Integrating telomeres into genome-wide ChIP-seq analyses

Miguel Vega-Palas (vega-palas@ibvf.csic.es)
Instituto de Bioquímica Vegetal y Fotosíntesis (CSIC - U. of Seville), Avda. Américo Vespucio nº 49, 41092 Seville, Spain

María Vaquero-Sedas
Instituto de Bioquímica Vegetal y Fotosíntesis (CSIC - U. of Seville), Avda. Américo Vespucio nº 49, 41092 Seville, Spain

Method Article

Keywords: telomeres, epigenetics, chromatin, ChIP-seq, genome-wide analyses

Posted Date: April 8th, 2013

DOI: https://doi.org/10.1038/protex.2013.039

Abstract

Telomeric chromatin organization studies using Chromatin Immunoprecipitation \(ChIP) might be challenged by the presence of Interstitial Telomeric Sequences \(ITSs). In this protocol we describe a method to study the chromatin structure of telomeres independently of ITSs by analyzing ChIP-seq data. This method has been used successfully in _Arabidopsis thaliana_ and could be applied to other model systems including humans. It could greatly speed the knowledge of telomeres biology and their influence in aging, illness or cancer. The method requires that the mean telomeric lengths of the biological sources to be studied are known as well as the built DNA sequence of their chromosomes. First, the numbers of sequences with four or more perfect tandem telomeric repeats are determined _in silico_ at ITSs. These numbers are normalized to the corresponding numbers of sequences of the same type present at telomeres. Those sequences for which the ITSs/telomeres ratios are lower than 2% can be selected to analyze telomeres in ChIP-seq experiments. The ChIP-seq experiments to be studied should be suitable for telomeric chromatin structure analyses and render reads longer than the telomeric sequences selected. Finally, the levels of telomeric enrichments are determined and represented together with the values corresponding to the rest of the genome. This is the first protocol described to integrate telomeres independently of ITSs into genome-wide studies of chromatin organization.

Introduction

Telomeres are essential for genome stability, which requires the integrity of their length and chromatin structure. Telomeric DNA usually contains tandem repeats of a short G-C rich motif that are also present at internal chromosomal loci in different model systems including Arabidopsis or humans[1-7]. Some caution should be taken when the chromatin structure of telomeres is analyzed. Telomeric chromatin organization studies are usually performed by immunocytolocalization or by chromatin immunoprecipitation \(ChIP)[7-11]. When the epigenetic status of telomeres is studied by immunocytolocalization both, telomeres and subtelomeric regions, are analyzed. In turn, ChIP studies focused on telomeric chromatin structure display simultaneously telomeres and ITSs. The combination of immunocytolocalization and ChIP experiments could help to differentiate between telomeres and ITSs. However, a specific epigenetic mark might be found associated with telomeric repeats by ChIP, with the end of chromosomes by immunocytolocalization and might also be required for telomere function and, still, not associate with true telomeres. Instead, the epigenetic mark could be associated with subtelomeric regions and ITSs or just with subtelomeric ITSs[7]. We have described two procedures based on ChIP experiments to analyze the chromatin structure of telomeres independently of ITSs in _Arabidopsis thaliana_[12,13]. In the first procedure, we performed ChIP experiments followed by digestion with the frequently cutting restriction enzyme Tru9I[12]. Since telomeres in Arabidopsis and in other model systems are essentially composed of perfect telomeric repeat arrays, they remain uncut after digestion with Tru9I[6]. In contrast, Arabidopsis ITSs are frequently cut because they are composed of short arrays of perfect telomeric repeats interspersed with degenerated repeats[6,14-16]. Thus, when Arabidopsis genomic DNA is digested with Tru9I and hybridized with a telomeric probe, most of the signals corresponding to

ITSs disappear[6]. The use of Tru9I has allowed us to discover that Arabidopsis telomeres exhibit euchromatic features. They are labelled with euchromatic marks like H3K4me2, H3K9Ac and H4K16Ac[12]. However, this procedure is laborious. More recently, we have developed a second method to analyze the chromatin structure of Arabidopsis telomeres independently of ITSs by analyzing ChIP-seq data[12]. Previous genome-wide chromatin structure studies performed by ChIP-seq did not differentiate between telomeres and ITSs. The application of this second method have confirmed our previous results and have shown that Arabidopsis telomeres associate with histone H3.3, which is incorporated in transcriptionally active regions, and are labelled with H3K27me3, a repressive euchromatic mark established by the PRC2 complex[12,17]. We interpreted that the presence of these epigenetic marks at telomeres might be due to their special chromatin organization or to the dynamic nature or their epigenetic marks[17]. Here, we describe a general and detailed protocol to analyze telomeric chromatin structure independently of ITSs by using ChIP-seq data. It can be applied to different model systems including humans.

Procedure

Following, we describe the step-by-step procedure and remark some key features that should be taken into consideration. All the steps of the protocol are summarized in Figure 1. Those steps marked with an asterisk include some comments in the troubleshooting section. In addition, we have applied part of this protocol to several model organisms \(see Table 1). Columns with red numbers in Table 1 correspond to the different steps. Step 1) Select an organism which genome had been sequenced and built. Sequencing should have been achieved near to completion. Step 2) Determine _in silico_ the number of sequences containing four perfect tandem telomeric repeats at ITSs. Similarly, determine the number of sequences containing five or more perfect tandem telomeric repeats. These numbers can be determined at the NCBI Mapviewer home \(http://www.ncbi.nlm.nih.gov/mapview/), running a megablast search and using the Build reference database. The megablast search parameters are the following: automatically adjust parameters for short input sequences; expect threshold=10; max matches in a query range=0; match/mismatch scores=1,-2; gap cost=linear; word size=256 and without filters or repeat masking. The output of the megablast searches should appear ordered by chromosomes and by degree of score. Download the alignments as a Hit Table \(text) and count the sequences with the highest evalues, which make perfect matches with the search sequences, using Microsoft Word. If the megablast search is performed with a sequence that contains four perfect tandem telomeric repeats, an specific ITS containing five perfect tandem repeats will reveal two overlapping sequences in the output and so on. The terminal perfect tandem telomeric repeat sequences should not be counted. Step 3) Determine the mean telomeric length of the biological source to be studied if it had not been previously published. The Terminal Restriction Fragments protocol \(TRFs)[18] could be used for that purpose. Different tissues of the same organism might have different telomeric lengths. In principle, if the length of telomeres varies in different tissues of the same organism, the mean telomeric length of the biological sources used to perform the ChIP-seq experiments should be determined. However, this might not be required for all the organisms \(see table 1). See troubleshooting 3. Step 4) Determine the number of sequences containing four or more perfect tandem telomeric repeats at telomeres. For a specific biological source, essentially

the same value is obtained with all the telomeric sequences independently of the number of repeats that they contain. This value is determined as follows: multiply the number of haploid chromosomes by 2 and by their average telomeric length expressed in bp. Then, divide the resulting value by the number of bp in each perfect telomeric repeat. Step 5) Divide the number of sequences containing four perfect tandem telomeric repeats at ITS between the number of the same type of sequences at telomeres. Repeat this step using the numbers obtained with sequences containing five or more perfect telomeric repeats. Select those sequences that render ratios equal or lower than 2%. See troubleshooting 5-1. A 2% ratio can be obtained with short stretches of perfect tandem telomeric repeats in different model organisms \(see Table 1). This is because, whereas telomeres are essentially composed of perfect telomeric repeats, ITSs are usually composed of short stretches of perfect repeats interspersed with degenerated repeats. However, there are some organisms, like hamster, that contain more perfect telomeric repeats at internal chromosomal loci than at telomeres[19,20]. This protocol cannot be applied to them. See troubleshooting 5-2. Step 6) Perform ChIP-seq experiments that render longer reads than the sequence or sequences selected to analyze telomeres. Check the quality of the results by standard methods. Previously published ChIP-seq experiments can also be used to analyze telomeric chromatin structure. In these cases, the number of telomeric reads can easily be calculated at NCBI, using the Sequence Reads Archive software \(http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?). We recommend to check the suitability of the ChIP-seq experiments for telomeric chromatin structure analyses. Specific quality controls should be performed. It might happen that a ChIP-seq experiment is fine to analyze genome-wide aspects of chromatin organization but not to analyze telomeres. This is because single sequences are used for the analysis of telomeres and some bias can arise along the ChIP-seq experiments during the cloning, amplification or sequencing of the telomeric sequences. The special architecture and repetitive nature of telomeres might favour those problems. The study of ChIP-seq experiments that have been repeated at least twice, starting with different samples of the same biological source, should help in this task. If similar results are obtained in the different repetitions of the same experiment, we consider the resulting ChIP-seq data suitable for telomeric chromatin structure analyses. See troubleshooting 6-1. If previously published ChIP-seq experiments are analyzed, we recommend calculating the percentage of four tandem telomeric repeat reads in the input samples. In addition, we also recommend to calculate the number of four tandem telomeric repeat reads versus the number of reads corresponding to another genomic sequence, also in the input samples. Both types of numbers should remain similar in different experiments. However, we would rather prefer to analyze repetitions of the same ChIP-seq experiment that had been started with different samples of the same biological source. In this way, the bias of the immunoprecipitated samples should also be controlled. See troubleshooting 6-2. We also recommend performing telomeric analyses using more than one telomeric sequence for each ChIP-seq run, if possible. Similar results should be obtained with sequences containing different number of perfect telomeric repeats, as far as the quality of the reads is good. See troubleshooting 6-3. Step 7) Count the number of telomeric reads in the input and immunoprecipitated DNA samples and normalize them against the total number of reads for each run. Determine the telomeric enrichments levels and represent them together with the results obtained for the rest of the genome. For example, the lg_2 IP/I could be represented for the whole genome, including telomeres. **Applications and target audience:** This method can be applied to

any model system, excluding those that contain many more perfect tandem telomeric repeats at ITSs than at telomeres. In addition, it could be used to analyze telomeric chromatin structure from previously published ChIP-seq experiments. **Advantages, limitations and adaptations** Our group has previously analyzed the chromatin structure of telomeres independently of ITSs using a frequently cutting restriction enzyme. However, the method described here is easier to achieve, can be applied to experiments already reported and to future ChIP-seq experiments. This method cannot be applied to model systems that contain many more perfect tandem telomeric repeats at ITSs than at telomeres. However, in those cases, it could be applied to analyze the chromatin structure of ITSs instead of the chromatin structure of telomeres.

Troubleshooting

The following considerations should be taken into account in the following steps of the procedure: **Step 3:** The mean telomeric length of different Arabidopsis organs have been determined by TRFs and have been found to be very similar[21]. **Step 5:** 5-1) We have selected a 2% ratio to play safe. There is always certain level of background when ChIP experiments are performed. For example, let's imagine that a promoter \(seq A) is associated with a transcription factor and another DNA sequence \(seq B) is not associated with it. If ChIP experiments are performed with antibodies against the transcription factor and the enrichment level of seq A versus seq B is analyzed by PCR, an enrichment level of 5-10 folds of seq A versus seq B would be a very good result. In these cases, it would be concluded that the transcription factor associates with the promoter \(seq A). However seq B is still being pulling down in the ChIP experiment because of the experimental background of the technique. Therefore, in ChIP-seq experiments, a sequence composed of perfect tandem telomeric repeats that is poorly found at ITSs should not reflect the epigenetic status of ITSs because the telomeric background would impaired it. This sequence should reflect the epigenetic characteristics of telomeres. 5-2) A stretch of perfect tandem telomeric repeats found more than 2% of the cases at ITSs might also be used for this procedure, whenever the level of background allows it. This background could be estimated for different organisms and experimental conditions by PCR[12]. **Step 6:** 6-1) Telomeric chromatin structure analyses using ChIP-seq data have been previously reported in humans but without considering ITSs[22-24]. In one study focused on CD4+ T-cells, telomeres were found to be enriched in H3K4me3 \(an euchromatic mark) and under-represented in H3K9me3 \(an heterochromatic mark)[22]. In another study focussed on different human cell lines, telomeres were found to be enriched in H3K9me3[23]. The different results obtained in these two studies might be due to the different biological sources studied or to the different number of telomeric repeats used to analyze telomeres, which were not clearly stated in any of the manuscripts. However, they could also be due to the bias generated during the experimental procedure followed to obtain the ChIP-seq data. This concern emphasizes the need to determine the suitability of the ChIP-seq studies for telomeric chromatin structure analyses. 6-2) We have analyzed the chromatin structure of Arabidopsis telomeres using some previously published ChIP-seq experiments but, first, we studied the suitability of these experiments to analyze telomeric chromatin structure[13,17,25-27]. We determined the percentage of the \(CCCTAAA)4 reads in the input samples of all these experiments and found that they were comprised

between 0.007% and 0.009%. We also determined the number of \(CCCTAAA)4 reads and the number of reads corresponding to a centromeric sequence \(CEN1: TTGGCTTTGTATCTTCTAACAAG) in the input samples of the same experiments. The \(CCCTAAA)4/CEN1 ratios that we obtained varied between 0.4 and 0.7. Therefore, we assumed that those experiments were suitable to analyze telomeric chromatin structure. Our selection criteria for the ChIP-seq experiments were supported by the fact that we observed similar telomeric epigenetic features in Arabidopsis by analyzing ChIP-seq experiments and by using a different experimental approach[12,13,17]. 6-3) We have obtained similar chromatin structure results in Arabidopsis using the sequence \(CCCTAAA)4 \(present at ITSs in 2% of the cases) than using the sequence \(CCCTAAA)6 \(present at ITSs in 1% of the cases)[17].

References

1. Blackburn, E. Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. FEBS Lett. 579, 859-862 \(2005). 2. Cech, T. Beginning to understand the end of the chromosome. Cell 43, 405-413 \(2004). 3. Greider, C. Telomerase RNA levels limit the telomere length equilibrium. Cold Spring Harb. Symp. Quant. Biol. 71, 225-229 \(2006). 4. Meyne, J. et al. Distribution of non-telomeric sites of the \(TTAGGG)n telomeric sequence in vertebrate chromosomes. Chromosoma 99, 3-10 \(1990). 5. Richards, E., Goodman, H., and Ausubel, F. The centromere region of Arabidopsis thaliana chromosome 1 contains telomere-similar sequences. Nucleic Acids Res. 19, 3351-3357 \(1991). 6. Gámez-Arjona, F., López-López, C., Vaquero-Sedas, M., and Vega-Palas, M. On the organization of the nucleosomes associated with telomeric sequences. Biochim. Biophys. Acta 1803, 1058-1061 \(2010). 7. Vaquero-Sedas, M., and Vega-Palas, M. On the chromatin structure of eukaryotic telomeres. Epigenetics 6, 1055-1058 \(2011). 8. Braunstein, M., Rose, A., Holmes, S., Allis, C., and Broach, J. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. Genes Dev. 7, 592-604 \(1993). 9. Hecht, A., Strahl-Bolsinger, S., and Grunstein, M. Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. Nature 383, 92-96 \(1996). 10. Orlando, V., and Paro, R. Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin. Cell 75, 1187-1198 \(1993). 11. Wu, C., et al. The chromatin structure of specific genes: Disruption of chromatin structure during gene activity. Cell 16, 807-14 \(1979). 12. Vaquero-Sedas, M., Gámez-Arjona, F., and Vega-Palas, M. Arabidopsis thaliana telomeres exhibit euchromatic features. Nucleic Acids Res. 39, 2007-2017 \(2011). 13. Vaquero-Sedas, M., Luo C., and Vega-Palas, M. Analysis of the epigenetic status of telomeres by using ChIP-seq data. Nucleic Acids Res. 40, e163 \(2012). 14. Lin, K., and Yan, J. Endings in the middle: current knowledge of interstitial telomeric sequences. Mutation Res. 658, 95-110 \(2008). 15. Regad, F., Lebas, M., and Lescure, B. ITSs within the Arabidopsis thaliana genome. J. Mol. Biol. 239, 163-169 \(1994). 16. Uchida, W., Matsunaga, S., Sugiyama, R., and Kawano, S. Interstitial telomere-like repeats in the Arabidopsis thaliana genome. Genes Genet. Syst. 77, 63-67 \(2002). 17. Vaquero-Sedas, M., and Vega-Palas, M. Differential association of Arabidopsis telomeres and centromeres with Histone H3 variants. Sci. Rep. 3, 1202 \(2013). 18. Kimura, M., et al. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. Nat. Protoc. 5, 1596-1607 \(2010). 19. Faravelli, M., Azzalin, C., Bertoni, L., Chernova, O., Attolini, C., Mondello, C., and Giulotto, E. Molecular organization of

internal telomeric sequences in Chinese hamster chromosomes. Gene 283, 11-16 \(2002). 20. Revaud, D., Mozziconacci, J., Sabatier, L., Desmaze, C., and Lavelle, C. Sequence-driven telomeric chromatin structure. Cell Cycle 8, 1099-100 \(2009). 21. Zentgraf, U., Hinderhofer, K., and Kolb, D. Specific association of a small protein with the telomeric DNA-protein complex during the onset of leaf senescence in Arabidopsis thaliana. Plant Mol. Biol. 42, 429-438 \(2000). 22. Rosenfeld, J., Wang, Z., Schones, D., Zhao, K., DeSalle, R. and Zhang, M. Determination of enriched histone modifications in non-genic portions of the human genome. BMC Genomics 10, 143 \(2009). 23. Ernst, J., Kheradpour, P., Mikkelsen, T., Shoresh, N., Ward, L., Epstein, C., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M. and Bernstein, B. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473, 43-49 \(2011). 24. Galati, A., Micheli, E. and Cacchione, S. Chromatin structure in telomere dynamics. Front Oncol. 3, 46 \(2013). 25. Stroud, H. et al. Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA 109, 5370-5375 \(2012). 26. Wollmann, H. et al. Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the Arabidopsis transcriptome. PLoS Genet. 8, e1002658 \(2012). 27. Luo, C., Sidote, D,, Zhang, Y., Kerstetter, R., Michael, T., and Lam, E. Integrative analysis of chromatin states in Arabidopsis identified potential regulatory mechanisms for natural antisense transcript production. Plant J. \(2012) doi: 10.1111/tpj.12017. 28. Siomos, M. and Riha, K. Telomeres and their biology. In Plant Genome Diversity Volume 1. J. F. Wendel et al., \(eds.) 71-81 \(2012). 29. Anchelin, M., Murcia, L., Alcaraz-Pérez, F., García-Navarro, E. and Cayuela, M. Behaviour of telomere and telomerase during aging and regeneration in zebrafish. PLoS One 6, e16955 \(2011).

Figures

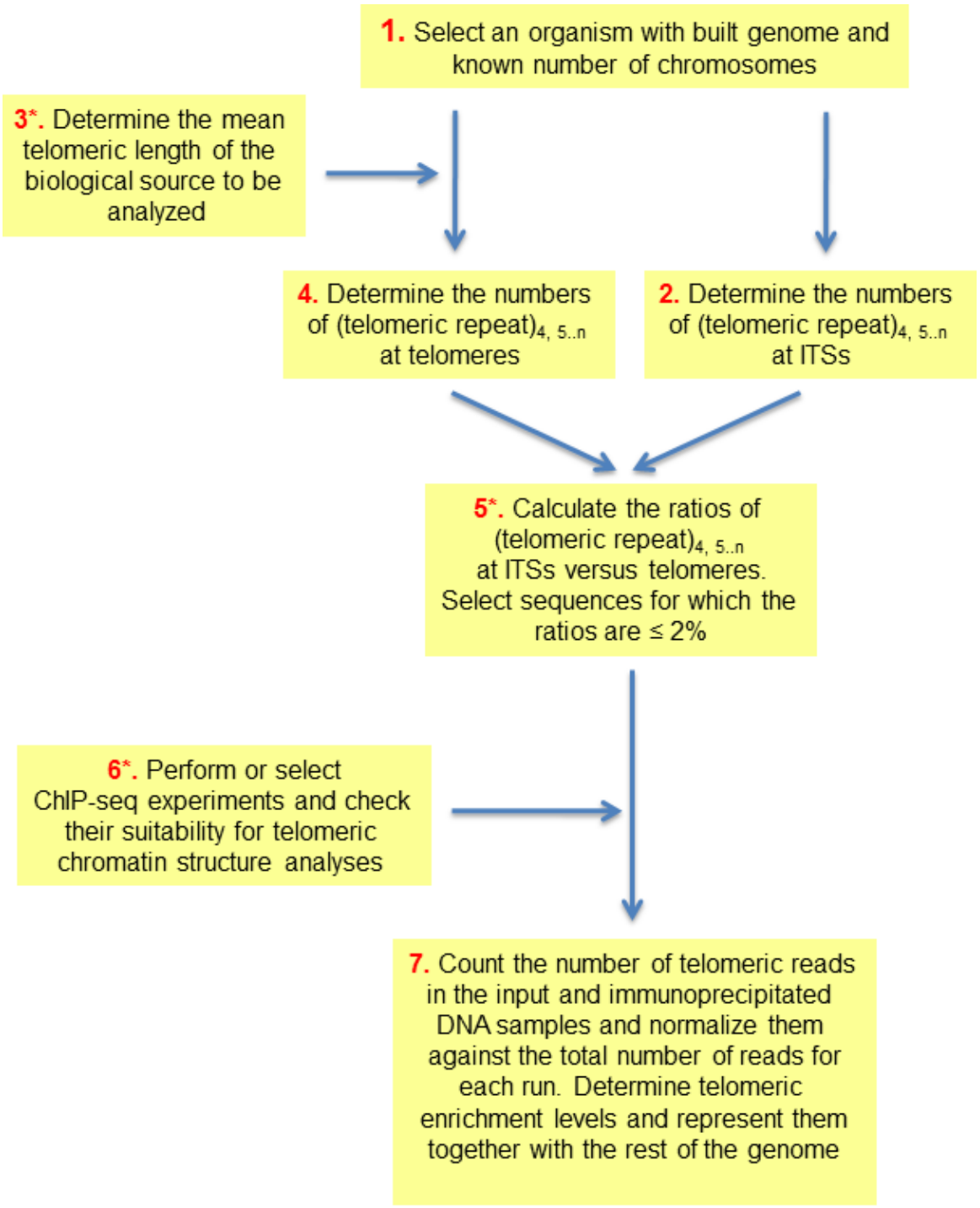

Figure 1

Figure 1

Schematic representation of the steps of the protocol Those steps marked with an asterisk include some comments in the troubleshooting section.

Table 1. Determination of the number of tandem telomeric repeats that could be used to analyze telomeric chromatin structure in different organisms. Crom indicates the number of haploid chromosomes (information obtained from NCBI), MTL indicates the mean telomeric length expressed in Kbp and Build indicates the genome reference used to perform analyses. The ratios of 4, 5 or 6 perfect tandem telomeric repeats at ITSs versus telomeres are indicated (ITSs/Tel). The telomeric repeats sequences and the mean telomeric length values were obtained from references 28 and 29. Calculations were performed as indicated in the text. Sequences at ITSs also include those present at telomeres in the build genomes, except for *C. elegans*. Therefore, the real ITSs/Tel ratios should be lower than those indicated in the table, except for *C. elegans*. Although telomeric repeats in *S. pombe* are degenerated [TTAC(A)G_{2-5}] we used the sequence TTACAGG as the repeat unit. Numbers in red indicate the step of the procedure to which each column is related.

1 Organism	1 Crom	1 Build	3 MTL	2 repeats at ITSs			4 repeats at Tel			5 ITSs/Tel		
				4	5	6	4	5	6	4	5	6
Homo sapiens	23	36.5	12.5	1162	953	829	aprox 96000			1.2	1.0	1.0
Mus musculus	20	38.1	35	2287	1920	1620	aprox 233000			1.0	0.8	0.7
Danio rerio	25	Zv9	16	150	118	108	aprox 133000			0.12	0.06	0.06
C. elegans	6	WS195	5.5	562	338	230	aprox 11000			5.1	3.0	2.1
C. reinhardtii	17	Build1.1	0.75	0	0	0	aprox 3200			<0.05	<0.05	<0.05
S. pombe	3	Build1.1	0.3	0	0	0	aprox 257			<0.5	<0.5	<0.5

Figure 2

Table 1 Determination of the number of tandem telomeric repeats that could be used to analyze telomeric chromatin structure in different organisms

Figure 3

Figure 1 as a powerpoint slide Schematic representation of the steps of the protocol Those steps marked with an asterisk include some comments in the troubleshooting section.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement0.doc