Aus dem Robert Koch-Institut eingereicht
über das Institut für Mikrobiologie und Tierseuchen des Fachbereichs
Veterinärmedizin der Freien Universität Berlin

Modalities of Transmission of Simian Foamy Virus in Wild Chimpanzees

Vorkommen und Übertragung von Simianen Foamy Viren
bei wildlebenden Schimpansen

Inaugural-Dissertation
zur Erlangung des Grades eines
Doktors der Veterinärmedizin
an der
Freien Universität Berlin

vorgelegt von
Anja Blasse
Tierärztin aus Berlin

Berlin 2013

Journal-Nr.: 3706

Gedruckt mit Genehmigung des Fachbereichs Veterinärmedizin

der Freien Universität Berlin

Dekan:	Univ.-Prof. Dr. Jürgen Zentek
Erster Gutachter:	Univ.-Prof. Dr. Lothar H. Wieler
Zweiter Gutachter:	Prof. Dr. Georg Pauli
Dritter Gutachter:	Prof. Dr. Benedikt Kaufer

Deskriptoren (nach CAB-Thesaurus):
Simian foamy virus, wild animals, chimpanzees, mixed infections, polymerase chain reaction, viral load

Tag der Promotion: 10. Juni 2014

Bibliografische Information der *Deutschen Nationalbibliothek*
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

ISBN: 978-3-86387-529-9
Zugl.: Berlin, Freie Univ., Diss., 2013
Dissertation, Freie Universität Berlin
D 188

Choriner Str. 85 - 10119 Berlin
verlag@menschundbuch.de – www.menschundbuch.de

für meine Familie

Contents

List of Figures

List of Tables

List of Abbreviations

+G	Rate heterogeneity
+I	Invariant sites
AICc	Akaike's information criterion, corrected for small samples
aLRT	Approximate likelihood ratio test
APOBEC	Apolipoprotein B mRNA editing enzyme
as	Antisense (strand)
AT	Annealing temperature
B1-B4	Individuals from Budongo Forest
BFV	Bovine foamy virus
bp	Basepair
bt	Bootstrap values
cDNA	Complementary DNA
CoeEFV	Coelacanth endogenous foamy-like virus
df	Degrees of freedom
DNA	Desoxyribonucleic acid
dsDNA	Double stranded DNA
EFV	Equine foamy virus
Env/*env*	Envelope protein/gene for envelope
EPD-PCR	End-point dilution PCR
FFV, FFVpc	Feline foamy virus, FFV *puma concolor*
FV	Foamy virus
G-to-A	Guanosine-to-adenosine
Gag/*gag*	Group-specific antigen/gene for group-specific antigen
GLMM	Generalized linear mixed model
GR-boxes	Glycine-arginine-rich basis sequences
GTR	Global time reversible
HFV	Human foamy virus
HIV-1	Human immunodeficiency virus type 1
HTLV	Human T-lymphotrophic virus
IP	Internal promoter
K80	Kimura 2-parameters
kb	Kilobase
LTR	Long-terminal repeats

mRNA	Messenger RNA
ML	Maximum likelihood
MT	Microtubule
MTOC	MT-organization centre
Mya	Million of years
NA	Not assessed
NHP	Non-human primate
OP	Outgroup probabilities
ORF	Open reading frame
p	*p*-value (probability)
P. t.	*Pan troglodytes*
PCR	Polymerase chain reaction
Pol/*pol*	Polymerase protein/*polymeras*e gene
pp	Posterior probability value
RaFV	Bat foamy virus
RNA	Ribonucleic acid
s	Sense (strand)
SE	Standard error
SFV	Simian foamy virus
SFVcpz	SFV chimpanzee
SFVmac	SFV macaque
SFVwrc	SFV red colobus monkey
SIV	Simian immunodeficiency virus
SloEFV	Sloth endogenous foamy-like virus
ssRNA	Single stranded ribonucleic acid
STLV	Simian T-cell leukemia virus
T1-T6	Individuals from Taï National Park, Côte d'Ivoire
Tas	Transcriptional transactivator
TCS	Statistical parsimony-based network building by Templeton, Crandall & Sing
VIF	Variance inflation factors
y	Years of age
χ^2	Chi-squared

1 Introduction

Retroviruses are arguably the best-studied microorganisms that infect wild non-human primates (NHPs) [90]. Simian immunodeficiency viruses (SIVs) and simian T-cell leukemia viruses (STLVs) have naturally received considerable attention due to their proximity to human pathogens, but presumably non-pathogenic simian foamy viruses (SFVs) have also been intensively studied. It was quickly realized that SFVs infect nearly all living wild NHPs, which applies at different scales. That is, not only do SFVs infect most NHP species in the wild, they also reach extremely high prevalence in wild NHP adults, *e.g.*, close to 100 % in Western red colobus monkeys (*Piliocolobus badius badius*) [79] or chimpanzees (*Pan troglodytes*) [89, 97]. Though, the modalities and routes of transmission are still relatively unclear.

In this respect, a key question is whether in their primate hosts, which are highly social, SFVs are transmitted horizontally, vertically, or both, vertical transmission being here understood as mother-offspring transmission (whenever it occurs). Increases in seroprevalence and/or SFV nucleic acid detection rate have been documented for several NHP colonies, pointing at horizontal transmission, *e.g.*, through aggressive contacts during adulthood, being a favored route of transmission. However, vertical transmission was only rarely directly investigated and not identified as major route for infection [22]. Given the high prevalence of SFVs, it is, however, likely that many individuals will get infected with different strains over their lifetime (*i.e.*, super-infections), a fact that might pass unnoticed when using bulk-PCR product sequencing (*i.e.,* amplification starts from a pool of molecules, indicating, that the resulting sequence is a mixture of various targets of the underlying viral population [40]). Further insights into in-host SFV variability might therefore lead to different conclusions.

While investigating co-infection with retroviruses belonging to distinct genera does not raise any specific technical difficulty – since independent, genus-specific serological/molecular methods will often be used (*e.g.,* [41, 79]) – studying super-infection appears more challenging. To date, end-point dilution (EPD) PCR is considered as gold standard for the depiction of intra-individual retroviral diversity. It was successfully applied to the investigation of human immunodeficiency virus 1 (HIV-1) micro-evolutionary trends (*e.g.*, [131]) including the detection of super-infection events (*e.g.*, [81]). EPD-PCR might therefore be considered a promising tool for the detection of super-infection cases among wild primate hosts. There is however at least one predictable barrier to its implementation in this context. As retrovirus-oriented studies of wild primate populations essentially rely on the analysis of non-invasively collected samples (most

often feces, *e.g.,* [67, 89]), it is conceivable that samples will only contain retroviral/proviral genomes in minute amounts [83]. Acquiring a substantial sample of EPD-PCR sequences might therefore require prohibitively large amounts of, by definition, rare starting material [77], which questions the use of the gold-standard EPD-PCR in the case of non-invasive samples.

Based on this background the aims of this study were:

1. To establish a diagnostic tool utilizing non-invasive samples, that allows the reconstruction of retrovirus diversity within a given host. Therefore, the aim was to compare costs and benefits of the standard techniques in order to develope a method (bulk-PCR cloning) using fecal samples from two distinct chimpanzee communities (Eastern chimpanzees *Pan troglodytes schweinfurthii* and Western chimpanzees *Pan troglodytes verus*).

2. To investigate the biology of SFVs in a social primate community: a wild group of habituated Western chimpanzees from the Taï forest, Côte d'Ivoire using fecal samples and adapted tools for powerful analyses. Beside SFV persistence and accumulation dynamics, modes of transmission (vertical, horizontal or both) were of particular interest. Therefore, the aim was to compare the influence of mother-offspring bond to relatively weaker social bonds linking offspring to their fathers or non-parental members of the same group on SFV transmission.

2 Literature

2.1 Foamy virus, the extraordinary retrovirus

Foamy viruses (FVs) constitute one of the seven genera of the family *Retroviridae* (Figure 1). All members of this family consistently feature the same morphology, which is characterized by a spherical to pleomorphic form, an envelope and a size of 80-100 nm (Figure 2). The FV genome is relatively large (12 kb) and encodes in addition to usual retroviral structure proteins Gag, Pol and Env two proteins with regulatory functions: Tas and Bet (reviewed in [85]).

However, FVs exhibit some unique biological features that justified the classification of the genera in its own sub-family, *Spumavirinae*, while all other retroviral genera are gathered into the *Orthoretrovirinae* sub-family [87]. An exceptional characteristic is the presence of dsDNA in some SFV virions instead of the usual ss (+) RNA genome. This results from the "unusual timing" of the reverse transcription during the assembly of novel virus particles. Another remarkable feature of FVs is the wide cell tropism including both cell-lines and tissues, which is probably related to an ubiquitous viral receptor. *In vitro*, FVs are extensively cytopathic causing multinucleated and foamy-like*, i.e.*, vacuolated cell syncytia, after which the virus was named [55]. In contrast, *in vivo* the virus seems to be non-pathogenic.

FVs are highly prevalent in a wide host range of mammals [85, 132] with whom they evolved in parallel. Recent studies on coelacanth yielded a co-divergence dating of even more than 400 Millions of years (Mya), which make FV the oldest known virus [48]. Co-divergence might explain the lack of the development of any disease during persistent infections [86, 96, 102]. Although FV are widespread and lead to lifelong infections, the knowledge about its biology, especially in natural hosts, is very limited.

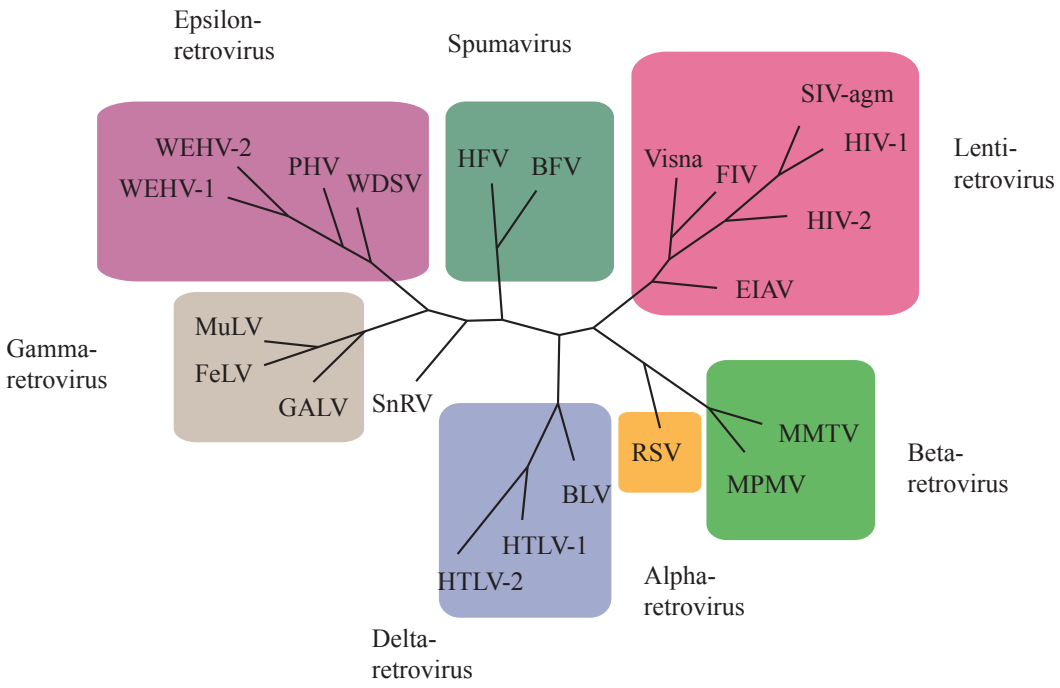

Figure 1: **Phylogenetic tree of *Retroviridae***. The phylogenetic tree was generated on the conserved regions of the polymerase genes using neighbor-joining algorithm. BFV: bovine foamy virus, BLV: bovine leukemia virus, EIAV: equine infectious anemia virus, FeLV: feline leukemia virus, FIV: feline immunodeficiency virus, GALV: gibbon ape leukemia virus, HFV: human foamy virus, HTLV: human T-lymphotrophic virus, MMTV: mouse mammary tumor virus, MPMV: Mason-Pfizer monkey virus, MuLV: murine leukemia virus, PHV: Perch hyperplasia virus, RSV: rous sarcoma virus, SFV: simian foamy virus, SIV: simian immunodeficiency virus, SnRV: snakehead retrovirus, SRV: simian retrovirus, STLV simian T-lymphotrophic virus, WDSV: walleye dermal sarcoma virus, WEHV: walleye epidermal hyperplasia virus ([69] modified).

2.1.1 Molecular biology of foamy virus

2.1.1.1 Virion structure Similar to other members of the family FV virions feature a spherical to pleomorphic form, an envelope and a size of 80-100 nm (Figure 2, Figure 3). The viral genome (dsDNA or ssRNA) is surrounded by the nucleocapsid, which again is surrounded by the capsid [127, 162]. Inside the capsid is the core housing the genome, the nucleocapsid and viral enzymes such as the protease, the integrase and the reverse transcriptase. Matrix proteins line the inner surface of the outer membrane of the virion, the envelope (Env). The envelope consists of particles from the host cell membrane and viral glycoproteins. FVs are specific for their 10-15 nm prominent spikes on the virion's surface. Similar to other *Retroviridae*, FV envelope proteins are build in trimeric complexes [85].

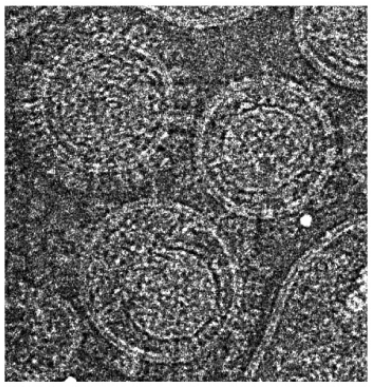

Figure 2: **Cryo-electron micrograph of foamy virus.** Image by courtesy of Steve Fuller, Thomas Wilk and Martin Löchelt.

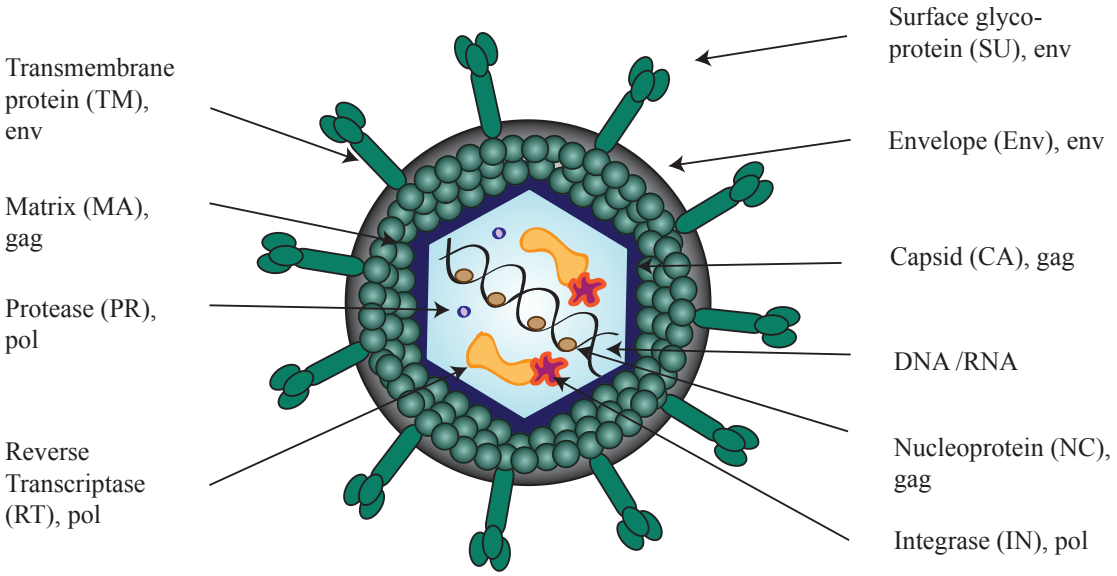

Figure 3: **Components of the foamy virus virion and its encoding region.**

2.1.1.2 Organization of the genome The FV genome composes of linear ssRNA or dsDNA with a 5'-cap and a 3' poly-A tail. With a size of 12 kb, FV genome is one of the largest among the retroviruses. The genome encodes the three structural and enzymatic genes [132], which

are common elements of the *Retroviridae*:

- The group-specific antigen (*gag)* coding core and structural proteins of the virus,
- The polymerase (*pol)* coding for reverse transcriptase, protease and integrase and
- The envelope (*env*) coding for the retroviral coat proteins.

In addition, FVs have up to three unique regulatory genes at the 3' end of the provirus expressing the transcriptional transactivator (Tas; previously called Bel1 for SFV) which is encoded from open reading frame (ORF) *tas/bel1* and Bet protein which is encoded from ORF *tas* and *bel2* (Figure 4). In the case of the prototype HFV another ORF called *bel3* could be identified, even though its function remains still unclear. The regulatory genes *rev* and *rex* of HIV and human T-lymphotrophic virus (HTLV), which are essential for the nuclear gene export, are absent, though the cis-acting element of the FV *pol* gene in the viral mRNA is considered to replace their role [26].

Within the FV genome *pol* is the most conserved region, especially the DD35E motif, which is essential for viral replication among all retroviruses [34, 95]. The conservation of the two other structural genes is unlike the *Orthoretrovirinae*: FV *env* is more conserved than *gag* [150].The accessory regulatory gene *bet* is highly divergent. However, FVs are very consistent within their FV species because of an extremely slow evolutionary rate with $1.7 \cdot 10^{-8}$ substitutions per site per year. FV is therefore the slowest-evolving RNA virus, which is known [148]. The genome keeps relatively stable, even after cross-species transmission as demonstrated for SFVs in humans [119, 142].

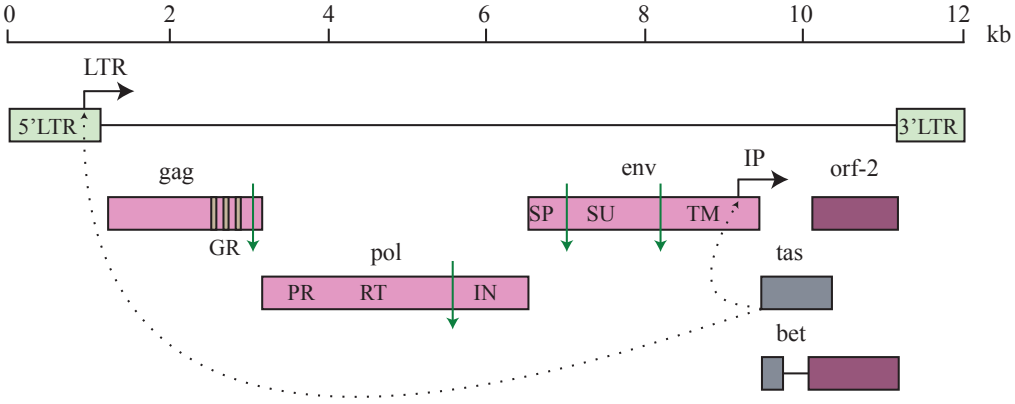

Figure 4: **Foamy virus genome with a size of 12 kb.** The foamy virus (FV) genome consists of three structural genes *gag*, *pol* and *env* and two accessory genes *tas* and *bet*, which are unique for FVs. The replication proceeds in two steps: the Internal Promoter (IP) with basal transcriptational activity expresses the transactivation protein Tas. This regulates the IP and in a later stage the "silent" promoter Long-terminal repeat (LTR), which indicates the transcription of viral RNA and structural proteins. The arrows indicate the major proteolytic cleavage within the translated polyproteins. The following domains are shown: three glycine-arginine-rich basis (GR-boxes) of the Gag protein; integrase (IN), reverse transcriptase (RT), protease (PR) of the polymerase; signal peptide (SP), surface (SU) and the transmembrane (TM) of the envelope; *tas* and *orf-2* of the Bet protein ([69] modified).

2.1.1.3 Gene expression The strategy of gene expression of FV being temporally regulated, is unusual in retroviruses, but is often the case for complex DNA viruses [26]. The FV transcription is regulated by the long-terminal repeats (LTRs) as well as the internal promoter (IP), located in the *env* gene [7]. After integration of the FV provirus into the host DNA, the high basal activity of IP induces the expression of Bet and Tas. In turn, the Tas protein transactivates the IP and therefore enhances its own production. In the late phase of an infection, when a certain level of Tas is reached, it also transactivates the "silent" LTR-promoter. LTR is responsible for the expression of viral structural and enzymatic proteins and the synthesis of the viral genomic RNA [91, 95]. The unspliced full length mRNAs will serve either as genomic RNA, integrated into virions or as a template for the translation of *gag*. The spliced mRNAs encode the genes (pro) *pol* and *env*. The splicing of the *tas* and *bel2* sequences is essential for the synthesis of the cytoplasmic Bet protein [85].

2.1.1.4 Viral proteins The FV **Gag** protein is encoded from genomic RNA, though the classical retroviral homology regions in the capsid domain and the cysteine-histidine box(es) in the nucleocapsid domain are absent [7]. Instead, FV *gag* contains three glycine-arginine-rich basis sequences (GR-boxes) in the carboxyl-terminus, binding viral nucleic acids (GR-box 1) and initiating the transport of FV to the nucleus soon after infection using the nuclear localization signal in the GR-box 2; [133, 138, 161]. The function of the GR-box 3 remains unknown. GR-boxes are involved in the packaging of the RNA genome as well as the expression, cleavage and packaging of Pol [144]. The Gag polyprotein is partially cleaved from the viral protease during assembly [32, 163]. After infection of a new host cell, these proteins are even further divided into smaller fragments, which is required for FV uncoating and subsequent integration of DNA into the genome [117]. The resulting components differ in their composition and specific function from the expected retroviral structure of matrix protein, capsid and nucleocapsid but are also involved in assembly and budding of FV [32].

Unlike other retroviruses, no Gag-Pol polyprotein is synthesized from the full-length genomic RNA and coassembled into the virion. Instead, **Pol** is independently expressed from a spliced, sub-genomic mRNA [13, 33, 161]. A unique feature of FVs is that the precursor of Pol is not completely cleaved and only the integrase is separated, which yields a protease-reverse transcriptase fusion protein or ribonuclease H protein [118]. All of these proteins are essential for effective replication of retroviruses: the protease is required for the protein cleavage, the integrase for the integration into the host genome and the reverse transcriptase for the transcription of RNA into DNA. The spliced mRNAs also encode the **Env** precursor protein, which is cleaved into the surface and the transmembrane glycoprotein by a cellular protease [82]. Similar to hepadnaviruses, envelope glycoproteins are important for FV assembly, release and infectivity [5, 29].

While the **Tas** protein is essential for the regulation of FV gene expression, the function of **Bet** is still unclear. A possible interaction with the host's immune system is actually intensively discussed [23, 25, 75]

2.1.1.5 Foamy virus life cycle Although FVs basically "act" as retroviruses during their replication, a variety of unique features has been observed for their life cycle. Probably the most important characteristic is the occurrence of either RNA or DNA in free viral particles (Figure 5 $\{1\}$). However, both forms, are able to attach to the host cell through their surface

glycoproteins. *In vitro*, all cells are susceptible for FV infections. This circumstance hampers the identification of a possible receptor on the surface of the host cell. Though, the recent study of Nasimuzzaman and Person, 2012 gives evidences that heparan sulfat proteoglycans conduce as FV receptor molecules [104].

After attaching, the FV transmembrane glycoprotein indicates the fusion with the cell membrane, which allows the cell entry through internalization. Following the release of the FV core, which includes the genome and viral enzymes, FVs migrate within eight hours to the nucleus. The FV core uses the microtubule (MT) network (Figure 5 $\{2\}$), which belongs to the cellular cytoskeleton, to traffic to the MT-organization centre (MTOC; Figure 5 $\{3\}$), that is the origin of the microtubules, placed in the center of the host cell. The process how FVs enter the nucleus is not fully understood. For FV uncoating a structural and functional alteration of the core is necessary. The essential Gag cleavage is therfore induced by the proteolytic impact of the FV protease. It is expected that the import of the FV genome is triggered by specific components of the Gag protein (GR-box 1, 2) instead of a passive transport after the dissolving of the nuclear membrane during mitosis [80]. Subsequently, the FV cDNA is randomly and covalently integrated into the host genome by the viral integrase and named "provirus" at that stage (Figure 5 $\{4\}$). In the case of FV particles with a RNA genome, the viral transcriptase is active early after infection to generate cDNA (Figure 5 $\{5\}$). Although, such early reverse transcription is common for retroviruses, it is much more complex in FVs than previously thought. There is evidence, that the actual stage of the cell cycle is critical for the process of early reverse transcription. Only replicating cells provide a high pool of nucleotides being essential for the viral cDNA synthesis. Because of the lack of nucleotides in resting cells, only FVs already containing a DNA genome can lead to productive infections [26]. However, as most of the entering viruses contain defective genomes, the number of productive infections resulting into effective viral replication is limited. Similar observations have been described for HIV-1, where maximal 10 % of all particles were infectious [116]. The replication of non-defective proviruses proceeds in the context of the cellular protein biosythesis. After the initial transcription of the proviral DNA (Figure 5 $\{6\}$), the gene expression is regulated by the two promoters IP and LTR and occurs two-folded. The transcription of the regulatory proteins Tas and Bet follows the encoding of structural proteins Gag, Pol and Env. Because these proteins are expressed as polyproteins (Figure 5 $\{7\}$), the viral protease is required for subsequent maturation to reach their functionality [26, 85]. Together with the viral RNA genome (Figure 5 $\{8\}$) they serve

as components of the novel virus particles. During the viral assembly, a step of late reverse transcription can occur (Figure 5 {9}). Therefore, about 20 % of the released virions contain a DNA genome, which is a remarkable feature of FVs [162]. The novel cDNA might also serve as a "proviral pool" to reintegrate into the host genome. Such intracellular retrotranspositions are known for *Hepadnaviridae* (Figure 5 {10}, [26]).

The process of encapsidation of Pol, Gag and the viral genome remains still unclear. The classical myristoylation of Gag as part of the retroviral Gag-Pol polyprotein is lacking for FV. It is suggested that the interaction between Pol and Gag or Pol and the viral genome (cis-acting RNA sequence) is required for Pol incorporation. The interaction of all three components, that form a ternary complex, seems to be essential for the complete encapsidation [26]. During FV assembly cellular glycoproteins are attached to build the viral envelope [85]. Simultaneously, the virion exits the cell. This process, called budding, occurs similar to *Hepadnaviridae* at intracellular membranes such as the endoplasmic reticulum (Figure 5 {11}) and only exceptionally at the extracellular membrane, which, however, is "usual" for *Retroviridae* (Figure 5 {12}, [26]). The process of budding varies between FV species clades which might result from a particular adaption to the host species and its immune system [68]. In contrast to the *Orthoretrovirinae*, maturation of the virions is even ongoing in the early infection phase of a new host cell [26]. Beside the extracellular spread of virions, FV can directly transfer to neighboring cells (Figure 5 {13}, [26]), even though this route seems to be cell specific [50].

Overall, the FV structure and life cycle clearly supports the classification in its own subfamily of the *Retroviridae*. Although, general elements are shared with those of *Orthoretrovirinae*, FVs are specific and in some points even more similar to the DNA virus family of *Hepadnaviridae* [85].

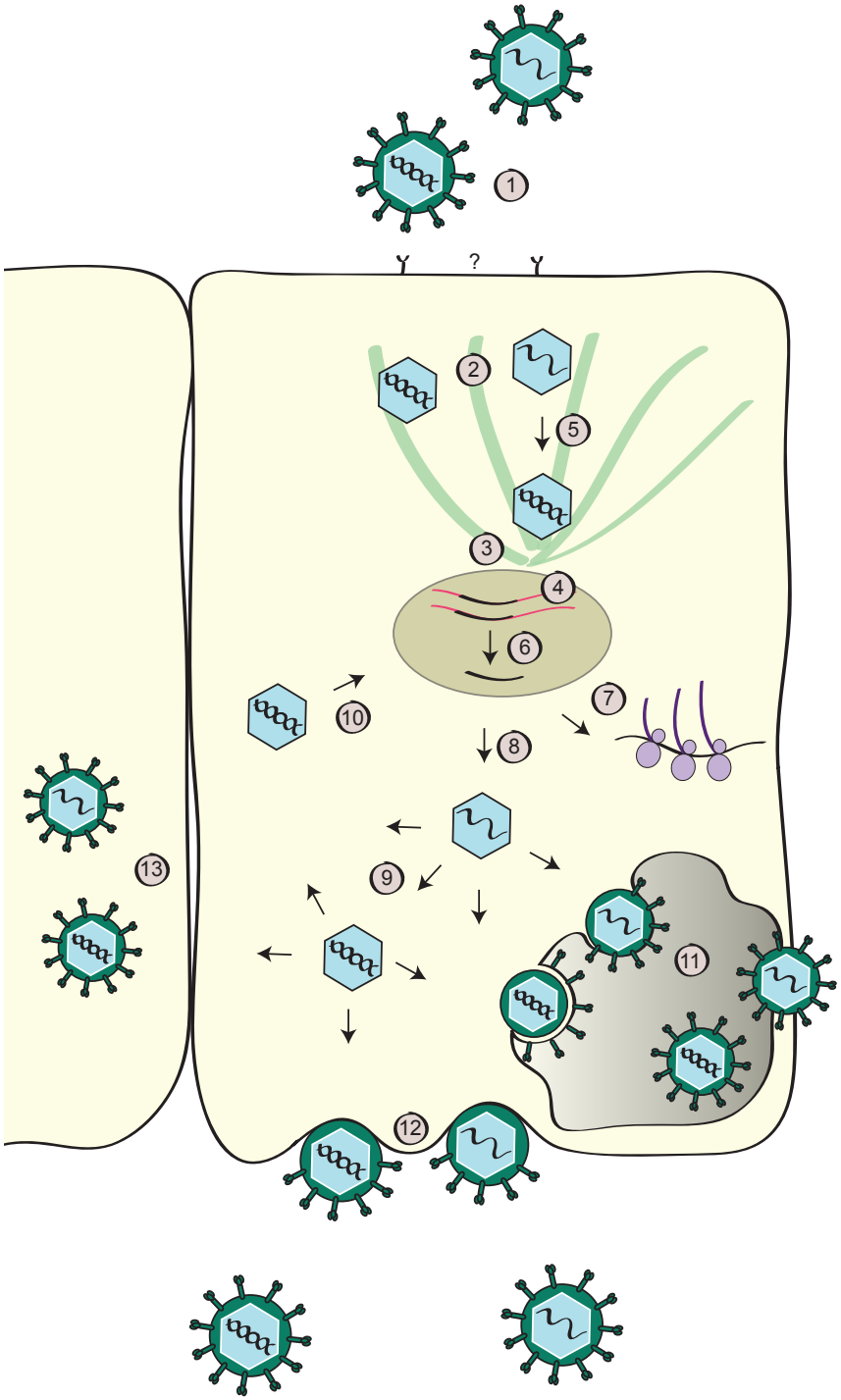

Figure 5: **Overview of the foamy virus (FV) life cycle.** FVs containing either RNA or DNA genome can enter the host cell. A receptor is so far unknown {1}. After uncoating, the FV core migrates among microtubule network (MT) {2} towards the MT-organization centre (MTOC) {3}. FV entry into the nucleus and FV DNA integrates into host's genome as provirus {4}. Early reverse transcription of FV RNA genome to enable the integration into the host's DNA {5}. Transcription of proviral DNA {6} and expression of polyproteins {7}. Messenger RNA serves directly as genomic RNA {8} or will be transcribed into DNA during late reverse transcription {9}. Novel cDNA might also serve as "proviral pool" to re-integrate into the host's genome {10}. FV assembly and budding occur predominatly at intracellular membranes {11} and only exceptionally at extracellular membranes {12}. In addition to extracellular spread, FV can directly infect neighboring cells {13}. This figure uses underlying structures from [26].

2.1.2 Foamy virus persistence

FVs are the slowest evolving retroviruses demonstrating a strong genetic stability during lifelong persistence [22, 142, 148]. Thus, FVs have certainly developed strategies to escape the host's immunity and infect vertebrates persistently [123].

Several studies on SFV and FFV provide evidences that the accessory protein Bet is involved in the inhibition of cellular APOBEC proteins [23, 92, 101, 115, 130]. These proteins are important components of the intrinsic immunity against several viruses (*e.g.,* HIV, Hepatitis B or endogenous retroviruses) and deaminate the viral cytidin to uridine, which indicates frequent proviral hypermutations (guanin-to-adenin in the plus strand) [25]. Infected cells highly express Bet proteins, which can be detected in both the cytoplasm and the nucleus [75]. Bet proteins even get secreted from infected cells and are able to enter neighboring naïve cells, whereby they target the nucleus [75]. While HIV inhibits the human APOBEC protein by the virion infectivity factor through degradation, primate FV's Bet protein inhibits APOBEC through the suppression of its dimerization [115]. Though, the influence of Bet proteins is still controversially discussed [25, 123]. For instance a study detecting similar sensitivities of wild type and *bet*-deleted FV questioned the effect of Bet on host immunity [25]. Therefore, alternative mechanisms, how FV escape the host's immunity, have been discussed. The detection of low viral loads *in vivo* provide evidences for low levels of viral replication intra-host, which might be an important requirement for virus persistence [36, 128].

2.1.3 Foamy virus prevalence

In 1971, FV was discovered for the first time in a human and henceforth described as the prototype of the human foamy virus (HFV) [1]. Further insights revealed that no human specific FV exists. Instead, humans become zoonotically infected with SFVs from Old World monkeys and great apes [51, 53]. The infection rate of SFVs ranges from 1-6 % in humans occupationally exposed to NHPs, *e.g.,* in research institutes, zoos or primate centers [17, 51, 135, 140, 147]. Populations in Sub-Saharan Africa are frequently exposed to NHPs because of hunting or the consumption of bushmeat. This resulted in a general SFV prevalence of 0.2-1 % in Africa [9, 69] and 2.6 % in Asia, where NHPs, especially macaques, are kept as pets or 'temple monkeys' [31, 59, 60, 137]. People, reporting bites or scratches from NHPs, get frequently infected with SFVs even up to a prevalence of 20 % [9].

Furthermore, FV have been described for a variety of vertebrates such as cattle, horses, cats, primates, sea lions, goats, sheep, hamsters [132], recently shrews [56] and an insectivorous bat *Rhinolophus affinis* from China [159]. However, most studies were performed on foamy viruses infecting feline (FFV), bovine (BFV) and primate (SFV) species [37, 132].

BFV, initially called bovine syncytial virus or bovine spumavirus, is present in 30-45 % of cattle [132]. FFV (initially called FSFV or FeSFV for feline syncytium-forming virus) is widespread among wild and domestic cats with 31-70 % seropositivity depending on age and geographic origin [154, 155, 156]. In non-human primates SFVs (simian foamy viruses) are highly frequent (44-100 %) in wild populations, which was shown for chimpanzees, colobus monkeys and Asian macaques [41, 59, 61, 79, 89, 97]. The infection rates are even higher when captive NHPs were investigated (80-100 %), probably because of close animal contact during housing [12, 17, 18, 22, 98, 69].

2.1.4 Evolution of foamy virus

Although parasite-host co-speciation is notably rare in RNA viruses, this mechanism applies to the FV evolution much more than previously thought [57, 148]. It could be demonstrated that SFVs have developed over the last 30 Mya in close association with their primate host. This pattern was substantiated by testing 44 species of old world primates: African and Asian monkeys as well as great apes [148].

Extreme co-speciation of virus and host can even result into the integration of virus elements into the host genome. The development of endogenous forms is a particular feature of retroviruses. The invasion of the virus into the host's germline provides a "fossil record" of the evolutionary history [54, 112]. The identification of the endogenous foamy virus in the lemur species aye-aye (*Daubentonia madagascariensis*) gives evidences for the extension of the co-speciation pattern to the *Strepsirrhini* lineage in addition to the *Haplorrhini* clade, which includes amongst others old and new world monkeys and apes [49]. The "new" foamy virus found in the strepsirrhine primate aye-aye is divergent from all known SFVs and evidence the presence of SFVs in primates for more than 85 Mya, before primates split into the haplorrhine and strepsirrhine clade [49].

Recently, the sequence analysis of FFVs from domestic cat species and FV isolated from free-living pumas (*Puma concolor*; FFVpc) gave evidences for the process of co-speciation

in non-primate FVs. Though, the evolutionary distance within FFVs is with 4 % in the highly conserved *pol* gene very low compared up to 36 % within the Old World monkeys SFVs. Therefore, a recent cross-species transmission within the family of *Felidae* cannot be ruled out. Much stronger is the evidence for co-speciation of FVs in the clade of *Bovidae/Equidae* because of the high evolutionary distance between BFV and EFV (35 %), which corresponds to the early split of the hosts from their common ancestor about 50 Mya ago [68].

The discovery of endogenous FV in sloths (*Choloepus hoffmanni*, SloEFV) provides evidences for the circulation of FV in ancestral mammals. Sloths [65] belong to Xenarthrans, a basal group of mammals, and diverged from other mammals more than 100 Mya ago [10]. Recently, endogenous foamy virus-like elements were discovered even outside of mammalian species, in particular in the genome of the West Indian Ocean coelacanth (*Latimeria chalumnae*, CoeEFV). While the sequence distance of the highly conserved *pol* gene was 40 % for SloEFV to other FVs, the distance of CoeEFV revealed even 51 % [68]. The coelacanth is a living fossil from Devonian times and belongs to lobe-finned fishes “that branched off near the root of all tetrapods” [48, 143]. Phylogenetic analyses demonstrated that the theory of co-speciation can therefore be extended to vertebrate hosts for at least 407 Mya (Figure 6).

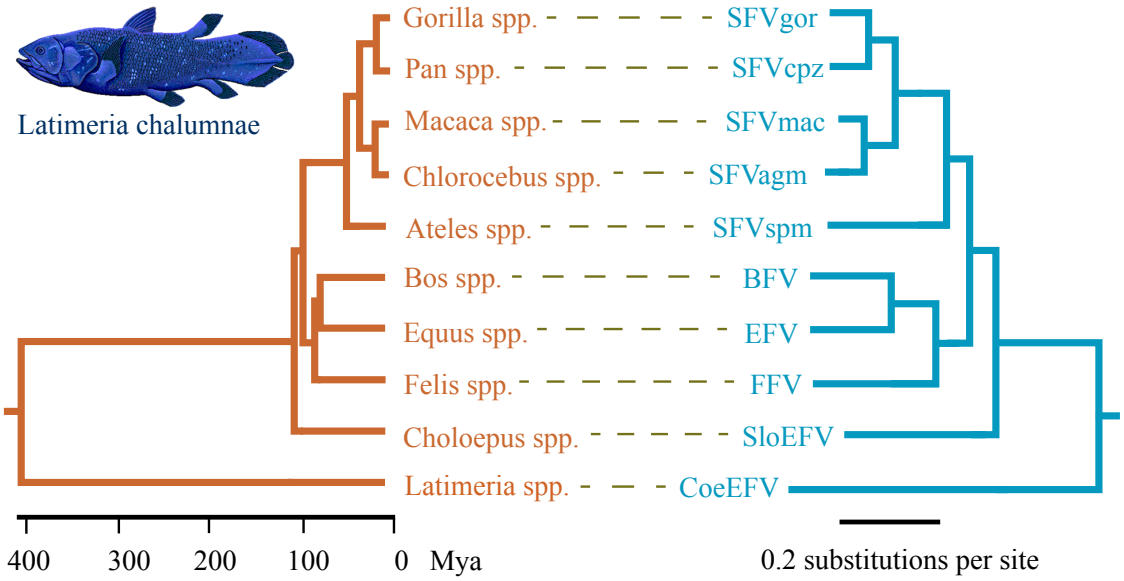

Figure 6: **Phylogenetic tree of foamy viruses and their hosts.** The phylogenetic trees demonstrate the congruence (dotted lines) between virus (right) and host (left) supporting a pattern of co-speciation for more than 400 Million of years (Mya) ago. The foamy virus phylogeny is the 50 % majority-rule consensus tree inferred from conserved region of the foamy virus and Class III retrovirus Pol protein alignment using Bayesian analysis. Host branch length (in Mya) are plotted versus virus branch length (in expected amino acid substitutions per site) for every branch. SFVgor, simian foamy virus gorilla; SFVcpz, simian foamy virus chimpanzee; SFVmac, simian foamy virus macaque; SFVagm, simian foamy virus african green monkey; SFVspm, simian foamy virus spider monkey; BFV, bovine foamy virus; EFV, equine foamy virus; FFV, feline foamy virus; SloEFV, sloth endogenous foamy-like virus; CoeEFV, coelacanth endogenous foamy-like virus ([48] modified).

This analysis postdates FV existence to at least the Early Devonian, making it to the oldest known virus. In this respect, the evolutionary conservation of foamy virus key features was of particular interest. A comparison of full genome sequences of all known FV clades (FVs of Figure 6 and bat FV: RaFV-1) demonstrated that FV specific motifs are present for at least 400 Mya. This includes the highly conserved motifs in the internal promotor (IP, located at the 3' end of *env*) and in the p3 cleavage site of Gag. These conservations support their indispensable function in the process of gene expression, infectivity (IP) and particle assembly (p3 cleavage motif). Interestingly, other motifs (Gag-Env interaction locus in the N-terminus of Env, WxxW motif for virus assembly and budding or the furin cleavage motif in Env for budding of virions) seem to be evolutionary younger because of their exclusive presence in mammal FVs and their lack in CoeEFV [68]. The discovery from Han & Worobey provides evidences that FV and therefore also retroviruses have a marine origin and underwent "remarkable evolutionary transition from water to land simultaneously with their vertebrate hosts" [48]. The recent identification of foamy virus-like sequences in platyfish (*Xiphophorus*), cod species and zebrafish

reinforced the marine origin of FVs. The discovery of "almost intact copies" of FV in the host genome provides evidence that exogenous foamy viruses have been introduced recently or even might circulate in the fish lineage [123, 136].

The long-term history of virus and host since the early Devonian might be the reason why FV infections are non-pathogenic, persistent and stable within their host [48, 57, 84, 96, 102, 148]. Further investigations in vertebrates such as reptiles, birds or amphibians but also non-vertebrates should help to understand FVs and also the retroviral evolution [48].

2.1.5 Modalities of foamy virus transmission in the natural host

Although FVs seem to be endemic in many host species, modalities of transmission are so far not fully understood. Only a handful of studies addressed this question.

In cats, the increase of seroprevalence with age, observed in domestic as well as in wild populations, point to horizontal transmissions between individuals [114, 156]. Frequent infections of feral female cats and the isolation of FFVs from saliva provide evidences for a non-aggressive transmission mode of FFVs through licking of social partners [113, 156, 160]. The detection of a vertical transmission event from queen to kitten remain an exceptional case [47].

In cows, modalities of BFV transmissions have been controversially discussed. The study of Johnson et al. hypothesized the exchange of saliva mainly through social licking as the predominant route for BFVs transmission based on infection trials [58]. Data on (sero)prevalence provided evidences for the protective effect of maternal antibodies for the first half year of life. Afterwards, BFVs spread horizontally within the herd demonstrated by a seroprevalence of 85 % in 2-3 years old cattle [58]. In contrast, other studies suggested vertical infections from dam to calve. BVFs were isolated from milk indicating transmission to the offspring during lactation even when maternal antibodies are present [93, 126]. Differences in livestock husbandry support nursing as influencing factor on BFV spread. Early separation of calves from their mothers – as is the practice in German dairy farming – highly decreased the number of BFV infections [126].

To my knowledge, nothing is known about FV transmission modalities in *Equidae*.

Although, primate FVs (SFVs) observed more attention than their relative foamy viruses, the routes of their transmission are also relatively unclear. Increases in seroprevalence and/or SFV nucleic acid detection rate with age have been documented in captive or semi-captive colonies of macaques (*Macaca tonkeana* [22]), baboons (*Papio* sp. [12, 18]), and mandrills (*Mandrillus*

sphinx [98]). For wild primates comparable data were produced only for chimpanzees (*Pan troglodytes schweinfurthii*) that showed a similar trend [89]. This was generally taken as indication of horizontal transmission being the most favored route, *e.g.*, through aggressive contacts during adulthood [22]. However, vertical transmission was only rarely directly investigated. Anecdotal reports were made concerning chimpanzees, with one captive [147] and one wild [89] mother-offspring pair being found to harbor indistinguishable SFV strains. Calattini et al. addressed the question more specifically, focusing on a captive colony of macaques [22]. Showing that 8 of 11 mother-offspring pairs harbored different SFV strains, these authors argued that vertical transmission was likely not a major route of infection [22]. Instead, seroconversion occurred contemporaneously to the onset of aggressive behavior in subadults and young adults. The seroconversion and the detection of an identical SFV strain of macaques, with a history of record frequent conflicts, evidenced severe biting as transmission route for SFVs.

Finally, these data provide evidences for horizontal transmission of FV through social interactions such as licking or biting. Vertical transmissions from mother to offspring seem to be an exception, even though it has been reported for FFV, BFV and SFV. To better understand the transmission routes of foamy viruses further insights are needed.

2.1.6 Cross-species transmission of foamy virus

Despite the fact that SFVs are specific for their host-species and even subspecies, several inter-species transmission events have been detected. Cross-species transmissions of SFV were documented in some hunter-prey systems in Sub-Saharan Africa. SFV inter-species transmissions were reported for two wild chimpanzee males (SFVcpz) from the Taï National Park, who were infected with SFV from red colobus monkeys (SFVwrc) [78]. These chimpanzees were identified as frequent hunters and therefore likely exposed to red colobus monkeys. Wounds, which were received during hunting or during chewing prey's bones [14], could provide an entry for the virus. The risk of cross-species transmission for females and infants is expected to be much lower since they hunt less and consume less meat than males. In contrast to the frequent transmission of species-specific SFV, cross-species transmission events seem to occur only rarely, also demonstrated by a broad screening of wild chimpanzees among Sub-Saharan Africa [89]. This might indicate that distinct super-infecting strains are not transmittable in the new species and result into a dead-end infections [78].

However, recently a high prevalence of FV was reported in wild tree shrews (*Tupaia be-*

langeri). The close relationship of their FV to SFVs of *Macaca mulatta* (99.3 %) was taken as an evidence for cross-species transmission, though transmission routes remain unclear [56]. Limited data and the low evolutionary distance of virus and host as observed for, e.g., FFV in Felidae or BFV in ungulate species so far prevented the identification of non-ambiguous inter-species transmission events in non-simian FVs [68].

Most findings of inter-species transmission have been reported for humans and their numerous primfate prey, more particularly great apes (*e.g.*, [9, 21, 157]). The infected persons were mainly men, who received bites from monkeys or great apes during hunting. Animal caretakers or veterinarians being occupationally in close contact with potentially infected animals, tissues, blood or body fluids also reported SFV infections [69].

Although humans are lifelong infected, no evidence for human-to-human transmission of SFVs exists [9, 69]. The observation of low viral load in blood and saliva and the lack of SFV viral gene expression were traced back to viral latency [129], which is in line with the dead-end infection of SFV in humans. Though, SFV transmission through blood transfusion is of particular concern, since this route could be demonstrated for a monkey model [17, 69]. So far, SFVs in NHP as well as in humans seem to be non-pathogenic, even though the turn into a pathogenic virus is of major health concern [69]. In the Northern hemisphere, the contact to non-primate animals carrying foamy virus (*e.g.,* BFV, FFV and EFV) is much higher than to NHPs. Direct contact to pets and livestock but also the consumption of animal products increase the risk for zoonotic infection with FV. So far, studies are very limited and do not provide any evidence for zoonotic transmission of non-primate FVs such as FFV, BFV or EFV [7]. Virus contact to BFV might be indicated by the recent finding of 7 % sero-reactivity in veterinarians and cattle farmers. Though, no DNA was detected in the samples to prove zoonotic transmission [68]. Beside the quantity and quality of exposure, the virus makeup by itself plays an important role for the risk of transmission across host barriers. Viral "structures determined by the host for instance the glycosylation pattern and host-derived, virion-associated proteins, are directly identified by the mechanisms of innate immunity or by passively adapted cross-reacting immune mechanisms" [7]. Therefore, the potential of cross-species transmission rises with the phylogenetic relation of natural and novel hosts [7]. This contributes that the transmission from NHPs to humans is more likely than from distantly related species like cats, horses or cattle [132].

2.1.7 Foamy virus super-infections – definition and diagnostics

Given the high prevalence rates of FV infections, in particular SFVs, it seems reasonable to predict that super-infections will occur frequently. Super-infections are here defined as simultaneous infection of the same individual host with several strains of the same virus. Although intra-individual diversity is a key parameter of infectious dynamics, no studies on FV actually provide these data. The only efforts were undertaken to clarify potential cases of SFV recombinants in wild chimpanzees [89]. Just recently, super-infections have been also found in humans. Four women from Bangladesh were dually infected with different SFV strains of macaques from various geographic origins [31].

Where it exists, intra-individual diversity can be determined using bulk-PCR and sequencing. In that case, PCR products can be expected to contain a mixture of sequences reflecting multiple targets in the underlying viral population. Even though the direct sequencing or heteroduplex mobility assays of bulk-PCR products enables a detection of heterogeneous retroviral populations of the same species (through multiple peaks in chromatograms or different mobility; [44, 110]) a truly complete depiction of bulk-PCR product contents will require using some kind of dissection method. For the latter purpose, cloning and subsequent sequencing or next generation sequencing can be performed. In the case of low intra-individual diversity, which is expected for SFVs, next generation sequencing results into an enormous dataset of highly redundant sequences. However, both methods are known to be prone to *Taq*-induced errors while cloning can result in selective biases since some sequences will be more likely to be cloned than others [134].

This recently prompted the rise of end-point dilution PCR (**EPD–PCR**), which is independent of preliminary bulk-PCR amplification. In theory, end-point dilution of the starting material to one template per positive PCR reaction prevent *Taq*-induced errors. Therefore, EPD-PCR provide the gold standard for the depiction of intra-individual diversity [111, 134]. Within this framework, EPD-PCR was successfully applied to the investigation of HIV-1 micro-evolutionary trends (*e.g.,* [131]), including the detection of super-infection events (*e.g.,* [81]). EPD-PCR might therefore be considered a promising tool for the detection of super-infection cases among wild primate hosts. There is however at least one predictable barrier to its implementation in this context. As retrovirus-oriented studies of wild primate populations essentially rely on the analysis of non-invasively collected samples (most often feces, *e.g.*, [67, 89]), it is conceivable that samples will only contain retroviral/proviral genomes in minute amounts

[83]. Acquiring a substantial sample of EPD-PCR sequences might therefore require prohibitively large amounts of, by definition, rare starting material [77]. Thus, applying initially biased methods (*e.g.,* bulk-PCR) might sometimes be preferable as long as the tools exist to extract biological signal from biased datasets. In this study the costs and benefits of EPD-PCR and multiple bulk-PCR cloning were assessed within the frame work of a case study focusing on simian foamy virus super-infection in wild chimpanzees (*Pan troglodytes*).

2.2 Chimpanzees

2.2.1 Overview about chimpanzees

Chimpanzees (genus *Pan)* belong to the great apes and are evolutionary the closest living relatives to humans, sharing a common ancestor about seven to eight Mya ago [73]. The common chimpanzee is classified into four subspecies: the West African subspecies *Pan troglodytes (P. t.) verus,* the Nigeria-Cameroon Chimpanzee *P. t. ellioti* (primary named *P. t. vellerosus*), the Central Chimpanzee *P. t. troglodytes* and the East African subspecies *P. t. schweinfurthii* (Figure 7) [108, 109]. Chimpanzee populations are distributed among Equatorial Africa between 13° North and 7° South and have a total population size of 173,000 to 300,000 [20].

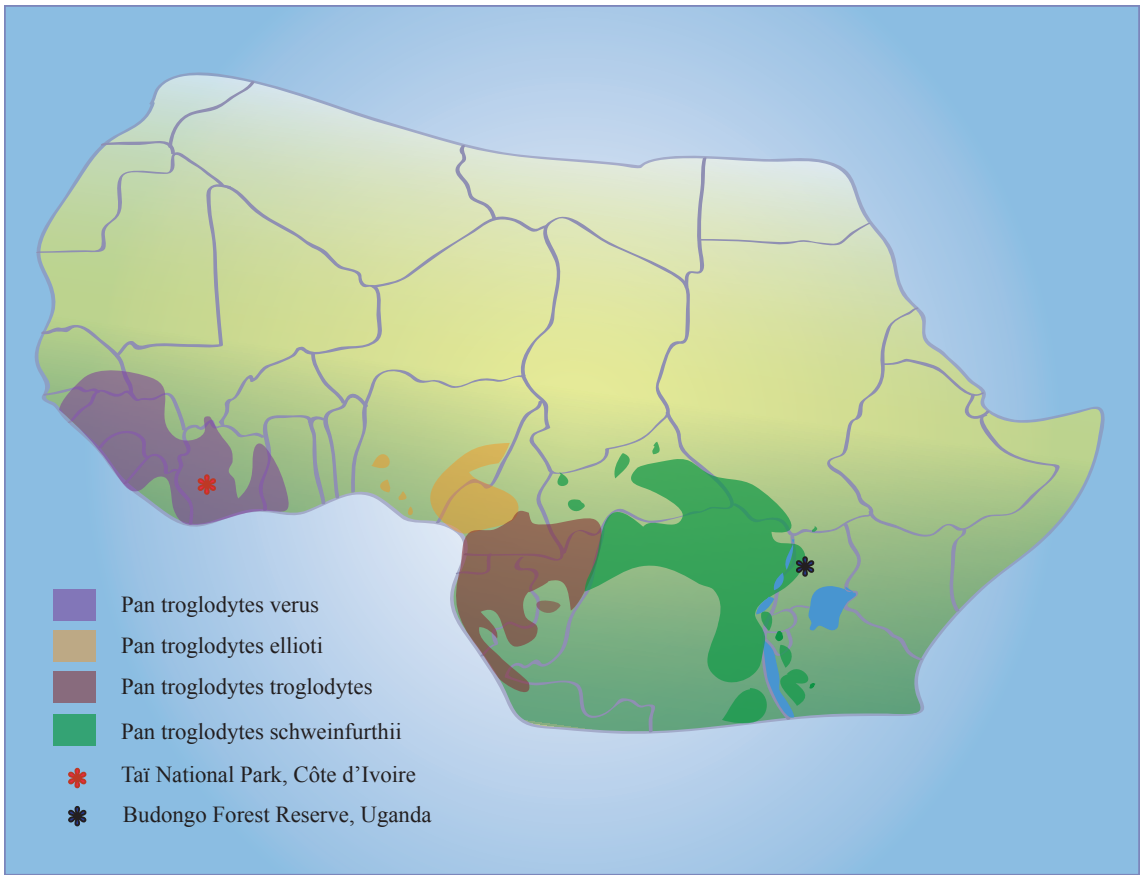

Figure 7: **Excerpt of Saharan and Sub-Saharan Africa highlighting chimpanzee distribution and study field sites.** This Figure uses underlying structures from http://www.greenpassage.org/chimp/indexE.html.

The habitat of chimpanzees consists of moist and dry forests as well as savanna woodlands. They live in large gender mixed groups comprising up to 150 individuals, called communities, with a complex social structure [109]. They live in so called fission-fusions where individuals can be absent from the group for a period of time [14]. Although chimpanzees are under protec-

tion in most countries and several conservation projects are ongoing, the communities suffered a dramatic decline during the last 20 years mainly due to poaching, habitat loss and diseases. Therefore, chimpanzees are ranked as "endangered species" from IUCN (International Union for Conservation of Nature) Red list of Threatened Species [108].

2.2.2 Field sites

2.2.2.1 Budongo chimpanzees The study site is located in the west of Uganda, within the Budongo Forest Reserve (1°37'-2°03'N, 31°22'-31°24'E), an East African semi-deciduous tropical forest of 435 km^2 size and home for 600-700 chimpanzees *P. t. schweinfurthii* (Figure 7) [125]. In 1960, the chimpanzees were originally studied by Professors Reynolds, Sugiyama and Suzuki (http://culture.st-and.ac.uk/bcfs/outreach/chimps.html, [124]). Long-term data are available for the Sonso group, which has been continuously monitored since 1990. In 2008, the Budongo Forest exhibited even four habituated groups comprising approximately 75 individuals in Sonso, ~80 in Busingiro, ~80 in Kaniyo-Pabidi and 15 in Kasokwa. All groups are exposed to different levels on human pressure such as logging, encroachment, ecotourism or neighboring sugarcane fields. Ongoing projects compare these field sites to evaluate the changes in the ecosystem on chimpanzee behavior and health (http://www.wildcru.org/research/research-detail/?project_id=28; [164]).

2.2.2.2 Taï chimpanzees The study site is located in the south west of Côte d'Ivoire, within the Taï National Park (5°15'-6°07'N, 7°25'-7°54'W), which is the largest forest in West Africa (3300 km^2) (Figure 7). In 1976, Christophe Boesch and his team began to habituate the first group of chimpanzees and investigated behavior and genetics of the chimpanzees [14]. To date, there are four habituated groups (South, North, Middle and East) with partly overlapping territories of up to 25 km^2 [52]. In 2010, the communities included approximately 102 individuals (North: 17, South: 38, East: approximately 47, Middle: 2) [70]. During the total observation time the size of the population decreased dramatically, although the long-term presence of researchers had a protective effect in the given habitat against forest destruction and poaching [71].

Sample collection and health monitoring Due to the number of unexplained deaths, the Taï Chimpanzee Health Project was established in cooperation with the Robert Koch-Institute in 2001. Under the permission of minimal human influence, including strict hygienic rules and

minimum distance to the chimpanzees, continuous monitoring for any clinical symptoms and sample collection were undertaken [77]. In the case of wild animals, non-invasive samples are of particular importance. Capturing and anesthetizing of wild great apes to collect blood samples is ethically questionable and encompass multiple risk factors for instances falling from the tree, over-dosage or other negative impacts on health through anesthetics [77]. Therefore, any biological material stemming from the chimpanzees like feces, urine, chimpanzee-meal-remains (*e.g.,* from their hunted prey or chewed fruits) and dead carcasses are collected and individually assigned. The study group is unique for its long-term data on genetics, behavior and sample collection. This background allowed for the identification of different pathogens as causative agents in a number of outbreaks in the Taï National Park. Frequent sudden deaths of chimpanzees could be traced back to a novel Ebola virus [74] but also a novel *Bacillus anthracis* [76]. In 2008, repeated waves of respiratory diseases affected the chimpanzees from the Taï National Park, leading to a 90 % morbidity rate and up to 20 % mortality rate. Contemporaneous worldwide circulating human paramyxoviruses (human metapneumovirus and human respiratory syncytial virus) with bacterial secondary infection (*Streptococcus pneumoniae* and *Pasteurella multocida*) were identified as causative agent and represent the first evidence for the introduction of human pathogens into an immunologically "naïve" chimpanzee population [71]. The close genetic relatedness between humans and great apes might facilitate cross-species disease transmission when humans are present in the natural habitat.

While acute disease contribution to population decline is very obvious, chronic infections are much more hidden and can lead to long-term health impacts as it was demonstrated for SIV in chimpanzees [35, 66]. Pathogens, such as retroviruses [62], hepatitis B virus (own studies, unpublished) or *Plasmodia spp.* [63], known to cause chronic diseases in humans, have been detected in Taï chimpanzees. These data demonstrate that numerous pathogens infect the Taï chimpanzees. As their impact on health is often recognizable only in late stage of infections surving as survival strategy of wild animals, regular monitoring is therefore of particular importance.

2.2.3 Simian foamy virus infection in chimpanzees – Subject of this study

Previous investigations revealed that SFVs are highly endemic in the Taï National Park [78, 79, 97] and even the first cases of cross-species transmission from chimpanzee prey to chimpanzees originate from this field site [78]. While many data have accumulated about patterns of

co-divergence with their hosts and cross-species transmission events, little is known about the modalities of SFV transmission within NHP species, especially in the wild. Therefore, the aim of the study was to investigate the dynamics of SFV circulation in a social primate community. The group of Western chimpanzees in the Taï National Park was ideal for this investigation as long-term behavioral and epizootic data as well as non-invasively collected samples were available. To increase the power of the analyses, the intra-individual diversity had to be determined. The gold standard for the detection of retroviral super-infections (end-point dilution PCR, EPD-PCR) has never been tested for the use of non-invasive material. Therefore, cost and benefits of EPD-PCR and a novel tool based on bulk-PCR had to be evaluated to provide recommendations for further investigations on material with expected low amounts of the target. These analyses should subsequently allow addressing the question of the transmission modalities of SFVcpz.

3 Material and Methods

3.1 Overview Materials

3.1.1 Chemicals

Product	Manufacturer
10 x Rxn-buffer (PCR)	Invitrogen^{TM}, Germany
6 x DNA Loading Dye	Fermentas GmbH, Germany[1]
Acetic acid	Carl Roth GmbH, Germany
Carrier RNA	Qiagen, Germany
Deoxynucleoside triphosphate (dNTP)	Invitrogen^{TM}, Germany
Desoxyuridine triphosphate (dUTP)	Fermentas GmbH, Germany
Dithiothreitol (DTT)	Invitrogen^{TM}, Germany
Ethanol (> 99 %)	Carl Roth GmbH, Germany
Ethidium bromide (10 mg/ mL)	Promega GmbH, Germany
Ethylendiaminetetraacetic acid (EDTA)	Carl Roth GmbH, Germany
GeneRuler^{TM} 1 kb DNA Ladder	Fermentas GmbH, Germany
GeneRuler^{TM} 100 bp DNA Ladder	Fermentas GmbH, Germany
Magnesium chloride ($MgCl_2$)	Invitrogen^{TM}, Germany
Nuclease free water	Applied Biosystems, Germany
Peq Gold Universal Agarose	Peq Lab, Germany
Primers and probes	Invitrogen^{TM}, Germany; TIB Molbiol, Germany
Random primer (Hexamer)	Metabion, Germany
Tris (tris (hydroxymethyl) aminomethane)	Carl Roth GmbH, Germany

[1] Fermentas GmbH, St. Leon-Rot belongs since 2010 to Thermo Fisher Scientific, USA.

3.1.2 Buffer

Product	Composition
Tris-acetate-EDTA-Buffer (50 x)	242.3 g Tris 0.4 M 60.1 mL acetic acid 0.2 M 18.61 g EDTA 10 mM ad 1 L ddH_2O; pH 8.0

3.1.3 Kits

Product	Manufacturer
BigDye® Terminator v3.1 Cycle Sequencing Kit	Applied Biosystems, Germany
ExoSAP-IT® For PCR Product Clean-Up	USB Corporation, USA
GeneMATRIX Stool DNA Purification Kit	Roboklon GmbH, Germany
Platinum® *Taq* DNA- Polymerase	Invitrogen^{TM}, Germany
SuperScript^{TM} II RT	Invitrogen^{TM}, Germany
TOPO TA cloning® Kit	Invitrogen^{TM}, Germany

3.1.4 Consumables

Product	Manufacturer
ABgene Clear Seal Diamond (Sealing foil for TaqMan® plates)	Thermo Fisher Scientific, USA
ABgene PCR Plates (96-well plates)	Thermo Fisher Scientific, USA
Falcon tubes (15 mL, 50 mL)	Neolab, Germany
Micro tubes (1.5 mL, 2 mL)	Sarstedt AG, Germany
Micropipettes	Eppendorf AG, Germany
Optical tubes and caps for TaqMan®-PCR	Applied Biosystems, Germany
Parafilm	American National Can, USA
Pipette tips for micro pipettes	Nerve plus GmbH, Germany
Reaction tubes (0.2 mL, 0.5 mL)	PeqLab, Germany; Thermo Fisher Scientific, USA; Carl Roth GmbH, Germany

3.1.5 Technical equipment

Product	Manufacturer
Biological Safety cabinet HeraSafe	Thermo Fisher Scientific, USA
FastPrep®-24 (high speed-homogenizer)	MP Biomedicals, USA
FlexCycler	Biozym Scientific GmbH, Germany
Gel electrophoresis chamber	Neolab, Germany
Micro-spoon spatula for extraction	Carl Roth GmbH, Germany
Microcentrifure for 1.5 mL reaction tubes	Neolab, Germany
Microwaves	SB-Großhandels GmbH, Quelle Gruppe, Germany
Nanodrop ND-1000 Spectrometer	Peq Lab, Germany
PCR Mastercycler epigradient	Eppendorf AG, Germany
Power supply gel electrophoresis	Neolab, Germany
Table-top centrifuge 5417C	Eppendorf AG, Germany
Table-top scale	Sartorius, Germany
TaqMan® Stratagene MX3005P	Agilent Technologies, USA
Thermal block incubator Thermomixer compact	Eppendorf AG, Germany
Transilluminator CN-1000	PeqLab, Germany
Vortexer (Labdancer)	Carl Roth GmbH, Germany

3.1.6 Software

Product	Inventor/ Webpage
Adobe Illustrater CS5, Photoshop CS5	Adobe Systems
ALTER (ALignment Transformation EnviRonment)	http://sing.ei.uvigo.es/ALTER/ [39]
BEAST v1.6.1, companion softwares	http://beast.bio.ed.ac.uk/
BLAST	http://blast.ncbi.nlm.nih.gov/Blast.cgi [2], NCBI, Bethesda, MD, USA
Corel Draw 12	Corel Corporation, USA
DNASTAR® Lasergene^{TM} (SeqMan®II)	DNAStar Inc., USA
FABOX (fasta sequence toolbox)	http://users-birc.au.dk/biopv/php/fabox/[153]
FigTree v1.4.0	http://tree.bio.ed.ac.uk/software/figtree/
Genenious Pro 5.4	Biomatters Ltd., New Zealand http://www.geneious.com/
JabRef 2.4.2	http://jabref.sourceforge.net/index.php
jModelTest v0.0.1, v0.1	[120]
LyX 1.6.1	http://www.lyx.org/
Mendeley	http://www.mendeley.com
Microsoft Office	Microsoft, USA
Network v4.610	http://www.fluxus-engineering.com/
PhyML 3.0	[45]
R package car	[38]
R package lme4	[8]
R v2.10.1	Development Core Team, 2010
RDP3 v3.42, v3.44	[94]
SeaView v4	[43]
Stratagene MxPro 4.1	Stratagene, USA
TCS	[24]
TREE-PUZZLE	http://www.tree-puzzle.de [139]

3.2 Sample description

The study was performed on non-invasive samples stemming from habituated chimpanzees. To establish a method, which detects retroviral super-infections, the first part of the study was conducted on a test dataset (**dataset A**), including individuals from two different chimpanzee communities of western and eastern Africa: *P. t. verus* (South group) from Taï National Park, Côte d'Ivoire and *P. t. schweinfurthii* (Sonso group) from Budongo Forest Reserve, Uganda. The second part of the study, investigating the biology of SFV within a chimpanzee community, refers to individuals from Taï National Park, Côte d'Ivoire only (**dataset B**).

In both field sites, chimpanzees have been continuously monitored for more than 20 years (Budongo-Sonso group: http://culture.st-and.ac.uk/bcfs/outreach/chimps.html, Taï -South group: [15]). This resulted in an initial behavioral determination of pedigrees, which were then further defined genetically [16, 106, 152]. Fecal samples were collected immediately after defecation by trained personnel, thereby allowing for immediate individual identification of the defecating chimpanzees. All necessary permissions were obtained for the described field studies – from the Ministry of the Environment and Forests, the Ministry of Research and the directorship of the Taï National Park for the study site in Côte d'Ivoire and from the Uganda Wildlife Authority and the Uganda National Council for Science and Technology for the study site in Uganda. Fecal samples were placed on ice (Taï samples) or soaked in RNAlater (Qiagen, Hilden, Germany; Budongo samples) directly after collection, before being stored in liquid nitrogen (Taï samples) or at -20 °C (Budongo samples), as previously described [72, 77].

The test dataset consisted of 10 samples (**dataset A**), which were selected from individuals older than ten year of age at the date of collection (range: 15-42) because primary infection with SFV is assumed to occur in early adulthood [89, 97]. These samples originated from four chimpanzees living in Budongo forest (two males and two females) and from six chimpanzees living in Taï National Park (two males and four females) (Figure 7). All details about individual samples are given in Table 1.

Long-term data on Taï chimpanzees, including a sample collection over the last 10 years, enabled detailed studies on SFV biology and transmission modalities. An initial screening of 208 fecal samples, collected from 37 chimpanzees showed that 32 individuals were SFV infected. The final **dataset B** consisted of 37 fecal samples stemming from 23 individuals (Table 2). Those were selected to maximize the number of mother-offspring and father-offspring pairs included in the study, ending up with 12 mother-offspring pairs and 7 father-offspring pairs

(Table 3); altogether individuals corresponding to the same sample formed a total of 257 non-mother-offspring non-father-offspring dyads. The 12 offspring had been sampled between 3 and 19 years of age, while the range within the complete dataset was 3 to 44 years. For 11 subjects, longitudinal sampling was available (8 subjects with 2 samples, 3 subjects with 3 samples). All experimental as well as analytical steps of the study were performed at the Robert Koch-Institute.

Table 1: **Individual sample characteristics of dataset A**. Individuals whose name start with a "T" are *P. t. verus* from Taï National Park, Côte d'Ivoire, with a "B" are *P. t. schweinfurthii* from Budongo Forest Reserve, Uganda. * The infection status single respectively super-infection (infection with different viral strains) of simian foamy virus was determined according to end-point dilution PCR.

Individual	Sex	Birth date	Sampling date	Age at sampling (years)	Infection status*
B1	male	1990	2007	17	super
B2	female	1990	2007	17	single
B3	female	1983	2007	24	single
B4	male	1982	2008	26	single
T1	female	1977	2002	25	super
T2	female	1970	2005	35	super
T3	male	1989	2006	17	single
T4	female	1965	2005	40	super
T5	female	1970	2004	34	super
T6	male	1964	2006	42	super

Table 2: **Individual sample characteristics of dataset B**. Samples were collected from (*P. t. verus*), the South group in the Taï National Park, Côte d'Ivoire and have been ordered according to the name of the chimpanzee. For each individual 1-3 samples[$] were collected, which were named from sample A (earliest sampling date) up to sample C (latest sampling date). For each individual were 25 bulk-PCR clone sequences generated, except for Sumatra 38 years because of technical limitations (15 sequences). * marks samples, which have been excluded from statistical analysis of super-infection status in relation to age and sex because of their ambiguous infection status.

Individual	Sex	Birth date	Sample[$]	Sampling date	Age at sampling (years)
Caramel	male	2002	A	2005	3
Celine	female	1995	A	2005	10
Coco	female	1980	A	2004	24
Coco*	female	1980	B	2005	25
Gogol	male	1991	A	2001	10

Table 2. Cont.

Individual	Sex	Birth date	Sample[$]	Sampling date	Age at sampling (years)
Gogol	male	1991	B	2008	17
Jacobo	male	1998	A	2004	6
Julia*	female	1970	A	2004	34
Kabisha	female	1977	A	2001	24
Kabisha	female	1977	B	2002	25
Kaos	male	1977	A	2006	29
Kinshasa	female	1990	A	2001	11
Kinshasa	female	1990	B	2007	17
Kiriku	male	2005	A	2008	3
Louise	female	1980	A	2005	25
Lula*	male	2003	A	2007	4
Rebecca	female	1995	A	2002	7
Romario*	male	1999	A	2004	5
Romario	male	1999	B	2008	9
Rubra	female	1970	A	2001	31
Rubra	female	1970	B	2005	35
Rubra	female	1970	C	2008	38
Sagu	male	1989	A	2002	13
Sagu	male	1989	B	2006	17
Sagu	male	1989	C	2008	19
Settut*	female	1996	A	2003	7
Shogun	male	2001	A	2008	7
Sumatra	female	1965	A	2003	38
Sumatra	female	1965	B	2005	40
Utan	male	1994	A	2001	7
Utan	male	1994	B	2005	11
Yao	male	1995	A	2003	8
Yucca	female	1970	A	2002	32
Yucca	female	1970	B	2004	34
Zyon	male	1964	A	2001	37
Zyon	male	1964	B	2006	42
Zyon	male	1964	C	2008	44

Table 3: **Parental relationship of the Taï chimpanzees from dataset B**. Members of one matrilineal line (here mother and offspring) have the same initial letter.

Individual	Mother	Father
Caramel	Coco	Sagu
Celine	Coco	Kaos
Coco		
Gogol		
Jacobo	Julia	
Julia		
Kabisha		
Kaos		
Kinshasa	Kabisha	
Kiriku	Kinshasa	
Louise		
Lula	Louise	Sagu
Rebecca	Rubra	
Romario	Rubra	Kaos
Rubra		
Sagu	Sumatra	
Settut	Sumatra	Kaos
Shogun	Sumatra	Zyon
Sumatra		
Utan		
Yao	Yucca	Zyon
Yucca		
Zyon		

3.3 Molecular biology analyses

3.3.1 Nucleic acids' extraction and cDNA generation

DNA and RNA were co-extracted with the GeneMATRIX Stool DNA Purification Kit (Roboklon GmbH, Berlin, Germany) using 80 mg feces (Taï samples) or 100 mL homogenate (Budongo samples) and 5 µL carrier RNA (Qiagen, Hilden, Germany) to enhance RNA yield. First-strand cDNA synthesis was then performed using SuperScript™II Reverse transcriptase (Invitrogen™, Karlsruhe, Germany) with random hexamer primers (Table 4). RNA genomes were not considered as specific targets. Therefore, extracts were not treated with DNAse so as to allow for the detection of SFV DNA (either packaged in viral particles or proviral), would it occur in fecal samples. Though published results suggest that SFV DNA will be shed much less frequently in feces than SFV RNA [89], it should be considered that any sequence produced here might come from RNA or DNA genomes found in viral particles or shed infected cells or from proviral DNA.

Table 4: **Single-stranded cDNA transcription protocol** (Invitrogen™).

	Composition		Protocol	
Mix 1:	Random hexamer primer (R6), 10 µM	1.0 µL	for Mix 1:	
	dNTPs, 25 mM	0.4 µL	65 °C	5'
	RNA extract	11.6 µL	4 °C	∞
	total	13 µL		
Mix 2:	SS 5x Buffer	4 µL	for Mix 2:	
	DTT	2 µL	37 °C	55'
	SuperScript II RT	1 µL	70 °C	15'
	Mix 1	13 µL	4 °C	∞
	total	20 µL		

3.3.2 Initial PCR screening

To identify SFV positive feces samples, a nested PCR assay using a set of generic primers targeting a 470 base pair (bp) fragment of the *integrase* (*int*) gene was employed [89, 141], using 1.5 µL cDNA as template in a 15 µL reaction mixture containing Platinum® *Taq* DNA-Polymerase (Invitrogen™, Karlsruhe, Germany) (Table 5, Table 6).

Table 5: **Simian foamy virus (SFV) nested PCR assay**, *integrase* fragment.

Composition		Protocol		
Rxn Puffer 10x	1.5 µL	95 °C	5’	
dNTPs (2.5 mM)	1.2 µL	95 °C	30”	
$MgCl_2$ (50 mM)	1.2 µL	56/60 °C [2]	45”	35 x
forward primer SFVint1s	0.3 µL	72 °C	60”	
reverse primer SFVint2as	0.3 µL	72 °C	10’	
Platinum *Taq* Polymerase	1.15 µL			
template	0.6/1.5 µL[3]			
nuclease free water	ad 15 µL			

Table 6: **Primer for nested PCR**, *integrase* fragment [89, 141]. AT: annealing temperature; s: sense; as: antisense.

Primer	Sequence 5’-3’	AT (°C)
SFVint 1s	GCCACCCAAGGGAGTTATGTGG	56
SFVint 2as	GCTGCACCCTGATCAGAGTG	56
SFVint 3s	CCTGGATGCAGAGTTGGATC	60
SFVint 4as	GAAGGAGCCTTAGTGGGGTA	60

3.3.3 DNA/RNA quantification

To determine the viral load in feces, a TaqMan probe-based quantitative assay targeting another short fragment of the polymerase gene (*pol*) was applied to the **dataset A**, using 2 µL cDNA as template in reactions otherwise prepared as in Table 7, Table 8 [100]. All six samples obtained from Taï National Park were tested, while Budongo samples had to be excluded from this analysis as samples were only available as RNAlater homogenate, which prevented determining the mass of fecal matter effectively used (fecal matter mass was not determined at the time of collection).

The TaqMan was also tested on dilutions of plasmids containing the corresponding sequence and could detect < 5 molecules per reaction, including when plasmids were mixed in “fecal” SFV-negative cDNA (Table 9).

[2] first round 56 °C; nested round: 60 °C
[3] first round 1.5 μL; nested round: 0.6 μL

Table 7: **Simian foamy virus (SFV) real-time PCR assay**, *polymerase* fragment.

Composition		Protocol		
Rxn Puffer 10 x	2.5 µL	95 °C	8’	
dNTPs (2.5 mM)	2.5 µL	95 °C	60”	
$MgCl_2$ (50 mM)	2.5 µL	58 °C	60”	50 x
forward primer SFVTs (10 pmol/ µL)	0.5 µL	72 °C	15”	
reverse primer SFVTas (10 pmol/ µL)	0.5 µL			
probe	0.5 µL			
Ampli*Taq*Gold (5 U/ µL)	0.2 µL			
template	2.0 µL			
nuclease free water	ad 25 µL			

Table 8: **Primer for real-time PCR assay**, *polymerase* fragment [100]. AT: annealing temperature; s: sense; as: antisense.

Primer	Sequence 5’-3’	AT (°C)
SFVT s	CTTCAACCTTTGCTGAATG	58
SFVT as	TAATACAGGGCTATAGGTGT	58
SFVT probe	6’-FAM-TTGGAATTCAGTACTCCTTATCACCC-3’BHQ1	58

Table 9: **Sensitivity of real-time PCR assay.** Different dilution of plasmids (Simian foamy virus, SFV standard and a plasmid from SFV of chimpanzee Sumatra; with and without additional fecal DNA) were compared based on the number of cycles at which the DNA-based fluorescence exceeds the threshold (threshold cycle = Ct). Please note that all dilutions of the plasmid came up with a detection less than 40 cycles.

Copy number of plasmids	Ct for 1 µL Plasmid-SFV Standard	Ct for 1 µL Plasmid SFV Standard + 2 µL SFV-negative fecal DNA	Ct for 1 µL Plasmid Sumatra	Ct for 1 µL Plasmid Sumatra + 2 µL SFV-negative fecal DNA
1	38	38	37	36
5	36	35	39	38
10	33	33	34	35
20	33	33	33	33
50	32	31	34	33
10^2	31	30	33	33
10^3	27	27	30	30
10^4	24	24	27	26
10^5	20	20	23	23
10^6	17	17	20	20

3.3.4 End-point dilution PCR

The standard technique to determine intra-individual diversity was applied to all samples from **dataset A**. Therefore, several dilutions of cDNA were tested for each isolate so as to identify the dilutions resulting in success rates < 30 %, conditions in which about 80 % of the sequences can be predicted to stem from a single starting template molecule [134]. For each of the ten individuals, 15 EPD-PCR products were sequenced on both strands according to the Sanger's method.

3.3.5 Bulk-PCR and cloning

Bulk-PCR and cloning were considered as alternative method to determine retroviral diversity and therefore initially applied to **dataset A** and in the second step to **dataset B**. Five positive bulk-PCR products were independently obtained from each individual, using the same aforementioned nested PCR assay, which was run using 1.5 µL undiluted cDNA as template (in the case of **dataset A**, cDNA was always derived from the same fecal sample as used for EPD-PCR). All resulting PCR products were sub-cloned using the Topo TA cloning kit (Invitrogen^{TM}, Karlsruhe, Germany) according to the manufacturer's instructions. Following colony PCR and visualization on a 1.5 % agarose gel, five positive colony PCR products from each of the amplicons were purified (ExoSAP-IT®) and sequenced on both strands according to the Sanger's method, thereby generating 25 clone sequences per individual.

3.4 Sequence analyses

3.4.1 Sequence processing

Chromatograms were analyzed using SeqMan®II (DNASTAR® Lasergene^{TM} DNAStar Inc., USA). Sequences comprising mixed bases were discarded. In total, 400 sequences (150 EPD-sequences and 250 bulk-PCR clone sequences) were generated for **dataset A** and 915 sequences (bulk-PCR clone sequences) for **dataset B**. The chimpanzee origin of these sequences (more precisely *P. t. verus* for **dataset A** and **B**, *P. t. schweinfurthii* for **dataset A**) was confirmed by BLAST [2]. Sequences were aligned using the MUSCLE algorithm [30] as implemented in SeaView v4 [43]. Modification of the sequence format or haplotyping to identify unique sequence types were performed using webpage FaBox v1.4.1 [153] and ALTER [39].

3.4.2 Recombination analysis

Aligned sequences were checked for evidence of recombination using the RDP, GENECONV, MaxChi, Chimaera, SiScan, and 3Seq tools, as available in RDP3 v3.42 [94] but did not detect any recombination event.

3.4.3 Sequence distance

The observed distance, *i.e.,* the observed number of differences within sequence alignments, were determined for each individual (**dataset A**; Table 11) and the overall community (**dataset B**) using SeaView v4 [43].

3.5 Phylogenetic analyses

3.5.1 Evolutionary relationship of simian foamy virus within hosts

To test if phylogenetic analyses are a suitable tool to identify super-infections, the phylogenetic signal comprised in the dataset, was determined. Therefore, the 10 individual alignments of clone sequences (**dataset A**) were investigated by TREE-PUZZLE (Table 11, [139]). TREE-PUZZLE is a maximum likelihood (ML) software generating phylogenetic trees. The implemented quartet puzzling algorithm (*i.e.*, likelihood mapping) provides weights of topologies from quartet derived ML-trees (relationships for each set of four out of N sequences), differentiating between resolved, partly resolved and unresolved quartets. Low phylogenetic signal is indicated by a high percentage of partly or unresolved quartets, in that case ML-values are too similar to identify only one of the three possible topologies.

On individual alignments (**dataset A** and **B**) phylogenetic analyses were performed using PhyML v3.0 [45] as implemented in SeaView, approximate likelihood ratio test values served as branch support values [3]. For all datasets and analyses, the same model of nucleotide substitution was employed (global time reversible plus rate heterogeneity; GTR+G; Figure S1, Figure S3).

3.5.2 Evolutionary relationship of simian foamy virus among great apes

To specify the origin of the SFV sequences, phylogenetic analyses were conducted on the generated clone sequences (**dataset A** separated by subspecies: Figure 13, Figure 14; **dataset B**

founder sequences only: Figure 19) adding representative sequences from all chimpanzee subspecies (*P. t. verus*, *P. t. ellioti*, *P. t. troglodytes*, and *P. t. schweinfurthii*) and rooting with bonobo sequences (*Pan paniscus*) [89, 151, 148].

Model selection by Akaike's information criterion score was first conducted using jModelTest v0.1 [121], which resulted for all alignments in selecting a global time-reversible (GTR) matrix of substitution together with across-site rate variation (+G). For **dataset B** the proportion of invariant sites (+I) was also selected. Tree reconstruction was then performed under the chosen model using PhyML (for the large dataset, *i.e.,* **dataset A** on a web server [46], otherwise PhyML v3.0 [45] implemented in SeaView). Branch robustness was assessed through non-parametric bootstrapping (500 pseudo-replicates). In addition, Bayesian analyses were performed for **dataset A** using BEAST v1.6.1 [28]. Output of BEAST analyses was examined with companion software (available at http://beast.bio.ed.ac.uk/ Main_Page), which notably allowed for the computation of branch posterior probabilities, which were taken as branch robustness measures.

3.5.3 Test for suspicious triple infections

A phylogenetic tree was also generated for three supspicious triple infection cases in **dataset B** (Figure 18). Initial model selection resulted in selecting Kimura 2-parameters (K80) together with across-site rate variation (+G). Branch support was performed by 500 pseudo-replicates to estimate bootstrap values.

3.6 Statistical Analyses

3.6.1 Mismatch distribution-based identification of single/super-infections

3.6.1.1 Reasoning line The first objective was to develop a statistical tool that provides an objective criterion for classifying individuals as single or super-infected, based on the analysis of a multiple bulk-PCR product clone alignment. The shape of the frequency distribution of the number of mismatches derived from an individual alignment (thereafter termed 'mismatch distribution') served as a way to investigate the nature of the underlying infection process. The assumption was that in case of single infection, mismatch distribution should be approximately unimodal and can be fit with "unimodal" distribution laws. On the contrary, super-infections were expected to result in bi- or multimodal distributions and should therefore better fit "bi-

modal" distribution laws. This approach was initially applied to sequences generated by the gold-standard method EPD-PCR **(dataset A)**. Therefore, mismatch distributions were produced for all individual EPD-PCR sequence alignments to identify the individuals as cases of single or super-infection (Figure 11). At that stage, assignment was made by eye only as the limited number of variants precluded any statistical assessment of the shape of the distributions. In contrast, mismatch distribution derived from bulk-PCR product clone alignments of the same individuals were much more variable (Figure 11) and therefore required the development of a statistical approach.

In detail, several models (one or two Poisson or normal distributions) were fitted to the observed mismatch distribution in a maximum likelihood framework. Models assuming a unimodal distribution (one Poisson or normal distribution; single infection) or a bimodal distribution (two Poisson or normal distributions; super-infection) were then compared using Akaike's information criterion, corrected for small samples (AICc) as a measure of model performance [19]. The model with the smallest AICc was selected as most likely mismatch distribution. However, when the smallest AICc-value differed by no more than two from the second smallest, the simplest model (*i.e.*, the one with the smallest number of estimated parameters) from all models with AICc-values differing from the smallest AICc-value by at most two was chosen (in case the model with one normal distribution and the model with two Poisson distributions both fulfilled this criterion, the one with the smaller AICc was chosen). The resulting diagnosis (single or super-infection) was confirmed by visual inspection of mismatch distributions and networks. It should be noted here that this method is by nature independent of the definition of a cutoff number of mismatches that would make two sequences unlikely to be derived from each other: it relies only on shape analysis. The shape analysis essentially allows for the identification of deviations from a single infection scenario and not for the determination of the number of SFV strains involved in multiple infections with > 2 SFV strains (Figure 8), but note that multiple infections with > 2 strains are properly identified as case of super-infection (tested by the simulation of triple infections).

The new technique was then applied to the bulk-PCR clone sequence data stemming from the Taï community (**dataset B**).

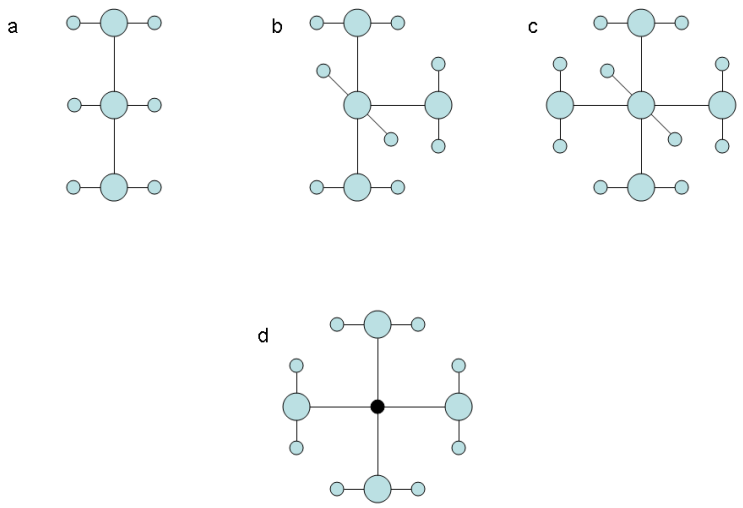

Figure 8: **Scheme of identifying super-infections by sequence analysis.** Hypothetical super-infections (a-d) illustrating that multiple modes (three in a-c; two in d) in mismatch distributions point to super-infection, but not to the number of underlying infections.
In example a, there is an obvious triple infection, which will result in building three modes in a mismatch distribution: one for local diversity around one founder sequence (large circles and their little satellites), one for comparisons involving neighboring clouds of sequences (*e.g.*, upper with middle cloud) and a last one for comparisons involving not-neighboring clouds (the upper and the lower).
Examples b and c illustrates further infections that are, by chance, implicating viruses, which are always more closely related to the initial central cloud of sequences and approximately as distant to it as the preceding infecting viruses. Adding them will not change anything to the overall shape of the mismatch distribution, which will still harbor the exact same three modes. Here, the same three mode distribution therefore describes a triple, quadruple or quintuple infection.
Example d further extends this reasoning line to a case of double mode distribution: it is enough to consider a case very close to example c but where the central cloud is not involved in the super-infection. There, only the first two abovementioned modes will appear, whatever the number of equidistant founder sequences were involved in super-infection.

3.6.1.2 Power analyses

3.6.1.2.1 Influence of the number of bulk-PCR products To determine the number of bulk-PCR products needed for an accurate determination of the infection status, analyses were performed for the use of one to five bulk-PCR products and compared to results using EPD-PCR. Therefore, mismatch distributions were produced for all ten individuals (**dataset A**) to any possible combination of bulk-PCR product clone alignments; *i.e.*, to the only possible combination of all five bulk-PCR product clone alignments (ABCDE), to the five possible combinations of four bulk-PCR product clone alignments (ABCD, ABCE, ABDE, ACDE and BCDE), and so on (Figure 15). In a few cases only two different numbers of mismatches were

found in a sample. In these cases, uni-/bimodality were assessed based by visual inspection of the distribution because it was impossible to fit the two means model to such a distribution.

3.6.1.2.2 Simulation of triple infections To demonstrate that infections more than two SFV strains will be identified as super-infection and their mismatch distributions better fit to a multimodal mode than an uni-modal mode, triple infection cases were simulated. For the sake of biological relevance, these simulated datasets were generated from the "real" data, randomly re-shuffling sequences of unambiguous "clouds" of sequences (supported by both phylogenetic and network analyses), forming 84 "triple infection clone alignments" with 24-26 sequences (from Taï specimens of **dataset A**). Most stand for triple infection with relatively distantly related "clouds" (n = 78) while in six cases two of the three "clouds" are really closely related (*e.g.*, some sequences belonging to these "clouds" only differ at one position).

3.6.2 Influencing factors on super-infection status

To test for influencing factors on the kind of infection (single or super-infection) of chimpanzees in natural settings, a generalized linear mixed model (GLMM) [4] with binominal error structure and logit link function was used. In particular, the influence of the combined effect of sex and age (categorical and continuous predictors) and their interaction (fixed effects) were analyzed for the Taï community (**dataset B**). Subject identity was included as a random effect because for several subjects samples from different ages were used in this study. The total number of data points analyzed was reduced to 32 samples from 20 subjects (Table 2, Figure S2), as the infection status of five samples was ambiguous. Into this model all variables were entered simultaneously. Various diagnostics were used to check for model validity and stability. For example, colinearity could be excluded by inspection of variance inflation factors (VIF) (both = 1.18, determined using the function vif of the R package car) [38]. Model stability was assessed by excluding data points one by one and comparing the derived coefficients. This revealed no obviously influential cases. To establish the significance of the full model a likelihood ratio test was used [27], comparing its deviance with that of the null model comprising only the intercept and the random effect. To test the significance of the interaction between age and sex, the standard z-test, provided by the function lmer [8] was used.

3.6.3 Effects of social relationships on simian foamy virus distribution

To test if close social relationships influence the virus circulation, the SFV pool of offspring to their mother, father, and non-parental group members within the Taï community was compared (**dataset B**). As a measure of dissimilarity between the SFV strains of samples from two individuals, the minimum number of mismatches between any pair of their respective SFV sequences was determined. To test whether this similarity measures corresponded to the familial relationship (mother-offspring, father-offspring, or non-mother-offspring non-father-offspring) between the dyad's members, a GLMM was used. In this model the similarity measures were considered between all samples of each of the 12 offspring, on the one hand, and samples from all group members, on the other hand, given they were collected prior to the respective sample of the offspring. In addition to the relation between the two individuals the following predictors were included, in order to control for their effects: the total time the two individuals have spent together in the same community until the first of the two samples was collected, assuming that longer times spent together might lead to increased similarity of their infecting strains; absolute difference between their birthdays (number of days), assuming that individuals born close to one another might be more similar; the sex of the offspring, assuming female and male offspring might differ in their social/play behavior; the sex of the other individual, assuming that interactions between offspring and other group members might be more likely when the other is a female; the difference between the two sampling dates, assuming that larger differences lead to decreased similarity; and, finally, offspring age, assuming that older offspring might have SFV strains more similar to those of others than younger ones. In addition, the two-way interactions between time spent together, on the one hand, and relation, difference between birthdays, and sex of the other individual, on the other hand, were included in the model. The interaction between time spent together and relation were included because it could be hypothesized that offspring SFV population would generally become more similar to that of any other individuals with increasing age but that this increase in similarity would be more pronounced in mother-offspring dyads. The other two interactions were included to control for their potential effects, assuming that a potentially increased similarity between offspring born close to one another in time and offspring with females would show up only after some time spent together. All these terms were included as fixed effects into the model. To control for the identity of the offspring and the other individual they were included as random effects into the model. Prior to running the model, all numerical predictors were inspected for their distribution and, as a consequence,

square root transformed the difference between birthdays to achieve an approximately symmetrical distribution. Subsequently, all numerical predictors were z-transformed to a mean of zero and a standard deviation of one. To test the overall effect of the factor relation and its interaction, the full model (as described above) was compared with a null model not comprising relation or its interaction with time spent together but including all other terms present in the full model. This comparison was based on a likelihood ratio test [27]. The model was fitted with Poisson error structure and log link function. Using a Poisson error, structure was justified, as the response (dissimilarity) comprised only integer numbers ≥ 0 and since overdispersion was not an issue (dispersion parameter: 0.97; $\chi^2 = 300.3$; df = 310; $p = 0.64$). Also, colinearity was no issue (largest VIF = 1.93).

All analyses were conducted in R (version 2.10.1; R Development Core Team, 2010). The GLMM was run using the function lmer of the R package lme4 [8], VIF were calculated using the functions vif of the R package car [38], and likelihood ratio tests were conducted with the function anova with the argument test set to chisq.

3.7 Network Analyses

3.7.1 Network-based identification of founder strains of simian foamy virus

3.7.1.1 Reasoning line Sequences generated with EPD-PCR are in principle *Taq*-error free sequences [134]. This is not the case for bulk-PCR clone sequences, which will on the contrary often harbor singleton changes due to *Taq* errors. Hence, much of the observed variation will be artifactual.

Therefore, it was of particular interest to develop a method that identifies biological sequences potentially comprised in clone alignments (including minor variants) and more particularly of those sequences which might be at the origin of the infection and/or have become main components of the overall retroviral population (hereafter called founder sequences).

Founder sequences with respect to their biological variants – but also biological sequences with respect to their *Taq*-modified variants – are expected to exhibit two distinctive properties: i) high connectivity and ii) high frequency. While phylogenetic analyses cannot capture these characteristics, tokogenetic analyses (*i.e.*, network building) can in principle allow for their simultaneous assessment [122]. TCS [24], which implements statistical parsimony-based network building as initially described in an article by Templeton, Crandall and Sing [149], offers the advantage to produce “outgroup probabilities” (OP), a statistic, which summarizes sequence

connectivity and frequency for all sequences examined. To identify founder sequences of SFV infection, networks were performed for all individual clone alignments of **dataset A** (Figure 16) and **dataset B** (Figure S2).

For all alignments, median-joining networks [6] were also reconstructed using Network v4.610 (www.fluxus-engineering.com). Those were always similar in shape to their statistical parsimony counterparts. Median-joining networks are presented in this document, as their lay-out could be much more easily reworked, using a dedicated tool, Network Publisher (www.fluxus-engineering.com).

3.7.1.2 Power analysis To investigate the power of network analyses in pointing at likely biological sequences, networks with according OP values were also generated for EPD-PCR derived sequences (**dataset A**, Figure 9). Sequences with highest OPs (either the very best if single infection or the two best ones if super-infection) were checked whether they match with the same founder sequence identified from clone alignments (Table 12).

In addition, it was checked whether the appearance of a sequence in several of the five PCR products (ability to replicate) was likely to point at it being a founder sequence. OP values were determined from all possible alignments (**dataset A**) derived from 3 or 4 PCR products only and in which no sequence was appearing in more than one PCR product so as to determine whether OPs were pointing at EPD-PCR founder sequences when ability to replicate could not be used as a criterion (Table 13).

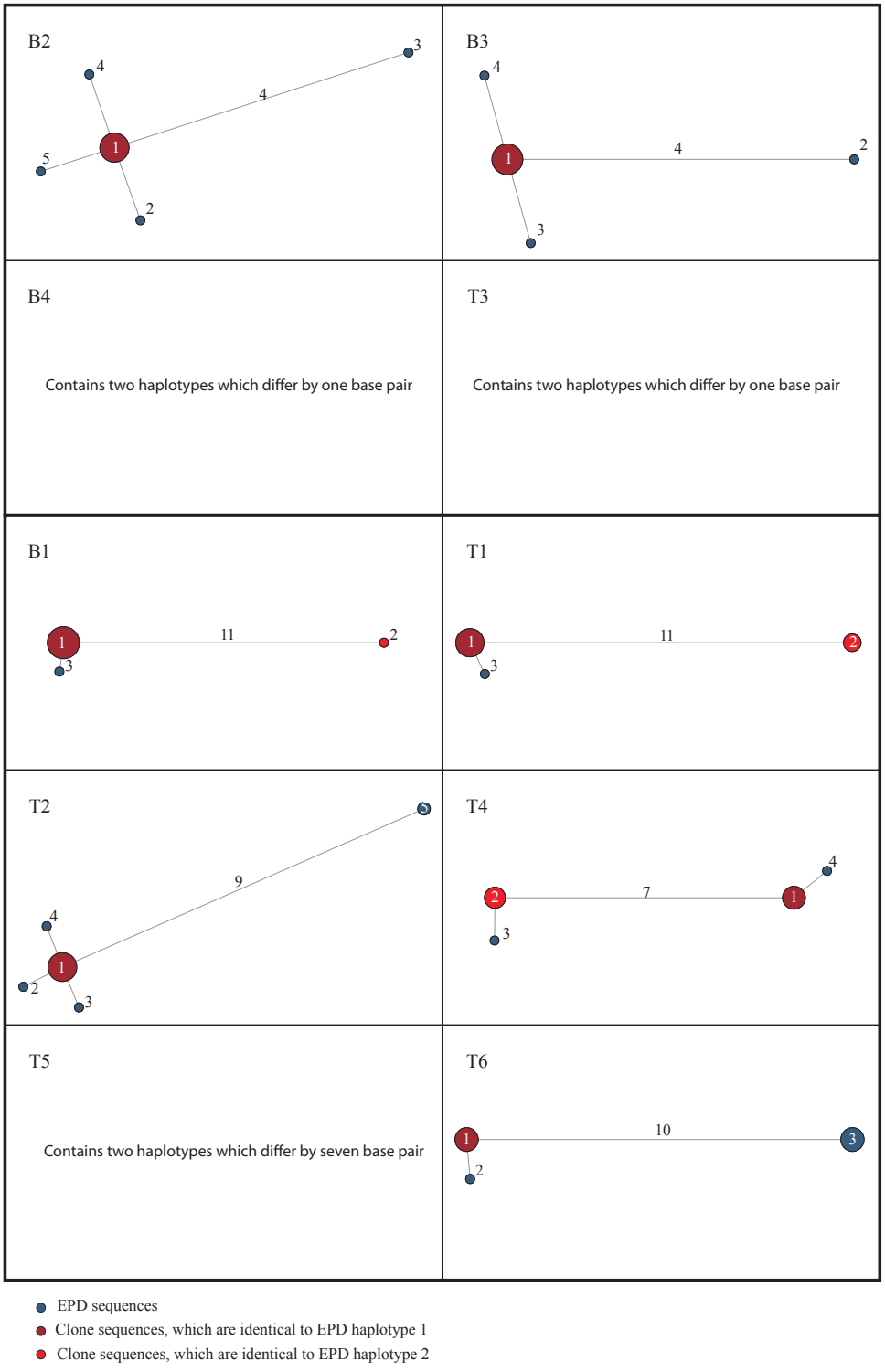

Figure 9: **Median joining network analysis of end-point dilution (EPD) PCR dataset.** The networks are ordered by infection status as determined by EPD-PCR (first four individuals above the bold bar are single infected; six individuals below the bar are super-infected). Individuals whose name start with a "T" are *P. t. verus* from Taï National Park, Côte d'Ivoire, with a "B" are *P. t. schweinfurthii* from Budongo Forest Reserve, Uganda. Within each network, node size is proportional to the frequency of sequence occurrence. Branch lengths are proportionate to the number of mutations between sequences, with values noted for differences greater than two base pairs. Haplotypes 1-5 (as shown in Table 12) are noted within or adjacent to their corresponding node. Networks generated by the parsimony-based network TCS were highly similar (data not shown).

3.7.2 Other networks

In addition, to the individual sample networks, further networks were produced for **dataset A** and **B**. Eight networks from simulated triple infections of **dataset A** were built to check, whether the three highest OPs were pointing to the appropriate "clouds" of sequences (Figure 17, Table 14). Different alignments of **dataset B** were also investigated using network analyses, including all specimen taken from multiply sampled individuals (9 networks; Figure 25, S4), all samples from mothers and their offspring (7 networks; Figure 22), and all samples included in the study of **dataset B** (1 global network out of 915 sequences; Figure 20A). Two alternative versions of the global network were also produced: (i) a streamlined network, which did not comprise clone sequences that exhibited minimal OP values in individual sample networks, that is, excluding these sequences more likely to be PCR-borne variants (128 sequences; Figure 20B), and (ii) a "founder" network including all founder sequences (n = 55; Figure 21). These networks were also computed with Network and TCS, which yielded again in very comparable structures.

3.8 Molecular evolution and substitution processes

To investigate the molecular evolution and substitution processes, clone sequences of **dataset B** were examined for the presence of frameshift-inducing indels (insertion or the deletion of bases), and in-frame stop codons using SeaView v4 [43]. In addition, the proportions of variable sites at the first and second codon positions and at the third codon position were also determined as a proxy of selective forces operating on these sequences (Table 16). Finally, mutations occurring as the first step away from founder sequences were recorded in all TCS-derived individual networks (n = 37) allowing for the determination of a complete non-reversible nucleotide substitution matrix including 471 mutational events (Table 17).

3.9 Estimation of virus persistence

For 9 individuals of the Taï community (**dataset B**), it was possible to unambiguously determine the infection status (single or super-infection) from two to three samples (n = 21) collected 1 to 7 years apart (mean, 3.3 years). Considering that some individuals were super-infected over multiple sampling points and that each of the underlying major strains constituted an independent observation, the number of viral years was determined (Table 15). Here viral years

were defined as the minimum number of consecutive years over which one SFV strain can be assumed to have infected an individual. Practically, this means that if strain A and strain B are detected during year n and strain A and strain B in year n = 1 while only strain A is detected in year n = 2, strain A will have persisted over 2 years (2 viral years) while strain B will have persisted over 1 year (1 viral year). It should be noted here that the assumption is made that two consecutive observations of the same strain are explained by viral persistence rather than other processes, *e.g.*, re-infection.

3.10 Nucleotide sequence accession numbers

All EPD-PCR sequences (all unique sequences per individual of **dataset A**, n = 33) have been deposited in the EMBL depository under accession numbers HE820059-HE820091. All founder sequences (**dataset B**, n = 55), standing most likely for biological variants, are also available in this database under the accession numbers HF568879-HF568933. The complete sequence dataset, comprising all clone sequences (**dataset B**, n = 915), has been deposited in DRYAD under doi: 10.5061/dryad.bb8r6.

4 Results

4.1 Detection of simian foamy virus diversity within a given host

The first objective of this study was to establish a suitable technique for the detailed estimation of intra-individual diversity, *i.e.*, single infection versus multiple infection (super-infection) of the given retrovirus. Therefore, ten SFV positive fecal samples collected from Eastern and Western chimpanzees (**dataset A**: samples B1-B4 and T1-T6) were investigated using bulk-PCR cloning and the gold-standard of EPD-PCR in comparison.

4.1.1 Quantification of retroviral loads in feces

Quantitative PCR performed on samples identified as positive by a nested PCR approach resulted in negative results for the six samples tested, even though the former test exhibited a high sensitivity (detection threshold < 5 molecules; Table 9). When extrapolating to the 80 mg feces typically used for extraction, this suggested that SFV retroviral/proviral loads in feces are in the range of at maximum a handful of copies per mg feces.

4.1.2 End-point dilution PCR

4.1.2.1 Sequence generation For each selected sample, a minimum of 15 EPD-PCR sequences were acquired, a point at which the average nucleotide distance most often reached a plateau, thereby suggesting appropriate sampling of the underlying viral population (Figure 10). Moderate dilutions (down to a maximum factor of 20) were always sufficient to lower the success rate to less than 30 % of the reactions (Table 10), [134]. On average, 69.3 µL of undiluted cDNA was necessary to gather the desired number of EPD-PCR sequences, though the range of used material was highly variable (median volume: 47 µL, range: 11–179 µL; Table 10).

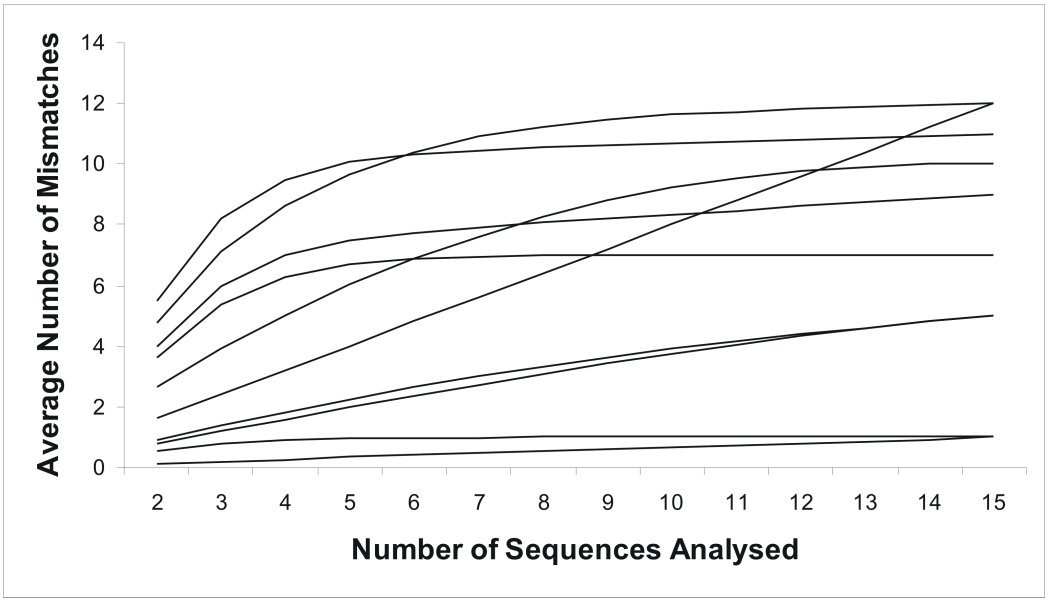

Figure 10: **Appropriate sampling using end-point dilution PCR**: average number of mismatches between end-point dilution-PCR sequences as a function of the number of sequences considered. All possible assemblages of two, three, and so on up to 15 sequences per sample were considered. For each of those, the mean pairwise distance was computed. The average of all mean pairwise distances was finally plotted. Appropriate sampling of the underlying sequence population can be expected to result in reaching a plateau phase.

Table 10: **Quantities of material used in the generation of end-point dilution (EPD) PCR sequences.** The total numbers of necessary experiments are reported in the left section, and those having finally been performed in EPD-PCR conditions are reported in the right section. Individuals whose name start with a "T" are *P. t. verus* from Taï National Park, Côte d'Ivoire, with a "B" are *P. t. schweinfurthii* from Budongo Forest Reserve, Uganda. * marks individuals for which it was necessary to re-extract from the same original fecal bolus to obtain a sufficient number of EPD-PCR sequences. †indicates the threshold for EPD conditions: a maximum 30 % of PCR products are simian foamy virus positive.

	Performed experiments in total			Performed experiments within EPD range		
Individual	Min/max of dilution	Number of PCRs performed	cDNA (µL)	Min/max of end-point dilution	Number of PCRs performed	Number of Positive PCR (%)†
B1*	original/ 1:20	168	71	1:3/ 1:20	104	21 (20)
B2*	original/ 1:5	218	179	original/ 1:5	191	17 (9)
B3	1:5/ 1:10	112	31	1:5/ 1:10	112	15 (13)
B4	1:3/ 1:5	146	56	1:4/ 1:5	130	17 (13)
T1	1:5/ 1:10	92	16	1:10	76	20 (26)
T2	1:4/ 1:5	93	34	1:4/ 1:5	93	22 (24)
T3*	original/ 1:10	298	175	original/ 1:10	266	18 (7)
T4*	original/ 1:5	176	82	1:5	152	17 (11)
T5	1:3	76	38	1:3	76	15 (25)
T6	1:5/ 1:20	92	11	1:20	76	18 (24)

4.1.2.2 Identification of single/super-infections In all cases of **dataset A**, several unique SFV sequences were identified by EPD-PCR (range: 2–5). These sequences were all assumed to represent authentic biological variants and therefore directly used to make a decision regarding chimpanzee-specific SFV super-infection, using mismatch distributions. In this way, eight of the ten individuals could be unambiguously assigned as single or super- infected (6/6 *P. t. verus*: T1 to T6, 2/4 *P. t. schweinfurthii*: B1 and B4; Figure 11). The two remaining individuals (B2 and B3) exhibited slightly ambiguous distributions but were ultimately assigned to the single infection category, based on a rather small maximum divergence of 5 bp (as compared to a minimum of 7 bp difference between different strains and a mean of 10 bp for unambiguously super-infected individuals).

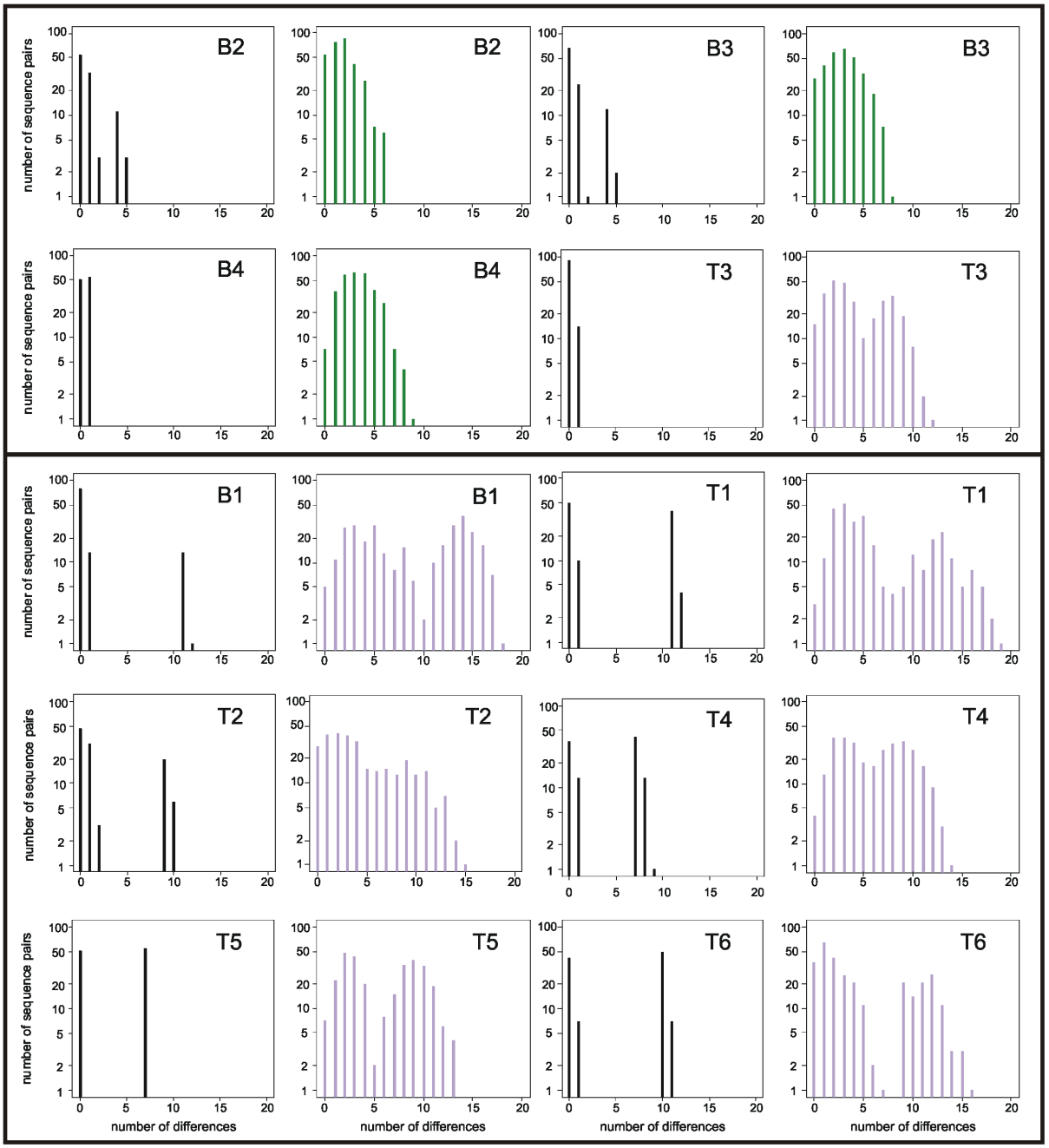

Figure 11: **Super-infection diagnostics using mismatch distribution analyses of end-point dilution (EPD) PCR and clone sequence.** As in Figure 9, individuals are ordered according their infection status as determined by EPD-PCR (first four individuals above the bold bar are single infected; six individuals below the bar are super-infected). Individuals whose name start with a "T" are *P. t. verus* from Taï National Park, Côte d'Ivoire, with a "B" are *P. t. schweinfurthii* from Budongo Forest Reserve, Uganda. Plots are paired for each individual with the EPD-PCR dataset on the left and the bulk-PCR clone dataset on the right. EPD-PCR distributions are black; bulk-PCR clone distributions identified as unimodal (Δ AICc < 2) are green; bulk-PCR clone distributions identified as bimodal are purple (Δ AICc > 2).

4.1.3 Bulk-PCR product cloning and sequencing

4.1.3.1 Sequence generation For each sample five bulk-PCR products were generated, consuming a median volume of 12.5 µL undiluted cDNA. All resulting PCR products (named A to E) were sub-cloned and five clones were sequenced per product. Clone alignments typically exhibited higher levels of polymorphism than the corresponding EPD-PCR sequence alignments, an expected consequence of the accumulation of random *Taq*-errors (*e.g.*, Figure 12). All sequences supported the expected scheme of subspecies specific distribution of SFV, demonstrated by maximum likelihood as well as Bayesian analyses (Figure 13, Figure 14, [89]).

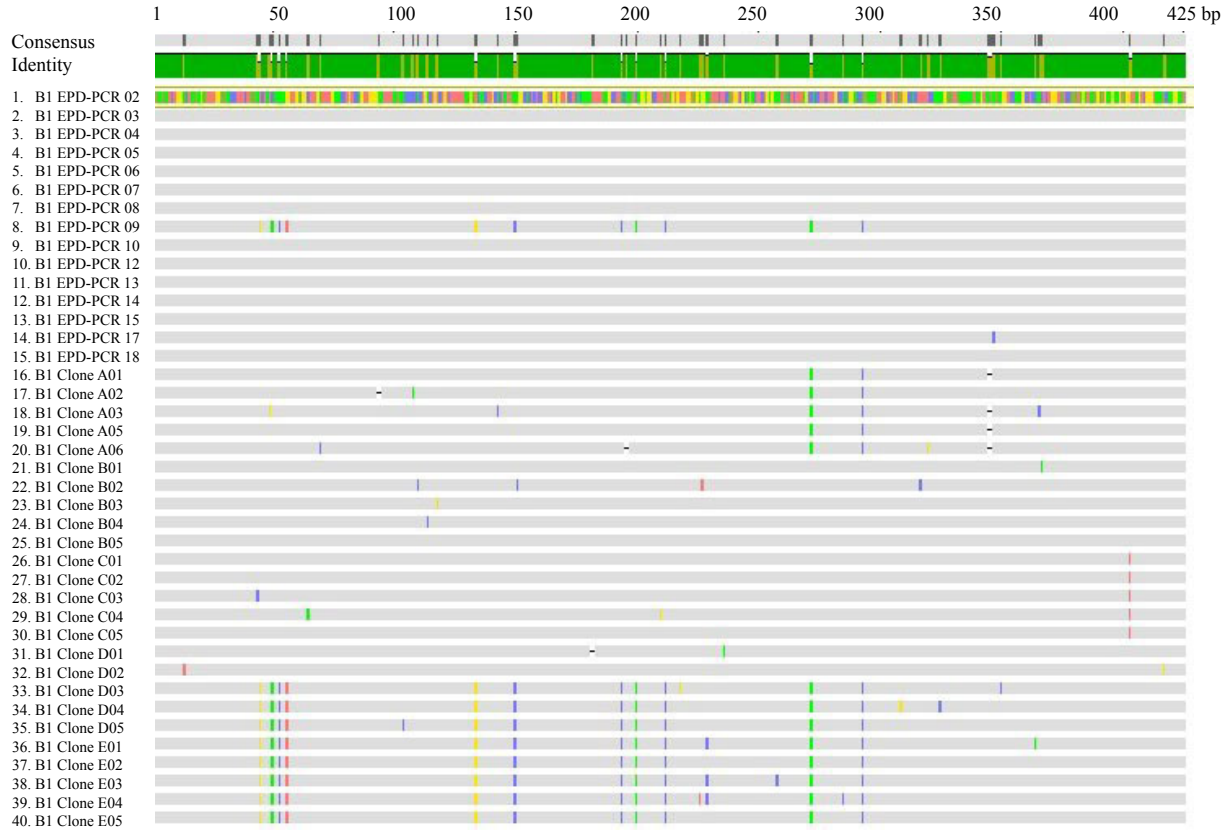

Figure 12: **Nucleotide alignment (425 bp of the Simian foamy virus *integrase* fragment) of end-point dilution (EPD) PCR and bulk-PCR clone sequences for individual B1.** Comparison of generated sequences to one EPD sequence (first sequence). In case of identical base pairs: gray bar, dissimilarities are shown as colored bar. The clone alignment exhibits a higher level of polymorphism than the EPD-PCR alignment.

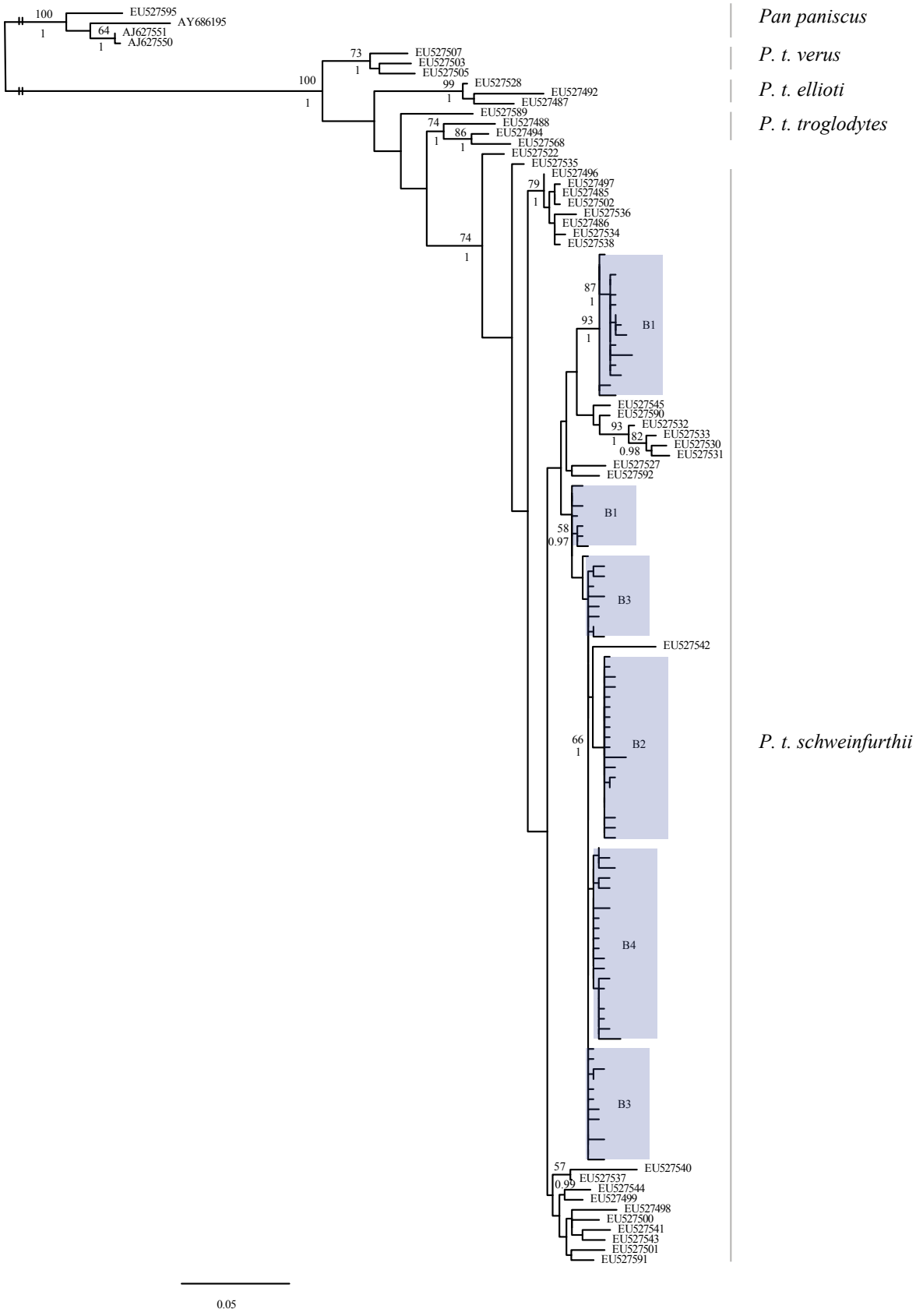
EU527595
AY686195
AJ627551
AJ627550
EU527507
EU527503
EU527505
EU527528
EU527492
EU527487
EU527589
EU527488
EU527494
EU527568
EU527522
EU527535
EU527496
EU527497
EU527485
EU527502
EU527536
EU527486
EU527534
EU527538
B1
EU527545
EU527590
EU527532
EU527533
EU527530
EU527531
EU527527
EU527592
B1
B3
EU527542
B2
B4
B3
EU527540
EU527537
EU527544
EU527499
EU527498
EU527500
EU527541
EU527543
EU527501
EU527591
0.05
Pan paniscus
P. t. verus
P. t. ellioti
P. t. troglodytes
P. t. schweinfurthii

Figure 13: **Maximum likelihood tree of all unique simian foamy virus sequences (clone and end-point dilution PCR sequences) derived from Budongo chimpanzees** (blue boxes refer to the given individual B1-B4). Reference sequences from all chimpanzee subspecies were included, *Pan paniscus* was used as outgroup [89, 151, 148]. The topology was similar when using Bayesian analyses. Branch robustness is shown as bootstrap values (Bp, produced from 500 pseudo-replicates) and posterior probability values (*pp*). Bp (numbers above the branches) and *pp* (italicized numbers below branches) are only presented where Bp $\geq$ 50 and *pp* $\geq$ 0.95. Scale bar indicates nucleotide substitutions per site.

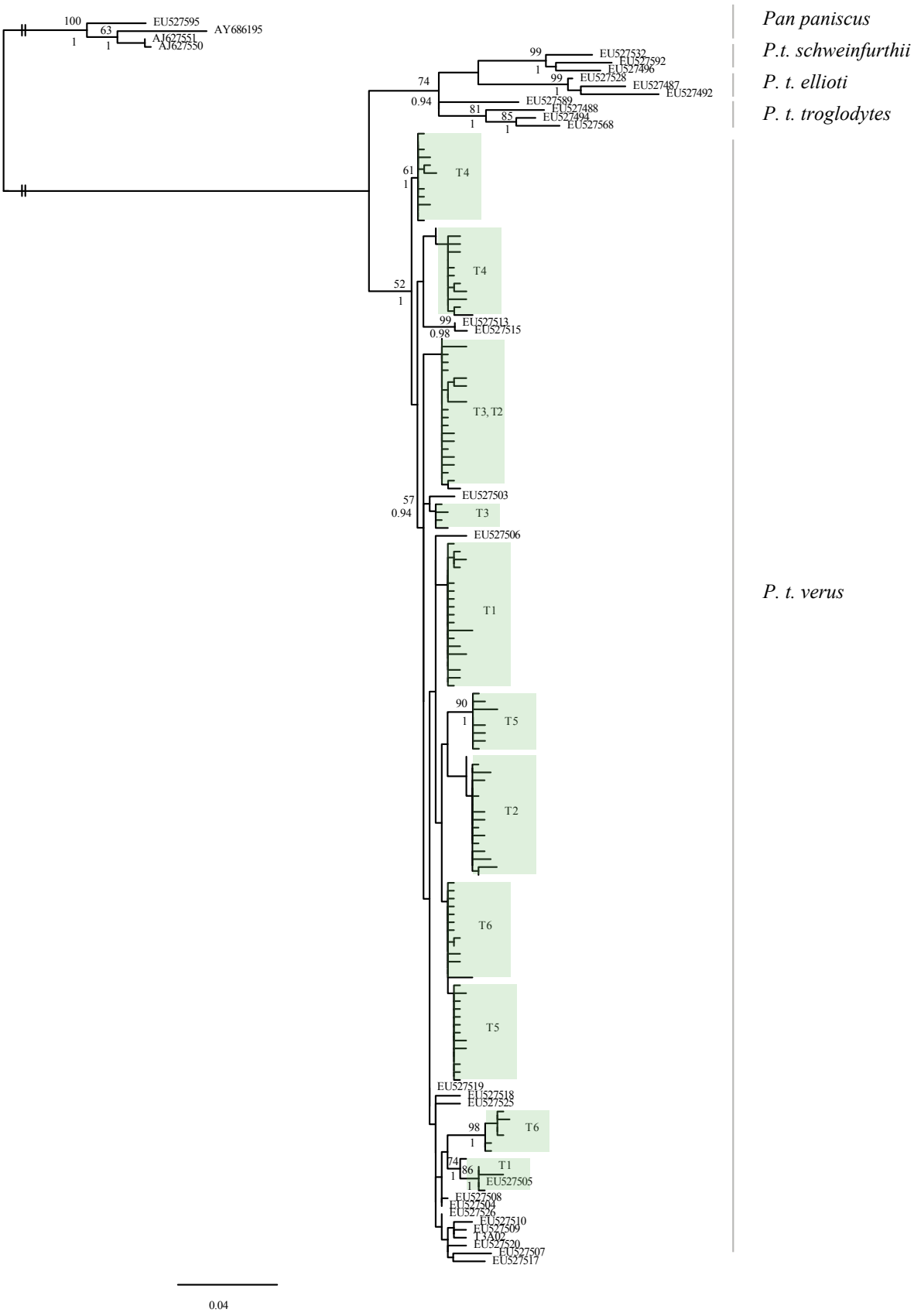
Pan paniscus
P.t. schweinfurthii
P. t. ellioti
P. t. troglodytes
P. t. verus
EU527595
AY686195
AJ627551
AJ627550
EU527532
EU527592
EU527496
EU527528
EU527487
EU527492
EU527589
EU527488
EU527494
EU527568
T4
T4
EU527513
EU527515
T3, T2
EU527503
T3
EU527506
T1
T5
T2
T6
T5
EU527519
EU527518
EU527525
T6
T1
EU527505
EU527508
EU527504
EU527526
EU527510
EU527509
T3A02
EU527520
EU527507
EU527517
0.04

Figure 14: **Maximum likelihood tree of all unique simian foamy virus sequences (clone and end-point dilution PCR sequences) derived from Taï chimpanzees** (green boxes refer to the given individual T1-T6). Reference sequences from all chimpanzee subspecies were included, *Pan paniscus* was used as outgroup [89, 151, 148]. The topology was similar when using Bayesian analyses. Branch robustness is shown as bootstrap values (Bp, produced from 500 pseudo-replicates) and posterior probability values (*pp*). Bp (numbers above the branches) and *pp* (italicized numbers below branches) are only presented where Bp ≥ 50 and $pp \geq 0.95$. Scale bar indicates nucleotide substitutions per site.

4.1.3.2 Identification of single/super-infections Simple statistics computed from individual clone alignments, such as the mean observed distance among clones or the percentage of resolved/unresolved topologies obtained using quartet puzzling analysis (*i.e.*, likelihood mapping), did not correlate well with infection status as determined through EPD-PCR sequence analysis, even though trends could be observed (mean and standard deviation of observed distance and the proportion of resolved quartets tend to be higher for super-infection than for single infections cases; Table 11). In addition, the proportion of unresolved quartets in the likelihood mapping analyses (32.5–98.2 %; Table 11) suggested alignments only comprised little phylogenetic information. This was also reflected by generally low support of many branches of the phylogenetic trees, though the main bipartition in super-infected individuals were always highly supported (Figure S1).

Assessing the shape of the mismatch distribution (uni- or bimodal) for the 10 complete clone alignments (*i.e.*, all 25 sequences stemming from five bulk-PCR products of a subject), allowed to properly infer individual status with respect to super-infection, as revealed by EPD-PCR results, in 9 out of ten cases (Figure 11).

Using the same test it was also possible to investigate the impact of the number of bulk-PCR products analyzed on the super-infection diagnostic. The latter revealed that the probability to infer single infection for super-infected individuals (*i.e.*, probability of a false negative) was zero as soon as three or more products were analyzed (Figure 15 A). The probability to infer super-infection for single infected individuals (*i.e.*, probability of a false positive) was markedly higher (up to 80 %; Figure 15 B).

Table 11: **Additional statistics computed from individual bulk-PCR clone sequence datasets.** The first four individuals (gray) have been identified as single infected using end-point dilution (EPD) PCR, others below as super-infected. Individuals whose name start with a "T" are *P. t. verus* from Taï National Park, Côte d'Ivoire, with a "B" are *P. t. schweinfurthii* from Budongo Forest Reserve, Uganda. [$] Likelihood mapping category "Partly resolved" is not shown here but it was always under 2.5 % and can be determined from the other values using the equation "Partly resolved" = 100 % - ("Unresolved"+ "Resolved"). [#] Branch support of phylogenetic tree is given as approximate likelihood ratio test (aLRT) values and equals to the main bipartition observed in corresponding networks. NA: Branch support for bipartition could not be assessed. In case these statistics are ideal for super-infection diagnostic, following results would have been expected: high values for observed distance in the alignment + high value for "Resolved", low value for "Unresolved" and "Partly resolved" + high branch support for main bipartition in the phylogenetic tree.

Individual	Observed distance (%)		Likelihood mapping[$](%)		Phylogeny
	Mean	Standard deviation	Unresolved Star-*like*	Resolved Tree-*like*	branch support for main bipartition[#]
B2	0.4	0.3	98.2	1.8	NA
B3	0.6	0.4	96.2	3.8	NA
B4	0.8	0.4	56	44	NA
T3	1	0.7	72.3	27	NA
B1	1.9	1.3	32.5	67.5	1
T1	1.5	1.1	57.6	42.3	0.99
T2	1	0.9	72.6	26.8	0.95
T4	1.4	0.8	32.6	64.9	0.97
T5	1.4	0.9	47.7	51.4	0.99
T6	1.2	1.1	77.3	22.7	1

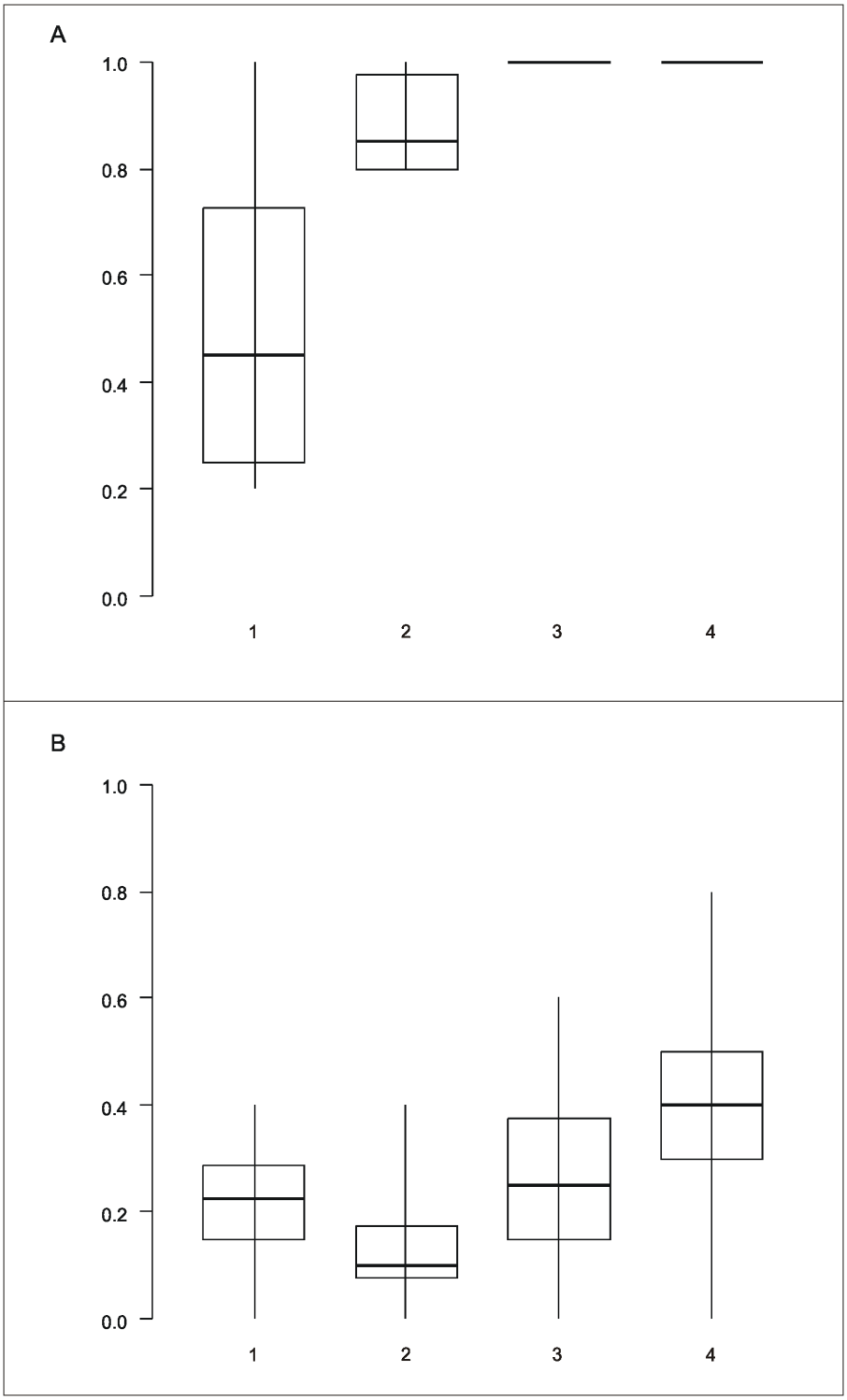

Figure 15: **Appropriate sampling using bulk-PCR: false negativity and positivity as a function of the number of bulk-PCR products analyzed.** Probability to detect bimodality (*i.e.*, super-infection) in the frequency distribution of the number of mismatches per pair of sequences (y-axis) as a function of the number of bulk-PCR products (x-axis) analyzed (A) for subjects for which end-point dilution (EPD) PCR revealed a super-infection (n = 6); (B) for subjects for which EPD-PCR revealed a single infection (n = 4). Shown are median, quartiles, minimum and maximum, of the respective probabilities per subject.

4.1.4 Identification of founder sequences

Sequences having likely founded the sampled retroviral populations (or at least standing for major variants at the time of sampling) were identified from EPD-PCR sequence alignments using outgroup probabilities (OPs) generated by TCS (Table 12). These EPD-PCR founder sequences were actually captured in 14 of 16 cases (here one case equals one founder sequence), when considering the total five bulk-PCR products generated for each sample (Table 12). From these 14 clone sequences, 10 had been replicated. Replication was generally not observed for other clone sequences, with two exceptions (B4 clone b and T1 clone e; Table 12). In the first case, two distinct clone sequences determined from chimpanzee B4 were replicated (B4 had been identified as single infected). Both sequences had been sampled through EPD-PCR and were very closely related (1 bp divergence). In the second case, a T1 clone sequence could be replicated which had not been identified through EPD-PCR, thereby revealing a possible example of erroneous replication-based identification of clone sequences or the sampling of a biological variant not sampled by EPD-PCR.

TCS networks were also computed from all clone alignments in order to assess the performance of OPs in pointing at EPD-PCR-determined founder sequences. The overall shape of the clone networks reflected expectations; "star-like" patterns were commonly observed, thereby highlighting founder effects (both biological and *Taq*-error dependent; Figure 16). Super-infected individuals also exhibited more complex networks (Figure 16), which were sometimes even disconnected (when computed using statistical parsimony TCS, marked by * in Table 12). The phylogenetic analyses of individual datasets always provided high support to the same branch defining distinct clouds of sequences in networks of super-infected individuals (Table 11). In all the 14 aforementioned cases where EPD-PCR-determined founder sequences had been sampled, the highest or two highest OP values, for single and super-infected individuals, respectively, highlighted the correct clone sequences as founders (Table 12). The co-occurrence of two very closely related sequences in B4 was also supported by OPs (B4 clone a and b; Table 12). In the two remaining cases – where one of the two EPD-PCR founder sequences had not been captured by cloning – selecting the sequences with the two highest OPs as potential founder sequences involved would have resulted in: i) misleadingly identifying a sequence closely related to the sampled EPD-PCR founder sequence as the source of the second infection (T2 clone b; Figure 16 and Table 12); and ii) being unable to identify a second founder sequence (T6; Figure 16 and Table 12), as all sequences differing from the first founder se-

quence were only exhibiting the “noise” OP, *i.e.*, the minimal OP value attributed to poorly represented/connected sequences.

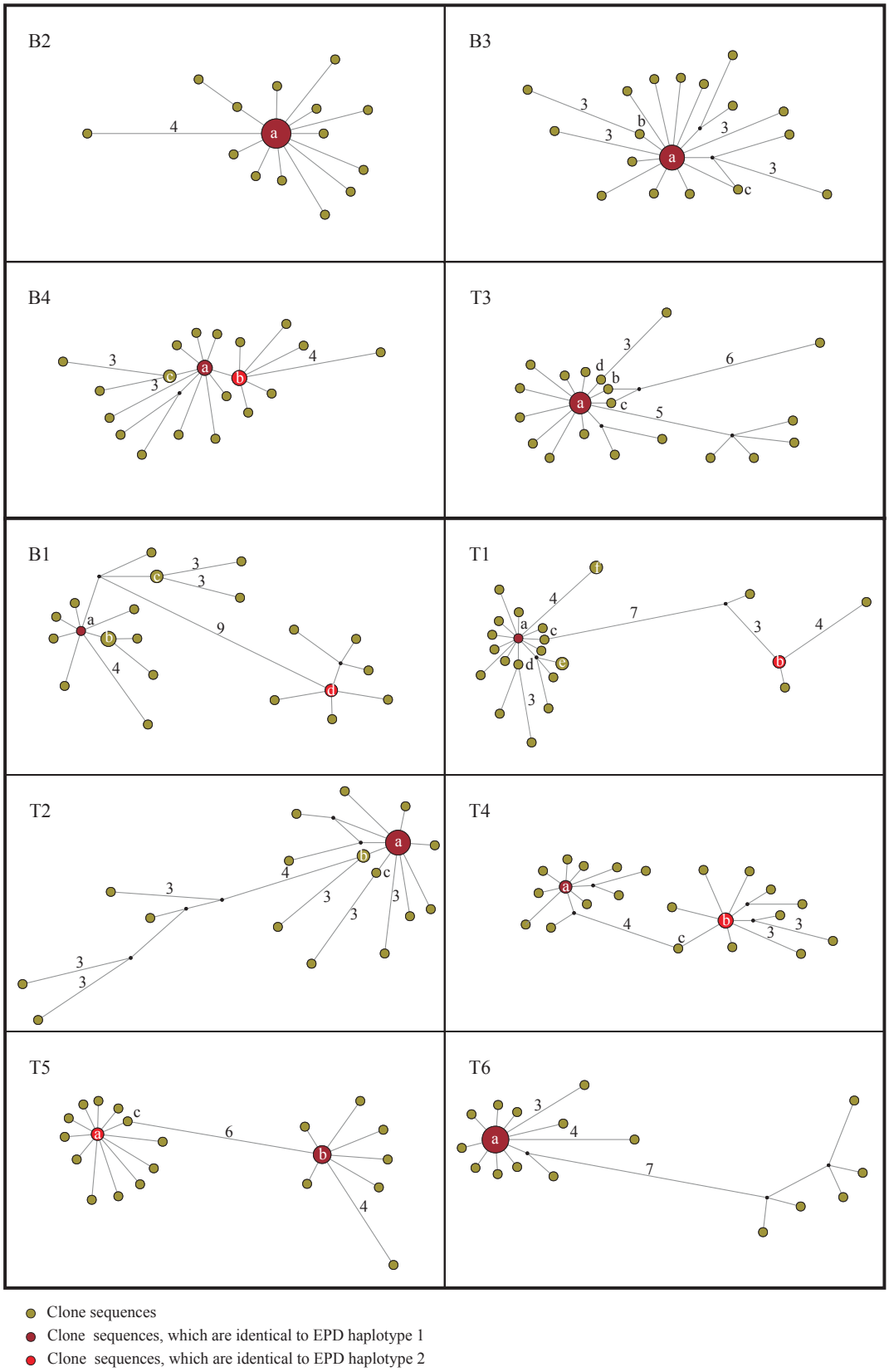

Figure 16: **Median joining network of bulk-PCR product clone sequences.** As in Figure 11, the individual networks are ordered by infection status. Within each network, node size is proportional to the frequency of sequence occurrence (total n = 25 for each individual). Branch lengths are directly related to the number of mutations between sequences, with values noted for differences greater than two base pairs. Clone haplotypes a-f (as shown in Table 12) are noted within or adjacent to their corresponding node. Networks generated by the parsimony-based network TCS were highly similar (data not shown).

Table 12: **Identification of founder sequences from end-point dilution (EPD) PCR and bulk-PCR clone sequence alignments.** All EPD-PCR haplotypes, *i.e.*, unique sequence types, are presented, while only those haplotypes reaching outgroup probability (OP) values above the minimum value are shown for bulk-PCR clone sequences. OPs are generated by the parsimony-based network TCS and summarize sequence connectivity and frequency for all sequences examined. Assumed founder sequences are highlighted in bold. # The origin of sequences refer to the bulk-PCR products from which they originate (*e.g.*, if haplotype a appeared two times in PCR product B and once in PCR product C, then 2 x B and 1 x C will appear in this column). * The individuals for which statistical parsimony analyses produced two separated networks; here the sum of all OPs will be greater than one as OPs will be calculated independently for each network. Individuals whose name start with a "T" are *P. t. verus* from Taï National Park, Côte d'Ivoire, with a "B" are *P. t. schweinfurthii* from Budongo Forest Reserve, Uganda.

Individual	EPD network			Clone network		
	EPD-PCR haplotypes	Number of sequences	Haplotype OP	Bulk-PCR clone haplotypes	Haplotype OP	Number and origin of sequences#
B1*	**1**	13	0.93	**a**	0.36	1 x B
	-	-	-	b	0.26	3 x C
	-	-	-	c	0.10	2 x A
	2	1	1	**d** *	0.50	2 x E
	3	1	0.07	-	-	-
B2	**1**	11	0.88	**a**	0.48	2 x A, 4 x B, 2 x C, 1 x D, 2 x E
	-	-	-	b	0.35	1 x E
	2	1	0.03	-	-	-
	3	1	0.03	-	-	-
	4	1	0.03	-	-	-
	5	1	0.03	-	-	-
B3	**1**	12	0.90	**a**	0.33	3x B, 1 x C, 2 x D, 2 x E
	-	-	-	b	0.23	1 x C
	-	-	-	c	0.23	1 x D
	2	1	0.03	-	-	-
	3	1	0.03	-	-	-
	4	1	0.03	-	-	-
B4	**1**	9	0.60	**a**	0.35	1 x B, 1 x C, 1 x E
	2	6	0.40	b	0.26	2 x D, 1 x E
	-	-	-	c	0.15	2 x A
T1*	**1**	10	0.90	**a**	0.37	1 x E
	2	4	1.00	**b**	0.11	1 x A, 1 x D
	-	-	-	c	0.07	1 x E
	-	-	-	d	0.07	1 x C
	-	-	-	e	0.04	1 x B, 1 x D
	-	-	-	f	0.04	2 x A
	3	1	0.09	-	-	-

Table 12. Cont.

Individual	EPD network			Clone network		
	EPD-PCR haplo-types	Number of sequences	Haplotype OP	Bulk-PCR clone haplotypes	Haplotype OP	Number and origin of sequences[#]
T2*	**1**	10	0.89	**a**	0.33	1 x B, 3 x D, 4 x E
	-	-	-	b	0.26	2 x B
	-	-	-	b	0.23	1 x A
	2	1	0.03	-	-	-
	3	1	0.03	-	-	-
	4	1	0.03	-	-	-
	5	2	1.00	-	-	-
T3	**1**	14	0.93	**a**	0.29	2 x C, 2 x D, 2 x E
	-	-	-	b	0.17	1 x B
	-	-	-	c	0.17	1 x B
	-	-	-	d	0.17	1 x E
	2	1	0.07	-	-	-
T4	**1**	7	0.50	**a**	0.32	2 x A
	2	6	0.44	**b**	0.19	2 x D, 1 x E
	-	-	-	c	0.05	1 x E
	3	1	0.03	-	-	-
	4	1	0.03	-	-	-
T5	**2**	6	0.40	**a**	0.33	1 x C, 1 X E
	1	9	0.60	**b**	0.22	2 x B, 2 x D
	-	-	-	c	0.11	1 x E
T6*	**1**	7	0.88	**a**	0.74	2 x A, 2 x B, 1 x C, 2 x D, 2 x E
	2	1	0.13	-	-	-
	3	7	1.00	-	-	-

In order to allow further comparison of replication-based and network/OP-based identification methods, networks for all the possible combinations of three or four PCR products for which EPD-PCR founder sequence(s) would not have been replicated (n = 41) were computed (Table 13). Using OPs, EPD-PCR founder sequences were properly identified in 63 % of the cases when only 100 % accurate identifications were considered, *i.e.*, both founder sequences were properly identified in cases of super-infection, (Table 13) and in 77 % of the cases when partial identifications were considered as well, *i.e.,* only one of the founder sequences was properly identified in cases of super-infection; data not shown.

Table 13: **Identification of end-point dilution (EDP)-PCR founder sequences from bulk-PCR clone sequence analyses where replication-based identification is not applicable.** # replicated clones *i.e.* clones with the highest OP value(s), which appear in more than one of the five PCR products; otherwise founder sequences are not represented in this column (*e.g.*, two founder sequences of B1). Only the most conservative view (that showing 100 % accuracy in identification) is presented here. Outgroup probabilities (OPs) were computed with the parsimony-based network TCS for all possible combinations of bulk-PCR clone alignments which would not have allowed using replication as a criterion for end-point dilution PCR founder sequence identification. * marks a case, where the founder sequences built separate networks consisting of less than 3 sequences; these were assumed to be negative. Individuals whose name start with a "T" are *P. t. verus* from Taï National Park, Côte d'Ivoire, with a "B" are *P. t. schweinfurthii* from Budongo Forest Reserve, Uganda.

Individual	Infection status according to EPD-PCR analysis	EPD-PCR founder sequence later identified as replicated clones[#] (number of PCR products)	Combination of three PCR products without replication of EPD-PCR founder sequence(s)		Combination of four PCR products without replication of EPD-PCR founder sequence(s)	
			Number of possible combinations	Occurrences of EPD-PCR founder sequences in the first, or first and second, OP	Number of possible combinations	Occurrences of EPD-PCR founder sequences in the first, or first and second, OP
B1	super					
B2	single	a (5)	-	-	-	-
B3	single	a (4)	-	-	-	-
B4	single	a (3)	3	2	-	-
T1	super	b (2)	6	1*	2	0
T2	super	a (3)	3	3	-	-
T3	single	a (3)	3	3	-	-
T4	super	b (2)	6	3	2	1
T5	super	a (2)	6	6	2	2
		b (2)	6	4	2	1
T6	super	a (5)	-	-	-	-
		Summary	33	22	8	4
		Success rate (%)	67		50	

4.1.5 Simulated triple infections

The shape of individual mismatch distributions, networks and phylogenetic trees did not provide evidence for triple infection cases in the experimental set of samples. Therefore 84 triple infection cases were simulated which were all recognized as super-infection cases, including six triplets consisting of two interspersed "clouds" plus one more distant one, *i.e.*, actually cases of double infection similar to those identified in the experimental panel. Eight randomly selected simulated triple infections cases were also examined using networks and TCS: the three highest OPs properly pointed to EPD-PCR sequences from the three "clouds" in all cases (Table 14, Figure 17).

Table 14: **Identification of founder sequences from simulated triple infections.** Triple infection cases (n = 84) have been simulated from randomly re-shuffled bulk-PCR clone sequences (n = 24-26), while eight randomly selected case were examined using the parsimony-based network TCS. The according networks and mismatch distributions are shown in Figure 17. [$] Name of infection cases, *e.g.,* "T1-A_T2-A_T4-A" refers to the origin of the three sequence clouds and is generated as follows: [individual T1-T6]-[phylogenetically defined sequence type: "cloud"A or B]. Sequence from pool A or B were randomly selected; therefore, *e.g.,* "T4-A" in infection case one and three are not totally identical. For all infection cases haplotypes reaching outgroup probability values (OPs) above the minimum value (a-e) are shown. [#] shows the sequence frequency of the haplotype and its origin (*e.g.*, haplotype "a" appears two times in infection case one "T1-A_T2-A_T4-A". Sequence frequency and connectivity are also demonstrated in the according network (Figure 17). * Marks the triple infection cases for which statistical parsimony analyses produced two separated networks; here the sum of all OPs will be greater than one as OPs will be calculated independently for each network.

Simulated triple infection case[$]	Bulk-PCR clone haplotypes	Haplotype OP	Number and origin of sequences[#]
T1-A_T2-A_T4-A*	a	0.29	2 x T2-A
	b	0.29	7 x T2-A
	c	0.23	1 x T4-A
	d	0.03	1 x T4-A
	e	0.5	1 x T1-A
T1-B_T3-A_T5-B	a	0.27	4 x T3-A
	b	0.21	3 x T1-B
	c	0.18	4 x T5-B
	d	0.15	1 x T1-B
T1-B_T4-A_T6-A	a	0.31	6 x T6-A
	b	0.24	3 x T1-B
	c	0.17	1 x T4-A
	d	0.03	2 x T1-B
T2-A_T5-A_T6-A	a	0.38	6 x T6-A
	b	0.25	5 x T2-A
	c	0.13	2 x T5-A
T3-A_T4-A_T4-B	a	0.35	5 x T3-A
	b	0.35	1 x T4-B
	c	0.21	1 x T4-A
	d	0.14	2 x T4-B
T3-A_T5-A_T5-B	a	0.29	5 x T3-A
	b	0.24	2 x T5-B
	c	0.18	3 x T5-A
	d	0.09	1 x T5-B
T4-B_T5-B_T6-A*	a	0.71	3 x T4-B
	b	0.86	4 x T5-B, 8 x T6-A
T5-A_T5-B_T6-A	a	0.46	1 x T5-B, 7 x T6-A
	b	0.23	1 x T5-B
	c	0.15	13 x T5-A

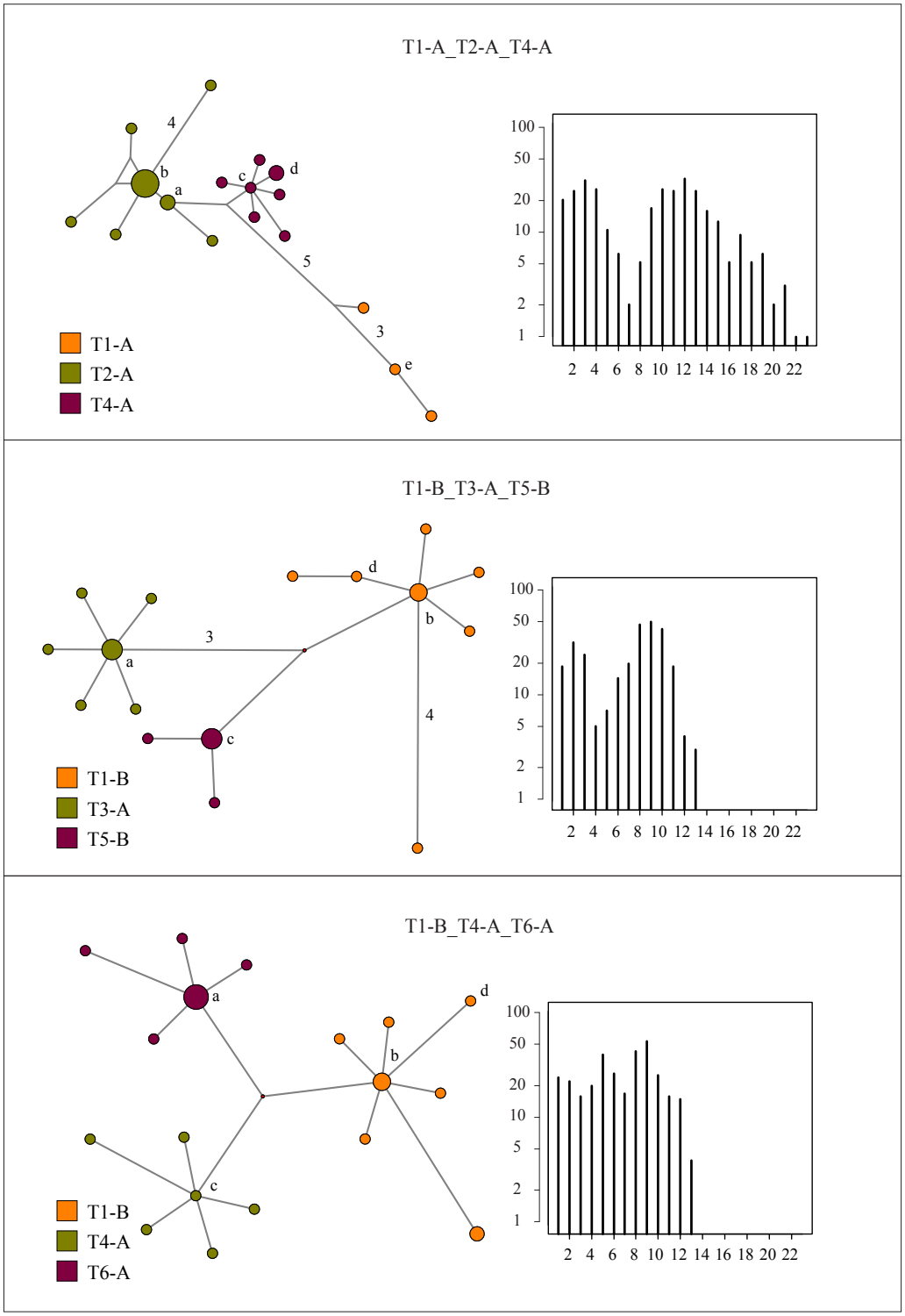

Figure 17: legend please refer to page 70.

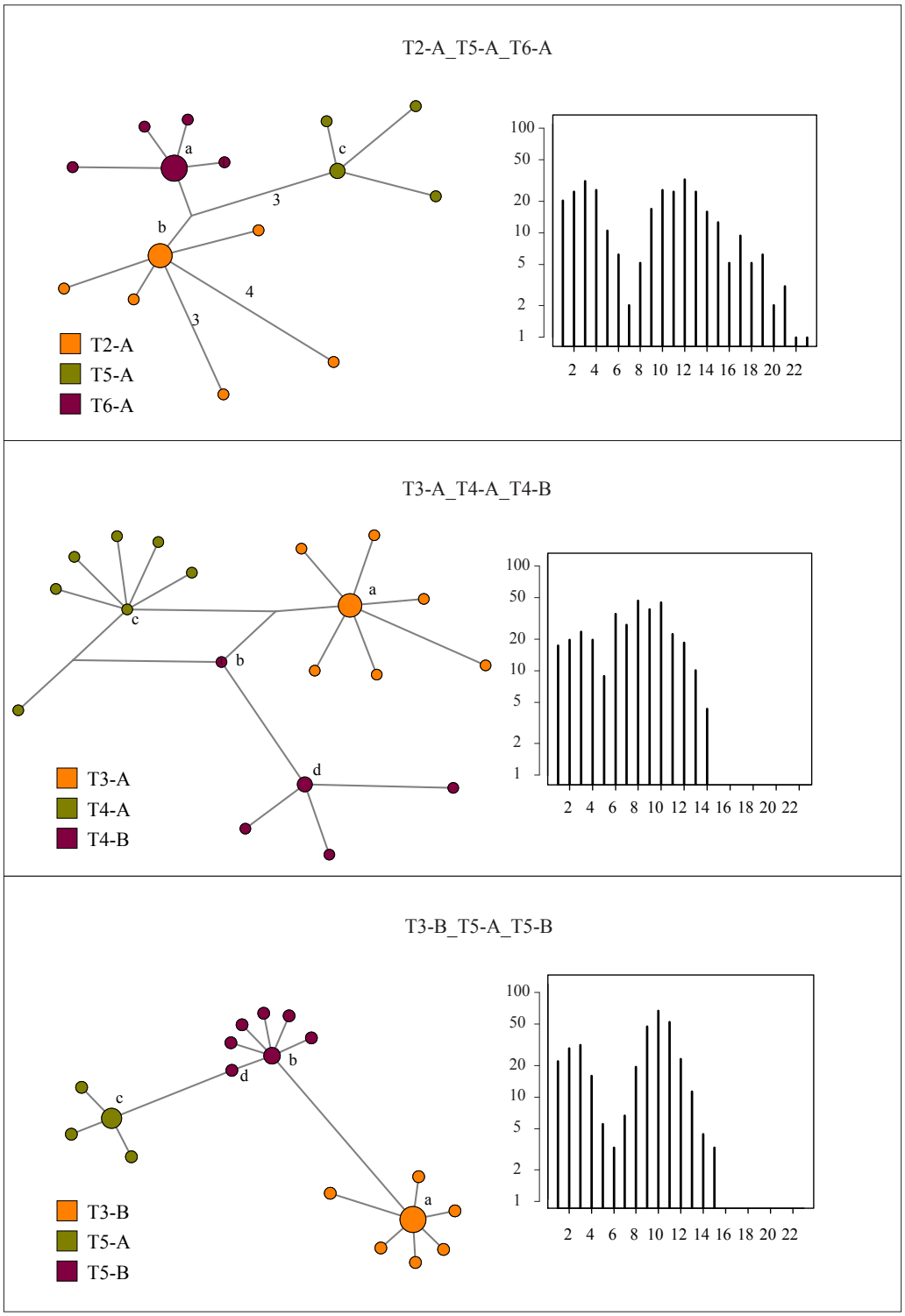

Figure 17: legend please refer to page 70.

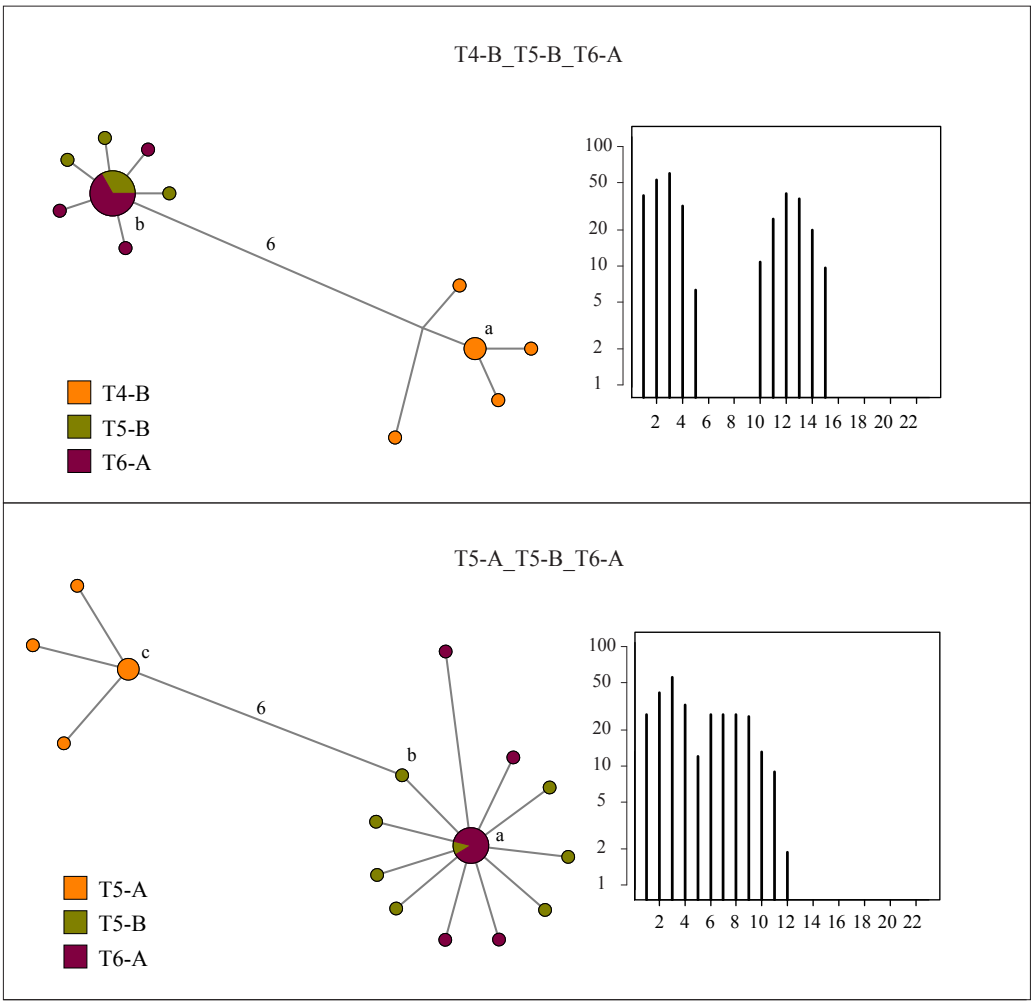

Figure 17: **Mismatch distributions and sequence networks for eight simulated triple infection cases.** Table 14 provides a detailed description of the simulated triple infection cases such as sequence origin: individual (T1-T6), sequence "cloud" A or B. Plots are paired for each case with mismatch distribution on the right and network on the left. Within each network, node color refers to the origin of the bulk-PCR clone sequence, node size is proportional to the frequency of the sequence occurrence (total n = 24-26 for each case). Branch lengths are directly related to the number of mutations between sequences, with values noted for differences greater than two base pairs. Clone haplotypes a-e (as shown in Table 14) are noted within or adjacent to their corresponding node. Mismatch distributions present the frequency of the number of mismatches (y-axis) according to the number of base pairs in the sequence alignment (x-axis).

4.2 Transmission modalities of simian foamy virus

The established method based on bulk-PCR cloning to estimate retroviral diversity was then applied to a larger dataset derived from a single community in the Taï National Park (**dataset B**). This allowed to study the biology of SFVs in a highly social community, investigating the possibility of vertical transmission of SFVs in the wild in particular.

4.2.1 Sequence generation

A total of 183 bulk-PCR products were generated from 37 fecal samples obtained from 23 chimpanzees (**dataset B**). All PCR products were cloned, resulting in a total of 915 clone sequences.

4.2.2 Sequence analyses

4.2.2.1 Identification of single/super-infections For 32 samples from 20 individuals, clear assignment to either of the categories was possible: 14 samples showed single infection, 18 samples showed super-infection (Figure S2). The five remaining samples could not be unambiguously classified and were therefore excluded from further analyses.

Beside the pattern of uni- or bimodality, three samples (Coco, 24 y; Rebecca, 7 y; Rubra, 38 y) showed some trend to trimodality. Therefore, a phylogenetic tree was recommitted including all corresponding clone sequences of the suspicious samples and all founder sequences of the community to check whether a pattern suggestive of triple infections was emerging. Only a handful of branches receive decent support and none of these branches allow for the definition of unambiguous triple infection (Figure 18). Similar findings were observed when individual networks and individual phylogenetic trees were examined (Figure S2, S3). All in all, there is no evidence for unambiguous instances of infection with more than two SFV strains within the Taï dataset.

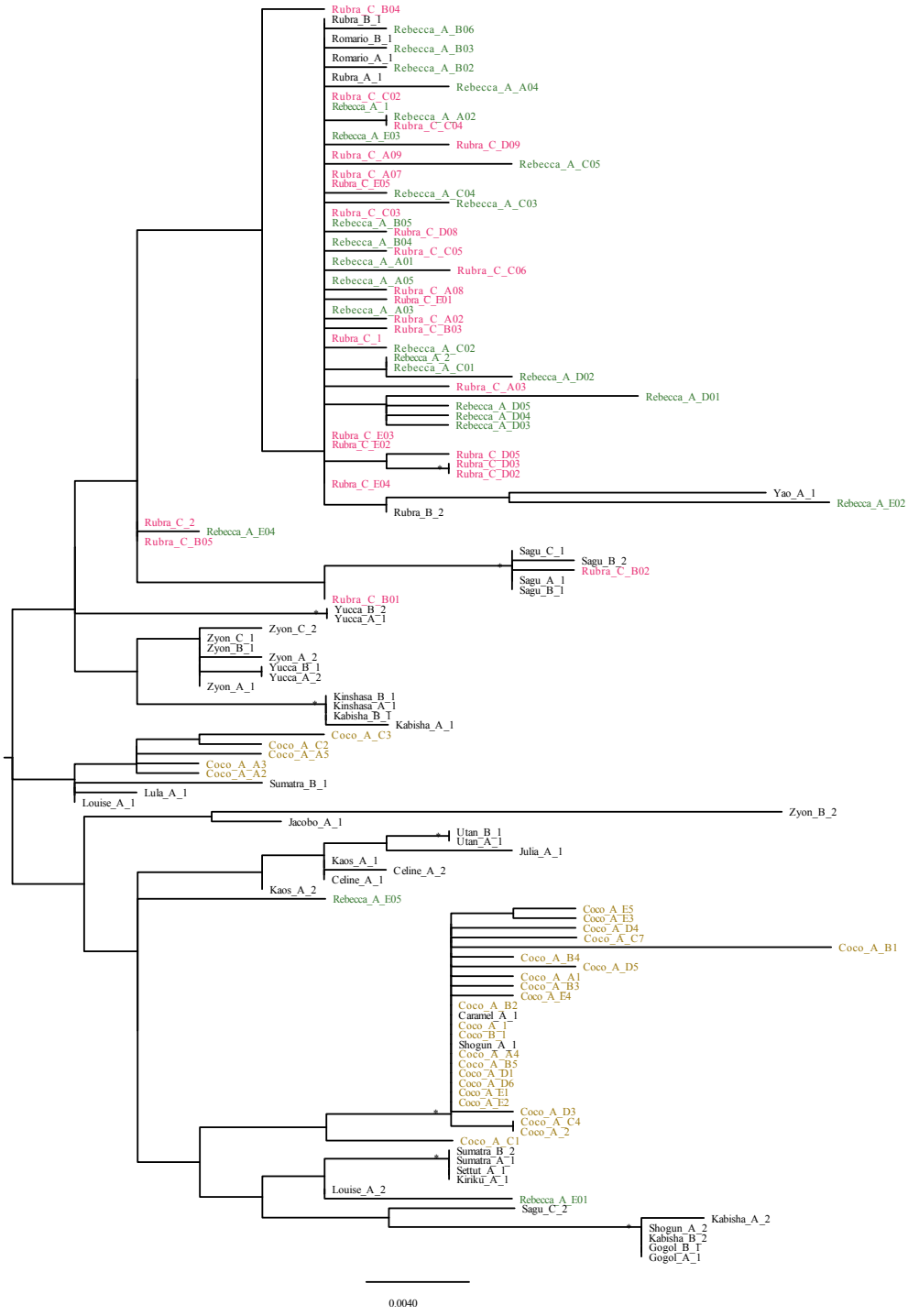

Figure 18: **Maximum likelihood tree of suspicious triple infections.** The phylogenetic tree includes all clones of the three individuals for which triple infection might be suspected (one color per individual) plus all founder sequences of the study (dataset B: 20 individuals from Taï National Park; Table 2). Sequence names are built as follows: [individual]_[sample "A" or "B"]_[founder sequence "1" or "2"] respectively [PCR product "A" to "E" and clone number "01" to "09"] for the bulk-PCR clone sequences (n = 25) of the three of suspicious triple infection cases. The tree is unrooted. Branch robustness is shown as bootstrap values (Bp; produced from 500 pseudo-replicates). Bp, indicated by an asterisk (*), are only presented where Bp $\geq$ 70. Scale bar indicates nucleotide substitutions per site. Please note that most branches are not statistically supported.

4.2.2.2 Identification of simian foamy virus diversity within the community To investigate the overall SFV diversity within the community, a phylogenetic tree was performed including founder sequences of all individuals. The analysis evidenced that most inner branches were poorly supported (for a definition of founder sequences please refer to Materials and Methods) (Figure 19). This was interpreted as reflecting the inappropriateness of phylogenetic methods in depicting reticulate evolution and most notably here the fact that some sequences are likely ancestral to others. Therefore, network analyses were implemented on three datasets: (i) a full dataset comprising the 652 unique sequences (Figure 20A), (ii) a streamlined dataset, from which sequences likely to stand only for "methodological noise" had been removed (Figure 20B), and (iii) a dataset only comprising founder sequences (Figure 21). All networks, including both streamlined versions, exhibited considerable complexity, despite a maximum observed pairwise distance of only 6 % (Figure 20A and B, Figure 21).

Figure 19: **Maximum likelihood tree of founder sequences** (n = 55). The phylogenetic tree includes all founder sequences of dataset B: 20 individuals (one color per individual) from Taï National Park (Table 2). Sequence names are built as follows: [individual]_[sample "A" or "B"]_[founder sequence "1" or "2"]. Reference sequences from all chimpanzee subspecies were included, *Pan paniscus* was used as outgroup [89, 151, 148]. Branch robustness is shown as bootstrap values (Bp; produced from 500 pseudo-replicates). Bp, indicated by an asterisk (*), are only presented where Bp $\geq$ 70. Scale bar indicates nucleotide substitutions per site. Please note that most branches are not statistically supported.

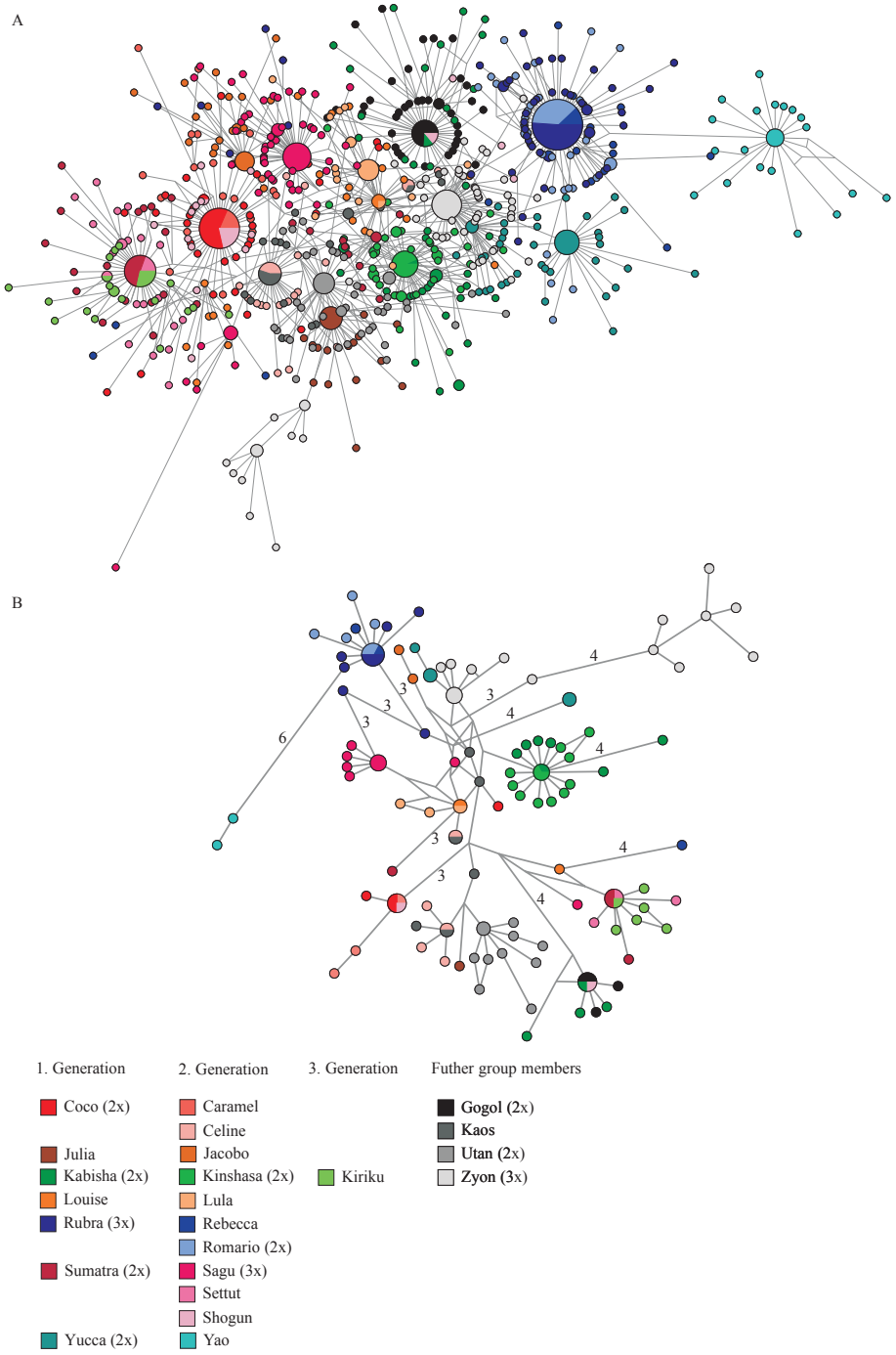

Figure 20: **Networks of all SFV sequences** (dataset B). A/Comprehensive version including all sequences generated; B/Streamlined version. Only those SFV sequences less likely to comprise "noise", *Taq*-induced mutations, were included. The legend illustrates matrilineal lines included in this dataset: mothers to the left, followed by their offspring (starting with the first generation) as well as additional group members. When individuals were repeatedly sampled, the number of samples are given between brackets in the legend. Within the networks, node size is proportional to the frequency of sequence occurrence. Branch lengths are directly related to the number of mutations between sequences [total length of aligned sequences: 432 bp (A), 426 bp (B)]. Please note that the streamlined network still comprise all but two of the shared SFV sequences, the two exceptions consisting of sequences closely related to other shared sequences (1 bp difference). Dataset reduction therefore did not result in any loss of information about SFV transmission.

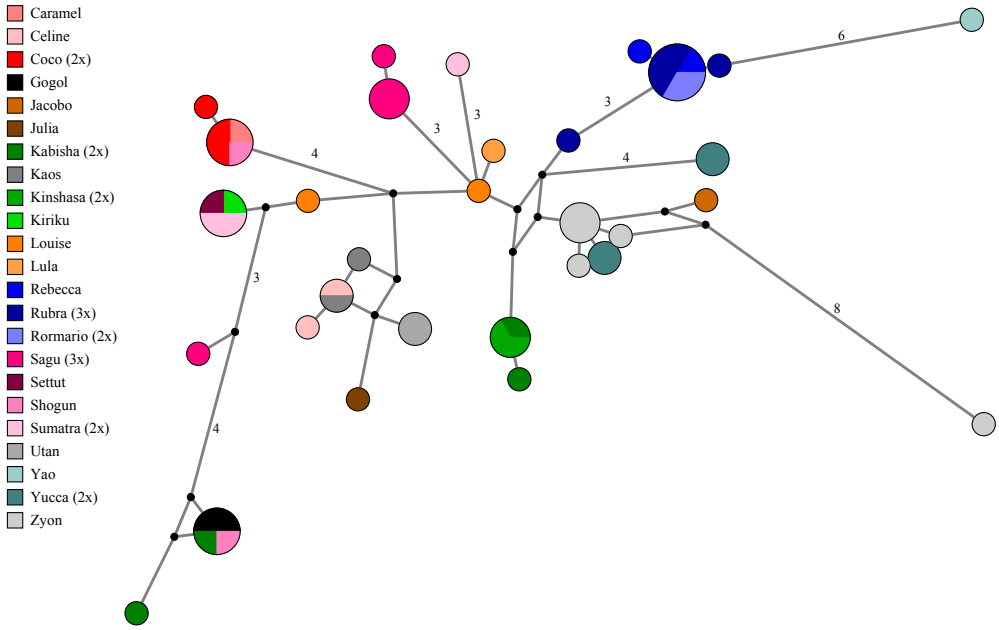

Figure 21: **Network of founder sequences** (n = 55). Node size is proportional to the frequency of sequence occurrence in the dataset. Branch lengths are directly related to the number of mutations between sequences, with values noted for differences greater than two base pairs. Legend conventions are the same as in Figure 20.

4.2.3 Transmission patterns of simian foamy virus

A number of individuals harbored identical SFV sequences/strains: 16/23 individuals did, 7 times with one other individual and 3 times with two other group members (Figure 20A and B). Among these, identical sequences were shared by six mother-offspring dyads (Figure 22). A GLMM revealed that similarity between the SFV strains found in samples was clearly influenced by the relatedness of the two respective individuals (likelihood ratio test comparing full model with null model not comprising relation or its interaction with time spent together but with all other terms present in the full model: $\chi^2 = 37.9$; df = 4; $p = 0.001$). No interaction was detected between relation and time spent together in the community ($\chi^2 = 0.1193$; df = 2; $p = 0.94$). After removing this interaction a clearly significant impact of the main effect relatedness was found ($\chi^2 = 37.8$; df = 2; $p = 0.0001$). Visual inspection of the data revealed that offspring SFV pools exhibited much more similarity to their mothers' than to their fathers' or unrelated individuals' (Figure 23). None of the other predictors nor their interactions (time spent together and difference between birthdays, time spent together, and sex of the other) appeared to significantly influence similarity measures (smallest $p = 0.14$).

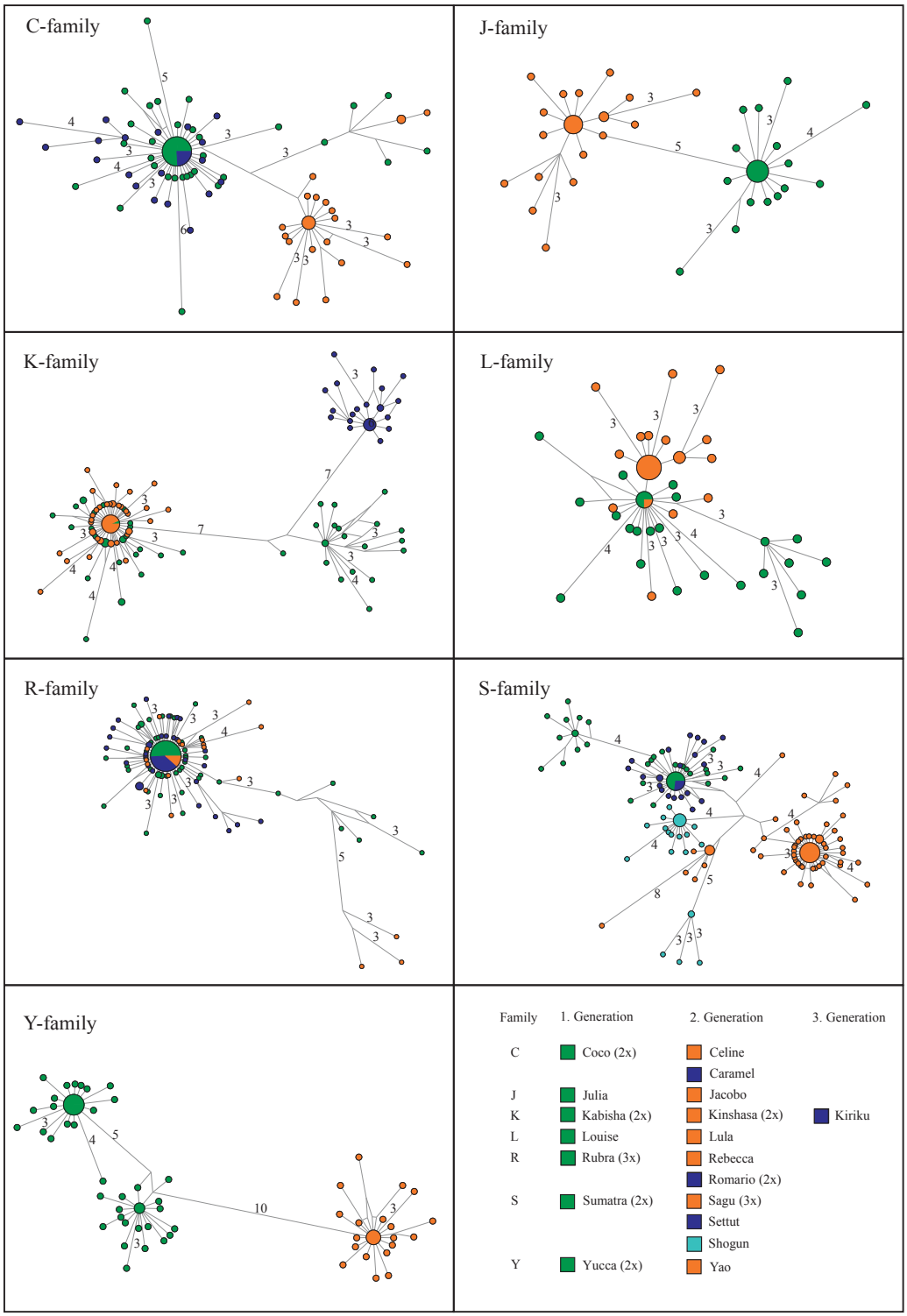

Figure 22: **Networks for all mother-offspring pairs.** The bottom right illustrates matrilineal lines included in dataset B. The column family marks the individuals belonging to the family with the same initial (letter). Legend conventions are the same as in Figure 20. Within each network, node size is proportional to the frequency of sequence occurrence. Branch lengths are directly related to the number of mutations between sequences, with values noted for differences greater than two base pairs (total length of aligned sequences: C, L, Y family, 425 bp; J, K family, 426 bp; R, S family, 428 bp).

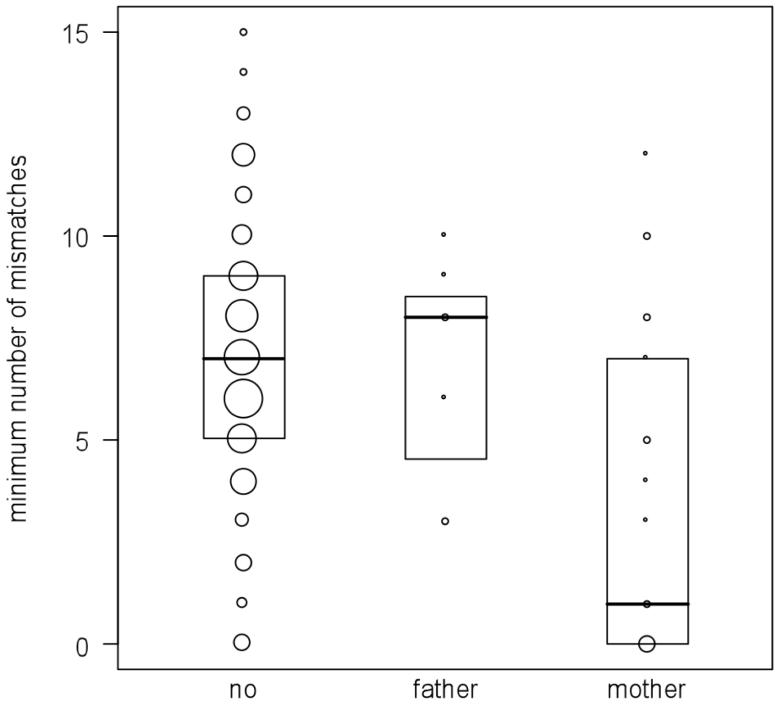

Figure 23: **Simian foamy virus similarity as a function of host relatedness.** Circle size represents the number of dyads exhibiting the corresponding relationship with the offspring (dataset B, Table 3).

4.2.4 Persistence of simian foamy virus

Nine individuals included in this study were sampled at multiple time points. Altogether, this provided a record of 42 years of chimpanzee lifetime and 57 viral years (for a definition of viral years please see Materials and Methods). Over this period, nearly identical sequences (exhibiting < 3 bp differences, *i.e.*, 0.7 % divergence) were found in consecutive samples spanning 55 viral years. Hence, 96 % of viral years were characterized by a close-to-perfect SFV sequence stability (Table 15).

Table 15: **Persistence of simian foamy virus.** Individuals for which multiple samples were available were tested for virus stability. Strains identified in the earlier samples were compared with strains found in the later corresponding sample. A virus strain was defined using network-based analysis [identification of the founder sequence(s)/group of closely related sequences in the absence of the founder sequence]. Strains were defined as stable in the consecutive sample(s) if sequences revealed a maximal distance of 0.7 % observed divergence (up to 3 bp difference). Of note, estimation of the number of independent infection events is a minimum estimate.

Individual	Age at sampling (years)	Infection status	Number of virus strains	Number of virus strains transmitted	Time between samplings (years)	Virus years based on number of strains and time between samplings	Virus stability (years)
Gogol	10	single					
Gogol	17	single	1	1	7	7	7
Kabisha	24	super					
Kabisha	25	super	2	2	1	2	2
Kinshasa	11	single					
Kinshasa	17	single	1	1	6	6	6
Rubra	31	single					
Rubra	35	super	1	1	4	4	4
Rubra	38	super	2	2	3	6	6
Sagu	13	single					
Sagu	17	super	1	1	4	4	4
Sagu	19	super	2	1	2	4	2
Sumatra	38	single					
Sumatra	40	super	1	1	2	2	2
Utan	7	single					
Utan	11	single	1	1	4	4	4
Yucca	32	super					
Yucca	34	super	2	2	2	4	4
Zyon	37	super					
Zyon	42	super	2	2	5	10	10
Zyon	44	super	2	2	2	4	4
Total					42	57	55
Persistence (%)							96

4.2.5 Accumulation dynamics of simian foamy virus

The design of this study included investigations of the accumulation of SFV strains through time based on the preliminary discrimination between single and super-infection cases. Most infected infants (0 to 4 year old), juveniles (5 to 9), and adolescents (10 to 14) were singly infected, while most adults showed evidence for super-infection (Figure 24). Correcting for multiple sampling of individuals, that is, considering only one sample per individual when consecutive samples had the same status within the age class, revealed that 3/12 non-adults and 9/11 adults were super-infected. Correspondingly, a GLMM revealed that super-infection status was clearly influenced by age, sex, and their interaction (likelihood ratio test comparing the full with the null model comprising only the random effect subject identity: $\chi^2 = 14.76$; df = 3; $p = 0.002$). As there was no interaction between age and sex (estimate + standard error [SE] = 0.258 + 0.169; z = 1.522; $p = 0.128$), it was removed from the model. The final model confirmed that super-infection was significantly more likely with increasing age (estimate + SE = 0.110 + 0.047; z = 2.337; $p = 0.019$) but did not support any obvious effect of sex (estimate + SE = -0.815 + 1.050; z = -0.776; $p = 0.438$) (Figure 24).

The cases of the 9 chimpanzees for which multiple samples were available also allowed for a longitudinal investigation of the dynamics of super-infection. Out of the six chimpanzees, which were initially single infected, a minimum of three super-infection events affecting Sagu, Rubra, and Sumatra, could be identified (Figure S4). Confirming the trend derived from the cross-sectional analysis, these super-infection events were recorded close to or during adulthood, between 13 and 19, 31 and 35, and 38 and 40 years of age, respectively. Of special note, two of these super-infections actually occurred in relatively old individuals (30 years old). Among individuals identified as super-infected over multiple sampling points, marked frequency shifts were observed in the number of clones belonging to one or the other clouds of sequences in two cases for which consecutive super-infected samples were available. For both Yucca and Kabisha, the smaller cloud of sequences doubled its relative size between the two samplings, with frequencies rising from 28 % to 56 % and from 40 % to 80 %, respectively (Figure 25).

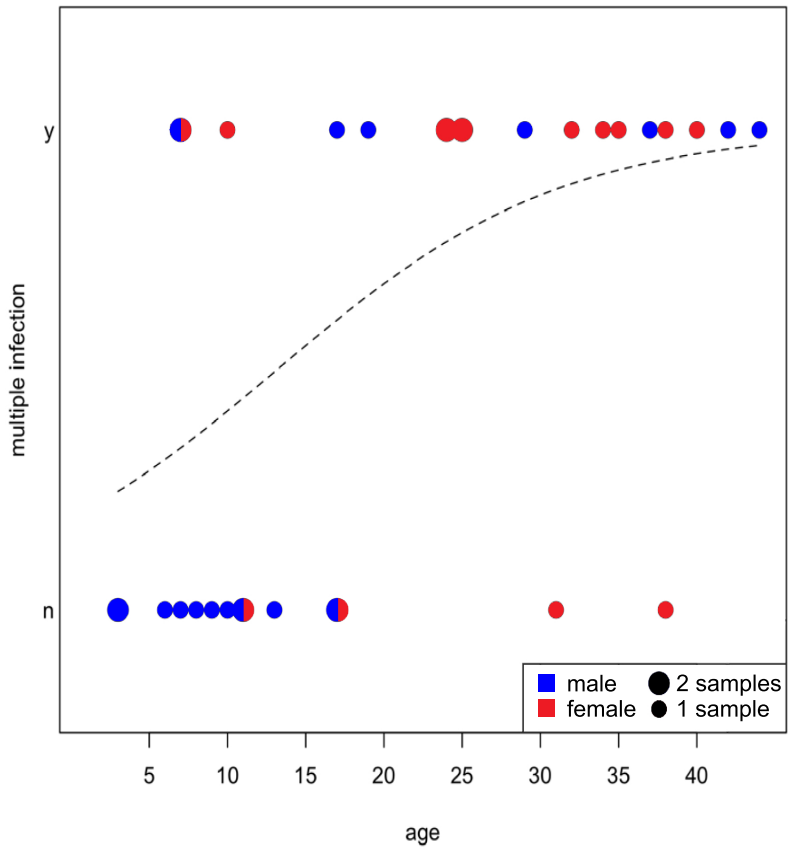

Figure 24: **Simian foamy virus accumulation dynamics.** Super-infection status (no/yes) is shown as a function of age. Circle size represents the number of samples at the corresponding combination of age and infection status; colors indicate the sex of the individuals (dataset B). The dashed line indicates the fitted model's prediction.

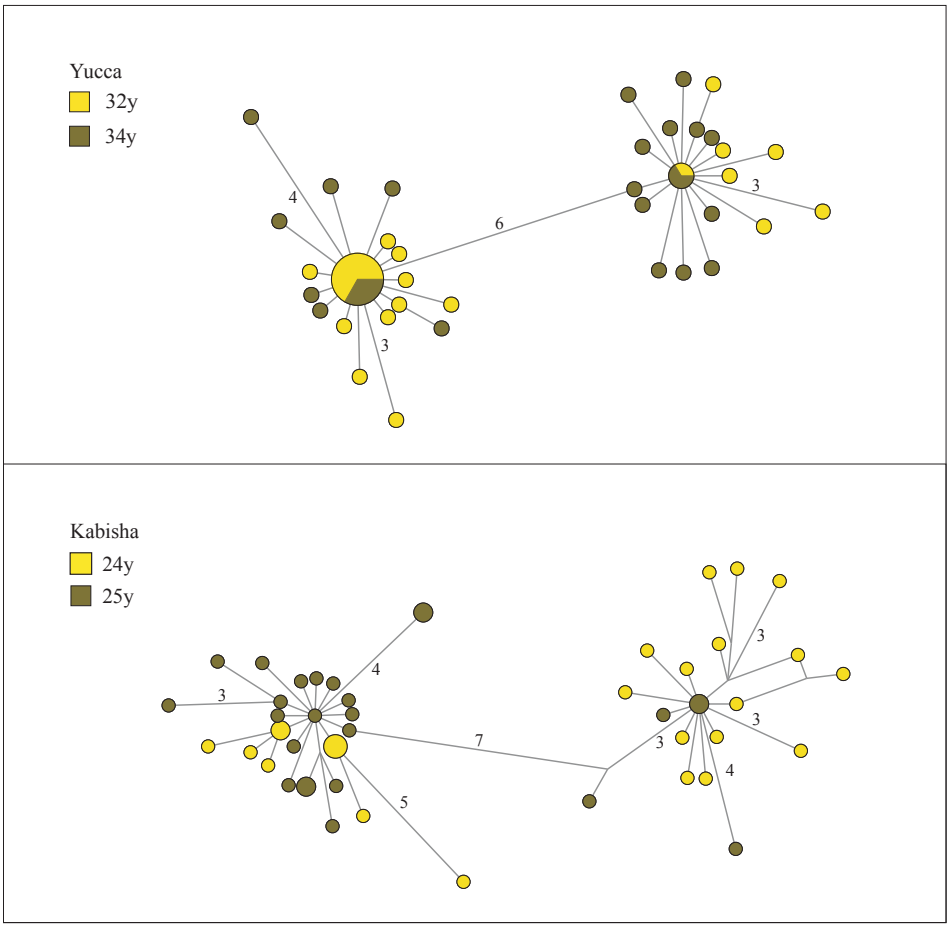

Figure 25: **In-host population dynamics of simian foamy virus.** Chimpanzees Yucca and Kabisha were sampled at different time points (y, years of age). Within the networks, node size is proportional to the frequency of sequence occurrence. Branch lengths are directly related to the number of mutations between sequences, with values noted for differences greater than two base pairs (total length of aligned sequences: 425 bp).

4.2.6 Molecular evolution of simian foamy virus

All clone sequences and the 55 founder sequences were examined for mutations indicating possibly defective viruses (*i.e.*, frameshift-inducing indels and in-frame stop codons). While 14.1 % of clone sequences were "defective," the proportion dropped to 1.8 % for founder sequences (Table 16), providing additional support in favor of these sequences' biological relevance. For founder sequences, variation occurred more often in the third codon position (mostly degenerate; 22.5 %) than in the first and second codon positions (mostly non-degenerate; 6.7 %, supporting an overall trend to purifying selection on this fragment (Table 16). Point mutations (other than nonsense) were also investigated to detect possible selective effects exerted by chimpanzee antiviral mechanisms, in particular through APOBEC proteins' editing activity (G-to-A

mutations). Analysis of the first mutational step away from the founder sequence(s) identified in individual networks supported a strong dominance of transitions (82 %) versus transversions (18 %). Among the former, G-to-A mutations were not particularly frequent (11 %; Table 17).

Table 16: **Defective sequences and variable sites.** All clone sequences and the 55 founder sequences (dataset B) were examined for mutations indicating possibly defective viruses and variable sites [number/total number (%)]. *Variable sites were calculated based on apparently non-defective viral sequences.

Group	Defective viruses		Variable sites*	
	Indels	Stop codons	1^{st}+ 2^{nd} codon position	3^{rd} codon position
All clones	90/915 (9.8 %)	39/915 (4.3 %)	217/283 (76.7 %)	121/142 (85.2 %)
Founder sequences	1/55 (1.8 %)	0/55 (0 %)	19/283 (6.7 %)	32/142 (22.5 %)

Table 17: **Nucleotide substitution matrix.** Point mutations (n = 471) occurring as the first step away from founder sequences were recorded for all individual parsimony-based network TCS networks (dataset B). Mean and standard deviation (SD) of percentages are shown. Transitions are shown in gray boxes.

Initial/ final state	Mean (SD) in %			
	A	C	T	G
A	-	2 (4)	6 (7)	28 (11)
C	1 (3)	-	7 (8)	0 (2)
T	5 (6)	36 (13)	-	2 (5)
G	11 (10)	1 (2)	1 (3)	-

5 Discussion

5.1 Evaluation of super-infection diagnostics

5.1.1 End-point dilution PCR

The first objective addresses the question of the applicability of EPD-PCR when samples only contain minute quantities of the targeted retrovirus/provirus, as present in fecal samples. Here it could be demonstrated that in such cases the amount of biological material required for proper EPD-PCR analysis can largely exceed the requirements of "classical" standard bulk-PCR assays (with a ca. 30-fold increase). This is in sharp contrast with applications of EPD-PCR on samples exhibiting high retroviremia, where requirements in undiluted cDNA might often be comparable to those of a single bulk-PCR assay. The magnitude of the increase poses a serious problem when one considers that limitation of biological material is an inherent characteristic of samples obtained from wild and endangered animal populations [77]. In these conditions, biological material used has to be optimized so as to allow investigating the biology and evolution of the broadest possible spectrum of pathogens. As an example, one could cite the recent use of the same collection of fecal samples from wild-living chimpanzees for the investigation of host [11], SIV [67], SFV [89] and Plasmodium [88] genetics. The implementation of EPD-PCR for the investigation of super-infection might therefore be unreasonable where "native" end-point dilution of pathogen DNA/RNA is to be encountered. In such situations, careful alternatives, which would still allow for a fair depiction of the underlying retroviral/proviral populations, should be favored.

There are two main alternatives to EPD-PCR analysis: bulk-PCR product cloning and sequencing and next generation sequencing. Though the latter strategy would be less labor-intensive and possibly even cheaper [110], it can be likely considered that highly redundant information will be produced where bulk-PCR amplification starts from only a few template molecules. Therefore focusing on bulk-PCR product cloning and sequencing, was considered appropriately scaled to the objectives of this study.

5.1.2 Bulk-PCR product cloning and sequencing

Here it could be shown that bulk-PCR product cloning is more parsimonious than EPD-PCR, even when gathering several bulk-PCR products per sample. Thus, in the present conditions, producing five bulk-PCR products required *ca.* 4-times less material than performing an equiv-

alent EPD-PCR analysis. Interestingly, using a comparable amount of material, EPD-PCR analyses would have resulted in gathering only 4 sequences, in which case three out of the six aforementioned cases of super-infection, namely B1, T1 and T2, would have faced a high probability of going unnoticed: 56 %, 29 % and 56 %, respectively. This clearly demonstrates that re-scaling EPD-PCR so as to spare material would significantly alter the probability to detect super-infection.

However, bulk-PCR product cloning is itself known to yield a biased view on within host retrovirus/provirus populations, at least when a detailed view is needed (*e.g.*, stoechiometries of viral genomes will likely be significantly skewed; [134]).

Here, the aim was to establish a coherent analytical framework likely to identify super-infection cases and the respective underlying main strains from bulk-PCR product cloning results, even where phylogenetic signal is low (Table 11).

The developed strategy was two-fold. First, a statistical method, which allows for the quick identification of super-infection cases from clone alignments had to be established. It could be shown here that using the shape of the mismatch distribution from clone alignments is a worthy approach. Using clones stemming from five independent bulk-PCR products, a satisfying 90% success rate, *i.e.,* rate of agreement with EPD-PCR results (T3 was the only case with disagreement) was reached, notably identifying the six super-infected individuals. In addition to yielding robust estimates of the expected results of an EPD-PCR experiment, the implementation of such a statistic also has the advantage to allow for a quick assessment of the number of samples needed to detect super-infection. Here, it can be suggested that using clones derived from fewer PCR products ($n \geq 3$) would still have allowed for reaching a high success rate (100%) but at the cost of a rather large false positive rate (up to 60 %; Figure 15). This high false positive rate is however entirely dependent on the inclusion of a single individual, T3, since running the same analysis with sequences stemming from the three other individuals results in a 0 % false positive rate (when ≥ 3 PCR products are considered). Interestingly, T3 was identified as single-infected by EPD-PCR analysis, while the global, five bulk-PCR product-based alignment rather supported a super-infection (the only faulty assignation in our sample). As the "second" strain evidenced by bulk-PCR cloning results only appears in one of the PCR products, T3 might well be a case where bulk-PCR cloning actually revealed a minor variant whose presence went undetected by EPD-PCR. Accordingly, the abovementioned can be interpreted as high false positive rate as likely resulting from a sampling artefact. All in all,

this case study therefore illustrates well how a careful use of the statistical approach could help in ascertaining the most efficient bulk-PCR based sampling method, *e.g.*, at the very beginning of larger projects.

The second step of the strategy consisted of identifying biological sequences with a particular focus on founder sequences, here understood as sequences having been at the origin of the infection or as sequences which ultimately became main components of the overall retroviral population. One of the most severe drawbacks of a bulk-PCR based method when compared to EPD-PCR lies in the fact that, contrary to EPD-PCR sequences, many clone sequences are expected to comprise artifactual mutations [134]. The ability to identify biological sequences from *Taq*-modified ones would further support the use of bulk-PCR derived clone alignments. Theoretically, in-host evolution of retroviruses/proviruses is likely to be well captured by network analyses, which should point at such sequences (assuming they have been sampled). It could be shown here that, in addition to providing an excellent visual support for the analysis of retroviral/proviral evolution, networks can also be used to identify truly biological (as opposed to *Taq*-error modified) founder sequences, through the use of OPs (as implemented in TCS; [24]). In this study, the highest OPs in clone networks always pointed at founder sequences, which were supported by EPD-PCR results, and, where super-infection was assumed, always identified the two founder sequences, when they had been sampled. Importantly, it could also be shown that network/OP-based analysis actually supersedes replication-based identification of founder sequences, notably exhibiting good performance where no replicated sequence is available. This good performance is in striking contrast with the results that would have provided the only possible alternative method, *i.e.*, phylogenetic analyses. Though it is sometimes argued that in phylogenetic trees ancestral sequences should appear as short branches basal to the longer branches of their descendants [122], it is indeed very unlikely that *Taq*-modified sequences will ever be separated from the biological sequence from which they are derived by a branch receiving strong statistical support (individual datasets only comprised weak phylogenetic information; Table 11). Therefore, networks and TCS-calculated OPs provide a unique, and so far under-explored, opportunity to identify founder sequences out of the *Taq*-induced noise.

The results obtained on simulated triple infections (no triple infection case could be identified from the experimental dataset) further support the robustness of the two-step algorithm. Mismatch distributions derived from simulations are indeed always a better fit to a bimodal dis-

tribution model, suggesting that the statistical tool actually detects deviation from a unimodal distribution model. On condition that it is known that super-infection involves more than two major type of sequences (which will require visual inspection of the networks), TCS-assessed OPs are also very efficient in identifying more than two likely founder sequences.

Nevertheless, it should be kept in mind that the method discussed here, even though all in all robust, remains bulk-PCR-based. As such, it is potentially heavily influenced by PCR- and/or sub-cloning-induced selective biases. The latter could presumably influence the results in both directions, *i.e.,* lead to false negatives when sequences of one of the infecting group of strains will be preferentially amplified and/or false positives when sequences of one minor infecting group of strains will be preferentially amplified. Selective biases are however unlikely to result in the ultimate identification of super-infection where individuals are truly single infected, *i.e.,* where no minor super-infection occurs. The main problem would therefore lie in accurately determining the frequency of biologically significant super-infection events, such as those resulting in several relatively distant strains becoming quantitatively important players in the overall retroviral population. This could however be corrected by applying the same approach to several genomic markers at one and the same time, since independent primer pairs are unlikely to result in congruent selective biases.

5.2 Transmission modalities of simian foamy virus

Using the above discussed method, the biology of SFV circulation within the Taï community could be investigated. The results provide unambiguous evidence that vertical transmission is an important modality of transmission in wild chimpanzees. Interestingly, this fits well with the assumed non-pathogenicity of SFVs. Because of their reliance on their host reproductive success, vertically transmitted microorganisms are indeed expected to evolve reduced virulence [145]. A consequence of reduced virulence is that microorganisms' selective pressure on their host decreases; that is, a very efficient immune response to the infection is not a selective advantage anymore. In this part of the study, two additional points are consistent with such a co-evolved "taming" process: (i) SFVs infecting Taï chimpanzees do not seem to undergo diversifying selection, as expected under strong immune pressure (but note that the fragment examined here encodes the *integrase*), and (ii) APOBEC editing does not seem to influence strongly these SFVs' mutational patterns (in contrast to previously published results obtained *in vitro*; [25]). It should, however, be noted that while mother-offspring SFV transmission is strongly supported here, other studies led on captive primates did not identify the same trend [12, 22], which might suggest that SFVs infecting distinct primate species are characterized by distinct transmission traits. The very long co-speciation history of SFV and their hosts would certainly have allowed for such divergence in SFV transmission patterns [148].

It is for the moment only possible to speculate about the possible medium or media supporting mother-offspring transmission. Mother-offspring dyads face numerous opportunities for body fluid exchanges. This includes intrauterine, perinatal, and/or breast-feeding-mediated transmission as observed for other exogenous retroviruses (*e.g.*, deltaretroviruses, lentiviruses), saliva-saliva contacts through fruit sharing, and saliva-blood contact through grooming on wounds or bites and nips received by infants during weaning. Although saliva has not yet been proven to be the main shedding site of SF viral particles in chimpanzees (in contrast to findings for other primates; [36, 103]), transmission of SFV from chimpanzees to humans following severe bites [9, 99, 148] and SF viral particle detection from chimpanzee fecal samples [40, 89] qualify it as the most likely medium for effective transmission of SFV among chimpanzees.

While it is shown here that mother-offspring transmission of SFV occurs frequently and therefore likely stands as a privileged route for primary infection, it is worth noting that not all siblings were infected with one of their mother's SFVs. This clearly points at the complexity of mother-offspring transmission patterns, which most likely depend on a variety of

parameters, *e.g.*, variation in maternal viral load and shedding as well as immune status of the offspring. Adding to the complexity of mother-offspring SFV transmission is the complexity of contemporary or subsequent horizontal transmissions. Assuming that saliva is the driving medium for infection, numerous opportunities of transmission will indeed also exist outside mother-offspring relationships. In this respect it is interesting to note that in this study father-offspring relationships did not measurably influence the circulation of SFVs. This does not come as a surprise, as in chimpanzees father-offspring relationships are notoriously weaker than mother-offspring relationships [15, 42, 107, 158]. Other relationships, *e.g.*, friendships, might be influential but could not be investigated here.

Whatever the quality of the relationships supporting their transmission, the data are also strongly supportive of a lifelong accumulation of SFV strains in chimpanzees. Some of the results support the possibility of significant SFV population shifts after super-infection and a natural extension of this is the possibility that some SFVs get extinct, *e.g.*, through the effects of genetic drift. Population shifts have recently been even identified in dually SFVmac infected humans during serial sampling of the variable *gag* gene [31]. However, a generally high persistence rate was recorded for the chimpanzee community investigated here based on the highly conserved *integrase* gene. Lifelong persistence of SFV is therefore likely to be the rule in wild chimpanzees, which mirrors observations made on a similar time series and similar gene region of samples taken from captive *Macaca tonkeana* [22]. In addition, the data provide very strong evidence that SFV super-infection becomes more common with increasing age, which suggests continuous acquiring of new strains through horizontal transmission events. Therefore, well-known pattern of increasing sero- and/or genoprevalence with age found in other chimpanzee subspecies and primates [22, 89] may hide very complex SFV dynamics as well.

An important question is that of the behaviors supporting horizontal transmission and therefore lifelong accumulation of SFVs, beyond a likely primary infection with mothers' SFVs. It has been hypothesized that truly violent aggressive behaviors, that is, those behaviors resulting in severe wounds providing opportunity for body fluid exchange, could fulfill the requirements for SFV transmission [22]. In primates in general and in chimpanzees in particular, such aggressive behaviors usually involve subadults and adults, for either the defense of social rank, territory, or access to food or reproductive partners. Interestingly, the onset of these behaviors fits well with our observation that super-infection events mainly occur during adulthood. The finding that SFV accumulation is not influenced by sex might at first seem somewhat con-

tradictory as males more frequently commit aggressive behaviors. However, it should not be considered that only males get involved into aggressive contacts or that aggressive contacts are systematically male-male or female-female events. For example, in Taï National Park, rates of male aggression toward females approach 0.08 per hour of female observation; that is, any female will be attacked by a male every 2 days [146]. During these frequent aggressive attacks, opportunities exist for both male-to-female and female-to-male SFV transmission since females also reply aggressively. Taken together with effective vertical transmission, this contributes to making marked sex differences in SFV transmission unlikely, although slight differences cannot be ruled out.

Another important factor that will influence SFV accumulation in chimpanzees is life expectancy. It is for example known that chimpanzees in Taï National Park have a relatively short life expectancy, with only 20 % individuals reaching adulthood and less than 5 % living to be older than 25, compared to 40 % individuals being expected to live at least 25 years in Gombe, Tanzania or Kibale, Uganda, for example *P. t. schweinfurthii*; [64]. Therefore triple or quadruple SFV infections could be expected to be more common in the latter subspecies/communities. De facto, no unambiguous triple or quadruple SFV infections were detected for in *P. t. verus* from Taï National Park, even for those individuals for which several samples collected during adulthood were identified as super-infected (n = 5). On the contrary, a quadruple infection was already reported in a wild Central chimpanzee (*P. t. troglodytes*, age unknown; [89]). It should however be noted that the ability to detect triple or quadruple infections depends on local SFV genetic diversity – the more divergent from each other are the strains circulating in a community, the more likely complex mixes of strains will be detected – and on the methodology deployed – depth of population description, analytic pathway, etc. Therefore, no clear-cut conclusion can be drawn at that stage.

5.3 Conclusions and outlook

This study presents an overall methodology that could stand for an acceptable alternative to EPD-PCR analysis for studies which involve the detection of retroviral super-infections from non-invasive samples. SFVs have the slowest evolutionary rates of all retroviruses [148], which should result in making the task of identifying super-infection harder than for other members of this viral family. The fact that this method is efficient in unveiling SFV super-infection instances thus offers a good perspective for its implementation for studies focused on other primate retroviruses such as STLV-1 and SIV. So far, this method was successfully applied to one community of chimpanzees in the Taï National Park to extensively study the biology of SFV in the wild. Super-infections with species-specific SFV could be frequently detected which have only been reported incidentally a couple of times, also in the case of chimpanzees [89]. Such a high preliminary estimate is both an expected result, since SFV reach very high prevalence rates in the species [89], and a puzzling finding as SFV, like other retroviruses, have evolved mechanisms aimed at restricting super-infection [105]. It clearly underlines how little we know about retroviral super-infection in the wild, and consequently, the necessity for further studies to address this question. The focus on in-host SFV diversity finally enabled for the development of a new model of SFV transmission in wild chimpanzee communities.

In this model (Figure 26), primary infection with SFV would likely occur through the mother-offspring relationship, a strong social bond, which is a major route of SFV circulation in this community. Subsequent SFV infections (that is, super-infections, since SFVs persist in wild chimpanzees) would occur during adulthood, possibly as a consequence of the onset of aggressive interactions with other members of the group. Within this framework, different strains of SFV might experience different population dynamics within individuals.

Further investigation of SFV transmission patterns in the wild are clearly needed. For chimpanzees, studies investigating the influence of other social bonds, such as other forms of kinship or friendship, would be highly desirable. Determining likely modes of SFV transmission in other wild NHP species would also be a welcome addition and would help determine whether the model of transmission proposed here for chimpanzees is applicable across parts or the entire phylogenetic tree of NHP and therefore whether SFV have maintained their transmission strategies over literally ten to hundred Mya. This could be further investigated by extending such studies to other tetrapods. This will also require gathering considerable amounts of behavioral and epizootic data. It might also provide useful information about possible patterns of FV

transmission within new host species, including humans.

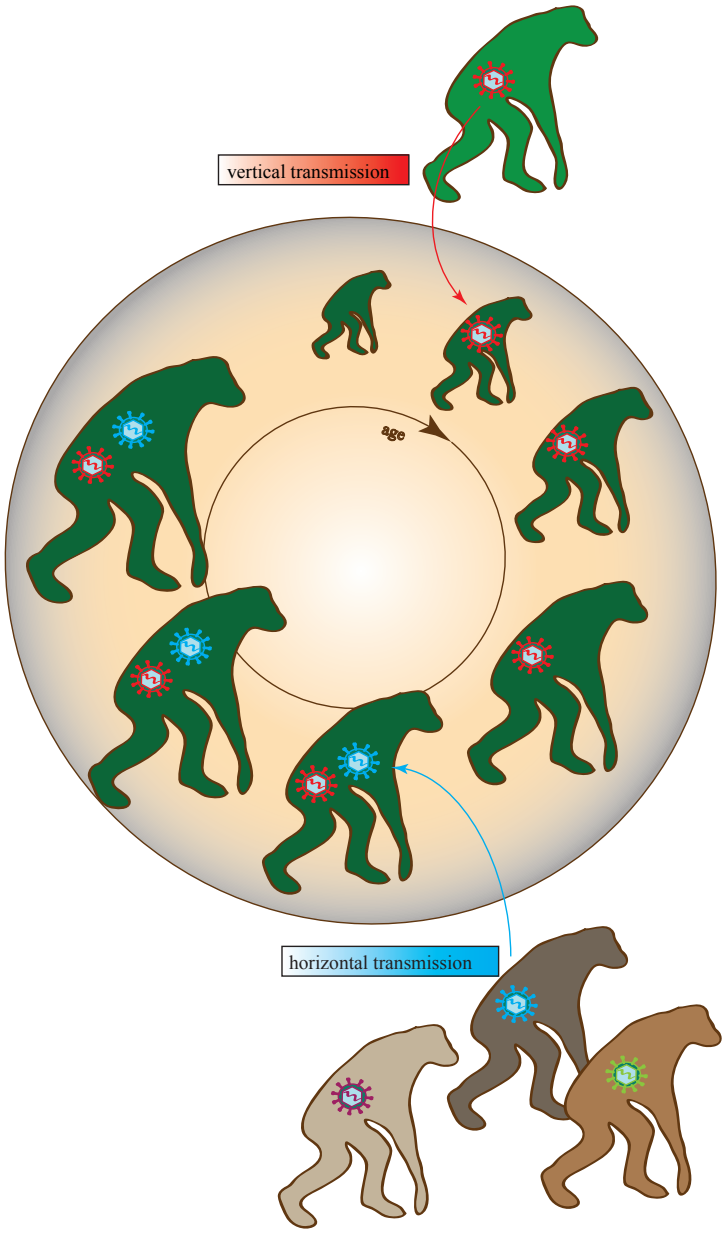

Figure 26: **Model of simian foamy virus (SFV) transmission.** The circle includes a chimpanzee (dark-green) becoming persistently infected with SFV (color corresponds to the SFV strain). The chimpanzee receives the primary infection from its mother (lime-green) due to their close social bond (vertical transmission). The individual receives subsequent infections (*i.e.* super-infections) from other group members during adulthood, possibly as a consequence of the onset of aggressive interactions.

6 Summary

Modalities of Transmission of Simian Foamy Virus in Wild Chimpanzees

(Anja Blasse)

The discovery of the zoonotic potential of retroviruses infecting non-human primates (NHPs) aroused the interest in simian retroviruses such as simian immunodeficiency virus, simian T-cell leukemia virus or simian foamy virus (SFV). However, investigations of retroviral circulation in wild primate communities are rare. Of particular interest are the transmission modalities of SFV, obviously successful strategies, which have led to frequent, persistent infections of NHPs for at least 40 million years. This work is an attempt to determine these modalities in wild chimpanzees from the Taï National Park, Côte d'Ivoire, where SFVs are highly endemic. These analyses presume a fine diagnostic tool to estimate the SFV diversity within each host and identify super-infection cases,*i.e.*, the simultaneous infection of the same individual host with several strains of the same virus. So far, methods possibly allowing such investigations from samples collected non-invasively (such as feces) have never been properly compared. Therefore, the costs and benefits of the gold standard (end-point dilution PCR, EPD-PCR) and multiple bulk-PCR cloning methods were assessed for SFV super-infections based on a case study using fecal samples of two different chimpanzee subspecies (*Pantroglodytes verus*and *P. t. schweinfurthii*). It could be shown that in these conditions EPD-PCR can lead to massive consumption of biological material. This constitutes a serious drawback in a field in which rarity of biological material is a fundamental constraint. In addition, data demonstrated that EPD-PCR results (single/multiple infection; founder strains) could be well predicted from multiple bulk-PCR clone experiments, by applying simple statistical and network analyses to sequence alignments. Therefore, the implementation of the latter method can be recommended, when the focus is put on retroviral super-infection and only low retroviral loads are encountered. Using this approach, SFV diversity was then estimated for each sample of the study community in Taï National Park to investigate dynamics of SFVs. The results indicate, that vertical transmission (being here understood as mother-offspring transmission) is a common route of SFV transmission within the community whereas previous studies so far only pointed at horizontal transmissions of SFVs. The strong bond between mother and offspring is likely to be responsible for primary infections. With increasing age subsequent infections with SFVs could be observed (super-infections). The

development of truly aggressive behavior during the onset of adulthood is hypothesized to result into frequent horizontal transmissions of SFVs between other members of the group. Finally, these data gives evidence for complex SFV dynamics in wild chimpanzees, even at a single community scale, and show that linking wild NHP social interactions and their microorganisms' dynamics is feasible.

7 Zusammenfassung

Vorkommen und Übertragung von Simianen Foamy Viren bei wildlebenden Schimpansen

(Anja Blasse)

Erst die Erkenntnis über das zoonotische Potential von simianen Retroviren, welche nichthumane Primaten (NHP) infizieren, lenkte das wissenschaftliche Interesse auf simiane Retroviren wie das Simiane Immunodefizienz Virus, das Simiane T-cell Leukämie Virus oder das Simiane Foamy Virus (SFV). Dennoch gibt es nur wenige Untersuchungen zur Verbreitung von Retroviren bei wildlebenden Primaten. Von besonderem Interesse sind dabei die Übertragungswege von SFV, welche offenbar eine so erfolgreiche Strategie verfolgen, dass sie bei NHPs seit über 40 Millionen Jahren verbreitet sind und diese persistent infizieren. Die vorliegende Arbeit verfolgt das Ziel die Übertragungswege bei wildlebenden Schimpansen aus dem Taï Nationalpark, Côte d'Ivoire, wo SFV endemisch vorkommt, aufzuzeigen. Solche Analysen setzen jedoch präzise diagnostische Methoden voraus, um die Diversität von SFV innerhalb des Wirtes und dadurch Superinfektionen, *i.e.,* ein Individuum ist parallel mit mehreren Stämmen eines Virus infiziert, zu bestimmen. Die zur Verfügung stehenden Methoden wurden jedoch bisher nicht für den Gebrauch von nicht-invasiv gesammeltem Probenmaterial (z.B. Kotproben) getestet. Daher wurden hier an einem Fallbeispiel mit Kotproben von zwei verschiedenen Schimpansensubspezies Vor- und Nachteile des Goldstandards (Endpunkt-Verdünnungs-PCR, EPD-PCR) und der klassischen PCR mit anschließender Klonierung zur Bestimmung von Superinfektionen verglichen. Dabei konnte gezeigt werden, dass unter diesen Konditionen EPD–PCR zu erheblichem Verbrauch von biologischem Material führt. Dies bedeutet eine große Einschränkung in Untersuchungsgebieten, wo biologisches Material nur in begrenztem Maße zur Verfügung steht. Weiterhin zeigen die Daten, dass Ergebnisse der EPD-PCR (Einfach-/ Superinfektion, Identifizierung von Gründersequenzen) auch mittels Experimenten der klassischer PCR und Klonierung erhoben werden können, soweit dabei die Sequenzalignments statistischen und Netzwerkanalysen unterzogen werden. Die Verwendung der letztgenannten Methode kann insbesondere empfohlen werden, wenn retrovirale Superinfektionen im Fokus der Untersuchung stehen und nur eine geringe Viruslast in den Proben vorliegt. Nach diesem Vorgehen wurde im Anschluss die SFV Diversität je Probe erfasst, um die Dynamik von SFV innerhalb der

Studiengruppe im Taï National Park zu untersuchen. Im Gegensatz zu bisherigen Studien konnte gezeigt werden, dass die vertikale Übertragung (hier definiert als Mutter-Kind-Übertragung) neben der horizontalen Übertragung eine große Rolle bei der Verbreitung von SFV spielt. Diese Übertragung ist sehr wahrscheinlich für die Primärinfektion mit dem Erreger verantwortlich, da Mutter und Kind zunächst eng miteinander verbunden sind. Mit zunehmendem Alter der Schimpansen wurden jedoch Infektionen mit weiteren SFV Stämmen beobachtet (Superinfektion). Diese sind vermutlich auf horizontale Übertragungen von SFV zwischen Gruppenmitgliedern durch aggressives Verhalten untereinander während des Erwachsenwerdens zurückzuführen. Insgesamt gibt es Hinweise auf eine sehr komplexe Dynamik von Foamy Viren bei wildlebenden Schimpansen bereits auf der Ebene einer einzelnen Schimpansengruppe. Es konnte außerdem gezeigt werden, dass die sozialen Interaktionen des Wirtes Einfluss auf die Dynamik der Mikroorganismen nehmen können.

8 References

A

[1] BG Achong, PW Mansell, and MA Epstein. A new human virus in cultures from a nasopharyngeal carcinoma. *J Pathol*, 103(2):18, 1971.

[2] SF Altschul, W Gish, W Miller, EW Myers, and DJ Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–410, 1990.

[3] M Anisimova and O Gascuel. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol*, 55(4):539–552, 2006.

B

[4] RH Baayen. *Analyzing linguistic data*. Cambridge University Press, Cambridge, United Kingdom., 2008.

[5] DN Baldwin and ML Linial. The roles of Pol and Env in the assembly pathway of human foamy virus. *J Virol*, 72(5):3658–3665, 1998.

[6] HJ Bandelt, P Forster, and A Rohl. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*, 16(1):37–48, 1999.

[7] P Bastone, U Truyen, and M Lo. Minireview potential of zoonotic transmission of non-primate foamy viruses to humans. *J Vet Med*, 423:417–423, 2003.

[8] D Bates, M Maechler, and B Bolker. lme4: linear mixed-effects models using S4 classes, R package, 2011.

[9] E Betsem, R Rua, P Tortevoye, A Froment, and A Gessain. Frequent and recent human acquisition of simian foamy viruses through apes' bites in central Africa. *PLoS Pathog*, 7(10):e1002306, October 2011.

[10] ORP Bininda-Emonds, M Cardillo, KE Jones, RDE MacPhee, RMD Beck, R Grenyer, SA Price, RA Vos, JL Gittleman, and A Purvis. The delayed rise of present-day mammals. *Nature*, 446(7135):507–512, March 2007.

[11] A Bjork, W Liu, JO Wertheim, BH Hahn, and M Worobey. Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. *Mol Biol Evol*, 28(1):615–623, 2011.

[12] EL Blewett, DH Black, NW Lerche, G White, and R Eberle. Simian foamy virus infections in a baboon breeding colony. *Virology*, 278(1):183–193, December 2000.

[13] J Bodem, M Löchelt, I Winkler, RP Flower, H Delius, and RM Flügel. Characterization of the spliced pol transcript of feline foamy virus: the splice acceptor site of the pol transcript is located in gag of foamy viruses. *J Virol*, 70(12):9024–9027, 1996.

[14] C Boesch and H Boesch-Achermann. *The Chimpanzees of the Taï Forest, Behavioural Ecology and Evolution*. Oxford University Press, Oxford, UK, 2000.

[15] C Boesch and H Boesch-Achermann. *Taï Forest*. Oxford University Press, New York, 2001.

[16] C Boesch, G Kohou, H Nene, and L Vigilant. Male competition and paternity in wild chimpanzees of the Taï forest. *Am J Phys Anthropol*, 130(1):103–115, 2006.

[17] JI Brooks, EW Rud, RG Pilon, JM Smith, WM Switzer, and PA Sandstrom. Cross-species retroviral transmission from macaques to human beings. *The Lancet*, 360(9330):387–388, 2002.

[18] SR Broussard, AG Comuzzie, KL Leighton, MM Leland, EM Whitehead, and JS Allan. Characterization of new simian foamy viruses from African nonhuman primates. *Virology*, 237(2):349–359, 1997.

[19] KP Burnham and DR Anderson. *Model selection and multi-model inference: a practical information-theoretic approach*. Springer-Verlag, New York, 2002.

[20] TM Butynski. The robust chimpanzee *Pan troglodytes* : taxonomy, distribution, abundance and conservation status. In M.I. Bakarr & T.M. Butynski R. Kormos, C. Boesch, editor, *Status Survey and Conservation Action Plan: West African Chimpanzees.*, pages 5–12. Switzerland, iucn, glan edition, 2003.

C

[21] S Calattini, EBA Betsem, A Froment, P Mauclère, P Tortevoye, C Schmitt, R Njouom, A Saïb, and A Gessain. Simian foamy virus transmission from apes to humans, rural Cameroon. *Emerg Infect Dis*, 13(9):1314–1320, 2007.

[22] S Calattini, F Wanert, B Thierry, C Schmitt, S Bassot, A Saïb, N Herrenschmidt, and A Gessain. Modes of transmission and genetic diversity of foamy viruses in a Macaca tonkeana colony. *Retrovirology*, 3:23, January 2006.

[23] S Chareza, Dragana Slavkovic L, Y Liu, AM Räthe, C Münk, E Zabogli, M Pistello, and M Löchelt. Molecular and functional interactions of cat APOBEC3 and feline foamy and immunodeficiency virus proteins: different ways to counteract host-encoded restriction. *Virology*, 424(2):138–146, March 2012.

[24] M Clement, D Posada, and KA Crandall. TCS: a computer program to estimate gene genealogies. *Mol Ecol*, 9(10):1657–1659, 2000.

D

[25] F Delebecque, R Suspene, S Calattini, N Casartelli, A Saïb, A Froment, S Wain-Hobson, A Gessain, JP Vartanian, and O Schwartz. Restriction of Foamy Viruses by APOBEC Cytidine Deaminases. *J Virol*, 80(2):605–614, 2006.

[26] O Delelis, J Lehmann-Che, and A Saïb. Foamy viruses–a world apart. *Curr Opin Microbiol*, 7(4):400–406, August 2004.

[27] AJ Dobson. *An introduction to generalized linear models*. Chapman and Hall/CRC, Boca Raton, USA, 2002.

[28] AJ Drummond and A Rambaut. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*, 7:214, January 2007.

E

[29] SW Eastman and ML Linial. Identification of a conserved residue of foamy virus gag required for intracellular capsid assembly. *J Virol*, 75(15):6857–6864, 2001.

[30] RC Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32(5):1792–1797, 2004.

[31] GA Engel, CT Small, K Soliven, MM Feeroz, X Wang, MK Hasan, G Oh, SMR Alam, KL Craig, DL Jackson, FA Matsen IV, ML Linial, and L Jones-Engel. Zoonotic simian foamy virus in Bangladesh reflects diverse patterns of transmission and co-infection. *Emerging Microbes and Infections*, 2, 2013.

[32] J Enssle, N Fischer, A Moebes, B Mauer, U Smola, and Rethwilm A. Carboxy-terminal cleavage of the human foamy virus Gag precursor molecule is an essential step in the viral life cycle. *J Virol*, 71(10):7312–7317, 1997.

[33] J Enssle, I Jordan, B Mauer, and A Rethwilm. Foamy virus reverse transcriptase is expressed independently from the Gag protein. *Proc Natl Acad Sci U S A*, 93(9):4137–4141, April 1996.

[34] J Enssle, A Moebes, M Heinkelein, M Panhuysen, B Mauer, M Schweizer, D Neumann-Haefelin, and A Rethwilm. An active foamy virus integrase is required for virus replication. *J Gen Virol*, 80 (6):1445–1452, June 1999.

[35] L Etienne, E Nerrienet, M LeBreton, GT Bibila, Y Foupouapouognigni, D Rousset, A Nana, CF Djoko, U Tamoufe, AF Aghokeng, E Mpoudi-Ngole, E Delaporte, M Peeters, ND Wolfe, and A Ayouba. Characterization of a new simian immunodeficiency virus strain in a naturally infected *Pan troglodytes troglodytes* chimpanzee with AIDS related symptoms. *Retrovirology*, 8(1):4, January 2011.

F

[36] V Falcone, J Leupold, J Clotten, E Urbanyi, O Herchenröder, W Spatz, B Volk, N Böhm, A Toniolo, D Neumann-Haefelin, and M Schweizer. Sites of simian foamy virus persistence in naturally infected African green monkeys: latent provirus is ubiquitous, whereas viral replication is restricted to the oral mucosa. *Virology*, 257(1):7–14, 1999.

[37] V Falcone, M Schweizer, and D Neumann-Haefelin. Replication of primate foamy viruses in natural and experimental hosts. *Curr Top Microbiol Immunol*, 277:161–180, 2003.

[38] A Field. *Discovering statistics with SPSS*. Sage Publications, London, United Kingdom, 2005.

G

[39] D Glez-Peña, D Gómez-Blanco, M Reboiro-Jato, F Fdez-Riverola, and D Posada. ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic Acids Res*, 38(Web Server issue):W14–18, July 2010.

[40] AS Goffe, A Blasse, R Mundry, FH Leendertz, and S Calvignac-Spencer. Detection of retroviral super-infection from non-invasive samples. *PLoS One*, 7(5):e36570, January 2012.

[41] TL Goldberg, DM Sintasath, CA Chapman, KM Cameron, WB Karesh, S Tang, ND Wolfe, IB Rwego, N Ting, and WM Switzer. Coinfection of Ugandan red colobus (*Procolobus [Piliocolobus] rufomitratus tephrosceles*) with novel, divergent delta-, lenti-, and spumaretroviruses. *J Virol*, 83(21):11318–11329, November 2009.

[42] J Goodall. *The chimpanzees of Gombe: patterns of behaviour*. Belknap Press, Cambridge, Massachusetts, 1986.

[43] M Gouy, S Guindon, and O Gascuel. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*, 27(2):221–224, 2010.

[44] J Grobler, CM Gray, C Rademeyer, C Seoighe, G Ramjee, SA Karim, L Morris, and C Williamson. Incidence of HIV-1 Dual infection and its association with increased viral load set point in a cohort of HIV-1 subtype C - infected female sex workers. *J Infect Dis*, 190:1355–1359, 2004.

[45] S Guindon, JF Dufayard, V Lefort, M Anisimova, W Hordijk, and O Gascuel. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*, 59(3):307–321, 2010.

[46] S Guindon, F Lethiec, P Duroux, and O Gascuel. PHYML Online–a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res*, 33(Web Server issue):W557–559, 2005.

H

[47] AJ Hackett, A Pfiester, and P Arnstein. Biological properties of a syncytia-forming agent isolated from domestic cats (feline syncytia-forming virus). *Proc Soc Exp Biol Med.*, 135(3):899–904, 1970.

[48] GZ Han and M Worobey. An endogenous foamy-like viral element in the coelacanth genome. *PLoS Pathog*, 8(6):e1002790, January 2012.

[49] GZ Han and M Worobey. An endogenous foamy virus in the aye-aye (*Daubentonia madagascariensis*). *J Virol*, 86(14):7696–8, July 2012.

[50] M Heinkelein, M Rammling, T Juretzek, D Lindemann, and A Rethwilm. Retrotransposition and cell-to-cell transfer of foamy viruses. *J Virol*, 77(21):11855–11858, 2003.

[51] W Heneine, WM Switzer, P Sandstrom, J Brown, S Vedapuri, CA Schable, AS Khan, NW Lerche, M Schweizer, D Neumann-Haefelin, LE Chapman, and TM Folks. Identification of a human population infected with simian foamy viruses. *Nat Med*, 4(4):403–407, 1998.

[52] I Herbinger, C Boesch, and H Rothe. Territory characteristics among three neighboring chimpanzee communities in the Taï National Park, Côte d'Ivoire. *Int J Primatol*, 22(2), 2001.

[53] O Herchenröder, R Renne, D Loncar, EK Cobb, KK Murthy, J Schneider, A Mergia, and PA Luciw. Isolation, cloning, and sequencing of simian foamy viruses from chimpanzees (SFVcpz): high homology to human foamy virus (HFV). *Virology*, 201(2):187–199, 1994.

[54] EC Holmes. The evolution of endogenous viral elements. *Cell Host Microbe*, 10(4):368–377, October 2011.

[55] JJ Hooks and CJ Gibbs. The foamy viruses. *Bacteriol Rev*, 39(3):169–185, September 1975.

[56] F Huang, W Yu, and Z He. Foamy virus in the tree shrew Tupaia belangeri is highly related to simian foamy virus in Macaca mulatta. *AIDS Res Hum Retroviruses*, 29(8):1177–1178, 2013.

J

[57] AP Jackson and MA Charleston. A cophylogenetic perspective of RNA-virus evolution. *Mol Biol Evol*, 21(1):45–57, January 2004.

[58] RH Johnson, J de la Rosa, I Abher, IG Kertayadnya, KW Entwistle, G Fordyce, and RG Holroyd. Epidemiological studies of bovine spumavirus. *Vet Microbiol*, 16(1):25–33, January 1988.

[59] L Jones-Engel, GA Engel, J Heidrich, M Chalise, N Poudel, R Viscidi, PA Barry, JS Allan, R Grant, and R Kyes. Temple monkeys and health implications of commensalism, Kathmandu, Nepal. *Emerg Infect Dis*, 12(6):900–906, June 2006.

[60] L Jones-Engel, CC May, GA Engel, KA Steinkraus, MA Schillaci, A Fuentes, A Rompis, MK Chalise, N Aggimarangsee, MM Feeroz, R Grant, JS Allan, A Putra, IN Wandia, R Watanabe, L Kuller, S Thongsawat, R Chaiwarith, RC Kyes, and ML Linial. Diverse contexts of zoonotic transmission of simian foamy viruses in Asia. *J Infect Dis*, 14(8):1200–1208, August 2008.

[61] L Jones-Engel, KA Steinkraus, SM Murray, GA Engel, R Grant, N Aggimarangsee, BP Lee, C May, MA Schillaci, C Somgird, T Sutthipat, L Vojtech, J Zhao, and ML Linial. Sensitive assays for simian foamy viruses reveal a high prevalence of infection in commensal, free-ranging Asian monkeys. *J Virol*, 81(14):7330–7337, July 2007.

[62] S Junglen, C Hedemann, H Ellerbrok, G Pauli, C Boesch, and FH Leendertz. Diversity of STLV-1 strains in wild chimpanzees (*Pan troglodytes verus*) from Côte d'Ivoire. *Virus research*, 150(1-2):143–147, June 2010.

K

[63] M Kaiser, A Löwa, M Ulrich, H Ellerbrok, AS Goffe, A Blasse, Z Zommers, E Couacy-hymann, F Babweteera, K Zuberbühler, S Metzger, S Geidel, C Boesch, TR Gillespie, and FH Leendertz. Wild Chimpanzees Infected with 5 Plasmodium species. *Emerg Infect Dis*, 16(12):16–19, 2010.

[64] PM Kappeler and CPV Schaik. Evolution of Primate Social Systems. *Int J Primatol*, 23(4):707–740, 2002.

[65] A Katzourakis, RJ Gifford, M Tristem, MT Gilbert, and OG Pybus. Macroevolution of complex retroviruses. *Science*, 325(5947):1512, September 2009.

[66] BF Keele, JH Jones, KA Terio, JD Estes, RS Rudicell, ML Wilson, Y Li, GH Learn, TM Beasley, J Schumacher-Stankey, E Wroblewski, A Mosser, J Raphael, S Kamenya, EV Lonsdorf,

DA Travis, T Mlengeya, MJ Kinsel, JG Else, G Silvestri, J Goodall, PM Sharp, GM Shaw, AE Pusey, and BH Hahn. Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature*, 460(7254):515–519, 2009.

[67] BF Keele, F Van Heuverswyn, Y Li, E Bailes, J Takehisa, ML Santiago, F Bibollet-Ruche, Y Chen, LV Wain, F Liegeois, S Loul, EM Ngole, Y Bienvenue, E Delaporte, JF Brookfield, PM Sharp, GM Shaw, M Peeters, and BH Hahn. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*, 313(5786):523–526, July 2006.

[68] T Kehl, J Tan, and M Materniak. Non-Simian Foamy Viruses: Molecular Virology, Tropism and Prevalence and Zoonotic/Interspecies Transmission. *Viruses*, 5(9):2169–2209, September 2013.

[69] AS Khan. Simian foamy virus infection in humans: prevalence and management. *Expert Rev Anti Infect Ther*, 7(5):569–580, June 2009.

[70] S Köndgen. *Detection and characterisation of respiratory pathogens among habituated, wild living chimpanzees (Pan troglodytes verus) of Taï National Park, Côte d'Ivoire*. Doctoral thesis, Technische Universität Berlin, 2011.

[71] S Köndgen, H Kühl, PK N'Goran, PD Walsh, S Schenk, N Ernst, R Biek, P Formenty, K Mätz-Rensing, B Schweiger, S Junglen, H Ellerbrok, A Nitsche, T Briese, WI Lipkin, G Pauli, C Boesch, and FH Leendertz. Pandemic human viruses cause decline of endangered great apes. *Curr Biol*, 18(4):260–264, February 2008.

[72] S Köndgen, S Schenk, G Pauli, C Boesch, and FH Leendertz. Noninvasive monitoring of respiratory viruses in wild chimpanzees. *Ecohealth*, 7(3):332–341, September 2010.

L

[73] KE Langergraber, K Prüfer, C Rowney, C Boesch, C Crockford, and K Fawcett. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *PNAS*, 109(39):15716 – 15721, 2012.

[74] B Le Guenno, P Formenty, M Wyers, P Gounon, F Walker, and C Boesch. Isolation and virus partial characterisation of strain of Ebola. *The Lancet*, 345:1271–1274, 1995.

[75] CH Lecellier, W Vermeulen, F Bachelerie, ML Giron, and A Saïb. Intra- and intercellular trafficking of the foamy virus auxiliary bet protein. *J Virol*, 76(7):3388–3394, 2002.

[76] FH Leendertz, H Ellerbrok, C Boesch, E Couacy-Hymann, K Matz-Rensing, R Hakenbeck, C Bergmann, P Abaza, S Junglen, Y Moebius, L Vigilant, P Formenty, and G Pauli. Anthrax kills wild chimpanzees in a tropical rainforest. *Nature*, 430(6998):451–452, 2004.

[77] FH Leendertz, G Pauli, K Maetz-Rensing, W Boardman, C Nunn, H Ellerbrok, SA Jensen, S Junglen, and C Boesch. Pathogens as drivers of population declines: The importance of systematic monitoring in great apes and other threatened mammals. *Biological Conservation*, 131(2):325–337, 2006.

[78] FH Leendertz, F Zirkel, E Couacy-Hymann, H Ellerbrok, VA Morozov, G Pauli, C Hedemann, P Formenty, SA Jensen, C Boesch, and S Junglen. Interspecies transmission of simian foamy virus in a natural predator-prey system. *J Virol*, 82(15):7741–7744, August 2008.

[79] SA Leendertz, S Junglen, C Hedemann, A Goffe, S Calvignac, C Boesch, and FH Leendertz. High prevalence, coinfection rate, and genetic diversity of retroviruses in wild red colobus monkeys (*Piliocolobus badius badius*) in Taï National Park, Côte d'Ivoire. *J Virol*, 84(15):7427–36, August 2010.

[80] J Lehmann-Che, M Giron, O Delelis, M Lo, P Bittoun, J Tobaly-Tapiero, H The, and A Saïb. Protease-dependent uncoating of a complex retrovirus. *J Virol*, 79(14):9244–9253, 2005.

[81] H Li, KJ Bar, S Wang, JM Decker, Y Chen, C Sun, JF Salazar-Gonzalez, MG Salazar, GH Learn, CJ Morgan, JE Schumacher, P Hraber, EE Giorgi, T Bhattacharya, BT Korber, AS Perelson, JJ Eron, MS Cohen, CB Hicks, BF Haynes, M Markowitz, BF Keele, BH Hahn, and GM Shaw. High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men. *PLoS Pathog*, 6(5):e1000890, May 2010.

[82] D Lindemann, T Pietschmann, A Berg, M Heinkelein, P Knaus, H Zentgraf, M Picard-Maureau, J Thurow, and A Rethwilm. A particle-associated glycoprotein signal peptide essential for virus maturation and infectivity. *J Virol*, 75(13):5762–5771, 2001.

[83] B Ling, ML Santiago, S Meleth, HM Mcclure, C Apetrei, B Gormus, BH Hahn, and doi = 10.1128/JVI.77.3.2214 file = :Users/corinna/Library/Application Support/Mendeley Desktop/Downloaded/Ling et al. - 2003 - Noninvasive Detection of New Simian Immunodeficiency Virus Lineages in Captive Sooty Mangabeys Ability To Amplify Virion RNA from Fecal Samples Correlates with Viral Load in Plasma Noninvasive Detection of New Simian Immunodefic.pdf:pdf journal = J Virol number = 3 pages = 2214–2226 title = Noninvasive detection of new simian immunodeficiency virus lineages in captive sooty mangabeys: ability to amplify virion RNA from fecal samples correlates with viral load in plasma volume = 77 year = 2003 Marx, PA.

[84] M Linial. Why aren't foamy viruses pathogenic? *Trends in Microbiology*, 8(6):284–289, June 2000.

[85] M Linial. *Foamy Viruses*. Knipe DM, Howley PM, Lippincott W and Wilkins MD, 5th edition, 2006.

[86] ML Linial. MINIREVIEW foamy viruses are unconventional retroviruses. *J Virol*, 73(3):1747–1755, 1999.

[87] ML Linial, H Fan, B Hahn, R Lwer, J Neil, S Quackenbush, A Rethwilm, P Sonigo, J Stoye, and M Tristem. *Index of Viruses - Retroviridae*, volume Classifica. Elsevier/Academic Press, classifica edition, 2005.

[88] W Liu, Y Li, GH Learn, RS Rudicell, JD Robertson, BF Keele, JB Ndjango, CM Sanz, DB Morgan, S Locatelli, MK Gonder, PJ Kranzusch, PD Walsh, E Delaporte, E Mpoudi-Ngole, AV Georgiev, MN Muller, GM Shaw, M Peeters, PM Sharp, JC Rayner, and BH Hahn. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature*, 467(7314):420–425, 2010.

[89] W Liu, M Worobey, Y Li, BF Keele, F Bibollet-Ruche, Y Guo, PA Goepfert, ML Santiago, JB Ndjango, C Neel, SL Clifford, C Sanz, S Kamenya, ML Wilson, AE Pusey, N Gross-Camp, C Boesch, V Smith, K Zamma, MA Huffman, JC Mitani, DP Watts, M Peeters, GM Shaw, WM Switzer, PM Sharp, and BH Hahn. Molecular ecology and natural history of simian foamy virus infection in wild-living chimpanzees. *PLoS Pathog*, 4(7):e1000097, July 2008.

[90] S Locatelli and M Peeters. Cross-species transmission of simian retroviruses: how and why they could lead to the emergence of new diseases in the human population. *AIDS*, 26(6):659–673, 2012.

[91] M Löchelt, RM Flugel, and M Aboud. The human foamy virus internal promoter directs the expression of the functional Bel 1 transactivator and Bet protein early after. *J Virol*, 68(2):638–645, 1994.

[92] M Löchelt, F Romen, P Bastone, H Muckenfuss, N Kirchner, YB Kim, U Truyen, U Rösler, M Battenberg, A Saïb, E Flory, K Cichutek, and C Münk. The antiretroviral activity of APOBEC3 is inhibited by the foamy virus accessory Bet protein. *Proc Natl Acad Sci U S A*, 102(22):7982–7987, May 2005.

M

[93] WA Malmquist, MJ Maaten, AD Boothe, and A Malmquist. Isolation, immunodiffusion, immunofluorescence, and electron microscopy of a syncytial virus of lymphosarcomatous and apparently normal cattle. *Cancer Research*, 29:188–200, 1969.

[94] DP Martin, P Lemey, M Lott, V Moulton, D Posada, and P Lefeuvre. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, 26(19):2462–2463, 2010.

[95] CD Meiering, KE Comstock, and ML Linial. Multiple integrations of human foamy virus in persistently infected human erythroleukemia cells. *J Virol*, 74(4):1718–1726, February 2000.

[96] CD Meiering and ML Linial. Historical perspective of foamy virus epidemiology and infection. *Clin Microbiol Rev*, 14(1):165–176, 2001.

[97] VA Morozov, FH Leendertz, S Junglen, C Boesch, G Pauli, and H Ellerbrok. Frequent foamy virus infection in free-living chimpanzees of the Taï National Park (Côte d'Ivoire). *J Gen Virol*, 90(Pt 2):500–506, February 2009.

[98] A Mouinga-Ondémé, E Betsem, M Caron, M Makuwa, B Sallé, N Renault, A Saïb, P Telfer, P Marx, A Gessain, and M Kazanji. Two distinct variants of simian foamy virus in naturally infected mandrills (Mandrillus sphinx) and cross-species transmission to humans. *Retrovirology*, 7(1):105, January 2010.

[99] A Mouinga-Ondémé, M Caron, D Nkoghe, P Telfer, P Marx, A Saïb, E Leroy, JP Gonzalez, A Gessain, and M Kazanji. Cross-species transmission of simian foamy virus to humans in rural Gabon, Central Africa. *J Virol*, 86(2):1255–1260, 2012.

[100] E Müllers, T Uhlig, K Stirnnagel, U Fiebig, H Zentgraf, and D Lindemann. Novel functions of prototype foamy virus Gag glycine-arginine-rich boxes in reverse transcription and particle morphogenesis. *J Virol*, 85(4):1452–1463, February 2011.

[101] C Münk, T Beck, J Zielonka, A Hotz-Wagenblatt, S Chareza, M Battenberg, J Thielebein, K Cichutek, IG Bravo, SJ O'Brien, M Löchelt, and N Yuhki. Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. *Genome Biology*, 9(3):R48, 2008.

[102] SM Murray and ML Linial. Foamy virus infection in primates. *J Med Primatol*, 35(4-5):225–235, August 2006.

[103] SM Murray, LJ Picker, MK Axthelm, K Hudkins, CE Alpers, and ML Linial. Replication in a superficial epithelial cell niche explains the lack of pathogenicity of primate foamy virus infections. *J Virol*, 82(12):5981–5985, June 2008.

N

[104] M Nasimuzzaman and DA Persons. Cell Membrane-associated heparan sulfate is a receptor for prototype foamy virus in human, monkey, and rodent cells. *Molecular Ther*, 20(6):1158–1166, June 2012.

[105] M Nethe, B Berkhout, and Antoinette C Kuyl. Retroviral superinfection resistance. *Retrovirology*, 2(52):1–13, 2005.

[106] NE Newton-Fisher, ME Thompson, V Reynolds, C Boesch, and L Vigilant. Paternity and social rank in wild chimpanzees (Pan troglodytes) from the Budongo Forest, Uganda. *Am J Phys Anthropol*, 142(3):417–428, 2010.

[107] T Nishida. Alloparental behavior in wild chimpanzees from the Mahale mountains, Tanzania. *Folia Primatologica*, (41):1–33, 1983.

O

[108] JF Oates. Is the chimpanzee, *Pan troglodytes* , an endangered species? It depends on what "endangered" means. *Primates*, 47(1):102–112, 2006.

[109] JF Oates, CEG Tutin, T Humle, ML Wilson, JEM Baillie, Z Balmforth, A Blom, C Boesch, D Cox, T Davenport, A Dunn, J Dupain, C Duvall, CM Ellis, KH Farmer, S Gatti, E Greengrass, J Hart, I Herbinger, C Hicks, KD Hunt, S Kamenya, F Maisels, JC Mitani, J Moore, BJ Morgan, DB Morgan, M Nakamura, S Nixon, AJ Plumptre, V Reynolds, EJ Stokes, and PD Walsh. *Pan troglodytes . IUCN Red List of Threatened Species. Version 2013.1.*, page www.iucnredlist.org, 2008.

P

[110] M Pacold, D Smith, S Little, PM Cheng, P Jordan, C Ignacio, D Richman, and SK Pond. Comparison of methods to detect HIV dual infection. *AIDS Res Hum Retroviruses*, 26(12):1291–1298, December 2010.

[111] S Palmer, M Kearney, F Maldarelli, EK Halvas, CJ Bixby, H Bazmi, D Rock, J Falloon, RT Davey, RL Dewar, JA Metcalf, S Hammer, JW Mellors, and JM Coffin. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J Clin Microbiol*, 43(1):406–413, 2005.

[112] MR Patel, M Emerman, and HS Malik. NIH Public Access. *Curr Opin Virol*, 1(4):304–309, 2012.

[113] NC Pedersen. *Feline syncytium-forming virus infection*. The W. B. Saunder Co. Philadelphia, Pa, diseases o edition, 1986.

[114] NC Pedersen, RR Pool, and T O'Brien. Feline chronic progressive polyarthritis. *Am J Vet Res*, 41(4):522–35, April 1980.

[115] M Perkovic, S Schmidt, D Marino, RA Russell, B Stauch, H Hofmann, F Kopietz, B-P Kloke, J Zielonka, H Ströver, J Hermle, D Lindemann, V Pathak, G Schneider, M Löchelt, K Cichutek, and C Münk. Species-specific inhibition of APOBEC3C by the prototype foamy virus protein bet. *J Bio Chem*, 284(9):5819–5826, March 2009.

[116] C Petit, ML Giron, J Tobaly-Tapiero, P Bittoun, E Real, Y Jacob, N Tordo, H De The, and A Saïb. Targeting of incoming retroviral Gag to the centrosome involves a direct interaction with the dynein light chain 8. *J Cell Sci*, 116(Pt 16):3433–3442, August 2003.

[117] KI Pfrepper, M Löchelt, M Schnölzer, H Heid, RM Flügel, and HR Rackwitz. Molecular characterization of proteolytic processing of the gag proteins of human spumavirus. *J Virol*, 73(9):7907–7911, 1999.

[118] KI Pfrepper, HR Rackwitz, H Heid, M Löchelt, and RM Flügel. Molecular characterization of proteolytic processing of the pol proteins of human foamy virus reveals novel features of the viral protease. *J Virol*, 72(9):7648–7652, 1998.

[119] HTT Phung, Y Ikeda, T Miyazawa, K Nakamura, M Mochizuki, Y Izumiya, E Sato, Y Nishimura, Y Tohya, E Takahashi, and T Mikami. Genetic analyses of feline foamy virus isolates from domestic and wild feline species in geographically distinct areas. *Virus Research*, 76(2):171–181, 2001.

[120] D Posada. jModelTest: phylogenetic model averaging. *Mol Biol Evol*, 25(7):1253–1256, 2008.

[121] D Posada. Selection of models of DNA evolution with jModelTest. *Methods Mol Biol*, 537:93–112, 2009.

[122] D Posada and KA Crandall. Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol*, 16(1):37–45, 2001.

R

[123] A Rethwilm and J Bodem. Evolution of foamy viruses: the most ancient of all retroviruses. *Viruses*, 5(10):2349–2374, January 2013.

[124] V Reynolds. *Budongo: An African Forest and Its Chimpanzees.* Natural History Press. Garden City, NY., 1965.

[125] V Reynolds. *The Chimpanzees of the Budongo Forest: Ecology, Behaviour and Conservation*. Oxford University Press, Oxford, UK, 2005.

[126] F Romen, P Backes, M Materniak, R Sting, TW Vahlenkamp, R Riebe, M Pawlita, J Kuzmak, and M Löchelt. Serological detection systems for identification of cows shedding bovine foamy virus via milk. *Virology*, 364(1):123–131, 2007.

[127] J Roy, W Rudolph, T Juretzek, K Gärtner, M Bock, O Herchenröder, D Lindemann, M Heinkelein, and A Rethwilm. Feline foamy virus genome and replication strategy. *J Virol*, 77(21):11324–11331, 2003.

[128] R Rua, E Betsem, S Calattini, A Saïb, and A Gessain. Genetic characterization of simian foamy viruses infecting humans. *J Virol*, 86(24):13350–13359, December 2012.

[129] R Rua, E Betsem, and A Gessain. Viral latency in blood and saliva of Simian Foamy Virus-infected humans. *PLoS One*, 8(10):e77072, January 2013.

[130] RA Russell, HL Wiegand, MD Moore, A Schäfer, MO Mcclure, BR Cullen, and A Scha. Foamy virus bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors. *J Virol*, 79(14):8724–8731, 2005.

[131] EG Ryland, Y Tang, CD Christie, and ME Feeney. Sequence evolution of HIV-1 following mother-to-child transmission. *J Virol*, 84(23):12437–12444, December 2010.

S

[132] A Saïb. Non-primate foamy viruses. *Curr Top Microbiol Immunol*, 277:197–211, 2003.

[133] A Saïb, F Puvion-Dutilleul, M Schmid, J Périès, and H De Thé. Nuclear targeting of incoming human foamy virus Gag proteins involves a centriolar step. *J Virol*, 71(2):1155–1161, 1997.

[134] JF Salazar-Gonzalez, E Bailes, KT Pham, MG Salazar, MB Guffey, BF Keele, CA Derdeyn, P Farmer, E Hunter, S Allen, O Manigart, J Mulenga, JA Anderson, R Swanstrom, BF Haynes, GS Athreya, BTM Korber, PM Sharp, GM Shaw, and BH Hahn. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J Virol*, 82(8):3952–3970, April 2008.

[135] PA Sandstrom, KO Phan, WM Switzer, T Fredeking, L Chapman, W Heneine, and TM Folks. Simian foamy virus infection among zoo keepers. *The Lancet*, 355(9203):551–552, 2000.

[136] M Schartl, RB Walter, Y Shen, T Garcia, J Catchen, A Amores, I Braasch, D Chalopin, J-N Volff, K-P Lesch, A Bisazza, P Minx, L Hillier, RK Wilson, S Fuerstenberg, J Boore, S Searle, JH Postlethwait, and WC Warren. The genome of the platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and several complex traits. *Nature genetics*, 45(5):567–572, May 2013.

[137] MA Schillaci, L Jones-Engel, GA Engel, Y Paramastri, E Iskandar, B Wilson, JS Allan, RC Kyes, R Watanabe, and R Grant. Prevalence of enzootic simian viruses among urban performance monkeys in Indonesia. *Trop Med Int Health*, 10(12):1305–1314, 2005.

[138] AW Schliephake and A Rethwilm. Nuclear localization of foamy virus Gag precursor protein. *J Virol*, 68(8):4946–4954, August 1994.

[139] HA Schmidt, K Strimmer, M Vingron, and A von Haeseler. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics (Oxford, England)*, 18(3):502–504, March 2002.

[140] M Schweizer, V Falcone, J Gänge, R Turek, and D Neumann-Haefelin. Simian foamy virus isolated from an accidentally infected human individual. *J Virol*, 71(6):4821–4824, June 1997.

[141] M Schweizer and D Neumann-Haefelin. Phylogenetic analysis of primate foamy viruses by comparison of pol sequences. *Virology*, 207(2):577–582, 1995.

[142] M Schweizer, H Schleer, M Pietrek, J Liegibel, V Falcone, and D Neumann-Haefelin. Genetic stability of foamy viruses: long-term study in an African green monkey population. *J Virol*, 73(11):9256–9265, November 1999.

[143] Y Shan and R Gras. 43 genes support the lungfish-coelacanth grouping related to the closest living relative of tetrapods with the Bayesian method under the coalescence model. *BMC research notes*, 4:49, January 2011.

[144] CR Stenbak and ML Linial. Role of the C Terminus of Foamy Virus Gag in RNA Packaging and Pol Expression. *J Virol*, 78(17):9423–9430, 2004.

[145] AD Stewart, JM Jr Logsdon, and SE Kelley. An empirical study of the evolution of virulence under both horizontal and vertical transmission. *Evolution*, 59(4):730–739, 2005.

[146] RM Stumpf and C Boesch. Male aggression and sexual coercion in wild West African chimpanzees, *Pan troglodytes verus* . *Anim Behav*, 79(2):333–342, 2010.

[147] WM Switzer, V Bhullar, V Shanmugam, M Cong, B Parekh, NW Lerche, JL Yee, JJ Ely, R Boneva, LE Chapman, TM Folks, and W Heneine. Frequent simian foamy virus infection in persons occupationally exposed to nonhuman primates. *J Virol*, 78(6):2780–2789, 2004.

[148] WM Switzer, M Salemi, V Shanmugam, F Gao, ME Cong, C Kuiken, V Bhullar, BE Beer, D Vallet, A Gautier-Hion, Z Tooze, F Villinger, EC Holmes, and W Heneine. Ancient co-speciation of simian foamy viruses and primates. *Nature*, 434(March):376–380, 2005.

T

[149] AR Templeton, KA Crandall, and CF Sing. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, 132(2):619–633, 1992.

[150] J Tobaly-Tapiero, P Bittoun, M Neves, MC Guillemin, CH Lecellier, F Puvion-Dutilleul, B Gicquel, S Zientara, ML Giron, H de Thé, and A Saïb. Isolation and characterization of an equine foamy virus. *J Virol*, 74(9):4064–4073, May 2000.

V

[151] EJ Verschoor, S Langenhuijzen, I Bontjer, Z Fagrouch, H Niphuis, KS Warren, K Eulenberger, and JL Heeney. The phylogeography of orangutan foamy viruses supports the theory of ancient repopulation of Sumatra. *J Virol*, 78(22):12712–12716, 2004.

[152] L Vigilant, M Hofreiter, H Siedel, and C Boesch. Paternity and relatedness in wild chimpanzee communities. *Proc Natl Acad Sci U S A*, 98(23):12890–12895, 2001.

[153] P Villesen. FaBox: an online toolbox for fasta sequences. *Molecular Ecology Notes*, 7(6):965–968, 2007.

W

[154] I Winkler, J Bodem, L Haas, M Zemba, H Delius, R Flower, RM Flügel, and M Löchelt. Characterization of the genome of feline foamy virus and its proteins shows distinct features different from those of primate spumaviruses. *J Virol*, 71(9):6727–6741, September 1997.

[155] IG Winkler, RM Flügel, M Löchelt, and RL Flower. Detection and molecular characterisation of feline foamy virus serotypes in naturally infected cats. *Virology*, 247(2):144–51, August 1998.

[156] IG Winkler, M Löchelt, and RL Flower. Epidemiology of feline foamy virus and feline immunodeficiency virus infections in domestic and feral cats: a seroepidemiological study. *J Clin Microbiol*, 37(9):2848–2851, September 1999.

[157] ND Wolfe, WM Switzer, JK Carr, VB Bhullar, V Shanmugam, U Tamoufe, AT Prosser, JN Torimiro, A Wright, E Mpoudi-Ngole, FE McCutchan, DL Birx, TM Folks, DS Burke, and W Heneine. Naturally acquired simian retrovirus infections in central African hunters. *The Lancet*, 363(9413):932–937, March 2004.

[158] EE Wroblewski, L Johnson, CM Murray, J Henderson, A Stanley, and AE Pusey. Father-Offspring relationships among the gombe chimpanzees (*Pan troglodytes schweinfurthii*). Kyoto, Japan, 2010.

[159] Z Wu, X Ren, L Yang, Y Hu, J Yang, G He, J Zhang, J Dong, L Sun, J Du, L Liu, Y Xue, J Wang, F Yang, S Zhang, and Q Jin. Virome analysis for identification of novel mammalian viruses in bat species from Chinese provinces. *J Virol*, 86(20):10999–11012, October 2012.

Y

[160] JK Yamamoto, E Sparger, EW Ho, PR Andersen, TP O'Connor, CP Mandell, L Lowenstine, R Munn, and NC Pedersen. Pathogenesis of experimentally induced feline immunodeficiency virus infection in cats. *Am J Vet Res*, 49(8):1246–1258, August 1988.

[161] SF Yu, DN Baldwin, SR Gwynn, S Yendapalli, and ML Linial. Human foamy virus replication: a pathway distinct from that of retroviruses and hepadnaviruses. *Science*, 271(5255):1579–1582, 1996.

[162] SF Yu, MD Sullivan, and ML Linial. Evidence that the human foamy virus genome is DNA. *J Virol*, 73(2):1565–1572, February 1999.

Z

[163] M Zemba, T Wilk, T Rutten, A Wagner, RM Flügel, and M Löchelt. The carboxy-terminal p3Gag domain of the human foamy virus Gag precursor is required for efficient virus infectivity. *Virology*, 247(1):7–13, July 1998.

[164] Z Zommers. *Impact of human disturbance on stress, disease and conservation of chimpanzees, Pan troglodytes, in Budongo Forest, Uganda*. PhD thesis, University of Oxford, 2010.

9 Appendix

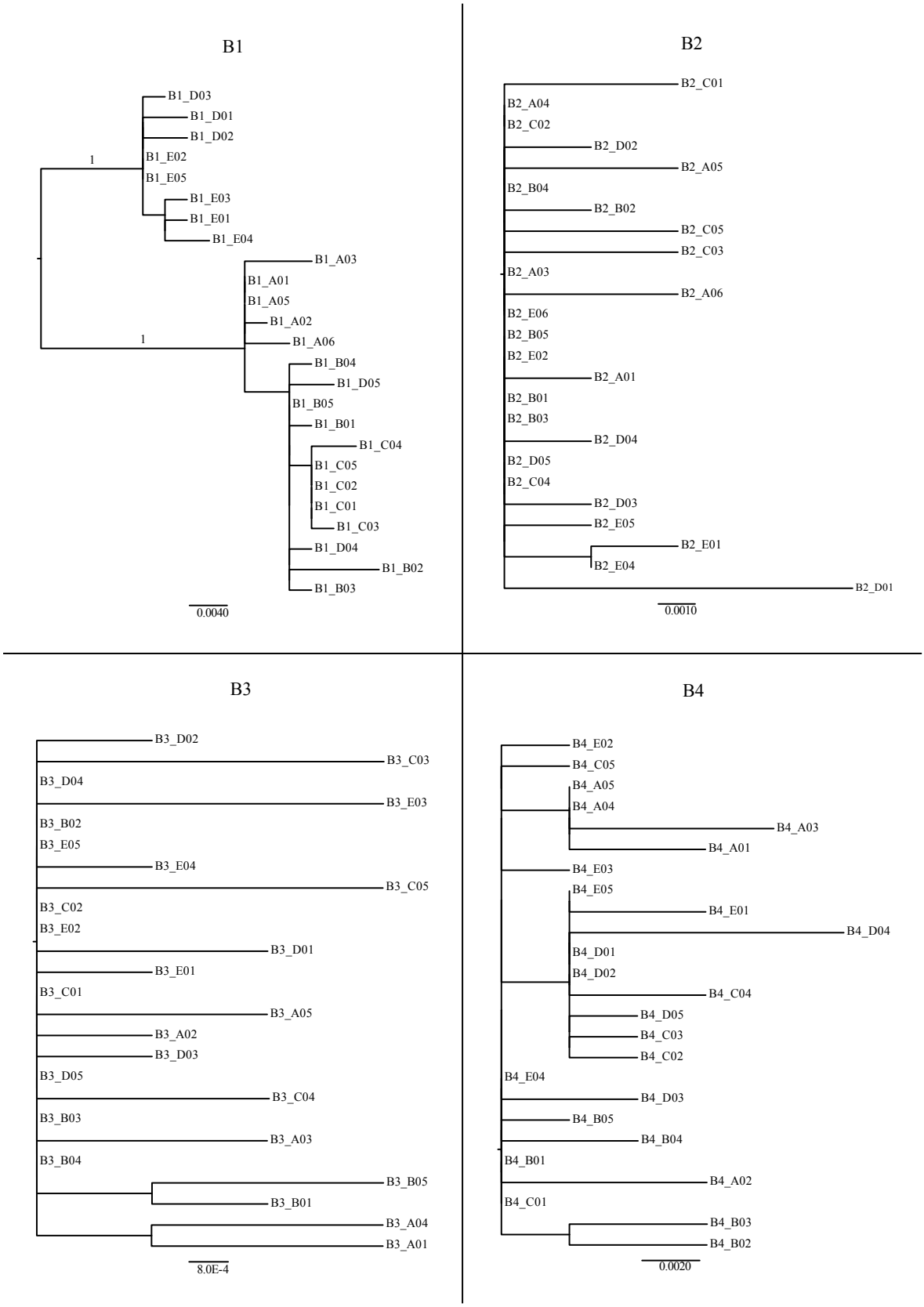

Figure S1: legend please refer to page 115.

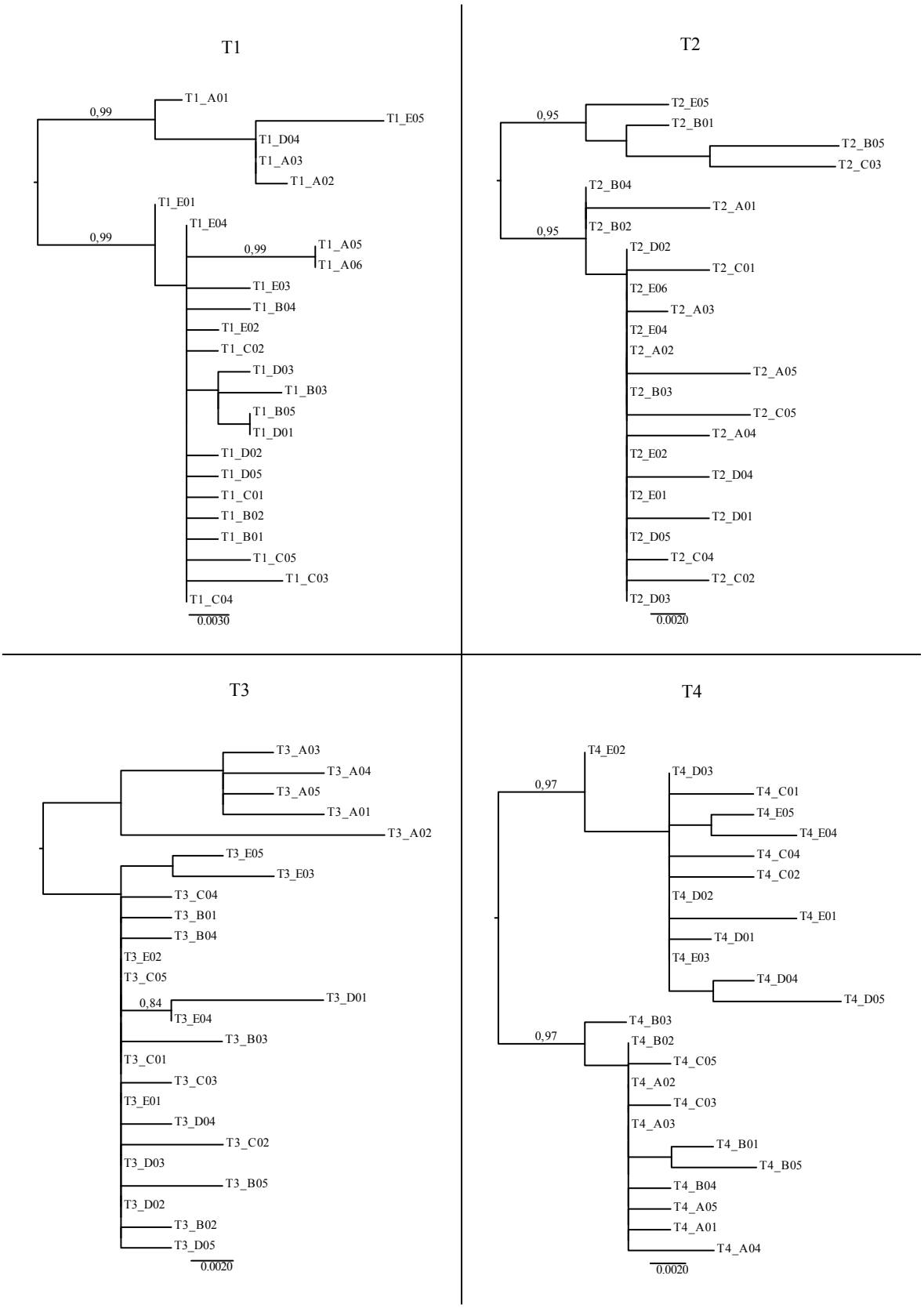

Figure S1: legend please refer to page 115.

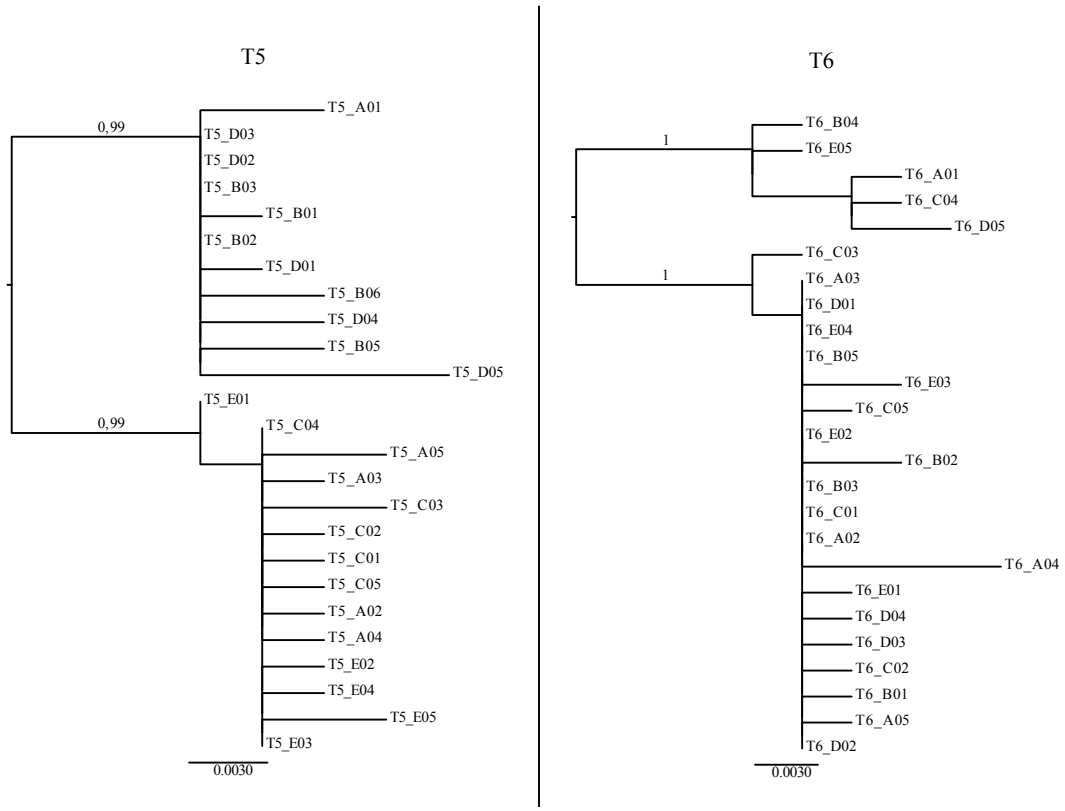

Figure S1: **Maximum likelihood trees of all individuals of dataset A**, sorted by community: *P. t. schweinfurthii* B1-B4 followed by *P. t. verus* T1-T6. Sequence names are built as follows: [individual]_[PCR product "A"or "B" and clone "01" to "09"]. The tree is not rooted. Branch robustness is shown as approximate likelihood ratio test (aLRT). Only aLRT values ≥ 0.95 are presented. Please note that most inner branches, except for main bipartition of super-infected individuals, are not statistically supported. The scale bar indicates nucleotide substitutions per site.

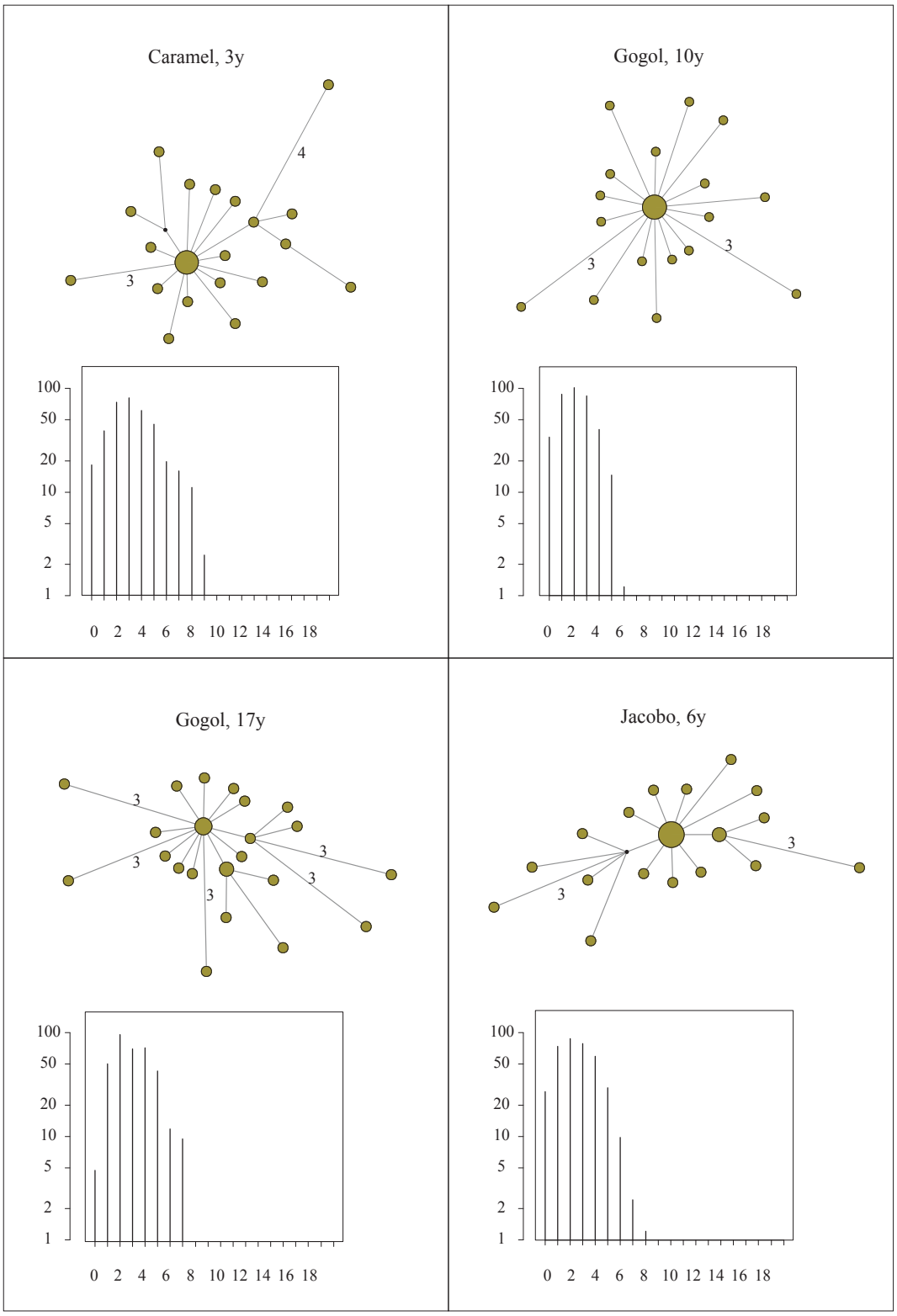

Figure S2/A: legend please refer to page 123.

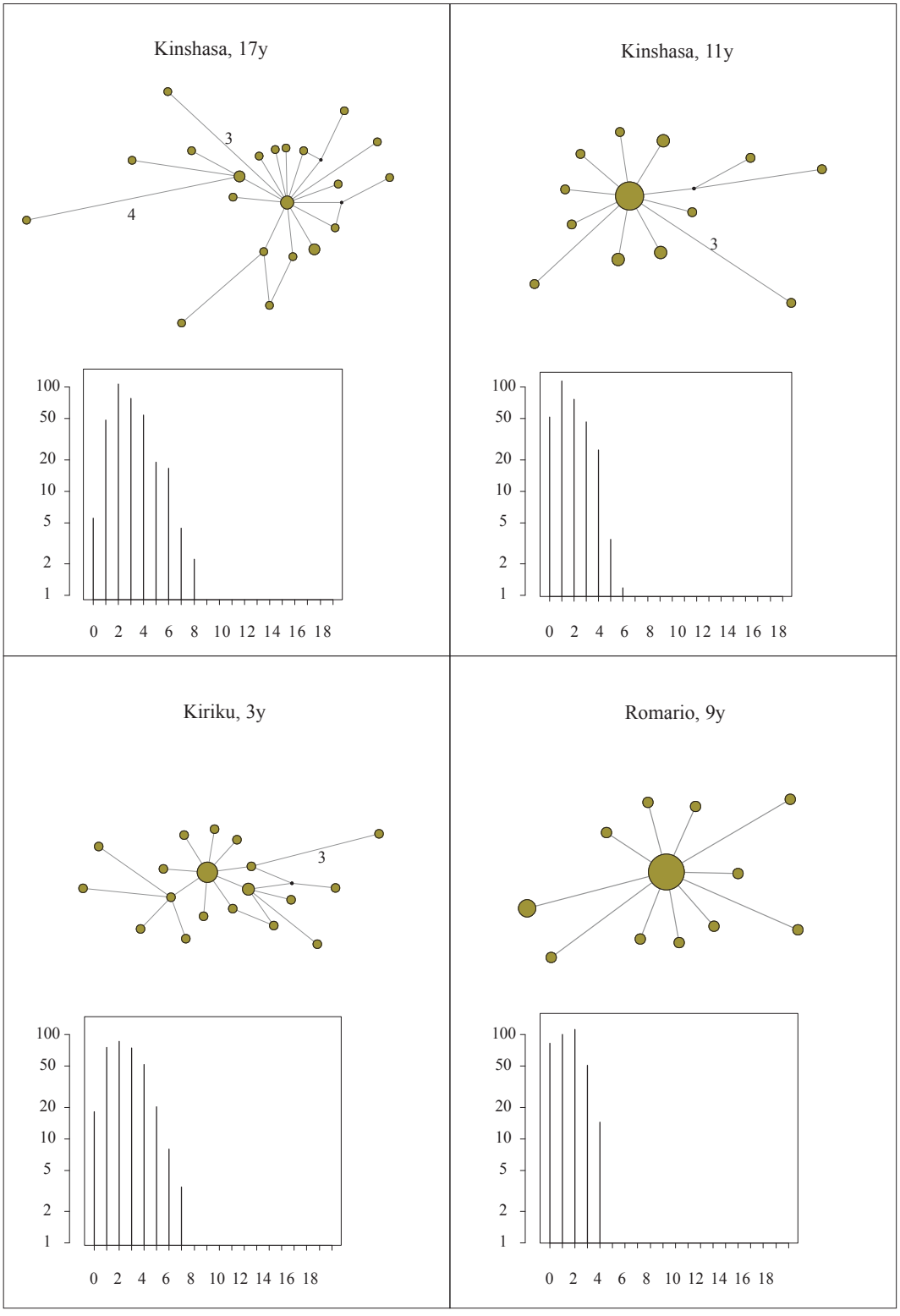

Figure S2/A: legend please refer to page 123.

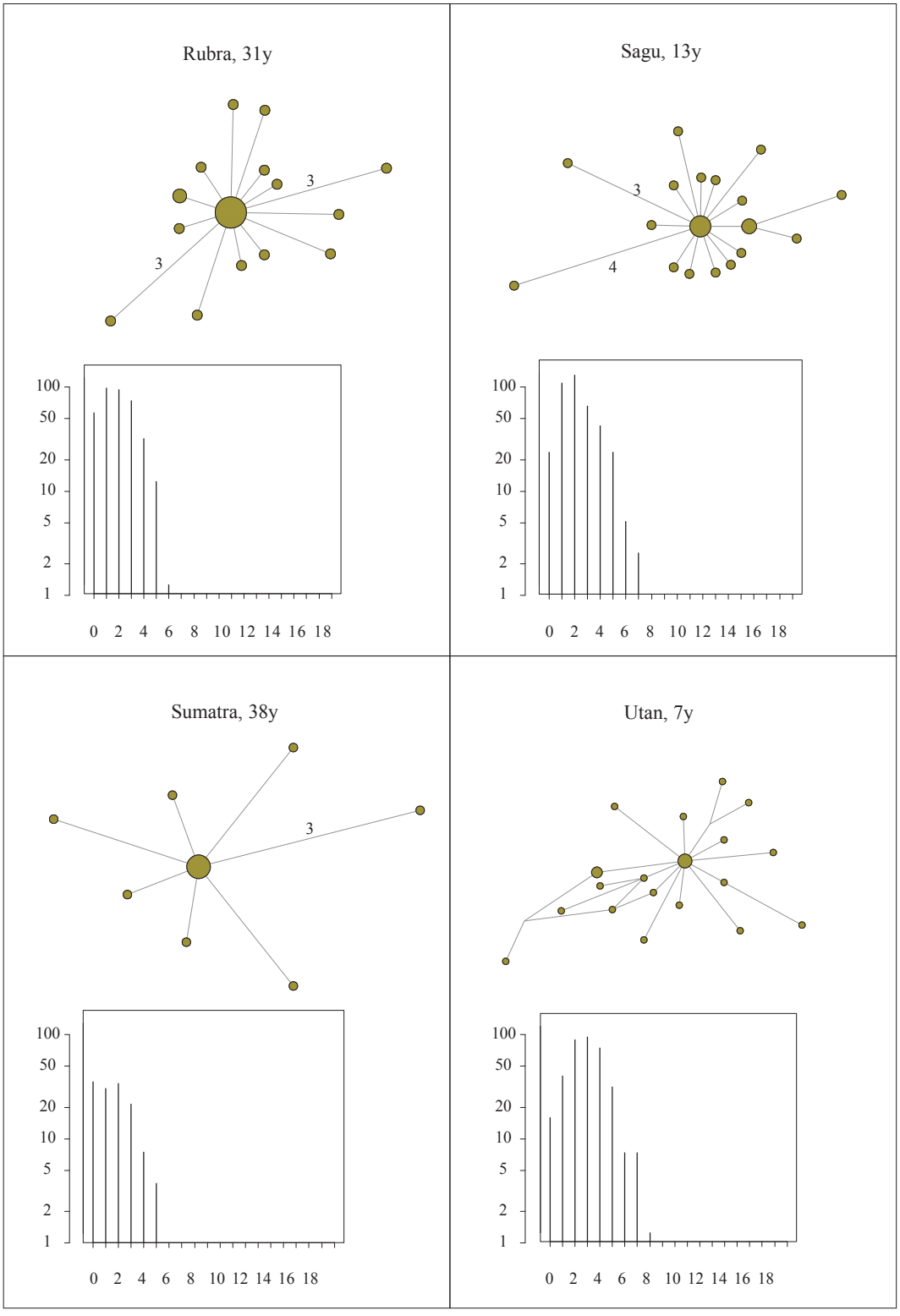

Figure S2/A: legend please refer to page 123.

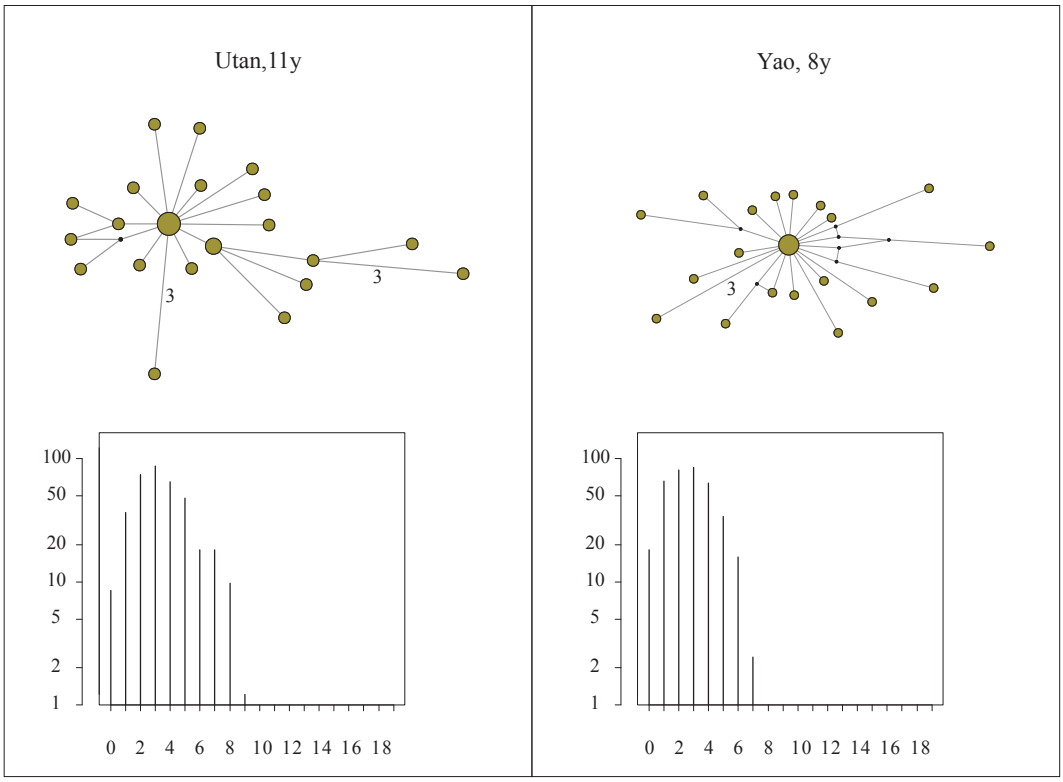

Figure S2/A: legend please refer to page 123.

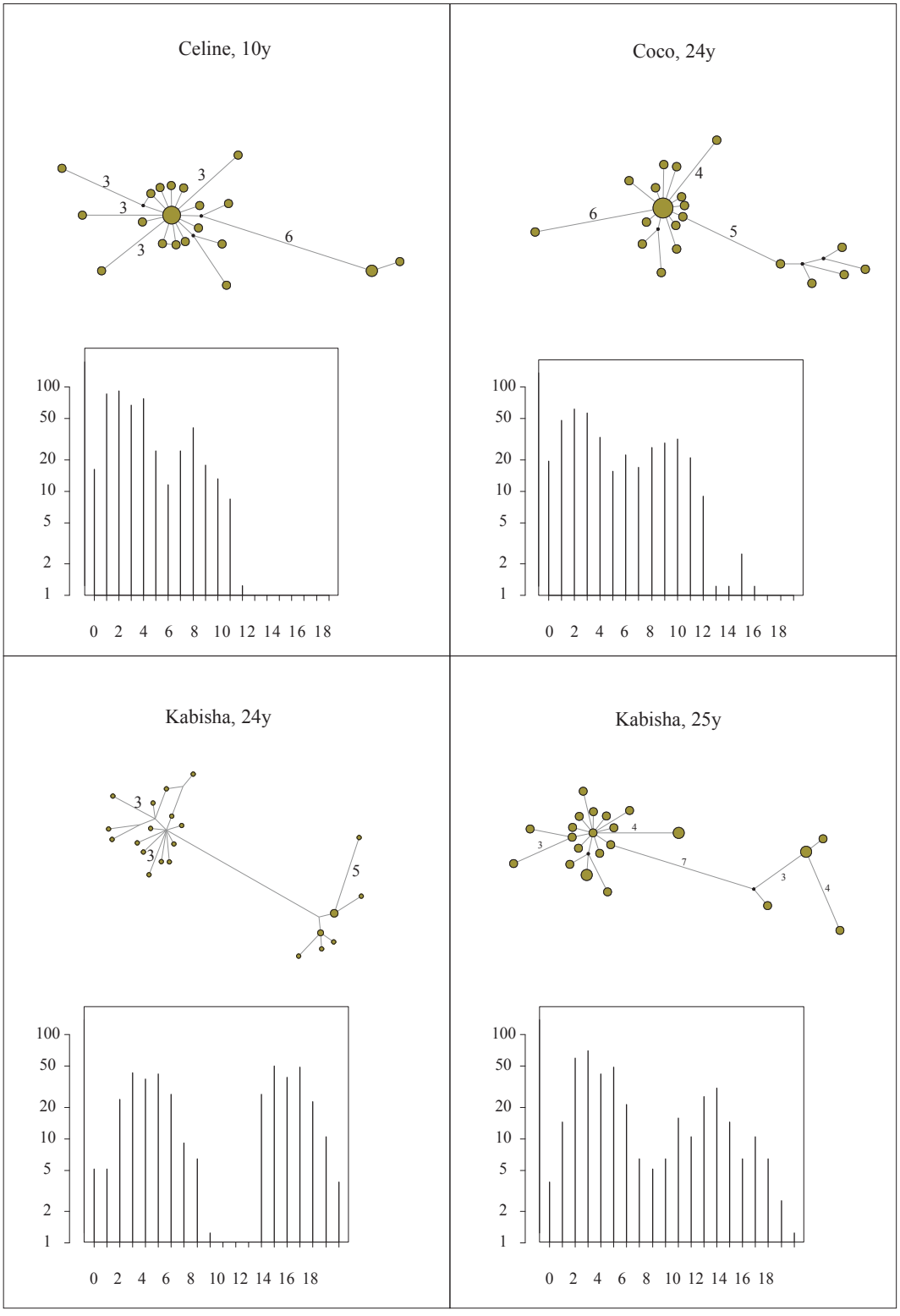

Figure S2/B: legend please refer to page 123.

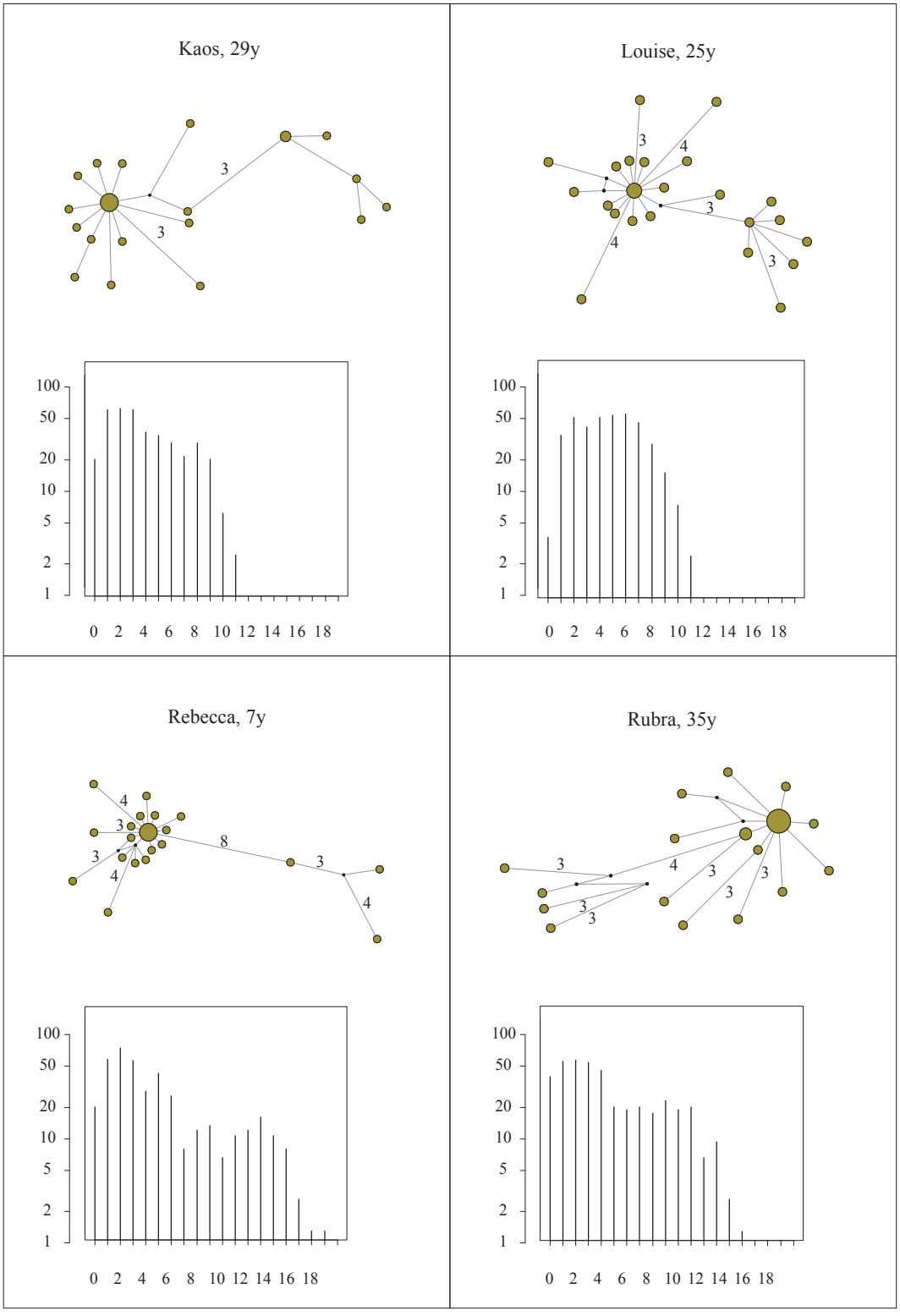

Figure S2/B: legend please refer to page 123.

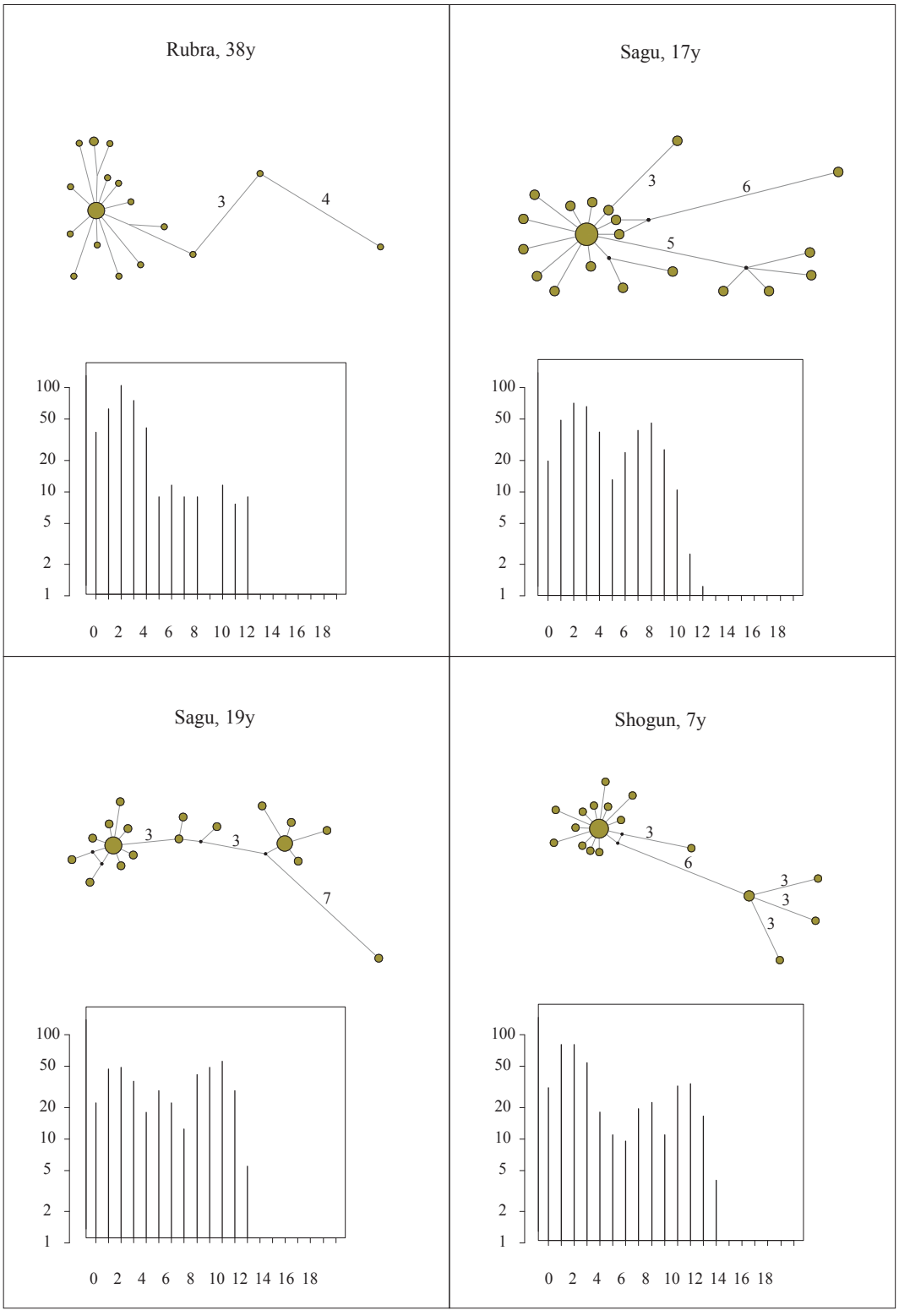

Figure S2/B: legend please refer to page 123.

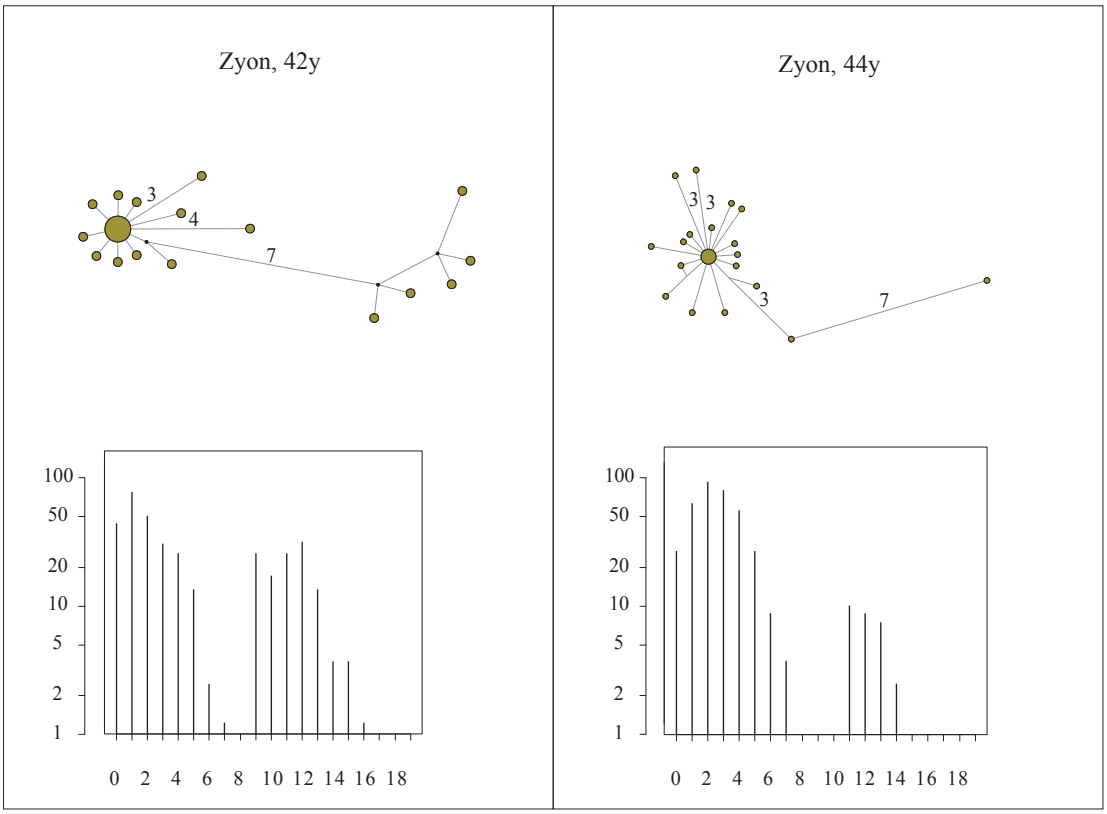

Figure S2: **Mismatch distributions and sequence networks of all individuals of Taï chimpanzees** (dataset B), sorted by infection status: A/single infection cases, B/super-infection cases, which were unambiguously identified. Plots are paired for each individual chimpanzee sample (age at time of sampling in years, y) with the mismatch distribution indicated on the bottom and the parsimony-based network TCS on the top. The mismatch distribution presents the frequency of the number of mismatches (y-axis) according to the number of base pairs in the sequence alignment (x-axis). Within each network, node size is proportional to the frequency of the sequence occurrence (25 for each individual, 15 for Sumatra 38 y). Branch lengths are directly related to the number of mutations between sequences, with values noted for differences greater than two base pairs.

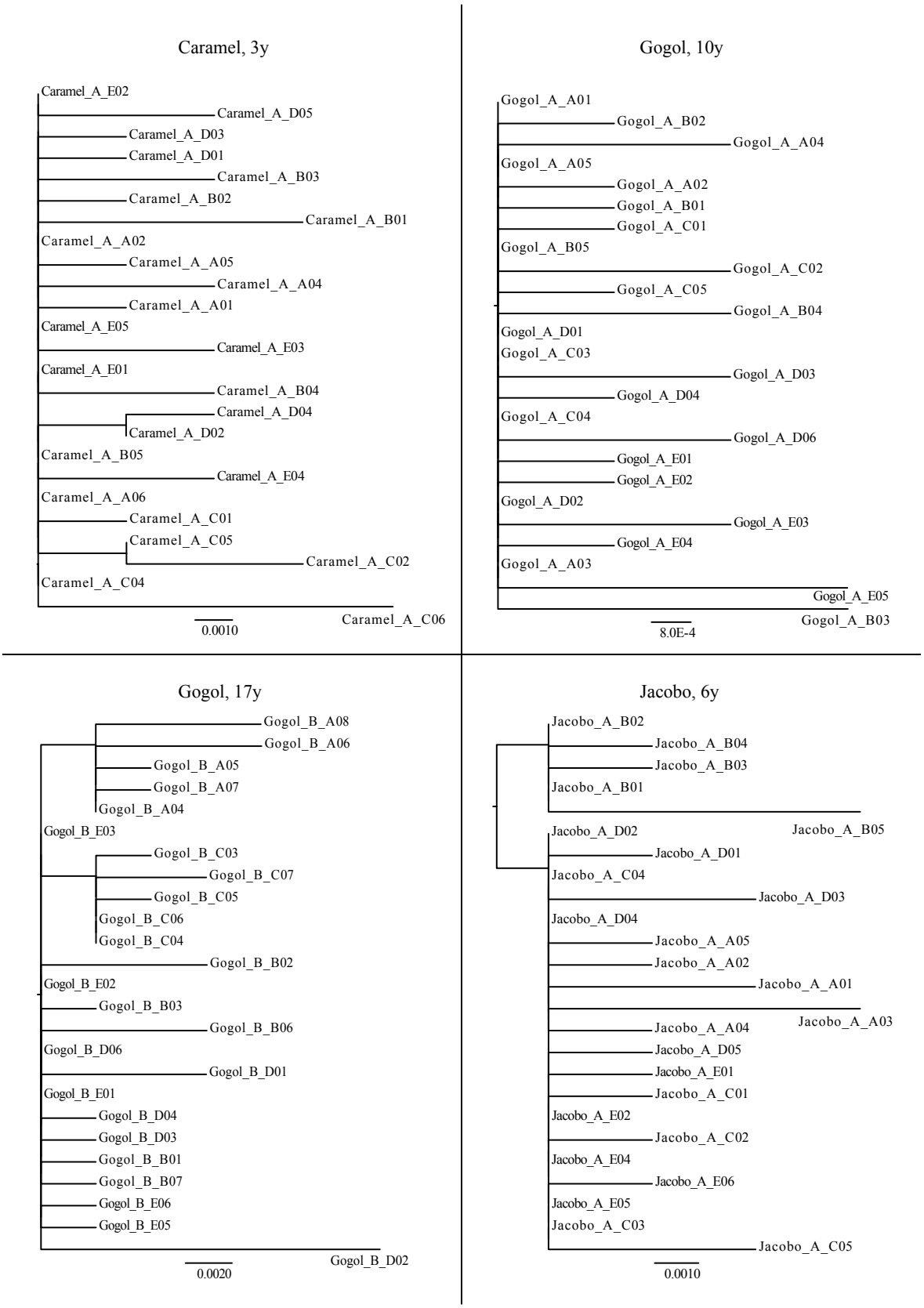

Figure S3/A: legend please refer to page 132.

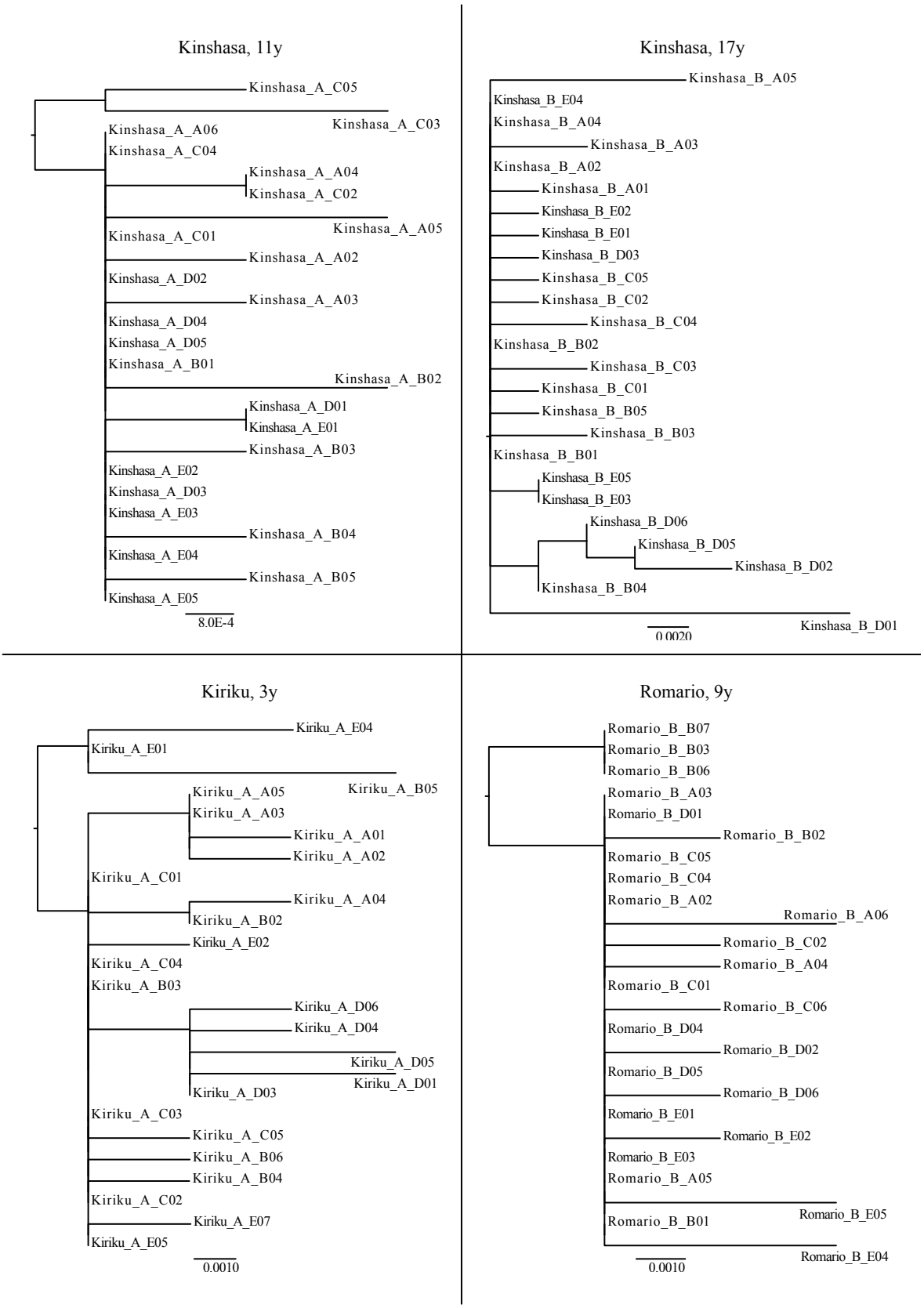

Figure S3/A: legend please refer to page 132.

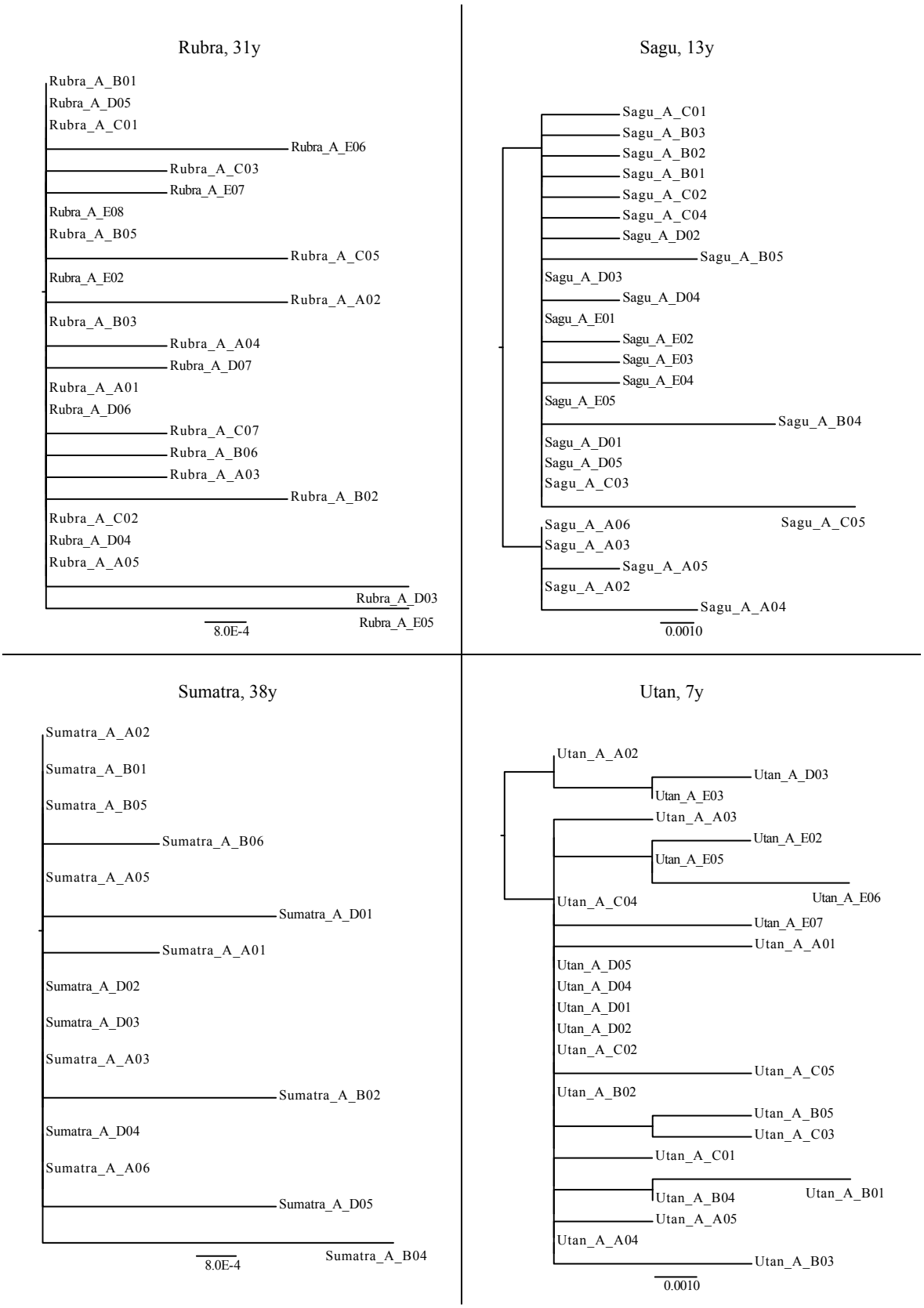

Figure S3/A: legend please refer to page 132.

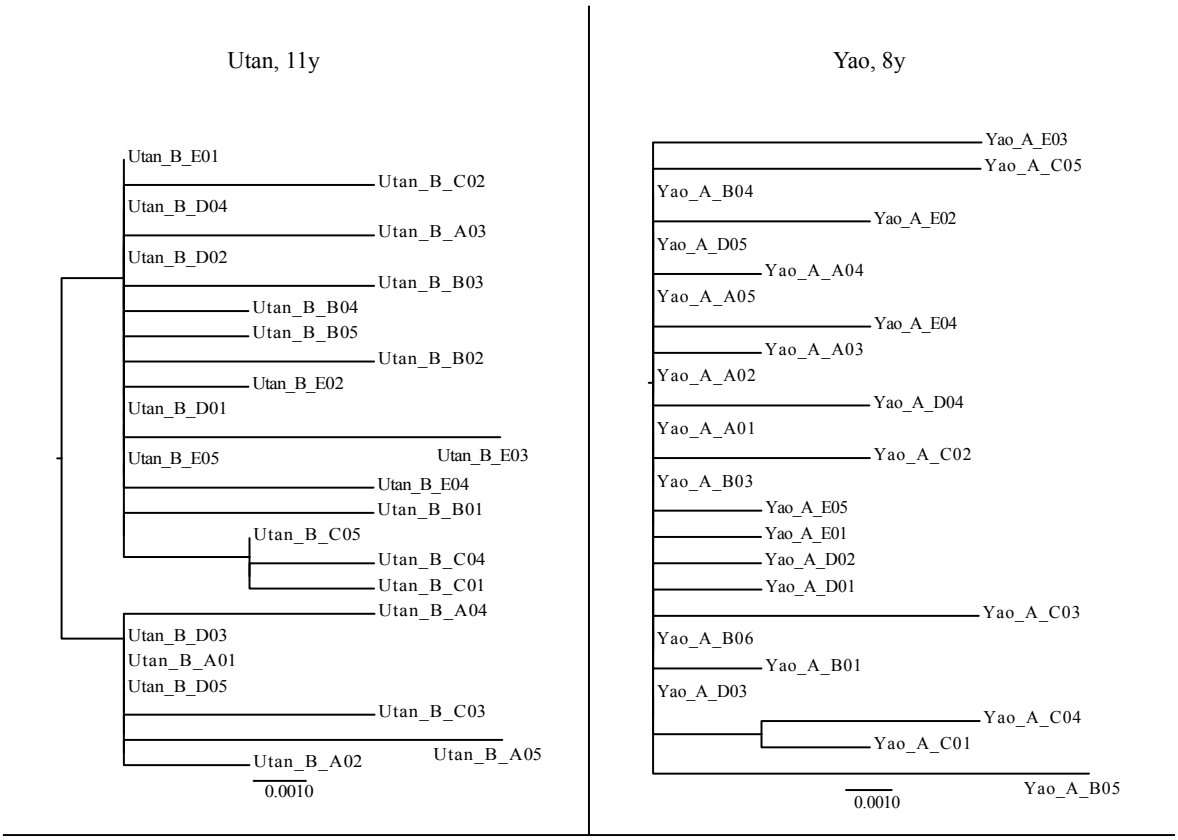

Figure S3/A: legend please refer to page 132.

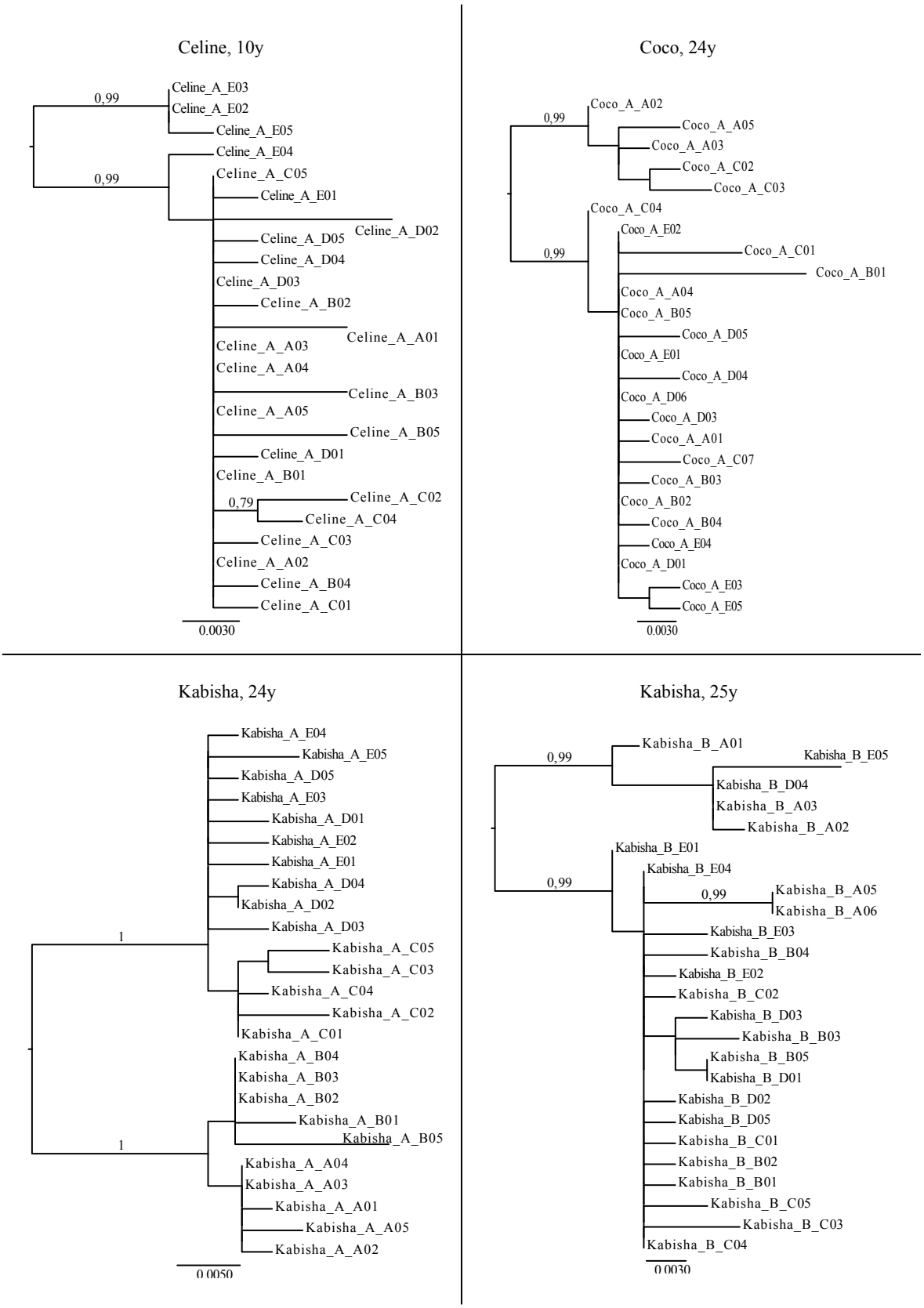

Figure S3/B: legend please refer to page 132.

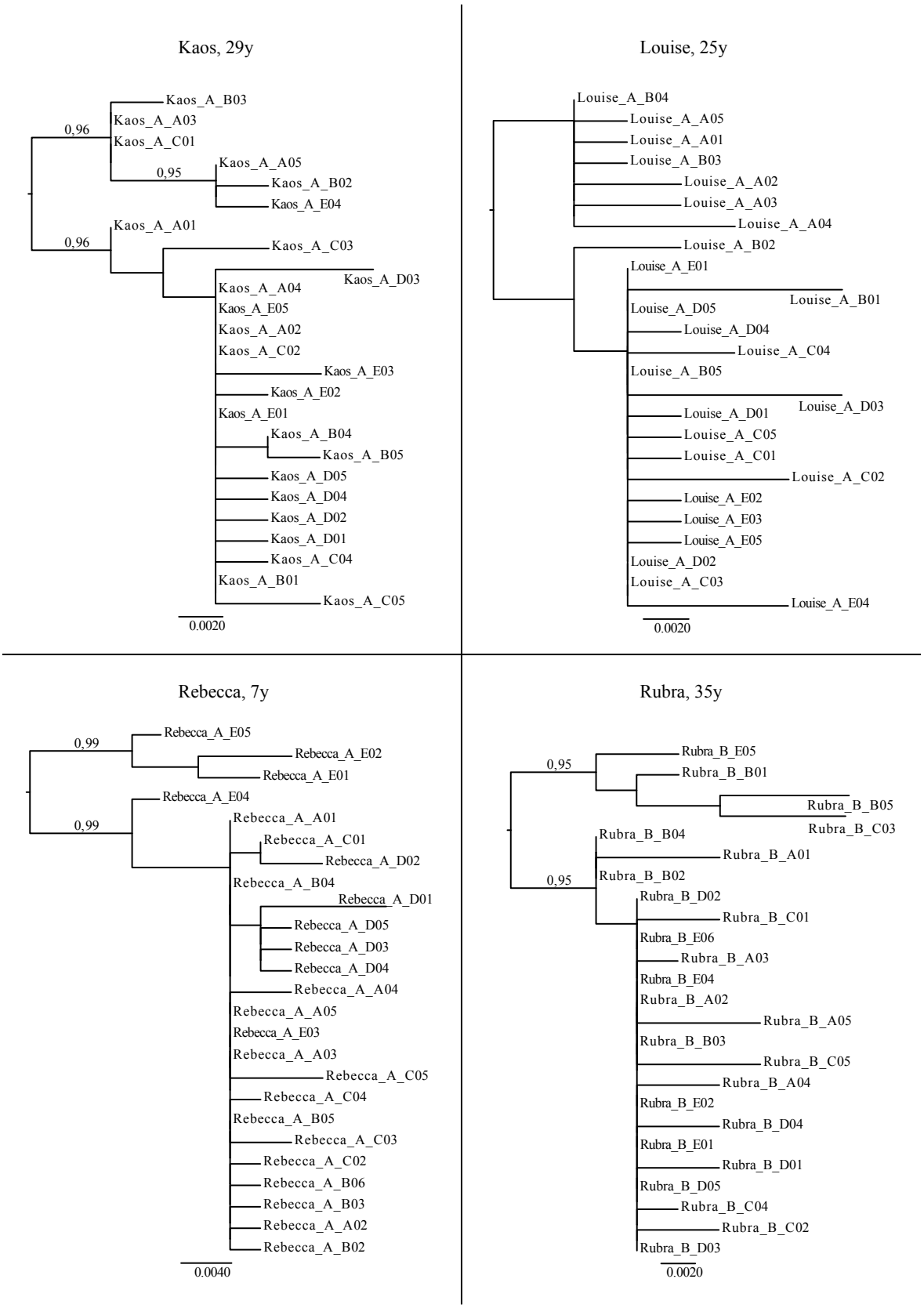

Figure S3/B: legend please refer to page 132.

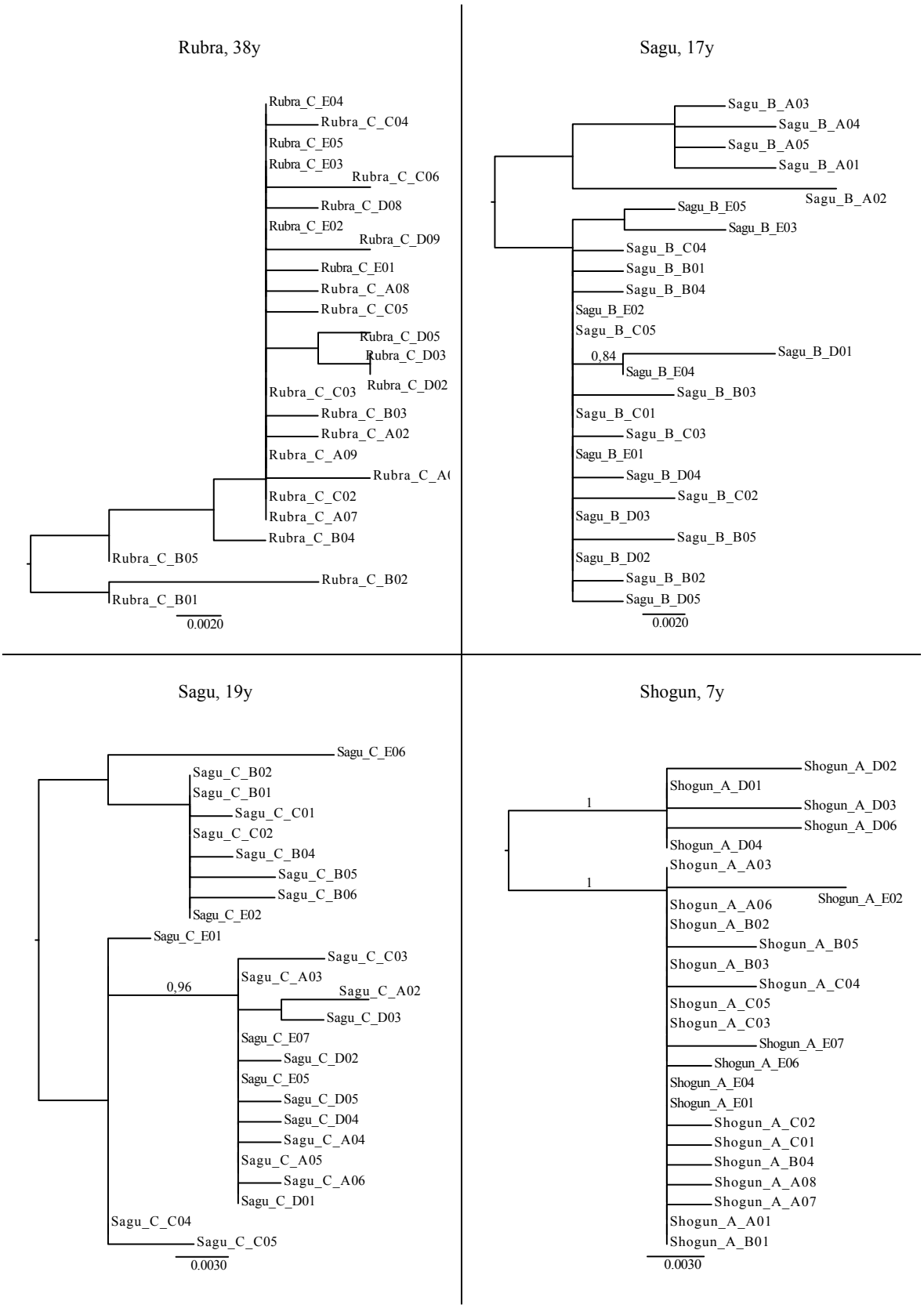

Figure S3/B: legend please refer to page 132.

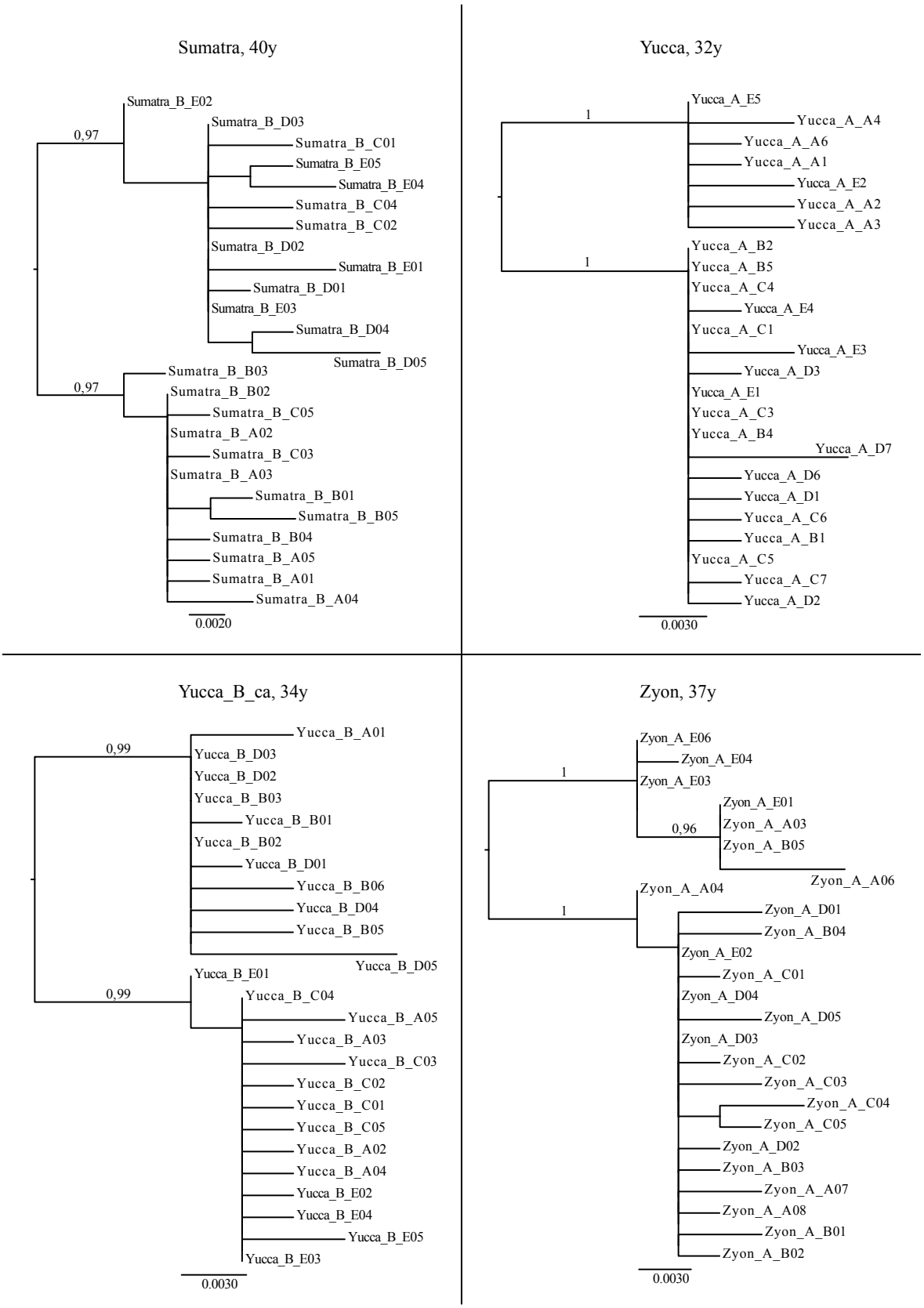

Figure S3/B: legend please refer to page 132.

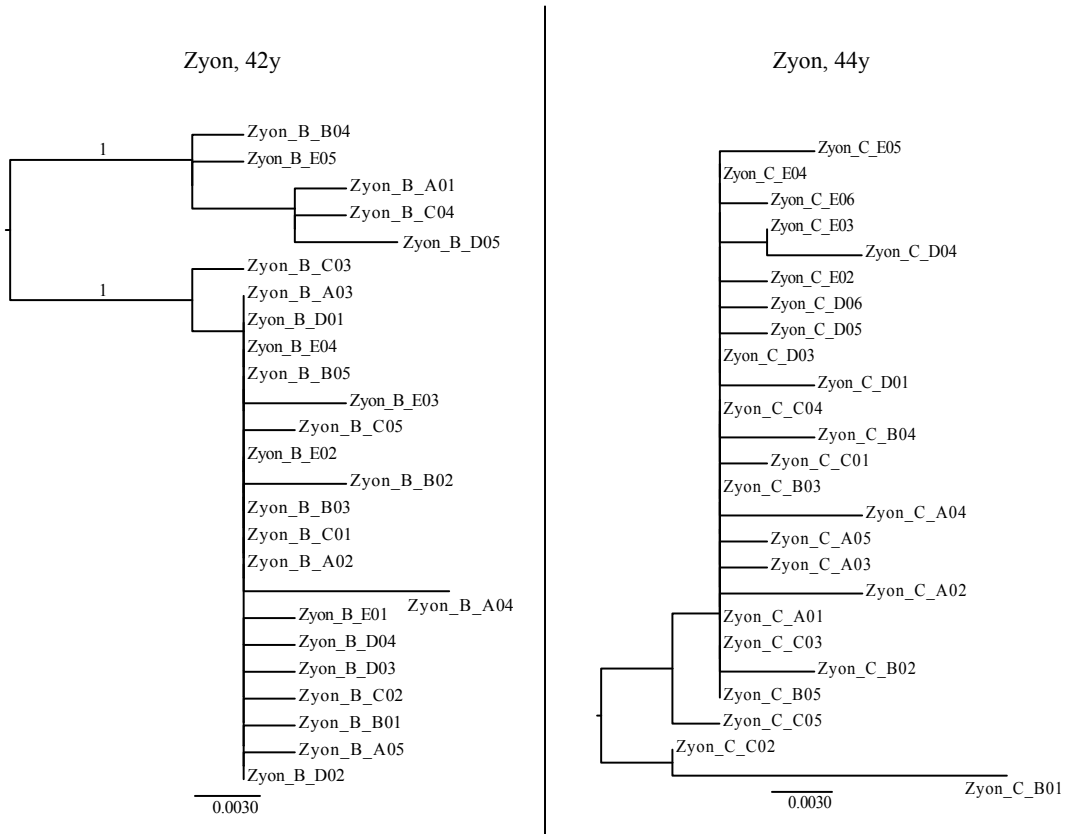

Figure S3: **Maximum likelihood trees of all individuals of Taï chimpanzees** (dataset B), sorted by infection status: A/single infection cases, B/super-infection cases, which were unambiguously identified. Phylogenetic trees include the complete set of bulk-PCR clone sequences (25 for each individual, 15 for Sumatra 38 y) of each individual (age at time of sampling in years, y) of dataset B. Sequence names are built as follows: [individual]_[sample "A" to "C"]_[PCR product "A" or "B" and clone "01" to "09"]. The tree is not rooted. Branch robustness is shown as approximate likelihood ratio test (aLRT). Only aLRT values ≥ 0.95 are presented. Please note that most inner branches are not statistically supported. The scale bar indicates nucleotide substitutions per site.

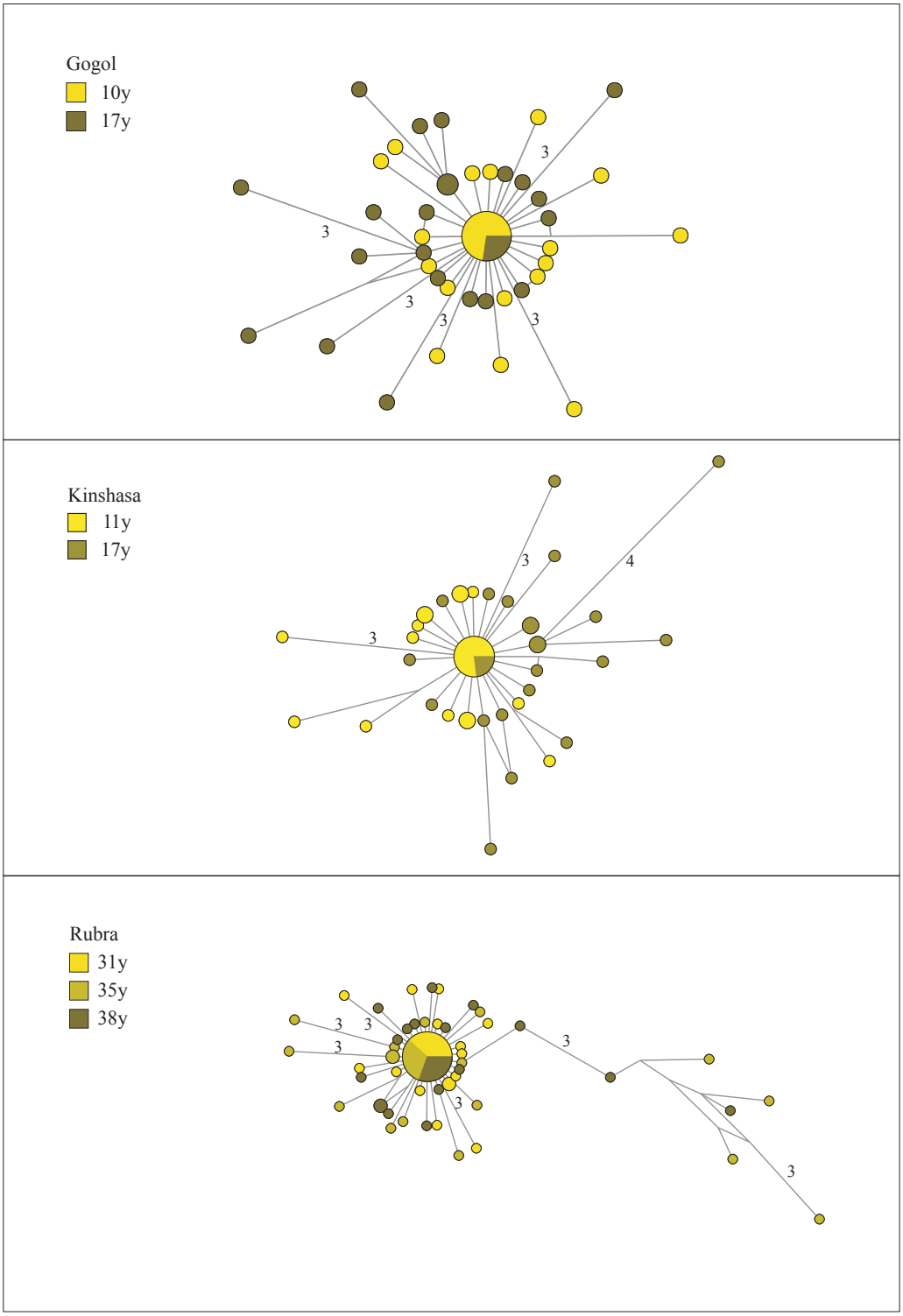

Figure S4: legend please refer to page 143.

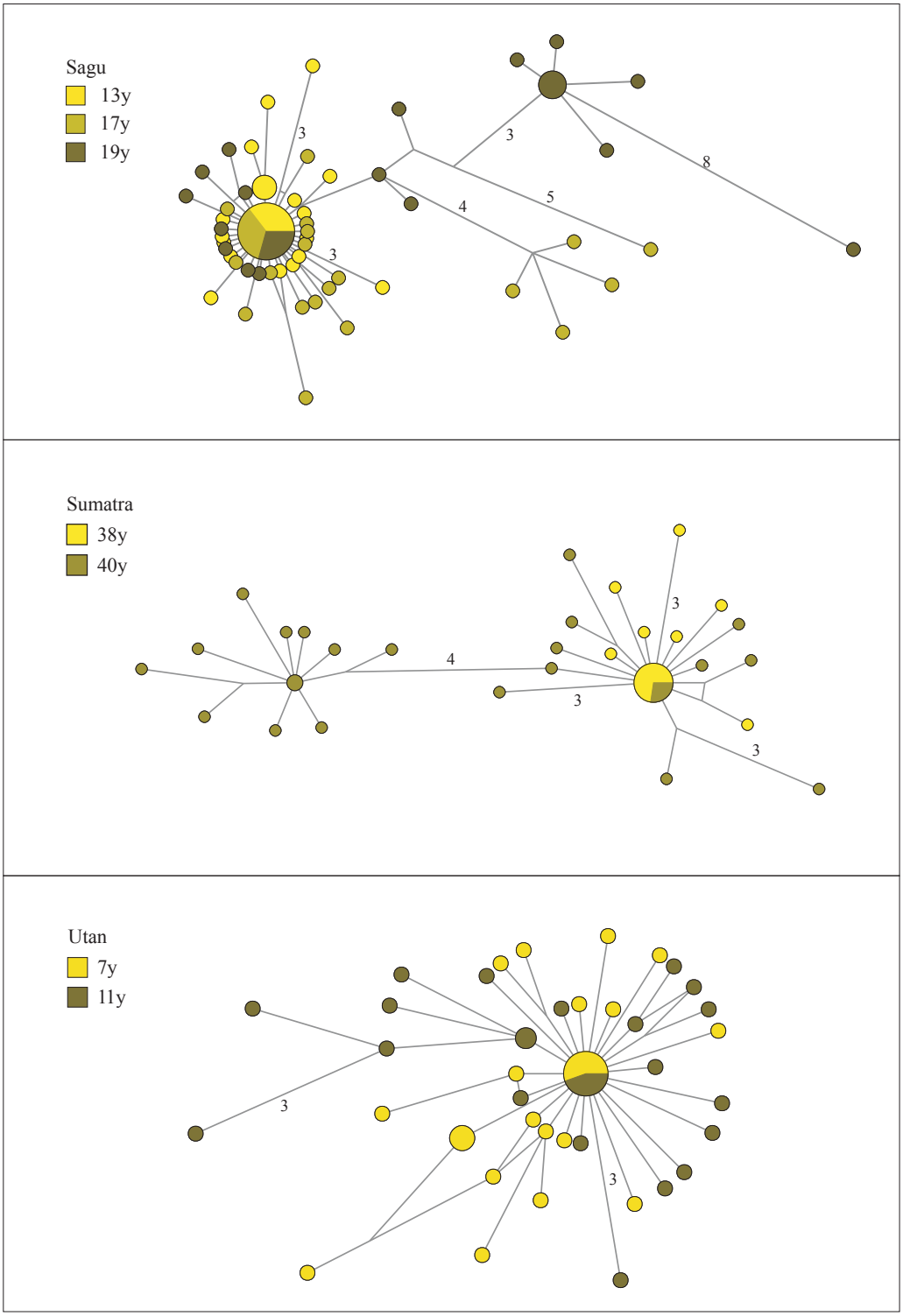

Figure S4: legend please refer to page 143.

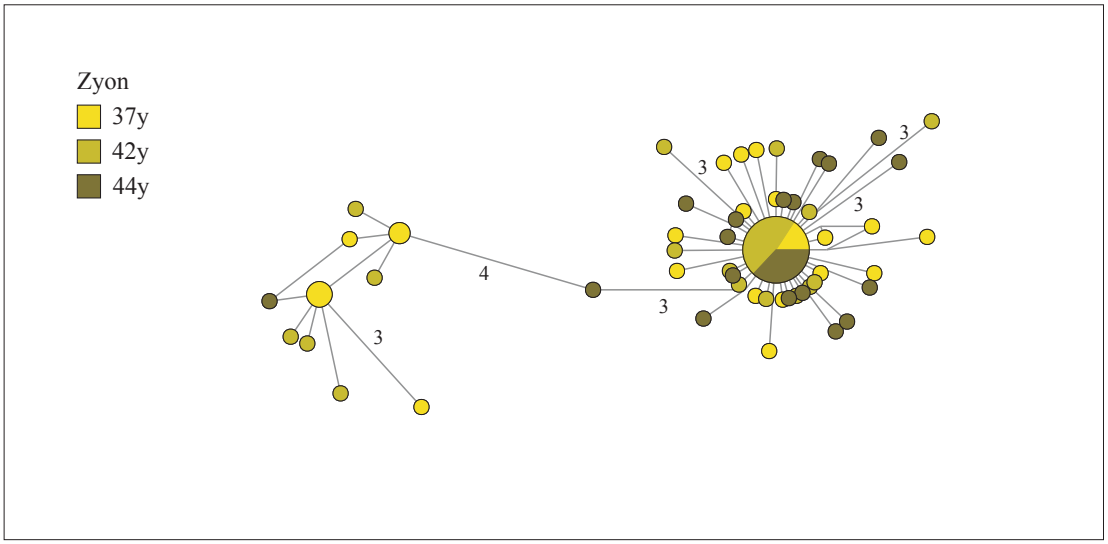

Figure S4: **Networks for all serially sampled chimpanzees** (dataset B). Within each network, the node color refers to individual age at time of sampling in years (y), the node size is proportional to the frequency of the sequence occurrence (25 for each individual, 15 for Sumatra 38 y). Branch lengths are directly related to the number of mutations between sequences, with values noted for differences greater than two base pairs.

Publication notes

Contents of this work are published in following articles and conference contributions:

Articles

- Blasse, A.*, Goffe, A.S.*, Mundry, R., Leendertz, F.H., Calvignac-Spencer, S., **Detection of retroviral super-infection from non-invasive samples.** PLoS ONE, May 7 2012 (5): e36570. doi:10.1371/journal.pone.0036570.

  * are contributed equally

- Blasse, A., Calvignac-Spencer, S., Merkel, K., Goffe, A.S., Boesch, C., Mundry, R., Leendertz, F.H., **Mother-offspring transmission and age-dependent accumulation of simian foamy virus in wild chimpanzees.** Journal of Virology, published ahead of print February 28 2013, doi: 10.1128/JVI.02743-12.

Conference contributions

- Blasse A., Goffe A.S., Mundry R., Leendertz F.H., Calvignac-Spencer S., **Investigating retroviral super-infection in wild chimpanzees (*Pan troglodytes verus*).** 15th International Conference on Human Retroviruses: HTLV and Related Viruses, June 2011, Leuven und Gembloux, Belgium.

- Blasse A., Kühl H., Boesch C., Leendertz F.H., **Simian foamy virus in wild-living chimpanzees.** 4th European Congress of Virology, April 2010, Cernobbio, Italy.

Contribution of the author

Anja Blasse and Adeelia Goffe performed viral RNA extractions from fecal samples, nested RT-PCR, cloning and Sanger's sequencing in parallel on clearly separated datasets: Anja Blasse *(P. t. verus* from Taï National Park), Adeelia Goffe (*P. t. schweinfurthii* from Budongo Forest Reserve, Uganda). These data were used in the publication Blasse et al., Plos One 2011.

In the second part of the study Kevin Merkel assisted performing the experiments (cloning and Sanger's sequencing) on the data of the Taï community. These data were used in the publication Blasse et al., Journal of Virology 2012.

All later analytical steps on these datasets were performed by Anja Blasse. This included statistical analyses, which were guided by Dr. Roger Mundry, phylogenetic analyses using maximum likelihood

and Bayesian methods guided by Dr. Sébastien Calvignac-Spencer, network analyses and simulation of triple infections.

The two published Foamy virus articles, which are the product of this thesis, were entirely drafted by Anja Blasse under the supervision of Dr. Sébastien Calvignac-Spencer and Dr. Fabian Leendertz and subsequently revised by other authors.

Acknowledgements

I warmly thank Dr. Fabian Leendertz for the opportunity to work on this interesting project, for his expert advices and inspirations. I appreciate his friendly relationship from the very first and his open ear during all phases of my thesis. I am grateful that he gave me the opportunity to participate in various courses, conferences and the Ph.D. Program of the Robert Koch-Institute. I appreciate his confidence in my work and the chance for a continuing project in collaboration with the Martin-Luther-University in Halle.

I want to express my gratitude to Prof. Lothar Wieler for his supervision and academic support of my phd even though I was not stuff in his institute. Nevertheless, he encouraged me, supported my application for a stipend and was an invaluable contact person over the whole time of my thesis.

Special thanks go to Dr. Sébastien Calvignac-Spencer for his guidance and his personal engagement in my work. He provided me excellent support and supervision, especially in the field of molecular biology, phylogenetics with hairy caterpillars as well as during writing of articles until final step of publication. I am very grateful that he always took the time for explanations and discussions and for his constructive corrections of my thesis.

I also thank Prof. Pauli and Prof. Kaufer for being reviewer of my thesis.

I would like to thank the department of primatology in Max Planck Institute for Evolutionary Anthropology and our African partners who facilitate the Taï chimpanzee project (the authorities in Côte d'Ivoire, especially the Ministry of Environment and Forests and the Ministry of Research, the directorship of the Taï National Park, the Office Ivoirien des Parcs et Réserves, the Swiss Research Center in Abidjan) and the chimpanzee project in Budongo Forest Reserve (Uganda Wildlife Authority and the Uganda National Council for Science and Technology). I especially thank the assistants and students of each of the chimpanzee projects for field observations, sample collection and sample management.

I really thank my colleagues in Project Group for Epidemiology of highly pathogenic microorganisms at Robert Koch-Institute (former NG2) for the great time and the nice atmosphere. I want to express my gratitude to Adeelia Goffe for the fun in the lab, the adventure with DGGE, the chance to improve my English, for her proof reading of my thesis and her constructive critism of previous articles. I especially thank Kevin Merkel for his great assistance in my project. I want to mention Ulla Thiesen, Anne Kopp, Verena Keil and Franziska Dahlmann (nee Kaup) for their support in the lab and the nice team of the sequencing lab. In addition, I would like to

thank Roger Mundry from Max Planck Institute for Evolutionary Anthropology for his guidance in statistics and his fresh mind and ideas during super long discussions.

I thank my current boss, Frau Prof. Nicole Kemper at the Institute of Agricultural and Nutritional Science, Martin-Luther-University Halle-Wittenberg for her understanding and support of an ongoing cooperation with the Robert Koch-Institute.

During the whole time of my phd, my family provided me with great support and motivation. I would like to emphasize my dear husband Thilo Blasse, my mum Beatrix Blasse, my sister Corinna Blasse and Ferdinand Häring. We are a wonderful team! My mum receives my deepest gratitude and love for her dedication and her boundless encouragement in everything I do. I thank my husband for his love, his understanding and his invaluable advices regarding my work, his tips for the use of Lyx or for being an insistent advocate of back-ups. I thank my dear sister for her interest in my studies and the contribution of her bioinformatical expertise and her constructive comments on my articles and my thesis. I am grateful for her patience with my endless questions about statistics or programming. I would like to acknowledge Ferdinand Häring for his long-lasting support and his encouragement during my studies. In addition, I thank my friends for their understanding and welcome distraction from time to time.

I am grateful to the FAZIT Stiftung mbH for their funding, which gave me the chance to work full time on the project and to participate to the 4th European Congress of Virology in Cernobbio, Italien.

Looking back, although I sometimes felt stuck, all experiences, the meeting of interesting personalities and forming friendships made the last five years to an invaluable time for me.

Selbständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig angefertigt habe. Ich versichere, dass ich ausschließlich die angegebenen Quellen und Hilfen in Anspruch genommen habe. Veröffentlichungen von Teilen der vorliegenden Dissertation sind angegeben. Weiter erkläre ich, dass ich nicht schon anderweitig einmal die Promotionsabsicht angemeldet oder ein Promotionsverfahren beantragt habe.

Halle, den 13. November 2013,

Anja Blasse