Improving the recovery of genomic information from complex samples

Jéssica Hernández Rodríguez

TESI DOCTORAL UPF / 2018

DIRECTORS DE LA TESI

Dr. Tomàs Marquès Bonet

Dr. Ferran Casals López

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

upf. Universitat Pompeu Fabra
Barcelona

A mis padres
A Ramón
A todos los que me hacéis feliz

"Don't be afraid of hard work. Nothing worthwhile comes easily. Don't let others discourage you or tell you that you can't do it. In my day I was told women didn't go into chemistry. I saw no reason why we couldn't."
Gertrude Belle Elion

Acknowledgments

Los que me conocéis sabéis que soy una sentimental y para mi esta es la parte más bonita de la tesis, tras estos años de altos y bajos, de verdad quiero agradecer mucho a mucha gente, porque han sido los años más difíciles de mi vida por muchos motivos y gracias a toda la gente a la que aquí cito he podido llegar hasta aquí (y muchos que están entre líneas, lo siento, pero tenía que dejar alguna página para la tesis de verdad).

A todo Bioevo, porque todos hacéis que sea un lugar genial para trabajar y sois nuestra pequeña familia para los que no la tenemos a mano. Gracias a todos los PIs y estudiantes que han pasado a lo largo de estos años, ha sido un placer coincidir con todos vosotros y compartir tantos "chocolates now".

A Tomàs, gracias por transmitir ese entusiasmo por lo que haces y que contagias a todos los que estamos en tu grupo. Gracias también por haber visto algo en mi curriculum que hizo que lo que empezó como un contrato de técnico de 4 meses han acabado siendo más de 6 años. Gracias también a Ferran por la oportunidad de hacer el doctorado con los dos como directores y poner siempre las cosas tan fáciles.

A todos los tomasinos, pasados, presentes y futuros (siempre habrá un nuevo estudiante al final del PhD score), Marta, Marc D., Belén, Marcos, Javi, Irene H., Tiago, Quilez, Raquel, Irene L., Claudia, Marc de M., Lukas, Aitor, Inna, Sojung, Martin, Meritxell, Manolo, Luis, Esther, Laura, Víctor, Paula y Marina, y a todos los visitantes y “mindundis” que han dado muchas historias que contar… Gracias por tantos años de aventuras, por haberme dejado ser “la mami”, ha sido un placer no sólo tener compañeros de trabajo, sino amigos con los que compartir muchas cosas fuera del IBE; lo que más difícil se me hace de acabar la tesis es separarme de vosotros.

Irene H. y Javi, echo de menos nuestras cenitas, domingos de museos y excursiones, sois uno de los motivos por los que me quedé a hacer la tesis; aunque estéis en Cambridge seguís formando parte de esto. Marcos, por hacerme creer que podía aprender a programar, nunca seré un crack como tú, pero tu formato de script estará conmigo para siempre; sabemos que el jessidioma no es lo tuyo, pero el esfuerzo es lo que cuenta. Quilez, te hiciste de querer en el añito que estuviste en el grupo, es una alegría seguir comiendo algún que otro día juntos, por palma-palma, las patatillas…

A mis niñas, Raquel e Irene, gracias de todo corazón por ser unas amigas increíbles, por todo lo que me habéis ayudado tanto personal y profesionalmente, por hacerme feliz dentro y fuera del trabajo, por comprenderme y animarme, no hubiese podido lograrlo sin vosotras; Sois extraordinarias tanto personal como profesionalmente, estoy deseando ver el increíble futuro que os espera.
Raquel, porque, aunque no lo quieras reconocer, eres un trocito de pan, los pequeños jesscansos han sido esenciales para “sobrevivir”.
Irene, eres un cielo de persona, siempre ahí para ayudar y hacerme sentir mejor, gracias por escucharme y darme tan buenos consejos; no sabes lo que agradezco que nuestros caminos se hayan cruzado, estoy segura de que incluso en la distancia seguiremos conectadas.

Gracias por haberte leído la tesis entera, por tus comentarios y tus correcciones.
¡Sois las dos unas masters en R!, os quiero en mi vida para siempre y no sé qué haré sin vosotras cuando acabe la tesis, no sabéis todo lo que os quiero.

A Claudia, por ser mi doble, compartir tantos proyectos juntas y por ese viaje tan inesperado y divertido a Bruselas. A Marc, por tu expresividad y exageración, me he reído mucho contigo. A Manolo, porque no hay nada cómo la alegría de un andaluz, hacías más divertido el zulo. A Luis, por esos paseítos por la playa, charlas sobre todo y nada, sabes que puedes contar conmigo siempre. Sojung, it was a pleasure to shear some tea from now and then.

A los que han compartido cada mediodía con los "tomasinos", Guillem e Ignasi, gracias por tantas risas y chistes, por ir siempre más allá… Ramón y José, fue una alegría compartir esos ratitos con vosotros; a todos los que han compartido "nuestra mesa", gracias por hacer que la hora de comer haya sido uno de los mejores momentos del día.

A Judit y Mónica, por toda la ayuda, apoyo moral y profesional, por hacer que bioevo siga siendo una familia para los que estamos "de paso". A Elena y David, por confiar en mi a los pocos meses de llegar para sustituir durante un tiempo a Mónica. A Oscar Ramírez, por ser un gran amigo y gran científico, ha sido un placer haberte conocido y se te echa mucho de menos, se nota que tienes algo de granaino (por eso me caes tan bien).

A Neus, mi hermana pequeña, gracias por estar siempre, sobre todo en los momentos malos, y por mejorarlos siempre. Tu apoyo estos años ha sido vital, gracias por ser la increíble persona que eres, es imposible no quererte, haces felices a todos los que están a tu lado. Eres mi primer recuerdo del laboratorio, siempre deseando coincidir en pre-PCR, sé que formarás parte de mi vida estemos donde estemos. A Íñigo, por la boda vasco-catalana con visita guiada por Vitoria, eres un genio. Siempre recordaré nuestro épico viaje a NYC.

Al despacho 412.04, Neus, Ignasi, Alicia y Carla, gracias por tantos ratitos en el lab y por ser mi confesionario. Ali, eres la ternura en persona, gracias por el privilegio de conocerte; Ignasi, gracias por hacerme reír siempre y por las calçotadas legendarias; Neus, mil cosas más te podría decir, pero, en resumen: t'estim; Carla, por tu alegría, mails all-bioevo y por los buenos ratos esos viernes trabajando tarde junto con André, Bárbara, Simone, Ana, Neus F….

A los "Arcadianos", por compartir group meetings y retreat con "the lady of the castle", Marco, gracias por transmitir ese positivismo tanto en el lab como en el despacho; Marina, gracias por tus retos en forma de puzzle 3D; y Diego, gracias por esos últimos ánimos.

David H., although I discovered you too late, you played an important role in this thesis and in my life, I wish you and Sylvia could come back and do more board-games' nights.

Mimi and all her team in the Max Planck, Leipzig, thanks for an incredible stay, I hope to visit you again.

A CSL: Eva, Raúl y Carlos R., por los buenos momentos en el lab, se nota mucho vuestra ausencia.

A la Endometriosis, esa enfermedad que me pone a prueba y marca cada paso en mi vida, pero que ha hecho que conociera a Sonia y Toño, gracias por compartir tantos partidos y sushi. Sonia, gracias por hacerme sentir comprendida (you feel me) y hacernos terapia, es muy difícil que la gente se ponga en nuestro lugar y tenerte a mi lado estos años me hace sentir menos sola. A Alonso, porque no hay nada más divertido que tirar todos los peluches del sofá.

En Jaén, a todas mis primas segundas, Lola, Pepi, Ana, Marta, Macarena, María José, Pepa y Antonio, porque a veces la familia más lejana es la que más cerca se siente, no importa el tiempo que pase sin veros o lo lejos que estemos, vuestro cariño me llega siempre. Me encantaría teneros más cerca, porque me alegráis el corazón.

En Granada, esa ciudad con embrujo que me cautivó nada más pisarla, con sus calles encantadas, sus teterías, las tapitas, dónde conocí tantas personas a las que quiero…

A mi familia política, por ser mi familia desde que conocí a Ramón hace más de 12 años. Rosa, siempre me he sentido como en casa, gracias por quererme tanto. Gracias por todo lo que has hecho por mí estos años, incluso cuando Ramón se fue a Barcelona, no había domingo en que no me vinierais a buscar para una comida familiar. Juanjo y Ramón, gracias por estar ahí y por tantos fines de semana en Talará. A la abuela Julia, allí donde estés, gracias por quererme como a una nieta más, no he visto abuela más alegre en mi vida, un poquito de Fujitsu…

Anita, mi pequeña, una constante en mi vida para siempre, la mejor amiga y compañera de piso, incluso cuando no vivíamos juntas seguíamos viéndonos casi a diario, junto a Fer sois nuestra visita esencial, cervecita en copa helada en Los Olivos, esos cumpleaños sin fin, Faragüit…, estar a vuestro lado me hace siempre infinitamente feliz, os quiero mucho, estoy deseando conocer a la pequeña Anita.

Ching, mi Zape, Christina Yang… que ardilla eres, somos tal por igual (Anthony), fue una experiencia extraordinaria pasar dos años de prácticas contigo, ya sabes lo que te quiero y que no importa lo lejos que estemos, que cuando nos juntamos no pasa el tiempo, podemos estar horas y horas, cambios de turnos,…, no puedo expresar lo que significas para mí; junto con Botecetes (Nacho), ese viaje mágico a Ibiza (tan especial que salimos hasta en los periódicos) y luego Mallorca, nos hizo endosimbiontes para siempre. Botecetes, no ha habido una visita a Mallorca en la que no quedásemos, gracias por ser como eres y hacerme sentir tan cómoda incluso hablando de cosas tan personales; si existe la reencarnación me pido ser tú. Sois dos de mis personas favoritas; os echo infinito de menos, me hacéis pasar momentos geniales, aunque sea sólo hablando. Gracias también a Luisja y Anthony por tantos buenos momentos en el laboratorio de genética forense y fuera de él.

A Cristóbal, porque donde quiera que estemos seguimos unidos, por tantos buenos ratos durante la carrera, nunca olvidaré esas tardes junto a J escuchando vuestras canciones o haciendo el tonto. Hemos pasado muchos años viviendo incluso en países diferentes, pero siempre has estado a mi lado. Tus sueños se cumplirán, porque alguien tan grande como tú se lo merece todo.

A Eva, por un año muy divertido viviendo juntas y por acabar trabajando las dos en el PRBB, qué falta me han hecho esos "tes" contigo y los buñuelos de tu madre.

En Mallorca, esa islita que me vio nacer, de la que un día marché y espero poder volver, porque… ¡qué bien se vive en Mallorca!, dónde está mi alma, mi familia…

A mis padres, sin duda sin vosotros no hubiese llegado hasta aquí, gracias por haber trabajado y luchado para que pudiera tener todo lo que he necesitado y, sobre todo, por haber sacado matrícula de honor en paternidad. Os lo debo todo, me habéis dado todo el amor que tenéis, vuestros consejos y apoyo, aun cuando me equivocaba, me han hecho ser una luchadora, me conformaría con ser sólo la mitad de increíbles que sois vosotros. Gracias por ser tan trabajadores y haber sido capaces de conseguir todo lo que tenéis levantando ladrillo a ladrillo. Mamá, sabes lo mucho que te quiero, no sólo porque te lo digo a diario, creo que te lo demuestro también. No pasa un día sin necesitar tus consejos, no importa lo mayor que me haga o las veces que me hayas repetido la misma receta, me gusta tenerte siempre a mi lado, aunque sea al otro lado del teléfono. Me has dado fuerza para seguir adelante, has estado a mi lado en los momentos más difíciles, pero también en los más felices. Eres la mejor persona que conozco y mi mejor amiga. Papá, gracias por enseñarme a dar lo mejor de mí, por enseñarme a ser mañosa (MacGyver a nuestro lado no es nadie) y por ir a comprar ensaimadas cada vez que hacemos una visita, da igual a la hora que te tengas que ir a trabajar. Aunque hayan pasado más de 15 años desde que me fui a estudiar a Granada, no ha pasado un día sin hablar y sin echaros de menos. Ojalá existiera el teletransporte para no tener que vivir separada de vosotros. No sólo os quiero, os adoro.

A mi hermano Carlos, aunque no os pueda visitar tanto como me gustaría, gracias por darme el mejor regalo que se le puede dar a una hermana, mi sobrina Nekane y el pequeño Diego. Diego, aún no has dicho tu primera palabra, pero estoy segura de que lo pasaremos en grande. Nekane, la primera vez que te vi te sentí muy mía (dichosa genética y sus parecidos razonables). Te he echado mucho de menos estos años y he disfrutado cada minuto contigo, verte crecer es un regalo y espero poderte guiar como solo las tías pueden hacerlo.

Al resto de mi familia, abuelos, tíos, primos, … porque tras todos estos años fuera, me alegra volver a casa, que nos reunamos y ver como crece la familia.

A Marga y Juan, porque existe familia que no es de sangre. Es un placer teneros en mi vida.

Y como siempre suele decirse, por último, pero no menos importante, gracias a Ramón. Está siendo una aventura hacer el doctorado a la vez, pero no hay nada mejor que discutir nuestros proyectos y aportarnos distintos puntos de vista. Gracias por hacerme reír cada día, por saber lo que pienso sin hablar, por conocerme como nadie y darme ese abrazo cuando más lo necesito que me quita todas las penas. Gracias por compartir estos 12 años tan maravillosos, por cuidarte de todo cuando yo no me puedo ni mover, por soportar tantas cosas que me han pasado; me "compraste" defectuosa pero nunca me has querido devolver. Es un lujo compartir vida contigo, sólo con estar juntos ya nos basta, siempre serás mi peke. Vaya donde vaya siempre echaré de menos a alguien, pero sé que estando a tu lado estoy siempre en casa. Han sido unos años muy difíciles a nivel personal y profesional, pero aun así has estado al pie del cañón e incluso cometiste la locura de casarte conmigo.
Te amo, cada día un poco más.

Jéssica Hernández Rodríguez

This research was supported by a PhD fellowship from the Ministerio de Economía, Industria y Competitividad "ayudas para contratos predoctorales para la formación de doctores, convocatoria 2013".

Abstract

The innovation of genomics has progressed at pace with the development of high-throughput DNA sequencing technologies. Prior to the outbreak of next-generation sequencing, retrieving genetic information was limited to a set of molecular markers. Additionally, acquiring genetic information from complex samples, due mostly to the limitation to extract good quality and quantity of DNA from them, was especially challenging. Nonetheless, over the last decade, there has been an enormous advancement in target enrichment methodologies that permits the improvement of the quantity of DNA obtained from the sample of interest, with the consequent increase in the amount of data recovered and the reduction in sequencing costs. All these advancements have also a great value in other fields such as population genetics, evolution, medicine and conservation. In this thesis I present an experimental method for library preparation and exome target enrichment and its application to chimpanzee faecal samples. This method may be appropriate for other researchers working with complex samples and/or focused in specific parts of the genome such as certain chromosomes, the exome, a set of SNPs or even the whole-genome.

Resumen

La innovación en genómica ha progresado al mismo ritmo que lo han hecho las tecnologías de secuenciación masiva de ADN. Antes del estallido de la secuenciación de última generación, la obtención de información genética se limitaba a un conjunto de marcadores moleculares. Además, la adquisición de información genética de muestras complejas, debido principalmente a la limitación para extraer de ellos suficiente cantidad de ADN de buena calidad, fue especialmente difícil. No obstante, en la última década, se ha avanzado mucho en las metodologías de enriquecimiento de regiones objetivo que permiten mejorar la cantidad de ADN obtenida de la muestra de interés, con el consiguiente aumento en la cantidad de información recuperada y la reducción en los gastos de secuenciación. Todos estos avances tienen también una gran utilidad en otros campos, como la genética de poblaciones, la evolución, la medicina y la conservación. En esta tesis presento un método experimental para la preparación de las bibliotecas de ADN y el enriquecimiento del exoma, y su aplicación a partir de muestras fecales de chimpancé. Este método puede ser adecuado para otros investigadores que trabajan con muestras complejas y/o se centran en partes específicas del genoma, como ciertos cromosomas, el exoma, un conjunto de SNPs o incluso el genoma completo.

Preface

The great apes are our kin. Like us, they are self-aware and have cultures, tools, politics, and medicines. They can learn to use sign language, and have conversations with people and with each other Sadly, however, we have not treated them with the respect they deserve, and their numbers are now declining, the victims of logging, disease, loss of habitat, capture, and hunting.

Nevertheless, there are signs of hope. In some places, governments have taken the lead in conservation efforts, often cooperating across national frontiers. It has become increasingly clear that whoever initiates actions, be it central governments, local governments, international nongovernmental organizations, or individual citizens, local communities need to be involved. It is they who live with the great apes, and it is they who need to have the incentives - such as sharing in revenues from tourism - to conserve them.

Often, people treat great apes better when they treat each other better, as a result of education, good governance, and reduced poverty. But saving the great apes is also about saving people. By conserving the great apes, we can also protect the livelihoods of the many people who rely on forests for food, clean water, and much else. Indeed, the fate of the great apes has both practical and symbolic implications for the ability of human beings to move to a sustainable future.

Kofi A. Annan

Secretary-General of the United Nations

The work I present here pretends to summarize the relationship between human and the rest of non-human primates. Also, to evaluate the relevance of the study of primates' genetics and the applications of the genetic information obtained.

As most of the work I developed during my thesis was experimental, I have tried to explain the most employed methods, samples used, and technologies that are applied nowadays in genomics research laboratories.

Index

Abbreviations

aDNA: Ancient DNA

bp: Base pair

BWA: Burrows-Wheeler Alignment

DNA: Deoxyribonucleic acid

fDNA: Faecal DNA

GAGP: Great ape genome project

GATK: Genome analysis toolkit

Gb: Gigabase

gDNA: Genomic DNA

Kb/Kbp: Kilobase pair

Kya: Thousand years ago

Mb/Mbp: Megabase pair

mtDNA: Mitochondrial DNA

Mya: Million years ago

NGS: Next generation sequencing

NI: Non-invasive

panAf: Pan Africa programme

PCA: Principal component analysis

PCR: Polymerase chain reaction

qPCR: Quantitative polymerase chain reaction

RNA: Ribonucleic acid

SNP: Single nucleotide polymorphism

STR: Short tandem repeat

PCA: Principal component analysis

PCR: Polymerase chain reaction

WGS: Whole genome sequencing

Ya: Years ago

1. INTRODUCTION

In this section I will explain the connection among the great apes and evolutionary medicine, the use of complex samples and the new technologies applied for this sort of samples and, finally, the genetic information that can be extracted from them.

1.1. Humans as Great apes

Great apes belong to the order Primates within the Hominidae family. The great apes or hominids are composed of eight living species grouped in four genera: *Pan, Gorilla, Pongo* and *Homo*. Within each genus we find different species: *Pan*, *Pan troglodytes* (common chimpanzee), and *Pan paniscu*s (bonobo); *Gorilla*, *Gorilla gorilla* and *Gorilla beringei* (Western and Eastern gorilla); *Pongo*, *Pongo pygmaeus* and *Pongo abelii* (the Bornean and Sumatran orangutan); and *Homo*, *Homo sapiens*, the species to which modern humans belong to (Figure 1) (Herron and Freeman, 2013).

Four subspecies of common chimpanzee are commonly accepted: *Pan troglodytes ellioti* (Nigeria-Cameroon chimpanzee), *Pan troglodytes schweinfurthii* (Eastern chimpanzee), *Pan troglodytes troglodytes* (Central chimpanzee) and *Pan troglodytes verus* (Western chimpanzee); and two subspecies of western gorilla: *Gorilla gorilla diehli* (Cross river gorilla) and *Gorilla gorilla gorilla* (Western lowland gorilla) (Caldecott et al., 2005).

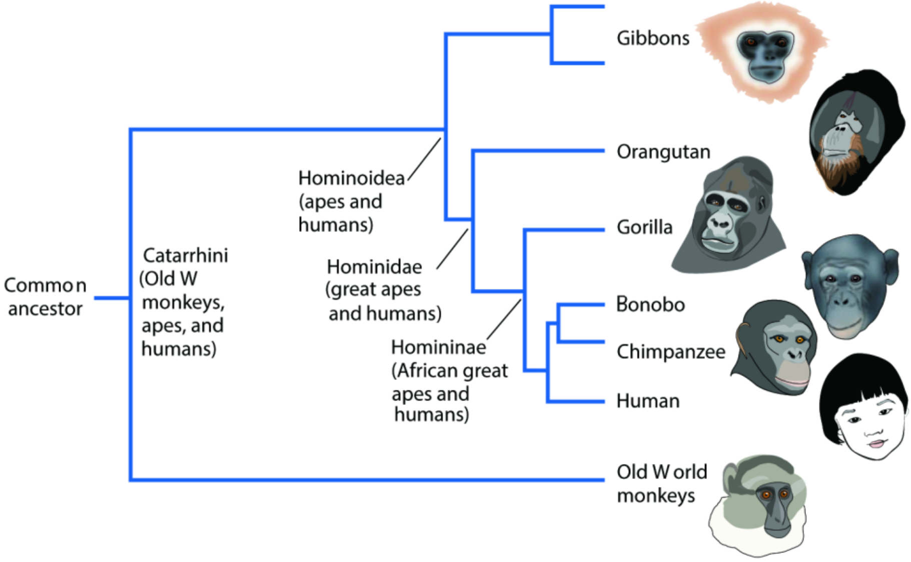

Gibbons
Orangutan
Gorilla
Bonobo
Chimpanzee
Human
Old W orld monkeys
Hominoidea (apes and humans)
Hominidae (great apes and humans)
Homininae (African great apes and humans)
Catarrhini (Old W monkeys, apes, and humans)
Commo n ancestor

Figure 1. Phylogeny of the apes. Relationships among the Old-World monkeys, represented by a rhesus monkey and the apes including humans. Among the apes, the gibbons branch off first, followed by the orangutans. The evolutionary relationships among gorillas, the two chimpanzee subspecies, and humans were long the subject of considerable dispute. (Adapted from Herron & Freeman, 2013)

Humans share with the other great apes abundant derived features (apomorphies) like the larger brain size, sexual dimorphism, a more vertical posture, similar gestation time, no external tail and y-shaped molar teeth among others (Herron and Freeman, 2013). Additionally to these evolutionary changes, the molecular analyses distinguish the great apes from the rest of the taxon Catarrhini (Goodman et al., 1998) pointing out that the apes descend from a common ancestor. There has been some controversy about the dating of the last common ancestor of great apes, a recent revision establishes it around 13 to 14 million years ago (Alba et al., 2015; Casanovas-Vilar et al., 2011; Nengo et al., 2017). When referring to the last common ancestor of

Old World Monkeys and apes (catarrhines), the two first primate species appeared as early as 25 million years ago (Figure 1) (Stevens et al., 2013). Moreover, recent studies of molecular data demostrate that primates split from other placental mammals approximately 76 million years ago (Steiper and Seiffert, 2012).

Despite the controversy with the different times to the common ancestor, which has become manifest is that humans, chimpanzees and bonobos are more tightly related to each other than with the other great apes and the rest of primates. All these evidences situate humans not only close to apes, but enclosed within the great apes.

Humans share around 98.77% of the DNA coding regions with the common chimpanzee; that is to say, the number of substitutions per synonymous site (functionally less important changes) is 1.23% for human-chimpanzee, only 1.06% or less corresponding to fixed divergence between them; indels occur in around 1.5% of the euchromatic sequence in each species (The Chimpanzee Sequencing and Analysis Consortium, 2005). On the other hand, between human and gorilla the number of nonsynonymous changes is 1.48% and 1.64% between chimpanzee and gorilla (Chen and Li, 2001; Wildman et al., 2003). These are absolutely minor differences compared for example with two species of birds, the red-eyed vireos and white-eyed vireos, who share barely 97.1% of their DNA (Rowe, 1996).

1.2. Relevance of primates' genetic research

In the last decades, there has been an immense development on the field of genetics and genomics, and these advances have also landed in the human research sector. One could ask what makes the human research so important, and the first obvious answer would be that, as humans, we are interested in the deeply understanding of our own species. From an anthropocentric-based point of view, we consider *Homo sapiens* the most important species on earth. And not only on a species level, but also focused on the human genetic variation, including the research of other species to better discern ourselves, contributing to the comprehension of genetics in general.

Another reason is the applicability of the study of human genetic variation to discover the genetic contribution to human diseases. These discoveries are a very effective instrument to understand how the different genes can contribute to the development of different diseases like Alzheimer, autoimmune diseases, cancer, or diabetes.

The third reason to focus on the study of human genetics is its relevance for evolutionary researchers. A few decades ago, the only data available for human evolution exploration was based on information about physical traits. Fortunately, in most cases, nowadays we have access to genetic data that can be integrated with palaeontology, providing a more feasible method for the studies on human evolution and, furthermore, relevant for primate studies.

All these arguments have led to the increment in the number of studies involving human genetics, by increasing the involvement of organizations and governments and their financial investment for the investigation and development of new studies.

The first populations of modern humans appeared in Africa around 200 thousand years ago (Kya) and extended out of Africa between 75 to 50 Kya, arriving to Europe almost 35 Kya; spreading afterwards to Asia and Australia (Lewis et al., 2012). Due to this recent spreading, we can considerate *Homo sapiens* as a quite young species; hence, not much time has passed for our species to acquire a huge amount of genetic variation compared to other species.

The total length of the human genome is over $3x10^9$ base pairs (bp), organized in 22 pairs of autosomal chromosomes plus two sexual chromosomes (X and Y). Although our species has a relatively young age, there has been a significant genetic variation accumulation, such that there cannot exist two individuals genetically identical (except for identical twins). This genetic variation can be measured among any two humans as around 0.1 percent, meaning that, for each 1,000 bp, there is one change. In total, two persons possess $6x10^6$ bp of difference, making these dissimilarities a considerably useful tool to analyse genetic variation.

Evolutionary studies have focused in some regions, as mitochondrial DNA (mtDNA), and/or molecular markers, the more frequently used are single nucleotide polymorphisms (SNPs) and microsatellites.

mtDNA

The DNA present inside the mitochondria is a haploid molecule with several copies per cell. mtDNA is inherited through the maternal lineage, has a high mutation rate and it is absent of recombination (Gustafsson et al., 2016). Some mtDNA genes are conserved enough to allow comparisons across species and to permit their use as markers to distinguish between closely related species.

Consequently, mtDNA has being employed to infer patterns of maternal relatedness, phylogeography and phylogenetics, and conservation among other uses (Ballard and Rand, 2005; Emery et al., 2015).

SNPs

SNPs are defined as a nucleotide change at a single site, generally found approximately every 1,000 bp (Kitts and Sherry, 2002). SNPs are located in noncoding regions, but even if they do not modify encoding proteins, they are a relevant marker for evolutionary, ecological and comparative genomics studies, as well as disease genetics and pharmacogenetics studies (Kim and Misra, 2007).

Microsatellites

Microsatellites are also known as short tandem repeats (STRs), simple sequence length polymorphisms (SSLPs), or simple sequence repeats (SSRs), and are typically composed of 1 to 6 nucleotide repeats. This short sequence of nucleotides in tandem is repeated consecutively, and can be located across all the genome. They are characterized by their high polymorphism, high frequency of distribution and co-dominance. Some of their applications include paternity and kinship studies, population genetic structure, genetic variation, and migration rates (Chistiakov et al., 2006; Launhardt et al., 1998).

1.2.1. Evolutionary medicine

As a result of the previously explained reasons, a new discipline called Evolutionary medicine or Darwinian medicine appeared. The field of evolutionary medicine consists in the application of modern evolutionary theory in order to understand health and disease. Using basic science from evolutionary biology it seeks forms of preventing and treating disease. It also explores why evolution has let the physiological and molecular mechanisms controlling health susceptible to disease. Evolutionary medicine is as well interested in the evolutionary history of humans and pathogens, focusing on preventing pathogens from becoming resistant to antibiotics, and creating new chemotherapy strategies for cancer treatment. (Gluckman, Beedle, & Hanson, 2009; Perlman, 2013; Stearns, 2012)

Nowadays, not only data from humans can be applied for studies in evolutionary medicine, since humans are primates, the shelter of wild primate populations guarantees the capacity to study the behaviour, health, physiology, ecology, genetic and sociality of our close relatives (Tung et al., 2010). Long time ago primates started to be contemplated as valuable to improve research in human health (Johnson et al., 1984; The Chimpanzee Sequencing and Analysis Consortium, 2005; Thung et al., 1981). Although in 2015, the US National Institutes of Health announced that they were not longer supporting biomedical research on chimpanzees due to ethical reasons, their physiological and genetic similarities with humans makes them relevant for the understanding of human illnesses (Bermejo et al., 2006; Formenty et al., 1999; Prado-Martinez et al., 2013; Solis-Moruno et al., 2017). Nevertheless, wild populations can be helpful for this purpose without the disadvantages of biomedical research (Walsh et al., 2017). For instance, it may contribute to the comprehension of future human diseases by analysing primate populations protected against some novel pathogens by natural immunity. However, most species of primates are endangered and close to disappearing, for this reason is their conservation crucial; their extinction would imply the loss of vital information to the survival of other species, including humans (Wich and Marshall, 2016).

1.2.2. Primate conservation

Primate conservation is a discipline focused on the preservation of non-human primates as well as the territory where they belong. As introduced in the previous paragraph, the attention paid in the last decades to primate conservation has increased considerably, a prove of that is the increment in the number of studies focused on this matter.

The IUCN Red List of Threatened Species is recognized as the most comprehensive global approach for evaluating the conservation status of animal and plant species. Through the IUCN Global Species Programme and the Species Survival Commission (SSC) the conservation status of species, subspecies, varieties, and several subpopulations has been evaluated globally for the past 50 years in order to underline the species in danger of extinction, to develop conservation programmes (International Union for Conservation of Nature and Natural Resources., 2000).

Around 200 species of primates were identified in 2000, from which, 31% were listed as threatened by IUCN; ever since, the number of primate species identified and the proportion of them classified as threatened have increased to a total number of 504 species from 79 genera, and 60% of them threatened with extinction (Estrada et al., 2017).

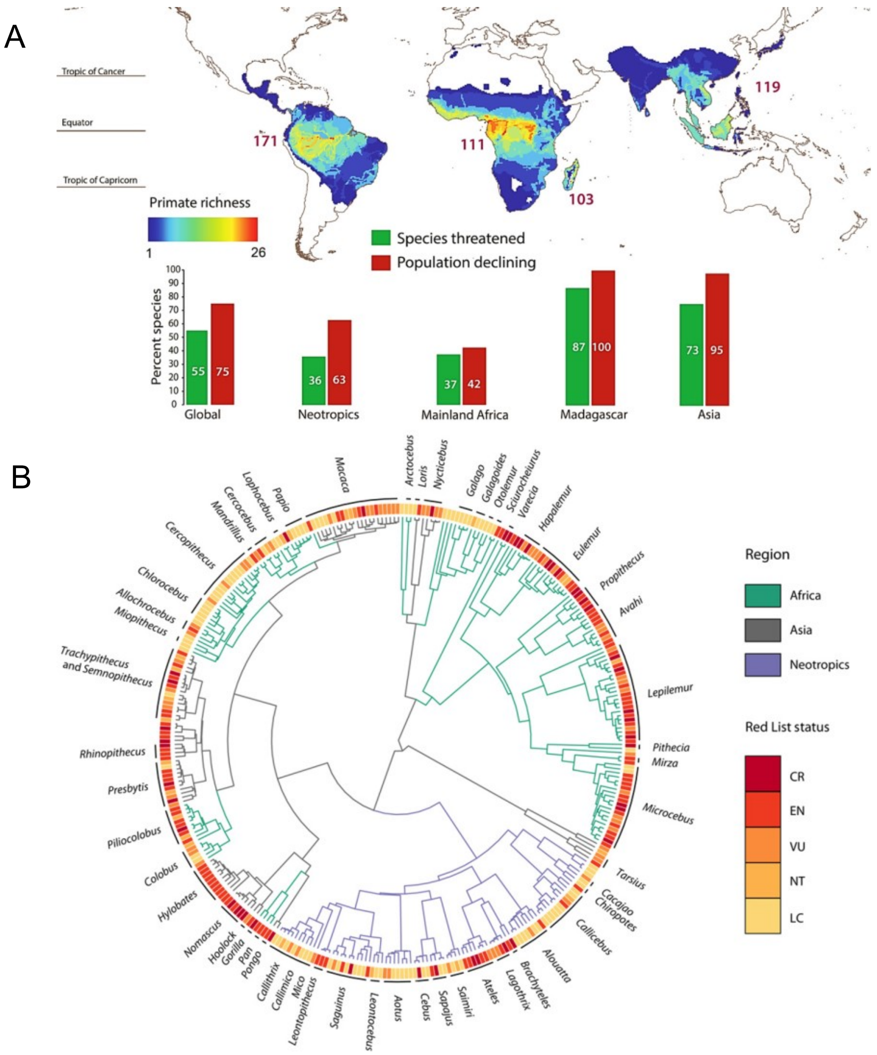


Figure 2. A) Geographic distribution of primate species. Richness and percentage of species threatened with declining populations. **B) Phylogeny from 340 primate species.** IUCN Red List Categories: CR (Critically Endangered), EN (Endangered), VU (Vulnerable), NT (Near Threatened), and LC (Least Concern). (Extracted from Estrada et al. 2017)

The increment in the number of endangered species has been explained by a set of threats, including habitat loss, disease, hunting and climate change (Figure 3).

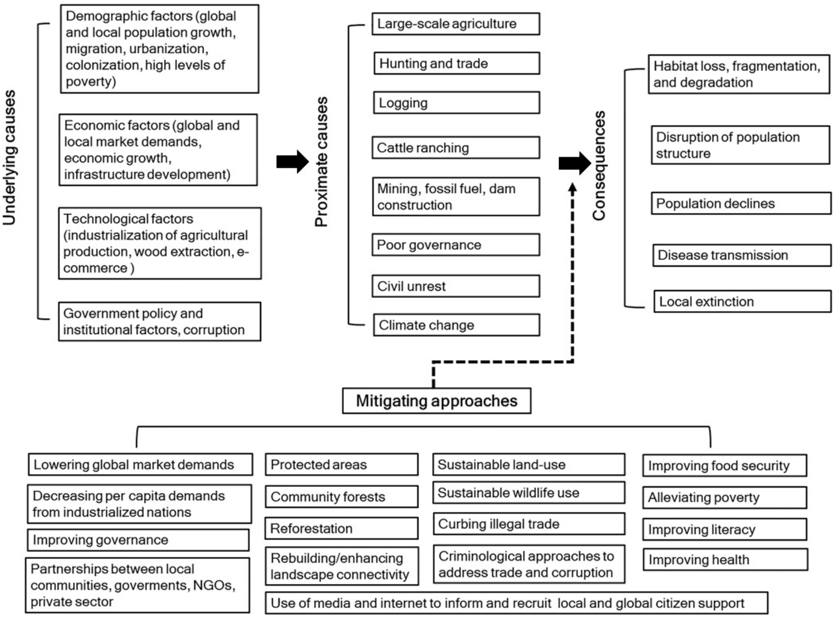

Underlying causes
Demographic factors (global and local population growth, migration, urbanization, colonization, high levels of poverty)
Economic factors (global and local market demands, economic growth, infrastructure development)
Technological factors (industrialization of agricultural production, wood extraction, e-commerce)
Government policy and institutional factors, corruption
Proximate causes
Large-scale agriculture
Hunting and trade
Logging
Cattle ranching
Mining, fossil fuel, dam construction
Poor governance
Civil unrest
Climate change
Consequences
Habitat loss, fragmentation, and degradation
Disruption of population structure
Population declines
Disease transmission
Local extinction
Mitigating approaches
Lowering global market demands
Decreasing per capita demands from industrialized nations
Improving governance
Partnerships between local communities, goverments, NGOs, private sector
Protected areas
Community forests
Reforestation
Rebuilding/enhancing landscape connectivity
Sustainable land-use
Sustainable wildlife use
Curbing illegal trade
Criminological approaches to address trade and corruption
Improving food security
Alleviating poverty
Improving literacy
Improving health
Use of media and internet to inform and recruit local and global citizen support

Figure 3. Causes of primate population declines. (Extracted from Estrada et al. 2017).

The habitat loss may be due to agriculture (for example, forest clearance for rice, sugar cane and oil palm plantations), logging, mining and fossil fuel extraction. The reduction of the territory is also produced by the construction of human transportation road networks, that aside from reducing and fragmenting their living area, produces abundant deaths because of run overs when primates cross from one location to another of their territory. Deforestation has concluded in shredding around 50% of the forests, primates have been forced to live in separate forest spots, leading to the decline in their number, loss of genetic diversity and population restructuration (Estrada et al., 2017).

Due to the phylogenetic relationship between humans and other primates there has been evidence of disease transmission by their contact through agriculture, hunting, tourism, or research. One known example would be the well documented Ebola virus outbreaks (Bermejo et al., 2006; Walsh et al., 2003), as well as the effect of the insecurity and poaching of apes by armed militias and rebel groups (Plumptre et al., 2016), producing a tragic decline of gorilla population.

Hunting takes place as a result of human bushmeat consumption, pet trade, biomedical research, zoo collections and for the sale of parts of the body (for traditional medicine, as trophies or talismans, and for magical purposes). The Convention on International Trade in Endangered Species (CITES) reported a global primate trade of around 450,000 live animals between 2005 and 2014 and another 11,000 individuals in the form of body parts (Estrada et al., 2017), demonstrating the high impact human activity imposes of great ape populations.

Lastly, climate change also affects primates. Resulting from the limitations of geographic distribution and their slow life history traits, primates are exposed to alterations on climate conditions, making them vulnerable. Ecological changes can force animals to move from their protected patches, and subsequently, leave them unprotected against hunting. Also, these changes may alter the food supplies that can derive in other consequences like competition with other species, exposure to new predators, disease and pathogens (Estrada et al., 2017; Wich and Marshall, 2016).

Besides sharing a close evolutionary and genetic history with humans, primates also participate to the cultural and biological richness of the countries where they are present. Most primates play an important role in ecosystem dynamics, affecting its function, structure, and elasticity, by their capacity to act as predator, prey, and mutualist species. They also have an impact in local and regional history and economy, and many societies protect and include primates as a key piece in their social structure.

1.3. Complex samples

Nowadays, the fields of genomics and genetics are used by a multitude of disciplines (Swenson et al., 2011; Wultsch et al., 2014, 2015). All of these disciplines are integrating new techniques to improve the datasets, by extending the data available, and that can be achieved by improving the diversity of samples used. So far, blood and other tissues have been the most used sources in genetic and population history studies (Lobon et al., 2016; de Manuel et al., 2016; Prado-Martinez et al., 2013; Rogers and Gibbs, 2014; Xue et al., 2016). Even though these types of samples have a crucial importance in molecular research, due to quality and quantity of DNA, other more complex samples have gained in importance in recent years. With these complex samples, such as non-invasive (NI) samples and ancient DNA (aDNA), we can obtain deep genomic data to be employed in genetic diversity studies of wild, living populations. With this replacement, we minimize the direct contact and interaction with the objects of study, that, as can be extracted from the previous section, may be of critical importance for primate conservation. But in exchange, other complications come into play when using these complex samples, that will be explained below.

1.3.1. Non-invasive samples

As a consequence of the endangered situation wherein many primates are involved, many researchers are concerned about the challenges presented by sample collection. The difficulty arises in finding the best way to collect samples causing the least possible disturbance to the animal but yielding a sufficient amount of good-quality DNA for the genetic analysis. By using invasive samples as blood or fresh tissue, the animal must be trapped or darted, causing several disadvantages, such as raising the individuals' stress, elevating the risk of infection, affecting their behaviour, and extremely causing the animals' death (Morin et al., 1993; Taberlet et al., 1999). Moreover, permits from CITES and other institutions are needed for the collection, transport and deliver of samples like tissue and blood between and within countries.

To avoid all these inconveniences, the fields of genetics and genomics are moving forward to the use of NI samples. The most frequent source of DNA from NI samples are faeces and hair (Goossens and Bruford, 2009; Ouborg et al., 2010; Shafer et al., 2015; Steiner et al., 2013), but we can also find studies of wild populations of primates that use other DNA sources such as saliva, urine, menstrual blood or male ejaculates (Goossens et al., 2011). Another added benefit from the use of NI samples, apart from not causing any injury to the animal, is that it does not have the limitation of using only samples collected from sanctuaries, zoos, museums, or hunted animals; NI samples from wild populations provide additional information from their geographic origin and are a better representation of the existent genetic diversity of the species (Hofreiter et al., 2003; Yu et al., 2004).

However, this type of samples presents two handicaps: the low proportion of endogenous DNA and the degree of degradation (Perry et al., 2010). For example, in the case of faecal samples, the low proportion of endogenous DNA is a result of the nature of the sample, composed by the genetic material from the individuals' own cells (endogenous DNA) and the microorganisms that constitute the gut flora as well as those that have adhered to the stool (exogenous DNA). Degradation is produced by the tropical conditions where these samples are collected, that accelerate degradation because of the warm and humid environment, and also by the effect of dung beetles and other scavengers, that decompose the excrement rapidly. In addition, the effect of the UV radiation and the degradation by the enzymes present in the faeces, influence highly the quality of the sample, reducing in this way the quality of the DNA that is going to be extracted.

Consequently, recovering enough DNA and of good quality from NI samples is a challenging task, which has led to some restrictions in their use for genetic studies and the resultant focus on targeting some regions of the genome. Numerous studies in great apes have utilized autosomal microsatellites (Fünfstück et al., 2014, 2015; Inoue et al., 2013; Kanthaswamy et al., 2006; Morin et al., 1993; Nater et al., 2013; Thalmann et al., 2007), other studies applied Y-chromosome microsatellites (Arandjelovic et al., 2011; Eriksson et al., 2006; Erler et al., 2004; Langergraber et al., 2014), the mitochondrial genome (Kawamoto et al., 2013; Thalmann et al., 2004a, 2004b) and also autosomal regions (Fischer et al., 2004, 2011; Hans et al., 2015; Thalmann et al., 2007) for genotyping NI samples.

There are different techniques for storage and DNA extraction from the different sources of NI samples (Wich and Marshall, 2016).

Hair

Samples from shed hair can be recovered from empty nests (Goossens et al., 2002; Jeffery et al., 2007; Morin et al., 1994), hanging from branches, or directly from the animal (in the case of individuals from zoos and sanctuaries). The best source of hair DNA is from pulled hair, because it preserves the roots, where mitotic cells can be found. In other cases, the hair has lost the root cells by apoptosis before shedding. When the root is present, a single hair provides enough mtDNA, and nuclear DNA; but whenever the root has been cut or is degraded only mtDNA is available for genetic analysis. For these reasons, there is another recommendation about the number of hairs to be collected, being more than 10 per individual a proposed amount (Goossens et al., 1998, 2002) on account of the variation in the completion of DNA amplification, affected by the environmental conditions under which the samples were collected and stored, and the method used for DNA extraction.

Specifically, storage is another very important aspect to consider. Generally, hair is stored in paper envelopes, dried with self-indicating silica granules, at room temperature or frozen at -80 °C (Goossens et al., 2011).

DNA extraction from hair is commonly performed with Chelex® 100 and Proteinase K (Walsh et al., 1991) , but other researches have obtained better results using taq polymerase PCR buffer as the extraction buffer (Allen et al., 1998). In another study by Jeffery et al.,(2007) (PCR buffer, Proteinase K, and H_2O in a small extraction volume) using gorilla's hair, the described method is used for shed and plucked hairs, with positive results.

Even though the use of hair over faeces has some advantages, like the lower presence of chemical inhibitors and contamination from other sources, some studies employing hair from great apes show that the amount of DNA is smaller, due to the reduced number of cells obtained from a shed-hair sample compared with a faecal sample (Broquet et al., 2006; Jeffery et al., 2007; Morin et al., 2001).

Faeces

Faecal samples are generally collected from nests, after the animals have left them, or collected directly on the ground. This last option is the only choice in non-nesting primates, which becomes a more demanding task in tropical forests where the vegetation is dense. To simplify these laborious task, the use of faeces detection dogs has been proposed (Orkin et al., 2016; Vynne et al., 2011; Wasser et al., 2004), but up until now this approach had not been tested for primates' stool (Orkin et al., 2016).

There are a few methods for the storage of faeces, depending on the diet, the species, and the habitat wherein the samples are collected (Piggott and Taylor, 2003; Waits and Paetkau, 2005). The most popular methods are the two-step, DETs (DMSO-EDTA-Tris-salt), and RNAlater (Bradley and Vigilant, 2002; Frantzen et al., 1998; Goossens et al., 2004; Nsubuga et al., 2004). The two-step method consists of a first step of desiccation with silica gel beads, followed by the freezing of the sample, and a second step of ethanol storage; this is the most reliable one to store faecal samples (Roeder et al., 2004), producing more high-quality DNA than methods where only silica gel is used.

As explained before, faecal samples are composed of a combination of the host DNA or endogenous DNA and the exogenous DNA (unabsorbed food, gut flora and enzymes). These exogenic sources may behave as chemical inhibitors that can negatively affect the succeeding reactions in DNA extraction and PCR reactions (Vallet et al., 2008). Besides the DNA level of degradation and inhibition, the quality and quantity of DNA may also diverge between species, and the ecological and environmental circumstances surrounding the sample: animals' diet, temperature, and humidity experienced by the sample since it was defecated and collected, method and time of preservation, and extraction method. The size of the faecal sample also varies between species, enabling the performance of multiple extractions from the same sample, obtaining more quantity of DNA than the amount that can be extracted from hairs. Faecal samples may as well be advantageous because not only genetic studies can result from them, but they can be used to monitor feeding patterns and microbiota (Quéméré et al., 2013; Schaumburg et al., 2013).

DNA extraction is a crucial step for this type of samples, and thus it is very important to use a method with the least possible number of steps. This way, we reduce the possibility of contamination and sample swapping. Even so, a minimal number of steps are inevitable due to the nature of these samples, to remove PCR inhibitors by several purification steps. The QIAamp Stool Mini Kit (Qiagen) protocol is one of the preferred methods used with verified results in primates (Bayes et al., 2000; Bradley et al., 2000; Eriksson J, Hohmann G, Boesch C, 2004; Goossens et al., 2004; Liu et al., 2009; Salgado-Lynn et al., 2010); there are also other methods available, as the CtAB/2PCI (Vallet et al., 2008) used in Lemuriformes (Quéméré et al., 2010a, 2010b), a standard phenol/chloroform extraction (Radespiel et al., 2008), CtAB (Lathuillière et al., 2001; Launhardt et al., 1998), diatomaceous earth (Gerloff et al., 1999), Chelex-100 (Walsh et al., 1991), and silica-based (Boom et al., 1990).

Apart from all the previous mentioned reasons to choose one method of storage and/or extraction over another, one more factor that may affect this selection is the cost of each technique. When the number of samples to be included in the study is considerable, the use of commercial kits may not be possible. Consequently, before starting a project, planning all the methods that will be applied before collecting the sample is utmost important.

Contamination is one of the most important factors to consider during collection, storage and processing of any type of sample, but it is more critical with NI samples. From collection to processing, at each of the steps where the sample is manipulated, contamination is a great concern. There are some preventing measures that can be applied: the use of latex gloves and sterile materials (pincers, tubes, bags, etc.) for handling samples in the field, changing gloves when collecting different samples, and the sterilization of mechanical tools with ethanol and fire. To limit human contamination the use of mask is recommended.

1.3.2. Ancient DNA samples

Ancient DNA samples are recovered from ancient specimens, examples include material recovered from archaeological and historical skeletons, being teeth, bones, tissues and hair the most used sources (van der Valk et al., 2017). Ancient DNA, alike DNA extracted from NI samples, is composed by endogenous and exogenous DNA. These external sources derive from the post-mortem colonization of the sample and consist of bacteria, fungi, plants and other microorganisms, and furthermore, from contamination from present-day environment, which occurs during the collection, storage, extraction or succeeding processing of the sample. The percentage of endogenous content in aDNA varies in a broad range, from samples with very low proportion (less than 5%; Green et al. 2010; Reich et al. 2010; Carpenter et al. 2013) to samples that exceed 70% (Keller et al., 2012; Miller et al., 2008; Rasmussen et al., 2010).

Due to the DNA repair mechanisms that are involved after the death of the organism, the enzymes present in the body and those present in the microorganisms that decompose the body, begin to digest immediately the biological material. Under perfect circumstances for DNA conservation, with cold, dry and low-radiation conditions, the survival of DNA is estimated to be around one million years (Millar and Lambert, 2013). The environmental conditions that can affect this survival are the presence of oxygen and humidity, temperature, pH and microorganisms present.

The consequences of degradation are the reduction of DNA length, due to breaks on single and double stranded DNA; the miscoding lesions, such as cytosine deamination; and the lesion that affects the replication of the DNA molecules, blocked by the polymerases (Dabney et al., 2013).

As a result of this DNA damage actions affecting aDNA sequences, we observe distinctive patterns. Recognizing and measuring them is a very important task when processing these samples at the laboratory and when computational analysing them. These damage patterns are, however, remarkably helpful when analysing the data, to discriminate the aDNA sequences from the contaminated sequences, which must not present these damage patterns.

As happens with NI samples, the low proportion of endogenous DNA has a huge impact on the amount of genetic information obtained by sequencing, requiring deeper sequencing and, consequently, increasing the costs of the experiments. To give an example, to obtain the draft sequence of the Neandertal genome, 1.5 billion reads were produced from samples with a percentage of endogenous content below 5% (Green et al., 2010).

1.3.3. Formalin-fixed and paraffin-embedded samples

Formalin-fixed paraffin-embedded (FFPE) tissue is the most common method of storage for tissue based molecular biological testing. In most pathology laboratories in a day-to-day base routine, all biopsies and surgical specimens are formalin-fixed and paraffin-embedded. FFPE tissue samples are extensively feasible, inexpensive in long term storage and, in many cases, are the only accessible materials for studies in retrospect. Moreover, FFPE tissue samples are also the most important material for standard diagnostics, used for some predictive and diagnostic tests in clinical routine with tumour tissues for genotyping somatic mutations. These analyses are essential to improve clinical management of cancer in the routine diagnostics, even more in the period of personalized medicine.

Degradation is also present in FFPE samples, fragmented primarily in small fragments, shorter than 300bp. Moreover, fixation with formalin produces DNA-protein crosslinks, which are not removed entirely by laboratory lysis protocols (Dietrich et al., 2013; Kuykendall and Bogdanffy, 1992). The sensitivity of DNA to mechanical stress increases with crosslinks, decreasing the accessibility for enzymes. Formalin is also oxidized to formic acid, originating DNA depurination and breaks in the DNA strand. The level of DNA degradation can be affected by various factors, such as type of fixative, composition of the fixative, duration of fixation, type of tissue and temperature.

DNA extraction method, as described in the previous complex samples, is a vital decision, affecting the final performance of DNA, by influencing the quality and quantity of DNA extracted. The first step for DNA extraction is the paraffin removal by dissolving it with xylene and it is followed by rehydration through subsequent washes with lower percentage of ethanol. Some researchers find the use of xylene time-consuming, and also, as it is a toxic chemical, prefer the use of mineral oil (Lin et al., 2009). For genomic DNA extraction there are different protocols, such as DNeasy Blood & Tissue Kit (Qiagen), Lysis buffer (10 mM Tris-HCl pH 8.0, 100 mM EDTA pH 8.0, 50 mM NaCl, 0.5% SDS and 200 μg/ml proteinase K) QIAamp DNA FFPE Tissue kit (Qiagen) and glycine-tris-ethylenediamine tetra-acetic acid buffer (100 mM glycine, 10 mM Tris-HCl - pH 8.0, 1 mM EDTA) (Ghatak et al., 2015; Lin et al., 2009; Pikor et al., 2011; Snow et al., 2014).

1.4. Target enrichment methods

Over the past years there has been an immense progression in sequencing technologies, which, along with the reduction in costs, have allowed the development of next-generation sequencing (NGS) platforms, also named second-generation or high-throughput sequencing techniques. Sequencing techniques have evolved from amplification of a single template by PCR to NGS techniques that allow parallel sequencing of millions of reads. This increase in the genetic information obtained by NGS allows the study of genetics at the genomic level compared with what was obtained previously. Some examples of these platforms are the HiSeq2500 (Illumina), obtaining 450-500 Gigabase (Gb) of data or up to 2 billion reads (2 x 125 bp) with a single flow cell, or the HiSeq4000, with 650-750 Gb or 2-2.5 billion reads (2 x 150 bp).

Whenever we apply NGS from NI samples, as mentioned in the preceding segment, there are some difficulties we have to address: the limited quality and quantity of endogenous DNA present in the sample. Resulting from these obstacles, the use of NI samples directly for high-throughput sequencing is not economically feasible. For example, some studies in great apes have proved that the amount of endogenous DNA that can be extracted from shed hair is lower than from other sources, because of the reduced number of cells (Jeffery et al., 2007; Morin et al., 2001). The same happens with faecal samples, but with the additional problematic of PCR inhibitors and contamination from exogenous sources.

However, recent target enrichment methodologies have provided methodological advances in acquiring more information from complex degraded samples like those used by zoologists and ecologists (Perry et al., 2010; Snyder-Mackler et al., 2016).

There are several target enrichment strategies available, such as hybridization-based enrichment, tagmentation, PCR-based enrichment and molecular inversion probes (Figure 4).

* Hybridization-based enrichment consists in preparing the DNA (known as library preparation) by first fragmenting it, followed by end-repair and indexed adapter ligation. Later, the DNA will be hybridized to probes complementary to the regions of interest (library preparation and hybridization further explained in Methods: 2.1.3. Library preparation, and Methods: 2.1.4. Target capture approach, pages 41-43 respectively).

* Tagmentation is based on transposon-mediated fragmentation. This strategy uses the same probes as hybridization-based enrichment, but the protocol followed to prepare the DNA is different. The fragmentation and labelling are performed in a single step using a transposase enzyme system.

* In PCR-based enrichment, by regular PCR technology, the regions of interest are amplified using primers. These amplicons are afterwards pooled in equimolar amount and prepared for sequencing through library preparation.

* Molecular inversion probes, also called MIPs, is a procedure similar to the previous approach. The probes are designed with approximately 20-nucleotide-long segment on both edges, complementary to each of the ends of the target region; these two tail segments are linked by a 40-nucleotide connection sequence. These probes hybridize to the target region and the central gap is filled-in by a ligase, creating a circle. All the non-circular DNA is removed, and this is followed by an amplification using primers complementary to part of the connection sequence that comprise indexed sample specific barcodes and other fragments necessary for the sequencing platforms (Kozarewa et al., 2015).

The application of these methods to different studies has proven to provide significant contributions to a variety of disciplines such as evolution, ecology and population genetics, and should prove invaluable to conservation efforts. The usage of target-enrichment strategies in general has been broad, focusing on the enrichment of SNPs (Perry et al., 2010), exomes (Chilamakuri et al., 2014; Guo et al., 2012), specific chromosomes and entire genomes (Carpenter et al., 2013; Snyder-Mackler et al., 2016). Some of this target enrichment methods have been developed for aDNA (Carpenter et al., 2013; Castellano et al., 2014; Fu et al., 2013; Gansauge and Meyer, 2014; Olalde et al., 2015, 2017). Due to this broad availability of capture methods, researchers use a variety of methods and technologies for obtaining genetic sequences depending on the target-regions of interest, considerations towards the specificity to their species of study, probe density, kit contents, adaptability of the method, sample multiplexing and cost (Ávila-Arcos et al., 2015).

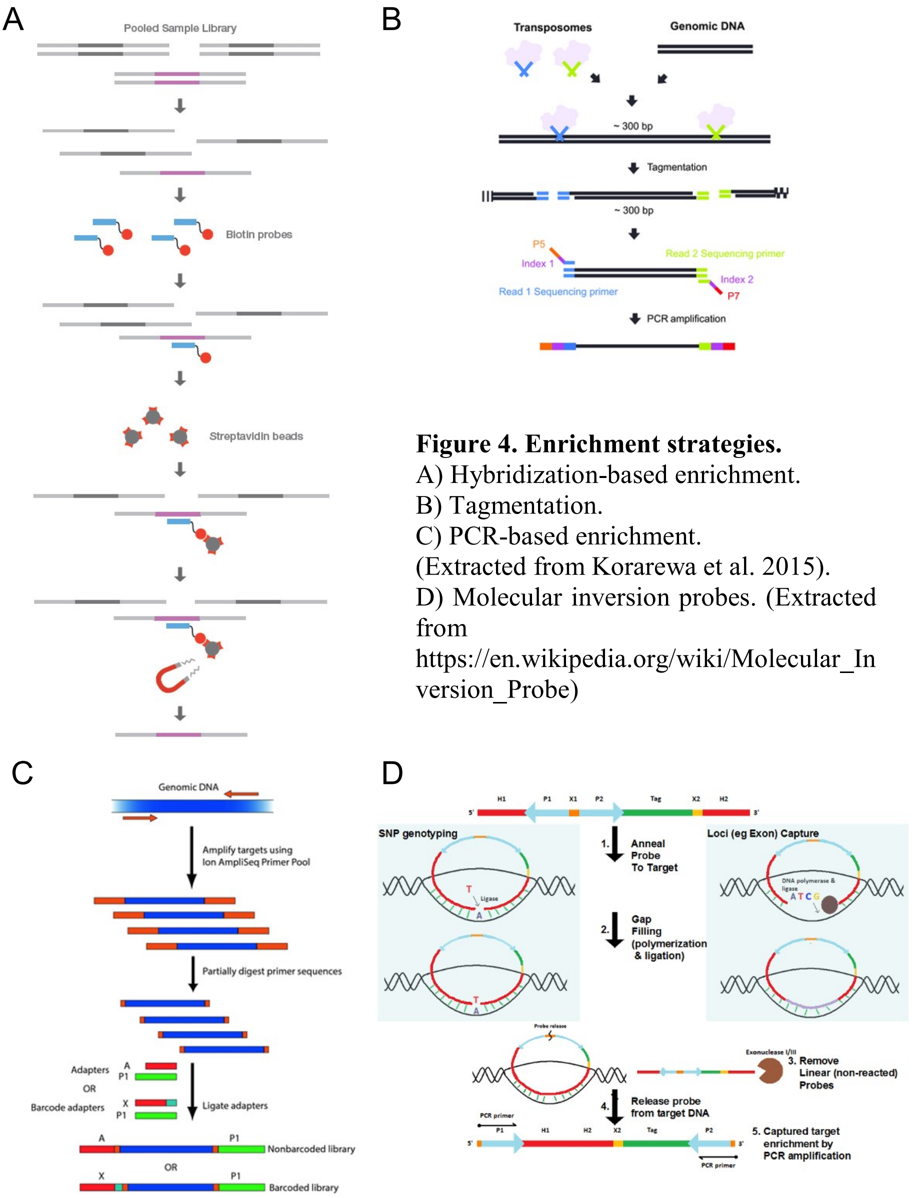
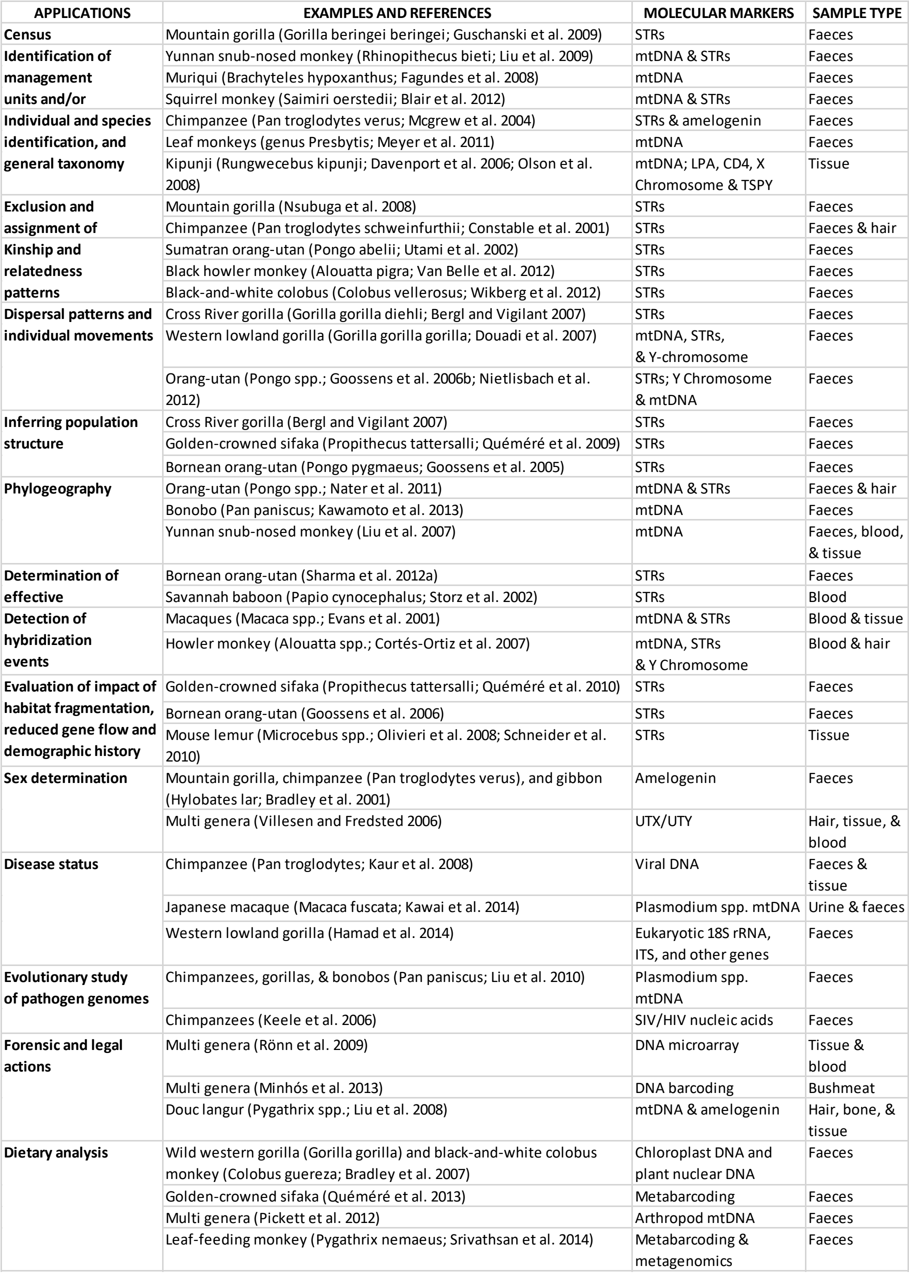


Figure 4. Enrichment strategies.
A) Hybridization-based enrichment.
B) Tagmentation.
C) PCR-based enrichment.
(Extracted from Korarewa et al. 2015).
D) Molecular inversion probes. (Extracted from https://en.wikipedia.org/wiki/Molecular_Inversion_Probe)

1.5. Application of the genetic information

The use of genetic markers has been exploited for innumerable purposes, since 2001, different molecular markers such as STRs, mtDNA, Y-chromosome and amelogenin, have been used for genetic studies. These markers can be used to describe general taxonomy, patterns of genetic differentiation and diversity, and to identify the ecological circumstances that conducted them (Liu et al., 2009; Quéméré et al., 2010b). They can be helpful for individual and species identification, sex determination, forensic and legal actions, disease status and dietary analysis, but also to estimate kinship and relatedness patterns and to determine effective population size and demographic history among others (Minhós et al., 2013; Vigilant et al., 2001; Wikberg et al., 2014) (Table 1).

Even though in the last two decades the two most commonly used markers for non-invasive genetics were microsatellites and mtDNA. Nowadays, genome-wide SNPs have become part of the genetic markers of choice, if not the most important genetic markers to study the ecology and conservation of wild populations (Wich and Marshall, 2016). Comparable to the experimental process, when planning an experiment, laboratories have to consider the amount of time and resources to bioinformatics training, data analysis, and data storage (Allendorf et al., 2010).



APPLICATIONS	EXAMPLES AND REFERENCES	MOLECULAR MARKERS	SAMPLE TYPE
Census	Mountain gorilla (Gorilla beringei beringei; Guschanski et al. 2009)	STRs	Faeces
Identification of management units and/or	Yunnan snub-nosed monkey (Rhinopithecus bieti; Liu et al. 2009)	mtDNA & STRs	Faeces
	Muriqui (Brachyteles hypoxanthus; Fagundes et al. 2008)	mtDNA	Faeces
	Squirrel monkey (Saimiri oerstedii; Blair et al. 2012)	mtDNA & STRs	Faeces
Individual and species identification, and general taxonomy	Chimpanzee (Pan troglodytes verus; Mcgrew et al. 2004)	STRs & amelogenin	Faeces
	Leaf monkeys (genus Presbytis; Meyer et al. 2011)	mtDNA	Faeces
	Kipunji (Rungwecebus kipunji; Davenport et al. 2006; Olson et al. 2008)	mtDNA; LPA, CD4, X Chromosome & TSPY	Tissue
Exclusion and assignment of	Mountain gorilla (Nsubuga et al. 2008)	STRs	Faeces
	Chimpanzee (Pan troglodytes schweinfurthii; Constable et al. 2001)	STRs	Faeces & hair
Kinship and relatedness patterns	Sumatran orang-utan (Pongo abelii; Utami et al. 2002)	STRs	Faeces
	Black howler monkey (Alouatta pigra; Van Belle et al. 2012)	STRs	Faeces
	Black-and-white colobus (Colobus vellerosus; Wikberg et al. 2012)	STRs	Faeces
Dispersal patterns and individual movements	Cross River gorilla (Gorilla gorilla diehli; Bergl and Vigilant 2007)	STRs	Faeces
	Western lowland gorilla (Gorilla gorilla gorilla; Douadi et al. 2007)	mtDNA, STRs, & Y-chromosome	Faeces
	Orang-utan (Pongo spp.; Goossens et al. 2006b; Nietlisbach et al. 2012)	STRs; Y Chromosome & mtDNA	Faeces
Inferring population structure	Cross River gorilla (Bergl and Vigilant 2007)	STRs	Faeces
	Golden-crowned sifaka (Propithecus tattersalli; Quéméré et al. 2009)	STRs	Faeces
	Bornean orang-utan (Pongo pygmaeus; Goossens et al. 2005)	STRs	Faeces
Phylogeography	Orang-utan (Pongo spp.; Nater et al. 2011)	mtDNA & STRs	Faeces & hair
	Bonobo (Pan paniscus; Kawamoto et al. 2013)	mtDNA	Faeces
	Yunnan snub-nosed monkey (Liu et al. 2007)	mtDNA	Faeces, blood, & tissue
Determination of effective	Bornean orang-utan (Sharma et al. 2012a)	STRs	Faeces
	Savannah baboon (Papio cynocephalus; Storz et al. 2002)	STRs	Blood
Detection of hybridization events	Macaques (Macaca spp.; Evans et al. 2001)	mtDNA & STRs	Blood & tissue
	Howler monkey (Alouatta spp.; Cortés-Ortiz et al. 2007)	mtDNA, STRs & Y Chromosome	Blood & hair
Evaluation of impact of habitat fragmentation, reduced gene flow and demographic history	Golden-crowned sifaka (Propithecus tattersalli; Quéméré et al. 2010)	STRs	Faeces
	Bornean orang-utan (Goossens et al. 2006)	STRs	Faeces
	Mouse lemur (Microcebus spp.; Olivieri et al. 2008; Schneider et al. 2010)	STRs	Tissue
Sex determination	Mountain gorilla, chimpanzee (Pan troglodytes verus), and gibbon (Hylobates lar; Bradley et al. 2001)	Amelogenin	Faeces
	Multi genera (Villesen and Fredsted 2006)	UTX/UTY	Hair, tissue, & blood
Disease status	Chimpanzee (Pan troglodytes; Kaur et al. 2008)	Viral DNA	Faeces & tissue
	Japanese macaque (Macaca fuscata; Kawai et al. 2014)	Plasmodium spp. mtDNA	Urine & faeces
	Western lowland gorilla (Hamad et al. 2014)	Eukaryotic 18S rRNA, ITS, and other genes	Faeces
Evolutionary study of pathogen genomes	Chimpanzees, gorillas, & bonobos (Pan paniscus; Liu et al. 2010)	Plasmodium spp. mtDNA	Faeces
	Chimpanzees (Keele et al. 2006)	SIV/HIV nucleic acids	Faeces
Forensic and legal actions	Multi genera (Rönn et al. 2009)	DNA microarray	Tissue & blood
	Multi genera (Minhós et al. 2013)	DNA barcoding	Bushmeat
	Douc langur (Pygathrix spp.; Liu et al. 2008)	mtDNA & amelogenin	Hair, bone, & tissue
Dietary analysis	Wild western gorilla (Gorilla gorilla) and black-and-white colobus monkey (Colobus guereza; Bradley et al. 2007)	Chloroplast DNA and plant nuclear DNA	Faeces
	Golden-crowned sifaka (Quéméré et al. 2013)	Metabarcoding	Faeces
	Multi genera (Pickett et al. 2012)	Arthropod mtDNA	Faeces
	Leaf-feeding monkey (Pygathrix nemaeus; Srivathsan et al. 2014)	Metabarcoding & metagenomics	Faeces

Table 1. Applications of non-invasive genetics and genomics in primatology, with a few examples from the literature. (Extracted from Wich and Marshall 2016).

Of course, there are also more difficulties when referring to genetics in primates, added to the problematic of the use of NI samples, is the lack of a reference genome for all the non-model species of primates. But besides these problems, the population genetics community has noticed the great transformation that population genetics studies have undergone with the evolution of genomic markers.

These molecular markers combined with statistical analyses can help understand primate populations and operate as a powerful tool in primate conservation, by evaluating threats and advising to wildlife managers and governing authorities on the conservation measures that can be applied in each case (Vigilant and Guschanski, 2009).

Some examples of genetic studies that can be applied to conservation genomic studies of primate species are Perry et al. (2010), that tested a DNA capture protocol on western chimpanzees (Pan troglodytes verus). The use of faecal samples provided genome-wide data and allowed the study of the species' genetic diversity. Later on, his team also extracted genomic data and performed population studies of the aye-aye (Perry et al., 2012, 2013).

With the new techniques for target enrichment, these genetic markers can also include whole chromosomes and exome. The study of exomes allows a characterization of the adaptive history of natural populations at coding sequences, while also permitting the evaluation of inter- and intra-group variation and population structure (Kidd et al., 2014). Some studies have been centred around exomes using different custom (Carpenter et al., 2013) and commercial target-enrichment methods (MYBAITS, Roche, Agilent, Illumina, etc.;

(Chilamakuri et al., 2014; Gnirke et al., 2009; Guo et al., 2012; Hodges et al., 2007; Kidd et al., 2014; Sulonen et al., 2011).

Overall, these new advances in target enrichment and sequencing technologies will allow to expand the genetic knowledge of a tremendous number of species using less common sources. Furthermore, embracing the progression of other fields such as ecology, conservation, evolution and medicine (Chiou and Bergey, 2018; Jones and Good, 2016; Kidd et al., 2014).

2. METHODS

2.1. Laboratory procedures

2.1.1. Sample collection and DNA extraction

For most of the faecal samples, the amount of DNA that belongs to the individual of study is very low (low quantity and quality of the DNA and deficient extract quality) (Taberlet et al., 1999). The endogenous DNA content is derived from the source individuals' intestinal wall epithelial cells, while the remaining DNA that can be found in the sample is a combination of the diet of the animal, the microbial flora and/or environmental contaminants (bacterial and other sources present on the ground) (Perry et al., 2010).

Due to the limitations described above, the efficiency of the collection methods, preservation of the samples and DNA extraction take on special relevance to try to meliorate as far as possible the quality and quantity of the obtained DNA (Ramón-Laca et al., 2015).

Sample Collection

The collection of the faecal samples for this thesis was performed by field teams conducting biomonitoring and ape habituation activities at Loango National Park, Gabon, and Kibale National Park, Uganda; the teams collected chimpanzee faecal samples from the ground beneath night nests and from areas where chimpanzees had defecated as they moved through the forest during the day, up to three-days-old (ape samples do not remain more than three days due to the presence of maggots, dung beetles and rain) (Arandjelovic et al., 2011).

Faeces were preserved using the two-step ethanol-silica procedure: approximately 5 g of faecal samples were conserved in 50 mL tubes containing 30 mL of 97% ethanol and mixed by inversion (faecal aliquots were placed into ethanol no later than 5 hours post-collection). After 24-36 hours the ethanol was poured off meticulously and the solid material left was shifted to new 50 mL tubes containing 25-30 mL of silica (Nsubuga et al., 2004; Roeder et al., 2004); these samples were stored in the field for up to 6 months and at 4ºC afterwards (Arandjelovic et al., 2010) (Two-step storage protocol available in Electronic Appendix, page 133).

DNA extraction

Faecal samples were extracted using the QIAmp DNA Stool kit (QIAGEN), from one month to one year after collection, with the following modifications (Figure 5) (QIAamp DNA stool mini kit protocol available in Electronic Appendix, page 135).

In the first step, around one fifth of the entire sample, 100 mg of dried sample, was mixed by vortexing at least 1 min with 1.7 mL of ASL buffer and left overnight (12-16 h) in a shaking heat block at 23ºC. This first step consists in the cell lysis, breaking the cell membranes in order to release the DNA present inside the cells. Vortexing thoroughly the mix takes on special relevance to ensure maximum DNA concentration in the final eluate. The subsequent steps followed the manufacturer’s protocol.

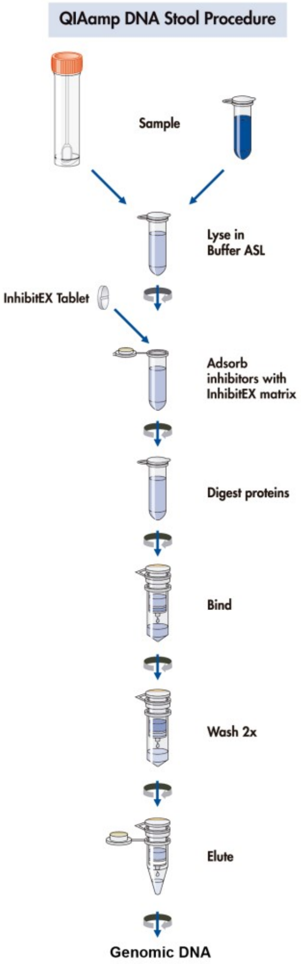


Figure 5. QIAamp DNA Stool Mini kit workflow. This kit is designed for rapid purification of total DNA from up to 220 mg stool (for both fresh and frozen samples). The fast and easy procedure comprises the following steps: lysis of stool samples in Buffer ASL, adsorption of impurities to InhibitEX matrix, and purification of DNA on QIAamp Mini spin columns. (Extracted from QIAamp DNA stool mini kit handbook, catalogue number: 51504).

The InhibitEX Tablet in the succeeding step is used to absorb the PCR inhibitors that can degrade DNA and inhibit enzymatic reactions. Once the DNA has been liberated and protected from the PCR inhibitors, the DNA must also be protected by the digestion of the proteins present in the mix, like DNases that degrade DNA, with the addition of Proteinase K. The addition of Buffer AL compromises the membrane integrity of the cell and improves the binding of the

DNA to the spin column. Lastly, the DNA purification is performed by adding ethanol; being the DNA insoluble in alcohols, such as ethanol or isopropanol, this step will produce DNA precipitation and agglomeration; the ethanol also contributes to the binding of the DNA to the spin column and washes the salts off the membrane.
Afterward, the Buffers AW1 and AW2 are added to wash all the components present in the lysate except the DNA.
At the final step, after adding 200 µL of Buffer AE, a 30 min incubation was included in the protocol and finished with a 2 min centrifugation. The purpose of this step is to elute the DNA present in the spin column to collect and store it; the additional incubation is performed to increase DNA yields by permitting a proper dilution of the DNA from the column, and the longer centrifugation to ensure that all the DNA has eluted.

Quantification of the faecal DNA (fDNA) obtained was performed by a quantitative PCR (qPCR) (Figure 6) (Morin et al., 2001) (qPCR protocol available in Electronic Appendix, page 138).

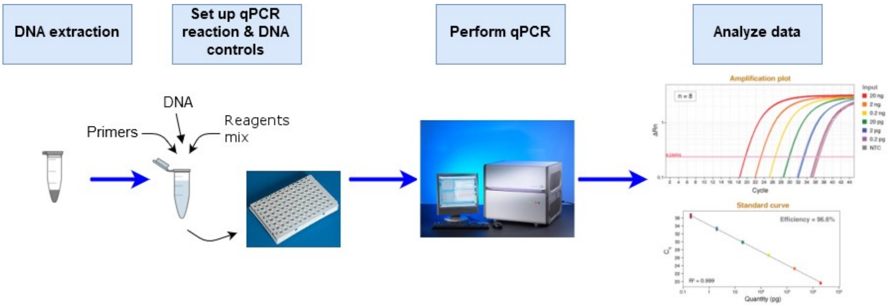


Figure 6. Quantitative PCR (qPCR) workflow. After DNA extraction the sample is dispensed to a plate and analysed in the qPCR machine.

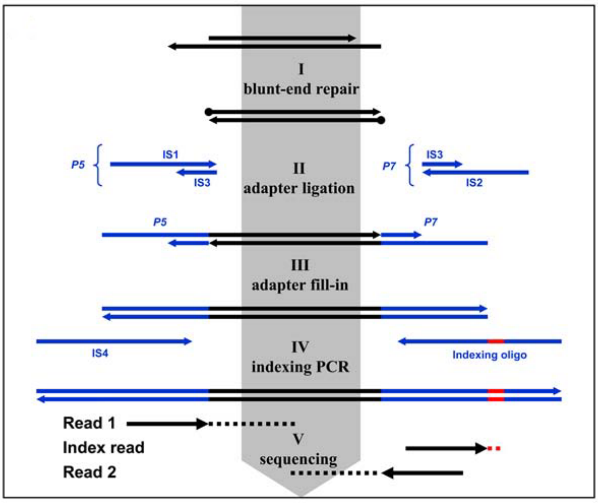
2.1.2. Sample quality control, endogenous content and level of degradation

As mentioned before, the percentage of endogenous DNA in a faecal sample is unpredictable and shows extreme variation within samples even from the same individual. For this reason, it is decisive to accomplish certain analysis to determine the quality of the sample and to detect the presence of PCR inhibitors that will interfere with the following lab procedures.

To confirm that the samples collected were of chimpanzee origin and not misidentified (gorilla faecal remains), a set of putative chimpanzee genotypes and 13 genetically identified gorilla genotypes from the study site (Arandjelovic et al., 2010) were incorporated in a STRUCTURE 2.1 Bayesian model-based clustering program analysis (Pritchard et al., 2000). The genotype results were considered representative of sample quality and absence of PCR inhibitors.

Endogenous content was predicted by qPCR and low-level shotgun sequencing of sample libraries, as reported in Meyer & Kircher, 2010. For shotgun sequencing, libraries were prepared following the in-house library preparation protocol published by Meyer & Kircher, 2010 (Figure 7) (Library preparation for shotgun sequencing protocol available in Electronic Appendix, page 144).

Figure 7. Double-stranded library preparation steps. Blunt-end repair, adapter ligation, adapter fill-in and indexing PCR. (Extracted from Meyer & Kircher, 2010, http://cshprotocols.cshlp.org/content/2010/6/pdb.prot5448.full)



The level of degradation was estimated by measuring the length distribution of DNA molecules, extracting the average fragment length. This measure was quantified by running samples on a Fragment analyzerTM (Automated CE System 96 capillary, Advanced Analytical Technologies, Inc.), an automated system for the quantification and qualification of NGS libraries, gDNA and RNA (Figure 8). For an accurate quantification, qualification, and sizing of genomic DNA we used the High Sensitivity Genomic DNA Analysis kit, following the manufacturer's protocol (Cat. Number

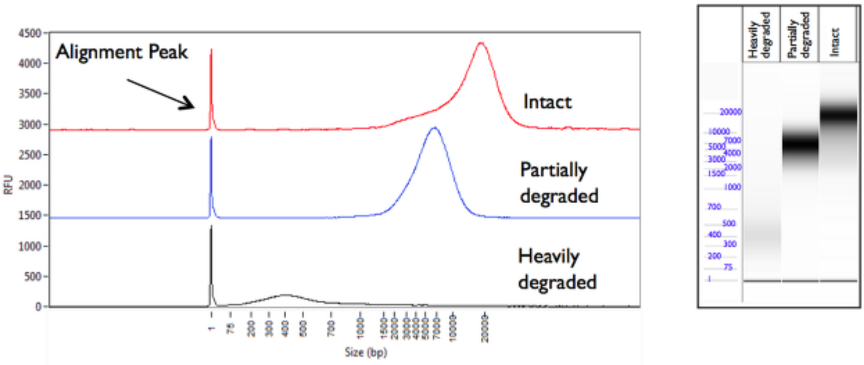


Figure 8. Quantification and qualification with Fragment Analyzer (AATI). **(**Extracted from http://geneer.tistory.com/category/Fragment%20Analyzer)

DNF-488) (High sensitivity genomic DNA analysis kit protocol available in Electronic Appendix, page 160). To assess the gDNA quality we used PROSize® Data Analysis Software. The smear analysis function in PROSize® reports the size (bp) and concentration of the gDNA smears.

2.1.3. Library preparation

Through the last 10-15 years, sequencing technologies have been broadly used by scientists, evolving and improving so fast that what was called next-generation sequencing has been converted into second-generation sequencing (SGS). At the same time, this new sequencing boost has led to the emergence of methods for preparing nucleic acids, as these NGS technologies require some previous modifications of the DNA before sequencing. This preparation of nucleic acids consists in attaching adapters (DNA fragments produced artificially) at the 3' and 5' ends of the DNA to be afterwards recognized by the sequencer; this method is known as library preparation.

DNA was sheared using a Covaris S2 focused ultrasonicator to 200 bp fragments to prepare it for library preparation, with the following settings: intensity 5, duty cycle 10%, cycles per burst 200, treatment time 120 s, temperature 7°C and water level 12.

In this work, we have used KAPA Library preparation kit (Cat. Number 07137923001) with slight modifications (Figure 9) (SeqCap EZ Library SR protocol available in Electronic Appendix, page 163).

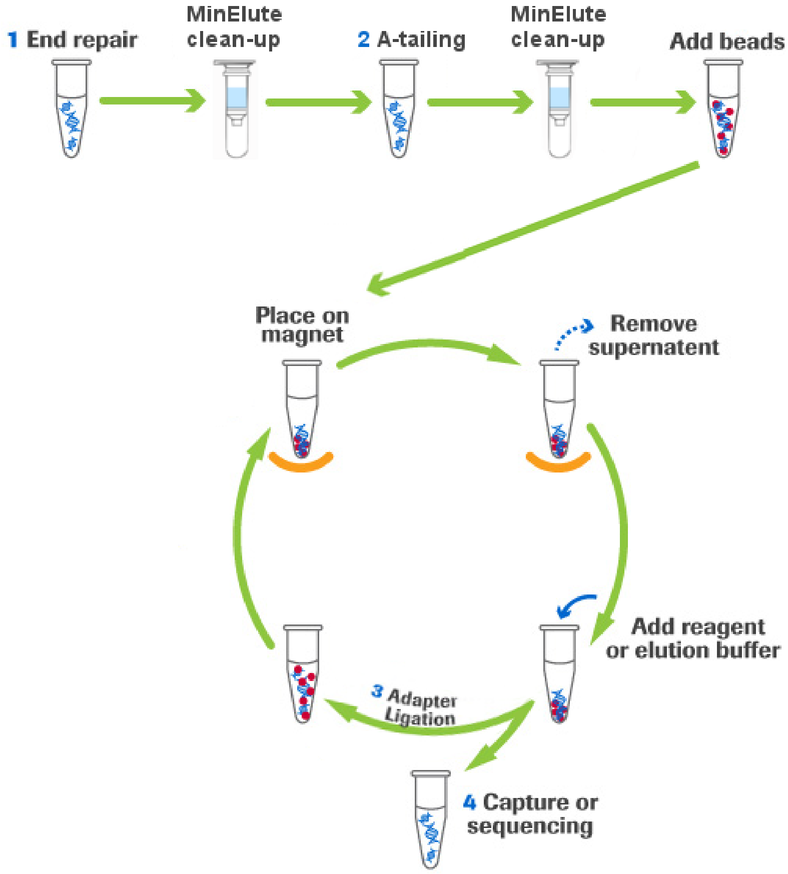


Figure 9. KAPA Library preparation workflow. Clean-ups after the steps 1 and 2 were modified, we used MinElute spin columns instead of SPRI-beads to recover smaller fragments of DNA. (Adapted from http://sequencing.roche.com/en/products-solutions/by-category/library-preparation/dna-library-preparation/kapa-htp-ltp.html)

The amount of starting material used was 40 µL in Experiment 1 and 20 µL for samples in Experiment 2 (variation of DNA concentration was between 1.31 and 4.03 µg). The reaction clean-ups for the end-repair and A-tailing were performed with MinElute Reaction clean-up spin columns, eluting in 20 µL of elution buffer (Cat. Number 28206) instead of Agencourt AMPure XP beads (solid-phase reversible immobilization (SPRI) paramagnetic bead technology). The reason for this change in the clean-ups was to retain molecules smaller than 100 bp. SPRI-beads retain molecules down to 100 bp and MinElute spins columns can retain molecules down to 50 bp, which is the expected size of most of the endogenous DNA present in the degraded samples. After the ligation reaction, the first bead clean-up was performed using 90 µL of Agencourt AMPure XP beads, the following steps were performed following the standard protocol. Finally, at the library amplification step, we used the pre-capture LM-PCR program with a total of 12 cycles.

2.1.4. Targeted capture approach

After the DNA has been processed in library preparation it can be sequenced straight away, a procedure called shotgun sequencing. This method has been used to study whole genomes, but also, as explained in a previous section, to obtain information about the endogenous content present in the sample as well as its degree of degradation. However, due to the low proportion of endogenous content present in non-invasive samples, and even though the sequencing costs have decreased in the last decade, the amount of sequencing required to obtain enough information is expensive and can not be afforded by a great majority of researchers.

Nowadays, thanks to the advances in target enrichment methodologies, this difficulty has been diminished. These targeted methodologies increase the proportion of endogenous DNA present in the library, acquiring more information from NI samples (Perry et al., 2010; Snyder-Mackler et al., 2016; Wall et al., 2016).

This target enrichment consists in the use of biotinylated RNA or DNA baits designed to hybridize complementarily with the DNA from the species of interest, which are pulled down with magnetic streptavidin-coated beads; these DNA libraries/biotinylated baits/streptavidin-coated beads are captured by a magnet, and after subsequent washes, these DNA fragments are isolated, eluted, and can be sequenced (Figure 10).

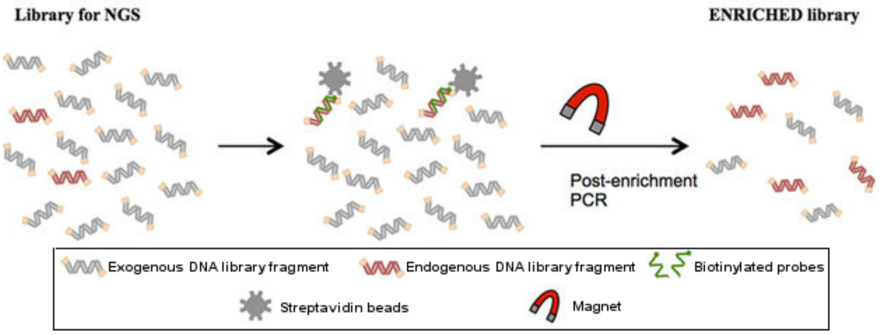


Figure 10. Target enrichment procedure. DNA libraries are hybridized with the biotinylated probes that bind to streptavidin-coated beads, after capturing them with a magnet and performing a PCR, a library enriched in endogenous DNA is obtained. (Adapted from Miyazato et al. 2016, https://doi.org/10.1186/s12864-016-2836-6).

For this thesis, we used Nimblegen baits (Roche) for the chimpanzee exome (57.5 Mb) designed using the panTro4 assembly (SeqCap EZ Developer Library, Cat Number 06740278001). We followed the manufacturer's protocol for the hybridization, except for the starting amount of Multiplex DNA sample library pool, adding 1.5 µg and 0.24 µg from Experiment 1 and 2 respectively, instead of 1 µg as the protocol suggested. This modification was introduced due to the number of samples included in each experiment. We also modified the hybridization time to 36 hours and reduced the number of PCR cycles to 12 in the post-capture PCR amplification. We increased the hours of hybridization to allow all baits to bind to their complementary DNA fragment. The number of PCR cycles was reduced to avoid raising the number of PCR duplicates sequenced, and also, because in some aliquots we performed a second round of capture and a second post-capture PCR amplification. This second PCR amplification increases exponentially the number of PCR duplicates sequenced, and ultimately, reduce the information obtained from our libraries. The second round of capture was performed following the same protocol used in the first hybridization, only altering the amount of starting material, using the entire volume obtained from the first hybridization. We reduced the number of PCR cycles also for this second post-capture PCR amplification, for the reasons explained beforehand (Figure 11).

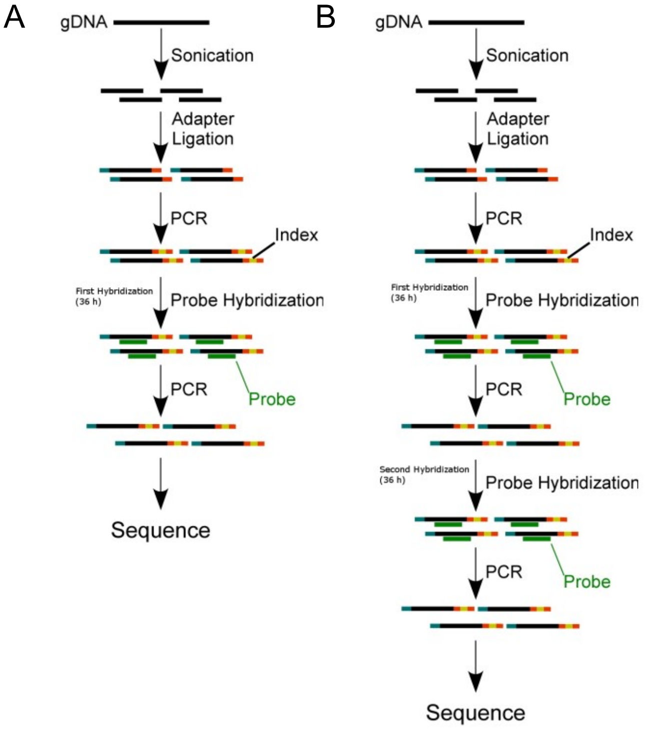


Figure 11. Workflow for SeqCap EZ Library preparation and probe hybridization experiments. A) One hybridization protocol. B) Two hybridizations protocol. (Adapted from Samorodnitsky et al. 2015, https://doi.org/10.1016/j.jmoldx.2014.09.009)

2.1.5. Sequencing

The high cost of traditional sequencing methods, such as Sanger sequencing, lead to the advancement in early 2005 of novel technologies that intended to extensively increase sequencing throughput and to reduce cost. These methods, usually known as NGS, entail the fragmentation of genomic DNA to shear it into ~200–500 bp fragments, and the succeeding immobilization of spatially separated template DNA fragments on a solid superficies previous to sequencing. These characteristics permit sequencing of millions or billions of fragments at the same time, for this reason, are often referred as massively parallel sequencing methods.

Once libraries are attached to the flow-cell surface, bridge amplification generates spatially separated clusters that contain around 1,000 identical molecules (Figure 12). This amplification is performed because the fluorescence detection method is not sensitive enough to recognize single-molecule fluorescence. After the annealing of the sequencing primers to each cluster template, a DNA polymerase incorporates the fluorescently labelled nucleotide/base that is complementary to the base present in the DNA fragment. The bases are added one at a time, with a 3-prime block to avoid the addition of a new base until the signal from the previous one has been detected. The nucleotides that have not been incorporated are washed away. Each of the four bases has a unique emission fluorescence that is recorded by imaging to determine the identity of the nucleotide that has been incorporated. The fluorescence and the 3-prime block are removed, allowing the enlargement of the next base. This process is repeated: adding just one base in each DNA fragment, detecting the signal, eliminating the block and the fluorescence, and washing. This cyclical process continues, and it is known as reversible terminator chemistry. It allows the addition of all 4 nucleotides in one reaction and gives regular read lengths of 100 bp, but depending on the machine and the kit used it can vary from 50 to 300 bp (Jobling et al., 2013).

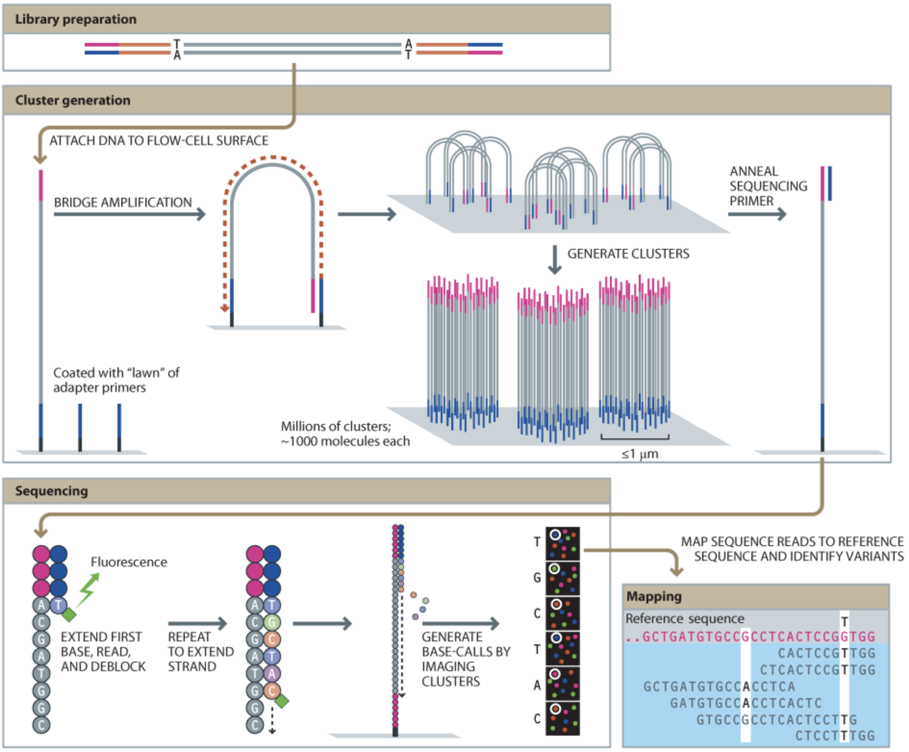

Library preparation
T
A
A
T
Cluster generation
ATTACH DNA TO FLOW-CELL SURFACE
BRIDGE AMPLIFICATION
Coated with "lawn" of adapter primers
GENERATE CLUSTERS
Millions of clusters; ~1000 molecules each
≤1 μm
ANNEAL SEQUENCING PRIMER
Sequencing
Fluorescence
A T C G A T G G C
EXTEND FIRST BASE, READ, AND DEBLOCK
REPEAT TO EXTEND STRAND
A T C G G C A T T A G C G C
GENERATE BASE-CALLS BY IMAGING CLUSTERS
T
G
C
T
A
C
MAP SEQUENCE READS TO REFERENCE SEQUENCE AND IDENTIFY VARIANTS
Mapping
Reference sequence
T
..GCTGATGTGCCGCCTCACTCCGGTGG
CACTCCGTTGG
CTCACTCCGTTGG
GCTGATGTGCCACCTCA
GATGTGCCACCTCACTC
GTGCCGCCTCACTCCTTG
CTCCTTTGG

Figure 12. Illumina Sequencing platform. (Adapted from Jobling, Hollox, Kivisild, &amp; Tyler-Smith, 2013).

The development of these high-throughput DNA sequencing technologies has enabled the whole-genome sequencing of different species' genomes. Certainly, one of the most important advances in genetics was the achievement of the Human Genome Project

(International Human Genome Sequencing Consortium, 2001; Venter et al., 2001): the sequencing of the entire human genome, conformed by approximately 21,000 protein-coding genes; this project began in 1990 and was completed in 2003.

From then on, whole-genome of hundreds of species have been sequenced, including mice (Mouse Genome Sequencing Consortium, 2002), chimpanzees (The Chimpanzee Sequencing and Analysis Consortium, 2005), western lowland gorillas (Scally et al., 2012), orangutans (Locke et al., 2011), bonobos (Prüfer et al., 2012), and rhesus macaques (Rhesus Macaque Genome Sequencing and Analysis Consortium, 2007). In May 2010, researchers finished sequencing the entire Neandertal genome (Green et al., 2010).(Extracted from Jurmain, Kilgore, Trevathan, & Ciochon, 2017).

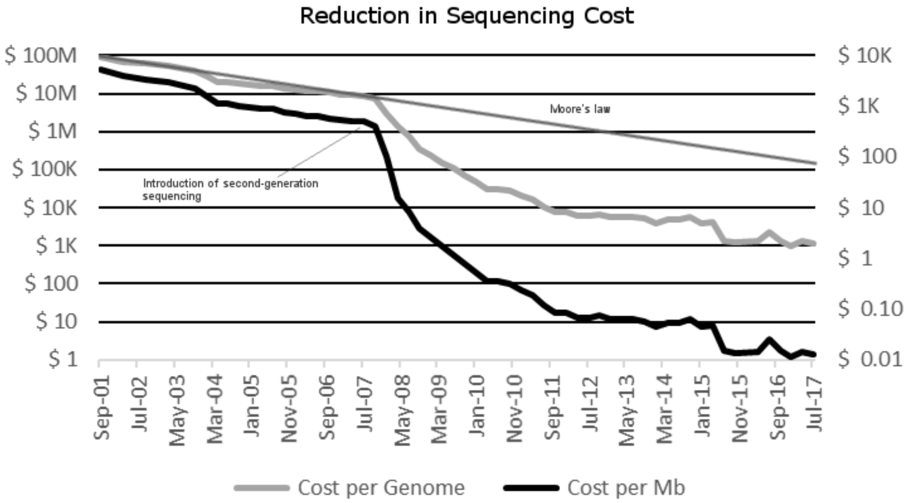


Figure 13. Dramatic reduction in sequencing cost. Moore's law is the observation that over the history of computing hardware, the number of transistors in an integrated circuit doubles approximately every 2 years. Since 2001, genome sequencing costs are decreasing at a rate that outpaces Moore's law. (Adapted from https://www.genome.gov/sequencingcostsdata/; National Human Genome Research Institute (NHGRI))

During the past decade there has been an incredible decline in the cost of DNA sequencing, since the development of NGS (Figure 13). This decline is frequently compared to Moore's law, which describes a long-term trend whereby computing power doubles every two years. In fact, the decay of the sequencing cost per megabase declines faster than the predictions from Moore's law.

This decline has resulted in the achievement of the target price of $1,000 per genome, which has been largely discussed to be the goal, not forgetting the quality and coverage to be reached (Jobling et al., 2013). The affordable sequencing opens the feasibility of studying the genomes of many species and individuals, making a great impact in science and biomedical research.

2.2. Bioinformatic processing

As I have already pointed out, the strategies for generating NGS data are broadly developed and are, day after day, becoming easier to produce and more affordable. After the data from NGS has been generated, the next step is to storage and process it to extract significant information, which is becoming an arduous task due to the high amount of data originated. Depending on the data used, the information that needs to be extracted, the programs used, etc., the pipeline followed may change. In this thesis we proceed with the pipeline represented in Figure 14.

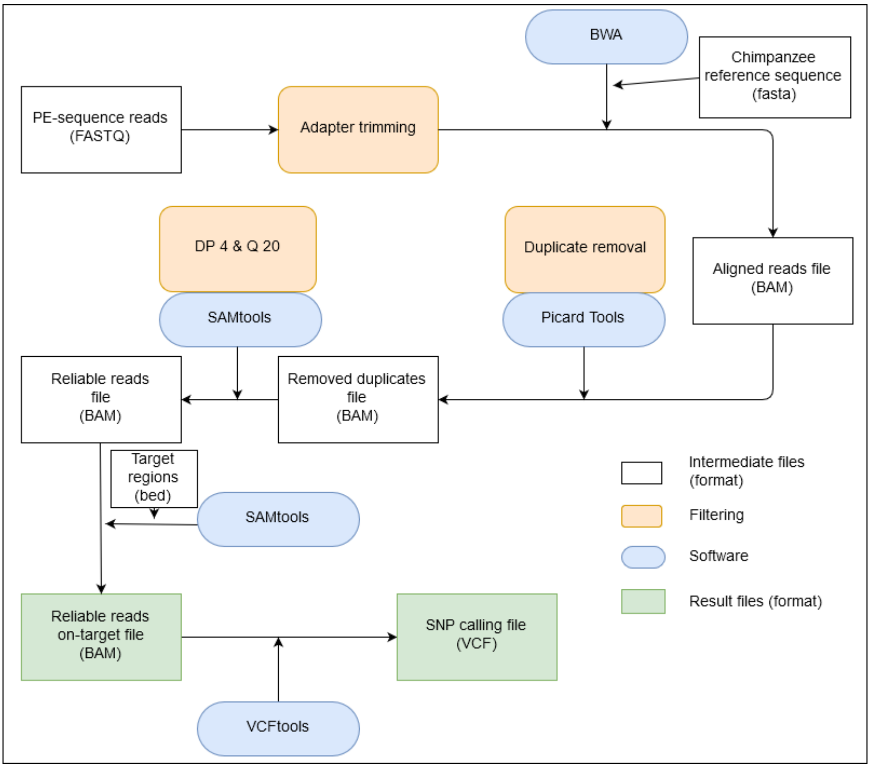


Figure 14. Bioinformatic pipeline followed for data processing.

The first step was carried out using Trim Galore (version 0.4.0) and Cutadapt software (version 1.8.3), that automatically and consistently apply quality and adapter trimming to FastQ files (Krueger, 2016; Martin, 2011).

2.2.1. Mapping

After removing the adapters, the reads sequenced have to be aligned against a reference genome. Nowadays, there are more than 60 software tools that can be used for mapping. These tools have had to adapt to manage the growing quantities of NGS data, exploiting technological improvements, and handling protocol advancements (such as the addition of new useful information with the development of paired-end library protocols). The elevated number of different mappers complicates the task of selecting the appropriate one for a specific application. The main purpose of a mapper is to find the location of each read sequenced in the reference genome, while admitting errors and structural variation (Fonseca et al., 2012).

The software used in this thesis is the Burrows-Wheeler Alignment tool (BWA, version 0.7.12), a read alignment package based on backward search with Burrows–Wheeler Transform (BWT) (Li and Durbin, 2009). As the samples used for this thesis were chimpanzee faecal samples and the target regions were designed using the panTro4 assembly (Feb. 2011, CSAC Pan_troglodytes-2.1.4 (GCA_000001515.4), the reads were aligned to that reference genome.

2.2.2. Duplicate removal and quality filtering

As explained in the library preparation and target enrichment sections, during the PCR amplifications, as the DNA is amplified to obtain the required amount for capture or sequencing, an excess of clonal molecules is produced. These molecules are defined as those that map to the same strand and have the same start and end coordinates as another molecule, and are known as PCR duplicates. The PCR duplicates must be recognized and removed, because they can seriously bias succeeding analyses such as effectiveness of the target enrichment or coverage achieved.
Using Picard Tools MarkDuplicates (version 1.95) with default parameters (http://broadinstitute.github.io/picard/) duplicates were removed.

To ensure the confidence of the data, quality filters were applied, identifying these filtered reads as “reliable reads”. Reliable reads are those that map to a single unique genomic location and have a mapping quality score of 30 or higher. When performing the analysis, reliable reads on-target were also defined, which are simply reliable reads that mapped to our target space. These reliable reads on-target were extracted by intersecting the target regions with the reliable read set and counting the number of reads for each condition using the function samtools -c (samtools version 0.1.19; Li et al., 2009).

2.2.3. Variant calling and principal component analysis

Similarly as what happens with the mappers, there are various software solutions that have been developed to identify genomic variants, such as SNPs and DNA insertions and deletions. The most extensively used callers in genomic variant analyses are Genome Analysis Tool Kit HaplotypeCaller (GATK-HC), Samtools mpileup, Freebayes, and Torrent Variant Caller (TVC) (Hwang et al., 2016).

In this thesis SNPs were called using Freebayes (version 0.9.20) (Garrison and Marth, 2012) with standard filters and no population priors. Sites with a quality score below 30 and a depth of coverage (DP) smaller than 4 were removed from further analysis, with the caveat that variants used in the principle component analysis were identified using a less stringent quality score of 20. The output file generated is in variant call format (VCF). Data from all libraries generated in this thesis was merged with whole-genome sequencing data derived from 59 country-referenced chimpanzees (de Manuel et al., 2016).

Principal component analysis (PCA) is one of the most popular statistical methods used in multiple scientific disciplines. PCA was invented in 1901 by Karl Pearson as an equivalent of the principal axis theorem in mechanics, and it has been subsequently developed depending on the field of application, being 1978 the year when it was used in the analysis of multilocus genetic data (Menozzi et al., 1978). This procedure transforms a number of potentially associated variables into a smaller number of different variables called principal components.

The VCF file generated in the variant calling (merged with the 59 chimpanzees with WGS) was processed by PLINK (version 1.90b) (Purcell et al., 2007) to ascertain the population structure among the individuals of study. The aim of this procedure is to acquire the most significant information from multivariate data where a visual representation is infeasible due to the multiple dimensionality. This method has been used to provide insight into further substructure within different ape populations (Hormozdiari et al., 2013). In this case, combining Alu and L1 insertions, the PCA visibly classifies the four chimpanzee subspecies, discriminating two groups of chimpanzees: Western-Nigerian from Central-Eastern. It also distinguishes Eastern lowland gorillas from Western lowland gorillas, and Sumatran from Bornean orangutans (Figure 15).

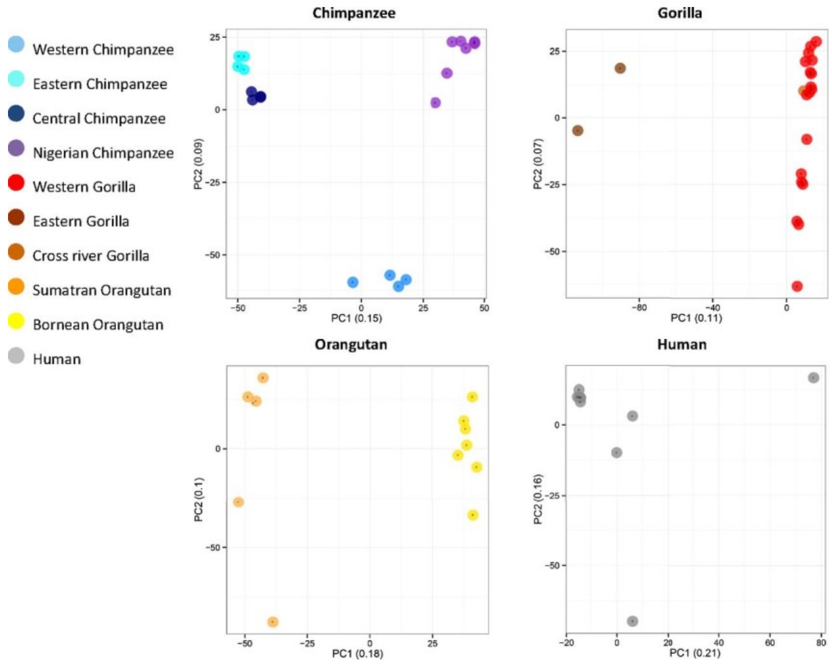


Figure 15. Principal component analysis. PCA using merged *Alu* and L1 insertions events on GRCh36 are depicted for chimpanzee, gorilla, orangutan, and human. (Extracted from Hormozdiari et al., 2013)

3. OBJECTIVES

The major objective of this thesis is to develop, refine and evaluate an experimental method to target enrich specific regions of the genome from complex samples. In this sense, the sample sources are faeces from chimpanzee and the target regions entangled the whole exome. Capturing the exome, the protein-coding portion of the genome, may be relevant for studies of natural selection, protein function, and evolution and yet, also remains useful in estimations of population ancestry, inbreeding, and potential geographic assignment. But this method can as well be adapted for other target regions as SNPs, specific chromosomes, etc., in consideration of the space selected and the desired coverage.

The empirical method developed here has the aim to provide essential knowledge, suggestions and guidelines for scientists in the usage of NI samples.

More explicitly, the aim of this work is to:

1. Assess the performance and replicability of a capture enrichment experiment involving a pool of multiple individuals.

2. Quantify the impact of wet lab technical variation on data acquisition and genotype discordance of a single sample.

3. Compare the differences of carrying out a single capture to that of a double capture.

4. Explore the information that may be gained by having faecal replicates, extract replicates, and/or library replicates in a study design. And measure discordance among these levels.

5. Quantify sample quality, defined as the endogenous DNA content and level of DNA fragmentation.

4. RESULTS

4.1. The impact of endogenous content, replicates and pooling on genome capture from faecal samples

Hernandez-Rodriguez J, Arandjelovic M, Lester J, de Filippo C, Weihmann A, Meyer M, et al. The impact of endogenous content, replicates and pooling on genome capture from faecal samples. Mol Ecol Resour. 2018 Mar;18(2):319–33. DOI: 10.1111/1755-0998.12728

5. DISCUSSION

Next-generation sequencing strategies have experienced a massive revolution over the last decade, introducing constantly new methodologies that can be applied to improve the results obtained by sequencing.

One of these methods is target enrichment. As presented in this work, this new approach allows the selection of the genomic regions of interest for each study. An advantage of this methodology is that other researchers can adapt it to their laboratory techniques, selecting the regions to study, whether is whole chromosome, a selected set of SNPs, or any target choice. Some adjustment must be made, taking into account the target space intended to be covered and re-adjusting the sequencing to obtain the desired coverage. In this work we also set the stage for pooling samples with different percentage of endogenous content, being a crucial step in any experiment of this kind, this has to be kept in mind while planning the experimental part of the study. This assumption was made on the basis of the percentage of endogenous DNA being the most important factor, affecting the number of raw reads obtained, that directly influences mean coverage (91.6% of variance explained) and proportion of target space covered (55% of variance explained).

With these precedents, we encourage researchers to, first, analyse DNA quality from as many specimens as possible to detect samples that will behave poorly. Second, when pooling multiple samples, to merge them by endogenous content or by generating equi-endogenous content (when the estimates of endogenous content are used to equilibrate the amount of each library added to the pool prior to hybridization and sequencing). This is important to minimize the negative effect that samples with higher content of endogenous DNA may have on the samples with lowest proportion; understanding this adverse outcome as the acquisition of most of the raw reads by the samples with higher proportion of endogenous DNA. To accurately estimate the percentage of endogenous DNA we suggest the performance of quantification assays and if possible, low level shotgun sequencing. And third, to perform multiple DNA extracts per specimen and/or create various libraries per extract when possible. By doing so, the molecule diversity will increase while the amount of sequencing necessary is reduced.

Another aim of this project was to produce enough data to allow the genotyping of our samples. A conclusion that can be extracted from our analyses is that the number of genotyped positions correlates with the proportion of target space covered at our minimum calling depth of four. At that depth, we estimate that, to cover around 80% of the target space we need a mean coverage of 20X, and 40X to cover around 95% of our target space (57.5 Mb).

We have demonstrated that at least 16 libraries can be pooled and captured together and still obtain enough amount of on-target reads. From our results it is clear that when performing two rounds of capture as opposed to just one, more genotype data is acquired for less sequencing data (by obtaining a 5-fold decrease in the off-target regions when performing two rounds of hybridization), when working with NI samples. Even though we discern some allele imbalance towards the reference allele present in the probe, the observed small difference is equivalent when comparing one versus two rounds of capture.

In summary, we have sequenced and captured chimpanzee exomes from a total of 24 libraries from 17 chimpanzees with a 4-reaction kit, executing two rounds of hybridization from most of the libraries and replicates. We have tried to make the protocol as cost-effective as possible, providing a methodology affordable for most laboratories using a commercial kit. We estimate that the cost from all Roche kits for the library preparation (24 libraries) and hybridization (without sequencing), including clean-up beads and purification columns, for the 72 experiments is around 450€ per library.

By applying capture technologies, the use of most uncommon samples, such as NI samples, has come into play. Their use had been restricted over the past decades to certain molecular markers, but nowadays, all the advances in the fields of genetics and genomics have enabled the expansion of the genetic information that can be obtained from this type of samples. Implementing all the advantages this entails, more information about the geographical origin of the sample, their habitat, diet habits, behaviour and other data can be compiled from wild populations. Moreover, individuals living in zoo and research organizations, as well as sanctuaries and other captive institutions, will benefit from the use on NI samples, avoiding the use of invasive samples like blood that can jeopardize their health because of the stress and infections the sampling can produce, not to mention the change in their behaviour that can affect the whole group of individuals.

The use of other complex samples such as aDNA (museum samples; van der Valk et al., 2017) can provide an immense advancement when seeking to understand how primate species have evolved since the expansion of humans to their territories. This has produced a decline in their populations due to the reduction of their habitat and hunting, bringing closer to extinction some species in several areas and sometimes even producing a complete extinction. Nowadays, when these extinction processes occur, we can rely on aDNA samples to sketch a map of which species could be found in those regions in the past and compare them with those we find these days, to evaluate if there has been a loss of diversity.

Furthermore, this genetic information can be of use and implemented in conservation programmes, not just for primates, but also for endangered species protection management in general. As reported by the International Union for the Conservation of Nature (IUCN) in their Red list of threatened species, more than 20.000 species are exposed to extinction worldwide, and 69 are already extinct in the wild (International Union for Conservation of Nature and Natural Resources., 2000).

In the case of primates, there are several threats that affect their survival: hunting, habitat loss, disease, climate change and road network. Hunting takes place when people seize them for trade or consumption and in most of the cases it also entangles killing other animals of the group. Habitat loss is mainly produced by the reduction of their habitat produced by the expansion of the land intended for agriculture, housing, roads, and plantations. There is also habitat degradation by forest exploitation, which increases the risk of being caught. Disease occurs mostly by the contact between humans and great apes through hunting, research, tourism and other incursions into their territory. Climate is likewise detrimental by diminishing their habitats or varying their alimentary sourcing. Road network, besides reducing their total area, produces the fragmentation of the habitat, increments the number of casualties and injures by vehicles impact and eases the approach of hunters.

Due to these threats, the species susceptible to extinction are grouped in 3 categories: vulnerable, endangered and critically endangered; the number of species included in each category changes each time the IUCN Red List is updated. This change of category may be produced by genuine or non-genuine reasons. The genuine reasons are divided in two: 1) absence of the main risks or conservation measures that have upgraded the species status to a less at risk category, and 2) continuance or escalation of the existing threats, or presence of new threats that have degraded the species status to a higher risk of extinction. The non-genuine categories may be a consequence of the availability of new information not present in the previous assessment, the detection of an error, and taxonomic revision.

With reference to primates, conservationists have played an important role in protecting primate populations, even overcoming unbeatable impediments. Some of the followed strategies range from law enforcement, protection of primates' territories and reintroduction programmes to payments for environmental benefits. In most countries there is an absence of a proper legislation of primate protection. This legislation is indispensable to handle illegal hunting or habitat transformation.

The management of protected areas is controversial because of the different conditions that exist across protected areas. This controversy springs from the role of local communities in the administration of protection areas and their influence in their efficiency. Reintroduction is a delicate procedure that must be held cautiously considering the exposition of wild populations to novel pathogens and the well-being of the introduced animal. This strategy should be performed as a last resource option and under rigorous guidelines (Campbell et al., 2017; IUCN, 2012).

Besides all these strategies adopted for conservation, there are still some limitations that could be circumvented with the use of genetic tools such as molecular genetics. This allows the collection of ecological and biological data that is not available with only field studies. The genetic information provides an approximation of the genetic diversity within a population, which corresponds to the phenotypic diversity. This information is of utmost importance, as the existence of diversity between populations would contribute to the survival even with environmental variance.

Molecular markers can be used to evaluate genetic diversity and its association with the surrounding environment. They are also effective to infer evolutionary history by measuring genetic and genotypic changes. As referenced in the introduction, some of these molecular markers include nuclear microsatellites, and mtDNA (Di Fiore, 2003), which have been largely significant in the study of NI samples. Microsatellites have been applied mostly for individual identity, parentage testing, to establish the degree of relatedness between individuals or populations, dispersal patterns and population structure. Given that mtDNA evolves considerable faster than most nuclear DNA, it is frequently used to evaluate maternal relatedness and sex-specific population structure for phylogenetic and phylogeographic studies. Yet, these markers present some discrepancies when comparing their respective phylogenies (Lobon et al., 2016; Shaw, 2002; Wiens et al.).

However, SNPs are the preferred molecular markers in current conservation and ecology studies, used for individual and population analysis, phylogeny inferences, population growth, patterns of gene flow, to evaluate models of selection, to assess patterns of migration in different populations and to examine novel adaptive mutations and diversity. Together with the exome and whole-genome, they are the sources of genetic information most commonly obtained through NGS for conservation genomics.

Along with population genetics, genomic knowledge will contribute to identifying genomic regions under selection, to reconstructing the demographic history of populations and to assessing the level of inbreeding (an important issue when referring to endangered species). All the genetic information collected in genomic research should be shared with conservationists, governments, and politicians, to devise a plan of action for the management and conservation of primate populations.

All things considered, we might wonder about the importance of non-human primate conservation. In the first place, for the mere fact of avoiding their decline and, ultimately, their extinction. Besides, their diminution may affect other species, for instance, humans. Primates have been decisive in human health research on account of the physiological and genetic closeness to humans (Carlsson et al., 2004; The Chimpanzee Sequencing and Analysis Consortium, 2005), and may contribute to the comprehension of human evolution. Moreover, to help understand human behaviour, for instance cognitive and linguistic adaptations, such as the theory of mind, thought to be exclusive from humans, but recently observed to be possessed by other apes as well (Buttelmann et al., 2017; Krupenye et al., 2016). They also play a key role in ecological functions like pollination and seed dispersion. Their conservation causes a benefit to local communities, as a tourist appeal, providing economic prosperity. Additionally, their conservation may contribute to the conservation of other species.

Given these points and thanks to the development of NGS, target enrichment technologies and the use of NI samples, the fields of genetics and genomics will contribute to a large extent in other fields. From an evolutionary medicine point of view, these data from non-human primates can be useful in studies applied to humans. As mentioned in the introduction, several studies have suggested non-human primates as animal models to study certain human diseases. Some of these conditions include hepatitis B virus infection, malaria, ebola, immunodeficiency virus and Arnold-Chiari malformation (Faust and Dobson, 2015; Klatt et al., 2012; Solis-Moruno et al., 2017; Thung et al., 1981; Walsh et al., 2017). All the data generated from target enrichment, as seen in this project, can be applied to ecological, population, evolutionary, and conservation genetic studies. Consequently, affecting primate conservation by fighting trafficking of primates as pet animals by identifying the origin of the individuals captured for trade, which allows punishment and the reintroduction of the animals in their region of origin.

6. Contributions to other publications

♦ Guillem de Valles-Ibáñez, Ana Esteve-Sole, Mònica Piquer, Azucena González-Navarro, **Jessica Hernandez-Rodriguez**, Hafid Laayouni, Eva González-Roca, Ana María Plaza-Martín, Angela Deyà-Martínez, Andrea Martín-Nalda, Mònica Martínez-Gallo, Marina García-Prat, Lucía del Pino, Ivon Cuscó, Marta Codina-Solà, Tomàs Marquès-Bonet, Elena Bosch, Eduardo Lopez-Granados, Juan Ignacio Aróstegui, Pere Soler-Palacín, Roger Colobrán, Jordi Yagüe, Laia Alsina, Manel Juan and Ferran Casals. (2018)

Evaluating the genetics of common variable immunodeficiency: monogenetic model and beyond.

Frontiers in Immunology, section Primary Immunodeficiencies (In press, 2018).

♦ Serres-Armero, A., Povolotskaya, I. S., Quilez, J., Ramirez, O., Santpere, G., Kuderna, L. F. K., **Hernandez-Rodriguez, J.**, Fernandez-Callejo, M., Gomez-Sanchez, D., Freedman, A. H., Fan, Z., Novembre, J., Navarro, A., Boyko, A., Wayne, R., Vilà, C., Lorente-Galdos, B. and Marques-Bonet, T. (2017)

Similar genomic proportions of copy number variation within gray wolves and modern dog breeds inferred from whole genome sequencing

BMC Genomics, 18(1). doi: 10.1186/s12864-017-4318-x.

https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-4318-x

♦ Solis-Moruno, M., de Manuel, M., **Hernandez-Rodriguez, J.**, Fontsere, C., Gomara-Castaño, A., Valsera-Naranjo, C., Crailsheim, D., Navarro, A., Llorente, M., Riera, L., Feliu-Olleta, O. and Marques-Bonet, T. (2017)

Potential damaging mutation in LRP5 from genome sequencing of the first reported chimpanzee with the Chiari malformation

Scientific Reports. Nature Publishing Group, 7(1), p. 15224. doi: 10.1038/s41598-017-15544-w.

https://doi.org/10.1038/s41598-017-15544-w

♦ de Valles-Ibáñez, G., **Hernandez-Rodriguez, J.**, Prado-Martinez, J., Luisi, P., Marquès-Bonet, T. and Casals, F. (2016)

Genetic Load of Loss-of-Function Polymorphic Variants in Great Apes

Genome biology and evolution, 8(3). doi: 10.1093/gbe/evw040.

https://academic.oup.com/gbe/article/8/3/871/2574144

♦ de Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V. C., Desai, T., Prado-Martinez, J., **Hernandez-Rodriguez, J.**, Dupanloup, I., Lao, O., Hallast, P., Schmidt, J. M., Heredia-Genestar, J. M., Benazzo, A., Barbujani, G., Peter, B. M., Kuderna, L. F. K., Casals, F., Angedakin, S., Arandjelovic, M., Boesch, C., Kuhl, H., Vigilant, L., Langergraber, K., Novembre, J., Gut, M., Gut, I., Navarro, A., Carlsen, F., Andres, A. M., Siegismund, H. R., Scally, A., Excoffier, L., Tyler-Smith, C., Castellano, S., Xue, Y., Hvilsom, C., Marques-Bonet, T., Kühl, H., Vigilant, L., Langergraber, K., Novembre, J., Gut, M., Gut, I., Navarro, A., Carlsen, F., Andrés, A. M., Siegismund, H. R., Scally, A., Excoffier, L., Tyler-Smith, C., Castellano, S., Xue, Y., Hvilsom, C. and Marques-Bonet, T. (2016)

Chimpanzee genomic diversity reveals ancient admixture with bonobos

Science, 354(6311), pp. 477–481. doi: 10.1126/science.aag2602.

http://www.ncbi.nlm.nih.gov/pubmed/27789843

♦ Lobon, I., Tucci, S., de Manuel, M., Ghirotto, S., Benazzo, A., Prado-Martinez, J., Lorente-Galdos, B., Nam, K., Dabad, M., **Hernandez-Rodriguez, J.**, Comas, D., Navarro, A., Schierup, M. H., Andres, A. M., Barbujani, G., Hvilsom, C. and Marques-Bonet, T. (2016)
Demographic History of the Genus Pan Inferred from Whole Mitochondrial Genome Reconstructions
Genome biology and evolution. Oxford University Press, 8(6), pp. 2020–30. doi: 10.1093/gbe/evw124.
http://www.ncbi.nlm.nih.gov/pubmed/27345955

♦ Ruiz-Orera, J., **Hernandez-Rodriguez, J.**, Chiva, C., Sabidó, E., Kondova, I., Bontrop, R., Marqués-Bonet, T. and Albà, M. M. (2015)
Origins of De Novo Genes in Human and Chimpanzee
PLOS Genetics, 11(12), p. e1005721. doi: 10.1371/journal.pgen.1005721.
http://dx.plos.org/10.1371/journal.pgen.1005721

♦ Xue, Y., Prado-Martinez, J., Sudmant, P. H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D. N., De Manuel, M., **Hernandez-Rodriguez, J.**, Lobon, I., Siegismund, H. R., Pagani, L., Quail, M. A., Hvilsom, C., Mudakikwa, A., Eichler, E. E., Cranfield, M. R., Marques-Bonet, T., Tyler-Smith, C. and Scally, A. (2015)
Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding
Science, 348(6231), pp. 242–245. doi: 10.1126/science.aaa3952
http://science.sciencemag.org/content/348/6231/242.long

♦ Ramirez, O., Olalde, I., Berglund, J., Lorente-Galdos, B., **Hernandez-Rodriguez, J.**, Quilez, J., Webster, M. T., Wayne, R. K., Lalueza-Fox, C., Vilà, C. and Marques-Bonet, T. (2014)

Analysis of structural diversity in wolf-like canids reveals post-domestication variants

BMC Genomics, 15(1). doi: 10.1186/1471-2164-15-465.

https://www.ncbi.nlm.nih.gov/pubmed/24923435

♦ Carbone, L., Alan Harris, R., Gnerre, S., Veeramah, K. R., Lorente-Galdos, B., Huddleston, J., Meyer, T. J., Herrero, J., Roos, C., Aken, B., Anaclerio, F., Archidiacono, N., Baker, C., Barrell, D., Batzer, M. A., Beal, K., Blancher, A., Bohrson, C. L., Brameier, M., Campbell, M. S., Capozzi, O., Casola, C., Chiatante, G., Cree, A., Damert, A., De Jong, P. J., Dumas, L., Fernandez-Callejo, M., Flicek, P., Fuchs, N. V., Gut, I., Gut, M., Hahn, M. W., **Hernandez-Rodriguez, J.**, Hillier, L. W., Hubley, R., Ianc, B., Izsvák, Z., Jablonski, N. G., Johnstone, L. M., Karimpour-Fard, A., Konkel, M. K., Kostka, D., Lazar, N. H., Lee, S. L., Lewis, L. R., Liu, Y., Locke, D. P., Mallick, S., Mendez, F. L., Muffato, M., Nazareth, L. V., Nevonen, K. A., O'Bleness, M., Ochis, C., Odom, D. T., Pollard, K. S., Quilez, J., Reich, D., Rocchi, M., Schumann, G. G., Searle, S., Sikela, J. M., Skollar, G., Smit, A., Sonmez, K., Ten Hallers, B., Terhune, E., Thomas, G. W. C., Ullmer, B., Ventura, M., Walker, J. A., Wall, J. D., Walter, L., Ward, M. C., Wheelan, S. J., Whelan, C. W., White, S., Wilhelm, L. J., Woerner, A. E., Yandell, M., Zhu, B., Hammer, M. F., Marques-Bonet, T., Eichler, E. E., Fulton, L., Fronick, C., Muzny, D. M., Warren, W. C., Worley, K. C., Rogers, J., Wilson, R. K. and Gibbs, R. A. (2014)

Gibbon genome and the fast karyotype evolution of small apes

Nature, 513(7517). doi: 10.1038/nature13679.

https://www.nature.com/articles/nature13679

♦ Lorente-Galdos, B., Bleyhl, J., Santpere, G., Vives, L., Ramírez, O., **Hernandez, J.**, Anglada, R., Cooper, G. M., Navarro, A., Eichler, E. E. and Marques-Bonet, T. (2013)

Accelerated exon evolution within primate segmental duplications

Genome biology, 14, p. R9. doi: 10.1186/gb-2013-14-1-r9.

http://genomebiology.com/content/14/1/R9

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

Great ape genetic diversity and population history

Nature. Nature Research, 499(7459), pp. 471–5. doi: 10.1038/nature12228.

http://www.nature.com/doifinder/10.1038/nature12228

7. Bibliography

Alba, D.M., Almécija, S., DeMiguel, D., Fortuny, J., Pérez de los Ríos, M., Pina, M., Robles, J.M., and Moyà-Solà, S. (2015). Miocene small-bodied ape from Eurasia sheds light on hominoid evolution. Science *350*, aab2625.

Allen, M., Engström, A.S., Meyers, S., Handt, O., Saldeen, T., von Haeseler, A., Pääbo, S., and Gyllensten, U. (1998). Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. J. Forensic Sci. *43*, 453–464.

Allendorf, F.W., Hohenlohe, P.A., and Luikart, G. (2010). Genomics and the future of conservation genetics. Nat. Rev. Genet. *11*, 697–709.

Arandjelovic, M., Head, J., Kühl, H., Boesch, C., Robbins, M.M.M., Maisels, F., and Vigilant, L. (2010). Effective non-invasive genetic monitoring of multiple wild western gorilla groups. Biol. Conserv. *143*, 1780–1791.

Arandjelovic, M., Head, J., Rabanal, L.I., Schubert, G., Mettke, E., Boesch, C., Robbins, M.M., and Vigilant, L. (2011). Non-invasive genetic monitoring of wild central chimpanzees. PLoS One *6*, e14761.

Ávila-Arcos, M.C., Sandoval-Velasco, M., Schroeder, H., Carpenter, M.L., Malaspinas, A.-S., Wales, N., Peñaloza, F., Bustamante, C.D., Gilbert, M.T.P., and Bunce, M. (2015). Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina libraries. Methods Ecol. Evol. *6*, 725–734.

Ballard, J.W.O., and Rand, D.M. (2005). The Population Biology of Mitochondrial DNA and Its Phylogenetic Implications. Annu. Rev. Ecol. Evol. Syst. *36*, 621–642.

Bayes, M.K., Smith, K.L., Alberts, S.C., Altmann, J., and Bruford, M.W. (2000). Testing the reliability of microsatellite typing from faecal DNA in the savannah baboon. Conserv. Genet. *1*, 173–176.

Bermejo, M., Rodríguez-Teijeiro, J.D., Illera, G., Barroso, A., Vilà, C., and Walsh, P.D. (2006). Ebola outbreak killed 5000 gorillas.

Science *314*, 1564.

Boom, R., Sol, C.J., Salimans, M.M., Jansen, C.L., Wertheim-van Dillen, P.M., and van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. *28*, 495–503.

Bradley, B.J., and Vigilant, L. (2002). False alleles derived from microbial DNA pose a potential source of error in microsatellite genotyping of DNA from faeces. Mol. Ecol. Notes *2*, 602–605.

Bradley, B.J., Boesch, C., and Vigilant, L. (2000). Identification and redesign of human microsatellite markers for genotyping wild chimpanzee (Pan troglodytes verus) and gorilla (Gorilla gorilla gorilla) DNA from faeces. Conserv. Genet. *1*, 289–292.

Broquet, T., Ménard, N., and Petit, E. (2006). Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. Conserv. Genet. *8*, 249–260.

Buttelmann, D., Buttelmann, F., Carpenter, M., Call, J., and Tomasello, M. (2017). Great apes distinguish true from false beliefs in an interactive helping task. PLoS One *12*, e0173793.

Caldecott, J.O., Miles, L., and UNEP World Conservation Monitoring Centre. (2005). World atlas of great apes and their conservation (University of California Press).

Campbell, C.O., Cheyne, S.M., and Rawson, B.M. (2017). Best practice guidelines for the rehabilitation and translocation of gibbons (IUCN).

Carlsson, H.-E., Schapiro, S.J., Farah, I., and Hau, J. (2004). Use of primates in research: A global overview. Am. J. Primatol. *63*, 225–237.

Carpenter, M.L., Buenrostro, J.D., Valdiosera, C., Schroeder, H., Allentoft, M.E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., et al. (2013). Pulling out the 1%: Whole-Genome capture for the targeted enrichment of ancient dna sequencing libraries. Am. J. Hum. Genet. *93*, 852–864.

Casanovas-Vilar, I., Alba, D.M., Garcés, M., Robles, J.M., and Moyà-Solà, S. (2011). Updated chronology for the Miocene hominoid radiation in Western Eurasia. Proc. Natl. Acad. Sci. U. S.

A. *108*, 5554–5559.

Castellano, S., Parra, G., Sánchez-Quinto, F.A., Racimo, F., Kuhlwilm, M., Kircher, M., Sawyer, S., Fu, Q., Heinze, A., Nickel, B., et al. (2014). Patterns of coding variation in the complete exomes of three Neandertals. Proc. Natl. Acad. Sci. U. S. A. *111*, 6666–6671.

Chen, F.C., and Li, W.H. (2001). Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am. J. Hum. Genet. *68*, 444–456.

Chilamakuri, C.S.R., Lorenz, S., Madoui, M.-A., Vodák, D., Sun, J., Hovig, E., Myklebost, O., and Meza-Zepeda, L. a (2014). Performance comparison of four exome capture systems for deep sequencing. BMC Genomics *15*, 449.

Chiou, K.L., and Bergey, C.M. (2018). Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. Sci. Rep. *8*, 1975.

Chistiakov, D.A., Hellemans, B., and Volckaert, F.A.M. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquaculture *255*, 1–29.

Dabney, J., Meyer, M., and Pääbo, S. (2013). Ancient DNA damage. Cold Spring Harb. Perspect. Biol. *5*.

Dietrich, D., Uhl, B., Sailer, V., Holmes, E.E., Jung, M., Meller, S., and Kristiansen, G. (2013). Improved PCR Performance Using Template DNA from Formalin-Fixed and Paraffin-Embedded Tissues by Overcoming PCR Inhibition. PLoS One *8*, e77771.

Emery, L.S., Magnaye, K.M., Bigham, A.W., Akey, J.M., and Bamshad, M.J. (2015). Estimates of Continental Ancestry Vary Widely among Individuals with the Same mtDNA Haplogroup. Am. J. Hum. Genet. *96*, 183–193.

Eriksson, J., Siedel, H., Lukas, D., Kayser, M., Erler, A., Hashimoto, C., Hohmann, G., Boesch, C., and Vigilant, L. (2006). Y-chromosome analysis confirms highly sex-biased dispersal and suggests a low male effective population size in bonobos (Pan paniscus). Mol. Ecol. *15*, 939–949.

Eriksson J, Hohmann G, Boesch C, V.L. (2004). Rivers influence the population genetic structure of bonobos (Pan paniscus). Mol. Ecol. *13*, 3425–3435.

Erler, A., Stoneking, M., and Kayser, M. (2004). Development of Y-chromosomal microsatellite markers for nonhuman primates. Mol. Ecol. *13*, 2921–2930.

Estrada, A., Garber, P.A., Rylands, A.B., Roos, C., Fernandez-Duque, E., Fiore, A. Di, Nekaris, K.A.-I., Nijman, V., Heymann, E.W., Lambert, J.E., et al. (2017). Impending extinction crisis of the world's primates: Why primates matter. Sci. Adv. *3*, e1600946.

Faust, C., and Dobson, A.P. (2015). Primate malarias: Diversity, distribution and insights for zoonotic Plasmodium. One Heal. *1*, 66–75.

Di Fiore, A. (2003). Molecular genetic approaches to the study of primate behavior, social organization, and reproduction. Am. J. Phys. Anthropol. *122*, 62–99.

Fischer, A., Wiebe, V., Pääbo, S., and Przeworski, M. (2004). Evidence for a Complex Demographic History of Chimpanzees. Mol. Biol. Evol. *21*, 799–808.

Fischer, A., Prüfer, K., Good, J.M., Halbwax, M., Wiebe, V., André, C., Atencia, R., Mugisha, L., Ptak, S.E., and Pääbo, S. (2011). Bonobos Fall within the Genomic Variation of Chimpanzees. PLoS One *6*, e21605.

Fonseca, N.A., Rung, J., Brazma, A., and Marioni, J.C. (2012). Tools for mapping high-throughput sequencing data. Bioinformatics *28*, 3169–3177.

Formenty, P., Boesch, C., Wyers, M., Steiner, C., Donati, F., Dind, F., Walker, F., and Le Guenno, B. (1999). Ebola virus outbreak among wild chimpanzees living in a rain forest of Côte d'Ivoire. J. Infect. Dis. *179 Suppl 1*, S120-6.

Frantzen, M.A., Silk, J.B., Ferguson, J.W., Wayne, R.K., and Kohn, M.H. (1998). Empirical evaluation of preservation methods for faecal DNA. Mol. Ecol. *7*, 1423–1428.

Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. Curr. Biol. *23*, 553–559.

Fünfstück, T., Arandjelovic, M., Morgan, D.B., Sanz, C., Breuer, T., Stokes, E.J., Reed, P., Olson, S.H., Cameron, K., Ondzie, A., et al. (2014). The genetic population structure of wild western lowland gorillas (Gorilla gorilla gorilla) living in continuous rain forest. Am. J. Primatol. *76*, 868–878.

Fünfstück, T., Arandjelovic, M., Morgan, D.B., Sanz, C., Reed, P., Olson, S.H., Cameron, K., Ondzie, A., Peeters, M., and Vigilant, L. (2015). The sampling scheme matters: Pan troglodytes troglodytes and P. t schweinfurthii are characterized by clinal genetic variation rather than a strong subspecies break. Am. J. Phys. Anthropol. *156*, 181–191.

Gansauge, M.-T., and Meyer, M. (2014). Selective enrichment of damaged DNA molecules for ancient genome sequencing. Genome Res. *24*, 1543–1549.

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing.

Gerloff, U., Hartung, B., Fruth, B., Hohmann, G., and Tautz, D. (1999). Intracommunity relationships, dispersal pattern and paternity success in a wild living community of Bonobos (Pan paniscus) determined from DNA analysis of faecal samples. Proceedings. Biol. Sci. *266*, 1189–1195.

Ghatak, S., sanga, Z., Pautu, J.L., and Kumar, N.S. (2015). Coextraction and PCR Based Analysis of Nucleic Acids From Formalin-Fixed Paraffin-Embedded Specimens. J. Clin. Lab. Anal. *29*, 485–492.

Gluckman, P.D., Beedle, A., and Hanson, M.A. (2009). Principles of evolutionary medicine (Oxford University Press).

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., et al. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat. Biotechnol. *27*, 182–189.

Goodman, M., Porter, C.A., Czelusniak, J., Page, S.L., Schneider, H., Shoshani, J., Gunnell, G., and Groves, C.P. (1998). Toward a Phylogenetic Classification of Primates Based on DNA Evidence Complemented by Fossil Evidence. Mol. Phylogenet. Evol. *9*, 585–598.

Goossens, B., and Bruford, M.W. (2009). Non-invasive genetic analysis in conservation. In Population Genetics for Animal Conservation, G. Bertorelle, M.W. Bruford, H.C. Hauffe, A. Rizzoli, and C. Vernesi, eds. (Cambridge: Cambridge University Press), pp. 167–201.

Goossens, B., Waits, L.P., and Taberlet, P. (1998). Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. Mol. Ecol. *7*, 1237–1241.

Goossens, B., Funk, S.M., Vidal, C., Latour, S., Jamart, A., Ancrenaz, M., Wickings, E.J., Tutin, C.E.G., and Bruford, M.W. (2002). Measuring genetic diversity in translocation programmes: principles and application to a chimpanzee release project. Anim. Conserv. *5*, 225–236.

Goossens, B., Chikhi, L., Jalil, M.F., Ancrenaz, M., Lackman-Ancrenaz, I., Mohamed, M., Andau, P., and Bruford, M.W. (2004). Patterns of genetic diversity and migration in increasingly fragmented and declining orang-utan (Pongo pygmaeus) populations from Sabah, Malaysia. Mol. Ecol. *14*, 441–456.

Goossens, B., Anthony, N., Jeffery, K., Johnson-Bawe, M., and Bruford, M.W. (2011). Collection, storage and analysis of non-invasive genetic material in primate biology. In Field and Laboratory Methods in Primatology, J.M. Setchell, and D.J. Curtis, eds. (Cambridge: Cambridge University Press), pp. 371–386.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710–722.

Guo, Y., Long, J., He, J., Li, C.-I., Cai, Q., Shu, X.-O., Zheng, W., and Li, C.-I. (2012). Exome sequencing generates high quality data in non-target regions. BMC Genomics *13*, 194.

Gustafsson, C.M., Falkenberg, M., and Larsson, N.-G. (2016). Maintenance and Expression of Mammalian Mitochondrial DNA. Annu. Rev. Biochem. *85*, 133–160.

Hans, J.B., Haubner, A., Arandjelovic, M., Bergl, R.A., Fünfstück, T., Gray, M., Morgan, D.B., Robbins, M.M., Sanz, C., and Vigilant, L. (2015). Characterization of MHC class II B polymorphism in multiple populations of wild gorillas using non-invasive samples and next-generation sequencing. Am. J. Primatol. *77*, 1193–1206.

Herron, J.C., and Freeman, S. (2013). Evolutionary Analysis (Pearson).

Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M.N., Smith, S.W., Middle, C.M., Rodesch, M.J., Albert, T.J., Hannon, G.J., et al. (2007). Genome-wide in situ exon capture for selective resequencing. Nat. Genet. *39*, 1522–1527.

Hofreiter, M., Siedel, H., Van Neer, W., and Vigilant, L. (2003). Mitochondrial DNA sequence from an enigmatic gorilla population (Gorilla gorilla uellensis). Am. J. Phys. Anthropol. *121*, 361–368.

Hormozdiari, F., Konkel, M.K., Prado-Martinez, J., Chiatante, G., Herraez, I.H., Walker, J.A., Nelson, B., Alkan, C., Sudmant, P.H., Huddleston, J., et al. (2013). Rates and patterns of great ape retrotransposition. Proc. Natl. Acad. Sci. U. S. A. *110*, 13457–13462.

Hwang, S., Kim, E., Lee, I., and Marcotte, E.M. (2016). Systematic comparison of variant calling pipelines using gold standard personal exome variants. Sci. Rep. *5*, 17875.

Inoue, E., Akomo-Okoue, E.F., Ando, C., Iwata, Y., Judai, M., Fujita, S., Hongo, S., Nze-Nkogue, C., Inoue-Murayama, M., and Yamagiwa, J. (2013). Male genetic structure and paternity in western lowland gorillas (*Gorilla gorilla gorilla*). Am. J. Phys. Anthropol. *151*, 583–588.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

International Union for Conservation of Nature and Natural Resources. (2000). The IUCN red list of threatened species (IUCN Global Species Programme Red List Unit).

IUCN (2012). Guidelines for Reintroductions and Other Conservation Translocations.

Jeffery, K.J., Abernethy, K.A., Tutin, C.E.G., and Bruford, M.W. (2007). Biological and environmental degradation of gorilla hair and microsatellite amplification success. Biol. J. Linn. Soc. *91*, 281–294.

Jobling, M., Hollox, E., Kivisild, T., and Tyler-Smith, C. (2013). Human Evolutionary Genetics (Taylor & Francis Inc).

Johnson, A.P., Ison, C.A., Hetherington, C.M., Osborn, M.F., Southerton, G., London, W.T., Easmon, C.S., and Taylor-Robinson, D. (1984). A study of the susceptibility of three species of primate to vaginal colonization with Gardnerella vaginalis. Br. J. Exp. Pathol. *65*, 389–396.

Jones, M.R., and Good, J.M. (2016). Targeted capture in evolutionary and ecological genomics. Mol. Ecol. *25*, 185–202.

Jurmain, R., Kilgore, L., Trevathan, W., and Ciochon, R.L. (2017). Introduction to physical anthropology (Wadsworth Publishing; 15 edition (February 2, 2017)).

Kanthaswamy, S., Kurushima, J.D., and Smith, D.G. (2006). Inferring Pongo conservation units: a perspective based on microsatellite and mitochondrial DNA analyses. Primates *47*, 310–321.

Kawamoto, Y., Takemoto, H., Higuchi, S., Sakamaki, T., Hart, J.A., Hart, T.B., Tokuyama, N., Reinartz, G.E., Guislain, P., Dupain, J., et al. (2013). Genetic Structure of Wild Bonobo Populations: Diversity of Mitochondrial DNA and Geographical Distribution. PLoS One *8*, e59660.

Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F., Leidinger, P., Backes, C., Khairat, R., Forster, M., et al. (2012). New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nat. Commun. *3*, 698.

Kidd, J., Sharpton, T., Bobo, D., Norman, P., Martin, A., Carpenter, M., Sikora, M., Gignoux, C., Nemat-Gorgani, N., Adams, A., et al. (2014). Exome capture from saliva produces high quality genomic and metagenomic data. BMC Genomics *15*, 262.

Kim, S., and Misra, A. (2007). SNP Genotyping: Technologies and Biomedical Applications. Annu. Rev. Biomed. Eng. *9*, 289–320.

Kitts, A., and Sherry, S. (2002). The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation.

Klatt, N.R., Silvestri, G., and Hirsch, V. (2012). Nonpathogenic simian immunodeficiency virus infections. Cold Spring Harb. Perspect. Med. *2*, a007153.

Kozarewa, I., Armisen, J., Gardner, A.F., Slatko, B.E., and Hendrickson, C.L. (2015). Overview of Target Enrichment

Strategies. In Current Protocols in Molecular Biology, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), p. 7.21.1-7.21.23.

Krueger, F. (2016). Babraham Bioinformatics - Trim Galore!

Krupenye, C., Kano, F., Hirata, S., Call, J., and Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. Science *354*, 110–114.

Kuykendall, J.R., and Bogdanffy, M.S. (1992). Efficiency of DNA-histone crosslinking induced by saturated and unsaturated aldehydes in vitro. Mutat. Res. Lett. *283*, 131–136.

Langergraber, K.E., Rowney, C., Schubert, G., Crockford, C., Hobaiter, C., Wittig, R., Wrangham, R.W., Zuberbühler, K., and Vigilant, L. (2014). How old are chimpanzee communities? Time to the most recent common ancestor of the Y-chromosome in highly patrilocal societies. J. Hum. Evol. *69*, 1–7.

Lathuillière, M., Ménard, N., Gautier-Hion, A., and Crouau-Roy, B. (2001). Testing the reliability of noninvasive genetic sampling by comparing analyses of blood and fecal samples in Barbary macaques (*Macaca sylvanus*). Am. J. Primatol. *55*, 151–158.

Launhardt, K., Epplen, C., Epplen, J.T., and Winkler, P. (1998). Amplification of microsatellites adapted from human systems in faecal DNA of wild Hanuman langurs (Presbytis entellus). Electrophoresis *19*, 1356–1361.

Lewis, R.B., Jurmain, R., and Kilgore, L. (2012). Understanding humans : introduction to physical anthropology and archaeology.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Lin, J., Kennedy, S.H., Svarovsky, T., Rogers, J., Kemnitz, J.W., Xu, A., and Zondervan, K.T. (2009). High-quality genomic DNA extraction from formalin-fixed and paraffin-embedded samples deparaffinized using mineral oil. Anal. Biochem. *395*, 265–267.

Liu, Z., Ren, B., Wu, R., Zhao, L., Hao, Y., Wang, B., Wei, F., Long, Y., and Li, M. (2009). The effect of landscape features on population genetic structure in Yunnan snub-nosed monkeys (Rhinopithecus bieti) implies an anthropogenic genetic discontinuity. Mol. Ecol. *18*, 3831–3846.

Lobon, I., Tucci, S., de Manuel, M., Ghirotto, S., Benazzo, A., Prado-

Martinez, J., Lorente-Galdos, B., Nam, K., Dabad, M., Hernandez-Rodriguez, J., et al. (2016). Demographic History of the Genus Pan Inferred from Whole Mitochondrial Genome Reconstructions. Genome Biol. Evol. *8*, 2020–2030.

Locke, D.P., Hillier, L.W., Warren, W.C., Worley, K.C., Nazareth, L. V., Muzny, D.M., Yang, S.-P., Wang, Z., Chinwalla, A.T., Minx, P., et al. (2011). Comparative and demographic analysis of orang-utan genomes. Nature *469*, 529–533.

de Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V.C., Desai, T., Prado-Martinez, J., Hernandez-Rodriguez, J., Dupanloup, I., Lao, O., Hallast, P., et al. (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science (80-.). *354*, 477–481.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.

Menozzi, P., Piazza, A., and Cavalli-Sforza, L. (1978). Synthetic maps of human gene frequencies in Europeans. Science *201*, 786–792.

Meyer, M., and Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448-prot5448.

Millar, C.D., and Lambert, D.M. (2013). Ancient DNA: Towards a million-year-old genome. Nature *499*, 34–35.

Miller, C.A., Campbell, S.L., and Sweatt, J.D. (2008). DNA methylation and histone acetylation work in concert to regulate memory formation and synaptic plasticity. Neurobiol. Learn. Mem. *89*, 599–603.

Minhós, T., Wallace, E., Ferreira da Silva, M.J., Sá, R.M., Carmo, M., Barata, A., and Bruford, M.W. (2013). DNA identification of primate bushmeat from urban markets in Guinea-Bissau and its implications for conservation. Biol. Conserv. *167*, 43–49.

Morin, P.A., Wallis, J., Moore, J.J., Chakraborty, R., and Woodruff, D.S. (1993). Non-invasive sampling and DNA amplification for paternity exclusion, community structure, and phylogeography in wild chimpanzees. Primates *34*, 347–356.

Morin, P.A., Moore, J.J., Chakraborty, R., Jin, L., Goodall, J., and Woodruff, D.S. (1994). Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science *265*, 1193–1201.

Morin, P.A., Chambers, K.E., Boesch, C., and Vigilant, L. (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (Pan troglodytes verus). Mol. Ecol. *10*, 1835–1844.

Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Nater, A., Arora, N., Greminger, M.P., van Schaik, C.P., Singleton, I., Wich, S.A., Fredriksson, G., Perwitasari-Farajallah, D., Pamungkas, J., and Krützen, M. (2013). Marked Population Structure and Recent Migration in the Critically Endangered Sumatran Orangutan (Pongo abelii). J. Hered. *104*, 2–13.

Nengo, I., Tafforeau, P., Gilbert, C.C., Fleagle, J.G., Miller, E.R., Feibel, C., Fox, D.L., Feinberg, J., Pugh, K.D., Berruyer, C., et al. (2017). New infant cranium from the African Miocene sheds light on ape evolution. Nature *548*, 169–174.

Nsubuga, A.M., Robbins, M.M., Roeder, A.D., Morin, P.A., Boesch, C., and Vigilant, L. (2004). Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. Mol. Ecol. *13*, 2089–2094.

Olalde, I., Schroeder, H., Sandoval-Velasco, M., Vinner, L., Lobón, I., Ramirez, O., Civit, S., García Borja, P., Salazar-García, D.C., Talamo, S., et al. (2015). A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. Mol. Biol. Evol. *32*, msv181.

Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Rohland, N., Mallick, S., Booth, T., Szécsényi-Nagy, A., Mittnik, A., et al. (2017). The Beaker Phenomenon And The Genomic Transformation Of Northwest Europe. bioRxiv 135962.

Orkin, J.D., Yang, Y., Yang, C., Yu, D.W., and Jiang, X. (2016). Cost-effective scat-detection dogs: unleashing a powerful new tool for international mammalian conservation biology. Sci. Rep. *6*, 34758.

Ouborg, N.J., Pertoldi, C., Loeschcke, V., Bijlsma, R. (Kuke) K., and Hedrick, P.W. (2010). Conservation genetics in transition to conservation genomics. Trends Genet. *26*, 177–187.

Perlman, R. (2013). Evolution and Medicine (Oxford University Press).

Perry, G.H., Marioni, J.C., Melsted, P., and Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. Mol. Ecol. *19*, 5332–5344.

Perry, G.H., Reeves, D., Melsted, P., Ratan, A., Miller, W., Michelini, K., Louis, E.E., Pritchard, J.K., Mason, C.E., and Gilad, Y. (2012). A Genome Sequence Resource for the Aye-Aye (Daubentonia madagascariensis), a Nocturnal Lemur from Madagascar. Genome Biol. Evol. *4*, 126–135.

Perry, G.H., Louis, E.E., Ratan, A., Bedoya-Reina, O.C., Burhans, R.C., Lei, R., Johnson, S.E., Schuster, S.C., and Miller, W. (2013). Aye-aye population genomic analyses highlight an important center of endemism in northern Madagascar. Proc. Natl. Acad. Sci. U. S. A. *110*, 5823–5828.

Piggott, M.P., and Taylor, A.C. (2003). Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. Wildl. Res. *30*, 1.

Pikor, L.A., Enfield, K.S.S., Cameron, H., and Lam, W.L. (2011). DNA extraction from paraffin embedded material for genetic and epigenetic analyses. J. Vis. Exp.

Plumptre, A.J., Nixon, S., Kujirakwinja, D.K., Vieilledent, G., Critchlow, R., Williamson, E.A., Nishuli, R., Kirkby, A.E., and Hall, J.S. (2016). Catastrophic Decline of World's Largest Primate: 80% Loss of Grauer's Gorilla (Gorilla beringei graueri) Population Justifies Critically Endangered Status. PLoS One *11*, e0162697.

Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., et al. (2013). Great ape genetic diversity and population history. Nature *499*, 471–475.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. Genetics *155*.

Prüfer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J.R., Walenz, B., Koren, S., Sutton, G., Kodira, C., Winer, R., et al. (2012). The bonobo genome compared with the chimpanzee and human genomes. Nature *486*, 527–531.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

Quéméré, E., Louis, E.E., Ribéron, A., Chikhi, L., and Crouau-Roy, B. (2010a). Non-invasive conservation genetics of the critically endangered golden-crowned sifaka (Propithecus tattersalli): high diversity and significant genetic differentiation over a small range. Conserv. Genet. *11*, 675–687.

Quéméré, E., Crouau-Roy, B., Rabarivola, C., Louis, E.J., and Chikhi, L. (2010b). Landscape genetics of an endangered lemur (Propithecus tattersalli) within its entire fragmented range. Mol. Ecol. *19*, 1606–1621.

Quéméré, E., Hibert, F., Miquel, C., Lhuillier, E., Rasolondraibe, E., Champeau, J., Rabarivola, C., Nusbaumer, L., Chatelain, C., Gautier, L., et al. (2013). A DNA Metabarcoding Study of a Primate Dietary Diversity and Plasticity across Its Entire Fragmented Range. PLoS One *8*, e58971.

Radespiel, U., Rakotondravony, R., and Chikhi, L. (2008). Natural and anthropogenic determinants of genetic structure in the largest remaining population of the endangered golden-brown mouse lemur, <i>Microcebus ravelobensis<i>. Am. J. Primatol. *70*, 860–870.

Ramón-Laca, A., Soriano, L., Gleeson, D., and Godoy, J.A. (2015). A simple and effective method for obtaining mammal DNA from faeces. Wildlife Biol. *21*, 195–203.

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature *463*, 757–762.

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature *468*, 1053–1060.

Rhesus Macaque Genome Sequencing and Analysis Consortium (2007). Evolutionary and Biomedical Insights from the Rhesus Macaque Genome. Science (80-.). *316*, 222–234.

Roeder, A.D., Archer, F.I., Poinar, H.N., and Morin, P.A. (2004). A novel method for collection and preservation of faeces for genetic studies. Mol. Ecol. Notes *4*, 761–764.

Rogers, J., and Gibbs, R.A. (2014). Comparative primate genomics: emerging patterns of genome content and dynamics. Nat. Rev. Genet. *15*, 347–359.

Rowe, N. (1996). The pictorial guide to the living primates (Pogonias Press).

Salgado-Lynn, M., Stanton, D.W.G., Sakong, R., Cable, J., Goossens, B., and Bruford, M.W. (2010). Microsatellite markers for the proboscis monkey (Nasalis larvatus). Conserv. Genet. Resour. *2*, 159–163.

Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., et al. (2012). Insights into hominid evolution from the gorilla genome sequence. Nature *483*, 169–175.

Schaumburg, F., Mugisha, L., Kappeller, P., Fichtel, C., Köck, R., Köndgen, S., Becker, K., Boesch, C., Peters, G., and Leendertz, F. (2013). Evaluation of Non-Invasive Biological Samples to Monitor Staphylococcus aureus Colonization in Great Apes and Lemurs. PLoS One *8*, e78046.

Shafer, A.B.A., Wolf, J.B.W., Alves, P.C., Bergström, L., Bruford, M.W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., et al. (2015). Genomics and the challenging translation into conservation practice. Trends Ecol. Evol. *30*, 78–87.

Shaw, K.L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proc. Natl. Acad. Sci. U. S. A. *99*, 16122–16127.

Snow, A.N., Stence, A.A., Pruessner, J.A., Bossler, A.D., and Ma, D. (2014). A simple and cost-effective method of DNA extraction from small formalin-fixed paraffin-embedded tissue for molecular oncologic testing. BMC Clin. Pathol. *14*, 30.

Snyder-Mackler, N., Majoros, W.H., Yuan, M.L.M.L.M.L., Shaver, A.O., Gordon, J.B., Kopp, G.H., Schlebusch, S.A., Wall, J.D., Alberts, S.C., Mukherjee, S., et al. (2016). Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. Genetics *203*, 699–714.

Solis-Moruno, M., de Manuel, M., Hernandez-Rodriguez, J., Fontsere, C., Gomara-Castaño, A., Valsera-Naranjo, C., Crailsheim, D., Navarro, A., Llorente, M., Riera, L., et al. (2017). Potential damaging mutation in LRP5 from genome sequencing of the first reported chimpanzee with the Chiari malformation. Sci. Rep. *7*, 15224.

Stearns, S.C. (2012). Evolutionary medicine: its scope, interest and potential. Proceedings. Biol. Sci. *279*, 4305–4321.

Steiner, C.C., Putnam, A.S., Hoeck, P.E.A., and Ryder, O.A. (2013). Conservation Genomics of Threatened Animal Species. Annu. Rev. Anim. Biosci. *1*, 261–281.

Steiper, M.E., and Seiffert, E.R. (2012). Evidence for a convergent slowdown in primate molecular rates and its implications for the timing of early primate evolution. Proc. Natl. Acad. Sci. U. S. A. *109*, 6006–6011.

Stevens, N.J., Seiffert, E.R., O'Connor, P.M., Roberts, E.M., Schmitz, M.D., Krause, C., Gorscak, E., Ngasala, S., Hieronymus, T.L., and Temu, J. (2013). Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. Nature *497*, 611–614.

Sulonen, A.-M., Ellonen, P., Almusa, H., Lepistö, M., Eldfors, S., Hannula, S., Miettinen, T., Tyynismaa, H., Salo, P., Heckman, C., et al. (2011). Comparison of solution-based exome capture methods for next generation sequencing. Genome Biol. *12*, R94.

Swenson, J.E., Taberlet, P., and Bellemain, E. (2011). Genetics and conservation of European brown bears Ursus arctos. Mamm. Rev. *41*, 87–98.

Taberlet, Waits, and Luikart (1999). Noninvasive genetic sampling: look before you leap. Trends Ecol. Evol. *14*, 323–327.

Thalmann, O., Serre, D., Hofreiter, M., Lukas, D., Eriksson, J., and Vigilant, L. (2004a). Nuclear insertions help and hinder inference of the evolutionary history of gorilla mtDNA. Mol. Ecol. *14*, 179–188.

Thalmann, O., Hebler, J., Poinar, H.N., Pääbo, S., and Vigilant, L. (2004b). Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. Mol.

Ecol. *13*, 321–335.

Thalmann, O., Fischer, A., Lankester, F., Pääbo, S., and Vigilant, L. (2007). The complex evolutionary history of gorillas: insights from genomic data. Mol. Biol. Evol. *24*, 146–158.

The Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. Nature *437*, 69–87.

Thung, S.N., Gerber, M.A., Purcell, R.H., London, W.T., Mihalik, K.B., and Popper, H. (1981). Animal model of human disease. Chimpanzee carriers of hepatitis B virus. Chimpanzee hepatitis B carriers. Am. J. Pathol. *105*, 328–332.

Tung, J., Alberts, S.C., and Wray, G.A. (2010). Evolutionary genetics in wild primates: combining genetic approaches with field studies of natural populations. Trends Genet. *26*, 353–362.

van der Valk, T., Lona Durazo, F., Dalén, L., and Guschanski, K. (2017). Whole mitochondrial genome capture from faecal samples and museum-preserved specimens. Mol. Ecol. Resour. *17*, e111–e121.

Vallet, D., Petit, E.J., Gatti, S., Levréro, F., and Ménard, N. (2008). A new 2CTAB/PCI method improves DNA amplification success from faeces of Mediterranean (Barbary macaques) and tropical (lowland gorillas) primates. Conserv. Genet. *9*, 677–680.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The Sequence of the Human Genome. Science (80-.). *291*, 1304–1351.

Vigilant, L., and Guschanski, K. (2009). Using genetics to understand the dynamics of wild primate populations. Primates *50*, 105–120.

Vigilant, L., Hofreiter, M., Siedel, H., and Boesch, C. (2001). Paternity and relatedness in wild chimpanzee communities. Proc. Natl. Acad. Sci. *98*, 12890–12895.

Vynne, C., Skalski, J.R., Machado, R.B., Groom, M.J., Jácomo, A.T.A., Marinho-filho, J., Neto, M.B.R., Pomilla, C., Silveira, L., Smith, H., et al. (2011). Effectiveness of Scat-Detection Dogs in Determining Species Presence in a Tropical Savanna Landscape.

Conserv. Biol. *25*, 154–162.

Waits, L.P., and Paetkau, D. (2005). Noninvasive Genetic Sampling Tools for Wildlife Biologists: A Review of Applications and Recommendations for Accurate Data Collection. J. Wildl. Manage. *69*, 1419–1433.

Wall, J.D., Schlebusch, S.A., Alberts, S.C., Cox, L.A., Snyder-Mackler, N., Nevonen, K.A., Carbone, L., and Tung, J. (2016). Genomewide ancestry and divergence patterns from low-coverage sequencing data reveal a complex history of admixture in wild baboons. Mol. Ecol. *25*, 3469–3483.

Walsh, P.D., Abernethy, K.A., Bermejo, M., Beyers, R., De Wachter, P., Akou, M.E., Huijbregts, B., Mambounga, D.I., Toham, A.K., Kilbourn, A.M., et al. (2003). Catastrophic ape decline in western equatorial Africa. Nature *422*, 611–614.

Walsh, P.D., Kurup, D., Hasselschwert, D.L., Wirblich, C., Goetzmann, J.E., and Schnell, M.J. (2017). The Final (Oral Ebola) Vaccine Trial on Captive Chimpanzees? Sci. Rep. *7*, 43339.

Walsh, P.S., Metzger, D.A., and Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques *10*, 506–513.

Wasser, S.K., Davenport, B., Ramage, E.R., Hunt, K.E., Parker, M., Clarke, C., and Stenhouse, G. (2004). Scat detection dogs in wildlife research and management: application to grizzly and black bears in the Yellowhead Ecosystem, Alberta, Canada. Can. J. Zool. *82*, 475–492.

Watts, D.P., and Amsler, S.J. (2013). Chimpanzee-Red Colobus Encounter Rates Show a Red Colobus Population Decline Associated With Predation by Chimpanzees at Ngogo. J. Primatol *75*, 927–937.

Watts, D.P., and Mitani, J.C. (2015). Hunting and Prey Switching by Chimpanzees (Pan troglodytes schweinfurthii) at Ngogo. Int. J. Primatol. *36*, 728–748.

Wich, S.A., and Marshall, A.J. (2016). An introduction to primate conservation (Oxford University Press).

Wiens, J.J., Kuczynski, C.A., and Stephens, P.R. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation.

Wikberg, E.C., Ting, N., and Sicotte, P. (2014). Kinship and similarity in residency status structure female social networks in black-and-white colobus monkeys (colobus vellerosus). Am. J. Phys. Anthropol. *153*, 365–376.

Wildman, D.E., Uddin, M., Liu, G., Grossman, L.I., and Goodman, M. (2003). Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus Homo. Proc. Natl. Acad. Sci. U. S. A. *100*, 7181–7188.

Wultsch, C., Waits, L.P., and Kelly, M.J. (2014). Noninvasive individual and species identification of jaguars (*Panthera onca*), pumas (*Puma concolor*) and ocelots (*Leopardus pardalis*) in Belize, Central America using cross-species microsatellites and faecal DNA. Mol. Ecol. Resour. *14*, 1171–1182.

Wultsch, C., Waits, L.P., Hallerman, E.M., and Kelly, M.J. (2015). Optimizing collection methods for noninvasive genetic sampling of neotropical felids. Wildl. Soc. Bull. *39*, 403–412.

Xue, C., Raveendran, M., Harris, R.A., Fawcett, G.L., Liu, X., White, S., Dahdouli, M., Rio Deiros, D., Below, J.E., Salerno, W., et al. (2016). The population genomics of rhesus macaques (Macaca mulatta) based on whole-genome sequences. Genome Res. *26*, 1651–1662.

Yu, N., Jensen-Seaman, M.I., Chemnick, L., Ryder, O., and Li, W.-H. (2004). Nucleotide Diversity in Gorillas. Genetics *166*.

8. Electronic Appendix

8.1. Supplementary information for section 4.1.

The impact of endogenous content, replicates and pooling on genome capture from faecal samples.

Hernandez-Rodriguez[1], J., Arandjelovic[2], M., Lester[2], J., de Filippo[2], C., Weihmann[3], A., Meyer[3], M., Angedakin[2], S., Casals[4], F., Navarro[1,5,6], A., Vigilant[2], L., Kühl[2,7], H.S., Langergraber[8], K., Boesch[2], C., Hughes[1,9§], D., Marques-Bonet[1,5,6§], T.

[1]Institut de Biologia Evolutiva (Universitat Pompeu Fabra/CSIC), Departament de Ciencies Experimentals i de la Salut, Barcelona, 08003, Spain
[2]Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Leipzig, 04103, Germany
[3]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, 04103, Germany
[4]Genomics Core Facility, Departament de Ciencies Experimentals i de la Salut, Universitat Pompeu Fabra, Parc de Recerca Biomèdica de Barcelona, Barcelona, 08003, Spain
[5]Centro Nacional de Análisis Genómico (CNAG), Barcelona, 08028, Spain
[6]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, 08010, Spain
[7]German Centre for Integrative Biodiversity Research (iDiv) Halle-Leipzig-Jena, 04103, Leipzig, Germany
[8]School of Human Evolution & Social Change, Arizona State University, Tempe, USA
[9]MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[§] Corresponding authors

Corresponding authors: David Hughes, Oakfield House, 6-10 Oakfield Grove, BS8 2BN, Bristol, UK; Tel: +44 (0117)3310142, +44 (0117)3310090, hughes.evoanth@gmail.com. Tomas Marques-Bonet, Carrer del Doctor Aiguader 88, 08003, Barcelona, Spain; Tel: +34 933160887, Fax: +34 933160901, tomas.marques@upf.edu.

File S1

Following the PCA analysis from the 17 samples we observed that the sample N189-10_LR16 was separated from the rest of samples by the first component, indicating that this sample may be from another species (Supplementary Figure S5). After performing an allele balance analysis (Supplementary Figure S6), we noticed that the proportion of heterozygotes from N189-10_LR16 was different from the rest of samples, pointing to a contamination (some of the samples are flat due to the low number of heterozygote SNPs available for them).

Even though we were capturing exome, we were able to retrieve mitochondrial reads to be used for a contamination analysis. We mapped the mitochondrial reads from all samples against different mitochondrial genomes of different species of primates and discovered that for this specific sample, the proportion of mitochondrial reads that mapped against the baboon (*Papio anubis*) mitochondrial genome was higher than for the chimpanzee genome (Supplementary Table S1).

Given these results we conclude that this sample may contain within it the digested remains and thus DNA source material from a member of the Cercopithecidae family, as has been previously describe in the literature as a hunting and predating behaviour observed in chimpanzees from Kibale National Park, Uganda (Watts and Amsler, 2013; Watts and Mitani, 2015), where the Eastern chimpanzees of this study belong.

Figure S1

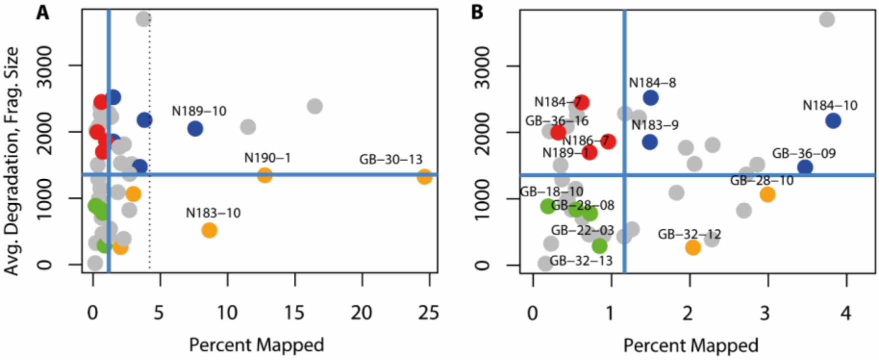


Figure S1. Sample distribution by average of degradation and percentage of endogenous DNA divided in 4 quadrants (blue line). Samples were selected in order to have the four possible combinations that represent faecal samples, samples that exhibit 1) low average of fragmentation and low percentage of endogenous content, 2) low average of fragmentation and high percentage of endogenous content, 3) high average of fragmentation and low percentage of endogenous content, and 4) high average of fragmentation and high percentage of endogenous content. Samples selected for this study are highlighted in colours and the name of the sample. **A)** All samples from the initial subset of 48 collected samples used as the initial screening pool. **B)** Zoomed version of the left side of the plot.

Figure S2

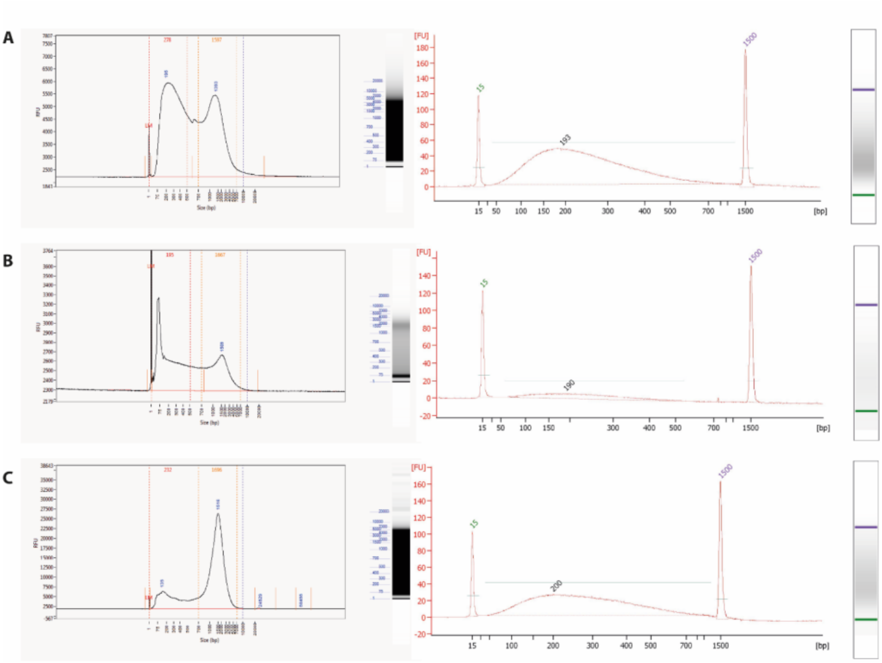


Figure S2. Fragment analyzer (before the shearing) and Bioanalyzer (after the shearing) with 3 types of samples. **A)** Fragment analyzer (left) and Bioanalyzer (right) from a sample with DNA below and above 500 bp. **B)** Fragment analyzer (left) and Bioanalyzer (right) from a sample with almost all the DNA below 500 bp. **C)** Fragment analyzer (left) and Bioanalyzer (right) from a sample with almost all DNA above 500 bp.

Figure S3

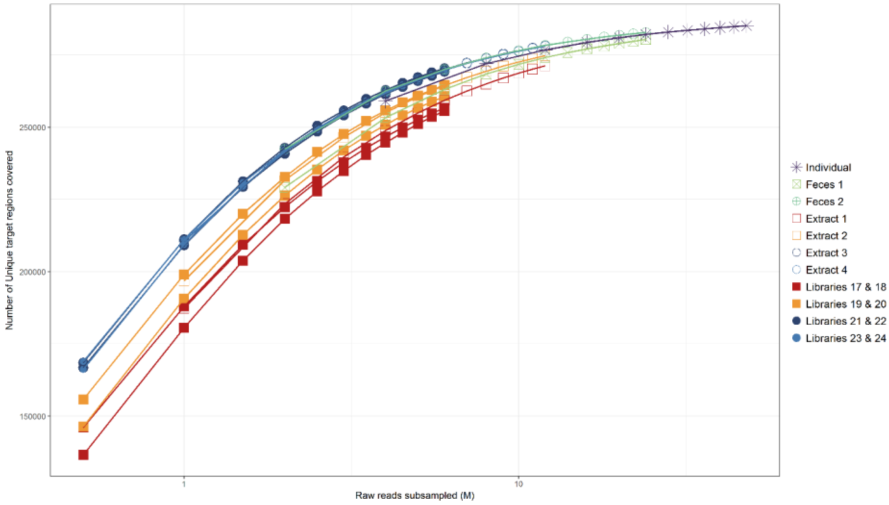


Figure S3. Molecule diversity of double capture in *Experiment 2*, comparing the unique target regions covered from different libraries, extractions and faeces from the same individual. The number of reliable reads on-target were merged by extract, faeces and individual, selected the unique target regions covered at least by one read.

Figure S4

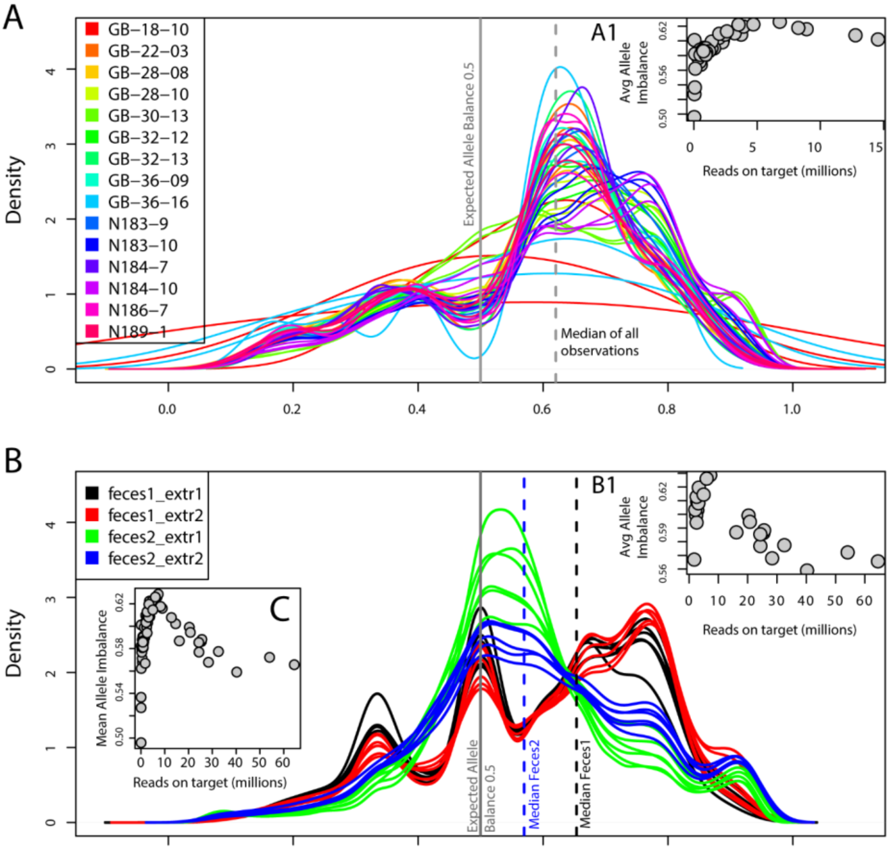


Figure S4. A) Analysis of the allele imbalance in *Experiment* 1. **A1)** Dot plot of the relationship between average allele imbalance and number of reliable reads on target in Experiment 1. **B)** Analysis of the allele imbalance in *Experiment* 2, **B1)** Dot plot of the relationship between average allele imbalance average and number of reliable reads on target in *Experiment* 2. **C)** Dot plot of the relationship between allele imbalance with all data from *Experiment* 1 and 2 (except the contaminated sample N189-10_LR16).

Figure S5

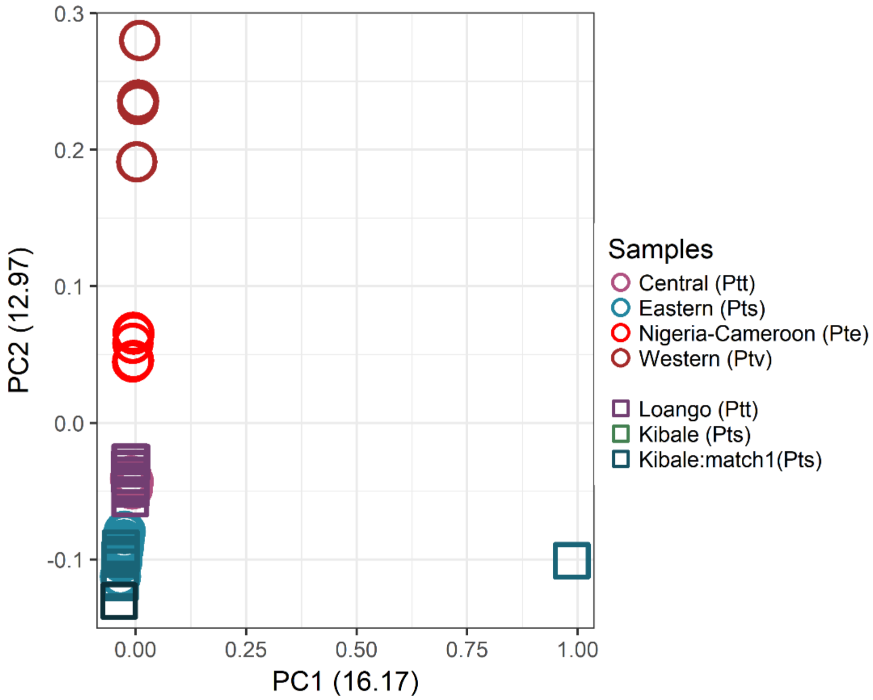


Figure S5. Principal component analysis with the 17 different individuals. The sample N189-10_LR16 is represented on the right, separated from the rest of the chimpanzee samples. The first principal component has the highest variability, in this case indicating that this isolated sample is more variable from the rest.

Figure S6

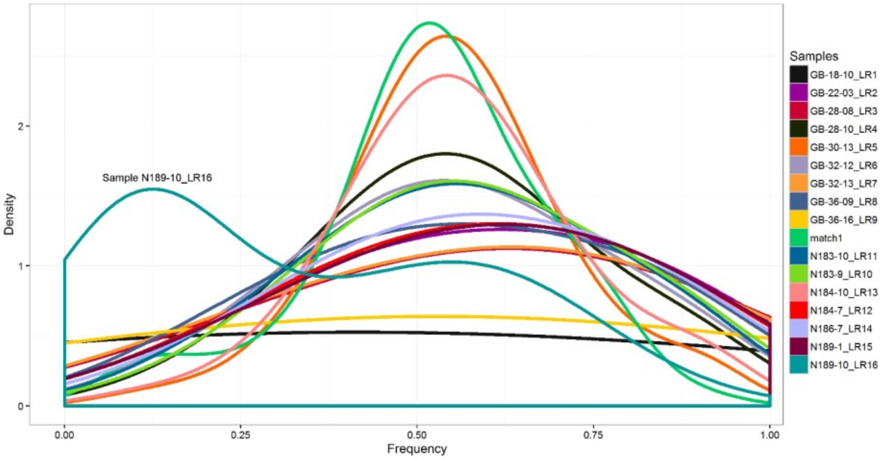


Figure S6. Analysis of the allele balance, merging all the reads of the same library from the 3 lanes, reveals differences in the proportion of heterozygosity. These results suggest that there has been some contamination in sample N189-10_LR16.

Table S1

	Green monkey	*Gorilla*	*Human*	*Crab-eating macaque*	*Rhesus macaque*	***Chimpanzee***	*Olive baboon*	*Hamadryas baboon*	*Orangutan*	*Squirrel monkey*	*Philippine tarsier*
GB-18-10_LR1	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%
GB-22-03_LR2	6%	5%	6%	5%	1%	55%	8%	4%	2%	5%	4%
GB-28-08_LR3	0%	20%	0%	13%	7%	53%	0%	0%	0%	7%	0%
GB-28-10_LR4	0%	18%	12%	5%	0%	39%	6%	6%	5%	11%	0%
GB-30-13_LR5	9%	6%	4%	6%	3%	43%	7%	4%	4%	8%	6%
GB-32-12_LR6	0%	6%	2%	2%	2%	70%	6%	6%	0%	6%	0%
GB-32-13_LR7	0%	12%	0%	0%	4%	85%	0%	0%	0%	0%	0%
GB-36-09_LR8	0%	31%	0%	12%	4%	38%	0%	8%	0%	8%	0%
GB-36-16_LR9	0%	0%	0%	6%	0%	88%	0%	0%	0%	6%	0%
N183-9_LR10	0%	11%	7%	3%	5%	59%	3%	0%	0%	11%	0%
N183-10_LR11	0%	26%	1%	13%	5%	41%	2%	1%	1%	9%	0%
N184-7_LR12	0%	2%	0%	4%	0%	89%	0%	0%	0%	5%	0%
N184-10_LR13	3%	9%	3%	6%	1%	48%	6%	6%	3%	16%	0%
N186-7_LR14	0%	16%	0%	9%	5%	50%	0%	0%	0%	20%	0%
N189-1_LR15	0%	18%	0%	9%	0%	55%	0%	0%	0%	18%	0%
N189-10_LR16	10%	3%	0%	1%	7%	9%	41%	18%	6%	4%	0%
N184-8_LR17	0%	11%	3%	5%	0%	61%	0%	1%	7%	9%	2%
N184-8-2_LR18	0%	28%	4%	15%	0%	33%	4%	0%	0%	15%	0%
N184-8-3_LR19	0%	15%	0%	2%	0%	65%	0%	0%	0%	19%	0%
N184-8-4_LR20	1%	17%	4%	9%	1%	50%	3%	2%	2%	10%	0%
N190-1_LR21	1%	16%	3%	6%	1%	50%	6%	5%	3%	9%	0%
N190-1-2_LR22	1%	11%	3%	4%	1%	61%	4%	4%	3%	7%	1%
N190-1-3_LR23	0%	20%	2%	6%	1%	52%	5%	4%	4%	6%	0%
N190-1-4_LR24	0%	14%	4%	5%	1%	60%	3%	3%	3%	5%	1%

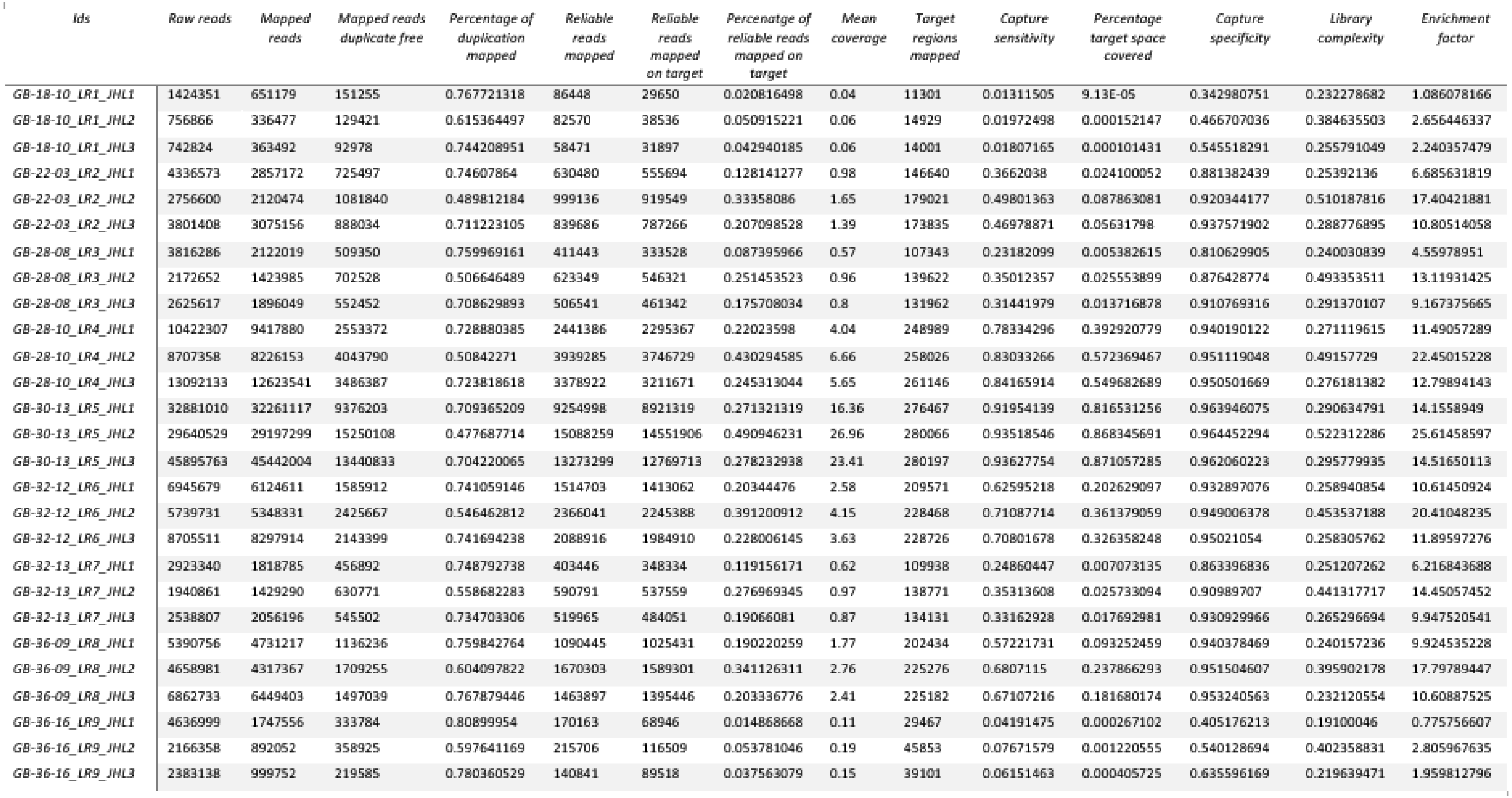
Table S1. Contamination analysis with mitochondrial (MT) genome. Mapping percentage of the MT reads of all samples to 11 primates' MT genomes. In the sample N189-10_LR16 the percentage of reads that mapped against Olive baboon were 41% compared with the 9% that mapped against Chimpanzee. Note that in all samples there are reads that mapped against other primates' genomes, due to the conservative regions in the MT genome.

Table S2. **Data from all samples with the information for statistical analysis and interpretation (continues).**

Ids	*Sample*	*Library*	*Subspecies*	*Sample park*	*Country*	*Coverage*	*Percentage endogenous DNA*	*Average of degradation*	*Initial ug of DNA*	*Experiment*	*Feces*	*Extract*	*Pool*	*Hybridization*	*Hybridization type*	*Lane*
GB-18-10_LR1_JHL1	GB-18-10	LR1	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.1853	887	3.95	exp1	fec1	ext1	pool1	hyb1	double	lane1
GB-18-10_LR1_JHL2	GB-18-10	LR1	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.1853	887	3.95	exp1	fec1	ext1	pool2	hyb2	double	lane2
GB-18-10_LR1_JHL3	GB-18-10	LR1	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.1853	887	3.95	exp1	fec1	ext1	pool3	hyb3	double	lane3
GB-22-03_LR2_JHL1	GB-22-03	LR2	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.72518	780	0.9	exp1	fec2	ext2	pool1	hyb1	double	lane1
GB-22-03_LR2_JHL2	GB-22-03	LR2	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.72518	780	0.9	exp1	fec2	ext2	pool2	hyb2	double	lane2
GB-22-03_LR2_JHL3	GB-22-03	LR2	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.72518	780	0.9	exp1	fec2	ext2	pool3	hyb3	double	lane3
GB-28-08_LR3_JHL1	GB-28-08	LR3	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.54823	837	1.2	exp1	fec3	ext3	pool1	hyb1	double	lane1
GB-28-08_LR3_JHL2	GB-28-08	LR3	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.54823	837	1.2	exp1	fec3	ext3	pool2	hyb2	double	lane2
GB-28-08_LR3_JHL3	GB-28-08	LR3	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.54823	837	1.2	exp1	fec3	ext3	pool3	hyb3	double	lane3
GB-28-10_LR4_JHL1	GB-28-10	LR4	Pan troglodytes troglodytes	Loango National park	Gabon	4	2.98666	1066	0.43	exp1	fec4	ext4	pool1	hyb1	double	lane1
GB-28-10_LR4_JHL2	GB-28-10	LR4	Pan troglodytes troglodytes	Loango National park	Gabon	4	2.98666	1066	0.43	exp1	fec4	ext4	pool2	hyb2	double	lane2
GB-28-10_LR4_JHL3	GB-28-10	LR4	Pan troglodytes troglodytes	Loango National park	Gabon	4	2.98666	1066	0.43	exp1	fec4	ext4	pool3	hyb3	double	lane3
GB-30-13_LR5_JHL1	GB-30-13	LR5	Pan troglodytes troglodytes	Loango National park	Gabon	4	24.6136	1323	0.36	exp1	fec5	ext5	pool1	hyb1	double	lane1
GB-30-13_LR5_JHL2	GB-30-13	LR5	Pan troglodytes troglodytes	Loango National park	Gabon	4	24.6136	1323	0.36	exp1	fec5	ext5	pool2	hyb2	double	lane2
GB-30-13_LR5_JHL3	GB-30-13	LR5	Pan troglodytes troglodytes	Loango National park	Gabon	4	24.6136	1323	0.36	exp1	fec5	ext5	pool3	hyb3	double	lane3
GB-32-12_LR6_JHL1	GB-32-12	LR6	Pan troglodytes troglodytes	Loango National park	Gabon	4	2.0424	265	1.62	exp1	fec6	ext6	pool1	hyb1	double	lane1
GB-32-12_LR6_JHL2	GB-32-12	LR6	Pan troglodytes troglodytes	Loango National park	Gabon	4	2.0424	265	1.62	exp1	fec6	ext6	pool2	hyb2	double	lane2
GB-32-12_LR6_JHL3	GB-32-12	LR6	Pan troglodytes troglodytes	Loango National park	Gabon	4	2.0424	265	1.62	exp1	fec6	ext6	pool3	hyb3	double	lane3
GB-32-13_LR7_JHL1	GB-32-13	LR7	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.84753	288	1.26	exp1	fec7	ext7	pool1	hyb1	double	lane1
GB-32-13_LR7_JHL2	GB-32-13	LR7	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.84753	288	1.26	exp1	fec7	ext7	pool2	hyb2	double	lane2
GB-32-13_LR7_JHL3	GB-32-13	LR7	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.84753	288	1.26	exp1	fec7	ext7	pool3	hyb3	double	lane3
GB-36-09_LR8_JHL1	GB-36-09	LR8	Pan troglodytes troglodytes	Loango National park	Gabon	4	3.4693	1471	2.21	exp1	fec8	ext8	pool1	hyb1	double	lane1
GB-36-09_LR8_JHL2	GB-36-09	LR8	Pan troglodytes troglodytes	Loango National park	Gabon	4	3.4693	1471	2.21	exp1	fec8	ext8	pool2	hyb2	double	lane2
GB-36-09_LR8_JHL3	GB-36-09	LR8	Pan troglodytes troglodytes	Loango National park	Gabon	4	3.4693	1471	2.21	exp1	fec8	ext8	pool3	hyb3	double	lane3
GB-36-16_LR9_JHL1	GB-36-16	LR9	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.31963	1998	1.45	exp1	fec9	ext9	pool1	hyb1	double	lane1
GB-36-16_LR9_JHL2	GB-36-16	LR9	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.31963	1998	1.45	exp1	fec9	ext9	pool2	hyb2	double	lane2
GB-36-16_LR9_JHL3	GB-36-16	LR9	Pan troglodytes troglodytes	Loango National park	Gabon	4	0.31963	1998	1.45	exp1	fec9	ext9	pool3	hyb3	double	lane3

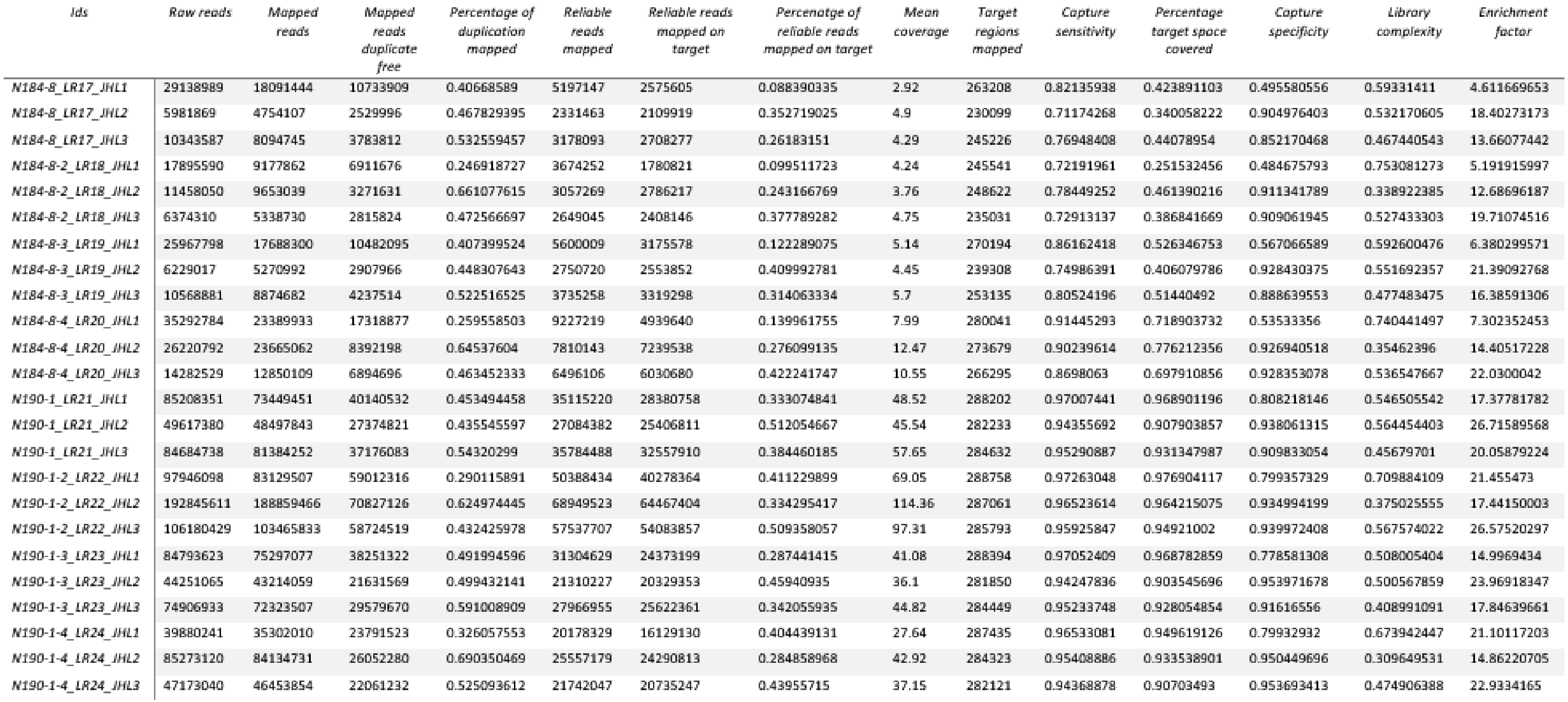
Ids	*Sample*	*Library*	*Subspecies*	*Sample park*	*Country*	*Coverage*	*Percentage endogenous DNA*	*Average of degradation*	*Initial ug of DNA*	*Experiment*	*Feces*	*Extract*	*Pool*	*Hybridization*	*Hybridization type*	*Lane*
N183-9_LR10_JHL1	N183-9	LR10	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	1.49119	1854	3.34	exp1	fec10	ext10	pool1	hyb1	double	lane1
N183-9_LR10_JHL2	N183-9	LR10	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	1.49119	1854	3.34	exp1	fec10	ext10	pool2	hyb2	double	lane2
N183-9_LR10_JHL3	N183-9	LR10	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	1.49119	1854	3.34	exp1	fec10	ext10	pool3	hyb3	double	lane3
N183-10_LR11_JHL1	N183-10	LR11	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	8.63515	517	0.65	exp1	fec11	ext11	pool1	hyb1	double	lane1
N183-10_LR11_JHL2	N183-10	LR11	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	8.63515	517	0.65	exp1	fec11	ext11	pool2	hyb2	double	lane2
N183-10_LR11_JHL3	N183-10	LR11	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	8.63515	517	0.65	exp1	fec11	ext11	pool3	hyb3	double	lane3
N184-7_LR12_JHL1	N184-7	LR12	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	0.6239	2452	1	exp1	fec12	ext12	pool1	hyb1	double	lane1
N184-7_LR12_JHL2	N184-7	LR12	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	0.6239	2452	1	exp1	fec12	ext12	pool2	hyb2	double	lane2
N184-7_LR12_JHL3	N184-7	LR12	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	0.6239	2452	1	exp1	fec12	ext12	pool3	hyb3	double	lane3
N184-10_LR13_JHL1	N184-10	LR13	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	3.83135	2176	1.43	exp1	fec13	ext13	pool1	hyb1	double	lane1
N184-10_LR13_JHL2	N184-10	LR13	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	3.83135	2176	1.43	exp1	fec13	ext13	pool2	hyb2	double	lane2
N184-10_LR13_JHL3	N184-10	LR13	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	3.83135	2176	1.43	exp1	fec13	ext13	pool3	hyb3	double	lane3
N186-7_LR14_JHL1	N186-7	LR14	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	0.95551	1860	4.46	exp1	fec14	ext14	pool1	hyb1	double	lane1
N186-7_LR14_JHL2	N186-7	LR14	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	0.95551	1860	4.46	exp1	fec14	ext14	pool2	hyb2	double	lane2
N186-7_LR14_JHL3	N186-7	LR14	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	0.95551	1860	4.46	exp1	fec14	ext14	pool3	hyb3	double	lane3
N189-1_LR15_JHL1	N189-1	LR15	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	0.71782	1699	2.13	exp1	fec15	ext15	pool1	hyb1	double	lane1
N189-1_LR15_JHL2	N189-1	LR15	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	0.71782	1699	2.13	exp1	fec15	ext15	pool2	hyb2	double	lane2
N189-1_LR15_JHL3	N189-1	LR15	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	0.71782	1699	2.13	exp1	fec15	ext15	pool3	hyb3	double	lane3
N189-10_LR16_JHL1	N189-10	LR16	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	7.58331	2051	2.27	exp1	fec16	ext16	pool1	hyb1	double	lane1
N189-10_LR16_JHL2	N189-10	LR16	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	7.58331	2051	2.27	exp1	fec16	ext16	pool2	hyb2	double	lane2
N189-10_LR16_JHL3	N189-10	LR16	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	7.58331	2051	2.27	exp1	fec16	ext16	pool3	hyb3	double	lane3

Ids	*Sample*	*Library*	*Subspecies*	*Sample park*	*Country*	*Coverage*	*Percentage endogenous DNA*	*Average of degradation*	*Initial ug of DNA*	*Experiment*	*Feces*	*Extract*	*Pool*	*Hybridization*	*Hybridization type*	*Lane*
N184-8_LR17_JHL1	N184-8-1	LR17	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	1.49824	2523	1.56	exp2	fec17	ext17	poolA	hyb4	single	lane1
N184-8_LR17_JHL2	N184-8-1	LR17	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	1.49824	2523	1.56	exp2	fec17	ext17	poolA	hyb5	double	lane2
N184-8_LR17_JHL3	N184-8-1	LR17	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	1.49824	2523	1.56	exp2	fec17	ext17	poolA	hyb6	double	lane3
N184-8-2_LR18_JHL1	N184-8-2	LR18	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	1.49824	2523	1.56	exp2	fec17	ext17	poolB	hyb7	single	lane1
N184-8-2_LR18_JHL2	N184-8-2	LR18	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	1.49824	2523	1.56	exp2	fec17	ext17	poolB	hyb8	double	lane2
N184-8-2_LR18_JHL3	N184-8-2	LR18	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	1.49824	2523	1.56	exp2	fec17	ext17	poolB	hyb9	double	lane3
N184-8-3_LR19_JHL1	N184-8-3	LR19	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.31	exp2	fec17	ext18	poolA	hyb4	single	lane1
N184-8-3_LR19_JHL2	N184-8-3	LR19	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.31	exp2	fec17	ext18	poolA	hyb5	double	lane2
N184-8-3_LR19_JHL3	N184-8-3	LR19	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.31	exp2	fec17	ext18	poolA	hyb6	double	lane3
N184-8-4_LR20_JHL1	N184-8-4	LR20	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.31	exp2	fec17	ext18	poolB	hyb7	single	lane1
N184-8-4_LR20_JHL2	N184-8-4	LR20	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.31	exp2	fec17	ext18	poolB	hyb8	double	lane2
N184-8-4_LR20_JHL3	N184-8-4	LR20	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.31	exp2	fec17	ext18	poolB	hyb9	double	lane3
N190-1_LR21_JHL1	N190-1-1	LR21	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	12.748	1346	4.03	exp2	fec18	ext19	poolA	hyb4	single	lane1
N190-1_LR21_JHL2	N190-1-1	LR21	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	12.748	1346	4.03	exp2	fec18	ext19	poolA	hyb5	double	lane2
N190-1_LR21_JHL3	N190-1-1	LR21	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	12.748	1346	4.03	exp2	fec18	ext19	poolA	hyb6	double	lane3
N190-1-2_LR22_JHL1	N190-1-2	LR22	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	12.748	1346	4.03	exp2	fec18	ext19	poolB	hyb7	single	lane1
N190-1-2_LR22_JHL2	N190-1-2	LR22	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	12.748	1346	4.03	exp2	fec18	ext19	poolB	hyb8	double	lane2
N190-1-2_LR22_JHL3	N190-1-2	LR22	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	12.748	1346	4.03	exp2	fec18	ext19	poolB	hyb9	double	lane3
N190-1-3_LR23_JHL1	N190-1-3	LR23	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.82	exp2	fec18	ext20	poolA	hyb4	single	lane1
N190-1-3_LR23_JHL2	N190-1-3	LR23	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.82	exp2	fec18	ext20	poolA	hyb5	double	lane2
N190-1-3_LR23_JHL3	N190-1-3	LR23	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.82	exp2	fec18	ext20	poolA	hyb6	double	lane3
N190-1-4_LR24_JHL1	N190-1-4	LR24	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.82	exp2	fec18	ext20	poolB	hyb7	single	lane1
N190-1-4_LR24_JHL2	N190-1-4	LR24	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.82	exp2	fec18	ext20	poolB	hyb8	double	lane2
N190-1-4_LR24_JHL3	N190-1-4	LR24	Pan troglodytes schweinfurthii	Kibale National park	Uganda	4	NA	NA	1.82	exp2	fec18	ext20	poolB	hyb9	double	lane3



Ids	Raw reads	Mapped reads	Mapped reads duplicate free	Percentage of duplication mapped	Reliable reads mapped	Reliable reads mapped on target	Percenatge of reliable reads mapped on target	Mean coverage	Target regions mapped	Capture sensitivity	Percentage target space covered	Capture specificity	Library complexity	Enrichment factor
GB-18-10_LR1_JHL1	1424351	651179	151255	0.767721318	86448	29650	0.020816498	0.04	11301	0.01311505	9.13E-05	0.342980751	0.232278682	1.086078166
GB-18-10_LR1_JHL2	756866	336477	129421	0.615364497	82570	38536	0.050915221	0.06	14929	0.01972498	0.000152147	0.466707036	0.384635503	2.656446337
GB-18-10_LR1_JHL3	742824	363492	92978	0.744208951	58471	31897	0.042940185	0.06	14001	0.01807165	0.000101431	0.545518291	0.255791049	2.240357479
GB-22-03_LR2_JHL1	4336573	2857172	725497	0.74607864	630480	555694	0.128141277	0.98	146640	0.3662038	0.024100052	0.881382439	0.25392136	6.685631819
GB-22-03_LR2_JHL2	2756600	2120474	1081840	0.489812184	999136	919549	0.33358086	1.65	179021	0.49801363	0.087863081	0.920344177	0.510187816	17.40421881
GB-22-03_LR2_JHL3	3801408	3075156	888034	0.711223105	839686	787266	0.207098528	1.39	173835	0.46978871	0.05631798	0.937571902	0.288776895	10.80514058
GB-28-08_LR3_JHL1	3816286	2122019	509350	0.759969161	411443	333528	0.087395966	0.57	107343	0.23182099	0.005382615	0.810629905	0.240030839	4.55978951
GB-28-08_LR3_JHL2	2172652	1423985	702528	0.506646489	623349	546321	0.251453523	0.96	139622	0.35012357	0.025553899	0.876428774	0.493353511	13.11931425
GB-28-08_LR3_JHL3	2625617	1896049	552452	0.708629893	506541	461342	0.175708034	0.8	131962	0.31441979	0.013716878	0.910769316	0.291370107	9.167375665
GB-28-10_LR4_JHL1	10422307	9417880	2553372	0.728880385	2441386	2295367	0.22023598	4.04	248989	0.78334296	0.392920779	0.940190122	0.271119615	11.49057289
GB-28-10_LR4_JHL2	8707358	8226153	4043790	0.50842271	3939285	3746729	0.430294585	6.66	258026	0.83033266	0.572369467	0.951119048	0.49157729	22.45015228
GB-28-10_LR4_JHL3	13092133	12623541	3486387	0.723818618	3378922	3211671	0.245313044	5.65	261146	0.84165914	0.549682689	0.950501669	0.276181382	12.79894143
GB-30-13_LR5_JHL1	32881010	32261117	9376203	0.709365209	9254998	8921319	0.271321319	16.36	276467	0.91954139	0.816531256	0.963946075	0.290634791	14.1558949
GB-30-13_LR5_JHL2	29640529	29197299	15250108	0.477687714	15088259	14551906	0.490946231	26.96	280066	0.93518546	0.868345691	0.964452294	0.522312286	25.61458597
GB-30-13_LR5_JHL3	45895763	45442004	13440833	0.704220065	13273299	12769713	0.278232938	23.41	280197	0.93627754	0.871057285	0.962060223	0.295779935	14.51650113
GB-32-12_LR6_JHL1	6945679	6124611	1585912	0.741059146	1514703	1413062	0.20344476	2.58	209571	0.62595218	0.202629097	0.932897076	0.258940854	10.61450924
GB-32-12_LR6_JHL2	5739731	5348331	2425667	0.546462812	2366041	2245388	0.391200912	4.15	228468	0.71087714	0.361379059	0.949006378	0.453537188	20.41048235
GB-32-12_LR6_JHL3	8705511	8297914	2143399	0.741694238	2088916	1984910	0.228006145	3.63	228726	0.70801678	0.326358248	0.95021054	0.258305762	11.89597276
GB-32-13_LR7_JHL1	2923340	1818785	456892	0.748792738	403446	348334	0.119156171	0.62	109938	0.24860447	0.007073135	0.863396836	0.251207262	6.216843688
GB-32-13_LR7_JHL2	1940861	1429290	630771	0.558682283	590791	537559	0.276969345	0.97	138771	0.35313608	0.025733094	0.90989707	0.441317717	14.45057452
GB-32-13_LR7_JHL3	2538807	2056196	545502	0.734703306	519965	484051	0.19066081	0.87	134131	0.33162928	0.017692981	0.930929966	0.265296694	9.947520541
GB-36-09_LR8_JHL1	5390756	4731217	1136236	0.759842764	1090445	1025431	0.190220259	1.77	202434	0.57221731	0.093252459	0.940378469	0.240157236	9.924535228
GB-36-09_LR8_JHL2	4658981	4317367	1709255	0.604097822	1670303	1589301	0.341126311	2.76	225276	0.6807115	0.237866293	0.951504607	0.395902178	17.79789447
GB-36-09_LR8_JHL3	6862733	6449403	1497039	0.767879446	1463897	1395446	0.203336776	2.41	225182	0.67107216	0.181680174	0.953240563	0.232120554	10.60887525
GB-36-16_LR9_JHL1	4636999	1747556	333784	0.80899954	170163	68946	0.014868668	0.11	29467	0.04191475	0.000267102	0.405176213	0.19100046	0.775756607
GB-36-16_LR9_JHL2	2166358	892052	358925	0.597641169	215706	116509	0.053781046	0.19	45853	0.07671579	0.001220555	0.540128694	0.402358831	2.805967635
GB-36-16_LR9_JHL3	2383138	999752	219585	0.780360529	140841	89518	0.037563079	0.15	39101	0.06151463	0.000405725	0.635596169	0.219639471	1.959812796

Ids	*Raw reads*	*Mapped reads*	*Mapped reads duplicate free*	*Percentage of duplication mapped*	*Reliable reads mapped*	*Reliable reads mapped on target*	*Percenatge of reliable reads mapped on target*	*Mean coverage*	*Target regions mapped*	*Capture sensitivity*	*Percentage target space covered*	*Capture specificity*	*Library complexity*	*Enrichment factor*
N183-9_LR10_JHL1	6545690	5561500	1606289	0.711177021	1522842	1406774	0.214916075	2.42	218173	0.64415908	0.189693238	0.923781981	0.288822979	11.21301259
N183-9_LR10_JHL2	5265439	4802379	2664164	0.445240786	2591452	2452039	0.465685577	4.29	234872	0.72900627	0.386929576	0.946202747	0.554759214	24.29663881
N183-9_LR10_JHL3	7526429	7110618	2182771	0.693026541	2118823	2008909	0.266913964	3.47	237683	0.73404402	0.32896503	0.948124973	0.306973459	13.92594595
N183-10_LR11_JHL1	11605821	11372826	2773007	0.756172564	2725191	2612479	0.225100749	4.57	254117	0.81078686	0.465024834	0.958640697	0.243827436	11.7443869
N183-10_LR11_JHL2	10721320	10590977	4125876	0.610434807	4077654	3923021	0.365908396	6.88	261407	0.84907038	0.619017673	0.96207795	0.389565193	19.09087281
N183-10_LR11_JHL3	16185919	16042266	3730103	0.767482786	3674096	3526263	0.217859919	6.15	264546	0.86043067	0.60675126	0.959763436	0.232517214	11.36660446
N184-7_LR12_JHL1	4799939	3399069	887090	0.739019714	774645	683853	0.142471186	1.16	162090	0.42574391	0.041106682	0.882795345	0.260980286	7.433279247
N184-7_LR12_JHL2	3255998	2601011	1354283	0.479324386	1262611	1167510	0.358572088	2.03	191243	0.55307725	0.148319454	0.924679098	0.520675614	18.70810892
N184-7_LR12_JHL3	4333029	3744756	1074118	0.713167427	1006604	945119	0.218119703	1.63	187601	0.52784793	0.089391311	0.938918383	0.286832573	11.38015843
N184-10_LR13_JHL1	18995165	18027569	5094269	0.717417861	4964735	4709402	0.247926354	8.33	268713	0.87965188	0.685434819	0.948570669	0.282582139	12.93528803
N184-10_LR13_JHL2	16705758	16145413	8800274	0.454936582	8640982	8251859	0.493952983	14.82	272104	0.89678361	0.770021672	0.954967734	0.545063418	25.77145999
N184-10_LR13_JHL3	25530974	24908788	7309546	0.706547504	7134078	6784296	0.265728053	12.03	274874	0.90882011	0.783725027	0.950970259	0.293452496	13.86407231
N186-7_LR14_JHL1	3981167	3054718	826395	0.7294693	744464	674296	0.169371443	1.19	163160	0.42996683	0.042858061	0.905746954	0.2705307	8.836770944
N186-7_LR14_JHL2	2891281	2493382	1345763	0.460266016	1277066	1198531	0.414532866	2.14	192973	0.56010305	0.161130214	0.93850357	0.539733984	21.62780172
N186-7_LR14_JHL3	4019664	3649662	1069671	0.706912311	1023056	966016	0.240322574	1.7	190541	0.53636477	0.099456667	0.944245476	0.293087689	12.53856909
N189-1_LR15_JHL1	3114755	2460343	674066	0.726027631	595860	530377	0.170278882	0.92	144407	0.3588568	0.022480534	0.89010338	0.273972369	8.884115598
N189-1_LR15_JHL2	2324823	2031500	1048081	0.484085159	990634	923471	0.397222068	1.63	176717	0.4921881	0.098466022	0.932202004	0.515914841	20.72462964
N189-1_LR15_JHL3	3222456	2959830	842287	0.715427237	799207	748672	0.232329627	1.31	171457	0.45856704	0.053751771	0.936768572	0.284572763	12.12154575
N189-10_LR16_JHL1	9996727	8796273	2767760	0.685348556	2625684	2478335	0.247914642	4.45	210538	0.64398665	0.351743095	0.94388167	0.314651444	12.934677
N189-10_LR16_JHL2	9103033	8364603	5082521	0.392377498	4907841	4676164	0.513692964	8.51	230263	0.72404629	0.491423993	0.952794518	0.607622502	26.80137202
N189-10_LR16_JHL3	11820501	11174239	3798254	0.660088351	3646994	3471939	0.293721814	6.25	222409	0.6932653	0.440458199	0.952000195	0.339911649	15.3246164



Ids	Raw reads	Mapped reads	Mapped reads duplicate free	Percentage of duplication mapped	Reliable reads mapped	Reliable reads mapped on target	Percenatge of reliable reads mapped on target	Mean coverage	Target regions mapped	Capture sensitivity	Percentage target space covered	Capture specificity	Library complexity	Enrichment factor
N184-8_LR17_JHL1	29138989	18091444	10733909	0.40668589	5197147	2575605	0.088390335	2.92	263208	0.82135938	0.423891103	0.495580556	0.59331411	4.611669653
N184-8_LR17_JHL2	5981869	4754107	2529996	0.467829395	2331463	2109919	0.352719025	4.9	230099	0.71174268	0.340058222	0.904976403	0.532170605	18.40273173
N184-8_LR17_JHL3	10343587	8094745	3783812	0.532559457	3178093	2708277	0.26183151	4.29	245226	0.76948408	0.44078954	0.852170468	0.467440543	13.66077442
N184-8-2_LR18_JHL1	17895590	9177862	6911676	0.246918727	3674252	1780821	0.099511723	4.24	245541	0.72191961	0.251532456	0.484675793	0.753081273	5.191915997
N184-8-2_LR18_JHL2	11458050	9653039	3271631	0.661077615	3057269	2786217	0.243166769	3.76	248622	0.78449252	0.461390216	0.911341789	0.338922385	12.68696187
N184-8-2_LR18_JHL3	6374310	5338730	2815824	0.472566697	2649045	2408146	0.377789282	4.75	235031	0.72913137	0.386841669	0.909061945	0.527433303	19.71074516
N184-8-3_LR19_JHL1	25967798	17688300	10482095	0.407399524	5600009	3175578	0.122289075	5.14	270194	0.86162418	0.526346753	0.567066589	0.592600476	6.380299571
N184-8-3_LR19_JHL2	6229017	5270992	2907966	0.448307643	2750720	2553852	0.409992781	4.45	239308	0.74986391	0.406079786	0.928430375	0.551692357	21.39092768
N184-8-3_LR19_JHL3	10568881	8874682	4237514	0.522516525	3735258	3319298	0.314063334	5.7	253135	0.80524196	0.51440492	0.888639553	0.477483475	16.38591306
N184-8-4_LR20_JHL1	35292784	23389933	17318877	0.259558503	9227219	4939640	0.139961755	7.99	280041	0.91445293	0.718903732	0.53533356	0.740441497	7.302352453
N184-8-4_LR20_JHL2	26220792	23665062	8392198	0.64537604	7810143	7239538	0.276099135	12.47	273679	0.90239614	0.776212356	0.926940518	0.35462396	14.40517228
N184-8-4_LR20_JHL3	14282529	12850109	6894696	0.463452333	6496106	6030680	0.422241747	10.55	266295	0.8698063	0.697910856	0.928353078	0.536547667	22.0300042
N190-1_LR21_JHL1	85208351	73449451	40140532	0.453494458	35115220	28380758	0.333074841	48.52	288202	0.97007441	0.968901196	0.808218146	0.546505542	17.37781782
N190-1_LR21_JHL2	49617380	48497843	27374821	0.435545597	27084382	25406811	0.512054667	45.54	282233	0.94355692	0.907903857	0.938061315	0.564454403	26.71589568
N190-1_LR21_JHL3	84684738	81384252	37176083	0.54320299	35784488	32557910	0.384460185	57.65	284632	0.95290887	0.931347987	0.909833054	0.45679701	20.05879224
N190-1-2_LR22_JHL1	97946098	83129507	59012316	0.290115891	50388434	40278364	0.411229899	69.05	288758	0.97263048	0.976904117	0.799357329	0.709884109	21.455473
N190-1-2_LR22_JHL2	192845611	188859466	70827126	0.624974445	68949523	64467404	0.334295417	114.36	287061	0.96523614	0.964215075	0.934994199	0.375025555	17.44150003
N190-1-2_LR22_JHL3	106180429	103465833	58724519	0.432425978	57537707	54083857	0.509358057	97.31	285793	0.95925847	0.94921002	0.939972408	0.567574022	26.57520297
N190-1-3_LR23_JHL1	84793623	75297077	38251322	0.491994596	31304629	24373199	0.287441415	41.08	288394	0.97052409	0.968782859	0.778581308	0.508005404	14.9969434
N190-1-3_LR23_JHL2	44251065	43214059	21631569	0.499432141	21310227	20329353	0.45940935	36.1	281850	0.94247836	0.903545696	0.953971678	0.500567859	23.96918347
N190-1-3_LR23_JHL3	74906933	72323507	29579670	0.591008909	27966955	25622361	0.342055935	44.82	284449	0.95233748	0.928054854	0.91616556	0.408991091	17.84639661
N190-1-4_LR24_JHL1	39880241	35302010	23791523	0.326057553	20178329	16129130	0.404439131	27.64	287435	0.96533081	0.949619126	0.79932932	0.673942447	21.10117203
N190-1-4_LR24_JHL2	85273120	84134731	26052280	0.690350469	25557179	24290813	0.284858968	42.92	284323	0.95408886	0.933538901	0.950449696	0.309649531	14.86220705
N190-1-4_LR24_JHL3	47173040	46453854	22061232	0.525093612	21742047	20735247	0.43955715	37.15	282121	0.94368878	0.90703493	0.953693413	0.474906388	22.9334165

Ids	*Total markers genotyped*	*Markers genotyped from chimpanzee*	*Markers genotyped from chimpanzee biallelic*	*Allele imbalance mean*	*Allele imbalance median*	*Allele imbalance var*	*Heterozygosity 72 libraries*	*Heterozygosity 69 libraries (without LR16)*	*Proportion Heterozygote SNPs*
GB-18-10_LR1_JHL1	300	138	119	0.495943556	0.6	0.090495837	0.6747	0.34106	0.05042017
GB-18-10_LR1_JHL2	672	325	253	0.527132	0.511905	0.036394696	0.65541	0.2174	0.05533597
GB-18-10_LR1_JHL3	462	215	171	0.601152913	0.6	0.019337284	0.25367	-0.68848	0.11695906
GB-22-03_LR2_JHL1	72739	29253	24692	0.583951636	0.6	0.029331678	0.57234	-0.22274	0.10351531
GB-22-03_LR2_JHL2	203303	79291	66274	0.580518267	0.6	0.034034808	0.5983	-0.17895	0.1043578
GB-22-03_LR2_JHL3	150434	61057	51994	0.590806862	0.6	0.030228562	0.56629	-0.21826	0.10252712
GB-28-08_LR3_JHL1	20643	8037	6737	0.579540841	0.6	0.027222635	0.60028	-0.11618	0.10568502
GB-28-08_LR3_JHL2	73068	27201	22324	0.567577253	0.6	0.033377763	0.62849	-0.0997	0.10602939
GB-28-08_LR3_JHL3	46279	18301	15341	0.571232968	0.6	0.029962161	0.57968	-0.18019	0.10716381
GB-28-10_LR4_JHL1	780079	322794	278946	0.600192595	0.636364	0.033790627	0.4928	-0.35259	0.11179104
GB-28-10_LR4_JHL2	1179102	496590	430941	0.608555857	0.666667	0.036253453	0.47204	-0.36265	0.11099122
GB-28-10_LR4_JHL3	1128804	473449	412381	0.610321648	0.666667	0.035934672	0.45719	-0.37916	0.11228179
GB-30-13_LR5_JHL1	1930261	793098	699467	0.616003512	0.636364	0.035512723	0.42566	-0.23639	0.10607744
GB-30-13_LR5_JHL2	2077544	850602	750598	0.602150007	0.6	0.033399732	0.4631	-0.14557	0.09922799
GB-30-13_LR5_JHL3	2086593	854196	754673	0.607756977	0.606061	0.034115851	0.4555	-0.14282	0.09997005
GB-32-12_LR6_JHL1	452839	189087	161261	0.590821559	0.6	0.033145268	0.49193	-0.2923	0.11241334
GB-32-12_LR6_JHL2	776008	328201	281600	0.598074203	0.631579	0.035969775	0.46608	-0.34528	0.11424594
GB-32-12_LR6_JHL3	713471	302910	260947	0.604034693	0.666667	0.034411928	0.44724	-0.35763	0.11517959
GB-32-13_LR7_JHL1	28688	12120	10113	0.576698167	0.6	0.028488082	0.55143	-0.20153	0.1024424
GB-32-13_LR7_JHL2	79354	32407	26981	0.584326439	0.6	0.031387803	0.54956	-0.23957	0.10885034
GB-32-13_LR7_JHL3	61714	26004	22058	0.587939018	0.6	0.027412464	0.54224	-0.21413	0.1016004
GB-36-09_LR8_JHL1	211050	84377	71939	0.585684288	0.6	0.030546228	0.55421	-0.26888	0.10842218
GB-36-09_LR8_JHL2	457979	186136	158790	0.589830743	0.6	0.033568271	0.52394	-0.33517	0.11375258
GB-36-09_LR8_JHL3	376395	154908	133094	0.589729855	0.6	0.031474577	0.51507	-0.32874	0.11103097
GB-36-16_LR9_JHL1	675	288	232	0.536493786	0.6	0.050560264	0.58785	0.1081	0.09913793
GB-36-16_LR9_JHL2	2985	1063	852	0.562154946	0.6	0.0384907	0.70441	0.09494	0.08920188
GB-36-16_LR9_JHL3	1265	462	372	0.582383433	0.6	0.032689042	0.76506	0.26637	0.06451613

Ids	*Total markers genotyped*	*Markers genotyped from chimpanzee*	*Markers genotyped from chimpanzee biallelic*	*Allele imbalance mean*	*Allele imbalance median*	*Allele imbalance var*	*Heterozygosity 72 libraries*	*Heterozygosity 69 libraries (without LR16)*	*Proportion Heterozygote SNPs*
N183-9_LR10_JHL1	366541	142851	121255	0.598887874	0.666667	0.030267869	0.60224	-0.17653	0.10058142
N183-9_LR10_JHL2	720789	290989	248308	0.608738799	0.666667	0.032528502	0.57424	-0.21806	0.10227053
N183-9_LR10_JHL3	619524	250133	214281	0.609731004	0.666667	0.031363302	0.56272	-0.24162	0.10294845
N183-10_LR11_JHL1	906320	376286	325567	0.613096941	0.666667	0.033152895	0.52683	-0.24049	0.10167888
N183-10_LR11_JHL2	1263644	532198	462841	0.619500495	0.666667	0.035651613	0.50197	-0.25836	0.10201489
N183-10_LR11_JHL3	1236353	518225	451841	0.622418713	0.666667	0.034838561	0.48561	-0.28492	0.10424966
N184-7_LR12_JHL1	92219	33778	28275	0.58540477	0.631579	0.030354443	0.67928	-0.00853	0.09033034
N184-7_LR12_JHL2	268203	99344	83170	0.586642403	0.625	0.032582327	0.65477	-0.07938	0.09728735
N184-7_LR12_JHL3	183125	68620	57933	0.590242672	0.666667	0.029029494	0.66236	-0.04622	0.09129197
N184-10_LR13_JHL1	1467945	612039	535133	0.622664735	0.666667	0.035106001	0.50915	-0.20616	0.09795498
N184-10_LR13_JHL2	1740777	721439	632472	0.617704331	0.642857	0.035464759	0.53238	-0.11037	0.09094028
N184-10_LR13_JHL3	1768692	731565	642535	0.626106746	0.666667	0.036224809	0.50809	-0.15299	0.09470993
N186-7_LR14_JHL1	98701	36235	30498	0.586798672	0.6	0.029590198	0.66616	-0.05492	0.09446521
N186-7_LR14_JHL2	296913	109177	91263	0.586080656	0.6	0.034372597	0.65796	-0.07943	0.09790945
N186-7_LR14_JHL3	202985	76460	30498	0.584140035	0.6	0.031506964	0.64877	-0.09511	0.09518571
N189-1_LR15_JHL1	57714	21275	17850	0.580834592	0.6	0.029088331	0.6732	-0.033	0.09019608
N189-1_LR15_JHL2	194683	70862	59048	0.581280729	0.6	0.033726063	0.65903	-0.0778	0.09898726
N189-1_LR15_JHL3	123767	46295	39186	0.585701999	0.6	0.031020022	0.64218	-0.10956	0.09667211
N189-10_LR16_JHL1	624315	NA	NA	NA	NA	NA	0.5181	NA	NA
N189-10_LR16_JHL2	1006965	NA	NA	NA	NA	NA	0.46711	NA	NA
N189-10_LR16_JHL3	856210	NA	NA	NA	NA	NA	0.46201	NA	NA

Ids	*Total markers genotyped*	*Markers genotyped from chimpanzee*	*Markers genotyped from chimpanzee biallelic*	*Allele imbalance mean*	*Allele imbalance median*	*Allele imbalance var*	*Heterozygosity 72 libraries*	*Heterozygosity 69 libraries (without LR16)*	*Proportion Heterozygote SNPs*
N184-8_LR17_JHL1	802207	327600	284171	0.594017484	0.607143	0.030095187	0.54424	-0.22506	0.09899112
N184-8_LR17_JHL2	631777	252486	215491	0.603280265	0.666667	0.029071325	0.57425	-0.25957	0.10361038
N184-8_LR17_JHL3	836772	344393	296462	0.609243152	0.666667	0.030684435	0.52807	-0.32183	0.1068847
N184-8-2_LR18_JHL1	469818	185436	159128	0.566848823	0.6	0.029978984	0.61562	-0.10232	0.09203225
N184-8-2_LR18_JHL2	899692	370386	319420	0.603017294	0.666667	0.032812795	0.54838	-0.24749	0.10110256
N184-8-2_LR18_JHL3	733667	295911	253164	0.599296784	0.636364	0.032813974	0.58263	-0.21074	0.09977367
N184-8-3_LR19_JHL1	1005086	416388	363040	0.607990789	0.666667	0.029529103	0.49282	-0.31072	0.10549511
N184-8-3_LR19_JHL2	757886	310057	265811	0.612847144	0.666667	0.028937142	0.53088	-0.34156	0.10848808
N184-8-3_LR19_JHL3	990292	413490	357596	0.619594509	0.666667	0.030240769	0.48009	-0.40843	0.11284676
N184-8-4_LR20_JHL1	1467926	610400	535461	0.614539409	0.666667	0.03306403	0.46652	-0.2898	0.10507513
N184-8-4_LR20_JHL2	1711795	711727	624587	0.62859534	0.666667	0.034162869	0.45617	-0.29809	0.10509362
N184-8-4_LR20_JHL3	1484732	622911	544526	0.626026711	0.666667	0.033422246	0.46807	-0.32966	0.10615913
N190-1_LR21_JHL1	2316750	940572	831866	0.567826605	0.555556	0.025077461	0.614	0.21776	0.0714873
N190-1_LR21_JHL2	2189650	887172	782635	0.586293084	0.575758	0.027633603	0.60575	0.15412	0.07320021
N190-1_LR21_JHL3	2247213	911243	804515	0.577397886	0.5625	0.025425864	0.60899	0.1787	0.07224968
N190-1-2_LR22_JHL1	2330645	946708	837443	0.559026371	0.548387	0.022087306	0.62085	0.23891	0.07044009
N190-1-2_LR22_JHL2	2315283	940361	831389	0.565582207	0.551282	0.019437536	0.62573	0.23897	0.06914402
N190-1-2_LR22_JHL3	2287191	928303	820225	0.572192319	0.559322	0.021899696	0.61986	0.21521	0.0701076
N190-1-3_LR23_JHL1	2311143	938322	829901	0.576838764	0.566667	0.02912779	0.59536	0.17827	0.07493485
N190-1-3_LR23_JHL2	2161874	876643	773140	0.599105346	0.6	0.030064776	0.57598	0.0823	0.07889155
N190-1-3_LR23_JHL3	2228498	903977	797945	0.588234463	0.576923	0.028539672	0.58436	0.12115	0.07686572
N190-1-4_LR24_JHL1	2260880	917517	811046	0.586838511	0.583333	0.033317751	0.57477	0.11685	0.07864599
N190-1-4_LR24_JHL2	2244403	910522	804010	0.585519843	0.571429	0.028222824	0.58938	0.13665	0.07582108
N190-1-4_LR24_JHL3	2174235	882011	777872	0.5944671	0.586207	0.030044747	0.57694	0.08859	0.07861446

8.2. Experimental protocols

8.2.1. Two-step storage protocol

Two-step storage of faeces for DNA analysis

Materials needed:

•50 mL tubes containing silica gel beads and topped with a Kim wipe

•Ethanol (pharmacy grade, 97%), 90% should also work

•Empty 50 mL tubes

Preparation:

1. Pour approximately 30 ml of ethanol into empty tubes for sample collection.

Collection:

2. Collect each fresh faeces sample (approx. 5 g – approximately the size of a small walnut) into a tube containing ~ 30 ml ethanol.

3. Label tube (but remember this tube will be discarded).

*** It is very rare, but occasionally a tube containing ethanol will leak and cause the writing to wear off itself and adjacent tubes. It is best to just put a few ethanol containing tubes together in any single plastic bag to minimize potential losses of information***

Processing (next day):

4. The faecal sample will either have maintained its shape and structure (faecal bolus) or have dissipated into the ethanol and have formed a sludge.

5. Carefully pour out as much ethanol as possible

a. If the faecal bolus is intact, it should be simple to pour off all of the ethanol and then transfer the bolus onto the Kim wipe in the silica tube, close lid.

b. If a faecal sludge has formed, let the sludge settle to the bottom of the tube and then decant as much ethanol from the tube as possible (it is OK to lose some faecal sludge at this step). Then, transfer the sludge to the Kim wipe in the silica tube, close lid.

6. The tube should be labelled, with a unique identifier and date (GPS location, collector name, species collected, field site of collection if possible).

7. Store at RT.

8. All samples and associated information should be entered into a spreadsheet and this spreadsheet should be sent with the samples.

Reference:

Nsubuga AM, Robbins MM, Roeder A, Morin P, Boesch C and Vigilant L (2004) Factors affecting the amount of genomic DNA extracted from ape feces and the identification of an improved sample storage method. Molecular Ecology 13: 2089-2094.

Protocol extracted from:

https://www.eva.mpg.de/fileadmin/content_files/primatology/Molecular_Genetics_Laboratory/pdf/protocols/2step_collection_protocol_2012.pdf

8.2.2. QIAamp DNA stool mini kit protocol

Fecal DNA Extraction

Materials needed:

•QIAGEN QIAamp DNA Stool Mini Kit (cat. no. 51504)

•4x 2ml tubes

•4x collection tubes (provided in kit)

•1x 1.5ml tubes

•100% EtOH

•Carrier RNA (Poly(A) carrier, Roche, cat. no. 0108626)

Required equipment:

Microfuge

Heat block/shaker

Processing:

1.Set heat block to 70°C.

2.Make sure that buffers ASL and AL are not precipitated, dissolve soln. at 70°C if necessary.

3.Add 100 mg desiccated faeces or 250 mg of fresh faeces to a 2 ml tube.

4.Add 1.6 ml ASL, vortex very well, and soak for 1 h at RT (fresh faeces) or 2-72 h (desiccated faeces) vortex occasionally while soaking.

5.Centrifuge full speed for 3 min to pellet faeces.

6.Transfer 1.4 ml of the supernatant into a new 2 ml tube, discard the pellet.

7.Add 1 InhibitEX tablet to each sample and vortex vigorously until the tablet is completely suspended.

8.Incubate the suspension for a few minutes at room temperature.

9.Centrifuge samples at full speed for 10 min.

10.Transfer the supernatant into a 2 ml tube.

11.Centrifuge the pellet at full speed for 3 min.

12.Transfer the supernatant into the tube from step 10, discard the pellet (you need 600µl of supernatant for step 15 (steps 11 and 12 may be repeated).

13.Centrifuge the supernatant at full speed for 6 min.

14.Pipet 25 µl proteinase K (provided in kit) into a new 2 ml tube.

15.Transfer 600 µl supernatant (avoid the white precipitate) from step 13 to the 2ml-tube containing proteinase K.

16.Add 600 µl of AL and vortex immediately (15 sec.).

17.Incubate at 70°C for 10 min (go to this step directly after vortexing).

18.Add 4 µl of carrier RNA and vortex immediately.

19.Add 600 µl of 100% EtOH to the lysate and mix by vortexing.

20.Carefully apply 600 µl of solution from step 19 to a QIAamp spin column.

21.Centrifuge at full speed for 2 min, place the spin column in a new 2 ml collection tube, discard the tube containing the filtrate.

22.Apply a second aliquot of 600 µl lysate and centrifuge at full speed for 2 min, place the spin column in a new 2 ml collection tube and discard the filtrate.

23.Apply the last aliquot of lysate (600 µl) and centrifuge at full speed for 2 min, place the spin column in a new 2 ml collection tube and discard the filtrate.

24.Wash the column with 500 µl AW1, centrifuge at full speed for 2 min, discard the filtrate and place column in a new collection tube.

25.Wash the column with 500 µl buffer AW2, centrifuge at full speed for 6 min, discard the collection tube with filtrate

26.Optional: place the spin column in a new collection tube and centrifuge at full speed for 2 min, discard the collection tube containing filtrate.

27.Transfer the spin column into a labelled 1.5 ml tube and pipet 200 µl buffer AE directly onto the membrane.

28.Incubate for 20-30 min at RT and then centrifuge at full speed for 2 min to elute the DNA.

Protocol extracted from:

https://swfsc.noaa.gov/uploadedFiles/Divisions/PRD/Projects/Research_and_Development/Molecular_Lab/3.fecal.extraction.pdf

8.2.3. qPCR protocol

Materials needed:

Starred (*) reagents are described in the 'standard laboratory solutions' document)

1.CMYC taqman assay reagents

•fwd primer = cMYC_E3_F1U1 (AGAGGAGGAACGAGCT)

•rv primer = cMYC_E3_R1U1 (GGGCCTTTTCATTGTTTTCCA)

•probe = cMYC_E3_TMV (TGCCCTGCGTGACCAGATCC)

•Eurogentec Taqman reagents (qPCR core kit/RT-QP73-05):

10 x master mix

$MgCl_2$ (25 mM)

dUTP's (2.5 mM)

Hotstar Gold Taq polymerase (5 U/µl)

•Bovine Serum Albumin (BSA) (20 mg/ml)

•Uracil–n–glycosylase (UNG) (1 U/µl)

2.Perkin Elmer (PE) DNA standard dilution plate for the 7700 sequence detector prepared following the qPCR Dilution Series protocol (Protocol #1).

3.PE optical 96-well PCR plates (N801-0560).

4.PE optical strip caps for sealing the plate. (4323032).

Required equipment:

7700 Sequence Detector ("Taqman" instrument).

Processing:

1.Design your experiment and enter the sample names into the Datasheet worksheet of the Experiment Workbook (see QPCR excel workbook and figure 1 below). For each set of samples to be quantified in one experiment, triplicate standards need to be run, along with several no-template controls and the samples. If more samples are to be run than can fit on a 96-well plate, they can be amplified sequentially in the QPCR instrument (using one PCR master mix) and analysed with the same standard.

2.In the no-DNA hood prepare a mastermix for the cMYC quantitative PCR assay using the following conditions (adjusted for the number of samples). You will need enough for 36 tubes of standards and 2 replicates of each of your DNA samples (plus ≥ 3 no template controls (NTCs)).

3.In the DNA hood, add 5 µl of sample DNA to wells in rows D to H of a pre-prepared PE Standard plate (see protocol # 1 on preparation of PE Standard plates). Ensure that you have duplicates of each of your samples and to include at least 3 NTC's. Samples that cannot go on the first plate can be put into additional plates or tube-strips (optical quality).

4.Use the multichannel pipette to add 15 µl of master mix into each well. Ensure that you change tips after each row to avoid cross-contamination. Cap the tubes with optical 8-strip lids.

Starting the QPCR run

1.Open a blank PEstd.sds template from the cMYC Taqman folder. This template is already set up with the PE standards in triplicate. Adjust the number of unknowns to reflect the number of samples in your plate. Highlight the empty wells and use the drop-down **Sample Type Menu** to change their status to “not in use”.

2.Check to see which reporter dye is on your taqman probe (usually VIC) and ensure that template is showing that dye layer for your assay set up.

3.Click on the box marked **Thermal Cycler Conditions** to make sure the proper conditions are listed. These should read as follows:

Step 1: 50°C, 2 min

Step 2: 95°C, 10 min

Step 3: 95°C, 15 sec

Step 4: 59°C, 30 sec

Go To Step 3, 49 times

Step 5: 20°C, 1 min

Step 6: End

4.Click on Show Analysis.

5.Load your sample tray into the taqman machine ensuring that well A1 is in the top left corner, and close and screw down the lid.

6.Ensure that Appletalk is disconnected and then click **Run** on the Show Analysis screen. After the cycles have begun, the time to completion will be shown (typically 1 hour and 58 minutes).

Analyzing Data after a Taqman Run

1.Immediately use the “save as” option from the **File** menu to save your run to an appropriate folder.

2.Pull down the **Analysis** menu and select **Analyze**.

3.When you see the amplification plot window, do the following:

Look at the figures in the **Baseline** boxes. Make sure the largest number is at least 3 cycles PRIOR to the beginning of the exponential amplification on the plot window. If it looks like it may be close, double click on the “ΔRn” (Y axis) and select “linear plot”, then check to see at what cycle the fluorescence values begin to rise above the baseline. If you need to change the maximum “baseline” cycle number, change it in the **Baseline** box, then change the plot back to “logarithmic”.

Establish the **Use Threshold** by clicking and holding on the black horizontal bar on the amplification plot. Move it to a point ABOVE the baseline and well WITHIN the area of exponential amplification. NOTE: If you ran your standards on a previous plate, record the value of the Use Threshold from the standard plate, and enter this number in the appropriate box at the lower left hand of the screen for each subsequent set of samples.

If you make any changes to the values in the box at the lower left, click on **Update Calculations.**

4.Pull down the **File** menu, select **Export** and then choose **Results**.

5.When prompted, save the exported file as **<Filename>.results**. Make sure that **Export All Wells** is selected and then click on **Export** to complete the process.

6.Open your **<Filename>.results** file and copy the entire table.

7.Open your Excel **Experiment Workbook** in which you entered all the sample names. Click on the tab labeled **7700 Results** at the bottom of the sheet and paste the results table here (be sure to start at the first cell (A1).

8.On the menu at the bottom of your **Experiment Workbook**, click on the **Standard Curve** worksheet tab to view an excel graph of the standard curve and the points used to construct the graph.

9.If the correlation coefficient (R^2) of the curve is not >0.98, check the graph for outlier points which may represent anomalous amplifications or failed reactions. Delete these outliers from the data table to the left of the graph to improve the R^2 value. As the curve is made up from 3 sets of identical standards, the removal of 1 or2 outliers will not unduly bias the plot because there are still 2 homologous points left from that exact same standard DNA.

10.Record the values for the slope, y-intercept and R^2 from the equation of the line for the standard curve.

11.On the menu at the bottom of your **Experiment Workbook**, click on the **Datasheet** tab. On the **Datasheet** fill in the three yellow boxes: Slope, y-int & R^2; with the values recorded from the standard curve.

12.As soon as you have entered the above values, the DNA quantities for your unknown samples Will be calculated automatically in the data sheet, columns J & K (pg and ng). (calculated from the equation: DNA amount = $10^{((Ct - Yint)/slope))}$.

13.SAVE YOUR RESULTS. Print out hard copies of the standard curve and datasheet for your lab notebook.

Protocol extracted from:

https://swfsc.noaa.gov/uploadedFiles/Divisions/PRD/Projects/Research_and_Development/Molecular_Lab/2.qPCRprotocol.pdf

8.2.4. Library preparation for shotgun sequencing protocol

Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing

Materials needed:

•Agarose gel (2%) and reagents for agarose gel electrophoresis.

•AMPure XP 60 mL Kit (Agencourt-Beckman Coulter A63881)

•ATP (100 mM) (Fermentas R0441)

•Bst

•DNA polymerase, large fragment (supplied with 10X ThermoPol reaction buffer) (New England BioLabs M0275S)

•DNA ladder (e.g., GeneRuler; Fermentas) (optional; see note before Step 6)

For unknown reasons, ladders from New England BioLabs do not work for this purpose.

•dNTP mix (25 mM each) (Fermentas R1121)

•EBT buffer

•Ethanol (70%, freshly prepared)

•H_2O (HPLC grade)

•Illumina reagents for DNA sequencing (Illumina, Inc.)

•Cluster generation kit (e.g., GD-103-4001 [Standard Cluster Generation Kit v4], PE-203-4001 [Paired-End Cluster Generation Kit v4])

•Multiplexing sequencing primer kit (PE-400-1002 [Multiplexing Sequencing Primers and PhiX Control Kit v1])

Alternatively, the following primers may be used for sequencing:
Read 1 Sequencing Primer:
5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
Index Read Sequencing Primer:
5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3'
Read 2 Sequencing Primer:
5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'

•Sequencing kit (FC-104-4002 [36 Cycle Sequencing Kit v4])
•MinElute PCR Purification Kit (QIAGEN) (optional)
•Oligo hybridization buffer (10X)
•Oligonucleotides (Sigma-Aldrich) (see Table 1)
•Phusion Hot Start High-Fidelity DNA Polymerase (New England BioLabs F-540L) (supplied with 5X Phusion HF buffer)
•Positive control DNA (200-to 300-bp fragment, generated via PCR using unmodified primers and a polymerase with terminal transferase activity, e.g., Taq DNA polymerase) (200-500 ng)
•Sample DNA

This protocol works reliably with as little as 100 pg and up to 1 µg of double-stranded sample DNA (e.g., genomic DNA, long-range PCR products, or cDNA). The amount of starting material should be chosen so that the representation of target molecules in the final library is sufficient. The final yield of the library preparation process is ~10%-20%. Therefore, a library prepared from 1 ng of human genomic DNA (about 300 copies of the haploid genome), will contain 30 to 60 copies of the human genome.

•Standard for quantitative PCR (qPCR) (see Steps 21.i-21.ii)

•SYBR Green qPCR master mix (e.g., DyNAmo Flash SYBR Green qPCR Kit; New England BioLabs)

•Tango buffer (10X; Fermentas BY5)

•T4 DNA ligase (5 U/µL; Fermentas EL0011) (supplied with 10X T4 DNA ligase buffer and 50%

•PEG-4000 solution)

•T4 DNA polymerase (5 U/µL; Fermentas EP0062)

•T4 polynucleotide kinase (10 U/µL; Fermentas EK0032)

•TET buffer

•Tween 20

Required equipment:

•Centrifuge for 96-well plates

•DNA shearing device (e.g., Bioruptor UCD-200 [Diagenode]; Covaris E210 [Covaris Inc]) (for high-molecular-weight DNA; see Step 3)

The Bioruptor UCD-200 can process 12 samples in parallel. Among the many alternative systems that are available for this step, the Covaris E210 system may be preferable, because it is compatible with the 96-well plate format.

Equipment for agarose gel electrophoresis:

•Ice

•Multichannel pipettes

•Multichannel reagent basins (e.g., Thermo Scientific 9510027)

•PCR plates (96-well, 200-µL capacity) and strip caps

•Real-time PCR cycler (e.g., Mx3005P QPCR System; Agilent Technologies-Stratagene)

•Sequencing machine (Genome Analyzer II/IIx/IIe or HiSeq2000; Illumina)

•Spectrophotometer for DNA quantification (e.g., NanoDrop; Thermo Scientific)

•SPRIPlate 96R-Ring Super Magnet Plate (Agencourt-Beckman Coulter A32782)

•Thermal cycler

•Tubes (microcentrifuge, 0.5-mL)

•Tubes (PCR)

•Vortex mixers for tubes and 96-well plates

Processing:

The protocol can be interrupted after Steps 3, 12, 16, 19, 24, and 26 by freezing the DNA at –20°C. Up to 94 samples can be processed in parallel on a 96-well reaction plate; two wells should be reserved for a blank and a positive control.

Seal each reaction plate with strip caps and centrifuge to 2000g in a plate centrifuge after setting up each reaction in order to collect the liquid in the bottom of the wells. This prevents cross-contamination while removing the caps.

•Preparation of Adapter Mix

This step produces sufficient adapter mix for 200 reactions. The adapter mix can be used repeatedly and stored at –20°C before and after usage.

1. Assemble the following hybridization reactions in separate PCR tubes:

Hybridization Mix for Adapter P5 (200µM):	Volume (µl)
IS1_adapter_P5.F (500µM)	40
IS3_adapter_P5+P7.R (500µM)	40
Oligo hybridization buffer (10x)	10
H_2O	10
	100µl

Hybridization Mix for Adapter P7 (200µM):	Volume (µl)
IS2_adapter_P7.F (500µM)	40
IS3_adapter_P5+P7.R (500µM)	40
Oligo hybridization buffer (10x)	10
H_2O	10
	100µl

2. Mix and incubate the reactions in a thermal cycler for 10 sec at 95°C, followed by a ramp from 95°C to 12°C at a rate of 0.1°C/sec. Combine both reactions to obtain a ready-to-use adapter mix(100 µM each adapter).

•Fragmentation and Purification of Sample DNA

This step in the method is not always required. Prior to library preparation, high-molecular-weight sample DNA must be sheared into fragments of suitable size for Illumina sequencing (<600bp). If samples other than high-molecular-weight DNA are used (e.g., short PCR products, highly degraded DNA, or short double-stranded cDNA), fragmentation may not be necessary. Step 3 describes DNA hearing by sonication using the Bioruptor UCD-200.

3. Shear the DNA as follows:

i. Transfer the samples to 0.5-mL tubes, and add H_2O to reach final volumes of 50 µL.

ii.Expose the DNA four times to sonication cycles of 7 min, using the energy setting "HIGH" and an "ON/OFF interval" of30 sec. If liquid spills to the tube walls, shake it down to the bottom of the wells after each sonication cycle.

This produces a fragment size distribution between 100 bp and 400 bp, with a mean around 200 bp.

iii.Transfer the sheared DNA samples to a 96-well PCR plate.

The fragment size distribution obtained from sonication is well-suited for sequencing. However, if a very narrow fragment size distribution is desired, the fragmented DNA may be separated on an agarose gel and isolated from a gel slice to obtain a more narrow distribution. In the example given in Figure 2, no gel excision was performed.

•Blunt-End Repair

If the sample DNA is not dissolved in H2O, Tris-Cl buffer (e.g., QIAGEN's Buffer EB), or TE buffer, purify the DNA as described in Steps 6-13 prior to beginning Step 4. If the sample volume exceeds 50 μL, purification can be used for concentrating the DNA. We strongly recommend carrying a positive and a blank control through Steps 4-18 of the protocol. As a positive control, 200-500 ng of a purified PCR product with a discrete size of 200-300 bp may be used. The product should be generated using unmodified PCR primers and a polymerase with terminal transferase activity (e.g., Taq DNA polymerase).

4. Add a blank control (50 µL of H2O) and a positive control to two empty wells of the reaction plate. Prepare a master mix as below for the required number of reactions. Mix carefully by flicking the tube with a finger. Avoid vortexing after addition of enzymes.

Reagent	Vol(µl/sample)	Mastermix 110x	Final conc.
Buffer Tango (10x)	7	770	1x
dNTPs (25mM each)	0.28	30.8	100µM each
ATP (100mM)	0.7	77	1mM
T4 PNK (10U/µl)	3.5	385	0.5U/µl
T4 DNA polymerase (5U/µl)	1.4	154	0.1U7µl
H_2O	7.12	783.2	
	20	2200	

5.Using a multichannel pipette, add 20 µL of master mix to 50 µL of sample. Mix and incubate in a thermal cycler for 15 min at 25°C followed by 5 min at 12°C. Place plate on ice or immediately proceed to the next step.

•Reaction Clean-Up Using Solid Phase Reversible Immobilization (SPRI)

Carboxyl-coated magnetic beads (SPRI beads) are ideally suited for reaction purification in a 96-well plate setup. However, under the conditions described here, SPRI purification does not retain molecules shorter than 100-150 bp. The exact size cutoff may vary among different batches of beads. If retention of short molecules is desired, the size cutoff can be adjusted by varying the volume of SPRI bead/buffer suspension added to the sample. The appropriate ratio of SPRI suspension to sample volume can be empirically determined using a DNA ladder (e.g., GeneRuler ladders). If retention of very short molecules is desired (30-80 bp), all SPRI purification steps should be replaced by spin column purification using the MinElute PCR Purification Kit.

6. Resuspend the stock solution of SPRI bead suspension (AMPure kit) by vortexing. To make subsequent pipetting easier, add Tween 20 to the bead suspension to a final concentration of 0.05% (i.e., add 1 µL of Tween 20 to 2 mL of bead suspension).

7. Add SPRI bead suspension to the reactions as follows:

i.Add a 1.8-fold volume of SPRI bead suspension to each reaction (e.g., add 126 µL of SPRI beads to a 70-µL sample or 72 µL of SPRI beads to a 40-µL sample).

ii.Seal the wells with caps and vortex for several seconds. Ensure the beads are properly suspended and repeat vortexing if necessary.

iii. Let the plate stand for 5 min at room temperature.

iv.Collect the liquid at the bottom of the wells by briefly centrifuging in a plate centrifuge to 2000g.

8. Place the plate on a 96-well ring magnetic plate, and let it stand for 5 min to separate the beads from the solution. Pipette off and discard the supernatant without removing the beads.

9. Leave the plate on the magnetic rack, and wash the beads by adding 150 µL of freshly prepared 70% ethanol. Let stand for 1 min and remove the supernatant.

10. Repeat Step 9.

11. Using a multichannel pipette, remove residual traces of ethanol. Let the beads air-dry for 20 min at room temperature without caps.

12. Elute as follows:

i.Add 20 µL of EBT to the wells and seal the plate with caps.

ii. Remove the plate from the magnetic rack, and resuspend the beads by repeated vortexing.

iii. Let stand for 1 min, and then collect the liquid in the bottom of the wells by briefly centrifuging the plate to 2000g.

Occasionally the beads may appear clumpy after vortexing; this does not have a negative effect on DNA recovery.

13. Place the plate back on the magnetic rack, let stand for 1 min, and transfer the supernatant to a new 96-well reaction plate.

Carryover of small amounts of beads will not inhibit subsequent reactions.

•Adapter Ligation

14. Prepare a master mix for the required number of ligation reactions as shown below. If white precipitate is present in the 10X DNA ligase buffer after thawing, warm the buffer to 37°C and vortex until the precipitate has dissolved. Since PEG is highly viscous, vortex the master mix before adding T4 DNA ligase and mix gently thereafter.

Reagent	Vol(μl/sample)	Mastermix 110x	Final conc.
T4 DNA Ligase Buffer (10x)	4	440	1x
PEG-4000 (50%)	4	440	5%
Adapter Mix (100μM each)	1	110	2.5μM each
T4 DNA Ligase (5U/μl)	1	110	0.125U/μl
H_2O	10	1100	
	20	2200	

When starting from low template quantities (50 ng or less), the amount of adapter mix can be reduced to 0.2 µL per reaction.

15.Add 20 µL of master mix to each eluate from Step 13 to obtain reaction volumes of 40 µL. Mix and incubate for 30 min at 22°C in a thermal cycler.

16. Perform reaction purification exactly as described in Steps 6-13. Elute in 20µL of EBT.

•Adapter Fill-In

17. Prepare a master mix for the required number of reactions.

Reagent	Vol(µl/sample)	Mastermix 110x	Final conc.
ThermoPol reaction buffer (10X)	4	440	1x
dNTPs (25 mM each)	0.4	44	250µM each
Bst polymerase, large fragment (8 U/µl)	1.5	165	0.3U/µl
H_2O	14.1	1551	
	20	2200	

18. Add 20 µL of master mix to each eluate from Step 16 to obtain reaction volumes of 40 µL. Mix well and incubate in a thermal cycler for 20 min at 37°C.

19. Perform reaction purification exactly as described in Steps 6-13. Elute the library in 20 µL of EBT.

•Library Characterization

In addition to agarose gel electrophoresis (Step 20), performance of qPCR (Step 21) prior to indexing PCR (Steps 22-24) is strongly recommended, particularly if little sample DNA was used for library preparation. This is the only option to directly measure the number of molecules in the library. If the mean average fragment length and the size of the genome are known, this number can be used to determine whether the average coverage of genomic targets in the library is sufficiently high for subsequent target capture or direct sequencing. Step 21 describes a qPCR assay using SYBR Green (for more details, see Meyer et al. 2008a).

20. To verify the success of the library preparation, load 10µL of the positive control library side-by-side with 100ng of the original positive control sample and a size marker on a 2% agarose gel and perform electrophoresis.

If all enzymatic reactions worked properly, the band produced by the control library should be shifted upward by 67 bp. See Troubleshooting.

21. Measure the number of molecules by qPCR:

i. Prepare a standard dilution series by incrementally diluting an indexed sequencing library of known molecular concentration 10-fold in TET buffer.

ii. If no such library is available, amplify 0.5 µL of the positive control in an indexing PCR (see Step 22). Purify the PCR product as described in Steps 6-13, determine its mass concentration on a spectrophotometer, calculate the molecular concentration, and use it as a standard as described in Step 21.i.

iii. In a real-time PCR machine, amplify in parallel 1 µL of each standard dilution and each sample using primer IS4 and one of the indexing oligos; we recommend using a commercial PCR master mix containing SYBR Green (e.g., DyNAmo Flash SYBR Green qPCR kit). Set the annealing temperature to 60°C, and otherwise follow the instructions provided by the manufacturers of the kit and the real-time PCR machine.

The concentration of molecules in the blank library (adapter dimers) should be at least one order of magnitude lower than in the sample libraries.

It is often necessary to measure dilutions of the samples (e.g., 1000-fold in EBT) to obtain values within the detection range of the qPCR system.

•Indexing PCR and Pooling

To avoid a downstream failure of Illumina's image analysis software, subsets of indexes must be chosen in a way that prevents unbalanced usage of the four nucleotides or the two laser channels during any cycle of index sequencing. The indexes provided with this protocol (see Supplemental Material [Indexing_Oligo_Sequences.doc]) are in an appropriate order to fulfill these requirements and should be used accordingly. For example, the first 22 indexes should be used if 22 indexes are needed. Fewer than four indexes should never be used in any experiment. Additional sets of indexes with different length and varying edit distance between indexes are provided on http://bioinf.eva.mpg.de. It will often not be necessary to use the entire library as template for indexing PCR. In this case, it is advisable to keep a backup that can be later used to add a different barcode to the sample.

Note that Phusion polymerase has proofreading activity. If this property is not desired (e.g., if deoxyuracil is present in the template DNA), another polymerase can be chosen for indexing PCR.

22. Prepare a PCR master mix for the required number of reactions. Dispense the master mix into a 96-well reaction plate, and then add

template DNA and a different indexing primer to each well using a multichannel pipette.

Reagent	Vol (µl/sample)	Final conc.
Phusion HF buffer (5X)	10	1x
dNTPs (25 mM each)	0.4	200 µM each
Primer IS4 (10 µM)	1	200 nM
Phusion Hot Start High-Fidelity DNA Polymerase (2U/µL)	0.5	0.02 U/µL
H_2O	37.1 - x	
Add separately to each well:		
Indexing primer (10µM)	1	200nM
Template DNA (library)	x	

If large amounts of sample DNA were used for library preparation (>>100 ng), only a fraction of the library containing the equivalent of ~100 ng of starting material should be used for indexing PCR in order to prevent saturation of the PCR with template DNA.

23. Mix and perform cycling using the following temperature profile:

Initial denaturation	98°C	30 sec
Denaturation/cycle	98°C	10 sec
Annealing/cycle	60°C	20 sec
Elongation/cycle	72°C	20 sec
Final extension	72°C	10 min

The optimal number of PCR cycles, that is, the number of cycles required to reach PCR plateau, will depend on the amount and concentration of template DNA and can be directly inferred from the amplification plots of the qPCR (Step 21). The cycle number can also be adjusted by rule of thumb according to the lowest amount of sample DNA that was used for library preparation:>100 ng→12 cycles; >10 ng→16 cycles, >1ng→20 cycles,>100 pg→24 cycles.

24. Perform reaction purification exactly as described in Steps 6-13. Elute the indexed libraries in 25 µL of EBT.

25. Load 3 µL of some of the PCR products on a 2% agarose gel to verify amplification success.

Indexed libraries prepared from sheared DNA should produce a smear. Due to the formation of heteroduplexes in the plateau phase of PCR (Ruano and Kidd 1992), the fragment size distribution inferred from the agarose gel may deviate slightly from the true distribution. However, no low-molecular-weight artifacts, such as primer dimers or adapter dimers, should be visible in the indexed sample libraries. See Troubleshooting.

26. Determine the DNA concentration, and pool the indexed libraries in equimolar ratios.

The pool of indexed libraries is now ready for target capture or direct sequencing on one of Illumina’s sequencing platforms. Due to the presence of heteroduplexes, qPCR is the only means of exactly determining the DNA concentrations in indexed libraries. However, concentration estimates derived from measurements with a spectrophotometer are sufficient in this step and more convenient. End product yield of indexing PCR is usually similar for all samples, particularly if there are no major differences in fragment size distribution. If this is the case, as can be confirmed by measuring DNA concentrations in a subset of indexed libraries, pooling equal volumes of all libraries will be sufficient.

•Target Capture and/or Sequencing on the Illumina Platform

27. For target capture on microarrays, follow, for example, the exact procedure given in the protocol of Hodges et al. (2009) with the following modifications:

i. Use a different set of blocking oligos (BO1-BO6).

ii. Use primers IS5 and IS6 at an annealing temperature of 60°C for amplifying the library pool after capture.

28. For sequencing and data analysis, use the recipes, kits, and analysis tools for multiplex sequencing provided by Illumina.

A tool for splitting up the qseq sequence files according to indexes is available in CASAVA 1.6 and later versions (demultiplex.pl). However, when using the 7-nt index sequences given in this protocol, the --qseq-mask parameter must be set to seven (the default is six). No modifications to the recipes provided by the Illumina machine control software (SCS) are required, because seven cycles of index sequencing are carried out by default. Additional software for data analysis on FastQ files (SplitFastQIndex.py), a file format created for example by the alternative base caller Ibis (Kircher et al. 2009), is provided on *http://bioinf.eva.mpg.de*. If single mismatches are allowed during index identification, the fraction of unidentified index sequences typically reduces to ~5%, as compared to ~15% when a perfect match is required. Using alternative base callers like Alta-Cyclic (Erlich et al. 2008), BayesCall (Kao et al. 2009), or IBIS (Kircher et al. 2009) may also increase the fraction of correctly identified indexes.

Indexed sequencing libraries are compatible with all capture methods requiring sequencing libraries. It is recommended to carry the blank library all the way through target capture and/or sequencing. To avoid cross-contamination of samples through jumping PCR (Meyerhans et al. 1990), pools of indexed libraries should be amplified with a minimum number of PCR cycles or sequenced without amplification if possible. See Troubleshooting.

Protocol extracted from:

http://www.protocol-online.org/forums/uploads/monthly_11_2010/msg-6470-047872000%201289836361.ipb

8.2.5. High sensitivity genomic DNA analysis kit protocol (DNF-488)

Materials needed:

•Genomic DNA Separation Gel, part # DNF-270

•Intercalating Dye, part # DNF-600-U030

•5X 930 dsDNA Inlet Buffer, part # DNF-355 (Dilute to 1X)

•5X Capillary Conditioning Solution, part # DNF-475(Refill as needed)

•0.25X TE Rinse Buffer, part # DNF-497

•High Sensitivity Genomic DNA Diluent Marker, part # DNF-375

•High Sensitivity Genomic DNA Ladder, part # DNF-377

•BF-25 Blank Solution, part # DNF-300

•Capillary Storage Solution, part # GP-440-0100(sold separately)

Gel guide: For 12 capillary Fragment Analyzer systems:

96 samples to be analysed → 4.5µL Intercalating dye + 45 mL Gel

Processing:

1.Mix fresh Gel and Dye. Refill 1X Conditioning Solution as needed.

2.Place a fresh 1X Inlet Buffer Tray on Fragment Analyzer.

3.Place Rinse Buffer plate in Marker Drawer location.

4.Mix Samples or Ladder with Diluent Marker in Sample Plate, add 24 µL of Blank Solution to unused wells.

Software:

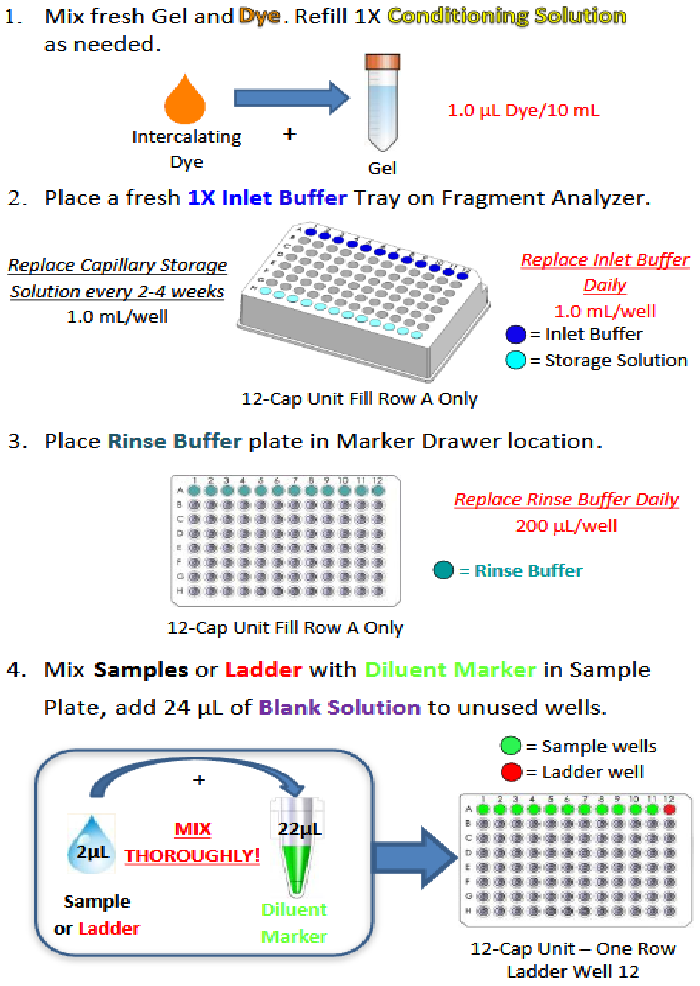
1.Select Tray and Row to run for 12-Cap.

2.Enter Sample ID and Tray ID (optional).

3.Select “Add to Queue”, select the DNF-488-(22, 33 or 55) -HS Genomic DNA method from the Dropdown menu.

4.Enter Tray Name, Folder Prefix, and Notes (optional), Select OK to add Method to the Queue.

5.Select to Start the Separation.



1. Mix fresh Gel and **Dye**. Refill 1X **Conditioning Solution** as needed.

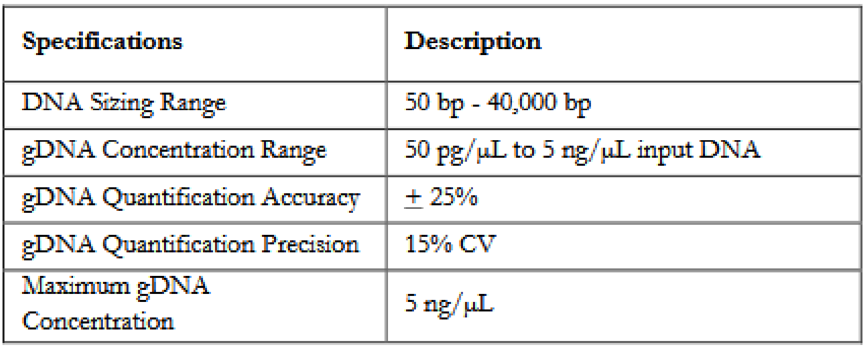
2. Place a fresh **1X Inlet Buffer** Tray on Fragment Analyzer.

Replace Capillary Storage Solution every 2-4 weeks
1.0 mL/well
Replace Inlet Buffer Daily
1.0 mL/well
● = Inlet Buffer
○ = Storage Solution
12-Cap Unit Fill Row A Only

3. Place **Rinse Buffer** plate in Marker Drawer location.

4. Mix **Samples** or **Ladder** with **Diluent Marker** in Sample Plate, add 24 µL of **Blank Solution** to unused wells.

Specifications:



Specifications	Description
DNA Sizing Range	50 bp - 40,000 bp
gDNA Concentration Range	50 pg/µL to 5 ng/µL input DNA
gDNA Quantification Accuracy	± 25%
gDNA Quantification Precision	15% CV
Maximum gDNA Concentration	5 ng/µL

Protocol extracted from:

https://www.aati-us.com/documents/quick-start-guides/dnf-488/dnf-488-quick-start-guide-12-capillary-11-03-2015.pdf

8.2.6. SeqCap EZ Library SR protocol

8.2.6.1. Prepare the Sample Library

Materials needed:

- KAPA Library Preparation Kit
- SeqCap Adapter Kit (A and/or B)
- SeqCap EZ Accessory Kit v2
- Agencourt Ampure XP Beads (warmed to room temperature prior to use)

Ensure that the following are available:

- Additional PCR-grade water for sample library preparation
- Freshly-prepared 80% ethanol: 1.6 ml per DNA sample
- Elution buffer (10 mM Tris-HCl, pH 8.0): 125 µl per DNA sample

⚠ If the sample library preparation protocol is split across two days, freshly prepare the required amount of 80% ethanol daily.

> 📝 Notes: protocol modifications performed in my experiments will be explained in boxes with this symbol 📝.

Sample Requirements

Roche NimbleGen recommends starting with 100 ng of input gDNA for sample library preparation; however, up to 1 ug of input gDNA has been validated and is supported for use in sample library preparation if desired (see Appendix E).

•Step 1. Resuspend the Index Adapters

Resuspension of the Index Adapters must be performed on ice. Care should be taken when opening tubes to avoid loss of the lyophilized pellet.

1. Spin the lyophilized index adapters, contained in the SeqCap Adapter Kit A and/or B, briefly to allow the contents to pellet at the bottom of the tube.

2. Add 50 µl cold, PCR-grade water to each of the 12 tubes labeled ‘SeqCap Index Adapter’ in the SeqCap Adapter Kit A and/or B. Keep adapters on ice.

3. Briefly vortex the index adapters plus PCR-grade water and spin down the resuspended index adapter tubes.

4. The resuspended index adapter tubes should be stored at -15 to -25°C.

•Step 2. Prepare the Sample Library

Instructions for preparing an individual sample library are included here in Step 2, based on v2.14 of the KAPA Library Preparation Kit Technical Data Sheet. When assembling a master mix for processing multiple samples, prepare an excess volume of ~5% to allow for complete pipetting (liquid handling systems may require an excess of ~20%). The KAPA Technical Data Sheet includes several specific scaling examples.

Prior to executing the sample library preparation, please carefully read the entire Technical Data Sheet (v2.14 or later). Ensure you are using the most recent version of the protocol, and contact support@kapabiosystems.com for technical assistance related to the library construction.

For guidelines on preparing sample libraries using amounts of input DNA other than 100 ng, or for using low quality DNA extracted from formalin-fixed paraffin-embedded (FFPE) tissues, see Appendix E, or contact your local Roche Technical Support (go to www.nimblegen.com/contact for contact information).

1. Pipette 100 ng of the gDNA sample of interest into a 1.5 ml tube.

For this project I did not took amount of DNA, instead I tested the performance from different samples with the same volume of DNA from each of them. In Experiment 1 I took 40µl of each sample (total DNA amount varied between 0.36 µg and 4.46 µg) and from Experiment 2 I took 20µl (total DNA amount varied between 1.31 µg to 4.03 µg).

2. Adjust the volume to a total of 52.5 µl using 1x TE (low EDTA) and transfer to a Covaris microTUBE for fragmentation.

3. Fragment the gDNA so that the average DNA fragment size is 180–220 bp.

Genomic DNA was fragmented by shearing using a Covaris S2 focused-ultrasonicator with the following settings for 200 bp fragments: Intensity 5, duty cycle 10 %, cycles per burst 200, treatment time 120 s, temperature 7°C and water level 12.

4. Following fragmentation, proceed with the End Repair Reaction Setup:

a. Transfer 50 µl of the fragmented DNA to a 0.2 ml PCR tube.

b. To each 50 µl fragmented sample add 20 µl of End Repair Master Mix, resulting in a total volume of 70 µl.

End repair master mix	Per individual sample library
PCR-grade water	8 µl
10X KAPA End Repair Buffer	7 µl
KAPA End Repair Enzyme Mix	5 µl
Total	**20 µl**

c. Mix the End Repair reaction by pipetting up and down.

d. Incubate the reaction at +20°C for 30 minutes.

e. Following the 30 minute incubation, proceed immediately to the next step.

5. Perform the End Repair Cleanup:

📝 Reaction clean-up for end-repair in my project was performed with MinElute Reaction clean-up spin columns (Cat. Number 28206) rather than Agencourt AMPure XP beads. This choice was made in an attempt to retain molecules smaller than 100 bp, that could be overly abundant because of initial sample degradation not sample fragmentation. In our hands, the MinElute Reaction kit retains molecules down to ~50 bp while SPRI-beads retained molecules down to ~100 bp. After each cleanup step, DNA was eluted in **20 µl** of elution buffer. If you want to follow MinElute reaction cleanup protocol is explained below with the symbol 📎

To follow SeqCap End repair clean-up:

a. To each 70 µl End Repair Reaction add 120 µl of room temperature Agencourt AMPure XP beads, resulting in a total volume of 190 µl.

End repair cleanup	**Per individual sample library**
End Repair Reaction	70 µl
Agencourt AMPure XP beads	120 µl
Total	**190 µl**

b. Mix thoroughly by pipetting up and down multiple times.

c. Incubate the tube at room temperature for 15 minutes to allow the DNA to bind to the beads.

d. Place the tube on a magnet to capture the beads. Incubate until the liquid is clear.

e. Carefully remove and discard the supernatant.

f. Keeping the tube on the magnet, add 200 µl of freshly-prepared 80% ethanol.

g. Incubate the tube at room temperature for ≥30 seconds.

h. Carefully remove and discard the ethanol.

i. Keeping the tube on the magnet, add 200 µl of freshly-prepared 80% ethanol.

j. Incubate the tube at room temperature for ≥30 seconds.

k. Carefully remove and discard the ethanol. Try to remove all residual ethanol without disturbing the beads.

l. Allow the beads to dry at room temperature, sufficiently for all the ethanol to evaporate.

Caution: Over-drying the beads may result in dramatic yield loss.

m. Remove the tube from the magnet.

MinElute® Reaction Cleanup Kit (Cat. Number 28206). Quick start protocol

Notes before starting

□ This protocol is for cleanup of up to 5 µg DNA (70 bp to 4 kb) from enzymatic reactions.

□ The yellow color of Buffer ERC indicates a pH of ≤7.5. Adsorption of DNA to the membrane is efficient only at pH ≤7.5.

□ Add ethanol (96–100%) to Buffer PE concentrate before use (see bottle label for volume).

□ All centrifugation steps are carried out at 17,900 x g (13,000 rpm) in a conventional tabletop microcentrifuge at room temperature (15–25°C).

□ Symbols: ● centrifuge processing; ▲ vacuum processing.

1. Add 300 µl Buffer ERC to the enzymatic reaction (sample volume 20–100 µl) and mix. If the enzymatic reaction is in a volume of <20 µl, adjust the volume to 20 µl. If the enzymatic reaction exceeds 100 µl, split your reaction, add 300 µl Buffer ERC to each aliquot, and use the appropriate number of MinElute columns.

2. Check that the color of the mixture is yellow (similar to Buffer ERC without the enzymatic reaction). If the color of the mixture is

orange or violet, add 10 µl 3 M sodium acetate, pH 5.0, and mix. The color of the mixture will turn to yellow.

3. Place a MinElute column ● in a provided 2 ml collection tube or ▲ into a vacuum manifold. See the MinElute Handbook for details on how to set up a vacuum manifold.

4. Apply sample to the MinElute column and ● centrifuge for 1 min or ▲ apply vacuum to the manifold until all samples have passed through the column. ● Discard flow-through and place the MinElute column back into the same collection tube.

5. Add 750 µl Buffer PE to the MinElute column and ● centrifuge for 1 min or ▲ apply vacuum. ● Discard flow-through and place the MinElute column back into the same collection tube.

6. Centrifuge the column in a 2 ml collection tube (provided) for 1 min. Residual ethanol from Buffer PE will not be completely removed unless the flow-through is discarded before this additional centrifugation.

7. Place each MinElute column into a clean 1.5 ml microcentrifuge tube.

8. To elute DNA, add **20 µl** Buffer EB (10 mM Tris·Cl, pH 8.5) or water to the centre of the MinElute membrane. (Ensure that the elution buffer is dispensed directly onto the membrane for complete elution of bound DNA). Let the column stand for 1 min, and then centrifuge the column for 1 min.

9. If the purified DNA is to be analyzed on a gel, add 1 volume of Loading Dye to 5 volumes of purified DNA. Mix the solution by pipetting up and down before loading the gel.

6. Perform the A-Tailing Reaction Setup:

a. To each tube of DNA plus beads add 50 µl of the A-Tailing Master Mix, resulting in a total volume of 50 µl.

A-tailing master mix	Per individual sample library
PCR-grade water	42 µl
10X KAPA A-tailing Buffer	5 µl
KAPA A-tailing Enzyme	3 µl
Total	**50 µl**

As End-repair cleanup was performed with MinElute Reaction Cleanup Kit and eluted in 20 µl, A-tailing master mix was modified: 22 µl of PCR-grade water were added instead of 42 µl.

b. Thoroughly resuspend the beads by pipetting up and down multiple times.

c. Incubate the A-Tailing reaction at +30°C for 30 minutes.

d. After incubation, proceed immediately to the next step.

7. Perform the A-Tailing Cleanup:

Reaction clean-up for a-tailing was performed with MinElute Reaction clean-up spin columns (Cat. Number 28206) rather than Agencourt AMPure XP beads. Eluted in 20 µl of Elution buffer

MinElute® Reaction Cleanup Kit (Cat. Number 28206).

(Explained in page 168)

a. To each 50 µl A-Tailing Reaction add 90 µl of thawed, room temperature PEG/NaCl SPRI Solution, resulting in a total volume of 140 µl.

A-tailing cleanup	Per individual sample library
A-tailing Reaction	50 µl
Agencourt AMPure XP beads	90 µl
Total	**140 µl**

b. Mix thoroughly by pipetting up and down multiple times.

c. Incubate the tube at room temperature for 15 minutes to allow the DNA to bind to the beads.

d. Place the tube on a magnet to capture the beads. Incubate until the liquid is clear.

e. Carefully remove and discard the supernatant.

f. Keeping the tube on the magnet, add 200 µl of freshly-prepared 80% ethanol.

g. Incubate the tube at room temperature for ≥30 seconds.

h. Carefully remove and discard the ethanol.

i. Keeping the tube on the magnet, add 200 µl of freshly-prepared 80% ethanol.

j. Incubate the tube at room temperature for ≥30 seconds.

k. Carefully remove and discard the ethanol. Try to remove all residual ethanol without disturbing the beads.

l. Allow the beads to dry at room temperature, sufficiently for all the ethanol to evaporate.

⚠ Caution: Over-drying the beads may result in dramatic yield loss.

m. Remove the tube from the magnet.

8. Proceed with the Adapter Ligation Reaction Setup:

a. To each tube of beads add 47 µl of the Ligation Master Mix, resulting in a total volume of 47 µl.

Ligation master mix	**Per individual sample library**
PCR-grade water	32 µl
5X KAPA Ligation Buffer	10 µl
KAPA T4 DNA Ligase	5 µl
Total	**47 µl**

As A-tailing cleanup was performed with MinElute Reaction Cleanup Kit and eluted in 20 µl, A-tailing master mix was modified: 12 µl of PCR-grade water were added instead of 32 µl.

b. Thoroughly resuspend the beads by pipetting up and down multiple times.

c. Add 3 µl of the SeqCap Library Adapter (with the desired Index) to the tube containing the Ligation Master Mix plus DNA and beads.

⚠ Ensure that you record the index used for each sample.

d. Pipette up and down 10 times to mix.

e. Incubate the Ligation reaction at +20°C for 15 minutes.

f. Following the incubation, proceed immediately to the next step.

9. Perform the First Post Ligation Cleanup as follows:

a. To each 50 µl Ligation Reaction add 50 µl of thawed, room temperature PEG/NaCl SPRI Solution, resulting in a total volume of 100 µl.

First post ligation cleanup	Per individual sample library
Ligation reaction	50 µl
PEG/NaCl SRPI solution	50 µl
Total	**100 µl**

As End-repair and A-tailing cleanup were performed with MinElute Reaction Cleanup Kit, in this step we added Agencourt AMPure XP beads. Beads were added in a 1.8x ratio. 50 µl Ligation reaction x 1.8 = 90 µl Agencourt AMPure XP beads added to the Ligation reaction (No need to add PEG/NaCl SRPI solution in this step).

b. Mix thoroughly by pipetting up and down multiple times.

c. Incubate the tube at room temperature for 15 minutes to allow the DNA to bind to the beads.

d. Place the tube on a magnet to capture the beads. Incubate until the liquid is clear.

e. Carefully remove and discard the supernatant.

f. Keeping the tube on the magnet, add 200 µl of freshly-prepared 80% ethanol.

g. Incubate the tube at room temperature for ≥30 seconds.

h. Carefully remove and discard the ethanol.

i. Keeping the tube on the magnet, add 200 µl of freshly-prepared 80% ethanol.

j. Incubate the tube at room temperature for ≥30 seconds.

k. Carefully remove and discard the ethanol. Try to remove all residual ethanol without disturbing the beads.

l. Allow the beads to dry at room temperature, sufficiently for all the ethanol to evaporate.

⚠ Caution: Over-drying the beads may result in dramatic yield loss.

m. Remove the tube from the magnet.

n. Thoroughly resuspend the beads in 100 µl of elution buffer (10mM Tris-HCl, pH 8.0 or PCR-grade water).

🔍 In this and subsequent steps, use buffer rather than PCR-grade water if the eluted sample will be stored for an extended period of time (> 24 hours).

o. Incubate the tube at room temperature for 2 minutes to allow the DNA to elute off the beads.

p. Proceed immediately to the next step.

10. Perform the Dual-SPRI Size Selection:

a. To each tube containing 100 µl resuspended DNA with beads add 60 µl of thawed, room temperature PEG/NaCl SPRI Solution, resulting in a total volume of 160 µl.

Dual-SRPI size cleanup	**Per individual sample library**
Resuspended DNA with beads	100 µl
PEG/NaCl SRPI solution	60 µl
Total	**160 µl**

b. Mix thoroughly by pipetting up and down multiple times.

c. Incubate the tube at room temperature for 15 minutes to allow library fragments larger than ~450 bp to bind to the beads.

d. Place the tube on a magnet to capture the beads. Incubate until the liquid is clear.

e. Carefully transfer 155µl of the supernatant containing library fragments smaller than ~450 bp to a new tube.

⚠ Do NOT discard the supernatant at this step. It is also critical to not transfer any beads with the supernatant.

f. To the 155 µl supernatant add 20 µl of room temperature Agencourt AMPure XP beads.

g. Thoroughly resuspend the beads by pipetting up and down multiple times.

h. Incubate the tube at room temperature for 15 minutes to allow library fragments larger than ~250 bp to bind to the beads.

i. Place the tube on a magnet to capture the beads. Incubate until the liquid is clear.

j. Carefully remove and discard the supernatant.

k. Keeping the tube on the magnet, add 200 µl of freshly-prepared 80% ethanol.

l. Incubate the tube at room temperature for ≥30 seconds.

m. Carefully remove and discard the ethanol.

n. Keeping the tube on the magnet, add 200 µl of freshly-prepared 80% ethanol.

o. Incubate the tube at room temperature for ≥30 seconds.

p. Carefully remove and discard the ethanol. Try to remove all residual ethanol without disturbing the beads.

q. Allow the beads to dry at room temperature, sufficiently for all the ethanol to evaporate.

⚠ Caution: Over-drying the beads may result in dramatic yield loss.

r. Remove the tube from the magnet.

s. Thoroughly resuspend the beads in 25 µl of elution buffer (10 mM Tris-HCl, pH 8.0 or PCR-grade water).

t. Incubate the tube at room temperature for 2 minutes to allow the DNA to elute off the beads.

u. Place the tube on a magnet to capture the beads. Incubate until the liquid is clear.

v. Transfer the clear supernatant to a new tube and proceed with the amplification of the sample library as detailed in Chapter 4.

8.2.6.2. Amplify the Sample Library Using LM-PCR

Materials needed:

- SeqCap EZ Accessory Kit v2
- SeqCap Adapter Kit A and/or B
- SeqCap Pure Capture Bead Kit

Ensure that the following is available:

- Freshly-prepared 80% ethanol: 0.4 ml per DNA sample

Sample Requirements

For each sample library to be captured, 20 μl of the sample library from Chapter 3 is amplified via Pre-Capture LM-PCR.

•Step 1. Resuspend the SeqCap Pre-LM-PCR Oligos

1. Briefly spin the lyophilized 'Pre-LM-PCR Oligos 1 & 2', contained in the SeqCap Adapter Kit A and/or B, to allow the contents to pellet at the bottom of the tube. Please note that both oligos are contained within a single tube.

2. Add 550 μl PCR-grade water to the tube of centrifuged oligos.

3. Briefly vortex the resuspended oligos.

4. Spin down the tube to collect contents.

5. The resuspended oligo tube should be stored at -15 to -25°C.

•Step 2. Prepare the Pre-Capture LM-PCR Master Mix

⚠ The Pre-Capture LM-PCR Master Mix is temperature sensitive. Thawing of components and preparation of LM-PCR reactions must be performed on ice.

We recommend the inclusion of negative (water) and positive (previously amplified library) controls in the Pre-Capture LM-PCR step.

Instructions for preparing an individual PCR reaction are shown here. When assembling a master mix for processing multiple samples, prepare an excess volume of ~5% to allow for complete pipetting (liquid handling systems may require an excess of ~20%).

1. To each PCR tube/well add 30 µl of Pre-Capture LM-PCR Master Mix, resulting in a total volume of 30 µl per tube.

Pre-Capture LM-PCR Master Mix	**Per individual sample library or negative control**
KAPA HiFi HotStart ReadyMix (2x)	25 µl
Pre LM-PCR Oligos 1 & 2, 5 µM*	5 µl
Total	**30 µl**

* Note: The pre-capture LM-PCR Oligos are contained within the SeqCap Adapter Kit A and/or B.

2. Add the 20 µl of sample library (or PCR-grade water for negative control) to the PCR tube or each well of the 96-well plate containing the LM-PCR Master Mix.

3. Mix well by pipetting up and down five times. Do not vortex.

•Step 3. Perform the Pre-Capture PCR Amplification

1. Place the PCR tube (or 96-well PCR plate) in the thermocycler.

2. Amplify the sample library using the following Pre-Capture LM-PCR program:

- Step 1: 45 seconds at +98°C
- Step 2: 15 seconds at +98°C
- Step 3: 30 seconds at +60°C
- Step 4: 30 seconds at +72°C
- Step 5: Go to Step 2, repeat eight times (for a total of nine cycles)
- Step 6: 1 minute at +72°C
- Step 7: Hold @ +4°C

> Amplification of each sample library was performed using the pre-capture LM-PCR program, with a total of 12 cycles.

3. Store the reaction at +2 to +8°C until ready for cleanup, up to 72 hours.

•Step 4. Purify the Amplified Sample Library using Agencourt AMPure XP Beads

Alternatively, samples can be purified using the Qiagen QIAquick PCR Purification Kit. If this purification method is chosen instead of the Agencourt AMPure XP Beads, follow the protocol detailed in Appendix D.

1. Allow the Agencourt AMPure XP Beads, contained in the SeqCap Pure Capture Bead Kit, to warm to room temperature for at least 30 minutes before use.

2. Transfer each amplified sample library (approximately 50 µl) into a separate 0.2 ml PCR tube (for use with a DynaMag-96 Side Magnet) or 1.5 ml tube (for use with a DynaMag-2 Magnet). Process the negative control in exactly the same way as the amplified sample library.

3. Vortex the Agencourt AMPure XP Beads for 10 seconds before use to ensure a homogenous mixture of beads.

4. Add 90 µl Agencourt AMPure XP Beads to the 50 µl amplified sample library.

5. Vortex briefly.

6. Incubate at room temperature for 15 minutes to allow the DNA to bind to the beads.

7. Place the tube containing the bead bound DNA in a magnetic particle collector.

8. Allow the solution to clear.

9. Once clear, remove and discard the supernatant being careful not to disturb the beads.

10. Add 200 µl freshly-prepared 80% ethanol to the tube containing the beads plus DNA. The tube should be left in the magnetic particle collector during this step.

11. Incubate at room temperature for 30 seconds.

12. Remove and discard the 80% ethanol and repeat Steps 4.9-4.11 for a total of two washes with 80% ethanol.

13. Following the second wash, remove and discard all of the 80% ethanol.

14. Allow the beads to dry at room temperature with the tube lid open for 15 minutes (or until dry).

⚠ Over drying of the beads can result in yield loss.

15. Remove the tube from the magnetic particle collector.

16. Resuspend the DNA using 52 µl of PCR-grade water.

⚠ It is critical that the amplified sample library is eluted with PCR-grade water and not buffer EB or 1X TE.

17. Pipet up and down ten times to mix to ensure that all of the beads are resuspended.

18. Incubate at room temperature for 2 minutes.

19. Place the tube back in the magnetic particle collector and allow the solution to clear.

20. Remove 50 µl of the supernatant that now contains the amplified sample library and transfer into a new 1.5 ml tube.

•Step 5. Check the Quality of the Amplified Sample Library

1. Measure the A260/A280 ratio of the amplified sample library to quantify the DNA concentration using a NanoDrop spectrophotometer and determine the DNA quality.

⚠ When working with samples that will be pooled for hybridization (i.e. multiplex Sequence Capture), accurate quantitation is essential. Alternative quantitation methods, such as those that are fluorometry-based, should be used in place of, or in addition to, the NanoDrop spectrophotometer. Slight differences in the mass of each sample combined to form the ‘Multiplex DNA Sample Library Pool’ will result in variations in the total number of sequencing reads obtained for each sample in the library pool.

- The A260/A280 ratio should be 1.7 - 2.0.
- The sample library yield should be > 1.0 µg.

- The negative control yield should be negligible. If this is not the case, the measurement may be high due to the presence of unincorporated primers carried over from the LM-PCR reaction and not an indication of possible contamination between amplified sample libraries.

2. Run 1 µl of each amplified sample library (and any negative controls) on an Agilent Bioanalyzer DNA 1000 chip. Run the chip according to manufacturer's instructions.

- The Bioanalyzer should indicate that average fragment size falls between 150 - 500 bp (Figure 2). The negative control should not show any significant signal within this size range, which could indicate contamination between amplified sample libraries. A sharp peak may be visible below 150 bp. This peak, which consists of unincorporated primers carried over from previous steps or the LM-PCR reaction, will not interfere with the capture process.
- The negative control should not show any signal above baseline within the 150 - 400 bp size range, which could indicate contamination between amplified sample libraries, but it may exhibit sharp peaks visible below 150 bp. If the negative control reaction shows a positive signal by the NanoDrop spectrophotometer, but the Bioanalyzer trace indicates only the presence of a sharp peak below 150 bp in size, then the negative control should not be considered contaminated.

3. If the amplified sample library meets these requirements, proceed to Chapter 5. If the amplified sample library does not meet these requirements, reconstruct the library.

8.2.6.3. Hybridize the Sample and SeqCap EZ Libraries

Materials needed:

- SeqCap EZ Library
- SeqCap Hybridization and Wash Kit
- SeqCap EZ Accessory Kit v2
- SeqCap HE Oligo Kit

The hybridization protocol requires a thermocycler capable of maintaining +47°C for 16 - 20 hours. A programmable heated lid is required.

Note: Instructions for using SeqCap HE-Oligo Kits A & B with automated liquid handling instruments for setting up hybridizations is described in Appendix A.

Note: In this chapter we use the term ‘Multiplex DNA Sample Library Pool’, however a single DNA sample library may be captured using the same instructions. It is not required to capture more than one library at a time.

Experiment 1: pool of 16 libraries; Experiment 2: Two pools each containing four libraries, one from each extract. (Scheme explained in Figure 1 - Section 4.1).

•Step 1. Prepare for Hybridization

1. Turn on a heat block to +95°C and let it equilibrate to the set temperature.

2. Remove the appropriate number of 4.5 µl SeqCap EZ Library aliquots (one per hybridization) from the -15 to -25°C freezer and allow them to thaw on ice.

•Step 2. Resuspend the SeqCap HE Universal and SeqCap HE Index Oligos

1. Briefly spin the lyophilized oligo tubes, contained in the SeqCap HE-Oligo Kits A and/or B, to allow the contents to pellet to the bottom of the tube.

2. Add 120 µl PCR-grade water to the SeqCap HE Universal Oligo tube (1,000 µM final concentration).

3. Add 10 µl PCR-grade water to each SeqCap HE Index Oligo tube (1,000 µM final concentration).

4. Vortex the primers plus PCR-grade water for five seconds and spin down the resuspended oligo tube.

5. The resuspended oligo tube should be stored at -15 to -25°C.

⚠ To prevent damage to the Hybridization Enhancing (HE) oligos due to multiple freeze/thaw cycles, once resuspended the oligos can be aliquoted into smaller volumes to minimize the number of freeze/thaw cycles.

•Step 3. Prepare the Multiplex DNA Sample Library Pool

1. Thaw on ice each of the uniquely indexed amplified DNA sample libraries that will be included in the multiplex capture experiment (generated in Chapter 4).

2. Mix together equal amounts (by mass) of each of these amplified DNA sample libraries to obtain a single pool with a combined mass of at least 1.25 µg. This mixture will subsequently be referred to as the ‘Multiplex DNA Sample Library Pool’. One µg of the multiplex DNA sample library pool will be used in the sequence capture hybridization step, and 60 ng will be used for measurement of enrichment using qPCR (Chapter 8).

⚠ To obtain equal numbers of sequencing reads from each component libraries in the Multiplex DNA Sample Library Pool upon completion of the experiment, it is very important to combine identical amounts of each independently amplified DNA sample library at this step. Accurate quantification and pipetting are critical.

🔍 Note: Store remaining 250 ng of Multiplex DNA Sample Library Pool at -15 to -25°C until use in measurement of enrichment using qPCR (Chapter 8).

•Step 4. Prepare the Multiplex Hybridization Enhancing Oligo Pool

1. Thaw on ice the resuspended SeqCap HE Universal Oligo (1,000 µM) and each resuspended SeqCap HE Index oligo (1,000 µM) that matches a DNA Adapter Index included in the Multiplex DNA Sample Library Pool from Step 2 of this section.

2. Mix together the HE oligos so that the resulting Multiplex Hybridization Enhancing Oligo Pool contains, by mass, 50% SeqCap HE Universal Oligo and 50% of a mixture of the appropriate SeqCap HE Index oligos. The total combined mass of the Multiplex Hybridization Enhancing Oligo Pool should be 2,000 pmol, which is the amount required for a single Sequence Capture experiment.

Example: If a Multiplex DNA Sample Library Pool contains four DNA sample libraries prepared with SeqCap Adapter Indexes 2, 4, 6, and 8, respectively, then the Multiplex Hybridization Enhancing Oligo Pool would contain the following:

Component	Amount
SeqCap HE Universal Oligo	1,000 pmol (1 µl of 1,000 µM)
SeqCap HE Index 2 Oligo	250 pmol (0.25 µl of 1,000 µM)
SeqCap HE Index 4 Oligo	250 pmol (0.25 µl of 1,000 µM)
SeqCap HE Index 6 Oligo	250 pmol (0.25 µl of 1,000 µM)
SeqCap HE Index 8 Oligo	250 pmol (0.25 µl of 1,000 µM)
Total	**2,000 pmol (2 µl of 1,000 µM)**

🔍 Due to the difficulty of accurately pipetting small volumes, it is recommended to either prepare a larger volume of the Multiplex Hybridization Enhancing Oligo Pool using the 1,000 µM stocks or dilute the 1,000 µM stocks and then pool. These pools can be dispensed into individual single-use aliquots that can be stored at -15 to -25°C until needed.

🔍 For optimal results, it is important that the individual SeqCap HE oligos contained in a Multiplex Hybridization Enhancing Oligo Pool are precisely matched with the adapter indexes present in the Multiplex DNA Sample Library Pool in a multiplexed Sequence Capture experiment.

•Step 5. Prepare the Hybridization Sample

Note: When working with non-human gDNA, consider using the SeqCap EZ Developer Reagent (catalog number 06684335001) in place of COT Human DNA. When using the SeqCap EZ Developer Reagent, add 10 µl of this reagent to each hybridization instead of COT Human DNA.

1. Add 5 µl of COT Human DNA (1 mg/ml), contained in the SeqCap EZ Accessory Kit v2, to a new 1.5 ml tube.

2. Add 1 µg of Multiplex DNA Sample Library to the 1.5 ml tube containing 5 µl of COT Human DNA.

Each pool hybridization reaction was performed by adding 1.5 µg of the equimolar pool of 16 DNA libraries in Experiment 1 and 0.24 µg of the equimolar pools of 4 DNA libraries in Experiment 2 (Scheme explained in Figure 1 - Section 4.1).

3. Add 2,000 pmol (or 2 µl) of the Multiplex Hybridization Enhancing Oligo Pool (1 µl of 1,000 pmol SeqCap HE Universal Oligo and 1 µl of the 1,000 pmol SeqCap HE Index Oligo pool) to the Multiplex DNA Sample Library Pool plus COT Human DNA.

The tube should now contain the following components:

Component	**Amount**	**Volume**
COT Human DNA	5 µg	5 µl
Multiplex DNA Sample Lib pool	1 µg	≤ 50 µl
SeqCap HE Universal Oligo	1,000 pmol	1 µl
SeqCap HE Index Oligo pool	1,000 pmol	1 µl
Total		≤ 57 µl

4. Close the tube’s lid and make a hole in the top of the tube’s cap with an 18 - 20 gauge or smaller needle.

🔍 The closed lid with a hole in the top of the tube’s cap is a precaution to suppress contamination in the DNA vacuum concentrator.

5. Dry the Multiplex DNA Sample Library Pool/COT Human DNA/Multiplex Hybridization Enhancing Oligo Pool in a DNA vacuum concentrator on high heat (+60°C).

🔍 Denaturation of the DNA with high heat is not problematic because the hybridization utilizes single-stranded DNA.

6. To each dried-down Multiplex DNA Sample Library Pool/COT Human DNA/Multiplex Hybridization Enhancing Oligo Pool, add:

- 7.5 µl of 2X Hybridization Buffer (vial 5)
- 3 µl of Hybridization Component A (vial 6)

The tube should now contain the following components:

Component	**Solution Capture**
COT Human DNA	5 µg
Multiplex DNA Sample Lib pool	1 µg
Multiplex Hybridization Enhancing Oligo Pool	2,000 pmol*
2X Hybridization Buffer (vial 5)	7.5 µl
Hybridization Component A (vial 6)	3 µl
Total	**10.5 µl**

*Composed of 50% (1,000 pmol) SeqCap HE Universal Oligo and 50% (1,000 pmol) of a mixture of the appropriate SeqCap HE Index oligos.

7. Cover the hole in the tube's cap with a sticker or small piece of laboratory tape.

8. Vortex the Multiplex DNA Sample Library Pool/COT Human DNA/Multiplex Hybridization Enhancing Oligo Pool plus Hybridization Cocktail (2X Hybridization Buffer + Hybridization Component A) for 10 seconds.

9. Centrifuge at maximum speed for 10 seconds.

10. Place the Multiplex DNA Sample Library Pool/COT Human DNA/Multiplex Hybridization Enhancing Oligo Pool/Hybridization Cocktail in a +95°C heat block for 10 minutes to denature the DNA.

11. Centrifuge the Multiplex DNA Sample Library Pool/COT Human DNA/Multiplex Hybridization Enhancing Oligo Pool/Hybridization Cocktail at maximum speed for 10 seconds at room temperature.

12. Transfer the Multiplex DNA Sample Library Pool/COT Human DNA/Multiplex Hybridization Enhancing Oligo Pool/Hybridization Cocktail to the 4.5 µl aliquot of EZ Library in a 0.2 ml PCR tube prepared in Chapter 2 (the entire volume can also be transferred to one well of a 96-well PCR plate).

13. Vortex for 3 seconds.

14. Centrifuge at maximum speed for 10 seconds.

The hybridization sample should now contain the following components:

Component	Solution Capture
COT Human DNA	5 µg
Multiplex DNA Sample Lib pool	1 µg
Multiplex Hybridization Enhancing Oligo Pool	2,000 pmol*
2X Hybridization Buffer (vial 5)	7.5 µl
Hybridization Component A (vial 6)	3 µl
EZ Library	4.5 µl
Total	**10.5 µl**

*Composed of 50% (1,000 pmol) SeqCap HE Universal Oligo and 50% (1,000 pmol) of a mixture of the appropriate SeqCap HE Index oligos.

15. Incubate in a thermocycler at +47°C for 16 - 20 hours. The thermocycler's heated lid should be turned on and set to maintain +57°C (10°C above the hybridization temperature).

Incubated in a thermocycler at +47 °C for 36 hours.

8.2.6.4. Wash and Recover Captured Multiplex DNA Sample

Washing and recovery of the captured multiplex DNA sample from the hybridization of the Multiplex DNA Sample Library Pool and SeqCap EZ Library. (Refer to Appendix C for instructions for increased throughput applications.) This chapter requires the use of components from the following kits:

- SeqCap Hybridization and Wash Kit
- SeqCap Pure Capture Bead Kit

Ensure that the following is available:

- Additional PCR-grade water for buffer preparation and elution.

⚠ It is extremely important that the water bath temperature be closely monitored and remains at +47°C. Because the displayed temperatures on many water baths are often imprecise, Roche NimbleGen recommends that you place an external, calibrated thermometer in the water bath.

⚠ Equilibrate buffers at +47°C for at least 2 hours before washing the captured Multiplex DNA sample.

• Step 1. Prepare Sequence Capture and Bead Wash Buffers

🔍 Volumes for an individual capture are shown here. When preparing 1X buffers for processing multiple reactions, prepare an excess volume of ~5% to allow for complete pipetting (liquid handling systems may require an excess of ~20%).

1. Dilute 10X Wash Buffers (I, II, III and Stringent) and 2.5X Bead Wash Buffer, contained in the SeqCap Hybridization and Wash Kit,

to create 1X working solutions. Volumes listed below are sufficient for one capture.

Component	Vol. Concentrated buffer	Vol. PCR-grade water	Total vol. 1X Buffer*
10X Stringent Wash Buffer (vial 4)	40 µl	360 µl	**400 µl**
10X Wash Buffer I (vial 1)	30 µl	270 µl	**300 µl**
10X Wash Buffer II (vial 2)	20 µl	180 µl	**200 µl**
10X Wash Buffer III (vial 3)	20 µl	180 µl	**200 µl**
2.5X Bead Wash Buffer (vial 7)	200 µl	300 µl	**500 µl**

*Store working solutions at room temperature (+15 to +25°C) for up to 2 weeks. The volumes in this table are calculated for a single experiment; scale up accordingly if multiple samples will be processed.

2. Preheat the following wash buffers to +47°C in a water bath:

• 400 µl of 1X Stringent Wash Buffer

• 100 µl of 1X Wash Buffer I.

•Step 2. Prepare the Capture Beads

1. Allow the Capture Beads to warm to room temperature for 30 minutes prior to use.

2. Mix the beads thoroughly by vortexing for 15 seconds.

3. Aliquot 100 µl of beads for each capture into a single 1.5 ml tube (i.e. for one capture use 100 µl beads and for four captures use 400 µl beads, etc.). Enough beads for six captures can be prepared in a single tube.

4. Place the tube in a DynaMag-2 device. When the liquid becomes clear (should take less than 5 minutes), remove and discard the liquid being careful to leave all of the beads in the tube. Any remaining traces of liquid will be removed with subsequent wash steps.

5. While the tube is in the DynaMag-2 device, add twice the initial volume of beads of 1X Bead Wash Buffer (i.e. for one capture use 200 µl of buffer and for four captures use 800 µl buffer, etc.).

6. Remove the tube from the DynaMag-2 device and vortex for 10 seconds.

7. Place the tube back in the DynaMag-2 device to bind the beads.

8. Once clear, remove and discard the liquid.

9. Repeat Steps 2.5 - 2.8 for a total of two washes.

10. After removing the buffer following the second wash, resuspend by vortexing the beads in 1x the original volume using the 1X Bead Wash Buffer (i.e. for one capture use 100 µl buffer and for four captures use 400 µl buffer, etc.).

11. Aliquot 100 µl of resuspended beads into new 0.2 ml tubes (i.e. one tube for each capture).

12. Place the tube in the DynaMag-2 device to bind the beads. Once clear, remove and discard the liquid.

13. The Capture Beads are now ready to bind the captured DNA. Proceed immediately to the next step.

⚠ Do not allow the Capture Beads to dry out. Small amounts of residual Bead Wash Buffer will not interfere with binding of DNA to the Capture Beads.

•Step 3. Bind DNA to the Capture Beads

1. Transfer the hybridization samples to the Capture Beads prepared in the previous step.

2. Mix thoroughly by pipetting up and down ten times.

3. Bind the captured sample to the beads by placing the tubes containing the beads and DNA in a thermocycler set to +47°C for 45 minutes (heated lid set to +57°C). Mix the samples by vortexing for 3 seconds at 15 minute intervals to ensure that the beads remain in suspension. It is helpful to have a vortex mixer located close to the thermocycler for this step.

•Step 4. Wash the Capture Beads Plus Bound DNA

1. After the 45-minute incubation, add 100 µl of 1X Wash Buffer I heated to +47°C to the 15 µl of Capture Beads Plus Bound DNA.

2. Mix by vortexing for 10 seconds.

3. Transfer the entire content of each 0.2 ml tube to a 1.5 ml tube.

4. Place the tubes in the DynaMag-2 device to bind the beads.

5. Remove and discard the liquid once clear.

6. Remove the tubes from the DynaMag-2 device and add 200 µl of 1X Stringent Wash Buffer heated to +47°C.

7. Pipette up and down ten times to mix. Work quickly so that the temperature does not drop much below +47°C.

8. Incubate at +47°C for 5 minutes.

9. Repeat Steps 4.4 - 4.8 for a total of two washes using 1X Stringent Wash Buffer heated to +47°C.

10. Place the tubes in the DynaMag-2 device to bind the beads.

11. Remove and discard the liquid once clear.

12. Add 200 µl of room temperature 1X Wash Buffer I and mix by vortexing for 2 minutes. If liquid has collected in the tube's cap, tap the tube gently to collect the liquid into the tube's bottom before continuing to the next step.

13. Place the tubes in the DynaMag-2 device to bind the beads.

14. Remove and discard the liquid once clear.

15. Add 200 µl of room temperature 1X Wash Buffer II.

16. Mix by vortexing for 1 minute.

17. Place the tubes in the DynaMag-2 device to bind the beads.

18. Remove and discard the liquid once clear.

19. Add 200 µl of room temperature 1X Wash Buffer III.

20. Mix by vortexing for 30 seconds.

21. Place the tubes in the DynaMag-2 device to bind the beads.

22. Remove and discard the liquid once clear.

23. Remove the tubes from the DynaMag-2 device.

24. Add 50 µl PCR-grade water to each tube of bead-bound captured sample.

25. Store the beads plus captured samples at -15 to -25°C or proceed to Chapter 7.

⚠ There is no need to elute DNA off the beads. The beads plus captured DNA will be used as template in the LM-PCR as described in Chapter 7.

8.2.6.5. Amplify Captured Multiplex DNA Sample Using LM-PCR

Amplification of captured Multiplex DNA sample, bound to the Capture Beads, using LM-PCR. A total of two reactions are performed per sample, and subsequently combined, to minimize PCR bias. This chapter requires the use of components from the following kits:

•SeqCap EZ Accessory Kit v2

•SeqCap Pure Capture Bead Kit

In addition, ensure that the following are available:

•Additional PCR-grade water for 80% ethanol preparation and elution

•Freshly-prepared 80% ethanol: 0.4 ml per DNA sample

<u>•Step 1. Resuspend the Post-LM-PCR Oligos</u>

1. Briefly spin the lyophilized ‘Post-LM-PCR Oligos 1 & 2’ oligos, contained in the SeqCap EZ Accessory Kit v2, to allow the contents to pellet at the bottom of the tube. Please note that both oligos are contained within a single tube.

2. Add 480 µl PCR-grade water to the tube of centrifuged oligos.

3. Briefly vortex the resuspended oligos.

4. Spin down the tube to collect the contents.

5. The resuspended oligo tube should be stored at -15 to -25°C.

•Step 2. Prepare the Post-Capture LM-PCR Master Mix

🔍 The Post-Capture LM-PCR Master Mix and the individual PCR tubes must be prepared on ice.

🔍 Instructions for preparing individual PCR reactions are shown here. When assembling a master mix for processing multiple samples, prepare an excess volume of ~5% to allow for complete pipetting (liquid handling systems may require an excess of ~20%). Note that each captured DNA sample requires two PCR reactions.

1. To each PCR tube/well (one pair per captured DNA sample) add 30 µl of Post-Capture LM-PCR Master Mix, resulting in a total volume of 30 µl per tube, or 60 µl per DNA sample.

Post-Capture LM-PCR Master Mix	**Per individual sample library**
KAPA HiFi HotStart ReadyMix	25 µl
Post LM-PCR Oligos 1 & 2, 5 µM*	5 µl
Total	**30 µl**

* Note: The post-capture LM-PCR Oligos are contained within the SeqCap EZ Accessory Kit v2.

⚠ Two LM-PCR reactions will be performed for each captured multiplex DNA sample. The total volume of the PCR Master Mix is 60 µl that will be distributed in two tubes (30 µl each).

2. Vortex the bead-bound captured DNA to ensure a homogenous mixture of beads.

3. Aliquot 20 µl of bead-bound captured DNA as template into each of the two PCR tubes/wells.

4. Mix well by pipetting up and down.

5. Add 20 µl of PCR-grade water to the negative control.

6. Mix well by pipetting up and down five times.

7. Store the remaining bead bound captured DNA at -15 to -25°C.

•Step 3. Perform the Post-Capture PCR Amplification

1. Place PCR tubes/plate in the thermocycler.

2. Amplify the captured DNA using the following Post-Capture LM-PCR program:

- Step 1: 45 seconds @ +98°C
- Step 2: 15 seconds @ +98°C
- Step 3: 30 seconds @ +60°C
- Step 4: 30 seconds @ +72°C
- Step 5: Go to Step 2, repeat 13 times (for a total of 14 cycles)
- Step 6: 1 minutes @ +72°C
- Step 7: Hold @ +4°C

Captured DNA was amplified using the Post-Capture LM-PCR program, with a total of 12 cycles.

3. Store reactions at +2 to +8°C until ready for purification, up to 72 hours.

•Step 4. Purify the Amplified Captured Multiplex DNA Sample using Agencourt AMPure XP Beads

🔍 Alternatively, samples can be purified using the Qiagen QIAquick PCR Purification Kit. If this purification method is chosen instead of the Agencourt AMPure XP Beads, follow the protocol detailed in Appendix D.

1. Allow the Agencourt AMPure XP Beads to warm to room temperature for at least 30 minutes before use.

2. Pool the like amplified captured Multiplex DNA Sample Libraries into a 1.5 ml microcentrifuge tube (approximately 100 µl). Process the negative control in exactly the same way as the amplified sample library.

3. Vortex the beads for 10 seconds before use to ensure a homogenous mixture of beads.

4. Add 180 µl Agencourt AMPure XP Beads to the 100 µl pooled amplified captured Multiplex DNA Sample library.

5. Vortex briefly.

6. Incubate at room temperature for 15 minutes to allow the DNA to bind to the beads.

7. Place the tube containing the bead bound DNA in a magnetic particle collector.

8. Allow the solution to clear.

9. Once clear, remove and discard the supernatant being careful not to disturb the beads.

10. Add 200 µl freshly-prepared 80% ethanol to the tube containing the beads plus DNA. The tube should be left in the magnetic particle collector during this step.

11. Incubate at room temperature for 30 seconds.

12. Remove and discard the 80% ethanol, and repeat Steps 4.9-4.11 for a total of two washes with 80% ethanol.

13. Following the second wash, remove and discard all of the 80% ethanol.

14. Allow the beads to dry at room temperature with the tube lid open for 30 minutes (or until dry).

⚠ Over drying of the beads can result in yield loss.

15. Remove the tube from the magnetic particle collector.

16. Resuspend the DNA using 52 µl of PCR-grade water.

17. Pipet up and down ten times to mix to ensure that all of the beads are resuspended.

18. Incubate at room temperature for 2 minutes.

19. Place the tube back in the magnetic particle collector and allow the solution to clear.

20. Remove 50 µl of the supernatant that now contains the amplified captured Multiplex DNA Sample Library Pool and transfer into a new 1.5 ml tube.

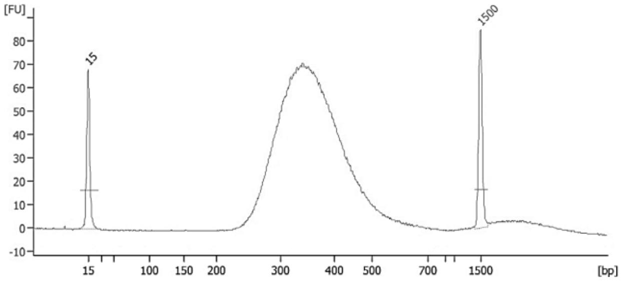
•Step 5. Determine the Concentration, Size Distribution, and Quality of the Amplified Captured Multiplex DNA Sample

1. Quantify the DNA concentration and measure the A260/A280 ratio of the amplified captured multiplex DNA and negative control using a NanoDrop spectrophotometer.

- The A260/A280 ratio should be 1.7 - 2.0.
- The LM-PCR yield should be ≥500 ng.
- The negative control should not show significant amplification, which could be indicative of contamination.

2. Run 1 μl of the amplified captured multiplex DNA sample and negative control using an Agilent Bioanalyzer DNA 1000 chip. Run the chip according to manufacturer's instructions. Amplified captured multiplex DNA should exhibit the following characteristics:

- The average fragment length should be between 150 - 500 bp.



Example of successfully amplified captured multiplex DNA analyzed using an Agilent Bioanalyzer DNA 1000 chip.

3. If the amplified captured multiplex DNA meets the requirements, proceed to Chapter 8.

If the amplified captured multiplex DNA does not meet the A260/A280 ratio requirement, purify again using the Agencourt

AMPure XP Beads (or alternatively, a second Qiagen QIAquick PCR Purification column).

A **second hybridization** was performed for the three pool replicates in Experiment 1 and four pool replicates in Experiment 2, as illustrated in Figure 1 - Section 4.1, following the same protocol as the first hybridization. Only the amount of starting material was altered, using for each of the second hybridizations all the material obtained after the PCR purification from the first hybridization. To limit the extent of PCR-duplicates the captured product of the second hybridization was amplified with 8 PCR cycles rather than 12.

References:

•KAPA Library Preparation Kit Technical Data Sheet, KR0935 – v2.14 (or later) (hard-copy included in the KAPA Library Preparation Kit or contact Kapa Biosystems Technical Support to obtain pdf, at support@kapabiosystems.com).

Protocol extracted from:

(SeqCap EZ Library SR User's Guide version 5.0 used in this thesis, but not available online)

http://netdocs.roche.com/DDM/Effective/06588786001_RNG_SeqCap_EZ_UGuide_v5.3.pdf