

## REVIEW

## Centromere repositioning in mammals

M Rocchi<sup>1</sup>, N Archidiacono<sup>1</sup>, W Schempp<sup>2</sup>, O Capozzi<sup>1</sup> and R Stanyon<sup>3</sup>

The evolutionary history of chromosomes can be tracked by the comparative hybridization of large panels of bacterial artificial chromosome clones. This approach has disclosed an unprecedented phenomenon: ‘centromere repositioning’, that is, the movement of the centromere along the chromosome without marker order variation. The occurrence of evolutionary new centromeres (ENCs) is relatively frequent. In macaque, for instance, 9 out of 20 autosomal centromeres are evolutionarily new; in donkey at least 5 such neocentromeres originated after divergence from the zebra, in less than 1 million years. Recently, orangutan chromosome 9, considered to be heterozygous for a complex rearrangement, was discovered to be an ENC. In humans, in addition to neocentromeres that arise in acentric fragments and result in clinical phenotypes, 8 centromere-repositioning events have been reported. These ‘real-time’ repositioned centromere-seeding events provide clues to ENC birth and progression. In the present paper, we provide a review of the centromere repositioning. We add new data on the population genetics of the ENC of the orangutan, and describe for the first time an ENC on the X chromosome of squirrel monkeys. Next-generation sequencing technologies have started an unprecedented, flourishing period of rapid whole-genome sequencing. In this context, it is worth noting that these technologies, uncoupled from cytogenetics, would miss all the biological data on evolutionary centromere repositioning. Therefore, we can anticipate that classical and molecular cytogenetics will continue to have a crucial role in the identification of centromere movements. Indeed, all ENCs and human neocentromeres were found following classical and molecular cytogenetic investigations.

*Heredity* (2012) **108**, 59–67; doi:10.1038/hdy.2011.101; published online 2 November 2011

**Keywords:** neocentromeres; mammals; centromere movements; evolutionary new centromeres; evolution

## INTRODUCTION

The centromere is a complex chromosomal structure responsible for proper chromosome/chromatid segregation at meiosis and mitosis. In almost all eukaryotes, centromeres are found at specific locations along chromosomes and are composed of, occasionally very large, blocks of satellite DNA. Evidence shows that in spite of the very high conservation of centromeric proteins (CENP), the satellite DNA sequences can substantially differ even among closely related species. Recently, two interconnected phenomena, human neocentromeres (HN) and evolutionary new centromeres (ENC, also called repositioned centromeres), have revolutionized our understanding of centromere function and its relationship to the underlying DNA sequences. HNs are centromeres that emerge in ectopic chromosomal regions and are devoid of aliphoid sequences, that is, the satellite DNA present at primate centromeres. ENCs are centromeres that move to a new position along the chromosome without any change in marker order (no inversion or other structural rearrangements).

## THE DISCOVERY OF EVOLUTIONARY NEW CENTROMERES

In the 1970s, chromosome banding triggered renewed interest in studies on karyotype evolution in primates and many other mammalian orders. The different position of the centromere along a chromosome was almost always interpreted as the result of a pericentric inversion or complex rearrangement. However, in Dutrillaux’s wide-ranging study of chromosomal evolution in 60 species of primates, ‘centromere translocation’ was given as a possible mechanism for the

evolution of chromosome 11 in some Cercopithecidae (Dutrillaux, 1979). He also hypothesized that in the case of Cercopithecidae with high diploid numbers, where the fission was not centromeric, there had to be a gain in centromeres. In the 1990 review of the evolution of human chromosomes, Clemente *et al.* (1990) hypothesized that differences in centromere position in homologs to chromosomes 4, 6 and 10 did not appear to be the result of inversions but seemed to result from the ‘activation/inactivation of centromeres’.

The advent of fluorescence in situ hybridization (FISH) technology, painting probes in particular, provided more solid, reliable tools to study karyotype evolution (Wienberg *et al.*, 1990; Jauch *et al.*, 1992). However, painting probes, even though very efficient in spotting chromosomal translocations, were not able to distinguish between an inversion and a centromere-repositioning event. Fortunately, the human genome-sequencing project produced large libraries of precisely mapped bacterial artificial chromosome (BAC) clones that can be very efficiently used in FISH experiments. Sequencing projects of other vertebrate species, using the shotgun approach, made extensive use of BACs and fosmids. The end sequences of these clones continue to be exploited to reliably anchor and close sequence contigs. The systematic use of BAC-FISH was very effective in guiding and disambiguating the sequence assembly of some genomes. See, for example, the cytogenetic frames that supported the sequence assembly in macaque (Gibbs *et al.*, 2007) as reported at <http://www.biologia.uniba.it/macaque>, and in orangutan (Locke *et al.*, 2011) as reported at <http://www.biologia.uniba.it/orang>. The BAC-FISH approach

<sup>1</sup>Department of Biology, University of Bari, Bari, Italy; <sup>2</sup>Department of Human Genetics, University of Freiburg, Freiburg, Germany and <sup>3</sup>Department of Evolutionary Biology, University of Florence, Florence, Italy

Correspondence: Professor M Rocchi, Department of Biology, University of Bari, Via Orabona 4, 70125 Bari, Italy.

E-mail: [roccchi@biologia.uniba.it](mailto:roccchi@biologia.uniba.it)

Received 3 June 2011; revised 29 July 2011; accepted 2 August 2011; published online 2 November 2011

represented a powerful, visual link in dealing with the cytogenetic organization of the species under study, with special reference to the centromeres whose position is almost impossible to spot from sequencing data.

Montefalcone *et al.* (1999) were the first to unequivocally demonstrate the existence of the evolutionary centromere-repositioning phenomenon. They traced the evolutionary and phylogenetic history of chromosome IX in primates by the FISH of cloned DNA. It became clear that if the position of the centromere was not taken into account, a much more parsimonious scenario of rearrangements could be hypothesized to account for between-species marker order differences. When the centromere was included, the analysis became an impossible

jigsaw puzzle. The centromere was therefore hypothesized to have repositioned along the chromosome, independently from the surrounding markers, with no need to hypothesize a seemingly endless series of inversions. Over the last decade numerous other studies have found ENC in primates and in other mammalian orders (Table 1). ENCs are now accepted as an important mechanism of genome evolution ranked on equal grounds with traditional chromosome rearrangements such as inversions, translocation, deletions and insertions.

### ENCs AND HNS

Several lines of evidence (see below) suggest that ENCs and HNs are related phenomena. As for any mutational event, a fixed ENC must

**Table 1** Described cases of centromere-repositioning events in mammals

Species	Chromosome	Reference
Wallaby tammar ( <i>Macropus eugenii</i> ) (marsupial)	MEU7	Ferreri <i>et al.</i> (2005)
Ryukyu spiny rat ( <i>Tokudaia osimensis</i> )	TOSX	Kobayashi <i>et al.</i> (2008)
Horse ( <i>Equus caballus</i> )	ECA11	Wade <i>et al.</i> (2009)
Donkey ( <i>Equus asinus</i> ), plains zebra ( <i>Equus burchelli</i> )	EBU12	Carbone <i>et al.</i> (2006)
	EAS8	
	EAS9	
	EAS11	
	EAS15	
	EAS19	
Donkey ( <i>Equus asinus</i> ), plains zebra ( <i>Equus burchelli</i> )	EAS18/EBU20	Piras <i>et al.</i> (2009)
	EAS16/EBU17	
Ring-tailed lemur ( <i>Lemur catta</i> )	LCAX (HSAX)	Ventura <i>et al.</i> (2001)
Black lemur ( <i>Eulemur macaco</i> )	EMAX (HSAX)	Ventura <i>et al.</i> (2001)
Squirrel monkey ( <i>Saimiri sciureus sciureus</i> ) (NWM)	SSCX (HSAX)	Present paper
Peruvian squirrel monkey ( <i>Saimiri boliviensis peruviansis</i> ) (NWM)	SB0p15(HSA8)	Stanyon <i>et al.</i> (2008)
Brown woolly monkey ( <i>Lagotrix lagotricha</i> ) (NWM)	LLA8 (HSA13)	Cardone <i>et al.</i> (2006)
	LLA9 (HSA1)	Stanyon <i>et al.</i> (2008)
	LLA10 (HSA17)	
	LLA11 (HSA5)	Unpublished
	LLA22 (HSA3)	Ventura <i>et al.</i> (2004)
	LLA28 (HSA1)	Stanyon <i>et al.</i> (2008)
Common marmoset ( <i>Callithrix jacchus</i> ) (NWM)	CJA3 (HSA4)	Stanyon <i>et al.</i> (2008)
	CJA9 (HSA12)	
	CJA15 (HSA3)	Ventura <i>et al.</i> (2004)
	CJA17 (HSA3)	
	CJA21 (HSA3)	
White-coated titi ( <i>Callicebus pallescens</i> ) (NWM)	CPA13 (HSA14)	Ventura <i>et al.</i> (2003)
	CPA18 (HSA13)	Cardone <i>et al.</i> (2006)
	CPA (HSA8)	Stanyon <i>et al.</i> (2008)
	CPA20 (HSA3)	Ventura <i>et al.</i> (2004)
	CPA21 (HSA13)	Cardone <i>et al.</i> (2006)
	CPA22 (HSA20)	Misceo <i>et al.</i> (2005)
Vervet monkey ( <i>Cercopithecus Aethiops</i> ) (OWM)	CAE22 (HSA3)	Ventura <i>et al.</i> (2004)
Rhesus macaque ( <i>Macaca mulatta</i> ) (OWM)	MMU4 (HSA6)	Eder <i>et al.</i> (2003)
	MMU14 (HSA11)	Cardone <i>et al.</i> (2007)
	MMU1 (HSA1)	Ventura <i>et al.</i> (2007)
	MMU2 (HSA3)	
	MMU12 (HSA2q)	
	MMU13 (HSA2p)	
	MMU17 (HSA13)	
	MMU18 (HSA18)	
	MMU15 (HSA9)	Montefalcone <i>et al.</i> (1999)
Lar gibbon ( <i>Hylobates lar</i> )	HLA11 (HSA11)	Misceo <i>et al.</i> (2008)
White-cheeked gibbon ( <i>Nomascus leucogenys</i> )	NLE15 (HSA11)	Roberto <i>et al.</i> (2007)

Abbreviations: NWM, new world monkeys; OWM, old world monkeys.

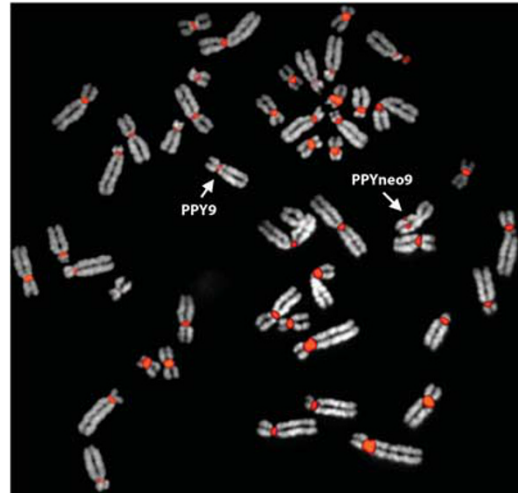
occur in a single chromosome and then spread in the population. Yet, large-scale cytogenetic studies at the population level are available, with few exceptions, only for humans (Bhasim, 2007). Our knowledge of the karyotypes of most species was usually gained by investigating just a few individuals. These small sample sizes were justified by the relatively high conservation of the karyotype in a species leading to the simplification that each species had one karyotype (Dutrillaux, 1979). As a consequence, the chances of spotting a polymorphic ENC at an early stage were very low. However, there is at least one exception, illustrated below, in the orangutan, which was thought to be a complex inversion polymorphism, but is now known to be an ENC (Locke *et al.*, 2011).

Cytogenetic studies of HNs have two enormous advantages. The first is that huge numbers of individuals go through a powerful clinical filter. The vast majority of HNs were seeded in acentric fragments generated by a fortuitous rearrangement. The acentric fragments would have been lost in normal circumstances, but, occasionally, a neocentromere is seeded, ensuring the rescue of the fragment. However, the fragment constitutes a supernumerary chromosome causing more or less severe phenotypic consequences, requiring medical attention. About 100 cases of HN have been described (for a review, see Marshall *et al.*, 2008). A second important advantage is the widespread practice of cytogenetic prenatal diagnosis because it can be regarded as a large ongoing population study. Prenatal diagnosis often discloses fortuitous events, like centromere shifts, that otherwise would have never been disclosed. These 'real-time' centromere-repositioning events in humans mimic seeding events that lead to the formation of an ENC, supporting the view that ENC and HN are two faces of the same coin. For this reason, in discussing neocentromeres in mammals, many hints and hypotheses came from what we have learned from HNs. Therefore, they deserve a short summary.

## HUMAN NEOCENTROMERES

The vast majority of HN was seeded in acentric fragments consisting of inverted duplications of a distal portion of a chromosome arm. They were classified as class I by Marshall *et al.* (2008): acentric duplicated fragments stabilized by a neocentromere function as supernumerary chromosomes with clinical manifestations. Class II neocentromeres are formed by acentric fragments that were excised to form linear or ring chromosomes. Clinical problems are due to accompanying deletions or gene disruption. Ring instability can also lead to the loss or duplication of the ring. Occasionally, class II neocentromeres were discovered following the malsegregation of a balanced rearrangement that did not cause phenotypic problems in the transmitting parent (Capozzi *et al.*, 2008).

A third type of very rarely reported neocentromeres is the most pertinent to our discussion. These neocentromeres arise in intact chromosomes and functionally replace the normal centromere. The old centromere appears unchanged but functionally inactivated, as proved by the absence of CENP-A and other centromere-specific proteins that are conversely present at the new centromeric site (Warburton *et al.*, 1997; Voullaire *et al.*, 1999). They do not cause clinical problems and, indeed, were discovered serendipitously, mostly through amniocentesis. Eight such cases have been described (see Table 1 in Hasson *et al.*, 2011): two were *de novo*, and another two segregated through at least three generations. Two *de novo* cases, one on chromosome Y (Bukvic *et al.*, 1996) and another on chromosome 7 (Liehr *et al.*, 2010), are relevant for understanding the timing of normal centromere inactivation. The chromosome 7 case did not show any old/new centromere mosaicism, thus suggesting that the



**Figure 1** Result of a fluorescence in situ hybridization (FISH) experiment on orangutan heterozygous for the evolutionary new centromeres (ENC) on chromosome 9, using the amplification products of a PCR experiment as a probe, the total orangutan DNA as template and primers specific for the orangutan alpha satellite DNA. Note the very small size of the centromeric alpha-satellite signal on both normal and variant chromosome 9. Forward primer: 5'-TCAACTCTGTGAGATGAATGCAAAC-3'; reverse primer: 5'-AAACATCTTTGTGATGTGTCATTC-3'. PCR conditions: 95 °C for 3 min; 35 times: 95 °C for 30 min, 60 °C for 60 min and 72 °C for 40 min. Primers were derived from a consensus sequence constructed using all the centromeric stretches of alpha satellite DNA of orangutan, available on the trace archive database (<http://www.ncbi.nlm.nih.gov/Traces/home/>).

neo-chromosome 7 was already present in one gamete. The chromosome Y case was mosaic 45,X/46,XY/46,XneoY, thus indicating that the event was post-zygotic. In the familiar or *de novo* case, no functionally dicentric chromosomes or mosaicism was observed. Instead mosaicism is relatively frequent in clinical neocentromeres, suggesting that neocentromeres might not be very efficient in ensuring mitotic segregation. The two contrasting views can be reconciled considering that there is probably a strong selection against the loss of normal neocentromeric chromosomes, which favors the loss of supernumerary neocentromeric markers.

An additional important point provided by HN is their clustering in specific chromosomal domains. Neocentromeres at 3q, 8p, 13q, 15q and Yq are especially frequent (see Figure 1a in Marshall *et al.*, 2008).

## ENCs IN MAMMALS

We can assume that neocentromeric chromosomes of class I and II are not limited to humans. However, because of the clinical problems they cause, these neocentromeres have no evolutionary perspective. Indeed, most ENCs belong to the third category of HN. A distinct ENC category, not yet reported clinically, is represented by neocentromeres that arise following a chromosomal fission with a breakpoint outside the centromere. In these cases a neocentromere forms in the acentric fragment (occasionally in both fragments), resulting from the fission. ENCs of chromosomes 15 and 14 were generated in this way, following the fission of an ancestral chromosome corresponding to chromosome 7 in macaque (Ventura *et al.*, 2003).

## FREQUENCY OF ENCS IN MAMMALS

Enough systematic data have now accumulated to provide information on the evolutionary rate and frequency of ENCs. In the macaque

(*Macaca mulatta*), 9 out of 20 autosomal centromeres were shown to be evolutionarily new: chromosomes 1, 2, 4, 12, 13, 14, 15, 17 and 18 (Ventura *et al.*, 2007). Comparative data show that these ENC are found in all the studied old world monkeys (OWM); therefore they accumulated during the ~14 million years span from the Hominoidea/Cercopithecoidea split (~32 million years ago, MYA) to the Cercopithecinae/Colobinae divergence (~18 MYA) (Perelman *et al.*, 2011). By comparison, six human centromeres are evolutionarily new. Centromeres of chromosomes 3, 6 and 11 were repositioned along the chromosomes (Ventura *et al.*, 2004; Cardone *et al.*, 2007; Capozzi *et al.*, 2009). Those of 14 and 15 were seeded, as mentioned, in the Hominoidea ancestor following the fission of a chromosome corresponding to chromosome 7 in macaque (Ventura *et al.*, 2003). A non-centromeric fission of the 3/21 association synteny in the Hominoidea ancestor generated chromosome 21 and its neocentromere.

A very informative additional example is provided by the evolutionary history of Equidae. Carbone *et al.* (2006) compared chromosomal marker order between Burchelli's zebra (*Equus burchelli*) and the donkey (*Equus asinus*), using the horse (*Equus caballus*) as an outgroup. Equidae, and these three species in particular, underwent a recent, rapid evolution and accumulated a large number of chromosomal changes (Trifonov *et al.*, 2008). Zebra and donkey diverged about 0.9 MYA, while their common ancestor diverged from the horse around 2 MYA (Oakenfull and Clegg, 1998; Oakenfull *et al.*, 2000). The study revealed that eight centromere-repositioning events took place during the evolution of this genus. Surprisingly, at least five cases occurred in the donkey after its divergence from zebra. 'At least', because some chromosomes are very small in these species, and marker order and the position of the centromere could not be established with certainty in these tiny chromosomes.

These examples show that ENC formation can be relatively frequent, on a par with other types of chromosome rearrangements. What about the frequency of centromere-repositioning seeding events in general, including those that were seeded, but disappeared from the population or remained at low frequency and have not yet been discovered? The data in humans support the idea that the fixed ENCs are just the tip of an iceberg.

### NEOCENTROMERES SEEDING

The vast majority of the ENCs possess a heterochromatic block similar to normal centromeres. This fact is particularly evident, for instance, in macaque, where all the nine ENCs have large blocks of alphoid DNA indistinguishable from other macaque centromeres (Ventura *et al.*, 2007). Mature repositioned centromeres are thought to have slowly acquired the large arrays of satellite DNA after their seeding in an anonymous sequence. The macaque-repositioning events occurred at least 18 MYA, providing plenty of time to 'mature'.

The possibility that an ENC could have resulted from a transposition of the functional centromere cannot be discarded with certainty. However, the following lines of evidence support the view that ENCs were the result of epigenetic events and not the transposition of particular sequences.

- All HNs are devoid of satellite DNA. FISH analysis did not detect any fluorescent signal at the neocentromeric loci, and, most importantly, in all class I cases the neocentromere seeding was an opportunistic event triggered by the acentric fragment formation. The probability, in these cases, of a simultaneous alpha-satellite sequence transposition can be reasonably assumed as unrealistic.

- Chromatin immunoprecipitation followed by hybridization on microarrays (ChIP-on-chip analysis) in HN cases, using anti-CENP-A and/or anti-CENP-C antibodies (see below), always showed that the centromeric function was associated with single-copy sequences. This circumstantial evidence was recently supported by data on horse and orangutan ENCs (Wade *et al.*, 2009; Locke *et al.*, 2011) (for details on orangutan see below). In both cases the neocentromere, precisely mapped by ChIP-on-chip analysis, was located in regions devoid of satellite sequences.

What about the features of the sequence underlying the neocentromere? ChIP-on-chip analysis has been performed in a number of mammalian neocentromere cases (Lo *et al.*, 2001a,b; Alonso *et al.*, 2003, 2007; Cardone *et al.*, 2006; Capozzi *et al.*, 2008, 2009; Wade *et al.*, 2009; Hasson *et al.*, 2011; Locke *et al.*, 2011). A comparison revealed no striking similarities, with only marginal shared features, like an occasional abundance of LINE1 repeats.

The cytogenetic mapping of neocentromeres showed that some of them cluster to specific chromosome domains, 3q, 13q and 15q, in particular (Marshall *et al.*, 2008). This finding suggests that at least some of them might be linked to a specific sequence, but ChIP-on-chip analysis pointed out that no two studied neocentromeres, apparently mapping to the same locus, shared the same seeding point sequence (Alonso *et al.*, 2003; Ventura *et al.*, 2004; Hasson *et al.*, 2011).

One ENC and one HN raised relevant points of discussion in this context.

The ENC found in macaque chromosome 18 (human 18) perfectly corresponded, in humans, to a clone gap, positioned at chromosome 18: 50 313 135–50 360 134 (UCSC genome browser, hg18 release) (Carbone *et al.*, 2009). We found that the gap was composed of non-alpha, satellite-like DNA. Sequence analysis of several primate species suggested that this sequence was present in the Cercopithecidae ancestor at the time of the neocentromere seeding. This satellite DNA was subsequently replaced by alpha satellite DNA.

A second, notable, HN case was reported by Hasson *et al.* (2011). These authors investigated an HN in chromosome 8. Different experimental approaches indicated that the neocentromere was seeded in a domain at 8q21, which consisted of a large array of tandemly repeated DNA with a monomer of 12 kb. This tandemly repeated DNA more closely resembled multiple segmental duplications (SDs) than classical satellite DNA, in which the repeats are usually much shorter. The presence of SDs at the seeding point was also found for other neocentromeres, those clustering at 15q24-26 in particular. This region is intriguing for an additional reason. The evolutionary history of chromosome 15 showed that chromosomes 15 and 14 originated, as mentioned, from a non-centromeric chromosomal fission (Ventura *et al.*, 2003). A neocentromere was formed in both derivative chromosomes, and the ancestral centromere, located to a region corresponding to 15q24-26, inactivated. The abundant SDs clustered at this domain are remains of the pericentromeric SDs that flanked the ancestral inactivated centromere. Capozzi *et al.* (2009) recently reported a similar example. The centromere of chromosome 6 in the primate ancestor was, very likely, at 6p22.1. It repositioned to the present day location in humans in the Hominoidea ancestor. The authors report a familial case in which the centromere repositioned back to its ancestral location. These findings raise additional points. Are there hidden sequence features, which are a legacy of the inactivated centromere? Is the legacy, if present, because of primary or secondary structures? Additionally, why do so many neocentromeres cluster at 15q, and a single one at 6p and none to 2q21.2, where an ancestral centromere recently inactivated following the telomere-

telomere fusion that generated human chromosome 2? One hypothesis is that trisomies/tetrasomies of the distal part of chromosome 15 are compatible with life, whereas trisomies for other regions are not. Indeed, the neocentromere on chromosome 6 was found in an otherwise normal chromosome.

Additional intriguing relationships between neocentromeres and ENC's have been reported for chromosome 13 and chromosome 3. In the case of chromosome 13, two novel ENC's were seeded in the same chromosomal domain in OWMs and pig (*Sus scrofa*), which diverged about 95 MYA (Cardone *et al.*, 2006). In human chromosome 3, a repositioned centromere (normal phenotype, found by chance) and a clinical neocentromere were seeded to the 3q26 chromosomal domain, the locus where the centromere repositioned in the OWM ancestor (Ventura *et al.*, 2004). Therefore, the same domain has been used as seeding point of an ENC and HNs.

Roizes (2006) hypothesized that centromere-repositioning events can be indirectly elicited by mutations, like retrotransposon insertions in the centromere, that could potentially affect functionality. Hasson *et al.* (2011) noticed that the alpha-satellite block of the inactivated centromere of the repositioned chromosome 8 showed a substantially reduced size of the alpha satellite array with respect to its homolog. In the orangutan, both the normal and repositioned chromosome 9 showed a very low amount of centromeric alpha-satellite heterochromatin (see below).

## AN ALTERNATIVE HYPOTHESIS

Zeitlin *et al.* (2009) demonstrated that CENP-A, a crucial component of the centromere, is rapidly recruited to DNA double-strand breaks, along with three components (CENP-N, CENP-T and CENP-U) associated with CENP-A at centromeres. The authors argue that, 'since cell survival after radiation-induced DNA damage correlates with CENP-A expression level, we propose that CENP-A may have a function in DNA repair'. These authors also hypothesized that a neocentromere could emerge because of the presence of CENP-A at the breakpoint. All class I and II neocentromeres were seeded after a break that generated an acentric fragment. The closeness of the breakpoint to the neocentromere location has been noticed in some studies (Ventura *et al.*, 2003). However, in other studies no relationship was found between neocentromeres and breakpoints (Warburton *et al.*, 2000).

## THE ENC POLYMORPHISM IN ORANGUTAN CHROMOSOME 9 (HUMAN 12)

Since the early days of comparative banding, cytogeneticists were aware that chromosome 9 in the orangutan had two forms. The difference was interpreted as an intra-chromosomal translocation and insertion of a segment containing the centromere (Turleau *et al.*, 1975) or a paracentric inversion within a intrachromosomal translocation (de Boer and Seuanez, 1982). Later, de Boer and Seuanez (1982) and Ryder and Chemnick (1993) showed that these variants demonstrated true polymorphism in orangutans by karyotyping numerous individuals, however, they never questioned that the different chromosome forms were due to complex structural rearrangements.

Now, we understand that this polymorphism is not a complex rearrangement but an ENC (Locke *et al.*, 2011). The heterochromatic blocks of aliphoid DNA of both repositioned and normal chromosome 9 are almost undetectable by FISH (Figure 1). As mentioned above, the reduced size could have negatively affected the centromere functionality of this chromosome, and thus indirectly favored neocentromere emergence.

Note that to facilitate comparison with human chromosomes, Locke *et al.* (2011) referred to this chromosome as orangutan 12.

However, here we preferred to follow the ICSN recognized standard nomenclature and will continue to refer to this chromosome as orangutan 9.

Over the years the laboratory at Freiburg had the opportunity to karyotype a total of 59 orangutans, which we report here for the first time (Supplementary Information). The publication of de Boer and Seuanez (1982) listed 71 individuals, according to their assigned species and stud book registration number. These authors previously reported on 11 out of 59 orangutans, studied by the Freiburg lab. Combining the two data sets (119 orangutans), we have the following distribution of karyotypes for chromosome 9:

- 51 Bornean orangutans, (*Pongo pygmaeus*) with 32 homozygous normal, 14 heterozygous and 5 homozygous ENC individuals providing a frequency of 0.235 for the ENC.
- 50 Sumatran orangutans, (*Pongo abelii*) with 26 homozygous normal, 22 heterozygous and 2 homozygous ENC individuals providing a frequency of 0.260 for the ENC.
- 18 hybrid orangutans, with 12 homozygous normal, 6 heterozygous and 0 homozygous ENC individuals providing a frequency of 0.167 for the ENC.

Ryder and Chemnick (1993) studied 141 orangutans, but they did not individually list each individual or clearly divide them into the two species. The frequency of the ENC in their total sample is 0.138. This figure is considerable lower than the ENC frequency of 0.235 in our total sample plus that of de Boer and Seuanez (1982).

It is not a simple task to relate these frequencies to those that would actually be found in natural populations. However, it is clear that the ENC frequency is notable. To understand clearly the implication and dating of the ENC origins, we need to briefly review what is known about the taxonomic, phylogenetic history of orangutans.

## DATING THE ORIGIN OF THE ORANGUTAN ENC

The taxonomic level of Bornean and Sumatran orangutans was debated for some time. Since the mid 1990s, it became ever more generally accepted that two species are present: *Pongo pygmaeus* in Borneo and *P. abelii* in Sumatra (Zhi *et al.*, 1996; Perelman *et al.*, 2011). In general biomolecular dates cluster around 1.5 million years for the separation of the two taxa. However, comparisons of the sequenced genome assemblies provided a much lower estimate of about 400 000 years. This low estimate may derive from the overall slowdown in genome evolution of the orangutan noted by these same authors (Locke *et al.*, 2011).

All dates of divergence between the two recognized species, whether early or late, are still amply before the final separation of Borneo and Sumatra into two islands (Steiper, 2006; Goossens *et al.*, 2008).

It seems highly likely, given the distribution in both species of orangutans, that the ENC emerged in their common ancestor after divergence from the line leading to the African apes and humans: between ~15 MYA and ~1 MYA. It may be that the emergence was closer to this last date, because it seems never to have acquired all the characteristics of a mature centromere (Locke *et al.*, 2011) (see also Figure 1). Additionally, we would expect that if it was old it would have been fixed or lost. A meiotic exchange in an individual heterozygous for an ENC within the region delimited by the old and the novel centromeres would result in dicentric and acentric chromosomes. Both derivatives are probably lost. However, a dicentric chromosome could inactivate one or the other centromere, thus reverting back to a normal or neocentric.

An important point is that the ENC polymorphism survived a fairly recent speciation event. Another point is that either the orangutan population was never particularly small or unknown selection factors are maintaining this polymorphism.

### ENCs IN THE X-CHROMOSOME OF NEW WORLD SQUIRREL MONKEYS

The X chromosome is probably the most conserved chromosome among mammals (Chowdhary *et al.*, 1998). Primate species have, with few exceptions (see Ventura *et al.* 2001), X chromosomes that are apparently identical in banding and centromere position to the human X. Therefore, the finding by Schempp *et al.* (1989) that the X chromosome of *Saimiri sciureus* (SSC) had undergone unclear intrachromosomal rearrangements, which had apparently moved the pseudoautosomal region to distal part of the long arm was of particular interest. Later Dumas *et al.* (2007) hypothesized that the X-chromosome in SSC either differed by a pericentric inversion or centromere shift.

If the X-chromosome of *S. sciureus* had an ENC it would raise a series of questions about its distribution and evolution in new world primates. We then proceeded to study the marker order of the X chromosome of squirrel monkeys, using a panel of appropriate BAC clones (see Table 2). The synteny of the two BAC clones RP11-552J9 (Xp11.22) and RP11-135B16 (Xq11.1), flanking the human centromere, was not disrupted, but the two markers mapped to the long arm of the SSC chromosome X (Figure 2). The analysis revealed that a segment delimited, in human, by BAC RP11-24M7 (HSAXq21.33) and BAC RP11-265K3 (HSAXq28; at chromosome X: 154 603 527–154 763 828, very close to the telomere chromosome X: 154 913 754) was inverted, and that a centromere was present at the breakpoint corresponding, in human, to Xq21.33 (Figure 2). The most parsimonious interpretation is that, in concomitance to the inversion, a centromere was seeded at the breakpoint at Xq21.33. The seeding event could have been favored by the break (see above) and/or by the presence of subtelomeric repetitive sequences. However, different temporal sequence of the inversion and centromere seeding events could not be discarded.

### DATING THE ORIGIN AND PHYLOGENETIC DISTRIBUTION OF THE X CHROMOSOME ENC

In order to understand better the origins and distribution of the neoX chromosome, we need to briefly summarize what is known about the taxonomy and phylogeny of squirrel monkeys. This is a controversial group of New World monkeys. Historically anywhere from 1 to 7 species and up to 16 subspecies of this new world primate were recognized. Prior to Hershkovitz (1982), squirrel monkeys were generally regarded as a single species. Hershkovitz (1982), considering morphology, geographic distribution and relying on cytogenetic data, divided squirrel monkeys into four species: *Saimiri boliviensis*, *S. sciureus*, *S. ustus* and *S. oerstedii*. An additional species, *S. vanzolinii*, was reported in 1985 (Ayres, 1985). Although Costello *et al.* (1993) minimized the importance of the cytogenetic data and recognized only two species, most workers have generally followed Hershkovitz with some slightly different arrangements. Groves, for instance, recognized five *Saimiri* species (Groves, 2001).

Cytogeneticists had long recognized that squirrel monkeys from various geographic regions all had 44 chromosomes, but differences were found in the number of acrocentric and biarmed chromosomes (Jones and Ma, 1975; Lau and Arrighi, 1976; Cambefort and Moro, 1978; Dutrillaux and Couturier, 1981; Moore *et al.*, 1990; Garcia *et al.*, 1995; Scammell *et al.*, 2001). The differences range

**Table 2 Human BACs used for FISH experiments on SSC to determine the marker order of the SSC chromosome X**

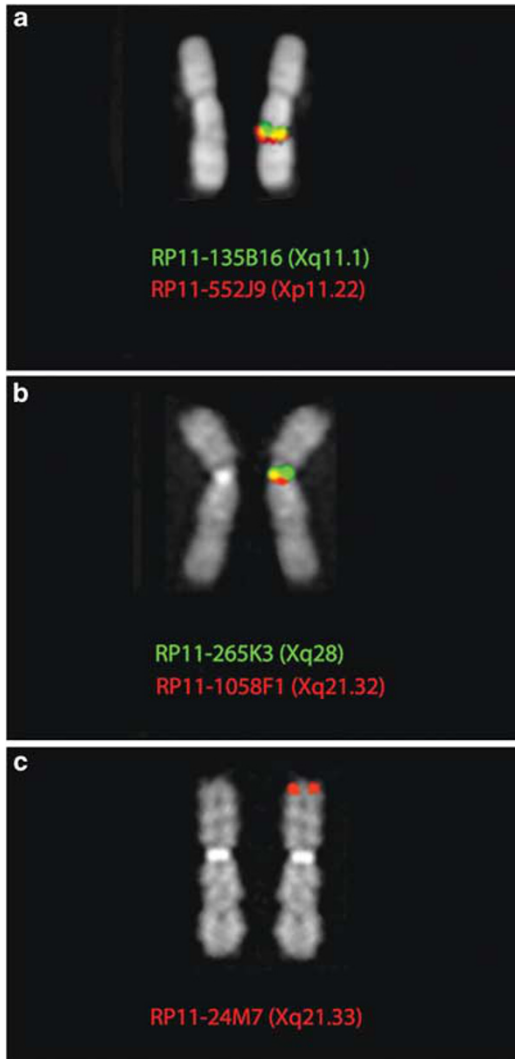
Human BAC	Acc. no.	hg18 position	Comment
RP11-24M7	BES	chrX:94,422,669-94,590,042	
RP11-39P2	BES	chrX:95,126,398-95,288,113	i
RP11-151D4	BES	chrX:96,818,325-96,964,772	n
RP11-138B3	BES	chrX:98,535,557-98,698,689	v
RP11-60B3	BES	chrX:99,388,268-99,563,848	e
RP11-1001A17	BES	chrX:101,042,960-101,235,524	r
RP11-426L6	BES	chrX:104,931,061-105,086,487	s
RP11-535K18	BES	chrX:135,051,131-135,233,538	i
RP11-1007I13	BES	chrX:150,877,314-151,051,253	o
RP11-207O16	BES	chrX:153,896,624-154,041,384	n
RP11-265K3	BES	chrX:154,603,527-154,763,828	
<i>SSC repositioned centromere</i>			
RP11-97P6	BES	chrX:93,899,349-94,056,485	No signal
RP11-32N6	BES	chrX:93,840,285-94,033,640	No signal
RP11-781G2	BES	chrX:93,294,116-93,485,026	
RP11-1058F1	BES	chrX:92,821,042-93,030,682	
RP11-483J19	BES	chrX:92,542,566-92,694,921	
RP11-776O14	BES	chrX:77,129,191-77,300,714	
RP11-157K21	BES	chrX:62,971,713-63,138,459	
RP11-135B16	BES	chrX:62,460,317-62,628,230	
<i>Human centromere</i>			
RP11-21E13	BES	chrX:57,294,374-57,500,705	
RP11-552J9	AL450023	chrX:52,617,556-52,803,087	
RP11-64P15	BES	chrX:33,526,698-33,692,695	
RP11-450P7	AL772392	chrX:21,533,785-21,657,970	
RP11-458E23	BES	chrX:10,225,422-10,401,851	

Abbreviations: BAC, bacterial artificial chromosome; BES, BAC end sequencing; FISH, fluorescence in situ hybridization; SSC, *Saimiri sciureus*.

The mapping of many of them was determined by the BAC end sequencing (BES), as reported by the UCSC genome browser (<http://genome.ucsc.edu>). BAC order, from top to bottom, reported in the table corresponds to the marker order, assessed by FISH, from SSCXpter to SSCXqter. Examples of the FISH experiments are reported in Figure 2.

from 5 acrocentric and 16 submetacentrics to 7 acrocentric and 14 submetacentric chromosomes (see Supplementary Information for a summary of taxonomy and karyotypes). In this paper given the confusing array of numbering systems, we prefer to follow the chromosome nomenclature adapted by Stanyon *et al.* (2000) and Dumas *et al.* (2007). Given that different chromosomes varied according to taxonomic designation and geographic distribution, we also wanted to test if the *Saimiri* neoX chromosomes might follow the same distinctions or was perhaps even polymorphic as in the orangutan. We also hypothesized that if the neoX was found in some squirrel monkey taxa and not others it might help date the origin of the ENC.

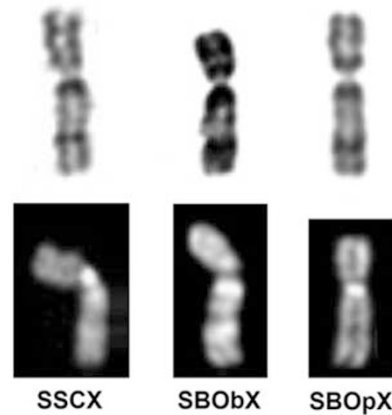
The most recent biomolecular studies generally identified four distinct clades *S. oerstedii*, *S. sciureus*, *S. boliviensis* and *S. ustus* (Lavergne *et al.*, 2010; Perelman *et al.*, 2011). In studies of both mtDNA (Chiou *et al.* 2011) and nuclear DNA (Perelman *et al.*, 2011), a sister relationship between *S. boliviensis* and other *Saimiri* taxa was found. Either *S.s. macrodon* (Chiou *et al.*, 2011) or *S. ustus* (Perelman *et al.*, 2011) was proposed as the sister lineage to *S. oerstedii/S.s. sciureus* (Chiou *et al.*, 2011). These studies found a very recent divergence of extant squirrel monkey species. *S. boliviensis* apparently diverged between 1.5 and 2.2 MYA, followed by a subsequent radiation of the other taxa between 0.7 and 1.2 MYA (Chiou *et al.*, 2011; Perelman *et al.*, 2011).



**Figure 2** Examples of FISH experiments, using human BAC clones (see Table 2), on squirrel monkey X chromosome, showing the position (a) of the human centromere and (b) of the squirrel monkey centromere. (c) shows a human BAC mapping, in humans, at Xq21.33, which, following the inversion, became telomeric. For detail, see text.

The repositioned centromere on chromosome X found in *Saimiri* is certainly present in *S. sciureus*, *S. boliviensis boliviensis* and in *S. boliviensis peruviansis* (Figure 3). A review of the literature shows that, when the X chromosome is illustrated with sufficient banding clarity, the repositioned centromere is evident in all squirrel monkeys regardless of the taxonomic designation (Jones and Ma, 1975; Lau and Arrighi, 1976; Cambefort and Moro, 1978; Garcia *et al.*, 1979, 1995; Dutrillaux and Couturier, 1981; Schempp *et al.*, 1989; Scammell *et al.*, 2001; Stanyon *et al.*, 2008). The seemingly anomalous q terminal position of the par of the two *Saimiri* in Schempp *et al.* (1989) is now easily explained by the presence of the neoX.

It is also noteworthy that no other neoX was found in any other new world primate, and in particular Cebinae, the sister group to *Saimiri*. Our conclusion is that the ENC in the X chromosome of *Saimiri* evolved in the common ancestor of all squirrel monkeys. The date for *Cebus/Saimiri* divergence was recently calculated at about 15 MYA (Perelman *et al.*, 2011), therefore the origin of the ENC has to be somewhere between 15 and 1.5 MYA.



**Figure 3** 4,6-Diamidino-2-phenyl indole (DAPI) banding (below), and trypsin-Giemsa banding (above) of chromosome X from squirrel monkeys *Saimiri sciureus* (SSCX), *S. boliviensis boliviensis* (SBObX) and *S. b. peruviansis* (SBPpX). The banding pattern appears identical, strongly indicating that both SBO share the same variant X with SSC (Figure 1).

There is some cytogenetic evidence that would favor that the centromere is relatively old. In Figures 2a and c the 4,6-diamidino-2-phenyl indole staining displays a consistent block of centromeric heterochromatin in SSCX-repositioned centromere. Additionally, two human BACs mapping in the region, where the novel centromere was seeded, failed to yield any FISH signals, indicating that the pericentromeric region was deeply restructured after the centromere repositioning event, further indicating that a long time elapsed since the neocentromere seeding (see below). Normally, only mature centromeres have these features. If the centromere is old it is also less likely that it is polymorphic in any *Saimiri* species. However, only additional research will conclusively answer these questions.

#### ENC EVOLUTIONARY MODIFICATIONS AFTER SEEDING

Mature eukaryotic centromeres, including ENC, are composed of arrays of satellite DNA frequently surrounded by clusters of SD (She *et al.*, 2004). ENCs, as illustrated above, reasonably emerge in anonymous sequences, and do not immediately affect the sequence itself. ENC fixation in the population is accompanied by the acquisition of species-specific arrays of centromeric satellite DNA as well as clusters of pericentromeric segmental duplications. FISH, using specific probes can easily test for the presence of satellite DNA. If unavailable, the total DNA of the species under study can be hybridized at very high stringency. The characterization of SDs around specific centromeres is definitely more complex. Detailed data on pericentromeric SDs are essentially limited to human and mouse, because all other genomes were sequenced using the shotgun approach. The methodology designed by Bailey *et al.* (2002) can efficiently detect duplicated sequences for whole-genome shotgun sequence, by calculating the relative depth of coverage in the raw shotgun sequence read pool, but it is not able to map them. The analysis of pericentromeric SDs in the six human ENCs (chromosomes 3, 6, 11, 14, 15, and 21) revealed that two of them (3 and 6) are among the most poor in SDs (She *et al.*, 2004). However, it has to be considered that SDs were already present in the seeding region of some of the other human ENCs at the time of the ENC emergence (see Cardone *et al.*, 2007). The only non-human pericentromeric region of an ENC examined in detail is that of macaque chromosome

4 (human 6q24.3) (Ventura *et al.*, 2007). Comparison of the sequence of the human 6q24.3 region to many other mammalian species indicated that, very likely, the seeding domain was devoid of satellite DNA and SDs. Following the ENC seeding, a 250-kb segment was extensively and imperfectly duplicated around the novel centromere. These duplications were strictly intrachromosomal. Interestingly, the two youngest ENCs yet studied, horse chromosome 11 and orangutan chromosome 9, are apparently devoid of satellite DNA. Therefore, the process leading the ENC toward the complexity of a normal centromere appears to be relatively slow, and forces acting to keep the region unaltered can oppose their restructuring. The presence of genes can be one of these forces. ENC maturation (and centromere deactivation) is still not clear. Perhaps in the future orangutan chromosome 9 might eventually provide some clues, once the neocentromere region is sequenced.

Some papers have examined the expression of genes embedded in neocentromeric regions (Saffery *et al.*, 2003; Nagaki *et al.*, 2004; Lam *et al.*, 2006; Yan *et al.*, 2006). Their conclusion was that neocentromeres do not affect gene expression *per se*. However, the accumulation of satellite DNA and the potential restructuring of the pericentromeric regions can negatively affect gene structure and, consequently, expression. The absence of (important) genes around the pericentromeric regions can therefore be seen as a condition favoring, or at least not opposing, ENC fixation. Lomiento *et al.* (2008) have found, in primates, that fixed ENCs were preferentially seeded in gene-deserts. Alternatively, those close to genes can be supposed to have remained poor in segmental duplication as in the case of human ENCs 3 and 6 (She *et al.*, 2004).

### TELOMERE/CENTROMERE INTERCHANGE

Ventura *et al.* (2004) have reported on the evolutionary history of chromosome 3 in primates. The ancestral chromosome 3 split, in the new world monkeys' ancestor, into three distinct acrocentric chromosomes. Marker order analysis confirmed that synteny and marker order was conserved in the three Platyrrhine families Cebidae, (*Callithrix jacchus*), Atelidae (*Lagothrix lagotricha*) and Pitheciidae (*Callicebus pallescens*). Strikingly, at least three centromere/telomere interchanges have occurred. The centromere position moved from one telomere to the other depending on the species examined. Subtelomeric repetitive sequences and/or SD could have a role in these exchanges. We can also note that Villasante *et al.* (2007) have proposed that, during the evolution of eukaryotic chromosomes, the centromeres were derived from telomeres.

### CONCLUDING REMARKS: CENTROMERES AND GENOME SEQUENCING

Over the last two decades sequencing technology had experienced quantum leaps. The 'parallel sequencing' era, initiated in 2005 (Margulies *et al.*, 2005), has allowed the sequencing of many human individuals (see the '1000 genomes' project; <http://www.1000genomes.org>). Concomitantly, the sequencing of entire non-human genomes has progressed exponentially (see <http://www.genome.gov/10002154>), and the sequencing of 10 K vertebrate genome has been proposed (Genome\_10K\_community\_of\_scientists, 2009). The giant panda was the first mammalian genome to be fully sequenced by parallel sequencing (Li *et al.*, 2010). However, this achievement also points to the weakness of these technologies in reliably assembling sequences into chromosomes. The panda sequence, in fact, is a collection of scaffolds, and, consequently, the position of the centromere along the chromosome was not considered at all. Additionally, the satellite DNA specific for the centromeres of the species under

study could be unknown or could be present in non-centromeric regions (see the stretches of alphoid sequences present, in humans, at 2q21.2; chromosome 2: 132 682 845–132 722 540; UCSC hg18). Furthermore, centromere-repositioning events can only be identified in evolutionary studies that compare a phylogenetic array of species to distinguish between the ancestral and derivative position of a specific centromere. Therefore, we can anticipate that classical and molecular cytogenetics will continue to have a crucial role in the identification of centromere movements, even in the era of massive genome sequencing. Indeed, all ENCs and HNs were found following classical and molecular cytogenetic investigations.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGEMENTS

This project was funded by MUR (Ministero della Universita' e della Ricerca). WS was funded by the DFG (SCHE 214/8).

- Alonso A, Mahmood R, Li S, Cheung F, Yoda K, Warburton PE (2003). Genomic microarray analysis reveals distinct locations for the CENP-A binding domains in three human chromosome 13q32 neocentromeres. *Hum Mol Genet* **12**: 2711–2721.
- Alonso A, Fritz B, Hasson D, Abrusan G, Cheung F, Yoda K *et al.* (2007). Co-localization of CENP-C and CENP-H to discontinuous domains of CENP-A chromatin at human neocentromeres. *Genome Biol* **8**: R148.
- Ayres J (1985). On a new species of squirrel monkey, genus *Saimiri*, from Brazilian Amazonia (Primates, Cebidae). *Papéis Avulsos Zool São Paulo* **36**: 147–164.
- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S *et al.* (2002). Recent segmental duplications in the human genome. *Science* **297**: 1003–1007.
- Bhasim MK (2007). Human population cytogenetics: a review. *Anthropologist* **3** (Special Volume): 435–506.
- Bukvic N, Susca F, Gentile M, Tangari E, Ianniruberto A, Guanti G (1996). An unusual dicentric Y chromosome with a functional centromere with no detectable alpha-satellite. *Hum Genet* **97**: 453–456.
- Cambefort Y, Moro F (1978). Cytogenetics and taxonomy of some south Bolivian monkeys. *Folia Primatol Int J Primatol* **29**: 307.
- Capozzi O, Purgato S, Verdun di Cantogno L, Grosso E, Ciccone R, Zuffardi O *et al.* (2008). Evolutionary and clinical neocentromeres: two faces of the same coin? *Chromosoma* **117**: 339–344.
- Capozzi O, Purgato S, D'Addabbo P, Archidiacono N, Battaglia P, Baroncini A *et al.* (2009). Evolutionary descent of a human chromosome 6 neocentromere: a jump back to 17 million years ago. *Genome Res* **19**: 778–784.
- Carbone L, Nergadze SG, Magnani E, Misceo D, Francesca Cardone M, Roberto R *et al.* (2006). Evolutionary movement of centromeres in horse, donkey, and zebra. *Genomics* **87**: 777–782.
- Carbone L, D'Addabbo P, Cardone MF, Teti MG, Misceo D, Vessere GM *et al.* (2009). A satellite-like sequence, representing a 'clone gap' in the human genome, was likely involved in the seeding of a novel centromere in macaque. *Chromosoma* **118**: 269–277.
- Cardone MF, Alonso A, Paziienza M, Ventura M, Montemurro G, Carbone L *et al.* (2006). Independent centromere formation in a capricious, gene-free domain of chromosome 13q21 in old world monkeys and pigs. *Genome Biol* **7**: R91.
- Cardone MF, Lomiento M, Teti MG, Misceo D, Roberto R, Capozzi O *et al.* (2007). Evolutionary history of chromosome 11 featuring four distinct centromere repositioning events in Catarrhini. *Genomics* **90**: 35–43.
- Chiou KL, Pozzi L, Lynch Alfaro JW, Di Fiore A (2011). Pleistocene diversification of living squirrel monkeys (*Saimiri* spp.) inferred from complete mitochondrial genome sequences. *Mol Phylogenet Evol* **59**: 736–745.
- Chowdhary BP, Raudsepp T, Fronicke L, Scherthan H (1998). Emerging patterns of comparative genome organization in some mammalian species as revealed by Zoo-FISH. *Genome Res* **8**: 577–589.
- Clemente IC, Ponsa M, Garcia M, Egozcue J (1990). Evolution of the Simiiformes and the phylogeny of human chromosomes. *Hum Genet* **84**: 493–506.
- Costello RK, Dickinson C, Rosenberg AL, Boinski S, Szalay F (1993). Squirrel monkey (genus *Saimiri*) taxonomy: a multidisciplinary study of the biology of species. In: Kimbel WH, Martin LB (eds). *Species, Species Concepts, and Primate Evolution*. Plenum Press: New York. pp 177–210.
- de Boer LEM, Seuanez H (1982). The chromosomes of the orang utan and their relevance to the conservation of the species. In: de Boer LEM (ed.). *The Orangutan. Its Biology and Conservation* vol. 1. W Junk: The Hague, Netherlands. pp. 135–170.
- Dumas F, Stanyon R, Sineo L, Stone G, Bigoni F (2007). Phylogenomics of species from four genera of new world monkeys by flow sorting and reciprocal chromosome painting. *BMC Evol Biol* **7**(Suppl 2): S11.



- Dutrillaux B (1979). Chromosomal evolution in primates: tentative phylogeny from *Microcebus murinus* (Prosimian) to man. *Hum Genet* **48**: 251–314.
- Dutrillaux B, Couturier J (1981). The ancestral karyotype of platyrrhine monkeys. *Cytogenet Cell Genet* **30**: 232–242.
- Eder V, Ventura M, Ianigro M, Teti M, Rocchi M, Archidiacono N (2003). Chromosome 6 phylogeny in primates and centromere repositioning. *Mol Biol Evol* **20**: 1506–1512.
- Ferreri GC, Liscinsky DM, Mack JA, Eldridge MD, O'Neill RJ (2005). Retention of latent centromeres in the mammalian genome. *J Hered* **96**: 217–224.
- Garcia M, Miro R, Ponsa M, Egozcue J (1979). Chromosomal polymorphism and somatic segregation in *Saimiri sciureus*. *Folia primatologica; international journal of primatology* **31**: 312–323.
- Garcia M, Borrell A, Mudry M, Egozcue J, Pons M (1995). Prometaphase karyotype and restriction-enzyme banding in squirrel monkeys, *Saimiri boliviensis boliviensis* (Primates: Platyrrhini). *J Mammal* **76**: 497–503.
- Genome\_10k\_community\_of\_scientists (2009). Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. *J Hered* **100**: 659–674.
- Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER *et al.* (2007). Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**: 222–234.
- Goossens B, Chikhi L, Fairus Jalil M, James SA, Ancrenaz M, Lackman-Ancrenaz I *et al.* (2008). *Orangutans: Geographic Variation in Behavioral Ecology and Conservation. Chapter 1. Taxonomy, Geographic Variation and Population Genetics of Bornean and Sumatran Orangutans*. Oxford University Press: Oxford.
- Groves C (2001). Why taxonomic stability is a bad idea, or why are there so few species of primates (or are there?). *Evol Anthropol* **10**: 192–198.
- Hasson D, Alonso A, Cheung F, Tepperberg J, Papenhausen P, Engelen JJ *et al.* (2011). Formation of novel CENP-A domains on tandem repetitive DNA and across chromosome breakpoints on human chromosome 8q21 neocentromeres. *Chromosoma* in press.
- Hershkovitz P (1982). Supposed squirrel monkey affinities of *Dolichocebus gaimanensis*. *Nature* **298**: 202.
- Jauch A, Wienberg J, Stanyon R, Arnold N, Tofaneli S, Ishida T *et al.* (1992). Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. *Proc Natl Acad Sci USA* **89**: 8611–8615.
- Jones TC, Ma NS (1975). Cytogenetics of the squirrel monkey (*Saimiri sciureus*). *Fed Proc* **34**: 1646–1650.
- Kobayashi T, Yamada F, Hashimoto T, Abe S, Matsuda Y, Kuroiwa A (2008). Centromere repositioning in the X chromosome of XO/XO mammals, Ryukyu spiny rat. *Chromosome Res* **16**: 587–593.
- Lam AL, Boivin CD, Bonney CF, Rudd MK, Sullivan BA (2006). Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. *Proc Natl Acad Sci USA* **103**: 4186–4191.
- Lau YF, Arrighi F (1976). Studies of the squirrel monkey, *Saimiri sciureus*, genome. I. Cytological characterizations of chromosomal heterozygosity. *Cytogenet Genome Res* **17**: 51–60.
- Lavergne A, Ruiz-Garcia M, Catzeflis F, Lacote S, Contamin H, Mercereau-Puijalon O *et al.* (2010). Phylogeny and phylogeography of squirrel monkeys (genus *Saimiri*) based on cytochrome b genetic analysis. *Am J Primatol* **72**: 242–253.
- Li R, Fan W, Tian G, Zhu H, He L, Cai J *et al.* (2010). The sequence and *de novo* assembly of the giant panda genome. *Nature* **463**: 311–317.
- Liehr T, Kosyakova N, Weise A, Ziegler M, Raabe-Meyer G (2010). First case of a neocentromere formation in an otherwise normal chromosome 7. *Cytogenet Genome Res* **128**: 189–191.
- Lo AW, Craig JM, Saffery R, Kalitsis P, Irvine DV, Earle E *et al.* (2001a). A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. *EMBO J* **20**: 2087–2096.
- Lo AW, Magliano DJ, Sibson MC, Kalitsis P, Craig JM, Choo KH (2001b). A novel chromatin immunoprecipitation and array (cia) analysis identifies a 460-kb cenp-a-binding neocentromere DNA. *Genome Res* **11**: 448–457.
- Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM *et al.* (2011). Comparative and demographic analysis of orangutan genomes. *Nature* **469**: 529–533.
- Lomiento M, Jiang Z, D'Addabbo P, Eichler EE, Rocchi M (2008). Evolutionary-new centromeres preferentially emerge within gene deserts. *Genome Biol (www)* **9**: R173.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA *et al.* (2005). Genome sequencing in microfabricated high-density picoliter reactors. *Nature* **437**: 376–380.
- Marshall OJ, Chueh AC, Wong LH, Choo KH (2008). Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet* **82**: 261–282.
- Misceo D, Cardone MF, Carbone L, D'Addabbo P, de Jong PJ, Rocchi M *et al.* (2005). Evolutionary history of chromosome 20. *Mol Biol Evol* **22**: 360–366.
- Misceo D, Capozzi O, Roberto R, Dell'Oglio MP, Rocchi M, Stanyon R *et al.* (2008). Tracking the complex flow of chromosome rearrangements from the Hominoidea ancestor to extant Hylobates and *Nomascus* gibbons by high-resolution synteny mapping. *Genome Res* **18**: 1530–1537.
- Montefalcone G, Tempesta S, Rocchi M, Archidiacono N (1999). Centromere repositioning. *Genome Res* **9**: 1184–1188.
- Moore CM, Harris CP, Abee CR (1990). Distribution of chromosomal polymorphisms in three subspecies of squirrel monkeys (genus *Saimiri*). *Cytogenet Cell Genet* **53**: 118–122.
- Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM *et al.* (2004). Sequencing of a rice centromere uncovers active genes. *Nat Genet* **36**: 138–145.
- Oakenfull E, Lim H, Ryder O (2000). A survey of equid mitochondrial DNA: implications for the evolution, genetic diversity and conservation of *Equus*. *Conservation Genet* **1**: 341–355.
- Oakenfull EA, Clegg JB (1998). Phylogenetic relationships within the genus *Equus* and the evolution of alpha and theta globin genes. *J Mol Evol* **47**: 772–783.
- Perelman P, Johnson WE, Roos C, Seuanez HN, Horvath JE, Moreira MAM *et al.* (2011). A molecular phylogeny of living primates. *PLoS Genet* **7**: e1001342.
- Piras FM, Nergadze SG, Poletto V, Cerutti F, Ryder OA, Leeb T *et al.* (2009). Phylogeny of horse chromosome 5q in the genus *Equus* and centromere repositioning. *Cytogenet Genome Res* **126**: 165–172.
- Roberto R, Capozzi O, Wilson RK, Mardis ER, Lomiento M, Tuzun E *et al.* (2007). Molecular refinement of gibbon genome rearrangement. *Genome Res* **17**: 249–257.
- Roizes G (2006). Human centromeric alphoid domains are periodically homogenized so that they vary substantially between homologues. Mechanism and implications for centromere functioning. *Nucleic Acids Res* **34**: 1912–1924.
- Ryder OA, Chemicnick LG (1993). Chromosomal and mitochondrial DNA variation in orangutans. *J Hered* **84**: 405–409.
- Saffery R, Sumer H, Hassan S, Wong LH, Craig JM, Todokoro K *et al.* (2003). Transcription within a functional human centromere. *Mol Cell* **12**: 509–516.
- Scammell JG, Wright JL, Tuck-Muller CM (2001). The origin of four squirrel monkey cell lines established by karyotype analysis. *Cytogenet Cell Genet* **93**: 263–264.
- Schempp W, Weber B, Muller G (1989). Mammalian sex-chromosome evolution: a conserved homoeologous segment on the X and Y chromosomes in primates. *Cytogenet Cell Genet* **50**: 201–205.
- She X, Horvath JE, Jiang Z, Liu G, Furey TS, Christ L *et al.* (2004). The structure and evolution of centromeric transition regions within the human genome. *Nature* **430**: 857–864.
- Stanyon R, Consigliere S, Muller S, Morescalchi A, Neusser M, Wienberg J (2000). Fluorescence *in situ* hybridization (FISH) maps chromosomal homologies between the dusky titi and squirrel monkey. *Am J Primatol* **50**: 95–107.
- Stanyon R, Rocchi M, Capozzi O, Roberto R, Miscio D, Ventura M *et al.* (2008). Primate chromosome evolution: ancestral karyotypes, marker order and neocentromeres. *Chromosome Res* **16**: 17–39.
- Steiper ME (2006). Population history, biogeography, and taxonomy of orangutans (genus: *Pongo*) based on a population genetic meta-analysis of multiple loci. *J Hum Evol* **50**: 509–522.
- Trifonov VA, Stanyon R, Nesterenko AI, Beiyuan FB, Perelman PL, O'Brien PCM *et al.* (2008). Multi-directional cross-species painting illuminates the history of karyotypic evolution in Perissodactyla. *Chromosome Res* **16**: 89–107.
- Turleau C, de Grouchy J, Chavin-Colin C (1975). Pericentric inversion of no. 3, homozygous and heterozygous, and centromeric transposition of no. 12 in a family of orangutans. Implications for evolution. *Ann Genet* **18**: 227–233.
- Ventura M, Archidiacono N, Rocchi M (2001). Centromere emergence in evolution. *Genome Res* **11**: 595–599.
- Ventura M, Mudge JM, Palumbo V, Burn S, Blennow E, Pierluigi M *et al.* (2003). Neocentromeres in 15q24-26 map to duplicons which flanked an ancestral centromere in 15q25. *Genome Res* **13**: 2059–2068.
- Ventura M, Weigl S, Carbone L, Cardone MF, Miscio D, Teti M *et al.* (2004). Recurrent sites for new centromere seeding. *Genome Res* **14**: 1696–1703.
- Ventura M, Antonacci F, Cardone MF, Stanyon R, D'Addabbo P, Cellamare A *et al.* (2007). Evolutionary formation of new centromeres in macaque. *Science* **316**: 243–246.
- Villasante A, Abad JP, Mendez-Lago M (2007). Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. *Proc Natl Acad Sci USA* **104**: 10542–10547.
- Voullaire L, Saffery R, Davies J, Earle E, Kalitsis P, Slater H *et al.* (1999). Trisomy 20p resulting from inverted duplication and neocentromere formation. *Am J Med Genet* **85**: 403–408.
- Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F *et al.* (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326**: 865–867.
- Warburton PE, Cooke CA, Bourassa S, Vafa O, Sullivan BA, Stetten G *et al.* (1997). Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Curr Biol* **7**: 901–904.
- Warburton PE, Dolled M, Mahmood R, Alonso A, Li S, Naritomi K *et al.* (2000). Molecular cytogenetic analysis of eight inversion duplications of human chromosome 13q that each contain a neocentromere. *Am J Hum Genet* **66**: 1794–1806.
- Wienberg J, Jauch A, Stanyon R, Cremer T (1990). Molecular cytogenetics of primates by chromosomal *in situ* suppression hybridization. *Genomics* **8**: 347–350.
- Yan H, Ito H, Nobuta K, Ouyang S, Jin W, Tian S *et al.* (2006). Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of a complex centromere. *Plant Cell* **18**: 2123–2133.
- Zeitlin SG, Baker NM, Chapados BR, Soutoglou E, Wang JY, Berns MW *et al.* (2009). Double-strand DNA breaks recruit the centromeric histone CENP-A. *Proc Natl Acad Sci USA* **106**: 15762–15767.
- Zhi L, Karesh WB, Janczewski DN, Frazier-Taylor H, Sajuthi D, Gombek F *et al.* (1996). Genomic differentiation among natural populations of orangutan (*Pongo pygmaeus*). *Curr Biol* **6**: 1326–1336.