MYCOBACTERIUM ULCERANS ACQUISITION BY AQUATIC HETEROPTERANS IS DRIVEN BY COMPLEX INTERPLAYS BETWEEN BIOGEOGRAPHICAL AND ECOLOGICAL CHARACTERISTICS

Solange Meyin à Ebong[1,2*], François Chevenet[1], Guillaume Castel[3], Philippe Le Gall[4], Sara Eyangoh[5], Jean-François Guégan[1,6]

[1]MIVEGEC, Université de Montpellier, CNRS, IRD, France

[2]Laboratoire de Microbiologie, Faculté des Sciences, Université de Yaoundé I, BP 812 Yaoundé, Cameroon

[3]CBGP, INRAE, CIRAD, IRD, Montpellier SupAgro, Univ Montpellier, Montpellier, France

[4]IRD UR 072, Laboratoire Evolution, Génomes et Spéciation, UPR 9034, CNRS, 91198 Gif sur Yvette Cedex, France et Université Paris-Sud 11, 91405 Orsay Cedex, France

[5]Service de Mycobactériologie, Centre Pasteur du Cameroun, BP 1274 Yaoundé, Cameroon

[6]ASTRE, INRAE, CIRAD, Université Montpellier, Montpellier, France

*Corresponding author

Telephone number: 677926583/656243314

E-mail adress: ebongsolange@yahoo.fr (MES)

Running title: *M. ulcerans* acquisition by aquatic Heteropterans

Acquisition de *Mycobacterium ulcerans* par les hétéroptères aquatiques facilitée par des interactions complexes entre les caractéristiques biogéographiques et écologiques

Résumé

Trois infra-ordres du sous-ordre des Hétéroptères ont développé des traits adaptatifs correspondant aux traits écologiques de la vie aquatique. Cette étude visait à montrer si l'acquisition de *Mycobacterium ulcerans* (MU) par ces insectes est facilitée par des interactions entre leurs caractéristiques biogéographiques et écologiques. À partir de 171 séquences du gène mitochondrial codant pour la première sous-unité de la cytochrome oxydase, nous avons reconstitué: 1) les relations phylogénétiques, entre les Hétéroptères aquatiques collectés au Cameroun en utilisant les approches Bayésiennes et de Maximum de Vraisemblance; 2) les caractères ancestraux de ces taxons selon les zones agroécologiques (biomes) du Cameroun où ils ont été récoltés, et ce en utilisant les méthodes bayésienne (BEAST) et de parcimonie (phylotype); et 3) nous avons utilisé les combinaisons de trois traits écologiques (type d'habitats, comportement alimentaire et capacité de vol) pour étudier le rôle de la biogéographie et des caractéristiques écologiques de ces taxons pour expliquer la présence de MU dans certaines lignées spécifiques d'hétéroptères aquatiques. L'étude révèle que: i) Nepomorpha apparaît comme un groupe monophylétique tandis que Gerromorpha considéré ici comme ancêtre de Nepomorpha constitue un groupe paraphylétique; ii) le biome 2 (zone 2) peut avoir facilité l'essor puis la dispersion des organismes aquatiques, y compris les Hétéroptères aquatiques, dans toute la région; et iii) cinq combinaisons d'annotations ancestrales montrent les signatures de MU, dans les lignées de taxons appartenant aux Naucoridae et Belostomatidae. En conclusion, la présence de MU chez ces familles serait facilitée par des traits géographiques et écologiques, et que ces taxons pourraient agir comme de potentiels hôtes contribuant à la dispersion à grande distance du bacille.

Mots clés: Hétéroptères aquatiques, traits ecologiques, ulcer de Buruli, *Mycobacterium ulcerans*, transmission.

Abstract

Three infraorders of the Heteroptera sub-order have developed adaptive traits which correspond to the ecological characteristics for living in aquatic environment. This study aimed to show whether *Mycobacterium ulcerans* (MU) acquisition by these insects is driven by complex interplays between their biogeographical and ecological characteristics. Based on 171 sequences of mitochondrial genes coding for the first subunit of cytochrome oxidase, we reconstructed: 1) the phylogenetic relationships, among aquatic Heteroptera collected in Cameroon using Bayesian and Maximum Likelihood approaches; 2) ancestral characters of these taxa according

to the agroecological zones (biomes) of Cameroun where they have been collected using Bayesian (BEAST) and parsimony (Phylotype) methods; and 3) we used the combinations of three ecological traits (type of habitats, feeding habits and flying capacity) to study the role played by biogeography and ecological characteristics of these taxa in explaining the presence of MU in some specific lineages of aquatic Heteroptera. Study reveals that: i) Nepomorpha appears as a monophyletic group while Gerromorpha considered here as the ancestors of Nepomorpha constitutes a paraphyletic group; ii) a biome with humid forest and monomodal rainfall (Zone 2) may have facilitated the dispersion of aquatic organisms including aquatic Heteroptera over all the Cameroonese region; and iii) five combinations of ancestral annotations show the MU signal, in the lineages of taxa belonging to the Naucoridae and Belostomatidae families. In conclusion, the presence of MU in Naucoridae and Belostomatidae is facilitated by geographic and ecological traits, and that these taxa could act as host vessels contributing to the large range dispersion of the bacillus.

Keys words: Aquatic Heteroptera, ecological traits, Buruli ulcer, *Mycobacterium ulcerans*, transmission.

Introduction

Buruli ulcer (BU), caused by *Mycobacterium ulcerans* (MU), is a neglected tropical disease associated with freshwater habitats (Garchitorena 2014). Aquatic bugs (Hemiptera: Heteroptera) have been hypothesized to be epidemiologically important host vectors for the transmission to human (Marsollier, Robert et al. 2002, Marion, Eyangoh et al. 2010, Garchitorena 2014). Heteroptera or true bugs, are divided into seven infraorders and three of them (Nepomorpha, Gerromorpha and Leptopodomorpha) are found in aquatic and sub-aquatic environments in all continents except Antarctica (Polhemus and Polhemus 2007). Even if the taxonomy and systematic of Heteroptera were subject of several works during the last two centuries, it is only in 1978 that Cobben has revised the scheme of the evolutionary history of this group of arthropods (Cobben 1978). Schuh presented the first documented higher-level phylogeny for the seven infraorders of Heteroptera chiefly based upon morphological characters (Schuh 1979). Other works carried out after analyzed more specifically the phylogeny of Gerromorpha and Nepomorpha on the basis of the morphology and structure of internal organs mainly in Australia and Malaisia (Schuh 1979, Andersen 1982, Andersen 2004, Chen 2005). Nowadays, molecular markers are increasingly used to study the evolutionary history of living organisms and notably insects among others. Concerning bugs, Damgaard has studied the phylogeny of many groups in Gerromorpha infraorders using molecular data set from nuclear and mitochondrial genes (Damgaard and Sperling 2001, Damgaard 2006, Damgaard 2008, Damgaard 2012). In the Afro-tropical region, these insects were ignored of the entomological studies for decades except the old works of Poisson on aquatic Heteroptera of Afro-tropical and Malagasy regions, (Poisson 1929, Poisson 1937, Poisson 1940, Poisson 1941, Poisson 1945, Poisson 1948). But since the 2000's, a resurging interest has grown documenting the species of aquatic Heteroptera inhabiting Cameroon because some species were suspected to be implicated in the ecology and transmission of MU (Portaels, Elsen et al. 1999, Marsollier, Robert et al. 2002, Marion, Eyangoh et al. 2010, Meyin A Ebong, Eyangoh et al. 2012, Esemu 2019). Recent work has studied the molecular taxonomy of aquatic Heteroptera from Cameroon (Meyin A Ebong, Petit et al. 2016, Esemu 2019).

In this study, and based on previous work (Meyin A Ebong, Petit et al. 2016) we used firstly two approaches which were the Bayesian (BEAST) (Drummond(a) 2012) and the maximum likelihood (PhyML) (Guindon 2010) methods to reconstruct the phylogenetic relationships among aquatic Heteroptera taxa collected in Cameroon, Central Africa. Secondly, we used Bayesian method (BEAST) (Drummond(a) 2012) and parsimony (PhyloType) (Chevenet 2013) to perform and study the reconstruction of ancestral characters of aquatic Heteroptera taxa according to the agro-ecological zones (biomes) of Cameroon where they have been collected, and the combinations of several ecological traits under study. Using comparative analyses, we then examined the role played by biogeography and ecological characteristics of the different Heteropteran taxa in explaining the presence of MU in some specific lineages of aquatic Heteroptera. Because two aquatic Heteroptera groups, i.e., Belostomatidae and Naucoridae, are suspected to be involved in the transmission of this bacillus to human and then caused the disease (Marsollier, Robert et al. 2002, Marion, Eyangoh et al. 2010), the present work is developing an innovative comparative approach to analyze the respective influence of phylogenetic relatedness, biogeography and ecological conditions in the contamination of the disease agent by these aquatic insects. These new findings will shed light on the important role played by both biogeography and ecological traits in the acquisition of the mycobacterium from the aquatic environment by these insects, thus supporting the

view that these aquatic insects may accidentally participate to the disease agent life-cycle as opportunistic hosts and participate in its wider spatial dissemination.

Material and methods

Sampling of aquatic Heteroptera across Cameroon, central Africa

All aquatic Heteroptera specimens used in this study were collected in different agroecological zones of Cameroon (Fig 1). Cameroon, a central African country, has a great diversity of soil conditions, climates and therefore a diversity of ecosystems that contain as rich as diversified animals and plants. On the agricultural plan, the diversity of landscapes, soil and climate conditions allow to subdivide Cameroon into 5 main agro-ecological zones known as biomes in this study (IRAD 2008). Aquatic Heteroptera were collected in the five biomes using two sampling methods: directly in the aquatic environment with hauling a deep net (32 × 32 cm and 1 mm in mesh size), and indirectly by using light trapping to capture winged imagoes (Meyin A Ebong, Petit et al. 2016). After collection, adults and nymphs were selected, counted, and preserved in 70% alcohol for further analysis (Meyin A Ebong, Petit et al. 2016).
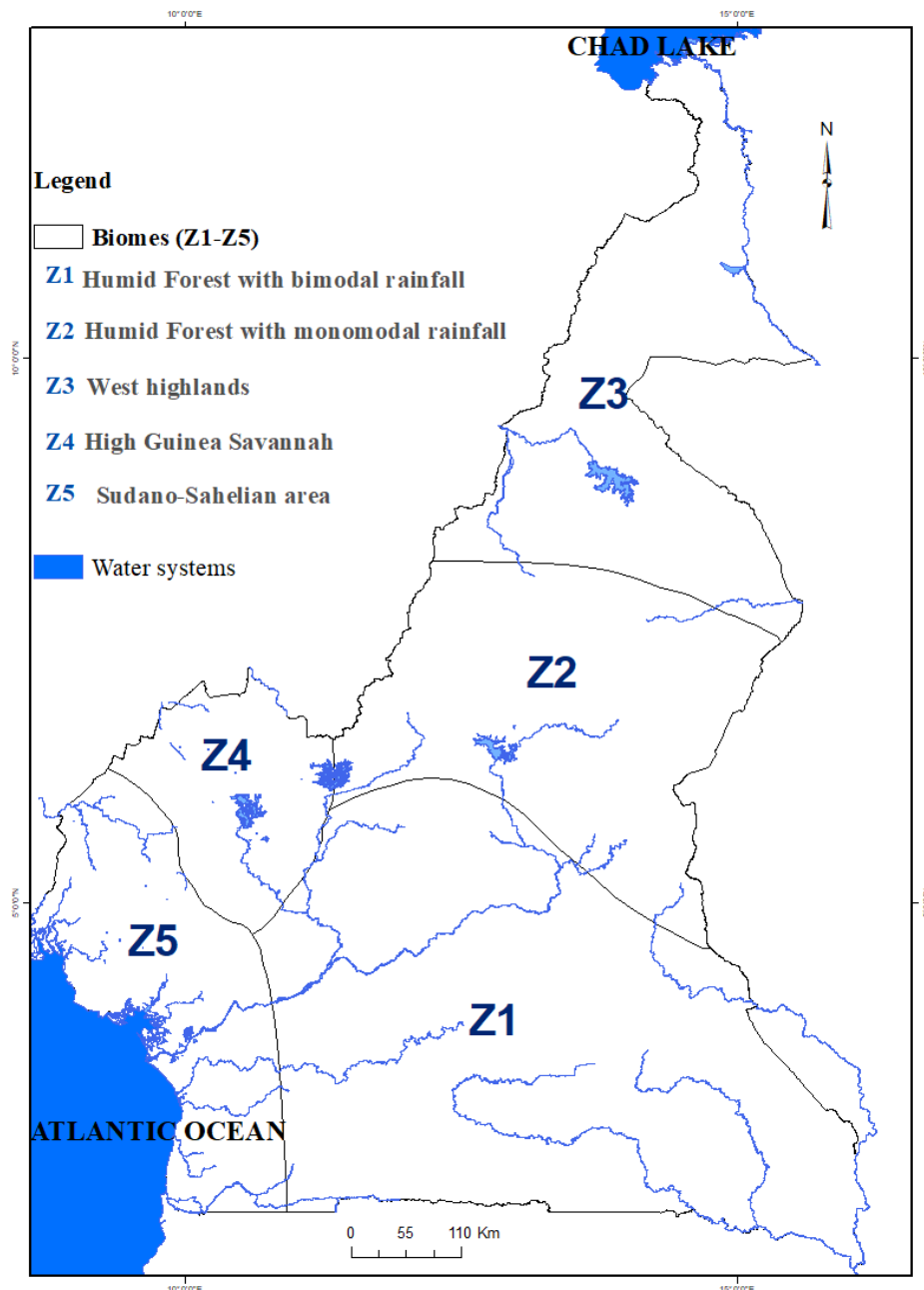

Fig 1. Map of Cameroon showing the different agro-ecological zones (biomes) according to the different sampling sites surveyed within the country

Phylogeny of aquatic Heteroptera

We used a fragment of the mitochondrial gene coding to the first subunit of cytochrome oxidase (COI). Total DNA was extracted from legs or full body for small insects, as previously described (Meyin A Ebong, Petit et al. 2016). PCR products were cleaned by Exosap-IT (Affimetrix® 2012), a single-step enzymatic clean up that eliminates unincorporated primers and dNTPs. The cleaned products were then sequenced to 690 bp nucleotides. Multiple alignments were made using Clustal W according to the default settings: full multiple alignments with bootstrap number equal to 1000. Finally, we obtained 171 COI homologous sequences data set of 669bp. We corroborated the (99%) homology of our sequences with COI sequences obtained from BLAST-GenBank (Genbank® 2013), and added to our data set three additional sequences external to Heteroptera to enable us to root the phylogenetic tree: *Cicadella* sp., *Graphocephala atropunctata* and *Graphocephala cytura*. References to these three genes in GenBank are respectively: gi262073319, gi227937213 and gi227937233.

We then reconstructed the phylogenetic relationships between aquatic Heteroptera taxa identified in this study using the 171 sequences data set described above, added to the three outgroup sequences from Genbank.

Phylogenetic tree calculation was firstly implemented with PhyML 3.0 (Guindon 2010) using the maximum likelihood method and default settings. The best substitution model of nucleotide sequences for these sequences data set was identified as GTR+I+G model (General Time Reversible + Invariant + Gamma) by using JModelTest (Posada 2008) and Akaike's criterion test. The reliability for internal branch was assessed using the aLRT test (approximate Likelihood Ratio Test) (Anisimova 2006). Phylogenetics trees were visualized using Treedyn (Chevenet 2006).

Secondly, phylogenetic tree was calculated using the Beast v1.8.0 program with Bayesian MCMC method (Drummond(a) 2012). BEAUTi v1.8.0 was used to define parameter settings. As with PhyML, a GTR+G+I substitution model was used. A lognormal relaxed clock allowing branch lengths to vary according to an uncorrelated lognormal distribution, a coalescent Bayesian skyline tree prior and a random tree as starting tree were used for this analysis (Drummond(b) 2012). All other priors were left to the default settings. We ran two independent analyses of 50 million generations with parameters sampled every 1000 generations. Parameters were evaluated using Tracer v1.5 and summary maximum clade credibility trees (MCC) were generated using TREE ANNOTATOR v1.8.0 after discarding the first 10% of the trees as burn-in as determined graphically using Tracer v1.5. Phylogenetic trees were visualized using FigTree v1.4.2.

Comparative analyses of ancestral annotations (geographical and ecological traits of aquatic Heteroptera) using Parsimony (Phylotype) and Bayesian (BEAST) methods

The study of the distribution of species according to their biogeographic criteria is part of the concept of phylogeography which describes principles and processes that govern the distribution of animal or plant lineages, especially at intra-specific level (Kidd and Ritchie 2006). The biogeographical character biomes (zones 1 to 5) were used to analyze the phylogeography of aquatic Heteroptera in Cameroon. Similarly, several ecological traits were analyzed in combination, thus forming what we will call next "primary annotations". These ecological traits are as follows: habitat characteristics: living on the aquatic vegetation (coded v), living near the bottom of the water column (coded f), living near or at the water surface (coded s), living near the river edge (coded r); feeding habits: macropredator that feeds on large preys or macropreys (fish, freshwater mollusc…) (coded M), micropredator that feeds on other aquatic insect larvae (coded m), phytozoophagous that (coded p) and scavenger/detritivorous that feed upon weakened or dead prey (coded c); flying capacity: able to fly (coded 1) or not (coded 0); and finally, the characterization of MU cells in that aquatic Heteroptera (coded a) or not (coded b). These primary annotations considered independently or in combination were used to infer ancestral traits reconstruction using BEAST and PhyloType programs and to perform comparative analyses of ancestral biogeographical and ecological traits for aquatic Heteroptera taxa.

A phylogenetic tree analysis with ancestral biomes was made in order to establish a phylogeographic scenario following which we build the primary annotations crossing the geographic biomes and the combinations of life history traits e.g., zone4×living near the bottom of the water column (f)×feeding on larger preys (M)×Flying capacity (1), or zone4×f×M×1; we simply used z4fM1 in following Figures.

The computation of ancestral traits with PhyloType is based on parsimony (Fitch, ACCTRAN and DELTRAN algorithms and variants). The PhyloType method is then based on the definition of the phylotype, as the most recent common ancestor of a set of taxa annotated in the same way along the evolutionary path, from the root of the phylotype to its underlying sequences. Relevant phylotypes are identified by applying combinatorial and statistical criteria to the tree with its ancestral traits (see below for details). Significant phylotypes are identified on the basis of their statistical support calculated by iteration of the

analysis after a randomization of the annotations associated with the sequences. This process of "shuffling" is repeated several thousands times and makes it possible to obtain a *p*-value for the phylotypes. The algorithmic complexity is in linear time, thus allowing multiple analyses of large data sets.

BEAST uses Bayesian method to calculate and display the probability of all ancestral annotations presented at each node. The spatial information and combinations were considered as discrete characters and their ancestral states were reconstructed by fitting a standard continuous-time Markov chain (CTMC) model. We used a Bayesian stochastic search variable selection (BSSVS) procedure (Lemey, Rambaut et al. 2009) allowing transition rates to be zero with some probability in order to achieve statistical efficiency and identify significant disease transmission linkage among all possible water bug migration pathways. For the "biomes" annotation, both a symmetric diffusion and an asymmetric diffusion models were tested. The asymmetric diffusion model was identified as better fitting our data by using the Model comparison function (which compares models AIC scores) implemented in the Tracer v1.6 software. Only the symmetric diffusion model was used for the combination.

Analysis was conducted using a Bayesian Skyline tree prior and the prior location clock rate was set as default (CTMC Rate Reference) since we do not have adequate prior information (Ferreira 2008). For each reconstruction we ran two independent analyses of 50 million generations with parameters sampled every 1000 generations. MCC were generated as previously described in the text and visualized using FigTree v1.4.2.

Searching for aquatic Heteroptera phylotypes according to their biogeographical and ecological traits and *M. ulcerans* presence

The phylotype analyses require a rooted tree of the studied Heteroptera taxa and primary annotations combining biogeographical and ecological traits. To compute the ancestral annotations the DELTRAN method was selected. To identify relevant phylotypes the following combinatorial and numerical criteria were used: Size (Sz=3), Size/Difference ratio (Sz/Df =1), Persistence (Ps=1), branches support (Sp $\geq$ 0.75). Significant phylotypes were identified with a shuffling process of 1000 iterations.

Visualization and analysis of the dissemination pathways

SPREAD v1.0.6 software (Bielejec 2011) was used to map annotated phylogenies and to export high-dimensional posterior summaries to keyhole markup language (KML) for visualization of the spatial diffusion of aquatic Heteroptera taxa

through time in Google Earth software. The MCC annotated tree calculated by BEAST was used as input and longitude and latitude data for each biome were chosen and included in the analyses as geographical coordinates.

Bayes factor calculation implemented in SPREAD was used to identify well-supported migration links between the different biomes based on the Phylogeographic Bayesian stochastic search variable selection procedure (BSSVS) estimates implemented in BEAST. Since we used an asymmetric model, we can infer directionality for a given route.

Results

Reconstruction of relationships between aquatic Heteroptera taxa using BEAST (Bayesian inference) and PhyML (Maximum Likelihood) models

Data set of 171 COI sequences from the aquatic Heteroptera taxa were used to reconstruct the ancestral relationships among these. In general, the two trees (PhyML/ BEAST) showed almost similar topology with a perfect grouping of taxa within the respective families, although significant differences were observed especially on the older nodes (Fig 2). Gerromorpha are positioned at the root of Nepomorpha. Both Mesoveliidae and Hydrometridae families are appearing as common old ancestors of all taxa for BEAST and for PhyML, and only Hydrometridae family is found as common old ancestor for all Heteroptera taxa. In this last case, Mesoveliidae is being the immediate descendant of Hydrometridae. There is one common ancestor for all Nepomorpha which forms monophyletic group in both cases unlike Gerromorpha which have 2 and 3 clusters with 2 and 3 different common ancestors respectively for BEAST and PhyML, and therefore forms a paraphyletic group. Within some families there are some differences in taxa grouping depending on the method used. The BEAST tree shows 8 monophyletic families with branches supported over 75% against 7 families for PhyML with branches supported over 75% as well. The Naucoridae and Veliidae are paraphyletic in the two trees, one Naucoridae cluster shares the same ancestor with a cluster of Notonectidae (monophyletic for PhyML and paraphyletic for BEAST) in both trees and the other cluster shares a common ancestor with Corixidae in the BEAST tree while the same cluster appears as direct ancestor of Corixidae in PhyML tree. Belostomatidae and Nepidae have a common ancestor in the two trees but in the BEAST tree, each of these families forms a monophyletic group and in the PhyML tree, each of them is a paraphyletic group. In this latter case, the clusters

of Nepidae are placed within clusters of Belostomatidae. One cluster of Veliidae shares a common origin with Hebridae in both trees and the other cluster appears just before and it is positioned at the root of the first cluster of Veliidae, Hebridae and Gerridae in the BEAST tree while the same cluster is not the root of Gerridae in PhyML tree.
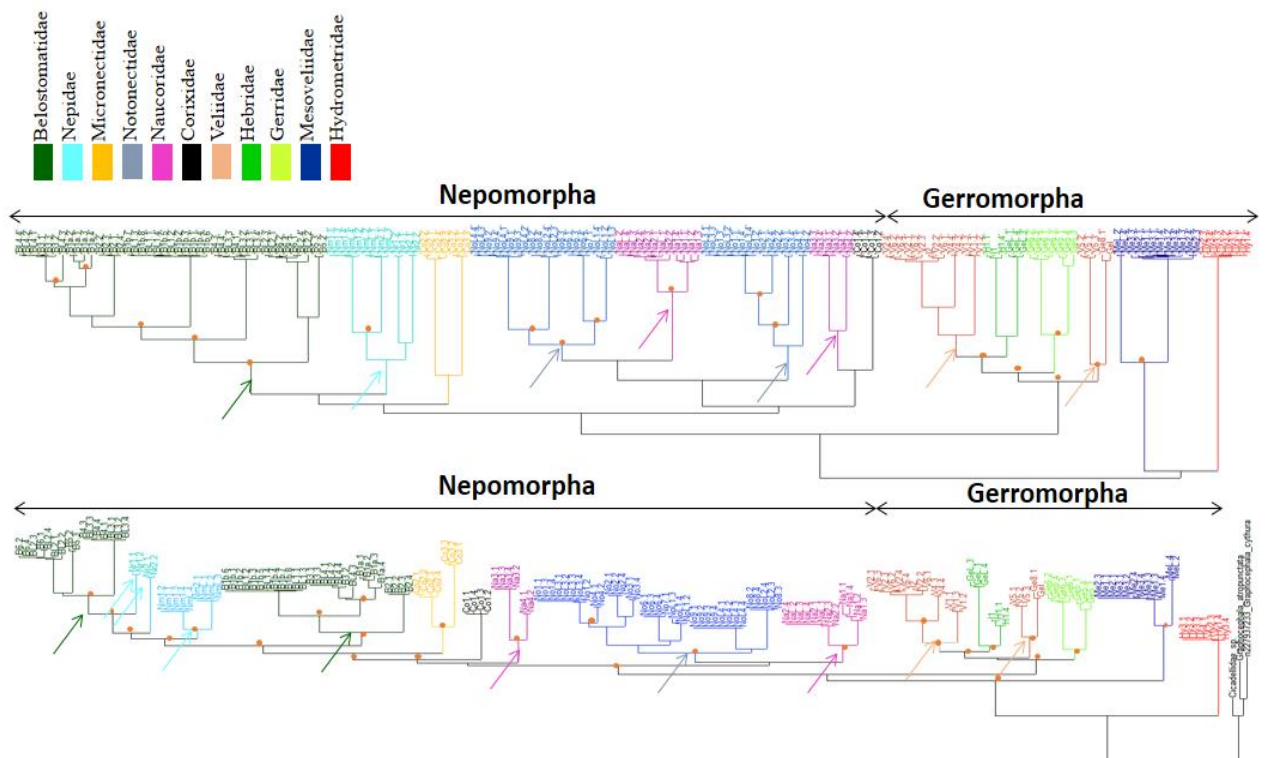

Fig 2. Setting facing phylogenetic trees based on the PhyML (left), and BEAST (right)

Trees were annotated according to taxonomic information (sub-trees color corresponding to families), the branches support values ≥ 75% (red points). Correspondence between Heteroptera taxa groups for one tree to another is represented by red arrows.

Aquatic Heteroptera of Cameroon: Phylogeography and comparative analysis of ancestral annotation using both Bayesian (BEAST) and Parsimony (PhyloType) methods

The Bayesian inference and parsimony approaches used in this study allowed tracing the historical scenario of aquatic Heteroptera taxa spread using the "biomes" factor (agro-ecological zones) of Cameroon where they were collected. Overall, the distribution of ancestral annotations is almost identical in both approaches and the ancestral annotation (zone2), corresponding to the high Guinea savannah zone of Cameroon, appears to be the origin center of spread of these Heteropterans within this region. This ancestral annotation persists in most of the internal nodes and even to the terminal taxa of some aquatic Heteroptera families (Fig 3). The taxa belonging to the 11 families identified in this work are represented in the biome with humid forest and monomodal rainfall (zone2) except for Hebridae in both methods and for Hydrometridae using the BEAST method (Table 1). These two approaches show in the final results certain annotations that are absent in one or the other method. This is the case of the annotations zone1 which is present in Mesoveliidae taxa (80% probability), zone3 present in Belostomatidae (90% probability) and Nepidae (80% probability) with Bayesian method, and which are absent using the parsimony method. Similarly the annotation zone2 present in Hydrometridae taxa for parsimony method was not retained for Bayesian method; in this case, the annotation "zone1" (68% probability) was retained at the expense of zone2 (only 29.5% probability) (Fig3).

Table 1: Ancestral annotations for biomes and their presence (apparition) in aquatic Heteroptera taxa and related families

| Families	Beast ancestral Annotations	Phylotype ancestral Annotations
Mesoveliidae	**zone1,** zone2, zone3	zone2, zone3
Gerridae	zone1, zone2	zone1, zone2
Veliidae	zone1, zone2, zone3, zone4	zone1, zone2, zone3, zone4
Hebridae	zone5	zone5
Naucoridae	zone2, zone5	zone2, zone5
Notonectidae	zone1, zone2, zone5	zone1, zone2, zone5
Micronectidae	zone2, zone5	zone2, zone5
Nepidae	zone2, **zone3,** zone5	zone2, zone5
Belostomatidae	zone1, zone2, **zone3**	zone1, zone2
Corixidae	zone2	zone2
Hydrometridae	zone1, zone4	zone1, **zone2,** zone4
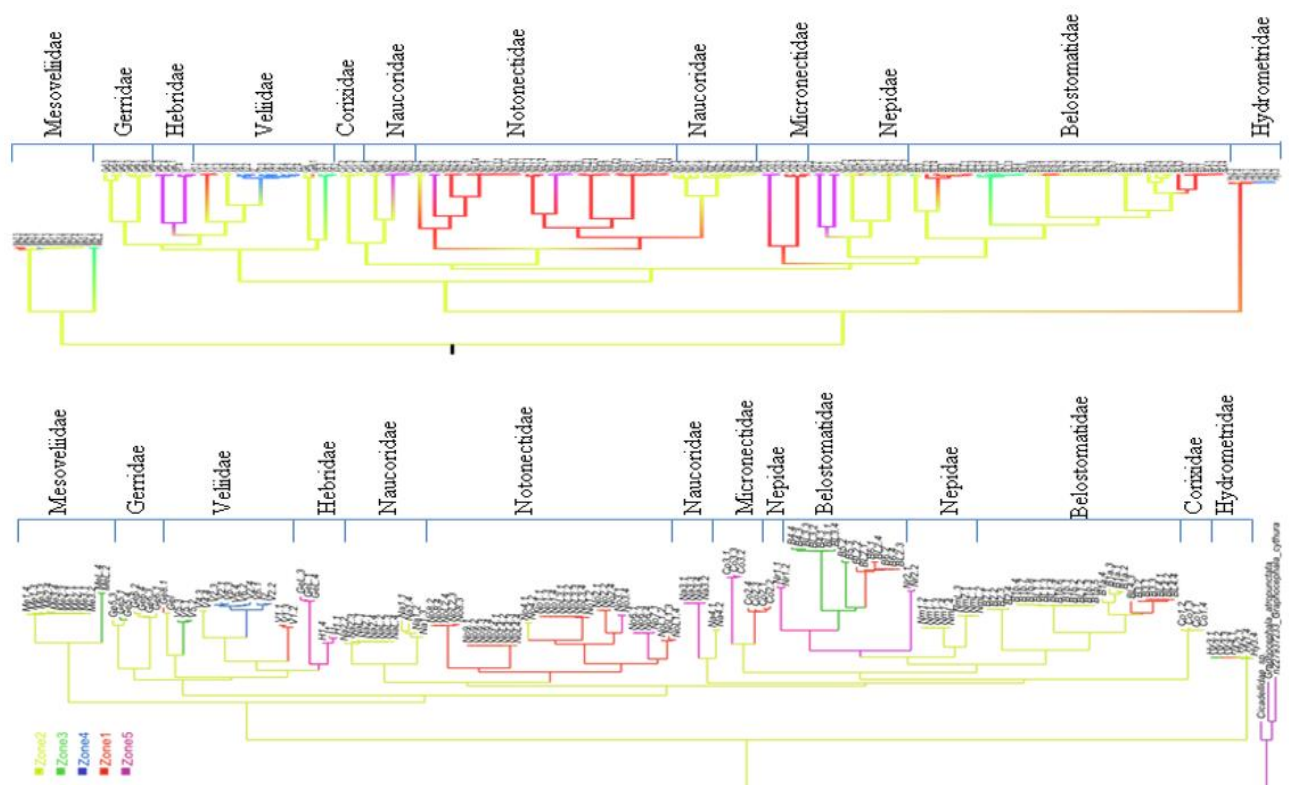
Fig 3. Setting facing ancestral annotation biomes reconstructed with PhyloType (left) and BEAST (right)
The 5 different geographical biomes are illustrated with different colors on the phylogenetic trees

Mapping of the dissemination pathways of aquatic Heteroptera across the different biomes

The asymmetrical model used for this analysis considers different transition rates for each transition. For each transition a Bayesian Factor (BF) is calculated to determine whether this transition is significant or not. Following significant BF were observed between the different transition states: Zone2 to Zone3 (BF=146765.708), Zone4 to Zone5 (BF= 1394.538), Zone2 to Zone1 (BF=908.346), Zone2 to Zone4 (BF=15.397), Zone1 to Zone2 (BF=14.643), Zone1 to Zone5 (BF=10.872). Biome zone2 was connected with all others biomes except biome zone5 which is connected with zone4 and zone1. There is also a double sense transition between zone1 and zone2 (Fig 4).
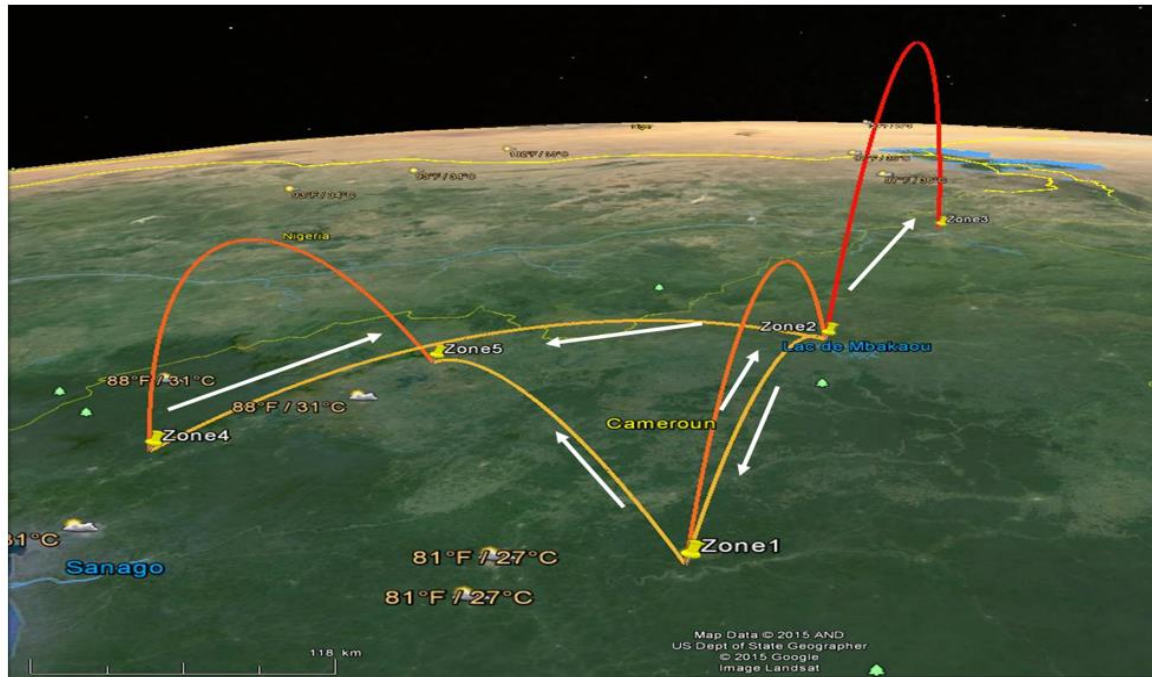

Fig 4. Aquatic bugs spread across the different biomes of Cameroon according to the asymmetrical transition model (see text).

The arrows show the direction of transition from one zone to another

Study of aquatic Heteroptera phylotypes based on agro-ecological biomes of Cameroon

The analysis based on the different biomes allowed to obtain 4 phylotypes including 99 sequences (57 % of all the molecular data set). The phylotype zone2 is the origin from which emerge directly phylotypes zone1 and zone4; the phylotype zone3 noted in red color seems to emerge indirectly from zone2 (Fig 5A). Indeed, there has been a break between annotations zone2 and zone3 including Zone5 that appears on the legend and having the same number as the phylotype zone3 (node 219 on Fig 5B) but there are no aquatic Heteroptera sequences associated at this ancestral annotation. This break is due to the presence of two ancestral annotations at this node (zone3 and zone5 ambiguity not resolved by ACTRANS) and annotation zone3 was retained as shown in the phylotype map (Fig 5B). These observations are confirmed by the BEAST results showing the best value (86.3%) for the ancestral annotation zone3 compared to the ancestral annotation zone5 (21.3%). Moreover, this node has the same topology in BEAST and PhyloType ancestral annotations trees (Fig 3). Fig 5C shows the potential dissemination pathways of aquatic Heteroptera across the different biomes found in Cameroon. Phylotype zone2 consists to 56 sequences (89% coverage) (S2 Table), and it includes taxa belonging to Corixidae, Belostomatidae, Nepidae, Naucoridae, Gerridae and Mesoveliidae families (Fig 5A). The phylotype zone1 is formed by 27 sequences (47% coverage) of taxa belonging to Notonectidae family while the phylotype zone4 is represented by 47 sequences (47% coverage) of taxa belonging to Veliidae family and zone3 formed by 8 sequences (53% coverage) of taxa belonging to Belostomatidae family (Fig 5A and S1 Table).
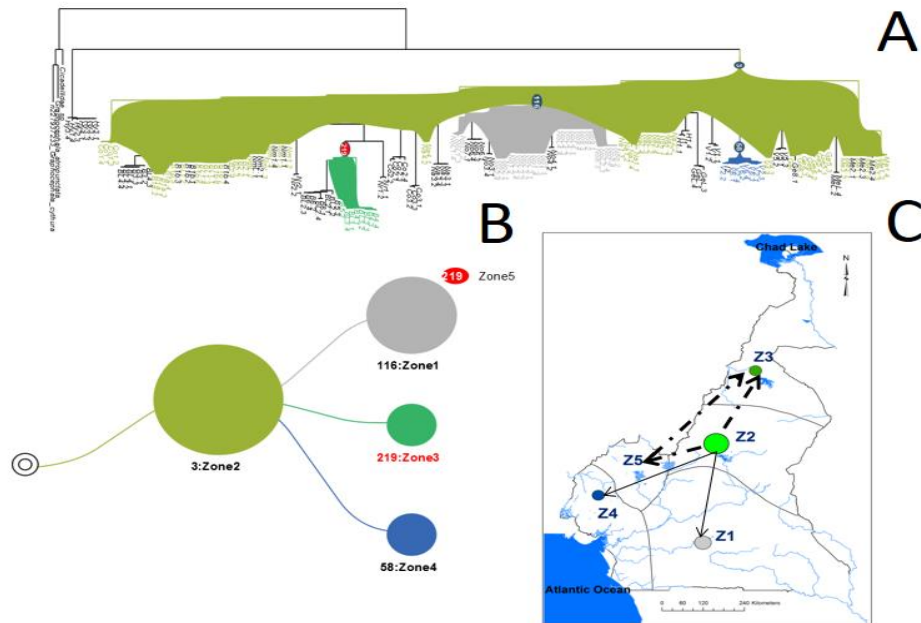
Fig 5. Distribution of aquatic Heteroptera phylotypes through the major biomes of Cameroon using the PhyloType approach: (A) Phylotypes distribution tree showing aquatic Heteroptera taxa involved; (B) phylotype distribution flow chart; (C) geographical distribution of the different Heteroptera phylotypes on the map of Cameroon

(Z_1 - Z_5) = major biomes of Cameroon, the arrows show the origin of phylotypes from Z2: full line = direct emergence, dashed = indirect emergence.

Analysis of aquatic Heteroptera ecological traits and search of *M. ulcerans* signature using BEAST (Bayesian) and Phylotype (Parsimony) methods

The analysis of the combination of aquatic Heteroptera ecological traits using Parsimony and Bayesian inference methods allowed obtaining respectively 20 and 30 combinations of ancestral annotations. Globally, these ancestral annotations are distributed in the same way in the phylogenetic trees calculated with both methods. This high number of ancestral annotation combinations makes the visualization and interpretation difficult. For this reason, we then selected only clusters showing a MU signature. The presence of MU signal (illustrated by the latter "a" in combination) initially absent in ancestral annotations appear in the lineages of some aquatic Heteroptera families and persists until the terminal taxa. Whatever the approach used (BEAST or PhyloType), the MU signature is detected in 5 following combinations of ancestral annotations: $z_1 \times v \times 1 \times M \times a$, $z_2 \times v \times 1 \times M \times a$, $z_3 \times v \times 1 \times M \times a$, $z_2 \times v \times 0 \times M \times a$ and $z_5 \times v \times 0 \times M \times a$ (Fig 6 and 7). These combinations include only taxa belonging to Naucoridae and Belostomatidae families (Fig 6 and 7). All biomes are represented in these combinations except z_4, the type of habitat present here is living on aquatic vegetation (v) and the retaining finding habit is macropredator (M) (Fig 6 and 7). We can note the presence of the flying capacity or not in these combinations of ancestral annotations. According to these results, the presence of MU signal in an ancestral annotation is closely associated with certain ecological traits of aquatic Heteroptera such as the type of habitats (living on aquatic vegetation) and feeding habits (being preferentially a macropredator).

Furthermore so that the MU signal is present in an ancestral annotation of aquatic Heteroptera taxa, it is necessary that the bacillus may contaminate aquatic Heteroptera specimens in an aquatic environment containing itself the mycobacterium. Therefore biomes Z_1, Z_2, Z_3 and Z_5 present in these ancestral annotations combinations would preferentially harbor MU in their aquatic ecosystems. On the contrary, our analysis shows that biome Z_4, i.e., High Guinea savannah, would nor harbour or favorize MU persistence in this type of aquatic environment. In contrast, aquatic Heteroptera flying capacity seems not to influence directly the presence of MU signal in ancestral annotations. However, aquatic Heteroptera taxa able of flying would potentially facilitate the dissemination of the mycobacterium in the environment by contaminating new free environments during their colonization and spreading the bacillus across other proximate biomes.
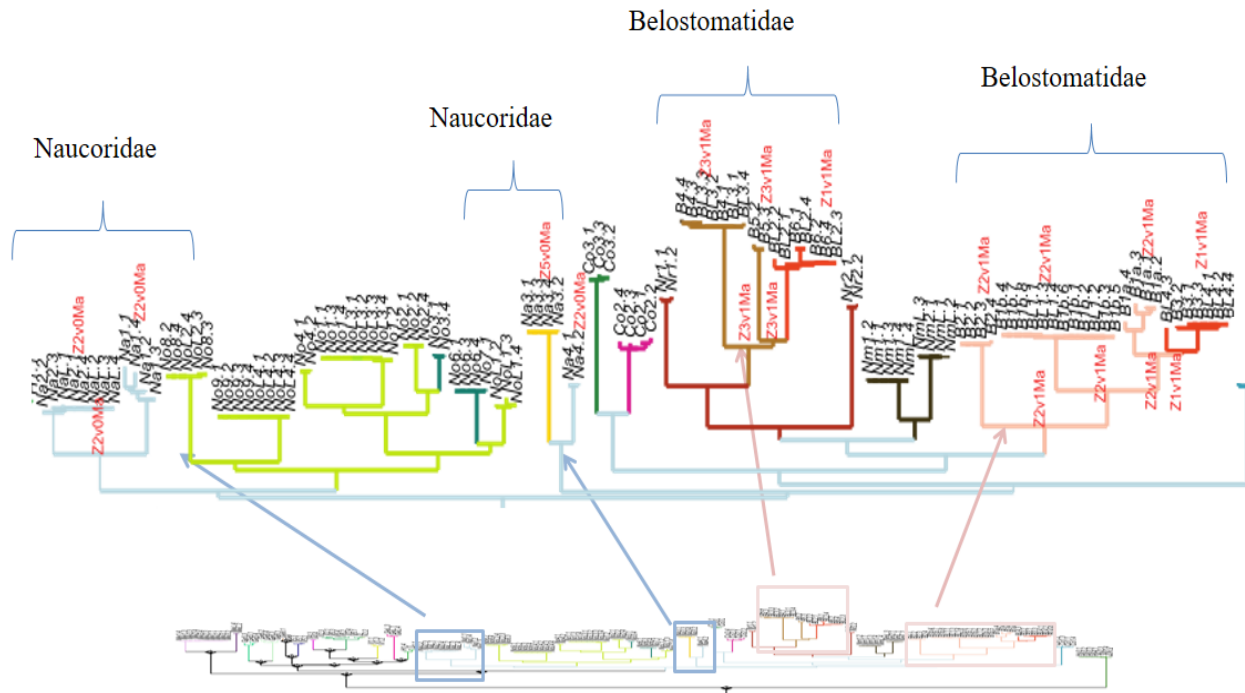
Fig 6. Zoom on clades of ancestral annotations with the *M. ulcerans* signal (Parsimony) "Z₁ to Z₅" = agro-ecological zones of Cameroon; "v" = type of habitat (living on aquatic vegetation); "1" corresponds to flight capacity versus "0" no flying capacity; "M" is macropredator; "a" is vector potential
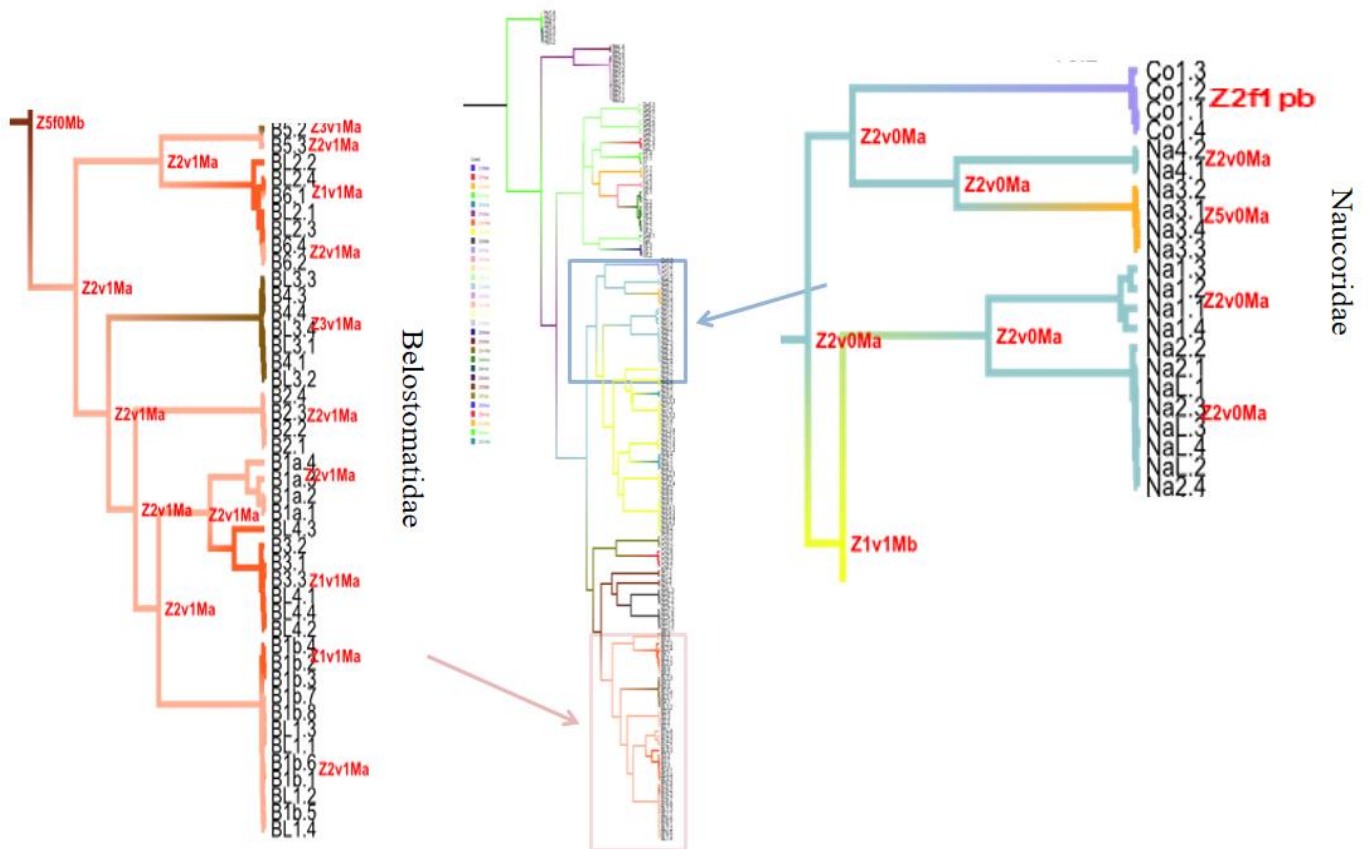

Fig 7. Zoom on clades of ancestrale annotations with the *M. ulcerans* signal (Bayesian inference) "Z₁ to Z₅" = agro-ecological zones of Cameroon; "v" = type of habitat (living on aquatic vegetation); "1" corresponds to flight capacity versus "0" no flying capacity; "M" is macropredator and "a" is potential vector

Searching of aquatic Heteroptera phylotypes based on combinations of their ecological traits and *M. ulcerans* signal

Aquatic Heteroptera ecological traits used in this study (biome, type of habitat, feeding habits, flight capacity or not, and potential host carrier or not) are combined and analyzed by the parsimony method to get 20 phylotypes associated to 19 ancestral annotation combinations including 135 sequences (77,6% of all molecular data set) (S3 table). Six phylotypes (91, 187, 270, 219, 312 and 237) represented respectively by the following ancestral annotation combinations z_1v1Ma, z_2v1Ma, z_3v1Ma, z_2v0Ma and z_5v0Ma contain the MU signature (phylotypes containing the red points in Fig 8B). The phylotype 91 consists of 14 sequences (100% coverage) of taxa belonging to the Naucoridae family (Fig 8A) that live on aquatic vegetation, which does not have a flying capacity,

feed on macro-prey and they were collected in the biome zone2. From this phylotype also emerge other phylotypes and 5 of them contain the MU signal (Fig 8B). The phylotype 187 consists of 4 sequences (100% coverage) of taxa also belonging to the Naucoridae family which were collected in the biome zone5. The phylotypes 270, 219, 237 and 312 consist respectively to 16, 4, 4 and 6 sequences (with respective coverage: 80%, 100%, 27% and 40%) of taxa belonging to the Belostomatidae family which are able to fly, live on aquatic vegetation, feed on macro-prey and were collected respectively in the biomes zone1, zone2, and zone3. There have been break to 219 annotation nodes due to the presence of two phylotypes annotated differently (z_3v1Ma and z_5f0Mb). In addition, the branches support value at this node is very low (0.75) relative to the other branches support we obtained.
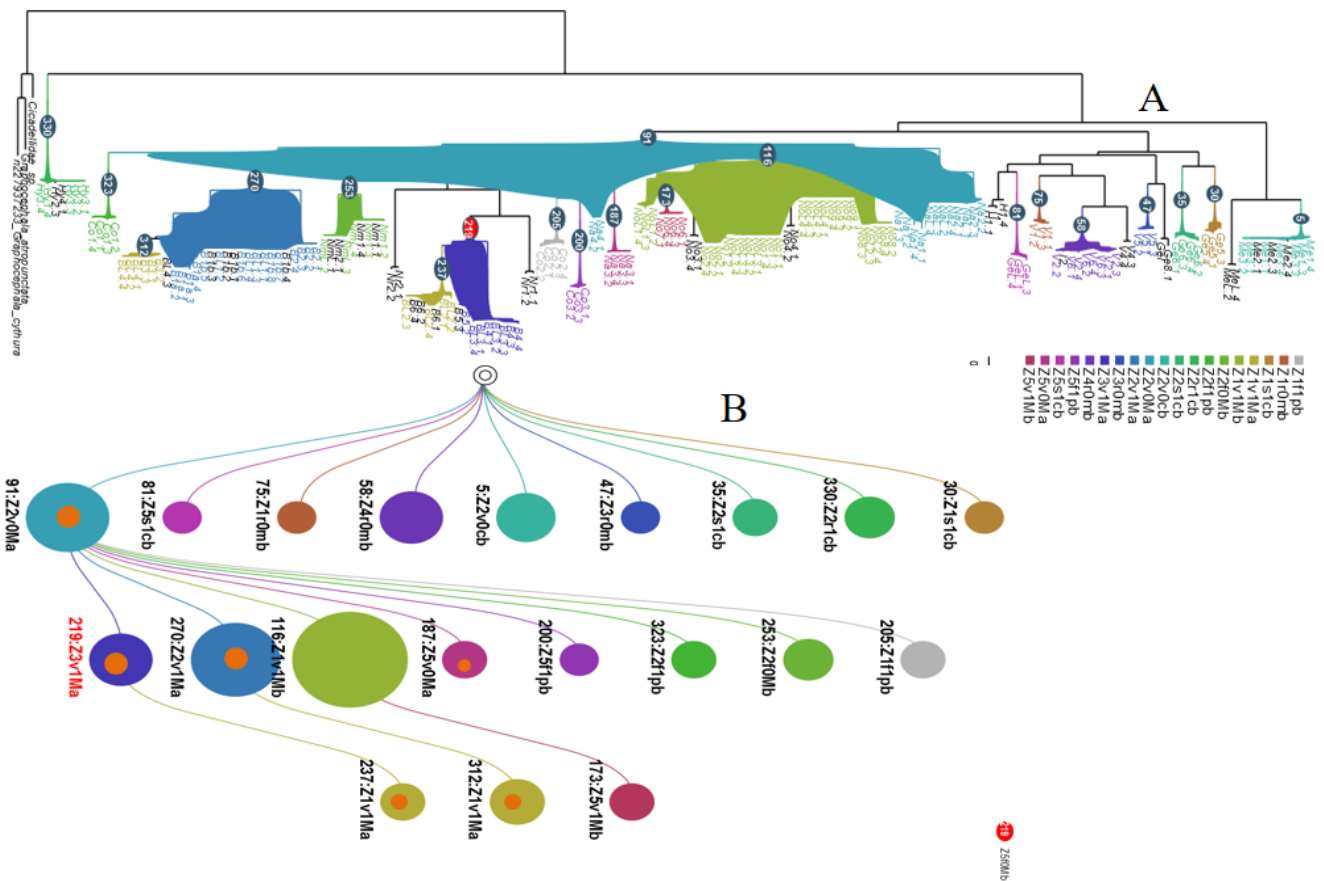

Fig 8. Aquatic Heteroptera phylotypes based on their biogeographical and ecological traits combinations and search of *M. ulcerans* signal (A) Phylotypes distribution tree showing aquatic bug taxa involved; (B) phylotype flow chart diagram; phylotypes with red points contain the *M. ulcerans* signal.

Biomes (z_1=zone1, z_2=zone2, z_3=zone3, z_4=zone4, z_5=zone5); type of habitat: living on (f= bottom of water column, r=river edge, v=aquatic vegetation, s=water surface), feeding habit (M=macropredator, m=micropredator, p= phytozoophagous, c=detrivorous), fligh capacity=1 or not=0, potential vector=a or not=b

Discussion

The comparative analysis of phylogenetic trees based on Maximum Likelihood (PhyML) and Bayesian (BEAST) inference methods from 171 sequences of COI shows the ancestral relationships among taxa of aquatic Heteroptera collected in different biomes of Cameroon, central Africa. Higher taxa (sub-orders and families) are distinguishable on both phylogenetic trees and sequences (taxa) are globally grouping within the respective families. All Nepomorpha included in this study have a common ancestor and form a monophyletic group in both cases unlike Gerromorpha which have 2 and 3 clusters with 2 and 3 different common ancestors respectively for BEAST and PhyML, and therefore form a paraphyletic group. In addition, Gerromorpha are positioned at the root of Nepomorpha on both phylogenetic trees. Despite the fact that the phylogenetic relationships remain highly discussed among Heteroptera, and particularly for aquatic taxa, our results corroborate with the revised scheme of the evolutionary history of Heteroptera presented by (Cobben 1978). This author describes Gerromorpha as a group at the basis of the lineages leading to other major taxa of modern Heteroptera and advocates the hypothesis of the Gerromorpha as a strongly paraphyletic group according to Ashlock (Ashlock 1974). Although the taxa are grouped in the respective families, significant differences observed here are due to the robustness of the trees calculation method and to the molecular marker used. Our study is performed on a portion of mitochondrial gene (COI) that seems not enough for phylogenetic analysis of aquatic Heteroptera. A robust phylogenetic study needs several nuclear and mitochondrial genes. For instance, to study the phylogeny of genus *Gerris* (Gerridae family), COI and a nuclear gene encoded elongate factor 1 alpha (EF-1α) has been used (Damgaard 2006).

Regardless of the Bayesian inference and parcimony approaches, the phylogeography of aquatic bugs of Cameroon is mainly represented by the ancestral annotation zone2, i.e., humid forest with monomodal rainfall that appears to be the origin center of aquatic Heteroptera spread in Cameroon. Despite the position of zone2 to root of ancestral annotations for higher taxa of aquatic bugs, (infra-orders and families), we have noted its wide geographical distribution at lower taxonomic levels. This is the case of taxa belonging to Veliidae, Notonectidae, Belostomatidae, Nepidae, Hydrometridae, Mesoveliidae and Micronectidae families which were present in three biomes on a total of five unlike the taxa of Corixidae and Hebridae families which were present in only in one biome respectively, i.e., Zone2 and Zone5 (Fig 2 and Table1). This geographical distribution of ancestral annotations of lower taxa of aquatic Heteroptera was illustrated by the phylotypes 58, 116 and 219 corresponding to ancestral annotations

zone1, zone3 and zone4 respectively emerging from phylotype 3 corresponding to ancestral annotation zone2; note however that phylotype 219 colored in red has indirect origin from this (Fig 5). According to Chevenet and collaborators' methodology (Chevenet 2013) the node 219 in our work was annotated differently by zone3 and zone5 but zone3 was retained by the shuffling using DELTRAN parsimony algorithm with 1000 iterations and P-v ≥ 0.05 and zone5 was displayed in the legend with the same phylotype number 219. Indeed, the hydrographic network of Cameroon forms a more or less communicating network of rivers and creeks whose starting point is the Zone2 containing the Adamawa region, which is the water tower that supplies the rivers of Cameroon thus facilitating the movement of aquatic animals including aquatic Heteroptera through these corridors and ripisylve between the different biomes. The starting point (zone2) appears as the origin center of spread of aquatic Heteroptera in Cameroon. These observations agree with the map obtained with Spatial Phylogenetic Reconstruction of Evolutionary Dynamic (SPREAD) showing higher number of connections between zone2 and all the other biomes (Fig 4).

In the study combining environmental and geographic traits of aquatic Heteroptera, both trees obtained show overall a similar topology of the distribution of ancestral annotation combinations. Five ancestral annotation combinations including taxa belonging to the Naucoridae and Belostomatidae families harbor the MU signal. The same combinations of ancestral annotations are found in phylotypes having the MU signal and also consisting of taxa belonging to the Belostomatidae and Naucoridae families (Fig 8). Belostomatidae and Naucoridae families have their preferred habitats on aquatic vegetation and they also share the same feeding habits as macropredators. These

two ecological traits seem to facilitate the contact between aquatic Heteroptera and mycobacteria or their biofilms which are present on aquatic vegetation (Marsollier, Robert et al. 2002, Plowright, Sokolow et al. 2008, Meyin A Ebong, García Peña et al. 2017). A study realized in southwest Cameroon (an emerging focus of Buruli ulcer) confirmed the absence of MU in Gerridae and Veliidae families (Esemu 2019).

Mycobacterium ulcerans growths and persists in aquatic environments having particular features like low oxygen and moderated *pH* conditions (Portaels and Pattyn 1982, Palomino, Obiang et al. 1998, Sanhueza, Chevillon et al. 2018). These particular environmental conditions exist in biomes Zone1, Zone2, Zone3 and Zone5 that are present in the ancestral annotation combinations having the MU signature (i.e., z_1v1Ma, z_2v1Ma, z_3v1Ma, z_2v0Ma and z_5v0Ma) and absent in the biome zone4 cooresponding to High Guinea Savanah that does

not exist in these combinations. Flying capacity does not seem to be directly involved in the contact between aquatic Heteroptera and MU but it would nevertheless facilitate the spread of this mycobacterium spatially by contaminating free-microbe environments through these waterbug dissemination and spread.

Conclusion

In conclusion, our findings tend to indicate that several biogeographical conditions favorizing the persistence of environmentally-free bacilli associated with some specific ecological traits in a couple of specific aquatic bug taxa, i.e., Belostomatidae and Naucoridae, might contribute to the acquisition of the bacillus from the aquatic environments by those arthropods in the central region of Africa. These aquatic bugs might then circumstantially and locally contribute to the disease agent transmission and spread in contaminating free aquatic environments from place to place, and act as host vessels better than true active vectors in human infections causing Buruli ulcer. Overall these results converge with pioneer investigation cited above.

28

References

Affimetrix® (2012). ""USB® ExoSAP-IT® Product Cleanup."." User manuel: 1-4.

Andersen, E. A. (1982). The semiaquatic bugs (Hemiptera, Gerromorpha): phylogeny, adaptations, biogeography and classification, Brill.

Andersen, N. M. W., Tom A (2004). Australian water bugs: their biology and identification (Hemiptera-Heteroptera, Gerromorpha & Nepomorpha), Apollo books.

Anisimova, M. G., Olivier (2006). "Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative." Systematic biology **55**(4): 539-552.

Ashlock, P. D. (1974). ""The uses of cladistics."." A. rev. Ecol. Syst **5**: 81-99.

Bielejec, F. R., Andrew Suchard, Marc A Lemey, Philippe (2011). "SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics." Bioinformatics **27**(20): 2910-2912.

Chen, P.-p. N., Nico Zettel, Herbert (2005). The aquatic and semi-aquatic bugs (Heteroptera: Nepomorpha & Gerromorpha) of Malesia, Brill.

Chevenet, F. B., Christine Bañuls, Anne-Laure Jacq, Bernard Christen, Richard (2006). "TreeDyn: towards dynamic graphics and annotations for analyses of trees." BMC bioinformatics **7**(1): 439.

Chevenet, F. J., Matthieu Peeters, Martine De Oliveira, Tulio Gascuel, Olivier (2013). "Searching for virus phylotypes." Bioinformatics **29**(5): 561-570.

Cobben, R. H. (1978). Evolutionary trends in Heteroptera. Part II Mouthpart-structures and feeding strategies, Veenman.

Damgaard, J. (2006). Systematics, historical biogeography and ecological phylogenetics in a clade of water striders, na.

Damgaard, J. (2008). "MtDNA diversity and species phylogeny of western Palaearctic members of the Gerris lacustris group (Hemiptera-Heteroptera: Gerridae) with implications for "DNA barcoding" of water striders." Insect Systematics & Evolution **39**(1): 107-120.

Damgaard, J. and F. A. Sperling (2001). "Phylogeny of the water strider genus Gerris Fabricius (Heteroptera: Gerridae) based on COI mtDNA, EF- 1α nuclear DNA and morphology." Systematic Entomology **26**(2): 241-254.

Damgaard, J. M., Felipe Ferraz Figueiredo Hayashi, Masakazu Weir, Tom A Zettel, Herbert (2012). "Molecular phylogeny of the pond treaders (Insecta: Hemiptera: Heteroptera: Mesoveliidae), discussion of the fossil record and a checklist of species assigned to the family." Insect Systematics & Evolution **43**(3-4): 175-212.

Drummond(a), A. J. X., Walter Heled, Joseph (2012). "Bayesian inference of species trees from multilocus data using* BEAST." Molecular Biology and Evolution **29**: 1969-1973.

Drummond(b), A. J. S., Marc A Xie, Dong Rambaut, Andrew (2012). "Bayesian phylogenetics with BEAUti and the BEAST 1.7." Molecular biology and evolution **29**(8): 1969-1973.

Esemu, S. N. D., Xiaofeng Kfusi, Achah J Hartley, Catherine S Ndip, Roland N Ndip, Lucy M Darby, Alistair C Post, Rory J Makepeace, Benjamin L (2019). "Aquatic Hemiptera in Southwest Cameroon: Biodiversity of Potential Reservoirs of Mycobacterium ulcerans and Multiple Wolbachia Sequence Types Revealed by Metagenomics." Diversity **11**(12): 225.

Ferreira, M. A. S., Marc A (2008). "Bayesian analysis of elapsed times in continuous- time Markov chains." Canadian Journal of Statistics **36**(3): 355-368.

Garchitorena, A. R., Benjamin Kamgang, Roger Ossomba, Joachim Babonneau, Jérémie Landier, Jordi Fontanet, Arnaud Flahault, Antoine Eyangoh, Sara Guégan, Jean-François (2014). "Mycobacterium ulcerans ecological dynamics and its association with freshwater ecosystems and aquatic communities: results from a 12-month environmental survey in Cameroon." PLoS neglected tropical diseases **8**(5).

Genbank® (2013). " "The NIH genetic sequence database, an annotated collection of all publicly available DNA sequences."."

Guindon, S. D., Jean-François Lefort, Vincent Anisimova, Maria Hordijk, Wim Gascuel, Olivier (2010). "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0." Systematic biology **59**(3): 307-321.

IRAD (2008). ""Rapport national sur l'état des ressources phytogénétiques pour l'alimentation et l'agriculture." " Rapport national: 1-83.

Kidd, D. M. and M. G. Ritchie (2006). "Phylogeographic information systems: putting the geography into phylogeography." Journal of Biogeography **33**(11): 1851-1865.

Lemey, P., A. Rambaut, A. J. Drummond and M. A. Suchard (2009). "Bayesian phylogeography finds its roots." PLoS computational biology **5**(9).

Marion, E., S. Eyangoh, E. Yeramian, J. Doannio, J. Landier, J. Aubry, A. Fontanet, C. Rogier, V. Cassisa and J. Cottin (2010). "Seasonal and regional dynamics of M. ulcerans transmission in environmental context: deciphering the role of water bugs as hosts and vectors." PLoS neglected tropical diseases **4**(7).

Marsollier, L., R. Robert, J. Aubry, J.-P. Saint André, H. Kouakou, P. Legras, A.-L. Manceau, C. Mahaza and B. Carbonnelle (2002). "Aquatic insects as a vector for Mycobacterium ulcerans." Appl. Environ. Microbiol. **68**(9): 4623-4628.

Meyin A Ebong, S., S. Eyangoh, E. Marion, J. Landier, L. Marsollier, J.-F. Guégan and P. Legall (2012). "Survey of water bugs in Bankim, a new Buruli ulcer endemic area in Cameroon." Journal of tropical medicine **2012**.

Meyin A Ebong, S., G. E. García Peña, D. Pluot-Sigwalt, L. Marsollier, P. Le Gall, S. Eyangoh and

J.-F. Guegan (2017). "Ecology and feeding habits drive infection of water bugs with Mycobacterium ulcerans." <u>EcoHealth</u> **14**(2): 329-341.

Meyin A Ebong, S., E. Petit, P. Le Gall, P.-P. Chen, N. Nieser, E. Guilbert, F. Njiokou, L. Marsollier, J.-F. Guégan and D. Pluot-Sigwalt (2016). "Molecular species delimitation and morphology of aquatic and sub-aquatic bugs (Heteroptera) in Cameroon." <u>PloS one</u> **11**(5).

Palomino, J., A. Obiang, L. Realini, W. Meyers and F. Portaels (1998). "Effect of oxygen on growth of Mycobacterium ulcerans in the BACTEC system." <u>Journal of clinical microbiology</u> **36**(11): 3420-3422.

Plowright, R. K., S. H. Sokolow, M. E. Gorman, P. Daszak and J. E. Foley (2008). "Causal inference in disease ecology: investigating ecological drivers of disease emergence." <u>Frontiers in Ecology and the Environment</u> **6**(8): 420-429.

Poisson, R. (1929). "Contribution à la faune du Cameroun Hémiptères aquatiques." <u>Faune Colon Fr</u> **3**: 135-164.

Poisson, R. (1937). "Voyage de MP-P. Grassé en Afrique Occidentale Française. Hemiptera aquatica." <u>Ann Soc Entomol Fr</u> **106**: 115-132.

Poisson, R. (1948). "Hydrocorises du Cameroun (Mission J. Carayon 1947)." <u>Rev Fr Entomol</u> **3**: 167-177.

Poisson, R. A. (1940). <u>Contribution à l'étude des Gerris de France et de l'Afrique du Nord (Hémiptères Gerridae)</u>.

Poisson, R. A. (1941). "Hémiptères aquatiques du Cameroun (Mission Lepesme, Paulian et Villiers 1939)." <u>Bulletin de la Société entomologique de France</u> **46**(2): 14-17.

Poisson, R. A. (1945). "Quelques hémiptères aquatiques africains et malgaches." <u>Bulletin de la Société entomologique de France</u> **50**(7): 87-94.

Polhemus, J. T. and D. A. Polhemus (2007). Global diversity of true bugs (Heteroptera; Insecta) in freshwater. <u>Freshwater animal diversity assessment</u>, Springer**:** 379-391.

Portaels, F., P. Elsen, A. Guimaraes-Peres, P.-A. Fonteyne and W. M. Meyers (1999). "Insects in the transmission of Mycobacterium ulcerans infection." <u>The Lancet</u> **353**(9157): 986.

Portaels, F. and S. Pattyn (1982). "Growth of mycobacteria in relation to the pH of the medium." <u>AnnInstitut Pasteur Microbiol (Paris)</u> **133**(2): 213-221.

Posada, D. (2008). "jModelTest: phylogenetic model averaging." <u>Molecular biology and evolution</u> **25**(7): 1253-1256.

Sanhueza, D., C. Chevillon, N. Bouzinbi, S. Godreuil and J.-F. Guégan (2018). "Chitin increases Mycobacterium ulcerans growth in acidic environments." <u>Microbes and environments</u> **33**(2): 234-237.

Schuh, R. (1979). "Review of RH Cobben, Evolutionary trends in Heteroptera. Part II. Mouthpart-structures and feeding strategies." <u>Systematic Zoology</u> **28**(4): 653-656.

Schuh, R. T. (1979). "Review of Evolutionary trends in Heteroptera Part II. Mouthpart-structures and feeding strategies, by R.H. Cobben." <u>Syst. Zool.</u> **28**: 653-656.

Supporting informations

S1 Table. Evolution model of nucleotide sequences the likelihood score was calculated with PhyML for 88 candidates models and number of substation schemes equal to 11. Akaike's test allowed to select the best model (GTR+I+G) with following settings: Partition = 012345, negative log likelihood (-InL) = 12489.6675, number of estimated parameters (K) = 356, P-inv = 0.0000, Gamma shape = 0.2310

S2 Table. Statistical table of phylotype analysis biomes of Cameroon total of 4 phylotype(s) related to 4 annotation(s) involving 99 strains (57 %) Shuffling: 1000 iterations; Sz p-value threshold: 50/1000

S3 Table. Statistical table of phylotype analysis on aquatic Heteroptera life traits combinations total of 20 phylotype(s) related to 19 annotation(s) involving 135 strains (77.6 %) Shuffling: 1000 iterations; Sz p-value treshold: all/1000