A Large Expansion of the HSFY Gene Family in Cattle Shows Dispersion across Yq and Testis-Specific Expression

Christine K. Hamilton, Tamas Revay*, Robin Domander, Laura A. Favetta, W. Allan King*

Department of Biomedical Sciences, University of Guelph, Guelph, Ontario, Canada

Abstract

Heat shock transcription factor, Y-linked (HSFY) is a member of the heat shock transcriptional factor (HSF) family that is found in multiple copies on the Y chromosome and conserved in a number of species. Its function still remains unknown but in humans it is thought to play a role in spermatogenesis. Through real time polymerase chain reaction (PCR) analyses we determined that the HSFY family is largely expanded in cattle (~70 copies) compared with human (2 functional copies, 4 HSFY-similar copies). Unexpectedly, we found that it does not vary among individual bulls as a copy number variant (CNV).

Using fluorescence in situ hybridization (FISH) we found that the copies are dispersed along the long arm of the Y chromosome (Yq). HSFY expression in cattle appears restricted to the testis and its mRNA correlates positively with mRNA markers of spermatogonial and spermatocyte cells (UCHL1 and TRPC2, respectively) which suggests that HSFY is expressed (at least in part) in early germ cells.

Citation: Hamilton CK, Revay T, Domander R, Favetta LA, King WA (2011) A Large Expansion of the HSFY Gene Family in Cattle Shows Dispersion across Yq and Testis-Specific Expression. PLoS ONE 6(3): e17790. doi:10.1371/journal.pone.0017790

Editor: Laszlo Orban, Temasek Life Sciences Laboratory, Singapore

Received October 30, 2010; Accepted February 10, 2011; Published March 7, 2011

Copyright: © 2011 Hamilton et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the National Sciences Engineering and Research Council, Canadian Research Chair program and L’Alliance Boviteq Incorporated. CKH was funded with a fellowship from the National Sciences Engineering and Research Council. TR was supported with a fellowship under the OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agricultural Systems. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: waking@uoguelph.ca

Current address: Research Institute for Animal Breeding and Nutrition, Herceghalom, Hungary

Introduction

The Y chromosome was originally thought to be devoid of functional genes because of its small size and heterochromatic nature [1]. This theory has since been disproven and now about 78 protein-encoding genes have been assigned to this chromosome in humans, most of which are involved in male growth, development and spermatogenesis [2,3]. The Y chromosome is unique in that the majority of its length does not pair with the X chromosome during meiosis to undergo homologous recombination [2]. This region is known as the male specific region or MSY [2]. The MSY is enriched with multi-copied genes and copy number variants (CNVs) [2,4]. CNVs are DNA segments of at least 1 kb in size that can vary in copy number among individuals through deletions and duplications and in many cases this variation has been linked to gene expression and phenotype [5–8]. Although there is still a lack of sequence data for the MSY, it has been fully sequenced in both human and chimpanzees and remains speculative and no attempt has been made to characterize the number of copies [9]. This diversity manifests itself in gene structure, content and number of gene copies.

An example of a multi-copied MSY gene is HSFY. It is a member of the heat shock transcriptional factor (HSF) family which are able to bind heat shock elements on heat shock protein (HSP) genes to regulate gene expression [10]. HSPs are important mediators of stress and act as molecular chaperones to regulate cellular homeostasis and promote cell survival in conditions that would otherwise be fatal [11,12].

HSFY contains a heat shock factor type A DNA-binding domain that is similar to that found in other HSF genes, including the X-homologue (LW-1) [13–15]. Its three-dimensional conformation, however, is altered so it is unknown if HSFY can act as a transcriptional regulator [13,14]. Its expression is reported to be mainly testis-specific in humans [15]. More specifically, HSFY expression seems restricted to Sertoli and spermatogenic cells [14]. It is likely that HSFY is involved in spermatogenesis but its exact function remains unknown [14–17].

The copy number of HSFY appears to vary between species. It has been measured in humans and felines and is present in 2 and about 8 copies, respectively [2,18]. HSFY orthologs have been found in a variety of other species including mouse, rat, rhesus macaque, and dogs and appears to be conserved, however, the gene copy number in these species has not yet been characterized [16,19]. Cattle have an HSFY ortholog (HSFY2) and it has been hypothesized that it is present in multiple copies but this theory remains speculative and no attempt has been made to characterize the number of copies [20].

The purpose of this study was to determine the exact genomic copy number of the HSFY family in different individuals, determine its chromosomal location and to determine its expression pattern in Canadian Holstein cattle. We found that bulls contain around 70 copies of the HSFY gene which are
dispersed along the long arm of the Y chromosome and we provide evidence that HSFY expression is testis-specific.

**Methods**

**HSFY gene analysis and sequencing**

A comprehensive search of the sequence database on the NCBI website was carried out in order to find and compare HSFY orthologs among different species. Structural similarities between deduced amino acid sequences among human HSFY sequences (hHSFY1: NP_149099.2; hHSFY2: NP_714927.1), as well as mouse (mHSFY1: NP_081937.1), rat (NP_001012132.1), rhesus macaque (ACL51668.1), cat (NP_001035212.1), and bovine (NP_001070474.1) were determined by multiple sequence alignments carried out using CLUSTAL W software [21]. The current bovine HSFY sequence (HSFY2: NW_001498001.1) is based on the Hereford breed of cattle and was assembled as part of the bovine genome project (Assembly Btau_4.2) [22]. We deduced the HSFY gene sequence in the Holstein breed by sequencing the PCR products that were generated throughout the study (as described below) and the resultant sequence was deposited into GenBank with accession number JF281100. Breed specific differences were analyzed by comparing the predicted Holstein amino acid sequence with the current predicted HSFY amino acid sequence derived from the Hereford breed using CLUSTAL W software.

**Sample collection and DNA/cDNA preparation**

A variety of tissues (blood, heart, kidney, liver, lung, ovary, testis) were obtained from a bank of tissues (L’Alliance Boviteq Inc., St Hyacinthe, Quebec, Canada) collected from a slaughtered Holstein heifer and from 24 slaughtered Holstein bulls. DNA was extracted using methods previously described [23]. Briefly, DNA was extracted from blood samples using standard phenol-chloroform methods. Total mRNA was extracted from the remaining tissues using a RNeasy Mini kit (QIAGEN Inc.) and treated with DNAse I (TURBO DNA-free, Ambion Inc.) following Table 1.

**Table 1. Primer sequences used for real time PCR experiments and FISH probe preparation.**

| Primer name | Primer sequences (5’→3’) | Size* (bp) | GenBank accession number | Gene name | Gene symbol | Annealing temp (°C) |
|-------------|--------------------------|------------|--------------------------|-----------|-------------|--------------------|
| GAPDH      | TTCCTGTTAAGCAATGATTGG    | 153        | NM_001034034.1           | glyceraldehyde-3-phosphate dehydrogenase | GAPDH | 66 |
| GAPDH      | GAGATGGGGCGAGCTC         |            |                          |           |             |                    |
| HSFY8F     | CAGAACAACCTGGATAGCTTC    | 111        | NW_001498001.1           | heat shock transcription factor, Y linked | HSFY2  | 67 |
| HSFY8R     | TCACAAGATCTCAAGACAAGC    |            |                          |           |             |                    |
| HSFY10F    | TGCTTTCCAGCTTTGTCTG      | 113        | NW_001498001.1           | heat shock transcription factor, Y linked | HSFY2  | 67 |
| HSFY10R    | CCAAGTTTTTGAGAATGTGA     |            |                          |           |             |                    |
| HSFY16F    | ATCTCGACTGTAGGTGGAG      | 102        | NW_001498001.1           | heat shock transcription factor, Y linked | HSFY2  | 67 |
| HSFY16R    | ACCAGTTTGCTCTCCTG        |            |                          |           |             |                    |
| HSFY8NAF   | GCTTCACTACTGTCTTCTGG     | 215        | NM_001077006.1           | heat shock transcription factor, Y linked | HSFY2  | 66 |
| HSFY8NAR   | TACTTCTGGTGACACTTG       |            |                          |           |             |                    |
| HSFY-E1F   | ATTCAGATGGGCGCTCCTAAG    | 1686       | NW_001498001.1           | heat shock transcription factor, Y linked | HSFY2  | 65 |
| HSFY-E2R   | GTTGACATCTCCCAAGTAACG    |            |                          |           |             |                    |
| HSFY-E1R   | GCCAATAAGTGAATATGTGGA    | 515        | NW_001498001.1           | heat shock transcription factor, Y linked | HSFY2  | 58 |
| HSFY-E2F   | AATCCAAAACTCAAGAGGTCA    | 752        | NW_001498001.1           | heat shock transcription factor, Y linked | HSFY2  | 58 |
| SRYF       | CCAATTAAGCCGGTCAAGCT     | 162        | NW_001505546.1           | sex determining region Y | SRY    | 67 |
| SRYR       | GACAGAAGGAAGCTCCGCTC     |            |                          |           |             |                    |
| TRPC2F     | AAGATGGAGGAGCAGTCTGAGC   | 180        | NM_174477.3              | transient receptor potential channel 2 | TRPC2  | 66 |
| TRPC2R     | TTTGCAACACGAGTGCCACAG    |            |                          |           |             |                    |
| UCHL1F     | ACCCGGAGTGCTGCAACAAAG    | 323        | NM_001046172.1           | ubiquitin carboxyl-terminal esterase L1 | UCHL1  | 66 |
| UCHL1R     | CCAAATGGTCCTGCTCAGA      |            |                          |           |             |                    |
| ZAR1F      | AAAGTGGCTATGTTGGTGTTG    | 191        | NC_007304.4              | zygote protein 1 | ZAR1 | 67 |
| ZAR1R      | CAGGTGATATCCTCCACTCG     |            |                          |           |             |                    |

*PCR product size in base pairs,

†primer used in combination with HSFY-E1F,

‡with HSFY-E2R.

The table shows the primer name, primer sequences, the predicted PCR product size (base pairs, bp), GenBank accession number of the sequence that was used as the target for primer design, gene name and symbol, and the unique annealing temperature (°C) that was used for the PCR amplification.

doi:10.1371/journal.pone.0017790.t001
manufacturers’ instructions. 1 μg of total mRNA was reverse transcribed with SuperScript II reverse transcriptase (Invitrogen Canada Inc.) using oligo(dT) primers (Invitrogen Canada Inc.) according to manufacturers’ instructions for all tissue samples. For each bull, the fertility status was obtained, as measured by 56-day non-return rates (NRRs). A 56-day NRR refers to the proportion of cows inseminated that were not reported to return to oestrus after 56 days and is used as an estimate of conception [24].

Quantitative PCR

The gene sequences used to design all primer sets were accessed from the sequence database on the NCBI website and their accession numbers are listed in Table 1. Primers were designed for these sequences using Primer3 software [25]. Three sets of primers (designated HSFY10F, HSFY10R, HSFY16F) targeting different regions of HSFY were used relative to the single copy reference gene, SRY, to measure HSFY copy number (Figure 1, Table 1). A separate set of primers targeting HSFY mRNA (designated HSFRNAF) was designed to span an intron to ensure cDNA specificity (Figure 1, Table 1). Primers targeting the mRNA specific markers for spermatogonia and spermatocytes, UCHL1 and TRPC2, respectively and the reference gene, GAPDH, were also used in the mRNA experiments [26,27]. To ensure male-specificity, both male and female DNA were run with the HSFY primers as well as an autosomal gene, ZAR1 (Table 1). Melting curve analyses, gel electrophoresis and PCR product sequencing (University of Guelph Laboratory Services, Guelph, ON, Canada) were used to verify amplification of the correct target genes.

1 ng of DNA and 50 ng of the reverse transcription product (cDNA) were analyzed by real time PCR as described previously [23]. Briefly, the real time PCR was performed using a LightCycler 1.5 apparatus (Roche Diagnostics) and a LightCycler Fast Start DNA Master SYBR Green 1 kit (Roche Diagnostics). The primer and MgCl2 concentrations were 0.5 μM and 3 mM, respectively. Samples were run in triplicate. The DNA samples were analyzed with two separate PCR runs that were averaged to minimize variability.

A calibrator sample was included in each run to minimize inter-run variability [28]. Primer efficiencies were measured according to the equation $E = 10^{-1/slope}$ [29]. Normalized ratios were determined for all runs using the $2^{-\Delta\Delta Ct}$ method (PE Applied Biosystems, Perkin Elmer, Forster City, CA) for DNA runs and the Pfaffl method of quantification for the cDNA runs [30]. HSFY copy number of the calibrator sample was determined with the $2^{-\Delta Ct}$ method using three sets of primers and averaged.

For statistical analyses all data first underwent the Kolmogorov-Smirnov test of normality. To determine differences between groups, we performed one way analysis of variance (ANOVA). All correlations were analyzed with a Pearson’s correlation coefficient.

Data is reported as the mean ± SEM unless otherwise stated. A

### Table 2. Comparison of the percent homologies (%) of the complete HSFY protein (A) and HSF-type DNA-binding domain (B), for all species analyzed.

|          | human (hHSFY2) | mouse (mHSFYL) | rat     | rhesus macaque | cat         | bovine |          |
|----------|----------------|----------------|---------|----------------|-------------|--------|---------|
|          | A   | B   | A   | B   | A   | B   | A   | B   | A   | B   |
| human (hHSFY1) | 100 | 100 | 43  | 71  | 44  | 71  | 81  | 87  | 64  | 78  | 51  | 72  |
| human (hHSFY2) | -   | -   | 43  | 71  | 44  | 71  | 81  | 87  | 64  | 78  | 51  | 72  |
| mouse (mHSFYL) | -   | -   | 43  | 71  | 95  | 100 | 43  | 69  | 47  | 80  | 40  | 76  |
| rat        | -   | -   | 43  | 69  | -   | -   | 46  | 80  | 39  | 76  |
| rhesus macaque | -   | -   | 58  | 73  | 43  | 69  | -   | -   | 58  | 81  |

In all cases, the percent homology was highest at the conserved HSF-type DNA-binding domain with the exception of an analysis of the two human HSFY copies which showed that both the complete protein sequence and conserved region were identical (100% homology). Bovine showed 51% and 72% homology with human HSFY for the complete protein and conserved domain, respectively. Bovine HSFY was most closely related to cat HSFY with a complete protein homology and conserved domain homology of 58% and 81%, respectively. The accession numbers for the proteins found in the NCBI database are as follows: human hHSFY1: NP_140909.2; hHSFY2: NP_714927.1, mouse NP_081937.1, rat NP_001012132.1, rhesus macaque ACLS1668.1, cat NP_001035212.1, bovine NP_001070474.1.

doi:10.1371/journal.pone.0017790.t002
p-value of less than 0.05 is considered significant. The data was analyzed using GraphPad Prism (version 5.02) software (GraphPad Software, Inc., San Diego, CA).

Fluorescence in situ hybridization (FISH)

Metaphase chromosome spreads were prepared from a bull’s peripheral blood sample cultivated in RPMI1640 (Invitrogen Canada Inc.) containing 10% FBS (Invitrogen Canada Inc.), 2% L-Glutamine (Sigma-Aldrich), 1% PenStrep (Invitrogen Canada Inc.), 1.5% Phytohemagglutinin (Invitrogen Canada Inc.) for 3 days. KaryoMax Colcemid (0.025 μg/ml, Invitrogen Canada Inc.) was added to the culture for the last 45 minutes. Slides were prepared from the hypotonized and MeOH/AcOH (3/1) fixed suspensions according to standard cytogenetic techniques and aged a minimum of 3 days before using for FISH. Three different biotinilated hybridization probes were prepared by PCR. The resulting PCR product were sequenced to confirm specificity of the primers. Primers were designed to anneal at the borders of both exons and were used in different combinations to amplify the whole HSFY gene (HSFY-E1F/HSFY-E2R) or exon1 (HSFY-E1F/HSFY-E1R) or exon 2 (HSFY-E2F/HSFY-E2R) selectively (Figure 1, Table 1). PCR mixes contained 80 ng bull genomic DNA, 2.5 Units AmpliTaq Gold DNA Polymerase (Applied Biosystems), 1 x PCR Buffer II, 1.5 mM MgCl2, 200 μM each dATP, dCTP, dGTP, 133 μM dTTP, 67 μM biotin-11-dUTP (Metkinen) and 0.5 μM of each primer. For the primer sets that amplified the whole HSFY gene (HSFY-E1F/HSFY-E2R), the PCR mixture underwent the following thermal profile: 95°C, 10 min; 30 cycles (94°C, 45 sec; 65°C, 45 sec; 72°C, 105 sec), 72°C, 10 min. For the other two primer sets which amplified exon

Figure 2. Sequence alignment of the conserved DNA-binding domain in various species. The conserved heat shock factor (HSF) type DNA binding-domain shows a high degree of sequence homology among different species. Amino acids conserved between at least 5 proteins are marked with shaded boxes. The accession numbers for the proteins deposited in the sequence database on the NCBI website are as follows: hHSFY1 NP_149009.2; hHSFY2 NP_714927.1; mHSFYL NP_081937.1; rat NP_00102132.1; rhesus macaque ACL51668.1; cow NP_001035212.1; bovine (Hereford) NP_001070474.1.

doi:10.1371/journal.pone.0017790.g002

Figure 3. Comparison of the predicted amino acid sequences of HSFY in Hereford and Holstein cattle. The mRNA sequences are 99% homologous and differ by only one base pair (c.1192A>C). This missense substitution mutation (p.Tyr266Ser) results in an amino acid change from tyrosine (Y) in Hereford to serine (S) in Holstein in the predicted protein sequence (shown by the box). This mutation is not found in the conserved HSF-type DNA-binding domain which is shown by the underlined sequence.

doi:10.1371/journal.pone.0017790.g003
Results

HSFY sequencing

A comprehensive search of the NCBI website revealed a bovine HSFY ortholog, HSFY2 (Entrez Gene ID: 767933). The genomic sequence of this gene was published as part of the bovine genome project (based on Btau_4.2; NW_001498001.1) and derived from the Hereford breed of cattle [22]. The complete mRNA coding sequence (NM_001077006.1) and predicted protein sequence (NP_0010770471.1) were also available for Hereford cattle in the database.

HSFY protein sequence alignments had been previously performed in human, mouse and cow, however, detailed percent homologies had not yet been described [16]. For this study, the alignments were expanded to include both copies of HSFY in human (hHSFY1 and hHSFY2) as well as the HSFY sequences of mouse, rat, rhesus macaque, cat and cow. Comparisons of sequence homologies for the complete HSFY protein as well as the conserved HSF-type DNA-binding domain for all species are presented in Table 2. Figure 2 shows the alignment results of the HSFY sequences of human, mouse, rat, rhesus macaque, cat and cow.

Table 3. HSFY copy number (n = 24), HSFY mRNA levels (n = 22) and percent 56-day non-return rate (NRR) for all bulls.

| Bull | HSFY copy number ± SEM | Relative HSFY mRNA ± SEM | 56-day non-return rate (NRR) (%) |
|------|------------------------|--------------------------|---------------------------------|
| 1    | 75.7 ± 8.5             | 0.851 ± 0.078            | 66.1                            |
| 2    | 73.2 ± 3.2             | 0.572 ± 0.046            | 66.9                            |
| 3    | 69.1 ± 9.0             | 0.490 ± 0.016            | 70.0                            |
| 4    | 77.5 ± 17.0            | 0.395 ± 0.007            | 49.6                            |
| 5    | 67.5 ± 5.6             | 0.283 ± 0.013            | 58.0                            |
| 6    | 67.5 ± 4.1             | 0.339 ± 0.010            | 72.6                            |
| 7    | 73.5 ± 1.5             | N/A                      | 74.5                            |
| 8    | 69.1 ± 4.3             | 0.295 ± 0.014            | 70.1                            |
| 9    | 80.8 ± 0.4             | 0.653 ± 0.060            | 58.8                            |
| 10   | 68.4 ± 11.1            | 0.568 ± 0.031            | 70.0                            |
| 11   | 75.9 ± 13.9            | 0.378 ± 0.014            | 57.3                            |
| 12   | 69.4 ± 1.9             | 0.285 ± 0.004            | 67.1                            |
| 13   | 74.8 ± 10.2            | 0.506 ± 0.006            | 69.0                            |
| 14   | 71.5 ± 16.5            | 0.308 ± 0.007            | 71.0                            |
| 15   | 74.3 ± 6.9             | 0.834 ± 0.028            | 66.1                            |
| 16   | 72.7 ± 0.3             | 0.373 ± 0.002            | 66.4                            |
| 17   | 76.5 ± 10.6            | 0.447 ± 0.001            | 65.0                            |
| 18   | 75.7 ± 7.9             | 0.762 ± 0.013            | 60.0                            |
| 19   | 73.1 ± 0.7             | 0.391 ± 0.010            | 57.4                            |
| 20   | 84.6 ± 21.3            | 0.278 ± 0.016            | 73.0                            |
| 21   | 73.0 ± 6.9             | 0.515 ± 0.006            | 74.8                            |
| 22   | 73.4 ± 3.8             | 0.706 ± 0.042            | 64.0                            |
| 23   | 72.0 ± 7.1             | 0.500 ± 0.027            | 57.0                            |
| 24   | 70.7 ± 5.7             | N/A                      | 77.3                            |
| Average | 73.3 ± 0.8             | 0.489 ± 0.038            | 65.9                            |

N/A = testis sample not available.

There were no significant differences in mean HSFY copy number among bulls and the average was 73.3 ± 0.8 copies. HSFY mRNA levels did vary significantly among bulls (p < 0.0001) with an average of 0.489 ± 0.038 relative to GAPDH. The 56-day non-return rates, a measure of field fertility, ranged from 49.6%–77.3%, with an overall average of 65.9%. doi:10.1371/journal.pone.0017790.t003

Figure 4. Gel electrophoresis of the real time PCR product of both male and female DNA samples for HSFY and ZAR1. HSFY shows male-specific amplification indicating that the target genes are on the Y chromosome. Zygote arrest 1 (ZAR1) is the autosomal control gene which shows amplification in both the male and female sample. N) Negative control M) 100 bp ladder – the top and bottom band represent 200 and 100 bp, respectively (Invitrogen Canada Inc.). doi:10.1371/journal.pone.0017790.g004

Figure 5. HSFY copy number in the calibrator sample measured by real time PCR with three different primers. There is no significant difference in mean HSFY copy number as measured by three different primer sets (HSFY8: HSFY8F/HSFY8R; HSFY10: HSFY10F/HSFY10R, and HSFY16: HSFY16F/HSFY16R) which target different regions of the gene. HSFY copy numbers for primer sets HSFY8, HSFY10 and HSFY16 were based on 20, 4 and 3 separate real time PCR runs, respectively. doi:10.1371/journal.pone.0017790.g005
conserved HSF-type DNA-binding domain among different species. Both copies of human *HSFY* (HSFY1 and HSFY2) were found to encode identical proteins (100% homology). Overall, cattle show 51% sequence homology with human HSFY and within the conserved DNA-binding domain the homology is increased to 72%. It showed its highest sequence identity with cat, with percent homologies of 58% and 81%, for the complete protein and HSF-type DNA-binding domain, respectively.

The complete *HSFY* gene sequence was deduced in Holstein and submitted to GenBank (accession number: JF281100). The Holstein *HSFY* genomic sequence was found to differ with the Hereford *HSFY* sequence by two nucleotides (c.56G>A, c.1192A>C). The published Hereford mRNA sequence, however, also shows the first discrepancy with the Hereford genomic sequence (c.56G>A) and therefore this one mismatch likely represents a genomic sequencing error. Our Holstein genomic sequence matches the Hereford mRNA sequence at this position. The second point mutation we found (c.1192A>C) is consistently different from both the Hereford genomic and mRNA sequences currently deposited on the NCBI database. This mutation results in a substitution from tyrosine (Hereford) to serine (Holstein) at position 266 of the predicted *HSFY* protein (p.Tyr266Ser). An alignment of the predicted amino acid sequences for both Hereford and Holstein cattle is shown in Figure 3.

**HSFY gene copy number**

Primer efficiencies were tested and were not found to be significantly different among targets (HSFY8, HSFY10, HSFY16) and the reference gene (SRY). We used the *HSFY* primers on both male and female DNA samples and as expected we found amplification in the male samples only whereas *ZAR1*, which is an autosomal gene, was present in both (Figure 4).

Three separate *HSFY* primers targeting different regions of the gene were used to measure *HSFY* copy number in the calibrator sample. There were no significant differences in *HSFY* copy number measured with the three different primers (Figure 5). The average *HSFY* copy number of the calibrator sample was 70.8 ± 0.9 copies. We selected the HSFY8 primers to analyze the full sample set (n = 24) and found that *HSFY* copy number did not vary significantly from bull to bull (Table 3). The average *HSFY* copy number amongst all bulls measured was 73.3 ± 0.8 and this value did not differ significantly with that of the calibrator. All data was normally distributed. The fertility status of the bulls ranged from an NRR of 49.6%–77.3% with an average of 65.9% but showed no significant correlation to *HSFY* copy number or mRNA levels (Table 3).

**Fluorescence in situ hybridization**

All three different FISH probes generated an intense, specific, yet dispersed signal on the Y chromosome. The signals were
almost chromosome painting probe-like and labeled the majority of the q-arm of the Y chromosome (Figure 6). There were no differences in the hybridization pattern between the probes.

**HSFY expression**

An analysis of a variety of different tissues (heart, kidney, liver, lung, ovary and testis) showed expression only in the testis for both HSFYRNA primers (spans intron 1) and HSFY8 primers (specific to exon 1) (Figure 7). GAPDH was present in all tissues.

For the bulls for which we had testis tissue samples (n = 22) we measured HSFY expression relative to GAPDH. We found significant bull-to-bull variation of HSFY mRNA levels (p < 0.0001). We also found significant correlations of HSFY mRNA levels to UCHL1 (r = 0.4352, p = 0.033) and TRPC2 (r = 0.6966, p = 0.0003) which are mRNA markers of spermatogonial and spermatocyte cells, respectively [26,27]. There was no correlation between HSFY copy number and HSFY mRNA levels (p = 0.3171).

**Discussion**

**HSFY copy number**

The sex chromosomes are thought to have originated from a pair of ancestral autosomes and although the X chromosome still remains highly homologous between species, the Y chromosome underwent a series of drastic inversions and species-dependent rearrangements that led to the current Y chromosome being highly heterogeneous in both size as well as in the genetic makeup among species [9,20,31–33]. Recently it has been shown that even closely related mammals such as humans and chimpanzees show enormous diversity in the MSY of their Y chromosomes for both structure and gene content [9].

One purpose of this study was to characterize the gene copy number of an MSY gene, HSFY, in bulls and determine whether or not it varies between individuals as a copy number variant (CNV). Since this study did not address how many of the copies are actually functional, we classify the copies found here under the common distinction of “HSFY family”. Since we found no amplification signal in female, we can be sure that the HSFY copies we are analyzing are specific to the Y chromosome and are not found on autosomes or the X chromosome (Figure 4).

This is the first study to estimate the number of HSFY family gene copies in cattle. We found that there are approximately 73 copies of the HSFY family genes on the Y chromosome. Our findings confirm the suggestion that HSFY is multi-copied in cattle [20]. The copy number is greatly expanded when compared to the expected copy number and this pattern is found among species. Not all genes, however, show expansion in cattle. There are some single copy genes present on the bulls’ Y chromosome such as SRY, ZFY and EIF1AY that are also found in single copy in humans [20,36,37].

We found that the genomic copy number of HSFY did not change significantly from bull to bull and therefore it does not appear to exist as a CNV. Deletions of HSFY copies in humans have been implicated in cases of male infertility [14,17,38,39]. The bulls in this study had differences in fertilities, as measured by non-return rates (NRRs), ranging from low fertility (NRR = 49.6%) to high fertility (NRR = 77.6%), but regardless of the bulls’ fertility status, their HSFY copy number stayed relatively constant. The difference may be that humans have only 2 copies of this gene and a deletion in one or both copies can have severe impacts on fertility. Cattle, on the other hand, have about 73 copies of HSFY which may allow for many “backups” in case one or more of the copies undergo mutations, thereby minimizing/eliminating harmful effects on fertility.

**Localization of HSFY by FISH**

Fluorescence in situ hybridization confirmed the results of the molecular copy number determination and localized it on the Y chromosome for the first time. The largest PCR produced probe, spanning the whole gene (1686 bp) generated an intense hybridization signal which covered almost the whole long arm of the Y chromosome (Figure 6). This suggests amplification of the targeted sequence and the presence of dispersed multiple copies. To test whether the gene specific sequences or an unknown highly repetitive element in the intron caused the observed signal distribution, we amplified the two exons specifically with PCR and used these as FISH probes as well. Both small probes (515 bp and 752 bp) generated the same hybridization pattern (Figure 6), thus we could exlude the contribution of intronic sequences.

**HSFY expression**

Notably, we found testis-specific expression of HSFY in cattle, as is seen predominantly for human HSFY [2,13]. Both human and feline studies have shown that many multi-copied Y chromosomal genes have acquired testis-specific expression and function so it is not surprising that we also found testis-specific expression of HSFY in cattle [2,19,40,41]. It is theorized that the evolution of multi-copied genes in the MSY allows for conservation of genes that enhance spermatogenesis as well as allow for a high level of expression of these genes through gene dosage effects [2,41]. Since we did not find evidence of a connection between HSFY mRNA levels and copy number, there is likely no dosage effects associated with this gene, but it is possible that its multi-copied nature may represent a mechanism of gene conservation. The mRNA levels of HSFY varied significantly between bulls (p < 0.0001) and correlated positively with mRNA markers of both spermatogonial and spermatocyte cells. This suggests that HSFY is being expressed in spermatogenic cells, as it is in humans, however further immunohistochemical analysis would be necessary to make firm conclusion about its distribution [14].

In conclusion, we found the HSFY family to be largely expanded as compared to other species studied to date and it is present in about 70 copies. While it does appear to be a multi-copy gene, it does not appear to be a copy number variant, because the gene copy number did not change within our Canadian Holstein bull sample set of over 20 bulls. FISH results showed that the copies are dispersed across the entire long arm of the Y chromosome. An analysis of mRNA showed its expression is testis-specific, like many other MSY genes, and its expression may be localized to spermatogenic cells as it is in humans.
Acknowledgments

The authors wish to thank Tom Kroetzch, Mathieu Boilard and Patrick Blondin for helping us obtain the bull samples. We would also like to thank Esther Semple, Ed Reyes and Liz St. John for their technical assistance.

References

1. Buhler EM (1980) A synopsis of the human Y chromosome. Hum Genet 55: 145–173.
2. Skalsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423: 825–837.
3. Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, et al. (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346: 240–244.
4. Jobling MA (2008) Copy number variation on the human Y chromosome. Cytoogenet Genome Res 123: 253–262.
5. Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. Nat Rev Genet 7: 85–97.
6. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) Global variation in copy number in the human genome. Nature 444: 444–454.
7. Ferrante A, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, et al. (2004) Detection of large-scale variation in the human genome. Nat Genet 36: 949–951.
8. Henn��chen CN, Chignard E, Reymond A (2009) Copy number variants, diseases and gene expression. Hum Mol Genet 18: R1–R8.
9. Hughes JJ, Humeaux F, Humeaux J, Humeaux M, Humeaux G, et al. (2010) The human Y chromosome. J Med Genet 47: 259–260.
10. Pirkkala L, Nykanen P, Sistonen L (2001) Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. FASEB J 15: 1118–1131.
11. Lindquist S (1986) The heat-shock response. Annu Rev Biochem 55: 1151–1191.
12. Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 381: 571–579.
13. Ferlin A, Mosso E, Rossi A, Dallapiccola B, Foresta C (2003) The human Y chromosome’s azoospermia factor b (AZFa/b) region: Sequence, structure, and deletion analysis in infertile men. J Med Genet 40: 18–24.
14. Shinka T, Sato Y, Chen G, Naroda T, Popa L, et al. (2004) Characterization of the human Y chromosome and implications for male infertility. Biol Reprod 71: 297–306.
15. Tesarik A, Salata E, Ferlin A, Bartolini L, Schioppo P, et al. (2004) Characterization of HSFY, a novel AZFa/b gene on the Y chromosome with a possible role in human spermatogenesis. Mol Hum Reprod 10: 233–238.
16. Kinosita K, Shiina T, Sato Y, Kuraishi H, Kowa H, et al. (2006) Expression analysis of a mouse orthologue of HSFY, a candidate for the azoospermic factor on the human Y chromosome. J Med Invest 53: 117–122.
17. Vinci G, Rauz F, Popa L, Popa O, Cocos R, et al. (2005) A deletion of a novel heat shock gene on the Y chromosome associated with azoospermia. Mol Hum Reprod 11: 293–298.
18. Pears Willermo A, Rauz F, Trinh V, Albrecht D, Warren W, et al. (2000) Gene discovery and comparative analysis of X-degenerate genes from the domestic cat Y chromosome. Genomics 59: 329–338.
19. Murphy WJ, Pears Willermo A, Rauz F, Agarawala R, Schaffer AA, et al. (2006) Novel gene acquisition on carnivore Y chromosomes. PLoS Genet 2: e53.
20. Liu W (2010) Comparative genomics of the Y chromosome and male fertility. In: Jiang Z, On ML, eds. Reproductive genomics in domestic animals. Amesoda, USA, Wiley-Blackwell. pp 129–156.
21. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
22. Bovine Genome Sequencing and Analysis Consortium, Elskig CG, Telliam RL, Wroley KC, Gibbs RA, et al. (2009) The genome sequence of taurine cattle: A window to ruminant biology and evolution. Science 324: 522–528.
23. Hamilton CK, Fawcett LA, Di Mes GP, Floriot S, Perucca A, et al. (2009) Copy number variation of testis-specific protein, Y-encoded (TSPY) in 14 different breeds of cattle (Bos taurus). Sex Dev 3: 205–213.
24. Koops WJ, Grossman M, Den Daas JH (1995) A model for reproductive efficiency of dairy bulls. J Dairy Sci 78: 921–929.
25. Rounen S, Skalsky H (2000) Primecr1 on the WWV for general users and for biologist programmers. Methods Mol Biol 132: 365–386.
26. Wissenbach U, Schroth G, Philipp P, Flocke W (1998) Structure and mRNA expression of a bovine trp homologue related to mammalian trp2 transcripts. FEBS Lett 429: 61–66.
27. Oatley JM, de Avila DM, Reeves JJ, McLean DJ (2004) Testis tissue explant culture supports survival and proliferation of bovine spermatogonial stem cells. Biol Reprod 70: 625–631.
28. Hellemans J, Moster G, De Parpe A, Spelman F, Vandesemelle E (2007) qBase relative quantitative framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 8: R19.
29. Ginzinger DG (2004) Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. Exp Hematol 30: 503–512.
30. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45.
31. Rauschfleisch F, Lee EJ, Kata SR, Brinkmeyer C, Michelson JR, et al. (2004) Exceptional conservation of horse-human gene order on X chromosome revealed by high-resolution radiation hybrid mapping. Proc Nail Acad Sci U S A 101: 2386–2391.
32. Graves JA (2006) Sex chromosome specialization and degeneration in mammals. Cell 124: 901–914.
33. O’Brien SJ, Menni-Raymond M, Murphy WJ, Nash WG, Wierzbicki J, et al. (1999) The promise of comparative genomics in mammals. Science 286: 450–481.
34. Jakubcza J, Schnieders F, Schmidtke J (1993) A bovine homologue of the human TSPY gene. Genomics 17: 732–735.
35. Nicklisch K, Bovine Genome Sequencing and Analysis Consortium, Tellam RL, et al. (2010) Y chromosome TSPY copy numbers and semen quality. Fertil Steril 94: 1744–1747.
36. Curry L, Pratt ML, Kelley DE, Lapin DR, Gibbons JR (2006) Use of a combined duplex PCR/Dot blot assay for more sensitive genetic characterization. Biochemistry Insights 1: 35–39.
37. Liu WS, de Leon FA (2004) Assignment of SRY, ANT3, and CSF2RA to the bovine Y chromosome by FISH and RH mapping. Anim Biotechnol 15: 103–109.
38. Li Z, Haines CJ, Han Y (2008) “Micro-deletions” of the human Y chromosome and their relationship with male infertility. J Genet Genomics 35: 193–199.
39. Repping S, Skaletsky H, Lange J, Sibler S, Van Der Veen F, et al. (2002) Recombination between palindromic P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am J Hum Genet 71: 906–922.
40. Lahn BT, Pearson NM, Jekajan K (2001) The human Y chromosome, in the light of evolution. Nat Rev Genet 2: 207–216.
41. Lahn BT, Page DC (1997) Functional coherence of the human Y chromosome. Science 278: 675–680.

Author Contributions

Conceived and designed the experiments: CH TR RD LA F WAK. Performed the experiments: CH TR RD. Analyzed the data: CH TR RD LF WAK. Contributed reagents/materials/analysis tools: WAK. Wrote the paper: CH TR RD LA F WAK.