Reassortment Network of Influenza A Virus

Xingfei Gong[1,2†], Mingda Hu[1†], Wei Chen[2†], Haoyi Yang[1,2], Boqian Wang[1], Junjie Yue[1], Yuan Jin[1*], Long Liang[1*] and Hongguang Ren[1*]

[1]Beijing Institute of Biotechnology, State Key Laboratory of Pathogen and Biosecurity, Beijing, China.

[2]College of Computer, National University of Defense Technology, Changsha, China.

*To whom correspondence should be addressed. Email: bioren@163.com. Correspondence may also be addressed to ll@bmi.ac.cn or jin_0220@sina.com

[†]These authors contributed equally to this paper.

Abstract

Influenza A virus (IAV) genomes are composed of eight single-stranded RNA segments. Genetic exchange through reassortment of the segmented genomes often endows IAVs with new genetic characteristics, which may affect transmissibility and pathogenicity. However, a comprehensive understood of the reassortment history of IAVs remains poorly studied. To this end, we assembled 40296 whole-genome sequences of IAVs for analysis. Using a new clustering method based on Mean Pairwise Distances in the phylogenetic trees, we classified each segment of IAVs into clades. Correspondingly, reassortment events among IAVs were detected by checking the segment clade compositions of related genomes under specific environment factors and time period. We systematically identified 1927 possible reassortment events of IAVs and constructed their reassortment network. Interestingly, minimum spanning tree of the reassortment network reproved that swine act as an intermediate host in the reassortment history of IAVs between avian and human. Moreover, reassortment patterns among related subtypes constructed in this study are consistent with previous studies. Taken together, our genome-wide reassorment analysis of all the IAVs offers an overview of the leaping evolution of the virus and a comprehensive network representing the relationships of IAVs.

Availability: Source code freely available for download at https://github.com/ElevenStr/IAVreassortment, implemented in python.

Lead Contact: bioren@163.com

1 Introduction

Influenza A Virus (IAV) is a highly infectious viral pathogen that causes seasonal epidemics, occasional pandemics, and zoonotic outbreaks, which may lead to substantial human morbidity and mortality and a considerable financial burden worldwide (Liu, et al., 2020). It has been more than one hundred years since the Spanish flu (H1N1) virus caused the first recorded influenza pandemic, which is considered to be the most lethal natural event in modern history (Liu, et al., 2018; Smith, 2011). Since then, there have been three other pandemics caused by A(H2N2), A(H3N2), and A(H1N1)pdm09 viruses—the 1957 Asian flu, the 1968 Hong Kong flu, and the 2009 swine-origin flu, respectively. There are currently 18 HA subtypes and 11 NA subtypes in IAVs, most of which spread in wild birds. Comparatively, only three combinations of HA and NA subtypes are known to be widespread in humans: H1N1, H2N2 and H3N2, of which H1N1 and H3N2 subtypes viruses cause seasonal epidemics (Petrova and Russell, 2018). The WHO estimated that seasonal influenza viruses infect 5-15% of the human population each year, causing approximately 500,000 deaths worldwide (Stohr, 2002).

As segmented RNA viruses, IAVs can exchange the gene segments through reassortment during co-infection. Specifically, when two or more IAVs infect the same cell, a hybrid virus can be produced by assembling the gene segments of the parental viruses into a nascent virion (McDonald, et al., 2016). Reassortment and mutation are both the main driving forces for the evolution of IAVs. However, the effect caused by mutations need to be accumulated for a long time, while the reassortment is often a leapfrog evolution. Reassortment also plays an important but unclear role in the emergence of the novel viruses and cross-species transmission (Dhanasekaran, et al., 2015). The virus that caused the 2009 H1N1 pandemic was generated from a triple-reassortment of H1N1 avian virus, H1N1 classical swine virus, and H3N2 human seasonal virus (Smith, et al., 2009). Furthermore, swine has been considered as the "mixing vessels" for IAVs since there are both α-2,3 and α-2,6 sialic acids (SAs) in their respiratory tracts, which provides conditions for the occurrence of reassortment (Ma, et al., 2009).

In recent years, a lot of studies for IAVs are based on mutations, especially in the receptor binding region on HA segment, while the study on reassortment detection method and the reassortment history of IAVs is relatively few. Rabadan et al. proposed a method to detect reassortment by comparing the sequence differences between two viruses (Rabadan, et al., 2008). This method considered that the number of differences in each segment sequence of two viruses should be

proportional if there is no reassortment. Another algorithm was proposed by Silva et al. based on the neighborhood of strains (Silva, et al., 2012), which determined the reassortant virus by the size of the common neighborhood of two segments. These rare researches on IAVs reassortment tend to focus on viruses causing a single epidemic or with a single subtype, resulting in a lack of overall understanding of the reassortment history of IAVs. In our work, we proposed a novel genotype nomenclature for IAVs and used it to detect the reassortment in IAV genomes. The reassortment history and gene flow network were constructed, by which we found the characteristics and patterns in frequent reassortment events of IAVs.

2 Methods

2.1 Data preparation and processing

We assembled whole-genomes sequences (as of October 13,2020) from the Influenza Virus Resources at the National Centre for Biotechnology Information (NCBI) ((https://ftp.ncbi.nih.gov/genomes/INFLUENZA/). The sequences with illegal characters in codon region or incomplete length were deleted. We also excluded sequences for which epidemiological information was unclear such as host, location, subtype and sampled date. In addition, the laboratory sequences were not in our consideration. After that, we obtained 40296 genomes of IAVs, in which each segment of the viruses met our requirements. Multiple sequence alignment for each segment was performed using MAFFT v7.037 (Katoh and Standley, 2013). In order to remove the bias and redundancy of sampling and reduce the computational complexity when reconstructing phylogenetic trees, we filtered the genome sequences with similarity (>99%) for viruses with the same host, location, subtype and sampled year. The similarity is simply defined as the percentage of the number of identical nucleotides to the genome sequence length. The virus correspondence before and after similarity filtering was shown in Supplementary Table S1.

The phylogenetic tree was reconstructed for each segment by IQ-TREE (Nguyen, et al., 2015) using PhyloSuite platform (Zhang, et al., 2020). In general, the phylogenetic analysis usually focuses on the two influenza surface proteins, hemagglutinin (HA) and neuraminidase (NA), since they can change rapidly in a short period of time (Lemey, et al., 2009; Wallace and Fitch, 2008; Wallace, et al., 2007). We analyzed the phylogenetic trees of HA and NA segments and then 46 genomes with wrong subtypes were deleted. After all data processing operations, we got 8932 representative high-quality IAV genomes. It is noteworthy that we divided the hosts of these IAVs into human, swine and avian since > 99% of the viruses are sampled from these hosts. The avian host was subdivided

into shorebirds, waterfowl, land birds and domestic birds. All avian hosts classification was shown in Supplementary Table S2. We also divided the locations into 22 areas according to the administrative division, including North America, Western Europe and East Asia et al. The correspondence between locations and areas was shown in Supplementary Table S3.

2.2 Segment detailed type determination

In order to define the detailed type of each segment for IAVs, we used a clustering method based on the Mean Pairing Distance (MPD) in the phylogenetic trees. MPD is the mean phylogenetic distance between the leaf nodes of an internal node on the phylogenetic tree. For example, the MPD of an internal node with n leaf nodes on PB1 segment phylogenetic tree is calculated as:

$$MPD = \frac{\sum_i^n \sum_j^n \delta_{i,j}}{\binom{n}{2}}$$

$\delta_{i,j}$ is the phylogenetic distance between the leaf node i and leaf node j. $\binom{n}{2}$ represents the combination number formula. If the number of leaf nodes of an internal node is less than the average number of leaf nodes of all internal nodes or its MPD is less than the average MPD, we consider that the leaf nodes of this internal node can be regarded as a detailed type, where each leaf node represents the PB1 segment of an IAV. We give a number index x to the type in the segment, namely segment_x.

2.3 Genotype nomenclature

The method of defining an IAV genotype is to give each segment a specific type. In the work of Lu et al. (Lu, et al., 2007), the genotype was defined as the combination of the lineages for each segment in an IAV genome. They assigned a letter to each lineage of PB2, PB1, PA, NP and MP. For HA, NA and NS, they assigned a number followed by a letter to each lineage, while the number represented the subtype or allele.

Here, we divided each segment into more detailed types according to the Mean Pairwise Distance (MPD) on the phylogenetic trees. The genotype was then defined as the sequential combination of the defined types for each segment. An index was assigned to each defined type of eight segments. For example, [PB2_292, PB1_415, PA_333, HA_422, NP_463, NA_405, MP_369, NS_406] is the genotype of a human H7N9 subtype virus with PB2 type 292, PB1 type 415, PA type 333, HA type 42, NP type 463, NA type 405, MP type 369 and NS type 406.

2.4 Reassortment definition and identification

Since the genomes of a reassortant virus are derived from different parents, the positions of different segments on the phylogenetic trees may be inconsistent. Therefore, the most intuitive way to detect the reassortment of IAVs is to compare the positions on phylogenetic trees of different segments (Nelson, et al., 2008). Based on this principle and our genotype nomenclature, we defined a reassortment as a mixation of the different genotype combinations. If the genotype of a virus can be produced by the recombination of the genotypes for two or more viruses and their epidemiological information is related, then we consider that the virus might be reassorted from the parental viruses.

To identify the reassortment events in IAVs, we first constructed the segment detailed type set for each year, which contains the segment detailed types of viruses that emerged in a year. If the eight segment detailed types of a virus could be found in the segment detailed type set of the previous five years, then we traversed the IAVs dataset of the previous five years to find whether the segment detailed types recombination between two viruses that appeared in the previous five years can produce this virus genotype. If the virus could be produced by two viruses, we calculated the reassortment cost as follows:

$$cost = locNum \cdot parNum + hostNum \cdot parNum + (maxYear - minYear)$$

$locNum$ and $hostNum$ are the numbers of locations and hosts of the viruses in reassortment event, respectively. $parNum$ is the number of the parental viruses. $maxYear$ and $minYear$ are separately maximum and minimum year of the viruses in reassortment. For each virus, we calculated the cost of all possible reassortments and finally selected the reassortment with minimum cost as our result. If no reassortment was found or reassortment cost was too large in the double-reassortment detection, the virus might be reassorted from three or more parental viruses. In our work, we only considered the double-reassortment and triple-reassortment because of the complexity limitation.

2.5 Gene flow network construction

For all viruses in our dataset, we calculated $dist$ and $diff$ between each pair of genotypes. $dist$ is the number of different segment detailed types between two genotypes, where $dist = 0$ means that the two genotypes are identical and $dist = 8$ means that each segment of the genotypes is different. $diff$ is the minimum difference in host, location and year between the viruses contained in the genotypes. Based on the $dist$ and $diff$ of each pair of genotypes, we constructed the gene flow network. Each genotype was represented as a node in the network, and we constructed an edge with $diff$ value between two nodes if their $dist$ value is less than 8. The minimum spanning tree for the network was built to simplify our network. Finally, we imported our network into cystoscope v3.6.0 (Shannon, et al., 2003) and used the yFile Hierarchical Layout to show it.

3 Results

3.1 Genotype definition for IAVs

We obtained 8932 high-quality and nonredundant representative IAV genomes from 40296 whole-genome database, where each segment of IAVs was divided into a detailed type. The segments of IAVs showed high diversity in Table 1, in which each segment was divided into hundreds of detailed types. As an example, we showed the detailed type determination result for PB1 segment in Figure 1. We listed the year range, hosts, locations, and subtypes where the detailed type emerged. The detailed type determination results for other segments were shown in Supplementary Figure SF1-7. After determining the detailed types for all segments of IAVs, we got the genotypes of all the IAV genomes, which consists of 6888 unique genotypes. In fact, these 6888 genotypes represent the genotypes of the previous 40296 viruses since each filtered virus can use the genotype of its representative virus as its genotype. The genotype of each virus was shown in Supplementary Table S4.

Table 1. The number of detailed types for each segment.

Segment	No. of detailed types
PB2	534
PB1	563
PA	561
HA	633
NP	609
NA	600
MP	493
NS	663

3.2 Reassortment detection

Combining the genotype and the epidemiological information of the virus, we immediately knew a segment detailed type had occurred in which year, which host, which location and which subtype. Meanwhile, for each segment of the IAV to be analyzed, we also knew which virus had the same segment detailed type. These viruses with identical segment detailed types may provide the reassortment source for the virus in analysis. In particular, we showed the result of reassortment analysis for the virus (A/WuXi/0409/2014) in Supplementary Figure SF8, where the viruses with identical segment detailed types were listed. The genotype of this H7N9 virus is [PB2_292, PB1_415, PA_333, HA_422, NP_463, NA_405, MP_369, NS_406]. Each segment detailed type had spread widely before the virus emerged, indicating that the virus had a high possibility to be a reassortant virus. Later we attempted to identify the reassortment source and found that the virus was might generated from a triple-reassortment of H7N9 virus (A/chicken/Huzhou/3791/2013), H9N2 virus (A/chicken/Suzhou/4954/2013) and H9N2 virus (A/chicken/Jiangsu/SIC11/2013) (Supplementary Figure SF9). The triple reassortant virus acquired its HA and NA segments from H7N9 virus (A/chicken/Huzhou/3791/2013). The PB2, PA, NP and MP segments were derived from H9N2 virus (A/chicken/Suzhou/4954/2013). Another H9N2 virus (A/chicken/Jiangsu/SIC11/2013) provided PB1 and NS segments for the triple reassortant virus. In a word, the H7N9 virus (A/WuXi/0409/2014) obtained HA and NA segments from its parental H7N9 virus, and internal segments from its parental H9N2 viruses. Obviously, with our segment detailed type determination and genotype nomenclature, it will be an easy task to detect reassortment for IAVs.

3.3 The reassortment history of IAVs

In order to construct the reassortment history of IAVs, we detected the reassortment events for all viruses in our dataset, using the method we proposed above. The result of reassortment history was shown in Figure 2, in which we displayed each reassortment event as the connections from the parental viruses to the reassortant strain. We also shown another form of the result in Supplementary Figure SF10. It should be noted that the reassortment event we detected first emerged in North America, but this does not mean that the origin of IAVs was in North American, since the early sequencing technology around the world was not as advanced as it is now. The reassortment history of IAVs generally reflects the reassortant trend, in which we can know which locations and which hosts are more likely to generate reassortment. In our work, we found that the frequency of inter-locations reassortment is high in various locations of Asia and Europe, which can be illustrated from

the large number of red lines in these locations in Figure 2. Although the number of reassortment events in North America is large, most of them are intra-locations reassortment. This finding was also reflected in Figure 3, in which we counted the number of reassortment events in different locations each year.

3.4 Swine act as an intermediate host in the reassorment history of IAVs

In reassortment history of IAVs above, we limited the number of reassortment sources due to the limitation of calculation. Actually, the reassortment of IAVs is complicated, since the genome of an IAV can come from eight different viruses at most. In order to analyze how the genes of IAVs are transmitted during the reassortment, we constructed the gene flow network as shown in Figure 4. The network was composed of nodes and edges, where each node was a genotype of IAVs and the edge represented gene segment flow by reassortment between two genotypes. It should be noted that we found some conclusions which is consistent with previous studies after we added the host information to the network. Swine as an intermediate host played an important role in the gene flow between avian and human IAVs, which can be seen from the swine host nodes in the figure as the hub connecting the avian host nodes and the human host nodes. Actually, it is universally acknowledged that swine are the “mixed vessels” for IAVs since they can be infected by both avian and human IAVs and facilitate the reassortment events of the IAVs (Amélie, et al., 2018). Meanwhile, due to the global live swine trade, swine also plays a crucial role in the global spread of IAVs, which provides the possibility of cross-locations reassortment for IAVs (Nelson, et al., 2015). Therefore, our result on swine as an intermediate host is consistent with previous studies on the role of swine.

3.5 Frequent reassortment between H9N2 and H7N9 viruses

Analyzing the subtypes of IAVs reassortment can also help us understand the patterns of the reassortment. The avian influenza viruses with H9N2 subtype are widely distributed in different regions of China, and provide internal segments for IAVs for other subtypes, which poses a serious threat to public health (Liu, et al., 2014). In our study, we also found a similar result after we added the subtype information to the gene flow network. It has been shown in Supplementary Figure SF11 that there are many intersections between H9N2 and H7N9 viruses, which means that frequent reassortments occurred in IAVs with such subtypes. Later we found that almost all of these viruses originated from domestic birds in East Asia (Supplementary Figure SF12). It has been demonstrated that H9N2 has replaced H5N6 and H7N9 as the dominant AIV subtype in both chickens and ducks of China, and the novel viruses generated from reassortment continued to appear and spread among the

avian hosts (Bi, et al., 2020). The internal segments of some H9N2 viruses have acquired fitness in human, while viruses with other subtypes such as H7N9 can acquire this fitness through reassortment with them (Liu, et al., 2013; Lu, et al., 2014; Zhang, et al., 2013). The reassortment analysis for the H7N9 virus (A/WuXi/0409/2014) (Supplementary Figure SF9) revealed that the virus obtained HA and NA segments from its parental H7N9 virus and internal segments from its parental H9N2 viruses. It is noteworthy that this reassortant virus was sampled from a human and its parental viruses originated from domestic birds. In other words, the parental viruses completed cross-species transmission through this reassortment patter.

4 Discussion

In order to verify the correctness of our clustering method, we used the unified nomenclature system for the highly pathogenic H5N1 avian influenza viruses (Organization, 2011) to cluster HA segments of H5N1 viruses in the database and compared with our result. The unified nomenclature system used three specific clade definition criteria: sharing of a common (clade-defining) node in the phylogenetic tree; monophyletic grouping with a bootstrap value of ⩾60 at the clade-defining node; and average percentage pairwise nucleotide distances between and within clades of >1.5% and <1.5%. We marked the clustering results of the two methods on the phylogenetic tree, which was shown in Supplementary Figure SF13. The inner circle was the clustering result of the unified nomenclature system, and the outer circle was our clustering result. The clustering result of the unified nomenclature system was more detailed than ours, but the clustering boundaries of the two methods were mostly synchronized.

IAVs are rapidly evolving through reassortment, posing a huge challenge to vaccine development and a serious threat to public health. Traditional studies in IAVs reassortment often focus on the specific virus that causes a certain epidemic, lacking a systematic understanding of IAVs reassortment history. In the present study, we divided each segment of IAVs into detailed types according to the Mean Pairwise Distance (MPD) on the phylogenetic trees. A new method was proposed to define the genotype of IAVs, and the genotype nomenclature was then exploited to analyze whole genomes of IAVs and detect the reassortment. Generally, all the segments of the virus should be compared comprehensively when detecting the reassortment source of IAVs, rather than considering only a single segment. If some segments of a virus have high homology with a parental virus and low homology with another parental virus, and the circumstance in other segments is opposite, the virus is likely to be generated from a reassortment of the two parental viruses (Tian, et

al., 2011). The reassortment detection method in our work not only compared the source of each segment but also considered the epidemiological information of the IAVs, which is more reasonable. Different from previous reassortment detection method (Niranjan and Carl, 2011; Suzuki, 2010; Svinti, et al., 2013; Yurovsky and Moret, 2011), we quantitatively represent the inconsistency of the phylogenetic trees as different segment detailed types. Meanwhile, the results produced by our method are consistent with previous studies, which confirms the effectiveness of our method. Therefore, the method we proposed can indeed effectively define the genotype of IAV and detect the reassortment, especially for large-scale data sets.

It should be pointed out that our result only listed the reassortment events we found. Some IAVs were generated from reassortment but we did not detect them because their parental viruses could not be found due to insufficient sampling or incomplete information. Our research was based on the sequence data and the epidemiological information, which benefits from the continuous monitoring and sequencing of IAVs by researchers around the world.

Taken together, we proposed a novel reassortment detection method based on our genotype nomenclature and systematically studied the history and patterns of reassortment in IAVs, which will be beneficial for the prevention and control of IAVs worldwide.

Figure Legends

Figure. 1. The detailed type result for PB1 segment. The year range, hosts, locations and subtypes are shown after each PB1 type, where the circles represent hosts and the rectangles represent locations. The hosts and locations are distinguished by different colors.

Figure. 2. The reassortment history of IAVs. The horizontal axis represents the locations, while the vertical axis represents the years. Each circle represents a virus, with different colors to indicate the hosts. The line from source to target represents the parental virus produces the reassortant virus. Intra-Locations and Inter-Locations reassortment are indicated by green and red lines, respectively.

Figure. 3. The number of reassortment events per year in different locations. The horizontal axis represents the years, and the vertical axis represents the number of reassortment events. The locations of the reassortant viruses are distinguished by different colors. The number of reassortment events per year was divided into three types for statistics: all reassortments, inter-location reassortments and cross-species reassortments.

Figure. 4. The gene flow network of IAVs (with host information annotated). Each node is a genotype of IAVs and the edge represents gene segment flow by reassortment between two genotypes. The hosts are distinguished by different colors, where red represents human, green represents avian, blue represents swine, and grey represent other hosts.

Funding

This work was supported by the National Natural Science Foundation of China [grant number 32070025, 31800136, 82041019]; and the Research Project from State Key Laboratory of Pathogen and Biosecurity [grant number SKLPBS1807].

Conflict of Interest*:* The authors declare no competing interests.

References

Amélie*, et al.* Spatio-temporal distribution and evolution of the A/H1N1 2009 pandemic virus in pigs in France from 2009 to 2017: identification of a potential swine-specific lineage. *Journal of virology* 2018.

Bi, Y.*, et al.* Dominant subtype switch in avian influenza viruses during 2016–2019 in China. *Nature Communications* 2020;11(1):5909.

Dhanasekaran, V., Mukerji, R. and Smith, G. RNA Virus Reassortment: An Evolutionary Mechanism for Host Jumps and Immune Evasion. *PLoS pathogens* 2015;11:e1004902.

Katoh, K. and Standley, D. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. *Molecular biology and evolution* 2013;30.

Lemey, P.*, et al.* Bayesian Phylogeography Finds Its Roots. *PLoS computational biology* 2009;5:e1000520.

Liu, D., Shi, W. and Gao, G.F. Poultry carrying H9N2 act as incubators for novel human avian influenza viruses. *Lancet* 2014;383(9920):869-869.

Liu, D.*, et al.* Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses. *The Lancet* 2013;381(9881):1926-1932.

Liu, W.*, et al.* On the Centenary of the Spanish Flu: Being Prepared for the Next Pandemic. *Virologica Sinica* 2018;33.

Liu, W.J.*, et al.* Emerging HxNy Influenza A Viruses. *Cold Spring Harbor Perspectives in Medicine* 2020.

Lu, G.*, et al.* FluGenome: a web tool for genotyping influenza A virus. *Nucleic Acids Research* 2007;35(Web Server):W275-W279.

Lu, J.*, et al.* Continuing Reassortment Leads to the Genetic Diversity of Influenza Virus H7N9 in Guangdong, China. *Journal of Virology* 2014;88(15):8297-8306.

Ma, W., Kahn, R.E. and Richt, J.A. The pig as a mixing vessel for influenza viruses: Human and veterinary implications. *Journal of Molecular & Genetic Medicine An International Journal of Biomedical Research* 2009;3(1):158-166.

McDonald, S.M.*, et al.* Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nature Reviews Microbiology* 2016;14(7):448-460.

Nelson, M.I.*, et al.* Multiple Reassortment Events in the Evolutionary History of H1N1 Influenza A Virus Since 1918. *Plos Pathogens* 2008;4(2):e1000012.

Nelson, M.I.*, et al.* Global migration of influenza A viruses in swine. *Nature Communications* 2015;6(1):6696.

Nguyen, L.-T.*, et al.* IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 2015;32(1):268-274.

Niranjan, N. and Carl, K. GiRaF: robust, computational identification of influenza reassortments via graph mining. *Nucleic Acids Research* 2011(6):e34.

Organization, W.H. Updated unified nomenclature system for the highly pathogenic H5N1 avian influenza viruses. *Influenza* 2011.

Petrova, V.N. and Russell, C.A. The evolution of seasonal influenza viruses. *Nature Reviews Microbiology* 2018;16(1):47-60.

Rabadan, R., Levine, A.J. and Krasnitz, M. Non‐random reassortment in human influenza A viruses. *Influenza and Other Respiratory Viruses* 2008;2.

Shannon*, et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* 2003;13(11):2498-2504.

Silva, U.*, et al.* A comprehensive analysis of reassortment in influenza A virus. *Biology open* 2012;1:385-390.

Smith, G.J.D. The emergence of pandemic influenza viruses. *International Journal of Infectious Diseases* 2011;15:S7-S8.

Smith, G.J.D.*, et al.* Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 2009;459(7250):1122-1125.

Stohr, K. Influenza-WHO cares. *The Lancet Infectious Diseases* 2002;2.

Suzuki, Y. A phylogenetic approach to detecting reassortments in viruses with segmented genomes. *Gene* 2010;464(1):11-16.

Svinti, V., Cotton, J.A. and McInerney, J.O. New approaches for unravelling reassortment pathways. *BMC Evolutionary Biology* 2013;13(1):1.

Tian, D., Wang, Y. and Tao, Z. A novel strategy for exploring the reassortment origins of newly emerging influenza virus. *Bioinformation* 2011;7(2):64-68.

Wallace, R. and Fitch, W. Influenza A H5N1 Immigration Is Filtered Out at Some International Borders. *PloS one* 2008;3:e1697.

Wallace, R.G.*, et al.* A statistical phylogeography of influenza A H5N1. *Proceedings of the National Academy of Sciences* 2007;104(11):4473.

Yurovsky, A. and Moret, B. FluReF, an automated flu virus reassortment finder based on phylogenetic trees. *Bmc Genomics* 2011;12(Suppl 2):S3-S3.

Zhang, D.*, et al.* PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources* 2020;20(1):348-355.

Zhang, L., Zhang, Z. and Weng, Z. Rapid Reassortment of Internal Genes in Avian Influenza A(H7N9) Virus. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2013;57.

Figure 1

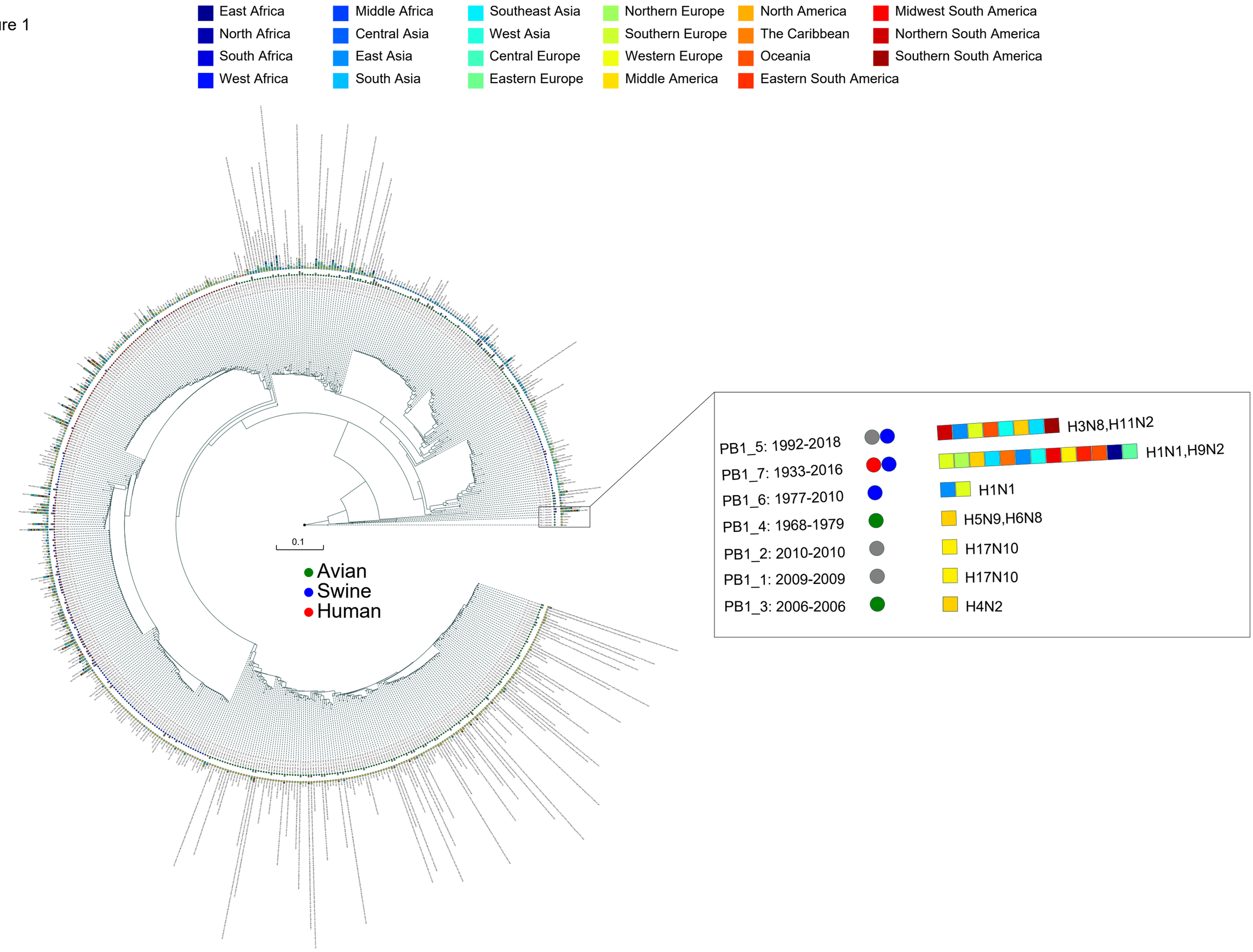

East Africa
North Africa
South Africa
West Africa
Middle Africa
Central Asia
East Asia
South Asia
Southeast Asia
West Asia
Central Europe
Eastern Europe
Northern Europe
Southern Europe
Western Europe
Middle America
North America
The Caribbean
Oceania
Eastern South America
Midwest South America
Northern South America
Southern South America
0.1
Avian
Swine
Human
PB1_5: 1992-2018
PB1_7: 1933-2016
PB1_6: 1977-2010
PB1_4: 1968-1979
PB1_2: 2010-2010
PB1_1: 2009-2009
PB1_3: 2006-2006
H3N8,H11N2
H1N1,H9N2
H1N1
H5N9,H6N8
H17N10
H17N10
H4N2

Figure 2

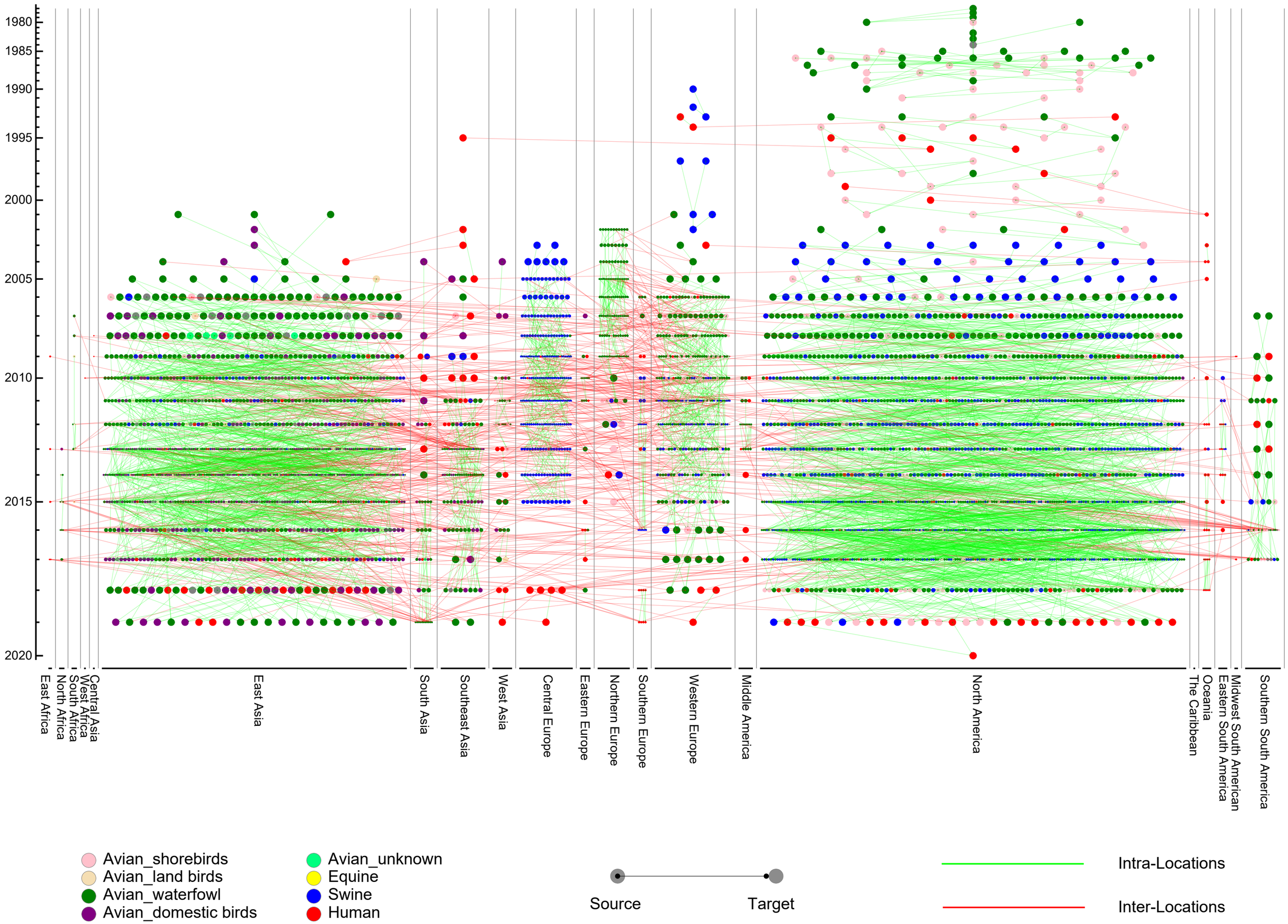
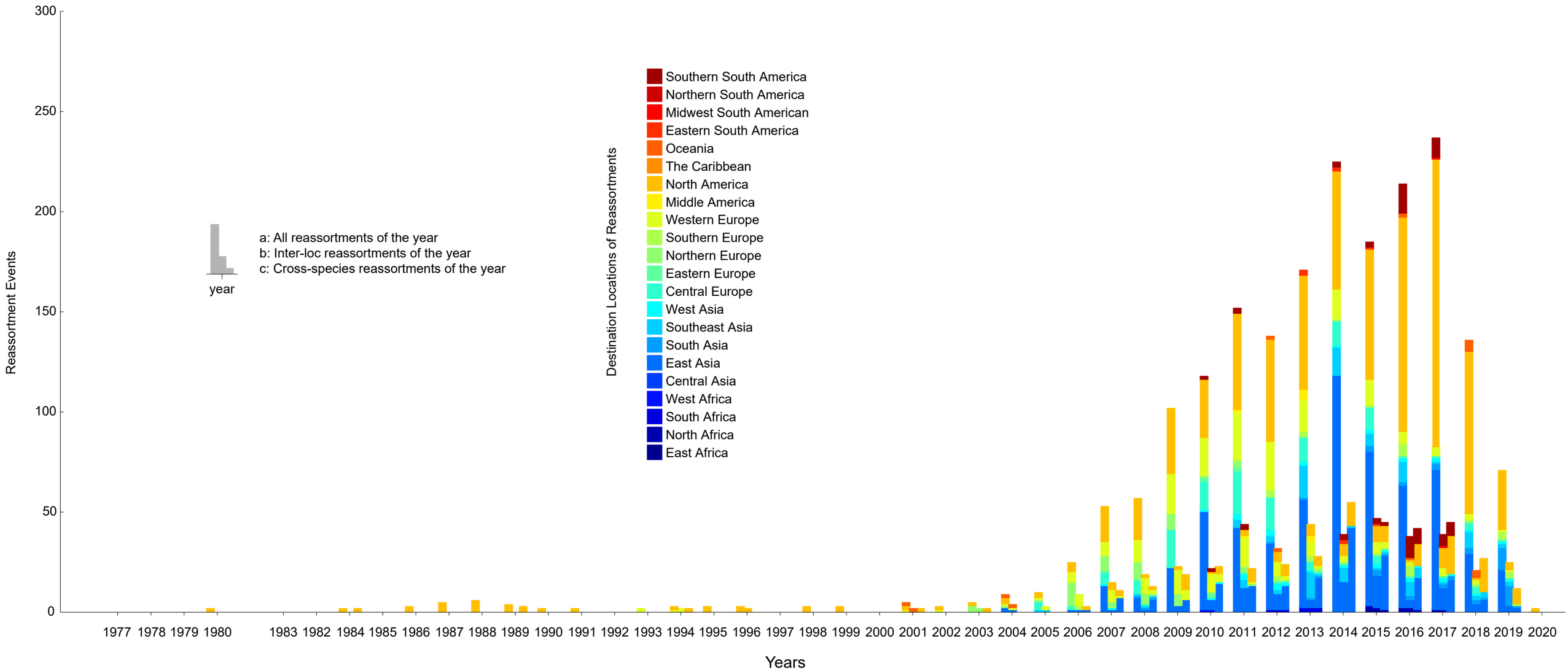

1980
1985
1990
1995
2000
2005
2010
2015
2020
East Africa
North Africa
South Africa
West Africa
Central Asia
East Asia
South Asia
Southeast Asia
West Asia
Central Europe
Eastern Europe
Northern Europe
Southern Europe
Western Europe
Middle America
North America
The Caribbean
Oceania
Eastern South America
Midwest South American
Southern South America
Avian_shorebirds
Avian_land birds
Avian_waterfowl
Avian_domestic birds
Avian_unknown
Equine
Swine
Human
Source
Target
Intra-Locations
Inter-Locations

Figure 3


Reassortment Events
300
250
200
150
100
50
0
year
a: All reassortments of the year
b: Inter-loc reassortments of the year
c: Cross-species reassortments of the year
Destination Locations of Reassortments
Southern South America
Northern South America
Midwest South American
Eastern South America
Oceania
The Caribbean
North America
Middle America
Western Europe
Southern Europe
Northern Europe
Eastern Europe
Central Europe
West Asia
Southeast Asia
South Asia
East Asia
Central Asia
West Africa
South Africa
North Africa
East Africa
1977 1978 1979 1980 1983 1982 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020
Years

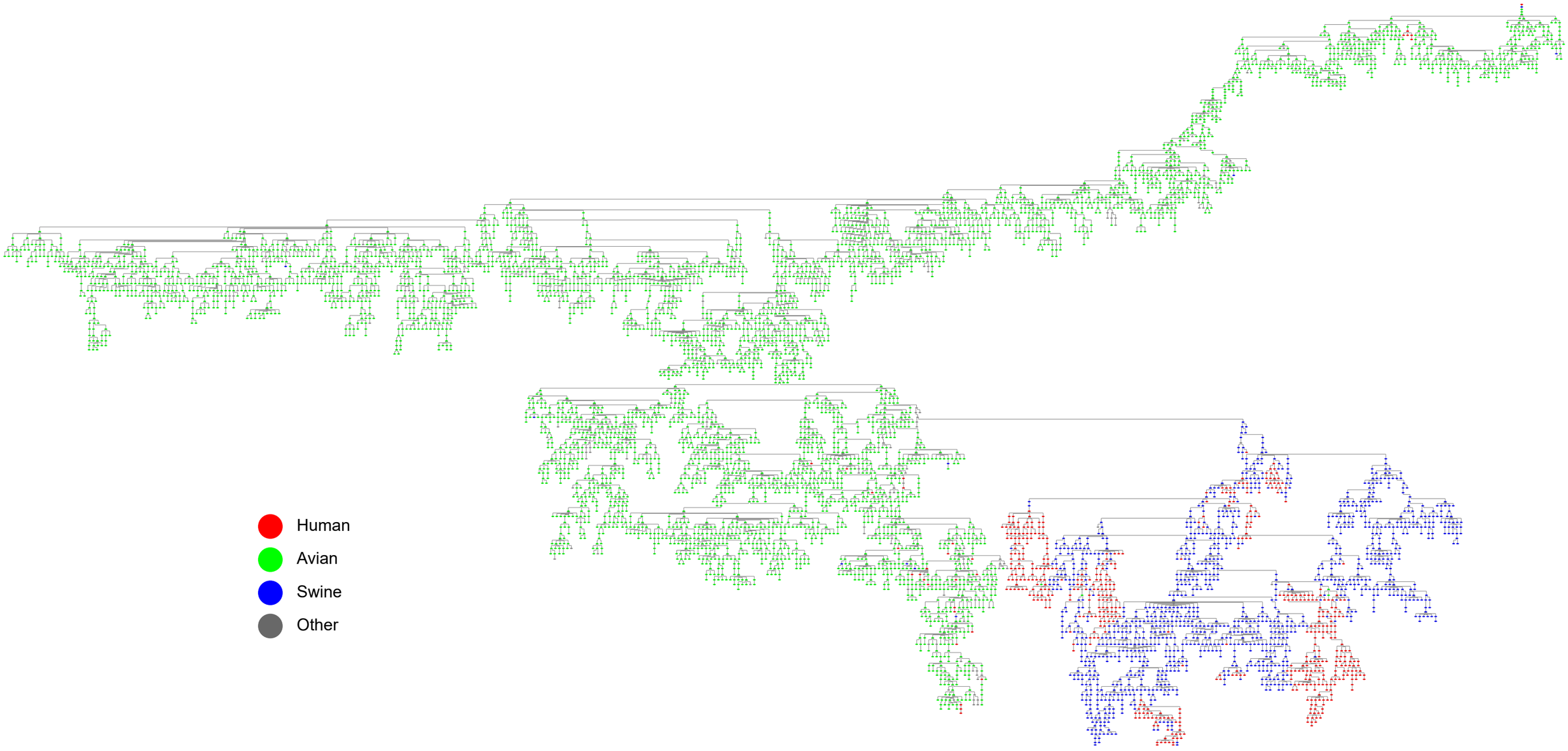



Figure 4