Phylogeographical Analysis and Visualization of Data of *Trypanosoma* Isolates from Different Foci in Nigeria Using Microreact and Phandango (2003 – 2018)

Akinseye Olanrewaju Roland [1*], Gbadamosi Folawiyo I.[1], Ale Ebenezer Morayo [2], Ojomo L. Joan [1], Adelabu Mustapha [1] and Olaleye Olusola [1]

[1] *Nigerian Institute for Trypanosomiasis & Onchocerciasis Research, Ibadan, Nigeria.*
[2] *Department of Biochemistry, Faculty of Pure and Applied Sciences, Federal University Wukari, Nigeria.*

Authors' contributions

This work was carried out in collaboration among all authors. Author AOR designed the study, performed the data retrieving and phylogenetic analysis, wrote the protocol, and wrote the first draft of the manuscript. Authors IGF, AEM and OLJ managed the analyses of the study. Author AEM managed the literature searches. All authors read and approved the final manuscript.

| Short Research Article |

ABSTRACT

Introduction: A combination of phylogenetic tree of trypanosome isolates and the geographical information of the epidemiologic foci can acquaint us with the occurrences of African trypanosomiasis outbreak necessitous for educating the spatial distribution of the parasites, vectors and drawing inferences for planning strategic control programs. Microreact and Phandango are free online tools that enables the visual representation and study of genomic epidemiology of infectious diseases. This study elicits the phylogenetic data of trypanosome isolates and the specific host along with geographical information

_____

Corresponding author: E-mail: Akinseyeroland@gmail.com;

Methods: A total of 46 isolates. Protein sequence data of trypanosomes isolated in Nigeria between 2003 and 2018 were extracted from National Center for Biotechnology Information (NCBI) database. Multiple sequence alignment and tree construction were carried out on sequence data files and Newick output was downloaded. Metadata .csv files were prepared accordingly and loaded for visualization in microreact and phandango.

Results: The preponderance of Glossina palpalis was reported in old Oyo National Park and they were majorly incriminated in the transmission of Trypanosoma gravi in *Glossina palpalis*. *Trypanosoma theileri* was isolated from the midgut of *Glossina morsitan* reported to be found in Kainji Lake National Park only. *Trypanosoma congolense* was reportedly isolated from cattle at Zaria, Glossina morsitans at Kainji Lake National Park, Glossina tachinoid at old Oyo National Park at Oyo and, *Glossina palpalis* at Yankari Game Reserve Bauchi. Trypanosoma brucei brucei was isolated from a rat in April 2018 at Federe and Vom, Plateau state. In Federe and Vom, *Trypanosoma brucei brucei* was found in rat in April 2018 while *Trypanosoma evansi* was obtained from animal in Vom December 2008. In Bida and Tsuaa, *Trypanosoma brucei gambiense* was isolated from unidentified animals (probably cattles) in October 2009. The analyses also revealed 7 protein sequence which includes 7 protein sequenced including cathepsin L-like protein (CPL) and glycosomal glyceraldehyde-3-phosphate dehydrogenase (G-GAPDH) from *Tryapanosoma vivi* and *Trypanosoma congolense*, glycoprotein (GP) from *Trypanosoma brucei gambiense*; trans-sialidase (TSs) from *Trypanosoma evansi*; phospholipase A2 (PLA2) and hexokinase 1 (HK1) from *Trypanosoma brucei brucei;* and also glycosomal glyceraldehyde-3-phosphate dehydrogenase (G-GAPDH) from *Trypanosoma theileri*.

Conclusion: The study generated a concise visual representation and trajectory of trypanosome isolates in Nigeria along with their genetic metadata and geographic location on a phylogenetic tree which will serve as a reference for information on trypanosomiasis outbreaks in Nigeria, provide blue prints for future outbreak predictions and guide decision-making on strategic control programs.

1. INTRODUCTION

Trypanosomes are members of protozoans of the genus Trypanosoma which proliferate in the body fluids mostly blood of the vertebrate hosts such as man, domestic and wild animals. They reside in the salivary glands or alimentary canal of the invertebrate hosts, precisely tsetseflies and other biting insects [1]. Trypanosomes infection often vary with hosts [2]. Notable trypanosomes species such as *T. congolense*, *T. vivax*, *T. brucei*, *T. grayi*, *T. theileri*, *T. simiae*, etc. infect both vertebrate and invertebrate hosts [3] [2] [4] [5].

Consequent to the suitable habitat provided by the national park for the flies and the pathogenic trypanosome, the wild animal in the park and domestic livestock grazing around are steadily exposed to the risk of trypanosomiasis [6]. Cases of tsetse flies has been documented in the national parks and game reserves where diverse species of living organisms are protected [7] [8] OIE, 213; Isaac *et al.*, [3] [9] [10] [5]. It is therefore necessary to keep track of the distribution of these flies in already existing foci to be able to predict their prevalence and assess the level of possible havoc they can wreck during an outbreak.

A combination of phylogenetic tree of trypanosome isolates and the geographical information of the epidemiologic foci can serve as a tool that provides critical information regarding the occurrences of African trypanosomiasis outbreak necessitous for educating the spatial distribution of the parasites, vectors as well as drawing inferences for planning strategic control programs. Locally and globally, phylogenetic serve as an essential tool for identifying disease outbreaks and monitoring the spread of a pathogen. This tool is capable of monitoring the spread or distribution of isolates as it can infer relationship between samples collected at various times, sources and places. Scientists have designed tools which can make use of these information (phylogenetic tree of trypanosome isolates and the geographical foci) and produce visual charts output. Examples of such software include microreact and phandango which were employed in this research.

Microreact is a free online tool that will enable you to visualize phylogenetic relationships of

isolates linked to geographic locations. This tool allows for dynamic visualization of the data with an interactive map using your own phylogenetic tree and metadata [11].

Additionally, phandango is an interactive visualisation tool for phylogenetics trees and can be viewed with other information that would benefit from a visual output such as metadata (e.g. resistance, virulence, serotyping and multi locus sequence typing or MLST data) and genomic structure information (e.g pangenome, recombination blocks, genome wide association studies or GWAS) [12].

This project is aimed at generating visual chart for the phylogenetic data combined with geographical information to navigate the vectors/hosts and *Trypanosoma* species in specific existing foci using microreact and phandango using protein sequence of *Trypanosoma* species found in Nigeria between 2003 – 2018 datasets obtained from National Center for Biotechnology Information (NCBI) website.

2. MATERIALS AND METHODS

2.1 Data Analyis

Protein sequence of Trypanosome isolates (represented by accession number) found in Nigeria between 2003 and 2018 were downloaded from National Center for Biotechnology Information (NCBI) database in 15th, September, 2020.

Multiple sequence alignment was performed using Multiple Alignment using Fast Fourier Transform (MAFFT). Phylogenetic tree was constructed and estimated by Phylogeny Maximum Likelihood (PhyML option in Seaview version 2.0), with blossom62 model and bootstrap set to 100 replicates as described by Guindon *et al.* [13]. The output was saved in NEWICK version (.nwk format).

2.2 Geographical Information of Isolate Origin

Geographic coordinates of epidemiological foci of the trypanosome isolates were generated with the web-based Google Earth application (Google Inc., California, USA).

2.3 Metadata File for Phandango Preparation

A metadata file (.csv format) was prepared (in excel readable by phandango server https://jameshadfield.github.io/phandango/#) from the protein sequenced data using information such as accession number, years, month and of collection, location of collection, amino acids sequenced and identity of protein sequenced. Protocols for entry data found in https://en.wikipedia.org/wiki/Newick_format were adhere to as describe by Hadfield *et al.* [12]. The .csv file was saved as comma delimited format.

2.4 Metadata File for Microreact Preparation

A metadata file (.csv format) prepared in excel readable by microreact server (https://microreact.org/) from the protein sequenced data which contains details such as accession number, years, month and of collection, location of collection, amino acids sequenced and identity of protein sequenced. Geographical information of the various foci was obtained from Google Earth software as described above. Protocols for entry and saving data found in https://microreact.org/instructions were adhere to as describe by Argimón *et al.* [11]. The .csv file was saved as comma delimited format.

2.5 Loading and Visualization of Data in Microreact and Phandango

Google Chrome application (Google Inc., California, USA) was used to access the microreact website (https://microreact.org/showcase). The folder containing the newick and metadata files were uploaded on the microreact web page to create the microreact project. Then, the folder where the phylogenetic tree (in .nwk format) and the metadata file were saved was navigated to. The two files were selected and drag and drop at the upload page. Then, the project was created. Some adjustments were effected to produce elegant output. The map panel is on the left, the phylogenetic tree and the legend on the right panel and the timeline at the bottom.

Fig. 1.1. Cladogram representation of Increasing order nodes of Phylogenetic tree

3. RESULTS

The cladogram represents the internal nodes (branch points) which diverge into different groups and the most recent common ancestor of all the groups descended from that branch point. The tips represent the Trypanosoma species in the tree evolved from a series of common ancestors.

3.1 Phylogenetic and Phandango Visualization

Phandango chart Fig. 1.2 consist of 46 isolates with 7 different sequence proteins, 7 different *Trypanosoma* species which were obtained from 3 isolation sources (blood, gut and proboscis) from 7 vectors/hosts found in 10 foci across the Nigeria.

Comparing the genetic relationship of isolates (represented by accession numbers) and the hosts/ *Trypanosoma*, isolates that are in group in the tree, have similar grouping in organism name (orgname).

Fig. 1.3 contain the visualized chart of phylogenetic tree in relation to protein sequenced from *Trypanosoma* species and we can deduced that cathepsin L-like protein (CPL) and glycosomal glyceraldehyde-3-phosphate dehydrogenase (G-GAPDH) are sequenced from *Tryapanosoma vivi* and *Trypanosoma congolensis;* glycoprotein (GP) is sequenced from *Trypanosoma brucei gambiens*; trans-sialidase (TSs) from *Trypanosoma evansi*; phospholipase A2 (PLA2) and hexokinase 1 (HK1) from *Trypanosoma brucei brucei;* and also glycosomal glyceraldehyde-3-phosphate dehydrogenase (G-GAPDH) from *Trypanosoma theileri*.

The chart consists of four components including the map panel on the left, the phylogenetic tree on the right panel, the legend and the timeline at the bottom panel.

The microreact and phandango chart show the relationship of phylogeny with various species of *Trypanosoma* isolated from sources in diverse hosts at different foci of unique geographical information and date as shown in Fig. 1.4. *Tryapanosoma vivi* (purple) are from isolates (ACJ02116, ACJ02117, ACJ02118, ACJ02119, ACJ02120, ALG02550, CAF04259, QFR04607, QFR04608, QFR04609, QFR04610, QFR04611, QFR04612, QFR04613, QFR04614, QFR04615, QFR04616, QFR04617,QFR04618) found in diverse hosts and different foci. They are isolated from the blood of cattle in Zaria in march 2009; proboscis of *Glossina morsitan* in Yankari Game Reserve, Ijah Gwari and Kainji Lake National

Park in [14] proboscis of *Glossina palpalis* in the Old Oyo National Park and Kainji Lake National Park in [14] and in the blood of Ovis aries (sheep) in Vom, Nigeria in 2003.

Trypanosoma evansi (deep green) from isolate (ACM43506) is obtained from the blood of unidentified host in Vom, Nigeria in December, 2008. *Trypanosoma brucei brucei* (dark blue) (AEX60759, AYK27210) are identified in the blood of rat in Vom in august, 2011; and in unidentified host in Federe, Nigeria in April, 2018. *Trypanosoma congolensis* (blue) (AHG56298, AHG56299, AHG56300, QFR04588, QFR04589, QFR04590, QFR04591, QFR04592, QFR04593) are as well isolated from the blood of cattle in Zaria in July, 2013; guts of *Glossina morsitan* and *Glossina tachnoid* in Yankari Game Reserve and Kainji Lake National

Park in 2014; guts of *Glossina palpalis* in Ijah Gwari and Old Oyo National Park, 2014. *Trypanosoma brucei gambiense* (light blue) (CBG92831, CBG92832) are obtained from unidentified host in Tsuaa and Bida, Nigeria October, 2009. *Trypanosoma grayi* (yellow) (QFR04594, QFR04595, QFR04596, QFR04597, QFR04598, QFR04599, QFR04600, QFR04601, QFR04602, QFR04603, QFR04604) are isolated from the guts of *Glossina palpalis* in Old Oyo National Park and Cross River Park in March, 2014; proboscis of *Glossina palpalis* in Old Oyo National Park, 2014; and also proboscis of *Glossina tachnoid* in Kainji Lake National Park in March [13]. *Trypanosoma theileri* (green) (QFR04605, QFR04606) are only obtained from the guts of *Glossina tachnoid* in Kainji Lake National Park in March [13].
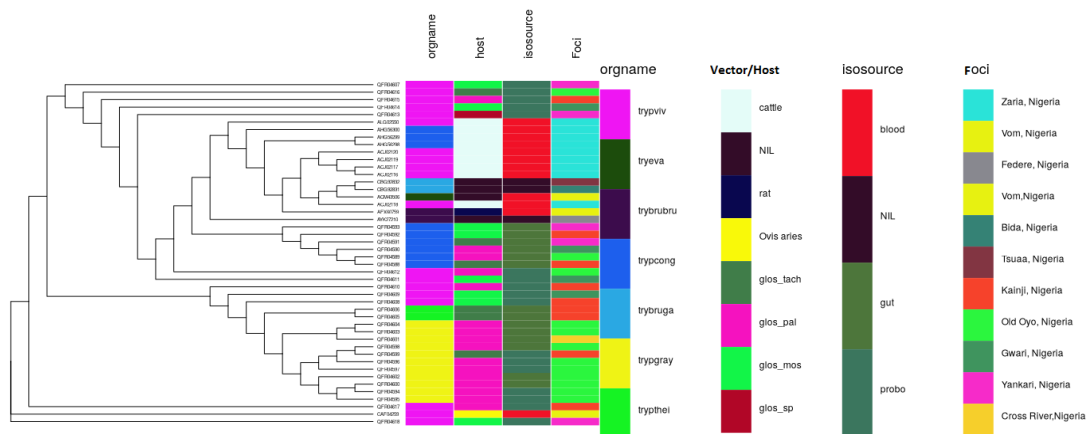

Fig. 1.2. Phandango chart of phylogenetic tree of isolates in relationship with *Trypanosoma* species, hosts, foci and isolation source
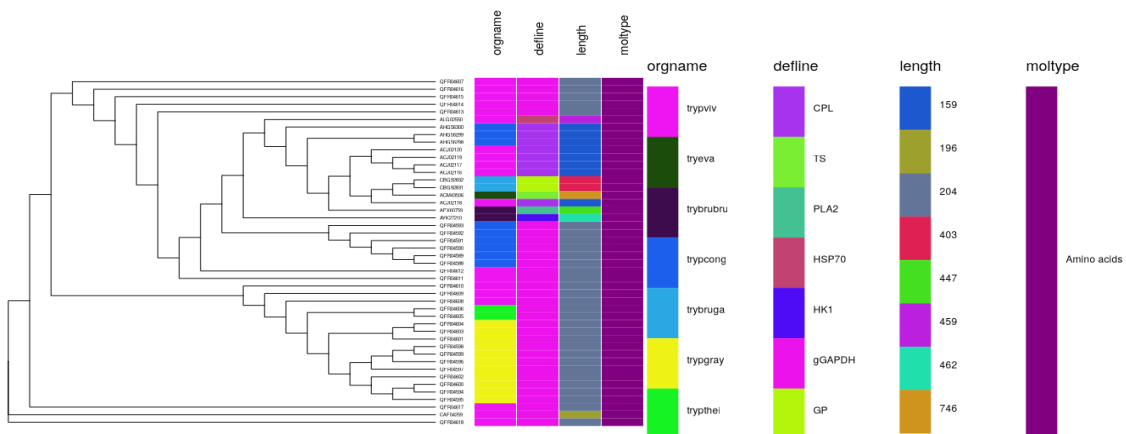

Fig. 1.3. Phandango chart of phylogenetic tree of isolates in relationship with *Trypanosoma* species, protein sequenced and others
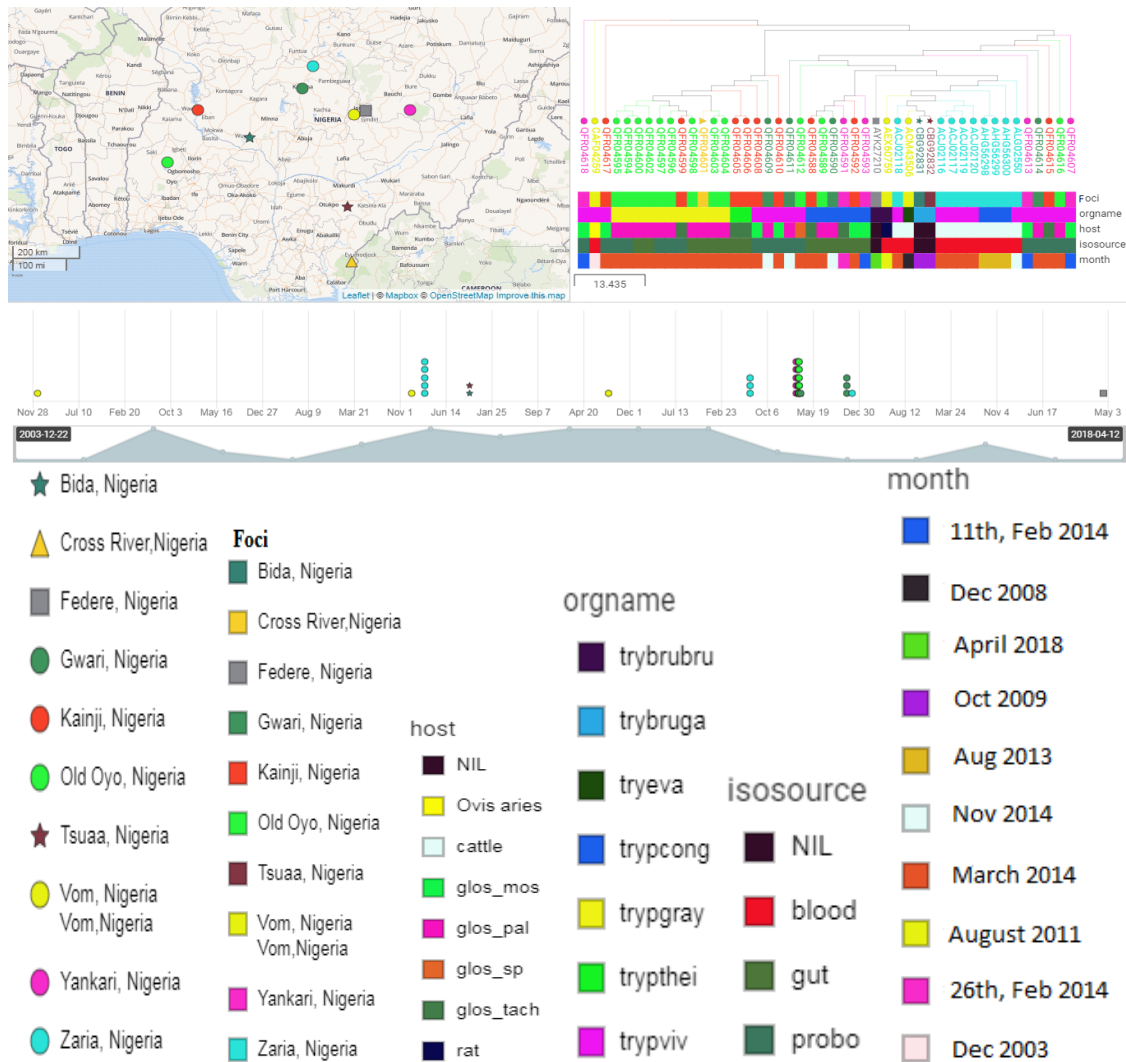
Fig. 1.4. Microreact chart embedded with phandango metadata information
Abbreviations: organism name (orgname); *Trypanosoma congolense* (trypcong), *Trypanosoma grayi*
(trypgray), *Trypanosoma brucei brucei* (trybrubru), *Trypanosoma evansi* (tryeva), *Trypanosoma theileri*
(trypthei), *Trypanosoma vivax* (trypviv), *Glossina sp.* (glos_sp), *Trypanosoma brucei gambiense*
(trybruga); Host– *Glossina tachenoid* (glos_tach), *Glossina palpalis* (glos_pal), *Glossina morsitans*
(glos_mos), *Glossina sp.* (glos_sp); isolation source (isosource), proboscis (probo)
https://microreact.org/project/4SM4tVyb4FHfAo6PsWqKfZ/87dcfa8d

Glossina palpalis are readily abundant in Old Oyo National Park and they transmitted mostly *Trypanosoma grayi*. The large grazing activities of cattle in Zaria could be responsible for the large number of *Trypanosoma vivi* parasites recorded. *Trypanosoma theileri* isolated from *Glossina morsitan* are only found in Kainji Lake National Park. *Trypanosoma congolensis* are identified to be readily isolated in the blood of cattle, *Glossina morsitan*, *Glossina tachnoides* and *Glossina palpalis* in Zaria, Old Oyo National Park and Yankari Game Reserve respectively.

4. DISCUSSION

Vectors are sensitive to climate change and other environmental factors which together affect Trypanosomiasis by influencing their distribution. Increased understanding in the population structure of Glossina species and other hosts in the different vegetation zones of Nigeria will aid decisions on appropriate control strategies [9].

In Fig. 1.4, the chart consist of three major informative panel. The phylogenetic tree panel describes the isolate phylogeny and coloured

based on the location/foci of isolation. The timeline panel at the bottom describe the (months, days and years of isolation) season of prevalence of vectors and the map panel gives the geographical information of the foci of vectors and isolates.

The chart in figure 1.4 mainly used the information on phylogeny isolates and foci to explain vector distribution through accurate interpretation of the data displayed on the panels.

The chart shown phylogeny isolates including (QFR04607, QFR04613, QFR04593, QFR04618) and (QFR04592, QFR04610, QFR04608, QFR04617) found in Yankari Game Reserve and Kainji Lake National Park foci respectively showed that *Glossina morsitan* vector was geographically restrained to the foci forming cluster on the timeline (February, 2014 – March 2014). The clustered explained that the vectors are in the same foci or region. Yankari Game Reserve is situated in Bauchi State within the Northern Guinea/Sudan savannah vegetation zone and is located between latitude 9° 45′ N and longitude 10° 30′ E. Kainji Lake National Park covers an area of 5340 km^2 across Niger and Kwara States (10° 22′ N and 4° 33′ E), within the southern guinea savannah vegetation zone. The morsitans tsetse fly formed the primary vector of trypanosomiasis that thrive mostly in savannah. The species identified in an earlier study [3] [9] was *Glossina morsitan submorsitans*, which corroborates our results.

The chart also displayed phylogenetic isolates including (QFR04614, QFR04590, QFR04611, QFR04609), (QFR04616, QFR04589, QFR04604, QFR04603, QFR04598, QFR04596, QFR0497, QFR04602, QFR04600, QFR04594, QFR04595), and (QFR04601) found in Ijah Gwari, Old Oyo National Park, and Cross River National Park respectively showed that *Glossina palpalis* vector was geographically restricted to specific foci forming clusters in the timeline (2013-2014 and 2014-2015). Old Oyo National Park is located in the northern part of Oyo State, South Western Nigeria. It has a total land area of 2512 km2. The Park lies within the derived savannah vegetation zone between latitudes (8° 10′- 9° 05′ N) and longitudes (3° 00′ - 4° 02′ E). Cross River National Park is located in the rainforest ecological zone in the extreme South East of Nigeria on the border with Republic of Cameroon. The park occupies a total land area of about 4000 km2 of tropical rain forest

ecosystem and lies between latitude (5° 05′ - 6° 29′ N) and longitude (8° 15′ - 9° 30′ E) [14]. Ijah Gwari (near Suleja) is located between latitude 9° 12′ N and 9° 24′ N and longitude 7° 12′ E and 7° 20′ E in Tafa Local Government Area of Niger State. It is characterized by several small streams traverse the area and the vegetation is riverine fringing forest forming a dense two-storey canopy. Timeline information explained that *Glossina palpalis* vector was readily found in Old Oyo National Park, and Cross River National Park around March, [14], but however was found abundant in November, 2014 in riverine fringing forest of Ijah Gwari due to seasonal fluctuation. The fact that *Glossina palpalis* vector is readily found in different habitat in distinct timeline may be attributed to variations in microclimatic conditions at the sampling spots [9].

The information in the chart shown the phylogeny isolates (QFR04615, QFR04588, QFR04606, QFR04605, QFR04599) and (QFR04591) found in the Kainji Lake National Park and Yankari Game Reserve respectively explained that the *Glossina tachinoides* vectors were geographically restrained as shown by the clustered in the timeline March, 2014. Kainji Lake National Park is characterized by thick riverine vegetation, gallery forest and game animals which formed suitable habitat for *Glossina tachinoides* vectors [15]. The coexistence of more than one vectors in the same spot could be attributed to variations in microclimatic conditions at the sampling spots [9]

The chart also displayed phylogeny isolates (ALG02550, AHG56298, AHG56299, AHG56300, ACJ02120, ACJ02119, ACJ02118, ACJ02117, ACJ02116), (ACM43506, AEX60759, CAF04259), (CBG92832), (CBG92831) and (AYK27201) found in Zaria, Vom, Tsua, Bida and Federe respectively which shown that the isolates from hosts such as cattle, sheep, etc were geographically spread across foci in timeline (December, 2003 – April, 2018), hence they were seen in different location in Nigeria due to seasonal variation of green pasture.

5. CONCLUSION

This study develop chart that can be used by field researchers to navigate (identify) the existing *Trypanosoma* foci in Nigeria and predict isolation source in vectors/hosts and parasites which can be found in diverse foci. The study generated a concise visual representation and trajectory of trypanosome isolates in Nigeria

along with their genetic metadata and geographic location on a phylogenetic tree which will serve as a reference for information on trypanosomiasis outbreaks in Nigeria, provide blue prints for future outbreak predictions and guide decision-making on strategic control programs.

6. LIMITATION IN THIS STUDY

The exact geographical locations were not known as the latitudes and longitudes of the foci were not parts of the information deposited in NCBI websites.

7. RECOMMENDATION

Geographical information system of isolation should be included among the information imputed in the NCBI website. https://microreact.org/project/4SM4tVyb4FHfAo6 PsWqKfZ/87dcfa8d

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Bouyer F, Seck MT, Dicko AH, Sall B, Lo M. Ex-ante Benefit-Cost Analysis of the Elimination of Glossina palpalis gambiensis population in the Niayes of Senegal. PLoS Negl. Trop. Diseases. 2012;8(8):e3112.
2. Odeniran PO, MacLeod ET, Ademola IO, Welburn SC. Molecular identification of bloodmeal sources and trypanosomes in Glossina spp., Tabanus spp. and Stomoxys spp. trapped from cattle farm settlements in south-west Nigeria. Med. Vet. Entomology. 2019;33(2):269-281.
3. Isaac C, Ciosi M, Hamilton A, Scullion MK, Dede P, Igbinosa BI, Nmorsi, OPG, Masiga DC, Michael R, Turner CRM. Molecular identification of different trypanosome species and subspecies in tsetse flies of Northern Nigeria Parasites and Vectors. 2016;9:301.
   Available: https://doi.org/10.1186/s13071-016-1585-3.
4. Weber JS, Ngomtcho SCH, Shaida SS, Chechet GD, Gbem TT, Nok JA, et al. Genetic diversity of trypanosome species in tsetse flies (Glossina spp.) in Nigeria. Parasit. Vectors. 2019;12:481.
5. Omonona AO, Abioye SA, Odeniran PO, Ademola IO. Distribution of tsetse flies and its Trypanosoma species infection in Old Oyo National Park, Nigeria; 2020. Available:http://dx.doi.org/10.4314/tzool.v18i1.5
6. Odeniran PO, Ademola IO, Jegede HO. A review of wildlife tourism and meta-analysis of parasitism in Africa's national parks and game reserves. Parasitol. Research. 2018;117(8):2359-2378.
7. Ajibade WA, Agbede SA. Tsetse fly species diversity in Kainji Lake National Park, Nigeria. Afr. J Agric. Resources. 2008;3:753-841.
8. Okoh KE, Ndams IS, Kogi E, Vajime CG. Catch composition of tsetse flies (Diptera: Glossinidae). Am. J Appl. Sciences. 2011;8(11):1067-1072.
9. Shaida SS, Weber JS, Gbem JJ, Ngomtcho SH, Musa UB, Achukwi MD, et al. Diversity and phylogenetic relationships of Glossina populations in Nigeria and the Cameroonian border region. BMC Microbiology. 2018;18(Suppl. 1): 180.
10. Wama BE, Naphtali RS, Houmsou RS, Joseph J, Alo EB. Study of distribution and abundance of tsetse fly (Glossina sp.) in Gashaka-Gumti National Park, Nigeria. Int. J Develpmt. Research. 2018;8(1):18459-18464.
11. Argimón S, Abudahab K, Goater R, Fedosejev A, Bhai J, Glasner C, Feil E, Holden M, Yeats C, Grundmann H, Spratt B, Aanensen D. 30/11/2016. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb. Genom. 2016;10199.2(11): DOI:10.1099/mgen.0.000093
12. James Hadfield, Nicholas J Croucher, Richard J Goater, Khalil Abudahab, David M Aanensen, Simon R Harris, 2018; Phandango: an interactive viewer for bacterial population genomics Bioinformatics. 2018;34(2):15,292–293, Available:https://doi.org/10.1093/bioinformatics/btx610 Published: 25 September 2017
13. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." Systematic Biology. 2010;59(3):307-21.
14. National Parks of Nigeria (NPN). A publication of National Park Service

Abuja. Nigeria: Haligraph printers Minna; 2004.

15. Stephen Saikiu Shaida, Judith SophieWeber, Thaddeus Terlumun Gbem, Sen Claudine Henriette Ngomtcho, Usman Baba Musa1, Mbunkha Daniel Achukwi, Mohammed Mamman, Iliya Shehu Ndams, Jonathan Andrew Nok and Soerge Kelm. Diversity and phylogenetic relationships of Glossina populations in Nigeria and the Cameroonian border region. BMC Microbiology. 2018;18(Suppl 1):180 Available: https://doi.org/10.1186/s12866-018-1293-6

_____