

Supporting Information Document

Optimizing Ancient DNA Recovery from Archaeological Plant Seeds

Giada Bolognesi ¹, Adriana Latorre ², Martina Marini ¹, Arianna Codato ^{2,3}, Francesco Fontani ², Fabio Saggioro⁴,
Donata Luiselli ², Patrizia Basso⁴, Elisabetta Cilli ^{2*}, Diana Bellin ^{1*}

Supplementary Material and Methods

Description of the archaeological sites sampled in this study

Nogara

Nogara (VR) medieval site was investigated through a multi-year archaeological excavation conducted from 2003 to 2009. The research campaigns focused on a medieval context, specifically examining the riverside area of a medieval village through extensive excavation. The origins of the medieval settlement can be traced back to the last decade of the 8th century and the first fifteen years of the 9th century. The inhabited area under study remained in use until the late 10th or early 11th century, undergoing transformations and changes to its buildings, but maintaining substantial continuity. Subsequently, the area was repurposed and occupied by production and artisanal spaces, possibly related to metallurgical activities. The stratigraphic context of Nogara is characterized by waterlogged deposits, likely linked to a fluvial harbour, preserving organic material as well as a significant portion of the wooden structures of the buildings. This represents one of the few examples of its kind for medieval Italy. The site consists of the riverside area and the village, which have been partially excavated, a castle (founded in the early 10th century and investigated through a series of test excavations), and a church that was built next to the castle in 10th/11th century. Among other plant remains many grape pips were found likely used for consumption. Pips come from stratigraphic units US3024 US3025 US3028 US3041 US3045(excavated in 2005) US7012 (excavated in 2007). Pips were not directly dated by radiocarbon (¹⁴C) but derive their chronological attribution from the closed and well-defined stratigraphic contexts in which they were recovered. These contexts have been dated through an integrated strategy (Tinazzi and Lerco, 2011). A dendrochronological approach relied over 100 wood samples taken from structural elements directly associated with the stratigraphic units. These have been analyzed and used to establish a dendrochronological curve. Radiocarbon dating was carried over on five ¹⁴C samples that were obtained from materials (e.g., charred wood, associated organic matter) directly linked to the buildings and their associated activity areas. Based on this multiproxy approach, the pips can be broadly assigned to between the second half of the 9th century and the 10th century.

Cologna Veneta

Cologna Veneta (VR) medieval site was investigated through an excavation conducted from 2016 to 2018. The excavation identified a defensive embankment, dated to the second half of the 13th century according to radiocarbon (¹⁴C) analysis of wood. A large wooden structure was uncovered consisting of a floorboard 6.5 meters wide and 16 meters long (although it extends beyond the excavation limits), characterized by several layers of wooden elements. On the surface of the floorboard, channels for water drainage were found, made with banks of interwoven wood, as well as wooden structures that likely served to hang/dry materials and objects related to artisanal work. A basket made of woven twigs was found. The structure indicates that elements associated with water were present, which must have been used in large and constant quantities. Grape seeds were found both in contact with the surface of the floor and in the silty layers. The large number of grape seeds and the usage of water indicate handcraft activities making use of grape seeds. Seeds were recovered from two distinct stratigraphic units (US 341 and US 383).

Extraction Methods

Phe-chl extraction method. The phenol-chlorophorm (Phe-chl) protocol was originally developed by Gilbert (2004) for the extraction of ancient hair samples. The crushed samples were placed in 1.5 mL tube with 750 μ L of digestion buffer consisted of: 10 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl) (pH 8.0), 10 mM NaCl, 2% w/v sodium dodecyl sulfate (SDS), 5 mM CaCl₂, 2.5 mM ethylenediaminetetraacetic acid (EDTA) (pH 8.0), 40 mM dithiothreitol (DTT), and 10% proteinase K. The samples were incubated overnight at 55°C. The following day, after the samples were equilibrated at room temperature, they were centrifuged at 13,000 rpm for 5 minutes. The supernatant was transferred in a new 1.5 mL tube and 1 volume of phenol was added, then the samples were gently rotated for 5 minutes at room temperature and centrifuged at 13,000 rpm for 5 minutes. The aqueous phase was collected and placed in a new tube, mixed with 1 volume of phenol, and the rotation and centrifuge steps were repeated. The aqueous layer was transferred into a new tube, and an equal volume of chloroform was added to it. The samples were vortexed for 5 minutes and centrifuged at 13,000 rpm for 5 minutes. The aqueous phase was collected and purified with MinElute PCR Purification kit (Qiagen, Valencia, CA).

CTAB extraction method. The CTAB method is a modified protocol from the Doyle and Doyle method (Doyle and Doyle, 1987). The crushed samples were placed in a 1.5 mL tube with 800 μ L of extraction buffer consisting of: 100 mM Tris (hydroxymethyl)aminomethane hydrochloride (Tris-HCl) (pH 8.9), 40 mM ethylenediaminetetraacetic acid (EDTA) (pH 8.0), 2M NaCl, 3.5% w/v Cetyltrimethylammoniumbromide (CTAB), 1.50% w/v polyvinylpyrrolidone (PVP) and 1.50% w/v B-mercaptoethanol. The samples were incubated at 60°C for 4 hours and the tubes were inverted every 30 minutes. After the incubation, 800 μ L chloroform was added to each sample and centrifuged at 10,000 rpm for 10 minutes. The supernatant was transferred in a new 1.5 mL tube; 1 volume of chloroform was added to each tube, and the centrifuge step was repeated. The supernatant was then placed in a new 1.5 mL tube and 0.6 volumes of cold isopropanol were added, and the tubes were gently inverted several times. The samples were incubated at -20°C overnight. The next day tubes were centrifuged for 30 minutes at 13,000 rpm (4°C). The supernatant was discarded and 400 μ L of cold 75% ethanol was added to precipitate nucleic acids, then the samples were centrifuged at 6,000 rpm for 10 minutes. The supernatant was carefully removed, and the pellet was air-dried to allow complete evaporation of residual ethanol. Finally, the pellet was resuspended in 50 μ L of ddH₂O for 20 minutes at 37°C.

DNeasy extraction method. The DNeasy Plant Mini Kit (Qiagen) was developed for extraction of fresh plant material. The ground samples were mixed with 500 μ L pre-heated buffer AP1 and incubated at 65°C for 10 minutes. After the incubation 163 μ L of buffer P3 was added and the samples were incubated in ice for 5 minutes. Then they were centrifuged at 13,000 rpm for 5 minutes and pipetted the lysate into a QIAshredder spin column placed in a 2 mL collection tube, that was then centrifuged for 2 minutes at 13,000 rpm. The eluate was transferred in a new tube and mixed with 1.5 volume of buffer AW1, then 650 μ L of this mixture was transferred into a DNeasy Mini spin column placed in a new 2 mL collection tube. The samples were then centrifuged for 1 minute at 8,000 rpm, the eluate was discarded, and the step was repeated with the remaining sample. The spin column was placed into a new 2 mL collecting tube, 500 μ L of AW2 were added, and then the tubes were centrifuged at 8,000 rpm. The eluate

was discarded and 500 µL of AW2 were added again. Then, the samples were centrifuged at 13,000 rpm for 2 minutes to eliminate ethanol residuals from the silicon filter. The spin column was placed in a new 1.5 mL tube, 50 µL of ddH₂O were added and incubated for 20 minutes at 37°C. The tubes were centrifuged at 8,000 rpm for 1 minute and this last step was repeated once.

S-PDE extraction method. The Silica-PowerBeads DNA Extraction (S-PDE) protocol was originally developed for the extraction of ancient DNA from soil sediments and coprolites (Rampelli et al., 2021, modified from Hagan et al., 2019). The protocol was slightly modified in this study, as detailed below. The cleaned and crushed samples were placed in a 1.5 mL tube with a digestion buffer composed of: 0.5 M ethylenediaminetetraacetic acid (EDTA) pH 8, 0.25 mg/mL of proteinase K and PowerBeads solution (Qiagen, Valencia, CA). The specimens were rotated and incubated overnight at 37°C. The following day, the samples were vortexed for 10 minutes and centrifuged at 15,000 rpm for 5 minutes. The supernatant was transferred in a 15 mL tube and 10 mL of buffer PB (Qiagen, Valencia, CA) were added, mixed and the whole mixture was transferred in 50 mL Roche tubes (High Pure Viral Nucleic Acid Large Volume kit, Roche, Basel, Switzerland). The samples were centrifuged at 15,000 rpm for 4 minutes, then the spin column was placed in a 2 ml collection tube and detached from the reservoir. The tubes were centrifuged for 1 minute at 6,000 rpm to dry the spin column filter, and then 720 µL of PE buffer (Qiagen, Valencia, CA) were added to the columns. The tubes were centrifuged at 6,000 rpm for 2 minutes, the eluate was discarded, and this step was repeated. Then empty columns were centrifuged at 14,000 rpm for 1 minute to dry the filter. The spin column was placed in a new 1.5 mL tube and 50 µL of pre-heated elution buffer (EB buffer, Qiagen) were added on the spin columns, and they were incubated for 5 minutes at 37°C. Then, the samples were centrifuged at 14,000 rpm for 1 minute, the eluate was recovered and re-placed back in the column. Finally, the incubation and centrifuge steps were repeated.

O. Tinazzi, E. Lerco, Indagine xilotomica e dendrocronologica su elementi lignei provenienti dallo scavo «Mulino di Sotto» di Nogara (VR): approccio metodologico e relazione preliminare, in Nogara. Archeologia e storia di un villaggio medievale (Scavi 2003–2008), edited by Fabio Saggioro, Rome, 2011, pp. 77–88 (ISBN: 9788876892615).

Supplementary Tables

Supplementary Table S1 DNA yields recovered from each sample. Samples are grouped according to the extraction method and archaeological site. DNA extract concentrations and estimated total nanograms of DNA extracted from each seed are reported.

Method	Archaeological site	Sample name	Concentration (ng/ μ L)	Total yield (ng)
CTAB	Cologna Veneta	CL108_A	0,46	23
		CL127_A	0,47	23,5
		CL128_A	0,812	40,6
		CL129_A	0,264	13,2
		CL171_A	2,14	107
		CL192_A	4,16	208
		CL208_A	4,48	224
		CL209_A	4,14	207
		CL217_A	4,4	220
		CL262_A	1,7	85
	CL274_A	3,64	182	
	Nogara	NG2_A	7,76	388
		NG6(24)_A	4,46	223
		NG6_A	3,92	196
		NG10_A	0,884	44,2
		NG32_A	0,576	28,8
		NG35_A	0,398	19,9
		NG38_A	0,592	29,6
		NG47_A	0,256	12,8
		NG224_A	0,318	15,9
		NG225_A	0,168	8,4
		NG231_A	0,45	22,5
		NG232_A	0,58	29
NG233_A		0,172	8,6	
NG234_A	0,406	20,3		
DNeasy	Cologna Veneta	CL83_B	0	0
		CL84_B	0	0
		CL85_B	0,148	7,4
		CL94_B	0,376	18,8
		CL99_B	0,38	19
		CL103_B	0,162	8,1
		CL104_B	0,592	29,6
		CL105_B	0,81	40,5
		CL107_B	0,476	23,8
		CL114_B	0,37	18,5
	CL119_B	0,384	19,2	
	Nogara	NG76_B	1,07	53,5
		NG77_B	0,49	24,5

		NG79_B	1,63	81,5
		NG80_B	0,73	36,5
		NG124_B	0,392	19,6
		NG133_B	0,44	22
		NG134_B	0,328	16,4
		NG135_B	0,642	32,1
		NG221_B	0,398	19,9
		NG223_B	0,386	19,3
S-PDE	Cologna Veneta	CL1_C	21,4	1070
		CL2_C	9,28	464
		CL3_C	1,34	67
		CL4_C	2,7	135
		CL5_C	3,52	176
		CL6_C	0,788	39,4
		CL7_C	4,76	238
		CL8_C	7,96	398
		CL9_C	3,96	198
		CL163_C	5,18	259
		CL164_C	4,94	247
		CL165_C	5,18	259
		CL166_C	2,5	125
	Nogara	NG1_C	9,58	479
		NG2_C	3,3	165
		NG4_C	3,64	182
		NG5_C	9,24	462
		NG6_C	5,94	297
		NG7_C	3,22	161
		NG8_C	1,65	82,5
		NG10_C	2,36	118
		NG11_C	1,85	92,5
		NG12_C	0,478	23,9
		NG13_C	3,9	195
		NG14_C	1,22	61
		NG29_C	2,66	133
NG30_C	1,95	97,5		
NG31_C	3,9	195		
NG32_C	11,6	580		
NG33_C	1,65	82,5		

Supplementary Table S2 Statistics summary table. Statistics tests are based on relative values originally presented in Supplemental Table 1 (DNA yields), Supplemental Table 2 (qPCR results) and Table 3 (NGS libraries sequencing metrics) and used for the different statistical test and comparisons as indicated. Asterisks indicate p-values significance (* p<0.05, ** p<0.01, *** p<0.001). Confidence intervals and size effects (according to Eta squared for Two-way ANOVA and Coehen’s D for Tukey and T-tests) are also reported. For Eta-squared size effects are as following: 0.01 small; 0.06 medium; 0.14 large. For Coehen’s D size effects are as following: 0.2 small; 0.5 medium; 0.8 large.

Test comparison	Statistic test	Parameters	p-value	95% CI	Size effect
DNA yields (ng)	Two-way ANOVA	Sites	0.207		0.02
		Methods	0.00000768***		0.28
		Sites:Methods	0.507		0.01
	Tukey Test among methods	DNeasy – CTAB	0.239	-174.88 / 33.04	-0.91
		S-PDE - CTAB	0.002**	45.76 /235.97	0.82
		S-PDE - DNeasy	0.0000091***	111.85 / 311.71	1.29
	T-test between archaeological sites (for each method)	CTAB	0.275	-39.52 / 132.37	-0.45
		DNeasy	0.04447*	-31.01 / -0.43	0.94
		S-PDE	0.3005	-77.53 / 242.16	-0.39
CT values	Two-way ANOVA	Sites	0.337		0.02
		Methods	0.120		0.07
		Sites:Methods	0.0051**		0.25
	Tukey Test among Sites:Methods	Nogara:CTAB - Cologna Veneta:CTAB	0.034*	-20.48 / -0.52	-7.00
		Cologna Veneta:DNeasy – Cologna Veneta:CTAB	0.0409*	-20.22 /-0.27	-5.15
		Cologna Veneta:S-PDE – Cologna Veneta:CTAB	0.0017**	-19.51 /-3.38	-2.41
Percentage of retained reads	Two-way ANOVA	Sites	0.891		6.08e-04
		Methods	0.0102*		0.37
		Sites:Methods	0.3471		0.03
	Tukey Test among methods	DNeasy – CTAB	0.78	-13.41 /7.74	-0.49
		S-PDE - CTAB	0.12	-1.41 / 14.74	1.30
		S-PDE - DNeasy	0.019*	1.41 /17.58	2.06
Deduplicated mapped reads	Two-way ANOVA	Sites	0.121		0.12
		Methods	0.651		0.04

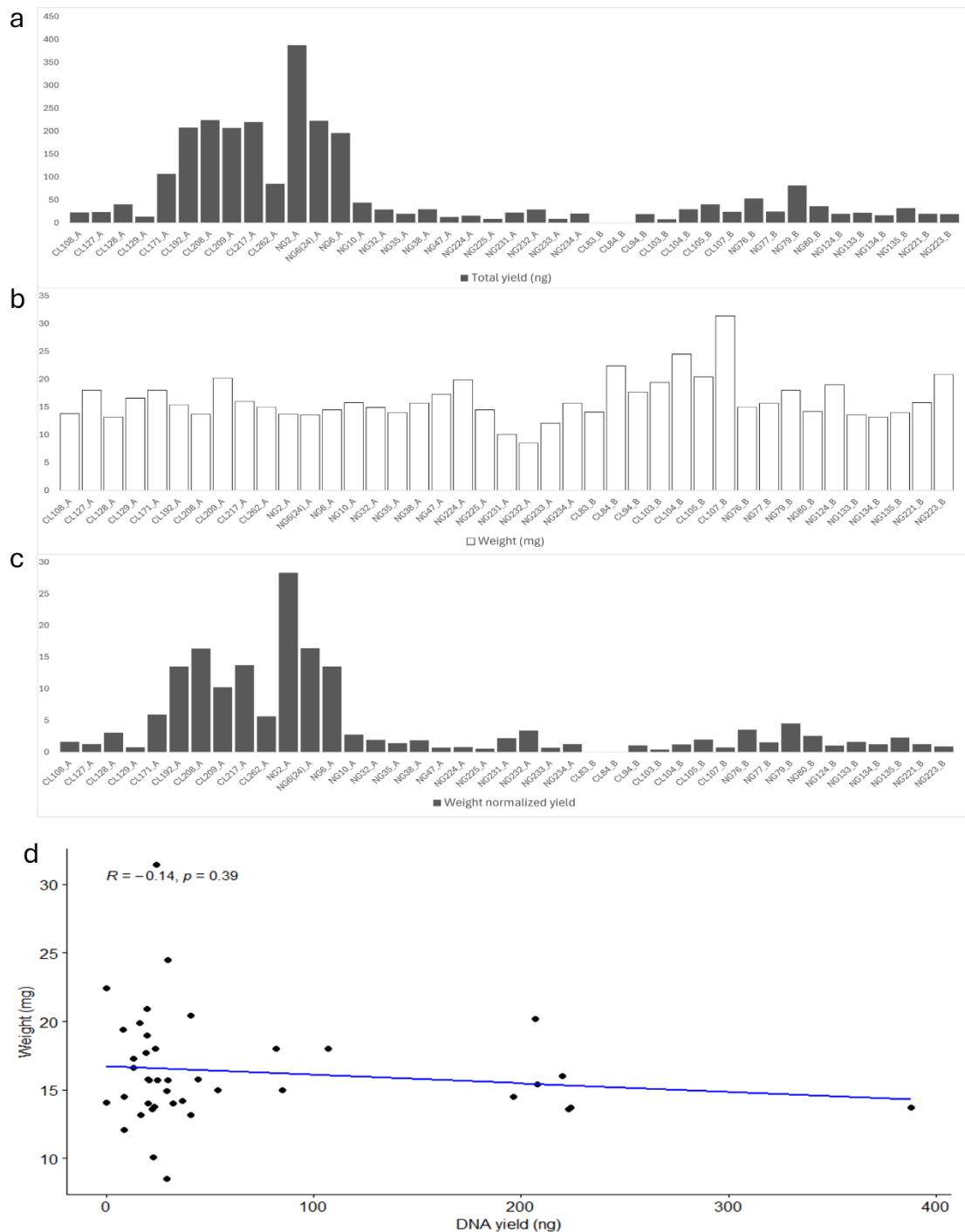
		Sites:Methods	0.524		0.02
Endogenous DNA content (%)	Two-way ANOVA	Sites	0.162		0.10
		Methods	0.9		9.72e-03
		Sites:Methods	0.468		0.03
Libraires clonality (%)	Two-way ANOVA	Sites	0.997		5.43 e-07
		Methods	0.218		0.15
		Sites:Methods	0.658		9.01 e -03
Libraries complexity (%)	Two-way ANOVA	Sites	0.373		0.04
		Methods	0.422		0.08
		Sites:Methods	0.474		0.02
GC content (%)	Two-way ANOVA	Sites	0.928		3.99e-04
		Methods	0.381		0.10
		Sites:Methods	0.868		1.35e-03
Mean length reads (bp)	Two-way ANOVA	Sites	0.00146**		0.20
		Methods	4.56e-05***		0.51
		Sites:Methods	0.469		7.93e-03
	Tukey test among methods	DNeasy – CTAB	4.02 e-05***	13.1 / 33.57	7.62
		S-PDE - CTAB	0.0325*	0.65 / 16.29	1.30
		S-PDE - DNeasy	0.0003275***	-22.69 / -7.05	-2.31
	Tukey test between sites	Nogara – Cologne Veneta	0.00281**	-11.13 / -2.69	-1.49
C > T in 1st. position	Two-way ANOVA	Sites	0.01843*		0.14
		Methods	1.20e-03**		0.40
		Sites:Methods	0.106		0.06
	Tukey Test among methods	DNeasy – CTAB	0.552	-2.58 / 6.25	1.99
		S-PDE - CTAB	0.00257**	1.82 / 8.57	2.01
		S-PDE - DNeasy	0.051	-0.012 / 6.73	1.29
	Tukey test between sites	Nogara – Cologne Veneta	0.0279*	0.25 / 3.89	0.34

Supplementary Table S3 Data on DNA library production and quality of produced libraries. Libraries with a qPCR CT-values higher than blank were considered unsuccessful and discarded. Following indexing and purification successful libraries were quantified (molarity given as nanomoles) and evaluated for dimer contamination (percentage value of dimers compared to whole library is given).

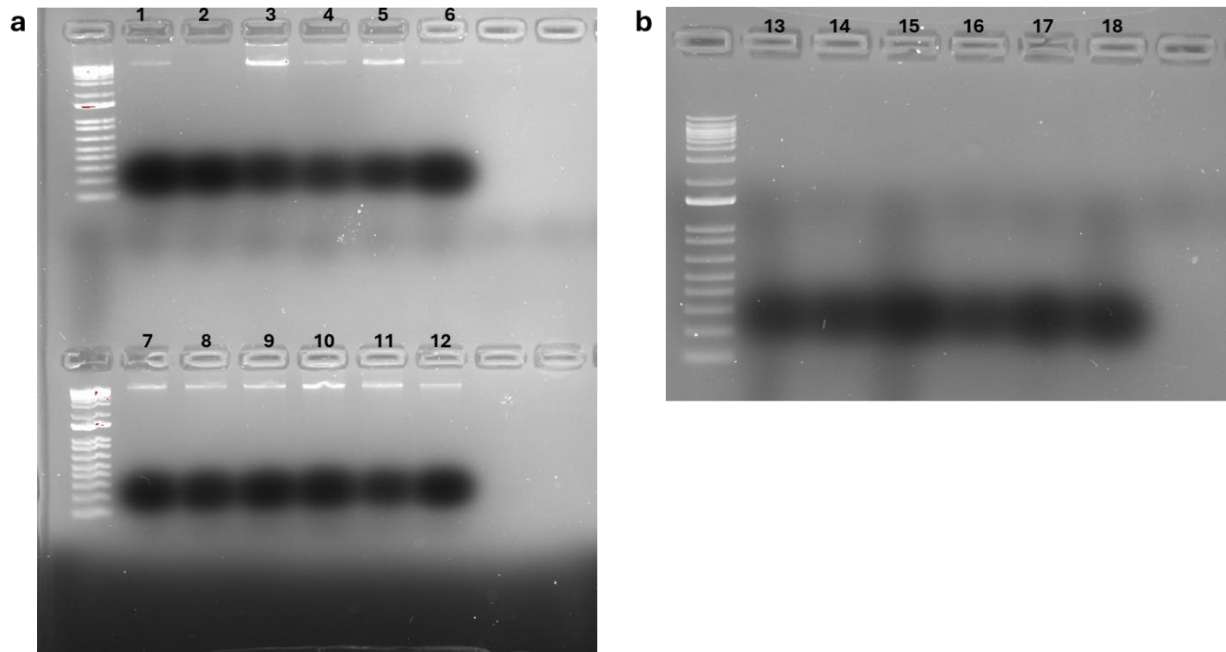
Method	Archaeological site	Sample name	CT	Molarity (nM)	Dimers (% of total)
CTAB	Cologna	CL192_A	27	-	-
		CL208_A	27	-	-
		CL209_A	24	-	-
		CL217_A	29	-	-
	Nogara	NG32_A	16	36.9	5.49
		NG35_A	16	12.8	2
		NG38_A	16	31	11.2
		NG47_A	17	36.3	18.5
DNeasy	Cologna Veneta	CL103_B	17	59.4	14.3
		CL104_B	15	40	3.8
		CL105_B	19	-	-
		CL107_B	15	51.9	7.02
	Nogara	NG133_B	26	-	-
		NG135_B	23	-	-
		NG221_B	16	17.6	16.32
		NG223_B	23	29.1	11
S-PDE	Cologna veneta	CL1_C	10	7.19	7
		CL2_C	8	13.7	8
		CL3_C	10	32.1	3
		CL4_C	11	13	12
		CL5_C	28	-	-
		CL6_C	16	20.6	10.7
		CL7_C	16	35	2.92
		CL8_C	16	22.5	1.36
		CL9_C	16	33.4	0.85
		CL163_C	21	-	-
		CL164_C	16	14.1	9.97
		CL165_C	15	45.9	1.21
	CL166_C	16	12.1	6.95	
	Nogara	NG1_C	12	5.9	23
		NG2_C	21	-	-
		NG4_C	19	-	-
		NG5_C	22	-	-
		NG6_C	14	15	5.31
		NG7_C	22	-	-
		NG8_C	19	-	-
		NG10_C	16	2.52	4.72
		NG11_C	16	20.2	7
NG12_C		19	3.4	15	
NG13_C	14	9.2	8		
NG14_C	11	33	6		

		NG29_C	22	16.2	10
		NG30_C	34	-	-
		NG31_C	22	5.82	2
		NG32_C	22	18.9	11
		NG33_C	22	54.1	6

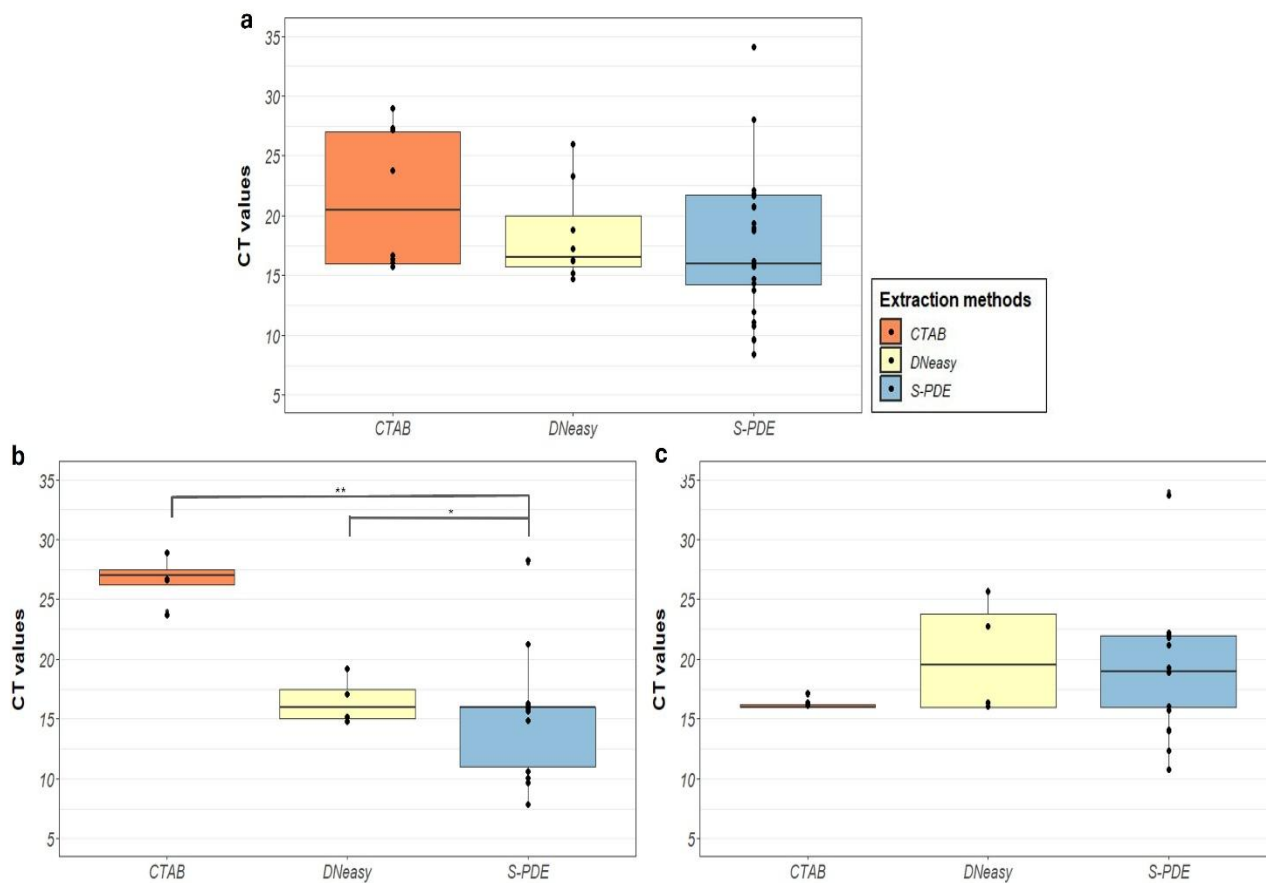
Supplementary Figures



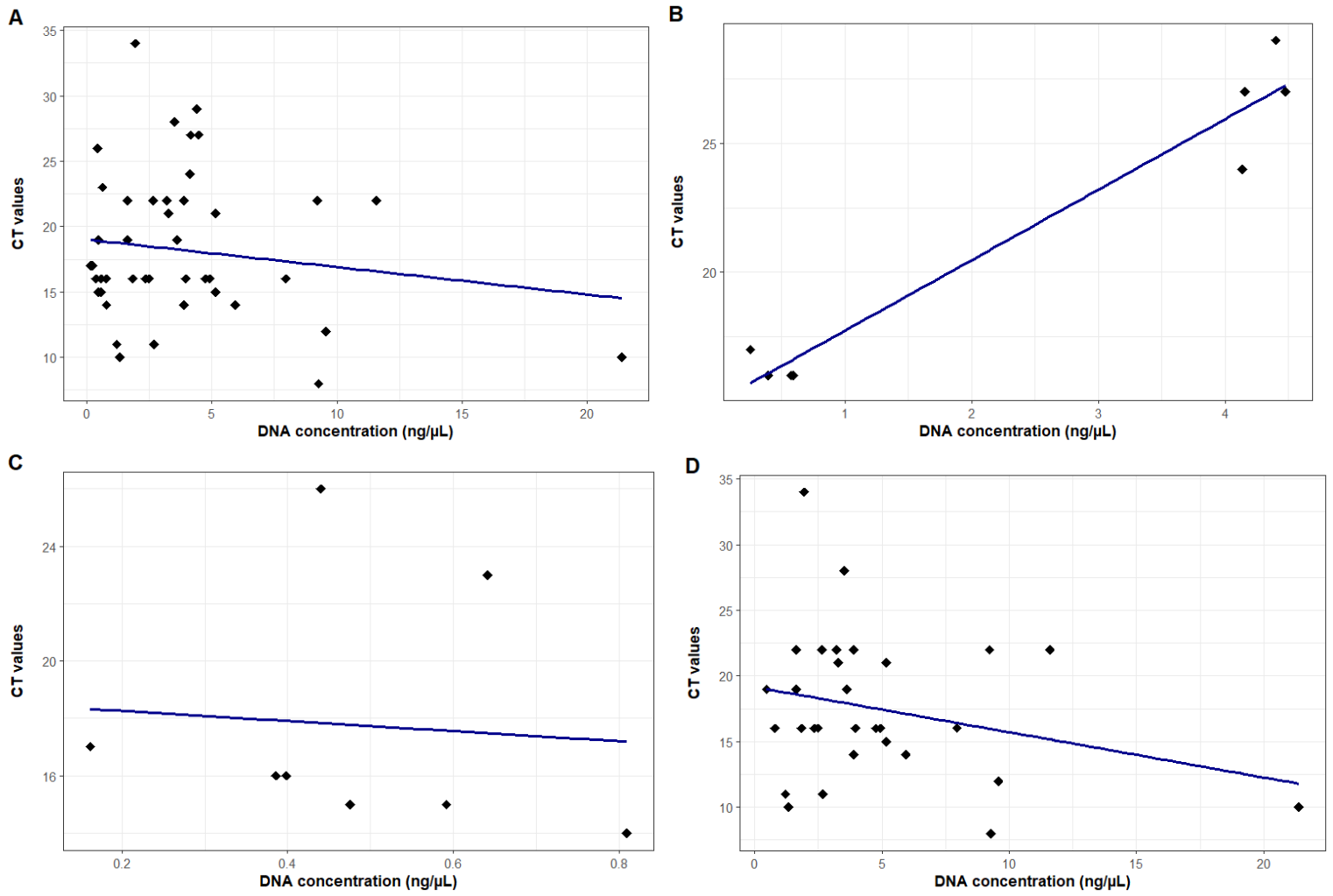
Supplementary Fig. S1 DNA yields, seed weights and normalized DNA yields comparisons and correlations between DNA yields and weights. (a) Histograms plot of raw DNA yields for different ancient seeds from different sites. Samples labelled _A were extracted with CTAB method and samples labelled _B were extracted with DNeasy method. (b) Histograms plot of seeds weight. (c) Histograms plot of DNA yields normalized to seed weight. The plot shows normalized yields are comparable to original raw yields and seed weights are relatively uniform. Yields across samples seem independent of mass differences. Other aspects than mass are likely affecting aDNA recovery. (d) Correlation between DNA yield and seed weights. The plot confirms previous observations showing no correlation between DNA yield and seed weight.



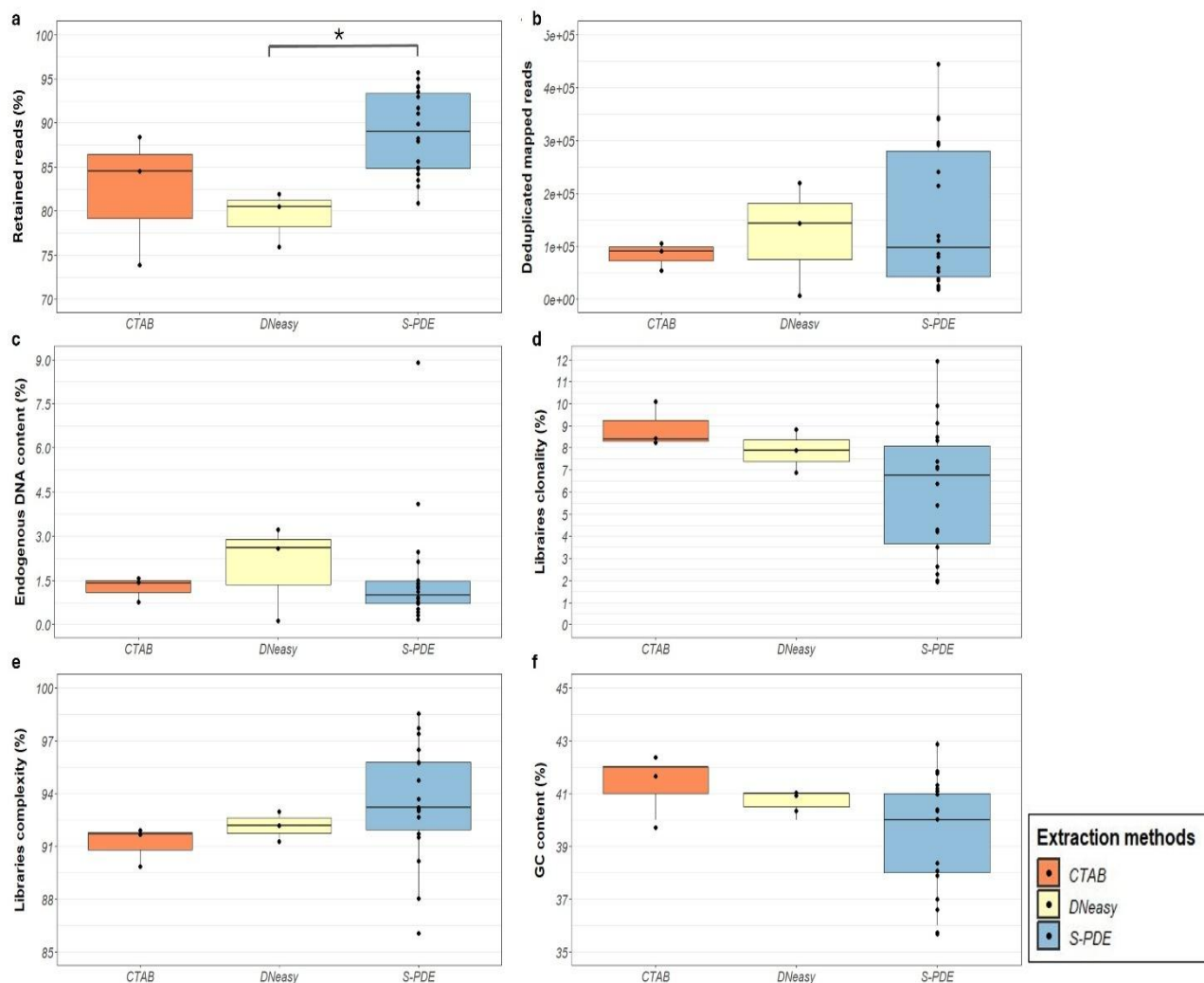
Supplementary Fig. S2 Gel electrophoresis of DNA extracts obtained using the Phe-Chl method. (a) Samples 1 to 12 are DNA extracts obtained by Phe-Chl protocol from independent modern seeds collected from clusters at ripening time used to preliminarily test the Phe-Chl protocol. DNA yields were also quantified by fluorometric analysis with following results 1=14.8 ng/ul, 2=14.8 ng/ul, 3=85.2 ng/ul, 4=16.6 ng/ul, 5=31,7 ng/ul, 6=17.9 ng/ul, 7=35,0 ng/ul, 8=30,1 ng/ul, 9=36,9 ng/ul, 10=49,3 ng/ul, 11=23,0 ng/ul, 12=47,4 ng/ul. (b) Samples 13 to 18 are DNA extracts obtained by Phe-Chl protocol from 6 independent ancient seeds collected at Nogara or Cologna Veneta sites. DNA yields of these extracts and further ancient extracts were also quantified by fluorometric analysis that provided not analysable results.

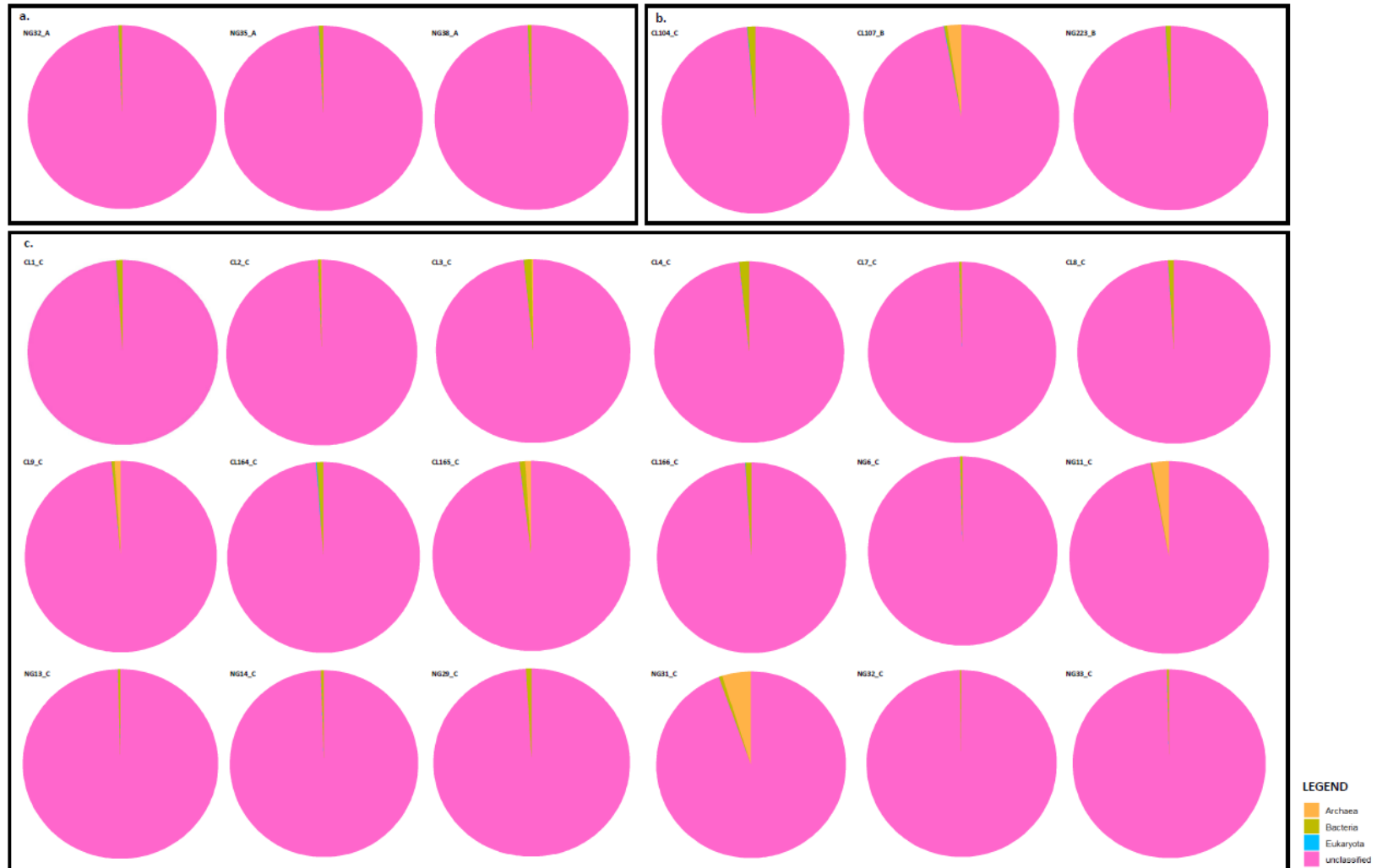


Supplementary Fig. S3. Comparison of the qPCR CT-values for the libraries produced from DNA extracted using the different extraction methods. (a) Boxplot for CT-values for all samples extracted either using the CTAB (n=8), the DNeasy kit method (n=8) and the S-PDE extraction method (n=30). (b) Boxplot for CT-values for samples from the Cologna Veneta archaeological site extracted using the CTAB, the DNeasy kit method and the S-PDE extraction method. (c) Boxplot for CT-values for samples from the Nogara archaeological site extracted using the CTAB, the DNeasy kit method and the S-PDE extraction method. Significance indicators test have been included in plots (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

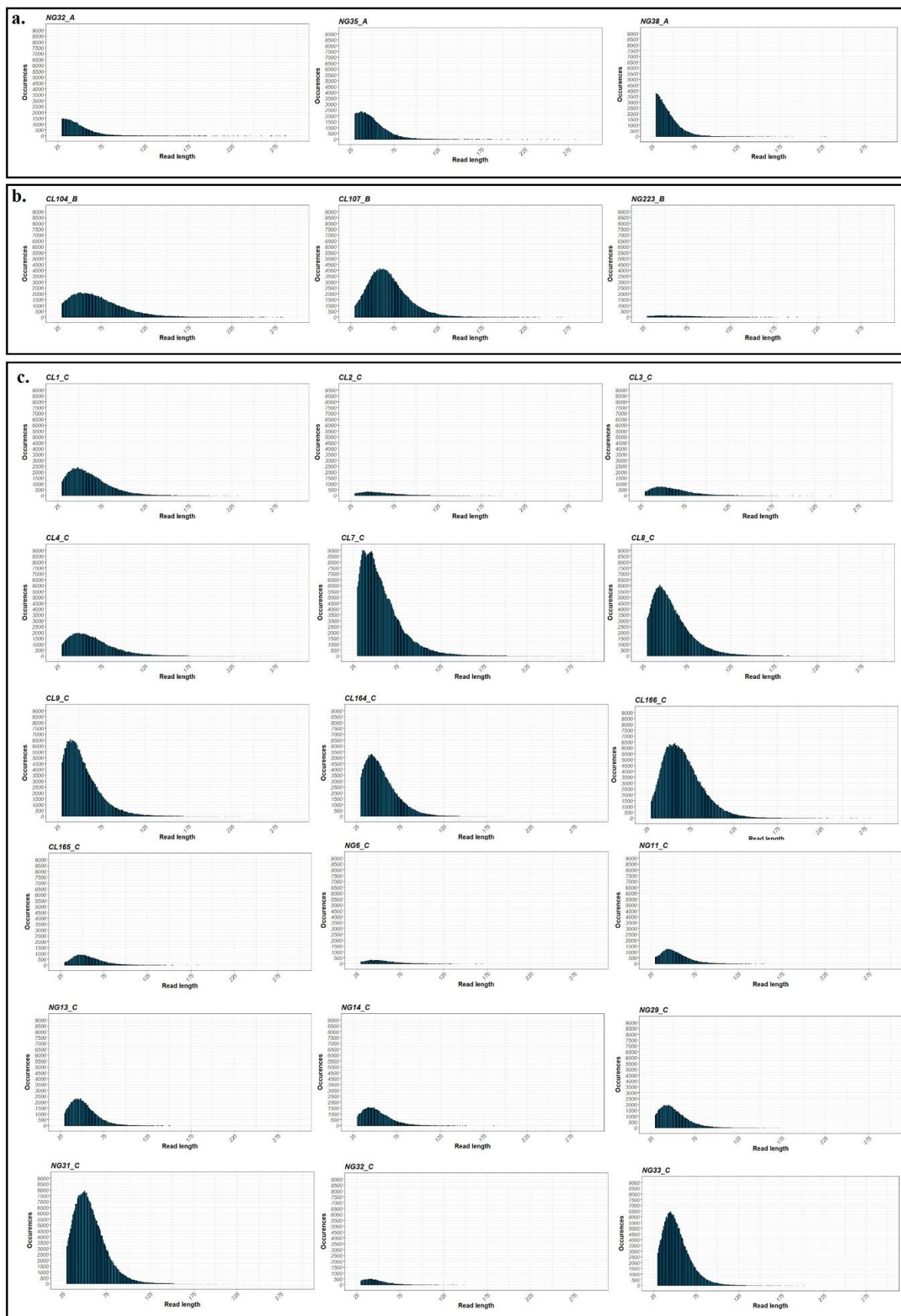


Supplementary Fig. S4. Correlations of qPCR CT values with DNA concentration for DNA extracts for (A) all samples, (B) for samples extracted by using the CTAB extraction method, (C) for samples extracted by using the DNeasy extraction method, (D) samples extracted by using the S-PDE protocol.

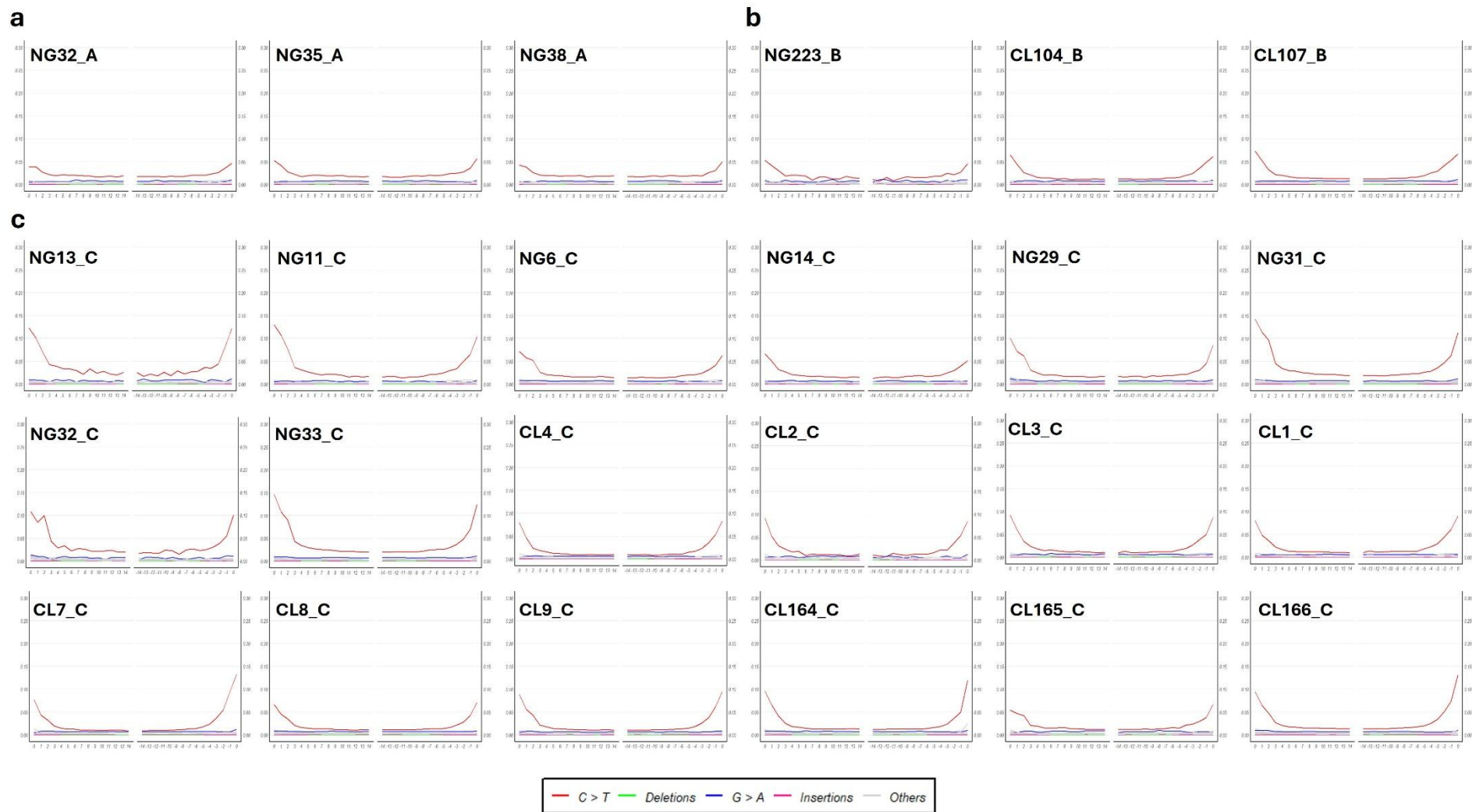




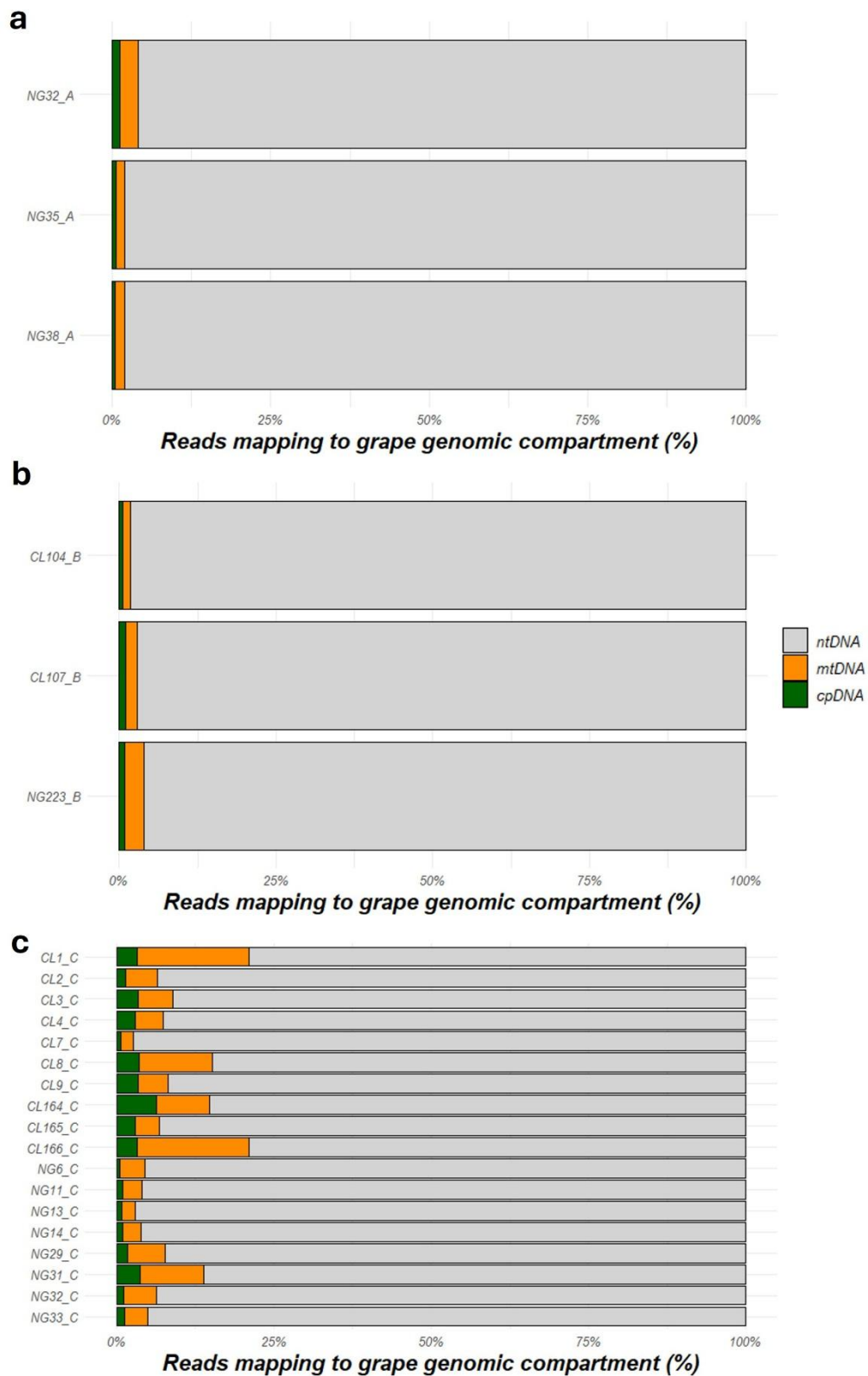
Supplementary Fig. S6. Taxonomic assignment of unmapped reads for the libraries produced from DNA extracted using the different extraction methods. Reads mapping was based on Kraken2 software. Outputs of taxonomic assignments as Archea, bacteria or Eukaryota as well as unclassified reads are given for samples extracted either using the CTAB (a), the DNeasy kit method (b) and the S-PDE extraction method (c).



Supplementary Fig. S7. Read lengths distribution for samples extracted using the different extraction methods. Distribution of fragment length for all samples grouped according to the extraction either using the CTAB (a), the DNeasy kit method (b) or the S-PDE extraction method (c).



Supplementary Fig. S8 Misincorporation patterns on mapped reads for archaeological samples extracted using (a) CTAB, (b) DNeasy, and (c) S-PDE. For each sample on left are represented nucleotide substitutions at 5' end of reads and on the right nucleotide substitutions at 3' end of reads.



Supplementary Fig. S9 Reads mapping to the different grape genomic compartments for archaeological samples extracted using (a) CTAB, (b) DNeasy, and (c) S-PDE. For each sample percentages of reads mapping to either the mitochondrial, chloroplast or nuclear genomes are shown.