Complementary Analyses-2 (CA-2)

Imperfect backsplicing for exonic circRNAs

The existence of imperfect backsplicing was examined using the data generated from 117 bovine samples [Robic et al. 2024 [13].

The first example retained was the *MORC3* (ENSBTAG0000000011876) gene. Among the circRNAs detected in this region, we focused (Figure A) on seven circRNAs without annotation and one exonic circRNA (Bov_MORC3(5,7), Figure A1) that share their downstream boundary (on the donor side, here the end of exon7). These seven circRNAs are hemi-exonic circRNAs (HE-MORC3) highlighted by CircDetector and three of them was also detected by CIRI2. The true exonic circRNA, Bov_MORC3(5,7) was not detected in the 117 samples (117T), but was previously detected by CE2+CIRI2 in other bovine samples (Robic et al. 2021 [29]). No tissue specificity was observed for these eight circRNAs (Figure A1). They may all be present together. For example, all eight were observed in the sample of subcutaneous fat, thyroid, uterus and adrenal glands from the 9-month-old female. The two most common (Figure A1, HE-MORC3#261 and HE-MORC3#265) were detected in all 117 samples by CD. We found that the upstream boundary of the seven HE-MORC3 was located in a region of about 30 nucleotides around the start of exon5 (Figure A1). The unique exonic circRNA of this region, Bov_MORC3(5,7) (Figure A2), was the result of a backsplicing not supported by a canonical splicing signal. Indeed, we know that the motifs flanking canonical introns are recognized by the large spliceosome. These consensus motifs are 'gt' at the 5′-Start of the intron and 'ag' at the 3′-End*. The hemi-exonic circRNA, HE-MORC3#265, which was the most abundant circRNA from this region (Figure A1), may involve backsplicing supported by a canonical splicing signal (SS) (Figure A3). The SS acceptor (or 3'SS) used in expected backsplicing (between exon-7 and exon-5) to generate Bov_MORC3(5,7) does not seem to be the most "appropriate" to use for the backsplicing. Therefore, there would be a competition to find another SS acceptor to perform backsplicing with the best efficiency. Nevertheless, one further hemi-exonic circRNA (Figure A, HE-MORC3#272) with a canonical splicing signal was observed, but with very low expression. We know that the consensus 3'SS actually spans 4 bases, including 1 exonic*. This consensus 3'SS, cagG, is found for HE-MORC3#265, but for HE-MORC3#272, we found tagA, which was rarely described as a 3'SS*. We can hypothesize that there may be competition between the different existing SSs for selecting the backsplicing targets, as there is for splicing. The same kind of competition may take place for the selection of the branch point (Robic et al. 2022 [9]). When we looked again at the eight circRNAs from *MORC3* that share their downstream boundary (end of exon7), we noted that they all were detected in the 63 total-RNA-seq datasets, but not in any of the 63 mRNA-seq datasets (Figure A). It can be assumed that they do exist *in-vivo*, but are created by imperfect backsplicing. They are generated by backsplicing, but to not fit into the exonic circRNA pattern.

A second example of imperfect backsplicing is reported in Figure B and concerns the *LIFR* gene (ENSBTAG00000010423). In this case, the hemi-exonic circRNA HE-LIFR#466 is more lowly expressed than the exonic circRNA bov_circLIFR(2,4) (Figure B1). Only the second is provided by backsplicing and supported by canonical splicing signals (Figure B2). We can hypothesize that HE-LIFR#466 was the result of an imperfect backsplicing and supported by non-canonical SS (Figure B3). Observations on *MORC3* and *LIFR* circRNAs support the idea that backsplicing is at least as demanding in terms of canonical SS as conventional splicing. (Jeck et al. 2013 [6]; Starke et al. 2015 [23]; Vromman et al. 2023 [34]).

* https://science.umd.edu/labs/mount/RNAinfo/consensus.html

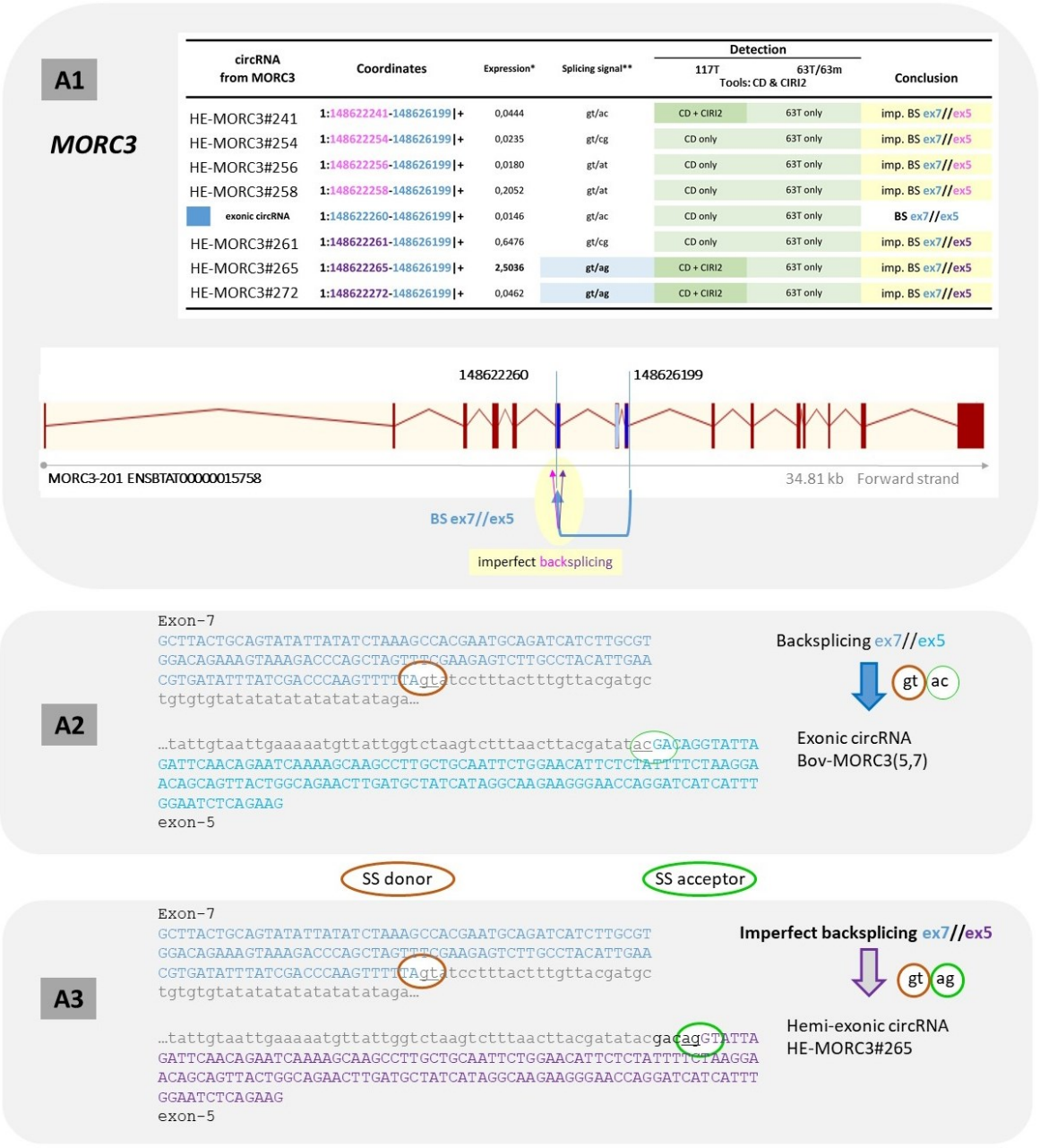

Figure A: Perfect and imperfect backsplicing in bovine *MORC3* region

(A1) In bovine *MORC3* region (ENSBTAG0000000011876; BTA1: 148,604-148,640 Mb), CircDetector (CD) identified 7 hemi-exonic circRNAs and 1 exonic circRNAs. Three of the seven hemi-exonic circRNAs are also detected by CIRI2. None of the eight circRNAs from *MORC3* are detected in 63m (mRNA-seq) by CircDetector. These 7 hemi-exonic circRNAs are probably generated by imperfect backsplicing (imp. BS). (A2) Details of the backsplicing (BS) leading to the exonic circRNA Bov_MORC3(5,7). (A3) Details of the imperfect backsplicing leading to the hemi-exonic HE-MORC3#265.

* Expression: average of the expression observed for 117 tissue samples
** Canonical motifs of a backsplicing exon7-exon5 are localized in the 5' boundary (gt) of the intron 7/8 and in the 3'boundary of the intron-4/5 (ag). See A2
SS splicing site

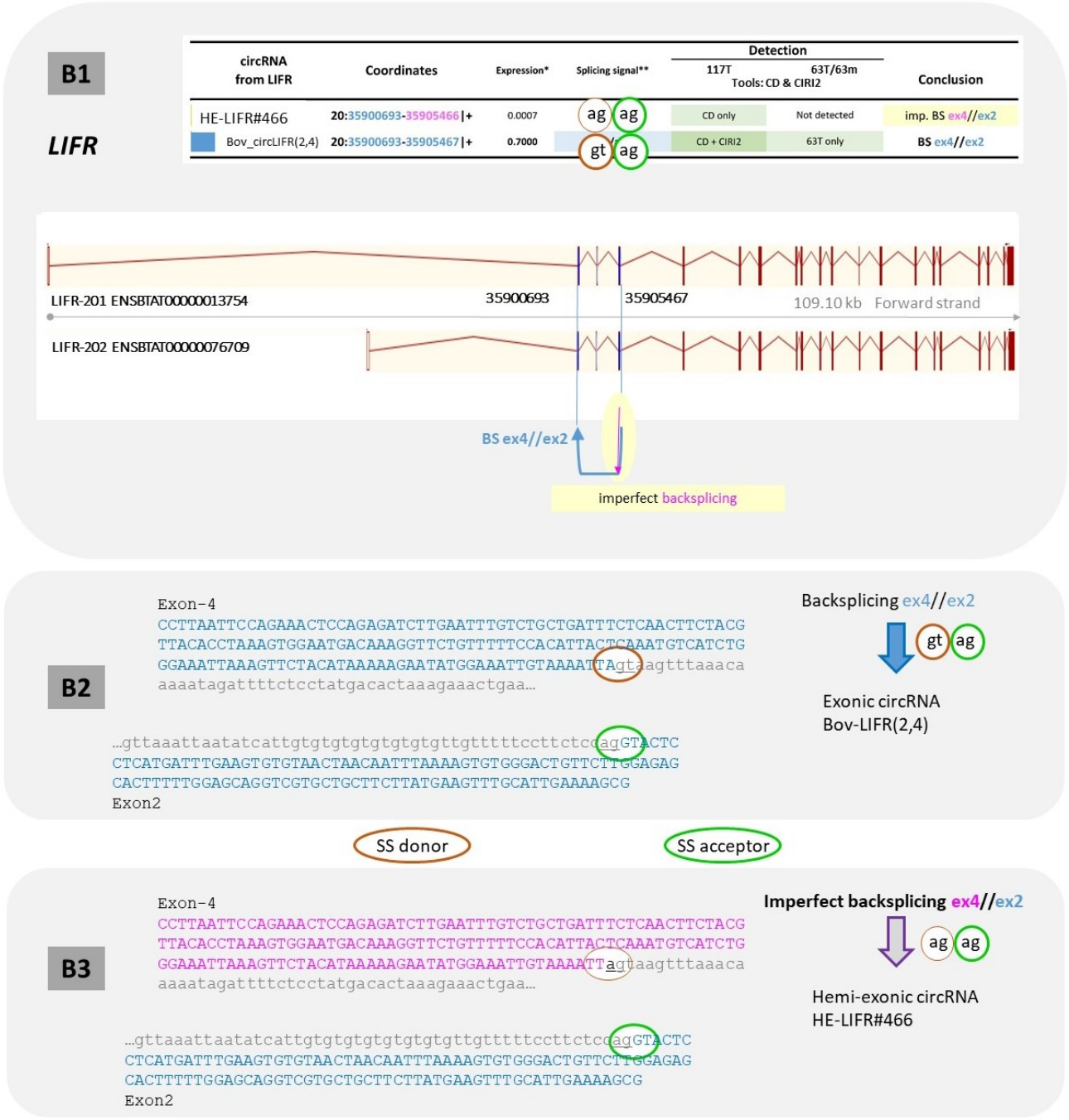

Figure B: Perfect and imperfect backsplicing in bovine *LIFR* regions.

In bovine *LIFR* region (ENSBTAG00000010423; BTA20: 35,840-35,950 Mb), CD identified 1 hemi-exonic circRNA and 1 exonic circRNA. Only the exonic circRNA is also detected by CIRI2. The exonic circRNA is generated by a backsplicing supported by canonical splicing signal (B2). The hemi-exonic circRNA is probably generated by imperfect backsplicing (B3).

BS: backsplicing imp. BS: imperfect backsplicing
* Expression: average of the expression observed for 117 tissue samples (117T)
** Canonical motifs of the backsplicing exon4-exon2 are localized in the 5' boundary (gt) of the intron 4/5 and in the 3'boundary of the intron-2/1 (ag). See B2
SS splicing site

Among the 36,215 circRNAs highlighted using CircDetector and without annotation, we detected 3,974 hemi-exonic circRNAs. Nevertheless, only 11.3% of them (450/3,974) have their second genomic coordinate in the region of the boundary of an exon (-30/+30 nt from the boundary). This statistic suggests that the generation of hemi-exonic circRNAs by imperfect backsplicing would be very rare.

The figures A and B were constructed from data and drawing available on Ensembl bovine genome database (http://www.ensembl.org/Bos_taurus/Info/Index) relative to the cow reference genome ARS-UCD1.2 and its latest annotation v-110 (April 2023). This was logical since these circRNA analyses we used the reference genome Bos_taurus.ARS-UCD1.2.
Since the end of 2023, a new bovine reference genome is available (ARS-UCD1.3).

- For *MORC3*, the circRNA coordinates determined on ARS-UCD1.2 cannot be transposed to ARS-UCD1.3. The structure of the exons concerned by this analysis remains unchanged. Unlike the *MORC3* exon-8 annotated on ARS-UCD1.3, the one annotated on ARS-UCD1.2 was similar to what can be found for human or porcine *MORC3*
- For *LIFR*, the circRNA coordinates determined on ARS-UCD1.2 cannot be transposed to ARS-UCD1.3. The first major difference in the *LIFR* exons involved in circRNAs between ARS-UCD1.3 and ARS-UCD1.2 is that exon-2 has become exon-1 and has lost 6 bases in 5'. It includes still the ATG (start of the translation). The second major difference is that two exon-1 previously reported (not containing ATG) have disappeared from the new annotation.
- The LIFR exon-ATG annotated on ARS-UCD1.3 has a similar length to what can be found for porcine *LIFR* (exon-ATG=exon-3). Nevertheless, the one annotated on ARS-UCD1.2 had a similar length to what can be found for human *LIFR* (exon-ATG=exon-2). We are not surprised by the uncertainty that may exist at the 5' end of an exon containing the ATG.

About CircDetector, see Robic et al. 2021 [29] and Robic et al. 2022 [9]