RESEARCH ARTICLE

Noise reduction in single time frame optical DNA maps

Paola C. Torche[1☯], **Vilhelm Müller**[2‡], **Fredrik Westerlund**[2‡], **Tobias Ambjörnsson**[1☯*]

1 Department of Astronomy and Theoretical Physics, Lund University, Lund, Sweden, **2** Department of Biology and Biological Engineering, Chalmers, University of Technology, Gothenburg, Sweden

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* tobias.ambjornsson@thep.lu.se

Abstract

In optical DNA mapping technologies sequence-specific intensity variations (DNA barcodes) along stretched and stained DNA molecules are produced. These “fingerprints” of the underlying DNA sequence have a resolution of the order one kilobasepairs and the stretching of the DNA molecules are performed by surface adsorption or nano-channel setups. A post-processing challenge for nano-channel based methods, due to local and global random movement of the DNA molecule during imaging, is how to align different time frames in order to produce reproducible time-averaged DNA barcodes. The current solutions to this challenge are computationally rather slow. With high-throughput applications in mind, we here introduce a parameter-free method for filtering a single time frame noisy barcode (snap-shot optical map), measured in a fraction of a second. By using only a single time frame barcode we circumvent the need for post-processing alignment. We demonstrate that our method is successful at providing filtered barcodes which are less noisy and more similar to time averaged barcodes. The method is based on the application of a low-pass filter on a single noisy barcode using the width of the Point Spread Function of the system as a unique, and known, filtering parameter. We find that after applying our method, the Pearson correlation coefficient (a real number in the range from -1 to 1) between the single time-frame barcode and the time average of the aligned kymograph increases significantly, roughly by 0.2 on average. By comparing to a database of more than 3000 theoretical plasmid barcodes we show that the capabilities to identify plasmids is improved by filtering single time-frame barcodes compared to the unfiltered analogues. Since snap-shot experiments and computational time using our method both are less than a second, this study opens up for high throughput optical DNA mapping with improved reproducibility.

OPEN ACCESS

Citation: Torche PC, Müller V, Westerlund F, Ambjörnsson T (2017) Noise reduction in single time frame optical DNA maps. PLoS ONE 12(6): e0179041. https://doi.org/10.1371/journal.pone.0179041

Editor: Mark Isalan, Imperial College London, UNITED KINGDOM

Received: November 16, 2016

Accepted: May 23, 2017

Published: June 22, 2017

Data Availability Statement: All relevant data are available from figshare. Kymographs: https://figshare.com/s/3d9d2f8840e9b8174a6d Matlab Function for Noise filtering: https://figshare.com/s/a63df3aac07777ed878d.

Funding: FW and TA acknowledge the EU Horizon 2020 program (Grant 634890, BeyondSeq, https://ec.europa.eu/programmes/horizon2020/). FW acknowledges the EuroNanoMed2 (project NanoDiaBac, http://www.euronanomed.net/) and Torsten Söderbergs Stiftelse (http://www.torstensoderbergsstiftelse.se/) for funding. TA

Introduction

In *optical DNA mapping* technologies, single DNA molecules are characterized by using fluorescent, sequence-specific, labeling. The fluorescently stained DNA molecules are stretched using surface adsorption or in nanochannels and visualized using fluorescence microscopy.

acknowledges the Swedish Research Council (grant number 2014-4305, http://www.vr.se/) for funding. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Optical DNA mapping allows to measure coarse intensity maps of DNA sequences, with resolution on the order of 1 kilobasepairs (kbp) [1]. Optical maps are often used as a complement to DNA sequencing, serving as a scaffold for assembling and validation [2–10]. Optical maps are also used in the characterization of structural variations (inversions, deletions, translocations, duplications, insertions) [11–16], which have shown in many studies correlation to human diseases [17–22]. Other applications are the detection of antibiotic resistance in plasmids from bacteria [1, 23–26], and rapid identification of bacterial species [27–31].

There are several types of fluorescent labeling techniques used in optical DNA mappings. The labeling methods may be roughly classified into two categories: (i) sparse labeling, and (ii) dense labeling. The word “label” is here used in its broadest sense and can represent a fluorescent signal, or a dark region, between fluorescent labels, for instance. Category (i) contains cases where each label can be visually (and algorithmically) identified in the optical map. Examples of sparsely labeled maps include restriction enzyme cut DNA fragmentation maps [32, 33]. In this method, fluorescently labeled DNA molecules are stretched, typically using surface adsorption, and cut at specific sequence dependent positions. When visualized in a microscope, the cut positions along the DNA appear as dark regions (the labels) in between bright regions (intact DNA). Another example of category (i) barcodes is sparse enzymatically labeled optical maps [34, 35] where the fluorescence of sequence-specifically bound molecules serve as labels. For type (ii) labeling, the sequence-dependent DNA fingerprint is instead a continuous (amplitude modulated) signal along the DNA and includes DNA melting maps [36, 37], DNA competitive binding maps [28, 38] and dense enzymatic labeling of DNA [39]. In the present study, DNA molecules are labeled using the competitive binding assay (dense labeling).

Competitive binding optical maps are created by introducing the DNA into a mixture of YOYO-1 and netropsin. Netropsin (non fluorescent) binds to DNA preferentially in AT-rich regions, while YOYO-1 (fluorescent) will bind in the regions that are left (GC-rich) since it is not sequence specific. As a consequence, GC-rich sequences are observed brighter than AT-rich sequences. The result is an alternating pattern of dark and bright regions which is a fingerprint, or barcode, of the DNA molecule. The DNA used in this article is plasmids from bacteria, which are circular DNA molecules separated from the chromosomal DNA, that can store, for instance the genes responsible for antibiotic resistance.

The stretching of DNA molecules in optical DNA mapping assays is achieved by surface adsorption or using nano-channels, where each of the two methods has its pros and cons. For instance, by using surface adsorption the DNA is perfectly still during imaging, but non-uniform stretching may occur, which introduces challenges for creating reproducible barcodes. Even though optical maps of surface adsorbed DNA has successfully used in high throughput applications [40], nanochannels are expected to be better suited for high throughput purposes, as an arbitrary number of molecules can be, essentially continuously, run through the experimental array and analyzed [1]. The main post-processing challenge for nanochannel-based methods is that the stretched DNA molecules are not completely still; thermal center-of-mass motion and conformational fluctuations still take place. These local and global longitudinal movements of the DNA introduces challenges for how to calculate reproducible time-averages of long measurements (typically, a few hundred time frames).

The currently most effective method for creating reproducible DNA barcodes from nanochannel based experiments is to use the time average of an *aligned kymograph*. A kymograph is a stack of, typically, a few hundreds of time-frames taken one after another (Fig 1a). Once such a raw kymograph has been obtained, the different time frames are aligned to each other using global shifting and local stretching [37, 41], resulting in an aligned kymograph (Fig 1b). Finally, an average over the time frames of the aligned kymograph produces a time-averaged

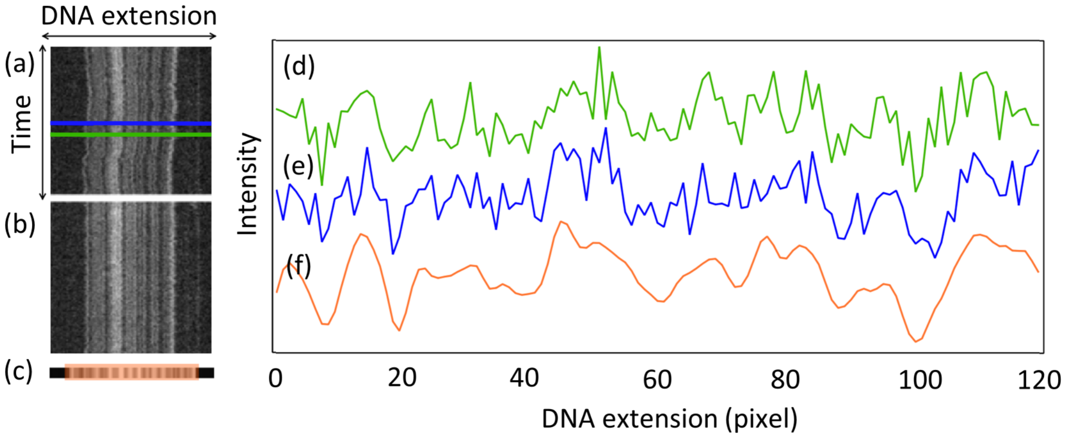

Fig 1. Examples of optical mapping kymographs (raw, aligned and time averaged) from a linearized plasmid DNA stretched in a nanochannel. (a) A raw kymograph (i.e. a stack of images of stretched and fluorescently labeled DNA) from a linearized plasmid obtained using the competitive binding assay described in [38]. The horizontal direction corresponds to the nano-channel extension (i.e., the direction of the stretched DNA) and vertical axes are different time points (0.1 s between time frames). The kymograph consists of 200 single time frame images (d, e). (b) The raw kymograph is aligned (using *WPAlign* from [41]) and subsequently averaged over all 200 time frames in order to produce (c, f) noise-reduced, time averaged DNA barcodes. The noisy curves in (d, e) represent the intensities along two single time-frame (snap-shot) barcodes, see (a). For visualization purposes, the snap-shot barcodes were shifted globally to the position where they have the maximum correlation coefficient with the time-averaged barcode. The challenge addressed in this study is how to make the noisy single-time frame barcodes of the form illustrated above resemble the (more reproducible) time-averaged barcode to a higher degree by using low-pass filtering. The barcodes shown are from plasmid *pEC005A*, see [24] for further information about the experiments.

https://doi.org/10.1371/journal.pone.0179041.g001

DNA barcode (Fig 1c and 1f), which has much less noise than any single time frame barcode (Fig 1d and 1e). Recent progress allows for reasonably efficient kymograph alignment using the WPAlign method where the computational time scales linearly with the number of pixels in the barcode [41]. For the plasmids used in this study, kymograph alignment using WPAlign takes a few minutes on a modern laptop computer, whereas for a whole human genome the alignment would take about an hour. These computational times may be prohibitive for high throughput measurements where thousands, or millions, of optical maps need to be aligned and averaged.

As an alternative to the strategy of recording movies of nano-channel stretched DNA (including subsequent alignment and time-averaging of kymographs as described above), we herein describe a new post-processing method for reducing the random noise in a *single time frame* DNA barcode (measured during 0.1s). The introduced methodology will allow experimentalists to turn single "snap-shot data" [39], into barcodes which are more similar to the time average of aligned kymographs. Thus, our method tackles the problem of how to make a single time frame barcode best "mimic" a barcode which would be obtained from a perfectly still, fluorescently labeled DNA molecule from which an infinite number of photons were collected. The method introduced is based on applying a low-pass filter on single time frame barcodes. In practice, we apply the filter by iteratively increasing the "strength" of the filtering, until all peaks have a spatial width of at least the width of the point spread function of the system (a known parameter). The use of filters for noise reduction on images has been studied numerous times [42–50], but such methods often require a calibration to find the best value for the parameters of the filter. Some other studies also aim at finding the optimal parameter, using only information from the image [51–54]. However, this study is the first to introduce a filtering method for snap-shot DNA barcodes and that is parameter free. Although the method developed herein is applied to competitive binding barcodes, our method is likely to apply essentially "as is" to all barcodes of type (ii), see above. We do not expect our method to be of use for restriction enzyme cut DNA fragmentation maps [32, 33]. On the other hand, for the case of sparse enzymatic labeling [35, 55] we believe that our method might also be effective for making a snap shot experiment better mimic the perfect time-average scenario (infinite number of photons). Possibly, however, minor modifications could be required for these barcode types.

Methods

Our method for reducing the random noise in a single time-frame DNA barcode is a post-processing method based on the application of a low-pass filter. In this section we describe and justify our method. We describe the data sets and the approach we use for validation. We also describe our approach for testing our method's plasmid identification (ID) capabilities (matching filtered single time frame barcodes to the correct plasmid from a database of plasmid theory barcodes).

Filtering method

Low-pass filters are commonly used in signal processing for reducing high-frequency noise. Examples of low-pass filters are the Window-Sinc, Moving Average and Gaussian filters. A brief description about these types of filter can be found in Section 1. Each filter has its own pros and cons but all of them require an input parameter that determines to which degree the signal, or barcode, is filtered, i.e. its filtering power. For instance, the Window-Sinc filter requires a cut-off frequency $f_{\text{cut-off}}$, the Moving Average filter needs a certain neighborhood for averaging b, and the Gaussian filter a standard deviation σ. The optimal choice for the parameters depends on the properties of the recorded signal, but is in general not immediately clear how to choose these parameter values without trial-and-error and visual inspection of the filtered signal. In order to circumvent the arbitrariness in the choice of the filtering parameter, we here base our approach on a measurable parameter, namely the FWHM (Full Width at Half Maximum) of the Point Spread Function (PSF) of the system. We here use the FWHM as the unique parameter in the filtering process by applying the low-pass filter to the single time-frame recursively, increasing its filtering power until all peaks become at least as wide as the FWHM of the PSF. The rationale behind this method is that the fluorescence from a single fluorescent point source is well-described by a Gaussian function with a standard deviation σ_{psf}. If the point source is kept perfectly still and an infinite number of photons are recorded (i.e., a perfect time-average) then this Gaussian is perfectly smooth and has a FWHM given by

$$\omega_{\text{psf}} = 2.355\sigma_{\text{psf}} \tag{1}$$

In our case $\sigma_{\text{psf}} \approx 300$ nm [28] and hence $\omega_{\text{psf}} \approx 707$ nm ≈ 4.5 pixels (one pixel is 159 nm in the experiments analyzed herein). We elaborate further on the point spread function at the end of this section. Based on the reasoning above, we note that a perfect time-average should have no peaks with FWHM smaller than ω_{psf}, which justifies our recursive filtering method as detailed below. In short, our method is:

1. Determine σ_{psf} for the experimental setup and set a threshold value, ω_{thresh}, equal to the associated FWHM of the PSF, see Eq (1), i.e., set $\omega_{\text{thresh}} = \omega_{\text{psf}}$.

2. Set the filter parameter ($f_{\text{cut-off}}$, b or σ) to its start value. The start value is chosen to correspond to essentially no filtering: for the Window-Sinc filter $f_{\text{cut-off}}$ is set to the largest allowed frequency (Nyquist frequency), for the Moving average b is set to 1 pixel and for the Gaussian filter we set σ to 1 pixel.

3. Apply the low-pass filter with the given value of the filter parameter.

4. Detect all peaks in the barcode and measure the width of each of them. The specific method we use for measuring the widths is discussed in S1 Fig.

5. Check if the FWHM ($\omega_{\text{peak},k}$) of each peak k in the filtered barcode is greater than the threshold ω_{thresh}. If indeed $\omega_{\text{peak},k} \geq \omega_{\text{thresh}}$ for all k, then we stop the algorithm. If this is not the case, update the filter parameter (decrease $f_{\text{cut-off}}$, increase b, or increase σ, respectively) and go back to step 3.

Let us make a few comments on our algorithm. First, we note that in our recursive method, the optimal parameters (e.g. $f_{\text{cut-off}}$, b or σ) are determined indirectly by setting the minimum FWHM that all peaks in the filtered barcode must have. Second, our method is not limited to the three filters presented here but any other low-pass filter can be used provided that its filtering power parameter can be controlled and increased gradually in each iteration. Third, a method for estimating the widths $\omega_{\text{peak},k}$ of the, often overlapping, peaks is needed in step 4 of our method. In S1 Fig we explain the algorithm we used for that purpose. Fourth, a perfect Gaussian form of the PSF (the expected emission pattern of a point source) does not necessarily apply to our data. For instance, during the 0.1 s of imaging for a single time frame each fluorescent molecule along the DNA will possibly move slightly due to center-of-mass motion and local conformational DNA fluctuations, thereby causing motional blurring. Moreover, the kymograph alignment procedure used here [41] is not perfect, which causes additional broadening of the peaks in time-averaged barcodes. The value used here, $\sigma_{\text{psf}} = 300$ nm, however, appears to well incorporate all these effects, as indicated by the good agreement between theory barcodes (calculated using $\sigma_{\text{psf}} = 300$ nm) and experiments [24]. Lastly, above we approximated the PSF by a Gaussian with standard deviation σ_{psf}. For such a function, Eq (1) relates the FWHM to the standard deviation of the PSF. However, high numerical aperture objectives could require a more complex expression for the PSF [56, 57]. In such a case our method still applies, but one must then use the known FWHM of this new PSF in our algorithm above.

Mathematical expressions for the filters used

Let us now describe mathematically the three low-pass filters which were introduced in Sec. Filtering method. Denote by $\boldsymbol{x}$ a vector with measured (noisy) intensity values (bold symbols denote vectors and non-bold symbols denote the components of a vector). We assume that for each pixel $j = 1, \ldots, N$, the measured intensities (the components of $\boldsymbol{x}$) can be written $x(j) = y(j) + n_y(j) + n_{\text{bg}}(j)$, where $\boldsymbol{n}_{\text{bg}}$ is a zero mean random noise vector (additive noise), $\boldsymbol{n}_y$ is signal-dependent noise, and $\boldsymbol{y}$ is the noise-free DNA barcode (perfect time average). The shot noise, $\boldsymbol{n}_y$, includes fluctuations due to the Poisson distribution of photons and electronic readout effects, whereas $\boldsymbol{n}_{\text{bg}}$ is background noise which is independent of $\boldsymbol{y}$. The shot noise is a Poisson distributed random variable [58] in general. Using the Gaussian approximation of the Poisson distribution (valid when the number of collected photons $\gg$1) we have that $n_y(j) = \eta(0, 1)\sqrt{y(j)}$, where $\eta(0, 1)$ is a Gaussian random variable with mean 0 and standard deviation 1. In the equations above, we neglected blurring effects due differences in a DNA molecule's lengths at two different times due to conformational fluctuations (these effect are reduced by stretching single-time frame barcodes to the same length as time-averaged barcodes, see Sec. Validation method).

The idea of applying a filter to the measured signal, $\boldsymbol{x}$, is to make the filtered signal, here denoted by $\tilde{\boldsymbol{y}}$, more closely resemble the perfect time average $\boldsymbol{y}$. We here only use linear filters, which mathematically are expressed as:

$$y(j) \approx \tilde{y}(j) = (\boldsymbol{h} * \boldsymbol{x})(j) = \sum_{k=1}^{N} h(j-k)\ x(k), \qquad j = 1, \ldots, N.$$

By the convolution theorem, in frequency domain this expression transforms to

$$\tilde{Y}(f_n) = H(f_n)\ X(f_n), \quad f_n = \frac{n}{N} \cdot \text{pixel}^{-1}, \quad n = -\frac{N}{2}, \ldots, 0, \ldots, \frac{N}{2} - 1,$$

where $\boldsymbol{H}$, $\boldsymbol{X}$, $\tilde{\boldsymbol{Y}}$ and $\boldsymbol{Y}$ are the Fourier transforms of $\boldsymbol{h}$, $\boldsymbol{x}$, $\tilde{\boldsymbol{y}}$ and $\boldsymbol{y}$, respectively, and $\boldsymbol{f}$ is a vector with allowed frequencies.

From the nature of the noisy (single time frame barcode) and noiseless (time averaged barcode) signals, see Fig 1, we observe that in order to improve the resemblance of the single time frame barcode to the time-averaged barcode, we must choose $\boldsymbol{H}$ to be a low-pass filter, i.e. the filter must act to reduce or eliminate the highest frequency components in $\boldsymbol{x}$. Information about the three filters used in this study can be found in Table 1.

Validation data set

We test our filtering method for reducing noise on single frame barcodes on the experimental data from [24] (DNA barcodes of bacterial plasmids obtained from competitive binding assays). The data set consists of 32 kymographs with a total 6400 single time frame barcodes. Eight of the plasmid molecules are type *pUUH* (barcodes of length 360 ± 23 pixels), eleven *pEC005A* (barcodes 127 ± 7 pixels long) and thirteen *pEC005B* (barcodes 247 ± 15 pixels long). Each single frame consists of 512 pixels, with the barcode signal in the center, and approximately the first and the last hundred pixels are the background (the number of pixels corresponding to the background depends on the length of the molecule measured). To extract the central region we use the Otsu method [59] to choose a threshold intensity to discriminate background ('0') from signal ('1'). Using this threshold we detect the edges of the signal region, as points where the signal makes '0'-'1' and '1'-'0' jumps, respectively. The intensity threshold in the Otsu method is determined by minimization of the intraclass variance of the intensity histograms (we use the histogram of the complete kymograph instead of the single frames for more robust values of the threshold). Some characteristic features of the experimental data set are summarized in S1 Table. More information about the experiment and the plasmids can be found in [24]. We apply the three filters specified in Table 1 to each of the 6400 single time frames in this data set.

Validation method

For validating that our filtering method indeed improves the resemblance of a filtered single time frame (*fst*) barcode to the associated time average of its kymograph (*ta*), we compare the Pearson correlation coefficient $\hat{C}$ between the single frame barcode and *ta* before and after applying the filter. The Pearson correlation coefficient between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined

Table 1. Summary of three low-pass filters used in this study: Gaussian, Moving average and Window-Sinc filter.

Filter	Gaussian	Moving average	Window-Sinc
Description	Assigns to each pixel the weighted average of all its neighbors, with weights defined by a Gaussian density function (standard deviation *σ*) centered at that pixel.	Assigns to each position the average of the intensity over its *b* nearest neighbors (included the position itself). By symmetry *b* is odd.	Sets to zero all the components of frequency higher than $f_{\text{cut-off}}$. Filtering function is a box-car function in frequency domain and a Sinc function in space domain.
Math. expr.	$\tilde{y}(j) = \frac{\sum_{k=1}^{N} x(k)\exp(-(k-j)^2/(2\sigma^2))}{\sum_{k=1}^{N}\exp(-(k-j)^2/(2\sigma^2))}$	$\tilde{y}(j) = \sum_{k=-(b-1)/2}^{(b-1)/2} \frac{x(j+k)}{b}$	$H[f_n] = 1$ if $\lvert f_n \rvert \leq f_{\text{cut-off}}$ and $H[f_n] = 0$ if $\lvert f_n \rvert > f_{\text{cut-off}}$
Parameter	Standard deviation, *σ*	Window size, *b*	Maximum frequency, $f_{\text{cut-off}}$

https://doi.org/10.1371/journal.pone.0179041.t001

in Eq (2).

$$\hat{C}_{x,y} = \frac{1}{N-1}\frac{\sum_{j=1}^{N}(x(d+j)-\mu_x)(y(j)-\mu_y)}{\sigma_x\sigma_y} \tag{2}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the two intensity vectors being compared (the single time-frame barcode and the kymograph time average), N is the length of the barcodes in pixels, $\mu_{\boldsymbol{x}}$ and $\mu_{\boldsymbol{y}}$ are the mean value of $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively, and $\sigma_{\boldsymbol{x}}$ and $\sigma_{\boldsymbol{y}}$ their standard deviations. As in [24] we use bit-weighting to mask the ends of the barcode, i.e. end pixels within a distance $3\sigma_{\text{psf}}$ from the edges are not included when calculating the correlation coefficient above. $\hat{C}_{x,y}$ takes values between -1 and 1, taking value 1 when the barcodes are identical, and 0 when they are uncorrelated. The quantity d is a relative global shift between the two barcodes $\boldsymbol{x}$ and $\boldsymbol{y}$ (a few pixels). This quantity is used herein since due to center-of-mass diffusion two different time frames will in general slightly displaced with respect to each other. Also, before comparing the two barcodes, we stretch/shrink the single time frame to the same length as the time average (using linear interpolation) to compensate for fluctuations in DNA barcode length between time frames. In the rest of the text, results for the Pearson correlation coefficient are at optimal shifts d (i.e., $\hat{C}$ is calculated as in Eq (2), where d is the position that maximizes $\hat{C}$) and with single time frames stretched to the same lengths as the time-averaged barcode.

Comparing experimental barcodes to database of plasmid theory barcodes

To address the question whether filtered barcodes are useful in plasmid "fingerprinting" applications we used the 3127 plasmid theory barcodes from [24]. The aim is to match an experimental plasmid barcode to this database and, ideally, identify the correct plasmid. To that end, we match experimental barcodes to the theory database and calculate $\hat{C}$ values. We only match to theory barcodes which have similar lengths as the experimental barcode (here, defined as a length within $\pm 3\sigma_{\text{length}}$, where σ_{length} is the standard deviation in the lengths of the experimental single time-frames). For every experimental barcode, we thus get a set of $\hat{C}$ values. We then turn this set into a histogram and fit to a Gumbel probability density, $\phi(\hat{C})$ [24]. Based on $\phi(\hat{C})$ we define a measure, the separability score (s-score), s-score = $\int_{\hat{C}_{\text{observed}}}^{\infty}\phi(\hat{C}')d\hat{C}'$, which quantifies how far out in the tail the observed correlation coefficient, $\hat{C}_{\text{observed}}$ (coefficient obtained by matching the experimental barcode to the "true" plasmid theory barcode), is. In [24] an analytical expression for the s-score is provided. The s-score is, by construction, in the

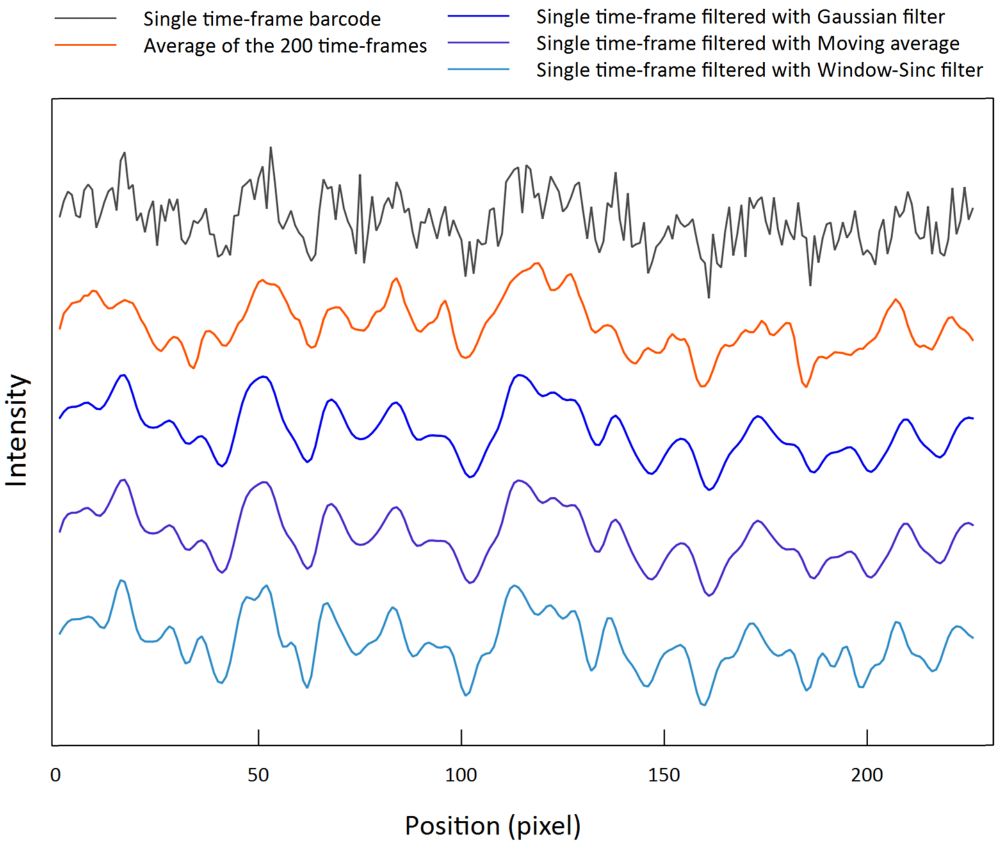

Fig 2. Example barcode filtered using our noise-reducing filtering method. (grey) A noisy single time-frame (snap-shot) barcode taken with 0.1s exposure time. (orange) The time average of the aligned kymograph. Such time-averages are used as reference ("true" barcode) throughout this study and used to judge the quality of the filtering process. (blue barcodes) From top to bottom Gaussian, Moving average and Sinc filter, respectively, is applied recursively to the single-time frame barcode (grey) until all peaks in the filtered barcode have a FWHM of at least that of the FWHM of the PSF of the system. Notice the visual similarity of the filtered barcodes to the time-averaged barcode. The Pearson correlation coefficient between the time average of the aligned kymograph and the barcode before and after the the filtering changes from 0.6 without the filtering, to 0.8 after filtering. The original raw kymograph consists of 200 single time-frame shots (exposure time 0.1s) from plasmid *pEC005B*.

https://doi.org/10.1371/journal.pone.0179041.g002

range [0, 1], and an s-score close to 0 means that the correct plasmid has a correlation coefficient far out in the tail of the histogram, and is well separated from the rest. In contrast, an s-score around 0.5 corresponds to bad separability. Note that the s-score is defined similarly to a p-value [24], and hence the s-score is an approximation to the expected number of false positive when matching the experiment to the database. As 'experimental barcodes' used for matching to the database we consider: (*ust*) unfiltered single time-frame barcodes, (*fst*) filtered single-time frame barcodes and (*ta*) time-averaged barcodes.

Results

In practice, single time-frame barcodes filtered with our method show a great degree of visual similarity to the associated time-averaged barcodes, see Fig 2 (*pEC005B* plasmid), where the three filters from Table 1 were applied to a single-time frame barcode. More examples can be found in S2 and S3 Figs. For the example in Fig 2 the Pearson correlation coefficient between the time-averaged barcode and the filtered single time frame barcode yields a value ≈ 0.8 for all the three filters. In contrast, the unfiltered and the time-averaged barcodes have a correlation coefficient of only ≈ 0.6. Thus, for this particular example, our method succeeds very well at making the filtered barcode better mimic the time-averaged barcode (33% improvement in the correlation coefficient).

In order to further quantify the apparent visual similarity between filtered and time-averaged barcodes, we applied our method to all 6400 single time-frame barcodes from the 32 kymographs (200 time frames each), see Sec. Validation data set. We find that the findings from Fig 2 hold generally, as illustrated in Fig 3, which shows the Pearson correlation coefficient (Eq (2)) between all 6400 single time frame barcodes, using filtering and no filtering, and the associated time-averaged barcodes. It is apparent that filtering does indeed significantly improve the agreement between single time frame barcodes and the true time-averaged barcode. As a minor remark, we notice (see S1 Table) that the mean intensity after background intensity subtraction (this quantity is proportional to the number of collected photons from the fluorescent labels on the DNA) is more than a factor 2 smaller for *pUUH* compared to *pEC005A* and *pEC005B*. The shot noise is hence more pronounced for the pUUH experiment, which is likely one of the major reasons that (see Fig 3) the *pUUH* molecules (first 1600 time frame barcodes) have an approximately 0.2 points lower original correlation than the *pEC005A* and *pEC005B* plasmids (time frame barcodes 1601-6400).

By averaging the results in Fig 3 over all barcodes and over the three filters we get a single number quantifying the effectiveness of our method, namely, the average change in correlation coefficient, $\Delta\hat{C}$, after and before filtering (Table 2). We find that the average value for $\Delta\hat{C}$ is 0.17 ± 0.06 points. It is noteworthy that the effectiveness of the three filter types is very similar. In S4 Fig we show results for $\Delta\hat{C}$ obtained by varying ω_{thresh}. For the choice $\sigma_{\text{psf}} = 300$ nm used here, we find that indeed the best improvement in correlation coefficient occurs close to $\omega_{\text{thresh}} \approx \omega_{\text{psf}}$ for the Gaussian and moving average filters (for the Window-Sinc filter the optimal is slightly shifted to $\omega_{\text{thresh}} \approx 1.5\omega_{\text{psf}}$). As argued in Sec. Filtering method, the choice $\sigma_{\text{psf}} = 300$ nm is a rough estimate, which includes motional broadening effects during exposure, effects due to imperfect kymograph alignment etc. Nevertheless, setting $\omega_{\text{thresh}} = \omega_{\text{psf}} = 300$ nm seems to be close to optimal.

The final (optimal) values of the filter parameters after recursive denoising are shown in S5 Fig, where we also list rough suggested values for these parameters. Note, however, that the rough suggestions must be used with care as they may be specific to the present data set (based on the competitive binding assay). In contrast, we expect the general parameter-free method introduced herein to be applicable to most types of optical maps.

Let us now investigate whether filtered single-time frame (*fst*) barcodes are useful in plasmid ID applications. To address this question we match experimental barcodes to the plasmid database and compute $\hat{C}$ and s-scores, see Sec. Comparing experimental barcodes to database of plasmid theory barcodes. The results are found in Fig 4 which show $\hat{C}$-histograms for *fst* (filtered single time-frame using the Gaussian filter) barcodes along with the correlation coefficients for the correct plasmid (vertical lines). Shown are also Gumbel fits, where for *ust* (unfiltered single time-frame) and *ta* (time averaged) barcodes we show only Gumbel fits for visual clarity. We find that that there are {320, 428, 300} theory barcodes which are similar in length (within 3 standard deviations in length compared to experiments, see Methods) for *pUUH*, *pEC005A* and *pEC005B*, respectively. For *ust* barcodes, the correlation coefficient for the correct plasmid is not always well-separated from the rest (s-scores = {0.0044, 0.10, 0.069}, which were obtained by converting the average $\hat{C}$ to a separability score, for the plasmids *pUUH*, *pEC005A* and *pEC005B*). Thus, *ust* barcodes cannot, for the plasmids considered here,

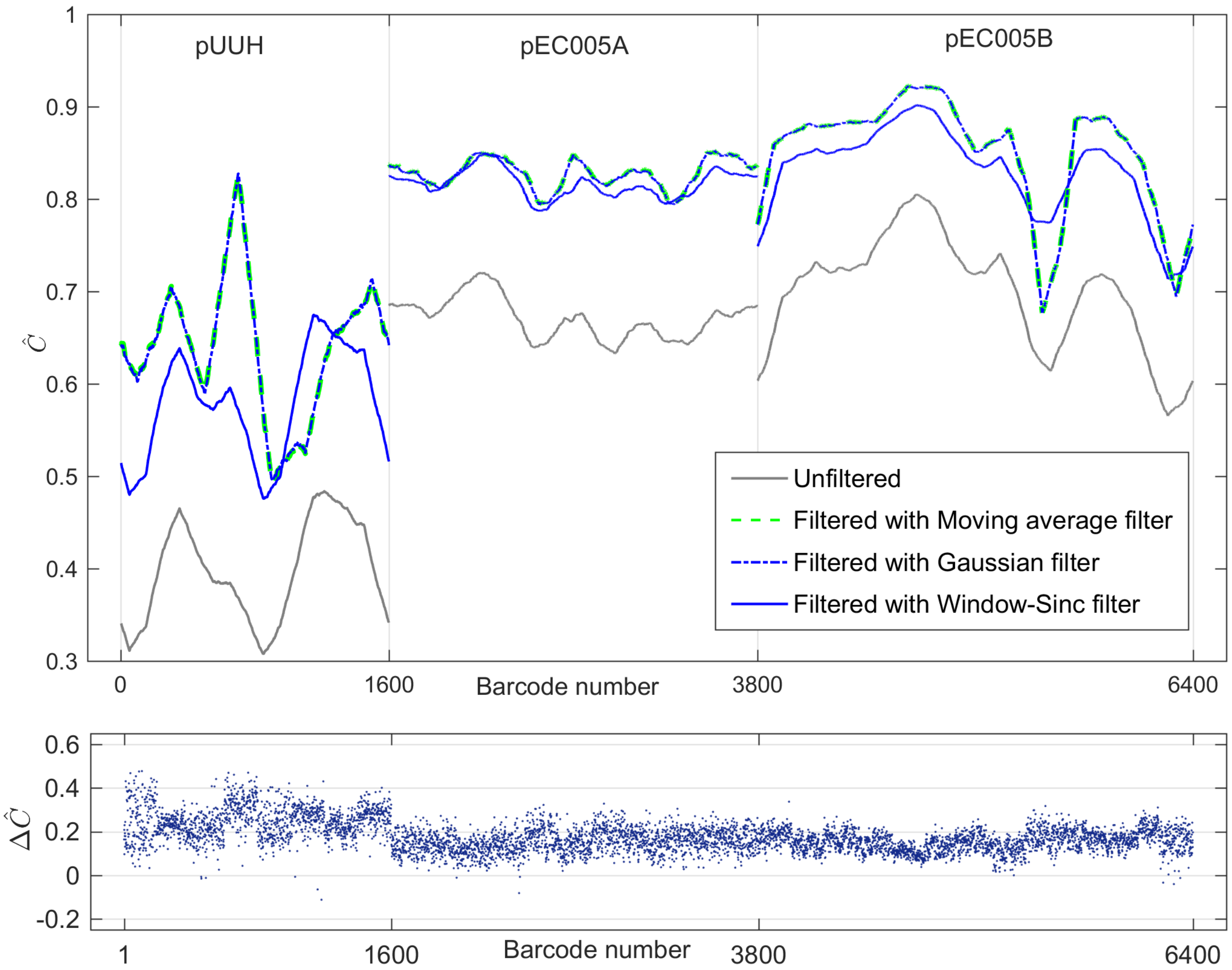

Fig 3. Change in the Pearson correlation coefficient between each single time-frame barcode and its aligned kymograph time average after application of the low-pass filter. The first 1600 barcodes originate from the *pUUH* plasmid, barcodes 1601-3800 are from plasmids *pEC005A*, and barcodes number 3801-6400 from plasmids (*pEC005B*). In grey is shown the correlation coefficients for the case without filtering, and in blue/green the correlation after filtering. The four lines in the top plot are smoothed for visualization purposes and are moving averages of the result over the 300 nearest neighbor barcodes (each of the three plasmid types treated separately). The bottom panel shows the change $\Delta\hat{C} = \hat{C}_{fst,ta} - \hat{C}_{ust,ta}$, for all 6400 single time-frames for the case of Gaussian filter. Here, $\hat{C}_{fst,ta}$ is the correlation coefficient between the filtered single time-frame (*fst*) barcode and the time averaged (*ta*) barcode, and $\hat{C}_{ust,ta}$ is the correlation coefficient between the unfiltered (noisy) single time frame barcode and the kymograph time average. Notice the rather dramatic improvement in correlation with the time-averaged ("true") barcodes after filtering. In all of the cases the filter parameter used was the (known) FWHM of the PSF of the system ω_{psf}. Here, $\omega_{psf} = 4.5$ pixels.

https://doi.org/10.1371/journal.pone.0179041.g003

reliably be used in plasmid ID applications. In contrast, *fst* barcodes are better at plasmid ID: visually the correlation coefficient for the correct plasmid ends up further out in the tail of the histogram and the associated s-scores = {0.0020, 0.027, 0.021} are indeed lower (expected fraction of false positives are thus 0.2%, 3% and 2% for *pUUH*, *pEC005A* and *pEC005B*, respectively). In fact, these s-scores are comparable to the s-scores for time-averaged barcodes ($\{5.4 \times 10^{-4}$, 0.030, 0.026$\}$).

Table 2. Increase in the Pearson correlation coefficient between filtered a single time frame and the aligned kymograph time average. The table shows the improvement, $\Delta\hat{C}$, in the Pearson correlation coefficient, $\hat{C}$ (see Eq (2)), between each single time frame barcode and its aligned kymograph time average after filtering the single frame. Correlation coefficients were averaged over all 6400 time-frame barcodes (see Fig 3) for each type of filter used (Gaussian, Moving Average and Windowed-Sinc filter). The improvement is defined as $\Delta\hat{C} = \hat{C}_{fst,ta} - \hat{C}_{ust,ta}$, where $\hat{C}_{fst,ta}$ is the correlation coefficient between the filtered single time-frame (*fst*) barcode and the time averaged (*ta*) barcode, and $\hat{C}_{ust,ta}$ is the correlation coefficient between the unfiltered (noisy) single time frame barcode and the kymograph time average. We see that all filters lead to a similar average improvement in the correlation, roughly 0.2 points. Results for $\Delta\hat{C}$ by type of plasmid (*pUUH*, *pEC005A*, *pEC005B*) are found in S1 Table in Supplementary Information.

Average improvement in $\hat{C}$ over all barcodes

Type of filter	$\langle\Delta\hat{C}\rangle = \langle\hat{C}_{fst,ta} - \hat{C}_{ust,ta}\rangle$	$\sigma_{\hat{C}_{fst,ta}-\hat{C}_{ust,ta}}$
Gaussian	0.18	0.07
Moving Average	0.18	0.07
Windowed-Sinc	0.16	0.05
Mean over all types	0.17	0.06

https://doi.org/10.1371/journal.pone.0179041.t002

Summary and discussion

We designed a post-processing method for reducing the random noise in single time frame optical DNA maps (DNA barcodes). Our method consists of applying a low-pass filter recursively until all features in the filtered signal have a width of at least the FWHM (Full Width at Half Maximum) of the optical PSF (Point Spread Function) of the system. For testing the method, we used 32 kymographs of 200 time frame barcodes each (in total 6400 noisy time frame barcodes). For quantifying the quality of the filtering, we compared each single noisy barcode and its filtered counterpart to the corresponding aligned kymograph time average ("noiseless", or "true", barcode). We used the Pearson correlation coefficient to quantify the similarity between barcodes, and we find that filtering improves the correlation coefficient by 0.17 ± 0.06 points (comparing single frames to the kymograph time average) and 0.11 ± 0.04 (comparing single frames the barcode from the theory, Fig 4, right) on average for the three different filters tested (Gaussian, Moving Average and Windowed-Sinc filter). The three low-pass filters display comparable effectiveness. We have here used three particular filters but the methodology can be applicable to other type of low-pass filters provided that one knows how to recursively increase the filtering power of the chosen filter.

Another commonly used filter for noise reduction is the Wiener filter (and different variants thereof) [60]. The Wiener filter achieves minimum squared error for additive, signal-independent noise and uses the spectral density of the background noise as input. In our case, there are *two* sources of noise. The background noise which is indeed additive and signal independent, and the shot noise, which is due to fluctuations in photon count (for finite number of photons there are deviations from the recorded signal and a perfectly smooth Gaussian) and is signal dependent. One of the conditions for the Wiener filter to achieve a least square error solution is that noise and signal have to be uncorrelated, which is not the case for shot noise. For that reason, we do not recommend the standard Wiener filter for noise reduction in single time frame DNA barcodes of the type considered here where shot noise effects are prominent. Possibly, the *Adaptative* Wiener filter [61] which uses the input parameter *neighborhood of influence* of a pixel, can be used within our framework.

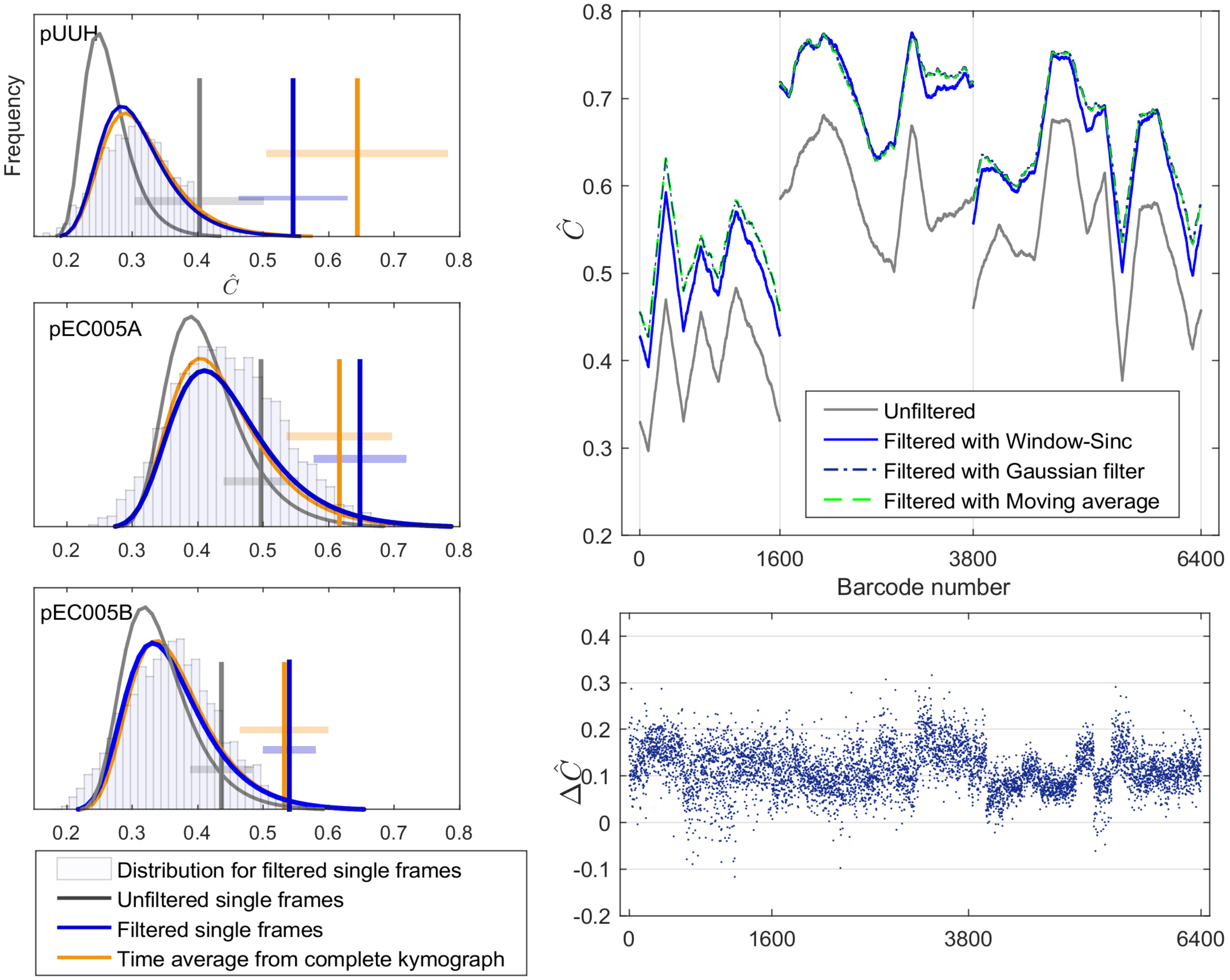

Fig 4. Plasmid ID by comparing individual experimental barcodes to a theory database. (Left) Pearson correlation coefficients between the experimental barcodes (*pUUH*, *pEC005A* and *pEC005B*) and the plasmid theory database (as in [24]) were calculated and turned into histograms. As input experimental barcode we used: unfiltered single-time frame (*ust*) barcodes, filtered (with Gaussian filter) single-time frame (*fst*) barcodes, and time-averaged (*ta*) barcodes. The vertical line gives the correlation coefficient for the correct plasmid. Notice that *ust* barcodes are not very good at plasmid ID (s-scores between 0.4-10%), but, *fsb* and *ta* barcodes are better (with s-scores less than 3%). Experiments were only matched to theory barcodes which had a length within ±3σ_{length} of the experimental barcode, with σ_{length} = {23, 7, 15} pixels for *pUUH, pEC005A and pEC005B*, respectively. The Gumbel fits to the histograms were done using the same method as in [24]. (Right) The panels on the right show the change in $\hat{C}$ between the experiment and the theoretical barcode (i.e., theory barcodes from *pUUH*, *pEC005A* or *pEC005B*, respectively) after filtering single time frame barcodes. On the top-right plot, in grey is shown the correlation coefficients for the case without filtering, and in blue/green the correlation after filtering. The four lines in the top plot are smoothed for visualization purposes and are moving averages of the result over the 300 nearest neighbor barcodes (each of the three plasmid types treated separately). The right-bottom panel shows the change $\Delta\hat{C} = \hat{C}_{filtered} - \hat{C}_{unfiltered}$ for all 6400 single time-frames for the case of Gaussian filter. On average, the single frames match to theory improves by 0.11 ± 0.04 points after filtering (average over all three filters).

https://doi.org/10.1371/journal.pone.0179041.g004

The increase in Pearson correlation coefficient using our method is rather significant but we do not reach a perfect match (correlation coefficient = 1) between time averaged barcodes and filtered barcodes. What fundamental limitations does our method have? First, we note that the DNA extension is a fluctuating quantity [62], i.e., different single time frame DNA barcodes will by necessity be of slightly different length. This effect will provide non-perfect similarity between filtered and time-averaged barcodes (we use linear interpolation to stretch single time frame barcodes to the same length as the time-average barcode). Second, local stochasticity is added to the single time frame DNA barcode since during the 0.1 s of imaging of a single frame, there will be minor local DNA fluctuations leading to random motion of the fluorescent molecules attached to it. Third, the kymograph alignment algorithm, WPAlign [41], does not produce perfectly aligned kymographs. Thus the time-averages (here used as "true" barcodes) contain deviations from the hypothetical scenario that the DNA was perfectly still and imaged for an infinitely long time. Fourth, noise fluctuations are often well described by "white" noise, i.e., all noise frequency components have equal weights. Thus, even if high frequency components are removed (as is done here), the final filtered barcode will always contain the low frequency components of the noise.

Our analysis further shows that filtered single time frame barcodes can be used for plasmid ID applications. We find that the expected fraction of false positive matches is less than 3% when matching filtered single time-frame experimental barcodes to a database of plasmid theory barcodes. This is a significant improvement compared to using unfiltered barcodes.

The results in this study originate from DNA barcodes from nano-channel based competitive binding assays. However, the method might also to be effective on single time frame DNA barcodes from other types of optical DNA mapping experiments, including barcodes with sparse enzymatic labels.

Finally, we point out that the filtering process takes less than 0.5 s per barcode on a standard laptop computer. Due to the speed with which experimental snap-shot data can be recorded (a single time frame is 0.1 s) and the real time character of our filtering methods (fraction of a second), we believe that the method we present herein will prove useful in high-throughput optical mapping applications.

Supporting information

S1 Table. Description of the experimental DNA barcode data. From left to right: I_s = mean intensity of the signal part of the barcodes (a.u.), where the errors are the standard deviation over different molecules, σ_s = the standard deviation in intensity for single time frame barcodes around the time-averaged barcode (a.u.), I_{bg} = mean intensity level of the background (a.u), σ_{bg} = standard deviation in the background intensity (a.u), length = average lengths of molecules (pixels), σ_{length} = average of standard deviations in length in a kymograph (pixels). Brackets, $\langle \ldots \rangle$, corresponds to an average over different molecules.
(PDF)

S2 Table. Increase in Pearson correlation coefficient $\hat{C}$ by type of barcode after applying our method. Change in $\hat{C}$ between the single time-frame and the aligned kymograph time average after reducing the noise with our method with the three filters (Gaussian, Moving average and Window-Sinc) independently. The values are an average over the three filters and over all time-frame barcodes of the same type (*pUUH*, *pEC005A* and *pEC005B*). $\hat{C}_{ust,ta}$ is the correlation between the unfiltered single time-frame (*ust*) barcode before any filtering and its aligned kymograph time average (*ta*). $\hat{C}_{fst,ta}$ is the correlation between the filtered barcode (*fst*)

and the aligned kymograph time average. We see that in *pUUH* barcodes (the longest DNAs) the correlation improves slightly more than for the other two barcode types (*pEC005A* and *pEC005B*).
(PDF)

S1 Fig. Estimated full width at half maximum (ω) of a peak in overlapping peaks. For estimating ω when peaks overlap, which is the case in our densely labeled barcodes, we use an algorithm implemented in Matlab Version 2015b in the function *findpeaks(PeakSig, x, 'WidthReference', 'halfheight')*, where *PeakSig* is the intensity signal, *x* the position (pixels in our case) and *'WidthReference', 'halfheight'* indicates the method for estimating the width of the peaks. In short, this function finds all maxima and estimates the full width at half maximum as follows: 1) Find all local maxima (peaks) (blue triangles). 2) The height of a peak is defined as the vertical distance between its maximum value and 0 (light blue line). 3) Detect the local minima on both sides of the peak. If a local minimum is not of intensity 0, draw a vertical line from that local minimum to 0 (blue line). 4) The estimated FWHM, ω, is then the distance, measured at half height of the peak, between the peak signal, or the drawn vertical line, from one side of the maximum of the peak to the other (red line). As shown in the examples in the figure, this method is rather successful at estimating the "true" width, i.e. 2.355σ, in a scenario of overlapping Gaussians.
(TIF)

S2 Fig. Noisy barcode from a plasmid *pEC005B* filtered using our method. (a) Time average of the aligned kymograph ("noiseless" barcode). (b) Frame 100 (measured during 0.1s) extracted from the kymograph (noisy barcode). The three barcodes (c)-(e) are the result of using our method to reduce the noise with: (c) a Gaussian filter, (d) a Moving average filter, (e) a Window-Sinc filter. The Pearson correlation coefficient between the time average of the aligned kymograph and the single frame barcode improves by $\approx$0.15 points after filtering, with any of the three filters (from 0.75 before filtering, to $\approx$0.9 afterwards).
(TIF)

S3 Fig. Noisy barcode from a plasmid *pUUH* filtered using our method. (a) Time average of the aligned kymograph ("noiseless" barcode). (b) Frame 100 (measured during 0.1s) extracted from the kymograph (noisy barcode). The barcodes (c)-(e) are the result of using our method to reduce the noise with: (c) a Gaussian filter, (d) Moving average filter, (e) a Window-Sinc filter. The Pearson correlation coefficient between the time average of the aligned kymograph and the single frame barcode improves in $\approx$0.2 points after filtering, with any of the three filters (from 0.41 before filtering, to $\approx 0.6 - 0.65$ afterwards).
(TIF)

S4 Fig. Improvement in $\hat{C}$ by ω_{thresh}. Mean value and standard deviation of the change, $\Delta\hat{C}$, in the Pearson correlation coefficient between the single time frame barcode and the aligned kymograph time average after filtering. Results are averages over the 6400 barcodes and the values of ω_{thresh} are in units of ω_{psf} (FWHM of the PSF of the system). We see that for the Gaussian and Moving average filters, using $0.5\omega_{\text{psf}} \leq \omega_{\text{thresh}} \leq 1.5\omega_{\text{psf}}$ produces the highest average improvement in the correlation, while for the Window-Sinc filter the optimal value are $\omega_{\text{psf}} \leq \omega_{\text{thresh}} \leq 2\omega_{\text{psf}}$.
(TIF)

S5 Fig. Final (optimal) values for the filter parameters. Distribution of the value of the parameters for the three filters and for $\omega_{\text{thresh}} = \omega_{\text{psf}}$. We observe that $\sigma_{\text{gaussian}} = 1.7 \pm 0.4$ pixels, $b = 3 \pm 0.4$ pixels and $f_{\text{cut-off}} = 0.18 \pm 0.03$ pixels^{-1}. For quick and simple filtering we suggest

the following choice of filter parameters: $\sigma_{\text{gaussian}} \approx \sigma_{\text{psf}}$, $b \approx 1.5\sigma_{\text{psf}}$ and $f_{\text{cut-off}} \approx 1/(\pi\sigma_{\text{psf}})$, for Gaussian, Moving Average and Window-Sinc filtering, respectively. The choice for $f_{\text{cut-off}}$ follows from a "two sigma rule" of the Fourier transform of a single Gaussian with standard deviation σ_{psf}, $\phi(x) = \exp[-x^2/(2\sigma_{psf})]/\sqrt{2\pi\sigma^2_{\text{psf}}}$: the Fourier transform of this function is $\Phi(f) = \int_{-\infty}^{\infty} \exp(2\pi ifx)\phi(x)dx = \exp[-f^2/(2S^2)]$ with $S = 1/(2\pi\sigma_{\text{psf}})$. Applying the two sigma rule, $f = f_{\text{cut-off}} = 2S$, gives our suggested value for the frequency cut-off for the Window-Sinc filter. The suggested values above are good approximations if one decides to not apply recursive filtering. However, beware that these suggested values may be specific to the present data set and not optimal for other data sets obtained using the competitive binding assay nor for other types of optical DNA maps.
(TIF)

Acknowledgments

Dr. Lena Nyberg is acknowledged for assisting with some of the experiments. The plasmids where a kind gift from Assoc. Prof. Linus Sandegren at Uppsala University. FW and TA acknowledges the EU Horizon 2020 program (Grant 634890, BeyondSeq). FW is grateful to the EuroNanoMed2 project NanoDiaBac and Torsten Söderbergs Stiftelse. TA acknowledges the Swedish Research Council (grant number 2014-4305) for funding.

Author Contributions

Conceptualization: Paola C. Torche, Vilhelm Müller, Fredrik Westerlund, Tobias Ambjörnsson.

Data curation: Paola C. Torche, Vilhelm Müller.

Formal analysis: Paola C. Torche, Tobias Ambjörnsson.

Funding acquisition: Fredrik Westerlund, Tobias Ambjörnsson.

Investigation: Paola C. Torche, Tobias Ambjörnsson.

Methodology: Paola C. Torche, Tobias Ambjörnsson.

Project administration: Tobias Ambjörnsson.

Resources: Fredrik Westerlund, Tobias Ambjörnsson.

Software: Paola C. Torche.

Supervision: Tobias Ambjörnsson.

Validation: Paola C. Torche, Tobias Ambjörnsson.

Visualization: Paola C. Torche, Vilhelm Müller.

Writing – original draft: Paola C. Torche, Tobias Ambjörnsson.

Writing – review & editing: Paola C. Torche, Vilhelm Müller, Fredrik Westerlund, Tobias Ambjörnsson.

References

1. Müller V, Westerlund F. Optical DNA Mapping in Nanofluidic Channels: Principles and Applications. Lab on a Chip. 2017;. PMID: 28098301
2. Howe K, Wood J. Using optical mapping data for the improvement of vertebrate genome assemblies. GigaScience. 2015; https://doi.org/10.1186/s13742-015-0052-y PMID: 25789164

3. Chamala S, Chanderbali S, De J, Lan T, Walts B, Albert V, et al. Assembly and Validation of the Genome of the Nonmodel Basal Angiosperm Amborella. Science. 2014; https://doi.org/10.1126/science.1241130

4. Shearer L, Anderson L, de Jong H, Smit S, Goicoechea J, Roe B, et al. Fluorescence in situ hybridization and optical mapping to correct scaffold arrangement in the tomato genome. G3 (Bethesda). 2014; https://doi.org/10.1534/g3.114.011197

5. Lewis N, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, et al. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nature Biotechnology. 2013; https://doi.org/10.1038/nbt.2624

6. Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nature Biotechnology. 2012; https://doi.org/10.1038/nbt.2478

7. Zhou S, Wei F, Nguyen J, Bechner M, Potamousis K, Goldstein S, et al. A single molecule scaffold for the maize genome. PLOS Genetics. 2009; https://doi.org/10.1371/journal.pgen.1000711

8. Nagarajan N, Read T, Pop M. Scaffolding and validation of bacterial genome assemblies using optical restriction maps. Bioinformatics. 2008; https://doi.org/10.1093/bioinformatics/btn102

9. Chen Q, Savarino S, Venkatesan M. Subtractive hybridization and optical mapping of the enterotoxigenic Escherichia coli H10407 chromosome: isolation of unique sequences and demonstration of significant similarity to the chromosome of E. coli K-12. Microbiology. 2006; https://doi.org/10.1099/mic.0.28648-0

10. Lim A, Dimalanta E, Potamousis K, Yen G, Apodoca J, Tao C, et al. Shotgun optical maps of the whole Escherichia coli O157:H7 genome. Genome Research. 2001; https://doi.org/10.1101/gr.172101

11. Lacroix J, Pelofy S, Blatche C, Pillaire M, Huet S, Chapuis C, et al. Analysis of DNA Replication by Optical Mapping in Nanochannels. Small. 2016; https://doi.org/10.1002/smll.201503795 PMID: 27624455

12. Pedersen J, Marie R, Bauer D, Rasmussen K, Yusuf M, Volpi E, et al. Fully Streched Single DNA Molecules in a Nanofluidic Chip Show Large-Scale Structural Variation. Biophysical Journal. 2013; https://doi.org/10.1016/j.bpj.2012.11.986

13. Neely R, Deen J, Hofkens J. Optical mapping of DNA: Single-molecule-based methods for mapping genomes. Biopolymers. 2011; https://doi.org/10.1002/bip.21682 PMID: 21207457

14. Conrad D, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2009; https://doi.org/10.1038/nature08516

15. Shukla S, Kislow J, Briska A, Henkhaus J, Dykes C. Optical Mapping Reveals a Large Genetic Inversion between Two Methicillin-Resistant Staphylococcus aureus Strains. Journal of Bacteriology. 2009; https://doi.org/10.1128/JB.00325-09

16. Redon R, Ishikawa S, Fitch K, Feuk L, Perry G, Andrews T, et al. Global variation in copy number in the human genome. Nature. 2006; https://doi.org/10.1038/nature05329 PMID: 17122850

17. Teo A, Verzotto D, Yao F, Nagarajan N, Hillmer A. Single-molecule optical genome mapping of a human HapMap and a colorectal cancer cell line. GigaScience. 2015; https://doi.org/10.1186/s13742-015-0106-1 PMID: 26719794

18. Weischenfeldt J, Symmons O, Spitz F, Korbel J. Phenotypic impact of genomic structural variation: insights from and for human disease. Nature Reviews Genetics. 2013; https://doi.org/10.1038/nrg3373 PMID: 23329113

19. Lee C, Scherer S. The clinical context of copy number variation in the human genome. Expert Reviews in Molecular Medicine. 2010; https://doi.org/10.1017/S1462399410001390

20. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong Association of De Novo Copy Number Mutations with Autism. Science. 2007; https://doi.org/10.1126/science.1138659 PMID: 17363630

21. Aitman T, Dong R, Vyse T, Norsworthy P, Johnson M, Smith J, et al. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature. 2005; https://doi.org/10.1038/nature04489

22. Feuk L, Carson A, Scherer S. Structural variation in the human genome. Nature Reviews Genetics. 2006; https://doi.org/10.1038/nrg1767 PMID: 16418744

23. Müller V, Rajer F, Frykholm K, Nyberg LK, Quaderi S, Fritzsche J, et al. Direct identification of antibiotic resistance genes on single plasmid molecules using CRISPR/Cas9 in combination with optical DNA mapping. Scientific Reports. 2016; 6.

24. Nyberg L, Quaderi S, Emilsson G, Karami N, Lagerstedt E, Müller V, et al. Rapid identification of intact bacterial resistance plasmids via optical mapping of single DNA molecules. Nature. 2016; https://doi.org/10.1038/srep30410

25. Müller V, Karami N, Nyberg L, Pichler C, Torche Pedreschi P, Quaderi S, et al. Rapid tracing of resistance plasmids in a nosocomial outbreak using optical DNA mapping. ACS Infectious Diseases. 2016; https://doi.org/10.1021/acsinfecdis.6b00017

26. Ramirez M, Adams M, Bonomo R, CentrÃn D, Tolmasky M. Genomic Analysis of Acinetobacter baumannii A118 by Comparison of Optical Maps: Identification of Structures Related to Its Susceptibility Phenotype. Antimicrobial Agents and Chemotherapy. 2011; https://doi.org/10.1128/AAC.01595-10 PMID: 21282446

27. Chapleau R, Baldwin J. Optical Whole-Genome Restriction Mapping as a Tool for Rapidly Distinguishing and Identifying Bacterial Contaminants in Clinical Samples. Journal of Clinical and Diagnostic Research. 2015; https://doi.org/10.7860/JCDR/2015/13983.6408

28. Nilsson A, Emilsson G, Nyberg L, Noble C, Stadler L, Fritzsche J, et al. Competitive binding-based optical DNA mapping for fast identification of bacteria-multi-ligand transfer matrix theory and experimental applications on escherichia coli. Nucleic Acids Research. 2014; https://doi.org/10.1093/nar/gku556

29. Lavigne J, Vergunst A, Goret L, Sotto A, Combescure C, Blanco J, et al. Virulence Potential and Genomic Mapping of the Worldwide Clone Escherichia coli ST131. PLoS ONE. 2012; https://doi.org/10.1371/journal.pone.0034294

30. Jackson S, Kotewicz M, Patel I, Lacher D, Gangiredla J, Elkins C. Rapid Genomic-Scale Analysis of Escherichia coli O104:H4 by Using High-Resolution Alternative Methods to Next-Generation Sequencing. Applied and Environmental Microbiology. 2012; https://doi.org/10.1128/AEM.07464-11

31. Schwan W, Briska A, Stahl B, Wagner T, Zentz E, Henkhaus J, et al. Use of optical mapping to sort uropathogenic Escherichia coli strains into distinct subgroups. Microbiology. 2010; https://doi.org/10.1099/mic.0.033977-0 PMID: 20378655

32. Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang YK. Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science. 1993; 262(5130):110–114. https://doi.org/10.1126/science.8211116 PMID: 8211116

33. Meng X, Benson K, Chada K, Huff E, Schwartz D. Optical Mapping of lambda bacteriophage clones using restriction endonucleases. Nature Genetics. 1995; https://doi.org/10.1038/ng0495-432 PMID: 7795651

34. Jo K, Schramm TM, Schwartz DC. A Single-Molecule Barcoding System using Nanoslits for DNA Analysis. Micro and Nano Technologies in Bioanalysis: Methods and Protocols. 2009;p. 29–42. https://doi.org/10.1007/978-1-59745-483-4_3

35. Neely RK, Deen J, Hofkens J. Optical mapping of DNA: Single-molecule-based methods for mapping genomes. Biopolymers. 2011; 95(5):298–311. https://doi.org/10.1002/bip.21579 PMID: 21207457

36. Østergaard P, Lopacinska-Jørgensen J, Pedersen J, Tommerup N, Kristensen A, Flyvbjerg H, et al. Optical mapping of single-molecule human DNA in disposable and mass-produced all-polymer devices. Journal of Micromechanics and Microengineering. 2015; https://doi.org/10.1088/0960-1317/25/10/105002

37. Reisner W, Larsen N, Silahtaroglu A, Kristensen A, Tommerup N, Tegenfeldt J, et al. Single-molecule denaturation mapping of DNA in nanofluidic channels. Proceedings of the National Academy of Sciences of the United States of America. 2010; https://doi.org/10.1073/pnas.1007081107 PMID: 20616076

38. Nyberg L, Persson F, Berg J, Bergström J, Fransson E, Olsson L, et al. A single-step competitive binding assay for mapping of single DNA molecules. Biochemical and Biophysical Research Communications. 2011; https://doi.org/10.1016/j.bbrc.2011.11.128 PMID: 22166208

39. Grunwald A, Dahan M, Giesbertz A, Nilsson A, Nyberg L, Weinhold E, et al. Bacteriophage strain typing by rapid single molecule analysis. Nucleic Acids Research. 2015; https://doi.org/10.1093/nar/gkv563 PMID: 26019180

40. Teague B, Waterman MS, Goldstein S, Potamousis K, Zhou S, Reslewic S, et al. High-resolution human genome structure by single-molecule analysis. Proceedings of the National Academy of Sciences. 2010; 107(24):10848–10853. https://doi.org/10.1073/pnas.0914638107

41. Noble C, Nilsson A, Freitag V, Beech J, Tegenfeldt J, Ambjörnsson T. A fast and scalable symograph alignment algorithm for nanochannel-based optical DNA mappings. PLoS ONE. 2015; https://doi.org/10.1371/journal.pone.0121905

42. Haider S, Cameron A, Siva P, Lui D, Shafiee M, Boroomand A, et al. Fluorescence microscopy image noise reduction using a stochastically-connected random field model. Scientific Reports. 2016; https://doi.org/10.1038/srep20640

43. Phelippeau H, Talbot H, Bara S, Akil M. Shot-noise adaptive bilateral filter. International Conference on Signal Processing. 2008; https://doi.org/10.1109/ICOSP.2008.4697265

44. Kervrann C, Boulanger J, Coupe P. Bayesian Non-Local Means Filter and Image Redundancy and Adaptive Dictionaries for Noise Removal. Proceedings of the 1st international conference on Scale space and variational methods in computer vision. 2007;p. 520–532.

45. Faraji H, MacLean W. CCD Noise Removal in Digital Images. IEEE Transactions on Image Processing. 2006; https://doi.org/10.1109/TIP.2006.877363 PMID: 16948312

46. Chambolle A. An algorithm for total variation minimization and applications. Journal of Mathematical Imaging and Vision. 2004; https://doi.org/10.1023/B:JMIV.0000011325.36760.1e

47. Pizurica A, Philips W, Lemahieu I, Acheroy M. A Versatile Wavelet Domain Noise Filtration Technique for Medical Imaging. IEEE Transactions on Medical Imaging. 2003; https://doi.org/10.1109/TMI.2003.809588 PMID: 12760550

48. Fiore L, Corsini G, Geppetti L. Application of non-linear filters based on the median filter to experimental and simulated multiunit neural recordings. Journal of Neuroscience Methods. 1996; https://doi.org/10.1016/S0165-0270(96)00116-1 PMID: 9007757

49. Alvarez L, Lions P, Morel J. Image Selective Smoothing and Edge Detection by Nonlinear Diffusion. II. SIAM Journal on Numerical Analysis. 1992; https://doi.org/10.1137/0729052

50. Pitas I, Venetsanopoulos A. Order statistics in digital image processing. Proceedings of the IEEE. 1992; https://doi.org/10.1109/5.192071

51. Okada M, Ishikawa T, Ikegaya Y. A Computationally Efficient Filter for Reducing Shot Noise in Low S/N Data. PLoS ONE. 2016; https://doi.org/10.1371/journal.pone.0157595

52. Boulanger J, Kervrann C, Bouthemy P, Elbau P, Sibarita J, Salamero J. Patch-Based Nonlocal Functional for Denoising Fluorescence Microscopy Image Sequences. IEEE Transactions on Medical Imaging and Institute of Electrical and Electronics Engineers. 2010; https://doi.org/10.1109/TMI.2009.2033991

53. Kervrann C, Boulanger J. Optimal Spatial Adaptation for Patch-Based Image Denoising. IEEE Transactions on Medical Imaging. 2006; https://doi.org/10.1109/TIP.2006.877529

54. Mrazek P, Navara M. Selection of optimal stopping time for nonlinear diffusion filtering. International Journal of Computer Vision. 2003; https://doi.org/10.1023/A:1022908225256

55. Vranken C, Deen J, Dirix L, Stakenborg T, Dehaen W, Leen V, et al. Super-resolution optical DNA Mapping via DNA methyltransferase-directed click chemistry. Nucleic acids research. 2014; 42(7):e50–e50. https://doi.org/10.1093/nar/gkt1406 PMID: 24452797

56. Stallinga S, Rieger B. Accuracy of the Gaussian Point Spread Function model in 2D localization microscopy. Optics Express. 2010; https://doi.org/10.1364/OE.18.024461 PMID: 21164793

57. Richards B, Wolf E. Electromagnetic diffraction in optical systems. II. Structure of the image field in an aplanatic system. Proceedings of the Royal Society London. 1959; https://doi.org/10.1098/rspa.1959.0200

58. Loudon R. The Quantum Theory of Light. OUP Oxford; 2000. Available from: https://books.google.de/books?id=AEkfajgqldoC

59. Otsu N. A threshold selection method from gray-level histograms. Automatica. 1975; 11(285-296):23–27. https://doi.org/10.1038/srep37938

60. Easton RL Jr. Fourier methods in imaging. John Wiley & Sons; 2010.

61. Ponomarev VI, Pogrebniak AB. Adaptive Wiener filter implementation for image processing. In: Mathematical Methods in Electromagnetic Theory, 1996., 6th International Conference on; 1996. p. 211–214. https://doi.org/10.1109/MMET.1996.565693

62. Iarko V, Werner E, Nyberg L, Müller V, Fritzsche J, Ambjörnsson T, et al. Extension of nanoconfined DNA: Quantitative comparison between experiment and theory. Physical Review E. 2015; 92(6):062701. https://doi.org/10.1103/PhysRevE.92.062701