

# Product Length, Dye Choice, and Detection Chemistry in the Bead-Emulsion Amplification of Millions of Single DNA Molecules in Parallel

Irene Tiemann-Boege,<sup>\*,†,§</sup> Christina Curtis,<sup>‡</sup> Deepali N. Shinde,<sup>§</sup> Daniel B. Goodman,<sup>‡</sup> Simon Tavaré,<sup>‡,§</sup> and Norman Arnheim<sup>§</sup>

Institute of Biophysics, Johannes Kepler University, Linz, 4040 Austria, Computational Biology Group, Department of Oncology, University of Cambridge, Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, CB2 0RE England, and Molecular and Computational Biology Program, University of Southern California, Los Angeles, California 90089

The amplification of millions of single molecules in parallel can be performed on microscopic magnetic beads that are contained in aqueous compartments of an oil-buffer emulsion. These bead-emulsion amplification (BEA) reactions result in beads that are covered by almost-identical copies derived from a single template. The post-amplification analysis is performed using different fluorophore-labeled probes. We have identified BEA reaction conditions that efficiently produce longer amplicons of up to 450 base pairs. These conditions include the use of a Titanium Taq amplification system. Second, we explored alternate fluorophores coupled to probes for post-PCR DNA analysis. We demonstrate that four different Alexa fluorophores can be used simultaneously with extremely low crosstalk. Finally, we developed an allele-specific extension chemistry that is based on Alexa dyes to query individual nucleotides of the amplified material that is both highly efficient and specific.

The amplification of single DNA molecules was first described two decades ago<sup>1,2</sup> and was traditionally performed in a large volume format that limited the number of feasible reactions to a few thousand. New developments involving the compartmentalization of reactions in picoliter-sized vesicles overcome this limitation and enable the individual amplification of millions of single DNA molecules in parallel.<sup>3,4</sup> In this method, single DNA molecules are compartmentalized using emulsions that isolate individual micro-PCR reactors in an oil phase (reviewed in ref 5). (Here “PCR” denotes polymerase chain reaction.) Emulsions for PCR

can be made using different components, all of which contain an oil phase, a surfactant that stabilizes the emulsion through the heating and cooling cycles of PCR, and an aqueous phase containing the PCR reagents.<sup>6–9</sup> Paramagnetic beads 1–34  $\mu\text{m}$  in diameter are covered with one of the PCR primers and are added to the reaction to capture the amplicons in a compartment.<sup>6,8–15</sup> The ratio of magnetic beads to DNA templates is such that there is, on average, a single DNA template in any compartment that contains a magnetic bead. After bead-emulsion amplification (BEA), the beads are coated with amplicons derived from a single DNA molecule.

Different chemistries that involve chemiluminescence or fluorescence have been used to characterize the DNA on the bead after the amplification. Researchers have used pyrophosphorolysis,<sup>9</sup> ligation,<sup>8</sup> single base extensions with and without a prior rolling circle amplification step,<sup>11,12</sup> simple probe hybridization,<sup>6,10</sup> labeled primers,<sup>15</sup> or fluorescently tagged antibody binding.<sup>13,14</sup> The chemiluminescent or fluorescent signal emitted by the beads

\* To whom correspondence should be addressed. Phone: ++43-732-2468-1619. Fax: ++43-732-2468-9270. E-mail address: Irene.Tiemann@jku.at.

<sup>†</sup> Institute of Biophysics, Johannes Kepler University.

<sup>‡</sup> Computational Biology Group, Department of Oncology, University of Cambridge.

<sup>§</sup> Molecular and Computational Biology Program, University of Southern California.

- (1) Jeffreys, A. J.; Wilson, V.; Neumann, R.; Keyte, J. *Nucleic Acids Res.* **1988**, *16*, 10953–10971.
- (2) Li, H. H.; Gyllenstein, U. B.; Cui, X. F.; Saiki, R. K.; Erlich, H. A.; Arnheim, N. *Nature* **1988**, *335*, 414–417.
- (3) Tawfik, D. S.; Griffiths, A. D. *Nat. Biotechnol.* **1998**, *16*, 652–656.
- (4) Nakano, M.; Komatsu, J.; Matsuura, S.; Takashima, K.; Katsura, S.; Mizuno, A. *J. Biotechnol.* **2003**, *102*, 117–124.
- (5) Griffiths, A. D.; Tawfik, D. S. *Trends Biotechnol.* **2006**, *24*, 395–402.

- (6) Diehl, F.; Li, M.; He, Y.; Kinzler, K. W.; Vogelstein, B.; Dressman, D. *Nat. Methods* **2006**, *3*, 551–559.
- (7) Mitra, R. D.; Butty, V. L.; Shendure, J.; Williams, B. R.; Housman, D. E.; Church, G. M. *Proc. Natl. Acad. Sci., U.S.A.* **2003**, *100*, 5926–5931.
- (8) Shendure, J.; Porreca, G. J.; Reppas, N. B.; Lin, X.; McCutcheon, J. P.; Rosenbaum, A. M.; Wang, M. D.; Zhang, K.; Mitra, R. D.; Church, G. M. *Science* **2005**, *309*, 1728–1732.
- (9) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y. J.; Chen, Z.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Godwin, B. C.; He, W.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L.; Jarvie, T. P.; Jirage, K. B.; Kim, J. B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S. M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.; Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. F.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P.; Begley, R. F.; Rothberg, J. M. *Nature* **2005**, *437*, 376–380.
- (10) Dressman, D.; Yan, H.; Traverso, G.; Kinzler, K. W.; Vogelstein, B. *Proc. Natl. Acad. Sci., U.S.A.* **2003**, *100*, 8817–8822.
- (11) Diehl, F.; Li, M.; Dressman, D.; He, Y.; Shen, D.; Szabo, S.; Diaz, L. A., Jr.; Goodman, S. N.; David, K. A.; Juhl, H.; Kinzler, K. W.; Vogelstein, B. *Proc. Natl. Acad. Sci., U.S.A.* **2005**, *102*, 16368–16373.
- (12) Li, M.; Diehl, F.; Dressman, D.; Vogelstein, B.; Kinzler, K. W. *Nat. Methods* **2006**, *3*, 95–97.
- (13) Kojima, T.; Takei, Y.; Ohtsuka, M.; Kawarasaki, Y.; Yamane, T.; Nakano, H. *Nucleic Acids Res.* **2005**, *33*, e150.
- (14) Kojima, T.; Yamane, T.; Nakano, H. *J. Biosci. Bioeng.* **2006**, *101*, 440–444.
- (15) Kumaresan, P.; Yang, C. J.; Cronier, S. A.; Blazej, R. G.; Mathies, R. A. *Anal. Chem.* **2008**, *80*, 3522–3529.

is then recorded by flow cytometry<sup>6,10–15</sup> or scanning devices.<sup>8,9</sup> If serial analysis of the amplicon is required, the beads are arrayed and immobilized to maintain the spatial information of the bead and the DNA is analyzed through a series of washing and probing cycles.<sup>8,9</sup>

The amplification of single DNA molecules with the BEA technology has revolutionized sequencing technologies by replacing bacterial cloning and has decreased the cost and increased the throughput several orders of magnitude. Currently, two different commercial sequencing platforms rely on the BEA technology: 454 GS-FLX sequencing<sup>9</sup> and SOLiD Systems<sup>16</sup> (Applied Biosystems) pioneered by the work in ref 8 and reviewed in ref 17. This technology has also impacted different protocols to measure genetic events. Each single original molecule is isolated during its amplification, and the products on each bead are PCR descendents of that molecule. Thus, the composition of the original pool of template molecules can be accurately characterized by studying the population of product-containing beads. This has enabled the quantification of rare genetic events, such as mutations in cancer tissues<sup>10,11</sup> or mutations introduced by polymerases.<sup>12</sup> Finally, because all the sequences amplified in a pool are queried by the BEA technology, DNA binding sites for transcription factors can be identified, including under-represented sites.<sup>13,14</sup>

The BEA technology offers a quick and simple procedure to assay simultaneously millions of molecules; however, there are still several shortcomings in the current protocols. First, long templates are amplified very inefficiently and the length of the amplicon is restricted to 100 to 250 bp in size (where the unit “bp” denotes base pairs) when using bulk emulsifying procedures, thus limiting the utility of this approach for certain studies.<sup>8,9</sup> We have addressed this issue by testing different reaction conditions for several different polymerases and have identified one set that can produce amplicons of 450 bp. Second, the simultaneous use of four different fluorophores is currently restricted to dyes with rapid decay rates of the fluorescent signal and considerable crosstalk between fluorescent channels. We developed an assay based on probes coupled to Alexa fluorophores (Molecular Probes) to query the DNA amplified on the beads. The absorption spectra of different Alexa dyes are sufficiently distinct to be excited and detected with common excitation sources and filter sets with minimal crosstalk. Finally, we have developed a chemistry to query the amplified DNA on the beads based on allele-specific extensions that is highly specific, even in GC-rich regions.

## EXPERIMENTAL SECTION

**Bead-Emulsion Amplification (BEA).** DNA templates were produced by amplifying human genomic DNA obtained from Clontech and Coriell Cell Depositories. Emulsion PCR was performed as described in refs 6 and 8 with modifications (see the Supporting Information for more details). The beads were washed after the BEA and the DNA on the beads was denatured with 0.1 M NaOH. The beads were arrayed on a microscope slide and immobilized with an acrylamide gel using a protocol modified from ref 8.

**Labeling the Beads.** Allele-specific extensions were conducted by incubating beads in 50  $\mu\text{L}$  with a 1 $\times$  Titanium buffer, 2  $\mu\text{M}$  of each Alexa labeled extension probes (sequence information provided in the Supporting Information), 200  $\mu\text{M}$  dNTPs each, and 1  $\mu\text{L}$  of 50 $\times$  Titanium Taq polymerase. The reaction was incubated for 2 min at 95  $^{\circ}\text{C}$ , cooled for 5 min at 66  $^{\circ}\text{C}$ , heated to 75  $^{\circ}\text{C}$  and then washed with 1E buffer (10 mM Tris pH 7.5, 50 mM KCl, 2 mM EDTA, 0.01% Triton X-100), followed by two TE washes. Allele-specific extensions were conducted in solution or directly on the arrayed beads using a Hybriwell chamber (Grace Biolabs). Ligations were performed as described in ref 8 with modifications described in the Supporting Information.

**Amplifying Different Product Lengths.** To compare the efficiency of the different polymerase amplification systems, as a function of template length, beads were coupled to a 982 bp amplicon. For the emulsion PCR, the beads were added to an aqueous phase Mastermix and divided into three different aliquots prepared with one of the three polymerases and the respective reaction buffer recommended by the manufacturer. Each Mastermix was further divided into four aliquots and a different forward primer (see Supporting Information for sequence details) was added to each one (12 aliquots in total). The final reagent concentration for each of the amplification system was the following:

**Platinum**—20 mM Tris pH 8.4, 50 mM KCl, 8 mM  $\text{MgCl}_2$ , 1 mM dNTPs, 0.05  $\mu\text{M}$  F-primer, 0.1  $\mu\text{M}$  R-primer, 0.3 U/ $\mu\text{L}$  Platinum Taq DNA Polymerase (Invitrogen);

**HiFi**—60 mM Tris- $\text{SO}_4$  pH 8.9, 18 mM  $(\text{NH}_4)_2\text{SO}_4$ , 8 mM  $\text{MgSO}_4$ , 1 mM dNTPs, 0.05  $\mu\text{M}$  F-primer, 0.1  $\mu\text{M}$  R-primer, 0.3 U/ $\mu\text{L}$  Platinum Taq DNA Polymerase High Fidelity (Invitrogen);

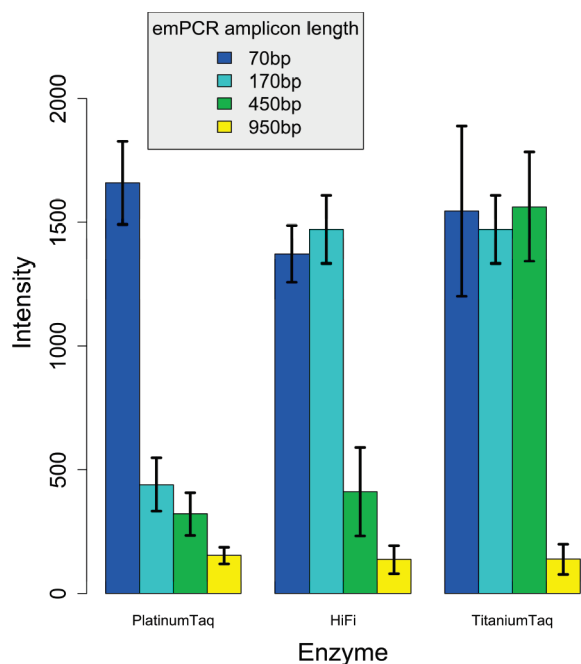
**Titanium**—40 mM Tricine pH 8.0; 16 mM KCl; 3.75  $\mu\text{g}/\text{mL}$  BSA, 11.5 mM  $\text{MgCl}_2$ ; 1 mM dNTPs; 0.05  $\mu\text{M}$  F-primer; 0.1  $\mu\text{M}$  R-primer; 0.06X/ $\mu\text{L}$  Titanium Taq Polymerase (Clontech).

The beads were washed and analyzed by allele-specific extensions of SNP1 and both SNP1 and SNP2, in the case of the 459 bp product analysis. To quantify the efficiency of the emulsion reaction using different polymerases, the intensities of beads captured in several hundred image frames was measured.

**Image Acquisition and Analysis.** The beads were visualized and scanned using an Eclipse TE2000E inverted microscope (Nikon) with a motorized stage at a 20 $\times$  magnification with a Plan Fluor objective. The light source was from a Nikon Intensilight Illumination System with a 120 W mercury short arc lamp. Each fluorophore was detected with individual filter cubes from Chroma: Chroma 41020 HQ480/20x; HQ510/20m Q495LP, Custom Alexa532 HQ530x; T550lpxr; ET565/25m, Chroma 31004 D560/40x; D630/60m; 595DCLP, and Chroma 41008 with a custom excitation filter: HQ630/20x; Q660LP; HQ700/75m. The images were taken with a 12-bit CCD camera. The monolayer formed by the beads allowed for scanning without an automatic focus function. For each experiment, one bright-field and two or four different fluorescent channel images were captured per raster position. Scans across 100–300 raster positions were performed to capture the majority of the arrayed beads. Images were acquired with Metamorph Software (MDS Analytical Technologies). Automated image analysis was performed using custom Matlab scripts. See Supporting Information for details.

(16) McKernan, K. J.; Blanchard, A. P.; Kotler, L.; Costa, G. U.S. Patent Application 20080003571, 2006.

(17) Shendure, J.; Ji, H. *Nat. Biotechnol.* **2008**, *26*, 1135–1145.



**Figure 1.** Plot representing the efficiency of three polymerase systems in the amplification of differently sized templates. Beads derived from a homozygous template were labeled using an allele-specific extension of SNP1. The data were calculated from  $\sim 300$  images. The mean intensity was calculated for the top 20% most intense beads (80th percentile) for each polymerase used to amplify a 70, 170, 450, and 960 bp product. Error bars denote one standard deviation from the mean.

## RESULTS

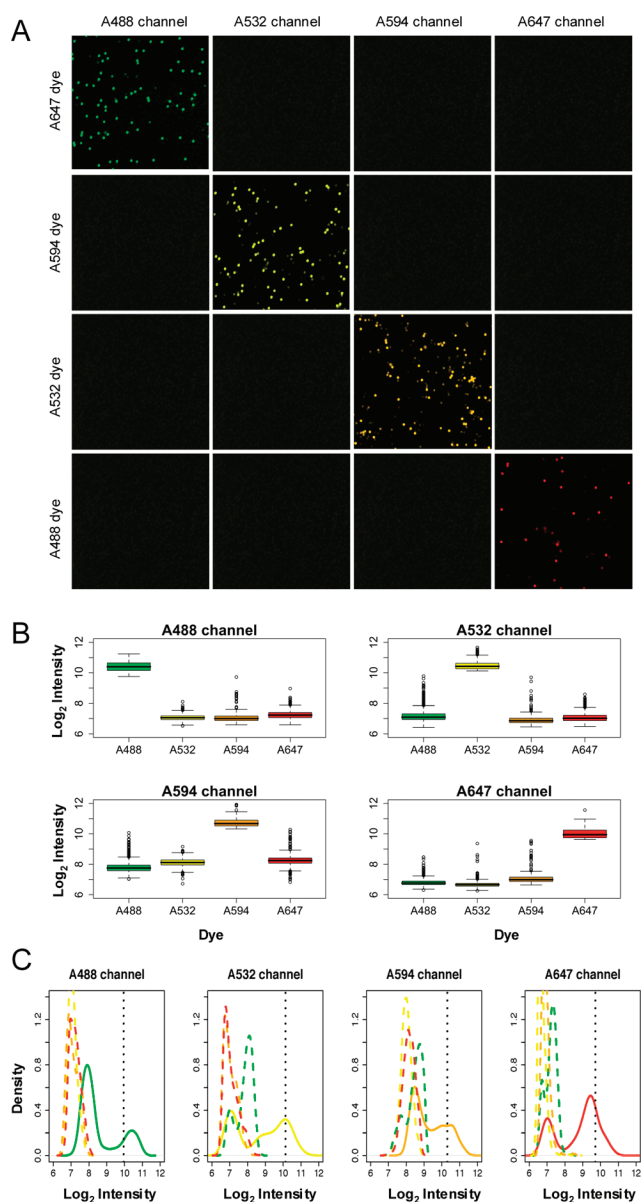
**Product Length.** We tested two different BEA conditions that have been published previously,<sup>6,8</sup> and we did not observe a variation in amplification yields or product length between the two conditions, despite the different reagent concentrations of the protocols (data not shown). We then tested whether using different polymerase amplification systems varying in salt, buffer, BSA, and enzyme composition rendered longer amplicons. Specifically, we used the Platinum, HiFi, and Titanium systems (see details in the Experimental Section). The performance of the different polymerase amplification systems was compared by amplifying the same beads coupled to a 982 bp template with four different forward primers that render either a 70, 170, 450, or a 950 bp product. Beads were queried by allele-specific extensions for SNP1. The reaction efficiency for different product lengths was measured as the average intensity of the most intense beads (80th percentile) across enzyme systems for different product lengths (see Figure 1). We confirmed that the 80th percentile is a conservative threshold for specifying the presence of a PCR product by comparing the intensity distribution of these beads with the distribution of the same beads in a different channel, which is equivalent to the signal of empty beads or background (see Figure S1 in the Supporting Information). As can be seen in Figure 1, the intensity of beads decreased significantly for the Platinum Taq system when the product length increased from 70 bp to 170 bp and longer. In contrast, both the HiFi and Titanium Taq amplification systems can produce high-intensity beads when amplifying 170 bp, but only the Titanium Taq system efficiently amplified products as long as 450 bp at the same intensity level as those obtained for a 70 bp or 170 bp product. None of the

enzyme systems could efficiently produce amplicons of 950 bp. We also added the same BSA concentration to the Platinum amplification system as that used for the Titanium system, and we failed to observe an increase in amplicon length (see Figure S2 in the Supporting Information). This suggests that the presence of BSA is not the only factor contributing to the higher amplification efficiency observed for the Titanium system.

To verify that the 450 bp product amplified by the Titanium Taq system is the full-length sequence and not an artifact, we amplified a different 459 bp sequence that could be probed at two different positions, located 35 and 436 nucleotides (SNP1 and SNP2) from the 5' end attached to the bead. Both positions were probed simultaneously, using different fluorophores, so that beads emitting a signal in more than one wavelength could be easily identified. The majority of single template beads ( $\sim 94.8\%$  of 11632 beads) were covered with the full length product and fluoresced with both fluorophores at SNP1 and SNP2. Approximately 5.2% of the total beads fluoresced with one color, of which 1.7% only emitted a signal with the probe specific for the 35 nucleotide site (SNP2) and 3.5% with the probe specific for the 436 nucleotide site (SNP1). These beads either contained a preliminary truncated product or some other PCR artifact recognized and extended by the probe.

**Imaging up to Four Fluorescent Colors Using Alexa Fluorophores.** We have explored fluorophores that give optimal signal intensity with minimal crosstalk between four different wavelengths using the standard light sources and filter sets described in the Experimental Section. A four-colored platform to analyze the amplified material on the magnetic beads has been published before using degenerate nonamer ligation probes coupled to either Texas Red, Cy5, Cy3, or FITC.<sup>8</sup> We obtained suboptimal results when using these dye combinations, because there was considerable crosstalk between the Cy3 and Texas Red signal, and the FITC dye exhibited rapid signal intensity decay (data not shown). We tested the performance of four Alexa fluorophores (Alexa 488, Alexa 532, Alexa 594, and Alexa 647) that have minimal overlap in their excitation and emission spectra. The fluorophores needed to be limited to the narrow spectral range of 488–647 nm, because the signal of the beads analyzed with dyes such as Cascade Blue, Alexa 350 (blue range), and Alexa 750 (far red range) was too weak to be detected with our standard equipment (data not shown).

Beads were prepared with DNA specific for each fluorescent probe and queried by allele-specific extensions. Figure 2A indicates that beads labeled with Alexa dyes can be easily differentiated, because each dye fluoresces brightly in only one channel, but not in the other channels. The average relative intensity of the fluorescent beads was measured to be  $\sim 1000$ – $1700$  (in a grayscale range of 0–4096), compared to a value of  $\sim 100$ – $300$  measured for the same beads in the other channels (see Table S1 in the Supporting Information). Because crosstalk between channels will most likely occur for the brightest beads, we plotted the signal of highly intense beads (80th percentile) across dyes for each channel in Figure 2B. The box plots in Figure 2B indicate that there is minimal crosstalk between channels, and the signal of the dye in its respective channel is significantly more intense ( $\sim 7$ – $10$  fold or 2.8–3.3 on the  $\log_2$  scale) than the signal from the other dyes imaged in that channel. Note that the bead

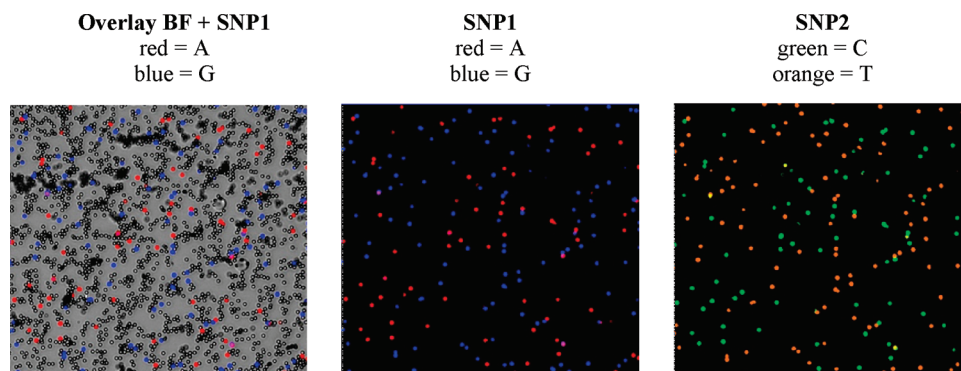


**Figure 2.** (A) Images of beads labeled with four different Alexa fluorophores acquired in all four fluorescent channels. Monocolored beads were produced using homozygous DNA specific for each fluorescent probe. The resulting beads were queried by allele-specific extension of probes coupled to one of the four fluorophores. (B) Boxplots of the signal emitted in each channel by high-intensity beads (80th percentile) for all four dyes imaged across 10 frames of beads prepared as described in panel A. The plots compare the  $\log_2$  signal in each channel measured for high-intensity beads labeled with one of the four different Alexa dyes. (C) Density plots of the  $\log_2$  signal intensities from beads imaged in the 10 frames used in panel B. Here, all beads (not only high-intensity beads) returned after mode-thresholding were plotted. The solid line represents the signal distribution for the dye in its respective channel (dyes are labeled as given in panel B), and the dashed lines represent the crosstalk signal from the other dyes in that same channel. The black dotted line indicates a threshold set at the 80th percentile of bead intensities for a given dye.

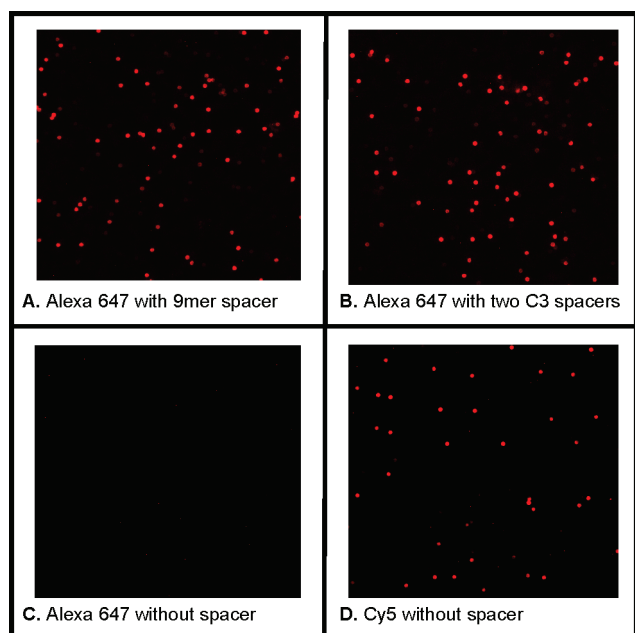
intensities measured in the A594 channel are higher overall, because of a greater exposure time. We also analyzed the intensity distribution of the beads in each channel (see Figure

2C). We found that all the dyes yield high signal-to-noise separation, as evidenced by the bimodal signal distributions for the dye in its respective channel, although the distributions and dynamic range vary across dyes (see Figure S3 in the Supporting Information for a view of Figures 2B and 2C on a per-dye basis). Because the majority of the beads do not have a PCR product, and most of these empty beads are filtered out during the image analysis stage, we expect to see a bimodal distribution consisting of a reduced population of empty beads or those with few PCR copies, and a separate population of bright beads with product. We also observed that the low-range intensity distribution for a given dye overlaps with the intensity distributions of the beads labeled with the other dyes imaged in that channel. This suggests that the crosstalk signal between channels differs little from the background signal, which is defined as that emitted by empty beads. Although it is difficult to distinguish between empty beads and those with only a few PCR products which emit little fluorescence, there is at least an  $\sim 5$ -fold difference between highly intense beads and beads that exhibit a low-intensity distribution (see Table S1 in the Supporting Information). Notably, only in the case of the A532 channel was the signal contribution from the A488 dye slightly larger than the contribution of the low-intensity beads captured in the A532 channel. Overall, based on these analyses, Alexa dyes have strong signal strength and up to four such fluorophores can be used simultaneously with minimal crosstalk.

**Allele-Specific Extensions and Ligations Combined with Alexa Dyes.** We also explored an alternative chemistry that was compatible with the Alexa dyes, based on allele-specific extensions of the Alexa labeled probes by a polymerase. This chemistry is particularly useful when the difference in the DNA sequence is only one nucleotide, as is the case for single nucleotide polymorphisms. The allele-specific probes have a different nucleotide at the 3' end and an identifier dye at the 5' end. Probes that perfectly match the queried DNA string are more likely to be extended by the polymerase lacking 3' to 5' exonuclease activity. A single nucleotide forming a mismatch at the 3' end of the probe is sufficient to reduce the extension efficiency of the probe. Non-extended probes have a lower melting temperature and can be washed away at increased temperature. The extensions can be performed by the Klenow fragment or its thermostable equivalent, Titanium Taq polymerase. We tested the efficiency of the Titanium Taq system for interrogating different alleles at two loci embedded in DNA sequences with 68% and 80% GC content, respectively. As shown in Figure 3, beads obtained from the amplification of a double heterozygote DNA exhibit the four different alleles, which are identified using allele-specific extensions with the corresponding labeled probes. Each bead with a product fluoresces only in one wavelength and there are few multicolored beads, which derive from amplifications beginning with more than one template molecule. There was no detectable cross-reaction of the probes with the mismatched template. Higher specificity is achieved with the extension reactions using the Titanium Taq system, because elevated temperatures can be used in comparison to Klenow extensions. In addition, the specificity of the extensions can be increased by shortening the length of the probes and including 3–4 phosphorothioate bonds at the 3' end of the probe. Interestingly, only the Titanium Taq or the Klenow fragment systems



**Figure 3.** Allele-specific extensions used to query the DNA in bead-emulsion amplification (BEA). Representative frames taken for two different SNPs depict pseudo-colored beads, which correspond to a double heterozygous sample. The overlay of the bright-field (BF) image and two fluorescent channels for SNP1 depict the fraction of beads containing a PCR product (~5%–10%). For both SNPs located in a ~70–80% GC-rich region, each bead hybridizes with only one of the two probes and, thus, fluoresces only in one channel. A few beads are multicolored and contain the product of more than one molecule.



**Figure 4.** Ligation after bead-emulsion amplification (BEA) of a 15mer anchor probe with a 10mer oligonucleotide coupled to different fluorophores and spacers. Panels A–C show the ligation of probes coupled to Alexa 647 separated from the 5' end of the ligation probe by one Sp9 (a 9mer triethylene glycol spacer), two C3 spacers, and no spacer, respectively. Panel D shows the ligation of a probe coupled to Cy5. The average fluorescent intensity from ~1000 visible beads estimated in panels A, B, and D is ~1200 (in a grayscale level of 0–4096).

efficiently extend our 15–18mer Alexa conjugated oligos. Other polymerase systems with 5' exonuclease activity, such as TaqGold, failed to extend these Alexa labeled oligos (data not shown), perhaps for stereochemical reasons, as shown by the Alexa ligation experiments described next.

We explored whether Alexa dyes are also compatible with the published ligation chemistry for degenerate nonamer ligation probes.<sup>8</sup> As shown in Figure 4, using the conditions published in ref 8, a probe of 10 nucleotides coupled to Cy5 will be ligated efficiently. The ligation is inhibited when the Cy5 dye is replaced by the spectrally equivalent Alexa 647 group. The reason for this is unclear, because Cy dyes and Alexa dyes do not differ much in

size.<sup>18</sup> However, the ligation of probes labeled with Alexa dyes is successful when a linker is added between the dye and the 5' end of the ligation probe (the probe that is ligated at its 3' end to the anchor probe carrying a 5' phosphate group). We tested a C3 hydrophilic spacer (phosphoramidite spacer) and Sp9 (triethylene glycol spacer). Our results show that adding two C3 groups or one Sp9 between the Alexa 647 dye and the oligo overcomes any inhibitory effect that this Alexa dye might have in the ligation of short probes (see Supporting Information for details on the ligation probes).

## DISCUSSION

In this report, we present several improvements in the amplification and analysis of single molecules using the BEA technology. A serious shortcoming of the BEA technology is the product length that can be amplified on paramagnetic beads.<sup>6,8,15</sup> Our approach produces BEA reactions containing amplicons up to 450 bp using a Titanium amplification system. To date, published protocols using a Platinum Taq polymerase system have efficiently produced 1–3  $\mu\text{m}$  beads covered with  $10^4$ – $10^5$  amplicons 100–150 bp (ideally 70–100 bp) in size.<sup>6,8,10–12</sup> Notably, these published optimized protocols with the highest yields use a wide range of salt, primer, dNTP, and concentrations of Platinum Taq, without an increase in product length (see Table 1). Another paper reported the amplification of  $10^7$  copies of slightly longer templates 200–250 bp in size, using a Titanium High-Fidelity system on 20- $\mu\text{m}$  beads.<sup>9</sup> Amplification on 34- $\mu\text{m}$  beads with a higher primer density contained in homogeneous droplet sizes created by microfluidics renders beads with  $10^7$  copies of ~600 bp products.<sup>15</sup> When amplifying products in emulsion without a solid phase, the template length produced can range from 1–2 kb (where “kb” denotes kilobases), using a system with the proofreading enzyme *Pfu* Turbo.<sup>19</sup> Reports on solid-phase PCR conducted on glass surfaces claim amplification of products ranging from 5.7 kb<sup>20</sup> to 16.5 kb.<sup>21</sup> It is striking that

(18) Jares-Erijman, E. A.; Jovin, T. M. *Nat. Biotechnol.* **2003**, *21*, 1387–1395.

(19) Williams, R.; Peisajovich, S. G.; Miller, O. J.; Magdassi, S.; Tawfik, D. S.; Griffiths, A. D. *Nat. Methods* **2006**, *3*, 545–550.

(20) Erdogan, F.; Kirchner, R.; Mann, W.; Ropers, H. H.; Nuber, U. A. *Nucleic Acids Res.* **2001**, *29*, E36.

(21) von Nickisch-Roseneck, M.; Marschan, X.; Andresen, D.; Abraham, A.; Heise, C.; Bier, F. F. *Biosens. Bioelectron.* **2005**, *20*, 1491–1498.

**Table 1. Reaction Conditions Used in Different Bead-Emulsion Amplification (BEA) Protocols<sup>a</sup>**

ref	PCR components	BSA ( $\mu\text{g/mL}$ )	polymerase	amplicon length (bp)	emulsion droplet size	bead size ( $\mu\text{m}$ )
10	67 mM Tris-HCl (pH 8.8), 16.6 mM $(\text{NH}_4)_2\text{SO}_4$ , 10 mM $\beta$ -mercaptoethanol, 6.7 mM $\text{MgCl}_2$ , 1 mM dNTPs, 0.05 $\mu\text{M}$ F-primer; 25 $\mu\text{M}$ R-primer	none	0.15 U/ $\mu\text{L}$ Platinum Taq	189–239	5–7 $\mu\text{m}$ ; bulk emulsification	1
8	20 mM Tris (pH 8.4), 50 mM KCl, 18.8 mM $\text{MgCl}_2$ , 3.5 mM dNTPs, 0.05 $\mu\text{M}$ F-primer, 25 $\mu\text{M}$ R-primer	none	0.28 U/ $\mu\text{L}$ Platinum Taq	135–230	3–5 $\mu\text{m}$ ; bulk emulsification	1
9	60 mM Tris (pH 8.9); 18 mM $(\text{NH}_4)_2\text{SO}_4$ ; 2.5 mM $\text{MgSO}_4$ ; 0.01% Tween-80; 0.003 U/ $\mu\text{L}$ thermostable pyrophosphatase; 1 mM dNTP; 0.625 $\mu\text{M}$ F-primer; 0.039 $\mu\text{M}$ R-primer	1000	0.15 U/ $\mu\text{L}$ Platinum HiFi	~200–250	100–150 $\mu\text{m}$ ; bulk emulsification	25–36
6	67 mM Tris-HCl (pH 8.8), 16.6 mM $(\text{NH}_4)_2\text{SO}_4$ , 10 mM $\beta$ -mercaptoethanol, 1.7 mM $\text{MgCl}_2$ , 0.2 mM dNTPs, 0.05 $\mu\text{M}$ F-primer; 8 $\mu\text{M}$ R-primer	none	0.3 U/ $\mu\text{L}$ Platinum Taq	70–110	10–50 $\mu\text{m}$ ; bulk emulsification	1
15	10 mM Tris (8.3), 50 mM KCl; 1.5 mM $\text{MgCl}_2$ ; 0.2 mM dNTPs; 0.01% Tween-80; 0.04 $\mu\text{M}$ F-primer, 0.4 $\mu\text{M}$ R-primer	1000	0.4 U/ $\mu\text{L}$ AmpliTaq Gold	$\geq 600$	~167 $\mu\text{m}$ ; monodispersed emulsification	34
this study	40 mM Tricine (pH 8.0); 16 mM KCl; 11.5 mM $\text{MgCl}_2$ ; 1 mM dNTPs; 0.05 $\mu\text{M}$ F-primer; 9 $\mu\text{M}$ R-primer	3.75	0.06 X/ $\mu\text{L}$ Titanium Taq	450	10–50 $\mu\text{m}$ ; bulk emulsification	3

<sup>a</sup> The BEA conditions of the first four rows vary considerably in salt, primer, dNTP, and BSA concentration, as well as in the type of polymerase used, although the reported product length does not differ considerably. Row 5 summarizes the results from experiments using ~6000 droplets with constant compartment sizes that yielded products of up to 600 bp in size.

the amplification efficiency of longer templates in BEA is so low, compared to the other microscopic PCR reactions. It is also surprising that a system using a Titanium Taq polymerase, which has low processivity<sup>22</sup> and no proofreading activity—both of which are factors that influence normal PCR product length—produced longer products than the polymerase systems with proofreading activity (HiFi) and higher processivity (Platinum Taq). It has been suggested that, in BEA,<sup>6</sup> molecular crowding on the bead's surface might interfere with the amplification of longer products.<sup>23</sup> It is possible that the efficiency of a Titanium amplification system, or other systems with polymerases lacking 5' exonuclease activity, are less sensitive to such inhibitory conditions, an observation we already made with the extension of Alexa-coupled probes. We wish to emphasize that we cannot distinguish whether the polymerase alone or the combination of reaction conditions used in the Titanium amplification system is responsible for the longer amplicons. It has been reported that bovine serum albumin (BSA) stabilizes the emulsions, reducing the leaking of small components;<sup>24</sup> however, the presence of BSA in a particular enzyme system does not necessarily increase the product length, as we showed for the Platinum system. Other factors, such as the volume of the emulsion compartment or the size of the beads, might also play a role in the PCR efficiency of longer templates.<sup>15</sup> Clearly, much more must be

learned about the molecular mechanisms that govern the amplification efficiency during BEA.

This report on the increased product size is an improvement in the current BEA protocols, because the small product size that could be obtained previously was considered a limitation of the technology.<sup>6</sup> The use of monodispersed droplets with a constant, relatively large, reaction volume has allowed the amplification of ~600 bp products efficiently, but the creation of the emulsion droplets requires an elaborate microfluidics setup.<sup>15</sup> In addition, creating millions of these homogeneous droplets requires a long preparation time of several days, compared to the quick bulk emulsification procedure used in this study and reported previously.<sup>6,8–10,12</sup> Increasing the amplicon size will have many benefits for current applications that employ BEA. This has been demonstrated for 454 sequencing protocols that have yielded unprecedented throughput with read lengths of 400–500 bp obtained from products amplified on beads. However, there are no published details about the necessary reaction conditions to produce these longer reads.

Our report also addresses the use of Alexa fluorophores for the DNA analysis that have several advantages over the dyes used in published protocols. Alexa fluorophores produce very intense beads with a high signal-to-noise ratio. In addition, the absorption spectra of different Alexa conjugates are separate enough to be excited and detected with common excitation sources and filter sets and up to four different Alexa fluorophores can be used simultaneously in our system with minimal crosstalk. Indeed, producing highly intense beads with minimal crosstalk is impor-

(22) Barnes, W. M. *Gene* **1992**, *112*, 29–35.

(23) Mercier, J. F.; Slater, G. W. *Biophys. J.* **2005**, *89*, 32–42.

(24) Courtois, F.; Olguin, L. F.; Whyte, G.; Theberge, A. B.; Huck, W. T.; Hollfelder, F.; Abell, C. *Anal. Chem.* **2009**, *81*, 3008–3016.

tant to increase the amount of accurate reads. The insufficient discrimination of the respective base emission spectra and rapid decay in fluorescent intensity can contribute to base call errors, as reported for Solexa sequencing for which several means of correction have been proposed.<sup>25–27</sup> Other means of correction that can further increase the amount of usable reads pertain to potential dye biases that arise when using several fluorophores with different quantum yields, as is the case for the Alexa dyes.<sup>28,29</sup>

We also report that Alexa fluorophores are compatible with allele-specific extensions of labeled probes and ligation of short labeled oligos, two very simple yet powerful chemistries to analyze the amplified DNA on the beads. The allele-specific extension chemistry is specific enough to distinguish a single nucleotide difference in very high GC-rich DNA. In published reports, DNA amplified on beads has been analyzed by extension reactions where the emission of light is recorded with the release of inorganic pyrophosphate during each of the four nucleotide flows. This method is highly accurate, with the majority of errors caused by longer runs of homopolymers, but requires very specialized equipment to conduct the reactions and measurements.<sup>9</sup> Other systems rely on single base extensions performed with labeled dideoxynucleotides and are specific enough to differentiate between two different allelic states, but rolling circle amplification precludes the labeling step to enhance the signal strength.<sup>12</sup> An alternative to extensions with ddNTPs is the hybridization of labeled probes<sup>6,10</sup> with the drawback that hybridizations are not as specific as enzymatic reactions and require optimization, especially if the difference is a single nucleotide in a GC-rich DNA region. Another chemistry involves the ligation of degenerate nonamers where the queried nucleotide is identified by the specific fluorophore attached covalently to the ligation probe. Only perfectly matched probes are ligated to the anchor probe, making this system quite specific.<sup>8</sup> In comparison, the allele-specific extension of labeled probes reported in this work is as simple as the ligation chemistry but can yield improved specificity for difficult DNA templates, because the reactions can be performed at high temperature. Because the allele-specific chemistry requires a labeled probe for each different queried DNA sequence, this chemistry is ideally suited for applications that interrogate a few loci at high specificity.

## CONCLUSIONS

In this work, we present useful modifications to the existing protocols for the amplification of millions of single molecules in

parallel in bead-emulsion amplification (BEA) reactors. Single molecule amplification confers information about each molecule that can be recorded as a digital readout and used for the precise characterization of all the molecules in a pool. The advantages of single molecule amplification in macroscopic volumes has been applied for quantitative purposes in a limited number of cases,<sup>30–37</sup> but the unprecedented throughput of single molecule amplification in BEA is well-suited for a variety of applications other than sequencing, mutation detection, and characterization of transcription sites. Areas such as epigenetics and gene expression variation might also see some benefit from this technology. The data obtained with BEA may change our understanding of certain biological processes that are otherwise difficult to study.

## ACKNOWLEDGMENT

This work was supported by National Institutes of Health (through Grant Nos. RO1 GM36745 and P50 HG002790), Cancer Research UK (Grant No. C14303/A4830), and the Austrian Science Fund Lise-Meitner (No. M1097-B12). The authors thank Jeremy Edwards for his highly valuable training and advice throughout the project.

## SUPPORTING INFORMATION AVAILABLE

A detailed description of experimental setup to create the BEA reactors, a description of image analysis used to analyze labeled beads, and primer and probe information are provided. Average relative fluorescence measured for differently labeled beads across channels (Table S1). Boxplots and density plots of the log<sub>2</sub> intensities of beads amplified with various polymerases for different BEA amplicon lengths (Figure S1). Effect of bovine serum albumin (BSA) in the amplification of different sized templates by the platinum amplification system (Figure S2). Boxplots and density plots of the log<sub>2</sub> intensities of beads labeled with different dyes imaged in all four channels (Figure S3). This material is available free of charge via the Internet at <http://pubs.acs.org>.

Received for review March 27, 2009. Accepted May 18, 2009.

AC900633Y

- (25) Erlich, Y.; Mitra, P. P.; delaBastide, M.; McCombie, W. R.; Hannon, G. J. *Nat. Methods* **2008**, *5*, 679–682.
- (26) Rougemont, J.; Amzallag, A.; Iseli, C.; Farinelli, L.; Xenarios, I.; Naef, F. *BMC Bioinform.* **2008**, *9*, 431.
- (27) Whiteford, N.; Skelly, T.; Curtis, C.; Ritchie, M.; Lohr, A.; Wait Zaranek, A.; Brown, C. Submitted for publication.
- (28) Fluorescence quantum yields (QY) and lifetimes ( $\tau$ ) for Alexa Fluor dyes; Table 1.5; Available via the Internet at <http://www.invitrogen.com/site/us/en/home/References/Molecular-Probes-The-Handbook/tables/Fluorescence-quantum-yields-and-lifetimes-for-Alexa-Fluor-dyes.html>.
- (29) Panchuk-Voloshina, N.; Haugland, R. P.; Bishop-Stewart, J.; Bhalgat, M. K.; Millard, P. J.; Mao, F.; Leung, W. Y.; Haugland, R. P. *J. Histochem. Cytochem.* **1999**, *47*, 1179–1188.

- (30) Hubert, R.; MacDonald, M.; Gusella, J.; Arnheim, N. *Nat. Genet.* **1994**, *7*, 420–424.
- (31) Jeffreys, A. J.; Murray, J.; Neumann, R. *Mol. Cell* **1998**, *2*, 267–273.
- (32) Buard, J.; Shone, A. C.; Jeffreys, A. J. *Am. J. Hum. Genet.* **2000**, *67*, 333–344.
- (33) Lien, S.; Szyda, J.; Schechinger, B.; Rappold, G.; Arnheim, N. *Am. J. Hum. Genet.* **2000**, *66*, 557–566.
- (34) Cullen, M.; Perfetto, S. P.; Klitz, W.; Nelson, G.; Carrington, M. *Am. J. Hum. Genet.* **2002**, *71*, 759–776.
- (35) Leeflang, E. P.; Tavare, S.; Marjoram, P.; Neal, C. O.; Srinidhi, J.; MacFarlane, H.; MacDonald, M. E.; Gusella, J. F.; de Young, M.; Wexler, N. S.; Arnheim, N. *Hum. Mol. Genet.* **1999**, *8*, 173–183.
- (36) Shinde, D.; Lai, Y.; Sun, F.; Arnheim, N. *Nucleic Acids Res.* **2003**, *31*, 974–980.
- (37) Brohede, J.; Arnheim, N.; Ellegren, H. *Mol. Biol. Evol.* **2004**, *21*, 58–64.