

PROCEEDINGS

Open Access

Origin of multiple periodicities in the Fourier power spectra of the *Plasmodium falciparum* genome

Miriam CS Nunes¹, Elizabeth F Wanner², Gerald Weber^{3*}

From 6th International Conference of the Brazilian Association for Bioinformatics and Computational Biology (X-meeting 2010)
Ouro Preto, Brazil. 15-18 November 2010

Abstract

Background: Fourier transforms and their associated power spectra are used for detecting periodicities and protein-coding genes and is generally regarded as a well established technique. Many of the periodicities which have been found with this method are quite well understood such as the periodicity of 3 nt which is associated to codon usage. But what is the origin of the peculiar frequency multiples $k/21$ which were reported for a tiny section of chromosome 2 in *P. falciparum*? Are these present in other chromosomes and perhaps in related organisms? And how should we interpret fractional periodicities in genomes?

Results: We applied the binary indicator power spectrum to all chromosomes of *P. falciparum*, and found that the frequency overtones $k/21$ are present only in non-coding sections. We did not find such frequency overtones in any other related genomes. Furthermore, the frequency overtones were identified as artifacts of the way the genome is encoded into a numerical sequence, that is, they are frequency aliases. By choosing a different way to encode the sequence the overtones do not appear. In view of these results, we revisited early applications of this technique to proteins where frequency overtones were reported.

Conclusions: Some authors hinted recently at the possibility of mapping artifacts and frequency aliases in power spectra. However, in the case of *P. falciparum* the frequency aliases are particularly strong and can mask the $1/3$ frequency which is used for gene detecting. This shows that albeit being a well known technique, with a long history of application in proteins, few researchers seem to be aware of the problems represented by frequency aliases.

Background

The detection and analysis of repetitions in genomes is one of the most recurrent problems in computational biology. It is of technological importance since it imposes significant limitations to next-generation sequencing technologies [1] and is related to numerous properties of the genome [2-5]. As a consequence it has sparked many approaches to detect and visualise genomic repeats [6]. Discrete Fourier transform (DFT) is one

of this approaches, which is employed for the detection of genome wide non-exact periodic structures such as tandem repeats. In essence, the symbolic genome sequence is translated into numeric series which are then handled as a biological signal. The method was originally proposed for detecting periodicities in proteins [7-11] but has since gained popularity for genomes [12-18]. The genome of *Plasmodium falciparum*, the malaria parasite, is quite remarkable: it has an unusually high AT-content (80% or more) [19] and lacks identified transposable elements [20]. Sharma *et al.* [21] reported a complete frequency comb of all overtones of the basic frequency $1/21 \text{ nt}^{-1}$ for chromosome 2 of *P. falciparum*,

* Correspondence: gweberbh@gmail.com

³Department of Physics, Federal University of Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil

Full list of author information is available at the end of the article

that is, all frequencies of type $k/21 \text{ nt}^{-1}$ with $k = 1, 2, \dots, 10$. Could this be yet another peculiarity of the *P. falciparum* genome? Such a frequency comb envisages several questions such as its biological origin and function, that were not addressed by Sharma *et al.* [21] and have remained unanswered, to the best of our knowledge. Additionally, it also gives rise to an important conceptual question: how should we understand fractional periodicities such as $21/2$ or $21/4$ given that there are no fractional nucleotides in our sequence? Or could it be that these multiples are so called frequency aliases [22,23]? In this work we analyse these questions more closely. First, we apply the Fourier transform to all chromosomes of *P. falciparum* and confirm that the $1/21 \text{ nt}^{-1}$ overtone frequency comb is found in all but one chromosomes and present its location in the genome. But perhaps most importantly, we have established that this frequency comb is indeed composed of frequency aliases, and as such is an artifact of the way the genome is converted from a symbolic into a numeric sequence. While this resolves our current conceptual problem, it poses an important warning for the blind use of this technique, especially if used to detect genes. Should one of the frequency multiples coincide with the $1/3$ frequency this could lead to important problems with the detection of coding regions [18,24-27]. We show that this is indeed the case for the $7/21$ frequency of *P. falciparum*, which could cause false positive candidates for coding regions.

Results and Discussion

In Fig. 1 we present the binary indicator power spectra for all chromosomes of *P. falciparum* separated by coding

(CDS) and non-coding (non-CDS) regions. The $k/21$ frequency comb can be seen only for the non-coding sections of chromosomes 1–13 in Fig. 1b and is notably absent in chromosome 14. No chromosomes show this particular frequency comb in the coding section, Fig. 1a. However, chromosome 5 displays a slightly different frequency comb of $k/18$. Strong $1/3 \text{ nt}^{-1}$ peaks in Fig. 1a and the lack thereof in Fig. 1b confirms that this genome appears to be well annotated [18,24-27].

Further refining the origin of the $k/21$ frequency comb shows that it comes mostly from regions at the extremes of the chromosomes, as shown in Fig. 2a, where we indicate the position dependent power spectra intensities for each of the $k/21$ overtones in chromosome 1. In contrast, the $k/18$ frequency comb originates from various regions of the genome as one would expect for CDS regions.

To identify the actual sequences involved in the $k/21$ frequency overtones, we extracted the first and last 4000 nt of each chromosome and analysed the repetitions with Tandem Repeats Finder [28]. The consensus sequences are shown in table 1, which indicates single occurrences of dimers CC or GG. For instance, chromosome 1 has one GG dimer, which is the only occurrence of guanine in this sequence. Therefore, this dimer will repeat with frequency $1/21 \text{ nt}^{-1}$ and explains the fundamental frequency $1/21 \text{ nt}^{-1}$ which we observe in Fig. 1b.

However, we still do not know where the multiples of this frequency originate. The answer to this lies in a closer inspection of the repeated consensus sequences in table 1. Let's take sequence TAAGACCTATGTTAG-TAAAG for chromosome 6 which has a double

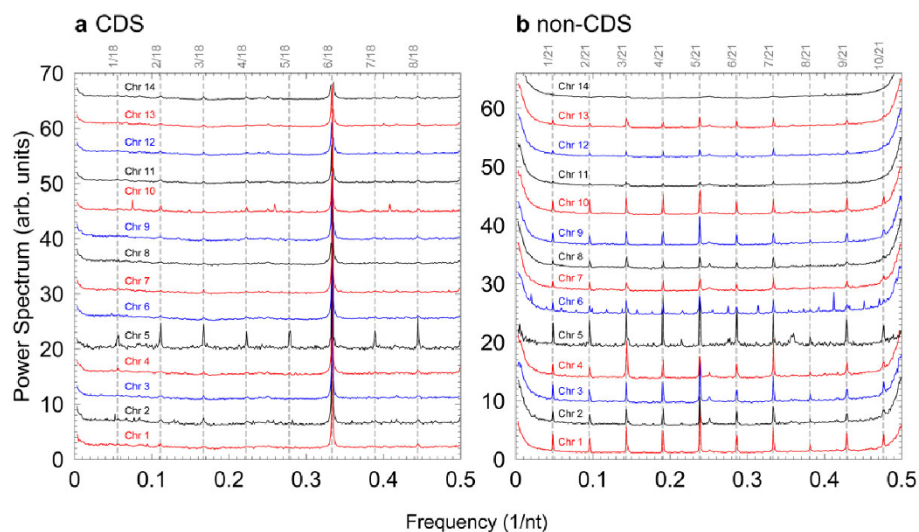


Figure 1 Binary indicator power spectra for all chromosomes of *P. falciparum*. The spectra were calculated separately for a) coding and b) non-coding region. The overlapping of the $1/3$ peaks results from enlarging the vertical axis scale to allow visualising details of non- $1/3$ peaks. Vertical dashed lines show the positions corresponding to a) $k/18$ and b) $k/21$ frequency multiples. For clarity the curves were shifted vertically by arbitrary amounts.

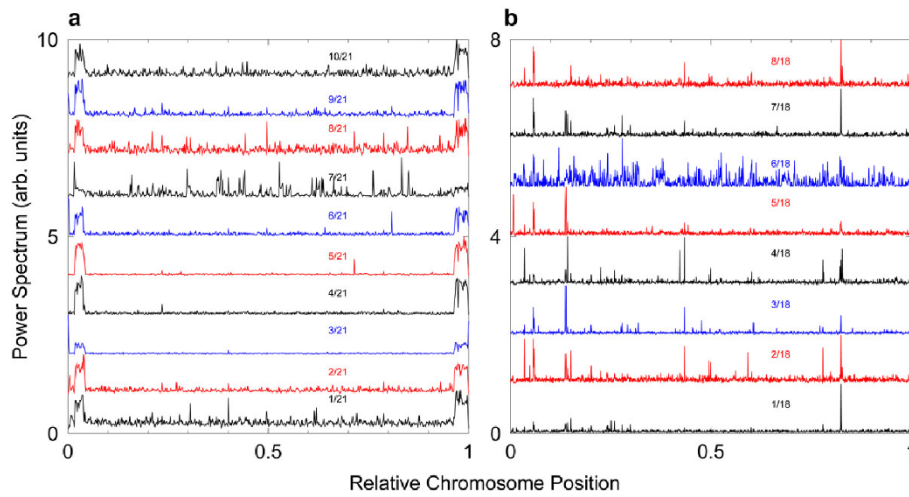


Figure 2 Position-dependent power spectra of *P. falciparum*. Part a) is for chromosome 1 and its $k/21$ frequency multiples and b) is for chromosome 5 and $k/18$ frequency multiples. A sliding window of $w = 1000$ nt was used. For clarity the curves were shifted vertically by arbitrary amounts.

cytosine. The four binary indicator sequences which result from converting the symbolic sequence are shown in table 2. The binary sequence for cytosine in table 2 displays two consecutive ones followed by zeroes only.

In other words, we have two digits one followed by 19 zeroes, repeated hundreds of times. This is essentially analogous to a Fourier integral of a train of Kronecker delta functions, that is, a sequence of pulses with period

Table 1 Consensus sequences for periodicity 21 nt.

Chromosome	Relative Position	Consensus sequence	Repetitions
1	0.0179–0.0383	CCTAACTAACATAGGTCTTAA	622.8
	0.965–0.997	AGTTAGTTAAGTTAAGACCTA	994.3
2	0.0146–0.0223	ACTAAGTCTAGGTCTTAACCTA	348.2
	0.981–0.986	TTAGTTAAGTTAAGACCTAA [†]	264.3
3	0.0126–0.0236	GGTCTTACTTCACTAACATA [†]	552.2
	0.0157–0.0234	ATAGGTCTTAACTTAACTAAC	417.6
	0.0121–0.0283	TAGGTCTTAATTAACCTAACT [†]	817.5
4	0.986–0.988	TAAGACCTAAGTTAGTGAAGT	78.6
	0.0139–0.0221	AGGTCTTACTTCACTAACTT [†]	472.4
5	0.986–0.991	TTAGTTAAGTTAAGACCTAAG	238.2
	0.0087–0.0129	ACTAACATAGGTCTTACTTC [†]	269.8
6	0.977–0.979	TAAGACCTATGTTAGTAAAG [†]	152.7
	0.982–0.990	AAAGTTAAGACCTAAGTTAGT	581.6
7	0.991–0.999	TTAAGACCTAAGTTAGTGAAG	543.8
	0.0077–0.0119	ACTAGGTCTTAACTTAACTA [†]	269.6
8	0.988–0.991	AAGACCTAAGTTAGTAAAGT [†]	185.4
	0.978–0.985	GACCTATGTTAGTAAAGTAA [†]	551.6
9	0.987–0.993	TAAGACCTAAGTTAGTGAAG [†]	413.4
	0.0073–0.0116	TAGGTCTTACTTTAACTAACT	344.4
10	0.981–0.990	GACCTATATTAGTTAAGTTAA	737.5
	0.0072–0.0093	TTACTAACATAGGTCTTAACT [†]	202.3
12	0.0049–0.0055	TAAGTTAGGTCTTAACTTAC	67.5
	0.991–0.995	TATGTTAGTTAAGTTAAGACC	354.0
13	0.0049–0.0069	ACTAACATAGGTCTTACTTC [†]	269.8
	0.990–0.991	AGTAAGACCTTAGTTAGTGA [†]	192.5
	0.991–0.993	GACCTATGTTAGTGAAGTTAA	345.2

Main tandem repetitions of periodicity 21 nt found in the *P. falciparum* genome, except for chromosome 14 from which no such periodicities were detected.

[†]Some consensus sequences provided by Tandem Repeats Finder [28] are of size 20 nt, yet they all correspond to the periodicity of 21 nt.

Table 2 Example of sequence binary mapping.

	T	A	A	G	A	C	C	T	A	T	G	T	T	A	G	T	A	A	A	G
n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$u_{A,n}$	0	1	1	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1	1	0
$u_{C,n}$	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
$u_{G,n}$	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1
$u_{T,n}$	1	0	0	0	0	0	0	1	0	1	0	1	1	0	0	1	0	0	0	0

T , which can be shown to result in another train of Kronecker delta functions with period $1/T$ in frequency space [29]. Therefore, the overtones $k/21$ do not correspond to real periods $21/k$ but are just artifacts of the way the symbolic sequence was mapped into binary sequences. Therefore, the frequencies $2/21, 3/21 \dots, 10/21$ are frequency aliases of the frequency $1/21$ [22,30].

In Fig. 3 we illustrated this effect by building an artificial sequence made of 80 repetitions of the basic unit of size 21 nt containing one or more consecutive cytosines. For ten cytosines we obtain exactly one peak at $1/21 \text{ nt}^{-1}$ as expected. However, as we gradually reduce the number of consecutive cytosines, while maintaining the size of the repeating unit constant to 21 nt, we also obtain more of the $k/21$ overtones. Eventually, for one or two cytosines, all overtones $k/21$ are present, that is, we obtain all fractional periodicities although only one cytosine repeats every 21 positions.

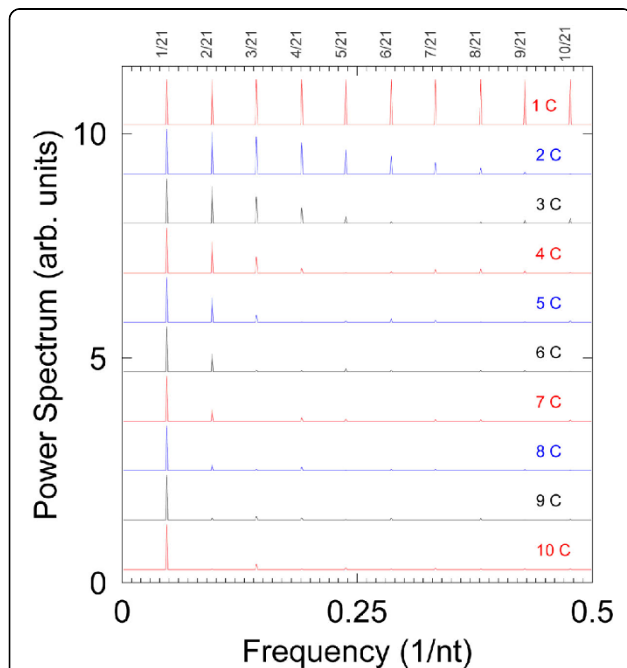


Figure 3 Binary indicator power spectra for an artificially constructed sequence. The sequence was constructed to contain 80 repeated units of size 21 nt. The spectra are for cytosine only ($\alpha = C$) with varying numbers of consecutive cytosine nucleotides while keeping the periodicity constant at 21 nt.

One simple way to verify if these overtones are artifacts of the numerical mapping is to change the mapping. As an example, we computed the power spectra using DNA flexibility mapping, as detailed in Methods section. In this case the symbolic sequence is converted into a single numerical sequence and instead of using only zeroes and ones we have now ten different numerical values. This particular mapping was selected because converting di-nucleotides allows us to choose over a larger set of numerical values resulting into an overall smoother numerical profile along the sequence. Therefore the occurrence of delta-like structures in the numerical sequence should be less likely. The power spectra for both coding and non-coding sections of *P. falciparum* are shown in Fig. 4. As expected, the frequency overtones are now barely noticeable for a few chromosomes only. Most chromosomes show little or no overtones at all. We have established that the occurrence of overtones is the result of one or two isolated nucleotides surrounded by a large number of different nucleotides and very frequently repeated. But what is the minimal size of the repeating unit for which overtones may start to appear? We carried out numerical tests using a simple repeating unit CA_n with $n = 1, 2, \dots, 9$, corresponding to a periodicity of $p = n + 1$ (data not shown), and overtones were present in all cases.

Conclusions

It was suggested that the mapping of the symbolic sequence into one or more numeric sequences could give rise to artifacts when signal processing techniques are applied [31]. In this case, the frequency comb results as an artifact of dividing one symbolic sequence into four separate sequences, that is, it is an artifact of the multichannel analysis. Indeed, for proteins fractional periodicities were already observed in the pioneering work by McLachlan and Stewart [7,9]. The sequence mapping used some amino acid properties and is similar to the flexibility mapping outlined in the Methods section. In this case, the authors merely state that the periodicity does not need to be an integer value [7,9]. While non-integral periods for certain mappings may be plausible, multiple overtones are harder to explain in this way. Subsequent works, such as McLachlan *et al.* [32], explicitly use the multichannel mapping and clearly

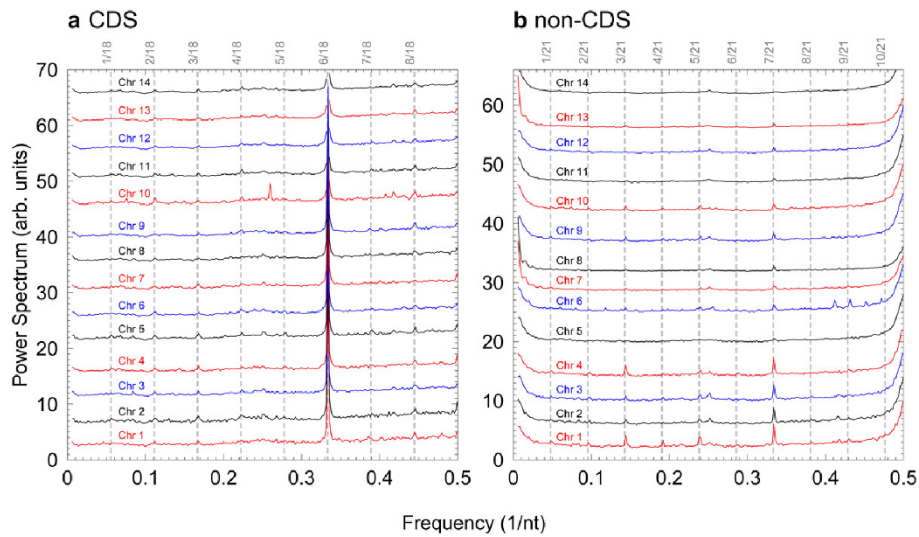


Figure 4 Flexibility power spectra for each chromosome of *P. falciparum*. Part a) is for coding regions and b) for non-coding regions. Also shown are the vertical dashed lines for the frequency multiples which are detected in Fig. 1. For clarity the curves were shifted vertically by arbitrary amounts.

show frequency combs, see Fig. 3 of [32]. Since only total spectra were published [32], we recalculated the multichannel Fourier transform for *Dictyostelium discoideum* for each amino acid. Fig. 5 shows the total Fourier spectrum and the detailed spectra for some amino acids with important contributions towards the frequency comb. We confirmed that the frequency comb is the same type of frequency alias as seen for the genome of *P. falciparum*. For instance, glutamic acid (E) alone is sufficient to account for the $k/28$ multiples of the power spectra presented by McLachlan *et al.* [32].

While nucleotide binary indicator mapping separates the symbolic genomic sequence into four binary sequences, they are separated into 20 sequences for amino acids. Therefore, the likelihood that an amino acid appears repeatedly and in isolation, is much higher than for nucleotides. It is therefore not surprising that for almost every work we encountered, power spectra in amino acid sequences were reported with multiple frequencies when binary mapping was used [7-9,32-35]. What remains somewhat of an open question is the absence of any explanation for the frequency multiples and fractional periodicities in all these works.

The presence of strong frequency aliases in the power spectra of so many chromosomes is presently unique to *P. falciparum*. However, considering the accelerated pace of genomic sequencing, we cannot rule out further occurrences of such frequency aliases in other genomes. In this work we stress the need for a careful evaluation of frequency multiples if they appear in Fourier power spectra since most available bioinformatics applications do not distinguish these aliases from genuine frequencies.

Methods

Binary indicator sequence

The original nucleotide sequence of length N is converted into four numerical sequences of binary indicators,

$$u_{\alpha,n} \begin{cases} 1 & \text{if } \alpha \text{ at } n \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

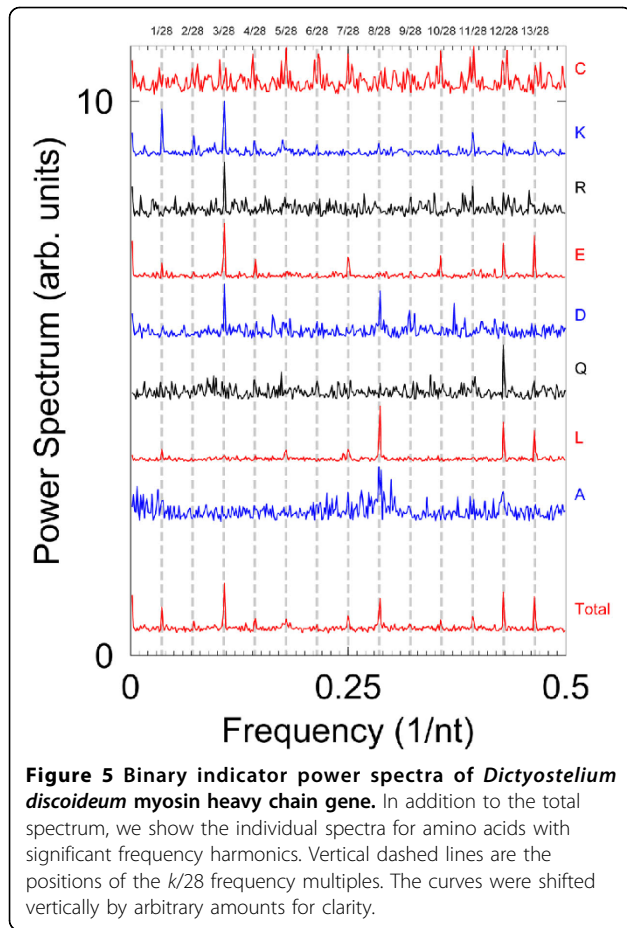
where the coefficient $u_{\alpha,n}$ indicates the presence or absence of a nucleotide of type α at position n with 1 or 0. This method is also sometimes called multichannel Fourier analysis where each α represents a channel [32].

For instance the sequence TAAGACCTATGTTAG-TAAAG would be represented as shown in table 2. This numerical representation removes any content bias of the original genomic sequences which may be an advantage or not depending on its intended applications [18,36].

For a sequence of amino acids this mapping is trivially extended by mapping the 20 amino acids into the same amount of binary indicator sequences [32].

Elastic constants

An alternative to the binary indicator sequence is to convert the genomic DNA sequence into a single numerical sequence using a conversion table. In this work, we use microscopic flexibilities for DNA calculated recently [37]. Instead of converting single nucleotides we convert pair-wise nearest-neighbour dimers following table 3. In this case the sequence



TAAGACCTATGTTAGTAAAG would be represented by the flexibility profile shown in figure 6.

Power spectrum

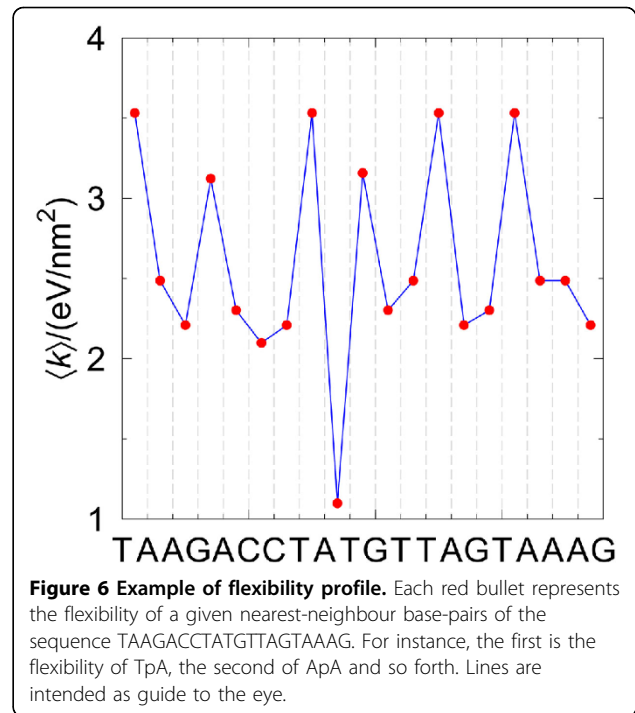
For the binary indicator sequences we calculate the discrete Fourier transform (DFT) for each channel independently

$$U_{\alpha,l} = \frac{1}{N} \sum_{n=0}^{N-1} u_{\alpha,n} e^{-i\frac{2\pi}{N}ln}, \quad l = 0, 1, 2, \dots, N-1 \quad (2)$$

where N is the length of the sequence, and then combine the four resulting power spectra [24]

Table 3 Flexibilities of the ten unique nearest-neighbours dimers in DNA [37].

dimer step	k (eV/nm ²)	dimer	k (eV/nm ²)
ApA	2.40	ApC	2.56
ApG	2.25	ApT	1.83
CpA	3.44	CpC	2.06
CpG	2.74	GpA	2.80
GpC	3.36	TpA	2.42



$$S_l = \sum_{\alpha=A,C,G,T} |U_{\alpha,l}|^2. \quad (3)$$

The DFT was implemented through the use of the efficient FFTW3 package [38].

For the flexibility profile we calculate only one Fourier transform and its corresponding power spectra

$$U_l = \frac{1}{M} \sum_{n=0}^{M-1} k_n e^{-i\frac{2\pi}{M}ln}, \quad S_l = |U_l|^2 \quad (4)$$

where M is the length of the nearest-neighbour sequence and $M = N - 1$.

Position dependent power spectrum

To identify the genomic origin of a given frequency f we apply the Fourier transform only to a section of size w starting at position p of the genome and divide this section along the genome monitoring only this specific frequency. In this way we are able to pinpoint the genomic sections which provide the strongest contribution to a specific frequency $f = l/w$. For the case of binary indicators this means taking Eq. (2) only over a window w

$$U_{\alpha,l}(p) = \frac{1}{w} \sum_{n=p}^{p+w} u_{\alpha,n} e^{-i\frac{2\pi}{N}ln}, \quad l = 0, 1, 2, \dots, N-1 \quad (5)$$

and then monitoring a specific frequency $f = 1/w$. In this work we used a window of size $w = 1000$ nt, unless noted otherwise.

Accession numbers

In this work we used the genomic sequences of *P. falciparum*, accession numbers, in increasing order of chromosomes: NC_004325(1), NC_000910(2), NC_000521(3), NC_004318(1), NC_004326(1), NC_004327(2), NC_004328(1), NC_004329(1), NC_004330(1), NC_004314(1), NC_004315(1), NC_004316(2), NC_004331(1), NC_004317(1). Version numbers are given in brackets. For *Dictyostelium discoideum* we used M14628(1).

Author contribution

MCSN wrote the Perl scripts, performed the calculations and prepared the figures. EFW and GW provided conceptual advice and supervised MCSN. All authors wrote the paper.

Acknowledgements

Funding CNPq, Capes and Fapemig.

This article has been published as part of *BMC Genomics* Volume 12 Supplement 4, 2011: Proceedings of the 6th International Conference of the Brazilian Association for Bioinformatics and Computational Biology (X-meeting 2010). The full contents of the supplement are available online at <http://www.biomedcentral.com/1471-2164/12?issue=S4>

Author details

¹Department of Biological Sciences, Federal University of Ouro Preto, 35400-000 Ouro Preto, MG, Brazil. ²Department of Computer Engineering, Federal Center for Technological Education of Minas Gerais, 30421-169 Belo Horizonte, MG, Brazil. ³Department of Physics, Federal University of Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil.

Competing interests

The authors declare that they have no competing interests.

Published: 22 December 2011

References

- Whiteford N, Haslam N, Weber G, Prügel-Bennett A, Essex JW, Roach PL, Bradley M, Neylon C: **An analysis of the feasibility of short read sequencing.** *Nucleic Acids Res* 2005, **33**(19):e171[<http://nar.oxfordjournals.org/content/33/19/e171.abstract>].
- Vinces MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ: **Unstable tandem repeats in promoters confer transcriptional evolvability.** *Science* 2009, **324**(5931):1213-1216[<http://www.sciencemag.org/content/324/5931/1213.abstract>].
- Cordaux R, Batzer MA: **The impact of retrotransposons on human genome evolution.** *Nat Rev Genet* 2009, **10**(10):691-703.
- Treangen TJ, Abraham AL, Touchon M, Rocha EPC: **Genesis, effects and fates of repeats in prokaryotic genomes.** *FEMS Microbiology Reviews* 2009, **33**(3):539-571.
- Huda A, Mariño-Ramírez L, Landsman D, Jordan IK: **Repetitive DNA elements, nucleosome binding and human gene expression.** *Gene* 2009, **436**(1-2):12-22.
- Whiteford N, Haslam N, Weber G, Prügel-Bennett A, Essex JW, Neylon C: **Visualising the repeat structure of genomic sequences.** *Complex Systems* 2008, **17**:381-398.
- Stewart M, McLachlan AD: **Fourteen actin-binding sites on tropomyosin?** *Nature* 1975, **257**:331-333.
- Parry DAD: **Analysis of the primary sequence of α -tropomyosin from rabbit skeletal muscle.** *J. Mol. Biol* 1975, **98**:519-535.
- McLachlan AD, Stewart M: **The 14-fold periodicity in alpha-tropomyosin and the interaction with actin.** *J Math Biol* 1976, **103**(2):271-298[<http://www.sciencedirect.com/science/article/pii/0022283676903132>].
- Dowling LM, Crewther WG, Parry DA: **Secondary structure of component 8c-1 of alpha-keratin. An analysis of the amino acid sequence.** *Biochem J* 1986, **236**(3):705-712.
- Makeev VJu, Tumanyan VG: **Search of periodicities in primary structure of biopolymers: a general Fourier approach.** *Comput Appl Biosci* 1996, **12**:49-54.
- Veljković V, Cosić I, Dimitrijević B, Lalović D: **Is it possible to analyze DNA and protein sequences by the methods of digital signal processing?** *IEEE Trans Biomed Eng* 1985, **32**(5):337-341.
- Silverman BD, Linsker R: **A measure of DNA periodicity.** *J Theor Biol* 1986, **118**(3):295-300[<http://www.sciencedirect.com/science/article/pii/S0022519386800601>].
- Benson DC: **Fourier methods for biosequence analysis.** *Nucl. Acids. Res* 1990, **18**(21):6305.
- Anastassiou D: **Frequency-domain analysis of biomolecular sequences.** *Bioinformatics* 2000, **16**(12):1073.
- Anastassiou D: **Genomic signal processing.** *IEEE Signal Processing Mag* 2001, **8**:20.
- Fukushima A, Ikemura T, Kinouchi M, Oshima T, Kudo Y, Mori H, Kanaya S: **Periodicity in prokaryotic and eukaryotic genomes identified by power spectrum analysis.** *Gene* 2002, **300**:203-211.
- Akhtar M, Epps J, Ambikairajah E: **Signal processing in sequence analysis: advances in eukaryotic gene prediction.** *IEEE Journal on Selected Topics in Signal Processing* 2008, **2**(3):310-321.
- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton J, Pain A, Nelson K, Bowman S, et al: **Genome sequence of the human malaria parasite *Plasmodium falciparum*.** *Nature* 2002, **419**(6906):498-511.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capi P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al: **A unified classification system for eukaryotic transposable elements.** *Nature Reviews Genetics* 2007, **8**(12):973-982.
- Sharma D, Issac B, Raghava GPS, Ramaswamy R: **Spectral Repeat Finder (SRF): identification of repetitive sequences using Fourier transformation.** *Bioinformatics* 2004, **20**(9):1405-1412.
- Brodzik A: **Quaternionic periodicity transform: an algebraic solution to the tandem repeat detection problem.** *Bioinformatics* 2007, **23**(6):694.
- Epps J: **A hybrid technique for the periodicity characterization of genomic sequence data.** *EURASIP J Bioinform Syst Biol* 2009, **924601**.
- Tiwari S, Ramachandran S, Bhattacharya A, Bhattacharya S, Ramaswamy R: **Prediction of probable genes by Fourier analysis of genomic sequences.** *Comput. Appl. Biosci* 1997, **13**(3):263-270.
- Issac B, Singh H, Kaur H, Raghava GPS: **Locating probable genes using Fourier transform approach.** *Bioinformatics* 2002, **18**:196-197[<http://bioinformatics.oxfordjournals.org/content/18/1/196.abstract>].
- Kotlar D, Lavner Y: **Gene prediction by spectra rotation measure: a new method for identifying protein-coding regions.** *Genome Res* 2003, **13**:1930-1937.
- Gao J, Qi Y, Cao Y, Tung WW: **Protein coding sequence identification by simultaneously characterizing the periodic and random features of DNA sequences.** *J Biomed Biotechnol* 2005, **2005**:139-146.
- Benson G: **Tandem repeats finder: a program to analyze DNA sequences.** *Nucleic Acids Res* 1999, **27**(2):573-580[<http://nar.oxfordjournals.org/content/27/2/573.abstract>].
- Brigham EO: **The fast Fourier transform and its applications.** London: Prentice-Hall International; 1988.
- Lobzin W, Chechetkin VR: **Order and correlations in genomic DNA sequences. The spectral approach.** *Physics-Uspekhi* 2000, **43**:55-78.
- Wang L, Schonfeld D: **Mapping Equivalence for Symbolic Sequences: Theory and Applications.** *IEEE Transactions on Signal Processing* 2009, **57**(12):4895-4905.
- McLachlan AD: **Multichannel Fourier analysis of patterns in protein sequences.** *J. Phys. Chem* 1993, **97**(12):3000-3006.
- McLachlan AD, Karn J: **Periodic features in the amino acid sequence of nematode myosin rod.** *J. Mol. Biol* 1983, **164**(4):605-626.
- Taylor WR, Heringa J, Baud F, Flores TP: **A Fourier analysis of symmetry in protein structure.** *Protein Eng* 2002, **15**(2):79-89.

35. Gruber M, Soding J, Lupas AN: **REPPER-repeats and their periodicities in fibrous proteins.** *Nucl. Acids. Res* 2005, **33**(Web Server Issue):W239.
36. Paar V, Pavin N, Basar I, Rosandić M, Glunčić M, Paar N: **Hierarchical structure of cascade of primary and secondary periodicities in Fourier power spectrum of alphoid higher order repeats.** *BMC Bioinformatics* 2008, **9**:466.
37. Weber G, Essex JW, Neylon C: **Probing the microscopic flexibility of DNA from melting temperatures.** *Nature Physics* 2009, **5**:769-773.
38. Frigo M, Johnson SG: **The Design and Implementation of FFTW3.** *Proceedings of the IEEE* 2005, **93**(2):216-231, Special issue on "Program Generation, Optimization, and Platform Adaptation".

doi:10.1186/1471-2164-12-S4-S4

Cite this article as: Nunes *et al.*: Origin of multiple periodicities in the Fourier power spectra of the *Plasmodium falciparum* genome. *BMC Genomics* 2011 **12**(Suppl 4):S4.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

