Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past

Article *in* Molecular Ecology Resources · April 2017

DOI: 10.1111/1755-0998.12679

CITATIONS
14

READS
582

11 authors, including:

Clio Der Sarkissian
University of Copenhagen
42 PUBLICATIONS **1,604** CITATIONS
SEE PROFILE

Vianney Pichereau
Université de Bretagne Occidentale
107 PUBLICATIONS **1,677** CITATIONS
SEE PROFILE

Catherine Dupont
French National Centre for Scientific Research
135 PUBLICATIONS **398** CITATIONS
SEE PROFILE

Peter ilsøe
University of Copenhagen
7 PUBLICATIONS **18** CITATIONS
SEE PROFILE

SUPPLEMENTARY FIGURES

Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past

Clio Der Sarkissian, Vianney Pichereau, Catherine Dupont, Peter C. Ilsøe, Mickaël Perrigault, Paul Butler, Laurent Chauvaud, Jón Eiríksson, James Scourse, Christine Paillard, Ludovic Orlando

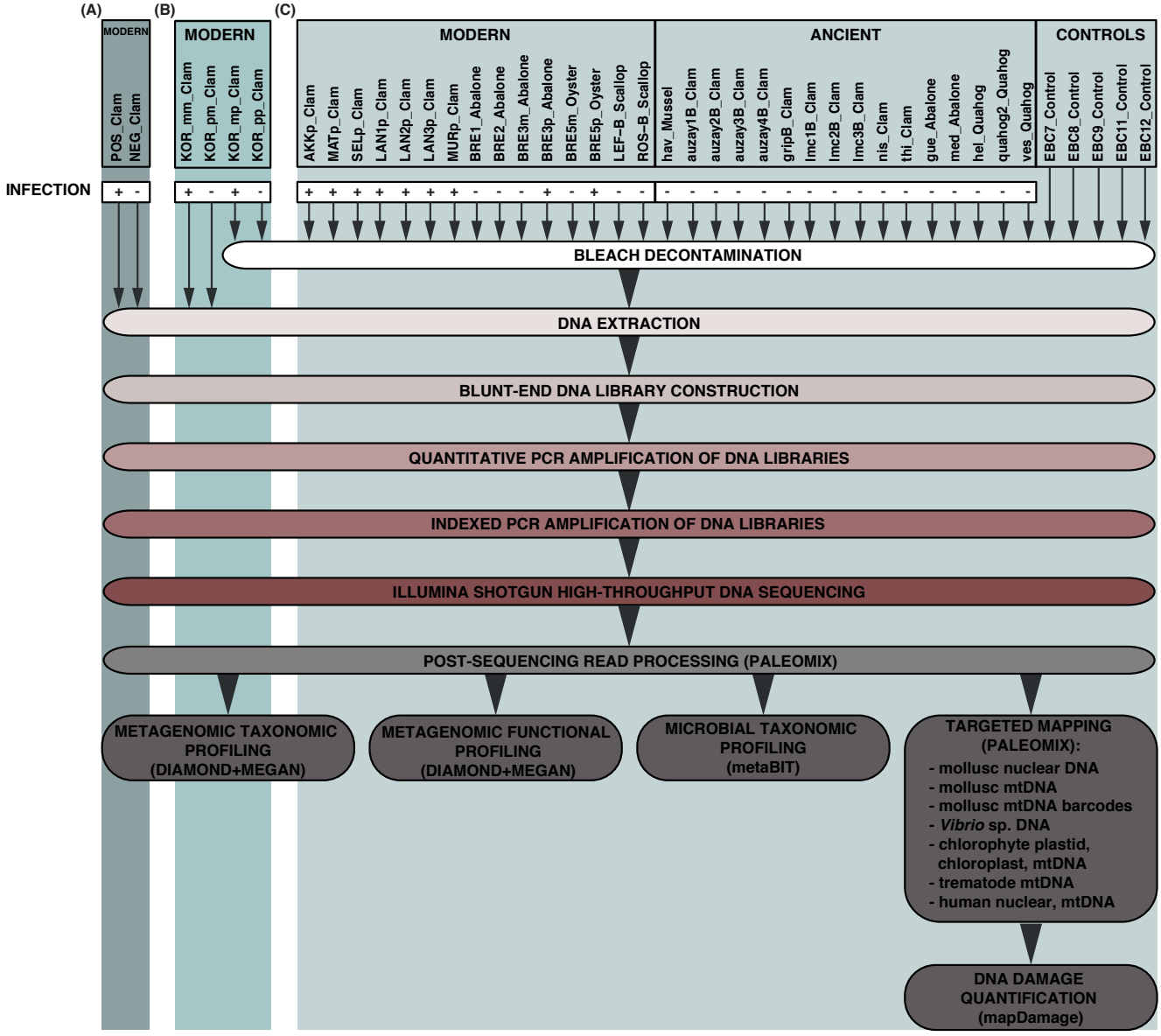

Figure S1. Experimental flowchart.
(A) Preliminary test on modern Manila clams. (B) Bleach decontamination test on the KOR Manila Clam. (C) Analyses on the rest of the shell samples and associated extraction blank controls (EBC). “mm”, non-infected non-bleached fragment; “pm”, infected non-bleached fragment; “mp”, non-infected bleached fragment; “pp”, infected bleached fragment.

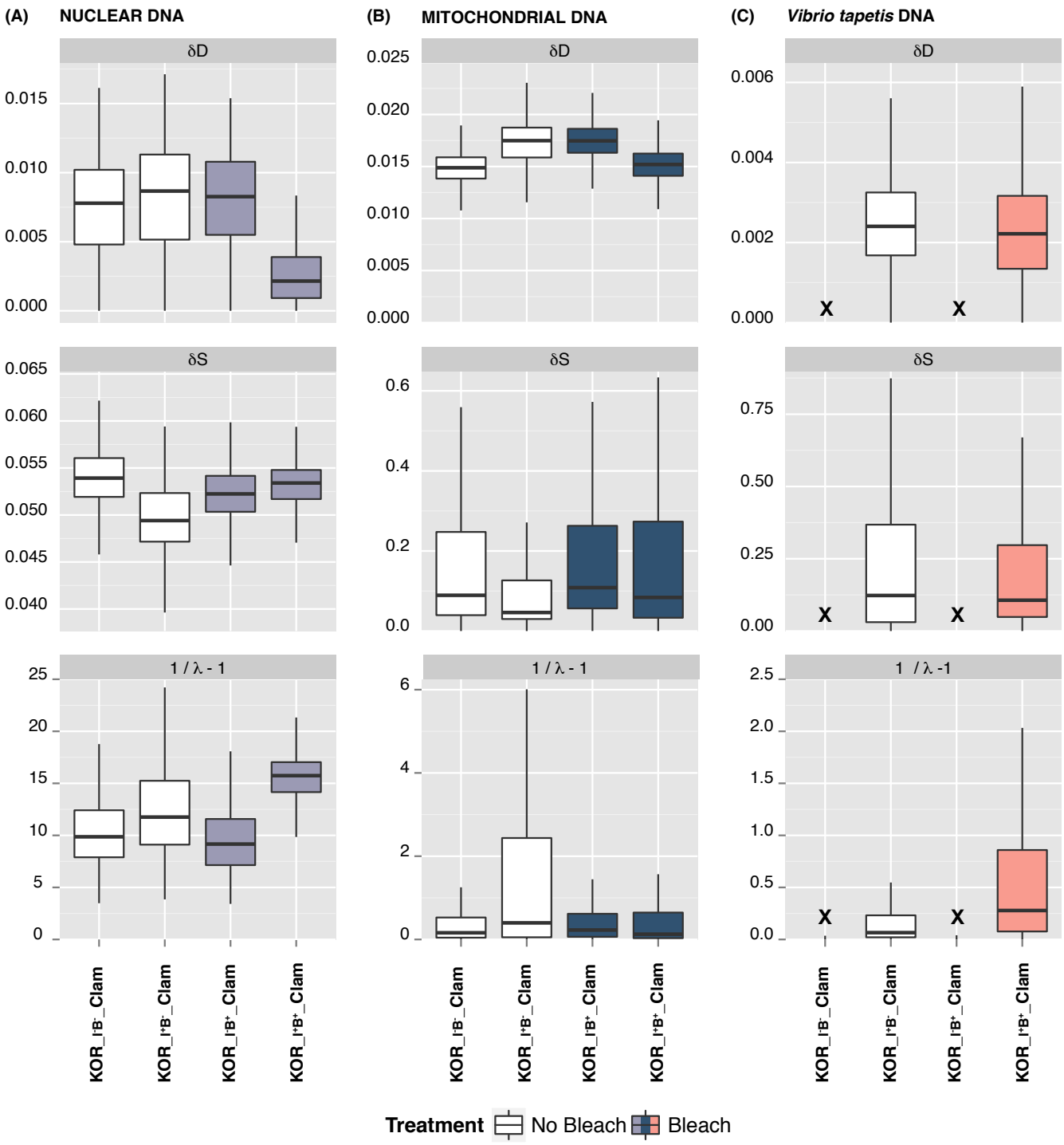

Figure S2. Impact of pre-extraction bleach decontamination on the posterior distributions of three damage parameters (δ_D, δ_S and $1/\lambda - 1$).

Damage parameters were estimated from 100,000 iterations for reads mapping to the (A) *Ruditapes philippinarum* nuclear transcripts (111,597 random reads), (B) the *R. philippinarum* mtDNA genome (2,850 random reads), and c) the *Vibrio tapetis* strain RP2-3 (512 random reads). Only unique high-quality (≥30) reads were considered in damage analyses. "δ_D", rate of cytosine deamination at double stranded regions; "δ_S", at single stranded overhangs; "$1/\lambda - 1$", proxy for overhang length. "I^-B^-", non-infected non-bleached fragment; "I^+B^-", infected non-bleached fragment; "I^-B^+", non-infected bleached fragment; "I^+B^+", infected bleached fragment.

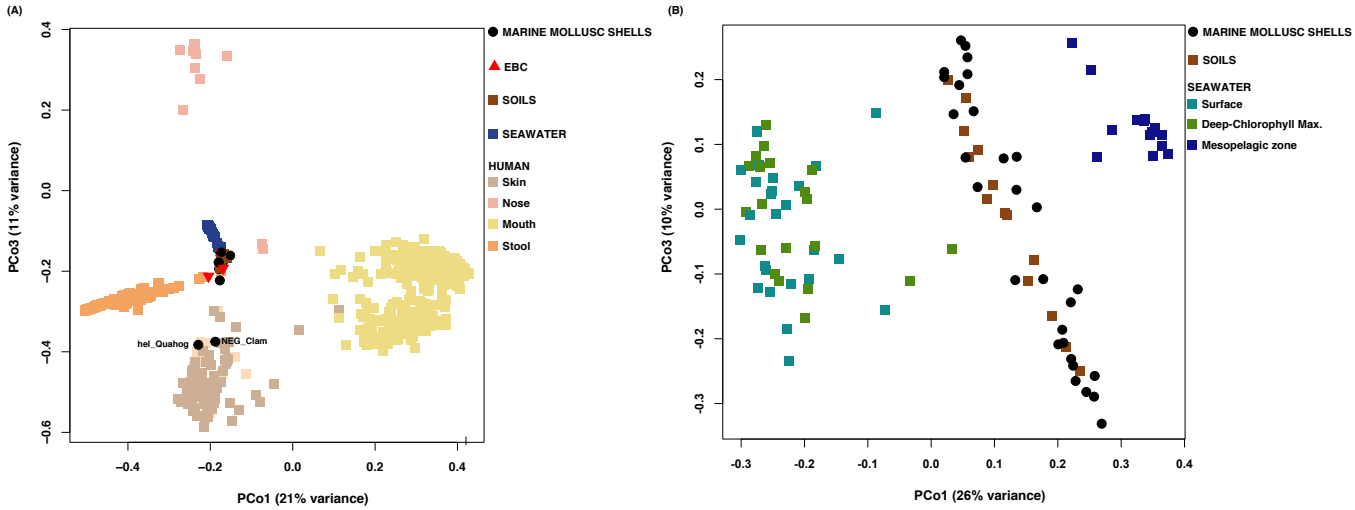

Figure S3. Microbial diversity in shells, human-associated, seawater, and soil samples.
PCoA of Bray-Curtis distances between genus-level microbial profiles. (A) Considering all taxa. (B) Considering only taxa absent from EBCs and after relative abundance re-adjustment.

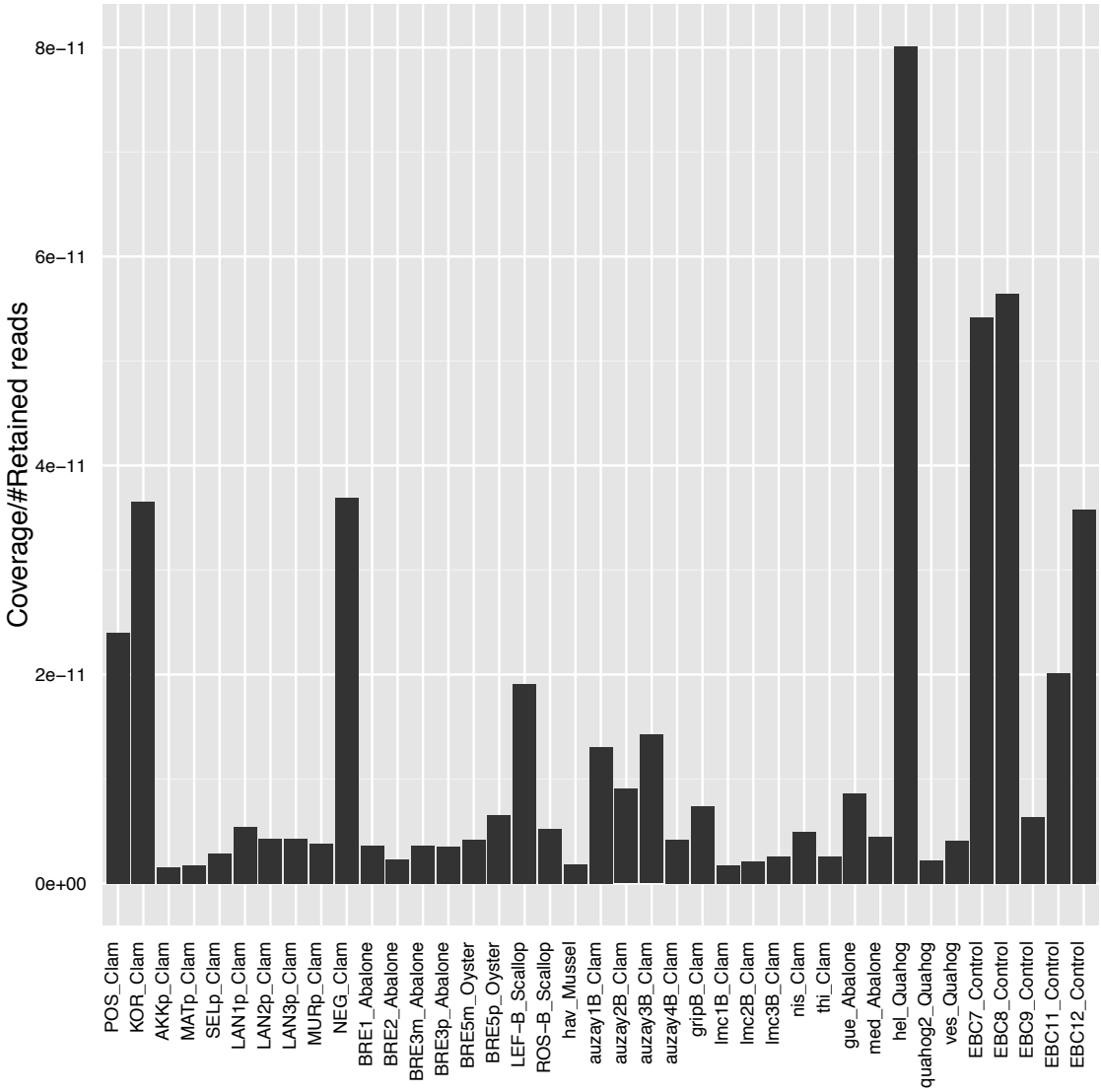

Figure S4. Human DNA content in shells.
Number of unique high-quality (MQ≥30) reads mapped to the *Homo sapiens* genome reference sequence (assembly hg19 build 37.1) normalized, for each sample, by the total number of reads retained after adapter trimming/quality filtering.

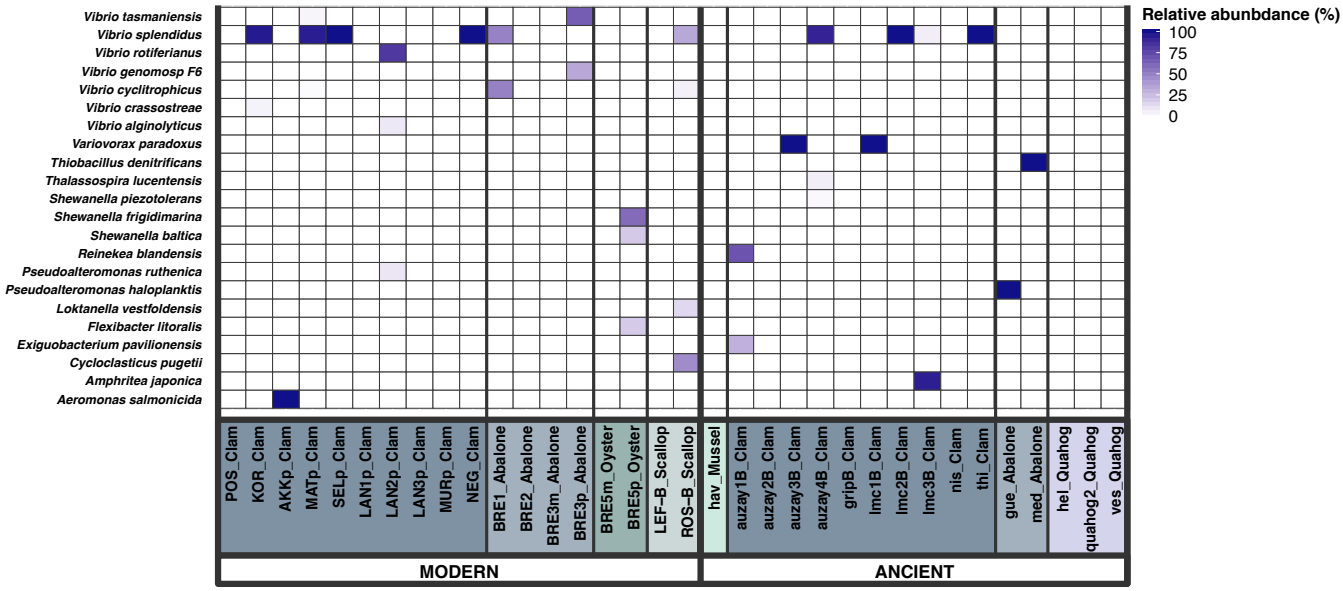

Figure S5. Marine seawater/sediment bacterial species in shells.
Only species that were not detected in extraction blank controls are shown. The colour gradient represents relative abundances as calculated by MetaPhlAn, as implemented in metaBIT.

Figure S6. Mapping of shell DNA against reference genome sequences representative of the known *Vibrio* species variability.

The colour gradient represents the coverage obtained when mapping against each *Vibrio* species reference genome sequence, considering unique high-quality (MQ≥30) reads, and normalized, for each sample, by the total number of reads retained after adapter trimming/quality filtering. The colour gradient was scaled by sample.

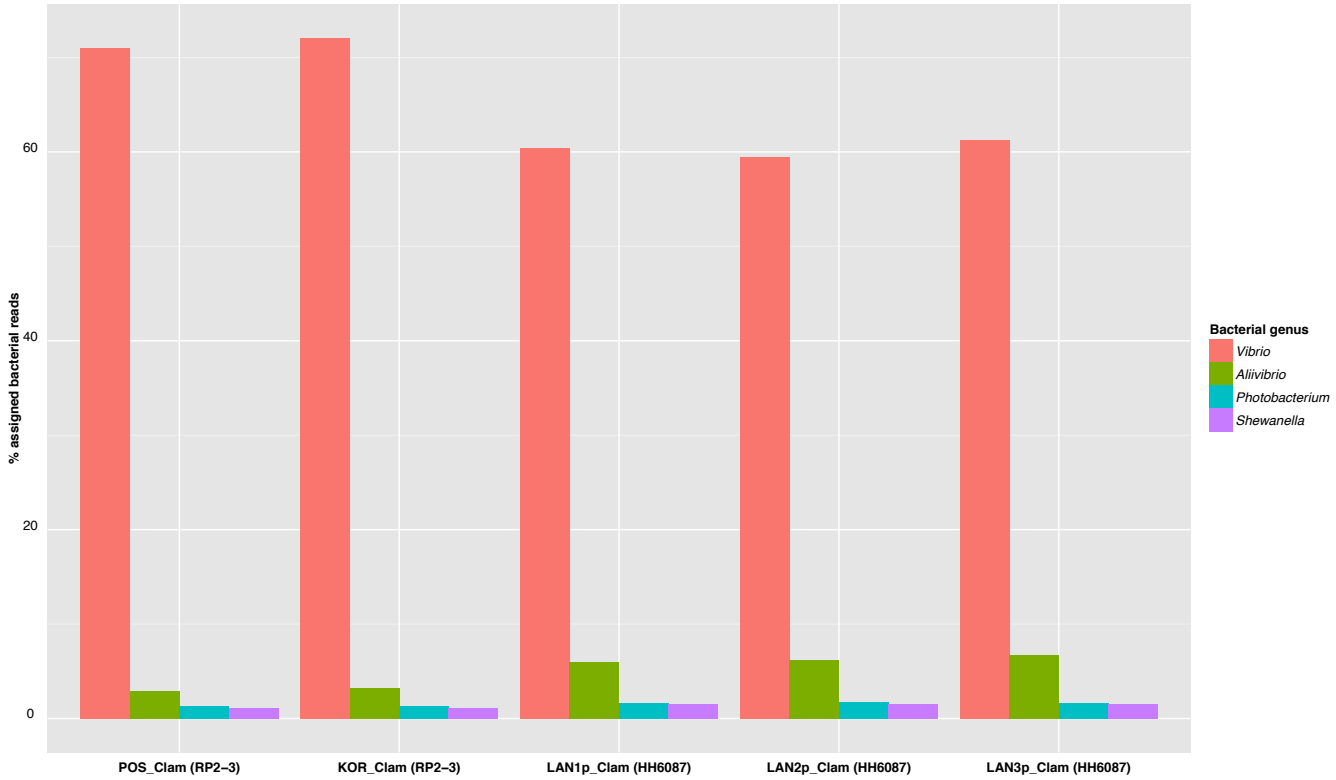

Figure S7. Most abundant bacterial genera assigned from reads mapping to the *Vibrio tapetis* genome reference sequences.

For samples showing maximal coverage for *Vibrio tapetis* genomes, reads mapping to the reference sequences of *V. tapetis* strains RP2-3 or HH6087 (indicated in parentheses) were aligned to the *nt* database using BLAST. The barplot represents the percentage of bacterial reads that were assigned at the genus level with frequencies ≥1%.

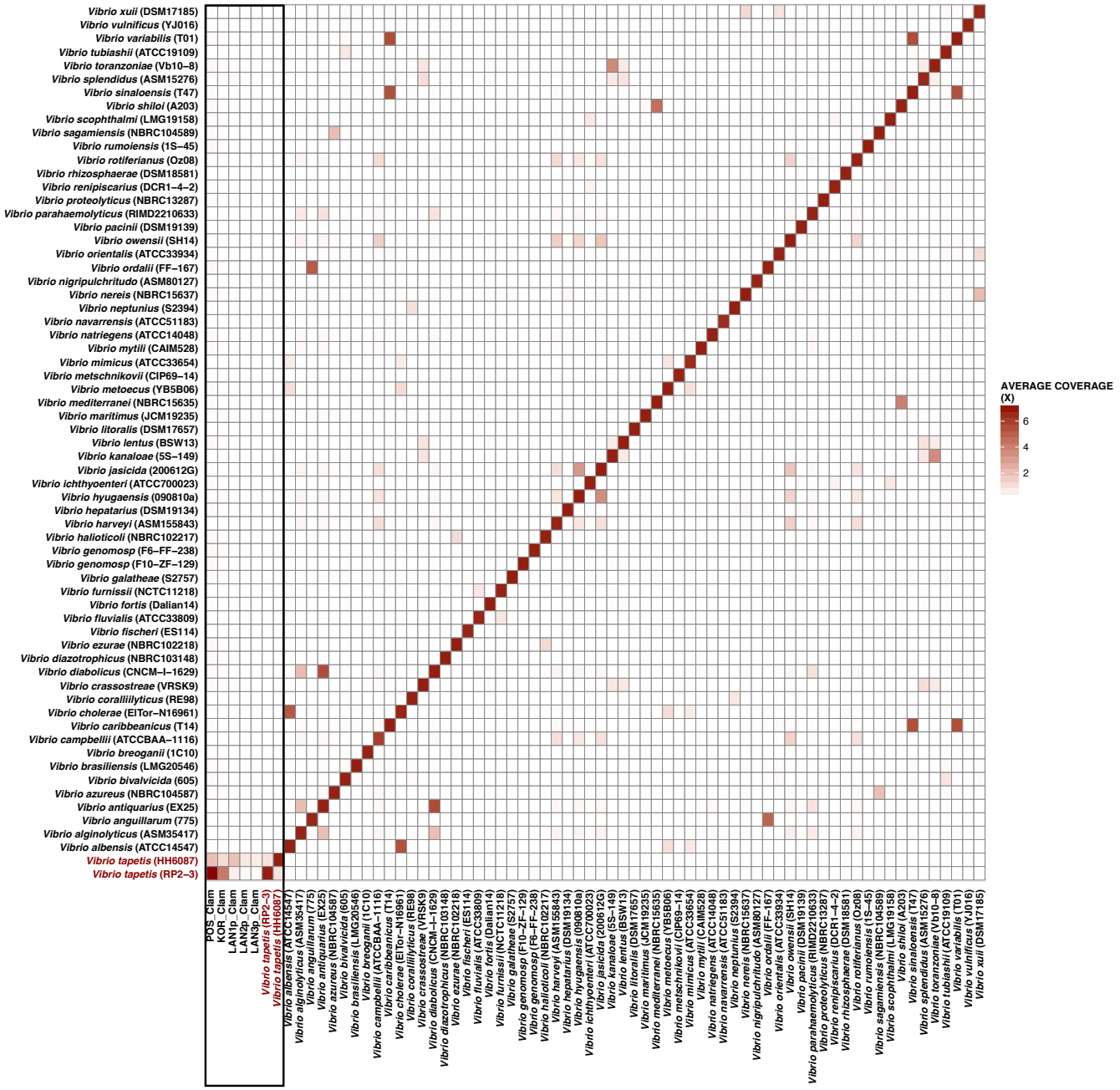

Figure S8. Mapping of shell and *Vibrio* DNA reads against reference genome sequences representative of the known *Vibrio* species variability.

Vibrio DNA reads were simulated from each *Vibrio* reference genome sequence to represent a depth-of-coverage of 7X, which is similar to the highest coverage obtained for POS_Clam when mapping shell DNA reads against *Vibrio tapetis* reference sequences. The colour gradient represents the coverage obtained when mapping against each *Vibrio* species reference genome sequence, considering unique high-quality (MQ≥30) reads, and normalized, for each sample, by the total number of reads retained after adapter trimming/quality filtering.

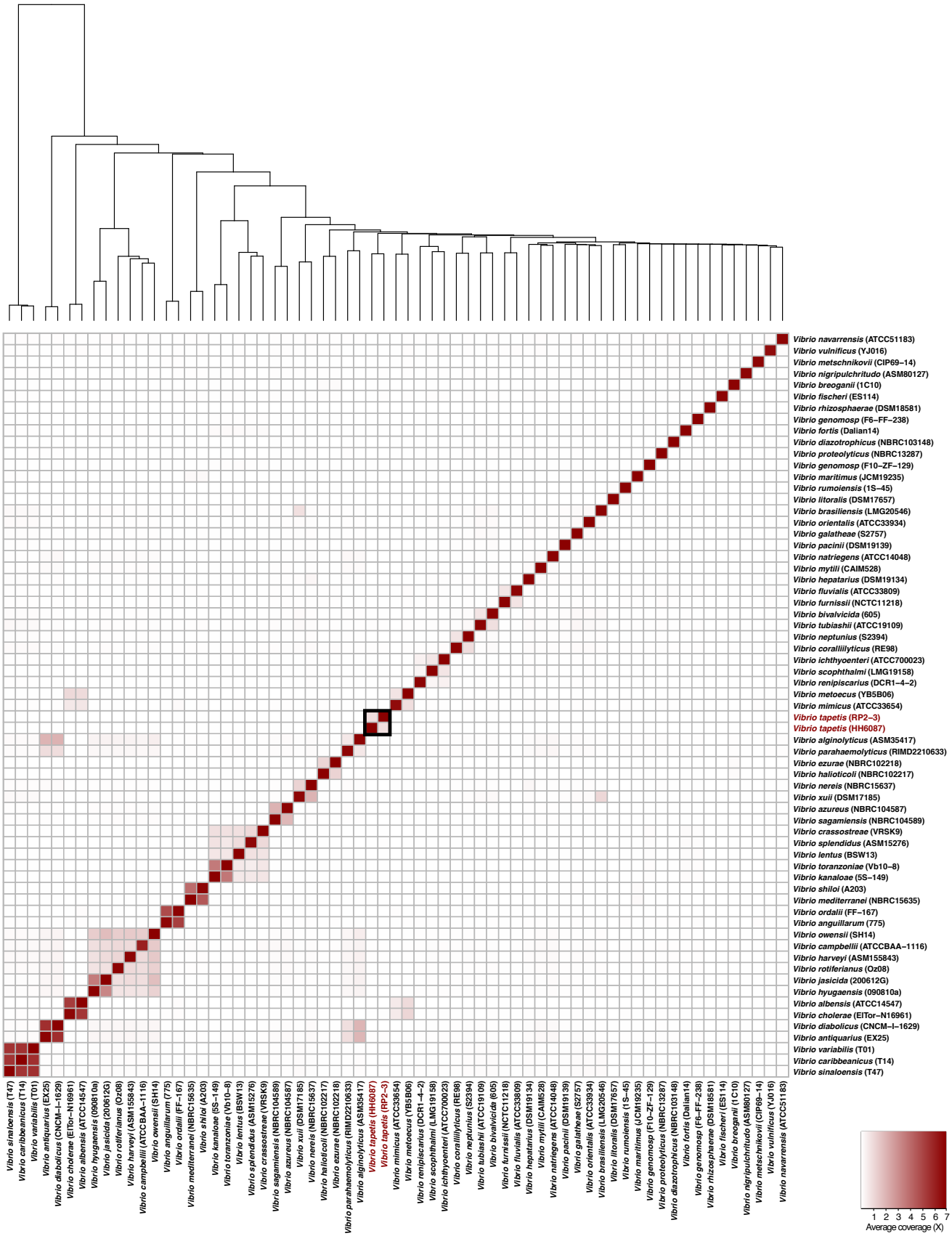

Figure S9. Mapping of *Vibrio* DNA reads against reference genome sequences representative of the known *Vibrio* species variability.

Vibrio DNA reads were simulated from each *Vibrio* reference genome sequence to represent a depth-of-coverage of 7X. The colour gradient represents the coverage obtained when mapping against each *Vibrio* species reference genome sequence, considering unique high-quality (MQ≥30) reads, and normalized, for each sample, by the total number of reads retained after adapter trimming/quality filtering. Samples are ordered by clustering based on Euclidean distances and a complete clustering method.

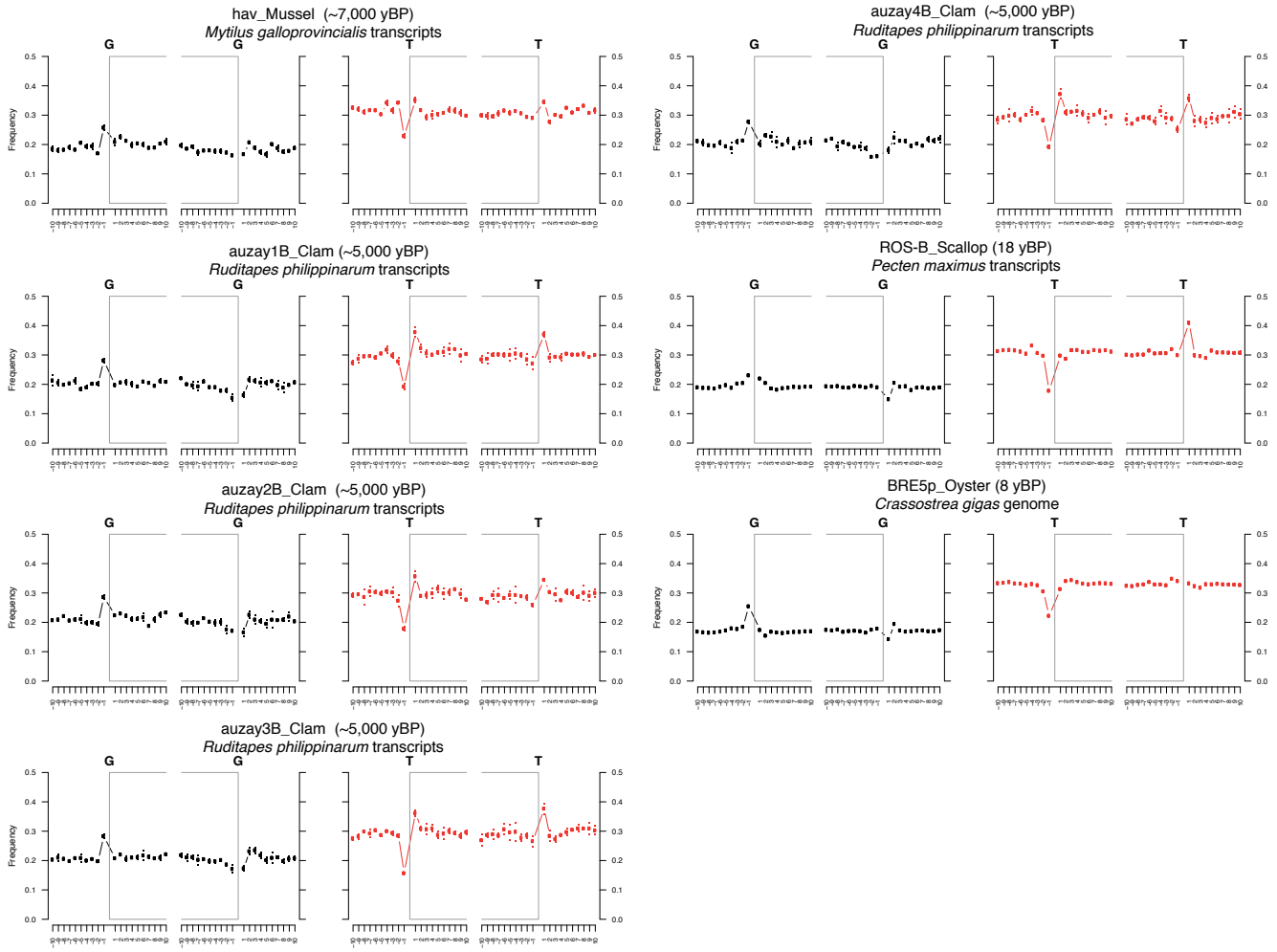

Figure S10. DNA fragmentation patterns in mollusc DNA from ancient and modern shells.
DNA fragmentation patterns are shown for reads mapping to the mollusc nuclear reference sequence showing the maximal depth-of-coverage for each sample. They are shown within 10 bp from read ends (post-adapter and quality trimming; positions 1-10 on the right panel of each base composition profile). Only unique high-quality (MQ≥30) reads were considered in damage analyses. mapDamage v.2 was run on the full alignment (100,0000 iterations).

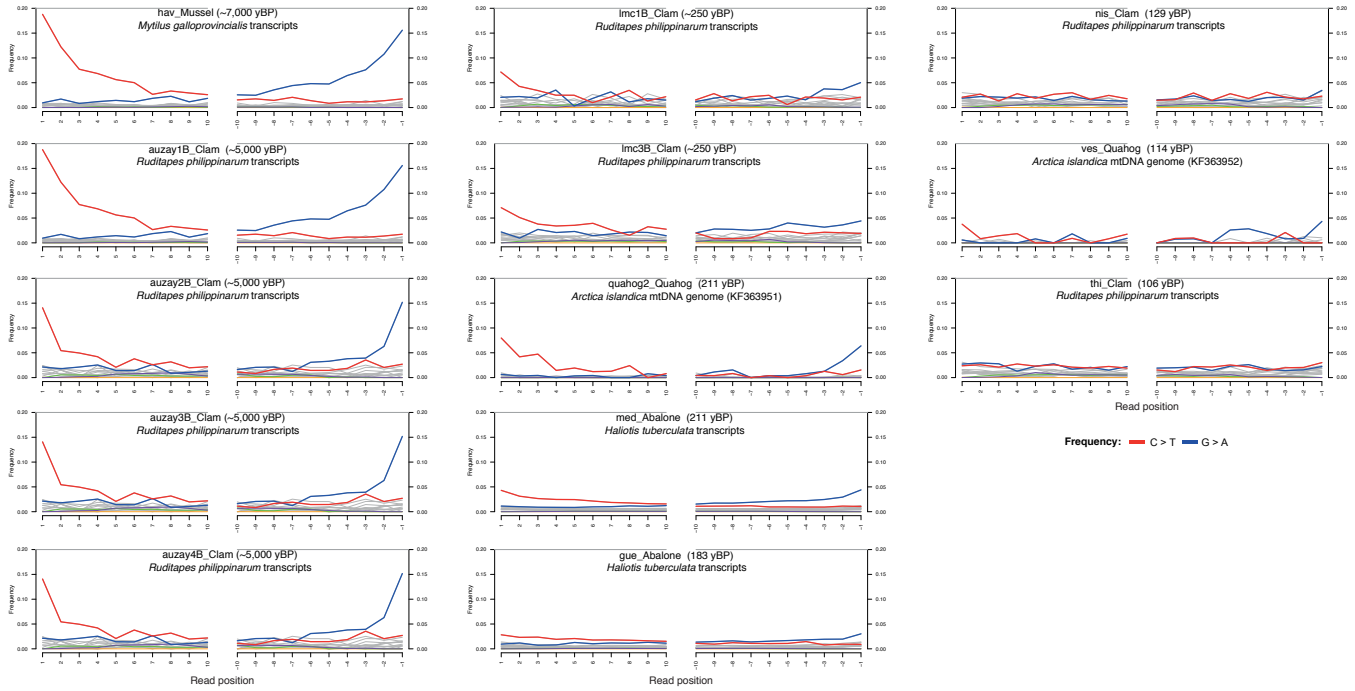

Figure S11. Nucleotide misincorporation patterns in mollusc DNA from ancient shells.
Nucleotide misincorporation patterns are shown for reads mapping to the nuclear (or mitochondrial) reference sequence showing the maximal depth-of-coverage for each sample. They are represented along the first and last 10 read positions. The x-axis provides read positions relative to read starts (positive numbers) and read ends negative numbers. Only unique high-quality (MQ≥30) reads were considered in damage analyses. mapDamage v.2 was run on the full alignment (100,0000 iterations).

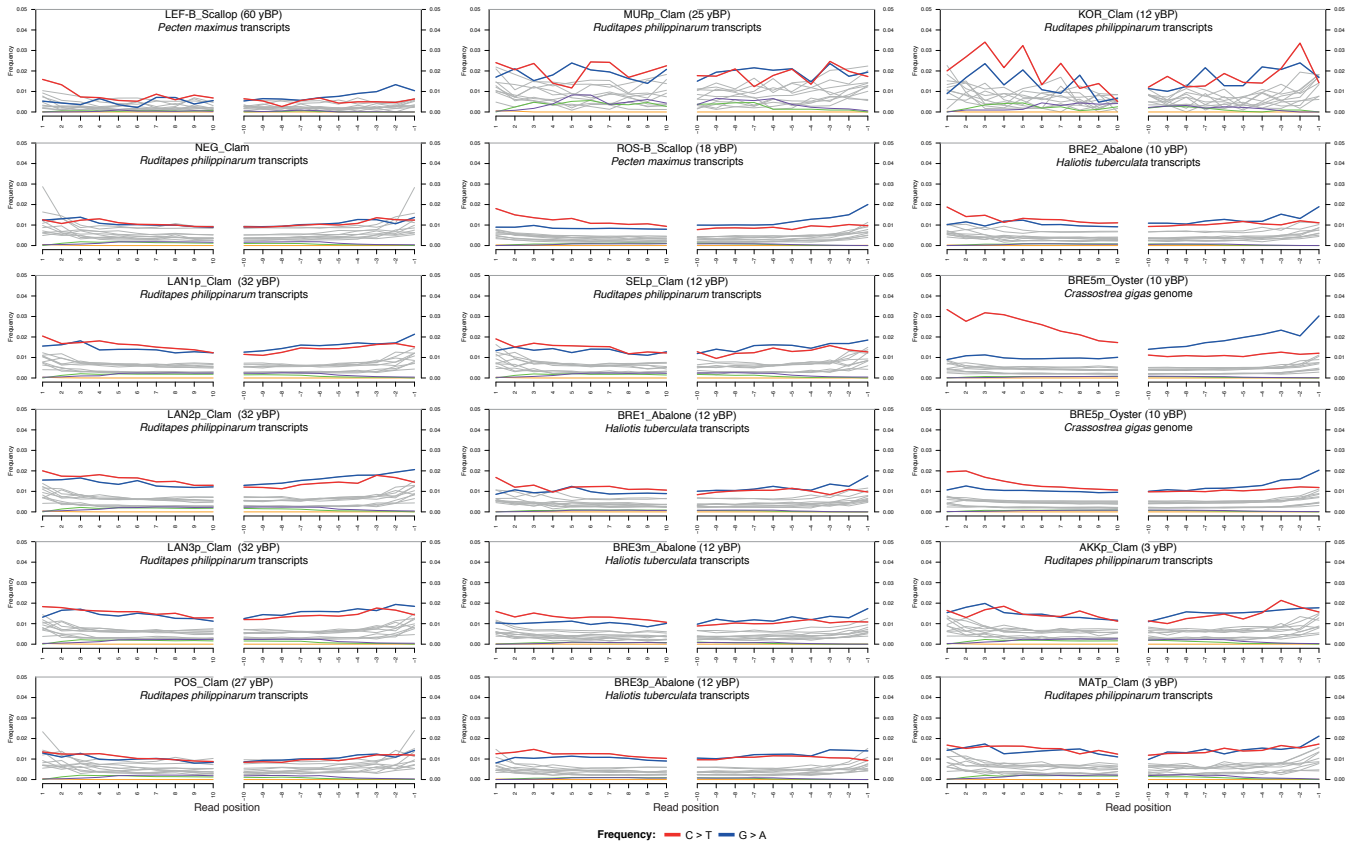

Figure S12. Nucleotide misincorporation patterns in mollusc nuclear DNA from modern shells.
Nucleotide misincorporation patterns are shown for reads mapping to the nuclear reference sequence showing the maximal depth-of-coverage for each sample. They are represented along the first and last 10 read positions. The x-axis provides read positions relative to read starts (positive numbers) and read ends negative numbers. Only unique high-quality (MQ≥30) reads were considered in damage analyses. mapDamage v.2 was run on the full alignment (100,0000 iterations).

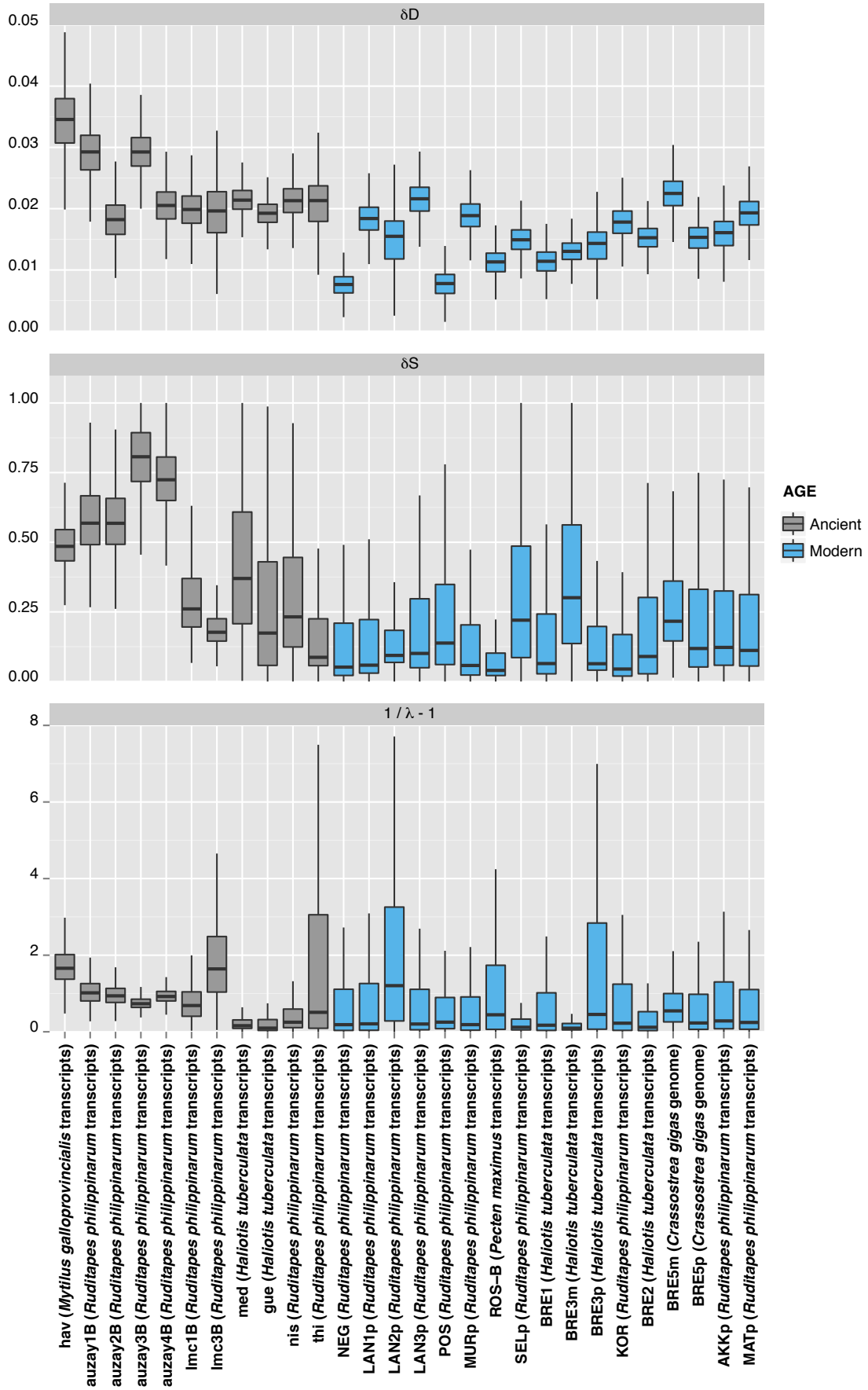

Figure S13. Posterior distributions of three damage parameters (δ_D, δ_S, and 1/ λ - 1) in mollusc nuclear DNA from shells.

Damage parameters were estimated from 100,000 iterations for 1,485 random reads mapping to the mollusc nuclear reference sequence showing the maximal depth-of-coverage for each sample (indicated in brackets). Only unique high-quality (MQ≥30) reads were considered in damage analyses. "δ_D", rate of cytosine deamination at double stranded regions; "δ_S", at single stranded overhangs; "1/ λ – 1", proxy for overhang length.

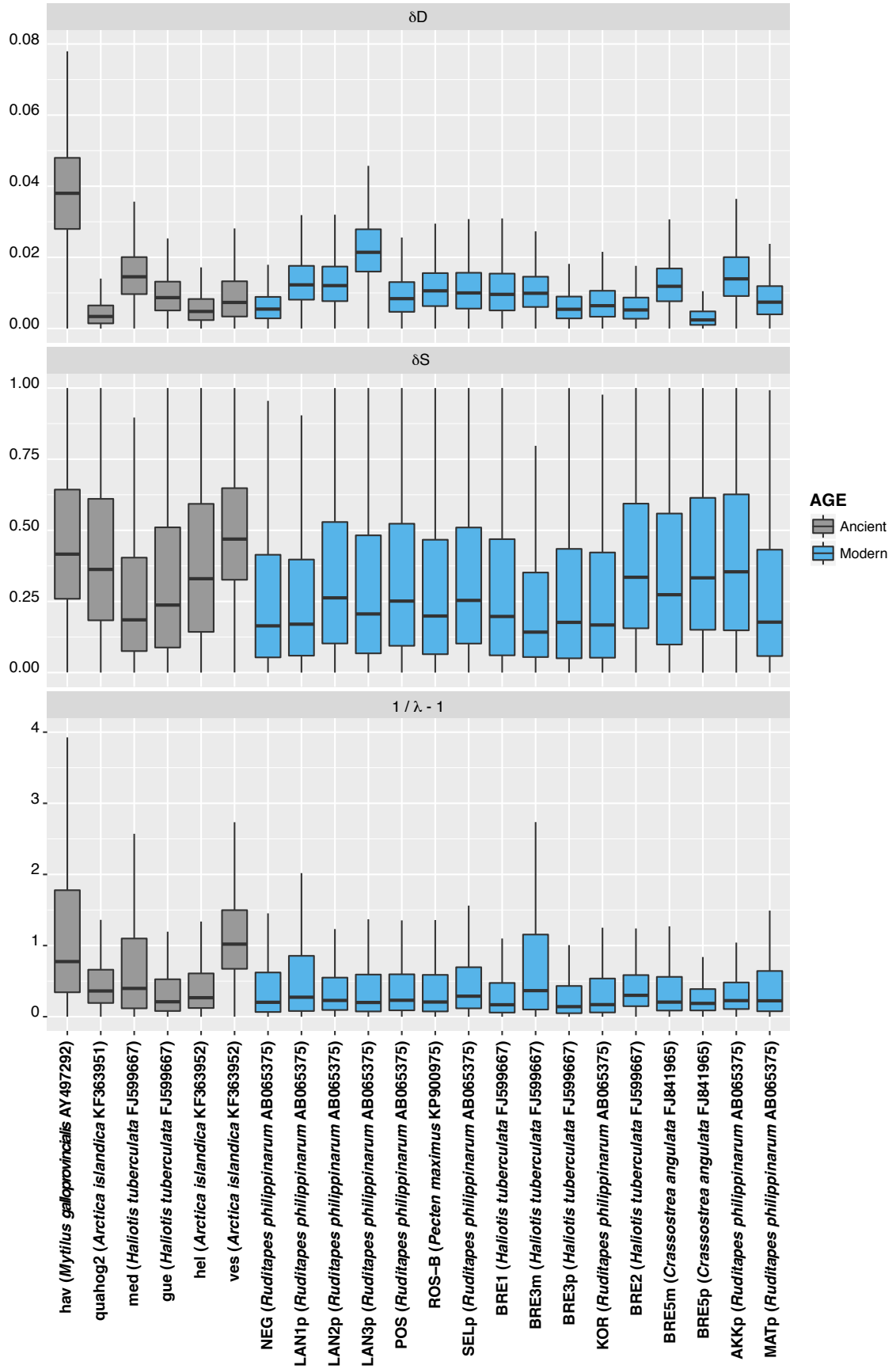

Figure S14. Posterior distributions of three damage parameters (δ_D, δ_S and 1/ λ - 1) in mollusc mitochondrial DNA from shells.

Damage parameters were estimated from 100,000 iterations for 80 random reads mapping to the mollusc nuclear reference sequence showing the maximal depth-of-coverage for each sample (indicated in brackets). Only unique high-quality (MQ≥30) reads were considered in damage analyses. "δ_D", rate of cytosine deamination at double stranded regions; "δ_S", at single stranded overhangs; "1/ λ – 1", proxy for overhang length.

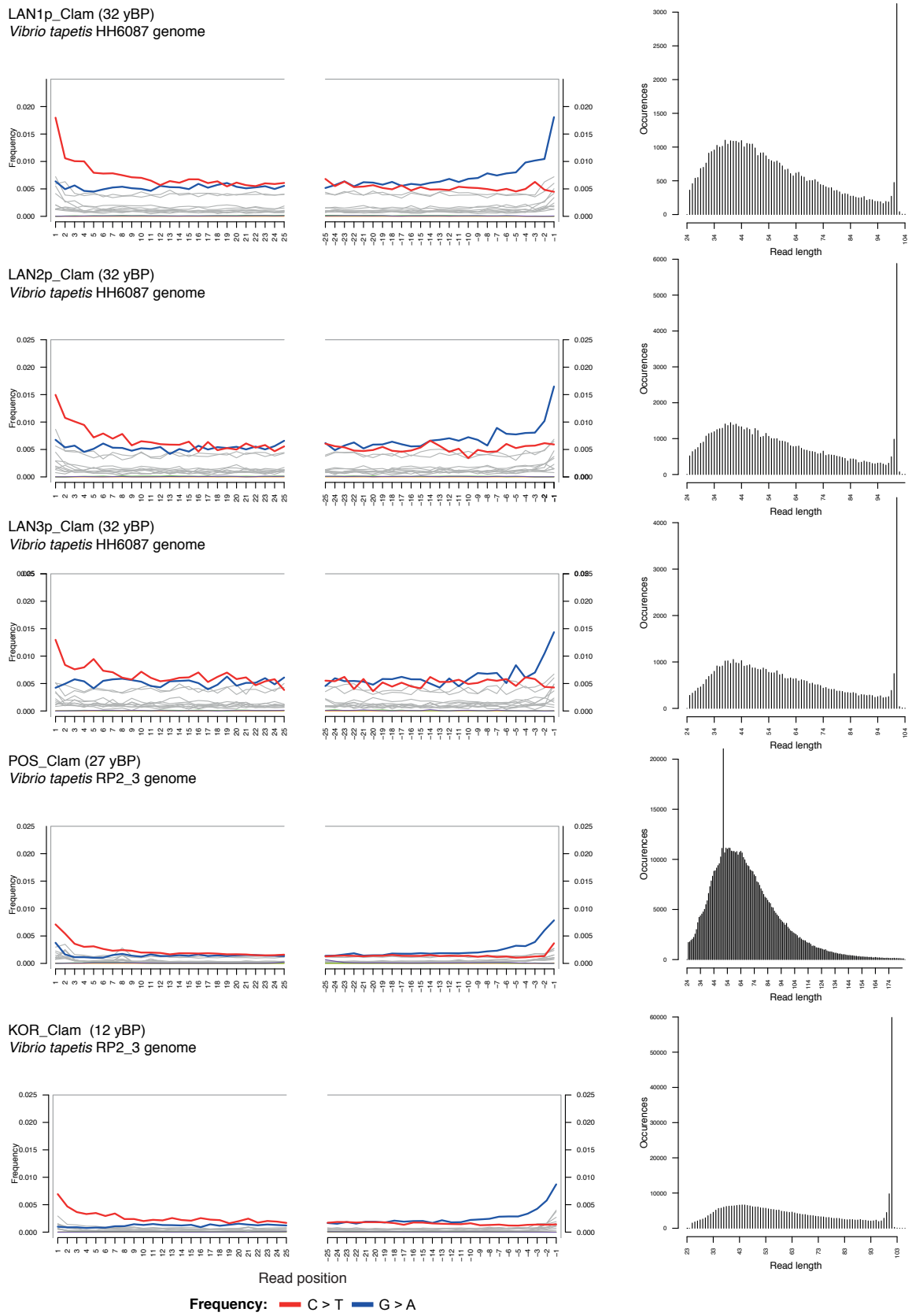

Figure S15. DNA damage patterns in *Vibrio tapetis* DNA recovered from infected shells.
Nucleotide misincorporation patterns and size distributions are shown for reads mapping to the *Vibrio tapetis* genome reference sequence showing the maximal depth-of-coverage for each sample. Nucleotide misincorporation is represented along the first and last 25 read positions. The x-axis provides read positions relative to read starts (positive numbers) and read ends (negative numbers). Only unique high-quality (MQ≥30) reads were considered in damage analyses. mapDamage v.2 was run on the full alignment (100,0000 iterations). Note, that the peak at 50bp in the fragment size distribution of POS_Clam is due to the presence of untrimmed 50 bp single-stranded MiSeq reads.

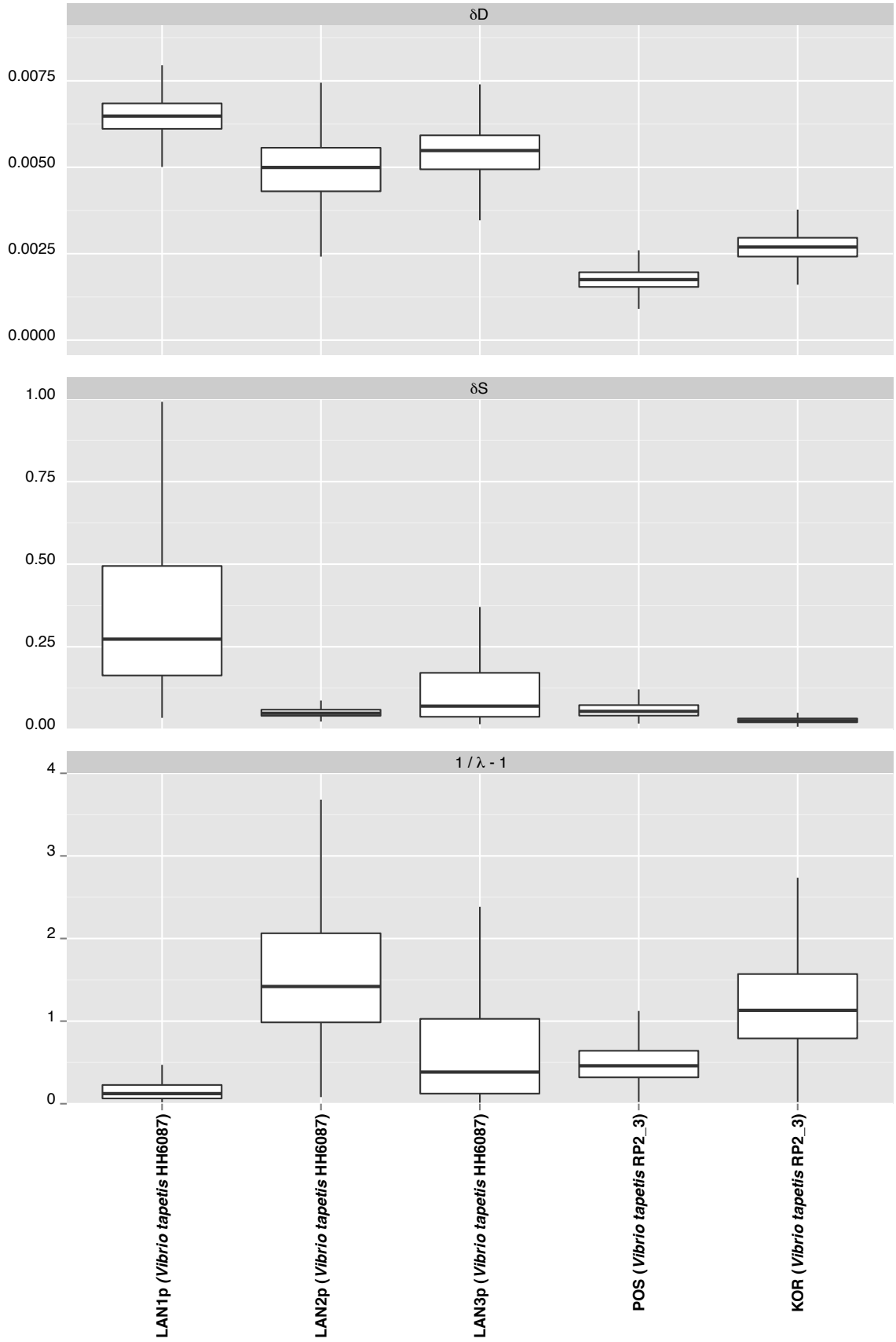

Figure S16. Posterior distributions of three damage parameters (δ_D, δ_S and 1/ λ - 1) in *Vibrio tapetis* DNA from modern infected shells.

Damage parameters were estimated from 100,000 iterations for 12,141 random reads mapping to the *Vibrio tapetis* genome reference sequence showing the maximal depth-of-coverage for each sample (indicated in brackets). Only unique high-quality (MQ≥30) reads were considered in damage analyses. "δ_D", rate of cytosine deamination at double stranded regions; "δ_S", at single stranded overhangs; "1/ λ – 1", proxy for overhang length.

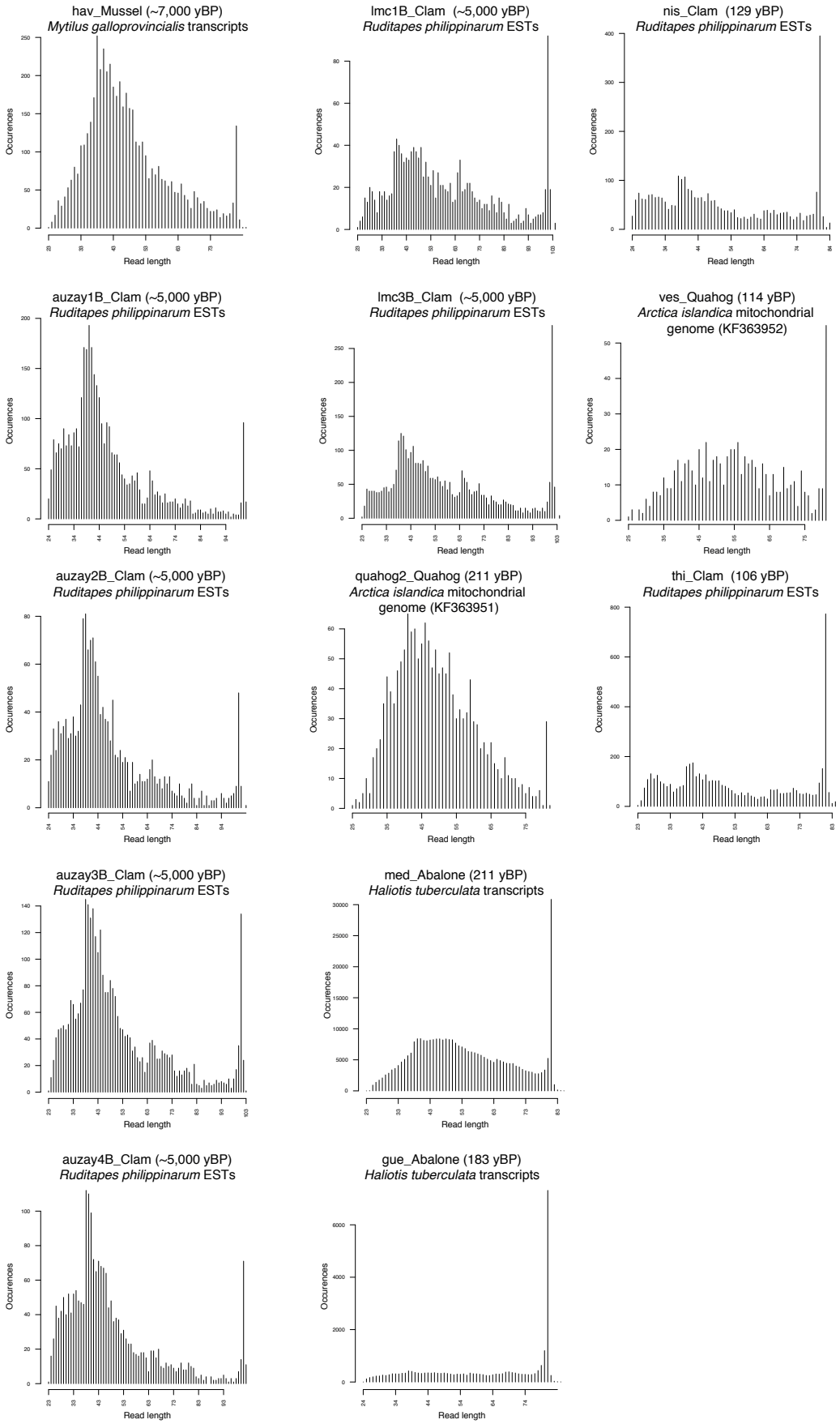

Figure S17. Fragment size distribution patterns in mollusc nuclear DNA from ancient shells.
Fragment size distributions are shown for reads mapping to the mollusc nuclear reference sequence showing the maximal depth-of-coverage for each sample. Only unique high-quality (MQ≥30) reads were considered in damage analyses. mapDamage v.2 was run on the full alignment (100,0000 iterations).

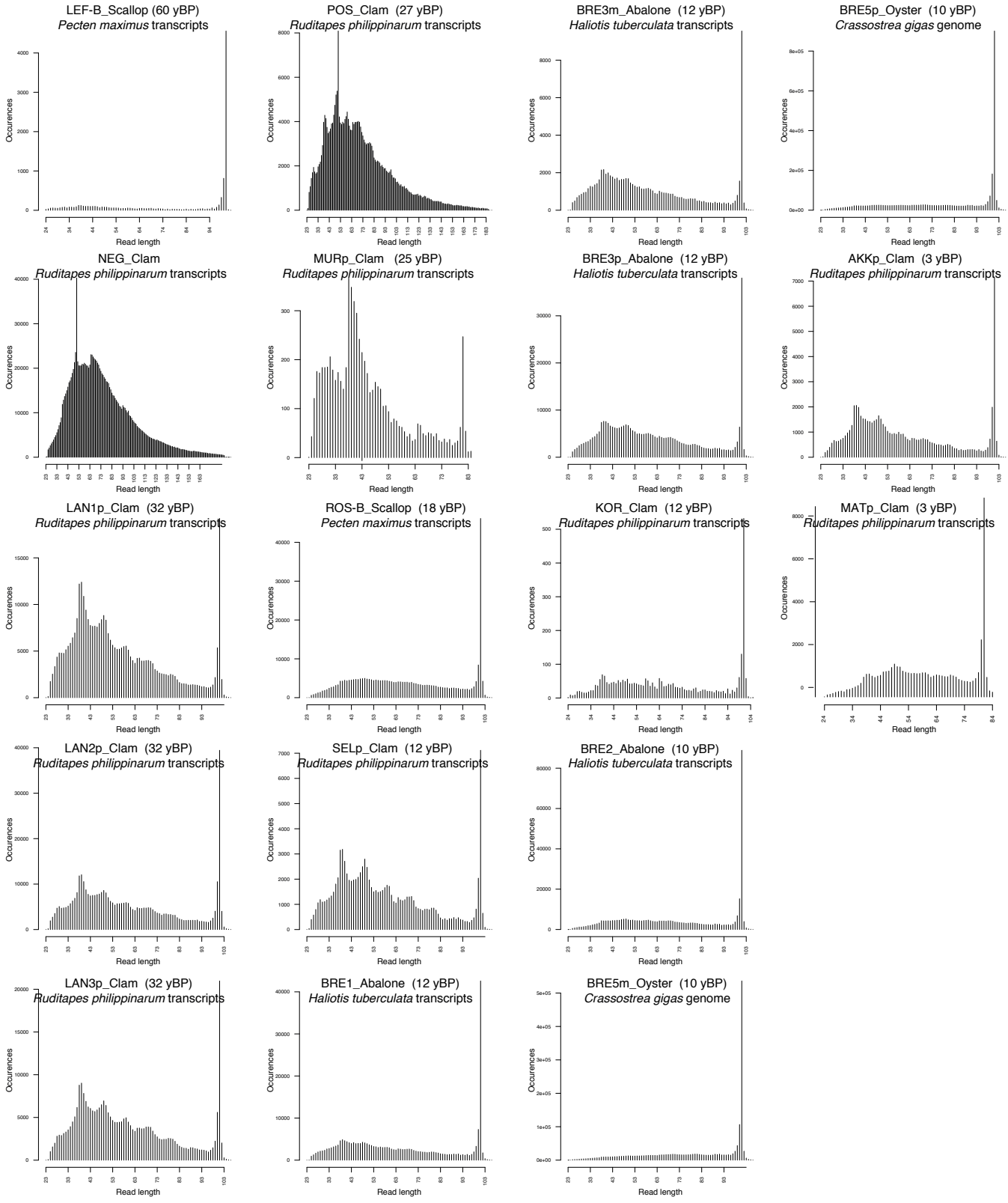

Figure S18. Fragment size distribution patterns in mollusc nuclear DNA from modern shells.
Fragment size distributions are shown for reads mapping to the mollusc nuclear reference sequence showing the maximal depth-of-coverage for each sample. Only unique high-quality (MQ≥30) reads were considered in damage analyses. mapDamage v.2 was run on the full alignment (100,0000 iterations). Note, that the peak at 50 bp in the fragment size distribution of POS_Clam and NEG_Clam is due to the significant presence of untrimmed 50 bp single-stranded MiSeq reads.

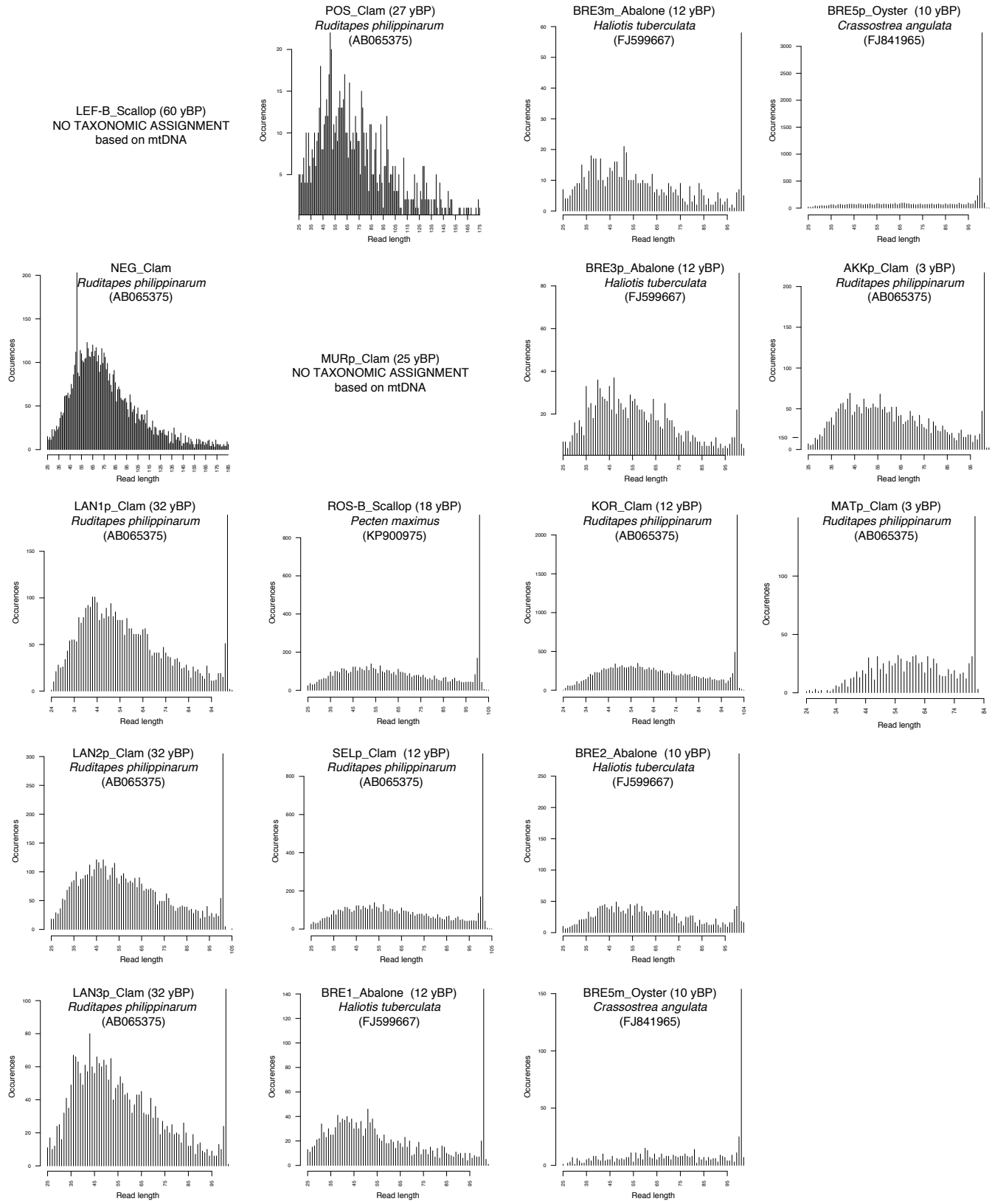

Figure S19. Fragment size distribution patterns in mollusc mitochondrial DNA from modern shells.
Fragment size distributions are shown for reads mapping to the mollusc mitochondrial reference sequence showing the maximal depth-of-coverage for each sample. Only unique high-quality (MQ≥30) reads were considered in damage analyses. mapDamage v.2 was run on the full alignment (100,0000 iterations). Note, that the peak at 50bp in the fragment size distribution of POS_Clam and NEG_Clam is due to the significant presence of untrimmed 50 bp single-stranded MiSeq reads.