Published online 15 February 2019 *Nucleic Acids Research, 2019, Vol. 47, No. 7* ***3619–3630***
doi: 10.1093/nar/gkz104

Engineering of a thermostable viral polymerase using metagenome-derived diversity for highly sensitive and specific RT-PCR

Ryan C. Heller*, Suhman Chung, Katarzyna Crissy, Kyle Dumas, David Schuster and Thomas W. Schoenfeld

Department of Research and Development, QIAGEN Beverly, 100 Cummings Center, Suite 407J, Beverly, MA 01915, USA

Received November 20, 2018; Revised February 07, 2019; Editorial Decision February 07, 2019; Accepted February 07, 2019

ABSTRACT

Reverse transcription is an essential initial step in the analysis of RNA for most PCR-based amplification and detection methods. Despite advancements in these technologies, efficient conversion of RNAs that form stable secondary structures and double-stranded RNA targets remains challenging as retroviral-derived reverse transcriptases are often not sufficiently thermostable to catalyze synthesis at temperatures high enough to completely relax these structures. Here we describe the engineering and improvement of a thermostable viral family A polymerase with inherent reverse transcriptase activity for use in RT-PCR. Using the 3173 PyroPhage polymerase, previously identified from hot spring metagenomic sampling, and additional thermostable orthologs as a source of natural diversity, we used gene shuffling for library generation and screened for novel variants that retain high thermostability and display elevated reverse transcriptase activity. We then created a fusion enzyme between a high-performing variant polymerase and the 5′→3′ nuclease domain of Taq DNA polymerase that provided compatibility with probe-based detection chemistries and enabled highly sensitive detection of structured RNA targets. This technology enables a flexible single-enzyme RT-PCR system that has several advantages compared with standard heat-labile reverse transcription methods.

INTRODUCTION

Reverse transcription PCR (RT-PCR) and variations like quantitative RT-PCR (RT-qPCR) (1–3) and reverse transcription droplet digital PCR (RT-ddPCR) (4) are indispensable tools that are widely used for monitoring expression levels of disease-specific mRNA biomarkers, profiling noncoding RNA, and detecting pathogens, especially RNA viruses, at high specificities and sensitivities from biological and environmental samples. A critical step in these technologies is the initial enzymatic synthesis of a cDNA molecule from the target RNA template. This step, typically catalyzed by a retroviral reverse transcriptase (RT) derived from either the Moloney murine leukemia virus (MMLV) or the avian myeloblastosis virus (AMV), is followed by PCR amplification of the cDNA using Taq DNA polymerase (Pol) (1). While the combination of a mesostable RT and a thermostable Pol generally provides an effective means for molecular detection of RNA, the lower temperature reaction conditions required for the retroviral reverse transcriptase (5) can limit its effectiveness in detecting certain sequences. This is especially true if the lower temperatures promote formation of unfavorable secondary structures such as hairpins, stem loops and G quadruplexes that can block primer binding and impede nascent strand synthesis on the RNA template (6). For highly structured RNA targets, especially common in viral genomes, it would be advantageous to perform the cDNA synthesis reactions at higher temperatures so that RNA secondary structures are destabilized and non-specific primer binding is minimized. The use of the two-enzyme systems has other inherent disadvantages. For example, the separate reverse transcription step delays the results by up to 30 min, low temperature DNA-dependent DNA polymerase activity of the reverse transcriptase can catalyze non-specific primer extension products that compete for and limit sensitivity of RT-PCR, and there is evidence that the RT and the Pol can compete for the template, resulting in efficiency anomalies especially apparent when detecting limiting amounts of target (7).

Over the past decades, extensive efforts have focused on improving the thermal stability and performance of the retroviral RTs (8–13). This work has resulted in impressive improvements in one-step RT-PCR systems based on incor-

*To whom correspondence should be addressed. Tel: +1 978 867 5743; Fax: +1 978 867 5724; Email: Ryan.Heller@qiagen.com

porating engineered mesophilic RTs in two-enzyme systems. However, limitations remain in the available enzyme systems impacting RT-PCR amplification (7,14) and current approaches to improving the retroviral RTs have not fully addressed these needs.

A single-enzyme alternative to the two-enzyme blends that would address some of these limitations depends on a DNA polymerase with both RT activity to generate cDNA templates from RNA targets and sufficient thermostability to survive repeated rounds of thermocycling to near boiling temperatures. Such a single-enzyme RT-PCR mix should provide advantages over the two-enzyme systems in performance on structured targets, simplicity of formulation, reduced time to result, improved efficiency, and would allow the generation of a single neutralizing antibody that blocks both RT and DNA polymerase activity in a hot-start format.

Various approaches have been explored to generate enzymes with a suitable combination of RT activity and thermostability. Certain bacterial family A DNA polymerases derived from thermophilic *Thermus* and *Thermatoga* species can be induced to reverse transcribe by mutagenesis of the enzyme and modification of the reaction buffer to include manganese (15–18). While Taq DNA polymerase possesses detectable innate RT activity (19), it is insufficient for robust RT PCR-based detection. Directed evolution using error-prone PCR mutagenesis and structural analysis have identified variants with markedly increased RT activity and have provided insight into mechanisms of template discrimination (20–22). In addition, intron-associated RTs from thermophilic bacterial strains have been explored as RT-PCR enzymes (23,24). Alternatively, mutagenesis of archaeal family B DNA polymerases has resulted in functional proofreading thermostable RTs (25). Despite successful engineering and directed evolution programs resulting in effective RT activity, the two-enzyme systems generally have been the methods of choice for most RT-PCR applications.

Viral metagenomics studies provide a means of accessing an especially fertile source of diversity for both discovery and *in vitro* evolution of a variety of enzymes, including those with thermostable RT activity. Viruses are believed to outnumber cellular life forms by at least a factor of ten (26), with a total nucleic acid diversity that vastly exceeds that of cellular life (27,28). From screens of thermophilic viral metagenomes from hot springs, a thermostable family A-type Pol known as PyroPhage Pol was previously identified and was shown to possess inherent reverse transcriptase activity without a manganese requirement, allowing it to be suitable for single-enzyme RT-PCR and for isothermal amplification and detection applications from RNA templates (29–32). However, impediments in consistency and overall performance limit its utility as a detection reagent and it has not found widespread use. We describe a program to improve the phage polymerase as an RT-PCR reagent using natural viral diversity as a starting point for directed enzyme evolution. This work was followed by additional protein engineering that further refined the functionality of the enzyme system and provided a single enzyme magnesium-based RT-PCR chemistry that is competitive with two-enzyme systems and has additional advantages in ease of formulation, time to result and displays flexibility with nucleic acid targets and detection chemistry without requiring modification of the reaction conditions.

MATERIALS AND METHODS

Reagents and equipment

DNA assembly reactions, including for the DNA shuffling method, were performed using the repliQa HiFi Assembly mix (Quantabio) according to the manufacturer's recommendation. Restriction enzymes were purchased from Thermo Scientific. Oligonucleotides were synthesized by IDT and the sequences are listed in Supplementary Table S1. A StepOnePlus (Applied Biosystems) or a QuantStudio 7 Flex (Applied Biosystems) was used for RT-qPCR and fluorescence measurements for activity tests. A 2720 thermal cycler (Applied Biosystems) was used for end-point PCR and microtube incubations. An ÄKTA Pure system (GE) was used for column chromatography.

Construction of viral polymerases, Magma fusion, and Magma exo mutants

Viral polymerases. The DNA sequences of previously identified viral metagenomic polymerases (33) were obtained from GenBank (AFN99405.1, AGL03982, AGL03983, AGL03985 AGL03987 and AGL03990), altered to eliminate 3′→5′ exonuclease activity as previously described (31), codon optimized, and synthesized by Synthetic Genomics. The DNA constructs were then assembled with a tac promoter-based expression vector and used to transform *Escherichia coli* strain BL21 (EMD Millipore).

Magma fusion. The 5′→3′ nuclease domain of the Taq polymerase gene (GenBank BAA06775) was PCR amplified using Veraseq 2.0 polymerase (QIAGEN) and primers Taqexo-F and Taqexo-R. The M160 gene was PCR amplified using primers M160-F and M160-R. The two fragments were DpnI treated to eliminate background plasmid DNA, purified using a QIAquick PCR purification kit (QIAGEN) and assembled with an NdeI enzyme-linearized M160 expression plasmid to create the Magma DNA polymerase.

Magma exonuclease mutants. Five PCR fragments were amplified from the Magma expression plasmid using the following primer combinations: Fragment 1, P6-SpeI-F/MagmaA339E-Rev; Fragment 2, MagmaA339E-F/Magma-Pml-R; Fragment 3, P6-SpeI-F/MagmaG46D-R; Fragment 4, MagmaG46D-F/MagmaA339E-R; Fragment 5, MagmaG46D-F/Magma-Pml-R. PCR fragments were DpnI treated, purified, and assembled with SpeI/PmlI restriction enzyme-digested Magma expression plasmid in the following combinations: M401, Fragments 1 and 2; M402, Fragments 2, 3 and 4; M403, Fragments 3 and 5.

Polymerase expression and purification

Cells transformed with expression plasmids were grown in TB medium at 37°C in the presence of kanamycin to an

OD_{600} of 2.0–2.5 and induced with 0.2 mM IPTG. At 3–4 h post-induction growth, cells were harvested, resuspended in 50 mM Tris–HCl, pH 8.0, 350 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 10% glycerol, 1 mM PMSF and lysed by two passes through a French press pressure cell at 700 psi. The cell suspension was heat treated at 65°C for 20 min to precipitate host proteins, then centrifuged to obtain clarified lysate. To precipitate nucleic acids, polyethyleneimine was added to 0.2% (v/v) and insoluble material was removed by centrifugation. Soluble protein was then precipitated by addition of saturated ammonium sulfate to 60%. Protein was resolubilized and purified by HiTrap heparin HP column, HiTrap HP SP column and HiTrap Q FF column (GE) and dialyzed against storage buffer (10 mM Tris–HCl, pH 7.5, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 50% glycerol). Concentrations were determined by optical absorbance at 280 nm.

Gene shuffling

Equimolar quantities of 3173, 967 and 1608 expression plasmids were combined and used as templates in nine individual PCR reactions using VeraSeq 2.0 polymerase (QIAGEN) according to manufacturer instructions with the following primer pairs: fragment A, P6-Xba-For/JxnA-Rev; fragment B, JxnA-For/JxnB-Rev; fragment C, JxnB-For/JxnC-Rev; fragment D, JxnC-For/JxnD-Rev; fragment E, JxnD-For/JxnE-Rev; fragment F, JxnE-For/JxnF-Rev; fragment G, JxnF-For/JxnG-Rev; fragment H, JxnG-For/JxnH-Rev; fragment I, JxnH-For/P6-Sal-Rev. Each fragment, containing from 21 to 26 bp of overlap between adjacent fragments, was treated with DpnI to remove residual parental plasmid and purified. DNA fragments A-C, D–F and G–I were each assembled into intermediately sized linear fragments, amplified by PCR and purified. The intermediate fragments were assembled with each other and with XbaI/SalI enzyme-linearized expression plasmid, then were directly used to transform E. cloni 10G (Lucigen) by electroporation. Eight individual clones were sequence-confirmed for the presence of independent, random inserts.

Lysate screening for RT-qPCR activity

Single colonies from the gene-shuffled library were picked for inoculation of individual 2 ml wells of a sterile 96-well plate (Nunc) filled with Overnight Express autoinduction TB-kanamycin media (EMD Millipore) and incubated at 37°C for 18–20 h with shaking at 450 rpm. Cells were pelleted and lysed in 50 mM Tris–HCl, pH 8.0, 100 mM NaCl, 0.5 mM EDTA, 0.1% Triton X-100, 5% glycerol, and 180 U/μl Ready-Lyse Lysozyme (Epicentre) for 15 min at room temperature. Lysates were mixed with an equal volume of DNase I buffer (0.1 U/μl DNase I (Roche) in 50 mM Tris–HCl, pH 8.0, 12 mM $MgCl_2$, 2 mM $CaCl_2$) and incubated at 37°C for 10 min, 65°C for 20 min and 75°C for 10 min. Plates were centrifuged at 4500 rpm for 10 min, and clarified lysate was tested in RT-qPCR reactions (20 μl) containing 1X ZipScript buffer (QIAGEN), 0.225× EvaGreen dye (Biotium), 1 ng MS2 phage RNA (Roche), 0.3 μM MS2–243-F primer, 0.3 μM MS2–160-R primer, and 2 μl crude lysate. Reactions were cycled at 94°C for 30 s; followed by 40 cycles of 94°C for 15 s and 60°C for 1 min with fluorescence data collection during the anneal/extension step. Cq measurements were calculated for ∼400 library variants in total.

Biochemical characterization of purified polymerases

Differential scanning fluorimetry. Reactions (25 μl) containing 50 mM Tris–HCl, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 5× SYPRO Orange (Invitrogen), and 2 μg enzyme were incubated in a real-time thermal cycler at 35°C for 1 min, then increased to 100°C at a ramp rate of 0.2°C/s. Fluorescence readings were taken at each interval, derivative values were plotted, and the peak minima corresponding to the polymerase melting temperature was reported.

DNA polymerase activity. Polymerase activity was tested in a buffer containing 20 mM Tris–HCl, pH 8.8, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$ and 0.1% Triton X-100. The M13-based reactions (20 μl) also contained 200 μM each dNTPs, 1X SYBR Green I (Thermo Fisher), 7.5 μg/ml M13mp18 DNA, 0.25 mM each of a mixture of three primers 24–33 nt in size, and 0.03–1 ng of enzyme. Reactions were incubated at the indicated temperature, fluorescence was measured every 15 s, and fluorescence initial slope values were calculated and compared. Calf thymus DNA-based reactions (50 μl) also contained 4 μg activated calf thymus DNA, 100 μM dNTPs, 7.5 μCi/ml ^{3}H-dTTP and 0.8–25 ng enzyme, were incubated at the indicated temperatures, and the TCA-insoluble radioactive counts were measured using a scintillation counter. The slopes of the initial rates of nucleotide incorporation were determined and compared. Activity measurements were based on four to seven datapoints and the standard error was determined.

Reverse transcriptase activity. Reactions (20 μl) containing 50 mM Tris–HCl, pH 8.3, 75 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.01% Tween-20, 2% trehalose, 0.4× EvaGreen dye (Biotium), 0.8 mM dTTP, 0.01 ug/μl Poly(A), 0.1 μM oligo$(dT)_{20}$ primer and 0–20 ng polymerase were incubated at the indicated temperatures and fluorescence was measured every 15 s. For the equivalent DNA polymerase activity measurements, the poly(A) RNA was replaced with poly(dA) DNA template. For the high temperature RT thermal profile, the Poly(A) was replaced with Poly(C) and the oligo$(dT)_{20}$ was replaced with 5′-GGGIIGGGIIGGGIIGGGIIG-3′. The initial slopes of fluorescence curves were calculated and compared for each polymerase. Activity measurements were based on four to seven datapoints and the standard error was determined.

Single-stranded exonuclease activity. Exonuclease reactions (50 μl) containing 50 mM Tris–HCl, pH 8.7, 75 mM KCl, 4 mM $MgCl_2$, 0.3 mM dNTPs, 0.04 mg/ml human serum albumin, 0.2 M trehalose, 50 nM ^{3}H-dTTP end-labeled single-stranded 59-mer oligonucleotide and 0.39–50 ng polymerase were incubated at 37°C for 60 min. Reactions were stopped by addition of salmon sperm carrier DNA and TCA-soluble radioactive counts were measured.

Polymerase fidelity. Nucleotide incorporation fidelity was measured using a modified *lacI*^{q} reversion assay as previously described (31), except that the sequences containing the *lacI*^{q} and *kan* genes were amplified in triplicate 50 μl reactions containing 1X ZipScript buffer, 0.25 μM primers Aat-Fidelity-F/Aat-Fidelity-R, 10 ng pENZK6 plasmid and 2 U KOD polymerase (EMD Millipore) or a quantity of a Magma variant with equivalent activity. The cycling conditions were 94°C for 30 s; followed by 25 cycles of 94°C for 5 s, 68°C for 30 s and 72°C for 2.5 min. Amplicons were then assembled with AatII-linearized pUC19 plasmid and the resulting DNA construct was used to transform 10G cells by electroporation. Cells were plated onto LB media containing kanamycin, ampicillin, and X-gal, grown at 37°C and the numbers of blue and white colonies were determined. 3000–5000 colonies were counted from each of the three independent PCR replicates and the error rate was determined according to the following formula: ER = mf/(bp X d), where mf is the mutation frequency (blue colonies divided by total colonies), bp is the effective target size (349 for *lacI*^{q}), and d is the number of PCR doublings ($\log_2$ of PCR fold amplification). The results were averaged and reported with the standard deviation.

Electrophoretic mobility shift assay. DNA binding assays were performed as previously described (34). Reactions (30 μl) containing 20 mM Tris–HCl, pH 8.8, 10 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.05 mg/ml BSA, 10% glycerol and 5 nM unlabeled DNA substrate were incubated with polymerase at 37°C for 10 min, then fractionated by 3–12% polyacrylamide gel electrophoresis. Gels were stained with 2X SYBR Gold (Invitrogen) and the band intensities were quantified.

RT-qPCR assays. Unless otherwise indicated, all amplifications were performed using a single polymerase enzyme without a pre-incubation step prior to thermal cycling. Reactions (20 μl) contained 1X ZipScript buffer, 0.3 μM forward and reverse primers, and either 0.225X EvaGreen dye or 0.2 μM probe oligonucleotide as indicated. Reactions comparing polymerase orthologs contained 200 ng polymerase and were cycled at 94°C for 15 s; followed by 40 cycles of 94°C for 10 s and 65°C for 30 s. Reactions comparing M160 with Magma contained 300 ng polymerase and were thermocycled at 94°C for 30 s; followed by 40 cycles of 94°C for 3 s and 64°C for 60 s. Reactions comparing dye, probe, DNA and RNA contained 100 ng Magma and were thermocycled at 94°C for 30 s; followed by 40 cycles of 94°C for 3 s and 72°C for 60 s. Reactions with a Taq Pol/MMLV RNAse H- enzyme mixture contained 1X ZipScript enzyme (QIAGEN) and were additionally pre-incubated at 50°C for 15 min prior to cycling. Amplification reactions from complex template mixes contained 100 ng Magma and were thermocycled at 94°C for 30 s; followed by 40 cycles of 94°C for 3 s and 64°C for 30 s. Reactions with MS2 RNA template all contained the MS2–160-R primer and a forward primer with a name indicating the amplicon size (Supplementary Table S1). All qPCR and RT-qPCR reactions were performed in triplicate unless otherwise indicated and average Cq values plotted with error bars indicating the standard deviation. The DNA construct corresponding to the MS2 genome was synthesized by IDT and total human reference RNA was purchased from Agilent.

To prepare the microbial cell lysates, *Vibrio natriegens* cells (Synthetic Genomics) and *Bacillus megaterium* cells (environmental isolate) were grown to early log phase in 1 ml 2× YT media. Cell densities were quantified by plating serial dilutions of the cultures to LB-agar and were grown overnight at 30°C for colony counting. Cells were washed in PBS, resuspended in 200 μl of a buffer containing 10 mM Tris–HCl, pH 7.5, 0.5 mM EDTA, 100 mM NaCl, 0.1% Triton X-100 and lysed by addition of 1 μl of Ready-Lyse Lysozyme and incubation for 15 min at room temperature. The lysate was briefly vortexed and serial dilutions were made using 0.01% Tween-20. Finally, 2 μl of this lysate was used directly in RT-qPCR reactions as described above.

RESULTS

Characterization of metagenomic thermostable polymerase orthologs

In addition to the PyroPhage Pol (referred to as 3173 in the published work and in this report), a clade including viral family A polymerase sequences of high sequence similarity were identified in metagenomic surveys of geochemically similar, but geographically dispersed hot springs (30); however, their biochemical characterization has not been reported. Six chosen orthologs, including 3173, had an overall sequence conservation (Figure 1A) that ranged from 93% amino acid identity (AAID) for the most similar polymerases (3173 and 1608, both isolated from the same hot spring over a 4-year period), to 45% AAID for the most divergent polymerases (3173 and 488, isolated from hot springs separated by ca. 1200 km) (Table 1). Similarity to 3173 Pol suggests that the orthologs might also have utility in single-enzyme RT-PCR. Based on sequence alignment (not shown), these polymerases all appear to have active 3′→5′ proofreading activity, as was seen previously with the 3173 Pol (31). To avoid potential degradation of primers and template during the screens and the characterization, the presumptive exonuclease activity was inactivated in all the sequence variants of this study by a point mutation incorporated during gene synthesis as described (31). The encoded proteins were expressed in *E. coli* and purified for biochemical characterization.

Enzyme stability is a critical property that would enable a polymerase to survive and maintain activity through the repeated rounds of high-temperature incubation that take place in the PCR reaction. To measure and compare polymerase thermostability, equivalent quantities of the enzyme preparations were tested by differential scanning fluorimetry (DSF) (35). The peak minimum of the fluorescence derivative plots (Figure 1B) reflects the temperature (Tm) at which half of the molecules exist in an unfolded state and was used as a comparative metric for the orthologs. The six orthologs showed wide variability in Tm values (Figure 1B and Table 1) with the highest thermostability seen in 3173 at 92.2°C and the closely related 1608 Pol at 90.2°C. The 488 Pol displayed the lowest Tm of the group of polymerases at 64.1°C.

Effective RT-PCR requires efficient high temperature synthesis on both DNA and RNA templates. DNA-

Table 1. Biochemical characterization of 3173 polymerase orthologs

Enzyme	Amino acid identity (%)	Relative Pol activity, 42°C	Relative RT activity, 42°C	Relative RT activity, 50°C	Melting temperature,°C
3173	100	1 ± 0.08	1 ± 0.02	1 ± 0.02	92.2
4B9	90.3	0.01 ± 0.002	0.42 ± 0.01	0.12 ± 0.004	81.9
1608	93.0	0.08 ± 0.01	0.49 ± 0.01	0.46 ± 0.02	90.2
488	45.0	0.46 ± 0.06	1.51 ± 0.07	1.98 ± 0.05	64.1
1440	49.6	0.31 ± 0.04	0.63 ± 0.02	0.35 ± 0.01	86.8
967	83.2	0.49 ± 0.07	1.64 ± 0.11	3.73 ± 0.10	87.2

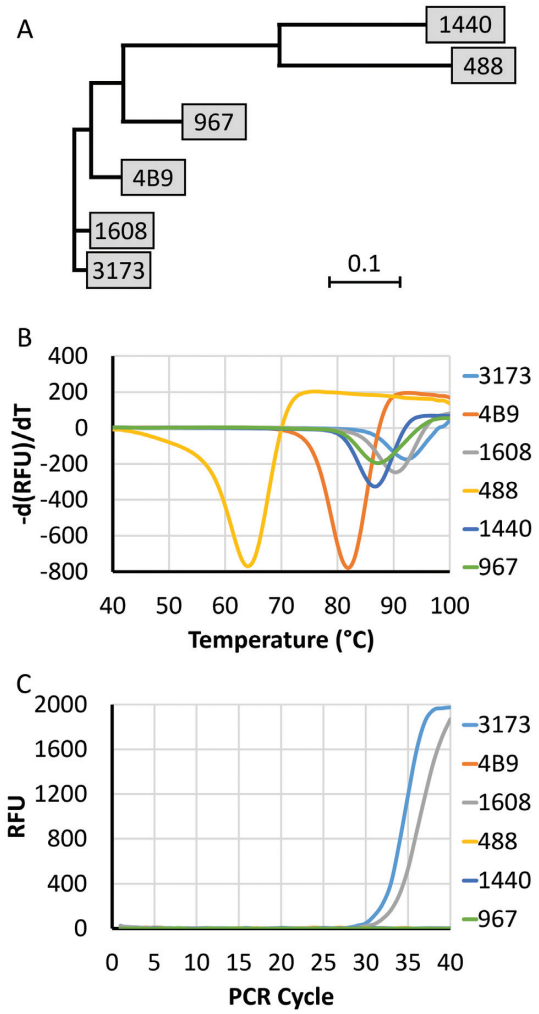

Figure 1. Analysis and characterization of polymerase orthologs. (**A**) Phylogenetic tree of metagenomic family A polymerase amino acid sequences. Scale bar represents the uncorrected pairwise distance. (**B**) Differential scanning fluorimetry of purified polymerase orthologs. –d(RFU)/dT indicates the derivative fluorescence values. (**C**) Dye-based RT-qPCR reactions using 10 pg total human RNA were catalyzed by the indicated polymerase. Amplicon was 89 bp in length and target was GAPDH mRNA sequence.

dependent synthesis was tested using a primed single-stranded M13-based assay in the presence of a dsDNA binding fluorescent dye. Without prior knowledge of thermal activity profile and considering the low thermostability of the 488 Pol, a relatively low temperature was used for all orthologs to establish a baseline activity for comparison. Initial reaction rates over an enzyme dilution series were measured at 42°C under steady-state conditions of substrate excess and values were compared and normalized to the 3173 Pol (Table 1). Whereas all the orthologs were active and displayed some level of DNA polymerase activity, the specific activity of the 3173 Pol was the highest; 488, 1440 and 967 Pols were moderately active; while 1608 and 4B9 Pols were markedly lower at 8% and 1% activity, respectively, compared with 3173. Relative RT activities were tested by measuring initial reaction rates on a primed poly(A) RNA template at 42°C and 50°C using dsDNA-specific fluorescent dye as a reporter (Table 1). Two orthologs, 488 and 967 Pols, exhibited higher RT activities than 3173 Pol and, in both cases, the relative activities were higher at 50°C than at 42°C, suggesting an elevated RT thermal optimum. 967 Pol exhibited the highest RT specific activity, with RT levels that were as much as 4-fold higher than 3173 Pol.

None of the native orthologs individually showed a suitable combination of thermostability, polymerase, and RT activities to improve RT-PCR compared to the 3173 parent (Figure 1C). Despite elevated RT specific activities, neither 488 or 967 Pols showed sufficient thermostability to support efficient PCR. Only the 3173 and 1608 orthologs were thermostable enough for RT-PCR; however, 1608 Pol showed no improvement over 3173 Pol in RT-qPCR cycle threshold, presumably due to its relatively low RT activity.

Diversity generation by gene shuffling and screening for single-enzyme RT-PCR activity

Gene shuffling was used to generate molecular diversity for the screening of novel polymerase variants that combined the relatively high native thermostability of the 3173 and 1608 Pols with the elevated RT activity of the 967 Pol. Like versions of this method that use DNase I to fragment a series of homologous gene family members (36), random reassembly results in diverse library of sequences containing individual segments originating from different starting genes. However, to create the diversity in this shuffled library, we used a pool of overlapping DNA sequences generated by PCR, then mixed and reassembled these fragments randomly at pre-defined crossover points. The sequences of the three genes were aligned and eight approximately equally spaced crossover sites were chosen that contained regions conserved over at least seven codons, i.e. 21 nucleotides. Using primers targeting these crossover sites, nine individual DNA segments were PCR amplified randomly from a pool of the three DNA templates. In two stages, DNA segments were combined and randomly reassembled into full chimeric sequences using enzymatic isothermal DNA assembly. Using this approach, the library size was

controlled, native gene segment order was preserved and formation of spurious assemblies was minimized. With the gene divided into nine individual DNA segments and three parental possibilities at each position, the theoretical library size was 3^9, or 19 683 variants in total. Notably, sequencing of individual clones (Figure 2) showed that the assembly process generated additional unintended crossover synthesis that resulted in a higher level of diversity than predicted.

The variants were functionally screened using Cq in RT-PCR as a metric. From *E. coli* cells transformed with the shuffled library, crude lysates were heat-treated to inactivate native host enzymatic activities. The lysates were DNase I treated to reduce viscosity from high molecular weight genomic DNA. The treated lysates were used directly in RT-PCR detection of MS2 viral RNA under time-limiting extension conditions. An intercalating fluorescent dye reporter was used to monitor synthesis in real time. Under these conditions, 3173 parental polymerase showed qPCR amplification efficiency of nearly 100% (not shown); therefore, the difference in Cq for detection of MS2 RNA between variants and the parental enzyme served as a direct measurement of RT efficiency in the conversion to cDNA prior to amplification. Approximately 400 variants were screened. Most variants showed no signal in this screen; however, 38 variants (∼10%) produced detectable signal. Of these, eight (∼2% overall) showed improvements in Cq number compared to the parental polymerases (Figure 2). The sequence schematic of the six highest performing variants are shown (Figure 2) and the amino acid sequences were aligned (Supplementary Figure S1). While most of the sequence compositions are varied with respect to the parental sequences and no pattern is discernible, there is a region from amino acid 400–472 near the conserved motif B (37) from the 967 parent that is seen in each of the variants (Supplementary Figure S2). Though it may have been possible to obtain an even further improved RT activity by screening more variants, from the set of 400 variants analysed the highest performing variant, M160, had a Cq value that was 5.2 cycles lower than for 3173, suggesting a significant improvement in RT function. This variant was used for further development.

Construction of Magma DNA polymerase by nuclease domain fusion

DNA hydrolysis probes, i.e. TaqMan® probes (Roche), are widely used in real-time RT-PCR to quantitatively detect specific DNA or RNA targets (1,38) and provide advantages in specificity, sensitivity and multiplex detection compared with dye-based methods. The basis of this approach is synthesis-dependent hydrolysis of a probe at each cycle by 5′→3′ nuclease activity to separate the fluorophore from its quencher and generate a fluorescent signal proportional to the amount of target in the sample. Typically, the nuclease activity is provided by a structural domain of Taq polymerase (39,40). Many PolA-type DNA polymerases, including the 3173 Pol and its variants, lack this nuclease domain and are inherently unable to support use of hydrolysis probes. To generate an enzyme construct compatible with hydrolysis probes, a clone was developed to express M160 with an amino terminal Taq 5′→3′ nuclease domain fused via a short, flexible linker of eight residues to the Pol domain (Figure 3A). After purification of this fusion molecule, hereafter termed 'Magma' DNA polymerase, we found that it showed a specific activity ∼3.5-fold higher than the parent M160 molecule (Figure 3A). We observed that the RT activity was also improved relative to the M160 (Figure 3B), which itself was five to six-fold higher than the original 3173 parent (Figure 3B). In addition, whereas the parental 3173 Pol displayed an ∼4-fold preference for a DNA template under otherwise identical assay conditions, the M160 variant and Magma showed significantly less activity difference between an RNA and DNA template (Figure 3B). The activity improvements observed with Magma compared with M160 may be analogous with Taq polymerase, where additional interactions between the nuclease domain and the DNA template appears to increase template affinity and improve processivity compared with the N-terminal nuclease deletion (41–43). Supporting this hypothesis, we observed a 12-fold increase in binding affinity to primed-template DNA with Magma (K_d = 78.9 nM) compared with the M160 polymerase (K_d = 946 nM) lacking the nuclease domain (Supplementary Figure S3).

While not essential for RT-PCR-based nucleic acid detection, high nucleotide incorporation fidelity of the polymerase would be beneficial for preparative applications, e.g. DNA/cDNA cloning and next-generation sequencing methods, where sequence accuracy is important. As noted above, Magma had its proofreading activity eliminated by mutagenesis. The error rate of Magma on a DNA template, measured using a standard blue-white screen of sequence errors in PCR-amplified *lacI*^{q} repressor was determined to be 1.91×10^{-4} (Table 2). In contrast, in preparations of altered versions of the Magma enzyme in which the 3′→5′ proofreading nuclease activity was reactivated with an A339E reversion, the measured error rate was nearly two orders of magnitude lower, similar to the error rate measured for KOD polymerase (Table 2, M401), though it is not clear whether these nuclease-active versions would show improved fidelity on an RNA template.

To discern whether the RT-PCR performance improvement of Magma compared to M160 was due to nuclease activity or simply the presence of the nuclease domain providing enhanced template binding affinity, two additional constructs were generated. In the first of these, M402, the 5′→3′ nuclease domain was present but inactivated by the G46D mutation. In the second, M403, both the 5′→3′ and the 3′→5′ activities were mutated. Elimination of the 5′→3′ nuclease did not have a measurable impact on the RT activity, regardless of the associated 3′→5′ exonuclease activity (Table 2), suggesting the improvement of RT-PCR function was dependent on biochemical attributes other than nucleolytic activity, perhaps modification of the binding affinity provided by the domain. In addition, the presence or absence of a 5′→3′ nuclease activity did not substantially affect fidelity.

In further investigation of biochemical properties, the DSF-determined melting temperature of the M160 variant at 92.3°C was determined to be very close to the 3173 Pol parent enzyme (Figure 4A and Table 1), suggesting that the numerous amino acid changes introduced by the gene shuffling process did not substantially reduce the enzyme stabil-

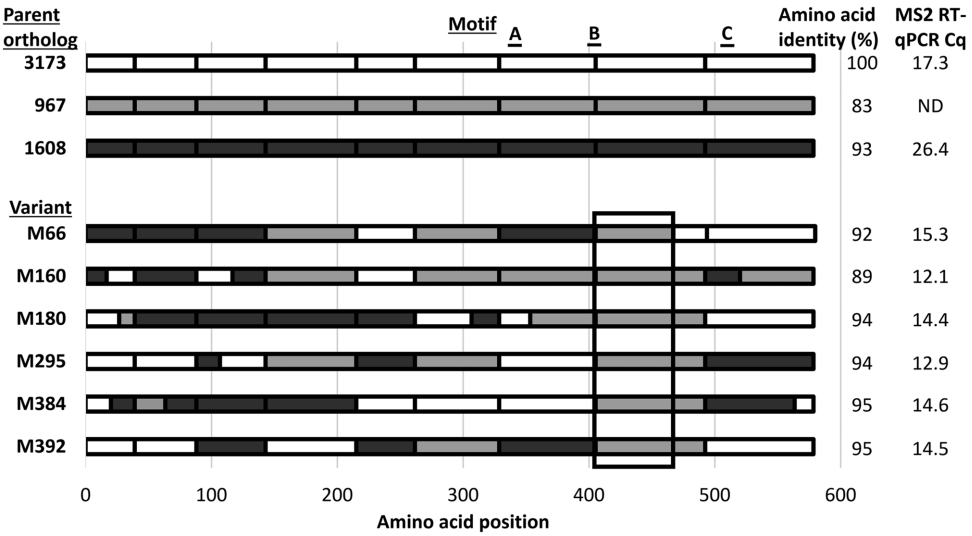

Figure 2. Gene shuffling and screening. Amino acid sequence schematic of the parent polymerases and the variants that displayed improvements in RT-qPCR amplification. Positions of the conserved motif A, B, and C motifs are indicated above the alignment. Amino acid identity values indicated are relative to the 3173 Pol parent. ND, no amplification detected.

Table 2. Magma nuclease mutations

Enzyme	Residue 46	Residue 339	5′→3′ exo	3′→5′ exo	Relative pol activity, 72°C	Relative RT activity, 50°C	Relative ssExo activity[a]	Error rate[b]	Relative error rate[c]
Magma	G	A	+	–	1 ± 0.01	1 ± 0.08	Not detected	$1.91 \times 10^{-4} \pm 0.196$	75.2 ± 9.9
M401	G	E	+	+	0.91 ± 0.04	1.02 ± 0.11	0.76 ± 0.13	$2.22 \times 10^{-6} \pm 0.02$	0.87 ± 0.07
M402	D	E	–	+	1.05 ± 0.03	0.98 ± 0.12	0.89 ± 0.17	$2.49 \times 10^{-6} \pm 0.33$	0.98 ± 0.15
M403	D	A	–	–	0.72 ± 0.02	1.03 ± 0.11	Not tested	Not tested	Not tested

[a]Single-stranded exonuclease activity measured relative to Pfu polymerase.
[b]Mutation frequency/bp/duplication.
[c]Polymerase mutation rate measured relative to KOD polymerase.

ity at least in the M160 variant. Additionally, the introduction of the Taq nuclease domain in the Magma fusion polymerase only very slightly affected the measured Tm. Consistent with the high melting temperature of Magma, the thermal activity profile showed peak polymerase activity from 65°C to 80°C using two distinct assay measurements (Figure 4B), which is similar to the thermal profile previously measured for Taq polymerase (44). The shifted peak seen with the primed M13 assay likely reflects the reduced hybridization of the primers at elevated temperature and not the inherent thermostability of the enzyme. In previous measurements of RT activity (Table 1, Figure 3B), a primed poly(A) RNA template was used, however the low melting temperature of the oligo(dT) primers to this template prevented accurate measurement at temperatures higher than ∼60°C. In contrast, use of an oligo(dG/DI)-primed poly(C) RNA template showed efficient RT activity at temperatures up to 85°C (Figure 4B).

Ostensibly, increased DNA polymerase and RT activities and high thermostability should translate into efficient performance in RT-qPCR detection. Indeed, even without a pre-incubation step prior to thermal cycling, Magma displayed significantly improved detection sensitivity and amplification of a large amplicon of MS2 viral RNA template at lower cycle numbers than M160 (Figure 5). The PCR design targeted a 531 bp amplicon, which generates a product significantly larger than that of a typical qPCR assay targeting amplicons of 75–150 bp. The larger amplicon was intentionally chosen to create a more stringent test that would magnify effects on amplification efficiency. In this assay, Magma detected as little as 100 copies of the viral RNA, which is 3 orders of magnitude more sensitive than the M160 variant polymerase. In addition to sensitivity improvements, the 5′→3′ nuclease domain of Magma was effective in probe hydrolysis during qPCR, supporting amplicon detection with equal efficiency and sensitivity compared to amplification reactions using a dye-based reaction chemistry (Figure 6A). The equivalent Cq values indicate both that adequate signal was generated by the nuclease activity and, conversely, that polymerase activity was not inhibited by the fluorescent dye under these conditions.

One of the anticipated benefits to using a single-enzyme thermostable polymerase for RNA detection is the ability to reverse transcribe the RNA at high temperature, relaxing any secondary structure that would impede or block primer extension through the region. Hydrolysis probe-based qPCR reactions were performed with dilutions of either a synthetic double-stranded DNA molecule corresponding to a portion of the MS2 phage genomic RNA sequence (Figure 6B) or with dilutions of single-stranded

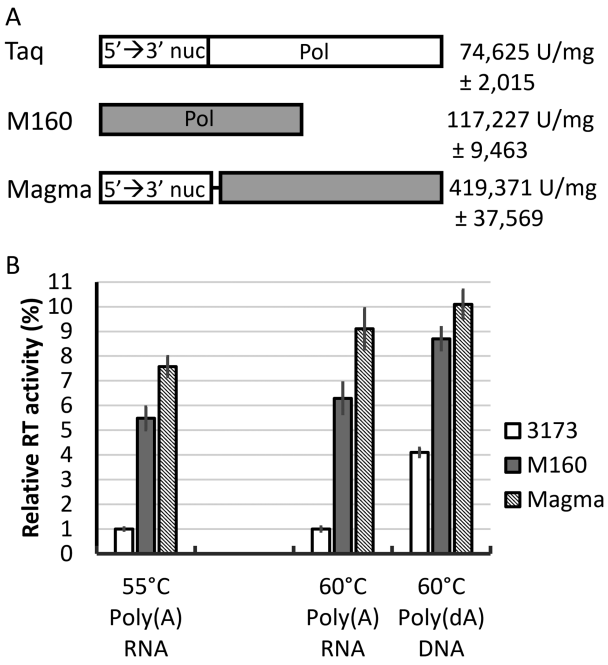

Figure 3. Taq polymerase 5′→3′ nuclease domain fusion. (**A**) Representation of the domain organization of Taq polymerase, the M160 variant, and Magma. The specific activity measured at 70°C is indicated to the right. (**B**) Reverse transcriptase activity using a primed poly(A) RNA template and DNA polymerase activity using a primed poly(dA) DNA template were measured for 3173, M160 and Magma at the indicated temperatures. 5′→3′ nuc, 5′→3′ nuclease domain.

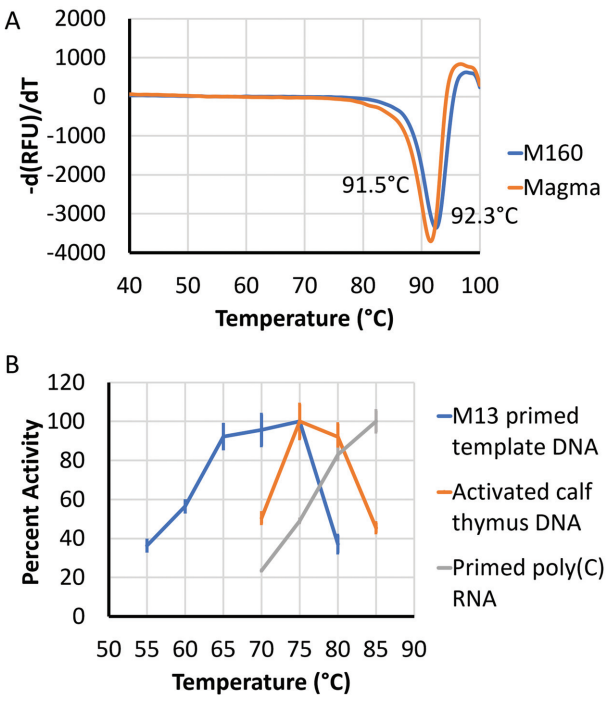

Figure 4. Magma thermal profile. (**A**) Differential scanning fluorimetry of M160 and Magma. (**B**) DNA polymerase and RT activites were measured at the indicated temperatures as described in Materials and Methods. Activities were normalized and plotted relative to the peak value determined for each assay method.

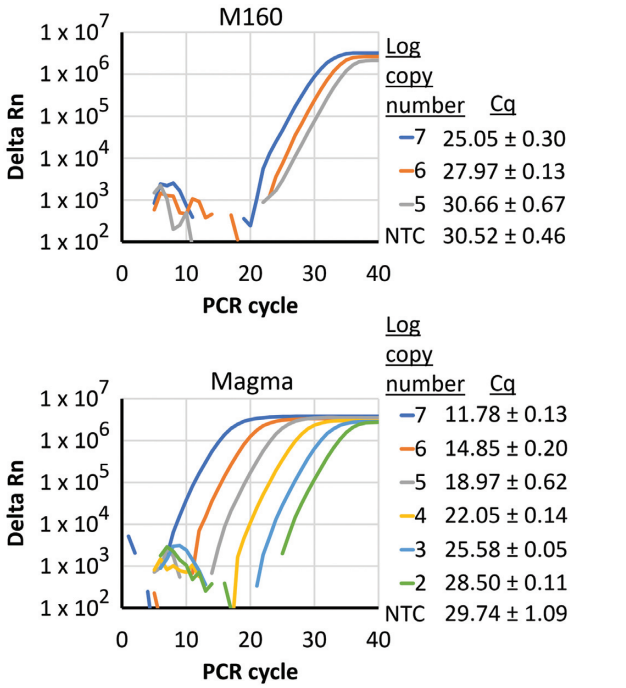

Figure 5. Taq nuclease fusion increases RT-qPCR detection sensitivity. Dye-based RT-qPCR reactions were catalyzed by either M160 or Magma and contained the indicated quantity of MS2 RNA. The amplicon size was 531 bp.

MS2 phage RNA (Figure 6C). Owing to the thermolabile RT of the Taq pol/MMLV RNase H(–) RT enzyme mixture, a pre-PCR incubation step (50°C, 15 min) in the RNA reactions was required in the two enzyme RT-PCR reaction for cDNA conversion. This step was dispensable in single enzyme Magma reactions and provided almost no benefit when tested (Supplementary Figure S4). The Cq values using the Taq pol/MMLV RNase H(–) RT mixture when amplifying this 531 nt segment were ∼5.5 cycles later than with Magma, indicating that Magma was significantly (∼45-fold) more efficient at reverse transcription of the highly structured MS2 RNA genome during the PCR cycling phase (72°C) compared with the MMLV RNase H(–) RT enzyme during the RT step at 50°C. The approximately equal Cq values for the Magma-catalyzed DNA- and RNA-based reactions (compare Figure 6B and C), suggests that RNA is nearly completely converted to cDNA in the first 1 min 72°C cycle of the PCR reaction. In addition to the Taq pol/MMLV RNase H(–) RT mixture, Magma was compared with commercial single enzyme and two-enzyme kits in RT-qPCR, where Magma showed benefits in sensitivity and lower Cq values in the amplification of a smaller, 243 nt segment of MS2 RNA (Supplementary Figure S5).

Magma catalyzes sensitive and specific amplification from complex nucleic acid mixes

Many samples, particularly clinical samples, contain RNA targets within complex mixtures of host nucleic acid or other heterologous sequences. To test the capacity of the Magma enzyme to detect individual mRNA transcripts in a

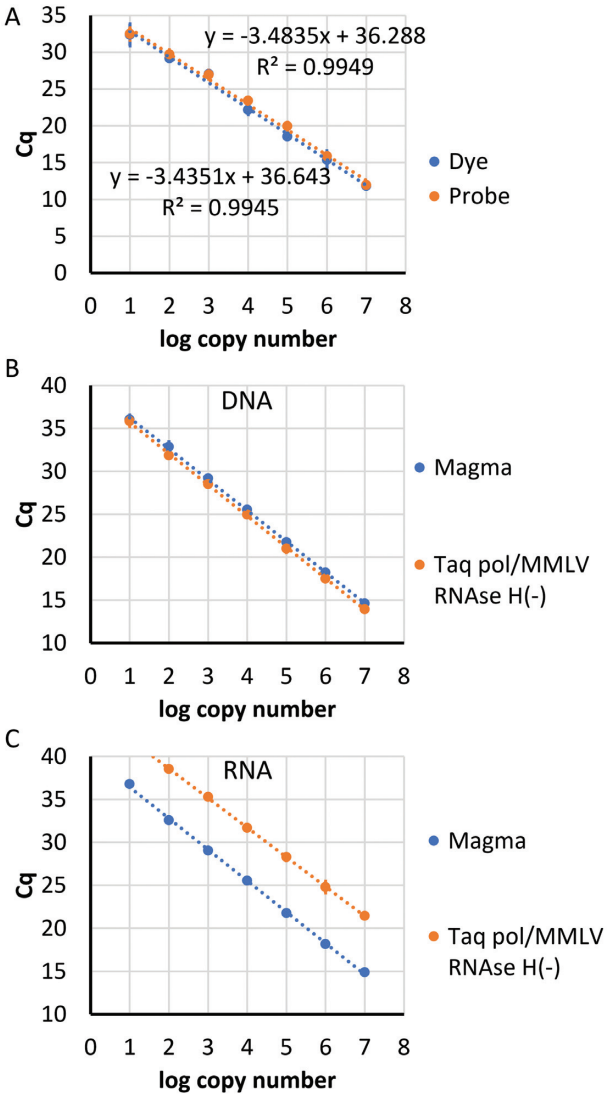

Figure 6. Magma efficiently generates fluorescent signal from structured targets in qPCR using hydrolysis probes. (**A**) Dye- and probe-based RT-qPCR reactions catalyzed by Magma contained the indicated quantity of MS2 RNA. The amplicon size was 362 bp. (B, C) Probe-based RT-qPCR reactions were catalyzed by Magma or a Taq pol/MMLV RNAse H(-) mixture as indicated. Reactions targeted an amplicon 531 bp in size from a synthetic double-stranded DNA molecule corresponding to a portion of the MS2 phage genomic RNA sequence (**B**) or from MS2 viral RNA (**C**). Four replicates were performed for each qPCR reaction and average Cq values were plotted with error bars indicating the standard deviation.

complex mixture containing nontarget RNA, Magma was used to amplify a 145 bp region of the LDHA mRNA from total human RNA using a FAM probe-based RT-qPCR assay (Figure 7). Detection sensitivity was 200 fg, which corresponds to approximately 5 copies as determined by ddPCR quantification (Bio-Rad, data not shown).

Ribosomal RNA (rRNA) tends to be highly structured and its amplification should be a test of the theory that higher temperature RT promotes synthesis through structured target RNA. Moreover, because rRNA is highly abundant in the cell, it represents a rich target for highly sensitive detection of trace quantities of cells. The capacity of

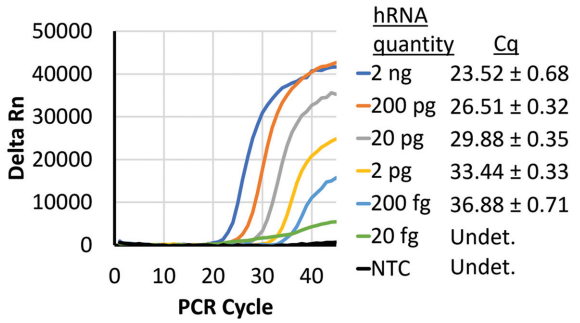

Figure 7. Magma supports specific amplification of mRNA transcript sequences from total human RNA. Magma was used to amplify a 145 bp region of the LDHA mRNA from total human RNA using a FAM probe-based RT-qPCR assay. NTC, no-template control.

Magma to detect rRNA targets directly from cell lysate without processing was tested in duplex RT-qPCR reactions (Figure 8). From mixtures of total cell lysates of a gram negative marine bacterium, *Vibrio natriegens,* and gram positive *Bacillus megaterium* cells, Magma was used in FAM and VIC probe-based RT-qPCR reactions to directly amplify a variable portion of the 16S rRNA. Cross reactivity between the cell types was not observed and successful duplex amplification from lysates of both cell types simultaneously indicates the lack of primer-primer interference between the six oligonucleotides present in the reaction. Detection by RT-PCR was seen over seven logs of lysate dilution, demonstrating efficient detection of the structured rRNA in the presence of cellular nucleic acids and lysate components, and the capacity of this method to reliably detect specific bacteria at limits of detection well below the single cell level.

DISCUSSION

Sensitive, specific amplification of RNA sequences by RT-PCR, RT-qPCR, RT-ddPCR and related methods provides critical means of detecting and monitoring transcription and pathogenesis. Although the available two-enzyme, one step RT-PCR systems are generally robust and reliable, the dependence on multiple enzymes for cDNA synthesis and DNA amplification in RT-PCR has an inherent consequence that reaction conditions are a compromise between those optimal for the two enzymes. This has a negative impact on sensitivity, specificity, amplification efficiency, time-to-result, ease of use and stability in storage. A viral metagenomics study had previously identified a promising enzyme, PyroPhage Pol, with sufficient thermostability and inherent RT activity that enabled it to function in RT-PCR. However, this enzyme has limitations affecting performance including the RT activity that is not competitive with retroviral RTs and the lack of 5′→3′ nuclease activity for compatibility with probe chemistries, making it less suited for routine RNA detection than the two-enzyme systems. To overcome these performance deficiencies, our approach explored the use of additional viral polymerases as a source of molecular diversity for gene shuffling, producing variants that could be screened for improved activity and thermostability, combining the best characteristics of parent enzymes.

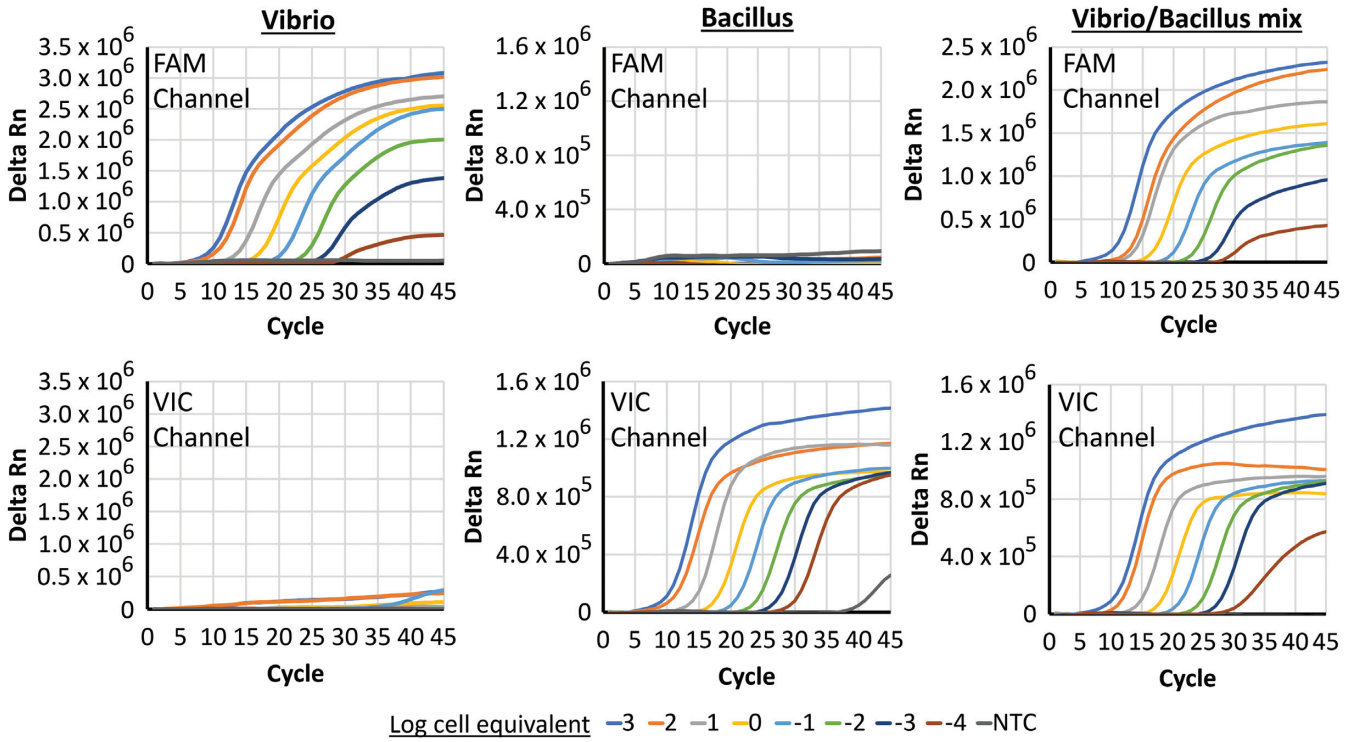

Figure 8. Duplex amplification of rRNA from microbial cell lysate. *Vibrio natrigiens* and *Bacillus megaterium* cells were combined and lysates were prepared and diluted as described in Materials and Methods. RT-qPCR reactions catalyzed by Magma contained a combination of target-specific primers, two of which were probe DNAs conjugated with different fluorophores. Amplification of the 159 bp rRNA sequences were detected in separate spectral channels on the thermocycler.

Variations of this artificial directed evolution approach have been used successfully in a number of industrial applications including vaccine development, biochemical production, metabolic engineering, and enzyme-based bioremediation (45). In addition, thermostable polymerase genes were previously shuffled with homologous genes derived from either family A *Thermus* species (46,47) or with metagenomic PCR fragments isolated from soil samples (34,48), resulting in variant polymerases with novel modified characteristics. An advantage of this randomized approach is that detailed structural and mechanistic information is not required. Since the library is based exclusively on natural diversity, an improved likelihood of generating functional variants compared with randomized approaches to library generation such as error-prone PCR was expected.

The diversity for this study was selected from orthologs of the PyroPhage Pol that were identified from viral metagenomes derived from geographically dispersed hot springs located in the western USA. The six native polymerases, including 3173 Pol, showed significant variability in both thermostability and RT activity compared to the parent. Interestingly, the thermostability does not correlate with the temperature of the hot spring. Nor does the variability in RT activity show an obvious pattern and its biological relevance is unclear.

Though none of these orthologs improved on the original parent enzyme when tested in RT-PCR, they were promising sources of molecular variation. Two orthologs showed either high thermostability or high RT activity, and after shuffling with the 3173 Pol, an enzyme was identified that better combines the optimal combination of the characteristics of the three enzymes. Rather than shuffling the gene fragments at random positions, a controlled approach was used based on PCR amplification of nine essentially equal-sized segments of the respective enzymes and joining at regions of sequence conservation using an enzymatic DNA assembly system. This approach was intended to limit the diversity, controlling the library size and allowing screening of a tractable number of variants. Compared with random fragmentation, preparing a library with pre-defined crossover points allowed for more efficient incorporation of less closely-related 967 sequences into the library because identical overlaps at each junction should allow each DNA segment to have an equal probability of being incorporated at each position. In addition, because of the high efficiency and accuracy of the isothermal assembly process, the incorrect formation of non-functional genes was minimized. With the gene divided into nine individual DNA segments and three parental possibilities at each position, the theoretical library size was approximately 3^9, or ~20 000 variants in total. Some unexpected diversity was introduced within the DNA segments, apparently as artefacts of the PCR amplification step. Since these spurious mutants preserve the overall gene order, they are unlikely to compromise the evolution program and may have a positive effect of moderately increasing the overall diversity in the variant library.

The sequences of the enhanced RT-PCR variants were compared to those of the ineffective variants to identify common sequence features that contribute to the improved RT activity. Although there were numerous positions that varied among the enhanced RT-PCR clones, a region spanning amino acids 400 to 472 (Figure 2) between motifs B and C (49) was fully conserved and apparently derived from 967 (Supplementary Figure S2). In better characterized family A DNA polymerases, this inter-motif region is characterized by two alpha helices, O Helix and P Helix and Beta Sheets 10 and 11, known to be in close contact with the template (50). This proximity to the template is consistent with the improved utilization of the non-natural RNA template. In all of the improved variants, the sequence outside this inter-motif region is more randomly distributed, with a majority of the sequence derived from parents 3173 and 1608 and containing fewer, more scattered conserved residues.

To allow use of hydrolysis probes, the Taq DNA polymerase 5′→3′ nuclease domain was fused to the amino terminus of the evolved M160 enzyme. While effective in providing this function, a fortuitous result of this fusion was increased activity and sensitivity of detection, even with dye-based detection chemistries. This improved performance suggests a secondary effect of this domain in increasing affinity for target RNA, allowing for efficient and complete cDNA synthesis during the initial cycle(s) of anneal/extension of the RT-PCR reactions.

This enzyme construct proved highly effective in detection of both viral RNA and transcription products and is competitive with the most effective two enzyme mixes in terms of sensitivity, specificity and time to result. The inherent thermostability of the single PolA enzyme also provides advantages over the retroviral RT-dependent mixes, which have shown diminished stability after aqueous formulation and lyophilization (data not shown). The ability of this enzyme to seamlessly and efficiently reverse transcribe highly structured RNA immediately following a high temperature (94°C) denaturation step should also prove beneficial for the detection and quantification of dsRNA viruses.

A less common application of RT-PCR is direct detection of rRNA. The high temperature of reverse transcription for minimizing secondary structure and the high abundance of rRNA in bacterial cells provided detection sensitivities that were several orders of magnitude lower than single cell levels. This could potentially be useful for detecting bacterial pathogens in very dilute samples. In addition, despite the focus on improving RT-PCR, the parent enzyme PyroPhage Pol has shown utility in isothermal amplification, specifically LAMP and it seems likely that the enzyme variants from this study will also find utility in this or similar applications involving cDNA synthesis.

In addition to RNA detection, RTs have important uses in preparative applications such as cDNA cloning and construction of RNA-seq libraries where amplification fidelity is critically important. For the purposes of this study, to simplify screening and characterization with unmodified primers, the innate proofreading activity was inactivated by mutagenesis. This activity was readily reconstituted by reversion of the wild-type exonuclease domain, which increased the fidelity by two orders of magnitude on a DNA template. This raises intriguing possibilities as this is a unique enzyme with high RT activity and an efficient proofreading function. The preparative use of the proofreading RT will be explored in detail in a future publication.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

QIAGEN. Funding for open access charge: QIAGEN.
Conflict of interest statement. Ryan C. Heller, Suhman Chung, Katarzyna Crissy, Kyle Dumas, David Schuster and Thomas W. Schoenfeld are employees of QIAGEN.

REFERENCES

1. Bustin,S.A. and Mueller,R. (2005) Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. *Clin. Sci. (Lond.)*, **109**, 365–379.
2. Sanders,R., Bustin,S., Huggett,J. and Mason,D. (2018) Improving the standardization of mRNA measurement by RT-qPCR. *Biomol. Detect Quantif.*, **15**, 13–17.
3. Tan,G.W., Khoo,A.S. and Tan,L.P. (2015) Evaluation of extraction kits and RT-qPCR systems adapted to high-throughput platform for circulating miRNAs. *Sci. Rep.*, **5**, 9430.
4. Sanders,R., Mason,D.J., Foy,C.A. and Huggett,J.F. (2013) Evaluation of digital PCR for absolute RNA quantification. *PLoS One*, **8**, e75296.
5. Yasukawa,K., Nemoto,D. and Inouye,K. (2008) Comparison of the thermal stabilities of reverse transcriptases from avian myeloblastosis virus and Moloney murine leukaemia virus. *J. Biochem.*, **143**, 261–268.
6. Malboeuf,C.M., Isaacs,S.J., Tran,N.H. and Kim,B. (2001) Thermal effects on reverse transcription: improvement of accuracy and processivity in cDNA synthesis. *BioTechniques*, **30**, 1074–1078.
7. Sellner,L.N., Coelen,R.J. and Mackenzie,J.S. (1992) Reverse transcriptase inhibits Taq polymerase activity. *Nucleic Acids Res.*, **20**, 1487–1490.
8. Arezi,B. and Hogrefe,H. (2009) Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucleic Acids Res.*, **37**, 473–481.
9. Baranauskas,A., Paliksa,S., Alzbutas,G., Vaitkevicius,M., Lubiene,J., Letukiene,V., Burinskas,S., Sasnauskas,G. and Skirgaila,R. (2012) Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. *Protein Eng. Des. Sel.*, **25**, 657–668.
10. Yasukawa,K., Mizuno,M., Konishi,A. and Inouye,K. (2010) Increase in thermal stability of Moloney murine leukaemia virus reverse transcriptase by site-directed mutagenesis. *J. Biotechnol.*, **150**, 299–306.
11. Konishi,A., Ma,X. and Yasukawa,K. (2014) Stabilization of Moloney murine leukemia virus reverse transcriptase by site-directed mutagenesis of surface residue Val433. *Biosci. Biotechnol. Biochem.*, **78**, 75–78.
12. Konishi,A., Yasukawa,K. and Inouye,K. (2012) Improving the thermal stability of avian myeloblastosis virus reverse transcriptase alpha-subunit by site-directed mutagenesis. *Biotechnol. Lett.*, **34**, 1209–1215.
13. Gerard,G.F., Potter,R.J., Smith,M.D., Rosenthal,K., Dhariwal,G., Lee,J. and Chatterjee,D.K. (2002) The role of template-primer in protection of reverse transcriptase from thermal inactivation. *Nucleic Acids Res.*, **30**, 3118–3129.
14. Chandler,D.P., Wagnon,C.A. and Bolton,H. Jr. (1998) Reverse transcriptase (RT) inhibition of PCR at low concentrations of template and its implications for quantitative RT-PCR. *Appl. Environ. Microbiol.*, **64**, 669–677.
15. Myers,T.W. and Gelfand,D.H. (1991) Reverse transcription and DNA amplification by a Thermus thermophilus DNA polymerase. *Biochemistry*, **30**, 7661–7666.

16. Schonbrunner,N.J., Fiss,E.H., Budker,O., Stoffel,S., Sigua,C.L., Gelfand,D.H. and Myers,T.W. (2006) Chimeric thermostable DNA polymerases with reverse transcriptase and attenuated 3′-5′ exonuclease activity. *Biochemistry*, **45**, 12786–12795.
17. Sano,S., Yamada,Y., Shinkawa,T., Kato,S., Okada,T., Higashibata,H. and Fujiwara,S. (2012) Mutations to create thermostable reverse transcriptase with bacterial family A DNA polymerase from Thermotoga petrophila K4. *J. Biosci. Bioeng.*, **113**, 315–321.
18. Okano,H., Baba,M., Kawato,K., Hidese,R., Yanagihara,I., Kojima,K., Takita,T., Fujiwara,S. and Yasukawa,K. (2018) High sensitive RNA detection by one-step RT-PCR using the genetically engineered variant of DNA polymerase with reverse transcriptase activity from hyperthermophilies. *J. Biosci. Bioeng.*, **125**, 275–281.
19. Jones,M.D. and Foulkes,N.S. (1989) Reverse transcription of mRNA by Thermus aquaticus DNA polymerase. *Nucleic Acids Res.*, **17**, 8387–8388.
20. Sauter,K.B. and Marx,A. (2006) Evolving thermostable reverse transcriptase activity in a DNA polymerase scaffold. *Angew. Chem. Int. Ed. Engl.*, **45**, 7633–7635.
21. Kranaster,R., Drum,M., Engel,N., Weidmann,M., Hufert,F.T. and Marx,A. (2010) One-step RNA pathogen detection with reverse transcriptase activity of a mutated thermostable Thermus aquaticus DNA polymerase. *Biotechnol. J.*, **5**, 224–231.
22. Blatter,N., Bergen,K., Nolte,O., Welte,W., Diederichs,K., Mayer,J., Wieland,M. and Marx,A. (2013) Structure and function of an RNA-reading thermostable DNA polymerase. *Angew. Chem. Int. Ed. Engl.*, **52**, 11935–11939.
23. Zhao,C., Liu,F. and Pyle,A.M. (2018) An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA*, **24**, 183–195.
24. Mohr,S., Ghanem,E., Smith,W., Sheeter,D., Qin,Y., King,O., Polioudakis,D., Iyer,V.R., Hunicke-Smith,S., Swamy,S. *et al.* (2013) Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA*, **19**, 958–970.
25. Ellefson,J.W., Gollihar,J., Shroff,R., Shivram,H., Iyer,V.R. and Ellington,A.D. (2016) Synthetic evolutionary origin of a proofreading reverse transcriptase. *Science*, **352**, 1590–1593.
26. Finke,J.F., Hunt,B.P.V., Winter,C., Carmack,E.C. and Suttle,C.A. (2017) Nutrients and other environmental factors influence virus abundances across oxic and hypoxic marine environments. *Viruses*, **9**, 152.
27. Paez-Espino,D., Eloe-Fadrosh,E.A., Pavlopoulos,G.A., Thomas,A.D., Huntemann,M., Mikhailova,N., Rubin,E., Ivanova,N.N. and Kyrpides,N.C. (2016) Uncovering Earth's virome. *Nature*, **536**, 425–430.
28. Wommack,K.E., Nasko,D.J., Chopyk,J. and Sakowski,E.G. (2015) Counts and sequences, observations that continue to change our understanding of viruses in nature. *J. Microbiol.*, **53**, 181–192.
29. Schoenfeld,T., Liles,M., Wommack,K.E., Polson,S.W., Godiska,R. and Mead,D. (2010) Functional viral metagenomics and the next generation of molecular tools. *Trends Microbiol.*, **18**, 20–29.
30. Schoenfeld,T., Patterson,M., Richardson,P.M., Wommack,K.E., Young,M. and Mead,D. (2008) Assembly of viral metagenomes from yellowstone hot springs. *Appl. Environ. Microbiol.*, **74**, 4164–4174.
31. Moser,M.J., DiFrancesco,R.A., Gowda,K., Klingele,A.J., Sugar,D.R., Stocki,S., Mead,D.A. and Schoenfeld,T.W. (2012) Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme. *PLoS One*, **7**, e38371.
32. Chander,Y., Koelbl,J., Puckett,J., Moser,M.J., Klingele,A.J., Liles,M.R., Carrias,A., Mead,D.A. and Schoenfeld,T.W. (2014) A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (LAMP). *Front. Microbiol.*, **5**, 395.
33. Schoenfeld,T.W., Murugapiran,S.K., Dodsworth,J.A., Floyd,S., Lodes,M., Mead,D.A. and Hedlund,B.P. (2013) Lateral gene transfer of family A DNA polymerases between thermophilic viruses, aquificae, and apicomplexa. *Mol. Biol. Evol.*, **30**, 1653–1664.
34. Yamagami,T., Ishino,S., Kawarabayasi,Y. and Ishino,Y. (2014) Mutant Taq DNA polymerases with improved elongation ability as a useful reagent for genetic engineering. *Front. Microbiol.*, **5**, 461.
35. Lavinder,J.J., Hari,S.B., Sullivan,B.J. and Magliery,T.J. (2009) High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *J. Am. Chem. Soc.*, **131**, 3794–3795.
36. Stemmer,W.P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, **370**, 389–391.
37. Loh,E. and Loeb,L.A. (2005) Mutability of DNA polymerase I: implications for the creation of mutant DNA polymerases. *DNA Repair (Amst.)*, **4**, 1390–1398.
38. Holland,P.M., Abramson,R.D., Watson,R. and Gelfand,D.H. (1991) Detection of specific polymerase chain reaction product by utilizing the 5′—-3′ exonuclease activity of Thermus aquaticus DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 7276–7280.
39. Kim,Y., Eom,S.H., Wang,J., Lee,D.S., Suh,S.W. and Steitz,T.A. (1995) Crystal structure of Thermus aquaticus DNA polymerase. *Nature*, **376**, 612–616.
40. Barnes,W.M. (1992) The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene*, **112**, 29–35.
41. Wang,Y., Prosen,D.E., Mei,L., Sullivan,J.C., Finney,M. and Vander Horn,P.B. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.*, **32**, 1197–1207.
42. Merkens,L.S., Bryan,S.K. and Moses,R.E. (1995) Inactivation of the 5′-3′ exonuclease of Thermus aquaticus DNA polymerase. *Biochim. Biophys. Acta*, **1264**, 243–248.
43. Murali,R., Sharkey,D.J., Daiss,J.L. and Murthy,H.M. (1998) Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: the Fab is directed against an intermediate in the helix-coil dynamics of the enzyme. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 12562–12567.
44. Lawyer,F.C., Stoffel,S., Saiki,R.K., Chang,S.Y., Landre,P.A., Abramson,R.D. and Gelfand,D.H. (1993) High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5′ to 3′ exonuclease activity. *PCR Methods Appl.*, **2**, 275–287.
45. Goda,S.K. (2012) DNA shuffling and the production of novel enzymes and microorganisms for effective bioremediation and biodegradation process. *J. Bioremed. Biodegrad.*, **3**, e116.
46. Baar,C., d'Abbadie,M., Vaisman,A., Arana,M.E., Hofreiter,M., Woodgate,R., Kunkel,T.A. and Holliger,P. (2011) Molecular breeding of polymerases for resistance to environmental inhibitors. *Nucleic Acids Res.*, **39**, e51.
47. d'Abbadie,M., Hofreiter,M., Vaisman,A., Loakes,D., Gasparutto,D., Cadet,J., Woodgate,R., Paabo,S. and Holliger,P. (2007) Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.*, **25**, 939–943.
48. Matsukawa,H., Yamagami,T., Kawarabayasi,Y., Miyashita,Y., Takahashi,M. and Ishino,Y. (2009) A useful strategy to construct DNA polymerases with different properties by using genetic resources from environmental DNA. *Genes Genet. Syst.*, **84**, 3–13.
49. Delarue,M., Poch,O., Tordo,N., Moras,D. and Argos,P. (1990) An attempt to unify the structure of polymerases. *Protein Eng.*, **3**, 461–467.
50. Li,Y., Korolev,S. and Waksman,G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase. I: structural basis for nucleotide incorporation. *EMBO J.*, **17**, 7514–7525.