Published in final edited form as:

Ann Hum Genet. 2010 March ; 74(2): 184–188. doi:10.1111/j.1469-1809.2010.00561.x.

Limited Distribution of a Cardiomyopathy-Associated Variant in India

Tatum S. Simonson¹, Yuhua Zhang¹, Chad D. Huff¹, Jinchuan Xing¹, W. Scott Watkins¹, David J. Witherspoon¹, Scott R. Woodward², and Lynn B. Jorde¹

¹Department of Human Genetics, University of Utah, Salt Lake City, Utah

²Sorenson Molecular Genealogy Foundation, Salt Lake City, Utah

Abstract

Heart failure is a leading cause of death of people in South Asia, and cardiomyopathy is a major cause of heart failure. Myosin binding protein C (MYBPC3) is expressed in the heart muscle, where it regulates the cardiac response to adrenergic stimulation and is important for the structural integrity of the sarcomere. Mutations in the *MYBPC3* gene are associated with hypertrophic or dilated cardiomyopathies. A 25-base-pair deletion in intron 32 causes skipping of the downstream exon and is associated with familial cardiomyopathy. To date, this deletion is found primarily in India and South Asia, although it is also found at low frequency in Southeast Asia. In order to better characterize the distribution of this variant, we determined its frequency in 447 individuals from 19 populations, including 10 populations from India and neighboring populations from Pakistan and Nepal. The deletion frequency is over 8% in some of our Indian samples, and it is not present in any of the populations we sampled outside of India. The differences in the deletion frequencies among populations in India are consistent with patterns of variation previously reported and with patterns we observed among Indian populations based on high-density SNP chip data. Our results indicate the *MYBPC3* deletion is primarily found among Indian populations, and that its distribution is consistent with genome-wide patterns of variation in India.

INTRODUCTION

Heart failure is a global disease. One important cause of heart failure, cardiomyopathy, leads to premature death among individuals over 40 years of age and affects an estimated 1–2% of the Western population (Berry *et al.*, 2001). Studies based on human subjects, animal models, and gene expression analyses have elucidated hundreds of cardiomyopathy-associated mutations in more than 20 genes (Liew & Dzau, 2004). Several of these genes are involved in pathways essential for the development and structural integrity of the heart muscle. Despite ongoing genetic investigation, little is understood about the causes and prevalence of cardiomyopathy among individuals in developing countries.

Studies of previously underrepresented populations have helped elucidate the genetic basis of cardiomyopathy (Kubo, 2005), and indicate that it is relatively common among individuals of South Asian ancestry (Joshi, 2007, Reddy *et al.*, 2005). Recently, Dhandapany *et al.* (2009) reported that a specific cardiomyopathy-associated genetic variant, a deletion in myosin binding protein C 3 (MYBPC3), is primarily limited to South Asia. As with most cardiomyopathy-associated mutations, many carriers of this deletion exhibit late onset of disease (Dhandapany *et al.*, 2009). *MYBPC3* is expressed in the heart muscle, where it regulates the cardiac response to adrenergic stimulation and is important for organization of the myocardium. The South-Asian-specific *MYBPC3* mutation is a 25-basepair (bp) deletion in the intron upstream of exon 33 (Figure 1). Transcriptional studies indicate that this deletion leads to skipping of exon 33 (Dhandapany *et al.*, 2009), and it is possible that accumulation of

the altered protein may contribute to late-onset cardiomyopathy (Kubo, 2005, Niimura, 1998).

In order to further characterize patterns of variation of this deletion, we determined its frequency in 447 Asians, including populations from India and surrounding regions, East Asia, and Southeast Asia (Figure 2). A previous report indicated a high prevalence of this mutation in India (approximately 4%) (Dhandapany *et al.*, 2009). Here, we provide a description of the South-Asian-specific distribution of the *MYBPC3* deletion in populations from and near India, highlighting differences in its prevalence among samples from 10 different tribal and caste populations from the southern state of Andhra Pradesh in India. To better place this variation in perspective, we compare it to genetic variation assayed using high-density single nucleotide polymorphism (SNP) microarrays.

METHODS

Samples

DNA samples were obtained from unrelated individuals with informed consent as previously described (Watkins *et al.*, 2005, Watkins *et al.*, 2003, Xing *et al.*, 2009). Subjects included in this study are permanent residents of their respective regions with ancestors having lived in that region. Each population is designated according to language or geographic location, and in the case of all Indian samples, by caste or tribal group. The DNA used for this study was collected from the following Indian populations from Andhra Pradesh: Brahmin (52), Kshatriya (11), Vysya (8), Kapu (51), Yadava (44), Madiga (27), Mala (24), Relli (19), Khonda Dora (25), and the Irula (26). Asian individuals from the following populations were included for additional analysis: Pakistan (25), Nepal (25), Iraq (25), Kyrgyzstan (25), Thailand (25), Vietnam (5), Cambodia (7), China (10), and Japan (13). We examined mitochondrial DNA and Y chromosome polymorphisms for each of the populations to verify ancestry. Genome-wide single nucleotide polymorphism (SNP) analysis was performed on Brahmin, Mala, Madiga, and Irula populations within India and all individuals from regions outside of India.

Deletion genotyping

We amplified the 25-bp deletion region using polymerase chain reaction (PCR) under standard conditions with 5'-GTTTCCAGCCTTGGGCATAGTC-3' and 5'-GAGGACAACGGAGCAAAGCCC-3' primers. The amplicons were run on 3% MetaPhor Agarose gel at 120 volts for two hours to determine presence or absence of the deletion. Allele frequencies were determined by direct counting and upper and lower binomial confidence limits were determined at the $\alpha = 0.05$ level. Fisher's exact tests were used to compare allele frequencies between each population and all the remaining populations.

Genome-wide analysis

Approximately 240,000 SNP genotypes for Andhra Brahmin, Madiga, Mala, and Irula and all populations outside of India were obtained from Xing *et al.* 2009. Per-locus pairwise F_{ST} was calculated for the set of SNPs genotyped in all populations as previously described (Xing *et al.*, 2009).

RESULTS

The deletion is present in most of our Indian population samples but absent in the populations we sampled outside of India. Interestingly, the allele frequency varies considerably within India. The deletion is absent in two of ten Indian populations and is present with an allele frequency greater than 8% in three of the populations. The allele frequency in the Yadava (middle caste) population are significantly higher than the remaining Indian populations

(Fisher's exact test, $P = 0.01$). However, the allele frequencies in the other Indian populations do not significantly deviate from the rest of the populations (Table 1).

The Khonda Dora tribal population, which exhibit genetic similarity to Southeast Asians, has a relatively high allele frequency of the deletion (10%), whereas the mutation is completely absent in the Irula sample, another tribal group from Andhra Pradesh. More than 8% of the Mala and Yadava (lower and middle castes, respectively) also have the deletion; all other Indian populations sampled, including upper-caste Brahmins, exhibit heterozygote frequencies less than five percent. The distribution of the deletion allele frequencies in all of our populations sampled from India is shown in Table 1.

In order to determine whether the variation in deletion allele frequencies within India is unusual, we calculated pairwise F_{ST} between four Indian populations using ~240,000 genome-wide SNPs and compared these values to the deletion pairwise F_{ST} in the same four Indian populations. All deletion F_{ST} values were less than those calculated for the genome-wide average, with the exception of comparisons made with the Mala. While pairwise F_{ST} values were higher than the genome average in all Mala comparisons (0.013–0.069), values fell within the 95% confidence interval of the genome-wide range of F_{ST} (the upper 95% confidence interval is between 0.10 and 0.165). Therefore, the F_{ST} for the deletion is not significantly outside the normal range of genetic variation for populations within India, even in the most extreme case of the Mala, and the population allele frequency distribution of the 25-bp deletion is typical of the variation observed throughout the genome in India.

DISCUSSION

The *MYBPC3* 25-bp deletion is similar in effect to other *MYBPC3* mutations, which manifest later in life, lead to mild hypertrophy, and exhibit incomplete penetrance (Kubo, 2005, Niimura, 1998). Based on a family study of this deletion, ~90% of the oldest family members who carried the deletion were symptomatic (Dhandapany *et al.*, 2009). Although individuals with this deletion typically present with cardiomyopathy at more than 30 years of age, homozygous individuals are affected with severe phenotypes approximately a decade earlier (Dhandapany *et al.*, 2009). The risk of developing cardiomyopathy is probably also increased by other genetic and environmental factors (Dhandapany *et al.*, 2009).

It is interesting that the *MYBPC3* 25-bp deletion is relatively common in India and its frequency varies across this limited region. The time to the most recent common ancestor (TMRCA) estimate, based on previous haplotype analysis, is ~33,000 (−/+23,000) years before present (Dhandapany *et al.*, 2009), suggesting that the deletion might have arisen in India and was possibly not present in the initial inhabitants that arrived in this region from Africa some 50,000 to 20,000 years ago (Misra, 1992). Previous reports indicate it reaches highest frequency in southern and western Indian states (2% to 8%) and is absent from India populations sampled in the Northeast, recent migrant populations originally from Africa, and a group from the Andaman Islands (Dhandapany *et al.*, 2009).

The *MYBPC3* deletion allele frequency we report here (~5%) and the variation we observe among our Indian samples is consistent with the previous study (Dhandapany *et al.*, 2009). One of the middle caste groups, the Yadava, has significantly different allele frequencies compared to other populations, potentially due to the effects of genetic drift. Interestingly, the Irula tribal group we sampled does not have the deletion, whereas the Khonda Dora, another tribal group from Andhra Pradesh, exhibits the second highest allele frequency among our ten populations (10%). The dramatic difference among tribal populations is possibly due to genetic drift and, in the case of the Irula samples, extreme isolation from other Indian populations.

The *MYBPC3* deletion patterns we describe here are also concordant with reported patterns of genetic variation and divergence based on neutral genetic markers. The genetic affinity of Indian populations has been previously analyzed using *Alus*, short-tandem repeats (STRs), and genome-wide SNP data (Bamshad *et al.*, 2001, Basu *et al.*, 2003, Roychoudhury *et al.*, 2001, Thanseem *et al.*, 2006, Watkins *et al.*, 2005, Watkins *et al.*, 2003, Xing *et al.*, 2009, Reich *et al.*, 2009). SNP microarray and STR studies show relative homogeneity among caste populations, whereas tribal Indian populations are more distinct from caste populations and each other (Watkins *et al.*, 2005, Xing *et al.*, 2009). The *MYBPC3* deletion F_{ST} values lie in the normal genome-wide range, suggesting that the same factors that shape genome-wide patterns are also shaping the distribution of this deletion.

It is somewhat surprising that a deleterious mutation would be found at such a high frequency. Considering the region surrounding the *MYBPC3* locus is gene-rich (over 15 genes in the 500 Kb region), it is possible that the *MYBPC3* deleterious variant hitch-hiked along with a linked region that was subjected to positive selection. Dhandapany *et al.* (2009) reported that there is no evidence of deviation from neutral evolution for the 25-bp deletion region. Their conclusion is based on two analyses performed in individuals both with and without the deletion, which indicate: 1) similar amounts of variation observed for five STR loci spanning 3.4 megabases (Mb) in more than 250 individuals examined, and 2) no significant difference in sequence variation 5 kilobases on either side of the deleted region in nine heterozygous individuals.

Interestingly, there are a relatively large number of mutations in the *MYBPC3* locus (over 150 identified thus far) that have deleterious effects in Western and Asian populations (<http://cardiogenomics.med.harvard.edu/home> Harvard CardioGenomics). The recombination rate in this region is exceedingly low (0.08 cM/Mb) compared to the genome-wide average (1.3cM/Mb) (McVean *et al.*, 2004). It is possible that the disadvantageous *MYBPC3* mutations accumulated in these populations faster than they can be removed by purifying selection due to the low recombination rate (Gillespie, 2004, Muller, 1964). While it is possible that variants in this region have been influenced by positive selection or have not been removed by purifying selection, such hypotheses require further investigation.

Acknowledgments

We thank J.M. Naidu, B.B Rao, B.V. Prasad, and M.J. Bamshad for assistance with sample collection and A.R. Rogers for helpful discussion. This work was supported by NSF grants SBR-9514733, SBR-9512178, and NIH grant GM-59290.

REFERENCES

- Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BVR, Reddy PG, Rasanayagam A, Papiha SS, Vilems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB. Genetic Evidence on the Origins of Indian Caste Populations. *Genome Research* 2001;11:994–1004. [PubMed: 11381027]
- Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya NP, Roychoudhury S, Majumder PP. Ethnic India: A Genomic View, With Special Reference to Peopling and Structure. *Genome Research* 2003;13:2277–2290. [PubMed: 14525929]
- Berry C, Murdoch DR, McMurray JJV. Economics of chronic heart failure. *Eur. J. Heart Fail* 2001;3:283–291. [PubMed: 11377998]
- Dhandapany PS, Sadayappan S, Xue Y, Powell GT, Rani DS, Nallari P, Rai TS, Khullar M, Soares P, Bahl A, Tharkan JM, Vaideeswar P, Rathinavel A, Narasimhan C, Ayapati DR, Ayub Q, Mehdi SQ, Oppenheimer S, Richards MB, Price AL, Patterson N, Reich D, Singh L, Tyler-Smith C, Thangaraj K. A common MYBPC3 (cardiac myosin binding protein C) variant associated with cardiomyopathies in South Asia. *Nat Genet* 2009;41:187–191. [PubMed: 19151713]
- Gillespie, JH. Maryland: The Johns Hopkins University Press; 2004.

- Joshi P. Risk factors for early myocardial infarction in South Asians compared with individuals in other countries. *J. Am. Med. Assoc* 2007;297:286–294.
- Kubo T. Lifelong left ventricular remodeling of hypertrophic cardiomyopathy caused by a founder frameshift deletion mutation in the cardiac myosin-binding protein C gene among Japanese. *J. Am. Coll. Cardiol* 2005;46:1737–1743. [PubMed: 16256878]
- Liew CC, Dzau VJ. *Molecular Basis of Cardiovascular Disease: A Companion to Braunwald's Heart Disease*. 2004
- Mcvean GAT, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. The Fine-Scale Structure of Recombination Rate Variation in the Human Genome. *Science* 2004;304:581–584. [PubMed: 15105499]
- Misra VN. Research on the Indus Civilization: A Brief Review. *Indus Civilization Special Number of the Eastern Anthropologist* 1992;45:1–19.
- Muller HJ. The Relation of Recombination to Mutational Advance. *Mutational Research* 1964;106:2–9.
- Niimura H. Mutations in the gene for cardiac myosin-binding protein C and late-onset familial hypertrophic cardiomyopathy. *N. Engl. J. Med* 1998;338:1248–1257. [PubMed: 9562578]
- Reddy KS, Shah B, Varghese C, Ramadoss A. Responding to the threat of chronic diseases in India. *Lancet* 2005;366:1744–1749a. [PubMed: 16291069]
- Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. *Nature* 2009;461:489–494. [PubMed: 19779445]
- Roychoudhury S, Roy S, Basu A, Banerjee R, Vishwanathan, Vishwanathan H, Usha R, Rani MU, Sil S, Mitra M, Majumder P. Genomic structures and population histories of linguistically distinct tribal groups of India. *Human Genetics* 2001;109:339–350. [PubMed: 11702215]
- Thanseem I, Thangaraj K, Chaubey G, Singh V, Bhaskar L, Reddy BM, Reddy A, Singh L. Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC Genetics* 2006;7:42. [PubMed: 16893451]
- Watkins WS, Prasad BVR, Naidu JM, Rao BB, Bhanu BA, Ramachandran B, Das PK, Gai PB, Reddy PC, Reddy PG, Sethuraman M, Bamshad MJ, Jorde LB. Diversity and Divergence Among the Tribal Populations of India. *Annals of Human Genetics* 2005;69:680–692. [PubMed: 16266407]
- Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington A-ME, Carroll ML, Nguyen SV, Walker JA, Prasad BVR, Reddy PG, Das PK, Batzer MA, Jorde LB. Genetic Variation Among World Populations: Inferences From 100 Alu Insertion Polymorphisms. *Genome Research* 2003;13:1607–1618. [PubMed: 12805277]
- Xing J, Watkins WS, Witherspoon DJ, Zhang Y, Guthery SL, Thara R, Mowry BJ, Bulayeva K, Weiss RB, Jorde LB. Fine-scaled human genetic structure revealed by SNP microarrays. *Genome Research* 2009;19:815–825. [PubMed: 19411602]

MYBPC3

Figure 1.
MYBPC3 region (Chromosome 11p11.2) and the 25-basepair deletion that leads to skipping of downstream exon 33.

Figure 2.

Map indicating regions from which DNA samples were collected. All Indian samples were collected from Andhra Pradesh. The Irula samples were collected from the southern region of Andhra Pradesh; all other samples were collected from the northern region of the state.

Frequency of the 25-basepair *MYBPC3* deletion in Indian samples. Comparisons between each population and the remaining populations were performed using Fisher's exact test. Indian samples genotyped with the Affymetrix NspI 250K SNP array are italicized.

Table 1

Caste or Tribal Affiliation	Population	Deletion count	Number of chromosomes	Allele frequency (Binomial 95% confidence limits)	Allele frequency previously reported by Dhandapani <i>et al.</i> 2009	P-value (Fisher's exact test)
Upper Caste	Brahmin	2	104	0.019 (0.002 – 0.068)		0.11
	Kshatriya	1	22	0.045 (0.001 – 0.228)		0.66
	Vysya	0	16	0.000 (0.000 – 0.206)	0.02	0.46
Middle Caste	Kapu	3	102	0.029 (0.006 – 0.084)		0.26
	Yadava	9	88	0.102 (0.048 – 0.185)		0.01
Lower Caste	<i>Malta</i>	4	48	0.083 (0.023 – 0.200)		0.18
	<i>Madiga</i>	1	54	0.019 (0.001 – 0.099)	0.04	0.26
	Relli	2	38	0.053 (0.006 – 0.178)		0.55
Tribal Populations	Khonda Dora	5	52	0.096 (0.032 – 0.210)		0.09
	<i>Irula</i>	0	52	0.000 (0.000 – 0.069)	0.08	0.07
Total in India			524	0.050 (0.074 – 0.034)		