Mol Genet Genomics (2016) 291:1219–1225
DOI 10.1007/s00438-016-1176-1

ORIGINAL ARTICLE

Integrated cytogenetics and genomics analysis of transposable elements in the Nile tilapia, *Oreochromis niloticus*

Guilherme Valente[1] · Thomas Kocher[2] · Thomas Eickbush[3] · Rafael P. Simões[1] · Cesar Martins[4]

Received: 21 August 2015 / Accepted: 21 January 2016 / Published online: 9 February 2016

Abstract Integration of cytogenetics and genomics has become essential to a better view of architecture and function of genomes. Although the advances on genomic sequencing have contributed to study genes and genomes, the repetitive DNA fraction of the genome is still enigmatic and poorly understood. Among repeated DNAs, transposable elements (TEs) are major components of eukaryotic chromatin and their investigation has been hindered even after the availability of whole sequenced genomes. The cytogenetic mapping of TEs in chromosomes has proved to be of high value to integrate information from the micro level of nucleotide sequence to a cytological view of chromosomes. Different TEs have been cytogenetically mapped in cichlids; however, neither details about their genomic arrangement nor appropriated copy number are well defined by these approaches. The current study integrates TEs distribution in Nile tilapia *Oreochromis niloticus* genome based on cytogenetic and genomics/bioinformatics approach. The results showed that some elements are not randomly distributed and that some are genomic dependent on each other. Moreover, we found extensive overlap between genomics and cytogenetics data and that tandem duplication may be the major mechanism responsible for the genomic dynamics of TEs here analyzed. This paper provides insights in the genomic organization of TEs under an integrated view based on cytogenetics and genomics.

Keywords Repeated DNA · Genomic organization · Bioinformatics · DNA mapping

Communicated by J. Cerdá.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-016-1176-1) contains supplementary material, which is available to authorized users.

✉ Guilherme Valente
valentegt@gmail.com; valentegt@fca.unesp.br

[1] Department of Bioprocess and Biotechnology, Faculty of Agronomic Sciences, UNESP, Sao Paulo State University, Botucatu, SP 18610-307, Brazil

[2] University of Maryland, Bioscience Research Building, 20742 College Park, MD, USA

[3] Department of Biology, University of Rochester, 14604 Rochester, NY, USA

[4] Morphology Department, Institute of Biosciences, UNESP, Sao Paulo State University, 18618-689 Botucatu, SP, Brazil

Introduction

The genomic studies of transposable elements (TEs) as well as other highly repeated DNAs have been hindered because of difficulties in determining the correct identification of their organization and copy number in the genome. While advances have been acquired concerning the application of new bioinformatics tools in the analysis of whole sequenced genomes, analysis of whole sequenced genomes lacks most TE copies due to difficulties in assembling those sequences. In this way, the completed sequenced genomes reported at this time have given a limited view of the true scenario of the repeated DNA fraction. On the other hand, the mapping of TEs by molecular cytogenetic tools provides important information on their genomic distribution and clustering, which are difficult to be obtained from sequenced genomes. However, cytogenetic mapping of TEs is usually able to recognize clusters sequences larger than 10 kb, thus small clusters or dispersed copies out of major clusters are not evidenced by cytological analysis. Finally, in both genome sequencing and cytogenetic approaches,

the genomic organization of TEs is underestimated and thus it is important to integrate both analyses to have a more realistic view on the genomic organization of TEs.

Primary studies on whole sequenced genomes revealed that most genomes contains a high frequency of TEs, as much as 50 % or more of the human genome for instance (International Human Genome Sequencing Consortium 2001; Venter et al. 2001), previously stated as junk DNA without function. Nowadays, these relegated DNA elements have moved into a position of major players in the genome architecture and function (for review, Shapiro 2010). TEs in cichlids have been classically studied by molecular cytogenetics methods (Bryden et al. 1998; Oliveira et al. 1999, 2003; Harvey et al. 2003; Mazzuchelli and Martins 2009; Teixeira et al. 2009; Gross et al. 2009, 2010; Ferreira et al. 2010; Fantinatti et al. 2011; Valente et al. 2011; Schneider et al. 2013) and now by genome sequencing (Brawand et al. 2014). Genomics data showed that East African cichlids (including the tilapiine *Oreochromis niloticus*) have about 16–19 % of TEs (Brawand et al. 2014). In reference to *O. niloticus*, the cytogenetic mapping of TEs reveled the organization of the CiLINE2, Ron1, Ron2, On2318, On239, Rex1, Rex3 and Rex6 elements (Oliveira et al. 1999, 2003; Harvey et al. 2003; Ferreira et al. 2010; Valente et al. 2011).

The *O. niloticus* species was chosen as model in this study because it has several TEs physically mapped on its chromosomes and a complete sequenced genome (Brawand et al. 2014). However, an integrated view is not available at this time, which was the motivation of this study. In this way, this paper reports the analysis of a specific set of TEs present in the *O. niloticus* genome based in an integrated view of genomics and cytogenetic data. The results obtained indicate that some TEs are not randomly distributed throughout the genome and that some TE-couples are genomic dependent from each other. Moreover, it is also reported that tandem duplications may be the major evolutionary process acting over these elements.

Materials and methods

Searching TEs in *O. niloticus* genome

The nucleotide sequences of the TEs CiLINE2 (AF016499; 1630 bp), On2318 (CC156516; 211 bp), On239 (CC156510; 237 bp), Rex1 (AJ288471, AJ288472, AJ288473, AJ288474, AJ288475, AJ288476 and AJ288477; 529 bp consensus length), Rex3 (AJ400459, AJ400368, AJ400369, AJ400370, AJ400371 and AJ400372; 420 bp consensus length), Rex6 (AJ293545, AJ293546 and AJ293547; 422 bp consensus length), Ron1 (AF097734, AF097735 and AF097736; 703 bp consensus length) and Ron2 (AF057520 and AF057521; 395 bp consensus length), were retrieved from the NCBI. Sequence alignments using ClustalW (Thompson et al. 1994) and consensus sequences construction were obtained using Geneious v. 4.8.5 software (Drummond et al. 2009) for TEs with more than one sequence available.

All final sequences were used as queries against the *O. niloticus* genome v.1 in a standard blast search in the Bouillabase database (bouillabase.org). Since there is not a gold-standard protocol to parse blast results (mainly concerning TEs), we selected hits with *E* values <0.78 (the first lowest values before 1.0) and lengths ≤10 % relative to the query length were excluded. Hits related to the same TE and that overlapped at the same direction in the genome were fused to give just one hit, since they probably correspond to the same copy of an element (Supplementary File 1). The same queries were searched against Repbase database (Jurka et al. 2005) at the Genetic Information Research Institute (Giri) (http://www.girinst.org/repbase/) using the CENSOR software (Kohany et al. 2006) to verify their mapping with known TEs.

TEs distribution in *O. niloticus* genome, gene mapping and statistical tests

The first analysis was conducted to address whether TE sequences are clustered or not in the genome. The common approach should be to calculate the average of DNA length between two or more TE copies. However, it is expected that larger scaffolds harbor more copies than smaller ones and consequently, longer spacer sequences among the copies could introduce bias in the results. Since we had a positive correlation between the number of copies and scaffold lengths (further discussed), the *O. niloticus* genome was split into 10 Kb length bins and subsequent TE copy counting was chosen as the procedure to analyze the distribution of TEs. The regression analysis was applied counting the amount of hits per scaffold and their scaffold lengths, to support the genome splitting decision aforementioned. The amount of repeats was counted per bin and hits present in two bins were counted twice (74 copies = 148 counting).

We defined: (1) TE-couples as two adjacent repeats into the same bin; (2) tandem copies as adjacent TEs (TE-couples) of the same element and orientation; (3) non-tandem repeats adjacent TEs of the same element but in different orientations; (4) adjacent TEs of different elements are not included in tandem and non-tandem definitions. Both tandem and non-tandem copies were counted for all bins and *T* test was applied over this counting.

The statistical Chi-square test for goodness-of-fit was applied over the observation of the distribution of TEs in the genome under the hypothesis that those elements are randomly distributed. The same test was applied to the

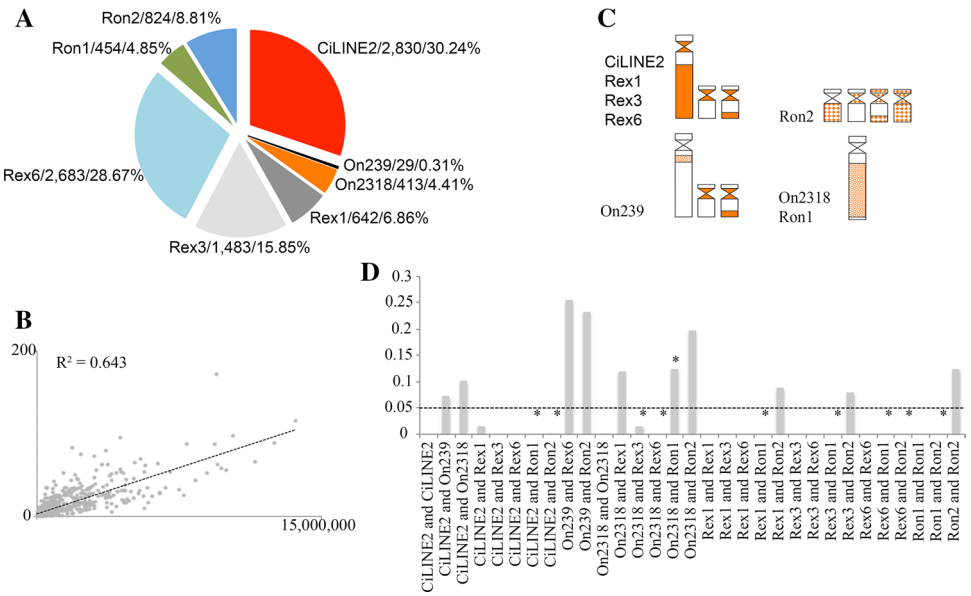

Fig. 1 Genome characteristics of analyzed TEs. **a** Name/number of copies/percentage (100 % is the sum of all copies here analyzed) of elements here investigated; **b** regression analysis between the length of *O. niloticus* scaffolds with hits (X axis) and the number of hits (Y axis); **c** physical mapping of each element (based in previous published data—see the topic "Materials and methods") in representative chromosomes with higher amount of hybridization signals. Orange color, presence of TEs onto the chromosomes; *full orange*, completely spread over chromosomal region; *small orange dots* light-spotted signals; *large orange dots* strong-spotted signals; **d** Fisher's exact test of independence for all TE-couples. Y axis, *P* value; *dashed line*, *P* value threshold; *stars*, results not correlated with cytogenetic mapping. Graphics were made on Excel and all figures were edited in Photoshop

hypothesis that the repeats with at least one neighbor copy in the same bin (TE-couples) is also randomly distributed. Both tests used degree of freedom = 5.

In attempt to check if TEs of some TE-couple have similar specific target site preferences, the genes were mapped for all bins with at least one TE-couple using *O. niloticus* RefSeq data and the regression analysis was performed using this counting and the Fisher's exact test of independence results (further detailed).

The Fisher's exact test of independence was applied over all TE-couples to test if those TEs copies are dependent on each other (*P* value $\leq$0.05). For this test, neither TEs orientation nor the positions of the pairs were considered (TE1 besides TE2 = TE2 besides TE1).

The comparisons of genomics and cytogenetics analyses were performed using data of TEs previously physically mapped in *O. niloticus* (Oliveira et al. 1999, 2003; Harvey et al. 2003; Valente et al. 2011). If TE1 and TE2 are statistically dependent upon each other (based on Fisher's exact test of independence results) and have similar cytogenetic mapping distribution (ideogram built by another researcher as a blind test considering the pictures and hybridization description along the papers), both results are considered correlated and vice versa.

The minimum value of Shannon entropy (Shannon 1948) was obtained for each TE used as query under an algorithm developed in Fortran. This algorithm uses the information theory to calculate the entropy for a single sequence or a single genome, without the necessity of comparisons among lots of sequences. It gives the minimum values of entropy and a map of organized regions for the target sequences (Simões and Valente, in preparation).

Results

After identification and filtering process, a total of 9,358 copies (Supplementary File 1) of the investigated TEs was retrieved with CiLINE2 and Rex6 elements the most abundant (Fig. 1a). All elements here studied are non-LTR elements, except the On2318 and Ron1 (Table 1; Supplementary Fig. 1) and the average of minimum values of Shannon entropy reported high values for all sequences (Table 1). These TEs are distributed through 873 scaffolds with different lengths and it is possible to see a positive correlation of the number of copies and scaffold length (Fig. 1b), thus the *O. niloticus* genome was fragmented into 10 Kb bins for further analysis, giving a total of 96,935 bins.

Concerning TE copies distributions, 74 are present in more than one bin and they were counted twice for Chi-square test. The TEs counting in each bin (9506 counts = 9358 + 148) showed that they are distributed from 1 to 6 copies per bin (Table 2) and the Chi-square test for goodness-of-fit applied over this data revealed that they

Table 1 Genomic data based on repbase searches and statistical testes employed in the TEs analyzed

TEs	*P* value	Class[a]	Similarity (%)[a]	Tandem copies (%)	SEV
CiLINE2	2.37E-23	DNA transposon; non-LTR	>0.77	>64	1.80
On239	1.000	Non-LTR[b]	>0.84	–	1.78
On2318	4.16E-05	DNA transposon	>0.88	>77	1.74
Rex1	3.70E-92	Non-LTR[b]	>0.83	>98	1.77
Rex3	1.24E-79	Non-LTR/LINE[b]	>0.84	>94	1.76
Rex6	7.45E-108	Non-LTR/site-specific LINE[b]	>0.77	>88	1.68
Ron1	0	–	–	100	1.77
Ron2	0.3939	Non-LTR/SINE[b]	>0.83	>66	1.79

The statistics include *P* value of Chi-square test for goodness-of-fit of hits with more than one neighbor, tandem copies counting and Shannon entropy values

SEV Shannon entropy values

[a] Searches against repbase database

[b] Retroelements

Table 2 Number of hits per bins

Maximum hits per bin	Number of bins	Total number of hits
0	89,134	0
1	6419	6419
2	1126	2252
3	202	606
4	43	172
5	9	45
6	2	12
Total	96,935	9506

were not randomly distributed throughout the genome (*P* value 0.0) (Supplementary Table 1).

A total of 1705 TE-couples (3087 TE copies) were obtained and the Chi-square test for goodness-of-fit was conducted to determine if copies with at least one neighbor have a random distribution and the results showed that the elements On239 and Ron2, with at least one neighbor, have a random distribution (Table 1; Supplementary Table 2). Interestingly, cytogenetic mapping of TEs also showed this pattern of distribution in the genome, being only both elements appeared randomly distributed (they do not share similar pattern with all other elements) (Fig. 1c, based in Oliveira et al. 1999, 2003; Harvey et al. 2003; Valente et al. 2011). Moreover, bins with at least one TE-couple and one or more genes in the same region (669 bins) were not correlated with the test of independence (R^2 0.07).

The test of independence applied in all TE-couples aforementioned (1705) showed that some elements are independently distributed from each other throughout the genome (*P* value >0.05) (Table 3; Supplementary Table 3). Comparing these results for all combinations of TE-couples and the cytogenetic data (see ideogram in Fig. 1c), we showed that a lot of TE-couples are distributed in the genome according to the cytogenetic data (Table 3; Fig. 1d; Supplementary Fig. 2). The TE-couples that have tandem repeats (1148 cases) and non-tandem repeats (127 cases) (see definitions in "Materials and methods") were completely dependently distributed from each other, the only exception being the Ron2 element, which is randomly distributed throughout the genome (Table 1; Fig. 1d). The latter conclusion cannot be applied to the On239 element, because it does not have itself as a neighbor (Supplementary Table 3). All elements (except the On239) have a high amount of tandem repeats (Table 1), accounting for >67 % of the TE-couples here analyzed. Moreover, the number of tandem repeats compared to non-tandem repeats is statistically significant (*P* value 0.0052).

Discussion

The test of independence (comparing genomics and cytogenetics data) resulted in a good fit between both datasets; thus, when TE-couple has elements not independently distributed (*P* value $\leq$0.05) from each other throughout the genome, their cytogenetic mapping shows similar pattern and vice versa. However, there was disagreement of the genomic and cytogenetic information, mainly for the element Ron1. This finding is interesting because we found a general good relationship between both data sets, which is not commonly reported in the literature (Froenicke et al. 2006). Furthermore, TE copies with one or more neighbors (TE-couples) are able to form clusters in the genome after the transposition event, with the exception of On239 and Ron2, which are randomly distributed. Our results are also in agreement with previous data that reported mobile elements are not randomly distributed, such as observed in some plants (Capel et al. 1993; Baucom et al. 2009),

Table 3 Summary of Fisher's exact test of independence on genomic and cytogenetic data

TEs	*P* value	Pattern of chromosome distribution	Relationship between both data
CiLINE2 and On239	0.073	Different	Correlated
CiLINE2 and On2318	0.101	Different	Correlated
CiLINE2 and Rex1	0.016*	Similar	Correlated
CiLINE2 and Rex3	0*	Similar	Correlated
CiLINE2 and Rex6	0*	Similar	Correlated
CiLINE2 and Ron1	0*	Different	Not correlated
CiLINE2 and Ron2	0.002*	Different	Not correlated
On239 and Rex6	0.255	Different	Correlated
On239 and Ron2	0.234	Different	Correlated
On2318 and Rex1	0.12	Different	Correlated
On2318 and Rex3	0.016*	Different	Not correlated
On2318 and Rex6	0*	Different	Not correlated
On2318 and Ron1	0.125	Similar	Not correlated
On2318 and Ron2	0.198	Different	Correlated
Rex1 and Rex3	0*	Similar	Correlated
Rex1 and Rex6	0*	Similar	Correlated
Rex1 and Ron1	0*	Different	Not correlated
Rex1 and Ron2	0.089	Different	Correlated
Rex3 and Rex6	0*	Similar	Correlated
Rex3 and Ron1	0*	Different	Not correlated
Rex3 and Ron2	0.079	Different	Correlated
Rex6 and Ron1	0*	Different	Not correlated
Rex6 and Ron2	0*	Different	Not correlated
Ron1 and Ron2	0*	Different	Not correlated

* *P* value of TE-couples with dependent elements

mammals (Wichman et al. 1992) and flies (O'Brochta et al. 1994).

Whichmand et al. (1992) reported that sequence-specific insertion, S-phase insertion, ectopic excision and recombinational editing may be the main mechanisms for non-random distribution of sequences. The sequence-specific insertion was the first hypothesis to explain our results concerning a general non-random distribution here observed; however, we did not find a correlation between TE distributions and genes localization in this genome, suggesting there is no specific target site for two elements. Thus, we discarded the sequence-specific insertion as a mechanism for the general non-random distribution here observed. After we considered that the sequences were intrinsically organized, however, the Shannon entropy calculation showed that the sequences here analyzed are not organized in terms of information. The application of this algorithm over genomes reports that lowest values (organized information) generally are over active sequences like genes (Simões and Valente, in preparation). Thus, we suppose that the genomic organization of these elements is not a consequence of their intrinsic structure (since the TEs experienced "loss" of organized information probably due to an ancient loss of activity) but it is a consequence of external factor or mechanisms (not related to their sequence information) such as the absence of selective pressure over TEs.

Despite reports of diverse fish species apparently having relatively recent active elements (Bouneau et al. 2003) (which include Rex1, Rex3, Rex6) (Volff et al. 1999, 2000, 2001), cichlid genomes have not had significant TE activities for a long period of time, at least after the split of South American and African cichlids. In the case of East African cichlids (which includes the *O. niloticus* species), this has been suggested to have occurred in the common ancestor of haplo-tilapiine cichlids. However, analyzing the TEs position in UTR regions of orthologous gene pairs, they found that elements near to the 5′ UTRs are responsible to increasing of gene expression in those species (Brawand et al. 2014). Moreover, there are no data at present suggesting the recent mobilization of the elements here investigated in *O. niloticus* genome. At the end, we did not consider recent transposition events as the main force to the TEs organization here observed and we suggest that probably unequal crossovers could duplicate adjacent copies of TE-couples and the dependence results are a cause of this genomic dynamics. Moreover, based on our conclusion that TE-couples with tandem and non-tandem repeats usually are completely dependently distributed, we suppose

that the tandem duplication events seem to be the main evolutionary force acting in the repeated DNA evolution.

In conclusion, a non-random distribution, the dependence of two TE elements in their genomic organization and an overlapping between genomics and cytogenetics were the major findings of this paper. Moreover, we suggest that external factors of those sequences, such as unequal crossover, could be the main evolutionary mechanism to their genomic dynamics. Despite of genomic analyses based on genome sequencing to be increasing in the last years and cytogenetics approach to be broadly applied to study mobile DNAs, both kind of approaches are prone to bias that could not generate a clear view of TEs genomic organization. For instance, cytogenetics mapping are not very efficient to detect small TEs blocks or dispersed copies and genome sequencing and assembling algorithms are not good enough to allocate all repeats on their correct location, which we suppose to be the causes of a non-complete overlapping between both data here analyzed for the *O. niloticus* TEs. On the other hand, our data are clearly showing that some overlapping of both results can be constructed in an integrative view and we suggest the use of both methods and statistical analyses to give a more realistic view concerning the genomic organization and dynamics of mobile elements.

Acknowledgments The authors thank Msc. Rafael Takahiro Nakajima for the ideogram construction as a blind experiment.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Funding This work was supported by grants from São Paulo Research Foundation (Grants 2011/03807-7, 2013/04533-3).
This article does not contain any studies with human participants or animals performed by any of the authors. This article did not use any personal data.

References

Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon JM, Westerman RP, Sanmiguel PJ, Bennetzen JL (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet 5:e1000732. doi:10.1371/journal.pgen.1000732

Bouneau L, Fischer C, Ozouf-Costaz C, Froschauer A, Jaillon O, Coutanceau JP, Körting C, Weissenbach J, Bernot A, Wolff JN (2003) An active non-LTR retrotransposon with tandem structure in the compact genome of the pufferfish *Tetraodon nigroviridis*. Genome Res 13:1686–1695

Brawand D, Wagner CE, Li YI, Malinsky M et al (2014) The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375–381. doi:10.1038/nature13726

Bryden L, Denovan-Wright EM, Wright JM (1998) ROn-1 SINEs: a repetitive DNA element from *Oreochromis niloticus* and its species-specific distribution in Old World cichlid fishes. Mol Mar Biol Biotech 7:48–54

Capel J, Montero LM, Martinez-Zapater JM, Salinas J (1993) Non-random distribution of transposable elements in the nuclear genome of plants. Nucleic Acids Res 21:2369–2373

Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R et al (2009) Geneious v4.8.5. http://www.geneious.com. Accessed Mar 2015

Fantinatti BEA, Mazzuchelli J, Valente GT, Cabral-de-Mello DC, Martins C (2011) Genomic content and new insights on the origin of the B chromosome of the cichlid fish *Astatotilapia latifasciata*. Genetica 139:1273–1282. doi:10.1007/s10709-012-9629-x

Ferreira IA, Poletto AB, Kocher TD, Mota-Velasco JC, Penman DJ, Martins C (2010) Chromosome evolution in African cichlid fish: contributions from the physical mapping of repeated DNAs. Cytogenet Genome Res 129:314–322. doi:10.1159/000315895

Froenicke L, Caldés MG, Graphodatsky A, Müller S, Lyons LA, Robinson TJ, Volleth M, Yang F, Weinberg J (2006) Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? Genome Res 16:306–310

Gross MC, Schneider CH, Valente GT, Porto JIR, Martins C, Feldberg E (2009) Comparative cytogenetic analysis of the genus *Symphysodon* (Discus fishes, Cichlidae): chromosomal characteristics of retrotransposons and minor ribosomal DNA. Cytogenet Genome Res 127:43–53

Gross MC, Schneider CH, Valente GT, Martins C, Feldberg E (2010) Variability of 18S rDNA locus among *Symphysodon* fishes: chromosomal rearrangements. J Fish Biol 76:1117–1127. doi:10.1159/000279443

Harvey SC, Boonphakdee C, Campos-Ramos R, Ezaz MT, Griffin DK, Bromage NR, Penman DJ (2003) Analysis of repetitive DNA sequences in the sex chromosomes of *Oreochromis niloticus*. Cytogenet Genome Res 101:314–319

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921. doi:10.1038/35057062

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467

Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: repbaseSubmitter and Censor. BMC Bioinformatics 7:474. doi:10.1186/1471-2105-7-474

Mazzuchelli J, Martins C (2009) Genomic organization of repetitive DNAs in the cichlid fish *Astronotus ocellatus*. Genetica 136:461–469. doi:10.1007/s10709-008-9346-7

O'Brochta DA, Warren WD, Saville KJ, Atkinson PW (1994) Interplasmid transposition of Drosophila hobo elements in non-drosophilid insects. Mol Gen Genet 244:9–14

Oliveira C, Chew JS, Porto-Foresti F, Dobson MJ, Wrigth JM (1999) A LINE2 repetitive DNA sequence from the cichlid fish, *Oreochromis niloticus*: sequence analysis and chromosomal distribution. Chromosoma 108:457–468

Oliveira C, Wang Y, Bryden LJ, Wright JM (2003) Short interspersed repetitive elements (SINEs) from the cichlid fish, *Oreochromis niloticus*, and their chromosomal localization by fluorescent in situ hybridization. Caryologia 56:181–189

Schneider CH, Gross MC, Terencio ML, do Carmo EJ, Martins C, Feldberg E (2013) Evolutionary dynamics of retrotransposable elements Rex1, Rex3 and Rex6 in neotropical cichlid genomes. BMC Evol Bio 13:152. doi:10.1186/1471-2148-13-152

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423

Shapiro JA (2010) Mobile DNA and evolution in the 21st century. Mobile DNA 1:4. doi:10.1186/1759-8753-1-4

Teixeira WG, Ferreira IA, Cabral-de-Mello DC, Mazzuchelli J, Valente GT, Pinhal D, Poletto AB, Venere PC, Martins C (2009) Organization of repeated DNA elements in the genome of the cichlid fish *Cichla kelberi* and its contributions to the knowledge of fish genomes. Cytogenet Genome Res 125:224–234. doi:10.1159/000230006

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Valente GT, Mazzuchelli J, Ferreira IA, Poletto AB, Fantinatti BEA, Martins C (2011) Cytogenetic mapping of the retroelements Rex1, Rex3 and Rex6 among cichlid fish: new insights on the chromosomal distribution of transposable elements. Cytogenet Genome Res 133:32–42. doi:10.1159/000322888

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG et al (2001) The sequence of the human genome. Science 291:1304–1351

Volff JN, Körting C, Sweeney K, Schartl M (1999) The non-LTR retrotransposon Rex3 from the fish *Xiphophorus* is widespread among teleosts. Mol Biol Evol 16:1427–1438

Volff JN, Körting C, Schartl M (2000) Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. Mol Biol Evol 17:1673–1684

Volff JN, Körting C, Froschauer A, Sweeney K, Schartl M (2001) Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. J Mol Evol 52:351–360

Wichman HA, Van den Bussche RA, Hamilton MJ, Baker RJ (1992) Transposable elements and the evolution of genome organization in mammals. Genetica 86:287–293