# Great ape genomics

## Diversity and evolution

Javier Prado Martínez

TESI DOCTORAL UPF / 2014

DIRECTOR DE LA TESI

Dr. Tomàs Marquès i Bonet,

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

upf. Universitat Pompeu Fabra Barcelona

A mi familia

*"With man gone, will there be hope for gorilla?"*
*"With gorilla gone, will there be hope for man?"*

*Ishmael* by Daniel Quinn

# Acknowledgments

Malgrat el dia a dia es fa palès que ens fem grans i que cada cop pesen més els anys, anem acumulant experiències, amics, viatges, lesions, malalties i algunes persones que estimem se'n van, d'altres continuen i moltes altres es fan a un costat a continuar un camí diferent del nostre. Totes aquestes persones i experiències ens emplenen i ens fan ser com som, millors o pitjors, però com som. També les persones que ens ho han fet passar malament són molt importants ja que aprenem les millors lliçons d'aquestes experiències malgrat que preferim recordar els bons moments amb les persones que estimem. Aquesta secció de la tesi és una oportunitat única per tal de fer repàs dels que estimem i agrair-vos tot el que m'heu aportat. També és un punt i a part en la vida i un motiu per tal de fer aquesta memòria que d'altra banda costa fer el dia a dia.

Després d'aquest rotllo i preveient que això es pot fer molt llarg intentaré ser breu, però no prometo res. No he obtingut cap manera lògica d'ordenar totes les seccions d'aquest apartat i sens dubte és de les parts més complicades de la tesi i de les poques que la gent pot arribar a llegint. Disculpar-me amb tots els que no nombraré en els següents apartats però espero que el primer paràgraf serveixi per agrair-vos tot el que m'heu aportat, tot i que ni us imagineu que pugueu influir en algú, segur que d'alguna manera o altra ho esteu fent.

Empezaré agradeciendo a mi familia sin los que no hubiera conseguido nada de lo que me he propuesto. A mis padres, Estrella y Daniel, que siempre me han apoyado en todo y me han permitido hacer siempre lo que yo he elegido. Estoy haciendo lo que me entusiasma gracias a ellos y a todo el esfuerzo de su trabajo. Parte de esta tesis está escrita en Galicia, tierra de mi padre que tan importante ha sido para mi. Gracias por todo lo que me habéis enseñado y entregado, gracias. A mis abuelos, yaya Manuela que me crío y

Visitación y Antonio, de los que tantos buenos recuerdos guardo y que tanto se han desvivido por darnos lo mejor. Las vacaciones que pasé con todos ellos fueron geniales y las comilonas de Navidad espectaculares, gracias Visi por las croquetas y demás manjares deliciosos y Antonio por tus sabios consejos. Muchas gracias por todo y por entender que no os he podido dedicar todo el tiempo que hubiera querido estos últimos años, ahora desde el exilio me esforzaré por dedicaros todo el tiempo que os merecéis, que es mucho. A mi hermano Sergio, que ha estado ahí y siempre ha sido una gran influencia para mi a la que siempre he querido imitar. Y el resto de primos y tíos, con los que he compartido tantas vacaciones y buenos momentos: Mariluz, Óscar, Mikel, Miguel y Ana. También agradecer a las familias de Arnuide y Cuenca todo el cariño que siempre nos han dado, especialmente la primera que tanto nos arropó en los momentos difíciles. Muchas gracias.

En el pla professional m'agradaria començar per remarcar els bons moments que m'han fet passar el clan *cenitas*, Jéssica i Irene. Suposo que hauria d'obviar incloure paraules del *jessidioma* però els jessayunos, dinars, jesscansos i jesspedides conjuntament amb els soparets (insuficients) que hem fet al llarg d'aquests anys han fet molt més fàcil el doctorat. Sense vosaltres aquests anys no hauria estat el mateix. Agrair també a tot el grup dels primats tot el que m'han ensenyat i el que he aprés de tots, des del primer fins l'últim, gràcies: Belén, Marcos, Marta, Marc, Guillem, Xavi, Tiago, Raquel, Lobón, Marc, Lukas i Clàudia; i la resta de visitants: Meritxell, Serena, Ann, Peter… Grup que va creixent i que estic convençut que farà grans coses.

Vull fer menció especial a la gent amb qui vaig començar aquesta aventura ara fa més de cinc anys, alguns dels quals encara compartim dinars, birres o viatges. Els avis Ramón i Jose, les Martes i el Gabriel. Des que vam començar l'aventura de doctorat hem seguit les nostres rutines amb els avis a les quals s'hi han afegit d'altres peces com

l'Ignasi i Guillem que han fet dels cafès unes experiències estranyes que tant trobaré a faltar ara que marxo. Moltes gràcies per haver entès la meva absència durant els dos primers anys i que tant poc temps us vaig dedicar. Agrair especialment al Gabriel tot el coneixement que hem compartit i de bon segur que les converses més productives de tot el doctorat van ser amb ell. Gràcies avis per tot el temps que hem compartit, tots els riures i tota la saviesa que de la vostra experiència, sens dubte he aprés moltíssim molt de vosaltres. I a la Marta amb la que hem compartit tantes aficions aquests anys i que tant bons moments hem passat.

Seguiré per bioevo, aquí segur que em deixaré a més de un i de dos, però doneu-vos per al·ludits tots, ja que aquest institut fa que cada dia sigui gran i que no faci mandra venir a *treballar*. Començo pels companys de lab, Diego i Oriol. No puc emfatitzar quant de temps ens hem perdut junts veritat Diego? Sens dubte el millor temps perdut de tot el doctorat l'he passat amb tu parlant de les nostres tonteries, d'esport i de tot una mica, gràcies tiu, ets molt gran. Se'm fa complicat fer una llista de bioevo perquè sou molts amb els que he passat bons moments, però, Gràcies: Neus, Marco, Pierre, Núria, Juan, Nino, Óscar, Mònica, Laura, Ixa, Urko, Iñigo, Lara, Marina, Joha, Fede, Marcs, Txema, Juanma, Jordi, Max, Alicia, Brandon, Haber, Elena, Gio, Maria... Sé que me'n deixo molts però doneu-vos per mencionats. Els dinars lingüístics amb la Ludo els trobaré molt a faltar, llàstima que no hem sigut gaire constants últimament. Ángel gracias por tu tiempo i tu sabiduría, eres un crack. Menció especial vull donar a com aquest institut fa pinya amb el futbol, voleï, l'equip de bàsquet amb Pierre, Diego, Ignasi i Carlos, que tant trobava a faltar durant el doctorat i les farres que hem compartit, algunes d'elles mítiques... Gràcies Bioevo! També donar les gràcies a tots els PIs que han fet que la meva tasca sigui més senzilla des de les diferents àrees de coneixement i que tant m'han ajudat a resoldre els meus dubtes.

I would also like to thank the Eichler lab for his kindness during my short stays in Seattle where I have probably been the most productive during this thesis. Thank you Fereydoun, you are a great person and scientist, you'll do great and you were very important during my stay there. Can, despite your kind rudeness you are still a great guy. Peter, we shared many experiences and I learned many things from you, thank you for all the time you spent with me. Michael, we still have some pending trips to do. And the rest, Laura, Tonia, Maika, Megan, Xander, Emre, Katy, Santosh… You all were very kind to me, thank you. I would like to thank especially Evan, the most outstanding scientist I have met so far and who put so much effort on the great ape project. To end this Seattle chapter, I would like to thank my adoptive family during the stay, Kevin, Mackenzie, Stephanie and Nabil, thanks! Also to mention my short stay in CNAG where I met so many people who helped me: Leonor, Emanuele, Jordis, Santiago, Paolo…

Al pla més personal vull agrair tot el recolzament que m'han donat tots els meus amics que m'acompanyen des de la infància i en major part des de l'època universitària. Gràcies per entendre que no us podia dedicar tot el temps que m'hagués agradat però que vaig intentar compensar els dos últims anys. Gràcies per les nostres quedades "improvisades". Manel, company de birres i de penes en tants moments, alguns d'ell dolents però que al final hem tirat endavant, et mereixes tot el bo que et passarà, ets un gran. Ignasi, tu també ets un gran i ho saps, hem compartit moltíssimes coses els últims anys i et trobaré a faltar. Miki, company d'aventures durant tants anys i ens separarem aviat per fer grans coses veritat? Trobaré a faltar les teves coses a Anglaterra. Abel, el xerraire del grup, sempre explicant les seves històries. L'Euli i tot el que hem compartit. La Roser i les nostres quedades. I tota la resta: Pau i Adriana, Adri, Marc, Miriam, Aníbal, Núries, Clàudia, Anna i Ari. Gràcies nois.

També vull agrair els bons moment que he passat amb amics més antics amb els que he reprès el contacte gràcies al Carles. Gràcies a vosaltres ha augmentat la freqüència de les meves visites a Puigcerdà. Els sopars i els pàdels han estat fantàstics i l'Interrail va ser genial. Gràcies Noe, Marc i Cris, David, Álvaro, Vinyes i Núria. Capítol especial li dono a tota la colla del meu germà amb els que tant bé m'ho he passat: Guillem, Uri, Jaume, Elena, Lluïsa, Edu... I als companys de pis amb els que he compartit tant de temps, Vanesa, Ari, Carlos, Jorge i Aníbal.

En un pla més impersonal m'agradaria agrair els col·laboradors que tant m'han ajudat en els meus projectes i dels que tant he aprés. La quantitat de col·laboracions que he gaudit durant aquests anys han estat molt enriquidores i sens dubte una gran experiència i entrenament les 'conference calls' setmanals. Podria afegir tot el que m'han aportat el futbol sala, tennis, bàsquet i d'altres esports, vitals en la vida sedentària de la ciència.

Per últim voldria agrair al meu director de tesi tot el temps i esforç que ha invertit en mi. Quan vaig començar aquesta aventura fa més de 4 anys gràcies a l'Arcadi, no podia imaginar-me que podria assolir el que he aconseguit. Vull agrair tota la confiança que vas dipositar amb mi posant-me al càrrec de projectes tant ambiciosos que he liderat i col·laborat i els fruits d'aquests projectes han estat excel·lents. No ha estat un camí fàcil i han requerit molt d'esforç, temps i estrès durant gran part de la tesi, uns elements sinònim de doctorat. Gràcies Tomàs.

This research was supported by a PhD fellowship "Formación de Personal Investigador" BES-2010-032251 and part of the Proyecto Nacional BFU2009-13409-C02-02 funded by El Ministerio de Competividad y Economía de España. Agrair a l'Arcadi el seu suport en aquesta tesi i el seu funding sense el qual no la podria haver realitzat.

## Abstract

Great apes are our closest relatives and as such they are our best resource to understand our recent origins. Through comparative genomics we can fully investigate this question, but the lack of great ape genomes have precluded to have a complete view on this and many other questions related to the Hominidae family. In the context of the current sequencing revolution, herein I present the contributions I have made in the study of great ape genomes. Starting from studies studying single genomes and following with the analysis of diversity in multiple great ape genomes, I summarize the findings in the most complete dataset of great ape genomes, covering all great ape species and most subspecies, providing an unprecedented view on diversity, demography and population structure in great apes. I finally discuss the most relevant implications of this work and how this can boost the conservation efforts in the protection of great apes.

# Resum

Els grans simis són els nostres parents evolutius més propers i com a tals són la millor eina per entendre els nostres orígens més recents. Mitjançant la genòmica comparativa ara podem estudiar en profunditat aquesta qüestió, però la manca de genomes de grans simis no ha permès una anàlisi completa d'aquesta i d'altres qüestions relacionades amb la família Hominidae. En el context de la revolució de la seqüenciació, en aquesta tesi presento les meves contribucions en l'estudi de la genòmica de grans simis. Començant des de l'estudi de genomes individualment, faig un resum de les troballes més importants del panell més complet de genomes de grans simis, on vam incloure totes les espècies i la majoria de subespècies d'aquesta família i vam proporcionar una visió sense precedents en diversitat genètica, demografia i estructura de la població. També discuteixo les implicacions més rellevants d'aquest treball i com aquest pot ser una eina important en la conservació dels grans simis.

# Preface

The origin of the human lineage along with our evolutionary relatives has been a longstanding question in biology. Taxonomy combined with the fossil record posed the foundations in this field but not until the advent of molecular data the question was clarified.

The study of the genome through DNA sequencing has been the most important tool in the study of the Hominidae phylogeny. Improvements in the sequencing technology have allowed the emergence of the complete genome assemblies and the whole field of comparative genomics. Now, more than a decade after the initial sequencing of the human genome (Lander et al., 2001), we have been able to assemble all the great ape genera genomes (Marques-Bonet, Ryder, & Eichler, 2009), providing an unprecedented view of human origins.

The last ten years have experienced the most dramatic improvements in the acquisition of DNA sequences. High throughput DNA sequencing has allowed the acquisition of genomes at an extraordinary rate. In this context, the understanding of human variation and population history has been boosted with big consortia such as the 1000 genomes (Abecasis et al., 2012). This has been a milestone in evolutionary genetics but still, the lack of genomic data from our evolutionary relatives has precluded the study of many questions related to our closest relatives.

This revolution in technology for sequencing combined with the importance to study our closest relatives has been the premise on the work I will present in this thesis. The availability of sequencing has boosted many areas of knowledge, including the study of our closest relatives.

## Table of Contents

## Abbreviations

Mya: million years ago

SNP: Single Nucleotide Polymorphism

Indels: Insertions and Deletions

mtDNA: mitochondrial DNA

MHC: major histocompatibility complex

GAGP: Great ape genome project

ROH: Runs of homozygosity

ILS: Incomplete lineage sorting

BAC: Bacterial artificial chromosome

WGS: Whole genome shotgun

kbp: kilobase pair

Mbp: mega base pair

AIM: Ancestry Informative Marker

PSMC: Pairwise Sequentially Markovian Coalescent

ABC: Approximate Bayesian computation

CoalHMM: Coalescent Hidden Markov Model

# 1. INTRODUCTION

## 1.1. Hominidae family

### 1.1.1. Great apes

There are currently four different extant genera among Hominidae family: *Homo, Pan, Gorilla and Pongo*. The former three are African great apes (despite humans have colonized all the world) and the latter is distributed in Southeast Asia limited in the islands of Sumatra and Borneo (Figure 1.1.1). There are in total seven different species and up to fourteen subspecies of great apes. Despite the sparse fossil record, the number of extinct species of great apes appears to be large.

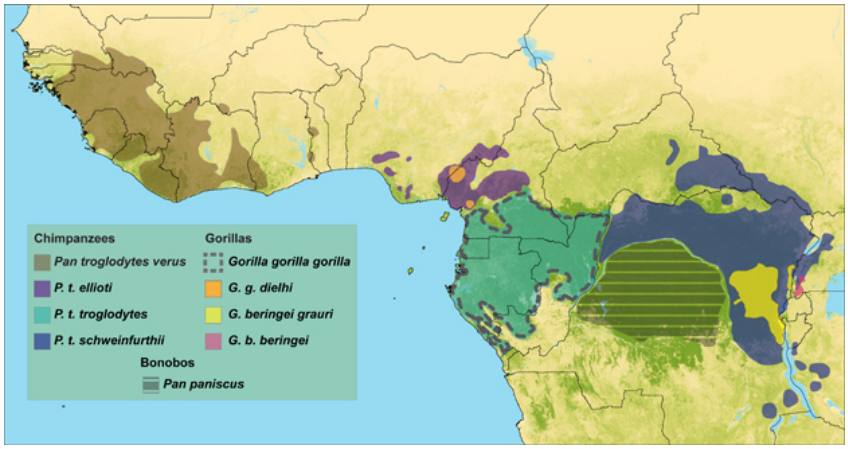


**Figure 1.1.1 –** Current distribution of the extant great apes. *Gorilla* and *Pan* genus inhabit the African forest with a wide distribution in equatorial Africa. Orangutan species are found in Southeast Asia in Borneo and Sumatra islands. (Mitchell & Gonder, 2013)

Orangutans (Malay word meaning the man of the forest) are our most distant relatives and are classified in two different species, the Sumatran orangutans (*Pongo abelii*) and Bornean orangutans (*Pongo pygmaeus*). This range has been changing over time since the area known as Sundaland (Malay Peninsula, Java Sumatra and Borneo) have been joining and separating repeatedly. There are no subspecies of the Sumatran orangutan but there are three in Borneo: *P.p. pygmaeus* in northwest Borneo, *P.p. wurmbii* in central Borneo and in northeast of the island ranges *P.p. morio*. The geographical conditions in Borneo have shaped this species and besides genetic support for this division there are size differences between the subspecies. Orangutans are characterized by its reddish-brown hair and grey-black skin. They spend most of their time hanging in trees and for this reason they are equipped with very long and strong arms opposed to short and bowed legs. Their feet are mostly adapted to the trees and allow them to grip on branches. Given this arboreal adaptation and unlike gorillas and chimpanzees, orangutans are not true knuckle-walkers, and they move in the ground using their fists. Both species are large and show great sex dimorphism. In terms of weight, adult males get to 75kg while females around 40kg. They are the least social species among great apes and live mostly a solitary lifestyle. Their diet is mostly based on fruits that they find opportunistically and changes from month to month.

Gorillas are the largest species of primates. Like orangutans they show extreme sex dimorphism and these black coated species can reach up to 180kg while females about half of that. Males reach human heights up to 1.8m and have larger arm spans and develop a silvering of the hair in their backs and sagittal crests, that gives them the 'silverback' name. Like Pan species they move using their knuckles in the ground, where they typically are found despite they nest using the trees as beds. The gorilla society is usually composed by groups of a single adult male and several females and their children. While this is strict in

western gorillas, eastern gorillas groups may be composed of multiple adult males, usually related among them. All gorillas inhabit the central equatorial Africa (Figure 1.1.1), and they are classified in two different species, both divided in two subspecies. The western species (*Gorilla gorilla*) are divided in the most abundant subspecies, western lowland gorillas (*G.g. gorilla*), and the critically endangered Cross River gorilla (*G.g. diehli*). The eastern species (*Gorilla beringei*) are classified in eastern lowland gorillas (*G.b. graueri*), and the most iconic of all great apes, the also critically endangered mountain gorillas (*G.b. beringei*). Eastern gorillas tend to be larger with longer and blacker hair than the western species, which have sleeker and greyer/browner hair. Their large body size allows them to consume a poor quality diet and it is mostly based on herbaceous leaves and shoots, but their diet is also seasonal and contains fruits when possible.

Chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*), also known as common chimpanzee and pygmy chimpanzee respectively, overlap in range with most of the distribution with gorillas but their areas are significantly larger. Bonobos are limited to the Democratic Republic of Congo (DRC), south of the Congo River while chimpanzees range from western African countries up to the most eastern countries. The latter species encompass the widest distribution among great apes, a fact that has helped to divide their population to four subspecies, western chimpanzees (*P.t. verus)*, Nigeria-Cameroon (*P.t. ellioti*), central (*P.t. troglodytes*) and eastern (*P.t. schweinfurthii*). This genus does not show such strong sex dimorphism but males are a bit larger. They are substantially smaller than gorillas, with male chimpanzees reaching weights up to 70kg. Like gorillas, chimpanzees are knuckle-walkers but they spend more time in the trees. Their diet is ample and includes a variety of fruits, seeds, leaves, insects and small mammals. Chimpanzees groups apply complex cooperative strategies in hunting (Pruetz & Bertolani, 2007) and bonobos, previously thought to not practice hunting, have also been reported (Surbeck & Hohmann,

2008). In this context orangutans have also been observed eating fish opportunistically (Russon, Compost, Kuncoro, & Ferisa, 2014). The two different species of Pan appear to share and differ in many aspects of their societies. While females leave the groups upon reaching adulthood, bonobo females have more power than in chimpanzees; bonobos are more peaceful than chimpanzees and there are marked differences in the sexuality between these species.

All the information used in this section have been obtained from these sources: (Caldecott, Miles, & eds, 2005; Harcourt & Stewart, 2007; Jurmain, Kilgore, Trevathan, & Ciochon, 2014).

### 1.1.2. Great apes as primates

Primate, "*any placental mammal of the order Primates, typically having flexible hands and feet with opposable first digits, good eyesight, and, in the higher apes, a highly developed brain"* is the definition given by the Collins English Dictionary. The *higher apes* are the species included under the Hominidae family and as such they belong to the Primate order. Despite being classified as order, they maintain many of the ancestral traits and they are not highly specialized mammals, but they still show the main characteristics mentioned above. Primates (order with 480 species in 78 genera) are mainly subdivided in Strepsirrhini (wet nosed prosimians, e.g. mouse lemur, aye-aye) and Haplorrhini (dry nosed tarsiers, monkeys and apes). Further grouping of the Haplorrhini suborder divides the Platyrrhini (New world monkeys, i.e. Marmoset) and Catarrhini. Great apes belong to the latter division, along with Old world monkeys (i.e. Macaque, Baboon) and the lesser ape, the gibbon (Figure 1.1.2). Great apes differ from monkeys in several traits: larger body size, no tail, generally more complex behaviour, more complex brain and enhanced cognitive abilities and increased period of infant development and dependency (Jurmain et al., 2014).

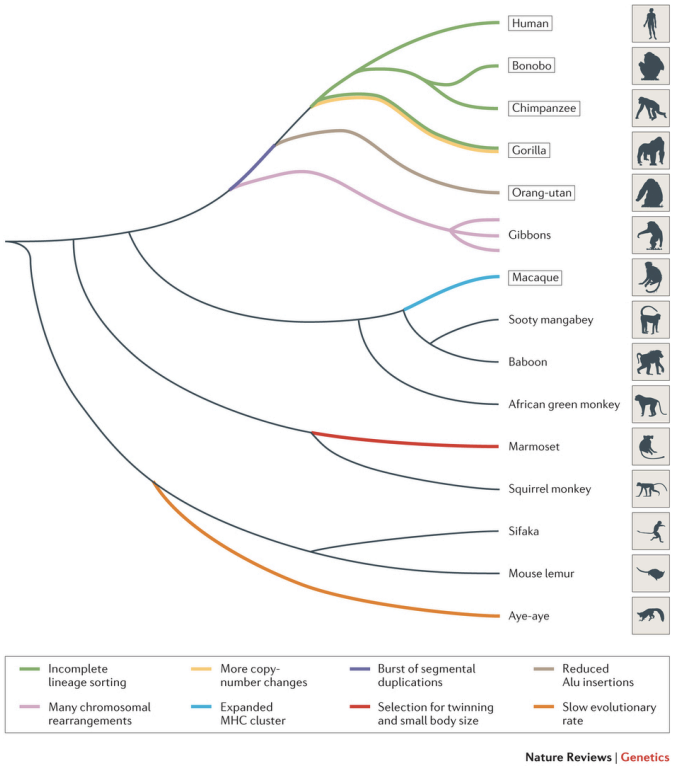


**Figure 1.1.2** – List of primates for which genomes are sequenced or in progress. The Hominidae family is shown on top of the tree along with the lesser ape, the Gibbon. Old-World monkeys are represented by the Macaque, Sooty mangabey, Baboon and African green monkey. Marmoset and Squirrel monkey are New-World monkeys and prosimians are represented by the Sifaka, Mouse lemur and the Aye-aye. (Rogers & Gibbs, 2014)

The dating of the most common ancestor of primates has been a subject of great interest among scientists. The consensus view is that the common ancestor of the order lived >80Mya (Chatterjee, Ho, Barnes, & Groves, 2009; Perelman et al., 2011). This was followed by an Eocene expansion until the different lineages diverged.

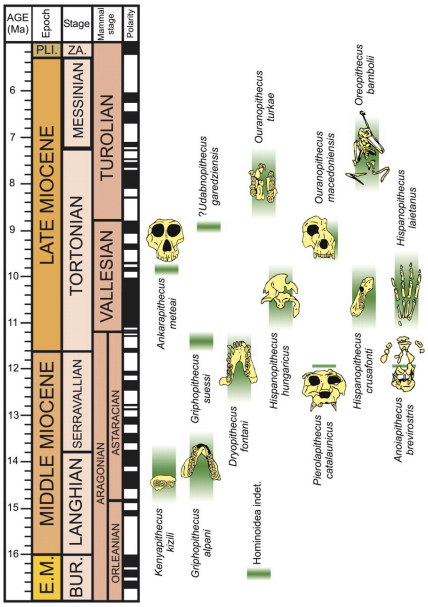


**Figure 1.1.3 –** Fossil record corresponding to the hominids from Western Eurasia from the Miocene period (Casanovas-Vilar, Alba, Garcés, Robles, & Moyà-Solà, 2011).

The common ancestor of great apes is also immersed in a heated debate. From the fossil record we know that the hominoid ancestor originated in Africa and was followed by an impressive early radiation during the Early Miocene with a decline in Africa (Begun, 2007; Harrison, 2010). The sparse but diverse fossil record (Figure 1.1.3) and uncertain dating has provided a plethora of different scenarios on the great ape origins. Some studies place a common origin in Eurasia followed by a back-to-Africa for the Homininae, something observed in other carnivores and hippos (Begun, Güleç, & Geraads, 2003). This attribution was given since the dating of the Eurasian hominids (*Griphopithecus* and *Kenyapithecus*) were wrongly dated ca. 16Mya, but a recent revision dated them around 14Mya (Casanovas-Vilar et al., 2011). A more plausible explanation given this scenario is that the most basal lineage originated from the African *Kenyapithecus* which expanded to Eurasia and the lineages leading to the pongins and hominines evolved independently in their respective continents

(Casanovas-Vilar et al., 2011). This process may have occurred between 14 and 12.5Mya.

### 1.1.3. Man among apes

Today, there is no scientific debate on man's place in evolution. But for centuries prior to the study of molecular biology, this issue has been of great controversy. First recorded observations on the position of human among primates was back in the Roman Empire where Galen of Pergamon (AD 129 – c. 200/c. 216) used Barbary macaques for biomedical research using dissections of these primates to understand physiology and translate the results into human medicine. Further attempts to place man in nature came from the botanist Carl Linnaeus (1707-1778). In his famous book, *Systema Naturae*, he placed humans in the same genus as chimpanzee, but he only classified individuals without an evolutionary perspective. Not until Thomas Henry Huxley (1825–1895) in his famous book *Evidence as to Man's Place in Nature* (1863) an evolutionary hypothesis was given on the common ancestor of human and great apes (Figure 1.1.4). This book was founded on Charles Darwin's masterpiece *On the Origin of Species* (1859) in which he did not tackle the human question. He would do so in the book *The Descent of Man*, a few years after Huxley's book.

At this point it was certain that a common ancestor between human and apes was the only explanation for the human origins, but a fierce debate has been in place for most of the XXth century. Initial analysis of morphological traits started to define that the apes were the closest relatives to human and classified these species together. But the defined relationship between these species was not clear. The sparse fossil record of great ape ancestors did not help either in the phylogeny. Despite morphological characters does not produce reliable phylogenies in accordance with molecular data (Gibbs,

Collard, & Wood, 2000), soft-tissue characters supports trees identical to that of molecular data (Gibbs et al., 2000). All in all, these studies did not help in the timing of the speciation events in the family because no reliable clock could be used with these data.

Further studies included the use of chromosome karyotypes. These studies clearly defined the phylogeny and helped in the macromolecular changes in the DNA of the Hominidae family, being able to reconstruct the ancestral karyotype of great apes (Yunis & Prakash, 1982). This resolved that orangutans are the most distant relatives to human, followed by gorillas and the Pan genus share the most recent ancestor with humans. But again, the lack of calibration in the evolutionary processes of chromosome evolution did not allow the proper estimation of the speciation times in the family.

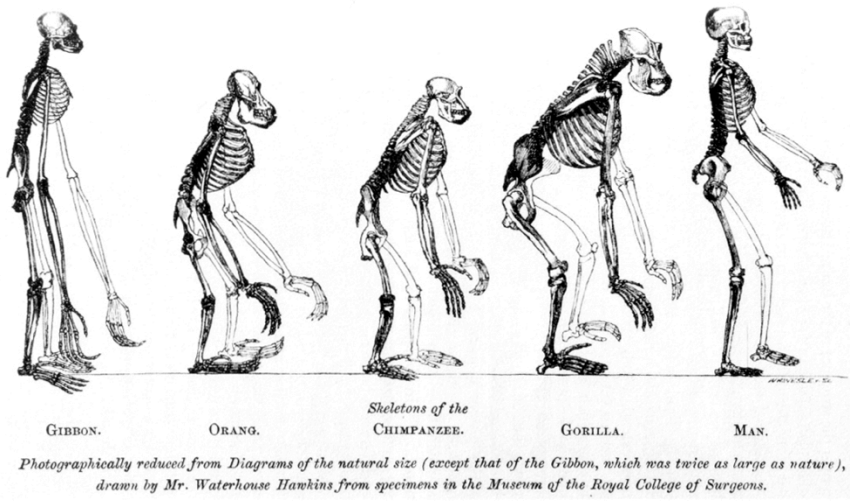


**Figure 1.1.4 -** This illustration was the frontispiece in Huxley's book *Evidence as to Man's Place in Nature*. It shows the anatomical similarities among the Hominoidea superfamily. Creative commons.

Finally with the usage of molecular techniques this was elucidated. Initial attempts included the usage of immunology to unravel the great ape tree. Allan Wilson may have been the main contributor in this

field by starting to shed light on the great ape phylogeny through immunological techniques (V M Sarich & Wilson, 1967; V. M. Sarich & Wilson, 1967), later followed by protein sequencing (King & Wilson, 1975) providing dating of the split events in the great ape phylogeny but these techniques provided low resolution in the correct relationship between African great apes. Later assays, using DNA-DNA hybridization and finally Sanger sequencing provided a more reliable tool in the final topology of the phylogenetic relationship between great apes (Jobling, Hollox, Hurles, Kivisild, & Tyler-Smith, 2013).

Despite there is enough evidence to support the clear phylogenetic tree shown in Figure 1.1.2, systematic analysis from phenotypical data in a recent study debated whether the correct great ape phylogeny placed orangutans as our closest evolutionary relatives (Grehan & Schwartz, 2009), a hypothesis highly criticized in (Stoneking, 2009) and contested in a reanalysis of that study combined with molecular data (Lehtonen, Sääksjärvi, Ruokolainen, & Tuomisto, 2011). This debate is far from finished and correspondences keep on the debate on who is right about the traits that can be analysed to support the Pongo-Homo clade (Grehan & Schwartz, 2011; Lehtonen, Tuomisto, Sääksjärvi, & Ruokolainen, 2012). But from molecular data it is clear that there is not such debate.

### 1.1.4. Hominidae phylogeny

Among the many applications that DNA sequencing has shed light on, the demographic history of great apes within species and the taxonomic classification of their populations has been the most important. Taxonomy in particular has been a discipline that has suffered from the problematic that the combination of phylogenetic, cladistics and systematics may not always agreed on the classification

of species, sometimes due to the lack of objectivity on the measurements assessed. Through the genetic study of these populations some light has been shed on their evolutionary relationships, further helping in the taxonomic classification of species.

In gorillas, initial genetic data between eastern and western gorillas (Garner & Ryder, 1996) suggested that eastern and western gorillas should have been elevated to the level of subspecies. Orangutans populations from the two major islands (Sumatra and Borneo) were also boosted as different species from the genetic study of mtDNA (Zhi et al., 1996). This also had repercussion in the captive management of this genus because for many years they hybrids were bred between the two species, nowadays hybrid populations are not bred anymore and the hybrid individuals are kept from breeding. Chimpanzees, due to their wide distribution across Africa and variation at many levels have been the most difficult species to tackle in their taxonomy. Currently, the common chimpanzee is classified in four subspecies but this has been changing over the last century. Over the last two decades, a heated debate has been the classification of the Nigeria-Cameroon chimpanzees as a subspecies (*P.t. ellioti*, initially named *P.t. vellerosus*). Initial work by (M K Gonder et al., 1997; Mary Katherine Gonder, Disotell, & Oates, 2006) studying the mtDNA of chimpanzees suggested that the Nigeria-Cameroon chimpanzees may be a differentiated population that could be elevated to the category of subspecies. Later assays of nuclear variation using microsatellite and SNP data were in line with this claim supporting the final classification of *P.t. ellioti* as a subspecies (Becquet, Patterson, Stone, Przeworski, & Reich, 2007; Bowden et al., 2012; Mary Katherine Gonder et al., 2011; Oates, Groves, & Jenkins, 2009).

Focusing on the phylogeny of great apes, chimpanzees have been the focal point in most of the phylogenetic and demographic studies

above all other taxa. Being the closest relative to our own and the huge population stratification in this species makes them the most interesting target to study relevant questions to population geneticists. Initial works studying within species relationships in great apes were rare (Morin et al., 1994), but later this question has been assessed deeper through the usage of all the molecular markers described in Section 1.3.1 (Becquet et al., 2007; Caswell et al., 2008; Fischer, Pollack, Thalmann, Nickel, & Pääbo, 2006; Fischer, Wiebe, Pääbo, & Przeworski, 2004; Kaessmann, Wiebe, Weiss, & Pääbo, 2001; Stone et al., 2010). These studies lacked a full representation of the genome to provide a more robust study of great ape phylogeny and demography. But recently, the genome assemblies of all genera have been released alongside whole genome sequencing of the taxa of orangutans (Bornean and Sumatran orangutans) (Locke et al., 2011), gorilla major species (Scally et al., 2012) and bonobo (Prüfer et al., 2012). These projects have provided the first whole genome view in the speciation processes of all major species but no genome-wide study has been performed between the subspecies of gorillas or between all chimpanzee taxa despite many comparisons have been carried in chimpanzees (Becquet et al., 2007; Bowden et al., 2012; Caswell et al., 2008; Mary Katherine Gonder et al., 2011). The whole genome comparisons between great ape species have provided a clearer picture on the complex evolutionary processes in the separation of these species. In gorillas, the split between western-eastern species was not clean and after an initial separation ~0.5Mya, gene-flow persisted until recently despite their current ranges are not contiguous (Scally et al., 2012). The separation between orangutan species occurred 400kya and as in gorillas, it was followed by low level gene-flow (Locke et al., 2011). The Pan genus has also been studied genome-wide pointing to a deeper split time compared to the other great ape genera, around 1Mya. In contrast to gorillas and orangutans, this split doesn't seem to have been followed by gene-flow, probably as a result of an allopatric

process driven by the formation of the Congo river (Prüfer et al., 2012).

Importantly, these studies have also contributed new methodologies in the study of population genetics and demography. (Becquet & Przeworski, 2007) developed a new Markov chain Monte Carlo method to estimate parameters of an isolation-migration model; multipopulation isolation-migration models (Hey, 2010a, 2010b); coalescent hidden Markov models applied in whole genomes (Mailund et al., 2012); pairwise sequentially Markovian coalescent (PSMC) to apply in a single whole genome (H. Li & Durbin, 2011). All these and many more have contributed largely to the basic understanding of the demography of our own species and that of our close relatives.

## 1.2. Great ape genomics

### 1.2.1. Studying great ape genomes

Great apes are amazing creatures that possess high instrumental and intrinsic values (Sandler, 2012). These species fascinate on many different grounds (Figure 1.2.1), but from a practical perspective great apes are the most suitable species to study many questions related to human biology as well as for the study of basic biological questions in great apes themselves.

The evolutionary position of great apes makes them a unique species in studying recent human evolution. They are the only extant species that can provide any clue on human origins in the last 15 million years. This has been a vital argument on the sequencing and assemblage of all great ape genera that I will summarize in the Section 1.2.3. Recently, the development of paleogenetics has boosted our knowledge in recent timescales through the sequencing of extinct hominins such as Denisovan, Neandertal and ancient modern humans (R. E. Green et al., 2010; Meyer et al., 2012).

Through comparative genomics of great apes and human we have also shed light on the genetic changes that have responsible of our *humanness*. The detection of unique genomic features to the human lineage has been the starting point in the study of the uniqueness of human traits. Despite limited, several genomic examples have been found to provide characteristic traits in humans. Some genes appear to have lost its function such as the MYH16 gene (Stedman 2004), expressed in muscles involved in chewing, or the loss of regulatory regions, one of them appear to have been the responsible of the loss of penile spines in humans (McLean et al., 2011). Some other genes appear to have gained important functions, two examples may be the

pivotal genes in the development of larger brains and complex language, SRGAP2 (Dennis et al., 2012) and FOXP2 (Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001) genes respectively. For a complete review on this topic see (Pääbo, 2014).

Several mechanisms have been proposed in the evolution of human traits. These include changes in the regulatory machinery that lead to differential gene expression, amino acid substitutions that alter protein function, gene duplication and the 'less-is-more' hypothesis (Olson, 1999). The latter was proposed in the late 90s by Maynard Olson and is based on the concept of loss of genes as an important evolutionary tool in human evolution. It is also founded in the idea that humans are a 'degenerate ape' with examples such as the loss of hair and muscle strength(Olson & Varki, 2003). This mechanism may have a quick effect in phenotype and could have important repercussions in the speciation process and genetic-loss events might outnumber the amount of genetic innovations. The idea seems feasible and straightforward. Since this hypothesis was proposed some examples of pseudogenization have been found, but these are still limited. Initial whole-genome scans using the chimpanzee genome started to identify human specific pseudogenes (HSP) (Hahn, Jeong, & Lee, 2007; Torrents, Suyama, Zdobnov, & Bork, 2003; Wang, Grus, & Zhang, 2006), in total 120 HSPs were reported but 14 of these were polymorphic in humans. (Kim, Igawa, Kawashima, Satta, & Takahata, 2010) reanalysed these HSPs and could find only 25 olfactory receptors (Gilad, Man, Pääbo, & Lancet, 2003) and 13 other pseudogenes. They conclude that these 38 examples may have not been enough to drive human evolution and that other kind of changes may have been more important. Studying human variation, (MacArthur et al., 2012) performed a systematic discovery of loss of function events using pilot data from the 1000 genomes. Then, they tested whether these gene disruption events correlated with signals of positive selection and they found no apparent deviation from

compared to non-synonymous events of the same allele frequency. They could still retrieve 20 loss-of-function events that could be under selection, supporting this mechanism as a driving force in evolution. Nevertheless, examples supporting the 'less-is-more' hypothesis keep showing up, for instance the pseudogenization of the Interferon Lambda 4 (*IFNL4*) has been associated with the adaptive maintenance of this gene inactivation because it improves the viral clearance of hepatitis C virus (Key et al., 2014).

An interesting and under-explored use of great ape genomics is their application to human biomedicine. As a consequence of being such a close evolutionary species we share many of the parasites with great apes. This represents a bidirectional risk of infection and many examples have been studied in detail of zoonotic diseases transferred from a great ape source into humans. Many of these pathogens have been studied in both humans and great ape counterparts. HIV and malaria parasites have been thoroughly studied and studies point to a zoonosis transmission from chimpanzees to human in the former while *Plasmodium falciparium*, the most common parasite in malaria infections; was transmitted from gorillas(Sharp, Rayner, & Hahn, 2013). But a large number of infectious pathogens have been found in great apes, posing a serious threat to global health(Calvignac-Spencer, Leendertz, Gillespie, & Leendertz, 2012). But this global health problem is in both directions, some great ape populations have been seriously compromised by the action of these zoonotic pathogens (Köndgen et al., 2008). And some of them such as the Ebola virus disease (EVD) has repeatedly affected to great apes following human outbreaks (Bermejo et al., 2006). We shall see whether the recent outbreak in west Africa (Gire et al., 2014), could also become a health threat to the chimpanzees neighbouring the affected areas.

And last but not least, conservation. All species of great apes are enlisted as endangered species from the IUCN Red List of Threatened

Species (IUCN 2014, 2014). This problem must be tackled from many different angles but conservation genetics must play an important role in the conservation of these species. Over the last decades the study of the mitochondrial genome and microsatellites have been crucial in this task. These have allowed the study of the biogeography of great apes, recent population declines and population stratification. Moreover, the use of genetics in breeding programs of captive populations and the origin identification of confiscated animals from illegal trade has to become widespread tools as it is currently happening in the ivory trade (Wasser et al., 2004). In fact, a UN organization devoted to the conservation of great apes (GRASP) has shown interest in applying same methodologies in the law enforcement in great apes. As I will explain in the next section, sequencing is becoming a very cheap technology that will can also be applied to conservation through conservation genomics (Avise, 2009; Kohn, Murphy, Ostrander, & Wayne, 2006; Primmer, 2009).

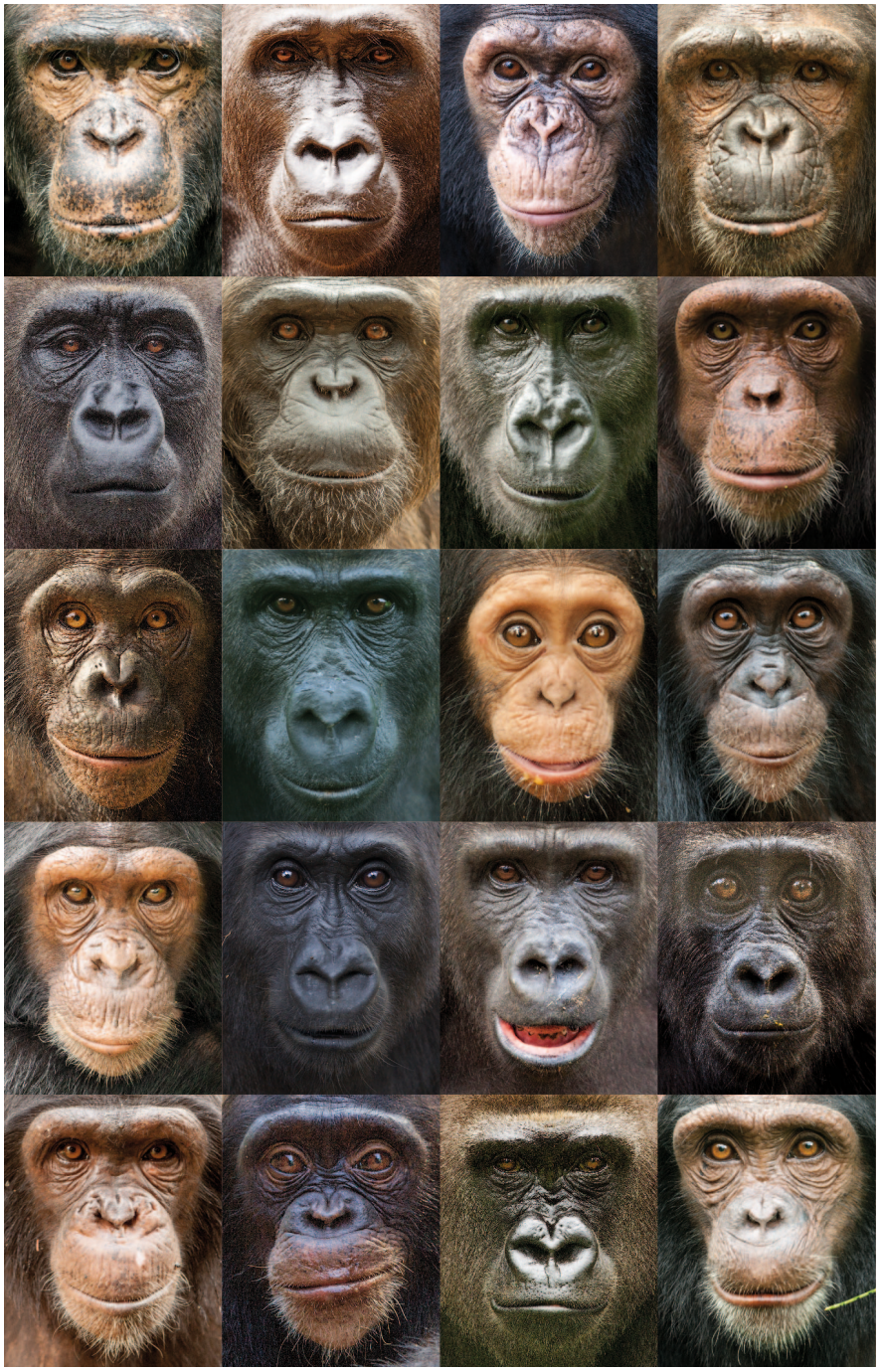

Figure 1.2.1 – Face diversity among African great apes. Pictures taken by Ian Bickerstaff.

### 1.2.2. Sequencing revolution

DNA sequencing has boosted many fields of biology over the last decade and the simple idea of studying whole genomes cannot be understood without the context of critical improvements in this technology. The early ages of DNA sequencing started with low-throughput technologies that allowed the first sequences to be available and start to understand small pieces of the genome. Not until 1981 the first human mitochondrial genome was available (Anderson et al., 1981), a decade after the first DNA sequencing was ever produced. These advances provided major insights and opened the field of genetics, but further improvements were still required to tackle the daunting task of sequencing full genomes, such as the human genome, several orders of magnitude larger than the mitochondria. Despite the first plans to tackle this issue started back in mid 80s, the project did not start until the next decade.

It took 15 years until it's final status (despite it is still being improved nowadays) with a cost around $3 billion (Hayden, 2014). By the end of the Human Genome Project, alternative methologies based on whole genome sequencing allowed to sequence a human genome for "only" $100 millions, with a quality trade-off (Venter et al., 2001). At this point the potential of the DNA sequencing in the genomic era kept on reducing the cost dramatically. In 2005, the 454 pyrosequencer, first high-throughput technology, plummeted the sequencing costs by an order of magnitude. This was followed by the Solexa/Illumina technology that plummeted the cost further. Many technologies have been competing in this race towards the so-called $1,000 genome (Figure 1.2.2)(Hayden, 2014). And along this competition, many large-scale projects were launched: ENCODE Project, 1,000 Genomes, Hapmap3, Human Microbiome Project (Mardis 2011). Despite the limitations of these technologies, they have opened the possibility to cheaply study genomes from many different perspectives and the

impact that this is having in science and our lives through biomedical research is unprecedented (Koboldt, Steinberg, Larson, Wilson, & Mardis, 2013).

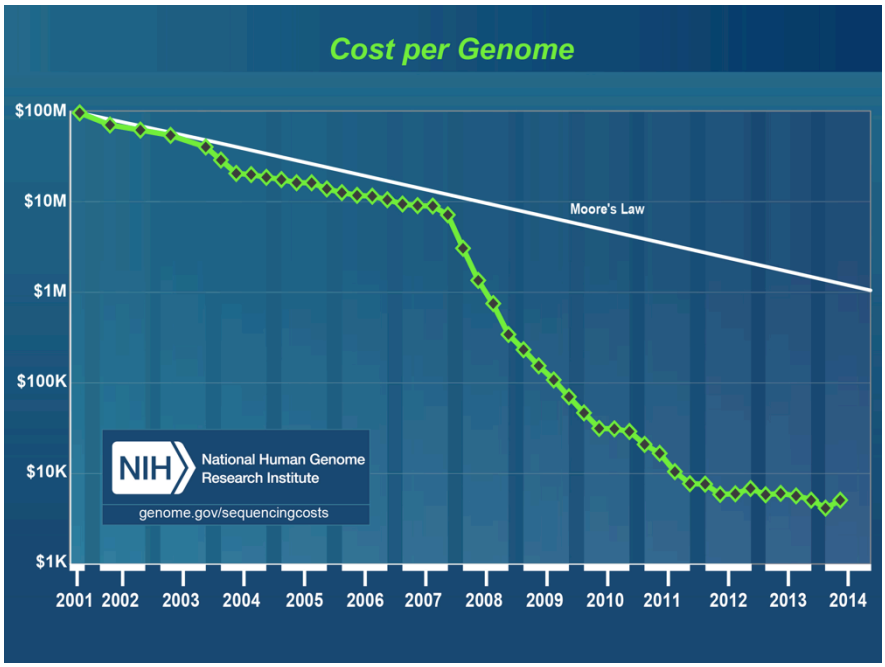


**Figure 1.2.2 –** Decrease in sequencing cost per genome in the last decade. From: http://www.genome.gov/sequencingcosts/

### 1.2.3. Great ape genomes

Back in 2000, when the Human Genome Project was about to enter in the final stages, the need to gain insights into the function of the genome was an imperative. Several approaches followed this path, but a major consensus in the scientific community was the sequencing of model organisms (mouse, *C. elegans*, fruit fly and yeast)(Edwin H. McConkey & Goodman, 1997). These genomes and the experiments carried in these organisms, gave major clues on the functions of many genes shared with human and therefore important information in the

understanding of our genome. But other basic questions on human biology were not accomplished with the sequencing of these species.

The study of the evolution of the human lineage (E H McConkey et al., 2000; Edwin H. McConkey & Goodman, 1997), the identification of the humanness in the human genome (E. H. McConkey, 2000) and fundamental implications on biomedical research, especially in Old World monkeys, i.e. Rhesus macaques and baboons but also in chimpanzee (VandeBerg, Williams-Blangero, Dyke, & Rogers, 2000; Varki, 2000) were questions that needed the sequencing of non-human primates, and the scientific community argued for the right species to sequence. But why all this controversy? These projects were really expensive, over $100 million at that time. While most scientists agreed on the sequencing of the chimpanzee genome as a priority along with other non-human primates used in biomedical research(Eichler & DeJong, 2002; E. H. McConkey, 2000), other scientists pointed in a more practical solution from a biomedical perspective to sequence the rhesus macaque and baboons genomes prior to that of an ape which cannot be considered an animal model (VandeBerg et al., 2000).

NIH gave high priority to the chimpanzee genome (Olson & Varki, 2003) and other projects were queued until the sequencing technology would make them more affordable. The approach used for this first genome was the whole genome shotgun (WGS) (Sanger, Coulson, Hong, Hill, & Petersen, 1982; Venter et al., 2001), a cheaper strategy compared to the clone-by-clone based approach (Lander et al., 2001) despite being a controversial issue (P. Green, 1997). The former methodology sequences the whole genome and then try to reassemble the complete puzzle while the latter chops the genome in smaller pieces (BACs, 150-350Kbp) and resolves each of these pieces individually to ensemble them all in a complete fully sequenced genome. The fact that the human reference was already available and would be used as a template for the assembly and the higher cost of

the clone-by-clone favoured the WGS strategy. A male captive-born individual (western chimpanzee, Clint) was chosen and he initial release of this genome had a 3.6X fold sequence redundancy in autosomes and the initial analysis was presented (Consortium, 2005) four years after the human genome (Lander et al., 2001) and three years after the mouse (Waterston et al., 2002). This represented a major landmark in the study of the human genome. The first comparison with a non-human primate genome allowed the first broad-scale analysis on the study of the primate genome. Genome-wide analysis on mutational processes, divergence rates, insertions and deletions, transposable elements and the complete analysis on gene evolution and further identification of genes under selection were the major findings on this first look at the chimpanzee genome. This also provided a framework in comparative genomics and an important resource in the study of human and chimpanzee biology.

But the list of primate genomes kept growing, the sequencing projects were revaluated by the funding agencies and some of them were prioritized. By 2002, macaque and baboon genome projects had started and orangutan, vervet and squirrel monkey were approved by the BAC Library Resource Network (Eichler & DeJong, 2002). Then European initiatives funded the remaining great ape genomes, *Gorilla gorilla gorilla* from the Wellcome Trust, UK, and *Pan paniscus* as a personal initiative funded with a ERC grant awarded to Svante Pääbo in 2008 (TWOPAN project).

Focusing on the great ape genome assemblies, they were made through different sequencing strategies according to the technological changes. While the first projects were done with the old Sanger WGS aiming to produce 6X coverage in the chimpanzee and orangutan, the latter assemblies were performed with a more adjusted budget. For this reason the gorilla genome started with the Sanger WGS approach (<2X) and finishing the assembly with a hybrid approach using high

throughput Illumina paired-end sequencing (56X) (Scally et al., 2012). The last genome was fully assembled with next-generation sequencing with the 454 sequencing platform (26X) (Prüfer et al., 2012), which provides longer reads than Illumina but with a higher indel error rate. Large consortiums were devoted to these projects and media released attracted huge attention providing a crucial resource for the scientific community in the study of great apes and primate evolution (Rogers & Gibbs, 2014).

High-throughput sequencing technologies have reduced the sequencing cost dramatically but only a handful of projects have been carried out in resequencing great apes genomes. Despite these data is crucial in studying the biology of the genomes, great ape genomes are still underrepresented. Only the orangutan and the gorilla genome projects (Locke et al., 2011; Scally et al., 2012) sequenced a few individuals to provide a deeper insight into the population history of these species and recent resequencing of chimpanzees have provided insights in recombination, mutation rate and balancing selection (Auton et al., 2012; Leffler et al., 2013; Venn et al., 2014). Finally the last resequencing projects of great ape genomes were devoted to study the structural variation of the genome (Gokcumen et al., 2013; Ventura et al., 2011). These are all the projects that have resequenced whole genomes of great apes (around 30 individuals), despite other capture approaches have been applied in great apes to provide a genome-wide perspective of the genome (Greminger et al., 2014; Christina Hvilsom et al., 2011; Scally et al., 2013).

## 1.3. Great ape diversity

### 1.3.1. Assessing genetic diversity

Genetic diversity is a fundamental tool in the evolutionary study of populations, the demographic history, population migrations, fluctuations in effective population size, geographic structure (Wall, 2013) are among the questions relevant to assess the genetic diversity but the fundamental question on how and why this variation is maintained is also a topic of interest (Leffler et al., 2012).

Over the last two decades the assessment of genetic variation in great apes has ranged a wide variety of strategies in terms of markers sampled and the technologies that were used in pace with the ongoing developments in the field. From uniparental markers such as mtDNA and Y chromosome, to autosomal chromosomes and the recombining X chromosome; some have been focused in neutral regions to have a less biased view on the genome while others focused in genes under high selective pressure such as the HLA genes (Lawlor, Ward, Ennis, Jackson, & Parham, 1988). Some studies sampled microsatellites while others focused in sequencing strategies (initially Sanger sequencing) or restriction fragment length polymorphism (RFLP) assays (Kaessmann & Pääbo, 2002); but currently, the advent of high-throughput sequencing will become the preferred choice.

Despite the widespread use of mtDNA markers in the study of great ape diversity, there are concerns on whether they are an ideal proxy in evolutionary studies (Melnick & Hoelzer, 2005) and the large number of nuclear insertions of mtDNA (Numts) has been a concern in studies of great apes (Thalmann et al., 2005; Thalmann, Hebler, Poinar, Pääbo, & Vigilant, 2004). The technologic advances in sequencing have allowed the possibility to cheaply sequence complete genomes or to sequence targeted regions of interest, e.g. exome

sequencing in clinical applications. This represents both a great opportunity to obtain high amounts of data but a problem since the analysis is becoming a more expensive part of the project. Now that all genera of great apes have been assembled (Section 1.2.3), the most common practice in the analysis of sequenced genomes is to use these references to align these short reads produced with the high throughput technologies. Most of these projects mainly study the most frequent and easier to sample genomic variation, the Single Nucleotide Polymorphisms (SNPs). Indels, microsatellite and other kind of structural variants such as large deletions or insertions or transposable elements are not as commonly assessed due to the technical difficulties in the detection of these variants. The shorter reads of high-throughput technologies poses the major problem in the detection of this variation but other concerns are known e.g. library preparation, genome complexity, coverage (Alkan, Coe, & Eichler, 2011; Medvedev, Stanciu, & Brudno, 2009).

The most common practice in the detection of any kind of variation starts with the mapping of the raw reads into a reference genome, then realignment and filtering of mapping and sequencing artifacts is crucial prior to the sampling of the variation of the sample of interest (Figure 1.3.1). This step can combine multiple samples belonging to a population to provide more power to the variants found or can be done in single genomes in case the coverage and the quality of the samples allow this. Further filtering is required to extract the meaningful variants from low quality variation.

The field of analysis of high throughput sequencing technologies is still under development. Currently, many teams are contributing in this field and the list of available tools keeps growing fast (Pabinger et al., 2014) providing fast and user-friendly tools that can be used by anyone, with an especial focus devoted to medical genetics studies where clinicians tend to lack computational skills (Altmann et al.,

2012) and tools such as galaxy are clearly conceived to be used in this direction (Goecks, Nekrutenko, & Taylor, 2010).

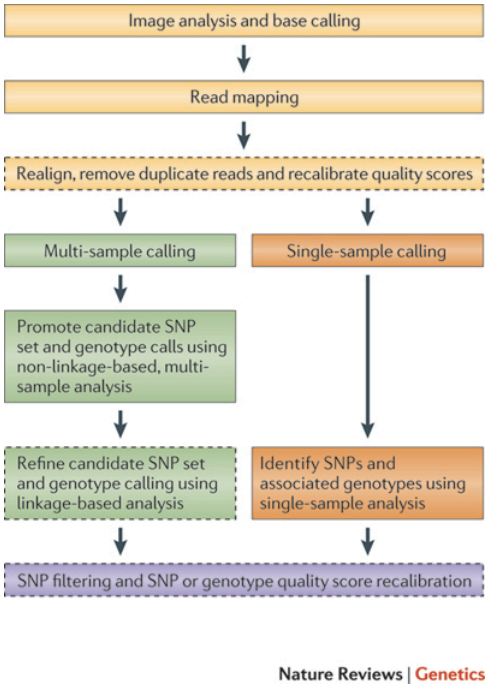


**Figure 1.3.1** – Typical workflow in the SNP calling strategy using a reference genome. From (Nielsen, Paul, Albrechtsen, & Song, 2011).

### 1.3.2. Genetic diversity in great apes

Most studies of genetic diversity in great apes have shown higher variability than humans, frequently using mtDNA. And by far the most studied group has been the chimpanzee. In general, these studies have shown that great apes bear higher levels of diversity (Kaessmann et al., 2001) despite the smaller census sizes of these species. Only a few studies have found lesser diversity in apes compared to humans: (Takahata, 1993) study suffered from poor sampling in the great ape individuals analysed and (Wise, Rubinsztein, & Easteal, 1997) with problems of ascertainment bias. But a large amount of studies have

found the opposite observation (Ferris, Brown, Davidson, & Wilson, 1981; Garner & Ryder, 1996; Kaessmann et al., 2001; Morin et al., 1994; Stone, Griffiths, Zegura, & Hammer, 2002; Warren et al., 2001; Zhi et al., 1996) (Figure 1.3.2).

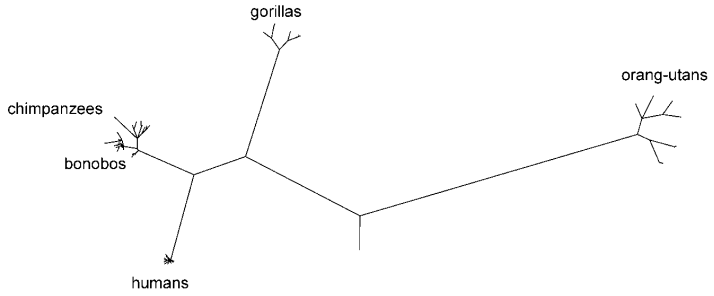


**Figure 1.3.2 –** Great ape genetic diversity and phylogeny in a 10kbp region of the X chromosome. All species appear to have higher diversity than the human lineage, evidenced by the longer and more divergent branches in great apes (Kaessmann et al., 2001).

But this story is different depending on the genetic marker assessed. While mtDNA appears to show greater genetic diversity in all great apes, nuclear markers are more modest in this respect. In chimpanzees, western chimpanzees (*P.t. verus*) show the greater diversity of all subspecies in the mtDNA, like central/eastern chimpanzees that still show greater genetic diversity to that of humans (Stone et al., 2010). But the sampling of nuclear DNA has provided different stories. Analyses of the non-recombining portion of the Y chromosome point to a higher diversity in central chimpanzees compared to western (Stone et al., 2002) and similar results were found in the Xq13.3 (Kaessmann, 1999). Furthermore, autosomal sequences show that the most variable subspecies are central chimpanzees, followed by eastern and western populations (Fischer et al., 2004; Yu et al., 2003), with the latter having comparable amount of variation to that of humans (Fischer et al., 2004). These differences have been attributed to possible scenarios of founding effects with

complex and skewed demographic histories between males and females (Stone et al., 2010).

### 1.3.3. Insights from genetic diversity

Despite the initial attempts in sampling diversity were focused on the importance of putting human genetic diversity in an evolutionary context, later assays tried to study the genetic diversity in great apes to solve evolutionary questions relevant in many different areas. Molecular ecology, a discipline tightly linked to conservation genetics, has been one of the many applications that genetic studies have focused to understand important aspects of ape biology. These studies include male and female dispersal patterns (Bradley, Doran-Sheehy, Lukas, Boesch, & Vigilant, 2004; Bradley, Doran-Sheehy, & Vigilant, 2007), population structure of natural populations of great apes (Fünfstück et al., 2014), population census using genetics (Guschanski et al., 2009), the effects of habitat fragmentation in critically endangered populations (Bergl, Bradley, Nsubuga, & Vigilant, 2008), among others studies (Vigilant & Guschanski, 2009). These and many more are critical in the recognition of genetics in field studies and it is important to acknowledge the important task of sampling and DNA recovery from non-invasive samples.

Recent assays of variations using a genome-wide approach have been focused in understanding basic population genetics processes such as mutation rates, recombination, selection and shared polymorphisms. In terms of recombination, a recent study found that humans and chimpanzees do not usually share the hotspots of recombination (Auton et al., 2012) despite humans share most of them between populations. The main contributor to these hotspots seems to be associated with the *PRDM9* gene and it seems to be extremely variable in chimpanzees, also changing the binding motifs of the resulting

protein. Using similar approaches, recombination is under current study in gorillas, as part of the GAGP (Wall, 2013). Another study has also tried to reconstruct the recombination map for the human-chimpanzee ancestor, finding that it has evolved more rapidly in humans since the split (Munch, Mailund, Dutheil, & Schierup, 2014). Mutation rate has also been studied through the sequencing of whole genomes of chimpanzee trios. A recent study found that despite human and chimpanzees share an identical mutation rate ($1.2 \times 10^{-8}$ per base pair per generation), chimpanzees have a stronger bias in the origin of these mutations compared with humans i.e. males contribute 7-8 times more mutations than females in chimpanzees, whereas humans 3-4 times. These crucial genetic processes in molecular evolution seem to differ in very close species such as human and chimpanzee, and it remains unclear how these values will be in more distant evolutionary relatives. But this has to been coupled with field work, a recent study based on long-term observations of social communities of African great apes focused in the generation times of these societies, finding generation intervals around 25 years, and 20 in gorillas, values larger than what was typically used in population genetics studies (Langergraber et al., 2012), and generation time is a crucial parameter to population geneticists.

In general, lack of whole genome data in most ape populations has lead to sample the variation from a limited set of genomes or even from human variation that has lead to problems such as ascertainment bias in the study of new populations. Without the proper sampling and variant discovery in a diverse set of populations, targeted studies of this variation may preclude from the correct conclusions in population genetics studies where new populations are studied. To overcome these limitations recent genomic have focused in the retrieval of proper genetic markers to study these populations. These studies include reduced representation sequencing in orangutans and gorillas (Greminger et al., 2014; Scally et al., 2013) as well as genome-

wide sequencing in the full exome in chimpanzees (Christina Hvilsom et al., 2011; Teixeira et al., 2014a) and whole genomes (Auton et al., 2012; Locke et al., 2011; Scally et al., 2012) are important to sample the variation in these species and provide a useful set of markers that can be applied to population genetics in great apes.

Finally, the possibility to study complete genomes and to retrieve diversity from our most close evolutionary relatives has also permitted the scan for variation that is shared between different lineages. These events are the result of ancient polymorphisms that are maintained due to selective pressures and the coalescence of the different haplotypes around these regions are able to predate the timing of speciation events between species. Examples of these are limited and initial resequencing of a few genes discovered some examples of shared polymorphism between humans and chimpanzees. The first gene discovered with this kind of signal was the major histocompatibility complex (MHC) a gene crucial in the immune system (Klein, Satta, O'hUigin, & Takahata, 1993), TRIM5 (Cagliani et al., 2010) and ABO blood group (Ségurel et al., 2012) are among the examples.

Currently the ability to scan the whole genome (Leffler et al., 2013) or the whole exome (Teixeira et al., 2014b), provide a unique scenario to find these rare signals. The former study found up to 125 regions with signals of trans-species haplotype sharing between human and chimpanzees, with only two examples of coding variation. But only six of those regions appear to conclusively point to ancestral polymorphisms that are still variable in human and chimpanzee populations. This was obtained through the whole genome sequencing of 10 chimpanzee individuals. The latter scans the exomes of 20 individuals of chimpanzees, bonobos and humans and finds a few examples of coding shared variation. As expected from the previous study, that showed that most trans-species polymorphisms are in the

non-coding regions of the genome, they could find a limited number of examples. Besides the omnipresent MHC locus, this study identifies a new example of balancing selection the *LAD1* gene, responsible maintaining cohesion at the dermal-epidermal junction and has been associated with linear IgA disease.

These and upcoming studies on different populations of great apes can help to identify how strong the evolutionary forces are shaping our genomes. And given the important implications of these kind of polymorphic adaptations have related with disease and as I have described in the Section 1.2.1, we share many diseases that we can study from this genomic perspective in the adaptation of great apes.

# 2. OBJECTIVES

1. Provide the first analysis of the complete great ape phylogeny using whole genomes in all the genera from the Hominidae family.

2. Study the unique case of albinism in gorillas and unravel the genetic cause for this disorder in gorillas using whole genome sequencing.

3. Through the use of high throughput sequencing, provide the most complete resource of whole genomes in great apes (both captive and wild populations) and genetic diversity.

4. Study the population history of all great apes genera, within and between species, producing the most comprehensive study of great ape evolution with whole genomes.

5. Analyse the genomic fingerprints of inbreeding in great ape populations and study how effective population size have an effect in selection.

6. Test the 'less-is-more' hypothesis for the first time combining polymorphism data on great apes.

# 3. RESULTS

## 3.1. Insights into hominid evolution from the gorilla genome sequence

A. Scally, J. Y. Dutheil, L. W. Hillier, G. E. Jordan, I. Goodhead, J. Herrero, A. Hobolth, T. Lappalainen, T. Mailund, T. Marques-Bonet, S. McCarthy, S. H. Montgomery, P. C. Schwalie, Y. A. Tang, M. C. Ward, Y. Xue, B. Yngvadottir, C. Alkan, L. N. Andersen, Q. Ayub, E. V Ball, K. Beal, B. J. Bradley, Y. Chen, C. M. Clee, S. Fitzgerald, T. A. Graves, Y. Gu, P. Heath, A. Heger, E. Karakoc, A. Kolb-Kokocinski, G. K. Laird, G. Lunter, S. Meader, M. Mort, J. C. Mullikin, K. Munch, T. D. O'Connor, A. D. Phillips, J. Prado-Martinez, A. S. Rogers, S. Sajjadian, D. Schmidt, K. Shaw, J. T. Simpson, P. D. Stenson, D. J. Turner, L. Vigilant, A. J. Vilella, W. Whitener, B. Zhu, D. N. Cooper, P. de Jong, E. T. Dermitzakis, E. E. Eichler, P. Flicek, N. Goldman, N. I. Mundy, Z. Ning, D. T. Odom, C. P. Ponting, M. A. Quail, O. A. Ryder, S. M. Searle, W. C. Warren, R. K. Wilson, M. H. Schierup, J. Rogers, C. Tyler-Smith, and R. Durbin. Insights into hominid evolution from the gorilla genome sequence. Nature. 2012;483(7388):169-175. DOI: 10.1038/nature10842

## 3.2. The genome sequencing of an albino Western lowland gorilla reveals inbreeding in the wild

J. Prado-Martinez, I. Hernando-Herraez, B. Lorente-Galdos, M. Dabad, O. Ramirez, C. Baeza-Delgado, C. Morcillo-Suarez, C. Alkan, F. Hormozdiari, E. Raineri, J. Estellé, M. Fernandez-Callejo, M. Valles, L. Ritscher, T. Schöneberg, E. de la Calle-Mustienes, S. Casillas, R. Rubio-Acero, M. Melé, J. Engelken, M. Caceres, J. L. Gomez-Skarmeta, M. Gut, J. Bertranpetit, I. G. Gut, T. Abello, E. E. Eichler, I. Mingarro, C. Lalueza-Fox, A. Navarro, and T. Marques-Bonet. The genome sequencing of an albino Western lowland gorilla reveals inbreeding in the wild. BMC Genomics. 2013;14:363.
DOI: 10.1186/1471-2164-14-363

## 3.3. Great ape genetic diversity and population history

J. Prado-Martinez, P. H. Sudmant, J. M. Kidd, H. Li, J. L. Kelley, B. Lorente-Galdos, K. R. Veeramah, A. E. Woerner, T. D. O'Connor, G. Santpere, A. Cagan, C. Theunert, F. Casals, H. Laayouni, K. Munch, A. Hobolth, A. E. Halager, M. Malig, J. Hernandez-Rodriguez, I. Hernando-Herraez, K. Prüfer, M. Pybus, L. Johnstone, M. Lachmann, C. Alkan, D. Twigg, N. Petit, C. Baker, F. Hormozdiari, M. Fernandez-Callejo, M. Dabad, M. L. Wilson, L. Stevison, C. Camprubí, T. Carvalho, A. Ruiz-Herrera, L. Vives, M. Mele, T. Abello, I. Kondova, R. E. Bontrop, A. Pusey, F. Lankester, J. a Kiyang, R. a Bergl, E. Lonsdorf, S. Myers, M. Ventura, P. Gagneux, D. Comas, H. Siegismund, J. Blanc, L. Agueda-Calpena, M. Gut, L. Fulton, S. a Tishkoff, J. C. Mullikin, R. K. Wilson, I. G. Gut, M. K. Gonder, O. a Ryder, B. H. Hahn, A. Navarro, J. M. Akey, J. Bertranpetit, D. Reich, T. Mailund, M. H. Schierup, C. Hvilsom, A. M. Andrés, J. D. Wall, C. D. Bustamante, M. F. Hammer, E. E. Eichler, and T. Marques-Bonet. Great ape genetic diversity and population history. Nature. 2013;499(7459):471-5. DOI: 10.1038/nature12228

# 4. DISCUSSION

## 4.1. Great ape genomes

Genomics has suffered a paradigm shift over the last decade, most importantly in the last five years with the advent of high-throughput sequencing technologies (Section 1.2.2). Over this period of time the costs of sequencing have plummeted and current sequencing projects have rapidly incorporated these technologies. In fact, the gorilla genome assembly (Section 3.1) did use high-throughput sequencing technologies to finish the assembly, a strategy pioneered in the assembly of the giant panda (R. Li et al., 2010). Despite the cost reduction in these kind of strategies it certainly has an important trade-off in quality (Alkan, Sajjadian, & Eichler, 2011). Most of the repetitive elements, segmental duplications and complex regions are likely to be misassembled due to the combination of whole genome sequencing and short read technology. These limitations are important for fine-scale analysis using these genomes, and undoubtedly the gene annotation of these genomes is far from perfect, but for broad evolutionary questions, these approaches are certainly the best option given the high costs of finished genome assemblies. With the advent of third-generation sequencing we have been promised that these caveats will be solved with longer reads at a lower cost (Schadt, Turner, & Kasarskis, 2010).

Related with the reduction of cost of the sequencing, both the orangutan and the gorilla consortia (Locke et al., 2011; Scally et al., 2012) incorporated the whole genome sequencing of the major species of these genera. Despite this is the common trend in the current era of genome projects, the human, mouse and chimpanzee genomes (Consortium, 2005; Lander et al., 2001; Waterston et al., 2002) did not include any additional full genome.

Following on the study of single genomes, I also studied the first whole genome of the only known albino gorilla (Section 3.2). This was not the first time that scientists have tried to unravel the genetic cause of his lack of pigmentation, back in 2000 the *TYR* gene in *Snowflake* and two other gorillas were assayed for the coding regions of this gene, without being able to find the causal mutation (Martínez-Arias et al., 2000). The main problem of this study was the lack of knowledge of albinism at that time because the gene we linked to the albino disorder was not yet associated with the albinism disorder, the first evidence came short after the *TYR* study was published (Newton et al., 2001). We have validated the mutation we found in *Snowflake*, but we could have had the same problem in our study. When we started the project only four genes were associated with albinism in humans but currently sis genes and an additional locus have been identified to cause albinism in humans (Montoliu et al., 2014). This project was also risky because we only could study a single individual with the disorder compared with two available gorilla genomes at the time. A similar project tried to unravel the genetics basis of white tigers, but in this case they had pedigrees of white tigers that they could analyse with linkage analysis through restriction-site-associated DNA sequencing (Xu et al., 2013), followed by WGS to pinpoint the concrete gene associated with their phenotype, *SLC45A2* like in *Snowflake*. It is probable that the scientific community won't devote many efforts in the whole-genome sequencing of great apes for biomedical efforts since there are other model organisms that serve better model organisms to understand the genetic basis of phenotypes, but we have been able to associate the unique pigmentation of *Snowflake* to its causal mutation. In fact, the whole genome sequencing of this gorilla was criticized in the reviewing process of the paper because the main finding on the *SLC45A2* could have been found through less expensive techniques. The reviewer was right, but at that moment this was only the fourth genome ever sequenced and we could unravel

another important characteristic on his recent origins that I discuss on (Section 4.4).

## 4.2. Great ape genetic diversity

As I have summarized in Section 1.3, the comparison of human versus great ape genetic variation has been a longstanding question to evaluate whether humans are typical apes. This has also been fundamental in the study of wild and captive populations through the study of genetic variation, mostly microsatellites and mtDNA. But the biases and problems associated with these have limited our understanding on these questions.

To overcome this limitation, the Great Ape Genome Project (Section 3.3) started as a collaborative work to sequence and analyse great ape samples from wild and captive origin, in the attempt to sample the most diverse panel of species and populations. We focused in the study of single nucleotide polymorphisms (SNPs) but we also sought for short insertions and deletions (indels) and companion studies focused in mobile element insertions (Hormozdiari et al., 2013) and large copy number variants (Sudmant et al., 2013) as well as ongoing projects studying selection and recombination in great apes. Regarding SNPs, we found extensive genetic variation in great apes that dwarves that of humans. In total, with the minimal sampling of 79 great apes, we discovered up to 80 million SNPs. To put this figure in context, the study of 1092 genomes in the *1000 Genomes Project* discovered 38 million SNPs (Abecasis et al., 2012). While this is an unfair comparison since we are comparing a very heterogeneous sampling of species that will increase the SNPs, if we just focus on *Pan troglodytes*, the sampling of 24 individuals unravels 25 million SNPs. To make this comparison less biased, in the GAGP we included nine individuals from HGDP and full genome sequenced (Meyer et al., 2012) sampled

from the most diverse regions in the planet. We compared these genomes sampling the heterozygosity of each individual and looking at the distribution within and between subspecies, showing a huge structure of diversity within species and certainly dwarfing the diversity in human populations. All but the three subspecies (western chimpanzees, bonobos and eastern lowland gorillas) show higher levels of diversity than that of humans, both Africans and non-African human populations.

These data has been fundamental to understand the genetic relationships in which these populations are structured. And not only to study wild populations of great apes but also to help to understand the genetic population structure in the zoo populations, a new metapopulation that now exists outside Africa and Southeast Asia. Through the lens of whole genomes using PCA (Patterson, Price, & Reich, 2006) and admixture (Alexander, Novembre, & Lange, 2009) approaches, we showed a clear population structure in chimpanzees, not only with respect to the *ellioti* classification (Bowden et al., 2012), but to point to subtle structure within subspecies, i.e. Gombe chimpanzees from Tanzania, presented a clear differentiation from the rest of the eastern chimpanzees and within central chimpanzees we could find signals of geographic structure within Gabon. Western lowland gorillas also showed a very interesting structure that correlated quite well with geographic origin at the country level and the only sample of Cross River gorilla presented a unique genetic background. The captive born gorillas had a composite of different genetic background belonging to the different population structures of western lowland gorillas in the wild, structure that has blended in the loss of phylogeography in captive populations. This last observation was also found in bonobos, where captive born individuals were a composite of wild born genetic structures.

Finally, this panel of diversity should become a reference for the study of natural populations. While most of the research in these species has relied in the study of few microsatellite markers or the hypervariable region of the mitochondrial genome, this set of SNPs combined with the technological advances that I have summarized in this thesis may be used to study the biology of great apes. Now more than ever a fruitful collaboration between field studies and genomic laboratories should be a starting point in studying these species.

## 4.3. Great ape evolution

The phylogeny of great apes at the species level have been fully resolved using whole genomes (Consortium, 2005; Locke et al., 2011; Prüfer et al., 2012; Scally et al., 2012). These studies have offered very important results on how these events occurred and have found extensive incomplete lineage sorting (ILS) in African great apes speciation event as well as in the Pan genus. They have proposed a Hominidae phylogeny without the ascertainment previously found in the study of a few loci and have tried to reconcile the fossil record with molecular data. (Scally et al., 2012)

Recent developments in the analysis of genomes in evolutionary biology have provided priceless tools such as CoalHMM, ILS CoalHMM (Hobolth, Christensen, Mailund, & Schierup, 2007; Mailund et al., 2012), ABC (Wegmann, Leuenberger, & Excoffier, 2009) and PSMC (H. Li & Durbin, 2011) that allow a deep understanding of the demographic and speciation histories of ancient populations. These algorithms rely on parameters such as mutation rate and generation times that can widely change the final estimates of effective population sizes and split times in these speciation events. Recent efforts have been devoted to both the estimation of the mutation rate in chimpanzee (Venn et al., 2014) and to the generation

times in great apes (Langergraber et al., 2012), studies that devote large resources in sequencing great ape trios and long-term field studies studying great apes in-situ respectively.

The main step forward of this work on great ape evolution has been the possibility to study all great ape species with uniform data and include all speciation events in the Hominoid phylogeny in a systematic way estimating all split times as well as the ancestral effective population sizes in all the speciation events. These was possible through the combination of these tools, since different timescales were only accessible to a given algorithm, but not to the others. Recent events were estimated using PSMC and ABC for the chimpanzee phylogeny while older events were analyzed using the CoalHMM framework. These data were combined in a figure that summarizes the complete phylogeny, effective population sizes in all branches, split times and divergence times are blended to provide the most comprehensive view of the Hominidae phylogeny (Section 3.3).

Of especial interest has been the reconstruction of the chimpanzee phylogeny. For the first time complete genomes from all four subspecies were sequenced and we could perform the most comprehensive study on the population history of these species. The phylogeny of the common chimpanzee has been object of debate for the last decade nicely discussed in (Stone et al., 2010), especially with respect to the last recognized subspecies, Nigeria-Cameroon chimpanzee (*P.t. ellioti*). We could find that *P.t. verus* and *P.t. ellioti* form a monophyletic clade and these diverged up to 1Mya from the central/eastern subspecies. Additional signals of gene flow were found between *schweinfurthii* and *ellioti* despite their current distributions are not adjacent. Similarly, we could also find the same signal between the Cross River gorilla and eastern lowland gorillas, both occupying similar ranges. These observations would point to a region of gene-flow between eastern and western populations through a wider range

of the tropical forest northern to the territories that these populations occupy.

Finally, we also wanted to study whether the 'less-is-more' hypothesis fitted with polymorphism data from all the extant great ape species. Through this variation we were able to pinpoint the fixed variation that have accumulated in these lineages since the common ancestor of the Hominidae family to avoid noise in the estimation of loss of function events that are still segregating in these populations. We annotated all these events throughout great ape evolution and we found a steady accumulation of loss-of-function events correlating with the speciation times in great apes. We couldn't find any excess of loss of genes in the human lineage as postulated in the hypothesis as a result of the phenotypic traits that are observed in humans (Olson, 1999). This does not imply that the loss of genes can have important roles of adaptation in a population as is observed with some examples maintained through balancing selection in the human genome, but as previously concluded (Section 1.2.1) our results indicate that this kind of evolutionary force has been the main driver in human evolution.

## 4.4. Implications for conservation

Conservation of great apes is a major concern for the scientific community (Caldecott et al., 2005). All great ape taxa are currently classified at least as endangered species and some are critically facing extinction. Mainly due to deforestation, but with several threats due to human action such as poaching; great ape populations are declining at alarming rates (Caldecott et al., 2005). Most of the work should be focused on stopping human action on these activities, but some attention has to be devoted into conservation genetics as a tool to assess the population fitness as well as to provide a way to manage captive populations or help law enforcement for illegal trading of great

apes. I have explored some of these applications in this thesis and the resources produced in this thesis are very important in the conservation genetics of great apes.

I started exploring this topic in the study of the genome of the albino gorilla. We initially discovered that the causal mutation of the albinism in *Snowflake* was found in homozygosis in the *SLC45A2* gene. Given the nature of Oculocutaneous albinism, autosomal recessive Mendelian disorder, pointed to a complex scenario of transmission. We first found that the region around this gene had no variation between the parental copies and we sought for this pattern throughout the genome. We scanned the genome seeking for these runs of homozygosity in the albino gorilla. We found that ~12% of the genome was autozygous (segments inherited through identity by descent). This level of consanguinity indicates that his parents were one of these three scenarios: grandparent–grandchild, half-siblings, or uncle–niece/aunt–nephew. Through simulations recreating the recombination rates in these different scenarios given the fact that males and females have different patterns of recombination and using the human estimates we obtained that the most probable scenario was the uncle/niece or aunt/nephew relationship. Despite this may seem a trivial finding; at the time of publication this was the first description of inbreeding in a natural population of western lowland gorillas. Long-term habitat loss and habitat fragmentation could be the underlying reason of this observation. These, combined with the phylopatric networks of male dispersals (Bradley et al., 2004) and the multiple group transfers that females experience throughout their lives (Stokes, Parnell, & Olejniczak, 2003), could result in an interconnected web of relatives that could increase the chances to permit inbreeding in wild populations. This hypothesis is in contraposition to the common inbreeding avoidance in the gorilla society through dispersal from the natal group, both males and females (Harcourt & Stewart, 2007).

After the analysis of this single individual we performed a systematic analysis of runs of homozygosity (ROH) to assess the inbreeding in natural and captive populations of all extant great ape species (Section 3.3). Through the systematic analysis of complete genomes we were able to determine the inbreeding levels on these populations. We found that both eastern lowland gorillas and diehli populations may be seriously affected, in accordance with the low population censuses in these subspecies. Additionally we could find sporadic cases of inbreeding in chimpanzees and more frequently in bonobos. The orangutan sampling could not provide any meaningful interpretation of this analysis. We also wanted to look at whether captive populations behaved differently from wild born populations. Limited by the sampling, we could compare this in western lowland gorillas and bonobos. Strikingly, we found that captive populations have a significant reduction of inbreeding in both species despite the low numbers of individuals present in captivity, the breeding programs have focused on maintaining as much diversity as possible. This study along with others (C Hvilsom et al., 2013; Nsubuga, Holzman, Chemnick, & Ryder, 2009) provides important clues in the management of captive population of great apes.

Two important outcomes from the study on the genetic variation of great apes are the capability to classify individuals into subspecies with a reduced set of ancestry informative markers (AIMs) and the ability to discriminate between subspecies at a fine-scale level, crucial in the management of captive great ape populations and determination of local populations in Africa for country determination. The resource we produced in this work provides a powerful tool in the identification of subspecies and critical in future management plans in captive populations and in reintroduction of apes. This panel of variation can also help in the identification and law enforcement in cases of illegal trading. Current technological improvements (Carpenter et al., 2013)

could also be applied in non-invasive samples that are typically used in field studies to assess population declines, inbreeding, recent gene-flow and relevant observations in great ape societies.

## 4.5. Future directions

The results I present in this thesis represent the most comprehensive study of variation in great apes, and by far the most complete set of polymorphism in these populations. We studied the most diverse set of populations where we could obtain high quality samples (invasive) to be able to fully sequence these genomes. Despite our best efforts, we could not sample all extant great ape taxa in this work. Ongoing projects are now focused on the study of isolated populations such as the mountain gorillas (*Gorilla beringei beringei*), and to provide a better sampling in orangutans that were certainly understudied in the GAGP. A major focus must also be devoted to the sequencing of great apes with known geographic origins, information that was not always available in the GAGP samples. This would provide a deeper understanding of the distribution of genetic variation within the ranges of these populations, providing a better characterization of population structure of these species.

Future studies should also be devoted in the development of sequencing techniques in non-invasive samples, the most used in the study of natural populations of great apes. Through the study of microsatellite markers and mtDNA, these low-quality samples have provided very important insights in the biology of these populations. Several problems such as DNA fragmentation, contamination and low percentage of endogenous DNA preclude them from whole genome sequencing. But recent developments of sequencing coupled with enrichment approaches (Carpenter et al., 2013; Gnirke et al., 2009) could offer the opportunity to the genomic study of non-invasive

samples. Either by targeting informative variants in the genome or by whole genome enrichment, the genomic study of non-invasive samples must be a priority in molecular ecology of great apes.

# 5. CONCLUSIONS

Through the sequencing of great ape genomes our view on the recent population and demographic history of the Hominidae family has been clarified. Despite the initial difficulties in the acquisition of these data, recent developments in DNA sequencing have provided the possibility to sequence complete genomes at an unexpected rate. In this thesis I present the result of this transition of the genomics field in the study of great apes. The initial assembly and analysis of the gorilla reference genome provided the first genomic view of the complete hominid evolution including all genera and shed light in the complex pattern of speciation of the African great apes. I have also analysed the complete genome of the only known albino gorilla, unravelling the genetic cause of his particular phenotype and reporting the first case of inbreeding in a natural population of western lowland gorillas.

I also present the whole genome sequencing of the most complete set of great apes to date with 79 individuals and covering all the species and most subspecies in the hominid family. In this work we provided the first genomic view on the population history and demography of all these genera. We reported the most complete set of variation in great apes and report an unbiased view on the diversity, population structure and levels of inbreeding in these populations. For the first time we elucidated the chimpanzee and gorilla phylogeny using whole genomes and shed light in the fluctuations of effective population sizes in these populations. Finally, we studied the 'less-is-more' hypothesis and we couldn't find support for this mechanism as the main contributor to human uniqueness. Finally, the dataset produced in this work has become the reference panel for the study of great ape variation and is a crucial resource for future studies in the

management, molecular ecology and conservation of these endangered species.

# 6. LIST OF COMMUNICATIONS

1. Scally A, Dutheil JY, Hillier LW, et al. Insights into hominid evolution from the gorilla genome sequence. Nature. 2012;483(7388):169-175.

2. Prado-Martinez J, Hernando-Herraez I, Lorente-Galdos B, et al. The genome sequencing of an albino Western lowland gorilla reveals inbreeding in the wild. BMC Genomics. 2013;14:363.

3. Prado-Martinez J, Sudmant PH, Kidd JM, et al. Great ape genetic diversity and population history. Nature. 2013;499(7459):471-5.

4. Sudmant PH, Huddleston J, Catacchio CR, et al. Evolution and diversity of copy number variation in the great ape lineage. Genome Res. 2013;23(9):1373-82.

5. Hormozdiari F, Konkel MK, Prado-Martinez J, et al. Rates and patterns of great ape retrotransposition. Proc Natl Acad Sci U S A. 2013;110(33):13457-62.

6. Hernando-Herraez I, Prado-Martinez J, Garg P, et al. Dynamics of DNA methylation in recent human and great ape evolution. PLoS Genet. 2013;9(9):e1003763.

7. Olalde I, Allentoft ME, Sánchez-Quinto F, et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature. 2014;507(7491):225-8.

# BIBLIOGRAPHY


Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., … McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews. Genetics*, (March), 15–18.

Alkan, C., Sajjadian, S., & Eichler, E. E. (2011). Limitations of next-generation genome sequence assembly. *Nature Reviews Genetics*, *8*(1), 61–65.

Altmann, A., Weber, P., Bader, D., Preuss, M., Binder, E. B., & Müller-Myhsok, B. (2012). A beginners guide to SNP calling from high-throughput DNA-sequencing data. *Human Genetics*, *131*(10), 1541–54.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., … Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, *290*(5806), 457–465.

Auton, A., Fledel-alon, A., Pfeifer, S., Venn, O., Ségurel, L., Street, T., … Mcvean, G. (2012). A Fine-Scale Chimpanzee Genetic Map from Population Sequencing. *Science*, *336*(April), 193–198.

Avise, J. C. (2009). Perspective: conservation genetics enters the genomics era. *Conservation Genetics*, *11*(2), 665–669.

Becquet, C., Patterson, N., Stone, A. C., Przeworski, M., & Reich, D. (2007). Genetic Structure of Chimpanzee Populations. *PLoS Genetics*, *3*(4), 10.

Becquet, C., & Przeworski, M. (2007). A new approach to estimate parameters of speciation models with application to apes. *Genome Research*, *17*(10), 1505–1519.

Begun, D. (2007). Handbook of Paleoanthropology, eds Henke W, Tattersall I.

Begun, D., Güleç, E., & Geraads, D. (2003). Dispersal patterns of Eurasian hominoids: implications from Turkey. *Deinsea*.

Bergl, R. a, Bradley, B. J., Nsubuga, A., & Vigilant, L. (2008). Effects of habitat fragmentation, population size and demographic history on genetic diversity: the Cross River gorilla in a comparative context. *American Journal of Primatology*, *70*(9), 848–59.

Bermejo, M., Rodríguez-teijeiro, J. D., Illera, G., Barroso, A., Vilà, C., & Walsh, P. D. (2006). Ebola Outbreak Killed 5000 Gorillas. *Science*, *314*(5805), 1564.

Bowden, R., Macfie, T. S., Myers, S., Hellenthal, G., Nerrienet, E., Ronald, E., … Mundy, N. I. (2012). Genomic Tools for Evolution and Conservation in the Chimpanzee : Pan troglodytes ellioti Is a Genetically Distinct Population. *PLoS Genetics*, *8 test*(3), 1–10.

Bradley, B. J., Doran-Sheehy, D. M., Lukas, D., Boesch, C., & Vigilant, L. (2004). Dispersed male networks in western gorillas. *Current Biology*, *14*(6), 510–513.

Bradley, B. J., Doran-Sheehy, D. M., & Vigilant, L. (2007). Potential for female kin associations in wild western gorillas despite female dispersal. *Proceedings. Biological Sciences / The Royal Society*, *274*(1622), 2179–85.

Cagliani, R., Fumagalli, M., Biasin, M., Piacentini, L., Riva, S., Pozzoli, U., … Sironi, M. (2010). Long-term balancing selection maintains trans-specific polymorphisms in the human TRIM5 gene. *Human Genetics*, *128*(6), 577–88.

Caldecott, J., Miles, L., & eds. (2005). *World Atlas of Great Apes and their Conservation*. *Environmental Conservation* (Vol. 33, p. 456). Berkeley, USA: Prepared at the UNEP World Conservation Monitoring Centre. University of California Press.

Calvignac-Spencer, S., Leendertz, S. a J., Gillespie, T. R., & Leendertz, F. H. (2012). Wild great apes as sentinels and sources of infectious disease. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *18*(6), 521–7.

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., … Bustamante, C. D. (2013). Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *American Journal of Human Genetics*, *93*(5), 852–64.

Casanovas-Vilar, I., Alba, D. M., Garcés, M., Robles, J. M., & Moyà-Solà, S. (2011). Updated chronology for the Miocene hominoid radiation in Western Eurasia. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5554–9.

Caswell, J. L., Mallick, S., Richter, D. J., Neubauer, J., Schirmer, C., Gnerre, S., & Reich, D. (2008). Analysis of chimpanzee history based on genome sequence alignments. *PLoS Genetics*, *4*(4), e1000057.

Chatterjee, H. J., Ho, S. Y. W., Barnes, I., & Groves, C. (2009). Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evolutionary Biology*, *9*(1), 259.

Consortium, C. S. and A. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, *437*(7055), 69–87.

Dennis, M. Y., Nuttle, X., Sudmant, P. H., Antonacci, F., Graves, T. A., Nefedov, M., … Eichler, E. E. (2012). Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. *Cell*, *149*(4), 912–22.

Eichler, E. E., & DeJong, P. J. (2002). Biomedical applications and studies of molecular evolution: a proposal for a primate genomic library resource. *Genome Research*, *12*(5), 673–8.

Ferris, S. D., Brown, W. M., Davidson, W. S., & Wilson, A. C. (1981). Extensive polymorphism in the mitochondrial DNA of apes. *Proceedings of the National Academy of Sciences*, *78*(10), 6319–6323.

Fischer, A., Pollack, J., Thalmann, O., Nickel, B., & Pääbo, S. (2006). Demographic history and genetic differentiation in apes. *Current Biology : CB*, *16*(11), 1133–8.

Fischer, A., Wiebe, V., Pääbo, S., & Przeworski, M. (2004). Evidence for a Complex Demographic History of Chimpanzees. *Molecular Biology and Evolution*, *21*(5), 799–808.

Fünfstück, T., Arandjelovic, M., Morgan, D. B., Sanz, C., Breuer, T., Stokes, E. J., … Vigilant, L. (2014). The genetic population structure of wild western lowland gorillas (Gorilla gorilla gorilla) living in continuous rain forest. *American Journal of Primatology*, *76*(9), 868–78.

Garner, K. J., & Ryder, O. a. (1996). Mitochondrial DNA diversity in gorillas. *Molecular Phylogenetics and Evolution*, *6*(1), 39–48.

Gibbs, S., Collard, M., & Wood, B. (2000). Soft-tissue characters in higher primate phylogenetics. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 11130–2.

Gilad, Y., Man, O., Pääbo, S., & Lancet, D. (2003). Human specific loss of olfactory receptor genes. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(6), 3324–7.

Gire, S. K., Goba, A., Andersen, K. G., Sealfon, R. S. G., Park, D. J., Kanneh, L., … Sabeti, P. C. (2014). Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*, *345*(6202), 1369–1372.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., … Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, *27*(2), 182–9.

Goecks, J., Nekrutenko, A., & Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, *11*(8), R86.

Gokcumen, O., Tischler, V., Tica, J., Zhu, Q., Iskow, R. C., Lee, E., … Korbel, J. O. (2013). Primate genome architecture in fl uences structural variation mechanisms and functional consequences.

Gonder, M. K., Disotell, T. R., & Oates, J. F. (2006). New Genetic Evidence on the Evolution of Chimpanzee Populations and Implications for Taxonomy. *International Journal of Primatology*, *27*(4), 1103–1127.

Gonder, M. K., Locatelli, S., Ghobrial, L., Mitchell, M. W., Kujawski, J. T., Lankester, F. J., … Tishkoff, S. A. (2011). Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations. *PNAS*, *108*(12), 4766–4771.

Gonder, M. K., Oates, J. F., Disotell, T. R., Forstner, M. R., Morales, J. C., & Melnick, D. J. (1997). A new west African chimpanzee subspecies? *Nature*, *388*(6640), 337.

Green, P. (1997). Against a Whole-Genome Shotgun. *Genome Res.*, *7*(5), 410–417.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, *328*(5979), 710–22.

Grehan, J. R., & Schwartz, J. H. (2009). Evolution of the second orangutan: phylogeny and biogeography of hominid origins. *Journal of Biogeography*, *36*(10), 1823–1844.

Grehan, J. R., & Schwartz, J. H. (2011). Evolution of human-ape relationships remains open for investigation. *Journal of Biogeography*, *38*(12), 2397–2404.

Greminger, M. P., Stölting, K. N., Nater, A., Goossens, B., Arora, N., Bruggmann, R., … Krützen, M. (2014). Generation of SNP datasets for orangutan population genomics using improved reduced-representation sequencing and direct comparisons of SNP calling algorithms. *BMC Genomics*, *15*, 16.

Guschanski, K., Vigilant, L., McNeilage, A., Gray, M., Kagoda, E., & Robbins, M. M. (2009). Counting elusive animals: Comparing field and genetic census of the entire mountain gorilla population of Bwindi Impenetrable National Park, Uganda. *Biological Conservation*, *142*(2), 290–300.

Hahn, Y., Jeong, S., & Lee, B. (2007). Inactivation of MOXD2 and S100A15A by exon deletion during human evolution. *Molecular Biology and Evolution*, *24*(10), 2203–12.

Harcourt, A. H., & Stewart, K. J. S. (2007). *Gorilla Society: conflict, compromise and cooperation between sexes* (Vol. 158, pp. 147–158). Chicago: The University of Chicago Press.

Harrison, T. (2010). Cenozoic Mammals of Africa, eds Werdelin L, Sanders WJ.

Hayden, E. C. (2014). Technology: The $1,000 genome. *Nature*, *507*(7492), 294–5.

Hey, J. (2010a). Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, *27*(4), 905–20.

Hey, J. (2010b). The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Molecular Biology and Evolution*, *27*(4), 921–33.

Hobolth, A., Christensen, O. F., Mailund, T., & Schierup, M. H. (2007). Genomic Relationships and Speciation Times of Human, Chimpanzee, and Gorilla Inferred from a Coalescent Hidden Markov Model. *PLoS Genetics*, *3*(2), 11.

Hormozdiari, F., Konkel, M. K., Prado-Martinez, J., Chiatante, G., Herraez, I. H., Walker, J. a, … Eichler, E. E. (2013). Rates and patterns of great ape retrotransposition. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(33), 13457–62.

Hvilsom, C., Frandsen, P., Børsting, C., Carlsen, F., Sallé, B., Simonsen, B. T., & Siegismund, H. R. (2013). Understanding geographic origins and history of admixture among chimpanzees in European zoos, with implications for future breeding programmes. *Heredity*, *110*(6), 586–93.

Hvilsom, C., Qian, Y., Bataillon, T., Li, Y., Mailund, T., Sallé, B., & Carlsen, F. (2011). Extensive X-linked adaptive evolution in central chimpanzees. *PNAS*.

IUCN 2014. (2014). *The IUCN Red List of Threatened Species. Version 2014.2. <http://www.iucnredlist.org>*.

Jobling, M., Hollox, E., Hurles, M., Kivisild, T., & Tyler-Smith, C. (2013). *Human Evolutionary Genetics, Second Edition* (p. 650). Garland Science.

Jurmain, R., Kilgore, L., Trevathan, W., & Ciochon, R. L. (2014). *Introduction to Physical Anthropology* (p. 556). Wadsworth, USA: Wadsworth, Cengage Learning.

Kaessmann, H. (1999). Extensive Nuclear DNA Sequence Diversity Among Chimpanzees. *Science*, *286*(5442), 1159–1162.

Kaessmann, H., & Pääbo, S. (2002). The genetical history of humans and the great apes. *Journal of Internal Medicine*, *251*(1), 1–18.

Kaessmann, H., Wiebe, V., Weiss, G., & Pääbo, S. (2001). Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nature Genetics*, *27*(2), 155–6.

Key, F. M., Peter, B., Dennis, M. Y., Huerta-Sánchez, E., Tang, W., Prokunina-Olsson, L., … Andrés, A. M. (2014). Selection on a Variant Associated with Improved Viral Clearance Drives Local, Adaptive Pseudogenization of Interferon Lambda 4 (IFNL4). *PLoS Genetics*, *10*(10), 12.

Kim, H. L., Igawa, T., Kawashima, A., Satta, Y., & Takahata, N. (2010). Divergence, demography and gene loss along the human lineage. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1552), 2451–7.

King, M., & Wilson, A. (1975). Evolution at two levels in humans and chimpanzees. *Science*, *188*(4184), 107–116.

Klein, J., Satta, Y., O'hUigin, C., & Takahata, N. (1993). The molecular descent of the major histocompatibility complex. *Annual Review of Immunology*, *11*, 269–95.

Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. (2013). The next-generation sequencing revolution and its impact on genomics. *Cell*, *155*(1), 27–38.

Kohn, M. H., Murphy, W. J., Ostrander, E. a, & Wayne, R. K. (2006). Genomics and conservation genetics. *Trends in Ecology & Evolution*, *21*(11), 629–37.

Köndgen, S., Kühl, H., N'Goran, P. K., Walsh, P. D., Schenk, S., Ernst, N., … Leendertz, F. H. (2008). Pandemic human viruses cause decline of endangered great apes. *Current Biology : CB*, *18*(4), 260–4.

Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*(6855), 519–23.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., … Szustakowki, J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921.

Langergraber, K. E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., … Vigilant, L. (2012). Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl Acad. Sci. USA*, *109*(39), 15716–21.

Lawlor, D. A., Ward, F. E., Ennis, P. D., Jackson, A. P., & Parham, P. (1988). HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature*, *335*(6187), 268–71.

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., … Przeworski, M. (2012). Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biology*, *10*(9), e1001388.

Leffler, E. M., Gao, Z., Pfeifer, S., Ségurel, L., Auton, A., Venn, O., … Przeworski, M. (2013). Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science (New York, N.Y.)*, *339*(6127), 1578–82.

Lehtonen, S., Sääksjärvi, I. E., Ruokolainen, K., & Tuomisto, H. (2011). Who is the closest extant cousin of humans? Total-evidence approach to hominid phylogenetics via simultaneous optimization. *Journal of Biogeography*, *38*(4), 805–808.

Lehtonen, S., Tuomisto, H., Sääksjärvi, I. E., & Ruokolainen, K. (2012). On cladistics and human-ape relationships. *Journal of Biogeography*, *39*(9), 1743–1744.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–6.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., … Wang, J. (2010). The sequence and de novo assembly of the giant panda genome. *Nature*, *463*(7279), 311–7.

Locke, D. P., Hillier, L. W., Warren, W. C., Worley, K. C., Nazareth, L. V., Muzny, D. M., … Wilson, R. K. (2011). Comparative and demographic analysis of orang-utan genomes. *Nature*, *469*(7331), 529–533.

MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., … Tyler-Smith, C. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science (New York, N.Y.)*, *335*(6070), 823–8.

Mailund, T., Halager, A. E., Westergaard, M., Dutheil, J. Y., Munch, K., Andersen, L. N., … Schierup, M. H. (2012). A New Isolation with Migration Model along Complete Genomes Infers Very Different Divergence Processes among Closely Related Great Ape Species. *PLoS Genetics*, *8*(12), e1003125.

Marques-Bonet, T., Ryder, O. a, & Eichler, E. E. (2009). Sequencing primate genomes: what have we learned? *Annual Review of Genomics and Human Genetics*, *10*, 355–86.

Martínez-Arias, R., Comas, D., Andrés, A., Abelló, M. T., Domingo-Roura, X., & Bertranpetit, J. (2000). The tyrosinase gene in gorillas and the albinism of "Snowflake". *Pigment Cell Research*, *13*(6), 467–70.

McConkey, E. H. (2000). A Primate Genome Project Deserves High Priority. *Science*, *289*(5483), 1295b–1296.

McConkey, E. H., Fouts, R., Goodman, M., Nelson, D., Penny, D., Ruvolo, M., … Wise, S. (2000). Proposal for a human genome evolution project. *Molecular Phylogenetics and Evolution*, *15*(1), 1–4.

McConkey, E. H., & Goodman, M. (1997). A human genome evolution project is needed. *Trends in Genetics*, *13*(9), 350–351.

McLean, C. Y., Reno, P. L., Pollen, A. a., Bassan, A. I., Capellini, T. D., Guenther, C., … Kingsley, D. M. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature*, *471*(7337), 216–219.

Medvedev, P., Stanciu, M., & Brudno, M. (2009). Computational methods for discovering structural variation with next-generation sequencing. *Nature Methods*, *6*(11), 13–20.

Melnick, D. J., & Hoelzer, G. A. (2005). What is mtDNA good for in the study of primate evolution? *Evolutionary Anthropology: Issues, News, and Reviews*, *2*(1), 2–10.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., … Pääbo, S. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science*, *338*(6104), 222–6.

Mitchell, M. W., & Gonder, M. K. (2013). Primate speciation: A case study of African apes. *Nature Education Knowledge*, *4*(2), 1.

Montoliu, L., Grønskov, K., Wei, A.-H., Martínez-García, M., Fernández, A., Arveiler, B., … Li, W. (2014). Increasing the complexity: new genes and new types of albinism. *Pigment Cell & Melanoma Research*, *27*(1), 11–8.

Morin, P. A., Moore, J. J., Chakraborty, R., Jin, L., Goodall, J., & Woodruff, D. S. (1994). Kin Structure , Social Selection , Evolution of the Flow , and Chimpanzees. *Science*, *265*(5176), 1193–1201.

Munch, K., Mailund, T., Dutheil, J. Y., & Schierup, M. H. (2014). A fine-scale recombination map of the human-chimpanzee ancestor reveals faster change in humans than in chimpanzees and a strong impact of GC-biased gene conversion. *Genome Research*, *24*(3), 467–74.

Newton, J. M., Cohen-Barak, O., Hagiwara, N., Gardner, J. M., Davisson, M. T., King, R. a, & Brilliant, M. H. (2001). Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of

oculocutaneous albinism, OCA4. *American Journal of Human Genetics*, *69*(5), 981–8.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Publishing Group*, *12*(6), 443–451.

Nsubuga, A. M., Holzman, J., Chemnick, L. G., & Ryder, O. a. (2009). The cryptic genetic structure of the North American captive gorilla population. *Conservation Genetics*, *11*(1), 161–172.

Oates, J. F., Groves, C. P., & Jenkins, P. D. (2009). The type locality of Pan troglodytes vellerosus (Gray, 1862), and implications for the nomenclature of West African chimpanzees. *Primates; Journal of Primatology*, *50*(1), 78–80.

Olson, M. V. (1999). MOLECULAR EVOLUTION ' 99 When Less Is More : Gene Loss as an Engine of Evolutionary Change, 18–23.

Olson, M. V, & Varki, A. (2003). Sequencing the chimpanzee genome: insights into human evolution and disease. *Nature Reviews. Genetics*, *4*(1), 20–8.

Pääbo, S. (2014). The human condition-a molecular approach. *Cell*, *157*(1), 216–26.

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., … Trajanoski, Z. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in Bioinformatics*, *15*(2), 256–78.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190.

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. a M., … Pecon-Slattery, J. (2011). A molecular phylogeny of living primates. *PLoS Genetics*, *7*(3), e1001342.

Primmer, C. R. (2009). From conservation genetics to conservation genomics. *Annals of the New York Academy of Sciences*, *1162*, 357–68.

Pruetz, J. D., & Bertolani, P. (2007). Savanna chimpanzees, Pan troglodytes verus, hunt with tools. *Current Biology : CB*, *17*(5), 412–7.

Prüfer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J. R., Walenz, B., … Pääbo, S. (2012). The bonobo genome compared with the chimpanzee and human genomes. *Nature*, *486*(7404), 527–31.

Rogers, J., & Gibbs, R. A. (2014). Comparative primate genomics: emerging patterns of genome content and dynamics. *Nature Reviews. Genetics*, *15*(5), 347–59.

Russon, A. E., Compost, A., Kuncoro, P., & Ferisa, A. (2014). Orangutan fish eating, primate aquatic fauna eating, and their implications for the origins of ancestral hominin fish eating. *Journal of Human Evolution*.

Sandler, R. L. (2012). *The ethics of species - An introduction* (First., p. 235). Cambridge: Cambridge University Press.

Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F., & Petersen, G. B. (1982). Nucleotide sequence of bacteriophage λ DNA. *Journal of Molecular Biology*, *162*(4), 729–773.

Sarich, V. M., & Wilson, A. C. (1967). Immunological Time Scale for Hominid Evolution. *Science*, *158*(3805), 1200–1203.

Sarich, V. M., & Wilson, A. C. (1967). Rates of albumin evolution in primates. *Proceedings of the National Academy of Sciences of the United States of America*, *58*(1), 142–8.

Scally, A., Dutheil, J. Y., Hillier, L. W., Jordan, G. E., Goodhead, I., Herrero, J., … Durbin, R. (2012). Insights into hominid evolution from the gorilla genome sequence. *Nature*, *483*(7388), 169–75.

Scally, A., Yngvadottir, B., Xue, Y., Ayub, Q., Durbin, R., & Tyler-Smith, C. (2013). A genome-wide survey of genetic variation in gorillas using reduced representation sequencing. *PloS One*, *8*(6), e65066.

Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A window into third-generation sequencing. *Human Molecular Genetics*, *19*(R2), R227–40.

Ségurel, L., Thompson, E. E., Flutre, T., Lovstad, J., Venkat, A., Margulis, S. W., … Przeworski, M. (2012). The ABO blood group is a trans-species polymorphism in primates. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(45), 18493–8.

Sharp, P. M., Rayner, J. C., & Hahn, B. H. (2013). Great apes and zoonoses. *Science (New York, N.Y.)*, *340*(6130), 284–6.

Stokes, E. J., Parnell, R. J., & Olejniczak, C. (2003). Female dispersal and reproductive success in wild western lowland gorillas ( Gorilla gorilla gorilla ). *Behavioral Ecology and Sociobiology*, *54*(4), 329–339.

Stone, A. C., Battistuzzi, F. U., Kubatko, L. S., Perry, G. H., Trudeau, E., Kumar, S., … Lin, H. (2010). More reliable estimates of divergence times in Pan using complete mtDNA sequences and accounting for population structure More reliable estimates of divergence times in Pan using complete mtDNA sequences and accounting for population structure. *Society*.

Stone, A. C., Griffiths, R. C., Zegura, S. L., & Hammer, M. F. (2002). High levels of Y-chromosome nucleotide diversity in the genus Pan. *PNAS*, *99*(1), 43–48.

Stoneking, M. (2009). Humans as second orangutans: sense or nonsense?: a reminder of what the molecular data indicate, and why they can't be ignored or dismissed. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *31*(10), 1010–2.

Sudmant, P. H., Huddleston, J., Catacchio, C. R., Malig, M., Hillier, L. W., Baker, C., … Eichler, E. E. (2013). Evolution and diversity of copy

number variation in the great ape lineage. *Genome Research*, *23*(9), 1373–82.

Surbeck, M., & Hohmann, G. (2008). Primate hunting by bonobos at LuiKotale, Salonga National Park. *Current Biology : CB*, *18*(19), R906–7.

Takahata, N. (1993). Relaxed natural selection in human populations during the Pleistocene. *Idengaku Zasshi*, *68*(6), 539–47.

Teixeira, J. C., de Filippo, C., Weihmann, A., Meneu, J. R., Racimo, F., Dannemann, M., … Andres, A. M. (2014a). *Long-term balancing selection in LAD1 maintains a missense trans-species polymorphism in humans, chimpanzees and bonobos*. *bio*R*xiv* (p. 006684). Cold Spring Harbor Labs Journals.

Teixeira, J. C., de Filippo, C., Weihmann, A., Meneu, J. R., Racimo, F., Dannemann, M., … Andres, A. M. (2014b). *Long-term balancing selection in LAD1 maintains a missense trans-species polymorphism in humans, chimpanzees and bonobos*. *bio*R*xiv* (p. 006684). Cold Spring Harbor Labs Journals.

Thalmann, O., Hebler, J., Poinar, H. N., Pääbo, S., & Vigilant, L. (2004). Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Molecular Ecology*, *13*(2), 321–335.

Thalmann, O., Serre, D., Hofreiter, M., Lukas, D., Eriksson, J., & Vigilant, L. (2005). Nuclear insertions help and hinder inference of the evolutionary history of gorilla mtDNA. *Molecular Ecology*, *14*(1), 179–88.

Torrents, D., Suyama, M., Zdobnov, E., & Bork, P. (2003). A genome-wide survey of human pseudogenes. *Genome Research*, *13*(12), 2559–67.

VandeBerg, J. L., Williams-Blangero, S., Dyke, B., & Rogers, J. (2000). Examining priorities for a primate genome project. *Science*, *290*(5496), 1504–5.

Varki, A. (2000). A Chimpanzee Genome Project Is a Biomedical Imperative. *Genome Research*, *10*(8), 1065–1070.

Venn, O., Turner, I., Mathieson, I., de Groot, N., Bontrop, R., & McVean, G. (2014). Strong male bias drives germline mutation in chimpanzees. *Science (New York, N.Y.)*, *344*(6189), 1272–5.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., … Zhu, X. (2001). The sequence of the human genome. *Science (New York, N.Y.)*, *291*(5507), 1304–51.

Ventura, M., Catacchio, C. R., Alkan, C., Marques-Bonet, T., Sajjadian, S., Graves, T. a, … Eichler, E. E. (2011). Gorilla genome structural variation reveals evolutionary parallelisms with chimpanzee. *Genome Research*, *21*(10), 1640–9.

Vigilant, L., & Guschanski, K. (2009). Using genetics to understand the dynamics of wild primate populations. *Primates; Journal of Primatology*, *50*(2), 105–20.

Wall, J. D. (2013). Great ape genomics. *ILAR Journal / National Research Council, Institute of Laboratory Animal Resources*, *54*(2), 82–90.

Wang, X., Grus, W. E., & Zhang, J. (2006). Gene Losses during Human Origins. *PLoS Biology*, *4*(3), e52.

Warren, K. S., Verschoor, E. J., Langenhuijzen, S., Heriyanto, Swan, R. a, Vigilant, L., & Heeney, J. L. (2001). Speciation and intrasubspecific variation of Bornean orangutans, Pongo pygmaeus pygmaeus. *Molecular Biology and Evolution*, *18*(4), 472–80.

Wasser, S. K., Shedlock, A. M., Comstock, K., Ostrander, E. A., Mutayoba, B., & Stephens, M. (2004). Assigning African elephant DNA to geographic region of origin: applications to the ivory trade. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(41), 14847–52.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., … Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915), 520–62.

Wegmann, D., Leuenberger, C., & Excoffier, L. (2009). Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*, *182*(4), 1207–18.

Wise, C. A., Rubinsztein, D. C., & Easteal, S. (1997). Comparative Nuclear and Mitochondrial Humans and Chimpanzees Genome Diversity in, *14*(7), 707–716.

Xu, X., Dong, G.-X., Hu, X.-S., Miao, L., Zhang, X.-L., Zhang, D.-L., … Luo, S.-J. (2013). The genetic basis of white tigers. *Current Biology : CB*, *23*(11), 1031–5.

Yu, N., Jensen-seaman, M. I., Chemnick, L., Kidd, J. R., Deinard, A. S., Ryder, O., … Li, W.-H. (2003). Low Nucleotide Diversity in Chimpanzees and Bonobos. *Genetics*, *164*(4), 1511–1518.

Yunis, J., & Prakash, O. (1982). The origin of man: a chromosomal pictorial legacy. *Science*, *215*(4539), 1525–1530.

Zhi, L., Karesh, W. B., Janczewski, D. N., Frazier-Taylor, H., Sajuthi, D., Gombek, F., … O'Brien, S. J. (1996). Genomic differentiation among natural populations of orang-utan (Pongo pygmaeus). *Current Biology*, *6*(10), 1326–1336.