On the Evolution of *Yeti*, a *Drosophila melanogaster* Heterochromatin Gene



Roberta Moschetti¹, Emanuele Celauro², Fulvio Cruciani², Ruggiero Caizzi¹, Patrizio Dimitri²*

1 Dipartimento di Biologia, Università degli Studi di Bari, Bari, Italy, 2 Dipartimento di Biologia e Biotecnologie "Charles Darwin" and Istituto Pasteur Fondazione Cenci-Bolognetti, Sapienza Università di Roma, Roma, Italy

Abstract

Constitutive heterochromatin is a ubiquitous and still unveiled component of eukaryotic genomes, within which it comprises large portions. Although constitutive heterochromatin is generally considered to be transcriptionally silent, it contains a significant variety of sequences that are expressed, among which about 300 single-copy coding genes have been identified by genetic and genomic analyses in the last decades. Here, we report the results of the evolutionary analysis of Yeti, an essential gene of Drosophila melanogaster located in the deep pericentromeric region of chromosome 2R. By FISH, we showed that Yeti maintains a heterochromatin location in both D. simulans and D. sechellia species, closely related to D. melanogaster, while in the more distant species e.g., D. pseudoobscura and D. virilis, it is found within euchromatin, in the syntenic chromosome Muller C, that corresponds to the 2R arm of D. melanogaster chromosome 2. Thus, over evolutionary time, Yeti has been resident on the same chromosomal element, but it progressively moved closer to the pericentric regions. Moreover, in silico reconstruction of the Yeti gene structure in 19 Drosophila species and in 5 non-drosophilid dipterans shows a rather stable organization during evolution. Accordingly, by PCR analysis and sequencing, we found that the single intron of Yeti does not undergo major intraspecies or interspecies size changes, unlike the introns of other essential Drosophila heterochromatin genes, such as light and Dbp80. This implicates diverse evolutionary forces in shaping the structural organization of genes found within heterochromatin. Finally, the results of $d_s - d_N$ tests show that Yeti is under negative selection both in heterochromatin and euchromatin, and indicate that the change in genomic location did not affected significantly the molecular evolution of the gene. Together, the results of this work contribute to our understanding of the evolutionary dynamics of constitutive heterochromatin in the genomes of higher eukaryotes.

Citation: Moschetti R, Celauro E, Cruciani F, Caizzi R, Dimitri P (2014) On the Evolution of Yeti, a Drosophila melanogaster Heterochromatin Gene. PLoS ONE 9(11): e113010. doi:10.1371/journal.pone.0113010

Editor: Igor V. Sharakhov, Virginia Tech, United States of America

Received August 11, 2014; Accepted October 22, 2014; Published November 18, 2014

Copyright: © 2014 Moschetti 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.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper.

Funding: This research was supported by grants from Ministero dell'Università e della Ricerca- Miur (Prin 2008), Istituto Pasteur-Fondazione Cenci-Bolognetti and Fondazione Roma terzo settore. 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.

* Email: patrizio.dimitri@uniroma1.it

Introduction

Constitutive heterochromatin is commonly found in large blocks near centromeres and telomeres; it consists mostly of repetitive DNA sequences and maintains its characteristic organization on both homologous chromosomes. It is a ubiquitous component of eukaryotic genomes and, in many species, comprises large chromosomal portions, or even entire chromosomes. For example, about 30% of the *Drosophila* and human genomes, and up to 70– 90% of certain nematode and plant genomes, are made up of constitutive heterochromatin [1,2,3], yet the reasons for its widespread occurrence are still unclear.

Heterochromatin was originally defined at cytological level as the chromosome portion that stains deeply at prophase and maintains a compact organization throughout all stages of the mitotic cell cycle [4]. Historically, distinctive antagonistic properties compared to the rest of the genome were identified: 1) strongly reduced level of meiotic recombination; 2) low gene density; 3) mosaic inactivation of the expression of euchromatic genes when moved nearby (position effect variegation, PEV); 4) late replication during S phase; 5) transcriptional inactivity; 6) enrichment in the so-called "junk" repetitive DNA, such as satellite sequences and truncated transposable element remnants.

Together, these properties led to the view of constitutive heterochromatin as a "desert" of genetic functions [5]. In the last three decades, however, studies primarily conducted in *Drosophila melanogaster* have shown that constitutive heterochromatin does in fact play roles in important cellular functions, such as chromosome organization and inheritance [6,7,8,9,10,11]. Although generally regarded as transcriptionally silent, constitutive heterochromatin has been found to contain actively transcribed genes [3]. For example, in *Drosophila melanogaster*, more than 40 genes essential for viability or fertility have been mapped to pericentric heterochromatin [12,13,14,15,16,17].

In the last decade, the release of *D. melanogaster* heterochromatin sequence by the Berkeley Drosophila Genome Project (http://www.fruitfly.org/) and Drosophila Heterochromatin Genome Project (http://www.dhgp.org/index_release_notes.html) has greatly facilitated studies of mapping, molecular organization and function of genes located in pericentromeric heterochromatin [18].



Figure 1. Cytogenetic mapping of heterochromatin genes of chromosome 2. The map was modified from that shown in previous papers [3,45]. The diagram shows the essential genes defined by mutational analyses (below) and annotated genes defined by the heterochromatin genome project (above). Shades of blue correspond to the intensity of DAPI staining, with the darkest blue blocks representing regions with strong fluorescence intensity and open blocks representing non fluorescent regions. The different cytological regions are numbered. doi:10.1371/journal.pone.0113010.g001

More recently, an improved whole genome shotgun assembly [19] has been produced, which includes 20.7 Mb of draft-quality heterochromatin sequence. In the last years, 15 Mb of this sequence have been further improved or completed [20] and a BAC-based physical map of 13 Mb of pericentric heterochromatin, together with the cytogenetic map that locates some 11 Mb to specific heterochromatin regions, have been constructed [20]. About 250 protein-coding genes were defined in the release 5.1 annotation of the currently sequenced heterochromatin DNA [21]. According to these results, the number of active genes in constitutive heterochromatin of *D. melanogaster* appears to be higher than defined by genetic analysis. Notably, these genes encode proteins involved in important cellular and developmental processes [3].

Further mapping of *D. melanogaster* heterochromatin was performed by comparative genomic hybridization [22]. The

transcription profiles of mapped sequences by microarray analysis also revealed region-specific temporal patterns of transcription within heterochromatin during oogenesis and in early embryonic development.

Evolutionary studies have shown that *D. melanogaster* heterochromatin genes, such as *light* and others, originated from progenitors that were originally located within euchromatin in the drosophilid lineage [23,24]. Here we have focussed our study on the evolutionary origin of *Yeti*, an essential heterochromatin gene of *D. melanogaster*, which encodes a protein belonging to the evolutionarily conserved BCNT family of chromatin remodellers [25,26]. We report that *Yeti* locates in euchromatin in distant species, e.g. *D. pseudoobscura* and *D. virilis*, similarly to what has been found for *light* and other genes [23,24]. Moreover, we found that the *Yeti* gene structure remains rather stable during the evolution of *Drosophila* species. In particular, the second exon that Table 1. List of the Yeti orthologs and their encoded proteins.

Species	Gene ID	Database Location	GenBank A.C. (position)	Uniprot reference	Amino acids
D.melanogaster	FBgn0128734	2Rh:13434031345119	NW_001848856.1	B4J7U2	241
D.simulans	FBgn0191193	3R:38,26939,056 [+]	NT_167061.1	B4QUX4	241
D.sechellia	FBgn0166002	2h;scaffold_170:38,6733,460 [-].	NW_001999858.1	B4IME9	241
D.yakuba	FBgn0236265	2h;v2_chr2h_random_005:668,463673,172.	NW_002052891.1	B4IT75	215
D.erecta	FBgn0103193	scaffold 4929	NW_001956548.1	B3N457	236
D.eugracilis	not available	not available	KB464511.1 (1971618949)	not available	235
D.biarmipes	not available	not available	KB462255.1 (1883919580)	not available	229**
D.takahashi	not available	not available	KB460683.1 (1104410283)	not available	235**
D.elegans	not available	not available	KB458177.1 (62046991)	not available	244**
D.ananassae	FBgn0088065	2R euchromatic region scaffold 13266	NW_001939294.1	B3MBR1 ^a	272 ^a
D.bipectinata	not available	not available	KB464371.1 (4801248773)	not available	232
D.pseudoobscura	FBgn0245984	2R euchromatic region	NC_009006.2	B5E0W2	275
D.persimilis	FBgn0148669	not available	NW_001985955.1	B4GBJ4	275
D.miranda	not available	not available	CM001519.2 (366522367496)	not available	275
D.willistoni	FBgn0212772	2R euchromatic arm	NW_002032340.1	B4MJ08	273
D.mojavensis	FBgn0143134	2R euchromatic arm	NW_001979114.1	B4KQ8	295
D.virilis	FBgn0209341	2R euchromatic arm	NW_002014420.1	B4LKP7	300
D.albomicans	not available	not available	JH859027.1 (58948055895706)	not available	279
D.grimshawi	FBgn0128734	2R euchromatic arm	NW_001961673.1	B4J7U2	285
C.quinquefasciatus	CPIJ018830	not available	NW_001888048.1	B0XH12	284
A.aegypti	AAEL007422	supercontig 1.255, euchromatic region 2p25 [46]	NW_001810963.1	Q0IF03	279
A.darlingi	not available	not available	ADMH02000690.1 (45710 44829)	not available	
A.gambiae	AGAP005152	2L:euchromatic region 21E	NT_078265.2	Q7PPY4	293
M.destructor	not available	not available	GL501532.1 (369211 368309)	not available	300

**ORF defective. Amino acids deducted from bestfit alignments.

^{a)}Bestfit protein is 236 aminoacids long.

doi:10.1371/journal.pone.0113010.t001

encodes the last 30 aminoacids of the conserved BCNT domain is invariably 91 bp-long. Finally, we found that the single intron of *Yeti* does not undergo major size changes in *D. melanogaster* and closely related species, unlike the introns of other essential *Drosophila* heterochromatin genes [27].

Results

Evolutionary repositioning of the *Yeti* gene from euchromatin to pericentric heterochromatin

The single-copy *Yeti* gene of *D. melanogaster* maps to the region h41 of chromosome 2R mitotic heterochromatin (Figure 1; Table 1), which corresponds to division 41A of salivary gland polytene chromosomes [25,26,28,29].

To characterize the chromosomal location of Yeti among Drosophila genus species, we performed fluorescent in situ hybridization (FISH) experiments on polytene chromosomes of D. simulans, D. sechellia, two sibling species of D. melanogaster, and on two distantly related species: D. pseudoobscura, belonging to the Sophophora subgenus and D. virilis, belonging to the Drosophila subgenus. These species cover nearly a 40 million years divergence time and thus represent a wide spectrum of the evolutionary history of Yeti.

To map Yeti in D. simulans and D. sechellia, we used the D. melanogaster Yeti cDNA probe (RE36623), while PCR species-specific probes were used in D. pseudoobscura and D. virilis. PCR

probes were amplified over a less conserved region located outside the C-terminal BCNT coding domain of YETI protein (see Materials and Methods).

The results of this analysis are shown in Figure 2. In D. melanogaster, D. simulans and D. sechellia the Yeti cDNA probe produces a signal mapping to the base of division 41A, in the right arm of chromosome 2 (Figure 2A,B,C). Notably, the signals show a large diffuse structure very different from the sharp hybridization signals usually seen with euchromatic probes; such a morphology is a distinctive mark for sequences derived from partially polytenized heterochromatin regions [30,31]. Together, our FISH results indicate that Yeti maintains a heterochromatic location in D. simulans and D. sechellia. The FlyBase localization of Yeti in D. sechellia is in 2 h (scaffold_170:38,673..3,460; Table 1), in accord with our mapping results, while that in D. simulans in 3R (38,269..39,056; Table 1) is apparently conflicting and may reflect an assembly error in the D. simulans genome sequences (see discussion), as reported by Schaffer et al. [32]. In both D. pseudoobscura and D. virilis a single FISH signal was observed in the euchromatic arms of polytene chromosomes, in agreement with FlyBase (Figure 2D, E; Table 1). In D. pseudoobscura, the Yeti PCR probe produced a sharp signal that maps to region 63C in the proximal euchromatin of chromosome 3, while in D. Virilis the Yeti signal is found at region 53E, in the distal euchromatin of chromosome 5. Thus, independently of their genome localization (heterochromatin or euchromatin), in the analysed species Yeti lies



Figure 2. Examples of FISH mapping of *Yeti* **probes to polytene chromosomes of** *Drosophila* **species.** Salivary gland polytene chromosomes were stained with DAPI and pseudocolorated in blue; fluorescent signals were pseudocolorated in red. In *D. melanogaster* (A) and in the closely related *D. simulans* (B) and *D. sechellia* (C) species, the *Yeti* cDNA probe maps to *2Rh* at the base of polytene division 41. The large and diffuse morphology of the *Yeti* signal found in these species, reflects the disorganized and poorly banded structure of the heterochromatin in the chromocenter. The arrows point the base of *2Rh*. In *D. pseudobscura*, the hybridization signal of *Yeti* PCR probe maps to region 63C in the proximal euchromatin (D). In *D. virilis* the *Yeti* hybridization signal maps to region 53E, in the distal euchromatin of chromosome 5 (E).

doi:10.1371/journal.pone.0113010.g002

in the syntenic chromosome Muller C, that corresponds to the 2R arm of *D*. *melanogaster* chromosome 2.

In silico reconstruction of *Yeti* gene organization in different sequenced genomes

To study the evolutionary conservation of the Yeti gene organization, we have characterized the structure of the Yeti orthologs in 19 Drosophila species (D. melanogaster, D. simulans, D. sechellia, D. yakuba, D. erecta, D. eugracilis, D. biarmipes, D takahashii, D. elegans, D. ananassae, D. bipectinata, D. pseudobscura, D. persimilis, D. miranda, D. willistoni, D. mojavensisi, D. virilis, D. albomicans, and D. grimshawi) and in five nondrosophilid dipterans (C. quinquefasciatus, A. Aegypti, A. darlingi, A. gambiae and M. destructor pest crop).

All the Yeti DNA sequences were retrieved from FlyBase. In case of annotated genes, the YETI protein sequences were extracted from the Ortho DB database [33], where they are reported as orthologous sequences belonging to the BCNT family complex [26,34]. For the recently sequenced genomes, the Yeti DNA sequences were recovered by the TblastN procedure, using the *D. melanogaster* YETI protein sequence. The alignments of the retrieved Yeti orthologs are shown in Figure 3 and their coordinates are reported in Table 1.

We were able to recover a deducted complete protein sequence except for the recently added genomes of *D. biarmipes*, *D. takahashi* and *D. elegans*, where frame-shift mutations were found in the 5' end of the gene, probably due to errors in sequencing that needs to be improved. However, in each case the 3'end of the gene, containing the BCNT domain-coding region, was detected (Figure 3). The protein sequence alignments show a strong conservation of the BCNT domain in the *Drosophila* genus and in non-drosophilid dipterans (Figure 4).

We next compared the reconstructed molecular organization of the Yeti gene in the above-mentioned species, to study whether it underwent substantial structural changes during evolution. The results of this analysis are shown in Figure 5. It appears that the structure of the Yeti orthologs, with two exons and one intron remained highly conserved during the divergence of the lineages in the Drosophila genus, the only exception being D. willistoni, where an additional intron of 75 bp is present. The gene size is identical among D. melanogaster, D. simulans and D. sechellia, the only detectable difference is represented by the intron that in D. melanogaster is 7 bp longer. In general, in the Drosophila genus both the first exon and the single intron undergo changes in size: the exon varies from 557 bp (D. yakuba) to 812 bp (D. virilis), while the intron spans from 51 bp (D. biarmipes) to 70 bp (D. willistoni). Notably, the size of the second exon, that encodes the last portion of the BCNT domain, shows a striking conservation in the Drosophila genus, being invariably 91 bp long. Finally, in the five genome species of non-drosophilid dipterans (C. quinquefasciatus, A. Aegypti, A. darlingi, A. gambiae and M. destructor) the organization of Yeti differs in that the intron disappears and a single exon is present.

Characterization of the Yeti intron in Drosophila species

Intraspecific and interspecific size polymorphisms of heterochromatin gene introns have been found in *Drosophila*, which are likely to be associated with *de novo* insertions of TE-related sequences [27]. We then asked whether the intron of *Yeti* is prone to TE insertions or to other gross changes in length. To answer this question, we PCR amplified a region of about 180 bp comprising the *Yeti* intron from genomic DNAs extracted from *D. melanogaster* and *D. simulans* strains. Most of the analyzed strains derived from geographically distant natural populations (see materials and methods for the complete list). We also included in the analysis a single strain of *D. sechellia* and *D. teissieri*. The rationale of these experiments is that if an insertion have targeted the intron, it would have in turn increased the expected size of the amplified region.

We analysed 25 wild type strains and 4 laboratory stocks of D. melanogaster, as well as 9 wild type strains of D. simulans (see materials and methods). As shown in Figure 6, PCR amplification of *Yeti* produces a prominent band of the expected size in all the different strains of D. melanogaster (Figure 6A) and of D. simulans (Figure 6B) and in both D. sechellia and D. tessieri strains (Figure 6C). Sequencing of the purified PCR products from Iso and Scansano (D. melanogaster), Chicharo and Death Valley (D. simulans), D. sechellia and D. tessieri confirmed that they correspond to the Yeti intron-containing region (Figure 7). In conclusion, the results of this analysis suggest that the small intron of Yeti does not frequently undergo significant increase in size, unlike other essential heterochromatin genes of Drosophila [27]. This conclusion is in agreement with the observation that the Yeti gene structure tends to be stable during the evolution of the Drosophila genus, with the single intron that retains its short size (Figure 5).

Yeti is under negative selection

We next asked whether *Yeti* has evolved under negative (purifying) or positive selection, and whether the change in

	1 10	20	30	40	50	60	70	80	90	100	110	120	130
PC Dwol			+		+	TTGOCOGTGC	+	COCOOTTCCOT	+	TTTOOOCO	+	T10000000	тетет
BC Dsin		GTATCGGT	ACCATGATA	ATAAATTTG	88886888886888	TTGACAGTGC	TGAAAGGT	CACAATTGGAT	TGGAAAAAC	TTTAAACA	ACCCARCCA	AGACGAACT	TCTGT
BC-Dsec	CTGTCCTCTGAAAAAA	GTATCGGT	ACCATGATA	ATAAATTTG	AAAAGAAAAAGAAA	TTGACAGTGC	TGAAAGGT	CACAATTGGAT	TGGAAAAAC	TTTAAACA	AGACGAAGGCA	AGACGAACT	TCTGT
BC_Dyak	CCATCCGCTAAAAAAC	GTATCGGC	TCTATTATGA	ATAAACTCG	AAAAGAAAAAAAAA	TTGACAGTGC	r tgagaggt	CACAATTGGAT	TGGAAAATA	TTTAAACA	AGACGAAGGTA	rag <mark>acg</mark> aact	TCTGT
BC_Dere	CCATCCGCTGAAAAAC	GTATTGGC	TCTATTATG	ATAAACTCG	AAAAGAAAAGAAAC	TTGTCAGTAC	TGAAAGGT	CACAATTGGAT	TGGAAAATA	TTTAAACA	AGACGAAGGCA	AGACGAACT	TCTGT
BC_Deug	CGGTTAAATGAGAGTC	CTATAGGT	TCTTTTATG	ATAATTTTC	CAAAAAAAAAGAAAA	TTATCTGTGC	rggaaaagt	CTCAACTGGAT	TGGAAGAAC	TTTAAAAT	AGATGAAGGGA	AGACGAACA	GCTGC
BC_DFPF				HIHHHLIGG	HHHHHHHHHHUHHHH COOOOCOOOOOOO	I I L I L HU I UL	I GOHHHHHO I	CTCOODTCCOT	TCCOODOOC		HUHLUHHUULH CCOCCOOCCCO	HUHLUHULH	CCTCO
BC Dose		GAGTTGGC	TCTTTTCTA		GGAAAAAAAAAGAAA	TIGTCTGTGTGC	IGGAAAAGT	CGCARATGGAT	TGGAAAACT	TTTAAAAAC	ACACCAACCE	TAATGAGCA	GCTGA
BC_Dper	CCGCAAATTGGAAGT	GAGTTGGC	TCTTTTCTAR	ATCAACTGG	GGAAAAAAAAAGAAA	TTGTCTGTGC	rggaaaagt	CGCAAATGGAT	TGGAAAACT	TTTAAAAC	AGACGAAGGTA	TAATGAGCA	GCTGA
BC_Dmir	CCGCAAATTGGAAGTO	GAGTTGGC	TCTTTTCTAF	ATCAACTGG	GGAAAAAAAAAGAAA	CTGTCTGTGC	rggaaaagt	CGCAAATGGAT	TGGAAAACT	TTTAAAAAC	AGACGAAGGTA	CAATGAGCA	GCTGA
BC_Dana	CCCACAGGACTGGGT	GTGTCGGT	TCTTTCATA	ATATATTGG	GTAAAAAGAAGAAA	ATGTCTGTGT	rggaaaagt	CGCAAATGGAT	TGGAATACG	TTTAAAAAG	CGACGAAGGCA	ragatgaaga	GTTAC
BC_Dbip	CCGTCACGTGTAGTAG	GTGTCAGT	TCTCTCATA	ATACATTGG	GTAAAAAAAAGAAA	ATGTCTGTTT	rggaaaaat	CGCAAGTGGAT	TGGAATACC	TTTAAAAAG	CGACGAAGGTA	AGATGAAGA	ATTAC
BC_DW11		CCTTCCCC		HIHLILIG	66444444444444444444444444444444444444	ATCTCCCTCC	I GGHHHHHH I		TCCOOLTCC		LUHIUHUUUUH CCOOCOOCCCO	I GH I GHGGH	OCTCC
BC_DHUJ BC Duir	001010100100100	GCTTGGGC	TETTTGETCE	ATCAGCTTG	GCAAGAAGAAGAAGAAA	ATGTCAGTGT	IGGAAAAGT	CACAGCTAGAT	TGGAAGTCG	TTTAAAAAG	TGAAGAGGGCAT	CGACGAGGA	TETEE
BC_Dalb	GGAGGCGGAGGCGGC	GATTAGGT	TCATTGCTC	ATCAGCTGG	GCAAGAAGAAAAAG	CTGTCAGTAC	rggagaagt	CGCAGATGGAT	TGGAAGTCT	TTCAAGAG	CGACGAGGGCA	CGATGAGGA	GCTGC
BC_Dgri	GGCGGCGGCGGCGGCGGCG	GTTTGGGT	TCATTACTCF	ATCAGCTGG	GCAAGAAGAAGAAA	ATGTCTGTGC	rggaaaaa t	CGCAAATGGAT	TGGAAAACA	TTCAAAAG	TGACGAGGGCA	ICGACGAGGA	TCTGC
BC_Dbia	CAGCTGGGTAAAGACA	AAATCGGT	TCTCTTATT	ATAAATCAA	Gacaaaaaaaaaaa	TTGTCAGTGC	ragaaatat	CTCAGATAGAT	TGGAAAAAT	TTTAAAAAA	AAATGAGGGCA	AGACGAGAA	ATTGC
BC_Cqua		GUUTUGGH	GCHGIHIIHU	ICCCHGIIGG	GCHHGHHGHHCHHH	CTORCCOCCC	GUHHHHUH	CCORPORTICIENT	TGGHHGTCC	TTTOOCCC	CHHCGHGGGCH	CGHGGHGGH	GUIGU
BC_Haeg		CTCTCGGG	TCACTATIO	DETCOOPTOC	GCAAGAAAAACCAA	CTCACCACCC	I HUHUHHUH	CCARACTCCAC	TECOOLCEC	TTCAACCC	CACCAACCAA	TCOOCOCCO	HL TUL
BC Agan		GATTGGGG	GCAGTATTA	ATCAACTCA	GTAAGAAAAACCAG	CTGAGCACGC	rggagaaaa	CGAAGCTCGAT	TGGACGAGC	TTCAAGCG	GCAGGAGGGGCA		GCTGC
BC_Hdes	CGAACCGGTGGCGGT	GTCTGTCA	TCGGTTTTA	GTCACATTG	GTAAAAAGAATAAA	TTAAGTGTTT	GGAAAAAA	CCCATTTAGAT	TGGAGCGGA	TTTAAGGA	AAAAGAGGGCA	CGATGAGGA	ATTGC
Consensus	ccctggagg.g	gt.TcGG.	tCt.TTaa	aTaaa.T.g	g.AAaAAaAa.AAa	.Tgtc.gtgc	l gGAaAagt	C.cAa.TgGAT	TGGAAaa	TTLAAaaga	agAcGAaGGcA	f.gAcGAgga	gcTGc
	101 140	450	460	470	400		200	010	000	020	0.40	050	000
	131 140	150	160	170	180	190 7	200	210	220	230	240	250	260
BC_Dnel	GCTCGCATAACAAAG	GCAA <mark>G</mark> GACG	GGTGAGTTTC	GAAGAAGAA	Gaagaagagtattt	A AATGGA T I	AAACTTAAA	TTTATTACCCA	Atgatttag	GTATTTGG	ACCGTCAAGAC	TTTTGGAGA	GAACĊ
BC_Dsim	GCTCGCATAACAAAG	CAAGGACG	GGTGAGTTTC	GAACAAGAC	TGTATTT	A <mark>A</mark> ATAGA T I	AACTTGAA	TTTATTACCCA	ACGATTTAG	GTATTTGG	ACCGTCAAGAC	TTTTGGAGA	GAACC
BC-Dsec	GCTCGCATAACAAAG	CAAGGACG	GGTGAGTTTC	GAACAAGAC	TGTATTT	A <mark>A</mark> ATAGATA	AACTTGAA	TTTATTACCCA	ACGATTTAG	GTATTTGG	ACCGTCAAGAC	TTTTGGAGA	GAACC
BC_Dyak	GCTCCCHTHHCHHHGU	ichhgghcg	GGTAAGTICU	IGATAAACAC		AHATTGAT	HHCTATHH	TTTOTTOOCTC	ACGATTIAG	GINIIGG	ACCGCCAAGAC	TCTTGCHGH	GAHCC
BC_Dere BC Deug	GTGCACATAACAGAG	SCAAGGAAG	GGTAAGTTC-	-67688888	CTTAAATTT	6TAAAGTTI	ACTTARAA	ACCATTTTC	ATATTTAG	ATATTTGG	ACCOLCHAGAC	TTTTGCAGA	GAACC
BC_Dele	GTACACATAACAAAG	CAAAGACG	GGTAAGTTC-	-TTGCGTTC	GTTTAGTTT		ACATAAAA	TGACTTTTGCC	ATGTTCTAG	ATATTTGG	AGCGTCAGGAA	TTTTGCAGA	GAACA
BC_Dtak	GAACCCATAACAAGG	GCAAATACG	GGTAAGTTA-	GAGATCAA	GCGCAACTA	AC	-ATCCAAAT	AATTTGTTTCC	ACATTTTAG	ATATTTTG	AGCGTCAAGAG	TTTTGCAGA	GAACC
BC_Dpse	GCACTCACAATAAAGO	6CAA <mark>g</mark> gatg	GGTACGTAC-	CTATGTTC	CTTAAATTT	CACC	-ATTTCATT	AATGTTGTTTT	ATCTTTCAG	TTATTTAG	AACGGCAGGAC	ITTTTACAGA	GAACC
BC_Dper	GCACTCACAATAAAGO	icaaggatg	GGTACGTAC-	CTATGTTC	CTTAAATTT	CACC	-ATTICATI	AATGIIGTITT	ATCTTTCAG	TTATTIAG	AACGGCAGGAC	TTTTACAGA	GAACC
BC_DM1C		COOCCOOC	GETACETTCC		00TTTTT	LHLL T00C00TT(HHIGIIGIIII OOTTOOOTTCT	HILIIILHU	OTOTTTCC	HHLUULHUUHL ACCCCCAACAC	TTTTTCCCC	GCOCO
BC Dhip	GTACTCACAATAAGG	CAAGGAAG	GGTAAGTTGA	TTTAAACTC	TTTAAATAT	TACCAATT	AGATAAGT	AATTAAACTTT	TGATTT-AG	ATATTTGG	AACGCCAAGACI	TTTTGCAGA	GAACG
BC_Dwil	GAACACATAATATGG	CAAAGATG	GGTAAGCAAG	TCAAGTTGG	CTTTTACTTCAATT	GCARATATCT	ATCTATTC	TCATTTTATTT	TCACTT-AG	CTACCTGG	AACGACAGGAT	TCTTACAGC	GCACT
BC_Dnoj	AAACTCACAATAAAGO	SCAAAGATG	GGTGAGTAAT	TT	GGACTAAATT	ATGAAAATCT	FGACTGATT	GAAGTATACCT	CT-TTCCAG	TTACTTGG	AACGCCAGGAC	TTTTGCAGC	GCACC
BC_Dvir	GTACACACAACAAAG	SCAAGGATG	GGTGAGTTCI	TTTTTTTTT	GATGGCTTTGCCAT	GTTAAATTCT	FGAATAATT	TAACTTTAT-T	CG-TTGCAG	TTACTTGG	AACGCCAGGAC	TTTTGCAGC	GCACC
BC_Dalb	GCOCCCOCCOCCOCCOCCO	COORCOTC	GCTODOTTOC	HHIGHHHGI	6HHII ттсоотото	HHUHHHIIHU	TOOTOTTT	THHIGHUILII	IGCI I HUHG		HULULLHUUHL	TCCTCCOOC	GCOCC
BC Dbia	GAACTTATAACAAAAG	SCAAGGACG	G			nninuniiii				ATACTTGG	ACCECCAAEAT	TTTTGGAAA	GAACA
BC_Cqua	AGACGCACAATCGTG	CCGGGATG	GG							-TACTTGG	AGAGACAGGAC	TTTTGCAGC	GCGCC
BC_Aaeg	AGACGCACAATAAAG	GCAAAGACG	GT							-TACCTGG	AAAGGCAAGAC	TCCTACAGA	GGGCT
BC_Adar	AAACGCACAACAAGG	icaaggatg	GC							-TTCCTCG	AGCGGCGCGAT	TTCTCGAGC	GAACG
BC_Hgan	HHHUGUHUHHUHHUGU	CHHGGHHG	66 CO								HUCGUCUCUHI	TTCTCCOOC	GIHUI
BL_NOES	R aC. CAcAAcAaaG	SCAAøGA.G	Get agtt		att	a. t	a.t.a.		tt ag	Tatt Tob	нононскионт А.сб. САабАс]		GaACc
consciisas	S+de+enemiendud	icinigan.•a	050+0500+4	********	***********	••••			****************	+ Tucci su		Teersensu	uunee
	261 270	280	290	300	310	320							
		+	+			1							
BC_Dmel		HHHIGGHH	HHUHHUIIU		GCOCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC								
BC_DSIN	GATCTTCGGCAGTTT	GAATTGAA	AAGAAGTTGI	TOTOTOTOTO C	GCAGGCCATACTAA								
BC_Dyak	GATCTTAGACAGTTT	GAAATTGAG	AAGAACTTGO	GGCTTTCTC	GCAGACCATACTAA								
BC_Dere	GATCTTAGGCAGTTT	GAAATTGAG	AAGAACTTGO	GCCAGTCTC	GCAGACCATACTAA								
BC_Deug	GATCTAAGACAGTTT	GAAATTGAG	AAGAACTTGO	GGCACCCTA	GCAGGTCAAATTGA								
BC_Dele	GHT (TTCGTCAGTTTC	HHATTGAG	HHGAAATTGO	GOCONTOTO	GIUGGCCAAATTGA								
BC_Dtak		COOTTCOC	HANNAL I HU	CTCOOOCCC	ULHUUILHHHIIGH CCCCCCQQQQCTQQ								
BC_Dpse	GATTTGCGACAATTT	AAATTAAA	AAAAACTTAC	GTCAAACGC	GCCGGCAAAACTAA								
BC_Dmir	GATTTGCGACAATTTC	GAAATTGAG	AAAAACTTAC	GTCAAACGC	GCCGGCAAAACTAA								
BC_Dana	GATATGCGTCAGTTTC	GAAATTGAG	AAGAGCTTG	GACAATCAC	GTCGGAAAAAATA								
BC_Dbip	GATTTGCGTCAGTTTC	GAAATCGAG	AAGAGCTTG	GACAATCAC	GACGGAATAATTGA								
BC_Uwil	UNITIGHUUUUUUUTCO	HUHIIUHG	HHHHHICIAC	COCOCOCCC	COCCTCOCCOCCTCO								
BC_DHOJ BC Dvir	GATGTGCGGCGATTT	GAGATCGAG	AAAAACTTG	CIACONICCION CONTRACTOR	GACGTTCAAACTAA								
BC_Dalb	GATCTGCGCCAGTTT	GAGATTGAG	AAAGGCATG	GACAATCGC	GTCGACAAAACTAA								
BC Deri	GATTIGCGGCGATTIC	AGATCGAG	AAAAACATGO	GTCAAACGC	GGCGTCAAAACTGA								

 BULUgri
 GATTIGCGGCANITIGAGATCGAGAAAAACATGCGTCAAACGCGGCGTCAAAACGAACGAA GATCTGCGACAGTTTGAAATCGAGAAGAAGCTTTAGGCAGTCCAAGAGCAAGCGATGA GATGTGCGACAGTTCGAAATAGAAAAATCCTTCCGTCAGACCACCAGGAGTAATCGATAA GATGTGCGGCAGTTTGAGATTGAAAAATCATTCCGACAGTCCAAACGGAGCAACCGATAG GATGTGCGGCAGTTCGAGATTGAAAAATCATTCCGACAGTCCAAACGGAGCAACCGATAG GATTTGCGTCAGTTCGAGATAGAGAAATCGTTCCGGCAAACGAAGCGAAGCAACTAA GATTIGAGACGTTTTGAAATTGAAAAAAATATGCGACAAACCACACGACAAAGTGA GATT,gcG,CAgTTtGAAATtGAAAAAAATATGCGACAAACCACACGACAAAGTGA GAT,TgcG,CAgTTtGAaATtGAgAAaaactgCG,CaatC,cg,cGgc,aaActaA... BC_Hdes

Consensus

Figure 3. Alignment of the Yeti ortholog sequences encoding the BCNT-C domain. The grey area corresponds to the intron present in the Drosophila species.

doi:10.1371/journal.pone.0113010.g003

location from euchromatin to heterochromatin may have affected the molecular evolution of the gene. To this aim we performed a $d_{\rm S}$ - $d_{\rm N}$ tests, a codon-based test of selection (for details see material and methods), using DNA sequences from five representative Yeti genes: three located in heterochromatin (Dmel\Yeti, Dsec\Yeti and Dere \Yeti) and two in euchromatin (Dpse \Yeti and Dvir \Yeti). The results suggest that Yeti is under purifying (negative) selection when present both in heterochromatin and euchromatin. Thus, the change in genomic location does not appear to have affected significantly the molecular evolution/function of Yeti (Table 2).

Discussion

In this paper, we have studied the evolutionary origin of Yeti, an essential gene of Drosophila melanogaster (Figure 1) located in the

	1	10	20	30	40	50.	60	70	81
	I	+	+	+	+	+-X	+	+	+1
Dmel_pro	SSAEKSI	[<mark>GTMIN</mark> KF	EKKKKLTYLE	RSQLDHKIFKO	DEGIDELLC	SHNKGKDGYLD	RQDFLERTO		L <mark>SRR</mark> PY
Dsi n_ pro	LSSEKG]	[<mark>gthin</mark> kf	EKKKKLTYLE	r <mark>sqldhknfk</mark> g	DEGIDELLC	SHNKGKDGYLD	RQDFLERTO		L <mark>SRR</mark> PY
Dsec_pro	LSSEKG]	[<mark>gthin</mark> kf	EKKKKLTYLE	rs <mark>qldhknfk</mark> (DEGIDELLC	SHNKGKDGYLD	RQDFLERTO		L <mark>SRR</mark> PY
Dyak_pro	PSAKNG]	[<mark>gsimnkl</mark>	EKKKNLTYLE	rsqldhkifk(DEGIDELLC	SHNKGKDGYLD	RQDFLQRTO		L <mark>SRR</mark> PY
Dere_pro	PSAENG]	[<mark>gsimnkl</mark>	EKKRNLSYLE	rs <mark>qldhkifk</mark> (DEGIDELLC	SHNKGKDGYLD	RQDFLQRTO	LRQFEIEKNLR	QSRRPY
Dbia_pro	QLGKDK]	[<mark>gslin</mark> ks	ROKKKLSYLE	ISQIDHKNFKK	KNEGIDEKLR	TYNKGKDGYLD	RODFLERTO	LRQFEHEKNYR	OSRRLN
Deug_pro	RLNESAJ	[<mark>gs</mark> fmnnf	PKKRKLSYLE	KSQLDHKNFK]	[DEGIDEQLR	AHNRGKEGYLE	RQDFLQRTD	LROFETEKNLR	HPS <mark>RSN</mark>
Dtak_pro	PYNEDKI	r <mark>gs</mark> lf n ks	RKEKKTSYLE	KSQMDHKNFKS	DEGIDEQLR	THNKGKYGYFE	RQEFLORTO	LROFETEKNLR	OSRRSN
Dele_pro	PLHQS <mark>g</mark> f	A <mark>gslidkl</mark>	EKKRKFSYLE	KSQMDWKNFKF	RDEGIDEQLR	THNKGKDGYLE	RQEFLQRTO	FROFETEKKLR	OSRRPN
Dana_pro	PTGLGG	/GSFINIL	GKKKKMSYLE	KSQMDHNTFKS	DEGIDEELR	THNKGKEGYLD	RODFLORTO	MRQFEIEKSLR	OSRRKK
Dbip_pro	PSRYY <mark>G</mark> Y	/S <mark>SLINTL</mark>	GKKKKMSYLE	KSQYDHNTFKS	DEGIDEELR	THNKGKEGYLE	RODFLORTO	LRQFEIEKSLR	OSRRNN
Dpse_pro	PQIGSG	/GSFLNQL	GKKKKLSYLE	KSQHDHKTFKT	DEGINEQLS	THNKGKDGYLE	RODFLORTO	LROFETEKNLR	OTRRON
Dper_pro	POIGSG	GSFLNQL	GKKKKLSYLE	KSOMDHKTFKT	DEGINEOLS	THNKGKDGYLE	RODFLORTO	LROFEIEKNLR	OTRRON
Dmir_pro	POIGSG	GSFLNQL	GKKKKLSYLE	KSQMDHKTFK1	DEGINEQLS	THNKGKDGYLE	RODFLORTO	LROFEIEKNLR	OTRRON
Dwil_pro	PANGGGL	GSFLNTL	GKKKKMSYLE	KSOLDHKTFKK	DEGIDEELR	THNMGKDGYLE	RODFLORTO	LROFEIEKNLR	OSRRON
Dno.j_pro	GYYGG <mark>G</mark> L	GSLLNOL	GKKKKMSYLE	KSQLDHKSFKS	SEEGIDEELQ	THNKGKDGYLE	RODFLORTO	LROFEIEKNLR	OTRRON
Dvir_pro	GGGGG <mark>G</mark> L	GSLLNOL	GKKKKMSYLE	KSOLDHKSFKS	SEEGIDEDLR	THNKGKDGYLE	RODFLORTO	VROFEIEKSLR	OTRRSN
Dalb_pro	GGGGG <mark>G</mark> L	GSLLNOL	GKKKKLSYLE	KSOMDHKSFKS	DEGIDEELR	THNKGKDGYLE	RODFLORTO	LROFEIEKGHR	OSRRON
Dgri_pro	GGGGG <mark>G</mark> L	GSLLNOL	GKKKKMSYLE	KSOMDAKTEKS	DEGIDEDLR	THNKGKDGFLE	RODFLORTO	LROFEIEKNMR	OTRRON
Cqua_pro	RSSGGGL	GAYLGOL	GKKNKLSYLE	KTOLDHKSFKF	RNEGIEEELQ	THNRGRDGYLE	RODFLORAD	LROFEIEKNER	OSKSKR
Aaeg_pro	RPSGG<mark>G</mark>L	GAYLGOI	GKKNKLSTLE	KTKLDHNSFKF	RSOGIEEELO	THNKGKDGYLE	RODFLORAD	VROFEIEKSFR	OTTRSN
Adar_pro	PARSGGL	GSYLNOI	GKKNOLSTLE	KTKLDHNRFKF		THNKGKDGFLE	RRDFLERTO	VROFETEKSFR	OSKRSN
Agan_pro	RSSSGGL	GAYLNOL	SKKNOLSTLE	KTKLDHTSFKF		THNKGKEGFLE	RRDFLERTO	LROFEIEKSFR	OTKRSN
Mdes_pro	RTGGG <mark>G</mark> L	_SSYLGHI	GKKNKLSYLE	KTHLDHSGFKE	KEGIDEELO	THNKGRDGFLE	RODFLERTO	LRRFEIEKNMR	OTTRRK
Consensus	•••• g	gsn.l	gKKkklsvLE	ksq10Wk.FK.	dEGI#E.L.	tHNkGk#G%L#	RQ#FL#RtD	1RQFEiEKn1R	qsr <mark>R₊n</mark>

Figure 4. Alignment of the BCNT domain of YETI proteins among species. The arrow points the intron position in the corresponding coding region of *Drosophila* species.

doi:10.1371/journal.pone.0113010.g004

deep heterochromatin of chromosome 2 [25,26,28,29] and found that it has evolved from a euchromatic ancestor in drosophilids.

Our FISH analysis shows that Yeti maintains a heterochromatin location in 2Rh, at the base of division 41A, in both D. simulans and D. sechellia sibling species of D. melanogaster (Figure 2). The FlyBase localization of Yeti in D. sechellia scaffold_170:38,673..3,460), is in accord with our mapping results, while D. simulans Yeti is reported to map to the 3R arm (3,765..73,764; Table 1) in a 70 kb gene-poor genomic region. However, we are confident about the heterochromatin location of Yeti in D. simulans for the following reasons: First, as discussed in the result section, the FISH signal morphology produced by the *Yeti* probe is different from that usually seen with euchromatic probes and represents a distinctive mark for sequences derived from polytenized heterochromatin [30,31]. Second, our FISH mapping of *Yeti* in *D. simulans* is based on three reproducible experiments, each carried out on several polytene chromosome figures obtained from at least 10 larvae. Finally, the paucity of genes around D. simulans Yeti is per se highly suggestive of heterochromatin localization. Thus, it is possible that the apparent discrepancy between our data and FlyBase may reflect an assembly error that occurred in the D. simulans genome sequence assembly, as reported by Schaffer et al. [32].

Our FISH analysis show that in two distantly related species, *D. pseudoobscura* and *D. virilis*, *Yeti* is located in euchromatin (Figure 2). In *D. pseudobscura, Yeti* maps to chromosome 3 at polytene division 63C, while in *D. virilis* it is found in chromosome 5, at polytene division 53E. Interestingly, *Yeti* lies in the syntenic chromosome Muller C that corresponds to the 2*R* arm of *D. melanogaster* chromosome 2. Together, the results of our analysis indicate that during the evolution of the *Drosophila* genus, *Yeti* has been resident on the same chromosomal element, but over time it progressively moved closer to the pericentric regions. Such movements would have occurred in about 40 million of years, the estimated divergence time between *D. melanogaster* and *D. virilis* (Figure 8). A similar evolutionary trend was reported for *light* and other neighboring genes in 2L heterochromatin [23] and for other genes of chromosome 3 heterochromatin [24].

A	ГG	S	ГОР
§	635	69 91	D.melanogaster
§	635	62 91	D.simulans
§	635	62 91	D.sechellia
§	557	62 91	D.yakuba
§	620	62 91	D.erecta
¥	617	60 91	D.eugracilis
≠	** 600	51 91	D.biarmipes
¥	** 616	55 91	D.takahashii
≠	** 643	54 91	D.elegans
§	620	61 91	D.ananassae
≠	608	63 91	D.bipectinata
§	737	57 91	D.pseudoobscura
§	737	57 91	D.persimilis
≠	737	57 91	D.miranda
§	155 75 576	70 91	D.willistoni
§	797	59 91	D.mojavensis
§	812	69 91	D.virilis
¥	749	62 91	D.albomicans
§	757	66 91	D.grimshawi
§	855		C.quinquefasciatus
§	840		A.aegypti
≠	882		A.darlingi
§	882		A.gambiae
≠	903		M.destructor

Figure 5. Comparison of the Yeti gene structure among sequenced genomes. Only the coding regions are showed. Exons are in boxes and numbers refer to nucleotides. Symbols: §, annotated genes; \neq , this study; **, defective ORF. The grey area at the 3' end represents the conserved BCNT-C domain in the protein. doi:10.1371/journal.pone.0113010.g005



Figure 6. PCR amplification of the genomic region containing the *Yeti* **intron in** *Drosophila* **species.** A single PCR product of about 180 bp was found, in *D. melanogaster* (A) and *D. simulans*, (B) *D. sechellia* (C) and *D. tessieri* (C) related species. M = Marker; Frib = Friburgo; Mal = Mali; Scan = Scansano; Iso = y^1 ; cn^1 bw¹sp¹ isogenic strain; OR = Oregon-R; Bej = Bejin; Chi = Chicharo; DV = Death Valley; Gen = Genoa; Mor = Moruya; Kyo = Kyogle; Arm = Armidale; Can = Canaries; Sech = D. Sechellia; tes = D. tessieri. Molecular weight is in bps. doi:10.1371/journal.pone.0113010.q006

A striking difference between heterochromatin and euchromatin genes lies in the generally larger size and complex molecular structure of the former compared to the latter. The example of the 'giant' Y-chromosome fertility factors of *D. melanogaster* mentioned above is paradigmatic in this respect [12]. Some of the essential heterochromatin genes of chromosomes 2 and 3 are also large due to the presence of large introns that harbour truncated TE copies (or TE "remnants") [27,35,36,37]. In this context, the *Yeti* gene of *D. melanogaster* with a 900 bp-long genomic region represents an exception [26]. The same is true for *RpL38*, *RpL5* and *RpL15*, three essential ribosomal protein-coding genes located in the heterochromatin of chromosomes 2 and 3, all of which are of relatively small size [24,26,37,38]. How might these observations be explained?

	1	10	20	30	40	50	60	70	80	90 1
Yeti-gDNA/1-134 Scansano-F/1-164 Chicharo-F/1-154 Death-Yalley-F/1-155 Sechellia-F/1-155 Tessieri-F/1-155 Consensus	GTGCAG GGGC-G GGTCT GGTG GGTCT	CATTCTT CGTTCTT CGTTCTT CGTTCTT CGTTCTT CGTTCTT c.ttctt	TA-TTTAAACT TTCTTTAAACT T-CTTTAAACT T-CTTTAAACT T-CTTTAAACT T-CTTTAAACT ttttaact	TAAGATCG GCCTAGATCG GCCGAGA-CG GCCGAGA-CG GCCGAGA-CG GCCGAGA-CG GCCGAGA-CG gcc.AGALCG	GTTCTCTCCA GTTCTCTCCA GTTCTCTCCA GTTCTCTCCA GTTCTCTCCA GTTCTCTCCA GTTCTCTCCA	AAAAGTCTTG AAA-GTCTTG AAA-GTCTTG AAA-GTCTTG AAA-GTCTTG AAA-GTCTTG AAA-GTCTTG	ACGGTCCAAA ACGGTCCAAA ACGGTCCAAA ACGGTCCAAA ACGGTCCAAA ACGGTCCAAA ACGGTCCAAA	IACCTAAATC IACCTAAATC IACCTAAATC IACCTAAATC IACCTAAATC IACCTAAATC IACCTAAATC	ATTGGGTAA -TTGGGTAA GTTGGGTAA GTTGGGTAA GTTGGGTAA GTTGGGTAA .TTGGGTAA	TAAATTT TAAATTT TAAATTC TAAATTC TAAATTC TAAATTC TAAATTC TAAATTC
	91 	100	110	120	130	140	150	160	167 	
Yeti-gUNH/1-134 Scansano-F/1-164 Chicharo-F/1-154 Death-Valley-F/1-155	AAGTTT AAGTTT AAGTTT	ATCCATT ATCCATT ATCTATT	THAHTHCTCTT TAAATACTCTT TAAATACA		TTCCAAACTC TTCCAAACTC TTCCAAACTC	ACCCGTCCTT	GCCTTTGTTA GCCTTTGTTA GCCTTTGTTA	I GCGHG I GCGAGCACA I GCGAGCACA I GCGAGCACA	AAA AA	
Sechellia-F/1-155 Tessieri-F/1-155 Consensus	AAGTTT AAGTTT AAGTTT	ATCTATT ATCTATT ATCCATT	TAAATACA Taaataca Taaatac <mark>eet</mark>	GTCTTG	TTCCAAACTC TTCCAAACTC TTCCAAACTC	ACCCGTCCTT ACCCGTCCTT ACCCGTCCTT	GCCTTTGTTA GCCTTTGTTA GCCTTTGTTA	IGCGAGCACA IGCGAGCACA IGCGAGcaca	AAA AA ••••	

Figure 7. Sequencing of the purified PCR products from *Drosophila* **species.** Sequence alignments from Iso and Scansano (*D. melanogaster*), Chicharo and Death Valley (*D. Simulans*), *D. sechellia* and *D. tessieri*. Sequence analysis confirmed that they correspond to the *Yeti* intron containing region. The *Yeti* intron is shown in normal text, the flanking exons are in bold. The *D. simulans* and *D. sechellia* intron lacks a 7 bp stretch (see the gap), in agreement with the genome sequence data (see results and Figure 5). The *D. Tessieri* intron sequence is identical to that of *D. simulans* and *D. sechellia*.

doi:10.1371/journal.pone.0113010.g007

Number of species compared	Number of codons ^a	d _s - d _N	P ^b						
5	232	10.2	<10 ⁻⁶						
3 (heterochromatic sequences)	233	5.3	<10 ⁻⁶						
2 (euchromatic sequences)	275	8.7	<10 ⁻⁶						

^aAll positions containing gaps were eliminated.

^bThe probability (P) of rejecting the null hypothesis of strict-neutrality ($d_N = d_S$) in favor of the alternative hypothesis (purifying selection, dN<dS) is shown. doi:10.1371/journal.pone.0113010.t002

One may imagine that, during evolution, genes increased their size by becoming targets for reiterated transposable-element insertions in the intronic regions, depending on their time of residence in heterochromatin. This, however, does not seem to have been the case. In fact, the *light* and *Yeti* genes, although having likely resided in heterochromatin for a comparable evolutionary time (less than 30 million of years), underwent a different molecular architecture; the *light* gene structure dramatically changed during the evolutionary transition from euchromatin to heterochromatin, due to a remarkable increase in the size of introns targeted by TEs [23]; Yeti retained its original organization in all analysed species, with a short genomic region carrying a single short intron (Figure 5). In addition, by PCR analysis we found that the Yeti intron does not undergo significant interspecies or intraspecies changes of its physical size (Figure 6). Similarly, the RpL15 gene shows a conserved structure among Drosophila species, independently from its genomic location [24].

How to explain the different behaviour of heterochromatin genes? In particular, there might have been a selective pressure to maintain some genes of short size (with few, short introns) despite of their genomic location, owing to their particular functional requirement: interestingly in that respect, highly expressed genes are known to harbour substantially shorter introns than genes that are expressed at low levels [39]. This may be the case of *Yeti* and *RpL38*, *RpL5* and *RpL15* heterochromatic genes of *Drosophila melanogaster*, which are all highly expressed and all have indeed short size and carry short introns [26,37,38]. *Yeti* itself encodes an important chromatin-remodeling factor required for development [26] and ribosomal protein coding genes are also essential for



Figure 8. Evolutionary repositioning of the *Yeti* **gene.** Schematic representation of *Yeti* gene transition of from euchromatin to heterochromatin. The arrows point the chromosomal position of *Yeti*. It appears that *Yeti* has been resident on the Muller C chromosomal element, but over evolutionary time it progressively approached to pericentric heterochromatin and in *D. melanogaster* it is found in the deep portion of *2Rh*.

doi:10.1371/journal.pone.0113010.g008

proper development. It is not unreasonable to speculate that these genes maintained the original structure, in spite of their transition to heterochromatin during massive chromosomal rearrangements that occurred over time, because of the requirement for their efficient expression during early development. In addition, it is possible that once in heterochromatin, a given sequence might be differentially targeted by transposable elements, with some sequences being more refractory than others. These observations suggest that the evolutionary forces that acted in shaping the structural organization of genes currently found in *D. melanogaster* heterochromatin are molecularly diverse.

Finally, the results of our $d_S - d_N$ tests, showing that *Yeti* is under negative selection both in heterochromatin and euchromatin (Table 2), are in accord with its evolutionary conserved function and suggest that the change in genomic location did not affected significantly the molecular evolution of the gene.

Together, these results contribute to expand our understanding of the molecular dynamics driving the evolution of the heterochromatin genome in higher eukaryotes.

Materials and Methods

Drosophila strains

Fly cultures were carried out at 25°C in standard commeal yeast medium. D. melanogaster strains derived from natural populations: Altamura I and Altamura II (South Italy), Beijing (Cina), California and California I (USA), Charolles (France), Charolles 1999 (France), Friburgo 1997 (Germany), HJ30 (France), Hobo (Grece) Israel-4 (Israel), Lanuvio (Italy), Luminy (France), Mali and Mali I (West Africa), Marrakesh (North Africa), Scansano (Central Italy), Vallecas (Spain), W15 (USA), W30 (USA), W90 (USA), W130 (USA), W135 (USA), Zimbabwe (Africa), Gaiano (North Italy). D. melanogaster laboratory strains (separated by comma): isogenized y^{I} ; cn^{I} , bw^{I} , sp^{I} strain (Iso) [40]. $l(2)LP2/SM1, Cy. Cy/Sp; Sb, Delta 2-3, ry^{506}/TM6, ry^{506}$. l(2)EMS-31/SM1, Cy. D. simulans strains derived from natural populations: Armidale (New South Wales, Australia), Canaries (Atlantic Spain), Can River (Australia) Chantal (France), Chicharo (Portugal), Death Valley (USA), Genoa (Italy), Kyogle (Australia) and Moruya (New South Wales, Australia). D. melanogaster and D. simulans wild-type strains were derived from natural populations collected in the wild before the end of year 2000 and are gifts of Sergio Pimpinelli, Nikolaj Junakovic, Chantal Vaury and Pierre Capy.

Cytology and fluorescent in situ hybridization

Polytene chromosomes prepared according to Pardue [41] were stained with DAPI. The *D. melanogaster* RE36623 cDNA *Yeti* probe was labelled by nick-translation with Cy3-dCTP (Amersham). Species-specific PCR probes were used for FISH in D. pseudobscura and D. virilis (see below). FISH procedures were performed according to Dimitri [42]. Digital images were obtained using an Olympus epifluorescence microscope equipped with a cooled CCD camera. Gray scale images, obtained separately recording Cy3 and DAPI fluorescence by specific filters, were pseudo colored and merged for the final image using the Adobe Photoshop software.

Nucleic acid manipulation and sequence alignments

Genomic DNA extraction and PCR were performed according to the protocol of Berkeley Drosophila Genome Project (http:// www.fruitfly.org/about/methods/index.html). To PCR amplify *Yeti* genomic fragments from *D. melanogaster*, *D. simulans* strains, *D. sechellia* and *D. tessieri* the following primers were used: F: 5'-TTAGTATGGCCTGCGAGACA-3'; R: 5'-TGTGCTCGCA-TAACAAAGGC-3'.

PCR cycle were: 40" at 98°C, 40 x (10" 98°C, 30" 58°C, 8" 72° C). Amplified fragments were gel purified and sequenced by Bio-Fab research s.r.l. The *Yeti* probes from *D. pseudoobscura* and *D. virilis* were generated by PCR over genomic DNA with the following primers:

Dpse-F 5'-GCGACGATGATAGCATCAAT-3'; Dpse-R 5'-GTGAGTGCTCAGCTGCTCAT-3'

Dvir-F 5'-AGCTAAACGTAGCACGCGTC-3'; Dvir-R 5'-TGTGTACGCAGATCCTCGTC-3'

PCR cycle were: 4' at 94° C, $35 \times (30'' 95^{\circ}$ C, $45'' 60^{\circ}$ C, $30'' 72^{\circ}$ C). Amplified fragments were cloned in pGEM-T vector (Promega) and verified by DNA sequencing.

Multiple sequence alignment were performed by ClustalW procedure available at EMBL-EBI (http://www.ebi.ac.uk) or with the multialin interface at http://multalin.toulouse.inra.fr.

References

- Moritz KB, Roth GE (1976) Complexity of germline and somatic DNA in Ascaris. Nature 259: 55–57.
- Peterson DG, Pearson WR, Stack SM (1998) Characterization of the tomato (Lycopersicon esculentum) genome using in vitro and in situ DNA reassociation. Genome 41: 346–356.
- Dimitri P, Caizzi R, Giordano E, Carmela Accardo M, Lattanzi, et al. (2009) Constitutive heterochromatin: a surprising variety of expressed sequences. Chromosoma 118: 419–35.
- 4. Heitz E (1928) Das Heterochromatin der Moose. Jb Wiss Bot 69: 762-818.
- 5. John B (1988) The biology of heterochromatin. In Heterochromatin: Molecular
- and Structural aspects ed. V. R. S., Cambridge University Press. p. 1–128.
 6. Williams SM, Robbins LG (1992) Molecular genetic analysis of Drosophila rRNA arrays. Trends Genet 8: 335–340.
- Dernburg AF, Sedat JW, Hawley RS (1996). Direct evidence of a role for heterochromatin in meiotic chromosome segregation. Cell 86: 135–146.
- Elgin SCR (1996) Heterochromatin and gene regulation in *Drosophila*. Current Opinion in Genetics and Development 6: 193–200.
- Karpen GH, Le MG, Le H (1996) Centric heterochromatin and the efficiency of achiasmate disjunction in *Drosophila* female meiosis. Science 273: 118–122.
- Pardue ML (2007) Following the chromosome path to the garden of the genome. Annu Rev Cell Dev Biol. 2007; 23: 1–22.
- Villasante A, Mendéz-Lago M, Abad JP, Montejo de Garcini E (2007) The birth of the centromere. Cell Cycle 6: 2872–2876.
- Gatti M, Pimpinelli S (1992). Functional elements in Drosophila melanogaster heterochromatin. Annu Rev Genet 26: 239–275.
- Hilliker AJ (1976) Genetic analysis of the centromeric heterochromatin of chromosome 2 of *Drosophila melanogaster*: deficiency mapping of EMS-induced lethal complementation groups. Genetics 83: 765–782.
- Marchant GE, Holm DG (1988) Genetic analysis of the heterochromatin of chromosome 3 in *Drosophila melanogaster*. II. Vital loci identified through EMS mutagenesis. Genetics 120: 519–532.
- Dimitri P (1991) Cytogenetic analysis of the second chromosome heterochromatin of *Drosophila melanogaster*. Genetics 127: 553–564.
- Koryakov DE, Zhimulev IF, Dimitri P (2002) Cytogenetic analysis of the third chromosome heterochromatin of *Drosophila melanogaster*. Genetics 160: 509– 517.
- Fitzpatrick KA, Sinclair DA, Schulze SR, Syrzycka M, Honda BM (2005) A genetic and molecular profile of third chromosome centric heterochromatin in *Drosophila melanogaster*. Genome 48: 571–584.
 Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, et al. (2000) The

genome sequence of Drosophila melanogaster. Science 287: 2185-2195.

Evolution of the Drosophila Yeti Gene

Testing for signatures of selection

A codon-based test of selection was conducted in MEGA 6.0 [43], using the Nei-Gojobori method [44]. The statistic test $(d_S - d_N)$ is expected to be zero under the null hypothesis of neutrality. d_S and d_N are the numbers of synonymous and nonsynonymous substitutions per site, respectively. Three analyses were performed. The first analysis involved five coding sequences of the *Yeti* gene from five different Drosophila species (*D. melanogaster*, *D. simulans*, *D. erecta*, *D. pseudoobscura* and *D. virilis*). The other two analyses involved either the three Drosophila species where the *Yeti* gene is heterochromatic (*D. melanogaster*, *D. simulans* and *D. erecta*) or the two species where the *Yeti* gene is euchromatic (*D. pseudoobscura* and *D. virilis*).

Acknowledgments

We are grateful to Patrizia Lavia for critical reading of the manuscript. We also thank Sergio Pimpinelli, Chantal Vaury and Pierre Capy for gift of *Drosophila* strains.

Author Contributions

Conceived and designed the experiments: RC EC FC RM PD. Performed the experiments: RC EC FC RM PD. Analyzed the data: RC EC FC RM PD. Contributed reagents/materials/analysis tools: RC FC PD. Wrote the paper: PD.

- Hoskins RA, Smith CD, Carlson JW, Carvalho AB, Halpern A, et al. (2002) Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. Genome Biology 3: research0085.1–0085.16.
- Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, et al. (2007) Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. Science 316: 1625–1628.
- Smith CD, Shu S, Mungall CJ, Karpen GH (2007) The Release 5.1 annotation of *Drosophila melanogaster* heterochromatin. Science 316: 1586–1591.
- He B, Caudy A, Parsons L, Rosebrock A, Pane A, et al. (2012) Mapping the pericentric heterochromatin by comparative genomic hybridization analysis and chromosome deletions in Drosophila melanogaster. Genome Res 22: 2507– 2519.
- Yasuhara JC, DeCrease CH, Wakimoto BT (2005) Evolution of heterochromatic genes of *Drosophila*. Proc Natl Acad Sci USA 102: 10958–10963.
- Schulze SR, McAllister BF, Sinclair DA, Fitzpatrick KA, Marchetti M, et al. (2006) Heterochromatic genes in Drosophila: a comparative analysis of two genes. Genetics 173: 1433–1445.
- Cenci G, Belloni G, Dimitri P (2003) 1(2)41Aa, a heterochromatic gene of Drosophila melanogaster, is required for mitotic and meiotic chromosome condensation. Genet Res 81: 15–24.
- Messina G, Damia E, Fanti L, Atterrato MT, Celauro E, et al. (2014) Yeti, an essential Drosophila melanogaster gene, encodes a protein required for chromatin organization. J Cell Sci 127: 2577–2588. doi:10.1242/jcs.150243 Epub 2014 Mar 20.
- Dimitri P, Junakovic N, Arcà B (2003) Colonization of heterochromatic genes by transposable elements in *Drosophila*. Mol Biol Evol 20: 503–512.
- Dimitri P, Corradini N, Rossi F, Verni F, Cenci G, et al. (2003) Vital genes in the heterochromatin of chromosomes 2 and 3 of Drosophila melanogaster. Genetica. 117: 209–215.
- Rossi F, Moschetti R, Caizzi R, Corradini N, Dimitri P (2007) Cytogenetic and molecular characterization of heterochromatin gene models in Drosophila melanogaster. Genetics 175: 595–607.
- Zhang P, Spradling AC (1995) The Drosophila salivary gland chromocenter contains highly polytenized subdomains of mitotic heterochromatin. Genetics. 1995 Feb; 139(2): 659–70.
- Berghella L, Dimitri P (1996) The heterochromatic rolled gene of Drosophila melanogaster is extensively polytenized and transcriptionally active in the salivary gland chromocenter. Genetics. 1996 Sep; 144(1): 117–25.
- Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, et al. (2008) Polytene chromosomal maps of 11 Drosophila species: the order of genomic scaffolds inferred from genetic and physical maps. Genetics 179: 1601–55.

- Waterhouse RM, Zdobnov EM, Tegenfeldt F, Li J, Kriventseva EV (2011) OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. Nucleic Acids Res. 39(Database issue): D283–8.
- Iwashita S, Osada N, Itoh T, Sezaki M, Oshima K, et al. (2003). A transposable element-mediated gene divergence that directly produces a novel type bovine Bent protein including the endonuclease domain of RTE-1. Mol Biol Evol 20: 1556–1563.
- Devlin RH, Bingham B, Wakimoto BT (1990a) The organization and expression of the *light gene*, a heterochromatic gene of *Drosophila melanogaster*. Genetics 125: 129–140.
- Tulin A, Stewart D, Spradling AC (2002) The *Drosophila* heterochromatic gene encoding poly (ADP-ribose) polymerase (PARP) is required to modulate chromatin structure during development. Genes and Development 16: 2108– 2119.
- Schulze SR, Sinclair DA, Fitzpatrick KA, Honda BM (2005) A genetic and molecular characterization of two proximal heterochromatic genes on chromosome 3 of *Drosophila melanogaster*. Genetics 169: 2165–2177.
- Marygold SJ, Coelho CM, Leevers SJ (2005) Genetic analysis of *RpL38* and *RpL5*, two minute genes located in the centric heterochromatin of chromosome 2 of *Drosophila melanogaster*. Genetics 169: 683–695.
- Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA (2002) Selection for short introns in highly expressed genes. Nat Genet 31: 415–418.

- Brizuela BJ, Elfring L, Ballard J, Tamkun JW, Kennison JA (1994) Genetic analysis of the brahma gene of Drosophila melanogaster and polytene chromosome subdivisions 72AB. Genetics 137: 803–813.
- Pardue ML (1986) In situ hybridization to DNA of chromosome and nuclei. In: Roberts DB (ed) Drosophila: a practical approach. IRL Press, Oxford, pp. 111– 137.
- Dimitri P (2004) Fluorescent in situ hybridization with transposable element probes to mitoticchromosomal heterochromatin of Drosophila. Methods Mol Biol 260: 29–39.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution 30: 2725–2729.
- Nei M, Gojobori T (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3: 418–426.
- Coulthard AB, Alm C, Cealiac I, Sinclair DA, Honda BM, et al. (2010) Essential loci in centromeric heterochromatin of Drosophila melanogaster. I: the right arm of chromosome 2. Genetics 185: 479–495. doi:10.1534/genetics.110. 117259 Epub 2010.
- Timoshevskiy VA, Kinney NA, deBruyn BS, Mao C, Tu Z, et al. (2014) Genomic composition and evolution of Aedes aegypti chromosomes revealed by the analysis of physically mapped supercontigs. BMC Biol 12: 27. doi:10.1186/ 1741-7007-12-27.