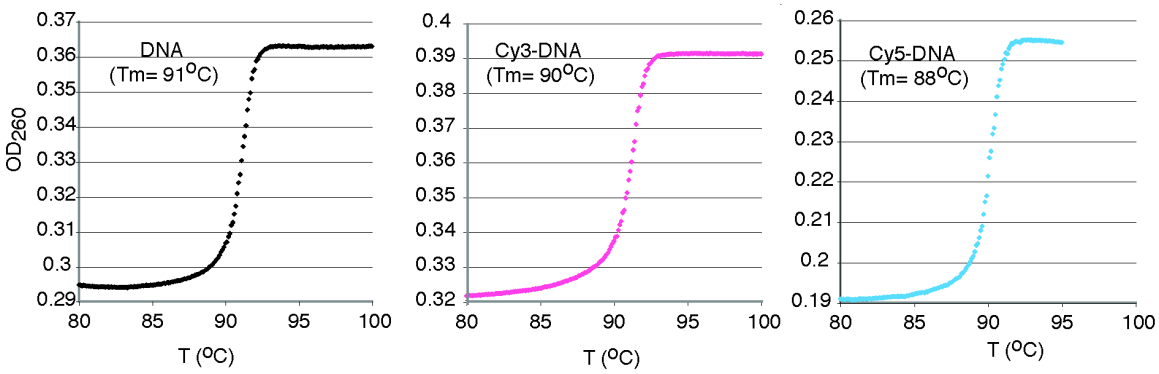
Supporting information

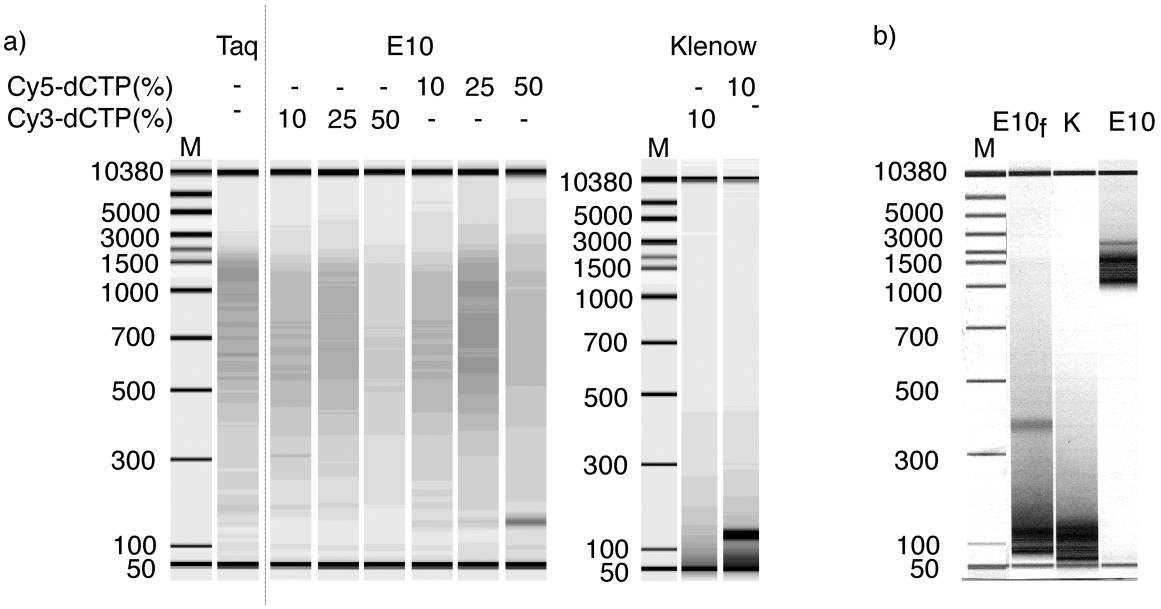
Synthesis and replication of highly substituted DNA with altered physico-chemical properties by an evolved polymerase

N. Ramsay, A.S. Jemth, A. Brown, N. Crampton, P. Dear and P. Holliger

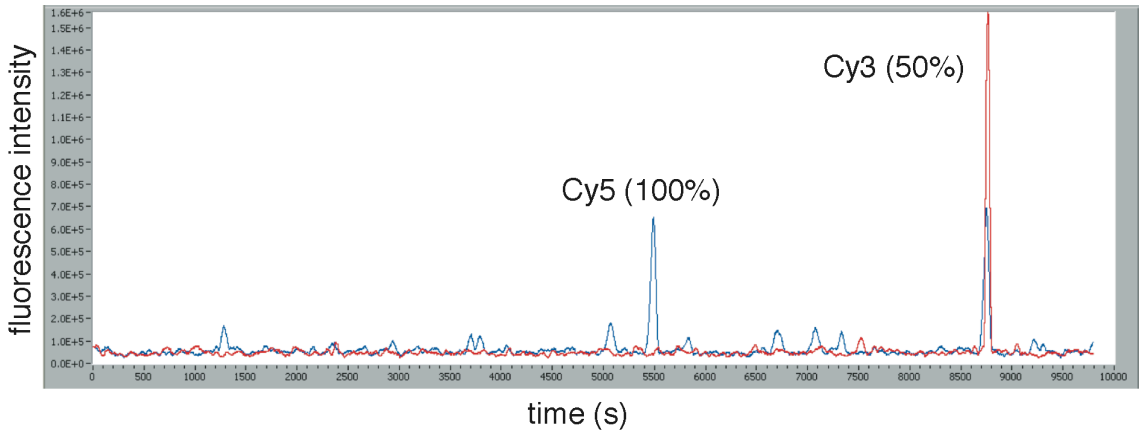
1. Supporting Figures



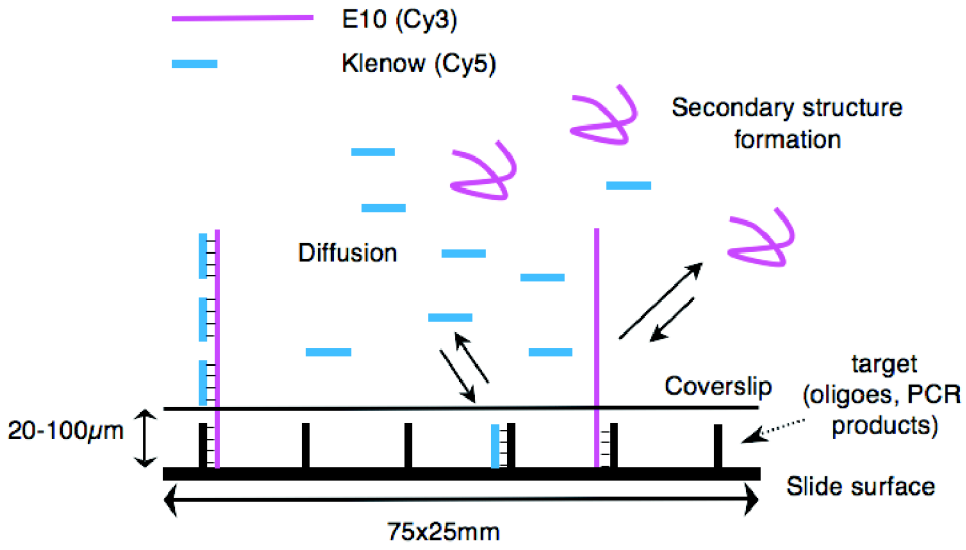
Supporting Figure 1: Melting curves. Melting curves for DNA (left panel), Cy3- (middle panel) and Cy5-DNA (right panel). CyDNA displays somewhat lower melting temperatures and slightly less cooperative melting.



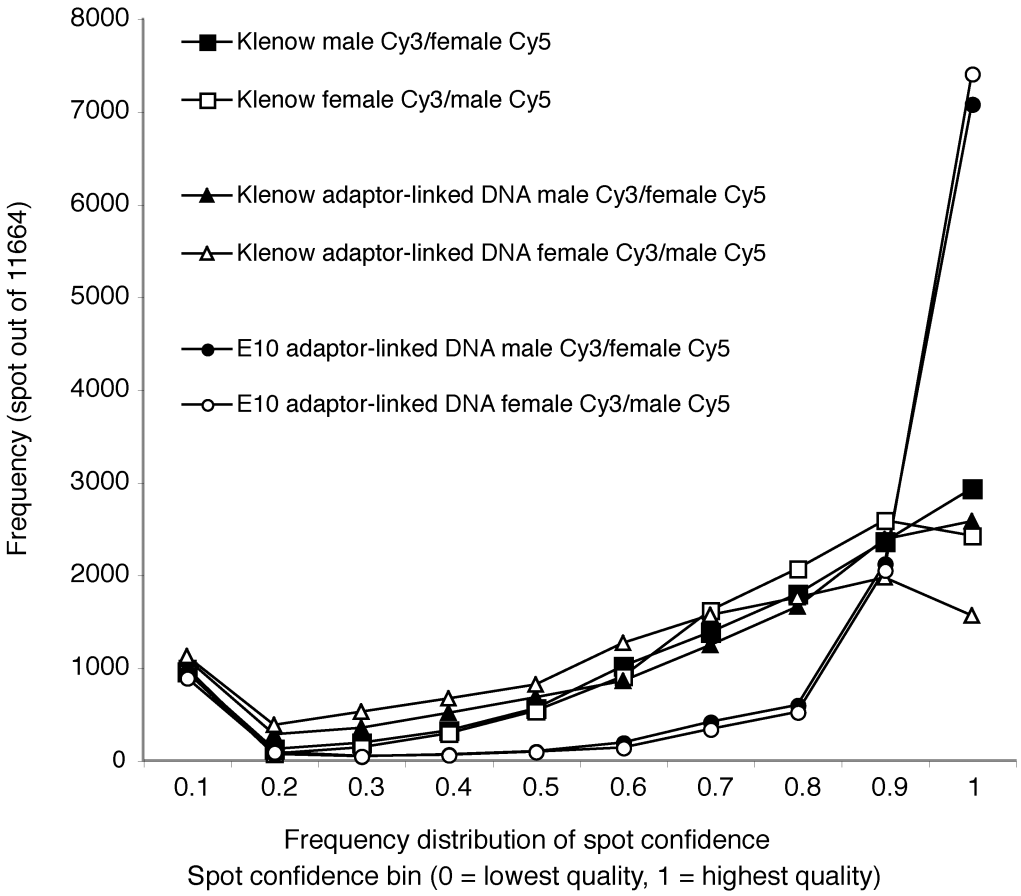
Supporting information Figure 2: a) Size distribution of randomly primed fluorescent probe synthesis (from human genomic DNA) by E10 (left panel) and Klenow (right panel) as resolved on an Agilent Bioanalyzer system. For comparison, the size distribution of randomly primed DNA synthesis by Taq is shown. Even at 50% Cy5-dCTP the average size of E10 probes is still is 0.5-1kb, while the majority of Klenow probes is < 0.1kb. The dark band at ca. 0.1kb in both the E10 and Klenow Cy5 reactions is an artefact caused by the slow mobility of unincorporated Cy5-dCTP. M is a 0.05-10.4kb size marker b) Klenow labelling of human genomic DNA (K) results in an average labelled DNA length of 50 – 150bp irrespective of the length of the DNA template, while E10 labelling of human genomic DNA (E10) yields labelled DNA of approximately 10x larger size (1-1.5kb). Incorporation of 10% dUTP and fragmentation with UDG and APE-1 ($E10_f$) yields fragments of similar yet still slightly larger size to Klenow labelling.



Supporting Figure 3: Single-molecule detection of Cy-DNA. A solution of 5pg/ul each of 0.27kb PCR products labeled 50% Cy3-DNA (50% of dC replaced by Cy3-dC, 51 Cy3 dyes / PCR product) and of 100% Cy5-DNA (102 Cy5 dyes / PCR product) were injected and pumped at 30μl/hr in a fused silica capillary in a microfluidic device past a sensitive fluorescence detection system calibrated using single-particle detection of quantum dots. Fluorescence was detected at a fixed point inside the lumen of the capillary. The representative output shows fluorescence (photon counts per second – vertical axis) at 670nm (Cy5 maximum – blue trace) and 570nm (Cy3 maximum – red trace) recorded in consecutive 20μsec periods (numbered along horizontal axis) and averaged over a sliding window (centre-weighted) of 40 periods. A representative section of the output, spanning a total of 0.2s, is shown. Since labeled molecules may pass through either the centre or the periphery of the Gaussian detection volume, peak heights are expected to be variable in this system. Also, the overlaps between the Cy3 and Cy5 emission spectra, and the finite bandwidth of the detection filters, means that Cy3 emission causes some signal in the Cy5 channel (though not *vice versa*). Nevertheless, the presence of mixed Cy3- / Cy5-DNA aggregates could be detected as coincident Cy3 and Cy5 peaks with altered signal ratios. This trace depicts strong detection of one Cy5-labelled molecule (at approximately interval 5500), and of one Cy3-labelled molecule (at approximately interval 8700); smaller peaks are visible, particularly for Cy5 – these are believed to represent labeled molecules passing through the edges of the Gaussian detection volume. The frequency and fluorescence signal of events was consistent with the detection of single Cy-DNA molecules.



Supporting Figure 4: Schematic of array probe hybridization. Two effects may contribute to superior signal derived from shorter array probes in array cohybridization experiments as shown here for short (Klenow, blue) and long probes (E10, magenta). The first contribution comes from faster diffusion of smaller probes and their smaller propensity to form secondary structures that need to equilibrate with hybridization to the array target. The second contribution comes from the binding of long probes, which by themselves create new hybridization sites for binding of more short probes (generated from random priming) (see left of figure) thus skewing the number of target sites on each spot towards the shorter probes.



Supporting Figure 5: Spot confidence scores in aCGH. E10 probes display a significantly increased signal to background ratio in all measurements (Supporting information Table 3), which is also reflected in superior spot confidences. Spot confidence is determined by whether a spot is detected in both the Cy3 and Cy5 channel, and whether those 2 spots correspond (i.e. do the signals overlap). Confidence is therefore largely determined by signal. Low confidence data (less than 0.3) would typically be excluded from an analysis.

Supporting Tables

Supporting Table 1 : AFM data

	$\langle R \rangle_{Measured}$ (nm)[1]	$\langle R \rangle^2_{Measured}$ (nm)	$\langle R \rangle^2_{Predicted}$ (nm)
Unmodified	62.4 ± 2.6, n=61	3800 ± 324	4033.0
Cy5 Modified	64.2 ± 2.4, n=61	4100 ± 308	3936.0

[1]<R> is the mean end-to-end distance (i.e. a straight line from on end to the other in a straight line, rather than following the DNA backbone). $\langle R \rangle^2_{Measured}$ is this value squared. $\langle R \rangle^2_{Predicted}$ is the value predicted for a DNA molecule of persistence length 53 nm (B-form DNA) and the contour lengths from the upper table.

Supporting Table 2: CGH array data summary (Fig. 9b).

Signal				
		Klenow genomic DNA	Klenow labeling adaptor-PCR DNA	E10 labeling adaptor-PCR DNA
	Cy3 female	2477.4	2204.49	5229.19
	Cy5 male	1424.38	1170.46	5028.35
	Cy3 female	2807.16	974.14	5258.46
	Cy5 male	1609.47	572.58	4180.89

Supporting Table 3: CGH array data summary 2.

Signal to background ratio (SBR)				
		Klenow genomic DNA	Klenow labeling adaptor-PCR DNA	E10 labeling adaptor-PCR DNA
	Cy3 female	2.53	2.12	4.04
	Cy5 male	0.47	0.45	1.2
	Cy3 female	2.68	1.88	4.3
	Cy5 male	0.51	0.48	1.06
Noise (standard deviation (SD) of autosome Log2ratio)				
		Klenow genomic DNA	Klenow labeling adaptor-PCR DNA	E10 labeling adaptor-PCR DNA[1]
	SD autosome Log2ratio	0.0736	0.1537	0.0993
Median X separation (χ) Log2 ratio[2]				
		Klenow genomic DNA	Klenow linker-PCR DNA	E10 linker-PCR DNA
	χ (Log2ratio)	0.648	0.7875	0.769
Median Standard Deviation Between Replicates				
		Klenow genomic DNA	Klenow linker-PCR DNA	E10 linker-PCR DNA
	female Cy3/male Cy5	0.09	0.09	0.07
	male Cy3/female Cy5	0.08	0.1	0.07

[1]To distinguish variability associated with E10 labelling from variability due to sample workup, E10 labelling is compared to both standard Klenow genomic DNA labelling and labelling of adaptor-PCR DNA; [2]Median X separation is a measure of how accurately the array measures copy number gains / losses. Chromosome X is a copy number gain of 2:1 for the female compared to the male, and chromosome Y is a copy number loss of 0:1 for the female. Ideal Log2ratios for X (and Y) for female vs. male DNA would be 0.7 (and -0.9)

Supporting Materials & Methods

1. 1 DNA manipulation

To allow the cloning of the Pfu sequence from pETPfu[1] the NdeI site in the vector backbone of pASKDpo4 (pASK75[2] with the Dpo4 gene cloned Xba/Sal from pET22b-Dpo4[3]) was removed by site-directed-mutagenesis (5´-GTT GTT TTT CTA ATC CGC ATG TGA TCA ATT CAA GGC CG-3´ and 5´-CGG CCT TGA ATT GAT CAC ATG CGG ATT AGA AAA ACA AC-3´) to give pASKDpo4ΔNde. The full length Pfu sequence was PCR amplified from pETpfu (5´-CAG GAA ACA GCT ATG ACC ATA TGA TTT TAG ATG TGG ATT ACA TAA CTG-3´ and 5´-AGT AGC GGC GTC GAC TTA GGA TTT TTT AAT GTT AAG CCA GGA AG-3´) and cloned NdeI/SalI into pASKDpo4ΔNde thereby generating pASKPfu. The 3´-5´ exonuclease function was disabled by mutation of D141A and E143A in the exonuclease motif I[4] by site-directed-mutagenesis using primers 5´-CTA AAG ATT CTT GCC TTC GCG ATC GCG ACC CTC TAT CAC GAA GGA GAA-3´and 5´-TTC TCC TTC GTG ATA GAG GGT CGC GAT CGC GAA GGC AAG AAT CTT TAG-3´ thereby generating pASKpfuexo$^-$. Silent mutations were introduced into the Pfu gene to remove the BsaI site at 1636bp (5´-CTA CAT TGA CAC TGA TGG CCT CTA TGC AAC TAT CCC A-3´ and 5´-TGG GAT AGT TGC ATA GAG GCC ATC AGT GTC AAT GTA G-3´; generating pASKpfuexo$^-$1), the XbaI site at 1683bp (5´-GAG GAA ATA AAG AAA AAG GCT CTC GAA TTT GTA AAA TAC ATA AAT TC-3´ and 5´-GAA TTT ATG TAT TTT ACA AAT TCG AGA GCC TTT TTC TTT ATT TCC TC-3´; generating pASKpfuexo$^-$2) and the BamHI site at 606bp (5´-CAG GAT TAT CAG GGA GAA GGA CCC TGA CAT TAT AGT TAC TTA T-3´ and 5´-ATA AGT AAC TAT AAT GTC AGG GTC CTT CTC CCT GAT AAT CCT G-3´; generating pASKpfuexo$^-$3). A silent mutation was introduced by site-directed-mutagenesis into the Pfu gene to include a unique BamHI (1416bp) restriction enzyme site (5´-GAC AAA AAT GAA GGA AAC TCA GGA TCC TAT AGA AAA AAT ACT CC-3´ and 5´-GGA GTA TTT TTT CTA TAG GAT CCT GAG TTT CCT TCA TTT TTG TC-3´) thereby generating pASKpfuexo$^-$4. The uracil stalling function[5] of the Pfu enzyme was removed by site-directed-mutagenesis (5´-GAA ACT TTA TTT GGA ACA TCC TCA GGA TCA GCC CAC TAT TAG AGA AAA AG-3´ and 5´-CTT TTT CTC TAA TAG TGG GCT GAT CCT GAG GAT GTT CCA AAT AAA GTT TC-3´) thereby generating pASKpfuexo$^-$5. Pfuexo-5 (i.e. Pfu: V93Q, D141A,

E143A) is the starting point for library construction and directed evolution and is referred to as wild-type Pfuexo- in the manuscript. To facilitate cloning of A-C motif library selections a silent mutation was introduced by site-directed-mutagenesis into the Pfu gene to include a unique silent XhoI (1726bp) restriction enzyme site and pASKpfuexo$^-$5 as a template (5´-CAA AGC TCC CTG GAC TGC TCG AGC TTG AAT ATG AAG GG-3´ and 5´-CCC TTC ATA TTC AAG CTC GAG CAG TCC AGG GAG CTT TG-3´) thereby generating pASKpfuexo$^-$6.

1.2 Library construction

pASKpfuexo$^-$5 was purified using a QIAprep Spin Miniprep kit (Qiagen) and used as a template in iPCR library construction. Briefly iPCR reactions were hot-started by the addition of 3.5U of Expand High Fidelity polymerase (Roche) to a PCR mix [10ng of plasmid template, primers (0.4μM), dNTPs (0.2mM), in 1x High Fidelity Expand buffer with $MgCl_2$ (Roche)]. Primers including diversity were PAGE purified. Reactions were thermocycled [94ºC for 4 min; 19 times (94ºC for 20 sec, 65ºC for 20 sec - 1ºC/cycle, and 68ºC for 10 min); 15 times (94ºC for 20 sec, 50ºC for 20 sec, 68ºC for 10 min); 15 times (94ºC for 20 sec, 50ºC for 20 sec, 68ºC for 10 min + 15 sec/cycle)]. iPCR reactions were purified with a QIAquick PCR purification kit (Qiagen) and eluted in 50μL of H_2O. Purified DNA was restriction digested overnight with DpnI (New England Biolabs) to remove the input plasmid template. Restricted DNA samples were purified with a QIAquick PCR purification kit (Qiagen) and eluted in 50μL of H_2O. Purified DNA was restriction digested overnight with BsaI (New England Biolabs) after which the reactions were purified with a QIAquick PCR purification kit (Qiagen) and eluted in 50μL of H_2O. 150μL of the restriction digested DNA was ligated overnight at 16ºC with 1200U T4 DNA ligase (New England Biolabs) in 1x T4 DNA ligase buffer (New England Biolabs) in a final volume of 200μL. Ligation reactions were electroporated into *E. coli* Ace6 or TG1^{TR} cells which were plated onto 2x TY/Amp 0.1mg/mL plates.

The Pfu exo$^-$ A motif library was generated using pASKpfuexo$^-$2 as template and primers 5´-GAG TAG GTC TCT ACC CCT TTT CTG GCT CTT TAA CG-3´ and 5´-GGA AAG GTC TCA GGG TTG TGG RAM RRC MTS RYY TMT CTA GAT TTT AGA GCC CTA TAT CCC TCG ATT ATA RTT ACC CAC AAT G-3´ the underlined

sequence is 90% wild-type and 10% mutation (i.e. G is G in 36 of 40 oligonucleotides and A, C, or T in 4/40). The library was transformed into electrocompetent *E. coli* Ace6 (MC1061, *endA*).

The Pfu exo$^-$ A-C motif library was prepared using the four best clones selected with Cy5-dCTP from the Pfu A motif library (23 (N400D, I401L, R407I), AH12 (E399D, N400G, I401L, V402A, R407I, Q572H), 55 (N400G, R407I) and in particular 15 (V337I, E399D, N400G, R407I)) as template and primers 5´-GGA AAG GTC TCT ATC CCA GGA GGA GAA AGT GAG GAG ATC AAG AAA AAG GCT CTA GA-3´ and 5´-GAG TAG GTC TCA GGA TAG TTG CGW GAR GAC CAT CAG TGT CAA TGT AGA GGA CTT TAA ATC CAA ACT TTT CTT C-3´ the underlined sequence is 90% wild-type 10% other sequence. The library was transformed into electrocompetent *E. coli* Ace6 (MC1061, *endA*) (library size $2x10^8$ cfu). Plasmid DNA was isolated from the Pfu exo$^-$ A-C rmotif library in *E. coli* Ace6 and transformed into electrocompetent *E. coli* TG1^{TR}. The library in *E. coli* TG1^{TR} was used in selections.

1.3 spCSR Selections

Principle of spCSR[6] (Fig. 1a)

Diversified polymerase genes are cloned and transformed into bacterial cell. Cells harbouring both expressed polymerase protein and their encoding genes are compartmentalized in a heat-stable water-in-oil emulsion[7, 8] together with primers and dNTPs such that each aqueous compartment (on average) contains a single bacterial cell. PCR thermocycling lyses the cells releasing both polymerases and their encoding genes into the emulsion compartments, where self-replication takes place. As individual compartments cannot exchange either DNA or protein molecules, the degree of amplification of a polymerase gene is proportional to the activity of the encoded polymerase. Here, only polymerases capable of utilizing Cy5-dCTP (instead of dCTP) are able to replicate, while those that cannot disappear from the gene pool. Self-replication is focused on the short-segment (a "patch") of the polymerase gene, which had been diversified (Fig. 2). After spCSR selection, the emulsion is broken, polymerase gene segments are isolated and either cloned for screening or for another round of selection.

Modifications to the previously described CSR and spCSR protocols[6, 8, 9] were required to enable selection for Pfu variants and especially variants of Pfu able to incorporate labeled nucleotide analogues. The aqueous phase of the emulsion had to be modified to include primers (1μM), 0.1mM each of dATP, dTTP, dGTP and Cy5-dCTP (GE Healthcare), RNase (10 μg/mL), glycerol (10% v/v), formamide (1% v/v), DTT (1mM) and 1x Pfu buffer (Stratagene).

For spCSR selection from the Pfuexo$^-$ A motif library Ace6 cells were induced for protein expression and induced cells were emulsified as described[8, 9] with the modified aqueous phase described above and primers 5´-CAG GAA ACA GCT ATG ACA AAC GGG AAA GAA TTC CTT CCA ATG G-3´ and 5´-GTA AAA CGA CGG CCA GTA CCT CTA TAG GAT CCT GAG TTT CCT TC-3´. The emulsions were thermocycled 94ºC for 5 min; 20 times at 94ºC for 30 sec, 50ºC for 1 min, and 72ºC for 18 min followed by a final elongation step at 65ºC for 10 min. Emulsions were extracted with 2 volumes of diethylether and purified using a PCR purification kit (Qiagen) including an additional wash step with 750μL 35% guanidium hydrochloride. Purified products were eluted in 50μL elution buffer. To remove parental plasmid DNA and primers 7μL of column eluate was digested with DpnI and ExoSAP-IT (USB). 2μL of DpnI and ExoSAP-IT treated sample was reamplified with a SuperTaq (HT Biotechnology)/PfuTurbo (Stratagene) blend (68 parts Taq: 8 parts PfuTurbo; 1μL per reaction) in 1x Taq buffer (HT Biotech) with primers 5´-CAG GAA ACA GCT ATG ACA AA-3´ and 5´-GTA AAA CGA CGG CCA GTA CCC TTT TCT TCG AGC TCC TTC CAT AC-3´ (1μM each), 1% (v/v) formamide and dNTPs (0.5mM each) and the cycling conditions: 94ºC 2 min; 31 times at 94ºC for 30 sec, 54ºC for 1 min, and 72ºC for 3 min followed by a final elongation step at 65ºC for 10 min. Reamplified selection products were digested with EcoRI and BamHI and subcloned into pASKpfuexo$^-$4 before transformation into electrocompetent *E. coli* Ace6.

spCSR Cy5-dCTP selections from the Pfuexo$^-$ A-C motif library (94ºC for 5 min; 20 times at 94ºC for 30 sec, 50ºC for 1 min, and 72ºC for 10 min) were carried out as described above using primers 5´-CAG GAA ACA GCT ATG ACA AAC GGG AAA GAA TTC CTT CCA ATG G-3´ and 5´-GTA AAA CGA CGG CCA GTA CCA TAT

TCA AGC TCG AGC AGT CCA GGG AG-3´. The aqueous phase was extracted as described above and purified selection products were reamplified by PCR (94ºC 2 min; 29 times at 94ºC for 30 sec, 54ºC for 1 min, and 72ºC for 1 min 30 sec) using primers 5´-CAG GAA ACA GCT ATG ACA AA -3´ and 5´-GTA AAA CGA CGG CCA GTA CCA TAT TCA AGC TCG AGC AGT CCA GGG AG -3´. Reamplified selection products were digested with EcoRI and XhoI and subcloned into pASKpfuexo$^-$6 and transformed into electrocompetent *E. coli* Ace6.

1.4 Polymerase ELISA of Cy5-dCTP selected Pfu variants

In order to rank the Cy5-dCTP selected Pfu variants their ability to incorporate Cy-dye labelled dCTP was assessed using an ELISA extension assay[6]. Briefly cells expressing Pfu or selected Pfu variants were washed twice in 1x Pfu buffer (Stratagene Ltd) and concentrated 10 fold in 1x Pfu buffer. Cells were lyzed by incubation at 85ºC for 10 min and cell debris pelleted by centrifugation at 2,000x g for 10 min. Cleared lysates were activity normalized to 0.1U of Pfuexo- in an ELISA extension reaction with primer 5´-GAC TTC CTT GCC TGC TCG TCG TCG GCA TCC GTC GCG ACC ACG TT*T* TTC GTG GTC GCG ACG GAT GCC G-3´ where *T* = Biotin-dUTP. Reactions were performed in 1x Pfu buffer (Stratagene) and included dATP (10μM), dGTP (10μM), dTTP (9μM), digoxigenin-11-dUTP (1μM; Perkin Elmer), dCTP (10μM), primer (1μM), 9μL H_2O, 2μL cleared cell lysate and incubated 94ºC for 5 min, 50ºC for 5 min, 72ºC for 5 min. The extension reactions (5µL) were captured with StreptaWell High Bind microplates (Roche) via the internal biotin in the primer. The primer requires that 20 consecutive nucleotides be incorporated prior to the insertion of digoxigenin-11-dUTP that is assayed for using anti-DIG antibody-peroxidase conjugate (Roche) according to standard ELISA protocols.

Activity normalized lysates (0.1U) were screened for the ability to incorporate either Cy5-dCTP or Cy3-dCTP with primer 5´-TAG CTA CCA <u>GGG GGG GG</u>C TCC GGC TTC CGT CGC GAC CAC GTT *T*TT CGT GGT CGC GAC GGA AGC CG-3´ where *T* = biotin dUTP as described above in an ELISA extension reaction where 100% of the dCTP was replaced by either Cy3-dCTP or Cy5-dCTP. The G_8 (underlined) template requires that 8 consecutive Cy3-dCTP or Cy5-dCTP nucleotides be incorporated prior to

the insertion of digoxigenin-11-dUTP that is assayed for using anti-DIG antibody-peroxidase conjugate (Roche) according to standard ELISA protocols.

1.5 Polymerase purification

Plasmid constructs or libraries were transformed into *E. coli* Ace6 or TG1^{TR} and expressed as described[8, 9]. Briefly, transformed Ace6 cells are grown overnight at 37°C in 2x TY, 0.1 mg/mL ampicillin. For expression overnight cultures were diluted 1:50 in 2x TY, 0.1 mg/mL ampicillin and grown to an OD_{595} of 0.6 at 37°C. Protein expression was induced by the addition of anhydrotetracycline (0.4 μg/mL). Protein expression was induced for 6 hours at 37°C. Cells were harvested by centrifugation, resuspended in 20mL (per litre of culture) of buffer A (50mM Tris pH8.0, 1% glucose, 1mM EDTA), Buffer B (10mM Tris pH8.0, 50mM KCl, 1mM EDTA, 0.5% NP40) was added to a final volume of 50mL and cells were lysed for 30 min at 75°C. Debris was pelleted by centrifugation and the NaCl was added to the supernatant to 0.25M final concentration. Then neutralized polyethyleneamine (PEI) was added to a final concentration of 0.1% v/v and precipitate pelleted by centrifugation. The cleared supernatant was diluted 5x with 20mM Tris pH7.5 and loaded onto a 6/10 Hi-Prep Heparin FF Column (Pharmacia) equilibrated with Column running buffer (CRB) (20mM Tris pH7.5, 50mM NaCl, 10% glycerol). The column was washed with 150mL of CRB and bound polymerase eluted with a NaCl gradient from 0 to 1M. Pfu eluted between 0.2M-0.3M NaCl, 10% glycerol. Eluted polymerase samples were concentrated using Ultra-15 centrifugal filter devices (Amicon), filter dialyzed into 2x Pfu storage buffer (100mM Tris pH8.0, 2mM DTT, 0.1% CHAPS) and glycerol was added to 50% final v/v. Samples were stored at -20°C.

1.6 E10 Fidelity (Table 1)

To determine the mutation rate of the E10 polymerase in comparison to the commercial Pfuexo$^-$ (Stratagene) in the absence of Cy-dyes a 314bp DNA fragment was amplified with E10 or Pfuexo$^-$ (Stratagene) and primers 5'-TAG CCC CCT TAT TAG CGT TTG CCA AGG GAA CCT TGT AGA GTG GT and 5'-TAG CCC CCT TAT TAG CGT TTG CCA CTT GAG GAG CGA TAT CAT AGT TC. Reactions were performed in 1x *Pfu* buffer (Stratagene) containing 10ng template (pASK*pfu*exo$^-$5), primers (1μM each), dNTPs (50μM each) and 1% formamide. E10 amplifications were, with the exception of

formamide which was included in all reactions, carried out in the presence and absence of additives required for CSR (10% glycerol, 10µg/mL RNase, 1mM DTT). Reactions were hot started with either E10 or Pfuexo$^-$ and thermocycled [94ºC 2 min; 25 x (94ºC 30 sec, 50ºC 30 sec, 72ºC 2 min)]. Reaction products were purified with a QIAquick PCR purification kit (Qiagen). The terminal transferase activity of Taq polymerase was used to append an adenosine at 3' end of the PCR products [0.25mM dATP, 1x Taq buffer (HT Biotech), 1.25U Taq polymerase (HT Biotech); 10min 72°C]. Reactions were purified with a QIAquick PCR purification kit (Qiagen) and cloned into a TOPO TA vector (Clontech) according to the manufacturers instructions. Clones were sequenced on both strands with M13F and M13R primers (5'-GTA AAA CGA CGG CCA G-3' and 5'-CAG GAA ACA GCT ATG AC-3´) and mutations scored. Analysis of the sequences of 15 (E10) respectively 8 (Pfuexo-) clones (4.7kb (E10), 2.5kb (Pfuexo-) (not including the primer sequences)) was used to calculate error rates.

To determine the error rate of E10 with 100% replacement of dCTP with Cy3-dCTP or Cy5-dCTP DNA fragments were amplified in reactions as follows: 10ng template pASK*pfu*exo$^-$5, primers (5'-TAG CCC CCT TAT TAG CGT TTG CCA AGG GAA CCT TGT AGA GTG GT and 5'-TAG CCC CCT TAT TAG CGT TTG CCA CTT GAG GAG CGA TAT CAT AGT TC, 1µM each), 1x *Pfu* buffer (Stratagene), 1% formamide, 10% glycerol, 10µg/mL RNase, 1mM DTT, dATP (50µM), dTTP (50µM), dGTP (50µM), Cy3- or Cy5-dCTP (50µM). Reactions were hot started and thermocycled [94°C 2min; 20 x (94°C 30sec, 50°C 30sec, 72°C 10min)]. Amplified fragments were purified with a QIAquick PCR purification kit (Qiagen) including an additional wash with 750µL of 35% guanidium hydrochloride to remove primer dimers and truncated amplification products. To remove plasmid template and primers 7µL of the column eluate was digested with *Dpn*1 (New England Biolabs) and ExoSAP (USB) (37°C 1 h 30 min, 85°C 15 min). Due to the fact that Cy-dye labeled DNA cannot be directly cloned the labeled fragments were re-amplified using E10 and primer 5'-TAG CCC CCT TAT TAG CGT TTG CCA-3' in reactions containing: 1µL Dpn1/ExoSAP treated sample, 2µM primer, 1x *Pfu* buffer, 1% formamide, 10% glycerol, 10µg/mL RNase, 1mM DTT, 0.2mM dNTPs. Cycling parameters were as follows: 94°C 2 min 25 times (94°C 30 sec, 50°C 30 sec, 72°C 2 min). Thus fragments had experienced a total of 45 cycles of PCR before

cloning. The terminal transferase activity of Taq polymerase was used to append an adenosine at 3' end of the PCR products [0.25mM dATP, 1x Taq buffer (HT Biotech), 1.25U Taq polymerase (HT Biotech); 10min 72°C]. Reactions were purified with a QIAquick PCR purification kit (Qiagen) and cloned into a TOPO TA vector (Clontech) according to the manufacturers instructions. Clones were sequenced on both strands with M13F and M13R primers (5'-GTA AAA CGA CGG CCA G-3' and 5'-CAG GAA ACA GCT ATG AC-3´) and mutations scored. For comparison the equivalent DNA fragment was amplified with Pfu exo$^-$ (2.5U; Stratagene) and the same thermocycling parameters (20 cycles) in the presence of 1% formamide but without the other additives required for E10. Amplified fragments were purified as above with a QIAquick PCR purification kit (Qiagen) including an additional wash with 750µL of 35% guanidium hydrochloride. To remove plasmid template and primers 7µL of the column eluate was digested with *Dpn*1 and ExoSAP (37°C 1 h 30 min, 85°C 15 min) reamplified with Pfu exo- and primer (5'-TAG CCC CCT TAT TAG CGT TTG CCA-3') in reactions containing: 1µL Dpn1/ExoSAP treated sample, 2µM primer, 1x *Pfu* buffer, 1% formamide, 10% glycerol, 10µg/mL RNase, 1mM DTT, 0.2mM dNTPs. Thermocycling parameters were as follows: 94°C 2 min 25 times (94°C 30 sec, 50°C 30 sec, 72°C 2 min). Thus fragments had experienced a total of 45 cycles of PCR (as with E10) before cloning with the only difference being that standard dNTPs were used throughout. The terminal transferase activity of Taq polymerase was used to append an adenosine at 3' end of the PCR products [0.25mM dATP, 1x Taq buffer (HT Biotech), 1.25U Taq polymerase (HT Biotech); 10min 72°C]. Reactions were purified with a QIAquick PCR purification kit (Qiagen) and cloned and sequenced as above and mutations scored. Analysis of the sequences of 20 clones each ((6.3kb), (not including the primer sequences)) were used to calculate error rates.

1.7 Atomic Force Microscopy (Fig. 5)

DNA fragments (314bp) for AFM analysis were amplified by PCR with primers 5'-TAG CCC CCT TAT TAG CGT TTG CCA AGG GAA CCT TGT AGA GTG GT-3' and 5'-TAG CCC CCT TAT TAG CGT TTG CCA CTT GAG GAG CGA TAT CAT AGT TC-3'. Reactions were performed in 1x *Pfu* buffer (Stratagene) containing 10ng template (pASKpfuexo$^-$5), 1% formamide (v/v), primers (1µM each) dNTPs (50µM each). Cy-dye

labelled DNA fragments were amplified with E10 and 100% of the dCTP was replaced by Cy3- or Cy5-dCTP (GE Healthcare). Reactions were hot-started with E10 and thermocycled as follows: 94°C 2 min; 50 times (94°C 10 sec, 50°C 10 sec, 68°C 20 min). Labelled DNA fragments were concentrated and purified by ethanol precipitation followed by additional purification with illustra Microspin G-50 columns (GE Healthcare). Reactions to generate unlabelled control DNA were hot-started with Pfu exo$^-$ (2.5U Stratagene) and thermocycled as follows: 94°C 2 min; 25 times (94°C 30 sec, 50°C 30 sec, 72°C 2 min). Unlabelled PCR products were purified with QIAquick PCR purification columns (QIAGEN Chatsworth CA).

For AFM imaging the modified and control DNA fragments were diluted to a concentration of 1nM in buffer X (10mM Tris-HCl pH7.4, 10mM $MgCl_2$, 10mM NaCl). Poly-L-lysine coated mica was formed by incubating 50μL of 0.001 % poly-L-lysine (Sigma, Poole, UK) on freshly cleaved mica (Goodfellows, Huntingdon, UK) for 10min, followed by rinsing with 10mL MilliQ water (Millipore, Billerica, MA), and blown dry with a stream of nitrogen. The DNA fragments were deposited on the poly-L-lysine coated mica by incubating a 10μL drop of DNA containing solution on the surface for 4min, the sample was then rinsed with 10mL MilliQ water and blown dry with a stream of nitrogen. AFM imaging was performed using a Veeco Multimode AFM with a Nanoscope IIIa controller (Veeco, Santa Barbara, CA) operated in Tapping mode, using Olympus AC160TS silicon nitride cantilevers with a resonant frequency of approximately 350 kHz.

The images were acquired at 512*512 pixel resolution with a scan size of 3μm, scan rates were 1.97 Hz. Images were flattened post capture to remove Z offsets and sample tilt. The DNA contour length of individual molecules was measured by approximating the DNA backbone as a series of straight lines using the Nanoscope software (Veeco). The DNA end-end distance was measured by tracing the straight line distance between the two DNA ends.

1.7 Phenol extraction

NaCl was added to equivalent volumes of Cy-dye labelled DNA to give final concentrations of 0mM NaCl, 100mM NaCl or 150mM NaCl or 200mM NaCl. Each sample was vortexed with 20μL of Tris-HCl (pH7.4) equilibrated phenol. Samples were subsequently centrifuged for 10 min at 13,000xg.

1.8 PCR analysis

Purified enzymes (Fig. 3b) were normalized for activity in PCR with primers 5'-AGG GAA CCT TGT AGA GTG GT-3' and 5'-CTT GAG GAG CGA TAT CAT AGT TC-3' and 10ng of pASK*pfu* as template. Activity normalized enzymes (2.5U) were added to PCR reactions containing: 10ng of plasmid template, 1x *Pfu* buffer (Stratagene), 50μM dGTP, 50μM dATP, 50μM dTTP, 50μM Cy3-dCTP or 50μM Cy5-dCTP, 1μM primers GGG TAC GTG GAG ACC CTC TTC and ACC ACC GAA CTG CGG GTG ACG CCA GGC G, 2% formamide, H_2O to 50μL. Cycling conditions were as follows: 94ºC 2 min, 50 times at 94ºC for 10 sec, 50ºC for 30 sec, and 72ºC for 20 min. Amplified fragments were analysed on 6% polyacrylamide gels.

Crude lysates of E10 (Pfuexo-: E399D, N400G, R407I, V337I, Y546H), clone 15 (E399D, N400G, R407I, and V337I) and Y546H (Y546H) (Fig. 10) were made by growing overnight cultures in 2x TY (100μg/mL ampicillin) at 30ºC. Protein expression was induced by the addition of anhydrous tetracycline (0.4μg/mL) to the overnight culture and the cultures allowed to grow for 5h at 37ºC. Induced cells were pelleted by centrifugation (13,000g, 5min) and 10x lysates were made by resuspension in 1/10th of the culture volume of 1x *Pfu* buffer (Stratagene). The cells were lysed by incubation at 85ºC for 10 min and cell debris was removed by centrifugation at 13,000g for 10 min. Crude lysates as well as *Pfu*exo$^-$ (5U; Stratagene) were activity normalized in ELISA using dNTPs and activity normalized lysates (0.1U) were screened for the ability to incorporate either Cy5-dCTP or Cy3-dCTP in polymerase ELISA as described in 1.4 Crude lysates and *Pfu*exo$^-$ (5U; Stratagene) were activity normalized in PCR with primers 5'-AGG GAA CCT TGT AGA GTG GT-3' and 5'- CTT GAG GAG CGA TAT CAT AGT TC-3' and 10ng of pASK*pfu*exo$^-$5 as template. Activity normalisation of the crude lysates in PCR was performed as follows: 1μM each primer, 1% (v/v) formamide, 100μM dNTPs, 1x Pfu buffer (Stratagene).

Reactions were thermocycled as follows: 94ºC for 2 min, 25 x [94ºC 30 sec, 55ºC 30 sec, 72ºC x min (x= 10sec, 30sec, 1 min or 2min)]. PCRs were set up with activity normalized crude lysates and *Pfu*exo$^-$ (5U; Stratagene) as described above.

1.9 Array manufacture and hybridisation

Dilutions series' (200, 100, 50, 25, and 12.5 ng/μl) of probe molecules (Pfu sequence, and sheared salmon testis genomic DNA) were prepared in 150mM $NaPO_4$ pH8.5/0.01% SDS and 20 replicates spotted onto GAPSII aminosilane-coated glass slides (Corning) using a *Bio*Robotics *Micro*Grid (Genomic Solutions Ltd). Printed slides were baked for 2h at 80°C, incubated with agitation for 30 min at 42°C in 5x SSC/1% BSA (Sigma-Aldrich)/0.1% SDS, boiled for 2min in ultrapure water, washed in 3 changes of ultrapure water at room temperature, rinsed in propan-2-ol and dried by centrifugation.

1.9.1.Model Array Hybridisations (Fig. 6-8)

Klenow DNA labelling:

Fragments of the *Pfu* polymerase gene of lengths 1.3kb (corresponding to the polymerase domain) and 270bp (a portion of the polymerase domain from 1047–1312bp) were amplified with primers 5'-CAG GAA ACA GCT ATG ACG AAC TCG GGA AAG AAT TC-3' and 5'-GTA AAA CGA CGG CCA GTA CCG TCG ACT TAG GAT TTT TTA ATG TTA AGC-3' and 5'-AGG GAA CCT TGT AGA GTG GT-3' and 5'-CTT GAG GAG TAT CAT AGT TC-3' respectively. DNA fragments were amplified with 5U of SuperTaq (HT Biotechnology Ltd) in reactions containing 1x PCR buffer (HT Biotechnology Ltd) 1μM each primer, 10ng pASK*pfu*exo$^-$5, 250μM dNTPs and H_20 to 50μL. Reactions were hotstarted and thermocycled as follows: 94ºC 2 min, 30 times at 94ºC for 30 sec, 55ºC for 30 sec and 72ºC for 3 min. The PCR products were purified using a Qiagen PCR purification kit. 250ng of purified PCR product of lengths 270bp or 1.3kb were used as templates in Klenow labelling reactions (Bioprime DNA labelling System; Invitrogen). Klenow labelling reactions containing 250ng of PCR product 1X random primer solution (50mM Tris-HCl pH6.8, 12.5mM $MgCl_2$, 10mM 2-mercaptoethanol, 300μg/mL random octamer oligodeoxyribonucleotide primers; Bioprime DNA labelling System Invitrogen) in a final reaction volume of 42μL were mixed by vortexing and heated to 95ºC for 5 min after which they were immediately placed on ice. To the reaction mix was

added 200μM (final concentration) of each of dATP, dTTP and dGTP, 100μM (final concentration) dCTP, 100μM (final concentration) Cy3- or Cy5-dCTP and 0.8U of Klenow Large Fragment of DNA polymerase 1 (Bioprime DNA labelling System Invitrogen). Reactions were incubated for 2h at 37ºC. In these standard reactions 20% of the dCTP is replaced by Cy-dye labelled dCTP. The size distribution of the Klenow labelled DNA populations was analysed using an Agilent 2100 Bioanalyser.

E10 DNA labelling

DNA fragments of lengths 1.3kb or 270bp were labelled with Cy3- or Cy5-dCTP by E10 in PCRs with primers 5'-CAG GAA ACA GCT ATG ACG AAC TCG GGA AAG AAT TC-3' and 5'-GTA AAA CGA CGG CCA GTA CCG TCG ACT TAG GAT TTT TTA ATG TTA AGC-3' and 5'-AGG GAA CCT TGT AGA GTG GT-3' and 5'-CTT GAG GAG TAT CAT AGT TC-3' respectively. PCRs were performed in 1x Pfu buffer (Stratagene), 1% (v/v) formamide, 1μM each primer, 50μM dATP, 50μM dGTP, 50μM dTTP, 25μM dCTP, 25μM Cy3- or Cy5-dCTP and 10ng pASK*pfu*exo$^-$5 in a reaction volume of 100μL H_2O. Reactions were hot started with E10 and thermocycled under the following conditions: 94ºC for 2 min 50 times at 94ºC for 10 sec, 50ºC for 10 sec and 68ºC for 20 min. CyDNA products were purified by ethanol precipitation (2.5x sample volume ethanol 1/10th sample volume sodium acetate pH5.2) followed by purification on Autoseq G50 columns (Amersham Biosciences). In these reactions 50% of the dCTP is replaced by Cy3- or Cy5-dCTP.

Labelled-target nomenclature is as follows:

$E10_{270}$ Cy3 or $E10_{270}$ Cy5 refers to a 270bp target DNA that has been labelled using the E10 enzyme where 50% of the dCTP present in the labelling reaction is Cy3-dCTP or Cy5-dCTP. $E10_{1300}$ Cy3 or $E10_{1300}$ Cy5 is a 1.3kb target DNA that has been labelled using the E10 enzyme where 50% of the dCTP present in the labelling reaction is Cy3-dCTP or Cy5-dCTP.

$Klenow_{270}$ Cy3 or $Klenow_{270}$ Cy5 refers to a 270bp target DNA that has been labelled using Klenow enzyme where 20% of the dCTP present in the reaction is Cy3-dCTP or Cy5-dCTP. $Klenow_{1300}$ Cy3 or $Klenow_{1300}$ Cy5 refers to a 1.3kb target DNA that has been

labelled using Klenow enzyme where 20% of the dCTP present in the reaction is Cy3-dCTP or Cy5-dCTP.

The following competitive hybridisations were performed:

1) $Klenow_{270}$ Cy3 vs $Klenow_{270}$ Cy5
2) $Klenow_{1300}$ Cy3 vs $Klenow_{1300}$ Cy5
3) $E10_{270}$ Cy3 vs $Klenow_{270}$ Cy5
4) $E10_{270}$ Cy5 vs $Klenow_{270}$ Cy3
5) $E10_{1300}$ Cy3 vs $Klenow_{1300}$ Cy5
6) $E10_{1300}$ Cy5 vs $Klenow_{1300}$ Cy3

Competitive hybridisations 1 and 2 were performed as control hybridisations and competitive hybridisations 3, 4, 5 and 6 were performed to measure the influence of fragment length on the level of fluorescent signal obtained following hybridisation.

10 ng of Cy3- and Cy5-labelled products were prepared in 20 μl of hybridisation buffer (1mM Tris-HCl pH7.4, 50mM tetrasodium pyrophosphate, 1x Denhardts solution, 40% deionised formamide, 1% NP-40, 10mM DTT, 100μg/ml sheared salmon sperm DNA). Each sample was heated to 95°C for 5min, centrifuged for 2min, applied to the surface of an array and covered with a 22 x 22 mm LifterSlip (Erie Scientific). Hybridisations were performed at 37°C for 16h in a hybridisation chamber (Telechem). Arrays were washed once with 2x SSC/0.1% SDS at room temperature for 5 min, and twice with 1x SSC at room temperature for 5 min. Slides were dried by centrifugation and scanned with an ArrayWoRx 'e' (Applied Precision Instruments). To ensure comparability between slides, identical scanning parameters were used for all arrays. Images were analysed and signals for each array feature extracted using BlueFuse for Microarrays v3.4 (BlueGnome Ltd)).

Calculation of fold-improvements in signal

Log_2(mean Cy3/mean Cy5) for each array feature was calculated. For each probe type (Pfu, or salmon testis DNA) in each experiment the mean log_2 ratio for all included array features (max. 100Pfu features i.e. 20 replicate features each of 5 Pfu probe dilutions; max. 100 salmon sperm DNA features ie 20 replicate features for each of 5 salmon sperm DNA probe

dilution; max. 20 spotting buffer features; max. 20 unspotted areas) was calculated. To normalise the data to the control experiment ($Klenow_{270}$ Cy3 vs $Klenow_{270}$ Cy5 or $Klenow_{1300}$ Cy3 vs $Klenow_{1300}$ Cy5), where equivalent signal is expected in both channels, the Klenow mean log_2 ratio from the control experiments was subtracted from that of every other comparison. If the resulting log_2 ratios are positive or negative, they represent signal improvements in Cy3 and Cy5, respectively, compared to labelling with Klenow at 20% Cy-dye concentration. Fold-improvements for Cy3 and Cy5 are therefore, respectively, the inverse of the resulting mean log2 ratio and the reciprocal of the inverse of the resulting mean log2 ratio.

1.9.2. Model Array Hybridisations (Fig. 9a):

A 1.3kb fragment of the *Pfu* polymerase gene (corresponding to the polymerase domain) was amplified with primers 5'-CAG GAA ACA GCT ATG ACG AAC TCG GGA AAG AAT TC-3' and 5'-GTA AAA CGA CGG CCA GTA CCG TCG ACT TAG GAT TTT TTA ATG TTA ACG-3'. DNA fragments were amplified with 5U of SuperTaq (HT Biotechnology Ltd) in reactions containing 1x PCR buffer (HT Biotechnology Ltd) 1μM primer 1 and 1μM primer 2, 10ng pASK*pfu*exo$^-$5, 250μM dNTPs and H_20 to 50μL. Reactions were hotstarted and thermocycled as follows: 94ºC 2min, 30 times at 94ºC for 30sec, 55ºC for 30sec and 72ºC for 3min. The DNA products of the PCR amplification were purified using a Qiagen PCR purification kit. 250ng of purified 1.3kb PCR product was used as a template in Klenow labelling reactions (Bioprime DNA labelling System; Invitrogen). Klenow labelling reactions containing 250ng of PCR product 1x random primer solution (50mM Tris-HCl pH6.8, 12.5mM $MgCl_2$, 10mM 2-mercaptoethanol, 300μg/mL random octamer oligodeoxyribonucleotide primers; Bioprime DNA labelling System Invitrogen) in a final reaction volume of 42μL were mixed by vortexing and heated to 95ºC for 5min after which they were immediately placed on ice. To the reaction mix was added 200μM (final concentration) of each of dATP, dTTP and dGTP, 100μM (final concentration) dCTP, 100μM (final concentration) Cy3- or Cy5-dCTP and 0.8U of Klenow Large Fragment of DNA polymerase 1 (Bioprime DNA labelling System Invitrogen). Reactions were incubated for 2h at 37ºC. In these standard reactions 20% of the dCTP is replaced by Cy-dye labelled dCTP. The size distribution of the Klenow labelled DNA populations was analysed using an Agilent 2100 Bioanalyser.

For UDG/APE1 digestion labelled 1.3kb Cy-DNA fragments were amplified with E10 in reactions where 10% of the dTTP was replaced by dUTP (Roche) and 50% of the dCTP was replaced by either Cy3- or Cy5-dCTP. PCRs were performed in 1x Pfu buffer (Stratagene), 1% (v/v) formamide, 1μM each primer, 50μM dATP, 50μM dGTP, 45μM dTTP, 5μM dUTP, 25μM dCTP, 25μM Cy3- or Cy5-dCTP and 10ng pASK*pfu*exo$^-$5 in a reaction volume of 100μL H_2O. Reactions were hot started with E10 and thermocycled under the following conditions: 94ºC for 2 min 50 times at 94ºC for 10 sec, 50ºC for 10 sec and 68ºC for 20 min. Cy-dye labelled DNA products were purified by ethanol precipitation (2.5x sample volume ethanol 1/10th sample volume sodium acetate pH5.2) resuspended in 30uL H_2O followed by purification on Autoseq G50 columns (Amersham Biosciences). In these reactions 50% of the dCTP is replaced by Cy3- or Cy5-dCTP. 20uL of Cy-DNA was incubated with 40U UDG (New England Biolabs) in a final volume of 40µL for 1h at 37ºC. UDG treated Cy-DNA samples were purified by ethanol precipitation (2.5x sample volume ethanol 1/10th sample volume sodium acetate pH5.2) resuspended in 30uL H_2O followed by purification on Autoseq G50 columns (Amersham Biosciences). UDG treated Cy-DNA samples were digested with 10U APE1 (New England Biolabs) in a final reaction volume of 50uL at 37ºC for 1h. Following digestion the UDG/APE1 treated Cy-DNA samples were ethanol precipitated (2.5x sample volume ethanol 1/10th sample volume sodium acetate pH5.2) and resuspended in 10uL of H_2O.

Model Array Hybridisations

Labelled-target nomenclature is as follows:

Klenow Cy3 and Klenow Cy5 refers to 1.3kb target DNA that has been labelled using Klenow enzyme where 20% of the dCTP present in the reaction is Cy3-dCTP or Cy5-dCTP. E10 Cy3 and E10 Cy5 refers to 1.3kb target DNA that has been labelled using E10 and treated with UDG/APE1) where 50% of the dCTP present in the reaction is Cy3-dCTP or Cy5-dCTP.

The following competitive hybridisations were performed:

1) Klenow Cy3 vs Klenow Cy5

2) Klenow Cy3 vs E10 Cy5

3) E10 Cy3 vs Klenow Cy5

Competitive hybridisations 2 and 3 were performed to measure the influence of fragment length on the level of fluorescent signal obtained following hybridisation.

10ng of Cy3- and Cy5-labelled products were prepared in 25µL of hybridisation buffer (50mM Tris-HCl pH7.4, 1.6mM tetrasodium pyrophosphate, 5x Denhardts solution, 5x SSC, 40% deionised formamide, 1% NP-40, 10mM DTT, 4µg Yeast tRNA). Each sample was heated to 95 °C for 5 min, centrifuged for 2 min, applied to the surface of an array and covered with a 22 x 22 mm LifterSlip (Erie Scientific). Hybridisations were performed at 37°C for 16h in a hybridisation chamber (Telechem). Arrays were washed once with 2x SSC at room temperature for 10 min, once with 1x SSC/0.05% Tween20 at room temperature for 10 min and once with 1x SSC at room temperature for 10min. Slides were dried by centrifugation and scanned with an ArrayWoRx 'e' (Applied Precision Instruments). To ensure comparability between slides, identical scanning parameters were used for all arrays. Images were analysed and signals for each array feature extracted using BlueFuse for Microarrays v3.4 (BlueGnome Ltd) Mean signal for each probe type/dilution was calculated and a fold-improvement in signal measured against Klenow labelled control for each target type.

1.9.3. Array CGH : Fig. 9b

Target Preparation

80µg of human female or male genomic DNA (Promega) was sonicated for 1 min 30 sec. 6µg of the male or female sonicated genomic DNA was blunted with T4 DNA polymerase (New England Biolabs) to which was ligated DNA linkers (according to Chua et al., 2004[9]). Ligation mediated PCR was performed on 250ng of the linker modified genomic DNA according to Chua et al., 2004[10]. 400ng of PCR (l-DNA) product was used as template for in Klenow labelling reactions (Bioprime DNA labelling System; Invitrogen). Klenow labelling reactions containing 400ng of PCR product 1x random primer solution (50mM Tris-HCl pH6.8, 12.5mM $MgCl_2$, 10mM 2-mercaptoethanol, 300μg/mL random octamer oligodeoxyribonucleotide primers; Bioprime DNA labelling System Invitrogen) in a final

reaction volume of 42μL were mixed by vortexing and heated to 95ºC for 5min after which they were immediately placed on ice. To the reaction mix was added 200μM (final concentration) of each of dATP, dTTP and dGTP, 100μM (final concentration) dCTP, 100μM (final concentration) Cy3- or Cy5-dCTP and 0.8U of Klenow Large Fragment of DNA polymerase 1 (Bioprime DNA labelling System Invitrogen). Reactions were incubated overnight at 37ºC. In these standard reactions 20% of the dCTP is replaced by Cy-dye labelled dCTP. As a control for amplification bias of genomic coverage and copy number representation both male and female sonnicated genomic DNA was labelled with Klenow (Bioprime DNA labelling system; Invitrogen) as described above.

For E10 labelling 250ng of the linker modified DNA was amplified with primer JW102[10] in reactions where 10% of the dTTP was replaced by dUTP (Roche) and 50% of the dCTP was replaced by either Cy3- or Cy5-dCTP. PCRs were performed in 1x Pfu buffer (Stratagene), 1% (v/v) formamide, 0.4μM each primer, 50μM dATP, 50μM dGTP, 45μM dTTP, 5μM dUTP, 25μM dCTP, 25μM Cy3- or Cy5-dCTP in a reaction volume of 100μL H_2O. Reactions were hot started with E10 and thermocycled under the following conditions: 94ºC for 2 min 50 times at 94ºC for 10 sec, 50ºC for 10 sec and 68ºC for 20 min. Cy-dye labelled DNA products were purified by ethanol precipitation (2.5x sample volume ethanol 1/10th sample volume sodium acetate pH5.2) resuspended in 30uL H_2O followed by purification on Autoseq G50 columns (Amersham Biosciences). In these reactions 50% of the dCTP is replaced by Cy3- or Cy5-dCTP. 20uL of Cy-DNA was incubated with 40U UDG (NEB) in a final volume of 40µL for 1h at 37ºC. UDG treated Cy-DNA samples were purified by ethanol precipitation (2.5x sample volume ethanol 1/10th sample volume sodium acetate pH5.2) resuspended in 30uL H_2O followed by purification on Autoseq G50 columns (Amersham Biosciences). UDG treated Cy-DNA samples were digested with 10U APE1 (New England Biolabs) in a final reaction volume of 50uL at 37ºC for 1h. Following digestion the UDG/APE1 treated Cy-DNA samples were ethanol precipitated (2.5x sample volume ethanol 1/10th sample volume sodium acetate pH5.2) and resuspended in 10uL of H_2O.

Samples were quantified for fluorescence incorporation and DNA yield using a NanoDrop ND-1000.

1) The following competitive hybridisations were performed:Female sonnicated DNA Klenow Cy3 vs Male sonnicated DNA Cy5
2) Female sonnicated DNA Klenow Cy5 vs Male sonnicated DNA Cy3
3) Female l-DNA Klenow Cy3 vs Male l-DNA Cy5
4) Female l-DNA Klenow Cy5 vs Male l-DNA Cy3
5) Female l-DNA UDG/APE1 E10 Cy3 vs Male l-DNA UDG/APE1 E10 Cy5
6) Female l-DNA UDG/APE1 E10 Cy5 vs Male l-DNA UDG/APE1 E10 Cy3

The female/male pairs were mixed, normalising the amount of DNA to the lowest DNA yield, and precipitated. Labelled target pairs were resuspended in 25 ul of hybridisation solution (50% formamide, 2x SSC, 0.1% SDS, 10% dextran sulfate, 2.5 mg/ml COT-1 DNA, 15mg/ml herring sperm DNA. Targets were denatured at 75°C for 10 min, spun, and hybridised to 1 Mb interval human BAC arrays (Centre for Microarray Resources, Department of Pathology, University of Cambridge) under a 22 x 22mm coverslip for 16 h. Slides were washed in 1x PBS/0.05% Tween20 at room temp for 10 min (3 times), 30 min in 50% formamide/2xSSC at 42°C, 1xPBS/0.05% Tween 20, 2x 5 min in 1x PBS, spun dry.

Slides were scanned with an ArrayWoRx e scanner. Scan settings were normalized to give unsaturated signal on the slide with the highest signal (E10-labelled targets in both channels)) and a 1:1 Cy3:Cy5 ratio on the control slide (Klenow labelled targets in both channels). The same scan settings were then used for all slides. Images were analysed as dyeswaps using BlueFuse for Microarrays v3.4 (BlueGnome Ltd). After normalisation, Log2ratio (Cy3:Cy5) vs chromosomal position was plotted for each probe on the slide. The aCGH method uses deviation from the Log2ratio = 0 to measure copy number gain or loss across the genome. In this experiment (pooled female DNA vs pooled male DNA) the autosomal chromosomes are copy number equal (Log2ratio = 0), and chromosomes X and Y are a copy number gain of 2:1 and a copy number loss of 0:1 for the female, respectively, compared to the male. Typical Log2ratios for X and Y are 0.7 and -0.9, respectively and do not typically reach the calculated theoretical Log2ratios.

1.10 Single-molecule Cy-detection in a microfluidic device (SI Fig. 3)

CyDNA fragments of length 270bp were labelled with Cy3- or Cy5-dCTP by E10 in PCRs with primers 5'-AGG GAA CCT TGT AGA GTG GT-3' and 5'-CTT GAG GAG CGA TAT CAT AGT TC-3' respectively. PCRs were performed in 1x Pfu buffer (Stratagene), 10ng pASK*pfu*exo$^-$5, 1% (v/v) formamide, 1μM each primer, 50μM dATP, 50μM dGTP, 50μM dTTP, 50μM Cy3- or Cy5-dCTP in a reaction volume of 50μL H_2O. Reactions were hot started with E10 and thermocycled under the following conditions: 94ºC for 2 min 50 times at 94ºC for 10 sec, 50ºC for 10 sec and 68ºC for 20 min. Cy-dye labelled DNA products were purified by ethanol precipitation (2.5x sample volume ethanol 1/10th sample volume sodium acetate pH5.2) followed by purification on AutoSeq G50 columns (Amersham Biosciences). For detection at single- or near-single-molecule levels, a solution was prepared containing 5pg/μl each of 100% Cy5-dCTP labelled and 50% Cy3-dCTP labelled 270bp PCR products in water. Based on the sequence of the PCR product, each double-stranded molecule should carry on average 102 Cy3 or 102 Cy5 labels. This solution was driven through a fused silica capillary (internal diameter ~40μm) at 30μl/hr using a syringe pump. Fluorescence was detected at a fixed point inside the lumen of the capillary, using a custom optical unit, which consisted of 532/635nm laser light fed in by a single mode (3/125μM) silica optical fiber. On emerging from the fiber the light was collected and focused onto the sample using a combination of 0.3NA and 0.65NA aspheric lenses, the latter being both the focusing and collecting objective for fluorescence. The working distance of the sample objective was approximately 1.5mm, and the intensity field in the focal volume had a Gaussian profile with a 1/e diameter of approximately 2.5μM (effective focal volume approximately 8μm^3, or 8fl). Collected light was filtered of reflected laser energy using a cascaded series of reflective/transmissive filters and then focused into a receiving fiber which conveyed the light to a Perkin Elmer SPCM module for the final quantitation of intensity. Counts for both Cy3 fluorescence (570nm; bandwidth ~25nm) and Cy5 fluorescence (670nm; bandwidth ~25nm) were collected in consecutive 20μsec periods and averaged over a sliding window (centre-weighted) of 40 periods.

References

1. Lu, C.; Erickson, H. P., *Protein Expr Purif* **1997,** *11*, 179-184.
2. Skerra, A., *Gene* **1994,** *151*, 131-135.
3. Boudsocq, F.; Iwai, S.; Hanaoka, F.; Woodgate, R., *Nucleic Acids Res* **2001,** *29*, 4607-4616.
4. Derbyshire, V.; Pinsonneault, J. K.; Joyce, C. M., *Methods In Enzymology* **1995,** *262*, 363-385.
5. Fogg, M. J.; Pearl, L. H.; Connolly, B. A., *Nat Struct Biol* **2002,** *9*, 922-927.
6. Ong, J. L.; Loakes, D.; Jaroslawski, S.; Too, K.; Holliger, P., *J Mol Biol* **2006,** *361*, 537-550.
7. Tawfik, D. S.; Griffiths, A. D., *Nature Biotechnol* **1998,** *16*, 652-656.
8. Ghadessy, F. J.; Ong, J. L.; Holliger, P., *Proc. Natl. Acad. Sci. USA* **2001,** *98*, 4552-4557.
9. Ghadessy, F. J.; Holliger, P., *Methods Mol Biol* **2007,** *352*, 237-248.
10. Chua, Y. L.; Mott, E.; Brown, A. P.; MacLean, D.; Gray, J. C., *Plant J* **2004,** *37*, 789-800.