Origin of modern syphilis and emergence of a pandemic *Treponema pallidum* cluster

Natasha Arora1,2,*, Verena J. Schuenemann3, Günter Jäger4, Alexander Peltzer3,4,†, Alexander Seitz4, Alexander Herbig3,4,‡, Michal Strouhal15, Linda Grillová5, Leonor Sánchez-Busó6,7, Denise Kühnert8, Kirsten I. Bos3,†, Leyla Rivero Davis19, Lenka Mikalová5, Sylvia Bruisten9, Peter Komeric̆ki10, Patrick French11, Paul R. Grant12, María A. Pando13, Lucía Gallo Vaulet14, Marcelo Rodríguez Ferme̓ pin14, Antonio Martínez15, Arturo Centurion Lara16, Lorenzo Giacani16, Steven J. Norris17, David Šmajs5, Philipp P. Bosshard18, Fernando González-Candelas6,*, Kay Nieselt4,†, Johannes Krause3,4,† and Homayoun C. Bagheri1,2,*,†

The abrupt onslaught of the syphilis pandemic that started in the late fifteenth century established this devastating infectious disease as one of the most feared in human history1. Surprisingly, despite the availability of effective antibiotic treatment since the mid-twentieth century, this bacterial infection, which is caused by *Treponema pallidum* subsp. *pallidum* (TPA), has been re-emerging globally in the last few decades with an estimated 10.6 million cases in 2008 (ref. 2). Although resistance to penicillin has not yet been identified, an increasing number of strains fail to respond to the second-line antibiotic azithromycin3. Little is known about the genetic patterns in current infections or the evolutionary origins of the disease due to the low quantities of treponemal DNA in clinical samples and difficulties in culturing the pathogen4. Here, we used DNA capture and whole-genome sequencing to successfully interrogate genome-wide variation from syphilis patient specimens, combined with laboratory samples of TPA and two other subspecies. Phylogenetic comparisons based on the sequenced genomes indicate that the TPA strains examined share a common ancestor after the fifteenth century, within the early modern era. Moreover, most contemporary strains are azithromycin-resistant and are members of a globally dominant cluster, named here as SS14-O. The cluster diversified from a common ancestor in the mid-twentieth century subsequent to the discovery of antibiotics. Its recent phylogenetic divergence and global presence point to the emergence of a pandemic strain cluster.

The first reported syphilis outbreaks in Europe occurred during the War of Naples in 1495 (ref. 5), prompting unresolved theories on a post-Columbian introduction6,7. Subsequently, the epidemic spread to other continents, remaining a severe health burden until treatment with penicillin five centuries later enabled incidence reduction. The striking present-day resurgence is poorly understood, particularly the underlying patterns of genetic diversity. Much of our molecular understanding of treponemes comes from the propagation of strains in laboratory animals to obtain sufficient DNA. The few published whole genomes were obtained after amplification through rabbit passage8–10 and represent limited diversity for phylogenetic analyses. These sequences suggest that the TPA genome of 1.14 Mb is genetically monomorphic. Its potential genetic diversity remains unexplored because clinical samples are mostly typed by PCR amplification of only 1–5 loci11,12. These epidemiological strain typing studies are motivated by the limitations of serological or microscopic tests to distinguish among TPA strains or among the subspecies *Treponema pallidum* subsp. *pertenue* (TPE) and *Treponema pallidum* subsp. *endemicum* (TEN), which cause the diseases yaws and bejel, respectively. All three diseases are transmitted through skin contact and show an overlap in their clinical manifestations, but syphilis is geographically more widespread and generally transmitted sexually. The precise relationships among the bacteria are still debated, particularly regarding the evolutionary origin of syphilis.

The paucity of molecular studies and the focus on typing of a few genes means that we have limited information regarding the
evolution and spread of epidemic TPA. In this study, we interro-
gated genome-wide variation across geographically widespread iso-
lates. In total, we obtained 70 samples from 13 countries, including
52 syphilis swabs collected directly from patients between 2012 and
2013, and 18 syphilis, yaws and bejel samples collected from 1912
onwards and propagated in laboratory rabbits (Supplementary
Table 1). Through comparative genome analyses and phyloge-netic
reconstruction, we shed light on the evolutionary history of TPA
and identify epidemiologically relevant haplotypes.

Due to the large background of host DNA, samples were en-
riched for treponemal DNA prior to Illumina sequencing13,14.
The resultant reads were mapped to the Nichols TPA reference
genome (RefSeq NC_021490; Supplementary Table 3)5,15.
Genomic coverage ranged from 0.13-fold to over 1,000-fold. As
expected, the highest mean coverage was found in strains propa-
gated in rabbits, while high variation in mean coverage was observed
in samples collected directly from patients (0.13-fold to 223-fold)
(Supplementary Table 2). This heterogeneity could potentially
affect our inferences. We therefore restricted the genome-wide ana-
lyses to the 28 samples where at least 80% of the genome was
covered by a minimum of three reads (highlighted in Supplementary
Table 2). Across the 28 samples, the average propor-
tions of genome coverage with at least 3-fold or 10-fold depth
were 97% and 82%, respectively (Supplementary Table 4).

De novo assemblies for the four highest covered syphilis swab
samples (NE17, NE20, CZ27 and AU15) and one Indonesian yaws
isolate (IND1) show no significant structural changes in the five
genomes (Fig. 1a; Supplementary Table 5), except for the deletion
in IND1 of gene TP1030, which potentially encodes a virulence-
factor16. The deletion was shared across all the yaws infection iso-
lates (Supplementary Methods), consistent with other studies17.

Before phylogenetic reconstruction we checked for signatures of
recombination. *T. pallidum* is considered to be a clonal species18,19,
but previous studies suggest recombinant genes in a Mexican syphilis
and a Bosnian bejel strain10,17. We screened for putative recombi-
nants across the 978 annotated genes in our 28 sequenced
samples, and the 11 publicly available genomes from laboratory
strains (Supplementary Table 3). Genes were selected as candidates
if they had unexpectedly high single nucleotide polymorphism
(SNP) densities, incongruent topologies with the genome-wide
tree and more than four homoplasies in a pair of branches
(Supplementary Methods). We identified four genes coding for
outer membrane proteins (Supplementary Table 6), one of which
(TP0136) is used in typing studies8.

After excluding the four putative recombinant genes, the genome
alignment for all 39 genomes contained 2,235 variable positions.
We used the Bayesian framework implemented in BEAST20 to
reconstruct a phylogenetic tree (Fig. 1b). The tree topology revealed
a marked separation between TPA and TPE/TEN (100% Bayesian
posterior support), with TPA forming a monophyletic lineage.

The distinction of the two lineages was robust, even with the
inclusion of putative recombinant genes (Supplementary Fig. 2). Anal-
yses of divergence between the two lineages yielded an
average mean distance of 1,225 nucleotide differences. By contrast,
within each of the lineages we found considerably less diversity
(124.6 average pairwise mutations within the TPA lineage and
200.2 within TPE/TEN). A heat map (Supplementary Fig. 3) to
show shared variation for pairs of samples with respect to the
Nichols reference genome confirms the divergence between the
lineages. The underlying SNP matrix yielded 443 SNPs specific to
TPA genomes and 1,703 to TPE/TEN genomes. Previous studies
have found cross-subspecies groupings when relying on a limited
set of markers21. Our results, incorporating genome-wide data
from clinical samples, not only establish a clear separation
between the two lineages, in agreement with studies examining
genomic data from rabbit propagated samples10,17, but also illustrate
the need for a careful choice of taxonomic markers when genome-
wide data are not available.

Using the sample isolation dates as tip calibration and applying the
Birth Death Serial Skyline model22, we obtained a mean evolu-
tionary rate of 3.6 × 10−8 (rate variance 3.8 × 10−9; 95% highest
posterior density (HPD) of 1.86 × 10−8 to 5.73 × 10−8). This estimate
is equivalent to a scaled mean rate of 6.6 × 10−7 substitutions per
site per year for the whole genome, in line with estimates for
other clonal human pathogens such as *Shigella sonnei* (6.0 × 10−7)
and *Vibrio cholerae* (O1 lineage; 8.0 × 10−7)23,24. Our divergence
analyses for TPA samples provide a time to the most recent
common ancestor (TMRCA) of less than 500 years ago (mean
calendar year 1744, 95% HPD 1611–1859; Fig. 1b).

Within the TPA lineage the samples group in two clades named
after the SS14 and Nichols reference genomes (with 100% and 82%
posterior probability values, respectively). The Nichols clade consists
almost exclusively of samples collected from patients in North America
from 1912 to 1986 and passed in rabbits before sequencing, with
the exception of one patient sample from 2013 (NE20). In contrast,
the SS14 clade has a geographically widespread distribution, encom-
passing European, North American and South American samples
collected from infections between 1951 and 2013. We investigated the
TPA clades further by generating a median-joining (MJ) network
to illustrate the mutational differences among the TPA samples
(Fig. 2a). As underscored by distances in the network, greater nucleo-
tide diversity is found within the Nichols clade (π = 0.05) than in
the SS14 clade (π = 0.01). Three closely related sequences derive
from the original Nichols sample isolated from the cerebrospinal
fluid of a patient in 1912 and propagated in laboratories in sub-
sequent decades: NIC_REF, the reference genome re-sequenced by
Petrovšová et al.23, and NIC-1 and NIC-2, which we sequenced follow-
ing independent propagation of the strains in Houston and Seattle,
respectively, during different time periods (Supplementary Table 1
and Supplementary Table 3). These three group together with
another three laboratory-propagated strains in a cluster labelled
Nichols-a (Fig. 2a), with a TMRCA in the mid-twentieth century
(Fig. 1a). The less diversified SS14 clade contains a dominant cen-
tral haplotype (labelled SS14-Ω) from which the other sequences
radiate (Fig. 2a). Critically, the cluster associated with the SS14-Ω
haplotype contains all but one of the recent patient samples from
1912 to 2013 (n = 17) that were captured and sequenced directly, in
addition to samples from 1977 (n = 1) and 2004 (n = 2). The genetic
variation within the SS14-Ω cluster is found primarily as single-
genot mutations (95.5%), with no evidence here for geographical
structuring. Bayesian analyses estimate coalescence for the SS14-Ω
center in 1964 (mean calendar year; 95% HPD 1950–1975; Fig. 1b), at a
time when incidence was reduced due to the introduc-
tion of antibiotics. The star-like topology of this cluster observed in
both the tree and the network is suggestive of a recent and rapid
clonal expansion.

To determine whether the dominance of SS14 clade sequences
applies across other countries for which genetic data are available,
we examined sequences from the widely typed TP0548 gene in
worldwide epidemiological studies11. Phylogenies for the TP0548
typing regions separate the SS14 from the Nichols clade for the
TPA samples, but do not distinguish the TPA and TPE/TEN
lineages (Supplementary Methods and Supplementary Fig. 4). Across 1,354 worldwide TP0548 sequences from clinical samples,
including the 78 from patients in this study, we found that 94% of
them grouped in the SS14 clade (Supplementary Tables 8 and 9
and Supplementary Fig. 5), consistent with a probable recent
spread of the epidemic cluster. The wide geographical distribution
of the SS14 clade establishes it as representative of the present
worldwide epidemic. Studies so far have focused on the Nichols
strain2,5,28, but our results indicate that further work on the SS14
clade is warranted.
Critically, typing of samples over multiple years in the Czech Republic, San Francisco, British Columbia and Seattle indicates that macrolide antibiotic resistance has increased over time1,12,27–29. We queried the presence of the two mutations (A2058G and A2059G) in the 23S ribosomal RNA (rRNA) genes associated with azithromycin resistance30,31. As observed in the MJ network, the resistance marker is a dominant characteristic of the SS14–Ω cluster (Fig. 2a), although it is also found in a recent patient sample (NE20) of the Nichols clade. Extending our analyses of the 23S rRNA gene to all sequenced samples from our study, with black circles for nodes with ≥96% posterior probabilities (PP), dark grey circles for nodes with 91–95% PP and light grey circles for nodes with 80–90% PP. Divergence date estimates (mean and 95% highest posterior density) for major well-supported TPA nodes are given in the legend.

The results here represent the first reported set of whole-genome sequences successfully obtained directly from syphilis patients, enabling us to disentangle evolutionary relationships at high resolution and paving the way for further clinical sequencing from current epidemics. Given our identification of putative recombinant genes in Treponema and previous reports on genes involved in homologous recombination32,33, further detailed analyses on the potential mechanisms of recombination will be necessary. Our phylogenetic reconstruction indicates that all TPA samples examined to date share a common ancestor that was infecting populations in the 1700s, within the early centuries of the modern era, and that was successful in leaving descendants until today. This date is posterior to the colonization of the Americas and therefore potentially compatible with the post-Columbian model for the emergence of syphilis in Europe. Nonetheless, our work does not exclude the possibility that older TPA lineages had previously existed in Europe but went extinct. Obtaining more patient sample genomes with high coverage could refine our detection of putative recombinants and our phylogenetic inferences. In addition, sequencing from ancient skeletal material would help to further ascertain the history of syphilis. Interestingly, we observed a time difference between the first reported syphilis outbreak in 1495 and the last common ancestor of modern strains dated to the 1700s. Although this difference could stem from imprecision in the divergence estimates, an alternative scenario is the eventual establishment of a specific lineage due to selection. For instance, it has been hypothesized that the symptoms of syphilis became less severe after the 18th century provided the context for the origin and propagation of a lineage that successfully outcompeted other lineages.

Critical to our epidemiological understanding of contemporary syphilis is our observation of an epidemic cluster (SS14–Ω) that emerged after the discovery of antibiotics. The relatively recent phylogenetic divergence of the SS14–Ω cluster and its global presence point to the emergence of a pandemic azithromycin-resistant cluster. The genome-wide data in this study will be useful to determine a suitable set of typing loci, because typing remains a more accessible method for most laboratories. Further characterization of the genomic diversity of TPA across the globe can prove

Figure 1 | De novo genome assemblies and phylogenetic reconstruction. a, De novo genome assembly for four syphilis patient samples and one yaws strain, with colour-coded geographic origin. Blank spaces correspond to gaps, overlapping with gene regions that are difficult to assemble from short reads such as the trp subfamilies and rRNA operons (regions shown in the outermost ring in grey). b, BEAST tree for the 39 genomes (excluding putative recombinant genes), with black circles for nodes with ≥96% posterior probabilities (PP), dark grey circles for nodes with 91–95% PP and light grey circles for nodes with 80–90% PP. Divergence date estimates (mean and 95% highest posterior density) for major well-supported TPA nodes are given in the legend.
instrumental in understanding the genetic and epidemiological basis for the spread of SS14-Ω strains.

Methods

Sample collection, DNA extraction and library preparation. Samples from 64 syphilis infections, 5 yaws infections and 1 bejel infection were collected from numerous countries across the globe (Supplementary Table 1). Syphilis infection samples were classified as either clinical or obtained from patients directly, or as laboratory strains if passaged in rabbits after isolation from patients. Clinical samples were obtained after swabbing lesions from patients at sexual health clinics, dermatological clinics or hospitals. Flocked swabs (from Copan Diagnostics) or nylon swabs were used according to local laboratory instructions. Laboratory strains were obtained as DNA extracts from Masaryk University (Brno, Czech Republic) and the University of Washington (Seattle, USA). DNA extractions were carried out in the participating laboratories using in-house protocols. At the University of Zurich the QIamp DNA mini kit and QIamp DNA blood min kit (Qiagen) were used following the manufacturer’s protocols.

Library preparation was conducted following a modified Illumina protocol for ancient DNA\textsuperscript{32,36}, at the University of Tübingen (Supplementary Methods). Libraries were barcoded with double indices.

Genome-wide enrichment and sequencing. Target enrichment for *Treponema pallidum* subsp. *pallidum* was carried out through two rounds of capture hybridization on a 1 million Agilent SureSelect array following the protocol detailed by Hodges and co-authors\textsuperscript{37}. The probes on the array were based on two reference genomes (Nichols, here abbreviated as NIC_REF, GenBank ID CP004010.2/RefSeq ID NC_010741.1 and SS14, GenBank ID CP000805.1/RefSeq ID NC_010741.1). High-throughput sequencing of the enriched libraries was performed on an illumina HiSeq 2500 platform.

Sequencing analyses and genome reconstruction. We applied EAGER\textsuperscript{38}, our own developed pipeline for read preprocessing (adapter clipping, merging of corresponding paired-end reads in the overlapping regions and quality trimming), mapping, variant identification and genome reconstruction, to all sequenced samples (for full details see Supplementary Methods). All reads (merged and unmerged) were treated as single-end reads and mapping was performed using the BWA-MEM algorithm\textsuperscript{34} with default parameters, using the Nichols genome as reference. Subsequently, we selected the samples that had at least 80% coverage of the Nichols genome and a minimum of three reads (n = 28 samples, Supplementary Table 2). For each of these samples, we used the Genome Analysis Toolkit (GATK)\textsuperscript{39} to generate a mapping assembly, applying the UnifiedGenotyper module of GATK to call reference bases and variants from the mapping. The reference base was called if the position was covered by at least three reads, the Phred-scaled genotype quality score of the call was at least 30, and at least 90% of the reads agreed with the reference. A variant position (SNP) was called if the position was covered by at least three reads, the genotype quality of the call was at least 30, and the minimum SNP allele frequency was 90% (detailed in the Supplementary Methods). If neither of the requirements for a reference base call nor the requirements for a variant call were met, the character ‘N’ was inserted at the respective position. For the generation of draft genome sequences we used an in-house tool (VCF2Genome), which reads a VCF file such as that produced by the GATK UnifiedGenotyper and incorporates—for each row and thus for each call—one nucleotide into the new draft sequence.

To apply our analysis pipeline also to those samples for which complete genomic sequences are available in GenBank (Supplementary Table 3), we produced artificial reads in these cases using an in-house tool (Genome2Reads) and then applied the same mapping, SNP calling and genome reconstruction procedure as for the sequenced samples to obtain consistent and comparable results.

To investigate conservation of structure and gene order in the genomes, in addition to the mapping assembly we also performed a de novo assembly for the five samples with highest coverage (Supplementary Table 5). Our de novo assembly pipeline started with the merged reads, and in a first step used the short read assembler software SOAPdenovo\textsuperscript{32} using ten different k-mer sizes (k = 37 + 10i, where i = 0, ..., 9). Different k-mer sizes were used because merging of read pairs into one single read results in very different lengths (between 30 and 190 bases). Next, all input reads were mapped back against the resulting contigs using
were examined and visualized through an MJ network analysis in Network 4.6 and obtained is in line with the number of mutations that differed between the samples of the focus of this study and therefore more extensively sampled, we report mean branch lengths for each of the genes, excluding the four putative recombinant genes (2,235 positions). Furthermore, we produced a maximum likelihood tree topology of the draft genome in TREE-PUZZLE v5.2 (refs 41,42). Genes for which both the expected likelihood weight and the Shimodaira–Hasegawa test rejected the genome (P < 0.05) were examined more closely. Third, genes within which we identified a minimum of five homoplasies (identical mutations in separate lineages) in at least two branches of the tree were marked as putative recombinants (Supplementary Table 6).

Genome-wide variation and phylogenetic analyses. We investigated genome-wide patterns of polymorphism and divergence using MEGA 6.0 (ref. 45) and DnaSP v.5.10 (ref. 46) to compute various measures of diversity including the average pairwise nucleotide differences, Nei’s pi (n) and the number of singletons in each group. We also estimated the number of SNPs private to particular groups. A comparison of the TPE and TPE/TEN genomes revealed between 1 (NIC1) and 339 (AR2) SNPs observed in the TPA samples and between 1,091 (GHAI) and 1,443 (Bosnia A) SNPs in the TPE/TEN-strains (Supplementary Table 4). Furthermore, we produced a heat map that displays the recombinant status of SNPs with respect to the Nichols reference genome that any two genomes share (Supplementary Fig. 3).

The molecular clock hypothesis was tested with maximum likelihood analysis in MEGA 6.0 (ref. 45). Tests were conducted for all TPA, TPE and TEN genomes (39 samples) using (1) multiple whole genome alignments (2) alignments with only the variable positions, in both cases excluding the four putative recombinant genes. The molecular clock hypothesis was rejected at the 5% significance level. Bayesian phylogenetic trees were produced in BEAST 2.3 (ref. 20) for the 28 sequenced samples and the 11 published samples. We compared the trees generated with the alignment of all variable positions in the TPA, TPE and TEN genomes (1,306 positions) and the tree generated with the set of variable positions after excluding the four putative recombinant genes (2,235 positions). Additionally, rooted trees were generated with maximum parsimony by including Treponema paradoxus strain strain (NC_015714) as the outgroup.

As a calibration for the BEAST (Bayesian evolutionary analysis sampling trees) trees that we obtained, this is that the isolation age of the modern T. pallidum (15 million years) is known with precision, we provided a range (for NIC, REF, NIC1, NIC2, and GAU). The two demographic models (coalescent tree prior under Constant Size and the Birth-Death Serial Skyline model (BDSS)) resulted in consistent parameter estimates. The relaxed clock model was chosen over the strict clock model based on marginal likelihood values obtained with PathSampler47,48. We report estimates for the five combined RDSS model runs, each with the following specifications: uncorrelated lognormal relaxed clock model, generalized time reversible plus gamma substitution model, and 100 million generations with parameter sampling every 10,000 generations. We used Tracer 1.6 (ref. 48) to assess convergence and suitable burn-in periods (see Supplementary Methods for full details). The maximum clade credibility tree was visualized and edited using Figtree v.1.4.2 (ref. 49). Because TPA samples are the focus of this study and therefore more extensively sampled, we report mean branch rate and divergence estimates for the TPA lineage. The mean branch rate estimate obtained is in line with the number of mutations that differed between the samples NIC, REF and NIC2 (n = 15), which were isolated 15–20 years ago and propagated continuously in rabbit propagation. We also checked that a run with the same specifications but with only TPA samples (n = 31) produced consistent results.

The phylogenetic relationships among the closely related TPA samples (n = 31) were examined and visualized through an MJ network analysis in Network 4.6 and Network Publisher (available at www.linux-engineering.com) using all variable positions after excluding the putative recombinant loci and sites with missing data (resulting in a total of 628 variable positions).

Clade classification

Samples from this study. From the 70 TPA, TPE and TEN samples sequenced in this study, 28 fulfilled our criteria for genome-wide analyses (minimum 80% genome covered with at least 3 reads). For the remaining 42 samples, we implemented two classification strategies (for details, see Supplementary Methods). First, we generated a new clade prediction strategy based on NGS reads to classify the genomes according to lineage (TPA or TPE/TEN) and within the TPE lineage, as part of the SS14 or the Nichols clade. Second, we used a classification scheme based on the TP0548 gene. For the TP0548 classification scheme we carried out PCR and Sanger sequencing of the TP0548 region following the protocols and primers of ref. 30. SNPs in the TP0548 typing regions enable the distinction of an SS14 clade versus a Nichols clade. Indels enable the classification of TPE and TEN. Our NGS prediction strategy was congruent with the TP0548 classification scheme wherever prediction strength was above 0.4, with the exception of one TEN sample (detailed in the Supplementary Methods).

Samples from typing studies. We put together all publicly available TP0548 sequences obtained in typing studies of syphilis infections around the world121–126. We also incorporated TP0548 sequences obtained for 36 Argentinian clinical samples by LGV at the University of Buenos Aires, Argentina (Supplementary Table 9). All TP0548 sequences were classified as part of the SS14 clade or part of the Nichols clade based on an ML tree (Supplementary Fig. 5). Subtypes were distinguished through visual inspection (Supplementary Table 9).

Antibiotic resistance. The two mutations associated with resistance to the macrolide azithromycin, A2058G and A2059G on the 23S ribosomal RNA operon (with positions referring to coordinates in the 23S ribosomal RNA gene of E. coli) were investigated in separate analyses. Because the operon contains two copies of the gene, mapping of reads with BWA was carried out independently for each of the genes, including a flanking region of 200 bases on both the 5′ and 3′ end of each gene. Following variant calling, the presence/absence of each of the two mutations was recorded for each sample. The two operons could not, however, be distinguished.

In addition, we used primers specific for each of the two operons to carry out PCR amplifications as well as Sanger sequencing on the samples, following the protocol in ref. 30. Details on the samples sequenced, as well as resistance or sensitivity to the macrolide as determined by the presence or absence of the associated mutations, are provided in Supplementary Table 8.

Data availability. All samples sequenced in this study are available in an NCBI Bioproject under accession code PRJA33497. Raw sequencing reads in FASTQ format were uploaded to the Short Read Archive (SRA). All accession codes are listed in Supplementary Table 2. Codes for the in-house scripts developed for some of the analyses are available upon request from the authors. All raw read files have been deposited in the trace archive of the NCBI Sequence Read Archive under accession code SRP072886.

Received 24 June 2016; accepted 3 November 2016; published 5 December 2016

References

6. Fernandez de Oviedo y Valdes, G. Sumario de la natural historia de las Indias (Fondo de Cultura Economico, 1526).


49. Rambaut, A. FigTree v.1.4.2. (2014). http://tree.bio.ed.ac.uk/software/figtree/