Reduction of 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.
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 *Arcella* (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′-GARGARCYC GYTDDTGTGAC-3′) and AhemAct-R (5′-TAYTYCTTYCDG- GRRGAGCAAT-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 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 × 10^6 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 × 10^5, 1.4 × 10^4, 1.4 × 10^3, 1.4 × 10^2, and 6.8 × 10^1 (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

**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), 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.

<table>
<thead>
<tr>
<th>Template Concentration</th>
<th>1.4 × 10^6</th>
<th>1.4 × 10^5</th>
<th>1.4 × 10^4</th>
<th>1.4 × 10^3</th>
<th>6.8 × 10^2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phusion – template mix 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 cycles (2x)</td>
<td>12, 10</td>
<td>23, 25</td>
<td>23, 22</td>
<td>NA</td>
<td>24, 24</td>
</tr>
<tr>
<td>30 cycles (1x)</td>
<td>27</td>
<td>29</td>
<td>24</td>
<td>15</td>
<td>failed</td>
</tr>
<tr>
<td><strong>Phusion – template mix 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 cycles (1x)</td>
<td>21</td>
<td>17</td>
<td>22</td>
<td>NA</td>
<td>15</td>
</tr>
<tr>
<td>30 cycles (2x)</td>
<td>17, 19</td>
<td>20, 22</td>
<td>22, 19</td>
<td>3, 11</td>
<td>failed</td>
</tr>
<tr>
<td><strong>Vent</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 cycles (2x)</td>
<td>7, 8</td>
<td>8, 3</td>
<td>failed</td>
<td>failed</td>
<td>failed</td>
</tr>
<tr>
<td>30 cycles (2x)</td>
<td>8, 10</td>
<td>10, 12</td>
<td>failed</td>
<td>failed</td>
<td>failed</td>
</tr>
</tbody>
</table>

We analyzed eight distinct experimental conditions to study, representing four initial template concentrations, each amplified using two different numbers of cycles. Numbers in bold refer to good-quality clones sequenced and analyzed for each experiment. For Phusion polymerase, each treatment was repeated three times, two of them with the same template mixing event, and one independent. For Vent polymerase, we used one mixture, independently assayed two times. We skipped the 1.4 × 10^3 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.
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 × 10^3 ng/μL or 6.8 × 10^2 molecules/μL). 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 polymerase, we only obtained satisfactory yields for downstream processing in the two highest concentrations analyzed (1 ng/μL and 10^1 ng/μL; 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) 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× HF buffer, 1.5 mM MgCl₂, 0.2 mM each dNTP, 0.5 μM each primer, and 0.01 U/μL 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 polymerase, the concentration for amplification mixtures was per manufacturer’s recommendations (1× Thermopol buffer, 1.5 mM MgSO₄, 0.2 mM each dNTP, 0.5 μM each primer, and 0.005 U/μL polymerase). Reaction mixtures were incubated at 95°C for 3 min; 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 reaction mixtures, the final concentrations of DNA were 1.1 × 10^4 μM, 1.1 × 10^5 μM, 1.1 × 10^6 μM, 1.1 × 10^7 μM, and 5.7 × 10^8 μM, corresponding to the experimental dilutions (in molecules/μL) 1.4 × 10³, 1.4 × 10⁴, 1.4 × 10⁵, and 6.8 × 10², 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 AmpliTag 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

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 polymerase, we only obtained satisfactory yields for downstream processing in the two highest concentrations analyzed (1 ng/μL and 10^1 ng/μL; 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) 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× HF buffer, 1.5 mM MgCl₂, 0.2 mM each dNTP, 0.5 μM each primer, and 0.01 U/μL 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 polymerase, the concentration for amplification mixtures was per manufacturer’s recommendations (1× Thermopol buffer, 1.5 mM MgSO₄, 0.2 mM each dNTP, 0.5 μM each primer, and 0.005 U/μL polymerase). Reaction mixtures were incubated at 95°C for 3 min; 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 reaction mixtures, the final concentrations of DNA were 1.1 × 10^4 μM, 1.1 × 10^5 μM, 1.1 × 10^6 μM, 1.1 × 10^7 μM, and 5.7 × 10^8 μM, corresponding to the experimental dilutions (in molecules/μL) 1.4 × 10³, 1.4 × 10⁴, 1.4 × 10⁵, and 6.8 × 10², 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 AmpliTag 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

Table 2. Total clones analyzed, average number of chimeras, original haplotypes recovered, and percentage of chimera haplotypes formed under varying conditions

<table>
<thead>
<tr>
<th>Cycle Number</th>
<th>50 cycles</th>
<th></th>
<th></th>
<th>30 cycles</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration</td>
<td>1.4 × 10³</td>
<td>1.4 × 10⁴</td>
<td>1.4 × 10⁵</td>
<td>6.8 × 10²</td>
<td>1.4 × 10³</td>
<td>1.4 × 10⁴</td>
<td>1.4 × 10⁵</td>
</tr>
<tr>
<td>Phusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total clones</td>
<td>43</td>
<td>65</td>
<td>67</td>
<td>63</td>
<td>63</td>
<td>71</td>
<td>65</td>
</tr>
<tr>
<td>Chimera Haplotypes</td>
<td>6</td>
<td>12</td>
<td>11</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Original Haplotypes</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Chimera %</td>
<td>65 ± 9</td>
<td>70 ± 11</td>
<td>64 ± 3</td>
<td>65 ± 2</td>
<td>67 ± 9</td>
<td>50 ± 10</td>
<td>5 ± 8</td>
</tr>
<tr>
<td>Vent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total clones</td>
<td>15</td>
<td>11</td>
<td>-</td>
<td>-</td>
<td>18</td>
<td>22</td>
<td>-</td>
</tr>
<tr>
<td>Chimera Haplotypes</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Original Haplotypes</td>
<td>3</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>Chimera %</td>
<td>38 ± 18</td>
<td>42 ± 12</td>
<td>-</td>
<td>-</td>
<td>46 ± 5</td>
<td>38 ± 31</td>
<td>-</td>
</tr>
</tbody>
</table>

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 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<sub>R</sub> 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 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.cit. PSU.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 MegaAlign (35) (DNASTAR). A similarity tree for all encountered chimeric haplotypes was used to search for the exact same chimeras in independent amplification events. Statistical analyses regarding distribution of breakpoints and one-way ANOVAs 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<sub>R</sub> 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 × 10<sup>7</sup>/µL: df = 5, F = 4.17, P = 0.11; for concentration 1.4 × 10<sup>8</sup>/µL: df = 5, F = 3.76, P = 0.12), but decreased to 5% at a starting concentration of 1.4 × 10<sup>9</sup> molecules/µL and no chimeras were recovered in the lowest concentration (1.4 × 10<sup>5</sup> molecules/µL), both significant decreases (one-way ANOVA: df = 11, F = 36.33, P = 0.00001). For Vent<sub>R</sub> polymerase, we had fewer conditions to compare as PCR failed at the lower DNA concentrations. At 50 cycles, Vent<sub>R</sub> 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<sup>7</sup> and 1.4 × 10<sup>8</sup> 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

![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.](image-url)
concentrations recovered four haplotypes on average. Vent₇polymerase 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 breakpoints 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 breakpoints 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 x 10³ 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 other studies. Some authors report that more than 90% of templates are equivalent. 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 ± 18% of chimeras were detected. Even more worrisome, there is a false-positive rate of 40 ± 31% (see Supplementary Material 5 for details).

Capturing the full diversity within a sample requires a combination of multiple PCR reactions that have been performed under different-reducing conditions. On average, PCR reactions with higher 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. 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. 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. 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.

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

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

---


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@nmss.umass.edu