View Online / Journal Homepage / Table of Contents for this issue
Dynamic Article Links

Lab on a Chip

Cite this: *Lab Chip*, 2012, **12**, 3314–3321

www.rsc.org/loc

PAPER

Denaturation mapping of *Saccharomyces cerevisiae*

Robert L. Welch,[a] Robert Sladek,[b] Ken Dewar[b] and Walter W. Reisner*[a]

Received 3rd May 2012, Accepted 20th June 2012
DOI: 10.1039/c2lc40504k

Downloaded by McGill University on 22 August 2012
Published on 23 July 2012 on http://pubs.rsc.org | doi:10.1039/C2LC40504K

Optical mapping of DNA provides large-scale genomic information that can be used to assemble contigs from next-generation sequencing, and to detect rearrangements between single cells. A recent optical mapping technique called denaturation mapping has the advantage of using physical principles rather than the action of enzymes to probe genomic structure. Denaturation mapping uses fluorescence microscopy to image the pattern of partial melting along a DNA molecule extended in a channel of cross-section 120 nm at the heart of a nanofluidic device. We used denaturation mapping to locate single DNA molecules on the yeast genome (12.1 Mbp) by comparing images to a computationally predicted map for the entire genome sequence. By locating 84 molecules we assembled an optical map of the yeast genome with > 50% coverage.

1 Introduction

While genome sequencing costs have continued to drop and we envisage routine human genome sequencing, there remain key research applications (assembling human and microbial genomes, capturing disease associated regions, phasing long-range haplotypes) that are unlikely to be solved by technological advances within a classical sequencing paradigm. In particular, there is a need for a technology that can efficiently manipulate large DNA fragments (100s of kbp to multi Mbp range), precisely assess genomic content to select target genomic regions and–most critically–to do this with single molecules, opening the door to single-cell genomic analysis[1] based on direct analysis of a single genome extracted from a single cell (with no need for intervening amplification).

Nanofluidics offers a possible route towards such a single-cell genomic technology. When a DNA molecule is introduced in a structure with dimension below its equilibrium coil size (*e.g.* radius of gyration), the molecule's equilibrium conformation will be altered as a consequence of confinement. In particular, confining a DNA molecule to a nanochannel will force the molecule to extend along the channel axis, linearly arraying the genome for analysis.[2] Nanochannel technology is highly parallel and permits continuous-throughput operation. Multiple nanochannels constructed side-by-side (or in serpentine arrays) allow many molecules (or long stretches of genome, 100s kbp – Mbp) to be analyzed in one camera aquisition. As the DNA is not surface bound, but free to move in the channels, fresh molecules can be continuously scrolled through the channel arrays.

Genomic maps are typically produced on the extended dsDNA through the use of sequence-specific enzymatic reactions. For example, nanochannel-based mapping approaches have been demonstrated using direct restriction of extended DNA[3] and (more recently) nicking endonucleases, which cleave a single strand at the recognition sequence and allow incorporation (by polymerase) of fluorescently labeled nucleotides.[4] These conventional enzymatic approaches require a multi-step biochemical preparation which is not ideal in a nanofluidic context. A recent approach called denaturation mapping relies on the physical principle of partial melting.[5] The probability that DNA will denature (melt from double to single strands) is sequence-dependent because regions of the sequence rich in G–C base pairs melt at a higher temperature than regions rich in A–T's. The DNA is uniformly stained with an intercalating dye (YOYO-1) that unbinds from single-stranded regions. The DNA is then partially melted under confinement in a nanochannel. The resulting pattern of fluorescence (with unmelted regions appearing bright and melted regions appearing dark), is a barcode that reflects the degree of denaturation along the sequence. Denaturation mapping has the advantage of not needing enzymatic labeling or reaction steps, requiring only a uniform stain. This simplicity makes denaturation mapping especially attractive as a cost-effective and scaleable genomic mapping tool, particularly applied to analyzing DNA extracted from cells lysed on-chip, where complex multi-step biochemical preparations may not be feasible.

We used denaturation mapping to locate single DNA molecules globally on a genome of size 12.1 Mbp (yeast, *Saccharomyces cerevisiae*). The successful genomic alignment of single molecules, rather than an average of an ensemble of molecules (as performed in Reisner *et al.*[5]), is a milestone in establishing the denaturation approach as an optical genomic mapping technology. We aligned 84 molecules to the genome to form an optical map with 56% coverage of the yeast genome,

[a] *Department of Physics, McGill University, 3600 rue University, Montreal, Canada. E-mail: reisner@physics.mcgill.ca; Fax: + 1(514) 398-8434; Tel: +1 (514) 398-3058*
[b] *The McGill University and Génome Québec Innovation Centre, 740 avenue du Docteur Penfield, Montreal, Canada*

This journal is © The Royal Society of Chemistry 2012

Downloaded by McGill University on 22 August 2012
Published on 23 July 2012 on http://pubs.rsc.org | doi:10.1039/C2LC40504K

with particularly strong coverage of chromosome 3 (96% of 317 kbp) and chromosome 7 (92% of 1.08 Mbp). Our results are the first demonstration that denaturation mapping of single DNA molecules can be used to form a comprehensive optical map of a eukaryotic genome.

We used a protocol for denaturation mapping with some technical refinements over previous work. By employing a SCODA-based DNA preparation procedure to extract DNA from agarose gel plugs we successfully obtained and mapped DNA molecules as long as 360 kbp, more than double the length of molecules mapped in previous experiments. We used more rigorous criteria for the statistical significance of genomic alignment results in order to systematically reject barcodes with experimental defects and prevent them from introducing errors into the optical map.

The ability to consistently record and genomically locate 100 kbp+ optical maps establishes denaturation mapping as a powerful probe of large-scale genomic structure lost in the genomic fragmentation required by sequencing reactions. The de novo assembly of genomes using short sequence read technologies (100-600 bp) remains a significant bioinformatics task.[6] Denaturation maps could overcome this challenge by providing a large-scale scaffolding over which to assemble contigs. As well, the ability to resolve single DNA molecules makes denaturation mapping applicable to important questions of genomic heterogeneity in populations of cells. Our technique could detect the large-scale rearrangements between single cells implicated in drug resistance in infectious strains of yeast and cancer.[7,8]

2 Methods

2.1 Fabrication

The device is similar to the design used in previous denaturation mapping experiments of Reisner *et al.*,[5] with some chips having minor variations in the dimensions of nanochannels. Fig. 1 is a schematic of the device design. A brief summary of the fabrication details follows. The device contains an array of nanochannels (of width and depth 120–150 nm, and length 200 nm, spaced 2 μm apart), supplied by loading microchannels (50 μm wide and 1 μm deep) that terminate in macroscale ports (of radius 1 mm). These features were patterned in fused silica using a multi-step fabrication process. Firstly, the nanochannel array was defined in ZEP520A resist using electron-beam lithography (JEOL) and etched into the silica using CF_4:CHF_3 reactive ion etching (RIE). Secondly, the loading microchannels were defined using contact UV photolithography and etched into the silica using RIE. Next the ports were sandblasted through the chip at the termini of the microchannels. Finally, the channels were sealed by direct silica-silica bonding on one side to a 150 μm thick fused silica cover glass (Valley Design).

2.2 DNA preparation

We obtained genomic DNA for denaturation experiments from whole chromosomes of a eukaryotic genome, as opposed to viral and bacterial artificial chromosome (BAC) DNA as in Reisner *et al.*[5] We used a SCODA development device provided by Boreal Genomics to extract chromosomal DNA from the yeast strain YPH80 to buffer from an agarose gel plug (chromosome sizes 225–1900 kbp, New England BioLabs). SCODA (synchronous coefficient of drag alteration) is a pulsed-field electrophoresis technique that exploits nonlinearity in the dependence of mobility on electric field in order to selectively migrate DNA molecules.[9] SCODA enables the concentration and recovery of high molecular weight DNA (100 kbp+) with high purity.[10] This approach allowed us to obtain DNA molecules longer (>300 kbp) than those previously observed by Reisner *et al.*[5] (40–166 kbp) for denaturation mapping experiments.

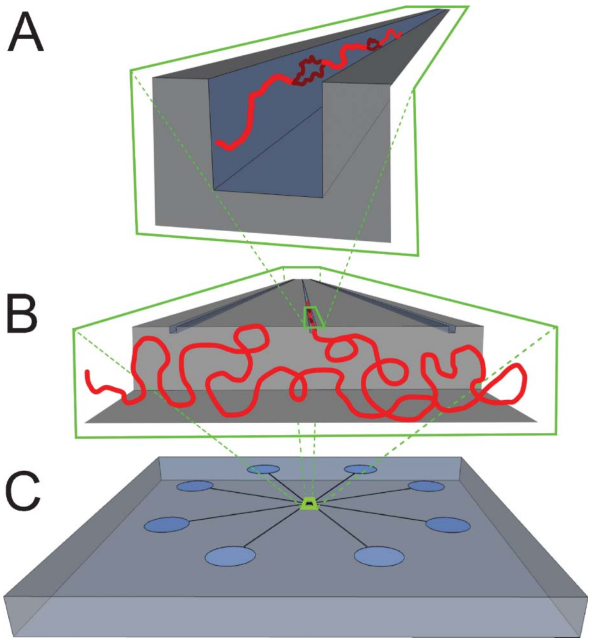

Fig. 1 Schematic of the principle of partial denaturation under confinement in nanochannels. DNA molecules are confined in an array of nanochannels at the centre of the device where they stretch out to adopt a linear conformation (*A*). The molecules are partially melted to form a sequence of double-stranded segments (red) and single-stranded bubbles (brown). DNA is brought into nanochannels from loading microchannels by applying a pressure gradient to overcome the entropic barrier to confinement (*B*). The microchannels are serviced by macroscale ports into which DNA is loaded by micropipetting (*C*).

The SCODA gel was made with 0.1 g 1% LMP agarose dissolved in 10 mL 0.25 × TBE. YPH80 gel plugs were embedded directly in the SCODA gel and DNA was collected in a reservoir in the centre of the gel. DNA was concentrated from the gel to 1 × TBE using SCODA runs of durations 20 h, voltage 12%, cycle length 2880 s, and no bias voltage. DNA extracted by SCODA was quantified using a PicoGreen assay (Invitrogen). Subsequently, DNA was stained with the intercalating dye YOYO-1 (Invitrogen) at a ratio of 1 dye molecule to 10 base pairs. The stained DNA was added to a running buffer containing 0.05 × TBE (4.5 mM Tris, 4.5 mM boric acid, 0.1 mM EDTA), 10 mM NaCl, and formamide at a ratio of 50% by volume (Sigma). The running buffer also contained a system to discourage photo-nicking, added before mixing with formamide, composed of β-mercaptoethanol at 3% per volume, as well as 0.2 mg mL^{-1} glucose oxidase, 0.04 mg mL^{-1} catalase and 4 mg mL^{-1} β-D-glucose.

Downloaded by McGill University on 22 August 2012
Published on 23 July 2012 on http://pubs.rsc.org | doi:10.1039/C2LC40504K

2.3 Experimental setup

The experimental setup for performing denaturation mapping using the nanofluidic devices is very similar to that of Reisner *et al.*[5] An overview is given in Fig. 2 and a brief summary follows. The device is mounted onto a custom-designed chuck machined in polyetheretherketone (PEEK). The chuck contains pressure lines that connect to loading ports on the device by an *o*-ring seal and terminate in luer locks. Pneumatic pressure is used to control the flow of DNA in the device. The chuck also contains a port that holds a resistive heating element and thermocouple in contact with the device in order to control temperature in the nanochannels. Lastly, the chuck fits to a custom-designed holder that positions the device above the objective of an inverted microscope for imaging. Fluorescence images were recorded using a Nikon Eclipse TE2000 inverted microscope with a 100× N.A. 1.4 immersion objective and electron multiplying CCD camera (Andor iXon).

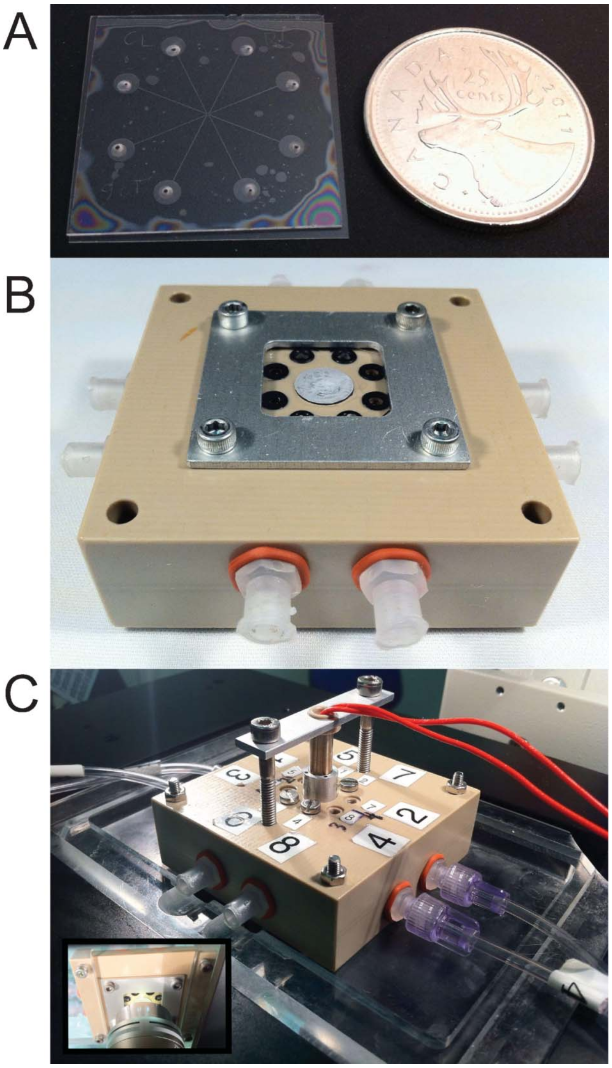

Fig. 2 Device and experimental setup. The device (*A*) is mounted onto a custom-designed chuck (*B*) which interfaces the device with a pump and heating element, and holds the device above the objective lens of an inverted microscope for imaging (*C* and *inset*).

The DNA-containing running buffer is pipetted into the loading ports on the device and allowed to wet the nanofluidic circuit. The device is mounted on the chuck, aligning the device's loading ports with the corresponding pressure lines, and fastened to the chuck using a metal bracket to form a seal around the *o*-rings. The resistive heating element is mounted on the chuck with thermal grease applied to create good thermal contact. The chuck, in turn, is mounted on the sample positioning stage of the inverted microscope and the objective lens is brought into oil immersion contact with the chip. The device is positioned so that the interface between the nanochannel array and a loading microchannel is in the field of view. The heating element is set to a temperature of 29 °C and allowed to reach thermal equilibrium with the chip for a few minutes. A pressure differential is applied across the loading microchannel to bring DNA molecules to the nanochannel array, and then a pressure differential is applied across the nanochannel array in order to load DNA molecules into the nanochannels. Finally, a confined DNA molecule is imaged in the nanochannel array and a movie is recorded of 50 images at 10 frames per second.

The procedure of loading and imaging DNA molecules is iterated to collect melting barcodes. Whereas in Reisner *et al.*[5] many DNA molecules were concentrated to the centre of the field of view, here concentration was not necessary because the longer DNA molecules occupied most of the field of view and so we flowed DNA through the channels processively.

2.4 Preparation of experimental denaturation barcodes

A DNA molecule in a nanochannel experiences Brownian fluctuations that create time-dependent spatial distortions in its fluorescence profile. We remove these distortions from movies of confined DNA using a custom algorithm in MATLAB, previously described in Reisner *et al.*[5] Fig. 3 is an example of a typical melting barcode that has been adjusted with these algorithms.

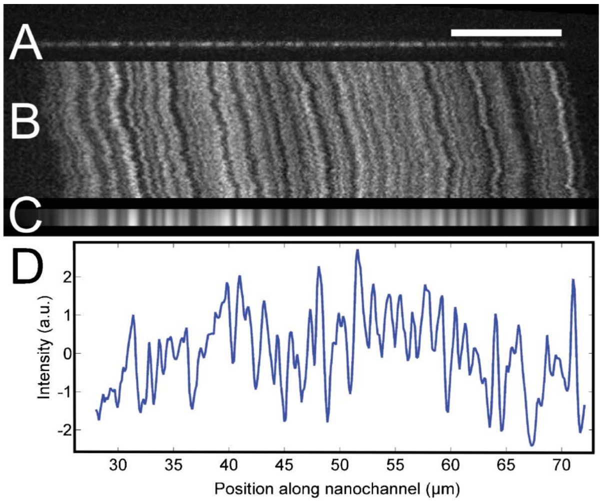

Fig. 3 Recording a denaturation barcode. (*A*) is a fluorescence image of a confined, partially melted DNA molecule dyed with YOYO-1 that exhibits a characteristic "barcode" pattern of fluoresence intensity. A time series of 50 images are recorded (*B*) and processed to remove distortions due to thermal fluctuation and produce an averaged barcode (*C*) and (*D*)). (Scale bar 10 μm).

This journal is © The Royal Society of Chemistry 2012

Downloaded by McGill University on 22 August 2012
Published on 23 July 2012 on http://pubs.rsc.org | doi:10.1039/C2LC40504K

A brief overview of the algorithm follows. The first frame of the movie is taken as a template. Firstly, centre-of-mass diffusion is corrected for by globally translating each of the remaining frames to maximize cross-correlation with the template.

Secondly, local variations in density are also corrected for by locally rescaling the resulting frames in segments. Each frame i contains a fluorescence profile $P_i(x_j)$ with respect to position in pixels x_j. We divide the profiles by position into a series of 50 segments. The length of each segment k is then dilated linearly by a factor d_k. The set of dilation factors $\{d_k\}$ for each frame i is chosen to minimize δ, the least squared difference between the dilated profile $\mathrm{P}_i'(x_j, \{d_k\})$ and the template $P_1(x_j)$:

$$\delta = \sum_{j=1}^{N} [P_i'(x_j,\{d_k\}) - P_1(x_j)]^2$$

The choice of dilation factors $\{d_k\}$ for each frame i represents a map $M(\{d_k\})_i$ for that frame that rescales the profile P_i to a modified profile $P_i' = M(\{d_k\})_i P_i$ that matches the local variation in density of the template profile P_1. If all the profiles are rescaled with their respective maps $M(\{d_k\})_i$, then the average of the rescaled profiles over all the frames $< P_i' >$ is a consensus profile P^* that has the same instantaneous conformation as the template. In addition, taking the average of the different maps over all the frames $<M(\{d_k\})_i>$ gives the average transformation M^* that rescales a profile to the conformation of the template profile. Conversely, the inverse of the average transformation M^{*-1} rescales the template profile to the average conformation of the profiles from all the other frames. Finally, by applying the inverse of the average map M^{*-1} to the consensus profile P^* we arrive at a melting barcode $P_f = M^{*-1}P^*$ that is a denaturation profile with the average conformation of the DNA molecule during the experiment.

Performing the mutli-segment rescale increases the signal-to-noise ratio of the melting barcode by allowing us to take an average profile over many frames despite fluctuations in conformation. As well, the fact that the resulting barcode has a conformation averaged over a duration of several seconds provides a barcode that is closer to the thermal equilibrium conformation than the profile of any given frame. We performed the rescaling process to produce a single molecule melting barcode for each DNA molecule we imaged during our experiments.

2.5 Calculation of the genomic denaturation barcode

We located DNA molecules on the genome by comparing experimentally obtained melting barcodes to a theoretically predicted barcode of the complete genomic sequence, as in Reisner *et al.*[5] The genomic barcode was predicted in three steps: by generating a melting probability map from the nucleotide sequence; predicting a barcode from the melting probability map; and then interpolating between barcodes predicted for a range of helicities.

The reference yeast genome nucleotide sequence was downloaded from NCBI Genbank (RefSeq numbers NC001133 to NC001148). The melting probability map was calculated from this sequence using version 0.9.3 of the software package bubblyhelix provided by Eivind Tøstesen.[11] Bubblyhelix uses the Poland-Scheraga model to compute melting probability as a function of base pair position for a given temperature and salt concentration. The program takes as inputs the respective enthalpies and entropies of the 10 possible base-pair duplex stacking interactions and takes into account the energetics of bubble formation, and distinguishes between a bubble and the melted end of a molecule.[12]

The genomic barcode is predicted from the melting probability map by accounting for two factors: local variation in DNA extension due to partial melting and the optical point-spread function (the diffraction limited profile of a point-source imaged with our microscope system). The intensity profile $P(s) \propto p_{ds}(s)$ and extension $r(s)$ as functions of sequence position s (in bp) are determined by

$$r(s) = r_{ds}p_{ds}(s) + r_{ss}{*}[1 - p_{ds}(s)]$$

where the quantity p_{ds} is probability of a base-pair remaining unmelted, the value of r_{ds} defines the extension of an unmelted base-pair along the nanochannel, and the ratio r_{ss}/r_{ds} determines the relative extension of a melted base-pair. From observations of λ-phage DNA molecules in similar nanochannel devices we find that values of $r_{ds} = 0.186$ and $r_{ss}/r_{ds} = 0.85$ are reasonable.[5] The series of extensions $r(s)$ forms a position co-ordinate in nanometers $x(s) = \sum_{s'=1}^{s} r(s')$ that describes the position of the DNA molecule along the length of the nanochannel. We simulate the effect of diffraction due to the microscope objective by convolving the expected intensity profile ($x(s)$, $P(s)$) with a point-spread function parameterized by a Gaussian of standard deviation $\sigma = 200$ nm. The net result is a simulated profile of fluorescence intensity with respect to position, or predicted melting barcode, of each of the chromosomes of the yeast genome at a given temperature. We compare experimental melting barcodes not just to a single genomic barcode, but to a matrix that interpolates between genomic barcodes predicted for a range of temperatures. We use helicity, measured as the average probability of remaining double-stranded of all the base-pairs of DNA molecule $h = <p_{ds}(s)>$, as the interpolation coordinate.

2.6 Alignment of experimental barcodes to the genome

We identified the position of single DNA molecules on the yeast genome by comparing experimentally obtained melting maps to the calculated genomic barcode helicity interpolation matrix. This approach is similar to that of the BAC alignment in Reisner *et al.*[5] with the important distinction that here we align melting barcodes from *single* DNA molecules to the genome, rather than barcodes that are a consensus of many (10–26) molecules. Alignment was performed by finding the genomic position that minimized the least-squared difference between them using a collection of custom MATLAB programs. In this work there are a few refinements to the software: many barcodes can now be aligned in a batch fashion, the total coverage of the mapped molecules is calculated after alignment, and the criteria for a statistically significant alignment result are more rigorous (as discussed later in this section).

We compare an experimental melting barcode $P_{\exp}$ to a segment of the same length from position i on the genomic barcode P_g. From these we subtract the mean and divide by the local standard deviation to obtain the adjusted barcodes:

Published on 23 July 2012 on http://pubs.rsc.org | doi:10.1039/C2LC40504K
Downloaded by McGill University on 22 August 2012

$$\delta P_{exp} = \frac{P_{exp} - \langle P_{exp} \rangle}{\langle (P_{exp} - \langle P_{exp} \rangle)^2 \rangle^{1/2}},$$

$$\delta P_g = \frac{P_g - \langle P_g \rangle_{i,N}}{\langle (P_g - \langle P_g \rangle_{i,N})^2 \rangle_{i,N}^{1/2}}$$

where i is the starting position of P_g on the genomic barcode and N is the length of both fragments. Using these definitions, for each position on the genome i we define the least-squares estimator:

$$\Delta(i) = \frac{1}{2N} \sum_{j=1}^{N} [\delta P_g(x_{i+j}) - \delta P_{exp}(x_j)]^2$$

To determine the global minimum of the estimator we perform a search on the genome barcode across four parameters. We vary the helicity on the genomic interpolation matrix from which P_g is taken, the sequence position i, and the relative orientation of the experimental barcode (forward or backward). As well we perform a global dilation to the experimental barcode, attempting alignment after elongation by a range of factors between 0.8 and 1.2 in order to allow for some variation in the dimensions of the nanochannels. We perform a coarse search varying these parameters widely, and then perform a finer search to find the global minimum of the estimator.

We assess the statistical significance of a location result by analyzing the distribution of estimators at different sequence positions i with the other search parameters held constant. Given the form of the adjusted barcodes δP_{exp} and δP_g we can write the estimator as

$$\Delta(i) = 1 - \frac{1}{N} \sum_{j=1}^{N} \delta P_g(x_{i+j}) \delta P_{exp}(x_j)$$

If δP_{exp} and δP_g are independent random variables, the set of values $\frac{1}{N} \sum_{j=1}^{N} \delta P_g(x_{i+j}) \delta P_{exp}(x_j)$ for each position i will also be a sequence of random numbers with a mean of zero. It arises from the central limit theorem that the distribution of estimators for all values of i where an experimental barcode cannot be successfully located will be a Gaussian centred about unity. Due to the finite length of the barcodes we add a quartic correction to obtain an expected distribution of estimators

$$P(\Delta) = Ae^{\left(-\frac{(\Delta-\Delta_0)^2}{2\sigma^2} - \frac{(\Delta-\Delta_0)^4}{24\eta^4}\right)},$$

where the mean value $\Delta_0 = 1$. During a location procedure we fit the distribution of estimators with the function $P(\Delta)$.

Integrating $P(\Delta)$ from $\Delta = 0$ to the global minimum Δ_f gives an estimate n_L of the number of locations more significant than the one obtained that would be expected to occur randomly between uncorrelated DNA sequences. We argue that the location is of statistical significance if the fit of $P(\Delta)$ is good and n_L is much less than one. The requirement that $P(\Delta)$ pass a goodness of fit test is an additional measure not applied in Reisner *et al.*[5] that is necessary to ensure the results derived from $P(\Delta)$ are meaningful. We assess the goodness of fit of $P(\Delta)$ by performing Pearson's chi-squared test for all bins of the estimator histogram between the global minimum Δ_f and $\Delta = 1$. Here we judge a location as successful whose estimator distribution can be fit by $P(\Delta)$with $n_L < 0.02$ that passes a chi-squared test with a p-value of >0.01.

3 Results

In total we aligned 84 DNA molecules to the yeast genome. Fig. 4 shows three typical melting barcodes that have been aligned *via* comparison to a theoretical barcode computed for the entire genomic sequence. For each molecule we identify one sequence position that represents a global minimum in the least-squares estimator Δ between the barcodes, Δ_f. Among the molecules aligned we find the number of sequence positions n_L, for which we would expect $\Delta \leqslant \Delta_f$, based on the distribution of values for Δ at other sequence positions, to be smaller than 0.003 on average.

These 84 melting barcodes form an optical map that covers 56.3% of the yeast genome sequence, about 6.8 Mbp in total. Fig. 5 shows the genomic positions mapped during our experiments. A handful of chromosomes are especially well covered: we mapped 96% of chromosome 3 (total length 317 kbp), 92% of chromosome 7 (total length 1.08 Mbp) and 91% of chromosome 2 (total length 813 kbp). Using SCODA to prepare DNA samples allowed us to record very long optical maps: from DNA molecules on average 167 kbp in length, and from three longer than 300 kbp. Our barcodes represent large fractions of chromosomes, on average 24% of the entire contour and often more in the case of shorter chromosomes. One 204 kbp fragment occupied 64% of chromosome 3.

Our results extend the first denaturation mapping results of Reisner *et al.*[5] in two significant ways. Firstly, where Reisner *et al.*[5] analyzed consensus barcodes formed by averaging barcodes from tens of DNA molecules (up to 40), we aligned single DNA molecules globally to a genomic barcode for the first time. Secondly, whereas Reisner *et al.*[5] mapped three viral genomes (48 kbp, 152 kbp and 166 kbp) and mapped a very small portion of human chromosome 12 (0.11%), we present a map with comprehensive (>50%) coverage of a eukaryotic genome (12.1 Mbp) for the first time.

4 Discussion

While overall there were 84 DNA molecules that could be aligned to the genome with statistical significance, there were others that could not be aligned. In total we recorded 124 barcodes. On the first alignment attempt, with no modification to the barcodes, 74 DNA molecules could be aligned but 50 could not. These 50 had distributions of the estimator Δ that did not meet our criteria of significance; either they gave a value of $n_L < 0.02$ (3 molecules) or, much more frequently, the histogram of Δ could not be well-fit to a quartic function $P(\Delta)$ and so n_L could not be meaningfully calculated (the remainder of cases). What factors might prevent the successful alignment of a melting barcode? It is important to distinguish between molecules which could not be aligned for experimental reasons, because a melting barcode was not correctly formed, and those which could not be aligned for more fundamental reasons. A brief discussion of the relative importance of these factors follows.

There are several reasons that images from a denaturation mapping experiment might fail to correctly reflect a melting pattern. These experimental defects have distinct visual signatures

Downloaded by McGill University on 22 August 2012
Published on 23 July 2012 on http://pubs.rsc.org | doi:10.1039/C2LC40504K

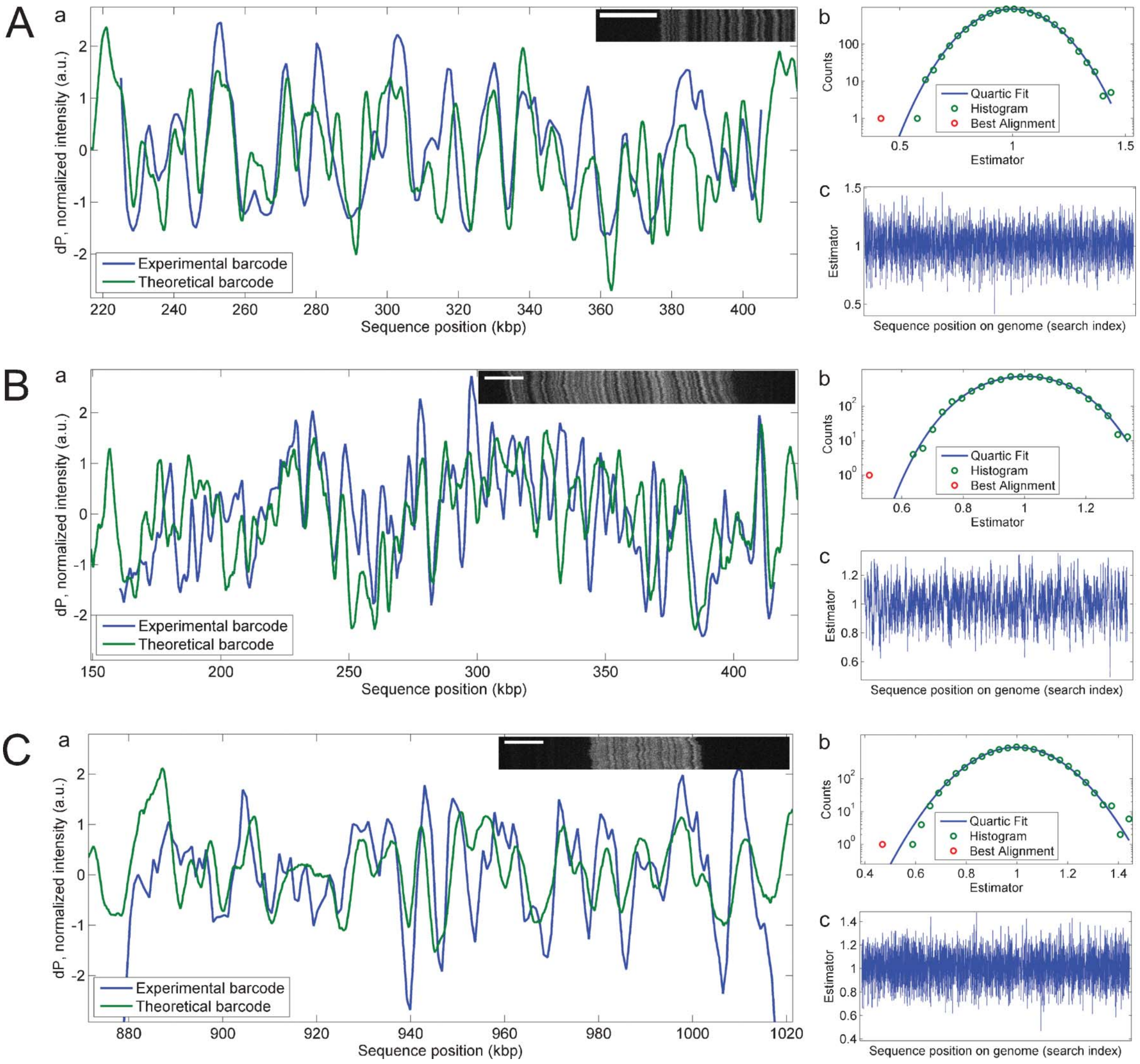

Fig. 4 Alignment of single DNA molecules to the yeast genome. Results from three DNA molecules, of length 200–250 kbp, which we locate on chromosome 10 (*A*), chromosome 16 (*B*) and chromosome 12 (*C*) respectively. (*a*) is a comparison of intensity traces between the experimental barcode (blue) and the computed barcode for the genome (green) at the sequence position where we locate the molecule. The blue trace arises from a time series of fluoresence images of a single DNA molecule (inset, scale bar 10 μm). The traces have been processed using a custom algorithm to remove background and thermal distortions to produce an adjusted barcode dP (see the Methods section for details). The profiles are first compared using a position scale of nanometres, and then plotted on a position scale of base-pairs inferred using the predicted genomic trace. (*b*) is a histogram of the distribution of least-squares estimators Δ calculated at each search position along the genomic sequence (green). The blue curve is a fit of the expected quartic distribution *P*(Δ). The red point is the global minimum of the estimator, taken at the sequence position where we locate the molecule. The fact that the red point lies above *P*(△)means the global minimum estimator is lower than would be expected by chance for uncorrelated experimental and genomic barcodes, and therefore that the location is significant. (*c*) displays the estimator calculated at each search position. Note that the search algorithm excludes genomic barcodes from chromosomes shorter than the experimental barcode. As a result long molecules are located by searching over fewer sequence positions in total, as in (*B*).

that can be diagnosed by inspecting barcodes. We observed DNA molecules that failed to melt before an image was recorded (10 molecules), entered a nanochannel while folded (9 molecules), tied into a knot before confinement (7 molecules), nicked into multiple fragments (3 molecules), and appeared faint due to photobleaching (3 molecules). In total, barcodes of 32 of the 50 DNA molecules that could not be aligned displayed these features. These molecules all had distributions of the estimator Δ that could not be fit well by the quartic function *P*(Δ). The fact that these fail to pass the goodness of fit test, which was added to the criteria for statistically significant alignment from Reisner *et al.*,[5] shows that the more rigorous approach of this work can systematically identify problematic DNA molecules and prevent them from introducing errors in an optical map.

A second alignment attempt was performed after manually removing portions containing knots, folds, or unmelted contour from some barcodes. Of the 32 poorly formed barcodes in total, 10 could be aligned successfully in the second attempt (4 molecules with knots, 3 with folds and 3 with unmelted contour). These 10 corrected barcodes have been added to the count of 74 that could be aligned on the first alignment attempt, to give the reported final count of 84 molecules.

Downloaded by McGill University on 22 August 2012
Published on 23 July 2012 on http://pubs.rsc.org | doi:10.1039/C2LC40504K

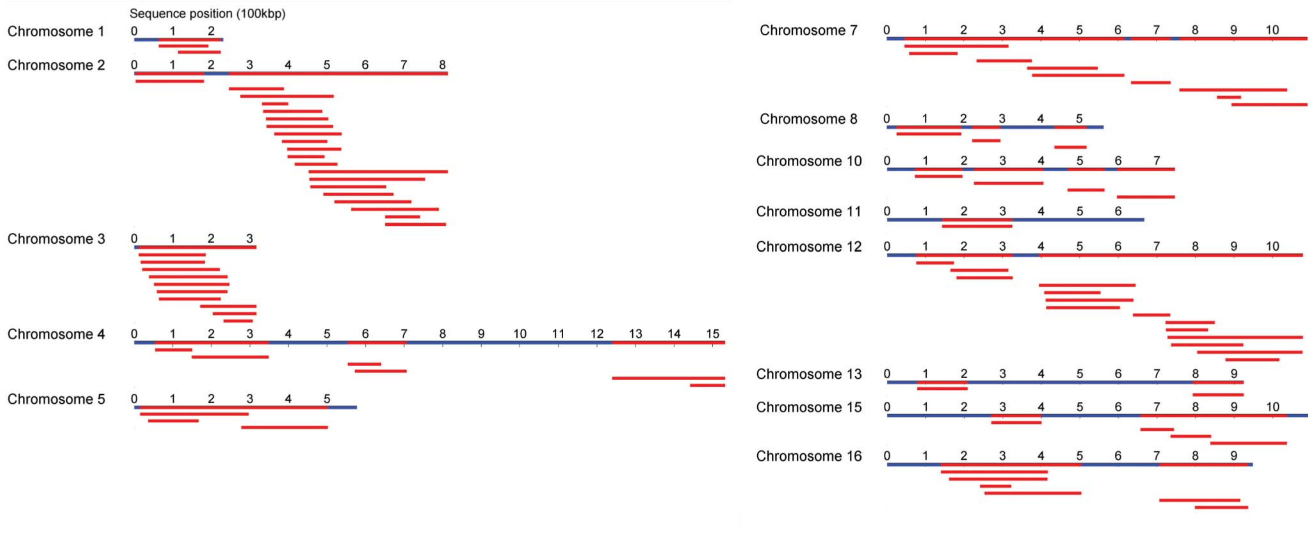

Fig. 5 Coverage of the yeast genome by alignment of melting barcodes. Blue lines represent the genomic sequence of each yeast chromosome on a position scale of 100 kbp. Overlapping red lines represent the set of sequence positions on that chromosome where a melting barcode has been located. The shorter red lines below represent the sequence positions occupied by individual DNA molecules that we locate. The optical map formed by 84 melting barcodes covers more than 50% of the yeast genome.

It may be possible to reduce the number of molecules which suffer experimental defects by making refinements in protocol and device design. Knotting and folding could be discouraged by introducing a post array before the nanochannels, nicking and photobleaching could be discouraged by optimising the oxygen scavenging system present in the buffer, and DNA molecules could be melted more reliably by preparing DNA for melting with formamide on the chip using a diffusive mixer.

Of the 50 barcodes that could not be aligned on the first attempt, 18 remain which cannot be aligned but have no obvious experimental defect that can be removed manually. This represents an alignment rate of 82% among the 102 melting barcodes which have been correctly formed (of which 84 can be aligned significantly and 18 cannot). The total alignment rate, out of all barcodes obtained (124), including those with clear experimental defects, is 67%. We suggest two reasons an apparently well-formed barcode may not align to the genome.

Firstly, the stochastic nature of melting a single DNA molecule may create discrepancies from the predicted melting profile. Occasionally discrepancies of about 5 kbp in size may appear between the two profiles that are in otherwise strong agreement. This need not necessarily prevent mapping; DNA molecules can often be aligned with statistical significance despite local discrepancies between denaturation barcodes and the theoretical prediction. Such details may reflect stochasticity in the melting process. The computed genomic barcode represents an ensemble average (in the thermodynamic sense). In Reisner *et al.*,[5] the theory is compared to a conensus-profile obtained from an average of about 40 single molecule profiles: these consensus-profiles agree well to the ensemble average profile (as we would expect). In this study, however, the theoretical prediction is compared directly to single-molecule fragments, which may have a denaturation pattern that does not exactly match the ensemble average profile. The denaturation process is co-operative and can occur in differing trajectories, each trajectory representing a different path through space of possible denaturation states (and giving rise to slightly different melting profiles). As discussed later, these discrepancies may be minimized with changes to the device design.

Secondly, there very likely exist differences in sequence between the strain of the reference sequence (S288c) and the strain used for these experiments (YPH80). The strain YPH80 is formed from crossing the strains AB972, YNN281 and A364A, with AB972 isogenic to S288C, the wild-type background used to form the sequence reference used in the comparison. Large-scale genomic changes, such as insertions, deletions and chromosomal re-arrangements, typically arise between different yeast strains.[13] These changes can be large enough to detect *via* the denaturation map and may contribute both to the number of observed DNA molecules which we cannot align and local discrepancies between theory and experimental barcodes. We believe that in light of possible genomic differences between YPH80 and S288c our current success in mapping more than half of the genome is a strong demonstration of the utility of denaturation mapping.

What might improve the number of correctly formed barcodes that align to the genome with statistical significance from the rate of 82% in our experiments? Computing a predicted barcode from the actual genomic sequence of the yeast strain YPH80, rather than from the reference strain sequence, should improve the number of experimental barcodes which align with statistical significance. We are currently sequencing YPH80 to determine the precise extent of the re-arrangements between the two strains. Extending the length of melting barcodes will also improve the statistical significance of alignments. The average fragment size imaged (167 kbp) is smaller than most of the yeast chromosomes. We believe that it should be possible to map entire yeast chromosomes with nanochannel technology given improvements in device design to discourage fragmentation of DNA molecules. For example, it may be necessary to directly extract DNA on-chip in proximity to the nanochannel inlet to prevent fragmentation of the large molecules due to hydrodynamic shear (*e.g.* induced *via* pipetting). Structures designed to make the entropic

Published on 23 July 2012 on http://pubs.rsc.org | doi:10.1039/C2LC40504K
Downloaded by McGill University on 22 August 2012

penalty to confinement more gradual, such as an array of posts of decreasing size or funnels at the interface to the nanochannels, may also help prevent breakage. Finally, it may be possible to reduce the stochastic variation in melting barcodes by controlling the conditions each DNA molecule is exposed to more precisely. Mixing DNA with stain and formamide on-chip using diffusive mixing structures, for a short and controlled time before confinement and imaging, would limit the number of melting pathways available to each DNA molecule.

In conclusion we show that it is possible to align single-molecule denaturation barcodes to a Mbp-size lower eukaryotic genome. We feel that the nanochanel-based denaturation mapping technology may play an important role in facilitating genome assembly and clarifying large-scale structural rearrangements at the level of single cells. In particular, the denaturation barcode could be used as the basis of a novel type of genomic physical partitioning (*e.g.* to isolate single fragments of DNA corresponding to a genomic region of interest). Single-molecule denaturation barcodes could be screened for identity with the genomic region of interest; molecules that match the theoretical profile could be extracted for further analysis (*e.g.* next-generation sequencing). Such a capability could substantially reduce sequencing costs in resequencing applications where only information from a target genomic loci is required.

Acknowledgements

The authors gratefully acknowledge funding from the Danish Council for Strategic Research, DSF, under grant agreement no 10-092322 (PolyNano) and support from the Natural Sciences and Engineering Research Council of Canada (NSERC) under the Discovery Grant program.

The authors also gratefully acknowledge material/financial support from the McGill-Génome Québec Innovation Center and use of the Danchip nanofabrication facility at the Technical University of Denmark. Finally, we would like to thank Said Attiya from the McGill University and Génome Québec Innovation Centre for his assistance with the SCODA apparatus.

References

1 R. K. Neely, J. Deen and J. Hofkens, *Biopolymers*, 2011, **95**, 298–311.
2 J. O. Tegenfeldt, C. Prinz, H. Cao, S. Chou, W. W. Reisner, R. Riehn, Y. M. Wang, E. C. Cox, J. C. Sturm, P. Silberzan and R. H. Austin, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 10979–10983.
3 R. Riehn, M. Lu, Y.-M. Wang, S. F. Lim, E. C. Cox and R. H. Austin, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10012–10016.
4 S. K. Das, M. D. Austin, M. C. Akana, P. Deshpande, H. Cao and M. Xiao, *Nucleic Acids Res.*, 2010, **38**, e177.
5 W. Reisner, N. B. Larsen, A. Silahtaroglu, A. Kristensen, N. Tommerup, J. O. Tegenfeldt and H. Flyvbjerg, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 13294–13299.
6 J. Shendure and H. Ji, *Nat. Biotechnol.*, 2008, **26**, 1135–1145.
7 A. Selmecki, A. Forche and J. Berman, Science (*Washington, DC, U.S.*), 2006, **313**, 367–370.
8 J. J. Salk, E. J. Fox and L. A. Loeb, *Annu. Rev. Pathol.: Mech. Dis.*, 2010, **5**, 51–75.
9 A. Marziali, J. Pel, D. Bizzotto and L. A. Whitehead, *Electrophoresis*, 2005, **26**, 82–90.
10 K. Engel, L. Pinnell, J. Cheng, T. C. Charles and J. D. Neufeld, *J. Microbiol. Methods*, 2012, **88**, 35–40.
11 E. Tøstesen, G. Jerstad and E. Hovig, *Nucleic Acids Res.*, 2005, **33**, W573–W576.
12 R. M. Wartell and A. S. Benight, *Phys. Rep.*, 1985, **126**, 67–107.
13 L. Carreto, M. F. Eiriz, A. C. Gomes, P. M. Pereira, D. Schuller and M. A. S. Santos, *BMC Genomics*, 2008, **9**, 524.