Supplemental Material

Reconstructing double-stranded DNA fragments on a single molecule level reveals patterns of degradation in ancient samples

Lukas Bokelmann, Isabelle Glocke and Matthias Meyer

Table of Contents

Figure S1: Background corrected reference base composition (sequence logo plots) and substitution frequencies in sequences from Vindija 33.19 in different structural contexts as indicated by schematic representations. 2

Figure S2: Background corrected C-to-T substitution frequencies and sequence logo plot of the reference base composition around putative 1-nt gaps in Neanderthal DNA fragments. 3

Figure S3: Background corrected C-to-T substitution frequencies and sequence logo plot of the reference base composition around Neanderthal molecules with 20-nt 5' overhangs. 4

Figure S4: Uncorrected C-to-T substitution frequencies and sequence logo plot of the reference base composition around Neanderthal molecules with 20-nt 3' overhangs. 5

Figure S5: Uncorrected C-to-T (black line) and CpG-to-TpG (green dashed line) substitution frequencies in sequences from Vindija 33.19 in different structural contexts as indicated by schematic representations. 6

Figure S6: Non-C-to-T substitutions observed in all sequences vs. sequences located in 1-nt 5' or 3' overhangs in the Vindija 33.19 library. 7

Figure S7: Reconstruction of double-stranded DNA fragments depending on sequencing depth in a Neanderthal library (A18947, Vindija 33.19). 8

Figure S8: C-to-T substitution frequency in 1-nt 5' overhangs vs. age in libraries from different Neanderthal specimens (N = 6). 9

Figure S9: C-to-T substitution frequency in 5' blunt ends vs. age in libraries from different Neanderthal specimens (N = 6). 10

Figure S10: Reference base composition in 1-nt 5' overhangs vs. age in libraries from different Neanderthal specimens (N = 6). 11

Figure S11: Background corrected C-to-T (black line) and CpG-to-TpG (green dashed line) substitution frequencies in sequences combined from 5 Neanderthal specimens (Hohlenstein-Stadel, Les Cottés Z4-1514, Scladina I-4A, Spy 94a and Sima de los Huesos) in different structural contexts as indicated by schematic representations. 12

Figure S12: Identification of sequences from DNA strands originating from the same double-stranded DNA molecules in a library constructed with the SRSLY library preparation method from cfDNA isolated from blood plasma. 13

Figure S13: Sequence logo plots of the reference base composition around molecule ends of a library constructed with the SRSLY library preparation protocol from cfDNA isolated from blood plasma. 14

Table S1: Sequencing summary statistics for the Vindija 33.19 Neanderthal DNA library. 15

Table S2: Number of sequences inferred to be located in a 1-nt 5'-overhang in sequence data from different Neanderthal specimens, as well as in the control data set down-sampled to an equal number of aligned sequences. 15

Table S3: Nucleotide substitution frequencies at the 5' terminus (position 0) and the penultimate position (position 1) of sequences from Vindija 33.19 in different structural contexts. 16

Table S4: Nucleotide substitution frequencies at the 3' terminus (position 0) and the penultimate position (position -1) of sequences from Vindija 33.19 in different structural contexts. 16

Table S5: Number of sequences inferred to be located in different contexts in previously published sequence data from cfDNA isolated from blood plasma (Troll et al. 2019), as well as in the control data down-sampled to an equal number of aligned sequences. 16

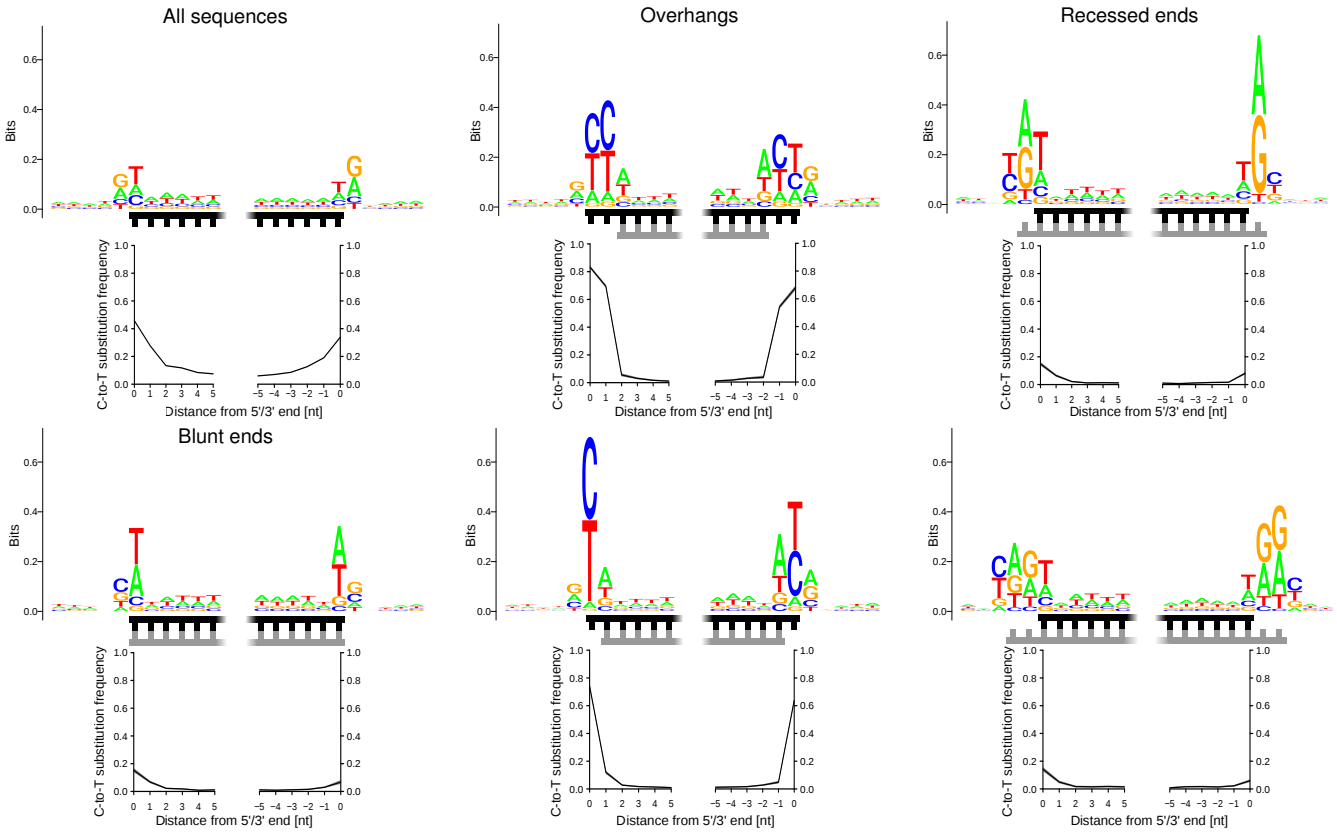

Fig. S1. Background corrected reference base composition (sequence logo plots) and substitution frequencies in sequences from Vindija 33.19 in different structural contexts as indicated by schematic representations.

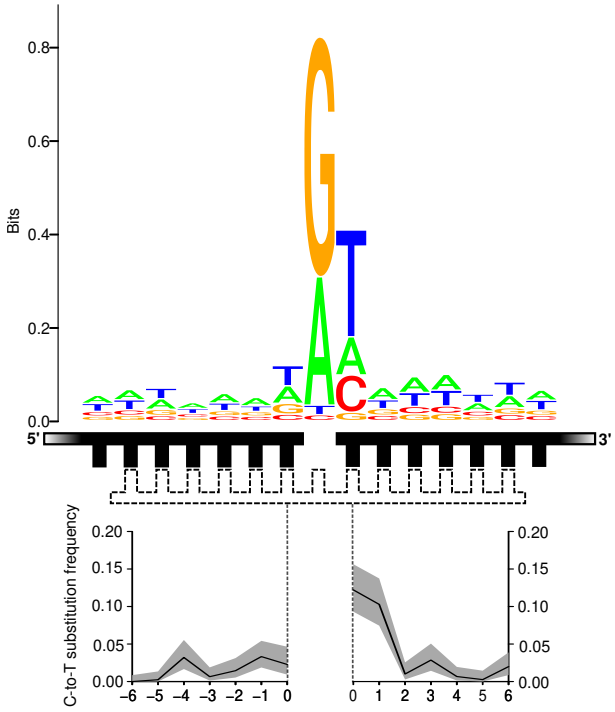

Fig. S2. Background corrected C-to-T substitution frequencies and sequence logo plot of the reference base composition around putative 1-nt gaps in Neanderthal DNA fragments. The strands used for calculating reference base compositions and substitution frequencies are marked black in the schematic drawing, the complementary strand that was putatively present in white. The grey area surrounding the C-to-T substitution frequencies denotes the 95% binomial confidence interval.

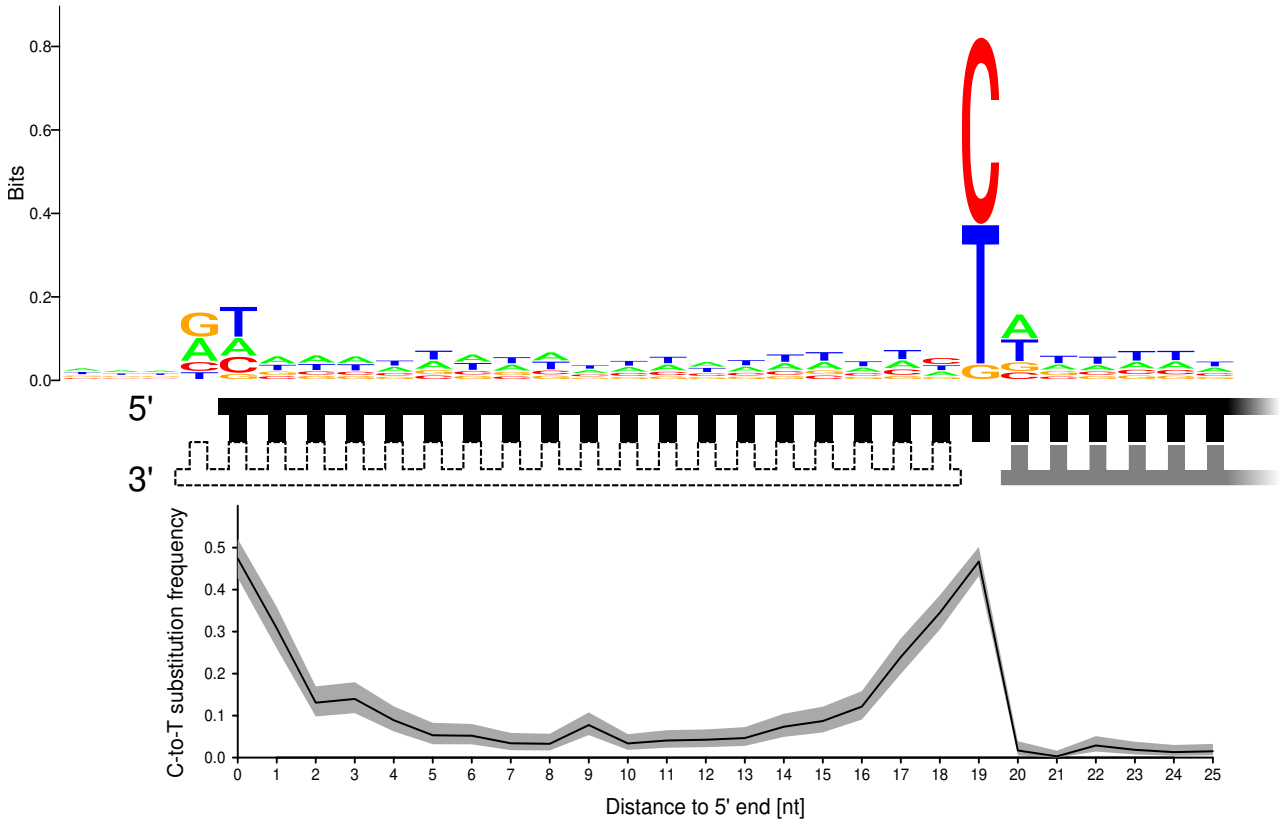

Fig. S3. Background corrected C-to-T substitution frequencies and sequence logo plot of the reference base composition around Neanderthal molecules with 20-nt 5' overhangs. A schematic representation of the alignments shows the strand used for calculating reference base composition and substitution frequencies (marked in black), the identified complementary strand (marked in grey) and a second complementary strand of unknown length that was putatively present (white fill). Note that the size of the gap is unknown. The grey area surrounding the C-to-T substitution frequencies denotes the 95% binomial confidence interval.

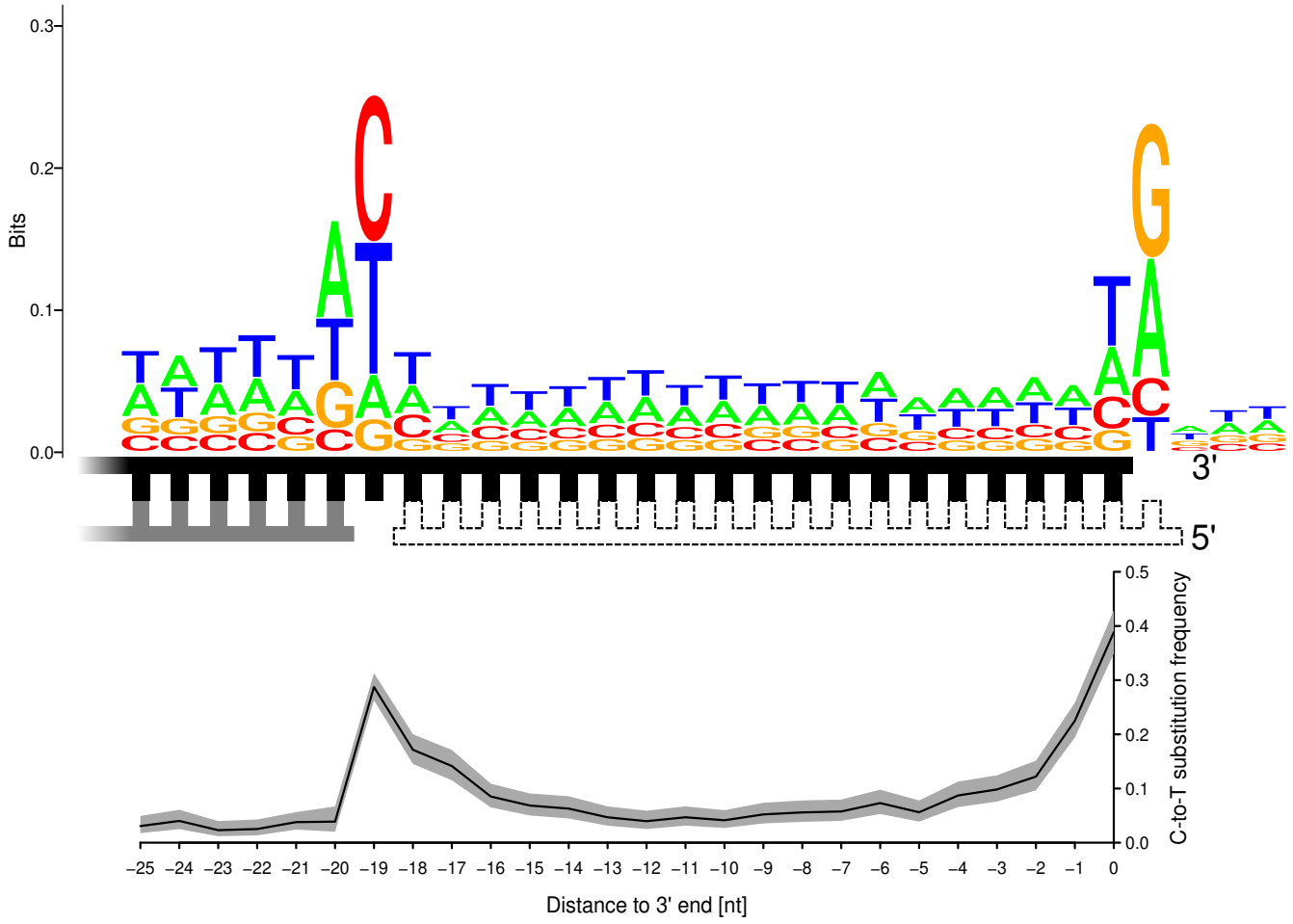

Fig. S4. Uncorrected C-to-T substitution frequencies and sequence logo plot of the reference base composition around Neanderthal molecules with 20-nt 3' overhangs. A schematic representation of the alignments shows the strand used for calculating reference base composition and substitution frequencies (marked in black), the identified complementary strand (marked in grey) and a second complementary strand of unknown length that was putatively present (white fill). Note that the size of the gap is unknown. The grey area surrounding the C-to-T substitution frequencies denotes the 95% binomial confidence interval.

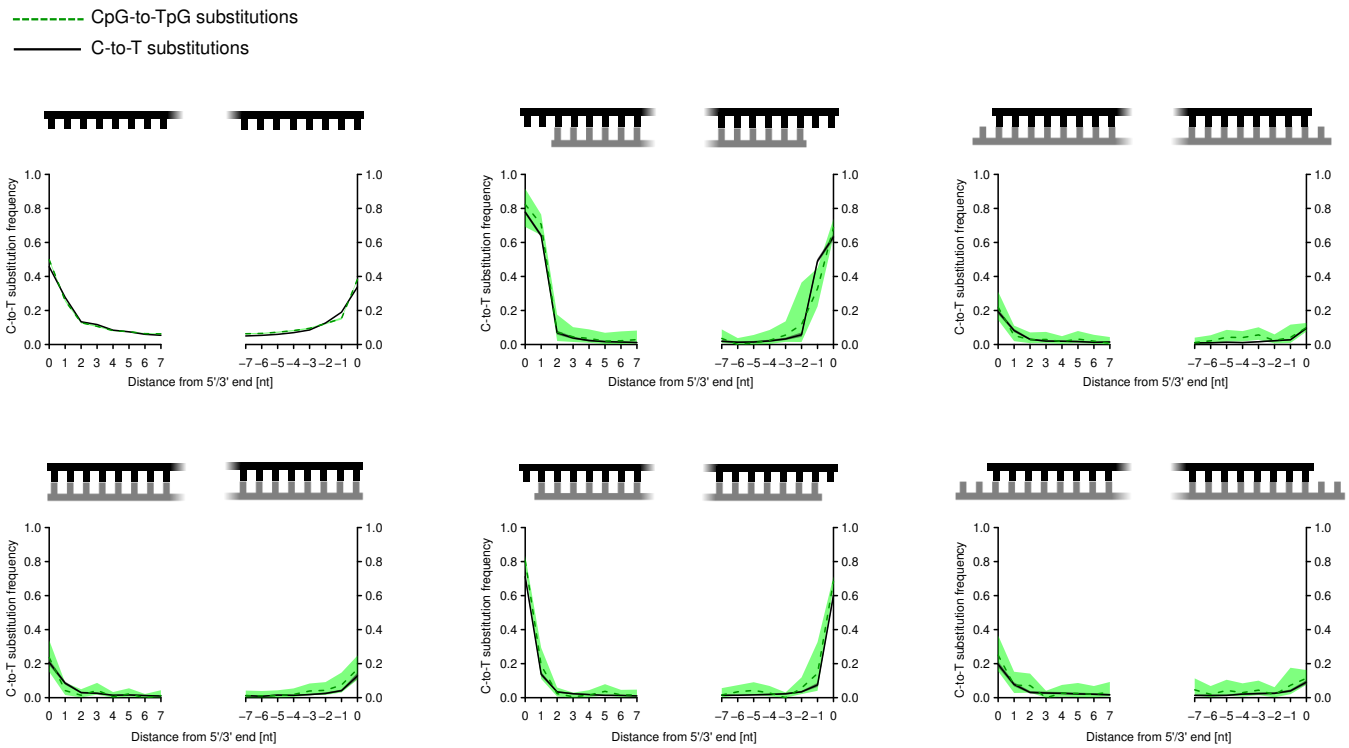

Fig. S5. Uncorrected C-to-T (black line) and CpG-to-TpG (green dashed line) substitution frequencies in sequences from Vindija 33.19 in different structural contexts as indicated by schematic representations. 95% binomial confidence intervals are indicated by shadings around the lines.

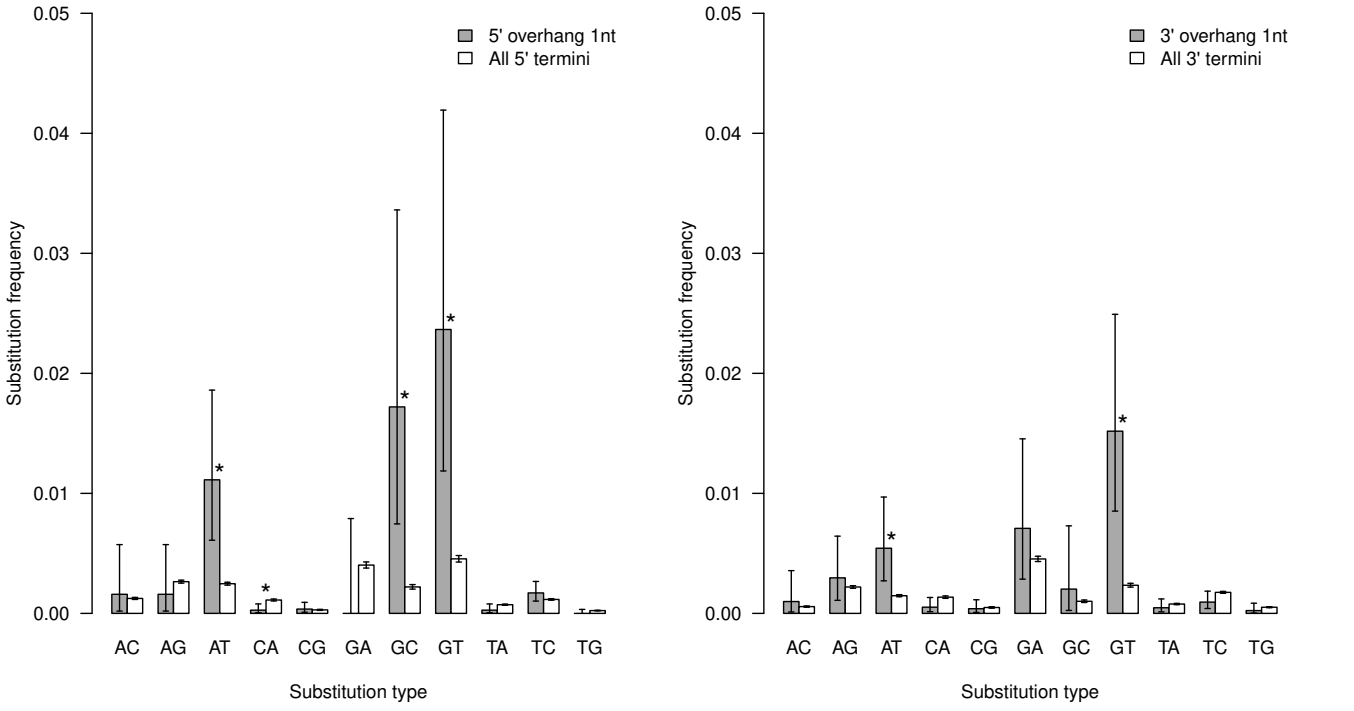

Fig. S6. Non-C-to-T substitutions observed in all sequences vs. sequences located in 1-nt 5' or 3' overhangs in the Vindija 33.19 library. Error bars indicated 95% binomial confidence intervals. Substitutions that are significantly more abundant in the respective overhang are indicated by an asterisk together with the p-value (two-tailed Fisher's test including Bonferroni correction for multiple testing (N = 11), p-values < 0.05).

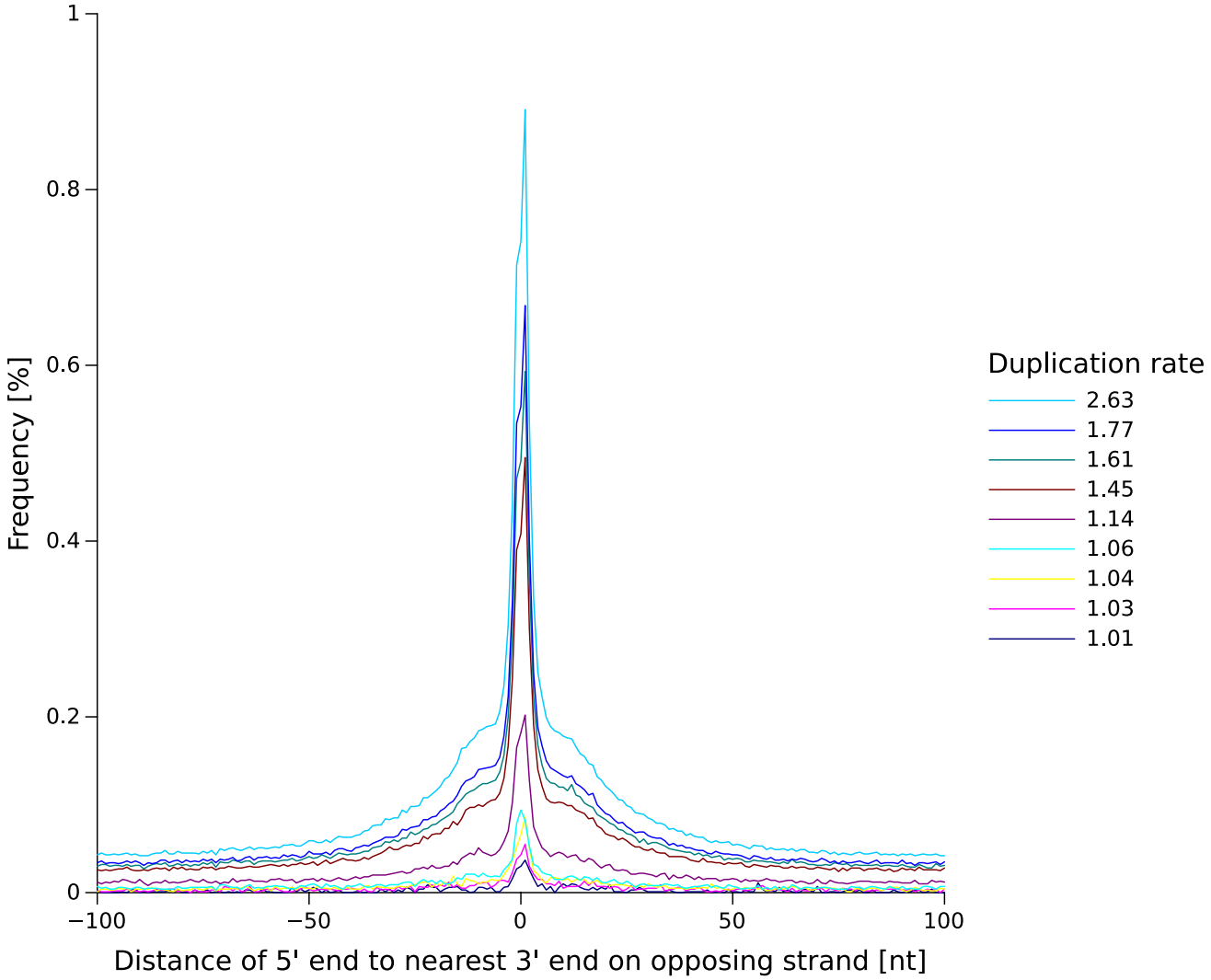

Fig. S7. Reconstruction of double-stranded DNA fragments depending on sequencing depth in a Neanderthal library (A18947, Vindija 33.19). Sequences from the Vindija 33.19 library were randomly subsampled to a fraction of the full dataset and the distribution of distances computed between the 5' end of each sequence and the nearest 3' end on the opposing strand. The duplication rate denotes the average number of observations for each unique hominin DNA sequence.

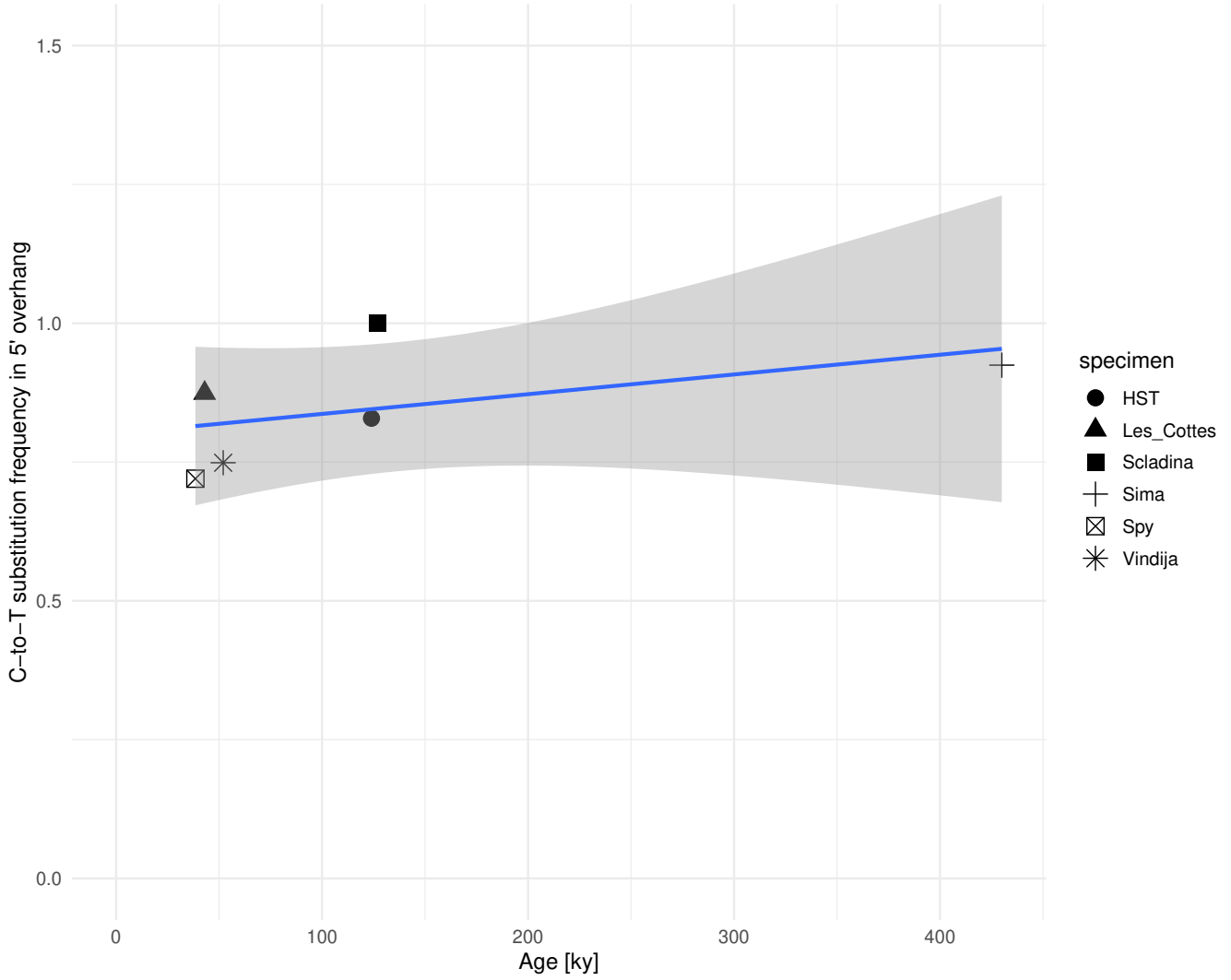

Fig. S8. C-to-T substitution frequency in 1-nt 5' overhangs vs. age in libraries from different Neanderthal specimens (N = 6). As some libraries show substantial levels of human DNA contamination, the analysis was restricted to putatively endogenous, deaminated fragments by requiring a C-to-T substitution at the 3' end of sequences. Background noise was subtracted using the signal/noise ratio determined via comparison to a modern control sample (see Table S1). For specimens for which more than one library were available, the library producing the largest number of DNA fragments with an inferred 1-nt 5' overhang was chosen. No significant correlation between deamination and sample age was found (one-tailed F-statistic linear regression on 1 and 4 degrees of freedom: 1.343, p-value: 0.31 (R2: 0.2514)).

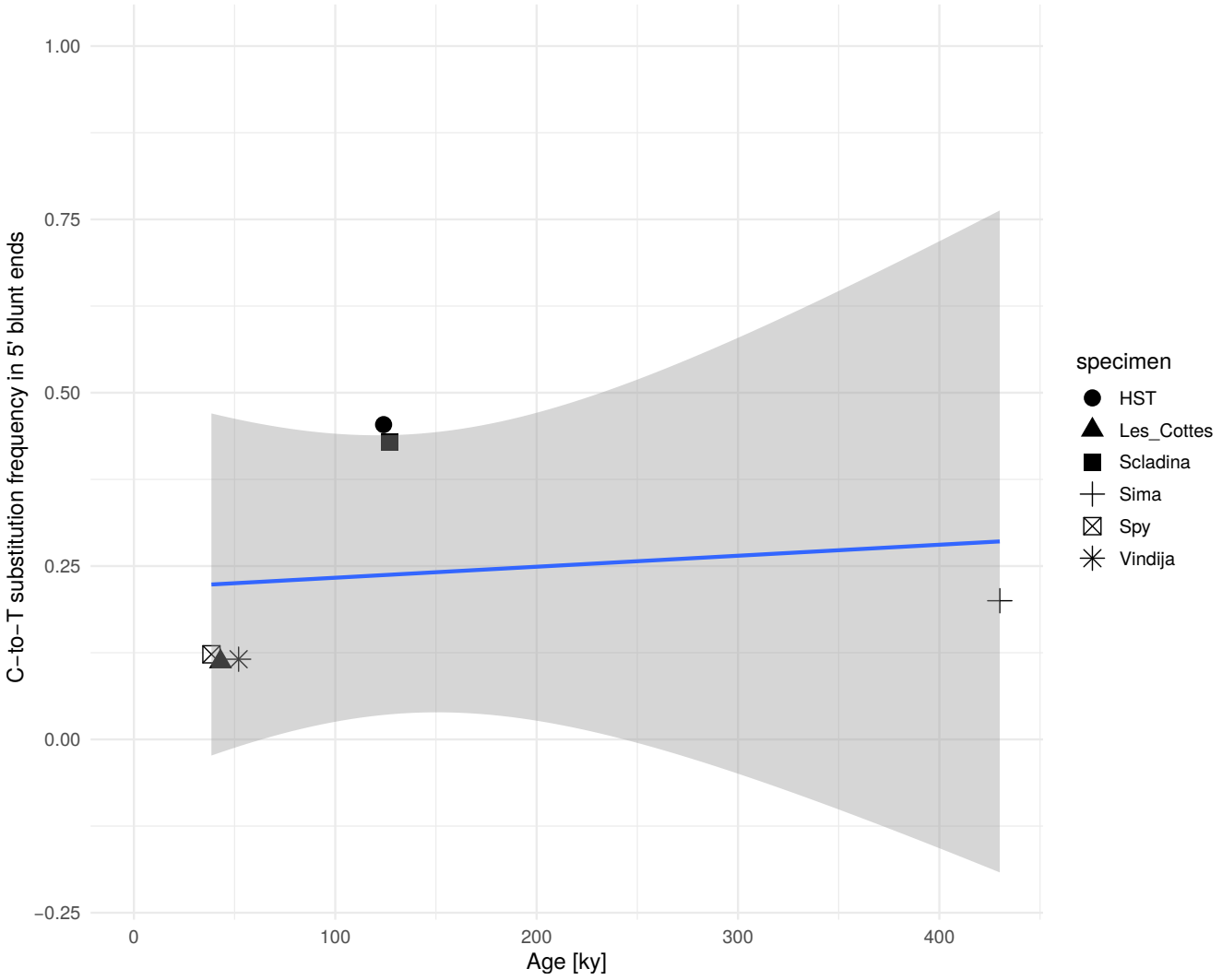

Fig. S9. C-to-T substitution frequency in 5' blunt ends vs. age in libraries from different Neanderthal specimens (N = 6). As some libraries show substantial levels of human DNA contamination, the analysis was restricted to putatively endogenous, deaminated fragments by requiring a C-to-T substitution at the 3' end of sequences. Background noise was subtracted using the signal/noise ratio determined via comparison to a modern control sample (see Table S1). For specimens for which more than one library were available, the library producing the largest number of DNA fragments with an inferred 1-nt 5' overhang was chosen. No significant correlation between deamination and sample age was found (one-tailed F-statistic linear regression on 1 and 4 degrees of freedom: 0.0896, p-value: 0.78 (R2: 0.0219)).

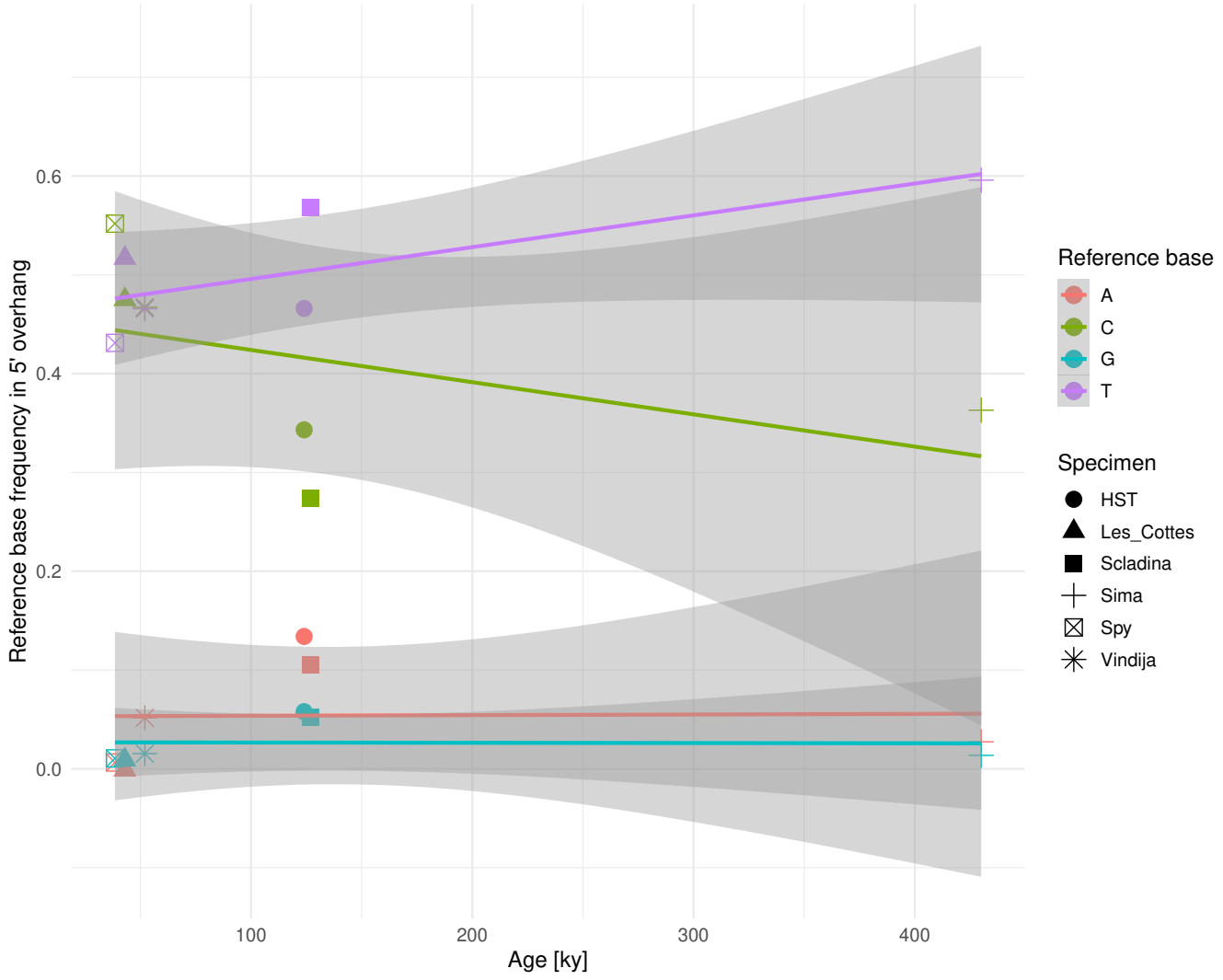

Fig. S10. Reference base composition in 1-nt 5' overhangs vs. age in libraries from different Neanderthal specimens (N = 6). As some libraries show substantial levels of human DNA contamination, the analysis was restricted to putatively endogenous, deaminated fragments by requiring a C-to-T substitution at the 3' end of sequences. Background noise was subtracted using the signal/noise ratio determined via comparison to a modern control sample (see Table S1). For specimens for which more than one library were available, the library producing the largest number of DNA fragments with an inferred 1-nt 5' overhang was chosen. No significant correlation between base composition and sample age was found (one-tailed F-statistic linear regression on 1 and 4 degrees of freedom, for "A": F = 0.0012, p-value: 0.97 (R2: 0.0003), for "C": F = 1.161, p-value: 0.34 (R2: 0.2249), for "G": F = 0.00108, p-value: 0.98 (R2: 0.00027), for "T": F = 4.979, p-value: 0.09 (R2: 0.5545)).

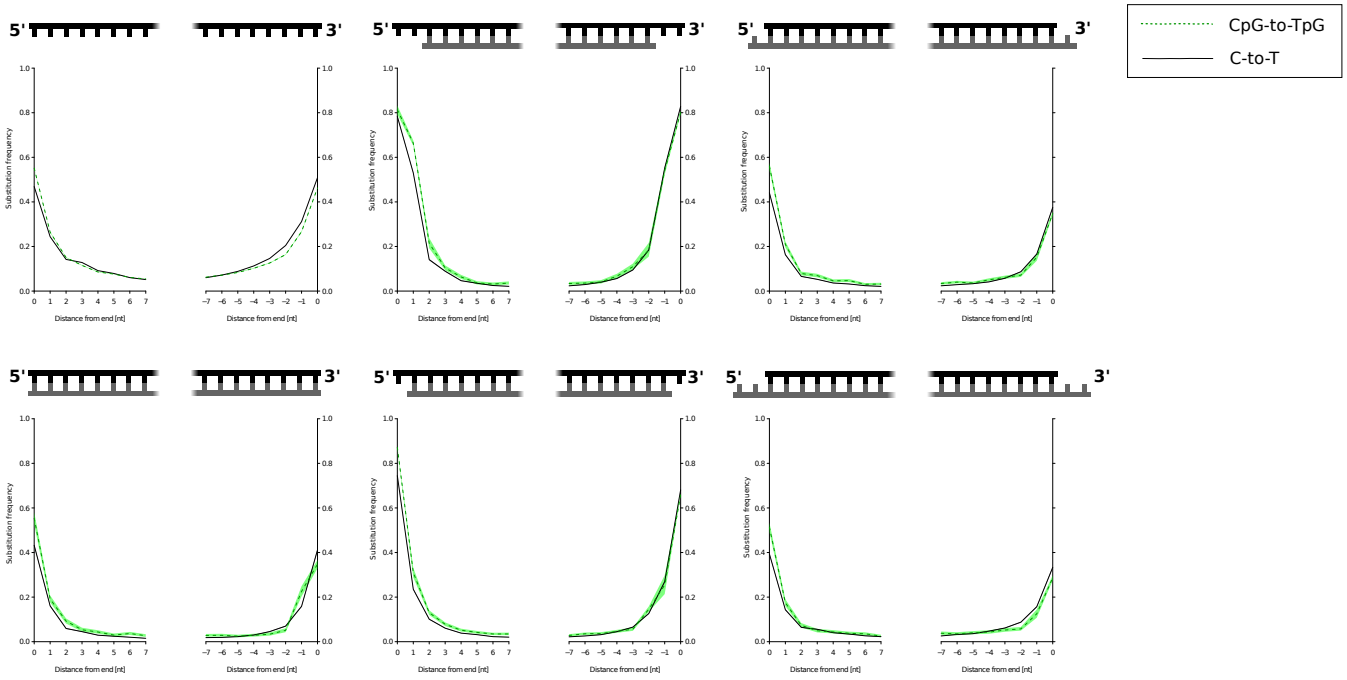

Fig. S11. Background corrected C-to-T (black line) and CpG-to-TpG (green dashed line) substitution frequencies in sequences combined from 5 Neanderthal specimens (Hohlenstein-Stadel, Les Cottés Z4-1514, Scladina I-4A, Spy 94a and Sima de los Huesos) in different structural contexts as indicated by schematic representations. 95% binomial confidence intervals are indicated by shadings around the lines.

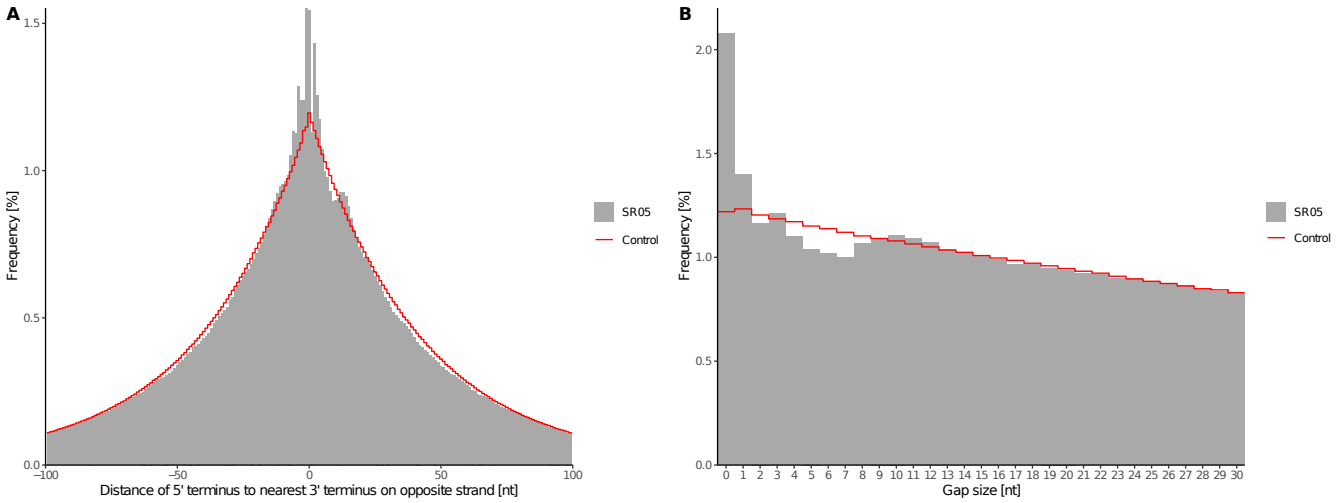

Fig. S12. Identification of sequences from DNA strands originating from the same double-stranded DNA molecules in a library constructed with the SRSLY library preparation method from cfDNA isolated from blood plasma. (A) Distribution of distances between the 5' end of each sequence to the nearest 3' end on the opposing strand. (B) Distribution of distances between sequence alignments on the same strand. Data from the cfDNA library (SR-05) are shown in grey bars. The red line shows the same analysis performed for an equal number of control sequences obtained from shallow sequences of artificially fragmented modern human DNA where sequences should be distributed randomly across the reference genome.

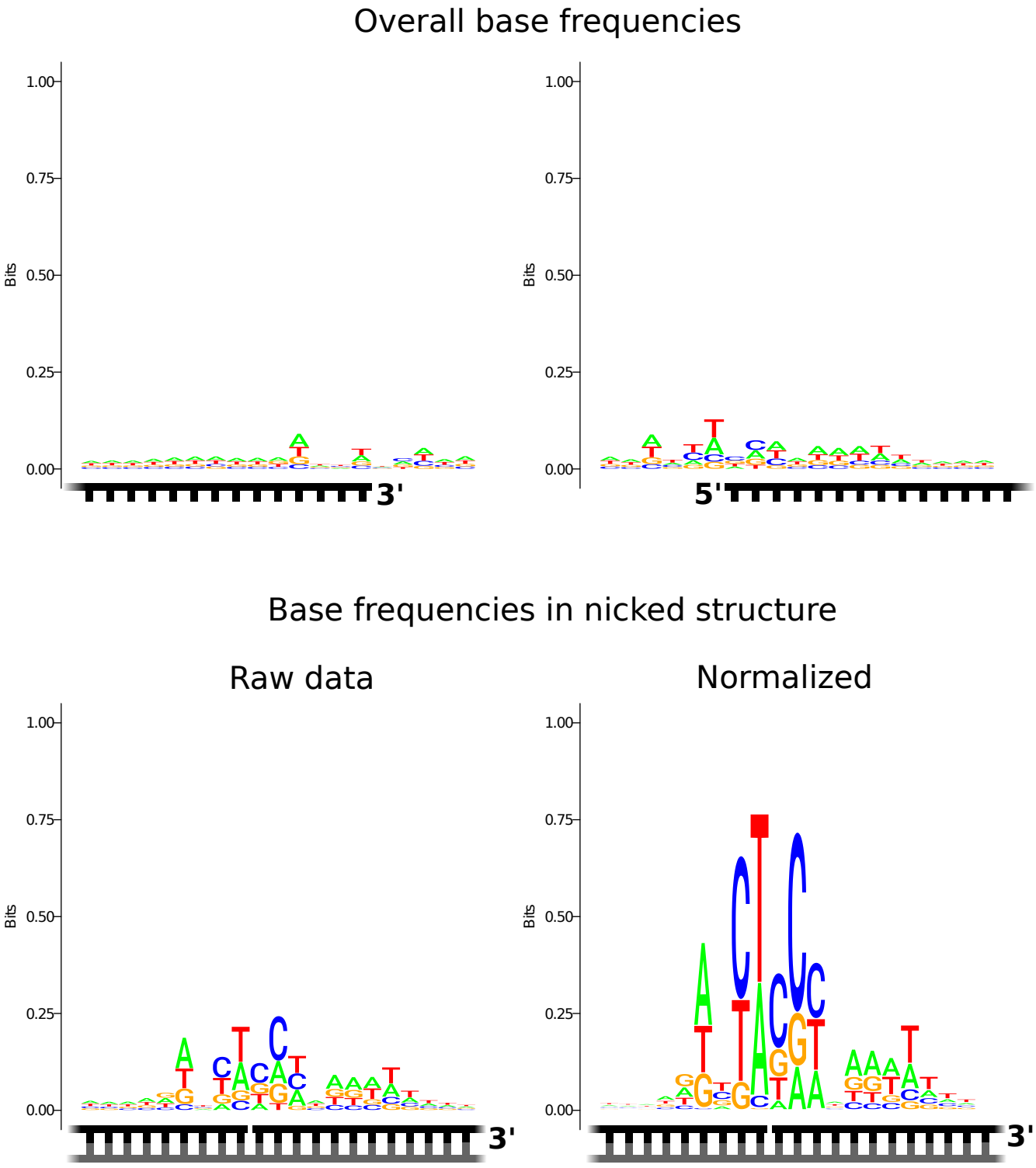

Fig. S13. Sequence logo plots of the reference base composition around molecule ends of a library constructed with the SRSLY library preparation protocol from cfDNA isolated from blood plasma. The base composition was computed for all DNA ends regardless of molecule structure (top) as well those ends that were inferred to be located in nicks, with and without background correction (bottom). A schematic representation of the alignments shows the strand used for calculating reference base composition (marked in black) and a second complementary strand of unknown length that was putatively present (marked in grey).

Table S1: Sequencing summary statistics for the Vindija 33.19 Neanderthal DNA library.

Description	Library ID	# Library molecules (qPCR)	# Sequences generated	# Sequences ≥ 35bp	# Sequences ≥ 35bp mapped to hg19 with mapping quality ≥ 25	% Sequences mapping to hg19	# Unique sequences mapping to hg19	Average sequence duplicates	Estimated number of hominin molecules in library
Library negative control	A18943	1.74E+08	1.82E+05	1,409	4	0.28	4	1	3,816
Vindija 33.19	A18947	5.56E+08	2.62E+06	111,764	7,387	6.61	7,319	1.01	1,571,115
Vindija 33.19 gel-excised	A18947 gel-excised	See above	2.35E+08	110,788,805	7,150,036	6.45	2,722,670	2.63	See above

Table S2: Number of sequences inferred to be located in a 1-nt 5'-overhang in sequence data from different Neanderthal specimens, as well as in the control data set down-sampled to an equal number of aligned sequences.

Specimen	Site	Library ID *	Genomic coverage generated [-fold]	Structural context	Sequences identified in sample DNA dataset	Sequences identified in down-sampled control dataset	Signal/noise ratio
Vindija 33.19 [this study]	Vindija cave, Croatia	A18947	0.0451	2-nt 5' overhang	14,450	1,379	10.5
			0.0451	1-nt 5' overhang	24,003	1,415	17.0
			0.0451	blunt	19,955	1,421	14.0
			0.0451	1-nt 3' overhang	19,227	1,464	13.1
			0.0451	2-nt 3' overhang	11,920	1,344	8.9
Goyet Q56-1	Troisiéme caverne, Goyet, Belgium	A9122	0.8146	1-nt 5' overhang	428,764	307,022	1.4
		A9229	0.5711	1-nt 5' overhang	153,879	123,256	1.2
		A9349	0.4978	1-nt 5' overhang	133,589	96,080	1.4
HST	Hohlenstein-Stadel cave, Germany	F7016	0.0175	1-nt 5' overhang	8,220	230	35.7
		F7017	0.0149	1-nt 5' overhang	6,043	168	36.0
		L5262	0.0001	1-nt 5' overhang	0	0	0.0
		L5263	0.0004	1-nt 5' overhang	1	0	0.0
		L5264	0.0026	1-nt 5' overhang	253	8	31.6
		L5265	0.0014	1-nt 5' overhang	13	1	13.0
		L5386	0.0230	1-nt 5' overhang	11413	446	25.6
		R5784	0.0015	1-nt 5' overhang	27	2	13.5
		R5785	0.0022	1-nt 5' overhang	43	4	10.8
Les Cottés Z4-1514	Les Cottés cave, France	A9230	0.6391	1-nt 5' overhang	891,212	273,635	3.3
		A9290	0.1838	1-nt 5' overhang	71,577	22,633	3.2
		A9291	0.1037	1-nt 5' overhang	21,048	6,740	3.1
		A9309	0.1578	1-nt 5' overhang	114,227	14,832	7.7
		A9350	0.4385	1-nt 5' overhang	436,137	132,255	3.3
		A9393	0.1890	1-nt 5' overhang	160,481	20,947	7.7
		A9394	0.1796	1-nt 5' overhang	143,471	18,777	7.6
		A9395	0.1907	1-nt 5' overhang	162,701	21,121	7.7
		A9420	0.1923	1-nt 5' overhang	174,883	21,747	8.0
Mezmaiskaya 1	Mezmaiskaya cave, Russia	R5661	0.6474	1-nt 5' overhang	266,177	310,727	0.9
		R5662	0.5521	1-nt 5' overhang	196,729	228,286	0.9
Mezmaiskaya 2	Mezmaiskaya cave, Russia	A9180	0.7403	1-nt 5' overhang	646,943	346,152	1.9
		A9288	0.3415	1-nt 5' overhang	129,027	71,157	1.8
		A9289	0.3196	1-nt 5' overhang	111,579	62,405	1.8
Scladina I-4A	Scladina cave, Belgium	F7490	0.0069	1-nt 5' overhang	1,097	23	47.7
		G2159	0.0073	1-nt 5' overhang	1,135	26	43.7
		G2160	0.0085	1-nt 5' overhang	1,558	49	31.8
		G2161	0.0083	1-nt 5' overhang	1,567	35	44.8
Sima de los Huesos	Sima de los Huesos, Spain	R1848	0.0037	1-nt 5' overhang	1,821	8	227.6
Spy 94a	Spy cave, Belgium	A9416	0.1567	1-nt 5' overhang	48,298	18,234	2.6
		A9417	0.1487	1-nt 5' overhang	43,367	16,580	2.6
		A9418	0.1576	1-nt 5' overhang	48,525	18,687	2.6
		A9419	0.1397	1-nt 5' overhang	43,889	14,624	3.0
		R5556	0.2134	1-nt 5' overhang	91,996	33,796	2.7

*Only libraries produced without uracil-DNA-glycosylase (UDG) treatment were investigated. Libraries highlighted in grey were chosen for further analyses.

Table S3: Nucleotide substitution frequencies at the 5' terminus (position 0) and the penultimate position (position 1) of sequences from Vindija 33.19 in different structural contexts. 95% bionomial confidence intervals are provided in parentheses.

Context	5p-blunt	5p-blunt	5p-overhang-1-nt	5p-overhang-1-nt	5p-overhang-2nt	5p-overhang-2nt	5p-overhang-2nt	5p-recessed-1-nt	5p-recessed-1-nt	5p-recessed-2nt	5p-recessed-2nt	5p-all	5p-all
Position	0	1	0	1	0	1	2	0	1	0	1	0	1
CT	0.2073 (0.1911-0.2243)	0.0852 (0.0762-0.095)	0.7113 (0.7028-0.7197)	0.1381 (0.1257-0.1512)	0.7781 (0.7673-0.7886)	0.6399 (0.6281-0.6517)	0.0689 (0.0578-0.0814)	0.1924 (0.1782-0.2072)	0.082 (0.0731-0.0916)	0.1953 (0.1786-0.2128)	0.0775 (0.0665-0.0896)	0.4573 (0.4561-0.4585)	0.277 (0.2758-0.2782)
AC	0.0007 (0.0002-0.0016)	0.0003 (0-0.0012)	0.0016 (0.0002-0.0057)	0.0004 (0.0001-0.0009)	0.0005 (0-0.0026)	0.0024 (0.0007-0.0062)	0.0727 (0.0202-0.1759)	0.0009 (0.0003-0.0024)	0.001 (0.0004-0.0022)	0.0017 (0.0006-0.004)	0.0003 (0-0.0015)	0.0012 (0.0012-0.0013)	0.0006 (0.0006-0.0007)
AG	0.002 (0.0011-0.0033)	0.0028 (0.0016-0.0045)	0.0016 (0.0002-0.0057)	0.0036 (0.0026-0.0049)	0.0078 (0.0046-0.0125)	0.0066 (0.0033-0.0118)	0.0005 (0.0001-0.0015)	0.0019 (0.0008-0.0037)	0.002 (0.001-0.0035)	0.0014 (0.0004-0.0035)	0.0016 (0.0006-0.0035)	0.0026 (0.0025-0.0028)	0.0031 (0.003-0.0032)
AT	0.0019 (0.001-0.0031)	0.0011 (0.0005-0.0024)	0.0111 (0.0061-0.0186)	0.0006 (0.0002-0.0012)	0.0032 (0.0013-0.0066)	0.0048 (0.0021-0.0095)	0.0019 (0.0009-0.0033)	0.0024 (0.0011-0.0043)	0.0003 (0-0.0012)	0.0031 (0.0014-0.0058)	0.0003 (0-0.0015)	0.0025 (0.0024-0.0026)	0.0009 (0.0008-0.0009)
CA	0.0017 (0.0005-0.0043)	0.0026 (0.0012-0.0049)	0.0003 (0.0001-0.0008)	0.0014 (0.0004-0.0035)	0.0005 (0.0001-0.0015)	0.0005 (0.0001-0.0014)	0.0003 (0-0.0012)	0.0003 (0-0.0019)	0.0009 (0.0002-0.0025)	0 (0-0.0017)	0 (0-0.0017)	0.0011 (0.001-0.0012)	0.0014 (0.0013-0.0015)
CG	0 (0-0.0016)	0.002 (0.0008-0.0041)	0.0004 (0.0001-0.0009)	0.0003 (0-0.0019)	0.0002 (0-0.0009)	0 (0-0.0006)	0.0011 (0.0001-0.0039)	0.0003 (0-0.0019)	0.0003 (0-0.0016)	0 (0-0.0017)	0.0014 (0.0003-0.0041)	0.0003 (0.0003-0.0003)	0.0006 (0.0005-0.0007)
GA	0.0086 (0.0045-0.015)	0.0019 (0.0008-0.0037)	0 (0-0.0079)	0.0027 (0.0014-0.0047)	0.0067 (0.0018-0.0172)	0.0011 (0-0.0061)	0.0011 (0.0001-0.0039)	0.0025 (0.0008-0.0059)	0.0017 (0.0007-0.0035)	0.0039 (0.0013-0.0091)	0.002 (0.0006-0.0046)	0.004 (0.0038-0.0043)	0.0024 (0.0023-0.0026)
GC	0.0014 (0.0002-0.0052)	0.0005 (0.0001-0.0017)	0.0172 (0.0075-0.0336)	0.0004 (0.0001-0.0016)	0 (0-0.0062)	0.0011 (0-0.0061)	0.0013 (0.0003-0.0037)	0.0025 (0.0008-0.0059)	0 (0-0.0009)	0.0016 (0.0002-0.0056)	0 (0-0.0014)	0.0022 (0.002-0.0024)	0.0004 (0.0004-0.0005)
GT	0.005 (0.002-0.0103)	0.0009 (0.0003-0.0024)	0.0237 (0.0119-0.0419)	0.0011 (0.0004-0.0026)	0.0118 (0.0048-0.0241)	0.0011 (0-0.0061)	0.0013 (0.0003-0.0037)	0.003 (0.0011-0.0066)	0.001 (0.0003-0.0025)	0.0039 (0.0013-0.0091)	0 (0-0.0014)	0.0045 (0.0043-0.0048)	0.0009 (0.0008-0.001)
TA	0.0005 (0.0001-0.0012)	0.0012 (0.0005-0.0024)	0.0003 (0.0001-0.0008)	0.0021 (0.0011-0.0036)	0 (0-0.0006)	0.0002 (0-0.001)	0.0013 (0.0003-0.0037)	0.0002 (0-0.0007)	0.0014 (0.0006-0.0028)	0.0002 (0-0.001)	0.0003 (0-0.0016)	0.0007 (0.0007-0.0008)	0.0011 (0.001-0.0012)
TC	0.0008 (0.0003-0.0017)	0.0005 (0.0001-0.0014)	0.0017 (0.001-0.0027)	0.0017 (0.0008-0.0032)	0.0009 (0.0003-0.002)	0.0004 (0-0.0013)	0.0005 (0.0001-0.0017)	0.0008 (0.0003-0.0016)	0.0011 (0.0004-0.0023)	0.0014 (0.0006-0.0028)	0.0014 (0.0005-0.0033)	0.0012 (0.0011-0.0012)	0.0012 (0.0011-0.0013)
TG	0.0003 (0.0001-0.001)	0 (0-0.0006)	0 (0-0.0003)	0.0003 (0-0.0012)	0 (0-0.0006)	0 (0-0.0007)	0.0007 (0.0001-0.002)	0.0001 (0-0.0006)	0.0002 (0-0.001)	0.0002 (0-0.001)	0.0003 (0-0.0016)	0.0002 (0.0002-0.0003)	0.0005 (0.0005-0.0006)
CG_TG	0.236 (0.1524-0.3378)	0.0417 (0.0154-0.0885)	0.8038 (0.769-0.8355)	0.1899 (0.1103-0.2938)	0.8235 (0.6913-0.916)	0.7061 (0.6424-0.7644)	0.0005 (0.0001-0.0017)	0.219 (0.1442-0.3103)	0.0526 (0.0196-0.111)	0.25 (0.1554-0.366)	0.0732 (0.0273-0.1525)	0.4982 (0.4917-0.5048)	0.258 (0.2513-0.2649)

Table S4: Nucleotide substitution frequencies at the 3' terminus (position 0) and the penultimate position (position -1) of sequences from Vindija 33.19 in different structural contexts. 95% bionomial confidence intervals are provided in parentheses.

Context	3p-blunt	3p-blunt	3p-overhang-1-nt	3p-overhang-1-nt	3p-overhang-2nt	3p-overhang-2nt	3p-overhang-2nt	3p-recessed-1-nt	3p-recessed-1-nt	3p-recessed-2nt	3p-recessed-2nt	3p-all	3p-all
Position	-1	0	-1	0	-2	-1	0	-1	0	-1	0	-1	0
CT	0.0414 (0.0356-0.0478)	0.1281 (0.1111-0.1467)	0.0746 (0.0634-0.087)	0.6015 (0.5905-0.6124)	0.0571 (0.045-0.0713)	0.4913 (0.4779-0.5048)	0.6322 (0.6149-0.6493)	0.0277 (0.0232-0.0328)	0.0961 (0.0876-0.1051)	0.0399 (0.033-0.0478)	0.0911 (0.0798-0.1034)	0.19 (0.189-0.191)	0.3386 (0.3373-0.3399)
AC	0.0003 (0-0.0012)	0.0006 (0.0002-0.0013)	0.0003 (0.0001-0.0009)	0.001 (0.0001-0.0036)	0.1176 (0.0146-0.3644)	0.0041 (0.0018-0.0081)	0.0004 (0-0.0021)	0.0004 (0.0001-0.0011)	0.0007 (0.0002-0.0018)	0 (0-0.0007)	0.0007 (0.0001-0.002)	0.0005 (0.0004-0.0005)	0.0006 (0.0005-0.0006)
AG	0.003 (0.0018-0.0047)	0.002 (0.0011-0.0031)	0.0016 (0.0009-0.0026)	0.003 (0.0011-0.0064)	0.0002 (0-0.001)	0.0046 (0.0021-0.0088)	0.0035 (0.0016-0.0066)	0.0016 (0.0009-0.0028)	0.0029 (0.0017-0.0047)	0.0012 (0.0004-0.0027)	0.0028 (0.0014-0.0049)	0.0023 (0.0022-0.0024)	0.0022 (0.0021-0.0023)
AT	0.0005 (0.0001-0.0014)	0.0011 (0.0006-0.0021)	0.0003 (0.0001-0.0009)	0.0054 (0.0027-0.0097)	0.0018 (0.0009-0.0033)	0.001 (0.0001-0.0037)	0.0027 (0.0011-0.0055)	0.0006 (0.0002-0.0014)	0.0023 (0.0012-0.0038)	0.0002 (0-0.0011)	0.0007 (0.0001-0.002)	0.0007 (0.0007-0.0008)	0.0015 (0.0014-0.0016)
CA	0.0012 (0.0004-0.0027)	0.0036 (0.0012-0.0083)	0.002 (0.0005-0.0052)	0.0005 (0.0001-0.0013)	0 (0-0.0007)	0.0009 (0.0003-0.0022)	0.0007 (0.0001-0.0023)	0.0006 (0.0001-0.0018)	0.0007 (0.0001-0.002)	0.0004 (0-0.002)	0.0017 (0.0005-0.0043)	0.0011 (0.001-0.0012)	0.0014 (0.0013-0.0015)
CG	0.0005 (0.0001-0.0017)	0.0021 (0.0004-0.0062)	0.001 (0.0001-0.0036)	0.0004 (0.0001-0.0011)	0.0031 (0.0009-0.008)	0 (0-0.0007)	0.0003 (0-0.0018)	0.0002 (0-0.0012)	0.0002 (0-0.0012)	0.0004 (0-0.002)	0 (0-0.0016)	0.0005 (0.0005-0.0006)	0.0005 (0.0004-0.0006)
GA	0.0011 (0.0003-0.0029)	0.006 (0.0033-0.01)	0.0055 (0.0031-0.0089)	0.0071 (0.0029-0.0145)	0 (0-0.0029)	0.0059 (0.0022-0.0127)	0.006 (0.0019-0.0138)	0.0027 (0.0014-0.0046)	0.0042 (0.0021-0.0072)	0.0049 (0.0027-0.0082)	0.0027 (0.0009-0.0062)	0.0034 (0.0033-0.0036)	0.0045 (0.0043-0.0048)
GC	0.0014 (0.0005-0.0033)	0.0013 (0.0003-0.0037)	0.0003 (0-0.0019)	0.002 (0.0002-0.0073)	0.0033 (0.0013-0.0068)	0.002 (0.0002-0.007)	0 (0-0.0044)	0.0009 (0.0002-0.0023)	0.0007 (0.0001-0.0025)	0.0003 (0-0.0019)	0.0011 (0.0001-0.0039)	0.0009 (0.0008-0.001)	0.001 (0.0009-0.0011)
GT	0.0026 (0.0012-0.0049)	0.0038 (0.0018-0.0073)	0.0017 (0.0006-0.004)	0.0152 (0.0085-0.0249)	0.0009 (0.0001-0.0034)	0.0059 (0.0022-0.0127)	0.006 (0.0019-0.0138)	0.002 (0.0009-0.0038)	0.001 (0.0002-0.003)	0.0003 (0-0.0019)	0.0005 (0-0.003)	0.0012 (0.0011-0.0013)	0.0023 (0.0022-0.0025)
TA	0.0016 (0.0008-0.003)	0.0008 (0.0003-0.0017)	0.0014 (0.0005-0.0031)	0.0005 (0.0001-0.0012)	0.0019 (0.0005-0.0048)	0.0017 (0.0006-0.0036)	0.0013 (0.0005-0.0027)	0.0008 (0.0002-0.0018)	0.0003 (0.0001-0.0008)	0 (0-0.001)	0.0002 (0-0.0009)	0.0011 (0.001-0.0012)	0.0008 (0.0007-0.0008)
TC	0.0028 (0.0016-0.0044)	0.0016 (0.0008-0.0028)	0.0005 (0.0001-0.0017)	0.0009 (0.0004-0.0019)	0.0014 (0.0004-0.0035)	0.0025 (0.0011-0.0047)	0.002 (0.001-0.0036)	0.0005 (0.0001-0.0013)	0.0023 (0.0015-0.0034)	0.0013 (0.0004-0.003)	0.002 (0.001-0.0035)	0.0015 (0.0014-0.0016)	0.0018 (0.0017-0.0018)
TG	0.0007 (0.0002-0.0017)	0.0003 (0-0.001)	0.0009 (0.0003-0.0024)	0.0002 (0-0.0008)	0.0007 (0.0001-0.0024)	0.0006 (0.0001-0.002)	0.0002 (0-0.001)	0.0009 (0.0003-0.002)	0.0006 (0.0002-0.0012)	0 (0-0.001)	0.0015 (0.0007-0.0029)	0.0007 (0.0006-0.0007)	0.0005 (0.0005-0.0006)
CG_TG	0.0745 (0.0305-0.1474)	0.1681 (0.1058-0.2476)	0.1429 (0.0403-0.3267)	0.6831 (0.6518-0.7132)	0.0014 (0.0004-0.0035)	0.3288 (0.2233-0.4487)	0.6877 (0.6335-0.7383)	0.0417 (0.0087-0.117)	0.0995 (0.0764-0.1267)	0.0727 (0.0202-0.1759)	0.114 (0.0759-0.1626)	0.153 (0.1468-0.1595)	0.3834 (0.3793-0.3874)

Table S5: Number of sequences inferred to be located in different contexts in previously published sequence data from cfDNA isolated from blood plasma (Troll et al. 2019), as well as in the control data down-sampled to an equal number of aligned sequences.

Specimen	Sample type	Library ID	Genomic coverage generated [-fold]	Structural context	Sequences identified in sample DNA dataset	Sequences identified in down-sampled control dataset	Signal/noise ratio
H-69	Cell-free DNA isolated from blood plasma (healthy donor)	SR-05	3.4	6-nt 3' overhang	752,682	682,360	1.10
				5-nt 3' overhang	748,044	699,186	1.07
				4-nt 3' overhang	854,355	715,714	1.19
				3-nt 3' overhang	823,298	733,351	1.12
				2-nt 3' overhang	761,151	824,024	1.08
				1-nt 3' overhang	769,410	1,031,929	1.34
				blunt	804,415	1,027,061	1.28
				1-nt 5' overhang	781,887	751,980	0.96
				2-nt 5' overhang	763,106	952,686	1.25