

# Genome-wide association analysis of body conformation traits in Mexican Holstein cattle using a mix of sampled and imputed SNP genotypes

M.E. González<sup>1</sup>, V.M. González<sup>2</sup>, M.F. Montaña<sup>2</sup>, G.E. Medina<sup>2</sup>,  
P. Mahadevan<sup>3</sup>, C. Villa<sup>1</sup> and R. Villa<sup>1</sup>

<sup>1</sup>Laboratory of Bioinformatics and Biofotonics, Engineering Institute, Autonomous University of Baja California, Mexicali, Baja California, Mexico

<sup>2</sup>Veterinary Science Research Institute, Autonomous University of Baja California, Mexicali, Baja California, Mexico

<sup>3</sup>Department of Biology, University of Tampa, Tampa, FL, USA

Corresponding author: M.E. González  
E-mail: mirvana.gonzalez@uabc.edu.mx

Genet. Mol. Res. 16 (2): gmr16029597

Received January 10, 2017

Accepted March 8, 2017

Published April 13, 2017

DOI <http://dx.doi.org/10.4238/gmr16029597>

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

**ABSTRACT.** In arid and semi-arid zones, animal health and production are closely correlated with body conformation traits. These selected traits, in turn, allow livestock to adapt unfavorable soil and environmental conditions. The primary objective of this study was to perform a genome-wide association analysis for a set of sampled and imputed SNPs with 16 conformation traits in a population of Holstein cows from a desert area of Northwestern Mexico. Imputation from 6K to 50K SNPs was performed as a low-cost optimization strategy. Results show eight SNPs associated with two conformation traits. The Udder Depth trait resulted in seven associated SNPs from chromosome 10, that related to Marbling Score, Milk Yield, Fat Yield, Protein

Yield, and Protein Percentage Quantitative Trait Loci (QTLs). The Body Depth trait resulted in one associated SNP from chromosome 2, although no QTL relation was found. The discovery of genes associated with conformation traits may be indicative of the adaptive selection pressures the Holstein breed has undergone in response to the extreme weather conditions found in the northwestern areas of Mexico. Results of this study indicate that traits such as stature and body depth may be used as indicators of cows' potential genetic merits for milk, fat, and protein production.

**Key words:** SNP genotypes; Association studies; Conformation traits; Mexican cattle

## INTRODUCTION

Association of cattle genes to economic traits is helping to generate genetic basis for different diseases and production. In particular, conformation traits are being selected to adapt livestock to unfavorable soil and environmental conditions. Hence, they are as important as other economic traits (Cole et al., 2011; Wu et al., 2013). In arid and semi-arid zones, such as northwestern Mexico, animal health and production are closely correlated with several traits, including body height, depth, and size; croup angle and width; and leg conditions among others. These characteristics are used as informative conformation features that facilitate reproduction and longevity, and form important variables in the generation of selection indices (Avendaño-Reyes et al., 2006; Avendaño-Reyes et al., 2010).

In this paper, we present an association study utilizing 6,912 SNPs obtained using a low-density bovine SNPchip, and 39,611 imputed SNPs along the bovine genome. Association was performed for 16 traits related to economic importance in Holstein cattle from a desert area in Northwest Mexico. First, an Infinium the BovineLD v2.0 BeadChip (Illumina, Inc., San Diego, CA, USA) was used to extract genotypes from 47 Holstein cows and then imputation was performed using an international reference panel. After applying quality control filters, we performed association analysis using two different methods. The results show eight SNPs associated with two different conformation traits. In addition, an analysis of QTLs covering significant SNPs shows association with milk delivery, fat delivery, and protein percentage in milk and marbling.

## MATERIAL AND METHODS

### Declaration of ethics

The Animal Care and Use Committee of the Autonomous University of Baja California (Universidad Autónoma de Baja California-UABC), considered ethical approval unnecessary for this study. All blood samples used for DNA extraction were collected under the direction of the UABC Veterinary Science Research Institute according to the laws on animal studies (NOM-003-ZOO-1994 and NOM-062-ZOO-1999).

Blood samples were obtained from 47 female Holstein-Friesian cattle registered with the Mexican Holstein Association. The herd was kept 3.5 km away from the San Felipe-

Mexicali highway, 32° -24-27 N, 115° 23 03 W at the Veterinary Science Research Institute of the Autonomous University of Baja California, Mexicali, Baja California, Mexico. This area is at an altitude of 8 m above sea level, and has a hot, arid climate, with an average maximum temperature of 50 °C in summer, and an average minimum temperature of -5°C in winter (García, 2004). All cows were born following artificial insemination, were between 47 and 120 months of age, and were between first and fourth lactation. All cows were clinically healthy and free of brucellosis and tuberculosis. Sixteen traits of interest were sampled using measurement ranges established by the Holstein Association USA. These traits are indicated as follows: Stature, Chest, Body Depth, Angularity, Rump Angle, Rump Width, Rear Leg Set Rear View, Rear Legs Set Side View, Podal Angle, Fore Udder Attachment, Anterior Nipple Placement, Front Teat Length, Udder Depth, Rear Udder Height, Suspensory Ligament, and Posterior Nipple Placement. Table 1 shows the phenotype statistics considered in this study.

**Table 1.** Summary of phenotype statistics.

Phenotype	N	Min	Max	Mean	SD	95%CI
Stature (cm)	46	135	159	146.36	4.85	144.95-147.75
Chest (cm)	46	30	36	34.62	2.45	33.91-35.33
Body Depth (1 to 9)	46	6	9	7.53	1.51	7.09-7.97
Angularity (1 to 9)	46	3	9	7.6	1.64	7.12-8.07
Rump Angle (1 to 9)	46	1	9	3.93	2.31	3.26-4.60
Rump Width (cm)	46	26	30	28.44	1.28	28.047-28.81
Rear Leg Set Rear View (1 to 9)	46	1	8	5.84	2.11	5.23-6.45
Rear Legs Set Side View (1 to 9)	46	1	3	1.86	0.50	1.72-2.01
Podal Angle (1 to 9)	46	1	9	4.91	2.31	4.24-5.57
Fore Udder Attachment (1 to 9)	46	1	3	2.13	0.62	1.95-2.31
Anterior Nipple Placement (1 to 9)	46	1	8	4.48	1.53	4.04-4.93
Front Teat Length (1 to 9)	46	3	9	6.04	1.93	5.48-6.60
Udder Depth (1 to 9)	46	1	8	5.6	2.73	4.81-6.38
Rear Udder Height (1 to 9)	46	0	6	0.26	1.25	-0.09-0.62
Suspensory Ligament (1 to 9)	46	1	9	5.44	2.98	4.58-6.30
Posterior Nipple Placement (1 to 9)	46	1	9	6.42	2.84	5.59-7.24

N = observation number; Min = minimum; Max = maximum; SD = standard deviation; CI = confidence interval.

## DNA extraction

Blood samples were collected in April 2015 by venipuncture of the coccygeal vein using Vacutainer tubes with Hemogard closure (Necton, Dickinson and Company, Franklin Lakes, New Jersey, USA) with an EDTA anticoagulant. DNA extraction and purification were performed using a QIAamp DNA Blood Mini Kit QIAGEN, Hilden, Germany. All DNA samples were analyzed using spectroscopy and agarose gel electrophoresis, and were quantified with a NanoDrop kit (NanoDrop Products, Wilmington, Delaware, USA) to ensure a minimum of 50 ng/ $\mu$ L DNA concentration.

## Genotyping

Extraction and purification of genotypes was carried out at the National Institute of Genomic Medicine in Mexico City, using the BovineLD v2.0 BeadChip, which samples 6912 SNPs along the 30 mapped chromosomes of the UMD 3.1 bovine genome (Zimin et al., 2009).

## Imputation

Genotype imputation was performed using an algorithm implemented in findhap

(version 4) software, as described by VanRaden (VanRaden et al., 2010). This algorithm analyzes data by relying on the fact that unknown genotypes can be elucidated (imputed) based on genotypes observed in the same loci (or near loci) of individuals from the same family using pedigree haplotyping, or by comparing allele patterns (without using the pedigree) via population haplotyping. In general, the algorithm generates a list of haplotypes, based on genotype, and through an iterative process, refines those haplotypes, taking into account pedigree or allele patterns. Genotypes are compared to the refined haplotypes to determine imputed value. In this study, real genotypes of our population (6,912 SNPs from 47 animals) were obtained using the BovineLD v2.0 BeadChip. Then, using a Holstein cattle reference panel provided by the Bovine HapMap Consortium (2017), we imputed 39,611 SNPs, mapped with the BovineSNP50 v2 BeadChip, to obtain 46,523 SNPs along the 29 autosomal chromosomes of the bovine genome.

### Quality control filters

Quality control filters were applied to the compiled dataset. Filters included the elimination of all SNPs with a genotyping error >0.05%, individuals with a missing values percentage >2%, SNPs that did not satisfy the Hardy-Weinberg principle ( $P < 1.0E-3$ ), and SNPs with minor allele frequency <0.05%. Ultimately, 46 of 47 initial samples along with 22,151 SNPs on the 29 autosomal chromosomes were examined in this study.

### Association analysis

Association analysis was performed via least squares regression using PLINK 1.07 software (Purcell et al., 2007), and via simple and multiple linear regression using R version 3.2.3 software (R Development Core Team, 2011). In general, regression models are mathematical methods used for modeling quantitative stochastic relationships between variables of interest and a set of explanatory variables (Rosner, 2011). These models can be expressed as follows:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i$$

where  $Y_i$  refers to the variable of interest, either dependent or regressor;  $X_{1i}, X_{2i}, \dots, X_{pi}$  are explanatory variables, independent or regressors;  $b_0$  is the intersection or constant term;  $b_1, b_2, \dots, b_p$  are parameters, that measure explanatory variables influence on regressors;  $p$  refers to the number of independent parameters considered;

$\varepsilon$  is the observation error that results from the absence of controlled variables; and  $i$  ( $1, 2, \dots, n$ ) is the observation number of the variables.

### Identification of SNP localization and gene annotation

Positions of significant SNPs were reported using the bovine UMD3.1 genome sequence, produced by the Center for Bioinformatics and Computational Biology at the University of Maryland, USA. Genes closest to the significant SNPs (within a distance of 1Mb) were determined using resources and tools provided by the National Animal Genome

Research Program (<http://www.animalgenome.org>) and the National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov>).

## RESULTS

To verify consistency of the results, association analysis was performed using two different algorithms. First, least squares regression was performed using PLINK software. Execution consisted of using the command line mode with the following arguments: *plink -bfile inputfileName -assoc -pheno qt.phe -out outputfileName*, where *inputfileName* is the file containing genotypes and pedigree in linkage format; *assoc* instructs the algorithm to perform an association with quadratic regression; *qt.phe* is a file containing vector phenotypes; and *outputfileName* is the output file. Next, association analysis was performed using Generalized Linear models, in R software. For the latter, a two-step procedure was used: In step one, a simple association was executed for each SNP; those with a moderated P value ( $\leq 0.1$ ) were selected. In the second step, a multiple association was executed using the previously selected SNPs, and a Bonferroni multiple-testing correction (Nakagawa, 2004) was applied to the P values before declaring associated SNPs.

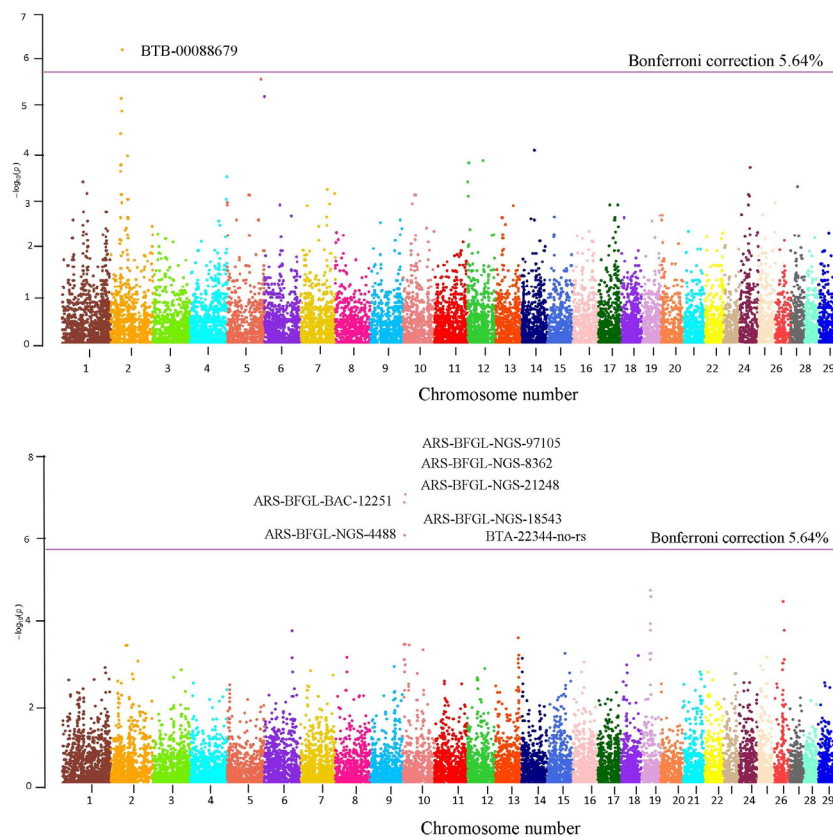
Based on association tests using PLINK for each phenotype, and following the application of Bonferroni correction (corrected P value  $\leq 2.25723E-06$ ), eight SNPs were found to be significant in two phenotypes. Table 2 shows the results of these associations.

**Table 2.** List of significant SNPs for Body Depth and Udder Depth traits.

Phenotype	SNP name	dbSNP-Identifier	Chr	Position (bp)	Most significant gene	Distance (bp)	-Log <sub>10</sub> (P)
Body Depth	BTB-00088679	rs43293152	2	34194876	GCA	Inside	6.11
Udder Depth	BTA-22344-no-rs	rs41606710	10	5598788	SFXN1	Inside	5.90
	ARS-BFGL-BAC-12251	rs109798047	10	5767015	LOC781028	414,403	6.83
	ARS-BFGL-NGS-4488	rs110012582	10	6476252	ANKRD31	Inside	5.90
	ARS-BFGL-NGS-18543	rs109007992	10	6521978	ANKRD31	Inside	5.90
	ARS-BFGL-NGS-8362	rs108984322	10	7943252	LOC100296562	35,024	7.06
	ARS-BFGL-NGS-97105	rs110898362	10	8255705	LOC107131388	15,532	7.06
	ARS-BFGL-NGS-21248	rs109191831	10	8553040	WDR41	Inside	7.06

Chr = Chromosome; bp = Base pair(s); P = P value.

For the Body Depth trait, the SNP BTB-00088679, located on chromosome 2, was significant with a  $-\log_{10}$  (P value =  $7.74E-07$ ) = 6.11. Its annotation was verified NCBI resources (<http://www.ncbi.nlm.nih.gov>). It is located in the GCA gene and encodes for a protein. Seven SNPs were significant for the Udder Depth trait, all located on chromosome 10 (BTA10), and identified as BTA-22344-no-rs, ARS-BFGL-BAC-12251, ARS-BFGL-NGS-4488, ARS-BFGL-NGS-18543, ARS-BFGL-NGS-8362, ARS-BFGL-NGS-97105, and ARS-BFGL-NGS-21248. Table 2 summarizes these two phenotypic traits in terms of chromosome location, start and end positions, gene and significance level for each SNP. Figure 1 shows a Manhattan plot of the  $-\log_{10}$  (P values). Thus, for Body Depth trait, only one SNP, located on chromosome 2, appears with a significance level above the threshold line, whereas for the Udder Depth trait seven SNPs appear as significant. No significant SNPs were detected for any of the other traits.



**Figure 1.** Manhattan plot of  $-\log_{10}(P)$  values for SNP - Phenotype association. The y-axis plots  $-\log_{10}(P)$  values and the x-axis plots the physical position of the SNPs sorted by chromosome and chromosome position. The horizontal line represents the Bonferroni-corrected significance threshold ( $-\log P = 5.64$ ). The plot are: (a) Body Depth, and (b) Udder Depth.

To verify if the significant SNPs were associated with quantitative trait loci (QTLs), a genomic QTL search was performed using data from the bovine genome database ([http://bovinegenome.org/bovineqtl\\_v2/findQTL.html](http://bovinegenome.org/bovineqtl_v2/findQTL.html)). The seven significant SNPs from chromosome 10 were associated with six QTLs related to Marbling Score, Milk Yield, Fat Yield, Protein Yield, and Protein Percentage. Table 3 shows a list of the QTLs associated with significant SNPs. No QTL relation was found in association with the significant SNP from chromosome 2.

**Table 3.** List of QTLs associated with significant SNPs.

Feature ID	Trait	Chromosome	Start position	End position
180	Marbling Score	BTA10	2225595	14961855
267	Milk Yield	BTA10	8292052	13636885
430	Milk Yield	BTA10	2225595	14270735
431	Fat Yield	BTA10	2225595	14270735
432	Protein Yield	BTA10	2225595	14270735
433	Protein Percentage	BTA10	2225595	14270735



To measure the consistency of these results, an alternative association analysis was performed using R version 3.2.3 software. The following procedure was implemented: 1) a simple linear regression was executed for each SNP, and those SNPs that had a P value  $\leq 0.1$  were selected; and 2) a multiple linear regression was executed for the selected set of SNPs. Following Bonferroni correction to the P values, we confirmed the same previous eight SNPs as significant.

## DISCUSSION

In this study, a genome-wide association analysis for a set of sampled and imputed SNPs with 16 conformation traits in a population of Mexican Holstein cows, revealed eight significant SNPs: one associated with the Body Depth trait, and seven associated with the Udder Depth trait. We used imputation from 6,912 to 46,523 SNPs as a low-cost optimization strategy, and although we did not specifically evaluate imputation accuracy because we only had low-density sampled data (6,912 SNPs), our results still indicated good associations. These findings support the validity of our study, especially given the fact that highly significant imputation accuracy has previously been reported using the same imputation tool (FindHap). More specifically (Boichard et al., 2012) and (Pimentel et al., 2015) reported 96.97% accuracy in Brown-Swiss cattle, and 97% accuracy in several other breeds when imputing from 6K to 50K, respectively.

SNP BTB-00088679 from chromosome 2 resulted in an association with the Body Depth trait. This SNP is located in the GCA gene, which codes for the protein Grancalcin. This calcium-binding protein plays a role in the adhesion of neutrophils to fibronectin, and may play a role in the formation of focal adhesions (see UniProtKB-P28676: <http://www.uniprot.org/uniprot/P28676>). We did not find any previous study reporting associations between the GCA gene and the Body Depth trait, however, a study by de Haas et al. (2007) reported that there was a moderate correlation between milk, fat, and protein yields and stature and Body Depth traits. Specifically, their study reported that genetically deeper, wider, and taller cows were predisposed for higher milk production.

In the case of Udder Depth trait, which was associated with seven SNPs from chromosome 10, four SNPs were located inside genes, whereas the rest were located within an average distance of 155 kb from the annotated genes (see Table 2). The specific functions of these genes vary; for example, gene SFXN1 codes for the protein Sideroflexin-1, which experimental evidence suggests may be involved in the transport of a required iron-utilization component into and out of the mitochondria. Similarly, the gene LOC781028 is a clathrin light chain A pseudogene, whereas the gene ANKRD31 codes for the putative ankyrin repeat domain-containing protein 31, and the gene LOC100296562 corresponds to a coiled-coil domain-containing protein 58-like. The gene LOC107131388 is a high affinity cAMP-specific and IBMX-insensitive 3',5'-cyclic phosphodiesterase 8B-like protein coding gene, and the gene WDR41 is a WD repeat-containing protein 41. All these genes, however, fall inside the five QTLs associated with Marbling Score, Milk Yield, Fat Yield, Protein Yield, and Protein Percentage.

## CONCLUSIONS

In this study, we present a genome-wide association analysis of a set of sampled and

imputed SNPs with 16 conformation traits in a population of Holstein cows from a desert area in Northwest Mexico. Imputation from 6K to 50K SNPs was performed as a low-cost optimization strategy. Results show eight SNPs associated with two conformation traits.

The Udder Depth trait resulted in seven associated SNPs from chromosome 10, and was related to Marbling Score, Milk Yield, Fat Yield, Protein Yield, and Protein Percentage QTLs. The Body Depth trait resulted in one associated SNP from chromosome 2, although no QTL relation was found.

Ultimately, the discovery of genes associated with conformation traits may be indicative of the adaptive selection pressures the Holstein breed has undergone in response to the extreme weather conditions found in northwest areas of Mexico. The results of this study indicate that traits such as stature and body depth may be used as indicators of cows' potential genetic merits for milk, fat, and protein production.

### Conflicts of interest

The authors declare no conflict of interest.

### ACKNOWLEDGMENTS

We are grateful to the National Council of Science and Technology of México for supporting a doctoral studies scholarship for M.E. González.

### REFERENCES

- Avendaño-Reyes L, Álvarez-Valenzuela FD, Correa-Calderón A, Saucedo-Quintero JS, et al. (2006). Effect of cooling Holstein cows during the dry period on postpartum performance under heat stress conditions. *Livest. Sci.* 105: 198-206. <http://dx.doi.org/10.1016/j.livsci.2006.06.009>
- Avendaño-Reyes L, Álvarez-Valenzuela FD, Correa-Calderón A, Algáandar-Sandoval A, et al. (2010). Comparison of three cooling management systems to reduce heat stress in lactating Holstein cows during hot and dry ambient conditions. *Livest. Sci.* 132: 48-52. <http://dx.doi.org/10.1016/j.livsci.2010.04.020>
- Boichard D, Chung H, Dasonneville R, David X, et al.; Bovine LD Consortium (2012). Design of a bovine low-density SNP array optimized for imputation. *PLoS One* 7: e34130. <http://dx.doi.org/10.1371/journal.pone.0034130>
- Cole JB, Wiggans GR, Ma L, Sonstegard TS, et al. (2011). Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics* 12: 408. <http://dx.doi.org/10.1186/1471-2164-12-408>
- de Haas Y, Janss LLG and Kadarmideen HN (2007). Genetic and phenotypic parameters for conformation and yield traits in three Swiss dairy cattle breeds. *J. Anim. Breed. Genet.* 124: 12-19. <http://dx.doi.org/10.1111/j.1439-0388.2007.00630.x>
- García E (2004). Modificaciones al sistema de clasificación climática de Köppen: (para adaptarlo a las condiciones de la República Mexicana). Instituto de Geografía. 6a ed. Universidad Nacional Autónoma de México, D.F. México.
- Holstein Association USA, Inc. (2017). Linear Descriptive Traits. available online at [http://www.holsteinusa.com/pdf/print\\_material/linear\\_traits.pdf](http://www.holsteinusa.com/pdf/print_material/linear_traits.pdf).
- Nakagawa S (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behav. Ecol.* 15: 1044-1045. <http://dx.doi.org/10.1093/beheco/arh107>
- Pimentel ECG, Edel C, Emmerling R and Götz K-U (2015). How imputation errors bias genomic predictions. *J. Dairy Sci.* 98: 4131-4138. <http://dx.doi.org/10.3168/jds.2014-9170>
- Purcell S, Neale B, Todd-Brown K, Thomas L, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81: 559-575. <http://dx.doi.org/10.1086/519795>
- R Development Core Team (2011) R: A Language and Environment for Statistical Computing. Vienna, Austria: The R Foundation for Statistical Computing. Available at [<http://www.R-project.org/>].



- Rosner B (2011). *Fundamentals of Biostatistics*, Cengage Learning. Independence, Kentucky.
- VanRaden PM, O'Connell JR, Wiggans GR and Weigel KA (2010). Combining different marker densities in genomic evaluation. *Interbull Bull* 42: 1-4.
- Wu X, Fang M, Liu L, Wang S, et al. (2013). Genome wide association studies for body conformation traits in the Chinese Holstein cattle population. *BMC Genomics* 14: 897. <http://dx.doi.org/10.1186/1471-2164-14-897>
- Zimin AV, Delcher AL, Florea L, Kelley DR, et al. (2009). A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10: R42. <http://dx.doi.org/10.1186/gb-2009-10-4-r42>