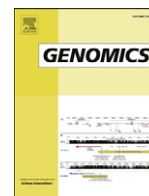




Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Afrotheria genome; overestimation of genome size and distinct chromosome GC content revealed by flow karyotyping



Fumio Kasai*, Patricia C.M. O'Brien, Malcolm A. Ferguson-Smith

Department of Veterinary Medicine, University of Cambridge, UK

ARTICLE INFO

Article history:

Received 21 June 2013

Accepted 13 September 2013

Available online 19 September 2013

Keywords:

Genome size

GC content

Afrotheria

Genome evolution

Chromosome profile

ABSTRACT

Afrotheria genome size is reported to be over 50% larger than that of human, but we show that this is a gross overestimate. Although genome sequencing in Afrotheria is not complete, extensive homology with human has been revealed by chromosome painting. We provide new data on chromosome size and GC content in four Afrotherian species using flow karyotyping. Genome sizes are 4.13 Gb in aardvark, 4.01 Gb in African elephant, 3.69 Gb in golden mole and 3.31 Gb in manatee, whereas published results show a mean of 5.18 Gb for Afrotheria. Genome GC content shows a negative correlation with size, indicating that this is due to differences in the amount of AT-rich sequences. Low genome GC content and small variance in chromosome GC content are characteristic of aardvark and elephant and may be associated with the high degree of conserved synteny, suggesting that these are features of the Afrotherian ancestral genome.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Comparative genomic studies on Afrotheria, the most basal mammalian lineage and distinct by molecular genetic studies from other eutherian mammals, can lead to an improved understanding of mammalian genome evolution [1]. Whole genome comparisons between human and Afrotheria have been made by chromosome painting, indicating extensive conservation and suggesting a presumptive eutherian ancestral karyotype of $2n = 44$ or $2n = 46$ [2–4]. However, the Afrotheria genome has not been fully characterised by sequencing, although this is currently underway in three species; lesser hedgehog tenrec, hyrax and African elephant.

Mammalian karyotypes are highly rearranged throughout evolution and show a great diversity in chromosome number and form [5]. Despite long divergence times, chromosome homology revealed by reciprocal painting displays almost one-to-one correspondence between human, aardvark and African elephant except for heterochromatic regions [2]. It is suggested that chromosome banding patterns are highly conserved within mammals and that all mammals have similar genome sizes after the subtraction of the heterochromatin component [6]. However, conventional cytogenetic studies based on qualitative analysis cannot detect intrinsic differences between homologies because of their low resolution.

Genome size is fundamental to genome structure and characterises each species. A large number of genome size data have been accumulated in the genome size database [7], showing extensive variation in mammals. A previous study on Afrotheria reported that the mean

genome size measured on Feulgen-stained nuclei was 5.5 pg, i.e. more than 50% higher than that of humans, suggesting that there were significant size differences between Afrotheria and other eutherian clades except for Xenarthra [8]. However, discrepancies between different estimates of genome size for the same species have been reported due to methodological problems [9]. This implies that the database contains inaccuracies and that more precise measurements of genome size are needed for meaningful comparisons.

An early analysis of aardvark chromosomes on photographic images measured the length of each chromosome arm [10], estimating, for example, that the proportion of chromosome 1 is approximately 20.5% of the genome. The aardvark genome size measured by microdensitometer was 1.67 times larger than that of human [10]. Applying this to the human genome size, 3.15 Gb as measured from the flow karyotype, aardvark chromosome 1 is calculated to be 1078 Mb, equivalent to the chicken total genome size. Chromosome painting has shown that aardvark chromosome 1 is homologous to 6 human chromosomes/regions 2pter-q13, 4, 8p, 13, 16q and 19q [2]. We estimate that the size of these homologous regions in the human karyotype is 520 Mb from sequence data, indicating that the aardvark chromosome 1 is 2.07 times larger than its human homologues if the previous data are accepted. It is questionable whether the extensive homology between human and Afrotheria is consistent with such a remarkably large genome size.

Flow karyotyping has revealed chromosome sizes in zebrafish [11], chicken, turtle, crocodile [12], bat [13], dog [14] and human [15], leading to estimations of genome size. Most data included in the database do not show the sex of samples and ignore the difference between males and females in each species. Chromosome measurements distinguishes sex differences in genome sizes by determining the size of individual sex

* Corresponding author at: Madingley Road, Cambridge, CB3 0ES, UK.
E-mail address: k-230@umin.ac.jp (F. Kasai).

chromosomes [12–15], allowing comparison between species, an important factor in considering X chromosome conservation during mammalian evolution.

Although some attention has been paid to Afrotherian genomes, genetic information is still lacking. In this study we determine the individual chromosome size and GC content of four Afrotherian species, armadillo (*Orycteropus afer*, OAF), African elephant (*Loxodonta africana*, LAF), manatee (*Trichechus manatus latirostris*, TMA) and golden mole (*Chrysochloris asiatica*, CAS) from flow karyotypes and find that their genome sizes have been seriously overestimated.

2. Results

The 11 OAF, 29 LAF, 25 TMA and 15 CAS chromosomes were resolved into 11, 23, 19 and 12 peaks, respectively, because some peaks included two or three chromosomes of similar size and GC content and so could not be distinguished in the flow karyotypes [2,16,17]. Each chromosome size and GC-content was estimated from the flow karyotype using human chromosomes 4, 17 and 19 as references (Table 1). Chromosome profiles for each species based on size and GC content are shown in Fig. 1. The variance in chromosome GC content excluding the Y chromosome is 1.83 for OAF, 0.75 for LAF, 1.81 for TMA and 3.13 for CAS.

The total genome sizes were 4132 Mb for OAF, 4006 Mb for LAF, 3308 Mb for TMA and 3690 Mb for CAS, calculated from the sum of individual chromosome sizes excluding the Y chromosome (Table 1), indicating that the previous study overestimates these sizes by 41, 10, 40 and 70% respectively. Our results show that the size relative to the human genome (3.15 Gb) is 131% for OAF, 127% for LAF, 105% for TMA and 117% for CAS. The sizes in human regions homologous to each OAF chromosome and the proportion of chromosomal DNA in each genome are shown in Table 2. The total GC content is 39.5% for

Table 1
Size and GC content of each chromosome and each total genome (TG). The lowest and highest GC content in each genome are shown in bold italics (excluding Y chromosome).

Chr	Armadillo		Elephant		Manatee		Golden mole	
	Size (Mb)	GC (%)	Size (Mb)	GC (%)	Size (Mb)	GC (%)	Size (Mb)	GC (%)
1	740	38.6	258	38.0	232	39.6	529	40.6
2	686	38.8	244	39.8	205	41.4	506	40.5
3	558	39.3	259	38.0	201	39.6	416	41.5
4	440	39.3	212	38.0	205	41.3	274	41.9
5	415	39.3	212	38.0	207	38.7	274	41.9
6	339	40.0	177	38.2	181	39.9	219	41.8
7	294	39.9	162	38.5	201	39.6	222	41.5
8	237	41.1	153	37.7	150	40.9	204	41.2
9	121	43.2	137	38.1	126	43.0	200	41.8
10			138	38.9	154	39.5	171	41.7
11			114	40.9	125	41.0	148	44.0
12			127	39.2	114	41.0	124	46.6
13			147	39.5	123	40.4	105	44.5
14			127	39.2	83	41.3	95	45.1
15			113	37.9	123	39.2		
16			127	37.9	111	40.0		
17			119	38.8	92	42.2		
18			106	39.4	87	40.0		
19			100	38.6	82	40.7		
20			106	39.4	86	44.6		
21			101	37.3	90	41.5		
22			100	38.6	90	39.2		
23			88	40.1	59	41.7		
24			65	39.5				
25			95	39.7				
26			95	37.6				
27			70	38.4				
X	301	40.4	198	38.2	181	39.9	202	41.8
Y	135	38.3	56	41.5	38	42.0		
TG	4132	39.5	4006	38.7	3308	40.5	3690	41.7
Variance		1.83		0.75		1.81		3.13

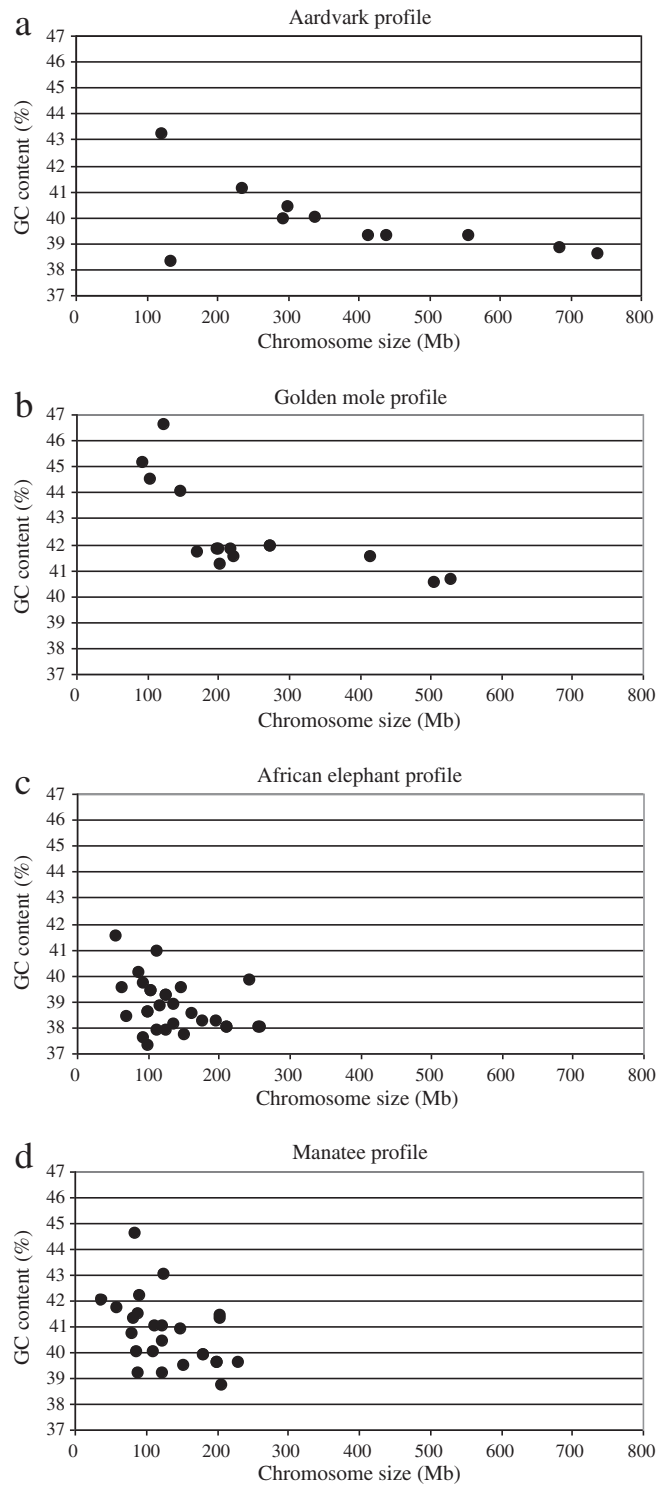


Fig. 1. Chromosome profiles showing the relationship between chromosome size and GC-content in the armadillo (a), golden mole (b), African elephant (c) and manatee (d) based on measurements from flow karyotypes. The golden mole and manatee are similar to the armadillo and elephant respectively, but have elevated GC content.

OAF, 38.7% for LAF, 40.5% for TMA and 41.7% for CAS (Table 1), and its relationship to genome size is shown in Fig. 2.

3. Discussion

The total genome sizes in Afrotherian species measured from flow karyotypes are between 3.3 and 4.1 Gb (Table 1), up to 30% larger

Table 2
Comparison of aardvark chromosomes with homologous regions in humans.

Aardvark			HSA homology		
Chr	Size (Mb)	Ratio (%)	Chromosomes	Size (Mb)	Ratio (%)
1	740	17.9	2pter-q13 + 4 + 8p + 13 + 16q + 19q	520	18.0
2	686	16.6	3 + 5 + 20 + 21	476	16.5
3	558	13.5	1 + 6 + 19p	424	14.7
4	440	10.6	10p + 11 + 12 + 22qdist	314	10.9
5	415	10.0	7a + 14 + 15	309	10.7
6	339	8.2	2q13-qter + 7b + 16p + 18	263	9.1
7	294	7.1	8q + 9	223	7.7
8	237	5.7	10q + 17	174	6.0
9	121	2.9	12qdist + 22qprox	29	1.0
X	301	7.3	X	155	5.4

than that in human. Our estimates are much smaller than the previous estimates of 5.96 Gb for aardvark, 4.51 Gb for elephant, 4.74 Gb for manatee and 6.42 Gb for golden mole [8], presented in the genome size database [7]. Genome sizes included in the database are determined by conventional methods based on measurement of nuclei DNA as a whole genome. The results of DNA size measurement from flow karyotyping in zebrafish, chicken, dog and human correlate well with estimates from genome sequence data in genomes with the most complete sequence [11,12,14,15,18–21] (Table 3). It has been suggested that all existing genome size values when assigned to the same human or other reference standards can be easily recalculated [9]. However, the differences in size between the 4 species in this study do not correlate with those in the previous study that used the Feulgen reaction [8], indicating that the inaccuracies are caused by using fixed nuclei which are affected by chromatin condensation or the inclusion of cytoplasm. It is now clear that the database contains questionable values, not only for reptiles as we have shown previously [12] but also for mammals in this study. We conclude that our measurement system is a reliable method for determining genome size and has many advantages over other techniques that yield contradictory estimations.

Our chromosome measurements provide quantitative data which enable us to compare sizes of homologous regions between OAF and human (Table 2). Chromosome 9 in the aardvark is of special interest as G-banding shows only one small dark band [2], possibly related to its high GC content. The proportion of OAF-9 in the genome is much higher than that of the corresponding human homologous regions, suggesting that OAF-9 has increased its size by additional GC-rich sequences. The DNA content of the Afrotherian X is also interesting as it is highly conserved throughout the evolution of eutherian mammals [22]. A gene mapping study shows that the gene content and order on the X are shared between African elephant and human [23]. However, OAF-X is distinctly larger in size than the X in the other three species (Table 1). This can be explained by a previous observation that L1 in

Table 3
Comparisons of genome size between different methodologies. C-values (pg) in the database are converted to base pairs (1 pg = 0.978 Gb).

Species	Genome size (Gb)			GC content (%)
	Genome size database	Flow karyotypes	Genome sequencing	
Zebrafish	1.68–2.28	1.45	1.33	
Chicken	1.22	1.15	1.05	45.2
Red-eared slider	1.89, 2.53, 2.59	1.21	N/A	47.4
Nile crocodile	2.78, 3.86	1.29	N/A	49.2
Manatee	4.57	3.31	N/A	40.5
Golden mole	6.18	3.69	N/A	41.7
African elephant	4.02, 4.34	4.01	N/A	38.9
Aardvark	5.73, 5.74	4.13	N/A	39.5
Mouse-eared bat	2.04, 2.10, 2.12	2.25	N/A	42.3
Dog	2.74–3.46	2.71	2.68	41.2
Human	3.42	3.15	3.02	41.0

the aardvark is uniformly increased across the entire X chromosome [24], indicating gain of L1 in the aardvark lineage. Except for OAF-9 and X, the proportion of each aardvark chromosome in the genome is similar to that of the homologous human regions, which represent about 90% of the aardvark genome. This suggests that the chromosomal content, with these two exceptions, has changed without obvious bias between the two species during divergence from a common eutherian ancestor.

The previous study showed that proportions of aardvark chromosomes 1, 2 and 3 in the genome were estimated to be 20.5, 18.4 and 14.3% respectively [10], and higher than our estimates. The measurements of larger chromosomes are greater when taken from photographic images rather than from base pair size calculated from the flow karyotype, indicating that larger chromosomes are less condensed structures compared to smaller chromosomes. Thus, the variation in chromosome condensation between chromosomes could be related to GC content.

The chromosome size and GC content of the four Afrotherian species (which belong to different phylogenetic orders) show two different chromosome profiles. In OAF chromosomes, except for chromosomes 9 and Y, the relative GC content is similar for each chromosome (Fig. 1a). The OAF X, while larger than the X in the other species, has a similar GC content (Table 1). A similar profile is observed in the golden mole except for its four smaller chromosomes in which the GC content is elevated compared to the aardvark (Fig. 1b). The LAF profile shows a more variable GC content (Fig. 1c), and this profile is similar to the manatee except that each manatee chromosome has a slightly higher GC content (Fig. 1d). Aardvark and golden mole karyotypes have 31 and 32 conserved segments with human chromosomes, in contrast to elephants and manatees which have 45 and 44 segments [6]. Thus, the similarities in chromosome profiles between the aardvark and golden mole, and between the elephant and manatee, are consistent with their place in the phylogenetic tree in Afrotheria [25]. The chromosomes with the highest GC content in the latter two species are LAF-11 and TMA-20, which are both homologous to human chromosome 17, a GC-rich human chromosome. Differences in size and GC content between the homologous LAF-11 and TMA-20 are greater than that of the homologous LAF-5 and TMA-5. The four profiles show lower variance in chromosome GC content than that found in the bat, dog and human [13]. Excluding OAF 9, the variance for OAF chromosomes drops to 0.63, similar to the elephant (Figs. 1a, c), and lower than the golden mole and manatee. The greater variation in chromosome GC contents in the golden mole and manatee may be related to their smaller GC-rich chromosomes and smaller genome size.

Up to now, data from chromosome measurements are limited to human, dog and bat in eutherians [13–15] and the four Afrotherian species studied here (Fig. 2). Comparisons show that the aardvark genome is 1.88 Gb larger than that of bat (*Myotis myotis*), 2.25 Gb, [13]. After

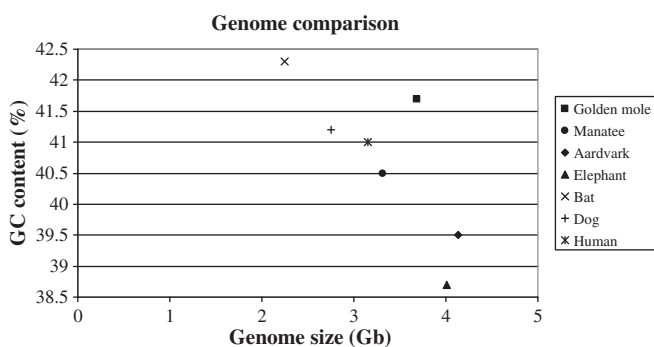


Fig. 2. Comparison of total genome size and GC-content between the four Afrotherian species and bat, dog and human. Genome size is correlated with GC content and Afrotherian genome sizes are larger than the other eutherian species.

subtraction of repetitive elements from the genome size of human (3.15 Gb), bat and dog (2.75 Gb), the non-repetitive component is calculated to be 1.62 Gb, 1.75 Gb and 1.81 Gb respectively [13], indicating that the larger genome sizes in Afrotheria are due to the greater amount of repetitive elements. Complete sequence data are not yet available for Afrotheria, partly as a result of their large genome size and partly due to the greater amount of repetitive elements. The variation in genome size within Afrotheria also can be explained by the amount of repetitive elements. The higher total GC content and the larger variation of chromosomal GC content in manatee and golden mole seems to be caused mostly by a reduction in AT-rich repetitive elements in small chromosomes. Although aardvark and elephant have different diploid numbers, their common genomic features characterised by genome size and GC content may indicate a close common ancestor.

In conclusion, measurements by flow karyotyping reveal major overestimations of genome size in previous studies and indicate the uniqueness of the Afrotherian genome. In addition to the evidence from chromosome homology, we show that patterns in chromosome GC content may provide clues to the direction of genome evolution. Similar studies on the size and GC content of Xenarthra chromosomes are likely to help in our understanding of the eutherian ancestral genome.

4. Material and methods

Chromosome preparations for sorting were made from fibroblast cultures from a male aardvark ($2n = 20, XY$), a male African elephant ($2n = 56, XY$), a male manatee ($2n = 48, XY$) and a female golden mole ($2n = 30, XX$) according to conventional protocols [26]. Normal male human lymphoblasts were used to provide a reference flow karyotype. For comparison, each chromosome sample was run on a flow cytometer (MoFlo, DAKO) separately but sequentially using the same settings. Chromosome size and GC content were calculated following the methods described previously [12]. The identity of chromosomes in each peak in the flow karyotype was based on previous studies [2,16,17]. According to homologies between aardvark and human determined by chromosome painting [2], sizes in human regions homologous to each aardvark chromosome were estimated at chromosome banding resolution from the human genome sequence database (<http://genome.ucsc.org>). Statistical analysis was performed using Microsoft Excel.

References

- [1] A.L. Roca, S.J. O'Brien, Genomic inferences from Afrotheria and the evolution of elephants, *Curr. Opin. Genet.* 15 (2005) 652–659.
- [2] F. Yang, E.Z. Alkalaeva, P.L. Perelman, A.T. Pardini, W.R. Harrison, P.C. O'Brien, B. Fu, A.S. Graphodatsky, M.A. Ferguson-Smith, T.J. Robinson, Reciprocal chromosome painting among human, aardvark, and elephant (superorder Afrotheria) reveals the likely eutherian ancestral karyotype, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 1062–1066.
- [3] L. Froenicke, M.G. Caldés, A. Graphodatsky, S. Müller, L.A. Lyons, T.J. Robinson, M. Volleth, F. Yang, J. Wienberg, Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? *Genome Res.* 16 (2006) 306–310.
- [4] M. Svartman, R. Stanyon, The chromosomes of Afrotheria and their bearing on mammalian genome evolution, *Cytogenet. Genome Res.* 137 (2012) 144–153.
- [5] M.A. Ferguson-Smith, V. Trifonov, Mammalian karyotype evolution, *Nat. Rev. Genet.* 8 (2007) 950–962.
- [6] A.S. Graphodatsky, V.A. Trifonov, R. Stanyon, The genome diversity and karyotype evolution of mammals, *Mol. Cytogenet.* 4 (2011) 22.
- [7] T.R. Gregory, Animal genome size database, <http://www.genomesize.com> 2013.
- [8] C.A. Redi, S. Garagna, M. Zuccotti, E. Capanna, Genome size: a novel genomic signature in support of Afrotheria, *J. Mol. Evol.* 64 (2007) 484–487.
- [9] J. Dolezel, J. Greilhuber, Nuclear genome size: are we getting closer? *Cytometry A* 77 (2010) 635–642.
- [10] K. Benirschke, D.H. Wurster, R.J. Low, N.B. Atkin, The chromosome complement of the aardvark, *Oryzopteropus afer*, *Chromosoma* 31 (1970) 68–78.
- [11] J.L. Freeman, A. Adeniyi, R. Banerjee, S. Dallaire, S.F. Maguire, J. Chi, B.L. Ng, C. Zepeda, C.E. Scott, S. Humphray, J. Rogers, Y. Zhou, L.I. Zon, N.P. Carter, F. Yang, C. Lee, Definition of the zebrafish genome using flow cytometry and cytogenetic mapping, *BMC Genomics* 8 (2007) 195.
- [12] F. Kasai, P.C. O'Brien, M.A. Ferguson-Smith, Reassessment of genome size in turtle and crocodile based on chromosome measurement by flow karyotyping: close similarity to chicken, *Biol. Lett.* 23 (2012) 631–635.
- [13] F. Kasai, P.C. O'Brien, M.A. Ferguson-Smith, The bat genome; GC-biased small chromosomes associated with reduction in genome size, *Chromosoma* (2013), <http://dx.doi.org/10.1007/s00412-013-0426-9>.
- [14] C.F. Langford, P.E. Fischer, M.M. Binns, N.G. Holmes, N.P. Carter, Chromosome-specific paints from a high-resolution flow karyotype of the dog, *Chromosome Res.* 4 (1996) 115–123.
- [15] B. Trask, G. van den Engh, B. Mayall, J.W. Gray, Chromosome heteromorphism quantified by high-resolution bivariate flow karyotyping, *Am. J. Hum. Genet.* 45 (1989) 739–752.
- [16] C. Gilbert, P.C. O'Brien, G. Bronner, F. Yang, A. Hassani, M.A. Ferguson-Smith, T.J. Robinson, Chromosome painting and molecular dating indicate a low rate of chromosomal evolution in golden moles (Mammalia, Chrysochloridae), *Chromosome Res.* 14 (2006) 793–803.
- [17] A.T. Pardini, P.C. O'Brien, B. Fu, R.K. Bonde, F.F. Elder, M.A. Ferguson-Smith, F. Yang, T.J. Robinson, Chromosome painting among Proboscidea, Hyracoidea and Sirenia: support for Paenungulata (Afrotheria, Mammalia) but not Tethytheria, *Proc. Biol. Sci.* 274 (2007) 1333–1340.
- [18] K. Howe, et al., The zebrafish reference genome sequence and its relationship to the human genome, *Nature* 496 (2013) 498–503.
- [19] International Chicken Genome Sequencing Consortium, Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution, *Nature* 432 (2004) 695–716.
- [20] K. Lindblad-Toh, C.M. Wade, T.S. Mikkelsen, et al., Genome sequence, comparative analysis and haplotype structure of the domestic dog, *Nature* 438 (2005) 803–819.
- [21] M.I. Jensen-Seaman, T.S. Furey, B.A. Payseur, et al., Comparative recombination rates in the rat, mouse, and human genomes, *Genome Res.* 14 (2004) 528–538.
- [22] S. Ohno, *Sex Chromosomes and Sex-linked Genes*, Springer-Verlag, New York, 1967.
- [23] C.L. Delgado, P.D. Waters, C. Gilbert, T.J. Robinson, J.A. Graves, Physical mapping of the elephant X chromosome: conservation of gene order over 105 million years, *Chromosome Res.* 17 (2009) 917–926.
- [24] P.D. Waters, G. Dobigny, A.T. Pardini, T.J. Robinson, LINE-1 distribution in Afrotheria and Xenarthra: implications for understanding the evolution of LINE-1 in eutherian genomes, *Chromosoma* 113 (2004) 137–144.
- [25] M. Kuntner, J.L. May-Collado, I. Agnarsson, Phylogeny and conservation priorities of afrotherian mammals (Afrotheria, Mammalia), *Zool. Scr.* 40 (2011) 1–15.
- [26] M.A. Ferguson-Smith, Genetic analysis by chromosome sorting and painting: phylogenetic and diagnostic applications, *Eur. J. Hum. Genet.* 5 (1997) 253–265.