Endogenous avian leukemia viral loci in the Red Jungle Fowl genome assembly

Bernhard Benkel\(^1\) and Katherine Rutherford

Dalhousie University, Department of Plant and Animal Sciences, Box 550, Truro, Nova Scotia, Canada, B2N 5E3

**ABSTRACT** The current build (galGal4) of the genome of the ancestor of the modern chicken, the Red Jungle Fowl, contains a single endogenous avian leukemia viral element (ALVE) on chromosome 1 (designated RSV-LTR; family ERVK). The assembly shows the ALVE provirus juxtaposed with a member of a second family of avian endogenous retroviruses (designated GGERV20; family ERVL); however, the status of the 3′ end of the ALVE element as well as its flanking region remain unclear due to a gap in the reference genome sequence. In this study, we filled the gap in the assembly using a combination of long-range PCR (LR-PCR) and a short contig present in the unassembled portion of the reference genome database. Our results demonstrate that the ALVE element (ALVE-JFevB) is inserted into the putative envelope region of a GGERV20 element, roughly 1 kbp from its 3′ end, and that ALVE-JFevB is complete, and depending on its expression status, potentially capable of directing the production of virus. Moreover, the unassembled portion of the genome database contains junction fragments for a second, previously characterized endogenous proviral element, ALVE-6.

**Key words:** Red Jungle Fowl, reference genome, avian leukosis virus, endogenous retrovirus

2014 Poultry Science 93:2988–2990
http://dx.doi.org/10.3382/ps.2014-04309

© 2014 Poultry Science Association Inc.

INTRODUCTION

Seventeen distinct families of endogenous retroviral elements have thus far been identified in the chicken genome (Huda et al., 2008). Of these, the most extensively studied groups are the avian leukemia viral elements (ALVE; Crittenden, 1991, review) and the endogenous avian viruses (EAV; Dunwiddie et al., 1986; Sacco and Nair, 2014). To date, more than 2 dozen ALVE have been unambiguously identified based on their insertion sites within the chicken genome (Benkel, 1998; Chang et al., 2006; Smith and Benkel, 2008, 2009; Rutherford and Benkel, 2013). The reference genome sequence of the wild ancestor of the modern chicken, the Red Jungle Fowl, contains a single ALVE element (designated ALVE-JFevB; Levin et al., 1994), which has so far not been reported in modern chickens. However, the current assembly (galGal4) contains a gap, estimated at roughly 1 kbp, which begins within the putative 3′ end of the ALVE-JFevB element and extends into its flanking region. Characterization of the insertion site of ALVE-JFevB (i.e., the identification of the regions flanking the element) is complicated by the fact that it appears inserted into a member of a second family of endogenous viral elements (GGERV20; Huda et al., 2008) of which there are multiple copies in the galGal4 assembly. The objectives of this study were to close the gap in the assembly and to determine the structural integrity of the ALVE-JFevB locus.

MATERIALS AND METHODS

**PCR and Sequence Analysis**

To fill in the sequence gap spanning the 3′ end of ALVE-JFevB, we performed long-range PCR (LR-PCR) on genomic DNA from a heritage breed individual that also carried the ALVE-JFevB element. An LR-PCR reaction was performed that targeted the 3′ long terminal repeat (LTR) of the ALVE-JFevB element using primer LTRC in combination with a primer designed to anneal within the unique (nonrepetitive) sequence nearest to the LTR (RJF-Rev; Table 1 and Figure 1). The ends of the approximately 6-kbp LR-PCR amplicon were sequenced, and the sequences compared with the reference genome assembly [galGal4 on University of California–Santa Cruz (UCSC) genome browser]. A second LR-PCR reaction was performed using a combination of primers that targeted the 5′ end of ALVE-JFevB (RJF-For and LTRD) to confirm the completeness of the 5′ LTR in the ALVE-JFevB provirus (Table 1 and Figure 1). A third LR-PCR reaction, carried out on DNA from an ALVE element-free bird
Using the primers RJF-For and RJF-Rev, produced an amplicon of approximately 7 kbp, representing the unoccupied site for ALVE-JFevB. Finally, we screened the unassembled portion of the reference genome sequence database for evidence of other ALVE/genome junction fragments (i.e., any additional ALVE elements that may not have been correctly assembled within the reference genome sequence), using an approach which involved a BLAT search of the unassembled sequences (UCSC genome browser) using the ALVE-1 sequence (GenBank accession no. AY013303) as a query, followed by a manual search for contigs that contained both ALVE and genomic sequences.

The LR-PCR conditions were as follows: 1× LA Buffer with 1.5 mM MgCl₂ (TaKaRa Biolabs, Mountain View, CA), 200 μM dNTP (TaKaRa), 50 to 250 ng of genomic DNA, 8 pmol of each primer, and 1 U of TaKaRa LA Taq in a 15-μL reaction mixture. The PCR profile consisted of an initial denaturation at 95°C for 5 min, followed by 35 cycles of 95°C for 20 s, 56°C for 15 s, and 72°C for 10 min (unoccupied site and downstream flank) or 2 min (upstream flank), followed by a final extension at 72°C for 10 min. The PCR products were separated on 1% agarose gels. Amplicons were sequenced using the BigDye Terminator Cycle Sequencing kit v3.1 (Life Technologies, Carlsbad, CA) and products separated on an ABI 3130 Nucleic Acid Analyzer (Life Technologies).

**RESULTS AND DISCUSSION**

The sequence that we derived for the 3′ junction fragment of ALVE-JFevB, based on the LR-PCR amplicon (GenBank accession no. KJ908690), overlaps a short contig in the unassembled portion of the galGal4 database (accession no. AADN03012812). Together, these sequences fill the gap at the 3′ end of the ALVE-JFevB element in the current reference assembly. We used this information to construct a composite sequence beginning 328 bp upstream of the 3′ end of the GGERV20 element (position chr1:32,560,388 in galGal4) and ending 142 bp downstream of the 5′ end of the GGERV20 element (position chr1:32,573,987). This composite sequence revealed that the ALVE-JFevB endogenous provirus (forward orientation) is inserted toward the 3′ end of a member of the GGERV20 family (Huda et al., 2008) of endogenous retroviruses (reverse orientation), approximately 1,020 bp from the end of the 3′ LTR of the GGERV20 element (Figure 1) at position chr1:32,561,737 of the galGal4 assembly. The hexanucleotide motif GGCTTG is present both immediately upstream and downstream of ALVE-JFevB; this duplication is generated upon insertion of the provirus (Hishinuma et al., 1981). Alignment of the ALVE-JFevB sequence with that of an ALVE element for which complete sequence is available, ALVE-1 (accession no. AY013303), showed that the 2 proviruses are 99% identical. Moreover, ALVE-JFevB is complete and intact, and because it lacks the single nucleotide insertion that typifies the defective ALVE-1 (Johnson and Heneine, 2001), ALVE-JFevB is potentially capable of directing the production of viral particles, if expressed. Finally, 2 additional contigs containing junction fragments were found in the unassembled portion of the reference genome database. The 2 contigs (GenBank accession no. AADN03013023 and AADN03021772) appear to contain the upstream and downstream flanking regions, respectively, for the previously identified element ALVE-

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Location in galGal4</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTRD</td>
<td>CGCCCCATATGTCCTTGGCCT</td>
<td>chr1:32,560,321-343</td>
</tr>
<tr>
<td>RJF-For</td>
<td>GCTCCAGTGGTGGATATAGATGG</td>
<td>chr1:32,560,321-343</td>
</tr>
<tr>
<td>LTRC</td>
<td>TGTAGTCAAATAGAGCCAGAAT</td>
<td>chr1:32,574,348</td>
</tr>
<tr>
<td>RJF-Rev</td>
<td>CCATCACTTTCAAGAAGTACATTCA</td>
<td>chr1:32,574,348</td>
</tr>
</tbody>
</table>

**Figure 1.** Schematic overview of avian leukemia viral element ALVE-JFevB insertion site. Part A, occupied site: ALVE-JFevB (dashed boxes and lines) shown inserted into a member of the GGERV20 family of endogenous viral elements (solid boxes and double lines) roughly 1 kbp from the end of the element. Boxes represent the long-terminal repeat (LTR) of the endogenous elements and indicate the direction of insertion of the proviruses with respect to chromosome 1 (ALVE-JFevB in the forward orientation, and GGERV20 in the reverse orientation). Annealing locations of PCR primers used in the study are indicated with solid arrowheads. The oval shows the location of a CR1 element just downstream of the 5′ LTR of the GGERV20 locus. Part B, unoccupied site: chromosome lacking the ALVE-JFevB element. Landmarks are as above for the occupied site.
6 (Rovigatti and Astrin, 1983). This strongly suggests that the individual used for construction of the reference genome sequence carried 2 elements, ALVE-6 and ALVE-JFevB, but that only ALVE-JFevB is represented in the current assembled draft of the reference genome sequence.

REFERENCES


