

Research Reports

Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase

Daniel J. G. Lahr¹ and Laura A. Katz^{1,2}

¹Graduate Program in Organismic and Evolutionary Biology, University of Massachusetts, Amherst, MA, USA, and

²Department of Biological Sciences, Smith College, Northampton, MA, USA

BioTechniques 47:857-866 (October 2009) doi 10.2144/000113219

Keywords: PCR chimera; template concentration; cycle number; gene family evolution; copy number

Supplementary material for this article is available at www.BioTechniques.com/article/113219.

PCR-mediated recombination can greatly impact estimates of diversity, both in environmental studies and in analyses of gene family evolution. Here we measure chimera (PCR-mediated recombinant) formation by analyzing a mixture of eight partial actin sequences isolated from the amoeba *Arcella hemisphaerica* amplified under a variety of conditions that mimic standard laboratory situations. We further compare a new-generation proofreading processivity-enhanced polymerase to both a standard proofreading enzyme and previously published results. Proofreading polymerases are preferred over other polymerases in instances where evolutionary inferences must be made. Our analyses reveal that reducing the initial template concentration is as critical as reducing the number of cycles for decreasing chimera formation and improving accuracy. Furthermore, assessing the efficiency of recovery of original haplotypes demonstrates that multiple PCR reactions are required to capture the actual genetic diversity of a sample. Finally, the experiments confirm that processivity-enhanced polymerases enable a substantial decrease in PCR-mediated recombination through reducing starting template concentration, without compromising the robustness of PCR reactions.

Introduction

Polymerase chain reaction (PCR)-based methods are the norm in molecular evolution studies of non-model taxa and in explorations of environmental DNA. For example, degenerate PCR is often used in systematic studies where numerous diverse taxa are to be sampled (1–6). Despite advances and dropping costs in mega- and metagenomic sequencing techniques (7–9), PCR methods remain key in hypothesis-driven environmental studies (10–14). PCR is used in such studies mainly because of its reproducibility, which enables the targeting of specific genes of interest from diverse taxa.

One worrisome aspect of PCR-based studies is the phenomenon of PCR-mediated recombination, or chimera formation (15,16). Chimeras are formed when incompletely extended DNA fragments anneal to closely related sequences generating recombinants between starting templates (17–19). It can be difficult to differentiate

original haplotypes from chimeras, leading to overestimation of biological diversity in environmental studies (20–22). Interpretations about the fate of genes in molecular evolution studies can also be compromised by the presence of chimeras, as has been shown in tests of positive selection in the major histocompatibility complex (MHC) in sticklebacks (23).

Most experimental work on PCR-mediated recombination has used traditional enzymes such as *Taq* polymerase to determine rates of chimera formation under conditions normally used in studies of environmental microbial samples [e.g., bacterial and archaeal 16S SSU-rDNA surveys (22,24–31)]. These studies determined that chimera formation can be reduced for most DNA polymerases when the cycle number is lowered and extension time increased (17,19,31–33). It is recommended that the lowest number of cycles be determined experimentally, which should be ~20 cycles or fewer. These suggestions can

be easily followed in experiments with high-quality DNA from organisms of known genome complexity (23,26). However, when dealing with DNA extracted from organisms that may have highly complex genomes, or preparations with chemical compounds that are not completely removed (e.g., environmental DNA from sediments), it is more difficult to optimize PCR for downstream applications such as cloning and sequencing (22,24,27,30). Additionally, in molecular evolution studies where single nucleotide polymorphisms are important for inferring evolutionary processes (e.g., population studies or analyses of rare biosphere), the high error rate of *Taq* polymerase is not desirable. To address such difficulties, a new generation of DNA polymerases has emerged that combine proofreading capabilities with enhanced DNA binding motifs (34), including Phusion (Finnzymes, Espoo, Finland), PfuUltra (Stratagene, La Jolla, CA, USA), and Pfx50 (Invitrogen, Carlsbad, CA, USA). These enzymes have not yet

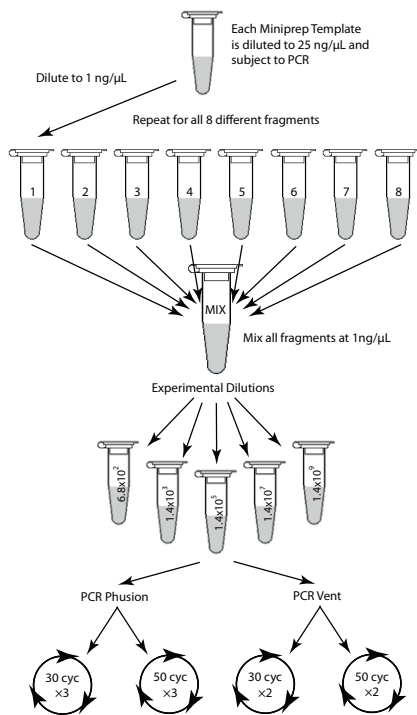


Figure 1. Experimental design used to amplify eight paralogous actin haplotypes in different cycling conditions using multiple initial template concentrations. The haplotypes, obtained from a previous study, were diluted and amplified individually. All eight were mixed to a concentration of 1 ng/μL and then diluted to five successive experimental concentrations. Each experimental concentration was amplified in triplicate using both a processivity-enhanced proofreading polymerase (Phusion) and a strict proofreading polymerase (Vent_R), in a low-cycle number (30) condition and a high-cycle number (50) condition. Each amplification was subsequently cloned, sequenced, and scanned for chimeras.

Table 1. Conditions examined for the formation of PCR-mediated recombinants (chimeras), with numbers of clones analyzed for each experiment

Template Concentration	1.4 × 10 ⁹	1.4 × 10 ⁷	1.4 × 10 ⁵	1.4 × 10 ³	6.8 × 10 ²
Phusion – template mix 1^a					
50 cycles (2x)	12, 10	23, 25	23, 22	NA	24, 24
30 cycles (1x)	27	29	24	15	failed
Phusion – template mix 2^a					
50 cycles (1x)	21	17	22	NA	15
30 cycles (2x)	17, 19	20, 22	22, 19	3, 11	failed
Vent_R^b					
50 cycles (2x)	7, 8	8, 3	failed	failed	failed
30 cycles (2x)	8, 10	10, 12	failed	failed	failed

We analyzed eight distinct experimental conditions to study, representing four initial template concentrations, each amplified using two different numbers of cycles. Numbers in body of table refer to good-quality clones sequenced and analyzed for each experiment. ^aFor Phusion polymerase, each treatment was repeated three times, two of them with the same template mixing event, and one independent. ^bFor Vent_R polymerase, we used one mixture, independently assayed two times. We skipped the 1.4 × 10³ concentration for the 50-cycle experiments. Amplification works normally under these conditions. NA, not available; failed, amplification did not yield a suitable product for cloning.

been analyzed for dynamics of chimera formation.

Our goal is to understand the formation of PCR-mediated recombinants when many closely related sequences are present in the same reaction, and when a high number of cycles is required to generate robust products. Low primer-to-target amplicon ratio is assumed to be the main reason for mismatch pairing in later cycles, which leads to chimera formation (15–17,24); thus, we also surveyed different initial DNA concentrations. Varying DNA concentrations is also relevant because in genomic DNA extractions, the absolute number of genome copies varies according to genome size and the subsequent high copy number of members of large gene families could lead to increased PCR recombination (23).

Here we analyze the formation of chimeras from a set of eight paralogous protein-coding genes by comparing the following experimental conditions: (i) a processivity-enhanced, proofreading polymerase to a traditional proofreading polymerase; (ii) high cycle number to standard cycle number; and (iii) a range of initial template concentrations. These sets of conditions are relevant to numerous research areas as parameters fall within recommendations and are likely to be used in standard laboratory practice.

Material and methods

1.1. Origin of templates

We chose to investigate a set of eight paralogous haplotypes of the actin gene extracted from the testate amoeba *Arcella hemisphaerica*. The eight haplotypes differ 2.4–20.5% in nucleotide sequence. Actin clones were obtained from previous work in *A. hemisphaerica* as described in Tekle et. al (5), except that resulting clones were purified using

the PureLink kit (Invitrogen). To generate templates for the experiment (Figure 1), we eliminated the vector by diluting each purification to 25 ng/μL and amplified them separately using *Arcella*-specific degenerate primers designed from an alignment with >30 actin paralogs from this taxon: AhemAct-F (5'-GARGARCAYC CYGTYTTTGTGAC-3') and AhemAct-R (5'-TAYTTYCTYTCDDG-GRGGAGCAAT-3'). Phusion Hot Start polymerase (Cat. no. F540; New England BioLabs, Ipswich, MA, USA) was used in the following conditions: 35 cycles of 98°C denaturing for 15 s, 56°C annealing for 15 s, and 72°C extension for 45 s. These primers yield an actin fragment that is 670 bp long. We performed these experiments using appropriate negative and positive controls and the amplified products were sequenced to check for quality (data not shown). Each amplified product was then purified using Microclean (The Gel Company, San Francisco, CA, USA). Finally, all haplotypes were individually diluted to 1 ng/μL and mixed (Figure 1).

1.2. Conditions

The conditions surveyed varied across a gradient of template concentrations at a high cycle number and a low cycle number (Table 1). We performed amplifications using both Phusion Hot Start polymerase (Cat. no. M0254; New England BioLabs) and Vent_R polymerase (Cat. No. F540; New England BioLabs). To assess varying template concentrations, we started with the mixture of eight haplotypes, each at 1 ng/μL measured in a NanoDrop (NanoDrop Products, Wilmington DE, USA), which is equivalent to 1.4 × 10⁹ molecules/μL of the amplified 670-bp fragment. We carried out six serial dilutions of this mixture in 10 mM Tris-HCl: five consecutive 1:10 dilutions followed by a final 1:2 dilution. We obtained the following concentrations: 1 ng/μL, 10⁻¹ ng/μL, 10⁻² ng/μL, 10⁻³ ng/μL, 10⁻⁴ ng/μL, 10⁻⁵ ng/μL, and 5 × 10⁻⁶ ng/μL. These concentrations correspond, respectively, to the following amounts of template molecules per microliter: 1.4 × 10⁹, 1.4 × 10⁷, 1.4 × 10⁵, 1.4 × 10³, and 6.8 × 10² (Table 1).

We chose to use 30 cycles as a reasonable number for standard PCR in molecular evolution and environmental studies. For a high cycle condition, 50 cycles was used as an upper extreme boundary where effects of high cycling would certainly be obtained [see Lenz et. al (23) for a brief discussion on commonly used cycle numbers]. For Phusion polymerase, every template concentration yielded enough

products in the 50-cycle condition, but the 30-cycle condition did not present sufficient yield for downstream analysis in the lowest template concentration (5×10^{-6} ng/ μ L or 6.8×10^2 molecules/ μ L). We decided to analyze the data from the 50-cycle condition to the lowest dilution possible, so for both cycle numbers we analyzed the lowest possible dilution. For Vent_R polymerase, we only obtained satisfactory yields for downstream processing in the two highest concentrations analyzed (1 ng/ μ L and 10^{-1} ng/ μ L; Table 1).

Except for number of cycles and initial template concentrations, all reactions were performed using the same cycling parameters. The recommended extension times (30 s/kb for Phusion, 60 s/kb for Vent_R) were increased threefold, taking into account previous claims that longer extension times decreases chimera formation (16,17,19,30,33). For experiments on Phusion polymerase, the concentrations for amplification mixtures followed manufacturer's protocol (1 $\times$ HF buffer, 1.5 mM MgCl₂, 0.2 mM each dNTP, 0.5 μ M each primer, and 0.01 U/ μ L polymerase). The cycling conditions were: 95°C for 3 min; followed by 30 (or 50) cycles at 98°C for 15 s, 56°C for 15 s, and 72°C for 90 s; and then a final extension at 72°C for 5 min. For Vent_R polymerase, the concentration for amplification mixtures was per manufacturer's recommendations (1 $\times$ Thermopol buffer, 1.5 mM Mg₂SO₄, 0.2 mM each dNTP, 0.5 μ M each primer, and 0.005 U/ μ L polymerase). Reaction mixtures were incubated at 95°C for 3min; followed by 30 (or 50) cycles at 95°C for 15 s, 56°C for 15 s, and 72°C for 3 min; and a final extension of 72°C for 5 min. In both Phusion and Vent_R reaction mixtures, the final concentrations of DNA were 1.1×10^{-4} μ M, 1.1×10^{-6} μ M, 1.1×10^{-8} μ M, 1.1×10^{-10} μ M, and 5.7×10^{-11} μ M, corre-

sponding to the experimental dilutions (in molecules/ μ L) 1.4×10^9 , 1.4×10^7 , 1.4×10^5 , 1.4×10^3 , and 6.8×10^2 , respectively.

1.3. Cloning

Each amplification reaction was run into 1% Seakem GTG agarose gel (Cambrex Bio Science, Rockland, ME, USA) made with modified TAE buffer (40 mM Tris-acetate, pH 8.0, 0.1 mM Na₂EDTA). The 670-bp band was visualized by staining with SYBR Safe (Invitrogen) at a dilution of 1:10⁴. We then excised the band from the gel and isolated DNA from agarose with the Millipore UltraFREE DA (Millipore, Billerica, MA, USA). The obtained product was further purified

using Microclean. The purified products were then ligated using Zero Blunt TOPO cloning kit (Invitrogen) and transformed into One Shot Competent Cells (Invitrogen) per the manufacturer's instructions. Cloned cells were plated in Luria-Bertani/kanamycin plates and colonies were screened for inserts by direct PCR using AmpliTaq Gold Polymerase (Invitrogen). Positive colonies were then purified in a 96-well format using a PureLink kit (Invitrogen), per the manufacturer's instructions. Sequencing reactions were performed in a 96-well format in an ABI 3100 automated sequencer (Applied Biosystems, Foster City, CA, USA) at the Pennsylvania State University Nucleic Acid Facility (University Park, PA, USA). We aimed to

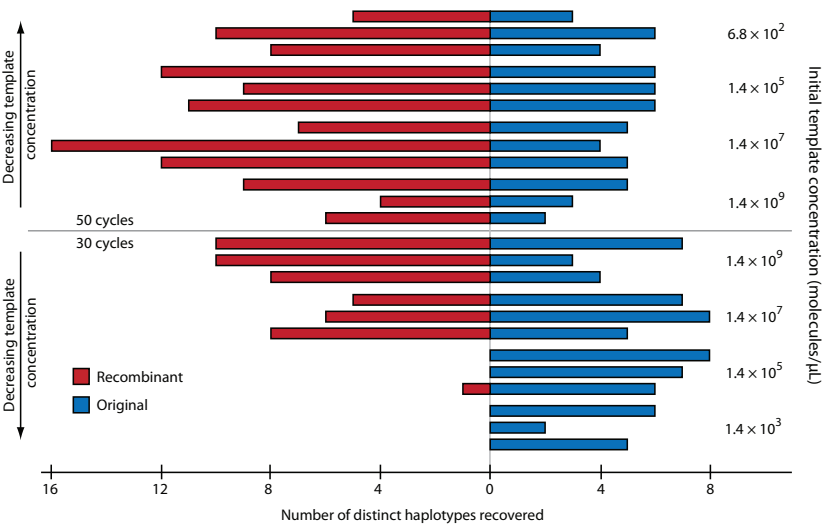


Figure 2. Number of chimeras and original haplotypes recovered across all concentrations and cycling conditions analyzed using a processivity-enhanced proofreading polymerase (Phusion). The three replicates for each condition are shown separately. Numbers on y-axis represent the initial concentration of molecules.

Table 2. Total clones analyzed, average number of chimeras, original haplotypes recovered, and percentage of chimera haplotypes formed under varying conditions								
Cycle Number	50 cycles				30 cycles			
Concentration	1.4×10^9	1.4×10^7	1.4×10^5	6.8×10^2	1.4×10^9	1.4×10^7	1.4×10^5	1.4×10^3
Phusion								
Total clones	43	65	67	63	63	71	65	29
Chimera Haplotypes	6	12	11	8	9	7	0	0
Original Haplotypes	3	5	6	4	5	7	7	4
Chimera %	65 ± 9	70 ± 11	64 ± 3	65 ± 2	67 ± 9	50 ± 10	5 ± 8	0
Vent _R								
Total clones	15	11	-	-	18	22	-	-
Chimera Haplotypes	2	2	-	-	3	2	-	-
Original Haplotypes	3	3	-	-	3	4	-	-
Chimera %	38 ± 18	42 ± 12	-	-	46 ± 5	38 ± 31	-	-
Numbers represent averages across all replicates for the particular experiment. Percentages are represented along with the standard deviation. For Phusion polymerase, the number of clones for each condition is averaged out of 3 replicate experiments; for Vent _R polymerase, out of 2 replicate experiments. A full detailed table is available in Supplementary Material 1.								

sequence 24 clones of each amplification/cloning event for comparative reasons, but the rate of sequencing failure varied across conditions (see Table 1).

1.4. Replicates and controls

In order to avoid stochastic effects, we replicated the experiment by first diluting and mixing the original templates two times independently, and then repeating a subset of conditions for each experiment (Table 1). For the Phusion experiments, templates were made two times independently, starting from the first amplification of haplotypes. Each condition was amplified up to two times for each independent making of templates. For Vent_R experiments, templates were made once and amplifications replicated twice. Negative controls were used throughout the experiment, and no contamination was detected. We also used a positive control for contamination by randomly choosing one of the original haplotypes to run through the whole protocol (dilution, amplification, gel isolation, cloning, and sequencing) side-by-side with the experimental mixed haplotypes. We sequenced at least four positive clones for each experiment and no cross-contamination was detected, as all sequenced clones were identical to the original haplotype. This positive control indicated that we did not have cross-tube contamination, which could cause concentration errors or template bias.

1.5. Determining PCR recombination events

We analyzed sequences for each experiment individually. The sequences were initially scanned for quality using Lasergene's SeqMan (35) (DNASTAR, Madison, WI, USA); poor quality sequences due to ambiguity at sites were discarded, less than 1% of the overall sample. This step presumably excludes heteroduplexes as well (36). All sequences from each PCR experiment were compiled and aligned manually using MacClade (37). Similarity trees were generated by neighbor-joining (NJ) algorithm using PAUP* (38) (<http://paup.csit.fsu.edu>) (data not shown) to tally cloned sequences as original or chimeric haplotype (Supplementary Material 1). Polymorphisms were confirmed by eye for sequences that were not 100% identical to the original haplotypes. Breakpoints were determined manually for chimeric haplotype (Supplementary Material 2), by aligning each chimera against all eight initial haplotypes in Lasergene's MegAlign (35) (DNASTAR). A similarity tree for all encountered chimeric haplotypes was used to search for the exact same chimera in independent amplification events. Statistical analyses regarding distribution of breakpoints and one-way ANOVAs

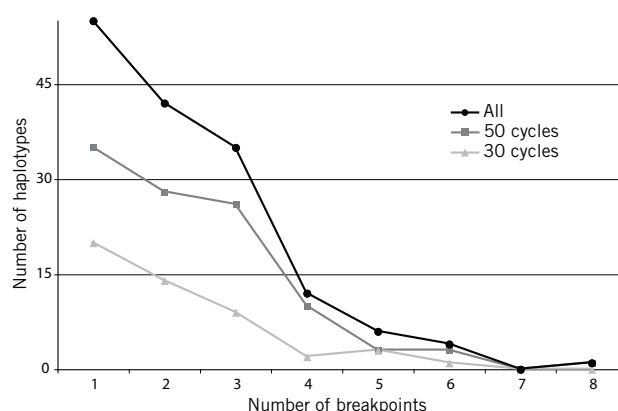


Figure 3. Distribution of chimeric haplotypes according to number of breakpoints. The majority of chimeric haplotypes have more than one breakpoint. There is no significant difference in distribution between cycle numbers.

were performed in STATA/SEM 9.1 (39) (STATA, College Station, TX, USA).

1.6. Software recognition of the chimeras obtained

We used the online software Bellerophon (40) (<http://foo.maths.uq.edu.au/~huber/bellerophon.pl>) to determine efficiency of automated chimera recognition (Supplementary Material 3). We chose to use this particular software as an exemplary system because (i) it is widely used; (ii) it is explicitly designed to recognize PCR-mediated recombination, and not historical recombination signal; and (iii) it does not rely on a database of confirmed sequences, which is useful to only a fraction of studies.

Results

PCR-mediated recombination between eight initial haplotypes (2.4–20.8% divergence) was investigated across a gradient of template concentrations in two different cycling conditions (Table 1). Extensive experiments using the Phusion polymerase were compared with a more limited data set using Vent_R polymerase (Table 2).

PCR recombination events across different treatments

We determined the number of chimeras recovered in each individual amplification/cloning event (Figure 2; see Supplementary Material 1 for a complete tally). For Phusion polymerase, 50-cycle amplification reaction yields significantly more recombinants than the 30-cycle amplification (one-way ANOVA: $df = 23$, $F = 9.03$, $P = 0.006$), with an average of at least 65% recombinant sequences for all initial template concentrations (Table 2). Reduction of concentration does not significantly reduce chimera formation for

the 50-cycle condition (one-way ANOVA: $df = 11$, $F = 2.20$, $P = 0.16$). Chimera formation remained high (~60%) at 30 cycles with higher initial template concentration not significantly different from the same concentrations at 50 cycles (one-way ANOVAs for concentration $1.4 \times 10^9/\mu\text{L}$: $df = 5$, $F = 4.17$, $P = 0.11$; for concentration $1.4 \times 10^7/\mu\text{L}$: $df = 5$, $F = 3.76$, $P = 0.12$), but decreased to 5% at a starting concentration of 1.4×10^5 molecules/ μL and no chimeras were recovered in the lowest concentration (1.4×10^3 molecules/ μL), both significant decreases (one-way ANOVA: $df = 11$, $F = 36.33$, $P = 0.00001$). For Vent_R polymerase, we had fewer conditions to compare as PCR failed at the lower DNA concentrations. At 50 cycles, Vent_R yielded an average of 40–44% chimeras at all observed concentrations, and at 30 cycles, 45% and 36% for higher and lower concentrations, respectively (Table 2).

Recovery of original haplotypes per treatment

The number of original haplotypes recovered out of the eight initially used varied for different treatments (Figure 2), and do not differ significantly with varying number of cycles (one-way ANOVA: $df = 23$, $F = 2.52$, $P = 0.12$) or starting template concentration (one-way ANOVAs; 30 cycles: $df = 11$, $F = 1.85$, $P = 0.21$; 50 cycles: $df = 11$, $F = 2.91$, $P = 0.1$). Nevertheless, a few trends emerge. For Phusion at 50 cycles, all template concentrations, on average, recover an assortment of four haplotypes out of the eight originals. At 30 cycles, the intermediate concentrations (1.4×10^7 and 1.4×10^5 molecules/ μL) recovered seven haplotypes on average, with individual PCR reactions actually being able to recover all eight (Figure 2, Supplementary Material 1); the higher and lower template

concentrations recovered four haplotypes on average. Vent_R polymerase recovered an average of five original haplotypes for all conditions. Some sequences were more prone to be recovered: original haplotypes 2 and 5 were recovered in almost all experiments; original haplotype 3 was recovered only in half the experiments (Supplementary Material 1).

Number of breakpoints in chimeric sequences

We determined breakpoints and participant original haplotypes for each chimeric haplotype. Both the distribution of numbers of breakpoints per sequence and the distribution of breakpoints along the sequence suggests that under PCR conditions, the recombination events are random. Chimeras varied from having a single breakpoint with two clear parental sequences to having eight breakpoint and six parental sequences alternating in participation (see Supplementary Material 2 for a complete tally). The majority of chimeras (65%) had more than one breakpoint and in most cases there were more than two parental sequences for each chimera (Figure 3). There is no clear pattern between number of breakpoints and template concentration or number of cycles: the distribution of sequences with breakpoints follows a Poisson distribution when taken together (Poisson regression likelihood ratio chi-squared = 7.83, $P = 0.02$, $df = 2$; Pearson goodness-of-fit chi-squared = 110.67, $P = 0.99$, $df = 152$), and follow that distribution when partitioned by concentration ($P = 0.01$) or cycling number ($P = 0.04$) (Figure 3). Additionally, we were unable to determine a correlation between sequence features (local similarity, conservation) and susceptibility to being a breakpoint. The distributions of breakpoints along the sequence are not significantly different from the expected in a normal distribution (Shapiro-Wilk $W = 0.95$, $P = 0.04$; Supplementary Material 4).

Haplotype participation in chimeric sequences

The frequency that a specific original haplotype was involved in chimeric events corresponds with the frequency with which that haplotype was recovered overall in PCR reactions (Supplementary Material 5). Original haplotype 2 was recovered the most times across experiments (51 clones overall), and it also was involved as a part of a chimera in 81 cases. Original haplotype 3 was the least recovered haplotype across experiments (16 clones overall) and was also the least likely haplotype to participate in chimeras (23 counts). Such a pattern in the composition of chimeras is further

evidence that recombination events are mostly random. The more readily available sequences are more likely to participate in recombination events, without any bias toward a particular haplotype or group of haplotypes.

Discussion

Chimeras are more likely to be observed when both a high-cycle condition and high initial concentration of templates are used (Table 2). In contrast, no recombinants were observed in the low cycle/low concentration conditions. Although our analyses corroborate the inference that high cycle numbers induce chimera formation in PCR (17–19,26,30,33), they also highlight the importance of initial template concentration in chimera formation (Figure 2). The effect of template concentration revealed here is due to the wider range (seven orders of magnitude) of concentrations analyzed compared with previous research (26). Such a range is enabled by the ability of Phusion—a processivity-enhanced polymerase—to amplify very low concentrations of template (Table 1).

Rates of recombination obtained for Phusion are highly variable across conditions. Previous research has proposed that enzymes with higher processivity yield more chimeric sequences (26). This is confirmed at high cycle/high template concentration conditions, with an average yield of 71% chimeras (Figure 2). However, the surprising result is that the same enzyme yields absolutely no chimeras when the initial template concentration is sufficiently low (1.4×10^3 starting molecules) and when the cycle number is reduced to 30 cycles (Figure 2). The processivity-enhanced enzyme makes up for its high rate of chimera formation by being able to amplify initial concentrations that are four to five orders of magnitude lower than what the strict proofreading polymerase can amplify (Table 1), effectively reducing chimera formation to zero (Table 2). Non-proofreading polymerases like *Taq* might also benefit from less concentrated initial templates but they are less desirable for molecular evolution studies. Furthermore, we attribute Phusion's ability to amplify low template concentrations to the enhanced processivity. Ordinary *Taq* polymerases might not be able to amplify concentrations that are low enough to reduce artifact formation.

The percentages of chimeras formed per reaction are higher for certain conditions in the present survey than have been reported in much of the literature, which we believe is due to the greater complexity of our starting templates (i.e., eight actin

haplotypes). The rates of recombination for a proofreading enzyme (Vent_R) under standard conditions (30 cycles) averaged ~40% recombinants in the present study, while the highest available literature reports are 32% for *Taq* polymerase after 30 cycles on 7 distinct initial haplotypes (30), 16% for the proofreading Expand H-F system (Boehringer, Mannheim, Germany) after 25 cycles using 8 initial haplotypes (27), and 31% recombinants using *Taq* polymerase across multiple loci in polyploid cotton (41). Since the reported chimera formation rates in other available literature ranges 1–5% across a variety of enzymes (15–19,24,26), we attribute the higher rates in our experiment [as well as in two others (27,30)] to the higher number of initial haplotypes: most studies used 2–4 initial haplotypes to test chimera formation. In contrast, ≤35% recombinants were reported for only two MHC loci (31), which may indicate a possible influence of the template itself. Furthermore, we demonstrated that there is an increase in chimera formation as diversity in the original sample increases. This reinforces the idea that molecular environmental studies might be plagued with a slew of artificial sequences (20). Moreover, initial template concentrations (i.e., abundances) in environmental samples will most likely be unequally distributed, which might influence the formation of artifacts. However, differential abundance of templates probably has a larger impact on the detection of true diversity and our analyses indicate that multiple (>2) PCR reactions will be required to capture the true diversity of a sample even when abundances of templates are equivalent.

We find that the majority of chimeras contain more than one breakpoint, indicating that more than two parental sequences can be involved in PCR-mediated recombination. This high rate of crossover is independent of cycle number or initial template concentration (Figure 3). This observation will create problems for chimera-detecting software that base search criteria in finding one breakpoint per sequence. For example, while using the online software Bellerophon (40) to detect the chimeras in the present data set, only an average of $65 \pm 18\%$ of chimeras were detected. Even more worrisome, there is a false-positive rate of $40 \pm 31\%$ (see Supplementary Material 3 for details).

Capturing the full diversity within a sample requires a combination of multiple PCR reactions that have been performed under chimera-reducing conditions. On average, PCR reactions at high cycle numbers are unable to recover all diversity (average 4 ± 1 out of 8 starting haplotypes, Figure 2),

even if all three replicates are combined (Supplementary Material 1), and have the added bias of generating false haplotypes. While low cycle number improves recovery (7 ± 1 out of 8 starting haplotypes; Figure 2), it is likely that a single PCR experiment will not capture all the diversity. For example, in chimera-reducing conditions with low cycle number and lowest initial template concentration possible, individual PCR reactions detected an average of four out of the eight haplotypes, but performing three replicates will certainly describe all diversity in this eight-haplotype system (Supplementary Material 1). Hence we reiterate the necessity of replicating PCR reactions to assess biodiversity (19,24,26), and we add the recommendation that these replicate PCR runs be undertaken with minimal DNA concentrations and cycle numbers, which need to be established on a sample-by-sample basis.

Acknowledgments

This work was supported by the National Institutes of Health (NIH; AREA Award no. 1R15GM081865-01) and the National Science Foundation (nos. OCE-0648713 and DEB 043115) to L.A.K., and the

Conselho Nacional de Pesquisa e Desenvolvimento, Brazil (no. GDE 200853/2007-4) to D.J.G.L.

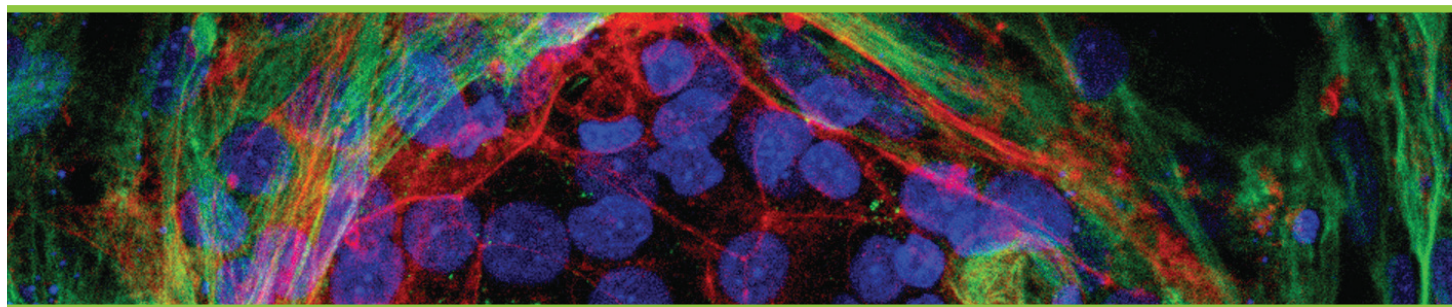
The authors declare no competing interests. This paper is subject to the NIH Public Access Policy.

References

1. Baptiste, E., H. Brinkmann, J.A. Lee, D.V. Moore, C.W. Sensen, P. Gordon, L. Durufle, T. Gasterlan, et al. 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl. Acad. Sci. USA* 99:1414-1419.
2. Grant, J., Y.I. Tekle, O.R. Anderson, D.J. Patterson, and L.A. Katz. 2009. Multigene evidence for the placement of a heterotrophic amoeboid lineage *Leukarachnion* sp. among photosynthetic stramenopiles. *Protist* 160:376-385.
3. Baldauf, S.L. 2003. The deep roots of eukaryotes. *Science* 300:1703-1706.
4. Nikolaev, S.I., C. Berney, J.F. Fahrni, I. Boliver, S. Polet, A.P. Mylnikov, V.V. Aleshin, N.B. Petrov, and J. Pawlowski. 2004. The twilight of Heliozoa and rise of Rhizaria, a new supergroup of amoeboid eukaryotes. *Proc. Natl. Acad. Sci. USA* 101:8066-8071.
5. Tekle, Y.I., J. Grant, O.R. Anderson, T.A. Nerad, J.C. Cole, D.J. Patterson, and L.A. Katz. 2008. Phylogenetic placement of diverse amoebae inferred from multigene analyses and assessment of clade stability within 'Amoebozoa' upon removal of varying rate classes of SSU-rDNA. *Mol. Phylogenet. Evol.* 47:339-352.
6. Yoon, H.S., J. Grant, Y. Tekle, M. Wu, B. Chaon, J. Cole, J. Logsdon, D. Patterson, et al. 2008. Broadly sampled multigene trees of eukaryotes. *BMC Evol. Biol.* 8:14.
7. Holt, R.A. and S.J. Jones. 2008. The new paradigm of flow cell sequencing. *Genome Res.* 18:839-846.
8. Allen, E.E. and J.F. Banfield. 2005. Community genomics in microbial ecology and evolution. *Nat. Rev. Microbiol.* 3:489-498.
9. Keller, M. and K. Zengler. 2004. Tapping into microbial diversity. *Nat. Rev. Microbiol.* 2:141.
10. Barns, S.M., C.F. Delwiche, J.D. Palmer, and N.R. Pace. 1996. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl. Acad. Sci. USA* 93:9188-9193.
11. Costas, B.A., G. McManus, M. Doherty, and L.A. Katz. 2007. Use of species-specific primers and PCR to measure the distributions of planktonic ciliates in coastal waters. *Limnol. Oceanogr. Methods* 5:163-173.
12. Dawson, S.C. and N.R. Pace. 2002. Novel kingdom-level eukaryotic diversity in anoxic environments. *Proc. Natl. Acad. Sci. USA* 99:8324-8329.
13. Doherty, M., B.A. Costas, G.B. McManus, and L.A. Katz. 2007. Culture-

BioTechniques®

The International Journal of Life Science Methods



Why publish in BioTechniques?

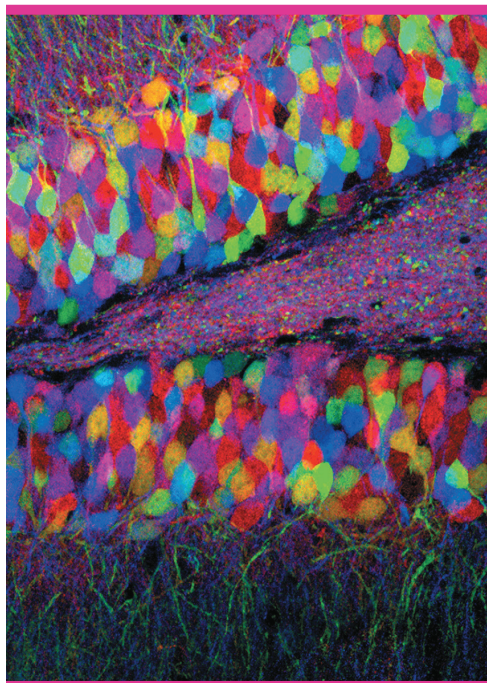
80,000 bench scientists in 70 countries receive personal print copies of *BioTechniques* every month, and *BioTechniques* is the most cited general methods journal.

We welcome reports describing innovative methods, substantive modifications to existing methods, and innovative applications of existing methods to new models or scientific questions.

More info: www.BioTechniques.com/authors

Submit manuscripts via: www.editorialmanager.com/bt





Author Reprints

Make the most of your hard work by ordering reprints of your article published in *BioTechniques*. Reprints are an inexpensive and easy way to distribute your findings to students and colleagues.

Corporate Reprints

Leverage *BioTechniques*, the most powerful brand in the market. Reprints help support your sales effort by utilizing articles that spotlight your brand/products to educate customers at meetings and industry trade events.

More info:

www.BioTechniques.com/reprints



- independent assessment of planktonic ciliate diversity in coastal northwest Atlantic waters. *Aquat. Microb. Ecol.* 48:141-154.
14. Edgcomb, V.P., D.T. Kysela, A. Teske, A. de Vera Gomez, and M.L. Sogin. 2002. Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proc. Natl. Acad. Sci. USA* 99:7658-7662.
 15. Brakenhoff, R.H., J.G.G. Schoenmakers, and N.H. Lubsen. 1991. Chimeric cDNA clones: a novel PCR artifact. *Nucleic Acids Res.* 19:1949.
 16. Meyerhans, A., J.-P. Vartanian, and S. Wain-Hobson. 1990. DNA recombination during PCR. *Nucleic Acids Res.* 18:1687-1691.
 17. Judo, M.S.B., A.B. Wedel, and C. Wilson. 1998. Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res.* 26:1819-1825.
 18. Bradley, R.D. and D.M. Hillis. 1997. Recombinant DNA sequences generated by PCR amplification. *Mol. Biol. Evol.* 14:592-593.
 19. Kanagawa, T. 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Biosci. Bioeng.* 96:317.
 20. Hugenholtz, P. and T. Huber. 2003. Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *Int. J. Syst. Evol. Microbiol.* 53:289-293.
 21. Berney, C., J. Fahrni, and J. Pawlowski. 2004. How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. *BMC Biol.* 2:13.
 22. von Wintzingerode, F., U.B. Gobel, and E. Stackbrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* 21:213.
 23. Lenz, T.L. and S. Becker. 2008. Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci—implications for evolutionary analysis. *Gene* 427:117.
 24. Acinas, S.G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M.F. Polz. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* 71:8966-8969.
 25. Liesack, W., H. Weyland, and E. Stackbrandt. 1991. Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. *Microb. Ecol.* 21:191-198.
 26. Qiu, X., L. Wu, H. Huang, P.E. McDonel, A.V. Palumbo, J.M. Tiedje, and J. Zhou. 2001. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Appl. Environ. Microbiol.* 67:880-887.
 27. Speksnijder, A.G.C.L., G.A. Kowalchuk, S. De Jong, E. Kline, J.R. Stephen, and H.J. Laanbroek. 2001. microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. *Appl. Environ. Microbiol.* 67:469-472.
 28. Suzuki, M., M.S. Rappe, and S.J. Giovannoni. 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. *Appl. Environ. Microbiol.* 64:4522-4529.
 29. Suzuki, M.T. and S.J. Giovannoni. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol.* 62:625-630.
 30. Wang, G.C. and Y. Wang. 1997. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl. Environ. Microbiol.* 63:4645-4650.
 31. Yu, W., K.J. Rusterholtz, A.T. Krummel, and N. Lehman. 2006. Detection of high levels of recombination generated during PCR amplification of RNA templates. *Biotechniques* 40:499-507.
 32. Zaphiropoulos, P.G. 1998. Non-homologous recombination mediated by *Thermus aquaticus* DNA polymerase I. Evidence supporting a copy choice mechanism. *Nucleic Acids Res.* 26:2843-2848.
 33. Kurata, S., T. Kanagawa, Y. Magariyama, K. Takatsu, K. Yamada, T. Yokomaku, and Y. Kamagata. 2004. Reevaluation and reduction of a PCR bias caused by reannealing of templates. *Appl. Environ. Microbiol.* 70:7545-7549.
 34. Wang, Y., D.E. Prosen, L. Mei, J.C. Sullivan, M. Finney, and P.B. Vander Horn. 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.* 32:1197-1207.
 35. Clewley, J.P. and C. Arnold. 1997. MegAlign. The multiple alignment module of LaserGene. *Methods Mol. Biol.* 70:119-129.
 36. Thompson, J.R., L.A. Marcelino, and M.F. Polz. 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res.* 30:2083-2088.
 37. Maddison, W.P. and D.R. Maddison. 1989. Interactive analysis of phylogeny and character evolution using the computer program MacClade. *Folia Primatol. (Basel)* 53:190-202.
 38. Wilgenbusch, J.C. and D.L. Swofford. 2003. Inferring evolutionary trees with PAUP*. *Current Protocols in Bioinformatics*. John Wiley & Sons, Malden, MA.
 39. Boston, R.C. and A.E. Sumner. 2003. STATA: a statistical analysis system for examining biomedical data. *Adv Exp Med Biol.* 537:353-369.
 40. Huber, T., G. Faulkner, and P. Hugenholtz. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* 20:2317-2319.
 41. Cronn, R., M. Cedroni, T. Haselkorn, C. Grover, and J.F. Wendel. 2002. PCR-mediated recombination in amplification products derived from polyploidy cotton. *Theor. Appl. Genet.* 104:482-489.

Received 29 April 2009; accepted 14 July 2009.

Address correspondence to Daniel J. G. Lahr, Graduate Program in Organismic and Evolutionary Biology, University of Massachusetts, 319 Morrill Science Center, Amherst, MA, 01003, USA. email: dlahr@nsm.umass.edu