

Trabajos presentados al *I Congreso Internacional de Biotecnología e innovación (ICBi)*, 9 - 12 de julio de 2018, Universidad Nacional Agraria La Molina, Lima, Perú.
Editoras:
Ilanit Samolski Klein
María Lucila Hernández-Macedo
Gretty Katherina Villena Chávez

Autores

F. Abel Ponce de León* 1

apl@umn.edu

<https://orcid.org/0000-0001-8645-553X>

Yue Guo 2

guoxx390@umn.edu

<https://orcid.org/0000-0001-7363-319X>

Brian Crooker 1

Crook001@umn.edu

<https://orcid.org/0000-0002-8433-9946>

Correspondencia

*Corresponding author

1 Department of Animal Science, 305 Haecker Hall, College of Food, Agricultural and Natural Resource Sciences, University of Minnesota, St. Paul, MN 55108, USA.

2 Department of Food Science & Nutrition, 205 Food Science and Nutrition, College of Food, Agricultural and Natural Resource Sciences, University of Minnesota, St. Paul, MN 55108, USA.

Citación

Ponce de León FA, Guo Y, Crooker B. 2020. Secuenciación de moléculas individuales de ácidos nucleicos en tiempo real (SMRT) para caracterizar transcriptomas e isoformas de ARNm. *I Congreso Internacional de Biotecnología e innovación (ICBi)*, Revista peruana de biología número especial 27(1): 091- 094 (Marzo 2020). doi: <http://dx.doi.org/10.15381/rpb.v27i1.17585>

SECTION III: INDUSTRIAL AND ENVIRONMENTAL BIOTECHNOLOGY ARTICLE

Secuenciación de moléculas individuales de ácidos nucleicos en tiempo real (SMRT) para caracterizar transcriptomas e isoformas de ARNm

Single molecule sequencing of nucleic acids in real time (SMRT) to characterize transcriptomes and mRNA isoforms

Abstract

Our efforts are oriented to assess bovine Y-chromosome gene expression patterns. One set of genes that are of interest are the so-called X-degenerate Y-chromosome genes that are located in the male-specific region of the Y-chromosome (MSY). This region contains 95% of the DNA of the Y chromosome. These genes are single copy and have an X-chromosome homolog. Both, the Y-encoded and X-encoded homologs have ubiquitous expression profiles. However, some genes, like *SRY* that regulates male sex determination, have functions that are more specific. Identifying DNA sequence differences between these homologs will allow evaluation of their spatial and temporal expression patterns. Identification of the Y-encoded mRNAs and their isoforms will allow our understanding of tissue specific expression of isoforms in male tissues. The latter will facilitate our evaluation of gene function in male sex differentiation and fertility. Hence, we hypothesized that each of these X-degenerate gene homologs generate isoforms and that differential expression patterns exist between sexes and across tissues. To investigate the latter we used a new generation sequencing (NGS) technology that generates long sequencing reads with a range between 1000 to 10,000 base pairs in length. Single molecule real time (SMRT) isoform sequencing (IsoSeq) of several tissues (liver, lung, adipose, muscle, hypothalamus and testis) was carried out. Transcript sequences were used for bioinformatics analysis and isoform characterization. Given the focus of this manuscript the SMRT technology we are only presenting results obtained with the analysis of the *BUTY* and *BUTX* genes.

Resumen

Nuestros esfuerzos están orientados a evaluar patrones de expresión génica del cromosoma Y bovino. Los genes de interés son los denominados genes X-degenerados que se encuentran en la región específica masculina del cromosoma Y (MSY). Esta región contiene el 95% del ADN del cromosoma Y. Estos genes son de copia única y tienen un homólogo en el cromosoma X. Ambos homólogos tienen perfiles amplios de expresión. Sin embargo, algunos genes, como el *SRY* que regula la determinación del sexo masculino, tienen funciones más específicas. La identificación de las diferencias de secuencia de ADN entre estos homólogos permitirá evaluar sus patrones de expresión espacial y temporal. La identificación de los ARNm codificados en el cromosoma Y y de sus isoformas permitirán analizar la expresión específica de sus isoformas en tejidos masculinos. Esto último facilitará nuestra evaluación de función génica en la diferenciación sexual masculina y la fertilidad. Por lo tanto, planteamos la hipótesis de que cada uno de estos genes homólogos degenerados del X genera isoformas y que existen patrones de expresión diferencial entre sexos y tejidos. Para investigar esto último, utilizamos una tecnología de secuenciación de nueva generación (NGS) que genera lecturas de secuenciación largas con un rango de longitud de 1000 a 10,000 pares de bases. Se secuenciaron los transcriptomas en varios tejidos (hígado, pulmón, adiposo, muscular, hipotálamo y testículo). Se utilizaron las secuencias generadas para el análisis bioinformático y la caracterización de isoformas. Siendo el foco de este manuscrito la tecnología SMRT, solo presentamos los resultados obtenidos con el análisis de los genes *BUTY* y *BUTX*.

Palabras clave:

cromosoma Y; genes degenerados del X; secuenciación de nueva generación; secuenciación SMRT.

Keywords:

Y-Chromosome; X-degenerate genes; Next generation sequencing; SMRT sequencing.

Introduction

X- and Y-chromosomes have two regions, the pseudoautosomal region (PAR) which is the recombining region and, the X-specific and Y-specific (MSY) regions that do not paired and therefore do not recombine during meiosis (Graves 2006). Absence of recombination at meiosis characterizes the MSY region that is poorly conserved. The MSY contains X-transposed genes (99% sequence similarity with X), X-degenerate genes (60 to 96% similarity with the X), ampliconic genes and sequences, and centromere repetitive sequences (Skaletsky et al. 2003). X-degenerate Y-chromosome genes are single copy genes (*SRY*, *RPS4Y1*, *ZFY*, *TBL1Y*, *PRKY*, *USP9Y*, *DDX3Y*, *UTY*, *TMSB4Y*, *NLGN4Y*, *CYorf15A* and *15B*, *JARID1D*, *EIF1AY*, and *RPS4Y2*), have an X-chromosome homolog, are described as housekeeping genes and have broad expression profiles. However, expression profile studies of these genes have not been oriented to assess differences of expression of transcript isoforms. Similarly, evaluation of expression patterns of the Y-encoded vs. X-encoded copies of the genes in males and their effect on maleness and male fertility has not been done. The underlying reasons for the latter is the paucity of sequence information available for bovine Y-chromosome genes and the difficulties inherent in generating complete gene sequences with short next generation sequencing reads. It has only recently been possible to generate long complete single molecule reads with acceptable sequencing error levels that allow transcript sequence comparisons and detection of possible isoforms.

Objectives

- Characterize X-degenerate Y-chromosome genes by comparing Iso-seq data to the available bovine genomic BTAY and BTAX and other available vertebrate transcriptomic sequences.
- Identify specific sequence priming sites to distinguish Y-encoded from X- encoded expression of X-degenerate genes in male tissues.
- Characterize the expression of these genes in somatic and gonadal tissues.

Material and methods

Animals and tissue samples: 14 tissue samples per animal including liver, kidney, spleen, cerebellum, hypothalamus, pituitary, adrenal gland, heart, longissimus dorsi, semitendinosus, lymph node, spinal cord, lung, testis and ovary were collected from four (two male and two females) one-week-old sire half-sibs Holstein calves. After collection, tissues were immersed in liquid nitrogen and transferred to -80 °C storage.

Single molecule IsoSeq sequencing: The protocol as described by Minoche et al. (2015) was used with some modifications. RNA was extracted with Trizol reagent followed by ethanol precipitation. RNA preparation was dissolved in water, and treated with DNase I to remove contaminating DNA. Treated preparation were purified using MinElute columns (Qiagen). RNA was converted to cDNA using the Clontech SMRTer cDNA synthesis kit

that includes a polyA purification step. cDNA was amplified by 8 – 12 cycles of PCR using primers for the polyA-end and the switching oligonucleotide to enrich for full-length transcripts. Amplified cDNA was separated into size fractions using an ELF device (Sage), with fractions combined into four pools of 1 – 2, 2 – 3, 3 – 5, and >5 kb. Individual size fractions were re-amplified as before, and the amplification products were transformed into standard libraries for sequencing using the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences, CA) as directed. For the 3 – 5 and >5 kb size fractions, the finished libraries were size selected via ELF a second time, to remove PCR-generated non-full-length molecules that were observed to interfere with efficient production of full-length sequence in these size classes.

Four size classes of transcripts in library form for each tissue were sequenced on the RSII platform using P6/C4 chemistry (Pacific Biosciences). Sixteen sequencing cells, four per size fraction, were used to generate sequencing reads. Average polymerase read lengths were 12 – 21 kb per library. Data was analyzed with the SMRT v2.3 software that uses the RS_IsoSeq.1 module to predict consensus isoforms followed by the use of the Quiver module to polish the isoforms.

Reverse transcription and PCR amplifications:

Specific primer pairs (not shown) were designed for each Y-encoded and X-encoded gene. The Reverse Transcript Polymerase Chain Reaction (RT-PCR SuperScript III one-step RT-PCR System with Platinum Taq DNA polymerase, Invitrogen, CA) was used to generate the first strand cDNA according to manufacturer's protocol. Gene expression analysis followed the protocol for SuperScript III one-step RT-PCR System with Platinum Taq DNA polymerase (Invitrogen, CA). PCR was performed in 20 µL reaction mixtures with 60 ng of RNA, 10 µM of each primer, 10 µL of 2X Reaction Mix, 0.8 µL of SuperScript III RT/Platinum Taq Mix and water to a total volume to 20 µL. RT-PCR conditions were an initial 30 min at 55 °C for cDNA synthesis, followed by 2 min at 94 °C denaturation and 40 cycles of: 15 sec at 94 °C, 30 sec at proper annealing temperature, based on primer melting temperature, 1 min per kilobase (kb) at 68 °C, followed by 5 min at 68 °C and a final hold at 4 °C. Beta-actin primers were used for positive control reactions. Negative controls included no primer control (C), no template control (Cp), no enzyme control (Ce), or only reaction mix plus water control (C0). PCR products were separated in 1% agarose gels run at 100 volts for 1 hour with 1 Kb plus DNA ladder marker (Invitrogen, CA).

Bioinformatics: Polished isoforms were aligned to bosTau7 (contains Y-chromosome sequences) for male tissues and bosTau8 (does not contain Y-chromosome sequences) for female tissues to identify all possible gene transcript locations and generate sorted SAM files. All unmapped reads/transcripts were added back to the SAM file. For genes for which there are bovine sequences available a FASTA file of those sequences was generated to search for similar genes in the transcriptome files with the Module FASTA36. For genes for which there is no bovine sequence available an amino acid FASTA file of

orthologous (human, mouse, etc.) sequences were created to search the transcriptome files with the tFASTx36 module. Multiple sequence alignments were carried out with Clustal Omega software (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). These alignments along with data visualization using IGV or IGB browsers facilitated the identification of gene isoforms.

Results

Sequence comparisons allowed the development of specific primers pairs (not shown) targeting gene se-

quence sites that differ between the Y encoded and the X encoded copies of these genes. Specific amplification of the X-degenerate Y-chromosome and X encoded copies was possible for 40% of the genes. Figure 1 shows the amplification observed for UTY and UTX in three different tissues. Similarly, based on the sequence transcript information we have identified the existence of four isoforms, two that were already predicted and two that are new for UTY (Fig. 2). Similarly, three new isoforms and two already predicted isoforms were identified for UTX (Fig. 3).

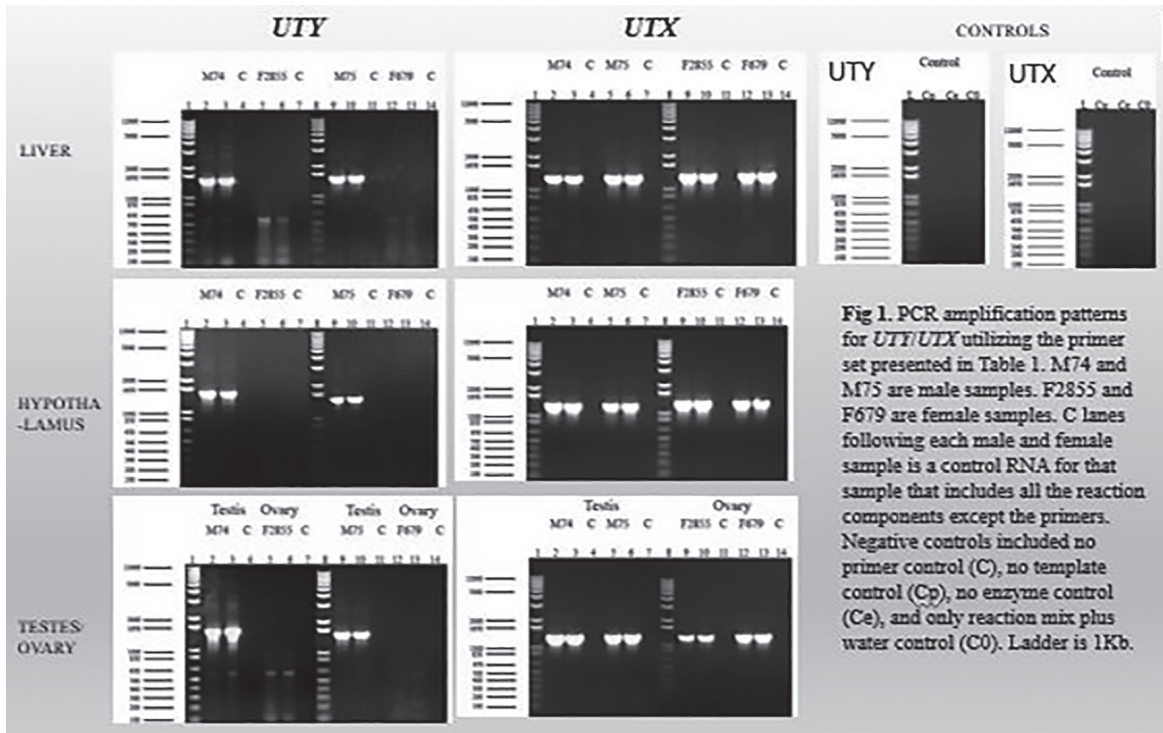
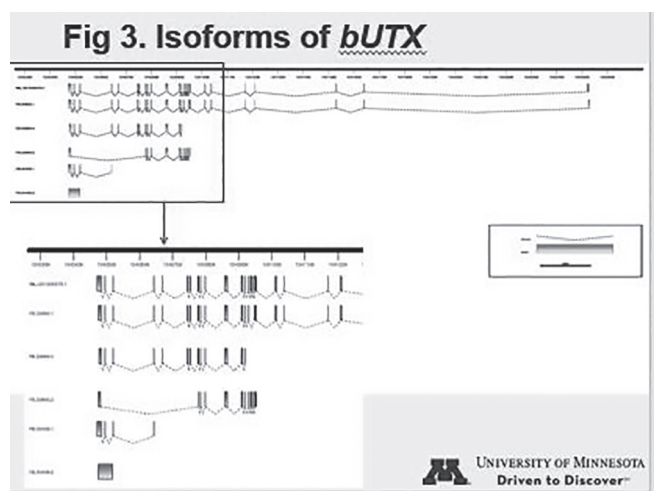
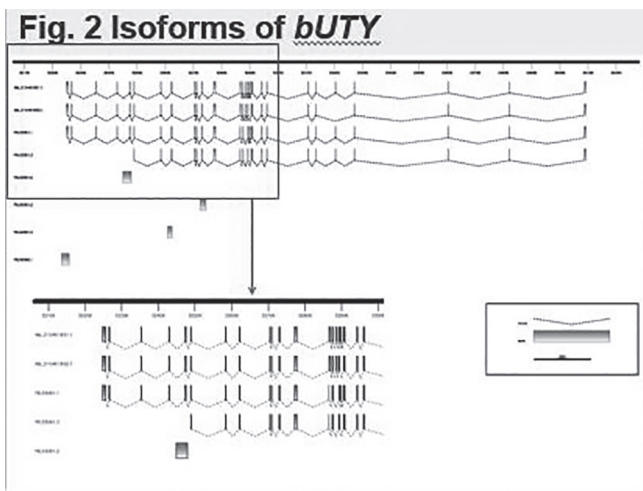


Fig 1. PCR amplification patterns for *UTY/UTX* utilizing the primer set presented in Table 1. M74 and M75 are male samples. F2855 and F679 are female samples. C lanes following each male and female sample is a control RNA for that sample that includes all the reaction components except the primers. Negative controls included no primer control (C), no template control (Cp), no enzyme control (Ce), and only reaction mix plus water control (CO). Ladder is 1Kb.



Conclusions

The use of single molecule sequencing in real time generated sequence reads between 1000 bp and greater than 7000 bp. The latter allowed the identification of mRNA isoforms for the genes under study and facilitated reliable sequence comparisons among them. For some of the X-degenerate Y-chromosome genes it has been possible to generate PCR primers that specifically amplify Y-encoded and X-encoded transcripts in male tissues. An assessment of differential amplification of those transcripts and their isoforms across somatic and gonadal tissues will be possible in the future, opening up the possibility to evaluate the effect of these genes in maleness and male fertility.

Literature cited

- Graves J A M. 2006. Sex chromosome specialization and degeneration in mammals. *Cell* 124: 901-14. <https://doi.org/10.1016/j.cell.2006.02.024>
- Minoche AE, JC Dohm, J Schneider, D Holtgräwe, P Viehöver, M Montfort, TR Sörensen, B Weisshaar, H Himmelbauer. 2015. Exploiting single-molecule transcript sequencing for eukaryotic gene prediction. *Genome Biology* 16:184-197. <https://doi.org/10.1186/s13059-015-0729-7>
- Skaletsky H., Kuroda-Kawaguchi T., Minx P. J., Cordum H. S., Hillier L., Brown L. G., 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423, 825-837, <https://doi.org/10.1038/nature01722>.

Agradecimientos / Acknowledgments:

Los autores apreciamos la colaboración del Dr. Timothy Smith (USDA-ARS, Nebraska) por compartir la base de datos de los transcriptomas utilizados en este trabajo.

Competing interests

The authors have declared that no competing interests exist.

Rol of authors:

FAPL: Conceptualización, Investigación, Escritura-Preparación del borrador original. YG: Investigación, Escritura-Preparación del borrador original. BC: Investigación, Redacción-revisión y edición.

Funding:

Research was supported by Minnesota Agricultural Experiment Station (MAES) MIN-16-103 and University of Minnesota Grant-in-Aid # 23132 to FAPDL.

Ethics / legals:

La metodología para la obtención de muestras de tejidos y beneficio de los animales usados fue aprobada por el Institucional Animal Care and Use Committee (IACUC) de la Universidad de Minnesota, Protocolo ID: 1408-31779A.