

Article

# Origin and Global Expansion of *Mycobacterium tuberculosis* Complex Lineage 3

Yassir A. Shuaib <sup>1,2,3,4,\*</sup> , Christian Utpatel <sup>2,3</sup> , Thomas A. Kohl <sup>2,3</sup>, Ivan Barilar <sup>2,3</sup>, Margo Diricks <sup>2</sup> , Nadia Ashraf <sup>2,3</sup>, Lothar H. Wieler <sup>4,5</sup>, Glennah Kerubo <sup>6</sup>, Eyob A. Mesfin <sup>7</sup>, Awa Ba Diallo <sup>8,9</sup> , Sahal Al-Hajoj <sup>10,11</sup>, Perpetua Ndung'u <sup>12</sup>, Margaret M. Fitzgibbon <sup>13,14</sup>, Farzam Vaziri <sup>15</sup>, Vitali Sintchenko <sup>16,17</sup>, Elena Martinez <sup>16</sup>, Sofia O. Viegas <sup>18</sup>, Yang Zhou <sup>19</sup>, Aya Azmy <sup>20</sup>, Khaled Al-Amry <sup>21</sup>, Sylvain Godreuil <sup>22</sup>, Mandira Varma-Basil <sup>23</sup>, Anshika Narang <sup>23</sup>, Solomon Ali <sup>24</sup>, Patrick Beckert <sup>2,3</sup>, Viola Dreyer <sup>2,3</sup>, Mwila Kabwe <sup>25</sup> , Matthew Bates <sup>26</sup>, Michael Hoelscher <sup>27,28</sup> , Andrea Rachow <sup>27,28</sup>, Andrea Gori <sup>29</sup> , Emmanuel M. Tekwu <sup>30</sup>, Larissa K. Sidze <sup>30</sup>, Assam A. Jean-Paul <sup>30</sup>, Veronique P. Beng <sup>30</sup>, Francine Ntouni <sup>31,32</sup>, Matthias Frank <sup>33</sup>, Aissatou Gaye Diallo <sup>9</sup>, Souleymane Mboup <sup>9</sup> , Belay Tessema <sup>34</sup> , Dereje Beyene <sup>35</sup>, Sadiq N. Khan <sup>36</sup>, Roland Diel <sup>37,38</sup>, Philip Supply <sup>39</sup>, Florian P. Maurer <sup>40,41</sup> , Harald Hoffmann <sup>42,43</sup>, Stefan Niemann <sup>2,3,†</sup> and Matthias Merker <sup>2,3,44,\*</sup> 

- 1 College of Veterinary Medicine, Sudan University of Science and Technology, P.O. Box 204 Hilat Kuku, Khartoum North 13321, Sudan
- 2 Molecular and Experimental Mycobacteriology, Research Center Borstel, 23845 Borstel, Germany; cutpatel@fz-borstel.de (C.U.); tkohl@fz-borstel.de (T.A.K.); ibarilar@fz-borstel.de (I.B.); mdiricks@fz-borstel.de (M.D.); nashraf@fz-borstel.de (N.A.); pbeckert@fz-borstel.de (P.B.); vdreyer@fz-borstel.de (V.D.); sniemann@fz-borstel.de (S.N.)
- 3 German Center for Infection Research, Partner Site Hamburg-Lübeck-Borstel-Riems, 23845 Borstel, Germany
- 4 Institute of Microbiology and Epizootics, Freie Universität Berlin, 14163 Berlin, Germany; wielerlh@rki.de
- 5 Robert Koch Institute, 13353 Berlin, Germany
- 6 Department of Medical Microbiology and Parasitology, School of Medicine, Kenyatta University, Nairobi 43844-00100, Kenya; glenakerubo@gmail.com
- 7 Ethiopian Public Health Institute, Addis Ababa 1242, Ethiopia; eyob2001@gmail.com
- 8 Mycobacteria Unit, Bacteriology-Virology Laboratory, CHU Aristide Le Dantec, Dakar 11000, Senegal; awa1.diallo@ucad.edu.sn
- 9 Institut de Recherche en Santé de Surveillance Epidémiologique et de Formation (IRESSEF), Diamniadio, Dakar 24600, Senegal; aissatou.gaye@iressef.org (A.G.D.); souleymane.mboup@iressef.org (S.M.)
- 10 King Faisal Specialist Hospital and Research Centre, Riyadh 11211, Saudi Arabia; hajoj@kfshrc.edu.sa
- 11 College of Medicine, Al Faisal University, Riyadh 11533, Saudi Arabia
- 12 Department of Medical Laboratory Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi 62000-00200, Kenya; perpetuaNdungu@yahoo.com
- 13 Irish Mycobacteria Reference Laboratory, St James' Hospital, D08 NHY1 Dublin, Ireland; mfitzgibbon@stjames.ie
- 14 Department of Clinical Microbiology, Trinity College Dublin, St James's Hospital, D08 NHY1 Dublin, Ireland
- 15 Department of Mycobacteriology and Pulmonary Research, Microbiology Research Center (MRC), Pasteur Institute of Iran, Tehran 1316943551, Iran; f\_vaziri@pasteur.ac.ir
- 16 Institute of Clinical Pathology and Medical Research, NSW Health Pathology, Sydney 2145, Australia; vitali.sintchenko@sydney.edu.au (V.S.); elena.martinez@health.nsw.gov.au (E.M.)
- 17 Marie Bashir Institute for Infectious Diseases and Biosecurity, University of Sydney, Sydney 2145, Australia
- 18 Instituto Nacional de Saúde, P.O. Box 1120, Marracuene 1220, Mozambique; viegas\_sofia@hotmail.com
- 19 Chinese Center for Disease Control and Prevention, Beijing 102206, China; carolynzy@hotmail.com
- 20 Central Public Health Laboratories, Clinical Microbiology Department, Ministry of Health, Cairo 11835, Egypt; ayazmy@yahoo.com
- 21 Department of Microbiology, Faculty of Veterinary Medicine, Cairo University, Giza 12211, Egypt; kelamry@daralfarouk.com.eg
- 22 Laboratoire de Bactériologie, CHU de Montpellier, MIVEGEC, Université de Montpellier—CNRS-IRD, 34295 Montpellier, France; godreuil@yahoo.fr
- 23 Department of Microbiology, Vallabhbhai Patel Chest Institute, University of Delhi, Delhi 110007, India; mandirav2001@yahoo.com (M.V.-B.); anshikanarang@gmail.com (A.N.)
- 24 St. Paul's Hospital Millennium Medical College, Addis Ababa 1000, Ethiopia; solomon.ali@ju.edu.et
- 25 Department of Pharmacy and Biomedical Science, La Trobe Institute for Molecular Science, La Trobe University, Bendigo 3552, Australia; mwila\_kabwe@yahoo.com
- 26 School of Life and Environmental Sciences, University of Lincoln, Lincoln LN6 7TS, UK; matthew.bates@ucl.ac.uk



**Citation:** Shuaib, Y.A.; Utpatel, C.; Kohl, T.A.; Barilar, I.; Diricks, M.; Ashraf, N.; Wieler, L.H.; Kerubo, G.; Mesfin, E.A.; Diallo, A.B.; et al. Origin and Global Expansion of *Mycobacterium tuberculosis* Complex Lineage 3. *Genes* **2022**, *13*, 990. <https://doi.org/10.3390/genes13060990>

Academic Editor: Isabel Henriques

Received: 2 April 2022

Accepted: 19 May 2022

Published: 31 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- <sup>27</sup> Division of Infectious Diseases and Tropical Medicine, University Hospital, LMU Munich, 80802 Munich, Germany; hoelscher@lrz.uni-muenchen.de (M.H.); rachow@lrz.uni-muenchen.de (A.R.)
- <sup>28</sup> German Centre for Infection Research (DZIF), Partner Site Munich, 80802 Munich, Germany
- <sup>29</sup> Infectious Diseases Unit, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Centre for Multidisciplinary Research in Health Science (MACH), University of Milan, 20122 Milan, Italy; andrea.gori@unimi.it
- <sup>30</sup> Laboratory for TB Research and Pharmacology, Biotechnology Centre of Nkolbisson, University of Yaoundé I, Yaoundé 8698, Cameroon; etekwu@yahoo.fr (E.M.T.); lsidze@yahoo.com (L.K.S.); assamjean@yahoo.fr (A.A.J.-P.); v.penlap@yahoo.fr (V.P.B.)
- <sup>31</sup> Faculty of Medicine, Institute for Tropical Medicine, University of Tübingen, 72074 Tübingen, Germany; fnttoui@hotmail.com
- <sup>32</sup> Fondation Congolaise Pour la Recherche Médicale (FCRM), Villa D6, Campus WHO-AFRO, Brazzaville 2672, Democratic Republic of the Congo
- <sup>33</sup> Allgemeinmedizin, Berliner Ring 39, 72076 Tübingen, Germany; matthias.frank@uni-tuebingen.de
- <sup>34</sup> Department of Medical Microbiology, College of Medicine and Health Sciences, University of Gondar, Gondar P.O. Box 196, Ethiopia; bt1488@yahoo.com
- <sup>35</sup> Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Addis Ababa P.O. Box 1176, Ethiopia; dereje.bd@gmail.com
- <sup>36</sup> Department of Medical Lab Technology, University of Haripur, Haripur 22621, Pakistan; snkhan@uoh.edu.pk
- <sup>37</sup> Institute for Epidemiology, University Medical Hospital Schleswig-Holstein, 24105 Kiel, Germany; roland.diel@epi.uni-kiel.de
- <sup>38</sup> Lung Clinic Grosshansdorf, Airway Research Center North (ARCN), German Center for Lung Research (DZL), 22927 Großhansdorf, Germany
- <sup>39</sup> CNRS, INSERM, CHU Lille, U1019-UMR 9017-CIIL-Center for Infection and Immunity of Lille, Institut Pasteur de Lille, Université de Lille, 59000 Lille, France; philip.supply@ibl.cnrs.fr
- <sup>40</sup> National and WHO Supranational Reference Center for Mycobacteria, Research Center Borstel, Leibniz Lung Center, 23845 Borstel, Germany; fmaurer@fz-borstel.de
- <sup>41</sup> Institute of Medical Microbiology, Virology and Hygiene, University Medical Center Hamburg-Eppendorf, 20251 Hamburg, Germany
- <sup>42</sup> Institute of Microbiology and Laboratory Medicine (IML), WHO/GLI-Supranational Reference Laboratory of Tuberculosis, Robert-Koch-Allee 2, 82131 Gauting, Germany; h.hoffmann@imlred.de
- <sup>43</sup> SYNLAB Gauting, SYNLAB MVZ Dachau, 82131 Gauting, Germany
- <sup>44</sup> Evolution of the Resistome, Research Center Borstel, 23845 Borstel, Germany
- \* Correspondence: vet.aboamar@gmail.com (Y.A.S.); mmerker@fz-borstel.de (M.M.)
- † These authors contributed equally to this manuscript.

**Simple Summary:** Tuberculosis still causes 1.5 million deaths annually and is mainly caused by *Mycobacterium tuberculosis* complex strains belonging to three evolutionary modern lineages (Lineages 2–4). While Lineage 2 and Lineage 4 virtually conquered the world, Lineage 3 is particularly successful in Northern and Eastern Africa, as well as in Southern Asia, the suspected evolutionary origin of these strains. Here, we sought to understand how Lineage 3 strains came to the African continent. To this end, we performed routine genotyping to characterize over 2500 clinical isolates from 38 countries. We then selected a representative collection of 373 isolates for a whole-genome analysis and a modeling approach to infer the geographic origin of different sublineages. In fact, the origin of Lineage 3 could be located in India, and we found evidence for independent introductions of four distinct sublineages into North/East Africa, in line with known ancient exchanges and migrations between both world regions. Our study illustrates that the evolutionary history of humans and their pathogens are closely connected and further provides a systematic understanding of the genomic diversity of Lineage 3, which could be important for the development of new tuberculosis vaccines or new therapeutics.

**Abstract:** *Mycobacterium tuberculosis* complex (MTBC) Lineage 3 (L3) strains are abundant in world regions with the highest tuberculosis burden. To investigate the population structure and the global diversity of this major lineage, we analyzed a dataset comprising 2682 L3 strains from 38 countries over 5 continents, by employing 24-loci mycobacterial interspersed repetitive unit-variable number of tandem repeats genotyping (MIRU-VNTR) and drug susceptibility testing. We further combined whole-genome sequencing (WGS) and phylogeographic analysis for 373 strains representing the global L3 genetic diversity. Ancestral state reconstruction confirmed that the origin of L3 strains is located in Southern Asia and further revealed multiple independent introduction events into

North-East and East Africa. This study provides a systematic understanding of the global diversity of L3 strains and reports phylogenetic variations that could inform clinical trials which evaluate the effectivity of new drugs/regimens or vaccine candidates.

**Keywords:** *Mycobacterium tuberculosis*; MTBC; Lineage 3; back to Africa

## 1. Introduction

With nearly 10 million incident cases and 1.5 million deaths worldwide in 2020, tuberculosis (TB) remains the leading cause of death among humans due to a single pathogen [1]. The vast majority of new TB cases each year occur in South-East Asia (43%), Africa (25%), and the Western Pacific (18%), posing a significant public health risk to the affected countries and health care systems [1]. As an additional challenge for TB control, in 2020, nearly half a million cases of rifampicin (RMP)-resistant (RR) TB were reported, out of which 78% were multidrug resistant (at least isoniazid (INH) and RMP resistant) [1].

Bacteria of the *Mycobacterium tuberculosis* complex (MTBC), the causative agents of TB, are classified into eight human-adapted (L1–L9) and animal-adapted phylogenetic lineages [2–5]. Globally, strains of L2, L3, and L4 are the most widely spread [6–8]. The epidemic success of L2 strains in Eurasia has been associated with drug resistance and hypervirulence, whereas L4 strains appear to be composed of ecologically distinct sublineages. These L4 sublineages include geographically widespread generalist clades causing disease in many different human populations, and geographically restricted specialist clades, closely associated with their sympatric host populations [6–9].

Although L3 strains constitute a major TB burden for high-incidence regions in South Asia, as well as North and East Africa, and are also potential drivers of the MDR TB epidemic in some parts of the world [2,3,10–14], a clear understanding of the population structure of this main MTBC lineage is lacking. Previous studies suggested an origin for L3 strains in South Asia, followed by dispersal to North and East Africa, likely via close economic and cultural interactions between South-East Asia and East Africa in the Common Era [10,15]. However, no study is available that combines a comprehensive phylogenetic framework with phylogeographic inference to provide robust insights into the genetic background and global spread of L3 strains.

To address this knowledge gap, we leveraged a collection of 2682 clinical L3 strains from 38 countries to define the global population structure and phylogenetic roots of L3 strains. We further investigated the geographical dispersal of L3 strains and determined their potential associations with drug-resistant TB.

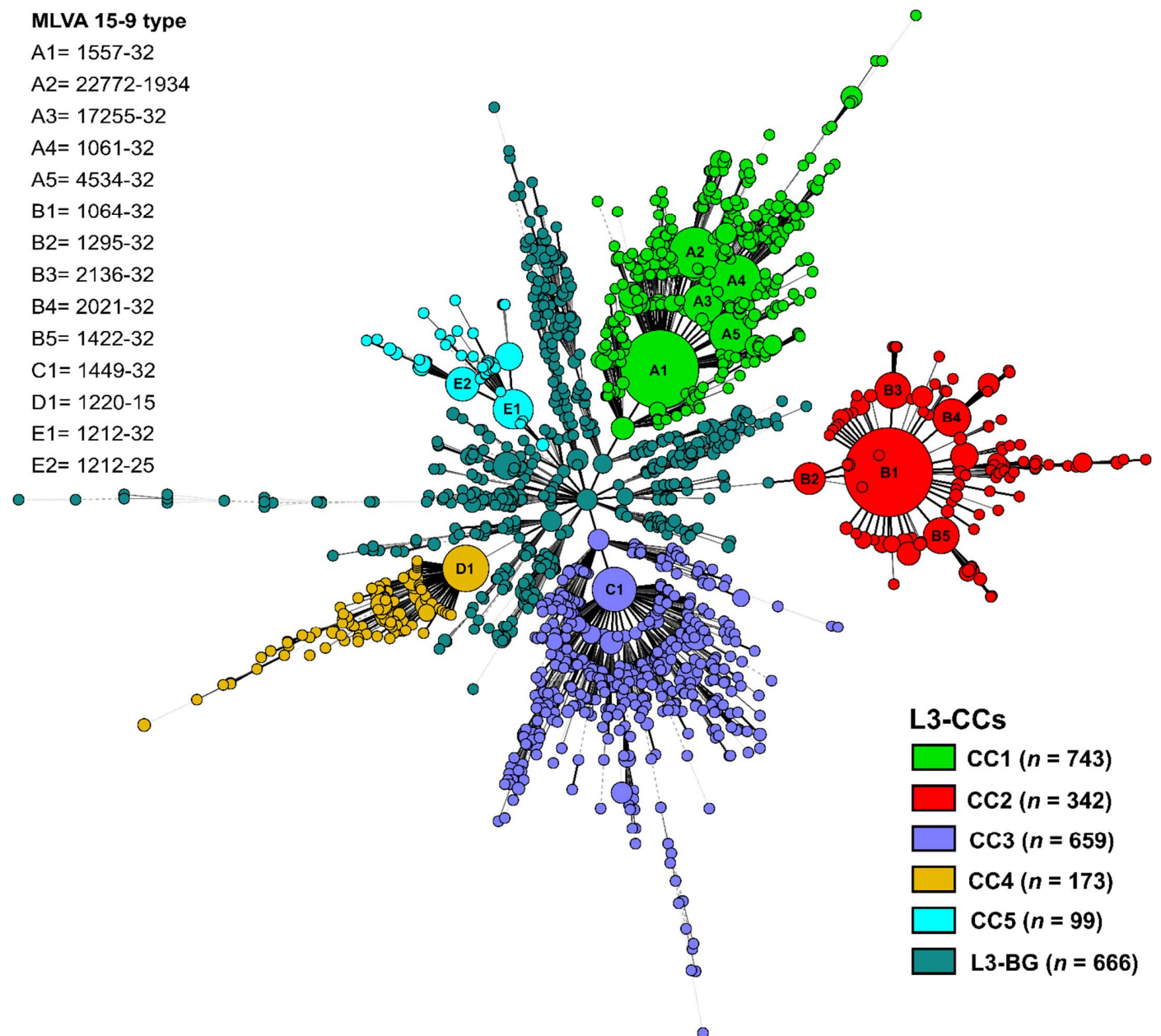
## 2. Results

### 2.1. Global Population Structure of MTBC L3 Strains

In this study, we analyzed a total of 2682 clinical L3 strains (one per patient) from 38 different countries, of which 25.9% ( $n = 695$ ) originate from 16 countries in Africa, 32.8% ( $n = 881$ ) from 12 countries in Asia, 18.9% ( $n = 506$ ) from 8 countries in Europe, 12.6% ( $n = 337$ ) from Canada, and 9.8% ( $n = 263$ ) from Australia (Supplementary Table S1). Inclusion criteria were a 24-loci MIRU-VNTR pattern classified as L3 (i.e., Delhi/CAS) using the MIRU-VNTR nomenclature at <http://www.miru-vntrplus.org>, accessed on 6 August 2020. Overall, 63.8% (1710/2682) of the L3 strains analyzed were either from population-based or cross-sectional studies, of which 1036/2,682 (38.6% overall) had phenotypic drug susceptibility testing (pDST) data available (Supplementary Table S1).

We first constructed a minimum spanning (MS) tree based on 24-loci MIRU-VNTR data of 2682 clinical strains to define major L3 clonal complexes (L3-CC), assigned a nomenclature code (MLVA MTBC 15–9), and estimated the clustering rate among identified L3-CCs (i.e., proportion of strains with identical genotypes) [16,17].

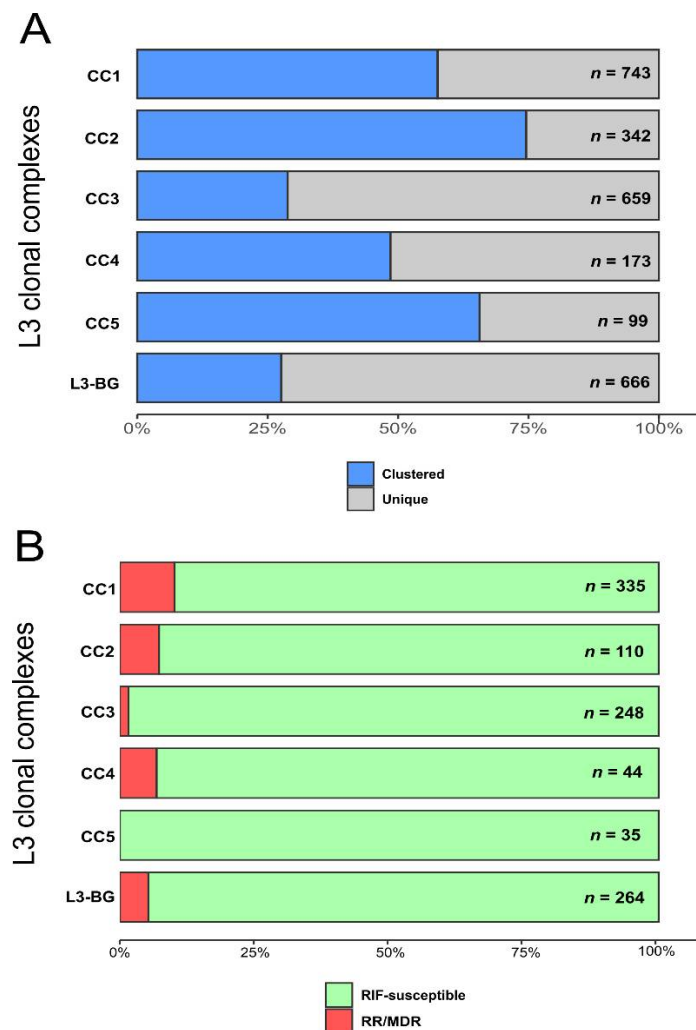
In total, we identified 1596 different MLVA 15-9 codes, of which 1322 were assigned to a single strain only (Supplementary Table S1). Overall, we identified 1206 strains in 274 clusters, with cluster sizes varying from 2 to 103 strains, including the largest cluster in each of the five L3-CCs: A1 ( $n = 78$ ), B1 ( $n = 103$ ), C1 ( $n = 25$ ), D1 ( $n = 29$ ), E1 ( $n = 21$ ) (Figure 1). We then defined L3-CCs based on the topology of the MS tree, which reveals the structure of relatedness of genotyping patterns by minimizing the weights of the edges between genotypes (Figure 1). Each of the five major above-defined L3-CCs appears as one parental node surrounded by successive layers of descending genotypes. Genotype groups showing a more scattered distribution in the MS tree and lacking a clearly visible structure likely represent the genetic background of L3 strains and were not further assigned to a clonal complex and in the following are termed L3-BG, i.e., L3 genetic background (Figure 1).



**Figure 1.** Clonal complexes (CCs) of MTBC Lineage 3 strains. Minimum spanning (MS) tree based on 24-loci MIRU-VNTR data of 2682 clinical MTBC L3 strains originating from 38 countries. Five L3 clonal

complexes (L3-CCs) and diverse strains representing the L3 background diversity (L3-BG) are color coded. Node size reflects the number of strains with identical 24-loci MIRU-VNTR patterns. Branch length proportional to the number of allele differences between two nodes. Solid lines indicate 1, 2, or 3 allele differences, gray dashed lines represent 4 allele differences, and gray dotted lines represent 5 or more allele differences.

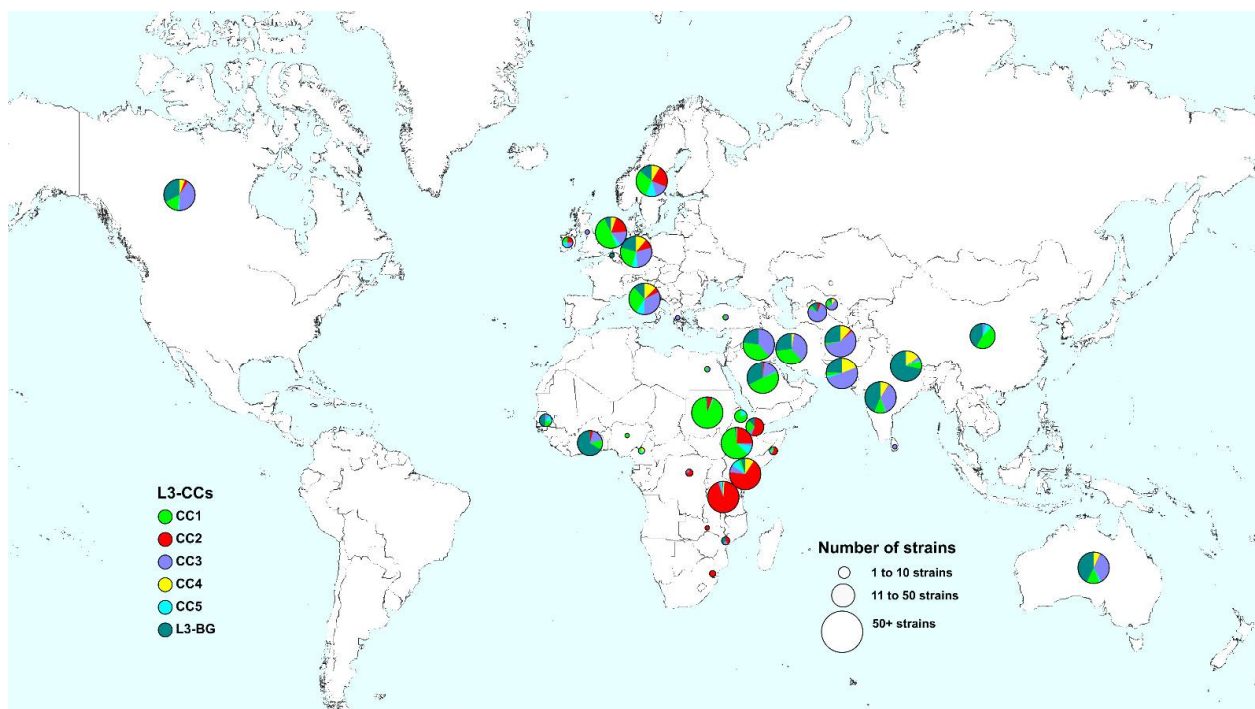
Stratified to the five L3-CCs, the proportions of clustered strains (i.e., two or more MTBC strains from different patients and exhibiting identical genotyping patterns [17]) ranged from 28.8% (190/659) in L3-CC3 to 7.6% (255/342) in L3-CC2 (Figure 2a). Strains of L3-CC1, L3-CC4, and L3-CC5 showed intermediate proportions of clustering rates, with L3-CC3 exhibiting the lowest proportion of clustered strains (Fisher's exact test  $p < 0.001$ ).



**Figure 2.** Molecular clusters and drug resistance among L3 clonal complexes (CC). Proportions of clustered strains (A) and rifampicin-resistant (RR) and multidrug-resistant (MDR) strains (B) within L3 clonal complexes (CC) and within diverse strains representing the genetic background (L3-BG).

We further considered pDST data for strains derived from population-based or cross-sectional studies (1036/2682, 38.6%) and with available pDST results for INH and RIF. Of these, 93.9% (973/1,036) were RIF susceptible, and 6.1% (63/1036) were either RR or MDR. Proportions of RR/MDR strains among L3-CC1 to L3-CC5 were as follows: 10.1% (34/335), 7.3% (8/110), 1.6% (4/248), 6.8% (3/44), and 0.0% (0/35), respectively, with no significant differences (Fisher's exact test  $p \geq 0.059$ ) (Figure 2b, Supplementary Table S1). Half ( $n = 31$ ) of the RR/MDR strains occurred in three countries in Northern and Eastern Africa, namely Ethiopia, Kenya, and Sudan.

Out of 274 clusters, 137 (50%) comprised strains that were found in 2 to 12 different countries, suggesting a relatively recent cross-border spread of different L3 strains (Supplementary Table S2). The spatial distribution of L3-CCs shows certain L3-CCs dominating in Southern Asia and Northern and Eastern Africa. While Europe shows a higher strain diversity, Australia and Northern America resemble the distribution of L3-CCs in Southern Asia (Figure 3, Supplementary Table S3). Specifically, we observed high proportions of diverse strains representing L3-BG, as well as L3-CC3, in Southern Asia, namely Afghanistan, India, Pakistan, and Nepal, as well as in Iran, Australia, and Canada. L3-CC1 strains are dominant in Northern Africa and the Arabian Peninsula, whereas L3-CC2 strains are highly prevalent in Eastern Africa. In Ghana and in Senegal, we identified mostly L3-BG strains (Figure 3, Supplementary Table S3).



**Figure 3.** Global geographical distribution of MTBC Lineage 3 clonal complexes (L3-CCs). Each pie chart represents a country, and the size of each pie chart is proportional to the number of strains. L3-CCs and L3-BG are color coded as in Figure 1.

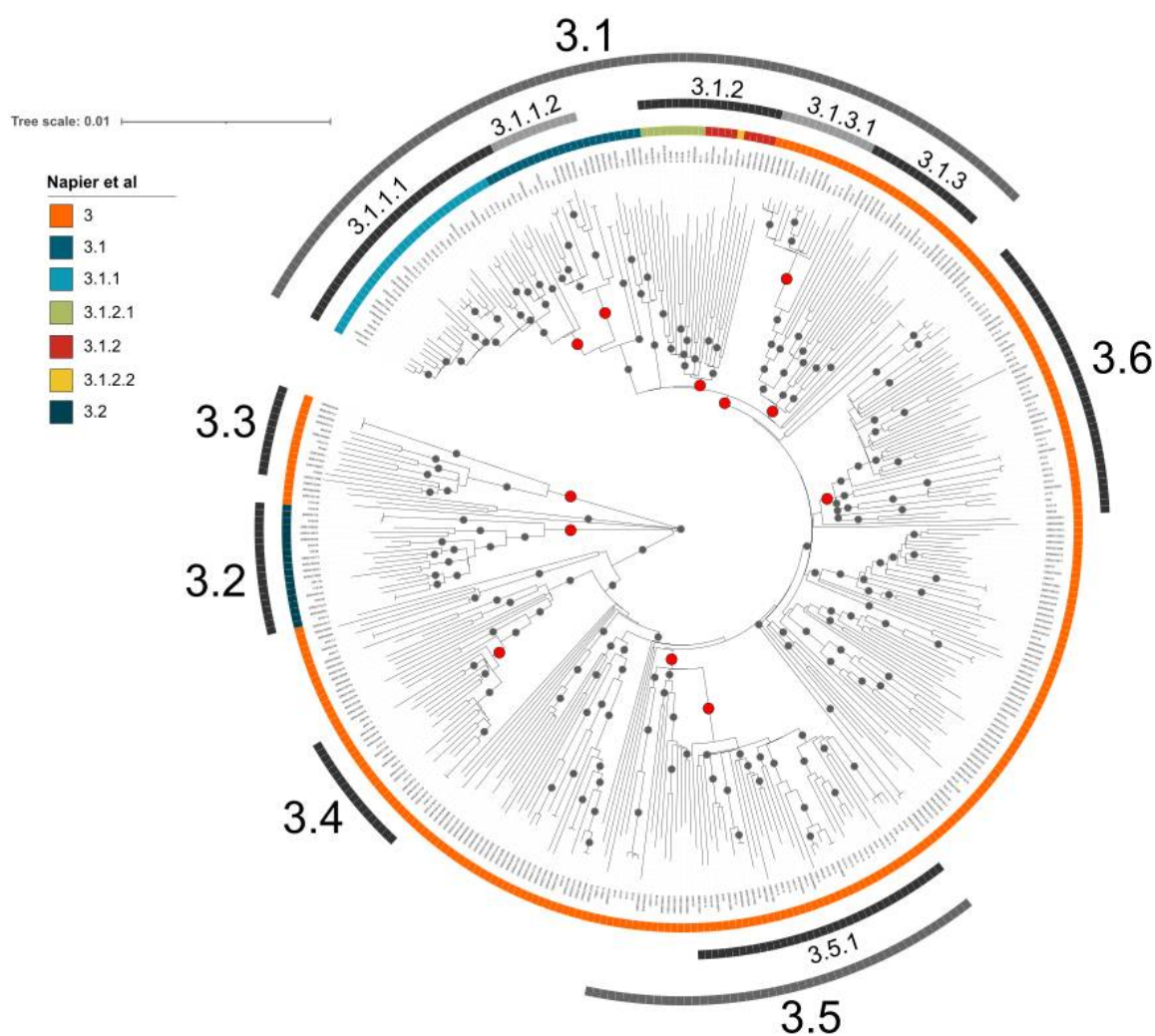
## 2.2. Genome-Based Phylogeny of MTBC L3

To analyze the global phylogeny and geographical origin of MTBC L3 strains at a higher resolution, we selected 152 representative strains from all L3-CCs and L3-BG for an initial WGS analysis (Supplementary Figure S1, Supplementary Table S4). The selected strains comprised 4.6% (34/743) of L3-CC1, 7.6% (26/342) of L3-CC2, 6.4% (42/659) of L3-CC3, 5.8% (10/173) of L3-CC4, 6.1% (6/99) of L3-CC5, and 5.1% (34/666) L3-BG strains (Supplementary Table S4). To represent the global diversity of the L3 population as best as possible, we included 51 strains from 10 African countries, 84 strains from 10 Asian countries, and 17 strains from Europe, and included five Lineage 2 strains as an outgroup. Next, we calculated a maximum-likelihood (ML) tree based on a concatenated sequence alignment of 12,262 single-nucleotide polymorphisms (SNPs) (Supplementary Figure S2).

In the ML phylogeny, we defined 12 clades that correlated well with the MIRU-VNTR-based L3-CCs and/or country of origin, and we defined signature SNPs specific for all strains of each clade (Supplementary Table S4). We then screened for these signature SNPs in a previously published L3 strain dataset from Napier et al. ( $n = 3406$ ) [18]. Overall, 2916/3406 (85.6%) L3 strains in this dataset could be classified with the signature SNPs defined above, while 490/3406 strains (14.4%) remained unclassified. To further extend

the genetic diversity in our L3 reference set and the related WGS-based phylogeny, we combined our initial L3-CC- and country-of-origin-based selection ( $n = 152$ ), with 221 strains from Napier et al., including 107 strains with and 114 strains without signature SNPs, respectively (Supplementary Table S5).

The resulting ML phylogeny then comprised 373 clinical L3 isolates from 38 countries (Figure 4). The L3 tree topology is characterized by a star-like expansion with short internal and long terminal branches and several multifurcations (i.e., more than two branches descending from one node). The highlighted clades and the overall topology remained stable using Lineage 2 strains as an outgroup (Supplementary Figure S2). There are very few indications for genetic bottleneck events (i.e., long internal branches) in the evolutionary history of L3 strains. The phylogeny is thus rather indicative of an ancient population expansion into different parts of the globe originating from a single common ancestor.

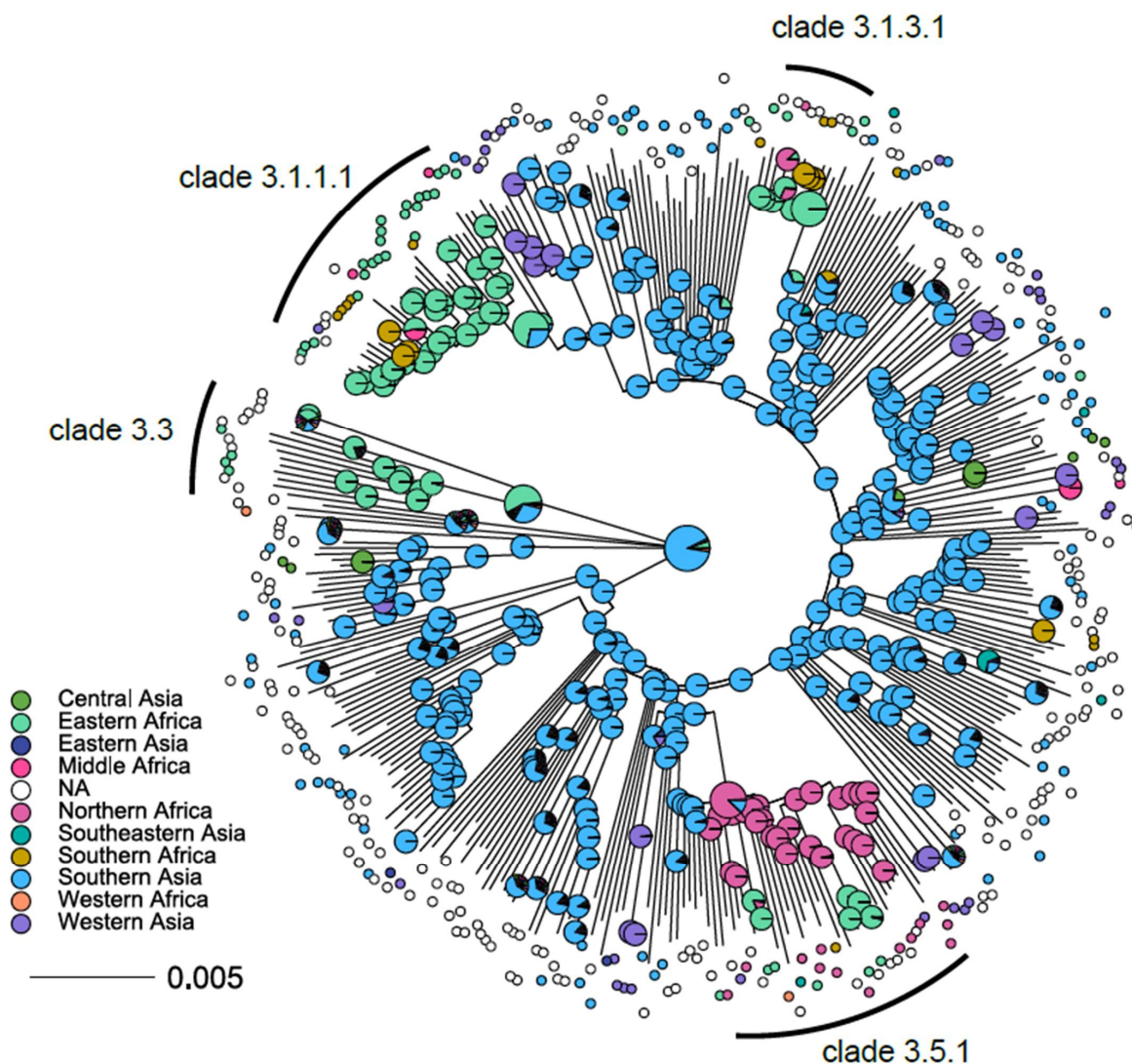


**Figure 4.** Global phylogeny of MTBC Lineage 3 strains. Midpoint-rooted maximum-likelihood (ML) tree based on 26,536 concatenated single-nucleotide polymorphisms (SNPs) and 373 L3 strains. The first outer ring provides a color code for the L3 subgroups defined by Napier et al. (18). Subsequent outer rings indicate clades for which we determined signature SNPs. Gray and red dots indicate branches with at least 99% bootstrap support; red dots further indicate branches for which we propose signature SNPs.

### 2.3. MTBC L3 Origin and Ancestral State Reconstruction

Lastly, we examined with a joint maximum likelihood estimation ancestral state reconstruction of the likely geographical origin of MTBC L3 strains and possible continent

transition events of particular L3 clades [19]. As expected, South Asia was predicted to be the origin of the common ancestor of all L3 strains, with a likelihood of 90.2% (Figure 5). However, one of the ancestral clades in the ML phylogeny, Clade 3.3, has a recent common ancestor that most likely originated in Eastern Africa (56.0%, Figure 5). This may suggest a very early introduction of L3 strains into Africa in the evolutionary history of this lineage. The observed geographical split of L3-CC1 and L3-CC2 around the Horn of Africa (Figure 3) is further confirmed by our ancestral state reconstruction. Clade 3.5.1 (associated with L3-CC1) shows a pronounced bottleneck effect in the ML phylogeny and a likely origin in North-East Africa (86.5%), particularly in Sudan (Figure 5). More recent introductions according to the ML phylogeny were observed for Lineage 3.1 strains. Specifically, Clade 3.1.3.1 (associated with L3-CC5), as well as Clade 3.1.1.1 (associated with L3-CC2), have predicted origins in Eastern Africa, with likelihoods of 99.8% and 71.9%, respectively (Figure 5). In summary, we confirm the likely origin of L3 strains in Southern Asia and found evidence of at least four independent introductions of individual L3 clades into East and North-East Africa.



**Figure 5.** Phylogeographic ancestral state reconstruction of MTBC L3 strains. Nodes are color coded by UN regions. Pie charts show likelihoods for ancestral geographic origins for the internal nodes. A South Asian origin for L3 progenitor is supported using the rooted maximum-likelihood phylogeny. The scale bar indicates nucleotide substitutions per site. Four clades with a predicted African common ancestor are indicated and their pie charts are shown enlarged for better visibility, as well as for the root.



### 3. Discussion

Based on a large-scale analysis of classical genotyping data (24-loci MIRU-VNTR), we developed a first view of the global population structure and spatial distribution of MTBC L3 strains. Whole-genome-based analysis further revealed the global phylogeny and geographical expansion patterns. While we confirm the South Asian origin of L3, we furthermore provide new evidence for at least four independent introduction events into East and North-East Africa in the evolutionary history of this important modern MTBC lineage. Moreover, our findings provide signature SNPs that describe up to 85% of the global L3 genetic diversity, which will be important for clinical trials and pharmaceutical studies that need to take the phylogenetic diversity of MTBC strains into account.

#### 3.1. Phylogeography of MTBC L3

Our ancestral state reconstruction confirms that the progenitor of L3 strains can be traced back to South Asia, as suggested earlier, but based on a much smaller, less diverse strain set [20]. Similar to other modern MTBC lineages, the L3 population likely began to expand in parallel with their human hosts and following ancestral movements out of Africa towards Eurasia [20]. The following long-term historic expansion in Southern Asia is reflected by the diverse MIRU-VNTR patterns of strains classified as L3-BG, which are dominating in India and Nepal [3,13].

Interestingly, our phylogeographic analysis further suggests four independent African introduction events of individual L3 clades. The earliest introduction event from Southern Asia into Eastern Africa is suggested for the evolutionarily ancestral clade 3.3 and is consistent with known ancient exchanges and migrations between both world regions. As suggested in a similar scenario for distinct L2 clades [19], three reintroductions occurred more recently, i.e., Clade 3.5.1 (associated with L3-CC1) mainly in Sudan, Clade 3.1.3.1 (associated with L3-CC5), and Clade 3.1.1.1 (associated with L3-CC2) in Kenya and Tanzania. Subsequent strong local expansion of L3-CC1 and L3-CC2 strains is supported by the fact that both clonal complexes also harbor the largest molecular clusters and the geographical separation of CC1 and CC2 around the Horn of Africa and particularly in Ethiopia.

The large geographic distribution of L3 strains beyond South Asia and East Africa likely reflects modern migratory movements. Strains from the largest L3-CC1 cluster (MLVA Code: 1557-32) were retrieved from patients in Australia, Canada, Eritrea, Ethiopia, Germany, India, Italy, Nigeria, Sudan, Sweden, and the Netherlands. Strains of the largest L3-CC2 cluster (MLVA Code: 1064-32) were also identified in 12 different countries in eight UN geo-regions across four continents, including Asia, Africa, Europe, and North America. Since the first half of the 20th century, North America, Europe, and Australia have been connected by extensive movement of workers from low-income countries in Asia and Africa [21–25]. In recent years, for instance, over half of the immigrant labor force of Canada and Australia were born in Asia, with nearly 60% of these immigrants born in China and India [26,27].

#### 3.2. Implication of (MTBC and) L3 Genomic Diversity

Despite the strongly clonal population structure of the MTBC, evidence is accumulating on phylogenetically associated biomedical consequences of the diversity of the pathogen. This genomic diversity can potentially lead to different pathobiological features, such as disease severity [28,29] and dormancy [28,30,31], or can affect treatment outcomes [32], bacterial fitness [33], or resistance levels to anti-TB drugs [34]. (Phylo)genetic determinants of such pathobiological differences have recently been pointed out. The loss of the TbD1 genomic region in modern MTBC lineages, comprising L2, L3, and L4, has been recently suggested to be a key driver of their global epidemic spread, as this loss causes an increased resistance against hypoxia, oxidative stress, and enhanced virulence in these strains, relatively to TbD1-intact strains, such as L1 and L7 [35]. Likewise, some phylogeny-linked SNPs in the *esxW* gene or deletion in the *ppe38* gene have been associated with increased transmissibility and hypervirulence of the Beijing lineage (L 2.2) [9,36].

Conversely, mutations in the *phoR* gene shared by the animal-adapted and *M. africanum* L6 lineages have been associated with lower virulence or transmissibility [37]. Here, we identified signature SNPs that were able to differentiate more than 85% of a global L3 strain collection. Our extended L3 nomenclature can be combined with existing SNP barcodes for Mtbc strains [18] and allows a more detailed classification of L3 strains for surveillance or clinical studies. Furthermore, these clade-specific SNPs can be important for the evaluation of phenotypic discrepancies between L3 strains such as response to individual antibiotics or potential new vaccine candidates.

### 3.3. Limitations

Molecular, phenotypic, and geographical data were collated from different studies and diverse sampling methods. Thus, the proportions of L3 strains and sizes of molecular clusters can differ in countries, especially those with only few available datasets. Moreover, to overcome the bias introduced by collections enriched with rifampicin-resistant strains, we only considered cross-sectional and population-based cohorts for the analysis of resistance determinants among clonal complexes.

## 4. Conclusions

Our findings extend the knowledge of the global genetic diversity of MTBC L3 strains. By employing large-scale 24-loci MIRU-VNTR data and WGS-based phylogeographic analysis, we offered a robust explanation of how L3 strains were independently introduced into North and East Africa and into other regions of the world. The evolutionary success of this important modern MTBC lineage is likely shaped by ancient cultural exchange and modern labor migrations from Africa and South Asia to Australia, North America, and Central Europe. The identified SNP signatures reveal an intriguing genetic diversity of L3 strains that needs to be considered in clinical trials that evaluate the effectivity of new drugs/regimens or vaccine candidates.

## 5. Materials and Methods

### 5.1. Data Collection

Our study is based on a global collection of 2682 clinical MTBC L3 strains from 38 countries. For patients in Western Europe with a migration background, we used country of birth as a surrogate. Datasets were compiled from 53 studies. Details on sampling time, sampling strategies, drug susceptibility patterns, and 24-loci MIRU-VNTR data with MLVA 15–9 genotypes and associated cluster numbers are presented in Supplementary Table S1.

### 5.2. 24-Loci MIRU-VNTR Typing

The 24-loci MIRU-VNTR typing was conducted according to standardized protocols published earlier [38]. Briefly, MIRU-VNTR alleles were amplified using the Quadruplex PCR Kit (Genoscreen, Lille, France) according to the manufacturer's instructions. Fragment analysis using the GeneScan™ 1200 LIZ dye as a size standard (Life Technologies, Darmstadt, Germany) was carried out on a capillary sequencer 3130xL and 3500xL for the genetic analyzer. GeneMapper software version 3.7 (Life Technologies, Darmstadt, Germany) was used to determine the copy number of MIRU-VNTR alleles. We considered strains with a maximum of one missing locus and classified them as L3 (Delhi/CAS genotype) by phylogenetic inference from a reference collection hosted at <http://www.miru-vntrplus.org>, accessed on 06 August 2020 [16]. Further, we assigned MLVA MTBC 15-9 codes for strains with complete MIRU-VNTR profiles. A minimum spanning (MS) tree was calculated based on 24-loci MIRU-VNTR typing data (a missing locus was treated as its own category) with BioNumerics v7.6 (Applied Maths (bioMérieux), Sint-Martens-Latem, Belgium) [39]. Strains with identical MLVA MTBC 15-9 codes were combined in a molecular cluster, as a surrogate marker for recently transmitted strains. Based on the MS tree topology, we selected major nodes that were considered as parental nodes for globally expanding L3 strain populations, i.e., L3 clonal complexes (L3-CCs). Strains descending from the parental nodes

(multilocus variants) were assigned to the respective L3 CCs. The remaining strains which did not show a clear structure were considered as having a diverse L3 genetic background (L3-BG).

For whole-genome sequencing (see below), we selected strains from the largest nodes and associated multilocus variants from the MS tree (Supplementary Figure S2). Selection criteria were equal representation of all defined L3-CCs and L3-BG, inclusion of different countries, and availability of cultures/DNA at the Research Center Borstel in Germany.

### 5.3. Whole-Genome Sequencing Analysis

WGS was performed with Illumina (San Diego, CA, USA) technology and Nextera XT library preparation [40], as instructed by the manufacturer. Strains were sequenced with a minimum average genome coverage of 50x. Raw read data were processed with MTBSeq [41] and mapped to the *M. tuberculosis* H37Rv genome (GenBank ID: NC\_000962.3). Briefly, variants were called if supported by at least four reads in forward and four reads in reverse orientation, and four reads calling the allele with at least a Phred score of  $\geq 30$  and a minimum allele frequency of 75%.

For phylogenetic analysis, identified single-nucleotide polymorphism (SNP) positions were combined into a concatenated sequence alignment, complemented with data from original mappings where necessary [42], excluding positions without a nucleotide call and positions that failed in the above-mentioned criteria for variant calling in more than 5% of the samples. Likewise, repetitive regions (e.g., PPE/PGRS genes) and drug-resistance-associated genes were excluded from the phylogenetic reconstruction [42].

### 5.4. Phylogenetic Analysis

Maximum-likelihood (ML) trees were calculated with FastTree using concatenated sequence alignment and a general time-reversible (GTR) substitution model [43]. Inspection and midpoint rooting of the ML tree were performed with FigTree v1.4.4 software, while graphical annotation was performed with the online tool EvolView [44] and iTol [45]. Major phylogenetic clades were defined on the basis of the tree topology, 100% bootstrap support, a previously defined L3 sublineage nomenclature [46], and considering L3-CCs based on 24-loci MIRU-VNTR data (see above) [16]. Clade-specific SNPs were extracted from the concatenated sequence alignment. WGS datasets were further inspected for the presence of the L3-specific deletion RD750 using Integrative Genomics Viewer (IGV) [47].

### 5.5. Ancestral State Reconstruction

Geographic origin was used as a discrete state in the ancestral state reconstruction and inferred from patients or sample country of origin and translated to the associated UN region. Origin of samples sequenced in Europe, the United States of America, and Australia lacking any patient information was set to "NA" in order to avoid geographical bias from larger sequencing collaborations. For an ML phylogeny of 373 L3 strains, we selected the best model of character evolution by calculating the Akaike information criterion for an equal-rates (ER) model, a symmetric, unordered model (SYM), and an all-rates-different (ARD) model with fitMk from the R package phytools [48] and computing their Akaike weights. Scaled likelihoods for each ancestral state were calculated with the best fitting model ER (aic.w = 0.96) and the ace function from the R package ape and mapped on the ML phylogeny [49].

### 5.6. Statistics

Differences between proportions of individual groups were calculated with pairwise Fisher's exact tests.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13060990/s1>, Table S1: Metadata and 24-loci mycobacterial interspersed repetitive units—variable number tandem repeat results for 2682 *Mycobacterium tuberculosis* Lineage 3 strains; Table S2: Mycobacterium tuberculosis Lineage 3 cross-border clusters; Table S3: Geographical distribution of L3 clonal complexes of 2682 Mycobacterium tuberculosis Lineage 3 strains; Table S4: Signature SNPs for 12 major Mycobacterium tuberculosis Lineage 3 clades; Table S5: Whole-genome sequencing data and accession numbers of 373 Mycobacterium tuberculosis Lineage 3 strains; Figure S1: Minimum spanning tree based on 24-loci mycobacterial interspersed repetitive units—variable number tandem repeat results from 2682 *Mycobacterium tuberculosis* Lineage 3 strains and selected L3 datasets colored in blue for whole-genome sequencing (WGS) and phylogenetic inference; Figure S2: Maximum-likelihood phylogeny of 152 Lineage 3 strains from this study with Lineage 2 strains as outgroup.

**Author Contributions:** Y.A.S., S.N. and M.M. conceived the study. Y.A.S., S.N. and M.M. curated the data. Y.A.S., C.U., T.A.K., I.B., M.D., N.A. and M.M. performed the formal analysis. S.N. acquired funding. Y.A.S., L.H.W., G.K., E.A.M., A.B.D., S.A.-H., P.N., M.M.F., F.V., V.S., E.M., S.O.V., Y.Z., A.A., K.A.-A., S.G., M.V.-B., A.N., S.A.-H., P.B., V.D., M.K., M.B., M.H., A.R., A.G., E.M.T., L.K.S., A.A.J.-P., S.A., V.P.B., F.N., M.F., A.G.D., S.M., B.T., D.B., S.N.K., R.D., P.S., F.P.M., H.H., S.N. and M.M. provided resources for this study. Y.A.S., C.U., T.A.K., I.B., N.A., S.N. and M.M. visualized the results. Y.A.S. and M.M. wrote the initial draft. All authors have read and agreed to the published version of the manuscript.

**Funding:** Parts of this work have been supported by grants from the European Community’s Seventh Framework Program (FP7/2007–2013) under Grant Agreement No. 278864 in the framework of the EU PathoNGenTrace project and Grant Agreement No. 223681 in the framework of the TB-PAN NET project. The first author is a DAAD (German: Deutscher Akademischer Austauschdienst) or the German Academic Exchange Service stipend recipient, Funding Program No. 57076385. Parts of this work have been supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy EXC 22167-390884018, the Leibniz Science Campus EvoLUNG, and the German Center for Infection Research. Part of this work has also been financially supported by CANTAM One- EDCTP Grant N° CB.2007.41700.006, offered to the University of Yaoundé I in Cameroon.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Whole-genome sequencing data (fastq files) were uploaded to the European Nucleotide Archive (ENA) under Project Accession Number PRJEB42771. Individual accession numbers are given in Supplementary Table S4.

**Acknowledgments:** The authors would like to sincerely thank all of the collaborators who generously provided their data. We would also like to deeply thank I. Razio, T. Niemann, A. Lüdemann, T. Struve-Sonnenschein, J. Zallet, and V. Mohr for providing excellent technical assistance at the Research Center Borstel in Germany. The contribution of the Research Center Borstel, Borstel, Germany, in accomplishing this work is appreciated.

**Conflicts of Interest:** P.S. is a consultant for Genoscreen. Other authors declare no conflict of interest.

## References

1. WHO. *Global Tuberculosis Report*; World Health Organization: Geneva, Switzerland, 2020; p. 232.
2. Niemann, S.; Merker, M.; Kohl, T.; Supply, P. Impact of Genetic Diversity on the Biology of *Mycobacterium tuberculosis* Complex Strains. *Microbiol. Spectr.* **2016**, *4*, 4–6. [[CrossRef](#)] [[PubMed](#)]
3. Gagneux, S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* **2018**, *16*, 202–213. [[CrossRef](#)] [[PubMed](#)]
4. Semuto, J.C.; Chloé, L.; Michael, M.; Agathe, J.; Fabrizio, M.; Oren, T.; Rudy, A.; Esdras, B.; Wim, M.; Kristina, F.; et al. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nat. Commun.* **2020**, *11*, 2917.
5. Coscolla, M.; Gagneux, S.; Menardo, F.; Loiseau, C.; Ruiz-Rodriguez, P.; Borrell, S.; Otchere, I.D.; Asante-Poku, A.; Asare, P.; Sánchez-Busó, L.; et al. Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *Microb. Genom.* **2021**, *7*, 000477. [[CrossRef](#)]

6. Merker, M.; Blin, C.; Mona, S.; Duforet-Frebourg, N.; Lecher, S.; Willery, E.; Blum, M.G.; Rusch-Gerdes, S.; Mokrousov, I.; Aleksic, E.; et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat. Genet.* **2015**, *47*, 242–249. [[CrossRef](#)]
7. Stucki, D.; Brites, D.; Coscolla, M.; Liu, Q.; Trauner, A.; Fenner, L.; Rutaihwa, L.; Borrell, S.; Luo, T.; Gao, Q.; et al. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat. Genet.* **2016**, *48*, 1535–1543. [[CrossRef](#)]
8. Merker, M.; Barbier, M.; Cox, H.; Rasigade, J.P.; Feuerriegel, S.; Kohl, T.A.; Diel, R.; Borrell, S.; Gagneux, S.; Nikolayevskyy, V.; et al. Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. *Elife* **2018**, *7*, e38200. [[CrossRef](#)]
9. Ates, L.S.; Dippenaar, A.; Ummels, R.; Piersma, S.R.; van der Woude, A.D.; van der Kuij, K.; Le Chevalier, F.; Mata-Espinosa, D.; Barrios-Payan, J.; Marquina-Castillo, B.; et al. Mutations in ppe38 block PE\_PGRS secretion and increase virulence of *Mycobacterium tuberculosis*. *Nat. Microbiol.* **2018**, *3*, 181–188. [[CrossRef](#)]
10. Galagan, J.E. Genomic insights into tuberculosis. *Nat. Rev. Genet.* **2014**, *15*, 307–320. [[CrossRef](#)]
11. Chihota, V.N.; Niehaus, A.; Streicher, E.M.; Wang, X.; Sampson, S.L.; Mason, P.; Kallenius, G.; Mfinanga, S.G.; Pillay, M.; Klopper, M.; et al. Geospatial distribution of *Mycobacterium tuberculosis* genotypes in Africa. *PLoS ONE* **2018**, *13*, e0200632. [[CrossRef](#)]
12. Shuaib, Y.A.; Eltahir, A.G.; Lothar, H.W.; Ulrich, E.S.; Mohammed, A.B.; Saad, E.M.; Mohamed, A.A.; Glennah, K.; Sönke, A.; Doris, H.; et al. *Mycobacterium tuberculosis* complex lineage 3 as causative agent of pulmonary tuberculosis, Eastern Sudan. *Emerg. Infect. Dis.* **2020**, *26*, 427–436. [[CrossRef](#)] [[PubMed](#)]
13. Brudey, K.; Driscoll, J.R.; Rigouts, L.; Prodinge, W.M.; Gori, A.; Al-Hajoj, S.A.; Allix, C.; Aristimuno, L.; Arora, J.; Baumanis, V.; et al. *Mycobacterium tuberculosis* complex genetic diversity: Mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol.* **2006**, *6*, 23. [[CrossRef](#)] [[PubMed](#)]
14. Kerubo, G.; Ndungu, P.; Shuaib, Y.A.; Amukoye, E.; Revathi, G.; Homolka, S.; Kariuki, S.; Merker, M.; Niemann, S. Molecular Epidemiology of *Mycobacterium tuberculosis* Complex Strains in Urban and Slum Settings of Nairobi, Kenya. *Genes* **2022**, *13*, 475. [[CrossRef](#)]
15. O'Neill, M.B.; Shockey, A.; Zarley, A.; Aylward, W.; Eldholm, V.; Kitchen, A.; Pepperell, C.S. Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. *Mol. Ecol.* **2019**, *28*, 3241–3256. [[CrossRef](#)] [[PubMed](#)]
16. Weniger, T.; Krawczyk, J.; Supply, P.; Niemann, S.; Harmsen, D. MIRU-VNTRplus: A web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. *Nucleic Acids Res.* **2010**, *38*, W326–W331. [[CrossRef](#)] [[PubMed](#)]
17. Mathema, B.; Kurepina, N.E.; Bifani, P.J.; Kreiswirth, B.N. Molecular epidemiology of tuberculosis: Current insights. *Clin. Microbiol. Rev.* **2006**, *19*, 658–685. [[CrossRef](#)]
18. Napier, G.; Campino, S.; Merid, Y.; Abebe, M.; Woldeamanuel, Y.; Aseffa, A.; Hibberd, M.L.; Phelan, J.; Clark, T.G. Robust barcoding and identification of *Mycobacterium tuberculosis* lineages for epidemiological and clinical studies. *Genome Med.* **2020**, *12*, 114. [[CrossRef](#)]
19. Rutaihwa, L.K.; Menardo, F.; Stucki, D.; Gygli, S.M.; Ley, S.D.; Malla, B.; Feldmann, J.; Farnov, S.B.; Beisel, C.; Middelkoop, K.; et al. Multiple Introductions of *Mycobacterium tuberculosis* Lineage 2–Beijing Into Africa Over Centuries. *Front. Ecol. Evol.* **2019**, *7*, 112. [[CrossRef](#)]
20. Comas, I.; Coscolla, M.; Luo, T.; Borrell, S.; Holt, K.E.; Kato-Maeda, M.; Parkhill, J.; Malla, B.; Berg, S.; Thwaites, G.; et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* **2013**, *45*, 1176–1182. [[CrossRef](#)]
21. Al-Hajoj, S.A.; Zozio, T.; Al-Rabiah, F.; Mohammad, V.; Al-Nasser, M.; Sola, C.; Rastogi, N. First insight into the population structure of *Mycobacterium tuberculosis* in Saudi Arabia. *J. Clin. Microbiol.* **2007**, *45*, 2467–2473. [[CrossRef](#)]
22. Varghese, B.; Supply, P.; Allix-Beguec, C.; Shoukri, M.; Al-Omari, R.; Herbawi, M.; Al-Hajoj, S. Admixed phylogenetic distribution of drug resistant *Mycobacterium tuberculosis* in Saudi Arabia. *PLoS ONE* **2013**, *8*, e55598. [[CrossRef](#)] [[PubMed](#)]
23. Al-Hajoj, S.; Varghese, B. Tuberculosis in Saudi Arabia: The journey across time. *J. Infect. Dev. Ctries.* **2015**, *9*, 222–231. [[CrossRef](#)]
24. Varghese, B.; Enani, M.; Alrajhi, A.; Al Johani, S.; Albarak, A.; Althawadi, S.; Elkhizzi, N.; AlGhafli, H.; Shoukri, M.; Al-Hajoj, S. Impact of *Mycobacterium tuberculosis* complex lineages as a determinant of disease phenotypes from an immigrant rich moderate tuberculosis burden country. *Respir. Res.* **2018**, *19*, 259. [[CrossRef](#)]
25. Al-Ghafli, H.; Kohl, T.A.; Merker, M.; Varghese, B.; Halees, A.; Niemann, S.; Al-Hajoj, S. Drug-resistance profiling and transmission dynamics of multidrug-resistant *Mycobacterium tuberculosis* in Saudi Arabia revealed by whole genome sequencing. *Infect. Drug Resist.* **2018**, *11*, 2219–2229. [[CrossRef](#)] [[PubMed](#)]
26. Phillips, J.; Simon-Davies, J. *Migration to Australia: A Quick Guide to the Statistics*; Government of Australia: Canberra, Australia, 2017; pp. 1–7.
27. Yssaad, L.; Fields, A. *The Canadian Immigrant Labour Market Recent Trends from 2006 to 2017*; Government of Canada: Ottawa, ON, Canada, 2018; pp. 1–26.
28. Amin, M.; Yanti, B.; Harapan, H.; Mertaniasih, N.M. The role of *Mycobacterium tuberculosis* lineages on lung tissue damage and TNF- $\alpha$  level among tuberculosis patients, Indonesia. *Clin. Epidemiol. Glob. Health* **2019**, *7*, 263–267. [[CrossRef](#)]
29. Sousa, J.; Ca, B.; Maceiras, A.R.; Simoes-Costa, L.; Fonseca, K.L.; Fernandes, A.I.; Ramos, A.; Carvalho, T.; Barros, L.; Magalhaes, C.; et al. *Mycobacterium tuberculosis* associated with severe tuberculosis evades cytosolic surveillance systems and modulates IL-1 $\beta$  production. *Nat. Commun.* **2020**, *11*, 1949. [[CrossRef](#)]

30. Domenech, P.; Zou, J.; Averback, A.; Syed, N.; Curtis, D.; Donato, S.; Reed, M.B. Unique Regulation of the DosR Regulon in the Beijing Lineage of *Mycobacterium tuberculosis*. *J. Bacteriol.* **2017**, *199*, e00696-16. [[CrossRef](#)]
31. Ofori-Anyinam, B.; Dolganov, G.; Van, T.; Davis, J.L.; Walter, N.D.; Garcia, B.J.; Voskuil, M.; Fissette, K.; Diels, M.; Driesen, M.; et al. Significant under expression of the DosR regulon in *M. tuberculosis* complex lineage 6 in sputum. *Tuberculosis* **2017**, *104*, 58–64. [[CrossRef](#)]
32. Schierloh, P.; Klepp, L.; Vazquez, C.; Rocha, R.V.; Blanco, F.C.; Balboa, L.; Lopez, B.; Ritacco, V.; Bigi, F.; Sasiain Mdel, C. Differential expression of immunogenic proteins on virulent *Mycobacterium tuberculosis* clinical isolates. *BioMed Res. Int.* **2014**, *2014*, 741309. [[CrossRef](#)]
33. Satta, G.; Witney, A.A.; Shorten, R.J.; Karlikowska, M.; Lipman, M.; McHugh, T.D. Genetic variation in *Mycobacterium tuberculosis* isolates from a London outbreak associated with isoniazid resistance. *BMC Med.* **2016**, *14*, 117. [[CrossRef](#)]
34. Merker, M.; Kohl, T.A.; Barilar, I.; Andres, S.; Fowler, P.W.; Chryssanthou, E.; Angeby, K.; Jureen, P.; Moradigaravand, D.; Parkhill, J.; et al. Phylogenetically informative mutations in genes implicated in antibiotic resistance in *Mycobacterium tuberculosis* complex. *Genome Med.* **2020**, *12*, 27. [[CrossRef](#)] [[PubMed](#)]
35. Bottai, D.; Frigui, W.; Sayes, F.; Di Luca, M.; Spadoni, D.; Pawlik, A.; Zoppo, M.; Orgeur, M.; Khanna, V.; Hardy, D.; et al. TbD1 deletion as a driver of the evolutionary success of modern epidemic *Mycobacterium tuberculosis* lineages. *Nat. Commun.* **2020**, *11*, 684. [[CrossRef](#)]
36. Holt, K.E.; McAdam, P.; Thai, P.V.K.; Thuong, N.T.T.; Ha, D.T.M.; Lan, N.N.; Lan, N.H.; Nhu, N.T.Q.; Hai, H.T.; Ha, V.T.N.; et al. Frequent transmission of the *Mycobacterium tuberculosis* Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nat. Genet.* **2018**, *50*, 849–856. [[CrossRef](#)]
37. Gonzalo-Asensio, J.; Malaga, W.; Pawlik, A.; Astarie-Dequeker, C.; Passemar, C.; Moreau, F.; Laval, F.; Daffe, M.; Martin, C.; Brosch, R.; et al. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11491–11496. [[CrossRef](#)] [[PubMed](#)]
38. Supply, P.; Allix, C.; Lesjean, S.; Cardoso-Oelemann, M.; Rusch-Gerdes, S.; Willery, E.; Savine, E.; de Haas, P.; van Deutekom, H.; Roring, S.; et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **2006**, *44*, 4498–4510. [[CrossRef](#)] [[PubMed](#)]
39. Kannan, L.; Wheeler, W.C. Maximum Parsimony on Phylogenetic networks. *Algorithms Mol. Biol.* **2012**, *7*, 9. [[CrossRef](#)] [[PubMed](#)]
40. Baym, M.; Kryazhimskiy, S.; Lieberman, T.D.; Chung, H.; Desai, M.M.; Kishony, R. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS ONE* **2015**, *10*, e0128036. [[CrossRef](#)]
41. Allix-Beguec, C.; Arandjelovic, I.; Bi, L.; Beckert, P.; Bonnet, M.; Bradley, P.; Cabibbe, A.M.; Cancino-Munoz, I.; Caulfield, M.J.; Chaiprasert, A.; et al. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *N. Engl. J. Med.* **2018**, *379*, 1403–1415.
42. Kohl, T.A.; Utpatel, C.; Schleusener, V.; De Filippo, M.R.; Beckert, P.; Cirillo, D.M.; Niemann, S. MTBseq: A comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex isolates. *PeerJ* **2018**, *6*, e5895. [[CrossRef](#)]
43. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)]
44. He, Z.; Zhang, H.; Gao, S.; Lercher, M.J.; Chen, W.H.; Hu, S. Evolview v2: An online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* **2016**, *44*, W236–W241. [[CrossRef](#)] [[PubMed](#)]
45. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [[CrossRef](#)]
46. Coll, F.; McNerney, R.; Guerra-Assuncao, J.A.; Glynn, J.R.; Perdigao, J.; Viveiros, M.; Portugal, I.; Pain, A.; Martin, N.; Clark, T.G. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* **2014**, *5*, 4812. [[CrossRef](#)] [[PubMed](#)]
47. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [[CrossRef](#)] [[PubMed](#)]
48. Revell, L.J. Phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **2012**, *3*, 217–223. [[CrossRef](#)]
49. Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **2004**, *20*, 289–290. [[CrossRef](#)]