Supplementary Tables

Name	Frequency (%)	Mean sequence identity (%)	Cenp-B (%)	PJalpha (%)
C1	84	95	0,04	96
C2	5	85	0,09	84
C5	7	95	0,02	96
C6	4	98	0,05	98

Table S1. Features of the four alpha satellite families identified within the CP XmnI monomer dataset. For each family, the frequency and sequence properties are shown. Sequence identities were calculated using a subset of 500 randomly selected sequences within each family. The percentages of sequences displaying the CENP-B or pJalpha fixation site were calculated for each family using all the sequences that were attributed to the family.

Name	Pattern (5'-3')	Label	Mismatch	C1	C2	C5	C6
C1a	TcCCtTtGcCaAtTcCAc	3'Cy3	0	63	1	64	-
			1	86	4	87	-
C1b	AcTgCtCtGtGtTcTGtTa	3'Digoxygenin	0	68	1	-	77
			1	91	29	1	95
C2a	TcACtTtGcAaAtTcCAc	5'AlexaFluor488	0	-	22	-	-
			1	5	57	1	-
C2b	AcTgCtTtGtGtTcTGtTa	5'Cy5	0	1	26	-	-
			1	69	58	-	78
C5a	TgAaTtCaGaGaAcAcAg	3'Biotin	0	-	-	77	-
			1	2	-	95	1
C6a	CaTTtTcCcTtCaAgAaTcC	3'Digoxygenin	0	-	-	-	74
			1	-	-	-	96
Cx	tctcagaaagctt	3'Biotin	0	77	40	76	74
			1	94	74	93	96

Table S2. Design and properties of the oligonucleotide probes for FISH experiments. The sequence of all the probes is indicated (LNA modifications are written in lower cases), as well as nature of the hapten used for detection. We also display the calculated percentage of sequences from each family that are expected to be recognized by each probe, either perfectly (0 mismatch) or with a single mismatch.

Id	Sequence	Number	Forward(%)
1	Consensus C1	2444	55
2	C158G	961	62
3	C137A-CC149AA-C2A-G17Del	468	51
4	C114Del	237	0
5	C116T-A3741T-G64A	210	49
6	T101Del	164	96
7	C116T	132	61
8	C114Del-C158G	114	100
9	C137A-C158G	104	58
10	A40C-C42G	103	60
11	A3741T-G64A	95	51
12	C2A	79	44
13	GC166TT	79	71
14	G64A	75	48
15	A86T	75	50
16	C116T-C158G	72	61
17	T121A	70	64
18	A110G	70	46
19	T101Del-C158G	67	94
20	A40C-C42G-G28T	60	67
21	G17C	59	56
22	C137A-CC149AA	52	48
23	C2A- C137A-CC149AA	50	52
24	A3741T	47	49
25	C137A	47	58
26	C2A-G17Del	46	66
27	C114A-C116A	46	43
28	T60C	45	58
29	T38G	41	55
30	T39G	40	60

Table S3. Analysis of alpha satellite sequences found in high copy number in the CP HindIII monomer dataset. The sequences are named according to the "Id" column. The "Sequence" column indicates how each sequence variant differs from the consensus sequence of the C1 family, using standard notations. Positions are numbered like the homologous positions in the XmnI phase. The "Number" column displays the number of identical copies of the sequence in the monomer dataset. The "Forward" column displays the percentage of reads obtained in the forward orientation (i.e. the orientation of our reference sequence). Strong biases for read orientation reveal artifactual sequences which are indicated on a grey background.

Supplementary Figures

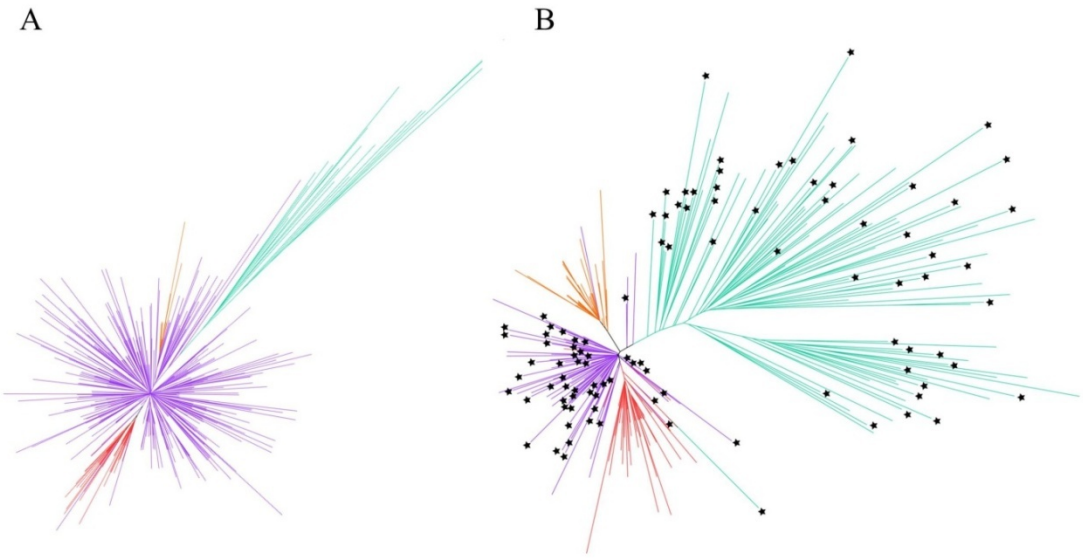

Fig. S1. Phylogenetic trees for the characterization of alpha satellite DNA families in CP. (*A*) Phylogenetic tree (Neighbor-joining method, K2P model) for 500 randomly selected sequences, colored according to their assignment to the C1 (purple), C2 (pastel green), C5 (red) or C6 (orange) alpha satellite family. (*B*) Phylogenetic tree (Neighbor-joining method, K2P model) for 50 randomly selected sequences from the CS C1 and C2 families (labeled with stars) and from the CP C1, C2, C5 and C6 families. The color code matches the one described for (*A*).

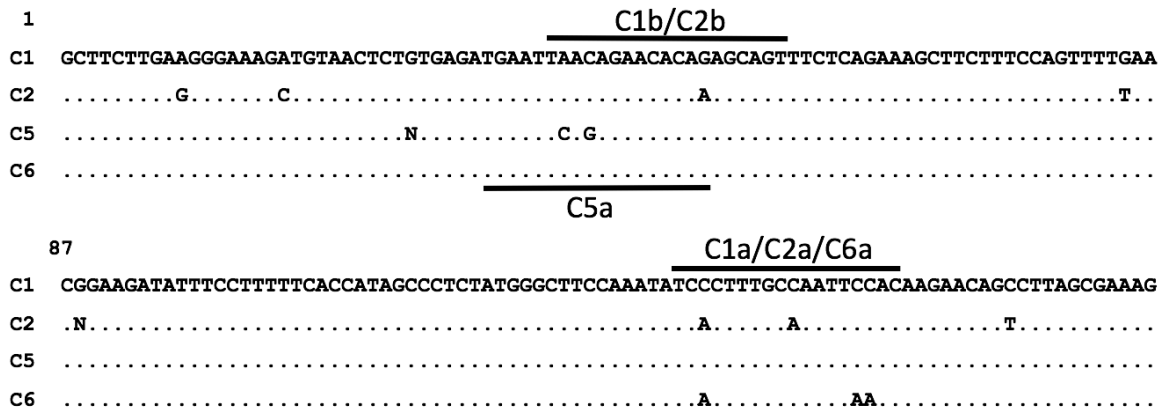

Fig. S2. Localization of the sequences that are recognized by the different probes used in the FISH experiments.

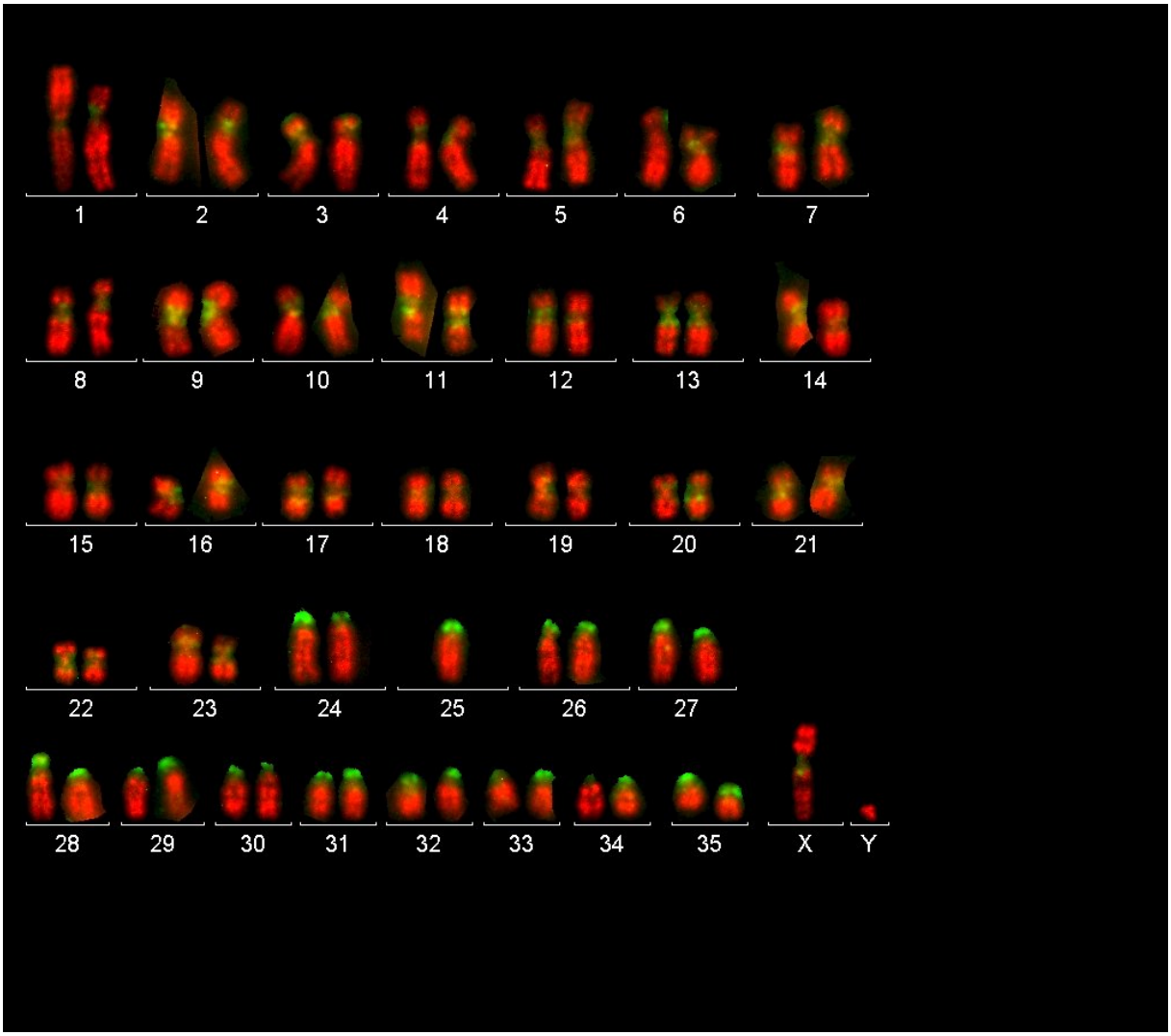

Fig. S3. Distribution pattern of the C2 alpha satellite family on CP chromosomes.

Metaphase chromosomes are counterstained with propidium iodide (2 µg/mL) and shown here in red, while the green colors stand for probe C2b. RBG chromosome banding techniques were employed as in Moulin et al. (2008). Observation was performed with an epifluorescent microscope (Microphot-FXA, Nikon) and images were captured using a cooled CCD camera (ProgRes MFcool, Jenoplik). The metaphase was karyotyped using the Isis 5.3 software (Metasystems, Altussheim, Germany) according to Dutrillaux et al. (1979).

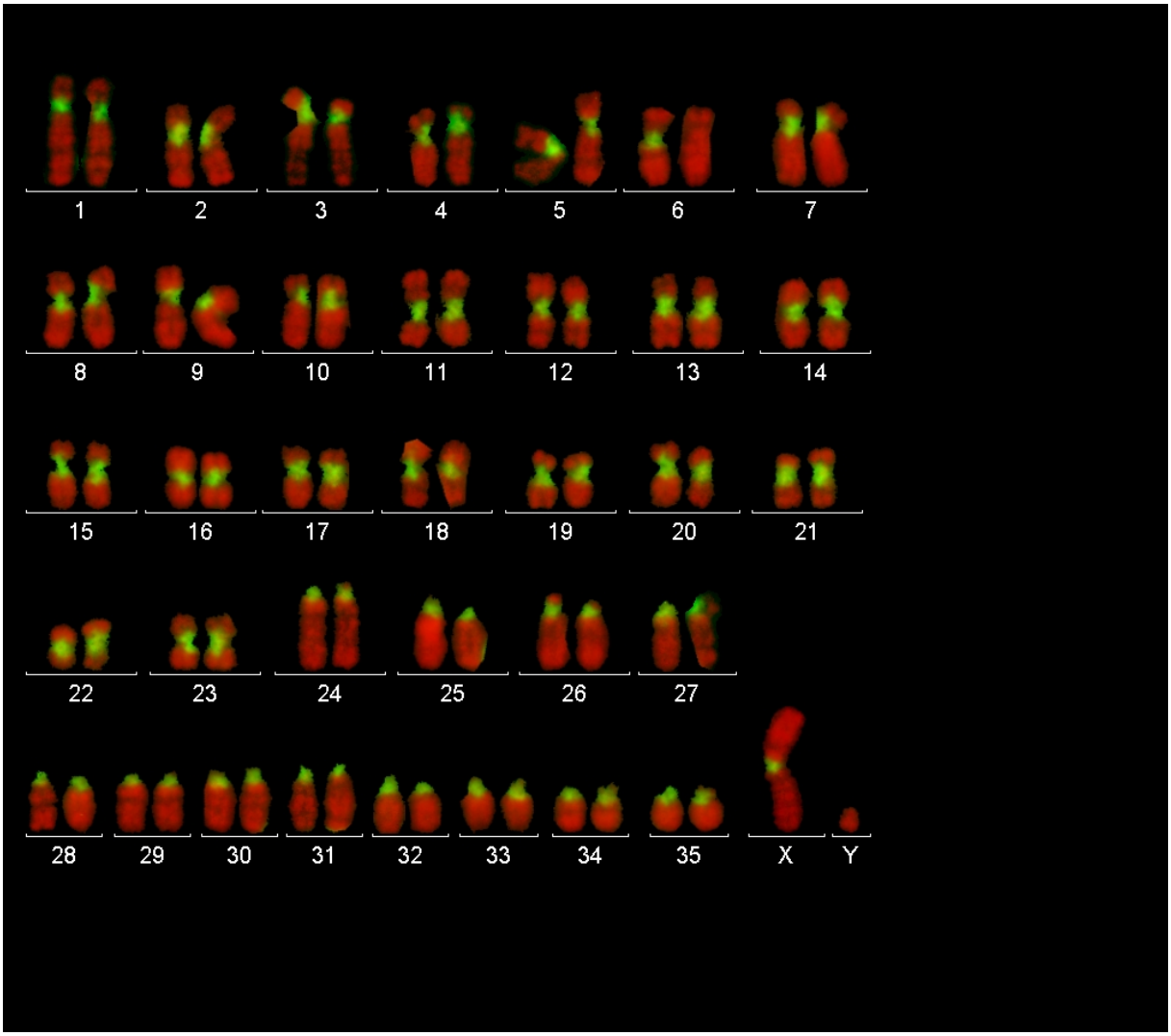

Fig. S4. Characterization of CP chromosomes that are devoid of alpha satellite DNA. Probe Cx (5'tctcagaaagctt3') was designed to target a short region that is highly conserved between all alpha satellite sequences from CS (Cacheux et al 2016). The same property is preserved with CP alpha satellites (see table S2). Hybridization of the probe is shown in green while chromosomes are shown in red. The Y chromosome and one chromosome 6 are not labeled (See the legend of supplementary fig. S3 for more information about methods)

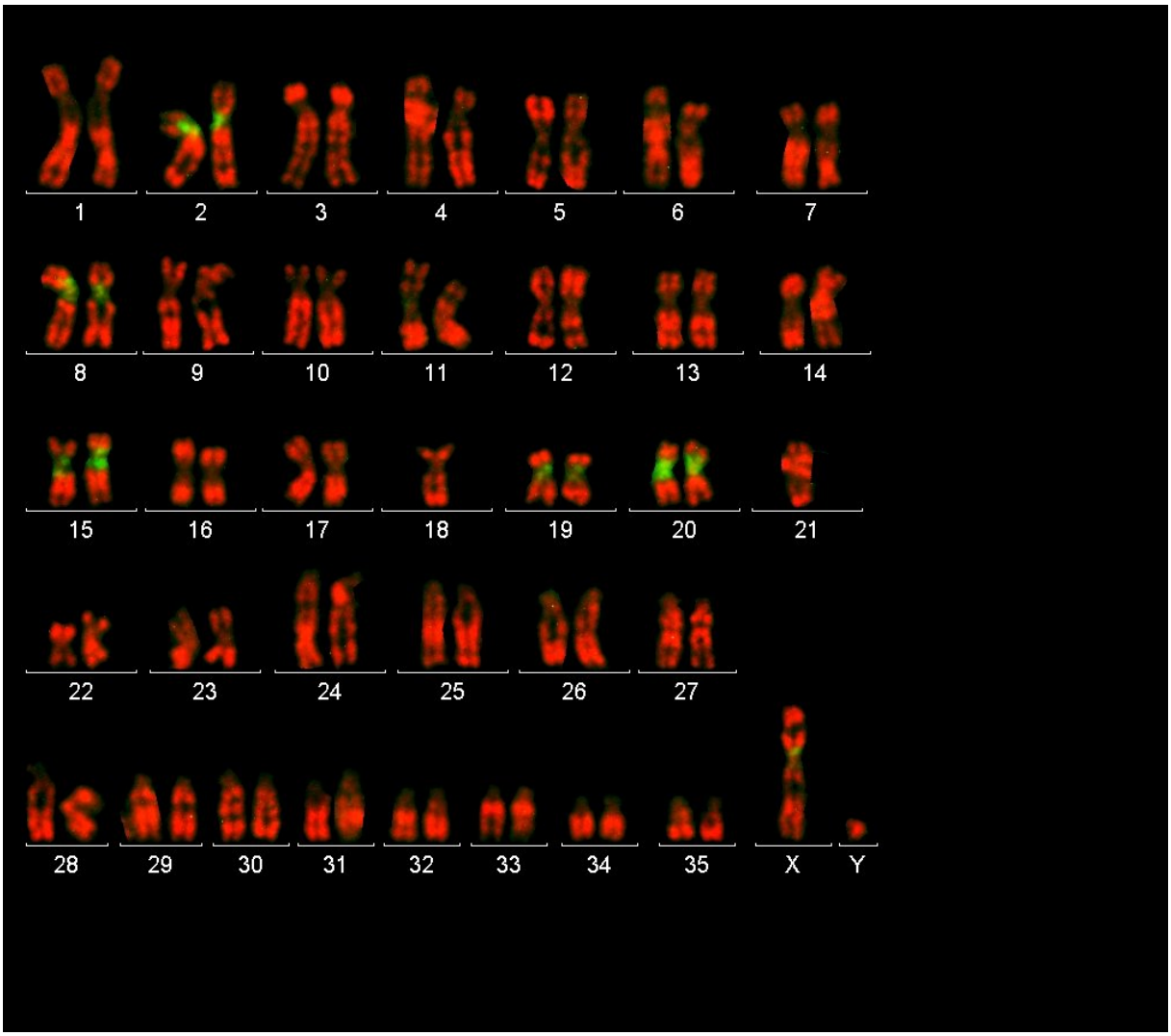

Fig. S5. Distribution pattern of the C5 alpha satellite family on CP chromosomes. Hybridization of probe C5a (green) on metaphase chromosomes (red). (See the legend of supplementary fig. S3 for more information about methods)

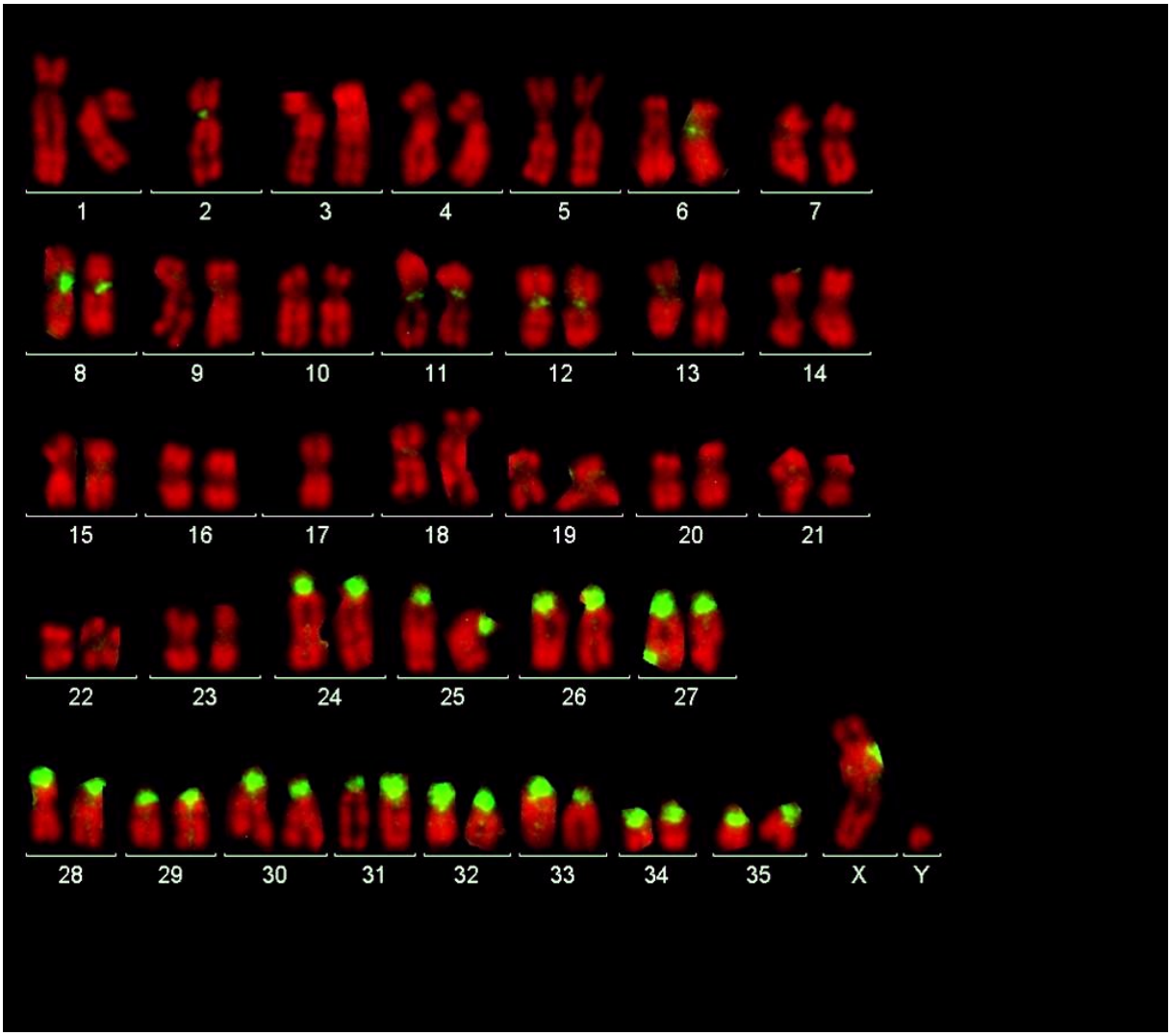

Fig. S6. Distribution pattern of the C6 alpha satellite family on CP chromosomes. Hybridization of probe C6a (green) on metaphase chromosomes (red). The analysis of several metaphases allowed determining that only one chromosome 2 and one chromosome 6 are labeled by probe C6a in the studied specimen. Apparent C6a signals on the long arm of left chromosome 27 and on chromosome X are artifacts due to chromosome superposition in the original metaphase. (See the legend of supplementary fig. S3 for more information about methods)

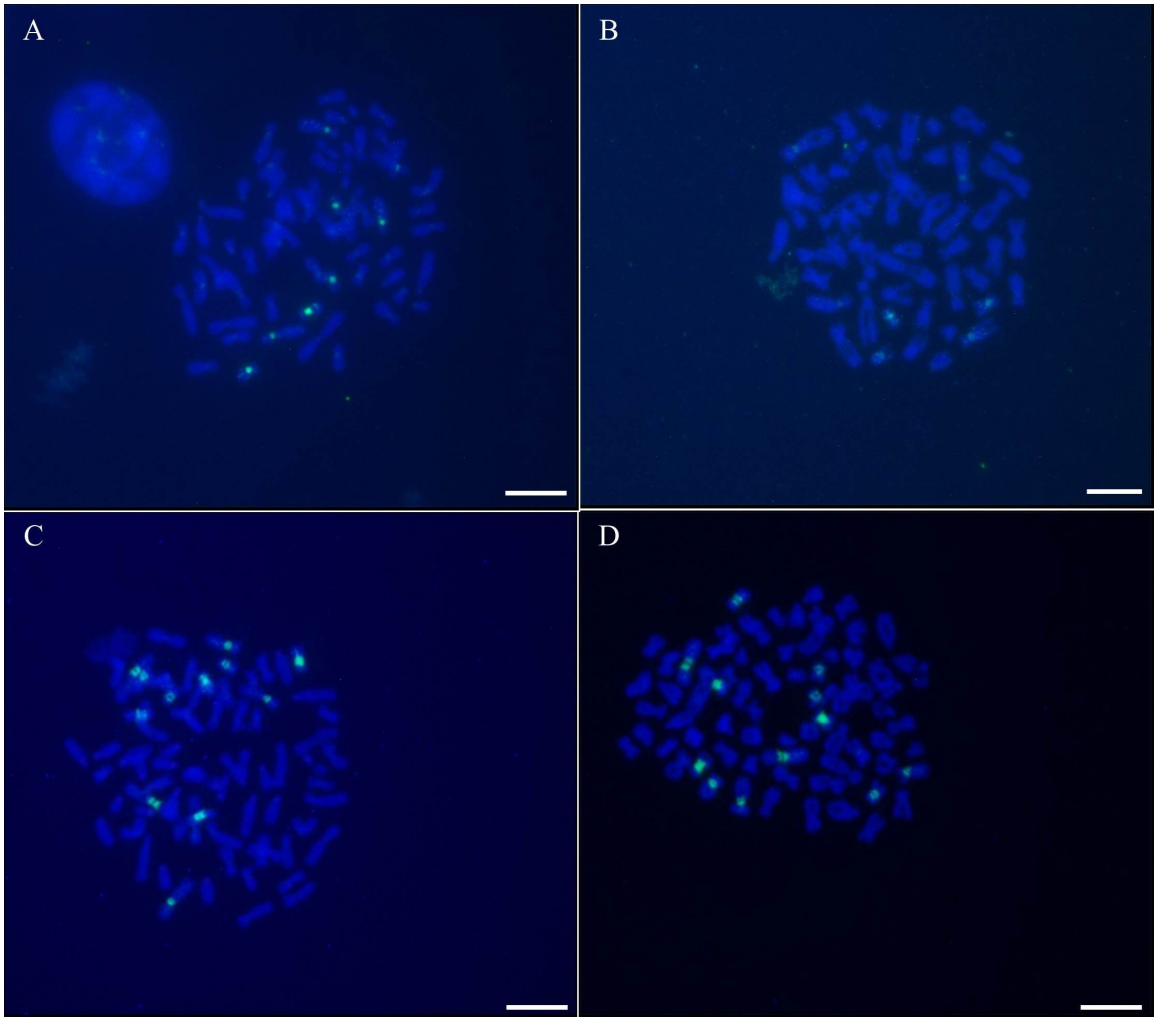

Fig. S7. Evidence for the absence of specific hybridization of the C5a probe on CS chromosomes. (*A*) and (*B*) Metaphase chromosomes from CS (ID: 2012– 028, male sample, as in our prevous study). (*C*) and (*D*) Metaphase chromosomes from CP (ID: 1979-013). A washing step is performed after probe hybridization, either at 63°C (*A* and *C*), or at 68°C (*B* and *D*). Chromosomes are colored in blue and signal from probe C5a is shown in green. A signal is observed on CS chromosomes when washing is performed at 63°C, but increasing this temperature to 68°C leaves only residual signal. In contrast, the signal observed on CP chromosomes isnot affected by the increase in temperature of the washing step. The residual signal on CS chromosomes may be explained by non-specific hybridization of the C5a probe on repeated sequences from the CS genome. Our previous work had identified the presence of a sequence variant (A40C) on 4 chromosome pairs of CS, that will be detected by the C5a probe if one mismatch is allowed. Taken together, these results suggest that the C5 family is absent from the CS genome.

```
      XmnI                                                                                HindIII
      1                                                                                       86
C6    GCTTCTTGAAGGGAAAGATGTAACTCTGTGAGATGAATTAACAGAACACAGAGCAGTTTCTCAGAAAGCTTCTTTCCAGTTTTGAA
C6'   -----------------------------------------------------------------AAGCTTCTTTCCAGTTTTGAA

      87                                                                                     172
C6    CGGAAGATATTTCCTTTTTCACCATAGCCCTCTATGGGCTTCCAAATATCACTTTGCCAATTAAACAAGAACAGCCTTAGCGAAAG
C6'   CGGAAGATATTTCCTTTTTCACCATAGCCCTCTATGGGCTTCCAAATATCACTTTGCCAATTAAACAAGAACAGCCTTAGCGAAAG

      1
C6    --------------------------------------------------------------------------------------
C6'   GATTCTTGAAGGGAAA-ATGTAACTCTGTGAGATGAATTAACAGAACACAGAGCAGTTTCTCAGA---------------------
```

Fig. S8 : Alignment of the consensus sequences of the C6 and C6' families.

The positions of the XmnI and HindIII restriction sites are highlighted in yellow and green, respectively. The differences between the homologous fragments are highlighted in red.

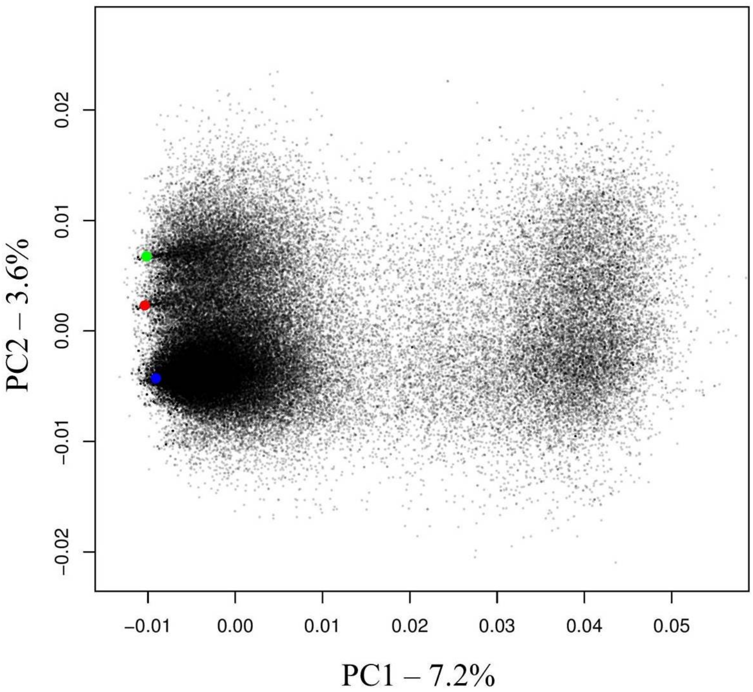

Fig. S9. Distinction of comet-like clusters within the CSXmnI monomer dataset. Prediction of the PCA projection of the normalized 5-mer frequency vectors from the CSXmnI monomer dataset within the axis system defined by the principal components 1 and 2 of the PCA from the CP XmnI monomer dataset. CS sequences corresponding to the C1 consensus and those containing the T39G variation or the A40C variations are spotted in blue, red and green, respectively (Cacheux et al., 2016).

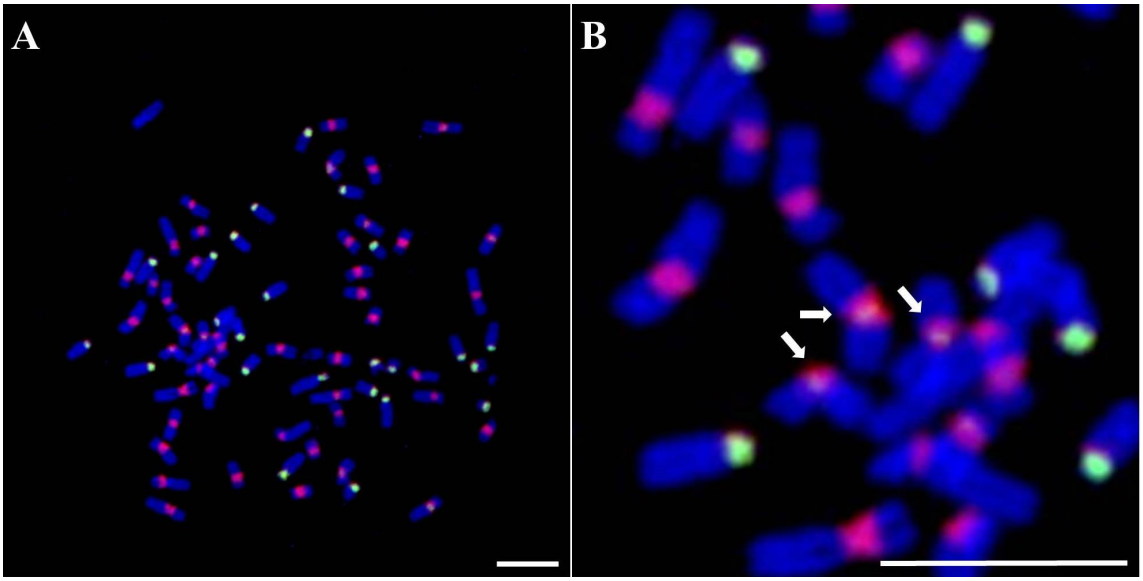

Fig. S10. Chromosomal distribution of the sequence variants containing a C or a G at position 158. (*A*) Probes 158C (red) and 158G (green) are hybridized simultaneously to CP chromosomes (blue). (*B*) Focus on image (*A*) showing that, in addition to strong signals at the centromere of acrocentrics, the 158G probe displays slighter signals (arrows) at the core centromere of several other chromosomes. Scale bar = 10 μm.