HAL
open science

Conserved DNA motifs, including the CENP-B Box-like, are possible promoters of satellite DNA array rearrangements in nematodes

Nevenka Mestrovic, Martina Pavlek, Ana Car, Philippe Castagnone-Sereno, Pierre Abad, Miroslav Plohl

To cite this version:

Nevenka Mestrovic, Martina Pavlek, Ana Car, Philippe Castagnone-Sereno, Pierre Abad, et al.. Conserved DNA motifs, including the CENP-B Box-like, are possible promoters of satellite DNA array rearrangements in nematodes. PLoS ONE, 2013, 8 (6), ⟨10.1371/journal.pone.0067328⟩. ⟨hal-02645975⟩

HAL Id: hal-02645975

https://hal.inrae.fr/hal-02645975v1

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Autorisation HAL

HAL Authorization

Conserved DNA Motifs, Including the CENP-B Box-like, Are Possible Promoters of Satellite DNA Array Rearrangements in Nematodes

Nevenka Meštrović[1]*[9], Martina Pavlek[1][9], Ana Car[1], Philippe Castagnone-Sereno[2,3,4], Pierre Abad[2,3,4], Miroslav Plohl[1]

1 Department of Molecular Biology, Rudjer Bošković Institute, Zagreb, Croatia, **2** French National Institute for Agriculture Research (INRA), Institut Sophia Agrobiotech, Sophia Antipolis, France, **3** University of Nice- Sophia Antipolis (UNSA), UMR Institut Sophia Agrobiotech, Sophia Antipolis, France, **4** Centre National se la Recherche Scientifique (CNRS), Institut Sophia Agrobiotech, Sophia Antipolis, France

Abstract

Tandemly arrayed non-coding sequences or satellite DNAs (satDNAs) are rapidly evolving segments of eukaryotic genomes, including the centromere, and may raise a genetic barrier that leads to speciation. However, determinants and mechanisms of satDNA sequence dynamics are only partially understood. Sequence analyses of a library of five satDNAs common to the root-knot nematodes *Meloidogyne chitwoodi* and *M. fallax* together with a satDNA, which is specific for *M. chitwoodi* only revealed low sequence identity (32–64%) among them. However, despite sequence differences, two conserved motifs were recovered. One of them turned out to be highly similar to the CENP-B box of human alpha satDNA, identical in 10–12 out of 17 nucleotides. In addition, organization of nematode satDNAs was comparable to that found in alpha satDNA of human and primates, characterized by monomers concurrently arranged in simple and higher-order repeat (HOR) arrays. In contrast to alpha satDNA, phylogenetic clustering of nematode satDNA monomers extracted either from simple or from HOR array indicated frequent shuffling between these two organizational forms. Comparison of homogeneous simple arrays and complex HORs composed of different satDNAs, enabled, for the first time, the identification of conserved motifs as obligatory components of monomer junctions. This observation highlights the role of short motifs in rearrangements, even among highly divergent sequences. Two mechanisms are proposed to be involved in this process, i.e., putative transposition-related cut-and-paste insertions and/or illegitimate recombination. Possibility for involvement of the nematode CENP-B box-like sequence in the transposition-related mechanism and together with previously established similarity of the human CENP-B protein and pogo-like transposases implicate a novel role of the CENP-B box and related sequence motifs in addition to the known function in centromere protein binding.

Citation: Meštrović N, Pavlek M, Car A, Castagnone-Sereno P, Abad P, et al. (2013) Conserved DNA Motifs, Including the CENP-B Box-like, Are Possible Promoters of Satellite DNA Array Rearrangements in Nematodes. PLoS ONE 8(6): e67328. doi:10.1371/journal.pone.0067328

Editor: Michael Freitag, Oregon State University, United States of America

Received February 8, 2013; **Accepted** May 17, 2013; **Published** June 27, 2013

Copyright: © 2013 Meštrović et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was financially supported by grants from the Ministry of Science, Education and Sports of the Republic Croatia (http://public.mzos.hr) (no. 098-0982913-2756) and INRA (Institut National de la Recherche Agronomique) (http://www.international.inra.fr/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: nevenka@irb.hr

[9] These authors contributed equally to this work.

Introduction

Satellite DNAs (satDNAs) can be briefly defined as DNA elements repeated in tandem. Often found as high-copy sequences underlying centromeres and broad pericentromeric regions, they rapidly achieve extreme diversity in nucleotide sequence, copy number, and organization in reproductively isolated groups of organisms (for review see [1]). Extensive studies of centromeric regions suggest coevolution of satDNAs and centromere-specific histone-like proteins leading to rapid evolution of centromeres and their rapid evolution is thought to be an intrinsic trigger of speciation [2].

SatDNAs evolve according to principles of concerted evolution, which is consequence of a 2-level process called molecular drive. At the first level, within the genome, mutations are homogenized among repeats of the satDNA [3]. Sequence homogenization results from a complex interplay of recombinational mechanisms, such as unequal crossing over and gene conversion. On the population level satDNA variants become fixed as a result of random assortment of genetic material in meiosis. The outcome of the whole process is higher homogeneity of repeats in a satDNA family within species than between species. Although turnover mechanisms in complex repetitive areas are difficult to explore, unequal crossing-over has been identified as the most widespread mechanism involved in satDNA dynamics in centromeric and pericentromeric regions [4], traditionally considered as regions of suppressed recombination (for review see [5]). Nevertheless, recent studies indicate gene conversion as the dominant mechanism in evolution of satDNAs [6]. As species diverge, satDNAs accumulate changes as a consequence of mutations and turnover mechanisms in separate lineages [3]. Rapidly accumulating differences in species satDNA profiles also can be accomplished by saltatory copy number changes and by emergence of new repeats in a

common set or a library of satDNAs shared by related genomes [7,8]. The library concept of satellite DNA evolution explains the occurrence of species-specific satellite DNA profiles as a result of differential amplifications and/or contractions within a pool of sequences shared by related genomes. In agreement with this concept, the study of 7 different satDNAs in six congeneric *Meloidogyne* species revealed the distribution of satDNAs consistent with lineage diversification and long term conservation (up to 45 Myr) of some satellite sequences [9].

However, the key question about satDNA evolution concerns the nature of mechanisms that drive formation and spread of novel tandem repeats in genomes. Although satDNAs can be extremely divergent, a common feature of many of them is irregular distribution of sequence variability along the monomer sequence and formation of conserved sequence segments, probably because of evolution under selective constraints (for review see [1]). The most prominent examples are found in rice [10], nematodes [8], *Arabidopsis* and human [11]. Among all detected conserved regions, the only function is assigned to the CENP-B box of alpha satDNA in human and other primates, which is proposed to act as a centromere protein binding site [12]. The possible role of other conserved sequence segments detected in satDNA monomers remains, however, obscure.

Meloidogyne are root-knot plant-parasitic nematodes that cause vast damage in agriculture. Although nematodes represent one large class of invertebrates, characterized by holocentric chromosomes with diffuse centromeres, evolutionary studies of satDNAs in this group are very limited. The recent completion of two root-knot nematode genomes *M. incognita* [13] and *M. hapla* [14] emphasized them as model organisms of metazoan plant parasitic species [15].

Recently separated but reproductively isolated nematodes *Meloidogyne fallax* and *M. chitwoodi* [16,17] offer an exceptional platform to explore mechanisms involved in satDNA formation and spread and possible requirements on their sequences. Previous work showed six satDNAs in *M. chitwoodi*, grouped according to sequence similarity in group 1 (1a, 1b, 1c and 1d satDNAs) and group 2 (2a and 2b satDNAs) [18]. The presence of the conserved 2a satDNA in *M. fallax* [19] indicates distribution of these satellites according to the principles of the satDNA library concept [7].

In this paper, we characterized five divergent satDNAs of the library shared by *M. fallax* and *M. chitwoodi* and one satDNA which is specific for *M. chitwoodi* only. We performed structural, organizational and phylogenetic analyzes which disclosed complex organization patterns of monomers in the form of simple and higher-order repeat (HOR) arrays. We also detected two short conserved domains in analyzed satDNA sequences. Interestingly, one of them appeared to be similar to the CENP-B box of human alpha satDNA. It was detected in sequence alignments as a conserved segment common for six divergent satDNAs. Our results suggest involvement of conserved domains in array rearrangements and onset of new sequence combinations. Proposed mechanisms act on short-segment tracts and indicate highly recombinogenic nature of satDNA arrays. Based on our findings we suggest an additional role of the CENP-B box and general involvement of conserved sequence motifs in rapid evolution of tandemly repeated sequences.

Materials and Methods

Sampling and DNA Isolation

The *Meloidogyne* spp. isolates used in this study were chosen from the living collection maintained at INRA, Sophia Antipolis, France. The geographic origin of the isolates was as follows: *M. chitwoodi* (Spijkenisse, The Netherlands), *M. fallax* (Baexem, The Netherlands), *M. javanica* (Pelotas, Brazil), *M. paranaensis* (Londrina, Brazil), *M. incognita* (Antibes, France), *M. arenaria* (Chappes, France) and *M. hapla* (La Môle, France). Nematodes were maintained on tomatoes (*Lycopersicon esculentum* cv. Saint Pierre) grown at 20°C in a greenhouse. They were specifically identified morphologically and according to their isoesterase electrophoretic pattern [20]. Eggs were collected from infested roots, according to the procedure described earlier [21]. Total genomic DNA was purified from 50–100 µl eggs using the DNeasy Tissue Kit (Qiagen) according to the manufacturer's instructions. Possibility of sample cross-contamination with other nematode DNA was excluded through PCR check of genomic DNA with SCAR (sequence characterized amplified region) primers specific for *Meloidogyne chitwoodi* and *M. fallax* species [22].

PCR Analyses, Cloning and Sequencing

The satellite sequences were amplified with specific primers derived from previously published data [18] and from sequences obtained in this work. Primer sequences and their positions on the HOR sequence are indicated in Table S1 and Fig. S1, respectively. The reaction mixture consisted of reaction buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 U GoTaq DNA polymerase (Promega), 0.4 µM of each primer and 20 ng of genomic DNA. The PCR cycling parameters used were as follows: 2 min initial denaturation at 94°C, followed by 30 cycles of: 95°C for 30 sec, 58°C for 30 sec, and 72°C for 1 min. Final extension was at 72°C for 10 min. PCR products were ligated in a pGEM T-Easy vector (Promega) and transformed in *Escherichia coli* DH5α-competent cells (Invitrogene). Recombinant clones with multimeric arrays of satellite DNA were sequenced by Macrogen (Korea). Monomers and HORs sequences from *M. chitwoodi* and *M. fallax* as well as Box 1-containing sequences from *M. incognita* sequenced genome were deposited in EMBL databank under Accession Numbers: JX186757-JX186849, JX186850-JX186855, JX186856 - JX186877, JX186878 - JX186996, KC968979 - KC969073.

Southern and Dot Blot Analyses

Standard procedures were used for restriction endonuclease digestions, electrophoresis, transfer to nylon membranes [23]. For genomic Southern hybridization analysis, 10 µg of genomic DNAs were partially digested with an appropriate restriction enzyme which cuts once in targeted repetitive unit. Gel electrophoresis was run in a 0.8% agarose, denatured and DNA transferred to Hybond N^+membrane (Amersham). Hybridizations were performed overnight under high stringency conditions (65°C) in the buffer containing 250 mM Na_2HPO_4 (pH 7.2), 7% SDS, 1 mM EDTA, 0.5% blocking reagent and 50 ng/ml of the probe. Posthybridization washes were done in 20 mM Na_2HPO_4/1 mM EDTA/1% SDS at the temperature 2°C lower than the hybridization temperature. Chemiluminescent detection was carried out using the alkaline phosphatase substrate CDP-Star (Roche Applied Science). Cloned satellite monomers labeled with biotin-dUTP by PCR were used as hybridization probes.

The abundance of satDNA sequences in *Meloidogyne* species was estimated by quantitative dot blot analysis using a series of genomic DNA dilutions ranging from 50 to 200 ng. Satellite monomers, excised from a plasmid, were dot-blotted in the range between 0.05 and 1 ng, and used as a calibration curve.

Sequence Analyses

DNA sequence data were compared to the GenBank databases by using the BLAST version 2.0 server at the National Center for Biotechnology Information.

The BLAST servers of *M. incognita* (http://meloidogyne.toulouse.inra.fr/blast/blast.html) and *M. hapla* (http://www.pngg.org/cbnp/index.php) genome were used to search for Box 1-containing sequences in the sequenced genomes. Initial sequence manipulations were done by using BioEdit v.7.0.5.3 [24]. Multiple alignments and pairwise sequence identity of monomers and HOR sequences were extracted from ClustalW Output (version 1.83) [25]. Lasergene software package v.7.0.0 (DnaStar) was used in further analyses of repetitive sequences including pairwise alignments, dot plot analyses and PCR primer design. Monomers from cloned multimeric arrays were extracted using Key-String Algorithm (KSA) [26]. KSA algorithm is based on the use of a freely chosen short sequence of nucleotides, called the key string, which cuts a given short sequence at each location within multimeric satDNA sequence. Distribution of monomer sequence variability was analyzed by using DnaSP v.4.10.9 [27]. The percent occurrence of the most frequent base at each site was calculated for all monomers repeats; this was plotted with the average percent occurrence and standard deviation (SD). A window length of 15 bp with a step size of 2 was used in the analysis. Due to the large number of monomers, neighbor-joining methods were used to construct phylogenetic tree by PAUP 4.0 (100 bootstrap iterations) [28]. Trees were displayed with MEGA 3.1 [29].

Results

Complex satDNA Arrays in *M. chitwoodi* and *M. fallax*

The first goal of this work was to characterize the structure and organization of satDNAs in *M. chitwoodi* and *M. fallax* genomes. PCR search for sequences related to *M. chitwoodi* satDNAs 1a, 1b, 1c, 1d, 2a and 2b revealed, except for 2b, orthologous counterparts in the closely related species *M. fallax* (Figure 1). However, it was not possible to detect any ortologous satDNA in other analysed *Meloidogyne* species (*M. incognita, M. javanica, M. arenaria, M. javanica, M. paranaensis*; data not shown).

Amplification of *M. chitwoodi and M. fallax* genomes with primers specific for satDNAs 1a, 2a and 2b produced ladder of bands based on the monomer size while other satDNA amplicons displayed different profiles (Figure 1A). PCR analysis of 1b showed bands of monomeric and dimeric size together with a fragment of about 1.5 kb in length, while amplification with 1c (Figure 1A) and 1d primers revealed complex but similar profiles (shown only for 1c). In order to perform detailed analyses of organizational patterns of these satDNA repeats, first we focused on sequences obtained by amplification of both genomes with 1c satDNA primers. In total, 20 cloned PCR fragments corresponding to multimeric size (i.e. ≥500 bp) were sequenced (Table S2). Two types of satDNA arrays were obtained, distinctive by the composition and organizational complexity of repeat subunits. The first type is characterized by arrays composed of alternating 1c and 1d satDNA monomers which together define the dimeric unit, 338 bp long (169 bp×2), organized in homogenous arrays (8 cloned fragments, $M_{1c}fa_n$ and $M_{1c}ch_n$; Table S1). Absence of a 170 bp based ladder in 1c PCR amplification supports this dimeric form as the basic repeating unit of these satDNAs. Multiple sequence alignment of another 12 fragments from *M. fallax* ($H_{1c}fa_n$) and *M. chitwoodi* ($H_{1c}ch_n$) (Table S2) revealed complex arrays composed of satDNA monomers 1a, 1b, a new 1b' variant, 1c, 1d and 2a together with U1, yet uncharacterized sequence segment (Figure 2A and Figure S1). BLAST searches did not indicate any relevant sequence homology of U1 with the studied satDNAs or any other sequence deposited in data bases. In the following experiment, U1 specific PCR primers were constructed in order to extend the segments of complex arrays. In both genomes, obtained PCR products revealed fragments of expected lengths (~1200 and ~1400 bp) but also generated a shorter fragment of about 700 bp (Figure 1A). Sequence alignment of fragments obtained with U1 primers (H_uch_n and H_ufa_n) and previously cloned complex arrays ($H_{1c}fa_n$ and $H_{1c}ch_n$) is consistent with tandem organization of the HOR unit (Figure S1 and Figure 2A). In contrast to homogeneity of HOR units (84–99% mutual sequence identity) neighboring monomers in HORs show a wide range of relationships: from relatively high sequence identity of 86% between 1b and 1b' variants to apparently unrelated sequences sharing only 32% identity, such as detected between 2a and 1c monomers (Figure 2A and Table 1). In addition, HOR segments revealed two variants which differ in the presence of 1b-type monomers (Figure 2A and Figure S1). Long HOR variants have two consecutive monomers, 1b and 1b', that share sequence identity of 86% (Table 1), while short HOR variants lack 1b monomer (Figure 2A). Genomic DNA cut with the REs specific for 1c monomer sequence and probed with the labeled 1c monomer fragment supports the proposed HOR tandem organization (marked with asterisks in Figure 1B). Southern hybridization of genomic DNA with 1c indicates that long HOR variants prevail in *M. fallax* genome, while short variants seem to be more abundant in *M. chitwoodi* (Figure 1B).

In addition to HORs, the alignment of the 700 bp-long complex fragments amplified with U1 primers revealed one additional homogenous group of sequences common for *M. fallax* and *M. chitwoodi*, named h_ufa_n and h_uch_n, respectively (Table S1). These sequences are composed of 1a and 1d complete monomers linked to a novel 170 bp long fragment named U2. The whole composite fragment is flanked by U1 sequences (Figure S2 and Figure 2B). It has to be noted that a 62 bp-long perfectly conserved fragment of U1 is also found as a part of U2 sequence. Additional PCR analyses using U2 specific primers could not prove tandem organization of the 700 bp complex fragment (data not shown). It can be therefore concluded that this fragment probably represents a particular combinatorial form of 1a and 1d satellite repeating units, present in the genome as an interspersed repeat.

Homogenous Monomeric Arrays

PCR with 1a-specific primers produced ladder-like profiles in both genomes, with fragments corresponding to multimers of 170 bp (Figure 1A). Cloning and sequencing (Table S2) revealed homogenous tandem arrays (94% mutual identity) composed of a variant of 1a satDNA sequence, indicated now as 1aM. This variant is different from the HOR variant detected above, which is therefore indicated as 1aH. Average sequence identity between 1aH and 1aM variants is 81% (Table 1). Southern blot hybridization of genomic DNA probed with cloned 1aM-type satDNA repeats confirmed tandem organization of 1aM variants (Figure 1B). In addition, 1aH-specific primers were constructed to check if 1aH builds independent tandem arrays. PCR reaction did not reveal any ladder-like profile (data not shown) indicating that these variants are exclusively present as subunits of HORs.

We also examined if 1b variants could be found in monomeric arrays or are an exclusive component of HOR elements. Southern blot analysis of genomic DNA showed hybridization signals only in bands corresponding to HOR arrays (Figure 1B). PCR with 1b primers revealed fragments whose length corresponds to HOR organization. According to primer position, fragments of monomeric and dimeric forms that appeared in the PCR reaction (Figure 1A) originate from HORs, as they were undetectable in

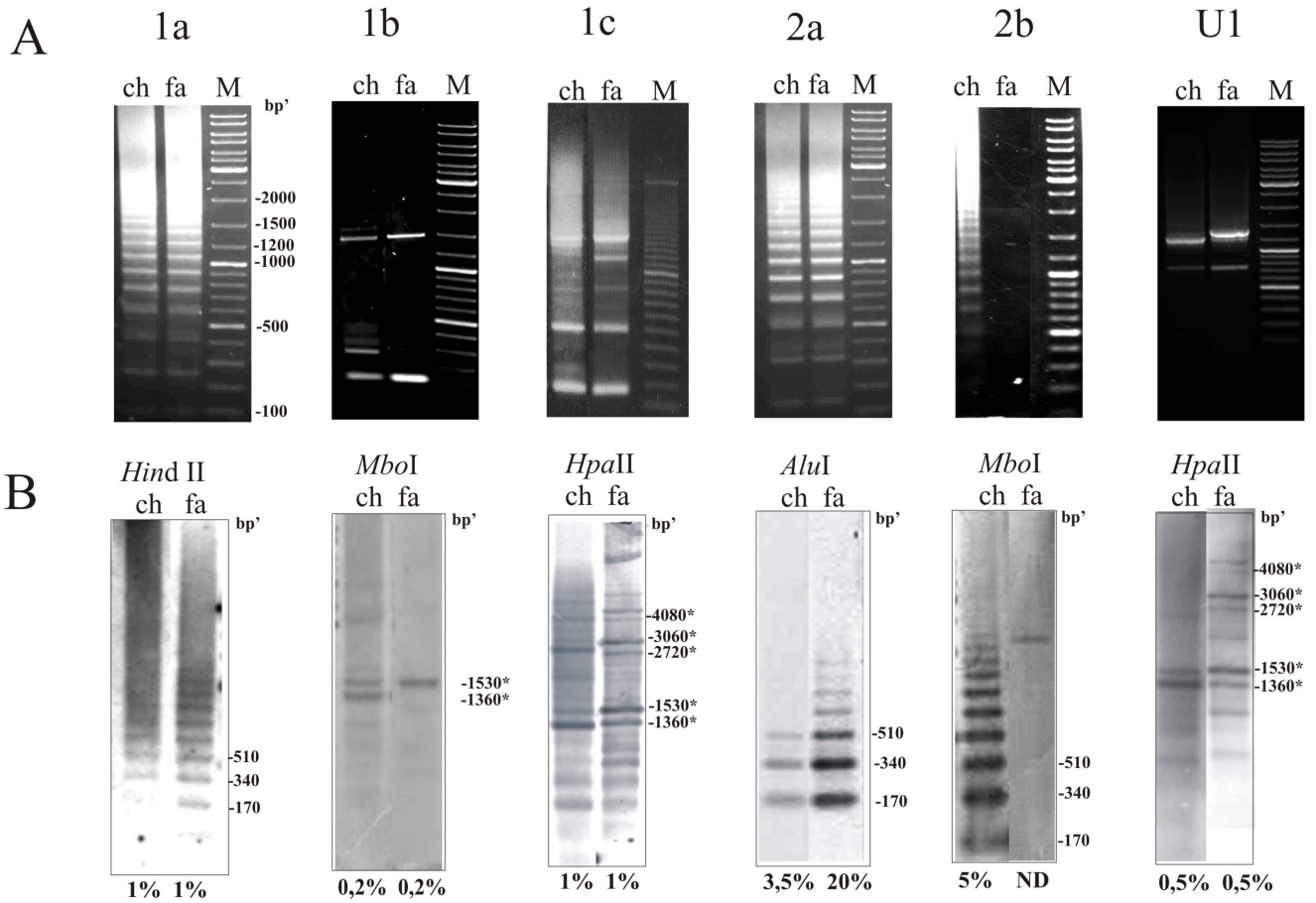

Figure 1. Organization of satDNAs in *M. chitwoodi* (ch) and *M. fallax* (fa) genome. (A) Electrophoretic separation of PCR products obtained by amplification of genomic DNAs using primers specific for 1a, 1b, 1c, 2a and 2b satDNAs and U1 sequence are shown on upper panel. (B) Southern hybridizations of genomic DNA partially digested with RE-s and probed with 1a, 1b, 1c, 2a and 2b satDNA monomers and with U1 sequence are shown on lower panel. Approximative contribution of particular sequence in the genome, estimated by dot blot, is shown as a percentage indicated below Southern blots. HORs are indicated with asteriks. M is the DNA ladder marker. ND-not detectable.
doi:10.1371/journal.pone.0067328.g001

genomic Southern blot. These results emphasize unique organization of 1b monomers exclusively in HORs in both genomes.

It has been published previously that 2a satDNA exists in the *M. fallax* and *M. chitwoodi* genome in tandem arrangement, in a high copy number, organized as homogenous monomeric arrays [19]. The results provided in this work show that 2a satellite also exists as the element of HORs in both genomes (Figure 2). No diagnostic sequence differences could be observed with respect to organizational pattern or species of origin (Figure 3 and Figure S3). The only difference is in abundance of 2a satDNA, 3.5% *in M. chitwoodi* and 20% in *M. fallax* (Figure 1). Examination of 2b satellite by PCR amplification and Southern blot (Figure 1) confirmed its exclusive presence in the *M. chitwoodi* genome in the form of high copy homogenous monomeric arrays. The only observed hybridization signal in *M. fallax* is the faint band (Figure 1B) which could represent a sporadic 2b sequence embedded in a longer DNA segment.

Phylogenetic Analyzes of Monomers

In an effort to assess sequence dynamics of repetitive units in the closely related *M. chitwoodi* and *M. fallax* genomes, we examined phylogenetic relationships of all monomers, regardless to their organizational pattern and species origin. A total of 212 monomeric units from *M. chitwoodi* and *M. fallax* were included in the multiple sequence alignment (Figure S3). Neighbor-joining phylogenetic analysis showed eight different clusters (1aH, 1aM, 1bH, 1b'H, 1cDH, 1dDH, 2aMH and 2bM; letters H, D, M indicate HOR, dimeric or monomeric organizational form, respectively) distributed in two main branches, satDNAs of group 1 and group 2 (Figure 3). Monomers within clusters could not be distinguished according to the species of origin nor was it possible to differentiate 1c, 1d and 2a monomers according to their array affiliation. In agreement with previous observation, 1a split in 1aM and 1aH according to their organizational origin, while 1b monomers form two distinct groups 1bH and 1b'H related to their position in HORs. It should be noted that 1aH further clusters in two subgroups, based on short and long HOR forms.

Sequence comparisons between monomer groups display three different levels of similarity (Table 1). Similarity is high within 1bH group (86%) and between 1aM and 1aH (81%) monomer variants. Similarities within other satDNAs of group 1 and within satDNAs of group 2 are moderate, ranging from 51 to 66%. Comparison between satDNAs of group 1 and 2 gives negligible similarities, 32–46% (Table 1), and it can be supposed that these two groups might represent sequences of unrelated origin.

Table 1. Mean of percent sequence identity between main groups of satDNAs monomers.

Monomer length(bp)	Monomers name	Group name	n	1aHch$_n$, 1aHfa$_n$	1aMch$_n$, 1aMfa$_n$	1bHch$_n$ 1bHfa$_n$	1b'Hch$_n$ 1b'Hfa$_n$	1dHch$_n$,1dHfa$_n$, 1dMch$_n$,1dMfa$_n$	1cHch$_n$, 1cHfa$_n$, 1cMch$_n$,1cMfa$_n$	2aHch$_n$,2aHfa$_n$ 2aMch$_n$, 2aMfa$_n$	2bMch$_n$
170	**1aHch$_n$, 1aHfa$_n$**	**1aH**	26	**94** (2.8)[a]							
169 (+5)	**1aMch$_n$, 1aMfa$_n$**	**1aM**	50	**81** (1.6)	**94** (2.1)						
170	**1bHch$_n$, 1bHfa$_n$,**	**1bH**	10	**63** (1.9)	**64** (0.8)	**99** (0.3)					
	1b'Hch$_n$ 1b'Hfa$_n$	**1b'H**	21	**64** (1.2)	**66** (0.9)	**86**(2.9)	**93** (4.7)				
169	**1dHch$_n$,1dHfa$_n$, 1dMch$_n$,1dMfa$_n$**	**1dMH**	31	**64** (1.9)	**63** (2.0)	**57** (0.5)	**56** (0.7)	**98** (1.7)			
169	**1cHch$_n$, 1cHfa$_n$, 1cMch$_n$,1cMfa$_n$**	**1cMH**	31	**56** (0.6)	**53** (1.0)	**52** (0.2)	**51** (1.2)	**62** (1.0)	**99** (0.5)		
180	**2aHch$_n$,2aHfa$_n$ 2aMch$_n$, 2aMfa$_n$**	**2aMH**	37	**39** (0.9)	**40** (1.0)	**46** (0.6)	**42** (1.0)	**36** (0.7)	**32** (0.5)	**97** (1.6)	
179	**2bMch$_n$**	**2bM**	10	**40** (1.1)	**40** (0.8)	**51** (0.5)	**46** (1.6)	**37** (0.6)	**35** (0.6)	**60** (0.9)	**96** (1.3)

n-number of analysed monomers.

[a]average percent identity scores for each pairwise comparison are indicated in bold, while standard deviation (SD) is indicated in bracket.

doi:10.1371/journal.pone.0067328.t001

Conserved Motifs and Junctions Between Monomers

In contrast to the very low overall sequence similarity between some of the monomer groups (Table 1), pairwise sequence alignment and sliding window analysis of all monomer sequences identified common domains of low variability (Figure 4A, B). The shaded domain in consensus sequences indicates the region of low variability shared among all satDNAs. Part of this region is a conserved 17 bp long segment, named Box 1. It is interesting to note that this sequence segment remains conserved among highly divergent satDNAs. For example, 1c and 2a satDNAs share only 32% identity while in the same time one single change characterizes the Box 1. Comparison of conserved Box 1 sequences (in Figure 4C presented as a reverse complement) with the human CENP-B box shows significant degree of similarity. Six of them have 10–12 out of 17 nucleotides conserved and if bases essential for CENP-B binding in human are considered, 4–5 out of 9 remain conserved. The lowest identity is in exclusively HOR-included elements, 1b'H and 1bH, in which sequences may represent degenerate variants of the motif. This analysis was extended with the search for related motifs in sequenced *M. incognita* and *M. hapla* genomes. Preliminary results recovered different repetitive sequences with the Box 1 in unassembled part of *M. incognita* sequenced genome (Figure S4). However, none of these repeats indicated any sequence similarity with satDNA sequences studied in this work in *M. chitwoodi* and *M. fallax*. In addition, detailed analysis of HOR elements in *M. chitwoodi* and *M. fallax* revealed that transitions from 1d to 2a monomer and from 2a to 1c are located exactly at the Box 1 (Figure 2A, see Discussion).

Another common region (Box 2) conserved in HOR-related monomers of group 1 satDNAs (1aH, 1bH, 1b'H, 1cH and 1dH) (Figure 4D and Figure S5). In order to refine alignment of this sequence motif, Box 2 segments were compared with their consensus sequence (Figure 4D). This region is 20 bp-long composed of T, C and A tracts and shows significant degree of mutual sequence identity with only few nucleotide changes (Figure 4D). It must be noted that the Box 2 region is always found in HORs as a transition region between monomers from group 1 (Figure 2A). In addition, detailed analysis of the so-called complex fragment revealed that 1a monomer extends into 1d monomer in the 50 bp long overlapping region shared by both monomers. This whole segment is highly conserved, with only 6 nucleotide substitutions (Figure S2).

Discussion

In the present study we performed a comprehensive analysis of five divergent satDNAs (1a, 1b, 1c, 1d, and 2a) shared as elements of the satDNA library of root-knot nematodes *M. chitwoodi* and *M. fallax*, the two species considered to become separated recently [16,17]. A distinctive element of this satDNA library is 2b satDNA, which turned to be present only in the *M. chitwoodi* genome. This observation supports our previous conclusion that presence of novel satDNAs in the library is accompany of speciation processes [9]. The distribution analysis data shows the absence of 1a, 1b, 1c, 1d, 2a and 2b counterparts in other congeneric *Meloidogyne* species thus indicating that satDNAs described in this work are specific for *M. chitwoodi* and M. *fallax*.

The exceptional attribute of studied satDNAs is complex organization of repeat units. Simple arrays are highly homogenous and composed of monomers or dimers, the later being built of two highly divergent monomers. Comparable dimeric organization based on monomers of low sequence similarity (50–60%) was reported in the marmoset (NewWorld monkeys) and it represents

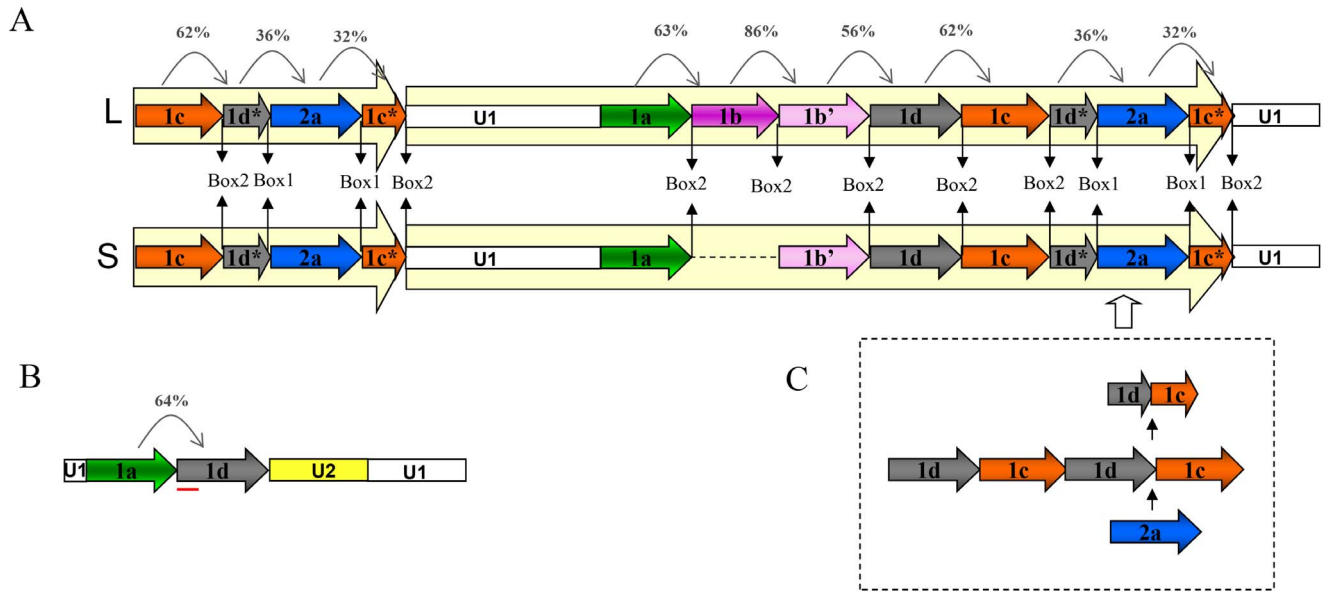

Figure 2. Schematic representation of complex satDNA structure. (A) The long-L and short-S HOR sequence and (B) complex fragment. The percent identity between monomers is written on arrows above the scheme. Box 1 and Box 2 in junction regions between different monomeres are indicated. 1d* and 1c* represent monomer parts which remain after 2a insertion. The red line below the complex fragment represents the overlapping segment of 1a and 1d monomers. (C) The scheme in the frame represents outcome of the proposed cut-and-paste mechanism of 2a insertion in HOR array.
doi:10.1371/journal.pone.0067328.g002

an ancient dimeric structure of alphoid sequences [30]. In our work, complex HORs are formed of monomers of divergent satDNAs that range from apparently unrelated (32% sequence identity) to those sharing up to 86% sequence identity. While the later can be considered as variants of a single satDNA, such as the 1b'H-1bH monomer pair, possible common evolutionary origin of the most divergent monomers is masked. Such a complex organization of monomers, described in details, is characteristic for alpha satDNA of human and great apes [30,31]. For the difference to characterized nematode satDNAs, alpha satDNA HORs are composed of monomers with relatively high mutual sequence similarity (75–88%) [32]. A significant difference in organization of simple arrays can be also observed; while simple arrays of *M. fallax* and *M. chitwoodi* are highly homogenous (94–97% sequence similarity), equivalent arrays of alpha satDNA exhibit sequence similarity comparable to that of monomers in alpha HORs [32]. Phylogenetic analyses of alpha satDNA monomers in primates and human chategorized HOR and monomeric forms as phylogenetically distinct and suggested evolution of both forms from ancestral arrays of monomeric repeats [33]. Similar analysis in *M. chitwoodi* and *M. fallax* revealed clustering of HOR units with those from simple arrays, indicating continuous shuffling of monomers between HORs and simple arrays. The only exception is grouping of 1aH and 1aM monomers, in accordance with array affiliation. This result suggests that mechanisms in addition to unequal crossover over and gene conversion [3,5] should be involved in creation of HORs (see below).

Irrespectively to the low level of sequence identity (32–64%) among studied satDNAs and the organizational pattern in which they were found, examined monomers share two conserved segments, named Box 1 and Box 2. Box 1 is a conserved 17 bp-long segment characteristic for all analyzed satDNAs. This particular motif is observed even in the divergent 2b satDNA, found only in homogeneous monomeric arrays of *M. chitwoodi*. One single deleted nucleotide was found in Box 1 of 1bH and 1b'H monomers which, curiously, appear exclusively as HOR-included elements. This raises the speculative possibility that conserved Box 1 participates in the formation of homogenous simple arrays. It was already proposed that abundant satDNAs may have been selected for amplification because of their ability to bind nuclear proteins [34]. Interestingly, conserved Box 1 shows significant homology with the human CENP-B box, with identity in 10–12 out of 17 nucleotides. The CENP-B box is a well-described sequence motif of human alpha satDNA which represents a binding site for the CENP-B protein in a subset of alpha satellite HORs [35]. It has been proposed that the CENP-B protein participates in human centromere assembly [35] but normal chromosome segregation in a mouse CENP-B protein null mutant and absence of CENP-B binding sites at the centromeres of human and mouse Y chromosome make its exact function unclear [36,37]. DNA sequence motifs similar to the CENP-B box were found in diverse mammalian species, although their satDNA sequences are completely unrelated among themselves and with the alpha satDNA [38,39]. For example, seven divergent horse satDNAs exibit CENP B box variants with identity in 9–12 out of 17 nucleotide of human CENP B box [39]. Presence of motifs similar to the CENP-B box has also been detected in a number of satDNAs from diverse species outside mammals [40,41]. In examined nematode species, homology of Box 1 with the human CENP-B box is in the same range found for the CENP-B box in diverse mammalian species [39]. Exceptional feature of the nematode CENP-B box-like motif is significant conservation in the six divergent satDNAs which emphasized it as the most prominent example of the CENP-B box-like sequence out of mammals.

Mechanisms of genetic exchange of satDNAs are hard to study because of repetitive nature of satDNAs arrays. However, our experimental system composed of complex HORs and their counterparts in simple arrays offers a convenient model in which

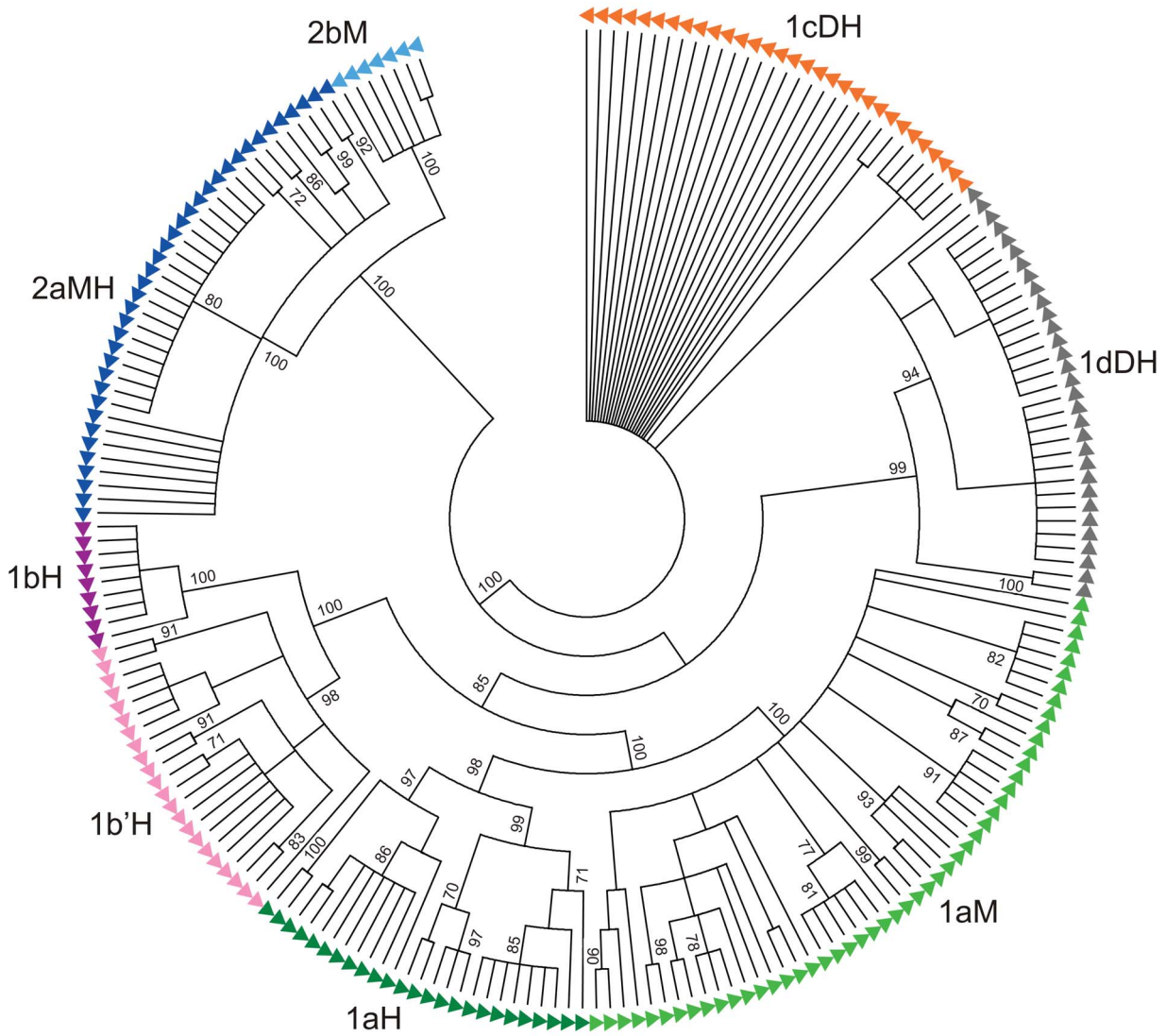

Figure 3. The phylogenetic tree of 1a, 1b, 1b' 1c, 1d, 2a and 2b monomers. Monomers from the HORs (H), dimeric (D) and monomeric arrays (M). Phylogenetic analysis of 212 monomers was performed by neighbor-joining method with bootstrap value of 100. Numbers at nodes indicate bootstrap values (100 replicates; only values greater than 70 % are shown.
doi:10.1371/journal.pone.0067328.g003

"beginning" and "end" of monomers can be precisely defined. Detailed analyses of *Meloidogyne* satDNA arrays led to observation that junctions between monomers are always located in conserved motifs. Box 1 is found at sites of insertion of the complete 2a monomer into highly divergent 1d and 1c monomers, while in turn, the corresponding segment of equivalent length in 1d and 1c, limited with Box 1, has been extruded (Figure 2C). This rearrangement event indicates novel cut-and-paste mechanism that involves the 17 bp-long CENP-B box-like motif and, probably, is related to mechanisms of transposition. It has been already hypothesized that the CENP-B box, in addition to its putative centromeric role, might have a function in satDNA sequence rearrangements [42]. This assumption is based on similarity of the CENP-B protein and transposases of the *pogo* family [43]. Accordingly, the CENP-B box might trigger illegitimate recombination in centromeric areas, in an epigenetically controlled process [44]. Highly conserved CENP-B protein homologs were detected in many mammalian species, but not in other metazoans [43]. In contrast, transposase-derived proteins related to the CENP-B and with putative ability to interact with satDNAs have been detected in diverse invertebrate and vertebrate species [43]. In support, a search in the genome sequence of related species *M. incognita* [13] allowed identification of an EST-supported gene encoding a protein with both CENP-B/Tc5 transposase DNA binding domains (Minc05185) (unpublished data) as well as the existence of different repetitive sequences that contain the CENP B box- like motif identical as that observed in this work.

The conserved Box 2 is a sequence motif composed of A/T/C tracts, found as a 20 bp- long transition region of all group 1 monomers in HORs. This indicates that homopolymeric tracts which have been found as a common feature of many satellites [1], participate in sequence recombination events in *Meloidogyne*. Since divergent monomers are involved, a mechanism of illegitimate recombination mediated by Box 2 can be assumed. Illegitimate recombination was previously proposed as a mechanism responsible for interspersion of long arrays generating abrupt switches between nonhomologous satDNAs in *Drosophila* [45]. While

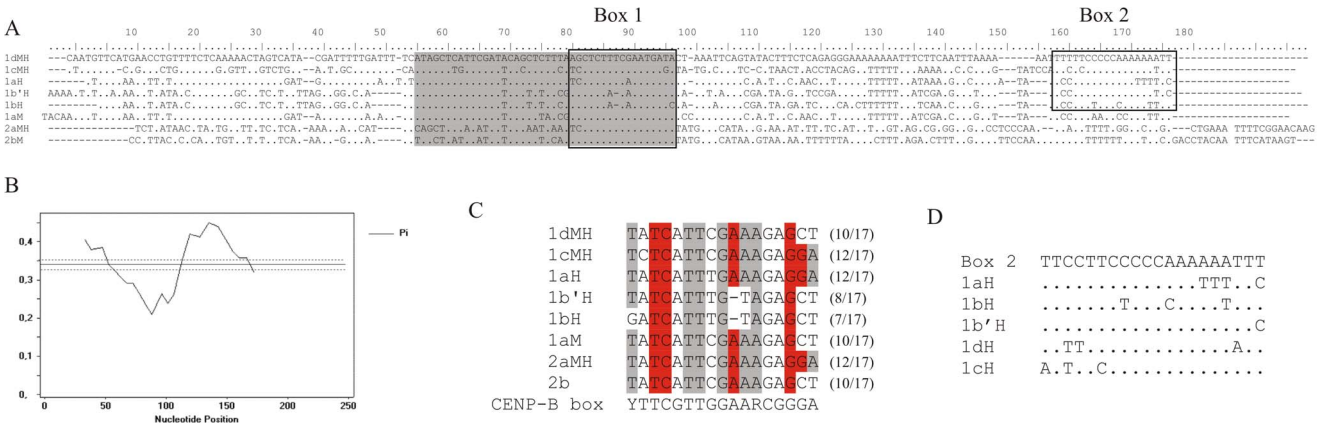

Figure 4. Conserved motifs in satDNAs. (A) Consensus sequences of 1dMH, 1cMH, 1aH, 1bH, 1b'H, 1aM, 2aMH and 2bM satDNAs, determined according to the 50% majority rule. Conserved Box 1 and Box 2 are indicated within the boxed area, and shaded part represents a region of low variability.(B) Identification of low variable domains by sliding window analysis by DnaSP. The average nucleotide variability P is shown by a solid line, and dashed lines represent 2-fold value of standard deviation. (C) Comparison of two variants of Box 1 with the consensus of human CENP-B box. The reverse complementary sequence of Box 1 is presented. Identities between sequences are highlighted in grey, and bases considered essential to bind the CENP-B protein in human [34] are highlighted in red. The number of total conserved bases is reported in brackets. (D) Aligment of Box 2 sequences from HOR related monomers; positions identical to the overall consensus are shown with dots.
doi:10.1371/journal.pone.0067328.g004

switches between unrelated arrays in *Drosophila* were detected as relatively rare events, our results nominate Box 2 as promoter of recombination acting frequently on DNA fragments of near monomer size. The minimal observed junction length of about 20 bp in both Box 1 and Box 2 is in accordance with the length of recombination breakpoints in human alpha-satellite [46]. In support to this, the role in satDNA shuffling can be assumed by presence of different conserved regions of similar length, as observed in the MEL 172 satDNA family identified in several *Meloidogyne* species [8] and in other, such as *Arabidopsis* [47].

In conclusion, we disclosed complex organization of monomers in two *Meloidogyne* species, characterized by highly homogenous simple arrays and by HORs, composed of highly divergent monomers. We propose that onset of this organizational pattern was mediated by conserved Box 1 and Box 2 sequence motifs. In principle, the two mechanisms are envisaged in this process, satDNA transposition and illegitimate recombination. Similarity of Box 1 with the CENP-B box of alpha satDNA and hypothesized transposase origin of the CENP-B protein [43] favor the role of transposition in formation and dynamics of satDNA arrays. These mechanisms act on short-segment tracts indicating the highly recombinogenic nature of repetitive environment which is in agreement with recent studies performed on mammalian centromere [44] and in other species [45,48]. Finally, HORs can also represent a template from which monomers with conserved CENP-B box-like segments can be amplified and form high copy number arrays. It can be hypothesized that parallelism in organizational patterns of nematode and human satDNAs and similar sequence motif may mirror similar mechanisms of genesis and sequence dynamics, presumably driven by the same family of transposase-related processes.

Supporting Information

Figure S1 Alignment of HORs from *M. fallax* (clone names in blue) and *M. chitwoodi* (clone names in green). H1cfa(n) and H1cch(n) represent fragments amplified with 1c primers. Hufa(n) and Huch(n) are amplified with primers specific for U1 sequence. All primer positions are marked above sequences and primers are listed in Table S1. SatDNA monomers are indicated in different colours; 1c, 1d, 2a, 1a, 1b and 1b'. Unlabeled part of the HOR is U1 sequence. Red boxes indicate Box A, and black boxes represent Box B. Sequences are deposited in EMBL databank under accession numbers: JX186856–JX186877.
(DOC)

Figure S2 Alignment of complex fragments from *M. fallax* (clone names in blue) and *M. chitwoodi* (clone names in green). Sequences are indicated in different colours; 1a monomer (green), 1d monomer (grey) and U2 sequence (yellow). Unlabeled part belongs to U1 sequence. Blue box represents overlapping region of 1a and 1d monomers. Box 1 is indicated in red, and Box 2 in black. Grey boxes represent perfectly conserved fragment common for U1 and U2 sequences. Primer positions for U2 are indicated above sequences. Sequences are deposited in EMBL databank under accession numbers: JX186850–JX186855.
(DOC)

Figure S3 Alignment of 1a, 1b, 1b', 1c, 1d, 2a and 2b monomers from *M. fallax* and *M. chitwoodi*. Monomers are extracted from monomeric and HOR arrays using KSA algorithm [26]. All monomers are compared with first sequence and positions identical to the first sequence are shown with dot. Monomer group are indicated on the right side. Monomer sequences are deposited in EMBL data bank under accession numbers: JX186757–JX186849 and JX186878–JX186996. Box 1 is shaded with yellow. Detail description of satellite monomers are indicated below alignment.
(DOC)

Figure S4 Alignment of Box 1-containing sequences extracted from unassembled part of *M. incognita* sequenced genome. All sequences are compared with first sequence and positions identical to the first sequence are shown with dot. Sequences are deposited in EMBL data bank under accession numbers: KC968979–KC969073. Box 1 is shaded with yellow.
(DOC)

Figure S5 Alignment of Box 2 from HOR related monomers of group 1 (1aH, 1bH, 1b'H1c, and 1dH).

(DOC)

Table S1 Primers used to amplify genomic sequences.
(DOC)

Table S2 Description of cloned satellite DNA arrays. In cloned satellite fragments, letters H, M and h indicate higher-order repeats, monomeric arrays, complex fragment, respectively. Then follow primer name (first subscript), species acronym and clone number (second subscript).
(DOC)

Acknowledgments

Authors would like to thank Barbara Mantovani and Brankica Mravinac for critical reading and useful suggestions during preparation of the manuscript.

Author Contributions

Conceived and designed the experiments: NM. Performed the experiments: M. Pavlek NM AC. Analyzed the data: M Pavlek NM. Contributed reagents/materials/analysis tools: M. Plohl PCS PA. Wrote the paper: NM M. Plohl. Intellectual support: PCS PA.

References

1. Plohl M, Luchetti A, Mestrović N, Mantovani B (2008) Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene 409: 72–82.
2. Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293: 1098–1102.
3. Dover GA (1986) Molecular drive in multigene families: How biological novelties arise, spread and are assimilated. Trends Genet 2: 159–165.
4. Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. Science 294: 109–115.
5. Talbert PB, Henikoff S (2010) Centromeres convert but don't cross. PLoS Biol 8: 1–5.
6. Shi J, Wolf SE, Burke JM, Presting GG, Ross-Ibarra J, et al. (2010) Widespread gene conversion in centromere cores. PLoS Biol 8: 1–10.
7. Meštrović N, Plohl M, Mravinac B, Ugarković ● (1998) Evolution of satellite DNAs from the genus *Palorus* - experimental evidence for the library hypothesis. Mol Biol Evol 15: 1062–1068.
8. Meštrović N, Castagnone-Sereno P, Plohl M (2006) Interplay of selective pressure and stochastic events directs evolution of the MEL172 satellite DNA library in root-knot nematodes. Mol Biol Evol 23: 2316–2325.
9. Meštrović N, Plohl M, Castagnone-Sereno P (2009) Relevance of satellite DNA genomic distribution in phylogenetic analysis: a case study with root-knot nematodes of the genus *Meloidogyne*. Mol Phylogenet Evol 50: 204–208.
10. Lee HRR, Neumann P, Macas J, Jiang J (2006) Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. Mol Biol Evol 23: 2505–2520.
11. Hall SE, Kettler G, Preuss D (2003) Centromere satellites from *Arabidopsis* populations?: maintenance of conserved and variable domains. Genome Res 13: 195–205.
12. Masumoto H, Nakano M, Ohzeki JI (2004) The role of CENP-B and alpha-satellite DNA: de novo assembly and epigenetic maintenance of human centromeres. Chromosome Res 12: 543–556.
13. Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EGJ, et al. (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. Nature Biotechnol 26: 909–915.
14. Opperman CH, Bird DM, Williamson VM, Rokhsar DS, Burke M, et al. (2008) Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. Proc Natl Acad Sci 105: 14802–14807.
15. Bird DM, Williamson VM, Abad P, McCarter J, Danchin EGJ, et al. (2009) The genomes of root-knot nematodes. Annu Rev Phytopathol 47: 333–351.
16. Van der Beek JG, Karssen G (1997) Interspecific hybridization of meiotic parthenogenetic *Meloidogyne chitwoodi* and *M. fallax*. Phytopathology 87: 1061–1066.
17. Van Megen H, Van den Elsen S, Holterman M, Karssen G, Mooyman P, et al. (2009) A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. Nematology 11: 927–950.
18. Castagnone-Sereno P, Leroy H, Semblat JP, Leroy F, Abad P, et al. (1998) Unusual and strongly structured sequence variation in a complex satellite DNA family from the nematode *Meloidogyne chitwoodi*. J Mol Evol 46: 225–233.
19. Castagnone-Sereno P, Semblat JP, Leroy F, Abad P (1998) A new AluI satellite DNA in the root-knot nematode *Meloidogyne fallax*: relationships with satellites from the sympatric species *M. hapla* and *M. chitwoodi*. Mol Biol Evol 15: 1115–1122.
20. Carneiro RMDG, Almeida MRA, Quénéhervé P (2000) Enzyme phenotypes of *Meloidogyne* spp. populations. Nematology 2: 645–654.
21. Castagnone-Sereno P, Leroy F, Abad P (2000) Cloning and characterization of an extremely conserved satellite DNA family from the root-knot nematode *Meloidogyne arenaria*. Genome/National Research Council Canada = Génome/Conseil national de recherches Canada 43: 346–353.
22. Zijlstra C (2000) Identification of *Meloidogyne chitwoodi*, *M. fallax* and *M. hapla* based on SCAR-PCR: a powerful way of enabling reliable identification of populations or individuals that share common traits. Eur J Plant Pathol 106: 283–290.
23. Sambrook J, Russell DW, editors (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. p.
24. Hall TA (1999) BioEdit- a user-friendly biological sequence alignment editor and analysis program for Windows 95–98-NT.pdf. Nucleic Acids Symp Ser 41: 95–98.
25. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876–4882.
26. Rosandić M, Paar V, Basar I (2003) Key-string segmentation algorithm and higher-order repeat 16mer (54 copies) in human alpha satellite DNA in chromosome 7. J Theor Biol 221: 29–37.
27. Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15: 174–175.
28. Swofford DL (2002) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts: 1–142.
29. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5: 150–163.
30. Cellamare A, Catacchio CR, Alkan C, Giannuzzi G, Antonacci F, et al. (2009) New insights into centromere organization and evolution from the white-cheeked gibbon and marmoset. Mol Biol Evol 26: 1889–1900.
31. Alkan C, Ventura M, Archidiacono N, Rocchi M, Sahinalp SC, et al. (2007) Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. PLoS Comput Biol 3: 1807–1818.
32. Rudd MK, Willard HF (2004) Analysis of the centromeric regions of the human genome assembly. Trends Genet 20: 529–533.
33. Rudd MK, Wray GA, Willard HF (2006) The evolutionary dynamics of alpha-satellite. Genome Res 16: 88–96.
34. Csink AK, Henikoff S (1998) Something from nothing: the evolution and utility of satellite repeats. Trends Genet 14: 200–204.
35. Masumoto H, Yoda K, Ikeno M, Kitagawa K, Muro Y, et al. (1993) Properties of CENP-B and its target sequence in a satellite DNA. In: Baldev KV, editor. In: Chromosome Segregation and Aneuploidy. NATO ASI Ser. H 72, Vol. 72. 31–43.
36. Earnshaw WC, Bernat RL, Cooke CA, Rothfield NF (1991) Role of the centromere/kinetochore in cell cycle control. Cold Spring Harbor Symp Quant Biol 56: 675–685.
37. Fowler KJ, Hudson DF, Salamonsen L a, Edmondson SR, Earle E, et al. (2000) Uterine dysfunction and genetic modifiers in centromere protein B-deficient mice. Genome Res 10: 30–41.
38. Kipling D, Mitchell AR, Masumoto H, Wilson HE, Nicol L, et al. (1995) CENP-B binds a novel centromeric sequence in the Asian mouse *Mus caroli*. Mol Cell Biol 15: 4009–4020.
39. Alkan C, Cardone MF, Catacchio CR, Antonacci F, O'Brien SJ, et al. (2011) Genome-wide characterization of centromeric satellites from multiple mammalian genomes. Genome Res 21: 137–145.
40. López CC, Edström JE (1998) Interspersed centromeric element with a CENP-B box-like motif in *Chironomus pallidivittatus*. Nucleic Acids Res 26: 4168–4172.
41. Fantaccione S, Pontecorvo G, Zampella V (2005) Molecular characterization of the first satellite DNA with CENP-B and CDEIII motifs in the bat *Pipistrellus kuhli*. FEBS letters 579: 2519–2527.
42. Kipling D, Warburton PE (1997) Centromeres, CENP-B and Tigger too. Trends Genet 13: 141–145.
43. Casola C, Hucks D, Feschotte C (2008) Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. Mol Biol Evol 25: 29–41.
44. Jaco I, Canela A, Vera E, Blasco M a (2008) Centromere mitotic recombination in mammalian cells. J Cell Biol 181: 885–892.
45. Kuhn GCS, Teo CH, Schwarzacher T, Heslop-Harrison JS (2009) Evolutionary dynamics and sites of illegitimate recombination revealed in the interspersion and sequence junctions of two nonhomologous satellite DNAs in cactophilic *Drosophila* species. Heredity 102: 453–464.
46. Warburton PE, Waye JS, Willard HF (1993) Nonrandom localization of recombination events in human alpha satellite repeat unit variants: implications for higher-order structural characteristics within centromeric heterochromatin. Mol Cell Biol 13: 6520–6529.

47. Hall SE, Luo S, Hall AE, Preuss D (2005) Differential rates of local and global homogenization in centromere satellites from *Arabidopsis* relatives. Genetics 170: 1913–1927.

48. Mravinac B, Plohl M (2010) Parallelism in evolution of highly repetitive DNAs in sibling species. Mol Biol Evol 27: 1857–1867.