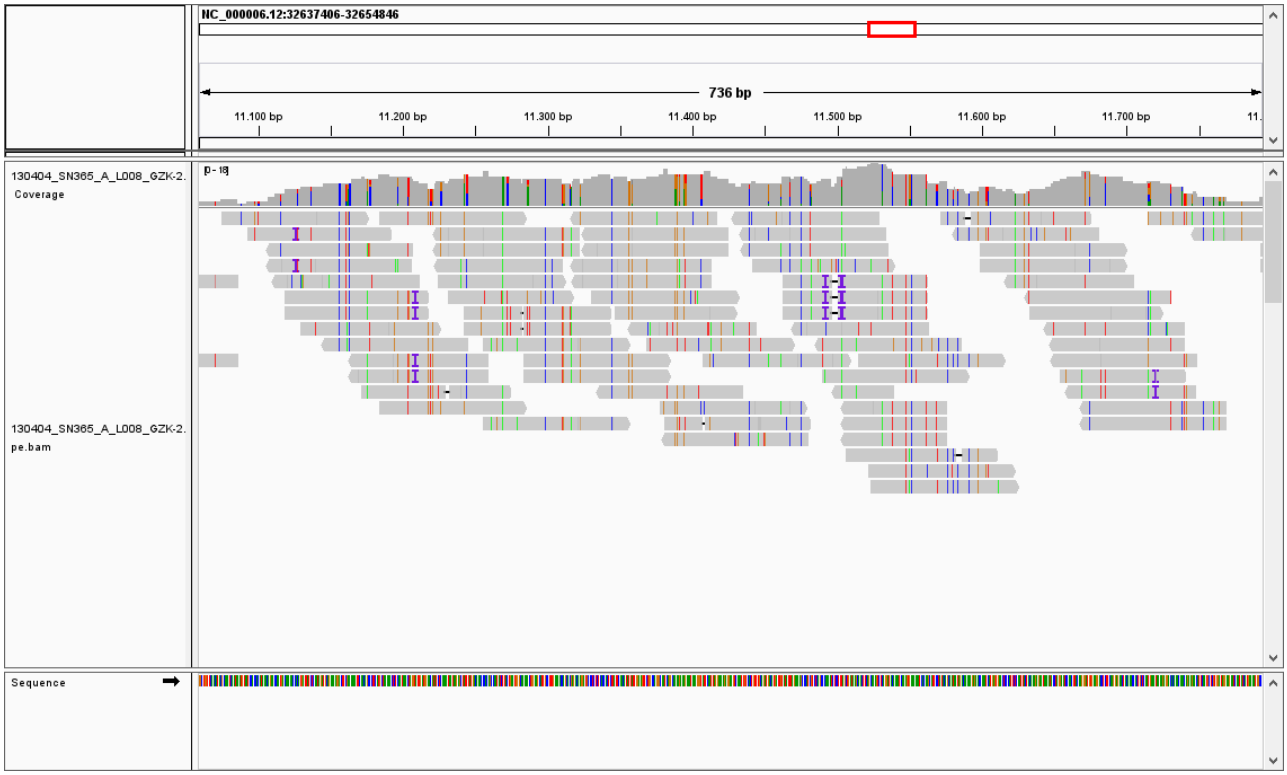

Fig. 1. HLADQA1 gene.

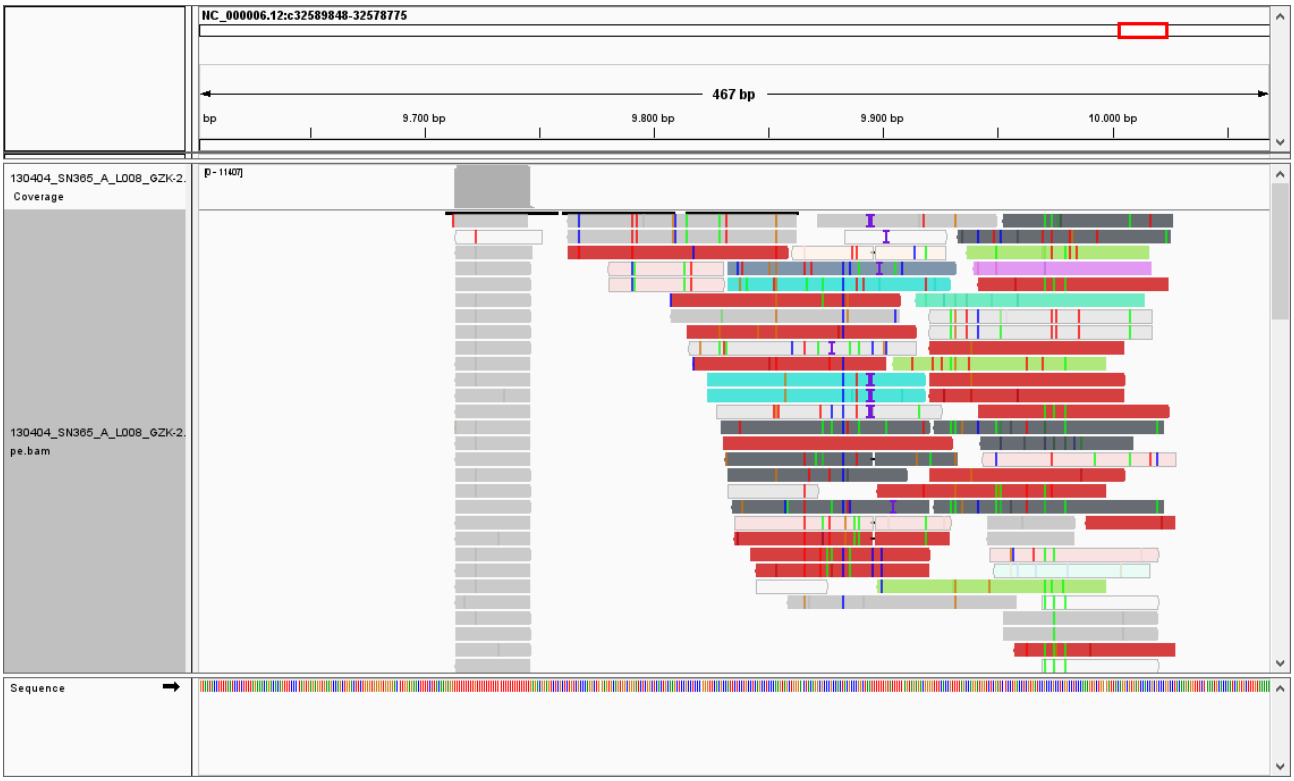

Fig. 2. HLADRB1 gene.

REFERENCES

[1] Leclerc M. (2020) Proteomics and Bioinformatics 2(1): 59-61.
[2] Li H. and Durbin R., (2010) Bioinformatics 26:589-95.
[3] Li H. and Durbin R. (2009) Transform 14:1754-60.