Electronic Supplementary Material (ESI) for RSC Chemical Biology.

Supporting Information

N^4-Allylcytidine: A New Nucleoside Analogue for RNA Labelling and Chemical Sequencing

Tengwei Li,[a] Xiao Shu, [a] Minsong Gao, [a] Chenyang Huang, [a] Ting Li, [a] Jie Cao, [ab] Xiner Ying, [a] Donghong Liu,[a] and Jianzhao Liu*[ab]

a. MOE Key Laboratory of Macromolecular Synthesis and Functionalization, Department of Polymer Science and Engineering, Zhejiang University, Yuhangtang Road 866, Hangzhou 310058, Zhejiang Province, China

b. Life Sciences Institute, Zhejiang University, Yuhangtang Road 866, Hangzhou 310058, Zhejiang Province, China

Table of Content

Section I: Characterizations of a^{4}C and a^{4}CTP and their synthetic intermediates

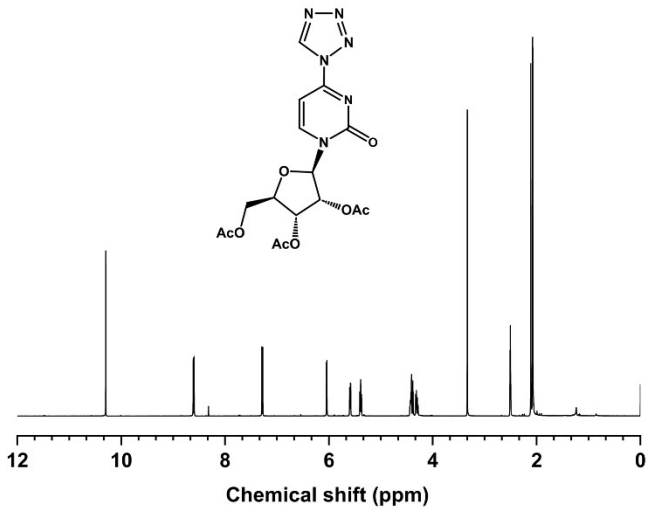


Figure S1. ^{1}H NMR spectrum of 4-(tetrazol-1-yl)-1-(2',3',5'-tri-*O*-acetyl-β-D-ribofuranosyl) pyrimidine-2-(1H)-one **(2)**. ^{1}H NMR (400 MHz, DMSO-d_6), δ (ppm): 10.30 (s, 1H), 8.60 (d, *J* = 7.3 Hz, 1H), 7.28 (d, *J* = 7.2 Hz, 1H), 6.04 (d, *J* = 3.4 Hz, 1H), 5.59 (dd, *J* = 3.3, 6.1 Hz, 1H), 5.39 (t, *J* = 6.5 Hz, 1H), 4.40 (d, *J* = 10.1 Hz, 2H), 4.27 – 4.35 (m, 1H), 2.11 (s, 3H), 2.08 (s, 3H), 2.07 (s, 3H).

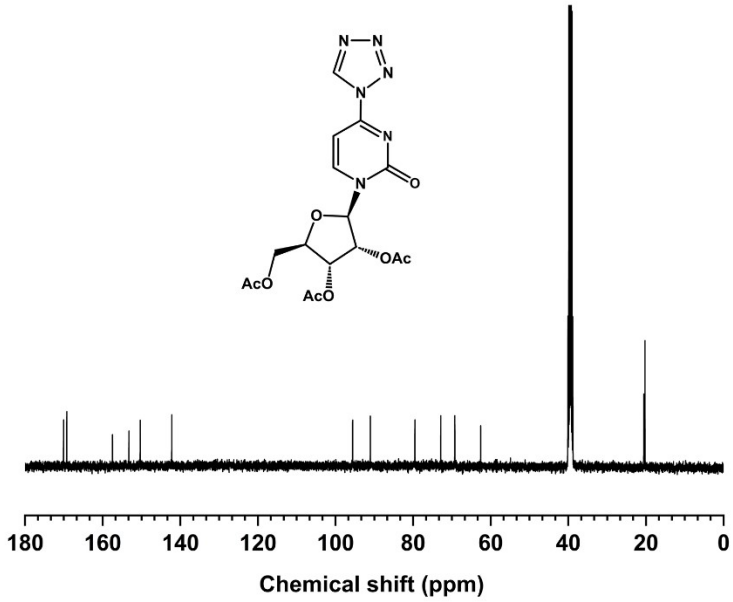


Figure S2. ^{13}C NMR spectrum of 4-(tetrazol-1-yl)-1-(2',3',5'-tri-*O*-acetyl-β-D-ribofuranosyl) pyrimidine-2-(1H)-one **(2)**. ^{13}C NMR (100 MHz, DMSO-d_6), δ (ppm): 170.1, 169.2, 157.5, 153.2, 150.3, 142.2, 95.6, 91.0, 79.5, 72.9, 69.3, 62.6, 20.6, 20.3

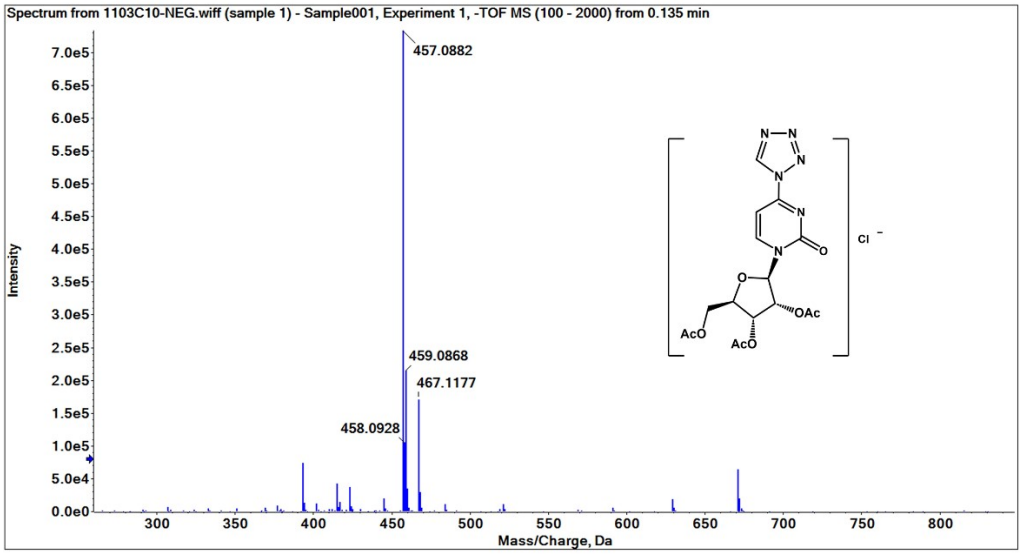


Figure S3. High-resolution mass spectrum (HRMS) of 4-(tetrazol-1-yl)-1-(2',3',5'-tri-*O*-acetyl-β-D-ribofuranosyl) pyrimidine-2-(1H)-one **(2)**. HRMS (ESI), *m/z* 457.0882 ([M+Cl]$^-$, calcd 457.0880).

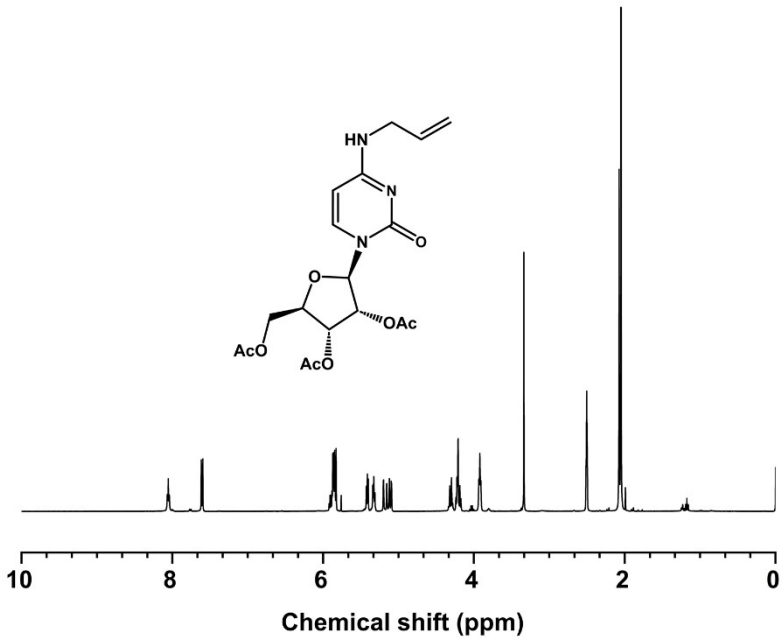


Figure S4. ^{1}H NMR spectrum of 2',3',5'-tri-*O*-acetyl-*N*^{4}-allylcytidine **(3)**. ^{1}H NMR (400 MHz, DMSO-d_6), δ (ppm): 8.05 (t, J = 5.6 Hz, 1H), 7.61 (d, J = 7.5 Hz, 1H), 5.76 – 5.97 (m, 3H), 5.37 – 5.47 (m, 1H), 5.33 (t, J = 5.9 Hz, 1H), 5.18 (dd, J = 1.7, 17.2 Hz, 1H), 5.11 (dd, J = 1.7, 10.2 Hz, 1H), 4.31 (d, J = 8.5 Hz, 1H), 4.13 – 4.26 (m, 2H), 3.86 – 3.98 (m, 2H), 2.07 (s, 3H), 2.05 (s, 6H).

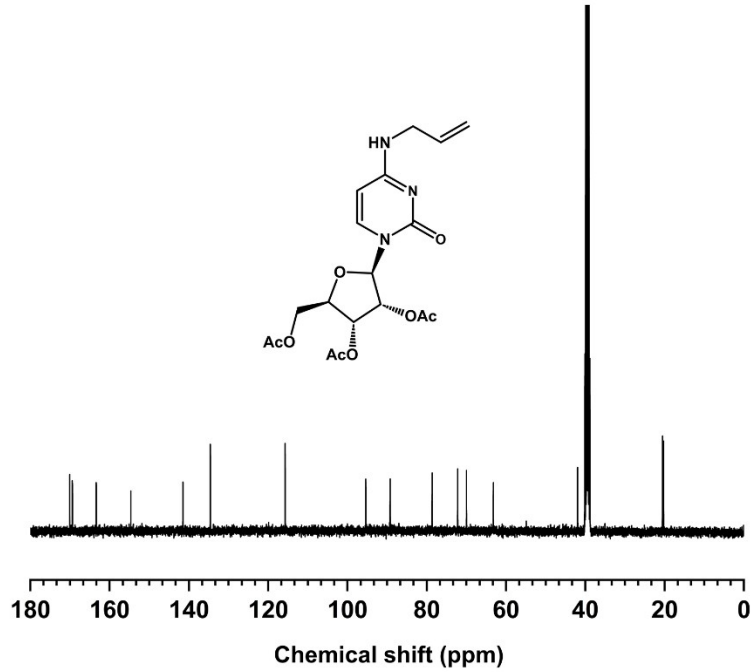


Figure S5. ^{13}C NMR spectrum of 2',3',5'-tri-*O*-acetyl-*N*^{4}-allylcytidine **(3)**. ^{13}C NMR (100 MHz, DMSO-d_6), δ (ppm): 170.1, 169.4, 169.3, 163.4, 154.6, 141.5, 134.6, 115.7, 95.4, 89.2, 78.6, 72.2, 70.0, 63.2, 42.0, 20.5, 20.3, 20.2

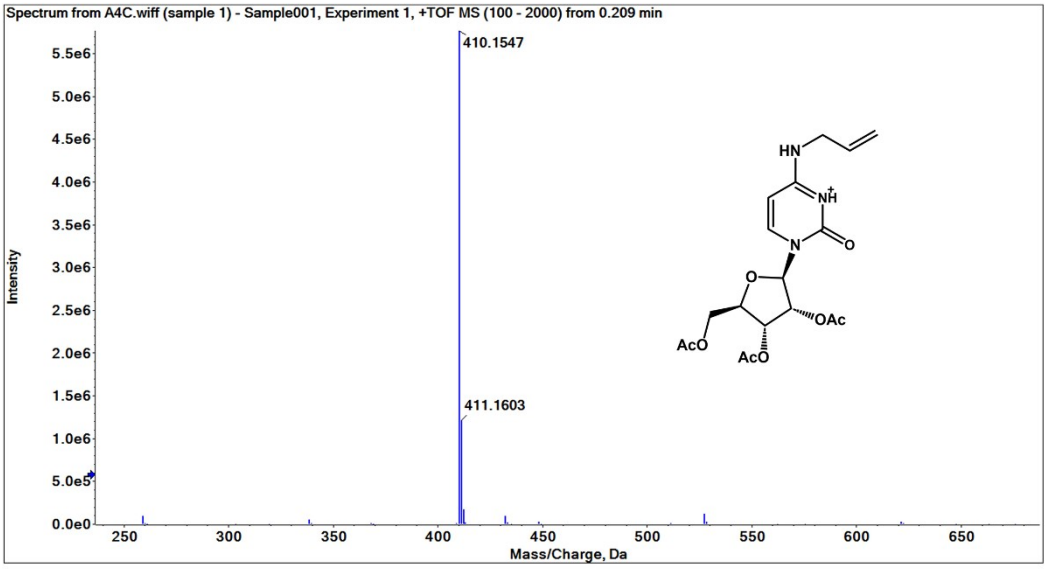


Figure S6. High-resolution mass spectrum of 2',3',5'-tri-O-acetyl-N^4-allylcytidine **(3)**. HRMS (ESI), *m/z* 410.1547 ([M+H]$^+$, calcd 410.1558).

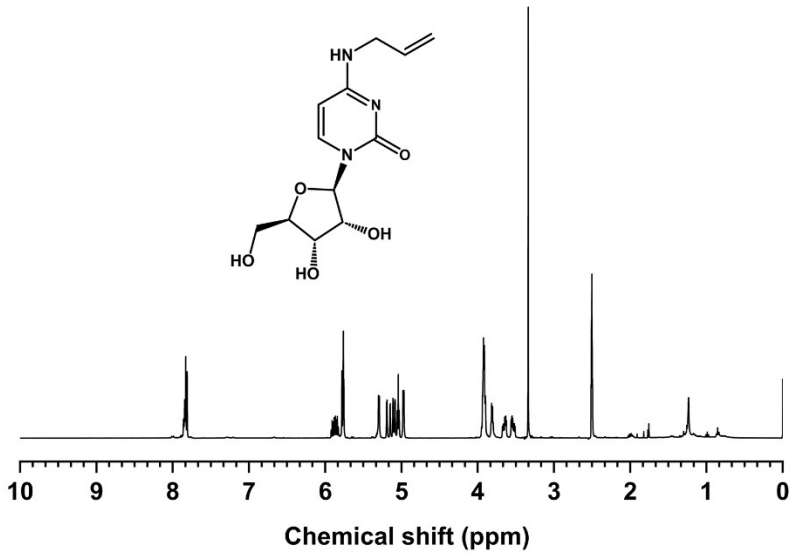


Figure S7. ^{1}H NMR spectrum of N^4-allylcytidine **(4)**. ^{1}H NMR spectrum (400 MHz, DMSO-d_6), δ (ppm): 7.74 – 7.90 (m, 2H), 5.87 (s, 1H), 5.76 (s, 2H), 5.30 (d, J = 5.2 Hz, 1H), 5.17 (dd, J = 1.8, 17.2 Hz, 1H), 5.10 (dd, J = 1.7, 10.2 Hz, 1H), 5.01 – 5.06 (m, 1H), 4.97 (d, J = 5.3 Hz, 1H), 3.92 (h, J = 4.4 Hz, 4H), 3.82 (t, J = 3.2 Hz, 1H), 3.64 (dt, J = 4.2, 8.4 Hz, 1H), 3.48 – 3.59 (m, 1H).

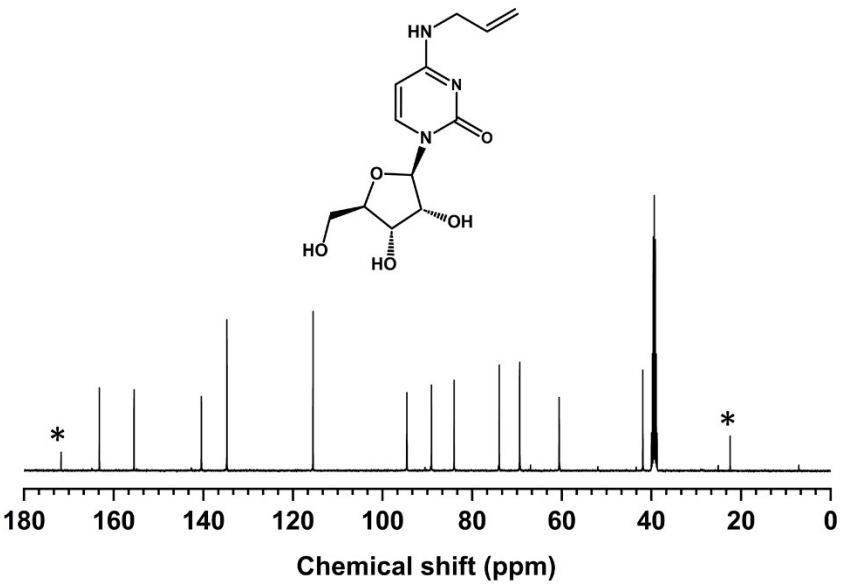


Figure S8. ^{13}C NMR spectrum of N^4-allylcytidine **(4)**. ^{13}C NMR spectrum (100 MHz, DMSO-d_6), δ (ppm): 163.2, 155.5, 140.4, 134.8, 115.5, 94.6, 89.1, 84.0, 74.0, 69.4, 60.6, 41.9. The asterisks (*) denote the signals from acetic acid.

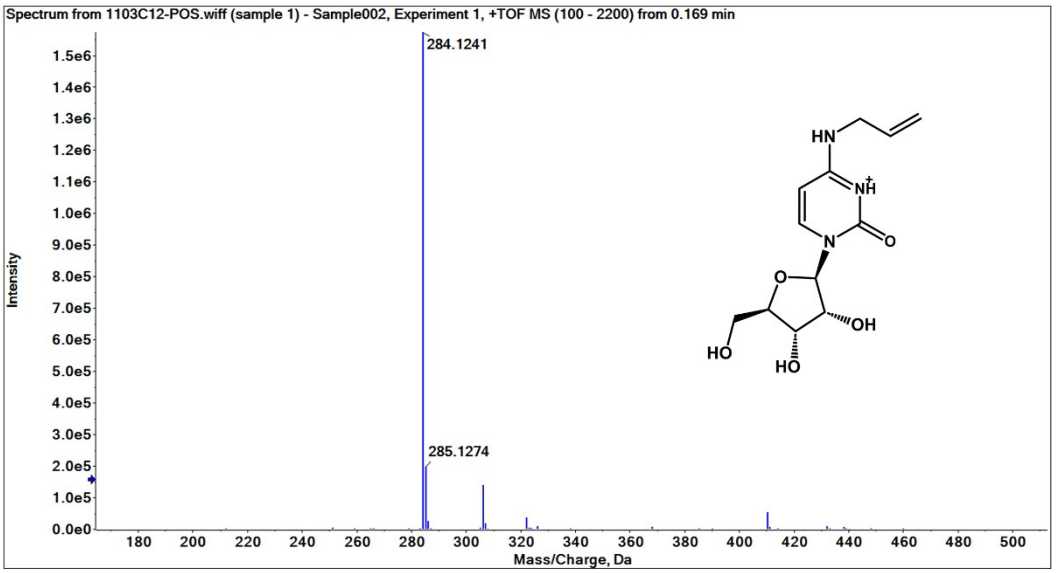


Figure S9. High-resolution mass spectrum of N^4-allylcytidine **(4)**. HRMS (ESI), *m/z* 284.1241 ([M+H]$^+$, calcd 284.1241).

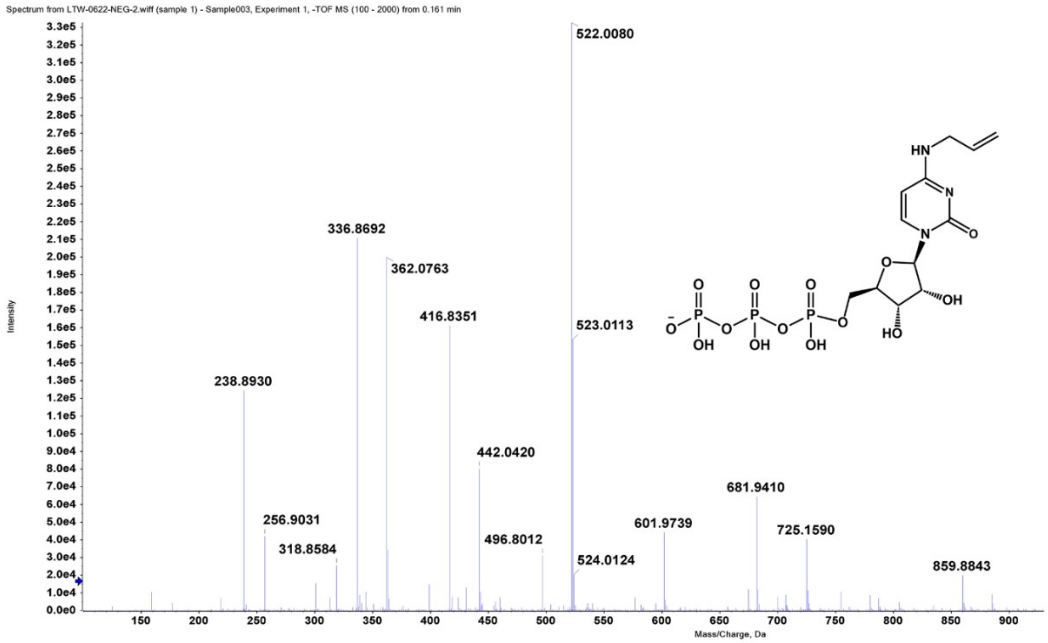


Figure S10. High-resolution spectrum of N^4-allylcytidine-5'-triphosphate **(5)**. HRMS (ESI), *m/z* 522.0080 ([M-H]$^-$, calcd 522.0085).

Section II: Model reactions of a^{4}C

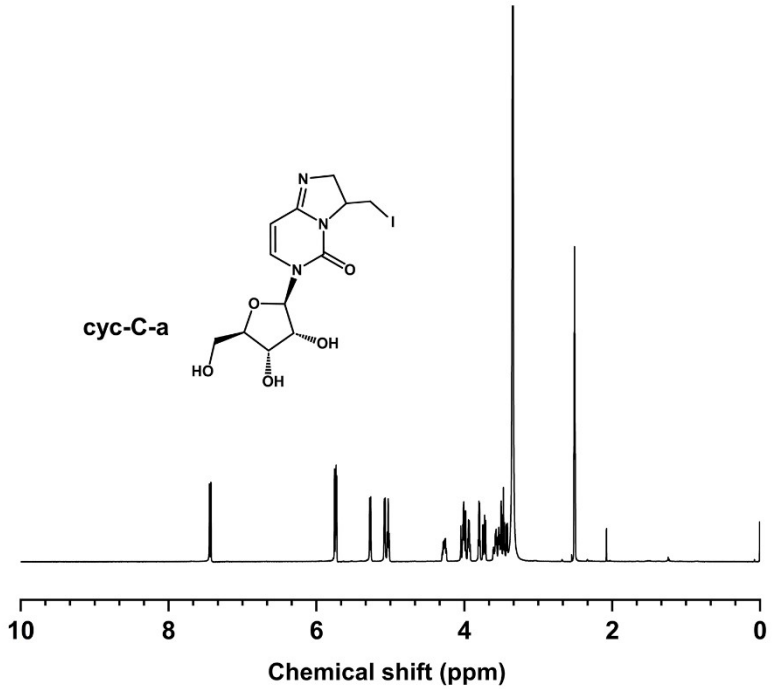


Figure S11. ^{1}H NMR spectrum (400 MHz, DMSO-d_6) of 3, N^4-cyclized cytidine (cyc-C-a). δ (ppm): 7.43 (d, J = 8.1 Hz, 1H), 5.80 – 5.66 (m, 2H), 5.26 (d, J = 6.3 Hz, 1H), 5.07 (d, J = 5.0 Hz, 1H), 5.02 (t, J = 5.1 Hz, 1H), 4.26 (dtd, J = 11.3, 5.7, 2.4 Hz, 1H), 4.06 – 3.96 (m, 2H), 3.96 – 3.90 (m, 1H), 3.79 (q, J = 3.4 Hz, 1H), 3.72 (dd, J = 10.1, 5.6 Hz, 1H), 3.61 – 3.48 (m, 2H), 3.48 – 3.39 (m, 2H).

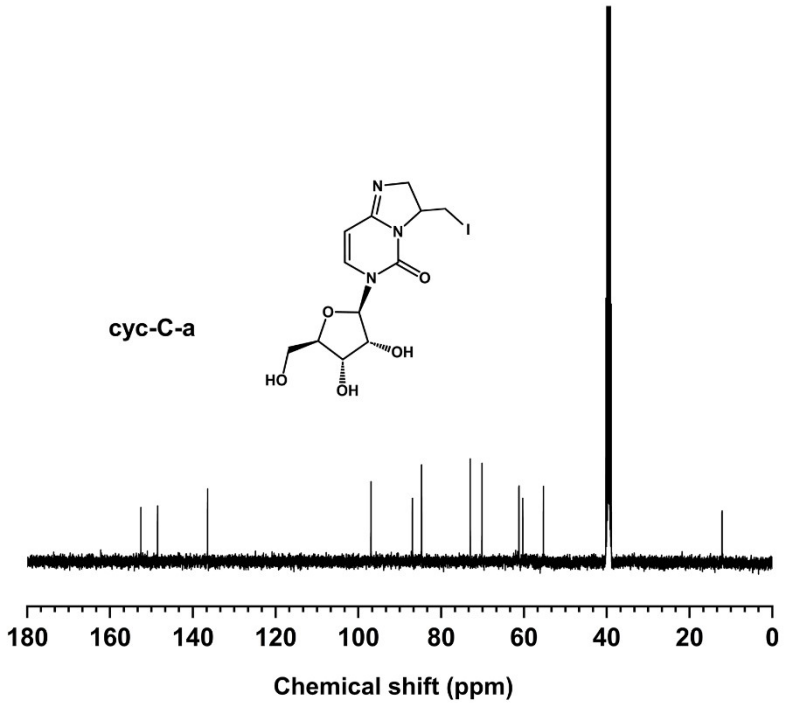


Figure S12. ^{13}C NMR spectrum (100 MHz, DMSO-d_6) of 3, N^4-cyclized cytidine (cyc-C-a). δ (ppm): 152.5, 148.5, 136.4, 96.9, 86.9, 84.7, 73.0, 70.1, 61.2, 60.3, 55.3, 12.1.

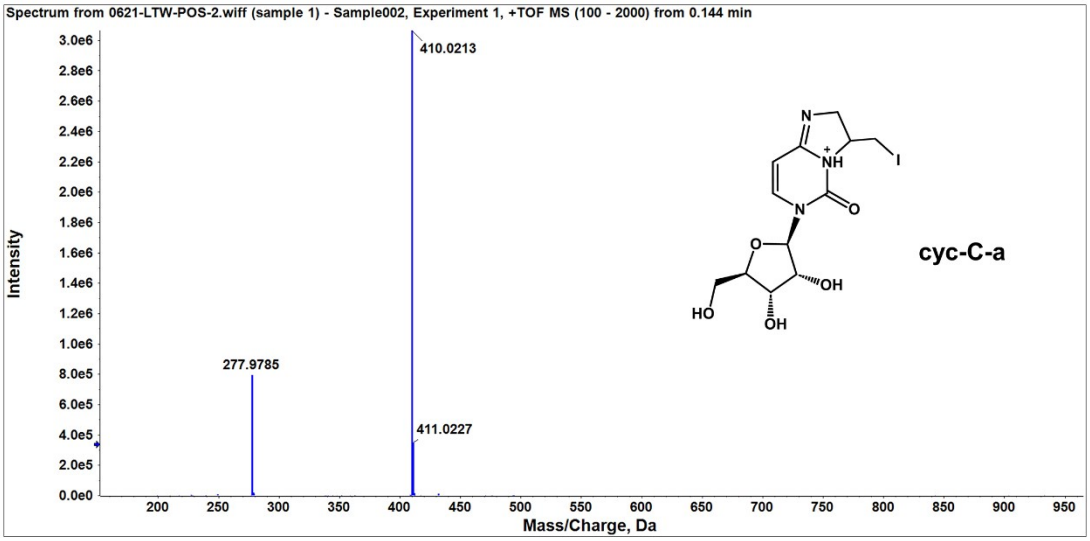


Figure S13. High-resolution mass spectrum of 3, N^4-cyclized cytidine (cyc-C-a). HRMS (ESI), *m/z* 410.0213 ([M+H]$^+$, calcd 410.0207).

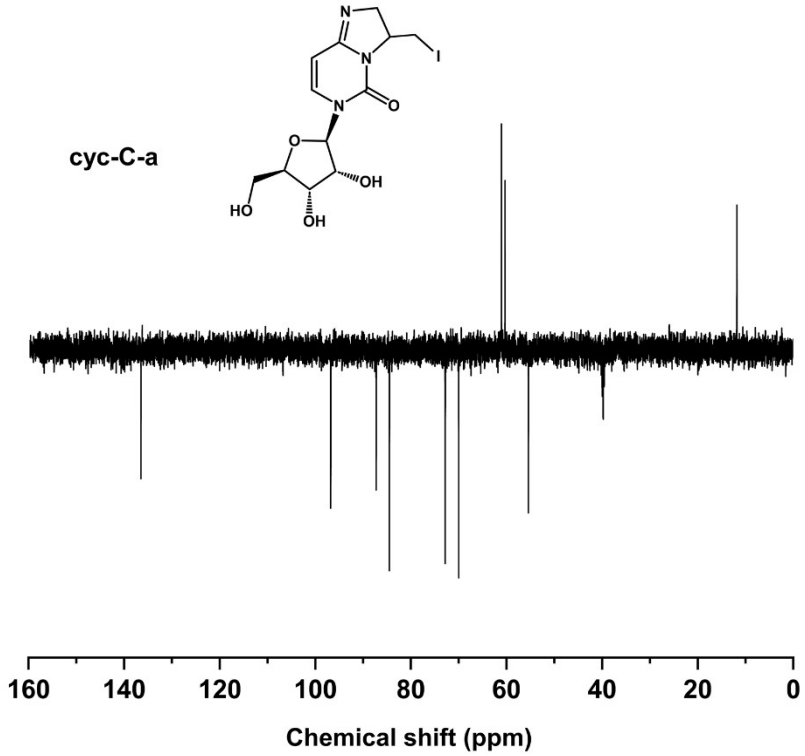


Figure S14. DEPT spectrum of 3, N^4-cyclized cytidine (cyc-C-a). δ (ppm): 135.9, 96.5, 87.0, 84.2, 72.6, 69.7, 60.8, 60.0, 55.2, 11.5

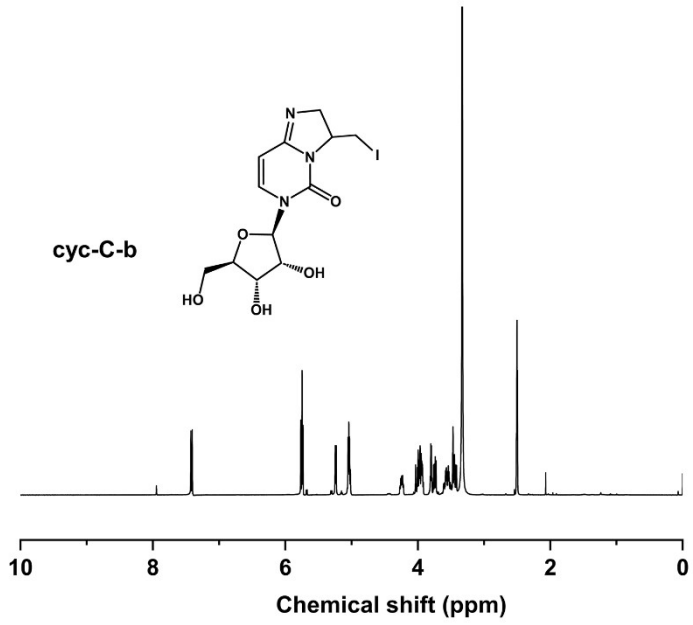


Figure S15. ^{1}H NMR spectrum (400 MHz, DMSO-d_6) of 3, N^4-cyclized cytidine (cyc-C-b). δ (ppm): 7.42 (d, J = 8.2 Hz, 1H), 5.84 – 5.69 (m, 2H), 5.24 (d, J = 5.8 Hz, 1H), 5.04 (p, J = 5.3 Hz, 2H), 4.25 (dtd, J = 10.6, 5.4, 2.1 Hz, 1H), 4.05 – 3.90 (m, 3H), 3.80 (q, J = 3.3 Hz, 1H), 3.75 (dd, J = 10.2, 5.3 Hz, 1H), 3.56 (qdd, J = 12.0, 5.1, 3.5 Hz, 2H), 3.47 (d, J = 2.3 Hz, 1H), 3.46 – 3.40 (m, 1H).

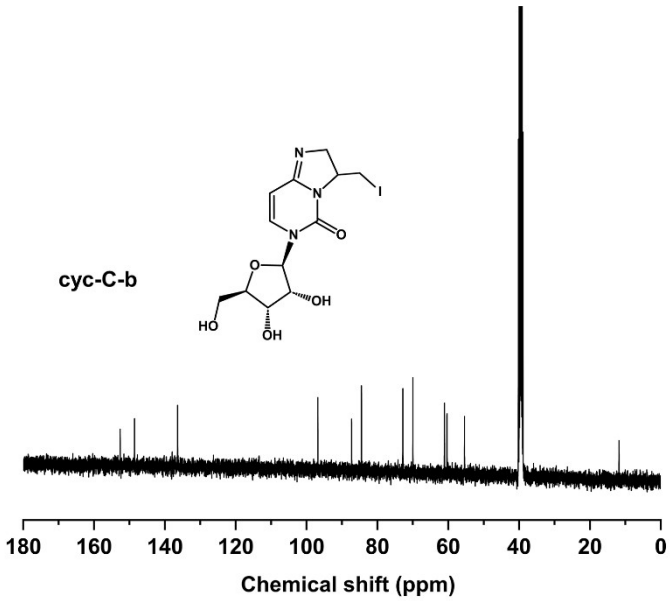


Figure S16. ^{13}C NMR spectrum (100 MHz, DMSO-d_6) of 3, N^4-cyclized cytidine (cyc-C-b). δ (ppm): 152.6, 148.5, 136.4, 96.8, 87.2, 84.4, 72.8, 70.0, 61.0, 60.3, 55.4, 11.7.

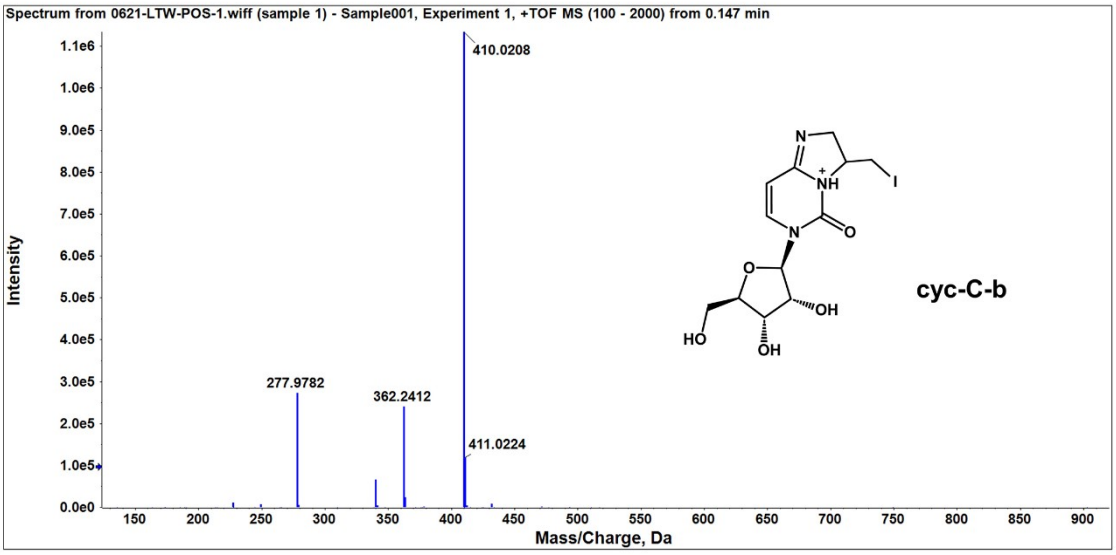


Figure S17. High-resolution mass spectrum of 3, N^4-cyclized cytidine (cyc-C-b). HRMS (ESI), *m/z* 410.0208 ([M+H]$^+$, calcd 410.0207).

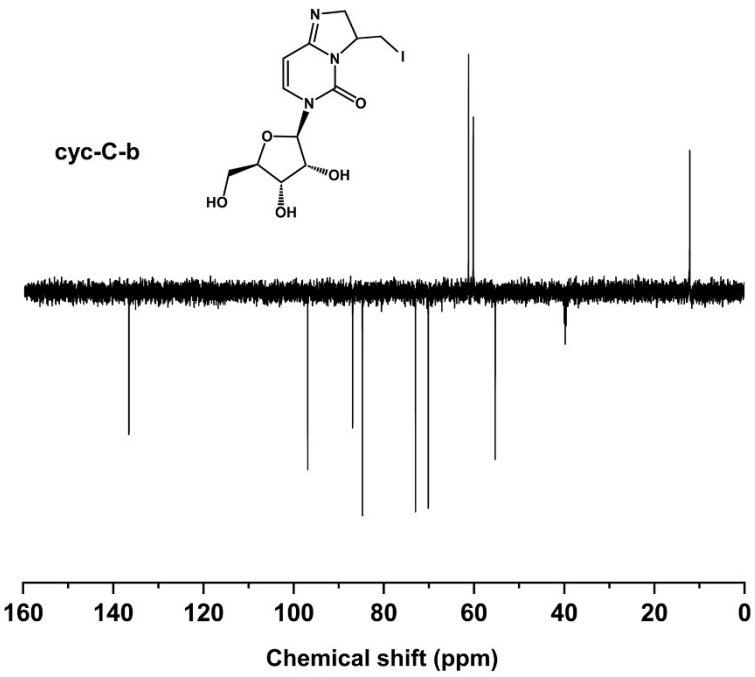


Figure S18. DEPT spectrum of 3, N^4-cyclized cytidine (cyc-C-a). δ (ppm): 136.3, 96.6, 86.7, 84.5, 72.7, 69.9, 60.9, 59.9, 55.1, 11.9.

Section III: Applications of a^{4}C in RNA sequencing

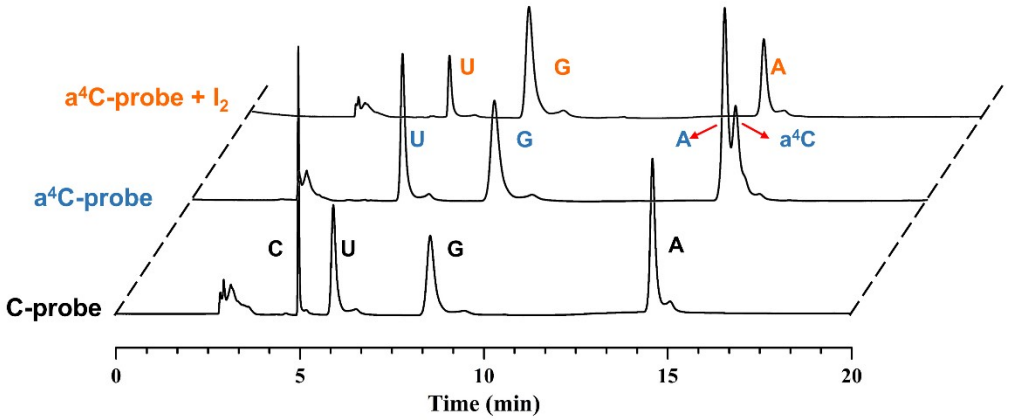


Figure S19. HPLC chromatograms of hydrolysed RNA C -probe,a^{4}C-probe and iodine-treated a^{4}C-probe .

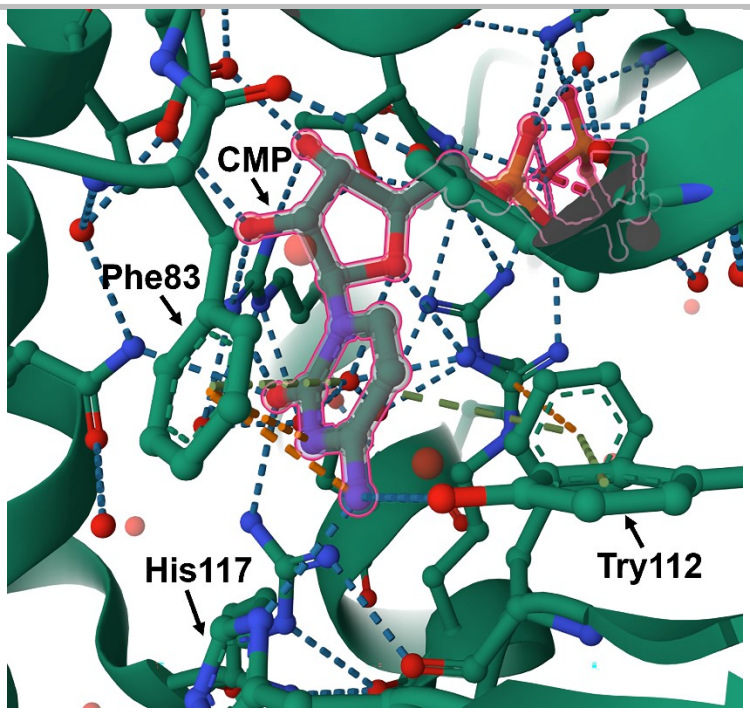


Figure S20. The X-ray crystal structure of UCK2 with the substrate cytidine monophosphate (CMP). This is adapted from the structure with PDB code 1XRJ (https://doi.org/10.2210/pdb1xrj/pdb).

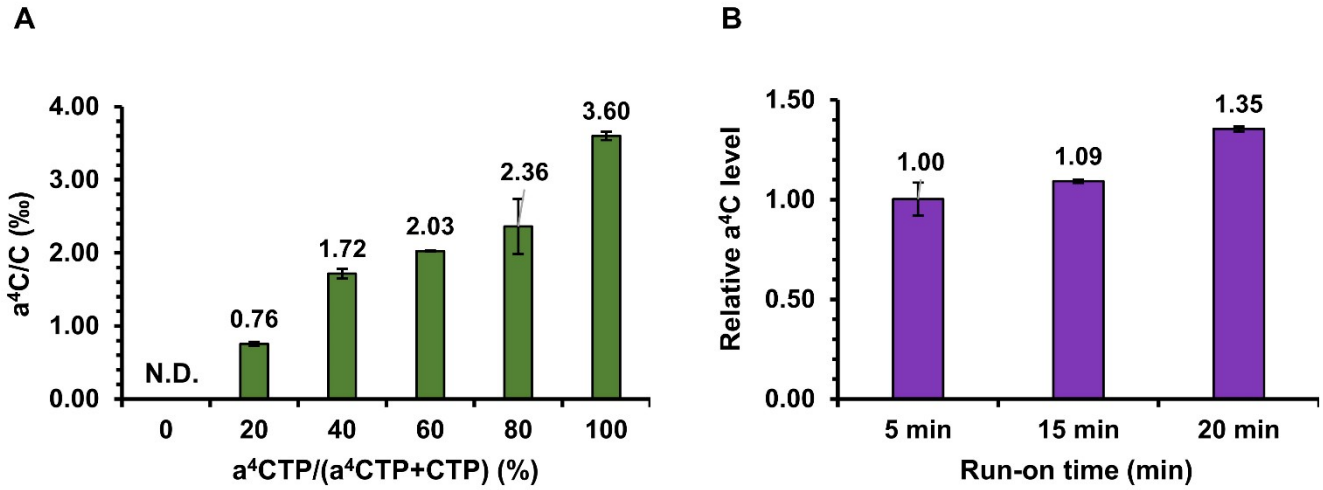


Figure S21. (A) The a^4C incorporation levels in RNA extracted from chromatin run-on reactions under different a^4CTP/CTP feeding ratios within 5 min. (B) The measured time course a^4C levels of RNAs extracted from chromatin run-on reactions in the presence of 100% a^4CTP.

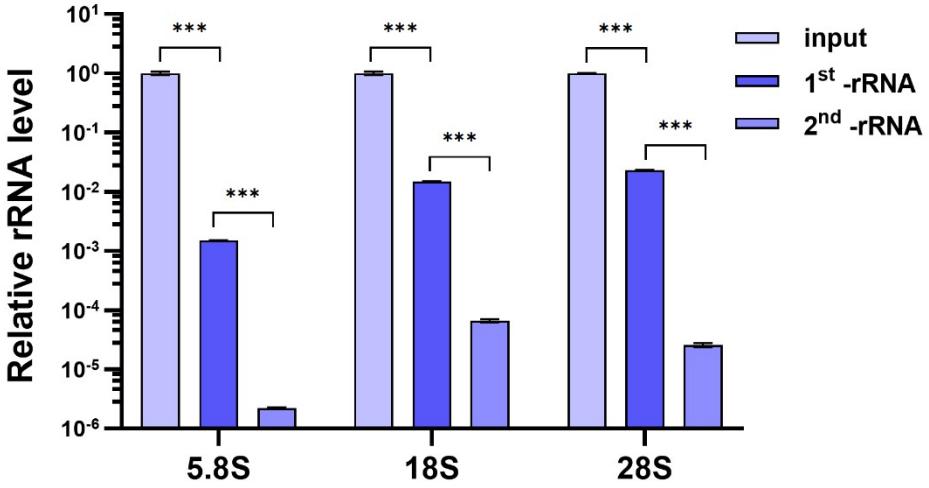


Figure S22. Quantification of different kinds of rRNAs after rRNA depletion by RT-qPCR.

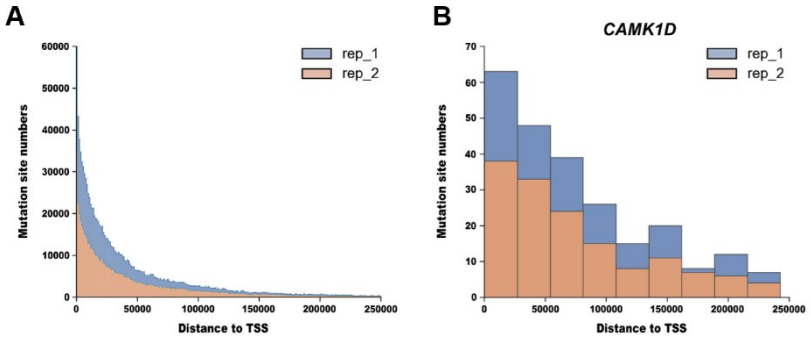
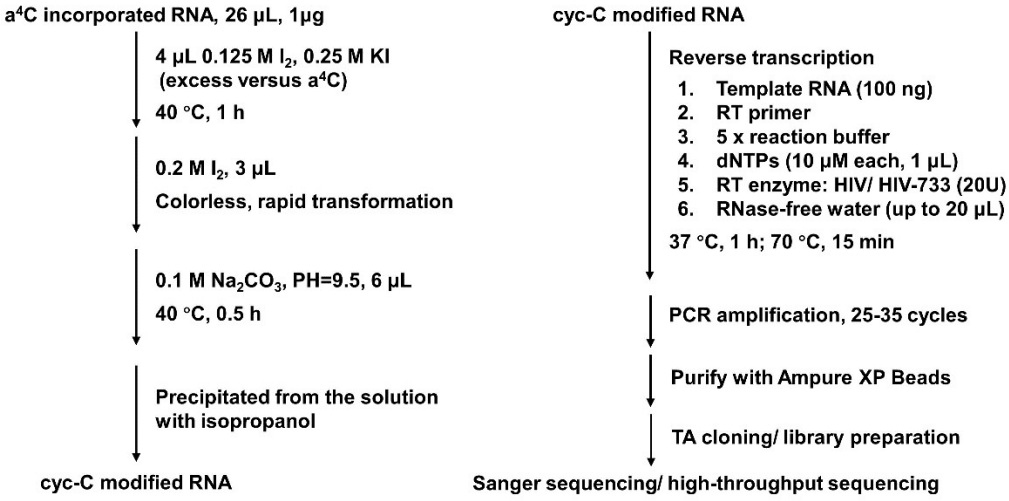


Figure S23. (A) The cumulative mutation site numbers (C-to-T) around RefSeq TSS by 50 bp-windows in a^{4}C-ChRO-seq using HIV 733 enzyme. (B) The cumulative mutation site numbers (C-to-T) of *CAMK1D* in a^{4}C-ChRO-seq around RefSeq TSS by 50 bp-windows.



Scheme S1. Flowchart of a^{4}C chemical sequencing in RNA.

Table S1. RNA probes used for *in vitro* chemical sequencing and enzyme-assisted labelling.IVT: in vitro transcription.

Name	Sequence (5'-3')	Length (nt)	Source	Purification
327nt-a^{4}C-probe	GGGAGATTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTT GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA TATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACG ACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTT CCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTG TATCATATGCCAAGTACGCCCC	327	IVT	Zymo Research RNA-Clean& Concentrator
150nt-12S-mt-rRNA	GGGCTAAGAGTAGAGTGCTTAGTTGAACAGGGCCCTGAAGCGCGTACACACCGCCCGTCA CCCTCCTCAAGTATACTTCAAAGGACATTTAACTAAAACCCCTACGCATTTATATAGAGGA GACAAGTCGTAACATGGTAAGTGTACTGG	150	IVT	Zymo Research RNA-Clean& Concentrator

Table S2. DNA primers used in this work.

Name	Sequence (5'-3')	Purification
327-PCR-F	CGTAATACGACTCACTATAGGGAGATTAGGGTTAGGCGTTTTGCGCTG	PAGE
327-PCR-F	GGGGCGTACTTGGCATATGATACACTTGAT	PAGE
150-PCR-F	TAATACGACTCACTATAGGGCTAAGAGTAGAGTGCTTAGTTGAACAGG	PAGE
150-PCR-F	CCAGTACACTTACCATGTTACGACTTG	PAGE
GAPDH-qF	AGAAGGCTGGGGCTCATTTG	PAGE
GAPDH-qR	AGGGGCCATCCACAGTCTTC	PAGE
rRNA5.8S-qF	TTAGCGGTGGATCACTCGG	PAGE
rRNA5.8S-qR	GCAAGTGCGTTCGAAGTGTC	PAGE
rRNA18S-qF	GCGCGCAAATTACCCACTC	PAGE
rRNA18S-qR	CTCCAATGGATCCTCGTTAAAGG	PAGE
rRNA28S-qF	CCAGGGGAATCCGACTGTTTA	PAGE
rRNA28S-qR	CGTTTACCCGCGCTTCATTG	PAGE