



Population collapse in viviparid gastropods of the Lake Victoria ecoregion started before the Last Glacial Maximum

Bert Van Bocxlaer, Catharina Clewing, Anne Duputié, Camille Roux,
Christian Albrecht

► To cite this version:

Bert Van Bocxlaer, Catharina Clewing, Anne Duputié, Camille Roux, Christian Albrecht. Population collapse in viviparid gastropods of the Lake Victoria ecoregion started before the Last Glacial Maximum. *Molecular Ecology*, 2020, 10.1111/mec.15599 . hal-02959771

HAL Id: hal-02959771

<https://hal.science/hal-02959771>

Submitted on 23 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

[This manuscript has been accepted for publication in Molecular Ecology. Published version available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15599>]

Population collapse in viviparid gastropods of the Lake Victoria ecoregion started before the Last Glacial Maximum

Bert Van Bocxlaer^{1,*}, Catharina Clewing², Anne Duputié¹, Camille Roux¹, Christian Albrecht^{2,3}.

¹ CNRS, Univ. Lille, UMR 8198 – Evo-Eco-Paleo, F-59000 Lille, France

² Department of Animal Ecology & Systematics, Justus Liebig University Giessen, D-35392 Giessen, Germany

³ Department of Biology, Mbarara University of Science and Technology, Mbarara, Uganda

* Corresponding author: bert.van-bocxlaer@univ-lille.fr; Tel: +33 3 20 43 40 15.

ORCID information: BVB: 0000-0003-2033-326X, CC: 0000-0001-7974-5037; AD: 0000-0003-2213-7940; CR: 0000-0001-9497-1446; CA: 0000-0002-1490-1825.

Running title: demographic collapse in African gastropods

Abstract

Ecosystems of Lake Victoria and riparian communities have been strongly disrupted by the introduction of the invasive Nile perch and its fishing industry. Beyond this invasion and other recent anthropogenic stressors, the Lake Victoria ecoregion also underwent phases of pronounced aridity over the Late Pleistocene, lastly during the Last Glacial Maximum (LGM). The consequences of recent and historic environmental change have been canvassed for the adaptive radiation of haplochromine cichlids occupying the ecoregion, but their effect on freshwater invertebrate diversity remains largely unknown. Here we use 15 microsatellite loci and approximate Bayesian computation to test whether viviparid gastropods experienced a population bottleneck during the LGM, as did cichlids. Clustering analyses support three viviparid gene pools in the Lake Victoria ecoregion, gathering specimens from 1) Lake Albert and the White Nile, 2) the Victoria Nile and Lake Kyoga and 3) Lake Victoria and tributaries. The last group contains the highest genetic diversity, but all groups have a considerable number of private alleles and are inferred to predate the LGM. Examinations of demographic history reveal a 190- to 500-fold population decline that started ~125-150 ka ago, thus substantially before the LGM bottleneck documented in haplochromine cichlids. Population collapses in viviparids are an order of magnitude more severe than declines in cichlids and have not been halted by the re-establishment of freshwater ecosystems since the LGM. Recent anthropogenic ecosystem deterioration is causing homogenization of previously diversified microhabitats, which may contribute to (local) extinction and enhanced gene flow among species within gene pools.

Keywords: Viviparidae, *Bellamya*, phylogeography, genetic diversity, approximate Bayesian computation

Introduction

Less than 1 % of the Earth's surface is covered by freshwater ecosystems, yet these habitats support ~10 % of all known species, including approximately a third of the global vertebrate biodiversity (Strayer & Dudgeon, 2010). The Lake Victoria ecoregion of the East African Rift System, which contains among others Lakes Albert, Kyoga and Victoria, has rapidly accumulated extraordinary freshwater biodiversity throughout the Quaternary, including almost 700 species of cichlid fishes (Danley et al., 2012; Salzburger, Van Bocxlaer, & Cohen, 2014). The region also underwent major geological and climatological changes during this period linked to the development of the East African Rift, including several phases of aridity (Bishop & Trendall, 1966; Johnson, Kelts, & Odada, 2000; Lehman, 2009; Tryon et al., 2016). Additionally, human activities have exerted multifactorial pressures on freshwater environments in the ecoregion over recent decades, through habitat degradation, pollution, exploitation, the introduction of invasive species, ecosystem modifications and climate change (Hecky, Mugidde, Ramlal, Talbot, & Kling, 2010; Johnson et al., 2000; Sayer, Máiz-Tomé, & Darwall, 2018; Verschuren et al., 2002). As such, understanding diversification dynamics in freshwater biota within the Lake Victoria ecoregion requires insight into the consequences of recent and historic environmental change. The interplay of environment and diversification dynamics have been canvassed for the remarkable evolutionary radiation of cichlid fishes in Lake Victoria while addressing how the lake accumulated hundreds of endemic species despite having been dry ~15 ka ago (Fryer, 2001, 2004; Johnson et al., 1996; Stager, Day, & Santini, 2004; Stager & Johnson, 2008; Verheyen, Salzburger, Snoeks, & Meyer, 2003). Apart from cichlids and the schistosome-transmitting snail genus *Biomphalaria* (Standley, Goodacre, Wade, & Stothard, 2014), the seemingly controversial interplay between environment and diversity in the Lake Victoria ecoregion has not generated detailed phylogeographic and demographic studies of freshwater taxa. This knowledge gap implies that the effects of ecosystem changes on the community composition and demography of benthic organisms remain poorly assessed, and that the effectiveness of currently implemented strategies for managing fisheries and sustaining biodiversity is unstudied.

Beyond cichlids, several other freshwater organisms display elevated diversity throughout the Lake Victoria ecoregion (Sayer et al., 2018), such as viviparid gastropods. Viviparidae has been proposed

73 to be particularly useful in reconstructing phylogeographic patterns due to poor dispersal in the absence
74 of hydrographic connections (Schultheiß, Van Bocxlaer, Riedel, von Rintelen, & Albrecht, 2014;
75 Stelbrink et al., 2020), and they are an important component of benthic ecosystems in general (Strong,
76 Gargominy, Ponder, & Bouchet, 2008; Van Bocxlaer & Strong, 2020), as illustrated in the Lake
77 Victoria ecoregion (Brown, 1994; Mandahl-Barth, 1954). The family has been studied on a rift-wide
78 scale at several occasions (Schultheiß et al., 2014; Schultheiß, Wilke, Jørgensen, & Albrecht, 2011;
79 Sengupta, Kristensen, Madsen, & Jørgensen, 2009), resulting in good knowledge on the
80 phylogeographic connectivity of drainage basins throughout the East African Rift. In contrast, the
81 relationships within clades occupying individual drainage basins remain largely unresolved in existing
82 phylogenetic analyses (Schultheiß et al., 2014; Sengupta et al., 2009). The observed polyphyly in
83 regional morphospecies and races may result from incomplete lineage sorting, from independent
84 recurrent mutations, or adaptive introgression between distant gene pools (Lee & Coop, 2017).
85 Studies of genetic diversity at the population level are thus required to obtain insight into regional
86 patterns of diversity and to link environmental and evolutionary history.

87 Here we study the population structure of the clade of *Bellamya* gastropods that occupies the Lake
88 Victoria ecoregion, i.e. the ‘Victoria clade’ (Schultheiß et al., 2014), to relate past environmental
89 change with demography and diversification dynamics. *Bellamya* is diploid, with separate sexes,
90 internal fertilization and ovovivipary (Van Bocxlaer & Strong, 2020), and the Victoria clade contains 13
91 geographical races, which have been grouped into 7 nominal species, 5 of which are endemic to Lake
92 Victoria (Brown, 1994; Mandahl-Barth, 1954). To examine the geographical structure of diversity we
93 first performed a network analysis using a fast evolving mitochondrial marker and increased
94 sampling. Subsequently, we used microsatellites to document how genetic diversity in Viviparidae of
95 the Lake Victoria ecoregion is structured in space and time. In light of the previous description of
96 geographical races and lake endemics, we hypothesized that the genetic diversity of viviparids in the
97 Lake Victoria ecoregion is geographically structured. Given the high morphological disparity and
98 elevated nominal species diversity in Lake Victoria, we hypothesized higher genetic diversity within
99 this lake and barriers to gene flow that reflect habitat divergence. Additionally, we used approximate
100 Bayesian computation (ABC) to test the hypothesis that individual gene pools of these benthic

gastropods experienced a population bottleneck during the Last Glacial Maximum (LGM), as did cichlids (Elmer et al., 2009).

Material and Methods

Sampling. Material was collected in 2008-2010 throughout the Lake Victoria ecoregion by CA and colleagues (Fig. 1), and was deposited in the Systematics and Biodiversity collection of the Justus Liebig University Giessen (UGSB). For the network analysis we used sequence data on cytochrome *c* oxidase subunit 1 (COI) from 60 specimens: one from Jørgensen, Kristensen, and Madsen (2008), 22 from Sengupta et al. (2009), 4 from Schultheiß et al. (2011), 16 from Schultheiß et al. (2014) and 17 sequences were newly generated (Fig. 1). Specimen information, including UGSB and NCBI GenBank accession numbers, is provided in Table S1. Our microsatellite dataset consists in total of 321 individuals from 39 localities at Lake Victoria, its tributaries, the Victoria Nile, Lake Kyoga, Lake Albert and the White Nile (Fig. 1). Specimen information is supplied in Table S2. Note that *Bellamya* is absent from three other large lakes in this ecoregion, i.e. Lakes George, Edward and Kivu (Brown, 1994). Because we recovered few *Bellamya* individuals at some of the sampling localities, we pooled localities with small samples sizes within a 5-km radius into a regrouped dataset (282 specimens from 21 populations with ≥ 10 individuals per population) from which some isolated, poorly-sampled localities were discarded. Depending on analytical requirements, we analysed the total or regrouped dataset for a robust characterisation of patterns of genetic variation within and between sampling localities. Regrouping did not alter the overall results and interpretation of our analyses compared to analyses of the total dataset.

Morphological identifications. Morphological identifications to race and morphospecies were performed with the identification key of Mandahl-Barth (1954), information in Brown (1994) and by evaluating type specimens. All *Bellamya* species and races currently recognized in the Lake Victoria ecoregion (Brown, 1994; Mandahl-Barth, 1954) are represented in our dataset (see SI Text). Construction of the regrouped dataset did not result in the exclusion of races or morphospecies, but implied that the rare *B. phthinotropis* is only represented by three specimens. Morphological

129 identifications were used to interpret patterns of genetic diversity, but not as an input variable to guide
130 molecular data analyses.

131
132 **COI data collection and analysis.** We extracted genomic DNA following the protocol of Schultheiß
133 et al. (2014). We used the forward primer LCO1490 (Folmer, Black, Hoeh, Lutz, & Vrijenhoek,
134 1994) and the reverse primer COX-B7R (Schultheiß et al., 2011) for PCR amplification. The PCR
135 conditions were an initial denaturation step at 95 °C for 60 s, followed by 29 amplification cycles
136 (denaturation at 95 °C for 30 s, annealing at 52 °C for 30 s and elongation at 72 °C for 30 s) and a
137 final extension step at 72 °C for 180 s. Sequencing was performed on an ABI3730XL sequencer at
138 LGC Genomics (Berlin, Germany). The first base pairs following the 3' end of each primer were
139 difficult to read and therefore trimmed, leaving a 593 bp fragment. The sequences were unambiguously
140 aligned with ClustalW (Thompson, Higgins, & Gibson, 1994) in BioEdit v. 7.2.5 (Hall, 1999) and
141 visually inspected. We constructed a haplotype network using this alignment and an infinite site
142 model with functions from the package PEGAS v. 0.12 (Paradis, 2010) in the statistical environment R
143 v. 3.6.1 (R Core Team, 2019).

144
145 **Microsatellite data collection.** The here analysed microsatellite loci have previously been described
146 by Clewing et al. (2013). Genomic DNA was extracted, amplified and genotyped as outlined in
147 Clewing et al. (2013). In summary, PCR reactions were performed separately for each locus, and loci
148 were subsequently pooled in four groups depending on the expected fragment length and fluorescent
149 dyes used (pool 1: Bel_L10, Bel_L13, Bel_L15, Bel_L53, Bel_L25; pool 2: Bel_L18, Bel_L23,
150 Bel_L50; pool 3: Bel_L45, Bel_L47, Bel_L63; pool 4: Bel_L16, Bel_L31, Bel_L32, Bel_L48,
151 Bel_L56). These pools were sent to LGC Genomics for precision electrophoresis on a 16-capillary
152 3130xl Genetic Analyzer (Applied Biosystems). Genotypes were called by CC using GeneMarker® v.
153 1.85 (SoftGenetics LLC). First, chromatograms were visually inspected and custom panels of allele
154 sizes per locus were created with prior knowledge on the length of repeat motives for each locus.
155 Subsequently, the software automatically called genotypes based on the created panels. Allele calls
156 that did not correspond to the associated panel were verified manually. For two markers, Bel_L18 and

Bel_L56, chromatograms were sometimes difficult to read or occasionally included more than two alleles, so these loci were excluded. Finally, individuals for which allele profiles were difficult to read were excluded from the dataset. All automated genotype calls were verified manually by CC and a subset of 15 % of individuals were genotyped twice, but as no inconsistencies occurred, we considered our genotype calling robust for subsequent analyses. Size data for each of the 321 specimens and each of the 15 microsatellite loci was tabulated for subsequent analyses. All data and analytical scripts have been uploaded to the Dryad Data Repository and are available via <https://doi.org/10.5061/dryad.q83bk3jg2> (Van Bocxlaer et al., 2020).

Examination of genetic variation within populations. We examined the 15 microsatellites for the presence of null alleles with the ApparentNulls method of the R package GENEPOP v.1.1.3 (Rousset, 2008) using the weighted average over all populations in the total and regrouped datasets. Subsequent analyses were performed both including and excluding loci with null alleles. We studied genetic variation within populations from the regrouped dataset. Genotypic linkage disequilibrium was documented pairwise for all possible pairs of loci with a log-likelihood ratio test as implemented in FSTAT v. 2.9.3.2 (Goudet, 2002). Significance was examined at a nominal level of $\alpha = 0.05$ and p -values were corrected for multiple tests using the false discovery rate (FDR; Benjamini & Hochberg, 1995). Subsequently, we estimated genetic variation for the microsatellite loci by examining the number of alleles, allelic richness, the observed and expected heterozygosity, and mean fixation indices (F_{IS} , F_{IT} , F_{ST}) in FSTAT and the R package HIERFSTAT v. 0.04-22 (Goudet & Jombart, 2015). Fixation indices were calculated following Weir and Cockerham (1984). The significance of departures from zero for F_{IS} , F_{IT} , F_{ST} were examined in FSTAT with 1000 randomizations of alleles within populations, alleles among populations, and individuals among populations, respectively. Subsequently, we tested whether our loci are in Hardy-Weinberg equilibrium with χ^2 tests and exact tests based on 1000 Monte Carlo permutations of alleles following Guo and Thompson (1992) with functions from PEGAS. Genetic diversity was also examined per sampling locality using SPAGeDi 1.5a (Hardy & Vekemans, 2002) and HIERFSTAT.

Spatial patterns of genetic structure. Spatial patterns of molecular differentiation were identified from the total dataset in three different ways. First, we subjected the centered matrix of genotypes to principal component analysis (PCA) with the function ‘indpca’ of HIERFSTAT. The coordinates for all individuals on PC1 and PC2 were then subjected to model-based clustering using Gaussian mixture models with functions of the R package MCLUST v. 5.4 (Scrucca, Fop, Murphy, & Raftery, 2016). MCLUST uses various models of variance-covariance structure to fit clusters to the data without *a priori* assumptions of group entities. The model-fit for the various models is subsequently evaluated with a Bayesian Information Criterion (BIC). Second, we performed Bayesian clustering analyses in STRUCTURE v. 2.3.4 (Pritchard, Stephens, & Donnelly, 2000), using an admixture model with correlated allele frequencies. Analyses were conducted for $K = 1-20$ with 30 independent runs per K . Each Markov chain ran for 5 000 000 iterations with a burn-in of 200 000 iterations. The underlying genetic structure was examined with the ΔK statistic (Evanno, Regnaut, & Goudet, 2005) in CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). Finally, we performed a discriminant analysis of principal components (DAPC) using the R package ADEGENET v. 2.1.1 (Jombart, 2008). We first evaluated the number of groups present in our dataset from the output of K -means clustering with a BIC and subsequently performed DAPC on scenarios retained from the clustering output. The results of DAPC were subsequently evaluated by examining the assignment probability of individuals to molecular groups.

Subsequently, we documented the proportions of shared and private alleles for the obtained molecular groups, i.e. gene pools, using the R package POPPR v. 2.8.1 (Kamvar, Tabima, & Grünwald, 2014) and the regrouped dataset. Because substantial differences existed in the number of individuals per molecular cluster, we resampled each molecular cluster to the smallest sample size with 1,000 replicates and tested for differences in the number of private alleles with a Kruskal-Wallis test followed by FDR-corrected pairwise Wilcoxon rank sum tests. Recent migration rates among populations in the regrouped dataset were assessed in BayesAss (Wilson & Rannala, 2003) from three independent runs of 20,000,000 Markov chain Monte Carlo iterations with default mixing parameters, a burn-in of 5,000,000 iterations and sampling every 10,000 iterations. We verified that the runs achieved stationary and that the variance among runs was low.

Finally, we examined the processes underlying geographical differentiation in the regrouped dataset by estimating the pairwise relatedness between individuals with a kinship coefficient, F_{ij} , in SPAGeDi. Kinship coefficients were regressed on the spatial distances among sampling localities in SPAGeDi, first for all localities together, and then for relevant subsets. This method is similar to regressing F_{ST} [or $F_{ST}/(1 - F_{ST})$] on spatial distances, with the difference that kinship coefficients can be established pairwise between individuals (used here) as well as between populations (see Fenster, Vekemans, & Hardy, 2003; Van Bocxlaer, 2017). Summary statistics are based on tests with 1000 permutations and jackknife procedures. Because *Bellamya* disperses very poorly in the absence of hydrographic connections (e.g. Schultheiß et al., 2014; Van Bocxlaer, Verschuren, Schettler, & Kröpelin, 2011) the distances between sampling localities were measured manually in Google Earth Pro v. 7.3.2.5776 as the shortest path of dispersal along the hydrographic network.

Examination of demographic history. All analyses of demographic history were performed with the regrouped dataset and ABC as implemented in the program DIYABC v.2.1.0 (Cornuet et al., 2014). Because null alleles are treated as missing data in DIYABC, demographic analyses included all 15 loci. First, we examined the extent to which five anagenetic scenarios of demographic changes in N_e were able to reproduce the patterns of genetic polymorphism observed in the empirical dataset (Fig. 2). Model 1) simulates constant N_e , 2) an increase in N_e , 3) a decline in N_e , 4) a bottleneck, and 5) population increase followed by decline. The priors for each parameter of N_e were set as a uniform distribution from 1 to 10,000,000 and for time parameters between 10 and 500,000 generations [with 1 generation \approx 1 year for *Bellamya* (Brown, 1994)]; additional constraints are indicated in Fig. 2. Subsequently, we designed cladogenetic models with all possible scenarios of hypothesized phylogenetic relationships between the recognized molecular groups, and three demographic models of changes in N_e upon divergence (Fig. S1): 1) divergence with constant N_e (3 time parameters, 1 parameters of N_e), 2) divergence with changing N_e upon each lineage split (3 time parameters, 5 parameters of N_e), and 3) divergence similar to 2) but allowing additional change in N_e along the longest branch (4 time parameters, 6 parameters of N_e). These models were implemented with the same priors as the anagenetic models and with additional constraints on the timing of population

splits: $t_2 > t_l$ and $t_2 > t_{lb}$. Change in N_e was unconstrained, so that increases or decreases were possible as time progressed. All calculation with DIYABC were implemented with the Generalized Stepwise Mutation model (Estoup, Jarne, & Cornuet, 2002), with the default parameters for microsatellite loci. For each model 1 million data sets were simulated and subsequently compared to our empirical data using 4 and 36 default summary statistics for anagenetic and cladogenetic simulations, respectively. Parameters were estimated for retained models from the posterior distribution of raw parameter values using the 1 000 ABC simulations that approximated the observed data best.

Some migration among gene pools was inferred in our analyses and, because the resulting gene flow could lead to inferential bias (see e.g. Heller, Chikhi, & Siegmund, 2013), we examined its potential effect on our ABC inferences in two ways. Because DIYABC cannot currently account for migration, we first repeated our analyses excluding specimens that showed signs of admixture, i.e. those with an assignment probability to their gene pool of <80% such as F_1 , F_2 or backcross hybrids. This strategy allows estimating the influence of migration from distant gene pools on ABC inferences. Second, we evaluated the effect of migration on our model comparisons by using a slightly different ABC approach that can account for migration. Specifically, we simulated datasets for three demographic scenarios (constant, declining and increasing population size, respectively) and 11 rates of migration into the studied population (0-10 migrants per generation). For each combination of demographic scenarios and migration rates ($n = 33$), we simulated 1 000 datasets with a coalescent simulator (Hudson, 2002), after which we used ABC to document the relative support for each demographic scenario under all 33 combinations. Population size was randomly drawn from a uniform distribution between 100 and 5,000,000 diploids; the start of population decline or increase was randomly drawn from a uniform distribution between 100 and 500,000 generations; for population decline model the current population was randomly drawn from a uniform distribution between 0.0 and $0.5 \times$ the ancestral N_e ; for population increase model the ancestral population was randomly drawn from a uniform distribution between 0.0 and $0.5 \times$ the current N_e . Each simulated dataset consisted of 500 sampled loci for each of 10 individuals, for which we calculated the following summary statistics: number of SNPs, diversity measured as π (Tajima, 1983), Watterson's θ (Watterson, 1975), Tajima's D (Tajima, 1989), and the number of SNPs in each category describing the site frequency spectrum.

Model comparison was performed by first training a random forest from 20 000 simulations without migration (RF; Pudlo et al., 2016) for each of the demographic models. This ABC RF is composed of 1 000 trained trees per demographic scenario and was used to estimate which of the demographic scenarios was best supported for each of the 33,000 simulated datasets.

Results

Haplotype network analysis. Our COI dataset contained 52 haplotypes and like previous phylogenetic studies (Schultheiß et al., 2014; Sengupta et al., 2009), our network analysis did not reveal geographic structure in the haplotype diversity of *Bellamya* from the Lake Victoria ecoregion (Fig. S2).

Linkage disequilibrium and genetic diversity. The total and regrouped datasets display low null allele frequencies (≤ 5.0 %) except for two of our 15 microsatellite loci, i.e. Bel_L50 and Bel_L45 (7.5 and 14.7 %, respectively). We removed these two loci from the dataset for the analyses reported below (except for demographic analyses, which were performed with 15 loci as mentioned above), although frequencies up to 8 % bias population genetic estimators only slightly (Putman & Carbone, 2014), and analyses with all 15 loci gave highly similar results. All analyses of genetic diversity have thus been performed with the regrouped dataset and 13 loci. Examining pairwise linkage among our loci did not reveal evidence for linkage disequilibrium, therewith enabling downstream demographic inferences that assume independence among loci.

Our dataset contains substantial allelic richness and genetic diversity (Table S3). Three loci display a deficit of heterozygotes (Bel_L23, Bel_L25 and Bel_L47; Table S3) and significant deviations from Hardy-Weinberg equilibrium were observed for all but two loci (Bel_L13 and Bel_L53; Table S4), implying the existence of population structure. Summary statistics of genetic diversity per sampling locality in the regrouped dataset are illustrated in Table S5 and Fig. S3. We found significantly higher allelic richness and genetic diversity in Lake Victoria ($A_r = 6.962$ among 14 gene copies; $H_E = 0.822$) than in other waterbodies ($A_r = 6.068$; $H_E = 0.761$; Wilcoxon rank sum tests: A_r : $W = 82$, $p = 0.004$; H_E : $W = 88.5$, $p < 0.001$; Fig. S3). Substantial differentiation exists among populations (global $F_{ST} =$

0.064, 95% CI = 0.048-0.082). Pairwise population differentiation ranges between $F_{ST} = 0.005$ and 0.150, and all but 9 of the 210 pairwise tests indicate significant differentiation after FDR-correction (Fig. S4).

Spatial patterns of genetic structure. PCA on the centered matrix of genotypes followed by model-based clustering with 14 models on the variance-covariance structure in the PCA revealed strong support for three gene pools (Fig. 3A). Solutions with a different number of groups received much less support ($\Delta BIC \geq 21.024$; Fig. 3B). Bayesian clustering with STRUCTURE and examination of the ΔK statistic supported two to three genetic clusters (Fig. 3C). Analysis of molecular structure with DAPC produced a four-group model, but with limited preference over models with three or five groups ($\Delta BIC < 2.03$; Fig. S5). Overall assignment probabilities decreased substantially when more than three groups were considered, however (99.7 %, 82.6 % and 68.5 % for models with three, four and five groups, respectively), so that the three group DAPC model is considered most representative, despite its slightly higher BIC score. In summary, our analyses of molecular structure jointly indicate a three-group model, which is moreover corroborated by high population pairwise differentiation among gene pools versus low differentiation within (Fig. S4; see below). These three groups display a clear geographical pattern: 1) sampling localities in Lake Albert and the White Nile, 2) those in Lake Kyoga and the Victoria Nile, and 3) those of Lake Victoria, with its tributaries, including Lake Muhazi in Rwanda, and localities directly at the outflow of Lake Victoria into the upper Victoria Nile, i.e. VN01 (Fig. 4). In what follows we label these gene pools as LAG, LKG, LVG, respectively, after the principal lake in each group.

Pairwise F_{ST} values between LAG and LVG, between LVG and LKG, and between LAG and LKG are 0.084, 0.078 and 0.108, respectively (regrouped dataset; all at $p < 0.001$). Beyond a higher allelic richness and genetic diversity (see above), LVG has more private alleles than LAG and LKG, even when resampled to identical sampling sizes (Kruskal-Wallis test; $\chi^2 = 2430.9$, $df = 2$, $p < 0.001$; Table S6, Fig. S6). The molecular clusters reflect morphospecies and geographical races to some extent (Fig. 3, 5): all individuals identified as *B. rubicunda* and most identified as *B. unicolor unicolor* belong to LAG ($n = 41$), whereas LKG ($n = 29$) consists almost exclusively of *B. costulata ugandae*.

The most genetically diverse group, LVG ($n=212$), also displayed the highest morphological disparity (11 races/morphospecies), but these races/morphospecies do not coincide with genetically distinct subgroups within LVG (Fig. 3C, D, 5). Overall, we found strong isolation-by-distance (IBD), i.e. a significant negative correlation between kinship coefficients and geographic distances ($r^2 = 0.916$; $p < 0.001$; Fig. 6A), in *Bellamya* throughout the Lake Victoria ecoregion. There is no IBD within LVG (Fig. 6B; 16 localities) and 113 of the 120 related F_{ST} values range between 0.00 and 0.05 (Fig. S4), revealing that overall limited genetic differentiation exists within LVG. Nevertheless, kinship coefficients within LVG are elevated at very small spatial scales compared to those at intermediate and large spatial scales (Fig. 6B). Recent migration rates estimated with BayesAss are low among gene pools: the fraction of individuals in a population that are inferred to be migrants from a population of another gene pool is usually $\leq 1.1\%$. Migration within each of the gene pools is elevated among at least some population pairs, however ($>10\%$; Table S7).

Demographic modelling indicates population collapse. Using anagenetic ABC models and the regrouped dataset we found that models simulating population decline (i.e. models 3 and 5 in Fig. 2) obtained high support for all three gene pools (joint posterior probability of 76.1-86.3 %; Table 1). Model 5, a past population expansion followed by population decline, consistently obtained the highest support for all three gene pools. Model 3 also consistently received substantial support and differs from model 5 by simulating a population at equilibrium, rather than under expansion, prior to the subsequent decline. These two models are difficult to distinguish because the ongoing decline is erasing the genetic signature of previous demographic events. The support for models of constant effective population size (N_e) was much smaller (13.3 – 21.5 %), whereas models that include a recent population expansion have very limited explanatory power (0.1 – 2.4 %). LKG experienced the strongest population decline, with a ~ 500 -fold reduction in N_e over the last ~ 102 ka (95% CI: 10-387 ka; Table 2). Nevertheless, LAG and LVG also endured 240-fold and 190-fold declines in N_e , respectively, over the same period (LAG: 140 ka, 95% CI: 18-426 ka; LVG: 173 ka, 95% CI: 26-452 ka). In simulations of model 5 the average value of t_2 , the onset of population expansion, was ~ 300 ka

(95% CI: 70-490 ka) for each molecular group. At that time, expansions in N_e were about 2-fold for each molecular group, and, hence, much more restricted than the subsequent population collapses. None of the cladogenetic models tested with ABC (Fig. S2) fitted the empirical data well, so we do not consider them further. In summary, the three gene pools display very similar demographic histories indicating a tight link since differentiation from their most recent common ancestor (MRAC). Their current population sizes represent only 0.2-0.6 % of those prior to ~125-150 ka ago. Some migration among gene pools occurs in our dataset (Table S7), which warranted an analysis of how the resulting gene flow may affect our ABC analyses. Specimens indicating admixture among gene pools were found in LAG ($n = 1$) and LVG ($n = 16$; regrouped dataset) but not in LKG. Model support remains very similar when these specimens are removed (Table S8), and also our estimates of model parameters are robust (Table S9). Running ABC on simulated datasets without migration, 99.8 %, 99.1 % and 99.2 % of simulations were assigned correctly to the models of population decline, population increase and a constant population size (Fig. S7). Including progressively increasing levels of migration, the demographic history of populations with declining and increasing N_e remained accurately reconstructed (correct assignment >97 %; Fig. S7A, B). Migration into populations at equilibrium size resulted in ~17 % of correct assignment only, with bias in the ABC inferences towards strongly inflated support for the population expansion model (~83 %; Fig. S7C). This bias results from migration causing long branches in the coalescent model, which mimics the expected coalescence under an expansion model. In conclusion, the model of population decline is not affected by migration nor did migration cause erroneous support for population decline under models of population increase and equilibrium. As such, our analysis of robustness indicates that the ABC support obtained for population decline cannot be attributed to bias caused by gene flow from distant gene pools.

Discussion

Our results indicate that three, geographically-structured gene pools exist in *Bellamya* gastropods of the Lake Victoria ecoregion, with the highest genetic diversity in LVG. We did not uncover barriers to gene flow within gene pools that reflect habitat divergence, which is surprising for LVG, given the

recognition of several endemic species (Brown, 1994; Mandahl-Barth, 1954). Our study presents the first examination of demographic history of freshwater invertebrates from the Lake Victoria ecoregion with ABC and it evidences population collapse, which started before the Last Glacial Maximum. Declines in viviparids are much more severe than in co-occurring haplochromine cichlids (Elmer et al., 2009), they have started earlier and they are evidenced in all studied gene pools. Whereas population collapse was initiated by historical environmental change, it appears to be intensified by ongoing anthropogenic ecosystem change. Below we discuss all these aspects in the ecological context of environmental change in the Lake Victoria ecoregion.

Spatial patterns of genetic structure. Our microsatellite dataset is the first to document three gene pools in viviparid gastropods of the Lake Victoria ecoregion, which occupy distinct geographic regions (see Fig. 4). Specimens of these gene pools had been included in rift-wide phylogenetic analyses based on 6 nuclear and mitochondrial gene fragments (Schultheiß et al., 2014; Sengupta et al., 2009), but these studies and our analysis of COI haplotypes (Fig. S2) did not uncover spatial patterns of genetic structure. Our data suggest that a few critical barriers have driven dispersal limitations. The main transition between LKG and LVG occurs ~20 km downstream of the outflow of Lake Victoria, between sampling localities VN01 and VN02, which are historically separated by a series of cascades, rapids and now also the Bujagali Hydropower Plant. Likewise, LAG and LKG are separated by a series of rapids that includes the Karuma and Murchison Falls, which also prevent fish migration (Hecky et al., 2010). Interestingly, all three gene pools contain substantial genetic diversity, including private alleles, although during the LGM, Lake Victoria was dry (Johnson et al., 2000; Johnson et al., 1996), Lake Albert experienced several low stands and potential desiccation (Beuning, Talbot, & Kelts, 1997), and the Victoria Nile became a permanent overflow of Lake Victoria only ~11.2 ka (Johnson et al., 2000).

Putative erosion of adaptive divergence. Genetic differentiation between the inferred viviparid gene pools seems to have been caused mainly by dispersal limitation but, given the consistent morphological differences among these gene pools, adaptive divergence may also have contributed

(Fig. 5). Adaptive divergence was previously only inferred within LVG, where morphological disparity is largest (Fig. 3D). Indeed, morphological differentiation within LVG has been interpreted as the result of ecological divergence related to different microhabitats within Lake Victoria (Brown, 1994; Mandahl-Barth, 1954), so that the Lake Victoria viviparids represent an evolutionary radiation (Schultheiß et al., 2014; Sengupta et al., 2009). Therefore, we anticipated that our neutral genetic markers would uncover clear traces of molecular divergence among morphospecies within LVG, which was not the case despite high molecular diversity and morphological disparity. Nevertheless, 112 out of 120 pairwise population F_{ST} values (~93 %) within LVG indicate limited albeit significant genetic differentiation. As such, subtle genetic differences exist in LVG, but they have not been picked up in our analyses of genetic structure. Interestingly, our sampling indicates that many of the morphospecies that were previously confined to different microhabitats within Lake Victoria currently occur in sympatry (Table S2, S5). This observation and the result that kinship coefficients within LVG are elevated at very small spatial scales (Fig. 6B) strongly suggest that gene flow occurs among sympatric morphospecies. To examine this hypothesis, we compared individual inbreeding coefficients for localities of LVG that contained one versus multiple morphospecies, which resulted in no significant differences ($F_i = 0.026 \pm 0.048$ vs. 0.030 ± 0.041 [mean \pm sd], respectively; Wilcoxon rank sum tests: $W = 32.5$, $p = 0.958$). Our data suggest that the anthropogenic homogenization of previously differentiated microhabitats (Sayer et al., 2018) is promoting gene flow among morphospecies (see below), which may also explain the deficit in heterozygotes observed for three loci. We would expect inbreeding to trigger similar patterns of heterozygosity across loci, however, and the uneven occurrence of null alleles in a few populations for these three loci (although being overall infrequent), may have contributed to heterozygote deficiency. Surprisingly, morphological differences currently persist among morphospecies of LVG that occur in sympatry, that share recent gene flow, and that are not molecularly differentiated for the here studied loci. Ecophenotypic plasticity is contradicted by regular sympatric co-occurrence and the observed morphological disparity exceeds that expected under such plasticity (see Van Bocxlaer, Van Damme, & Feibel, 2008). If morphological differentiation indeed was adaptive then we would expect it to be linked to regions of the genome that are/were under selection and where differentiation may persist in the face

of gene flow. If the size of such regions is restricted in comparison to that of the viviparid genome, we would unlikely have sampled such regions with our microsatellite loci. More in-depth investigations into the molecular diversity of the viviparids of Lake Victoria is warranted to further test this hypothesis, which would best be undertaken with genome-wide, rapidly-evolving molecular markers and samples that cover all shorelines of the lake (see e.g. Wagner et al., 2013).

Comparison of phylogeography and demography with haplochromine cichlids. The largest intraspecific genetic differentiation in the mega-diverse species flock of haplochromine cichlids of the Lake Victoria ecoregion occurs in Lake Kivu (Verheyen et al., 2003), from which other parts of the ecoregion were subsequently colonized (Elmer et al., 2009; Nagl et al., 2000). Lake Kivu does not harbor extant viviparid populations, and neither do Lakes Edward and George (Brown, 1994), but otherwise phylogeographic patterns are highly similar in viviparids and cichlids. For example, cichlids from tributaries of Lake Victoria are genetically more closely related to those of Lake Victoria than to those of Lakes Albert and Kyoga (Elmer et al., 2009). Our results suggest that the ancestral stock of viviparids from the Lake Victoria ecoregion either occurred in the geographic area currently covered by LVG, or that it colonized this area well before colonizing the regions occupied by LAG and LKG. The increase in viviparid N_e in ABC model 5 ~300 ka (95% CI 70-490 ka) coincides broadly with the formation of the Lake Victoria Basin ~400 ka ago (Bishop & Trendall, 1966; Johnson et al., 2000; Lehman, 2009). That demographic expansion was only two-fold, which indicates that large populations of *Bellamyia* were already existing in the ecoregion before the origin of the Lake Victoria Basin. Indeed, the MRCA of the viviparids from the Lake Victoria ecoregion was estimated to ~2.5 Ma (95% CI 1.5-4.3 Ma) (Schultheiß et al., 2014), which broadly overlaps with estimates on the MRCA of the Lake Victoria haplochromines ~1.4 Ma (Nagl et al., 2000) or ~3.0 Ma (95% CI 2.3-3.9 Ma; Elmer et al., 2009). As such, some genetic lineages of viviparids in the ecoregion clearly predate the Late Pleistocene desiccation of Lake Victoria, which has, beyond cichlids, also been reported for cyprinid fish (Rutaisire, Booth, Masembe, Nyakaana, & Muwanika, 2004), catfish (Mwita & Nkwengulila, 2008) and unionid bivalves (Ortiz-Sepulveda et al., 2020). Furthermore, the current genetic differentiation, estimated as population pairwise F_{ST} values, is very

similar in viviparids (0.005-0.150) and cichlids (0.013-0.104; Elmer et al., 2009) of the Lake Victoria ecoregion.

After formation of the basin, aquatic environments experienced three low stands, potentially related to 100,000-year glacial-interglacial cycles (Johnson et al., 2000). As mentioned, the last megadrought in the Lake Victoria ecoregion occurred at the LGM ~18-15 ka and included a substantial interval during which Lake Victoria, with a current surface area of almost 70,000 km², was completely desiccated. Lake Albert experienced simultaneous low stands and potential desiccation, and the Victoria Nile was only formed ~11.2 ka ago (Beuning et al., 1997; Johnson et al., 2000). Interestingly, none of the three viviparid gene pools evidence a population expansion related to the re-establishment of aquatic habitats in the Lake Victoria ecoregion after the LGM. Instead, our demographic modelling suggests that the *N_e* of each of these viviparid gene pools remained relatively stable during recent ecosystem calamities (Table 2). This diversity was potentially maintained in mountain-fed river systems along the western branch of the East African Rift (Johnson et al., 2000). The penultimate phase of megadrought in the Lake Victoria ecoregion transpired at the end of the last interglacial ~125 ka (Tryon et al., 2016), and whereas the extent of aridity during this period is more uncertain in comparison to the LGM (Johnson et al., 2000; Johnson et al., 1996), it corresponds broadly to the onset of viviparid population decline estimated in our ABC simulations, suggesting at least a similar level of aridity to that at the LGM.

Whereas we anticipated scenarios of early population expansion and subsequent decline to perform well in simulations, based on earlier demographic studies of cichlids (Elmer et al., 2009), the early onset of such decline and the magnitude of population collapse were unexpected. Indeed, the here documented collapses in viviparids (Table 3) are an order of magnitude more severe than those recorded in haplochromine cichlids (30- to 50- fold population declines), and they also started 100 ka earlier (Elmer et al., 2009). Additionally, our evidence indicates that they affected all viviparid gene pools in the Lake Victoria ecoregion simultaneously. Given the persistence and extent of population collapses, it is likely that viviparid species have gone extinct in the Lake Victoria ecoregion over the Late Pleistocene.

Paleoenvironmental studies in the Lake Victoria ecoregion focused primarily on the megadrought at the LGM because it is geologically better constrained than previous phases of aridity (Johnson et al., 1996), and because desiccation seemed counter-intuitive with cichlid evolution (Elmer et al., 2009; Verheyen et al., 2003). Consequently, cichlid research addressing this seeming contradiction likewise focused on the LGM, and responses in cichlid evolution to earlier ecological calamities may not have been fully considered. The increase in haplotype diversity observed for haplochromines since the penultimate phase of aridity ~125 ka (Nagl et al., 2000) suggests a simultaneous bottleneck for cichlids and viviparids. The alternative explanation, i.e. that the LGM megadrought affected cichlids more severely than viviparids, is unlikely given the limited dispersal abilities of viviparids compared to cichlids. A search for genomic imprints of earlier megadroughts in other freshwater organisms including other benthic invertebrates is warranted, especially because high-throughput sequencing technology now allows the reconstruction of population history with unprecedented resolution.

Recent ecological degradation. The historical population decline in viviparid gastropods is intensified by the ongoing ecological degradation resulting from multifactorial anthropogenic pressures on freshwater ecosystems of the Lake Victoria ecoregion (Hecky et al., 2010; Sayer et al., 2018; Van Damme, Seddon, & Carr, 2018). Indeed, as for cichlids (Elmer et al., 2009), no signs for a potential recovery of allelic diversity have been observed in viviparids since the LGM. Algal blooms and widespread bottom anoxia are becoming increasingly common in Lake Victoria (Hecky et al., 1994; Stager & Johnson, 2008), which could result in a shoaling of gastropods (Christian Albrecht, pers. obs.), as is occurring in Lakes Malawi and Tanganyika (Van Bocxlaer, Schultheiß, Plisnier, & Albrecht, 2012). Viviparids were previously (19th and early 20th centuries) found to maximum depths of 30 and 33 m in Lake Victoria (*B. trochlearis* and *B. phthinotropis*, respectively), and *B. rubicunda* was recovered from a maximum depth of 18 m in Lake Albert (Brown, 1994; Mandahl-Barth, 1954). In our surveys, no living *Bellamya* specimens were retrieved from depths >6 m in Lake Albert. In Lake Victoria *B. trochlearis* and *B. phthinotropis* were obtained from <15 m and usually in sympatry with other *Bellamya* species (see Table S5). At two localities living viviparids, i.e. *B. costulata ugandae* and *B. unicolor meta*, which were previously reported up to depths of 10 and 12 m,

respectively, were collected between 15 and 23 m. Beyond mass mortality in fishes, as occasionally reported (Ochumba, 1990), anoxia would asphyxiate large proportions of zoobenthos communities, especially of taxa with low mobility such as viviparids. The toll of such events on mollusc populations is evidenced by deposits of dead and moribund individuals along the shorelines of Lake Victoria (Van Bocxlaer and Albrecht, pers. obs.). The consequences of ongoing degradation may differ among shorelines, among the various habitats and among taxa. Some viviparid species, such as *B. phthinotropis*, which was predominantly found in deeper waters (see above), are rare in recent surveys and are reported to be critically endangered (Sayer et al., 2018).

The ongoing ecological degradation of Lake Victoria is homogenising environmental conditions through the elimination of microhabitats, which may cause the reduction of endemic diversity in various freshwater taxa through extinction and faunal homogenization (Sayer et al., 2018). Our data indicate that habitat destruction has forced putatively differently adapted (incipient) viviparid species to occupy increasingly more similar habitats, and that this renewed (ecological and geographical) proximity is causing gene flow among them, which may erode genetic distinctiveness. Estimating the extent to which reproductive barriers among viviparids have already been obliterated is currently difficult, especially because it is unclear to what level such barriers existed prior to the onset of ecological degradation and given the lack of dense genomic sampling. Nevertheless, reverse speciation has previously been hypothesized for co-occurring cichlids (Seehausen, 2006), so investigations as to how reproductive barriers have changed over time are warranted for other freshwater biota in the Lake Victoria ecoregion. Our data suggests inbreeding among morphospecies and therewith faunal homogenisation in aquatic invertebrates and we document the strongest demographic collapse observed thus far in the Lake Victoria ecoregion. Given the current rate of ecosystem change throughout the entire Lake Victoria ecoregion, and given the anticipated additional anthropogenic pressure expected over coming decades (Sayer et al., 2018), further in-depth studies of the viviparid species flock from Lake Victoria, and benthic invertebrate diversity in general, are highly pressing. Indeed, understanding the consequences of ongoing, multifactorial ecosystem transformations on freshwater communities will be critical for sustaining biodiversity and managing fisheries in the Lake Victoria ecoregion over the coming decennia.

548

549 **Acknowledgements**

550 We thank Daniel Engelhard, Thies Geertz, Charles Lange and Roland Schultheiß for collecting
551 material, Casim Umba Tolo and Julius B. Lejju of Mbarara University of Science and Technology for
552 logistical support, Sophie Gallina for help running analyses on a university server, and Jean-Francois
553 Arnaud and Isabelle De Cauwer for discussions that improved the paper. Sampling in Rwanda was
554 organized in the framework of a memorandum of understanding with Umutara Polytechnic University
555 and in Uganda and Kenya with permits from the National Councils of Science and Technology. This
556 project was funded by project ANR-JCJC-EVOLINK of the French Agence Nationale de la
557 Recherche to BVB and grants of the Deutsche Forschungsgemeinschaft to CA (DFG AL 1076/5-2,
558 DFG AL 1076/6-2). This work is a contribution to the CPER research project CLIMIBIO and BVB,
559 AD and CR thank the French Ministère de l'Enseignement Supérieur et de la Recherche, the Hauts-
560 de-France Region and the European Funds for Regional Economic Development for financial support.

561

References

- Abell, R., Thieme, M. L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., . . . Petry, P. (2008). Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *BioScience*, 58(5), 403-414 (<https://doi.org/410.1641/B580507>).
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57(1), 289-300.
- Beuning, K. R. M., Talbot, M. R., & Kelts, K. (1997). A revised 30,000-year paleoclimatic and paleohydrologic history of Lake Albert, East Africa. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 136, 259-279.
- Bishop, W. W., & Trendall, A. F. (1966). Erosion-surfaces, tectonics and volcanic activity in Uganda. *Quarterly Journal of the Geological Society*, 122, 385-420.
- Brown, D. S. (1994). *Freshwater snails of Africa and their medical importance. Second edition.* London: Taylor & Francis.
- Clewing, C., Schultheiß, R., Wilke, T., & Albrecht, C. (2013). Isolation, characterization and cross-species amplification of 15 microsatellite markers for the African species of the freshwater gastropod genus *Bellamya*. *Conservation Genetics Resources*, 5, 51-54.
- Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Lebois, R., . . . Estoup, A. (2014). DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30(8), 1187-1189.
- Danley, P. D., Husemann, M., Ding, B., DiPietro, L. M., Beverly, E. J., & Peppe, D. J. (2012). The impact of the geological history and paleoclimate on the diversification of East African cichlids. *International Journal of Evolutionary Biology*, 2012, e574851 (doi: 574810.571155/572012/574851).
- Elmer, K. R., Reggio, C., Wirth, T., Verheyen, E., Salzburger, W., & Meyer, A. (2009). Pleistocene desiccation in East Africa bottlenecked but did not extirpate the adaptive radiation of Lake Victoria haplochromine cichlid fishes. *Proceedings of the National Academy of Sciences of the U.S.A.*, 106(32), 13404-13409.
- Estoup, A., Jarne, P., & Cornuet, J.-M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, 11, 1591-1604.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14, 2611-2620.
- Fenster, C. B., Vekemans, X., & Hardy, O. J. (2003). Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (Leguminosae). *Evolution*, 57(5), 995-1007.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5), 294-299.
- Fryer, G. (2001). On the age and origin of the species flock of haplochromine cichlid fishes of Lake Victoria. *Proceedings of the Royal Society B-Biological Sciences*, 268, 1147-1152.
- Fryer, G. (2004). Speciation rates in lakes and the enigma of Lake Victoria. *Hydrobiologia*, 519, 167-183.
- Goudet, J. (2002). Fstat version 2.9.3.2. University of Lausanne, Switzerland: Institute of Ecology, Biology Building.
- Goudet, J., & Jombart, T. (2015). hierfstat: Estimation and tests of hierarchical *F*-statistics. R package version 0.04-22. <https://CRAN.R-project.org/package=hierfstat>.
- Guo, S. W., & Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48, 361-372.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95-98.
- Hardy, O. J., & Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population level. *Molecular Ecology Notes*, 2, 618-620.
- Hecky, R. E., Bugenyi, F. W. B., Ochumba, P., Talling, J. F., Mugidde, R., Gophen, M., & Kaufman, L. (1994). Deoxygenation of the deep water of Lake Victoria, East Africa. *Limnology and Oceanography*, 39(6), 1476-1481.

- Hecky, R. E., Mugidde, R., Ramlal, P. S., Talbot, M. R., & Kling, G. W. (2010). Multiple stressors cause rapid ecosystem change in Lake Victoria. *Freshwater Biology*, 55(suppl. 1), 19-42.
- Heller, R., Chikhi, L., & Siegismund, H. R. (2013). The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *Plos One*, 8(5), e62992 (doi: 62910.61371/journal.pone.0062992).
- Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, 18(2), 337-338.
- Johnson, T. C., Kelts, K., & Odada, E. (2000). The Holocene history of Lake Victoria. *Ambio*, 29(1), 2-11.
- Johnson, T. C., Scholz, C. A., Talbot, M. R., Kelts, K., Ricketts, R. D., Ngobi, G., . . . McGill, J. W. (1996). Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science*, 273, 1091-1093.
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403-1405.
- Jørgensen, A., Kristensen, T. K., & Madsen, H. (2008). A molecular phylogeny of apple snails (Gastropoda, Caenogastropoda, Ampullariidae) with an emphasis on African species. *Zoologica Scripta*, 37, 245-252.
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281 (doi: 210.7717/peerj.7281).
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15, 1179-1191.
- Lee, K. M., & Coop, G. (2017). Distinguishing among modes of convergent adaptation using population genomic data. *Genetics*, 207, 1591-1619.
- Lehman, J. T. (2009). Lake Victoria. In H. J. Dumont (Ed.), *The Nile: Origin, environments, limnology and human use* (Vol. 89, pp. 215-242). New York: Springer.
- Mandahl-Barth, G. (1954). The freshwater mollusks of Uganda and adjacent territories. *Annals of the Royal Museum of Belgian Congo Tervuren, series 8 - Zoological Sciences*, 32, 1-206.
- Mwita, C. J., & Nkwengulila, G. (2008). Molecular phylogeny of the clariid fishes of Lake Victoria, Tanzania, inferred from cytochrome b DNA sequences. *Journal of Fish Biology*, 73(5), 1139-1148.
- Nagl, S., Tichy, H., Mayer, W. E., Takezaki, N., Takahata, N., & Klein, J. (2000). The origin and age of haplochromine fishes in Lake Victoria, East Africa. *Proceedings of the Royal Society B-Biological Sciences*, 267, 1049-1061.
- Ochumba, P. B. O. (1990). Massive fish kills within the Nyanza Gulf of Lake Victoria, Kenya. *Hydrobiologia*, 208, 93-99.
- Ortiz-Sepulveda, C. M., Stelbrink, B., Vekemans, X., Albrecht, C., Riedel, F., Todd, J. A., & Van Bocxlaer, B. (2020). Diversification dynamics of freshwater bivalves (Unionidae: Parreysiinae: Coelaturini) indicate historic hydrographic connections throughout the East African Rift System. *Molecular Phylogenetics and Evolution*, 148, e106816 (doi: 106810.101016/j.ympev.102020.106816).
- Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26, 419-420.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.
- Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., & Robert, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics*, 32(6), 859-866.
- Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, 4, 4399-4428.
- R Core Team. (2019). R: A language and environment for statistical computing. v. 3.6.1 (Version 3.2.1). Vienna, Austria: R foundation for statistical computing. Retrieved from <http://www.R-project.org/>
- Rousset, F. (2008). Genepop'007: a complete re-implementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, 8, 103-106.

- Rutaisire, J., Booth, A. J., Masembe, C., Nyakaana, S., & Muwanika, V. B. (2004). Evolution of *Labeo victorianus* predates the Pleistocene desiccation of Lake Victoria: evidence from mitochondrial DNA sequence variation. *South African Journal of Science*, 100(11-12), 607-608.
- Salzburger, W., Van Bocxlaer, B., & Cohen, A. S. (2014). Ecology and evolution of the African Great Lakes and their faunas. *Annual Review of Ecology, Evolution and Systematics*, 45, 519-545.
- Sayer, C. A., Máiz-Tomé, L., & Darwall, W. R. T. (2018). *Freshwater biodiversity in the Lake Victoria Basin: Guidance for species conservation, site protection, climate resilience and sustainable livelihoods*. Cambridge, UK and Gland, Switzerland: IUCN.
- Schultheiß, R., Van Bocxlaer, B., Riedel, F., von Rintelen, T., & Albrecht, C. (2014). Disjunct distributions of freshwater snails testify to a central role of the Congo system in shaping biogeographical patterns in Africa. *BMC Evolutionary Biology*, 14, 42.
- Schultheiß, R., Wilke, T., Jørgensen, A., & Albrecht, C. (2011). The birth of an endemic species flock: demographic history of the *Bellamya* group (Gastropoda, Viviparidae) in Lake Malawi. *Biological Journal of the Linnean Society*, 102, 130-143.
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1), 205-233.
- Seehausen, O. (2006). Conservation: Losing biodiversity by reverse speciation. *Current Biology*, 16(9), R334-R337.
- Sengupta, M. E., Kristensen, T. K., Madsen, H., & Jørgensen, A. (2009). Molecular phylogenetic investigation of the Viviparidae (Gastropoda: Caenogastropoda) in the lakes of the Rift Valley area of Africa. *Molecular Phylogenetics and Evolution*, 52, 797-805.
- Stager, J. C., Day, J. J., & Santini, S. (2004). Comment on "Origin of the superflock of cichlid fishes from Lake Victoria, East Africa". *Science*, 304, 963b.
- Stager, J. C., & Johnson, T. C. (2008). The late Pleistocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia*, 596, 5-16.
- Standley, C. J., Goodacre, S. L., Wade, C. M., & Stothard, J. R. (2014). The population genetic structure of *Biomphalaria choanophala* in Lake Victoria, East Africa: implications for schistosomiasis transmission. *Parasites & Vectors*, 7, e524 (510 pages).
- Stelbrink, B., Richter, R., Köhler, F., Riedel, F., Strong, E. E., Van Bocxlaer, B., . . . von Rintelen, T. (2020). Global diversification dynamics since the Jurassic: low dispersal and habitat-dependent evolution explain hotspots of diversity and shell disparity in river snails (Viviparidae). *Systematic Biology*, (doi: 10.1093/sysbio/syaa011).
- Strayer, D. L., & Dudgeon, D. (2010). Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society*, 29(1), 344-358.
- Strong, E. E., Gargominy, O., Ponder, W. F., & Bouchet, P. (2008). Global diversity of gastropods (Gastropoda; Mollusca) in freshwater. *Hydrobiologia*, 595, 149-166.
- Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, 105(2), 437-460.
- Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics*, 123(3), 597-601.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673-4680.
- Tryon, C. A., Faith, J. T., Peppe, D. J., Beverly, E. J., Blegen, N., Blumenthal, S. A., . . . Sharp, W. D. (2016). The Pleistocene prehistory of the Lake Victoria basin. *Quaternary International*, 404, 100-114.
- Van Bocxlaer, B. (2017). Hierarchical structure of ecological and non-ecological processes of differentiation shaped ongoing gastropod radiation in the Malawi Basin. *Proceedings of the Royal Society B-Biological Sciences*, 284, e20171494 (doi: 10.1098/rspb.2017.1494).
- Van Bocxlaer, B., Clewing, C., Duputié, A., Roux, C., & Albrecht, C. (2020). Data and scripts supporting: Population collapse in viviparid gastropods of the Lake Victoria ecoregion started before the Last Glacial Maximum. Dryad, <https://doi.org/10.5061/dryad.q83bk3jg2>.

- Van Bocxlaer, B., Schultheiß, R., Plisnier, P.-D., & Albrecht, C. (2012). Does the decline of gastropods in deep water herald ecosystem change in Lakes Malawi and Tanganyika? *Freshwater Biology*, 57, 1733-1744.
- Van Bocxlaer, B., & Strong, E. E. (2020). Viviparidae Gray, 1847. In C. Lydeard & K. S. Cummings (Eds.), *Freshwater mollusks of the world - A distribution atlas* (pp. 43-50). Baltimore, USA: Johns Hopkins University Press.
- Van Bocxlaer, B., Van Damme, D., & Feibel, C. S. (2008). Gradual versus punctuated equilibrium evolution in the Turkana Basin molluscs: Evolutionary events or biological invasions? *Evolution*, 62(3), 511-520.
- Van Bocxlaer, B., Verschuren, D., Schettler, G., & Kröpelin, S. (2011). Modern and early Holocene mollusc fauna of the Ounianga lakes (northern Chad): implications for the palaeohydrology of the central Sahara. *Journal of Quaternary Science*, 26(4), 433-447.
- Van Damme, D., Seddon, M., & Carr, J. A. (2018). The status and distribution of freshwater molluscs in the Lake Victoria Basin. In C. A. Sayer, L. Máiz-Tomé, & W. R. T. Darwall (Eds.), *Freshwater biodiversity in the Lake Victoria Basin: Guidance for species conservation, site protection, climate resilience and sustainable livelihoods* (pp. 65-81). Cambridge, UK and Gland, Switzerland: IUCN.
- Verheyen, E., Salzburger, W., Snoeks, J., & Meyer, A. (2003). Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science*, 300, 325-329.
- Verschuren, D., Johnson, T. C., Kling, H. J., Edgington, D. N., Leavitt, P. R., Brown, E. T., . . . Hecky, R. E. (2002). History and timing of human impact on Lake Victoria, East Africa. *Proceedings of the Royal Society B-Biological Sciences*, 269, 289-294.
- Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., . . . Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, 22, 787-798.
- Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical population biology*, 7(2), 256-276.
- Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, 38(6), 1358-1370.
- Wilson, G. A., & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, 163, 1177-1191.

Data accessibility

Various metadata are provided in Supplemental Information, including data on specimen vouchers and NCBI GenBank accession numbers. All data files and the R script used for the analysis have been deposited in the Dryad Data Repository and are available via <https://doi.org/10.5061/dryad.q83bk3jg2>.

Author contributions

BVB and CA designed research; CA coordinated sampling; CC performed molecular biology; BVB, AD and CR performed analyses; BVB wrote the paper, with help of all authors.

Tables

Table 1. Relative model support for the five anagenetic demographic models included in our ABC simulations. Scenarios of effective population size through time for each of these models are displayed in Fig. 2.

Group	Model 1	Model 2	Model 3	Model 4	Model 5	Model 3+5
Lake Albert Group	0.1468	0.0003	0.3591	0.0009	0.4930	0.8521
Lake Kyoga Group	0.1328	0.0000	0.3571	0.0041	0.5060	0.8631
Lake Victoria Group	0.2147	0.0030	0.3443	0.0214	0.4166	0.7609

Table 2. Estimates on the principal demographic parameters for the best supported ABC models. 95% CI indicates the 95% confidence intervals. Given the high similarities in models 3 and 5, their averaged outcome is presented under Model 3+5. N_e represents the number of diploid individuals in the population and times t , t_1 and t_2 are expressed in number of generations [1 generation is ~ 1 year for *Bellamya* (Brown, 1994)].

	mean N_e	mode N_e	95% CI		mean t or t_1	mode t or t_1	95% CI		mean $N_{e_{big}}$	mode $N_{e_{big}}$	95% CI		mean t_2	mode t_2	mean $N_{e_{s_Anc}}$	mode $N_{e_{s_Anc}}$
Lake Albert Group																
Model 3	23 800	12 700	5 830	60 300	186 000	72 900	25 700	460 000	4.36×10^6	1.36×10^6	259 000	9.63×10^6	n.a.	n.a.	n.a.	n.a.
Model 5	19 800	9 220	4 840	51 400	98 600	51 600	15 000	308 000	6.06×10^6	8.60×10^6	1.12×10^6	9.82×10^6	291 000	323 000	2.81×10^6	503 000
Model 3+5	21 500	11 200	5 310	56 000	140 000	66 000	17 500	426 000	5.22×10^6	8.11×10^6	397 000	9.77×10^6	n.a.	n.a.	n.a.	n.a.
Lake Kyoga Group																
Model 3	9 860	4 080	2 210	27 100	116 000	28 500	11 000	411 000	4.26×10^6	2.49×10^6	239 000	9.58×10^6	n.a.	n.a.	n.a.	n.a.
Model 5	12 900	5 110	3 300	35 00	81 600	30 000	11 400	265 000	6.73×10^6	9.83×10^6	1.82×10^6	9.88×10^6	293 000	439 000	3.50×10^6	2.17×10^6
Model 3+5	11 700	5 570	2 750	32 700	102 000	40 700	9 690	387 000	5.58×10^6	7.75×10^6	484 000	9.82×10^6	n.a.	n.a.	n.a.	n.a.
Lake Victoria Group																
Model 3	27 800	12 800	6 800	72 900	213 000	85 000	35 900	470 000	4.19×10^6	1.50×10^6	158 000	9.63×10^6	n.a.	n.a.	n.a.	n.a.
Model 5	30 600	10 200	7 000	69 900	136 000	94 500	22 000	360 000	6.04×10^6	9.41×10^6	1.12×10^6	9.72×10^6	345 000	414 000	2.82×10^6	722 000
Model 3+5	27 000	11 700	6 870	70 000	173 000	78 300	25 500	452 000	5.18×10^6	8.95×10^6	348 000	9.71×10^6	n.a.	n.a.	n.a.	n.a.

Figure captions

Figure 1. Geographical context of this study with indication of the major lakes and sampling localities (triangles for the COI dataset, circles for the microsatellite dataset). The Lake Victoria ecoregion is demarcated. Lakes Albert and Kivu are currently considered to be part of the Upper Nile and Lake Tanganyika ecoregions, respectively (see Abell et al., 2008), but their historic biogeography is strongly connected with the Lake Victoria ecoregion. Therefore, we here consider them to be part of the Lake Victoria ecoregion for practical reasons (dashed lines).

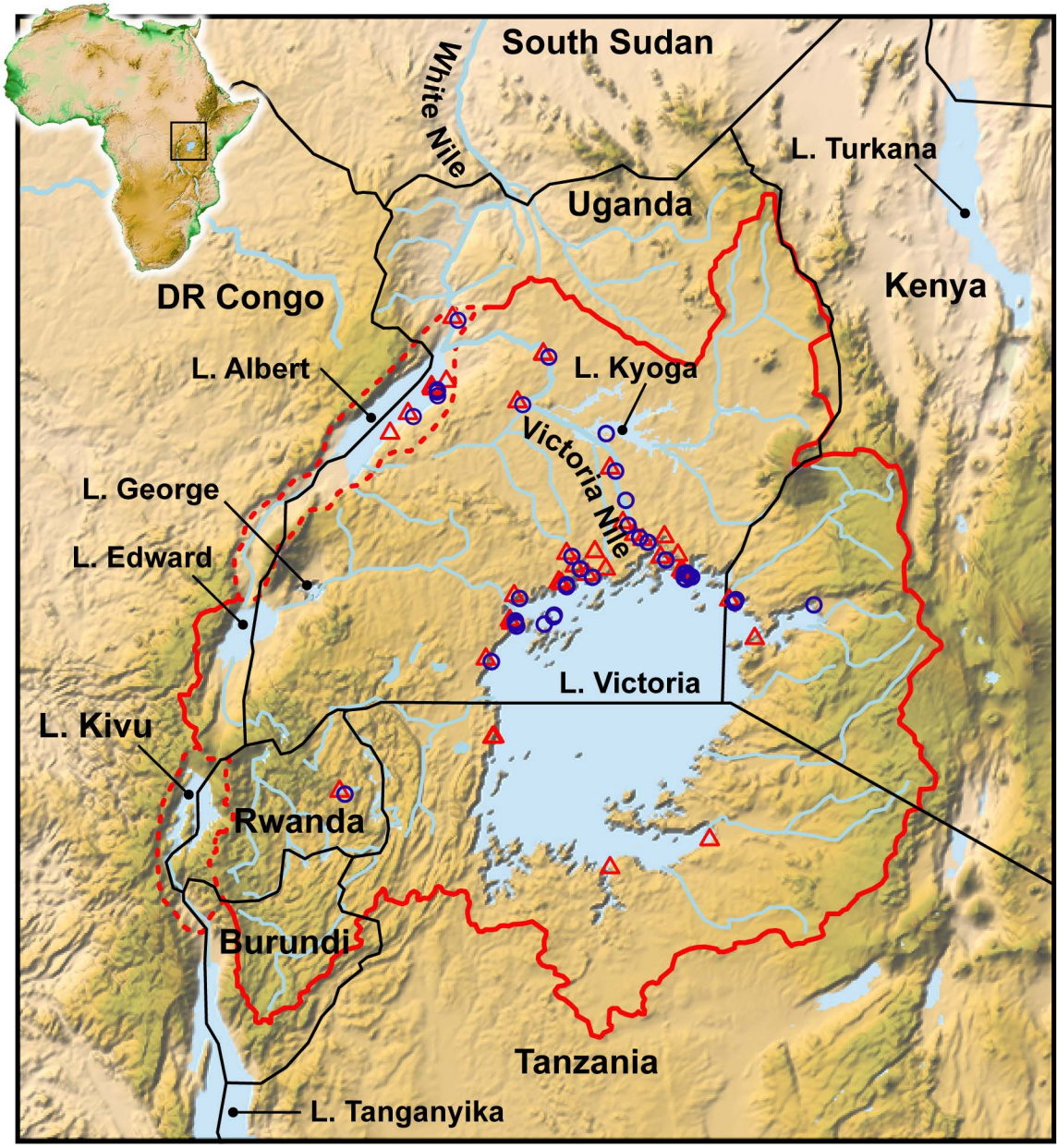
Figure 2. Five scenarios of anagenetic demographic change in effective population size (N_e) over time (t), as implemented in approximate Bayesian computations. Scenario 1) indicates constant N_e , scenario 2) an expansion in N_e , scenario 3) a decline in N_e , scenario 4) a population bottleneck, and scenario 5) population expansion followed by population decline. These scenarios are reflected in the specified model conditions (see Methods).

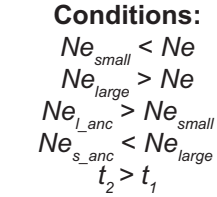
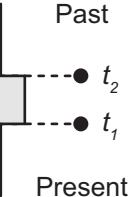
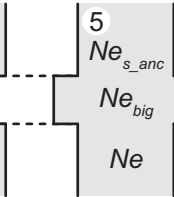
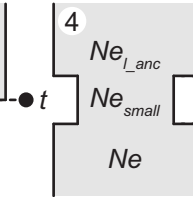
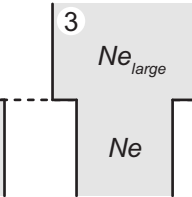
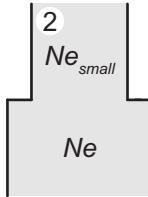
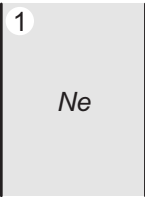
Figure 3. Structure of molecular diversity in *Bellamyia* from the Lake Victoria ecoregion. A) Principal component analysis on the centered matrix of genotypes for the total dataset (39 populations; 321 individuals) with results of model-based clustering using 14 variance-covariance models (B; see Methods). C) results of Bayesian clustering with STRUCTURE for $K = 2, 3$; the individuals of one cluster that approach another cluster in A, correspond to the individuals that have mixed assignments in C under scenario $K = 3$. D) the morphospecies belonging to each molecular group. LAG: *B. unicolor unicolor* (a) and *B. rubicunda* (b); LKG: *B. costulata ugandae* (c) and *B. u. elatior* (d); LVG: *B. c. ugandae* (e), *B. c. dagusiae* (f), *B. c. costulata* (g), *B. u. elatior* (h), *B. u. meta* (i), *B. phthinotropis* (j), *B. jucunda altior* (k), *B. j. kisumiensis* (l), *B. j. jucunda* (m), *B. trochlearis* (n-p), *B. constricta* (q).

Figure 4. Geographical distribution of sampling localities in the total dataset with pie charts indicating the average assignment of individuals per sampling localities to the three gene pools following the output of STRUCTURE.

Figure 5. Principal component analysis on the centered matrix of genotypes in our total dataset with symbols indicating the race or morphospecies to which individuals have been assigned. Specimens from LAG have been assigned to *B. rubicunda* and *B. unicolor unicolor*, whereas most specimens from LKG were assigned to *B. costulata ugandae*. Most morphological disparity was found in LVG, but these races/morphospecies do not coincide with genetically distinct subgroups (*B. c.* = *Bellamyia costulata*, *B. j.* = *Bellamyia jucunda*, *B. u.* = *Bellamyia unicolor*).

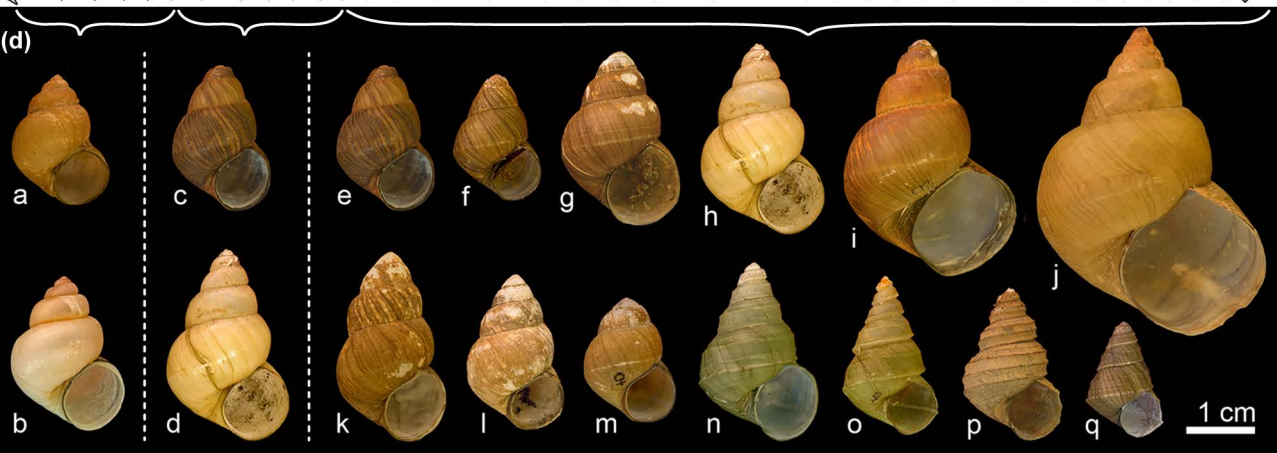
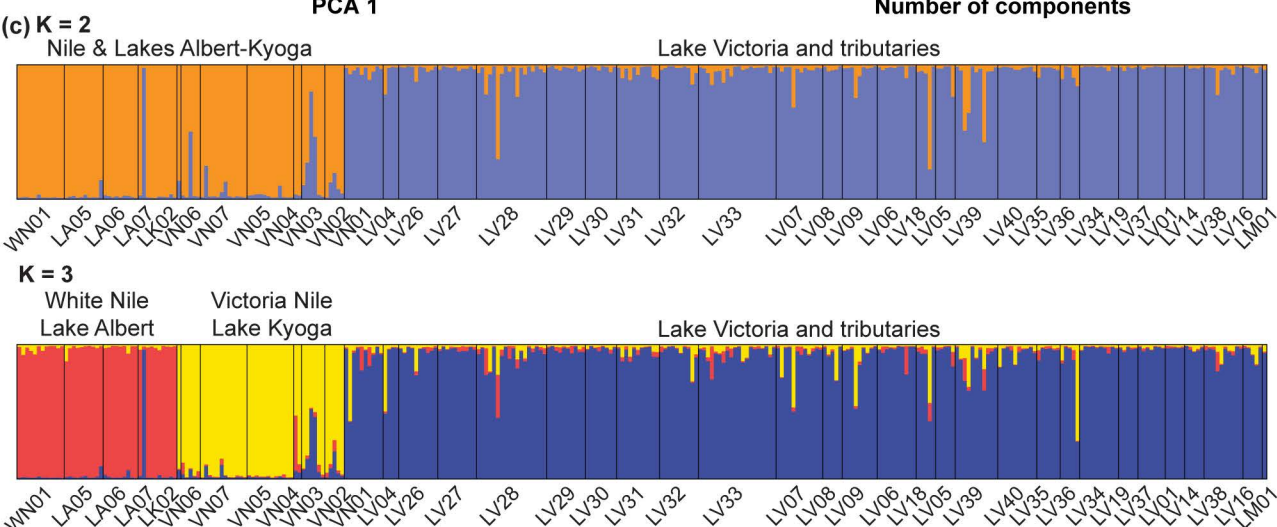
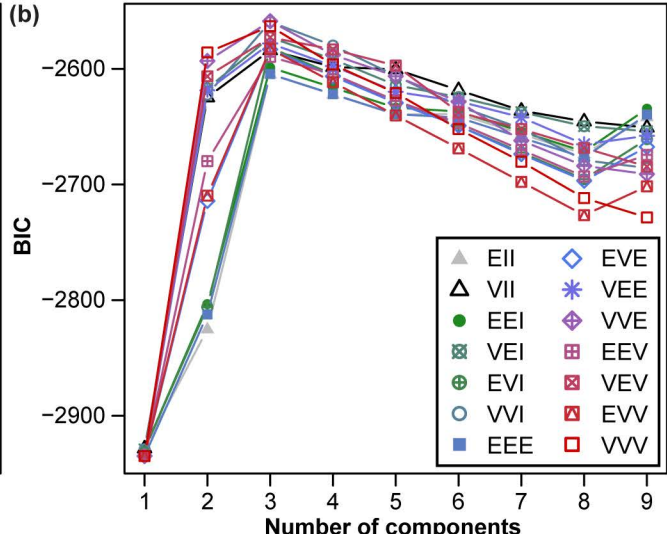
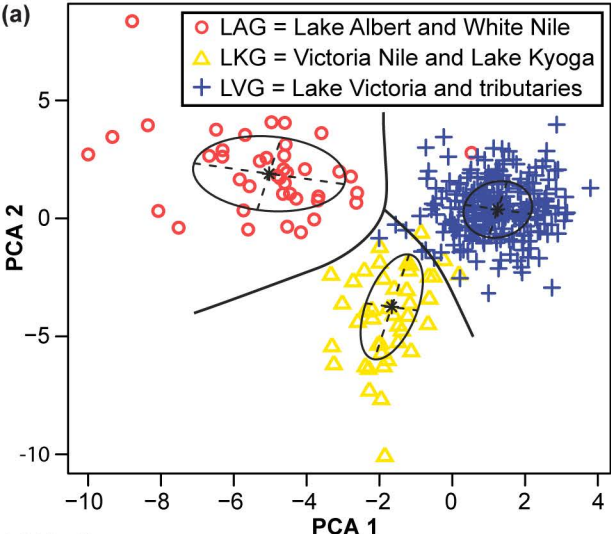
Figure 6. Regression of kinship coefficients on geographic distance for the regrouped dataset indicates a pattern of isolation-by-distance (IBD) over the entire ecoregion (A), but not for the Lake Victoria group (LVG; 16 localities; B), although kinship coefficients are elevated over very short geographical distances for LVG (blue vertical bar).

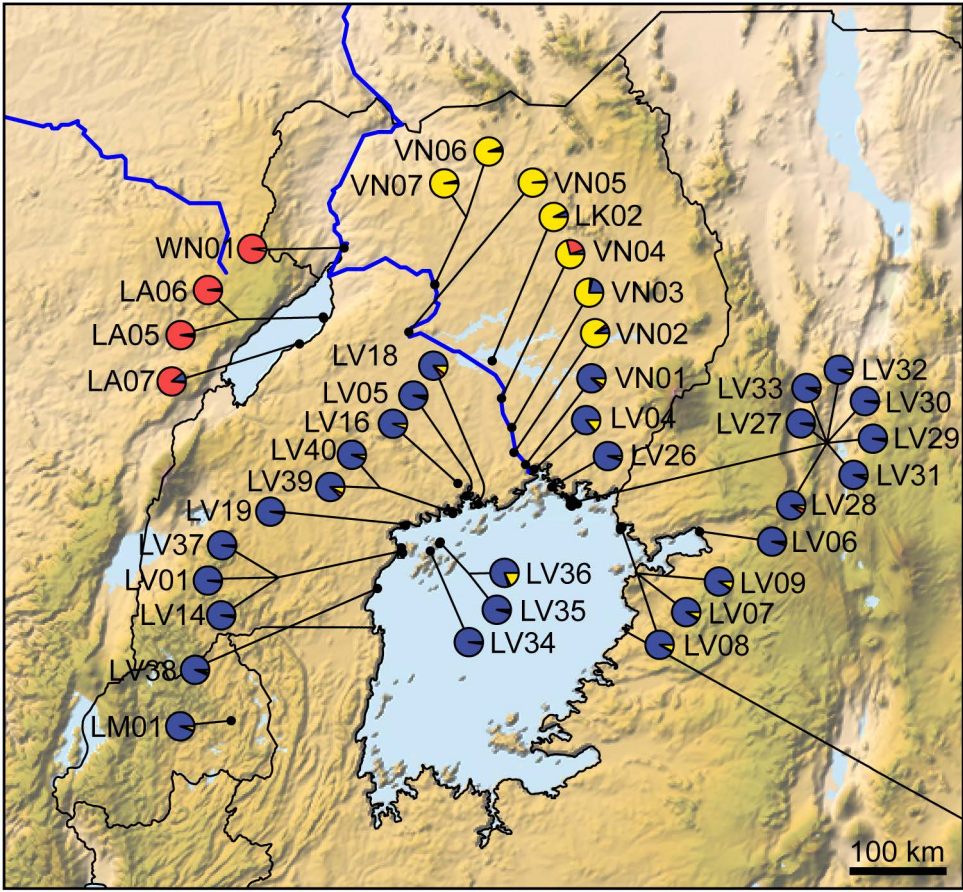


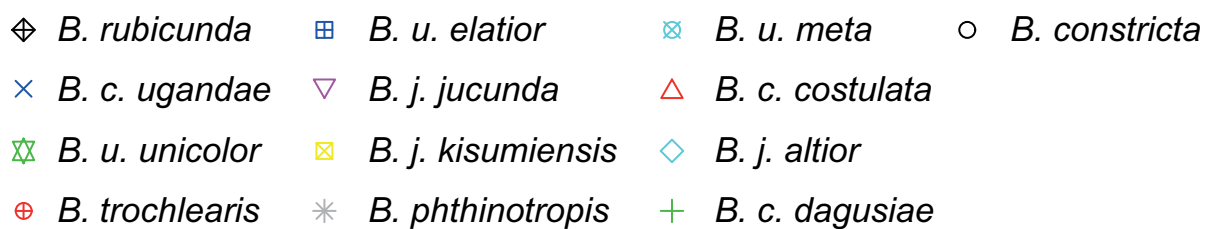
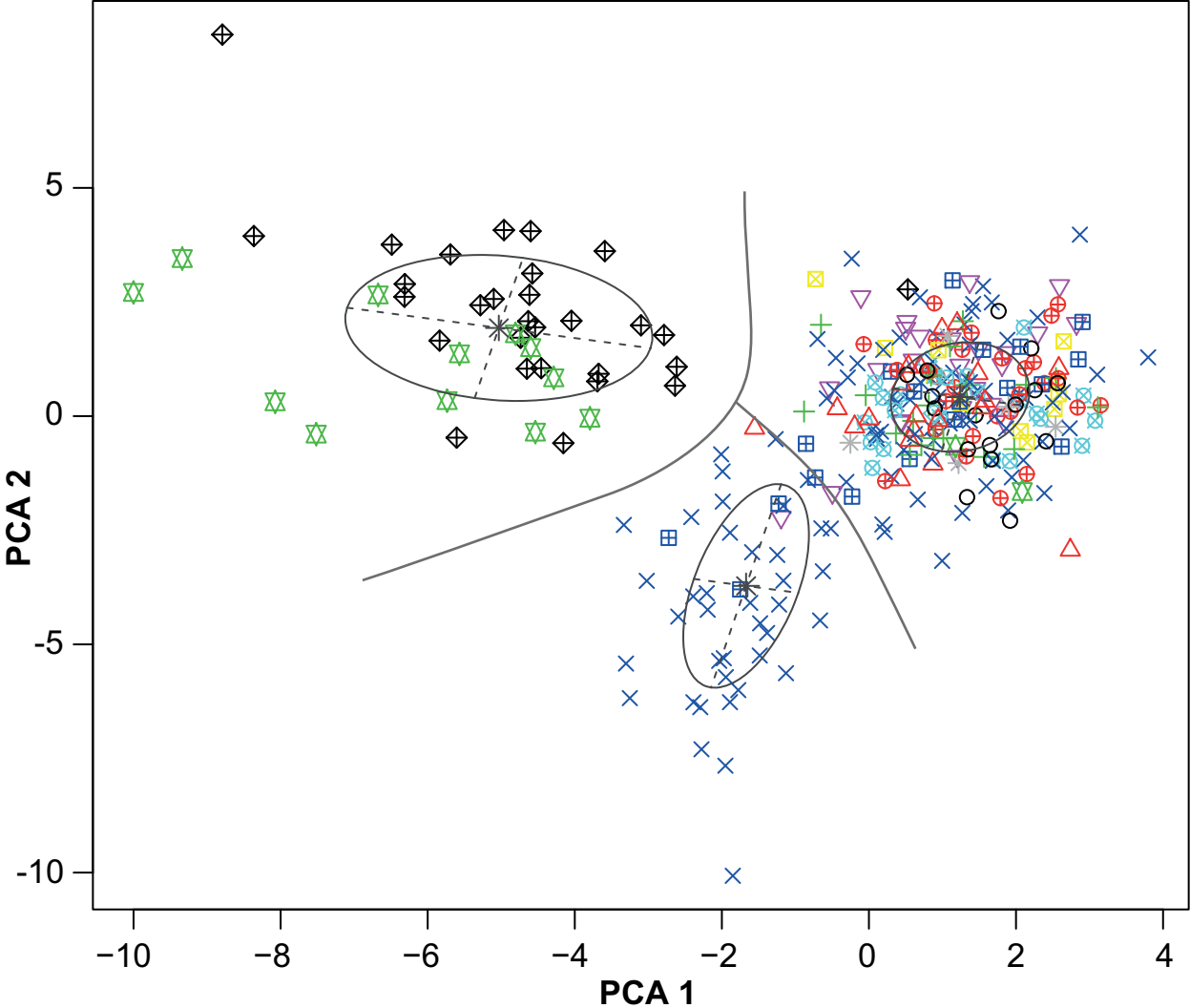


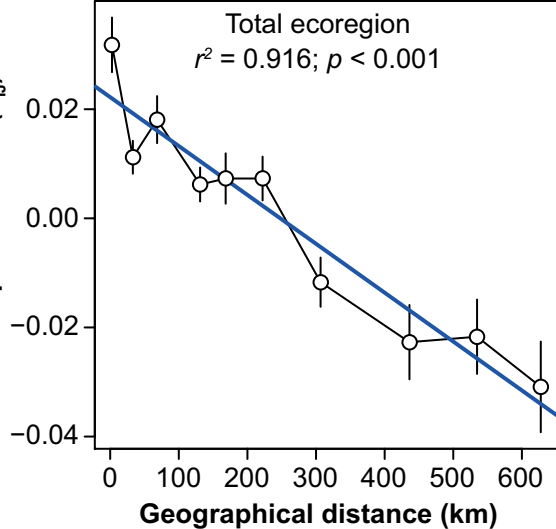
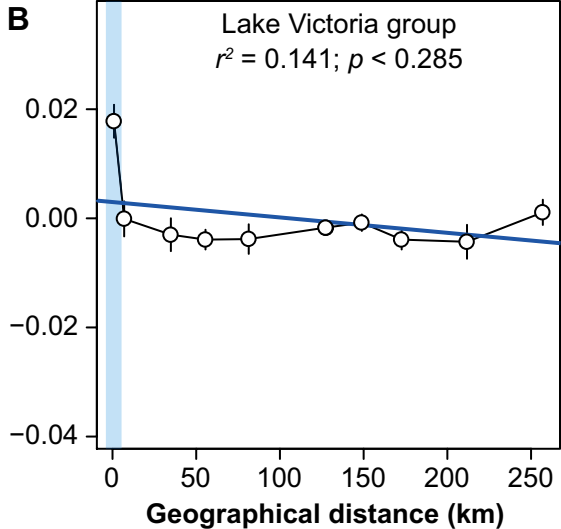
Conditions:

- $Ne_{small} < Ne_{large}$
- $Ne_{l_anc} > Ne_{s_anc}$
- $Ne_{l_anc} > Ne_{small}$
- $Ne_{s_anc} < Ne_{large}$
- $t_2 > t_1$







AKinship coefficient (F_{IJ})**B**

Supplemental Information for:

Population collapse in viviparid gastropods of the Lake Victoria ecoregion started before the Last Glacial Maximum

Bert Van Bocxlaer, Catharina Clewing, Anne Duputié, Camille Roux, Christian Albrecht

SI text

Morphological identifications

Morphological identifications to species and/or races were performed with the identification key of Mandahl-Barth (1954) and these identifications were subsequently updated following Mandahl-Barth (1973) and Brown (1994). Because the identification key has some ambiguous steps, we additionally cross-validated our identifications using type specimens from the Natural History Museum in London and the Zoological Museum, Copenhagen, which houses the Mandahl-Barth collection. As multiple morphological transitions exist, the status of geographical races, nominal species and the assignment of races to species requires further taxonomic research. Nevertheless, all currently recognized races and species were represented in our total dataset, although some morphological variants of *B. trochlearis*, *B. phthinotropis* and *B. unicolor* are missing. Construction of the reduced dataset from the total dataset did not result in the exclusion of morphospecies, but it implied that *B. phthinotropis* is only represented by three specimens. This regrouping of the populations also implied three cases of lumping several morphospecies into joint populations. More specifically, it affects specimens of *B. constricta*, *B. costulata costulata* and *B. trochlearis* in the merge of LV01, LV14 and LV37; specimens of *B. costulata ugandae*, *B. jucunda altior*, *B. jucunda jucunda*, and *B. unicolor elatior* in the merge of LV29 and LV30; and specimens of *B. costulata costulata*, *B. jucunda altior*, *B. cf. phthinotropis*, *B. costulata ugandae*, *B. sp.* (juvenile) in the merge of LV35 and LV36. In none of these cases, however, did the merge result in drastically altered summary statistics.

References

- Brown, D. S. (1994). *Freshwater snails of Africa and their medical importance. Second edition.* London: Taylor & Francis.
- Jørgensen, A., Kristensen, T. K., & Madsen, H. (2008). A molecular phylogeny of apple snails (Gastropoda, Caenogastropoda, Ampullariidae) with an emphasis on African species. *Zoologica Scripta*, 37, 245-252.
- Mandahl-Barth, G. (1954). The freshwater mollusks of Uganda and adjacent territories. *Annals of the Royal Museum of Belgian Congo Tervuren, series 8 - Zoological Sciences*, 32, 1-206.
- Mandahl-Barth, G. (1973). *A field guide to African freshwater snails. 2. East African Species.* Charlottenlund, Denmark: Danish Bilharziasis Laboratory.
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89(3), 583-590.
- Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26, 419-420.
- Paradis, E. (2018). Analysis of haplotype networks: The randomized minimum spanning tree method. *Methods in Ecology and Evolution*, 9, 1308-1317.
- Schultheiß, R., Van Bocxlaer, B., Riedel, F., von Rintelen, T., & Albrecht, C. (2014). Disjunct distributions of freshwater snails testify to a central role of the Congo system in shaping biogeographical patterns in Africa. *BMC Evolutionary Biology*, 14, 42.
- Schultheiß, R., Wilke, T., Jørgensen, A., & Albrecht, C. (2011). The birth of an endemic species flock: demographic history of the *Bellamya* group (Gastropoda, Viviparidae) in Lake Malawi. *Biological Journal of the Linnean Society*, 102, 130-143.
- Sengupta, M. E., Kristensen, T. K., Madsen, H., & Jørgensen, A. (2009). Molecular phylogenetic investigation of the Viviparidae (Gastropoda: Caenogastropoda) in the lakes of the Rift Valley area of Africa. *Molecular Phylogenetics and Evolution*, 52, 797-805.

- 53 Weir, B. S., & Cockerham, C. C. (1984). Estimating F -statistics for the analysis of population structure.
54 *Evolution*, 38(6), 1358-1370.
- 55 Wilson, G. A., & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus
56 genotypes. *Genetics*, 163, 1177-1191.
- 57

58 Tables

59 **Table S1.** Sampling and collection information on all specimens in the cytochrome *c* oxidase subunit I
60 dataset: population code (Pop.), specimen code (Ind.), morphological identification, accession number
61 in the Systematics and Biodiversity collection of the Justus Liebig University Giessen (UGSB), the
62 DNA voucher number (DNA#), the NCBI GenBank accession number and sequence code. For each
63 locality sampling information is provided as well. The reference (Ref.) for the sequence data is: 1 =
64 Jørgensen, Kristensen, and Madsen (2008); 2 = Sengupta, Kristensen, Madsen, and Jørgensen (2009);
65 3 = Schultheiß, Wilke, Jørgensen, and Albrecht (2011); 4 = Schultheiß, Van Bocxlaer, Riedel, von
66 Rintelen, and Albrecht (2014); 5 = this study. Identifications of specimens from Schultheiß et al. (2014)
67 were updated: Specimens indicated with an * under identification were previously identified as *B.*
68 *unicolor*, that with ** as *B. trochlearis*. Individuals with bold specimen codes were included in the
69 microsatellite dataset.

Pop.	Ind.	Identification	UGSB	DNA#	GenBank	Seq_code	Lat.	Long.	Waterbody	Country	Locality	Ref.
LV01	LV01/1	<i>B. trochlearis</i>	7933	n.a.	FJ405837	Btr837vi	-0.29431	32.02575	Lake Victoria	Uganda	Bukakata	2
LV02	LV02/1	<i>B. costulata/jucunda/unicolor</i>	n.a.	n.a.	FJ405872	Bsp872vi	0.48827	33.38524	Lake Victoria	Uganda	Buluba	2
LV03	LV03/1	<i>B. costulata/jucunda/unicolor</i>	n.a.	n.a.	FJ405862	Bsp862vi	0.05388	32.47012	Lake Victoria	Uganda	Entebbe	2
LV03	LV03/2	<i>B. costulata/jucunda/unicolor</i>	n.a.	n.a.	FJ405866	Bsp866vi	0.05388	32.47012	Lake Victoria	Uganda	Entebbe	2
LV04	LV04/1	<i>B. unicolor elatior</i>	7935	n.a.	FJ405848	Bun848vi	0.44008	33.23195	Lake Victoria	Uganda	Jinja	2
LV04	LV04/2	<i>B. unicolor elatior</i>	n.a.	n.a.	FJ405853	Bun853vi	0.44008	33.23195	Lake Victoria	Uganda	Jinja	2
LV05	LV05/1	<i>B. costulata/jucunda/unicolor</i>	7928	n.a.	FJ405849	Bsp849vi	0.19892	32.61540	Lake Victoria	Uganda	Kaazi	2
LV08	LV08/1	<i>B. costulata ugandae</i>	7064	15686	JX489237	Bco237vi	-0.10900	34.02031	Lake Victoria	Kenya	Usenge	2
LV08	LV08/2	<i>B. costulata ugandae</i>	7063	15685	MT834851	Bsp685V1	-0.10900	34.02031	Lake Victoria	Kenya	Usenge	5
LV10	LV10/1	<i>B. costulata/jucunda/unicolor</i>	n.a.	n.a.	FJ405850	Bsp850vi	0.19364	32.85216	Lake Victoria	Uganda	Kibanga Port	2
LV11	LV11/1	<i>B. costulata/jucunda/unicolor</i>	n.a.	n.a.	FJ405869	Bsp869vi	0.32119	33.50837	Lake Victoria	Uganda	Kityerea	2
LV12	LV12/1	<i>B. costulata/jucunda/unicolor</i>	n.a.	n.a.	FJ405868	Bsp868vi	0.35247	32.75374	Lake Victoria	Uganda	Kivindi	2
LV12	LV12/2	<i>B. trochlearis</i>	n.a.	n.a.	FJ405870	Btr870vi	0.35247	32.75374	Lake Victoria	Uganda	Kivindi	2
LV14	LV14/1	<i>B. trochlearis</i>	7930	n.a.	FJ405846	Bco846vi	-0.32093	32.03923	Lake Victoria	Uganda	Lambu	2
LV15	LV15/1	<i>B. sp.</i>	n.a.	n.a.	FJ405854	Bsp854vi	n.a.	n.a.	Lake Victoria	Uganda	Lomba	2
LV16	LV16/1	<i>B. trochlearis</i>	7934	n.a.	FJ405847	Btr847vi	0.31341	32.54089	Lake Victoria	Uganda	Lugala	2
LV17	LV17/1	<i>B. costulata/jucunda/unicolor</i>	n.a.	n.a.	FJ405852	Bsp852vi	-0.43713	34.20606	Lake Victoria	Uganda	Mbita	2
LV18	LV18/1	<i>B. costulata/jucunda/unicolor</i>	7929	n.a.	FJ405839	Bsp839vi	0.12303	32.72698	Lake Victoria	Uganda	MPunge	2
LV19	LV19/1	<i>B. trochlearis</i>	7932	n.a.	FJ405855	Btr855vi	-0.06967	32.06489	Lake Victoria	Uganda	Ngombe	2
LV20	LV20/1	<i>B. costulata</i>	n.a.	7125	HQ012712	Bco712vi	-2.26658	33.79654	Lake Victoria	Tanzania	Speke Bay	3
LV21	LV21/1	<i>B. costulata/jucunda/unicolor</i>	n.a.	n.a.	FJ405845	Bsp845vi	n.a.	n.a.	Lake Victoria	Uganda	Nile Source	2
LV22	LV22/1	<i>B. costulata costulata</i>	13203	20330	MT834852	Bsp330V1	-2.52111	32.89134	Lake Victoria	Tanzania	Mwanza	5
LV23	LV23/1	<i>B. costulata costulata</i>	13204	20331	MT834853	Bsp331V1	-1.33525	31.82467	Lake Victoria	Tanzania	Bukoba	5
LV24	LV24/1	<i>B. constricta</i>	13206	20333	MT834854	Bsp333V1	-1.33076	31.83954	Lake Victoria	Tanzania	Bukoba	5
LV24	LV24/1	<i>B. sp. (juvenile)</i>	13207	20334	MT834855	Bsp334V1	-1.33076	31.83954	Lake Victoria	Tanzania	Bukoba	5
LV25	LV25/1	<i>B. costulata costulata</i>	13209	20336	MT834856	Bsp336V1	-1.33605	31.82361	Lake Victoria	Tanzania	Bukoba	5
LV26	LV26/1	<i>B. unicolor meta</i>	6261	15170	MT834857	Bsp170V1	0.28161	33.39454	Lake Victoria	Uganda	Lufu Island	5
LV26	LV26/2	<i>B. unicolor meta</i>	6262	15171	MT834858	Bsp171V1	0.28161	33.39454	Lake Victoria	Uganda	Lufu Island	5
LV26	LV26/3	<i>B. unicolor meta</i>	6264	15173	MT834859	Bsp173V1	0.28161	33.39454	Lake Victoria	Uganda	Lufu Island	5
LV30	LV30/1	<i>B. unicolor elatior</i>	6518	15261	JX489229	Bun229vi	0.14067	33.60258	Lake Victoria	Uganda	Simu Island	4
LV30	LV30/2	<i>B. costulata ugandae</i>	6517	15260	MT834860	Bsp260V1	0.14067	33.60258	Lake Victoria	Uganda	Simu Island	5
LV32	LV32/1	<i>B. trochlearis</i>	6283	15192	JX489227	Btr227vi	0.16176	33.58264	Lake Victoria	Uganda	Dagusi Island NE	4
LV32	LV32/2	<i>B. trochlearis</i>	6284	15193	JX489228	Btr228vi	0.16176	33.58264	Lake Victoria	Uganda	Dagusi Island NE	4
LV38	LV38/1	<i>B. jucunda jucunda</i>	6578	15321	JX489231	Bju231vi	-0.64564	31.80791	Lake Victoria	Uganda	Ddimio	4
LV38	LV38/2	<i>B. jucunda jucunda</i>	6577	15320	MT834861	Bsp320V1	-0.64564	31.80791	Lake Victoria	Uganda	Ddimio	5
LV39	LV39/1	<i>B. unicolor elatior**</i>	6889	15494	JX489235	Btr235vi	0.03892	32.49402	Lake Victoria	Uganda	Entebbe S	4
LV39	LV39/2	<i>B. unicolor elatior</i>	6890	15495	JX489236	Bun236vi	0.03892	32.49402	Lake Victoria	Uganda	Entebbe S	4
LM01	LM01/1	<i>B. unicolor unicolor</i>	7123	15745	JX489240	Bun240mu	-1.84843	30.47826	Lake Muhazi	Rwanda	Gahini	4
VN01	VN01/6	<i>B. costulata ugandae</i>	6253	15162	MT834862	Bun162NI	0.48544	33.15524	Victoria Nile	Uganda	Bujagali Falls	5
VN01	VN01/9	<i>B. costulata ugandae*</i>	6256	15165	JX489226	Bun226ni	0.48544	33.15524	Victoria Nile	Uganda	Bujagali Falls	4
VN02	VN02/1	<i>B. unicolor elatior</i>	7931	n.a.	FJ405851	Bun851vi	0.59448	33.05258	Victoria Nile	Uganda	Kumucanga	2
VN04	VN04/1	<i>B. costulata ugandae*</i>	7085	15707	JX489238	Bun238ni	1.09002	32.93648	Victoria Nile	Uganda	Budali	4
LK01	LK01/1	<i>B. unicolor</i>	n.a.	n.a.	FJ405863	Bun863ky	n.a.	n.a.	Lake Kyoga	Uganda	Lyingo	2
VN05	VN05/1	<i>B. costulata ugandae*</i>	7111	15733	JX489239	Bun239ni	1.69255	32.09473	Victoria Nile	Uganda	Masindi Port	4
VN06	VN06/1	<i>B. costulata ugandae*</i>	7124	15746	JX489241	Bun241ni	2.12373	32.33107	Victoria Nile	Uganda	Mutunda	4
VN07	VN07/1	<i>B. costulata ugandae</i>	6692	15458	MT834863	Bsp458NI	2.12852	32.32919	Victoria Nile	Uganda	Mutunda	5
VN07	VN07/2	<i>B. costulata ugandae</i>	6695	15461	MT834864	Bsp461NI	2.12852	32.32919	Victoria Nile	Uganda	Mutunda	5
LA01	LA01/1	<i>B. rubicunda</i>	n.a.	6612	HQ012714	Bru714al	1.42904	30.89265	Lake Albert	Uganda	Kiabahamba	3
LA01	LA01/2	<i>B. rubicunda</i>	n.a.	6611	HQ012713	Bru713al	1.42904	30.89265	Lake Albert	Uganda	Kiabahamba	3
LA01	LA01/3	<i>B. rubicunda</i>	n.a.	6615	HQ012715	Bru715al	1.42904	30.89265	Lake Albert	Uganda	Kiabahamba	3
LA02	LA02/1	<i>B. rubicunda</i>	n.a.	n.a.	FJ405861	Bru861al	1.90188	31.40362	Lake Albert	Uganda	Bugoigo	2
LA03	LA03/1	<i>B. rubicunda</i>	n.a.	n.a.	FJ405836	Bru836al	1.80911	31.32299	Lake Albert	Uganda	Butiaba	2
LA04	LA04/1	<i>B. rubicunda</i>	n.a.	n.a.	EU274556	Bru556al	n.a.	n.a.	Lake Albert	Uganda	Rwangala	1

MOLECULAR ECOLOGY

Pop.	Ind.	Identification	UGSB	DNA#	GenBank	Seq_code	Lat.	Long.	Waterbody	Country	Locality	Ref.
LA05	LA05/1	<i>B. rubicunda</i>	6303	15212	MT834865	Bru212AL	1.83060	31.31484	Lake Albert	Uganda	Butiaba	5
LA06	LA06/1	<i>B. rubicunda</i>	6877	15482	JX489233	Bru233al	1.81842	31.32021	Lake Albert	Uganda	Butiaba	4
LA06	LA06/2	<i>B. rubicunda</i>	6878	15483	JX489234	Bru234al	1.81842	31.32021	Lake Albert	Uganda	Butiaba	4
LA07	LA07/1	<i>B. rubicunda</i>	6569	15312	JX489230	Bru230al	1.58649	31.10057	Lake Albert	Uganda	Kiriamboga	4
WN01	RN07/1	<i>B. unicolor unicolor</i>	6710	15476	JX489232	Bun232ni	2.45933	31.50487	White Nile	Uganda	Pakwach	4
WN01	RN07/2	<i>B. unicolor unicolor</i>	6711	15477	MT834866	Bsp477NI	2.45933	31.50487	White Nile	Uganda	Pakwach	5
WN01	RN07/3	<i>B. unicolor unicolor</i>	6712	15478	MT834867	Bsp478NI	2.45933	31.50487	White Nile	Uganda	Pakwach	5

70

71

Table S2. Sampling and collection information on all individuals included in our microsatellite dataset: population code (Pop.), specimen code (Ind.), morphological identification, accession number in the Systematics and Biodiversity collection of the Justus Liebig University Giessen (UGSB), the DNA voucher number (DNA#), and locality information. Information specific to the sampling event is identical for all specimens in a population and, therefore, it is only indicated on the first row for each population. Populations that have been pooled in the regrouped dataset are indicated with identical locality names, however, their geographic coordinates specify precise sampling locations.

Pop.	Ind.	Identification	UGSB	DNA#	Lat.	Long.	Waterbody	Country	Locality
LA05	LA05/1	<i>B. rubicunda</i>	6303	15212	1.83060	31.31484	Lake Albert	Uganda	Butiaba
	LA05/2	<i>B. rubicunda</i>	6295	15204					
	LA05/3	<i>B. rubicunda</i>	6296	15205					
	LA05/4	<i>B. rubicunda</i>	6297	15206					
	LA05/5	<i>B. rubicunda</i>	6298	15207					
	LA05/6	<i>B. rubicunda</i>	6299	15208					
	LA05/7	<i>B. rubicunda</i>	6301	15210					
	LA05/8	<i>B. rubicunda</i>	6302	15211					
	LA05/9	<i>B. rubicunda</i>	6305	15214					
	LA05/10	<i>B. rubicunda</i>	6306	15215					
LA06	LA06/1	<i>B. rubicunda</i>	6877	15482	1.81842	31.32021	Lake Albert	Uganda	Butiaba
	LA06/2	<i>B. rubicunda</i>	6878	15483					
	LA06/3	<i>B. rubicunda</i>	6879	15484					
	LA06/4	<i>B. rubicunda</i>	6880	15485					
	LA06/5	<i>B. rubicunda</i>	6881	15486					
	LA06/6	<i>B. rubicunda</i>	6882	15487					
	LA06/7	<i>B. rubicunda</i>	6884	15489					
	LA06/8	<i>B. rubicunda</i>	6886	15491					
	LA06/9	<i>B. rubicunda</i>	6887	15492					
	LA06/10	<i>B. rubicunda</i>	6888	15493					
LA07	LA07/1	<i>B. rubicunda</i>	6569	15312	1.58649	31.10057	Lake Albert	Uganda	Kiriamboga
	LA07/2	<i>B. rubicunda</i>	6567	15310					
	LA07/3	<i>B. rubicunda</i>	6568	15311					
	LA07/4	<i>B. rubicunda</i>	6570	15313					
	LA07/5	<i>B. rubicunda</i>	6571	15314					
	LA07/6	<i>B. rubicunda</i>	6572	15315					
	LA07/7	<i>B. rubicunda</i>	6573	15316					
	LA07/8	<i>B. rubicunda</i>	6574	15317					
	LA07/9	<i>B. rubicunda</i>	6575	15318					
	LA07/10	<i>B. rubicunda</i>	6576	15319					
LK02	LK02/1	<i>B. costulata ugandae</i>	7122	15744	1.43136	32.85360	Lake Kyoga	Uganda	Bukungu
LM01	LM01/1	<i>B. unicolor unicolor</i>	7123	15745	-1.84843	30.47826	Lake Muhazi	Rwanda	Gahini
LV01	LV01/2	<i>B. trochlearis</i>	7961	16549	-0.29431	32.02575	Lake Victoria	Uganda	Bukakata
	LV01/3	<i>B. trochlearis</i>	7962	16550					
	LV01/4	<i>B. trochlearis</i>	7963	16551					
	LV01/5	<i>B. trochlearis</i>	7964	16552					
	LV01/6	<i>B. trochlearis</i>	7965	16553					
	LV01/7	<i>B. trochlearis</i>	7966	16554					
LV04	LV04/4	<i>B. unicolor elatior</i>	7972	16560	0.44008	33.23195	Lake Victoria	Uganda	Jinja
	LV04/5	<i>B. jucunda jucunda</i>	7973	16561					
	LV04/6	<i>B. jucunda jucunda</i>	7974	16562					
	LV04/7	<i>B. jucunda jucunda</i>	7975	16563					
LV05	LV05/2	<i>B. jucunda jucunda</i>	7936	16524	0.19892	32.61540	Lake Victoria	Uganda	Kaazi
	LV05/3	<i>B. jucunda jucunda</i>	7937	16525					
	LV05/4	<i>B. jucunda jucunda</i>	7938	16526					
	LV05/5	<i>B. jucunda jucunda</i>	7939	16527					
	LV05/6	<i>B. jucunda jucunda</i>	7940	16528					
	LV05/7	<i>B. jucunda jucunda</i>	7941	16529					
LV06	LV06/1	<i>B. jucunda kisumiensis</i>	6493	15236	-0.12739	34.74232	Lake Victoria	Kenya	Kisumu
	LV06/2	<i>B. jucunda kisumiensis</i>	6494	15237					
	LV06/3	<i>B. jucunda kisumiensis</i>	6495	15238					
	LV06/4	<i>B. jucunda kisumiensis</i>	6497	15240					
	LV06/5	<i>B. jucunda kisumiensis</i>	6498	15241					
	LV06/6	<i>B. jucunda kisumiensis</i>	6499	15242					
	LV06/7	<i>B. jucunda kisumiensis</i>	6500	15243					
	LV06/8	<i>B. jucunda kisumiensis</i>	6502	15245					
	LV06/9	<i>B. jucunda kisumiensis</i>	6503	15246					
	LV06/10	<i>B. jucunda kisumiensis</i>	6504	15247					
LV07	LV07/1	<i>B. costulata ugandae</i>	6505	15248	-0.09034	34.02106	Lake Victoria	Kenya	Usenge
	LV07/2	<i>B. costulata ugandae</i>	6506	15249					
	LV07/3	<i>B. costulata ugandae</i>	6507	15250					
	LV07/4	<i>B. costulata ugandae</i>	6508	15251					
	LV07/5	<i>B. costulata ugandae</i>	6509	15252					

Pop.	Ind.	Identification	UGSB	DNA#	Lat.	Long.	Waterbody	Country	Locality
	LV07/6	<i>B. costulata ugandae</i>	6510	15253					
	LV07/7	<i>B. costulata ugandae</i>	6511	15254					
	LV07/8	<i>B. costulata ugandae</i>	6512	15255					
	LV07/9	<i>B. costulata ugandae</i>	6513	15256					
	LV07/10	<i>B. costulata ugandae</i>	6514	15257					
	LV07/11	<i>B. costulata ugandae</i>	6515	15258					
	LV07/12	<i>B. costulata ugandae</i>	6516	15259					
LV08	LV08/1	<i>B. costulata ugandae</i>	7064	15686	-0.10900	34.02031	Lake Victoria	Kenya	Usenge
	LV08/2	<i>B. costulata ugandae</i>	7063	15685					
	LV08/3	<i>B. costulata ugandae</i>	7065	15687					
	LV08/4	<i>B. costulata ugandae</i>	7066	15688					
	LV08/5	<i>B. costulata ugandae</i>	7068	15690					
LV09	LV09/1	<i>B. costulata ugandae</i>	7069	15691	-0.08075	34.03286	Lake Victoria	Kenya	Usenge
	LV09/2	<i>B. costulata ugandae</i>	7070	15692					
	LV09/3	<i>B. costulata ugandae</i>	7071	15693					
	LV09/4	<i>B. costulata ugandae</i>	7072	15694					
	LV09/5	<i>B. costulata ugandae</i>	7073	15695					
	LV09/6	<i>B. costulata ugandae</i>	7074	15696					
	LV09/7	<i>B. costulata ugandae</i>	7075	15697					
	LV09/8	<i>B. costulata ugandae</i>	7076	15698					
	LV09/9	<i>B. costulata ugandae</i>	7077	15699					
LV14	LV14/2	<i>B. costulata costulata</i>	7946	16534	-0.32093	32.03923	Lake Victoria	Uganda	Bukakata
	LV14/3	<i>B. costulata costulata</i>	7947	16535					
	LV14/4	<i>B. costulata costulata</i>	7948	16536					
	LV14/5	<i>B. costulata costulata</i>	7949	16537					
	LV14/6	<i>B. costulata costulata</i>	7950	16538					
LV16	LV16/2	<i>B. unicolor elatior</i>	7966	16554	0.31341	32.54089	Lake Victoria	Uganda	Lugala
	LV16/3	<i>B. unicolor elatior</i>	7967	16555					
	LV16/4	<i>B. costulata ugandae</i>	7968	16556					
	LV16/5	<i>B. costulata ugandae</i>	7969	16557					
	LV16/6	<i>B. costulata ugandae</i>	7970	16558					
LV18	LV18/2	<i>B. jucunda jucunda</i>	7941	16529	0.12303	32.72698	Lake Victoria	Uganda	Mpunge
	LV18/3	<i>B. jucunda jucunda</i>	7942	16530					
	LV18/4	<i>B. jucunda jucunda</i>	7943	16531					
	LV18/5	<i>B. jucunda jucunda</i>	7944	16532					
	LV18/6	<i>B. jucunda jucunda</i>	7945	16533					
LV19	LV19/2	<i>B. phthinotropis</i>	7956	16544	-0.06967	32.06489	Lake Victoria	Uganda	Ngombe
	LV19/3	<i>B. phthinotropis</i>	7957	16545					
	LV19/4	<i>B. trochlearis</i>	7958	16546					
	LV19/5	<i>B. trochlearis</i>	7959	16547					
	LV19/6	<i>B. trochlearis</i>	7960	16548					
LV26	LV26/1	<i>B. unicolor meta</i>	6261	15170	0.28161	33.39454	Lake Victoria	Uganda	Lufu Island
	LV26/2	<i>B. unicolor meta</i>	6262	15171					
	LV26/3	<i>B. unicolor meta</i>	6264	15173					
	LV26/4	<i>B. unicolor meta</i>	6260	15169					
	LV26/5	<i>B. unicolor meta</i>	6263	15172					
	LV26/6	<i>B. unicolor meta</i>	6265	15174					
	LV26/7	<i>B. unicolor meta</i>	6266	15175					
	LV26/8	<i>B. unicolor meta</i>	6267	15176					
	LV26/9	<i>B. unicolor meta</i>	6268	15177					
	LV26/10	<i>B. unicolor meta</i>	6270	15179					
LV27	LV27/1	<i>B. costulata costulata</i>	6541	15284	0.11627	33.56262	Lake Victoria	Uganda	Dagusi Island SW
	LV27/2	<i>B. costulata costulata</i>	6542	15285					
	LV27/3	<i>B. costulata costulata</i>	6543	15286					
	LV27/4	<i>B. unicolor elatior</i>	6544	15287					
	LV27/5	<i>B. unicolor meta</i>	6545	15288					
	LV27/6	<i>B. costulata ugandae</i>	6546	15289					
	LV27/7	<i>B. costulata ugandae</i>	6547	15290					
	LV27/8	<i>B. costulata ugandae</i>	6549	15292					
	LV27/9	<i>B. unicolor elatior</i>	6551	15294					
	LV27/10	<i>B. unicolor elatior</i>	6552	15295					
LV28	LV28/1	<i>B. costulata ugandae</i>	6271	15180	0.10460	33.57583	Lake Victoria	Uganda	Dagusi Island S
	LV28/2	<i>B. costulata ugandae</i>	6272	15181					
	LV28/3	<i>B. costulata ugandae</i>	6274	15183					
	LV28/4	<i>B. costulata ugandae</i>	6275	15184					
	LV28/5	<i>B. costulata ugandae</i>	6276	15185					
	LV28/6	<i>B. costulata ugandae</i>	6277	15186					
	LV28/7	<i>B. costulata ugandae</i>	6280	15189					
	LV28/8	<i>B. costulata ugandae</i>	6282	15191					

Pop.	Ind.	Identification	UGSB	DNA#	Lat.	Long.	Waterbody	Country	Locality
	LV28/9	<i>B. sp.</i> (juvenile)	6553	15296					
	LV28/10	<i>B. sp.</i> (juvenile)	6554	15297					
	LV28/11	<i>B. sp.</i> (juvenile)	6555	15298					
	LV28/12	<i>B. sp.</i> (juvenile)	6556	15299					
	LV28/13	<i>B. sp.</i> (juvenile)	6557	15300					
	LV28/14	<i>B. sp.</i> (juvenile)	6558	15301					
	LV28/15	<i>B. sp.</i> (juvenile)	6559	15302					
	LV28/16	<i>B. sp.</i> (juvenile)	6560	15303					
	LV28/17	<i>B. sp.</i> (juvenile)	6561	15304					
	LV28/18	<i>B. sp.</i> (juvenile)	6562	15305					
LV29	LV29/1	<i>B. jucunda jucunda</i>	6644	15410	0.13707	33.60149	Lake Victoria	Uganda	Simu Island
	LV29/2	<i>B. jucunda altior</i>	6645	15411					
	LV29/3	<i>B. jucunda jucunda</i>	6646	15412					
	LV29/4	<i>B. jucunda jucunda</i>	6647	15413					
	LV29/5	<i>B. jucunda jucunda</i>	6649	15415					
	LV29/6	<i>B. jucunda jucunda</i>	6650	15416					
	LV29/7	<i>B. jucunda jucunda</i>	6651	15417					
	LV29/8	<i>B. jucunda jucunda</i>	6653	15419					
	LV29/9	<i>B. jucunda jucunda</i>	6654	15420					
	LV29/10	<i>B. jucunda jucunda</i>	6655	15421					
LV30	LV30/1	<i>B. unicolor elatior</i>	6518	15261	0.14067	33.60258	Lake Victoria	Uganda	Simu Island
	LV30/2	<i>B. costulata ugandae</i>	6517	15260					
	LV30/3	<i>B. costulata ugandae</i>	6519	15262					
	LV30/4	<i>B. unicolor elatior</i>	6520	15263					
	LV30/7	<i>B. costulata ugandae</i>	6525	15268					
	LV30/8	<i>B. unicolor elatior</i>	6526	15269					
	LV30/9	<i>B. unicolor elatior</i>	6527	15270					
	LV30/10	<i>B. unicolor elatior</i>	6528	15271					
	LV31/1	<i>B. unicolor meta</i>	6656	15422	0.12456	33.62969	Lake Victoria	Uganda	Kaza Island E
	LV31/2	<i>B. unicolor meta</i>	6657	15423					
	LV31/3	<i>B. unicolor meta</i>	6658	15424					
	LV31/4	<i>B. unicolor meta</i>	6659	15425					
	LV31/5	<i>B. unicolor meta</i>	6660	15426					
	LV31/6	<i>B. unicolor meta</i>	6661	15427					
	LV31/7	<i>B. unicolor meta</i>	6662	15428					
	LV31/8	<i>B. unicolor meta</i>	6663	15429					
	LV31/9	<i>B. unicolor meta</i>	6665	15431					
	LV31/10	<i>B. unicolor meta</i>	6666	15432					
LV32	LV31/11	<i>B. unicolor meta</i>	6667	15433	0.16176	33.58264	Lake Victoria	Uganda	Dagusi Island NE
	LV32/1	<i>B. trochlearis</i>	6283	15192					
	LV32/2	<i>B. trochlearis</i>	6284	15193					
	LV32/3	<i>B. trochlearis</i>	6285	15194					
	LV32/4	<i>B. trochlearis</i>	6286	15195					
	LV32/5	<i>B. trochlearis</i>	6287	15196					
	LV32/6	<i>B. trochlearis</i>	6289	15198					
	LV32/7	<i>B. trochlearis</i>	6290	15199					
	LV32/8	<i>B. trochlearis</i>	6292	15201					
	LV32/9	<i>B. trochlearis</i>	6293	15202					
LV33	LV32/10	<i>B. trochlearis</i>	6294	15203	0.15662	33.56414	Lake Victoria	Uganda	Dagusi Island NW
	LV33/1	<i>B. costulata dagusiae</i>	6668	15434					
	LV33/2	<i>B. costulata dagusiae</i>	6669	15435					
	LV33/3	<i>B. costulata dagusiae</i>	6671	15437					
	LV33/4	<i>B. costulata dagusiae</i>	6672	15438					
	LV33/5	<i>B. costulata dagusiae</i>	6673	15439					
	LV33/6	<i>B. costulata dagusiae</i>	6674	15440					
	LV33/7	<i>B. costulata dagusiae</i>	6675	15441					
	LV33/8	<i>B. costulata dagusiae</i>	6676	15442					
	LV33/9	<i>B. costulata dagusiae</i>	6678	15444					
	LV33/10	<i>B. costulata dagusiae</i>	6679	15445					
	LV33/11	<i>B. costulata dagusiae</i>	6680	15446					
	LV33/12	<i>B. costulata dagusiae</i>	6681	15447					
	LV33/13	<i>B. costulata dagusiae</i>	6682	15448					
	LV33/14	<i>B. costulata dagusiae</i>	6683	15449					
	LV33/15	<i>B. costulata dagusiae</i>	6685	15451					
	LV33/16	<i>B. costulata dagusiae</i>	6687	15453					
	LV33/17	<i>B. costulata dagusiae</i>	6688	15454					
	LV33/18	<i>B. costulata dagusiae</i>	6689	15455					
	LV33/19	<i>B. costulata dagusiae</i>	6690	15456					
	LV33/20	<i>B. costulata dagusiae</i>	6691	15457					

Pop.	Ind.	Identification	UGSB	DNA#	Lat.	Long.	Waterbody	Country	Locality
LV34	LV34/1	<i>B. trochlearis</i>	6307	15216	-0.30371	32.28927	Lake Victoria	Uganda	Kalangala
	LV34/2	<i>B. trochlearis</i>	6308	15217					
	LV34/3	<i>B. trochlearis</i>	6309	15218					
	LV34/4	<i>B. trochlearis</i>	6310	15219					
	LV34/5	<i>B. trochlearis</i>	6311	15220					
	LV34/6	<i>B. trochlearis</i>	6312	15221					
	LV34/7	<i>B. trochlearis</i>	6313	15222					
	LV34/8	<i>B. trochlearis</i>	6315	15224					
	LV34/9	<i>B. trochlearis</i>	6317	15226					
	LV34/10	<i>B. trochlearis</i>	6318	15227					
LV35	LV35/1	<i>B. cf. phthinotropis</i>	7091	15713	-0.23731	32.37333	Lake Victoria	Uganda	Serinya Island
	LV35/2	<i>B. cf. phthinotropis</i>	7092	15714					
	LV35/3	<i>B. cf. phthinotropis</i>	7093	15715					
	LV35/4	<i>B. costulata costulata</i>	7094	15716					
	LV35/5	<i>B. jucunda altior</i>	7095	15717					
	LV35/6	<i>B. costulata costulata</i>	7096	15718					
LV36	LV36/1	<i>B. sp. (juvenile)</i>	7097	15719	-0.22237	32.37996	Lake Victoria	Uganda	Serinya Island
	LV36/2	<i>B. sp. (juvenile)</i>	7098	15720					
	LV36/3	<i>B. costulata ugandae</i>	7099	15721					
	LV36/4	<i>B. costulata ugandae</i>	7100	15722					
	LV36/5	<i>B. costulata ugandae</i>	7102	15724					
LV37	LV37/1	<i>B. constricta</i>	7103	15725	-0.27263	32.02691	Lake Victoria	Uganda	Bukakata
	LV37/2	<i>B. constricta</i>	7104	15726					
	LV37/3	<i>B. constricta</i>	7105	15727					
	LV37/4	<i>B. constricta</i>	7106	15728					
	LV37/5	<i>B. constricta</i>	7107	15729					
	LV37/6	<i>B. constricta</i>	7108	15730					
	LV37/7	<i>B. constricta</i>	7109	15731					
LV38	LV38/1	<i>B. jucunda jucunda</i>	6578	15321	-0.64564	31.80791	Lake Victoria	Uganda	Ddimo
	LV38/2	<i>B. jucunda jucunda</i>	6577	15320					
	LV38/3	<i>B. costulata costulata</i>	6579	15322					
	LV38/4	<i>B. costulata costulata</i>	6580	15323					
	LV38/5	<i>B. costulata costulata</i>	6581	15324					
	LV38/6	<i>B. costulata costulata</i>	6582	15325					
	LV38/7	<i>B. costulata costulata</i>	6583	15326					
	LV38/8	<i>B. costulata costulata</i>	6585	15328					
	LV38/9	<i>B. jucunda kisumiensis</i>	6496	15239					
	LV38/10	<i>B. costulata costulata</i>	6584	15327					
LV39	LV39/1	<i>B. unicolor elatior</i>	6889	15494	0.03892	32.49402	Lake Victoria	Uganda	Entebbe S
	LV39/2	<i>B. unicolor elatior</i>	6890	15495					
	LV39/3	<i>B. unicolor elatior</i>	6891	15496					
	LV39/4	<i>B. costulata costulata</i>	6892	15497					
	LV39/5	<i>B. costulata costulata</i>	6894	15499					
	LV39/6	<i>B. costulata costulata</i>	6895	15500					
	LV39/7	<i>B. costulata costulata</i>	6896	15501					
	LV39/8	<i>B. costulata costulata</i>	6897	15502					
	LV39/9	<i>B. costulata costulata</i>	6899	15504					
	LV39/10	<i>B. costulata costulata</i>	6900	15505					
	LV39/11	<i>B. costulata costulata</i>	6893	15498					
LV40	LV40/1	<i>B. constricta</i>	6529	15272	0.05525	32.49198	Lake Victoria	Uganda	Entebbe N
	LV40/2	<i>B. constricta</i>	6530	15273					
	LV40/3	<i>B. constricta</i>	6531	15274					
	LV40/4	<i>B. constricta</i>	6532	15275					
	LV40/5	<i>B. constricta</i>	6533	15276					
	LV40/6	<i>B. constricta</i>	6534	15277					
	LV40/7	<i>B. unicolor unicolor</i>	6535	15278					
	LV40/8	<i>B. constricta</i>	6537	15280					
	LV40/9	<i>B. constricta</i>	6539	15282					
	LV40/10	<i>B. constricta</i>	6540	15283					
VN01	VN01/1	<i>B. costulata ugandae</i>	6247	15156	0.48544	33.15524	Victoria Nile	Uganda	Bujagali Falls
	VN01/2	<i>B. costulata ugandae</i>	6249	15158					
	VN01/3	<i>B. costulata ugandae</i>	6250	15159					
	VN01/4	<i>B. costulata ugandae</i>	6251	15160					
	VN01/5	<i>B. costulata ugandae</i>	6252	15161					
	VN01/6	<i>B. costulata ugandae</i>	6253	15162					
	VN01/7	<i>B. costulata ugandae</i>	6254	15163					
	VN01/8	<i>B. costulata ugandae</i>	6255	15164					
	VN01/9	<i>B. costulata ugandae</i>	6256	15165					
	VN01/10	<i>B. costulata ugandae</i>	6257	15166					

Pop.	Ind.	Identification	UGSB	DNA#	Lat.	Long.	Waterbody	Country	Locality
VN02	VN02/2	<i>B. unicolor elatior</i>	7951	16539	0.59448	33.05257	Victoria Nile	Uganda	Kumucanga
	VN02/3	<i>B. jucunda jucunda</i>	7952	16540					
	VN02/4	<i>B. unicolor elatior</i>	7953	16541					
	VN02/5	<i>B. unicolor elatior</i>	7954	16542					
	VN02/6	<i>B. unicolor elatior</i>	7955	16543					
VN03	VN03/1	<i>B. costulata ugandae</i>	7078	15700	0.82298	33.03183	Victoria Nile	Uganda	Bulemuti
	VN03/2	<i>B. costulata ugandae</i>	7079	15701					
	VN03/3	<i>B. costulata ugandae</i>	7080	15702					
	VN03/4	<i>B. unicolor elatior</i>	7081	15703					
	VN03/5	<i>B. costulata ugandae</i>	7082	15704					
	VN03/6	<i>B. costulata ugandae</i>	7083	15705					
VN04	VN04/1	<i>B. costulata ugandae</i>	7085	15707	1.09002	32.93648	Victoria Nile	Uganda	Budali
	VN04/2	<i>B. costulata ugandae</i>	7084	15706					
VN05	VN05/1	<i>B. costulata ugandae</i>	7111	15733	1.69255	32.09473	Victoria Nile	Uganda	Masindi Port
	VN05/2	<i>B. costulata ugandae</i>	7110	15732					
	VN05/3	<i>B. costulata ugandae</i>	7112	15734					
	VN05/4	<i>B. costulata ugandae</i>	7113	15735					
	VN05/5	<i>B. costulata ugandae</i>	7114	15736					
	VN05/6	<i>B. costulata ugandae</i>	7115	15737					
	VN05/7	<i>B. costulata ugandae</i>	7116	15738					
	VN05/8	<i>B. costulata ugandae</i>	7117	15739					
	VN05/9	<i>B. costulata ugandae</i>	7118	15740					
	VN05/10	<i>B. costulata ugandae</i>	7119	15741					
	VN05/11	<i>B. costulata ugandae</i>	7120	15742					
	VN05/12	<i>B. costulata ugandae</i>	7121	15743					
VN06	VN06/1	<i>B. costulata ugandae</i>	7124	15746	2.12373	32.33107	Victoria Nile	Uganda	Mutunda
	VN06/2	<i>B. costulata ugandae</i>	7125	15747					
	VN06/3	<i>B. costulata ugandae</i>	7126	15748					
	VN06/4	<i>B. costulata ugandae</i>	7127	15749					
	VN06/5	<i>B. costulata ugandae</i>	7128	15750					
VN07	VN07/1	<i>B. costulata ugandae</i>	6692	15458	2.12852	32.32919	Victoria Nile	Uganda	Mutunda
	VN07/3	<i>B. costulata ugandae</i>	6693	15459					
	VN07/4	<i>B. costulata ugandae</i>	6694	15460					
	VN07/2	<i>B. costulata ugandae</i>	6695	15461					
	VN07/5	<i>B. costulata ugandae</i>	6696	15462					
	VN07/6	<i>B. costulata ugandae</i>	6697	15463					
	VN07/7	<i>B. costulata ugandae</i>	6698	15464					
	VN07/8	<i>B. costulata ugandae</i>	6699	15465					
	VN07/9	<i>B. costulata ugandae</i>	6700	15466					
	VN07/10	<i>B. costulata ugandae</i>	6701	15467					
	VN07/11	<i>B. costulata ugandae</i>	6702	15468					
	VN07/12	<i>B. costulata ugandae</i>	6703	15469					
WN01	WN01/1	<i>B. unicolor unicolor</i>	6710	15476	2.45933	31.50487	White Nile	Uganda	Pakwach
	WN01/2	<i>B. unicolor unicolor</i>	6711	15477					
	WN01/3	<i>B. unicolor unicolor</i>	6712	15478					
	WN01/4	<i>B. unicolor unicolor</i>	6704	15470					
	WN01/5	<i>B. unicolor unicolor</i>	6705	15471					
	WN01/6	<i>B. unicolor unicolor</i>	6706	15472					
	WN01/7	<i>B. unicolor unicolor</i>	6707	15473					
	WN01/8	<i>B. unicolor unicolor</i>	6708	15474					
	WN01/9	<i>B. unicolor unicolor</i>	6709	15475					
	WN01/10	<i>B. unicolor unicolor</i>	6713	15479					
	WN01/11	<i>B. unicolor unicolor</i>	6714	15480					
	WN01/12	<i>B. unicolor unicolor</i>	6715	15481					

Table S3. Estimates of nuclear genetic variation for our 15 microsatellite loci in the regrouped dataset (21 populations; 282 specimens). Note that loci Bel_L045 and Bel_L050 (in grey) are discarded from the results reported in the main text due to the presence of null alleles. A_n = number of alleles; A_r = allelic richness standardized to a sample size of 5 diploid individuals, H_o = observed heterozygosity; H_E = expected heterozygosity. The mean fixation indices F_{IS} , F_{ST} , F_{IT} are indicated with their standard error following Weir and Cockerham (1984).

Locus	A_n	A_r	H_o	H_E	F_{IS}	s.e. F_{IS}	F_{ST}	s.e. F_{ST}	F_{IT}	s.e. F_{IT}
<i>Bel_L53</i>	19	6.9	0.888	0.880	-0.005	0.020	0.038***	0.009	0.034*	0.022
<i>Bel_L25</i>	23	6.5	0.800	0.859	0.067**	0.034	0.033***	0.010	0.098***	0.032
<i>Bel_L48</i>	17	5.1	0.738	0.727	-0.014	0.027	0.107***	0.032	0.095**	0.046
<i>Bel_L16</i>	18	7.2	0.870	0.861	-0.016	0.020	0.067***	0.025	0.052**	0.026
<i>Bel_L63</i>	15	4.8	0.702	0.670	-0.036	0.036	0.108***	0.032	0.076**	0.050
<i>Bel_L13</i>	10	3.4	0.571	0.602	0.026	0.037	0.023*	0.017	0.048	0.043
<i>Bel_L10</i>	19	6.3	0.755	0.769	0.021	0.027	0.117***	0.035	0.136***	0.045
<i>Bel_L50</i>	24	7.4	0.734	0.883	0.147***	0.034	0.037***	0.010	0.178***	0.036
<i>Bel_L32</i>	30	8.1	0.907	0.924	0.010	0.018	0.031***	0.007	0.040**	0.017
<i>Bel_L23</i>	27	7.6	0.788	0.907	0.118***	0.033	0.037***	0.011	0.151***	0.030
<i>Bel_L47</i>	33	7.2	0.817	0.853	0.038*	0.025	0.064***	0.019	0.099***	0.029
<i>Bel_L15</i>	23	6.5	0.763	0.791	0.040	0.034	0.105***	0.026	0.141***	0.034
<i>Bel_L45</i>	28	6.3	0.574	0.852	0.314***	0.039	0.025***	0.011	0.332***	0.036
<i>Bel_L31</i>	33	7.3	0.833	0.864	0.035	0.023	0.050***	0.009	0.083***	0.022
<i>Bel_L17</i>	16	5.1	0.728	0.764	0.041	0.036	0.058***	0.019	0.096**	0.033

Significance scores for F_{IS} , F_{ST} and F_{IT} : * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S4. Tests whether genotype frequencies in the 13 retained microsatellite loci follow Hardy-Weinberg equilibrium. These tests, i.e. χ^2 tests and exact tests based on 1000 Monte Carlo permutations of alleles, were run on the regrouped dataset.

Locus	χ^2	df	Pr(χ^2)	Pr.exact
<i>Bel_L53</i>	152.7	171	0.839	0.073
<i>Bel_L25</i>	395.0	253	<0.001	0.002
<i>Bel_L48</i>	142.1	136	0.343	0.002
<i>Bel_L16</i>	182.6	153	0.051	0.031
<i>Bel_L63</i>	166.2	105	<0.001	<0.001
<i>Bel_L13</i>	52.0	45	0.220	0.225
<i>Bel_L10</i>	262.1	171	<0.001	<0.001
<i>Bel_L32</i>	507.8	435	0.009	0.012
<i>Bel_L23</i>	608.5	351	<0.001	<0.001
<i>Bel_L47</i>	920.1	528	<0.001	<0.001
<i>Bel_L15</i>	649.7	253	<0.001	<0.001
<i>Bel_L31</i>	708.9	528	<0.001	0.002
<i>Bel_L17</i>	194.3	120	<0.001	0.002

Table S5. Locality information for the regrouped dataset (13 loci) including statistics of genetic diversity and the identification of specimens to species and/or race. A_n = number of alleles; A_r = expected allelic richness among 14 gene copies ($K = 14$), H_E = expected heterozygosity, i.e. gene diversity corrected for sample size (Nei, 1978), H_O = observed heterozygosity, F_i = individual inbreeding coefficient with asterisks indicating the level of significance for $F_i \neq 0$ from a two-sided test with 1000 randomizations of gene copies among individuals (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Code	Morphospecies	Waterbody	Country	Latitude	Longitude	# ind.	A_n	A_r	H_E	H_O	F_i
LA05_LA06	<i>B. rubicunda</i>	Lake Albert	Uganda	1.82451	31.31753	19	10.00	6.68	0.787	0.765	0.029
LA07	<i>B. rubicunda</i>	Lake Albert	Uganda	1.58649	31.10057	10	7.77	6.65	0.786	0.729	0.076
LV01_LV14_LV37	<i>B. constricta</i> ; <i>B. costulata costulata</i> ; <i>B. trochlearis</i>	Lake Victoria	Uganda	-0.29596	32.03063	17	10.85	7.01	0.800	0.790	0.013
LV06	<i>B. jucunda kisumiensis</i>	Lake Victoria	Kenya	-0.12739	34.74232	10	7.85	6.63	0.791	0.754	0.050
LV07_LV08_LV09	<i>B. costulata ugandae</i>	Lake Victoria	Kenya	-0.09336	34.02474	26	12.62	7.14	0.820	0.782	0.048*
LV26	<i>B. unicolor meta</i>	Lake Victoria	Uganda	0.28161	33.39454	10	7.54	6.44	0.786	0.703	0.110**
LV27	<i>B. costulata costulata</i> ; <i>B. costulata ugandae</i> ; <i>B. unicolor elatior</i> ; <i>B. unicolor meta</i>	Lake Victoria	Uganda	0.11627	33.56262	10	7.15	6.45	0.808	0.774	0.044
LV28	<i>B. costulata ugandae</i> ; <i>B. sp. (juvenile)</i>	Lake Victoria	Uganda	0.10460	33.57583	18	10.92	7.18	0.831	0.853	-0.027
LV29_LV30	<i>B. costulata ugandae</i> ; <i>B. jucunda altior</i> ; <i>B. jucunda jucunda</i> ; <i>B. unicolor elatior</i>	Lake Victoria	Uganda	0.13887	33.60204	18	10.31	7.08	0.838	0.859	-0.027
LV31	<i>B. unicolor meta</i>	Lake Victoria	Uganda	0.12456	33.62969	11	8.38	6.98	0.831	0.788	0.056
LV32	<i>B. trochlearis</i>	Lake Victoria	Uganda	0.16176	33.58264	10	8.54	7.32	0.845	0.804	0.051
LV33	<i>B. costulata dagusiae</i>	Lake Victoria	Uganda	0.15662	33.56414	20	10.85	7.24	0.843	0.842	0.001
LV34	<i>B. trochlearis</i>	Lake Victoria	Uganda	-0.30371	32.28927	10	7.38	6.41	0.797	0.815	-0.024
LV35_LV36	<i>B. costulata costulata</i> ; <i>B. costulata ugandae</i> ; <i>B. jucunda altior</i> ; <i>B. phthinotropis</i> ; <i>B. sp. (juvenile)</i>	Lake Victoria	Uganda	-0.22984	32.37665	11	8.92	7.39	0.847	0.810	0.046
LV38	<i>B. costulata costulata</i> ; <i>B. jucunda jucunda</i> ; <i>B. jucunda kisumiensis</i>	Lake Victoria	Uganda	-0.64564	31.80791	10	8.00	6.78	0.798	0.747	0.067
LV39	<i>B. costulata costulata</i> ; <i>B. unicolor elatior</i>	Lake Victoria	Uganda	0.03892	32.49402	11	8.62	7.19	0.852	0.855	-0.005
LV40	<i>B. constricta</i> ; <i>B. unicolor unicolor</i>	Lake Victoria	Uganda	0.05525	32.49198	10	8.62	7.19	0.842	0.782	0.075*
VN01	<i>B. costulata ugandae</i>	Victoria Nile	Uganda	0.48544	33.15524	10	6.92	5.97	0.752	0.777	-0.035
VN05	<i>B. costulata ugandae</i>	Victoria Nile	Uganda	1.69255	32.09473	12	6.54	5.36	0.729	0.724	0.006
VN06_VN07	<i>B. costulata ugandae</i>	Victoria Nile	Uganda	2.12613	32.33013	17	7.85	5.74	0.747	0.709	0.052
WN01	<i>B. unicolor unicolor</i>	White Nile	Uganda	2.45933	31.50487	12	7.00	6.01	0.765	0.747	0.025

101

102

Table S6. Locus-specific percentage of private alleles for each of the three molecular groups, i.e. LAG = the Lake Albert Group, LVG = the Lake Victoria Group and LKG = the Lake Kyoga group.

Locus	% LAG-private	% LVG-private	% LKG-private	% common
<i>Bel_L53</i>	0.00	0.32	0.00	0.68
<i>Bel_L25</i>	0.00	0.21	0.00	0.79
<i>Bel_L48</i>	0.00	0.42	0.05	0.53
<i>Bel_L16</i>	0.11	0.11	0.00	0.79
<i>Bel_L63</i>	0.05	0.26	0.00	0.68
<i>Bel_L13</i>	0.05	0.05	0.00	0.89
<i>Bel_L10</i>	0.05	0.21	0.00	0.74
<i>Bel_L32</i>	0.00	0.53	0.00	0.47
<i>Bel_L23</i>	0.05	0.26	0.05	0.63
<i>Bel_L47</i>	0.00	0.84	0.05	0.11
<i>Bel_L15</i>	0.21	0.42	0.00	0.37
<i>Bel_L31</i>	0.16	0.32	0.00	0.53
<i>Bel_L17</i>	0.00	0.42	0.05	0.53

Table S7. Estimates of contemporary relative gene flow between pairwise populations of *Bellamyia* from the Lake Victoria ecoregion as inferred with BayesAss (Wilson & Rannala, 2003). Source populations appear in rows and receiving populations in columns, with diagonal values indicating the proportion of non-migrants. Other values indicate the fraction of individuals in the receiving population that are migrants from the source population. Represented values are averages from three independent runs, each with 20,000,000 Markov chain Monte Carlo iterations, a burn-in of 5,000,000 iterations and sampling every 10,000 iterations. Standard deviations among runs were negligible (<0.11 %). The greyscale background color of cells is proportional to values.

	LA05			LV01	LV07			LV29					LV35									
	WN01	06	LA07	14_37	LV06	08_09	LV26	LV27	LV28	30	LV31	LV32	LV33	LV34	36	LV38	LV39	LV40	VN01	VN05	VN06_07	
WN01	67.7 %	12.2 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	2.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	
LA05_06	0.8 %	82.4 %	0.8 %	0.9 %	0.8 %	0.9 %	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	1.7 %	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	0.9 %	
LA07	1.1 %	10.8 %	67.7 %	1.1 %	1.1 %	2.2 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	
LV01_14_37	0.9 %	0.9 %	0.9 %	67.5 %	0.9 %	15.7 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	
LV06	1.1 %	1.1 %	1.1 %	1.1 %	67.8 %	11.8 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	
LV07_08_09	0.7 %	0.7 %	0.7 %	0.7 %	0.7 %	84.2 %	0.7 %	0.7 %	0.7 %	0.7 %	0.7 %	0.7 %	1.5 %	0.7 %	0.7 %	0.7 %	0.7 %	0.7 %	0.7 %	0.7 %	1.5 %	
LV26	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	11.8 %	67.7 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	
LV27	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	8.9 %	1.1 %	67.8 %	1.1 %	1.1 %	1.1 %	1.1 %	4.0 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	
LV28	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	16.2 %	0.9 %	0.9 %	67.5 %	0.8 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	
LV29_30	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	13.7 %	0.9 %	0.9 %	0.9 %	67.5 %	0.9 %	0.9 %	3.4 %	0.9 %	0.9 %	0.9 %	0.9 %	0.8 %	0.9 %	0.9 %	0.9 %	
LV31	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	12.6 %	1.0 %	1.0 %	1.1 %	1.0 %	67.7 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	
LV32	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	11.8 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	67.7 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	
LV33	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	14.6 %	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	69.9 %	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	0.8 %	
LV34	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	11.8 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	67.7 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	
LV35_36	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	12.4 %	1.0 %	1.0 %	1.1 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	67.7 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.1 %	
LV38	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	11.8 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	67.7 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	
LV39	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	11.4 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	2.1 %	1.0 %	1.0 %	1.0 %	67.7 %	1.0 %	1.0 %	1.0 %	1.1 %	
LV40	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	11.9 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	67.7 %	1.1 %	1.1 %	1.1 %	
VN01	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	11.8 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	1.1 %	67.7 %	1.1 %	1.1 %	
VN05	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	1.0 %	67.7 %	13.1 %	
VN06_07	0.9 %	1.0 %	0.9 %	0.9 %	0.9 %	1.5 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	0.9 %	81.7 %	

Table S8. Relative support for the five anagenetic demographic models in our ABC simulations for LAG and LVG, excluding individuals that indicate admixture ($n = 1$ and 16 , respectively; in LKG all individuals have assignment probabilities to the gene pool $\geq 80\%$). Scenarios of effective population size through time for each of these models are displayed in Fig. 2.

Group	Model 1	Model 2	Model 3	Model 4	Model 5	Model 3+5
Lake Albert Group	0.1700	0.0004	0.3933	0.0006	0.4357	0.8290
Lake Victoria Group	0.2378	0.0109	0.3492	0.0199	0.3822	0.7314

Table S9. Estimates on the principal demographic parameters for the best supported ABC models on LAG and LVG excluding individuals that indicate admixture. 95% CI indicates the 95% confidence intervals. Given the high similarities in models 3 and 5, their averaged outcome is presented under Model 3+5. N_e represents the number of diploid individuals in the population and times t , t_1 and t_2 are expressed in number of generations [1 generation is ~ 1 year for *Bellamyia* (Brown, 1994)].

	mean N_e	mode N_e	95% CI		mean t or t_1	mode t or t_1	95% CI		mean $N_{e_{big}}$	mode $N_{e_{big}}$	95% CI		mean t_2	mode t_2	mean $N_{e_{s_Anc}}$	mode $N_{e_{s_Anc}}$
Lake Albert Group (no admixture)																
scenario 3	16 400	7 500	4 360	43 400	148 000	48 600	18 700	433 000	3.92×10^6	1.11×10^6	203 000	9.46×10^6	n.a.	n.a.	n.a.	n.a.
scenario 5	21 100	8 270	4 450	43 400	100 000	39 500	14 800	308 000	6.20×10^6	9.69×10^6	1.38×10^6	9.84×10^6	299 000	339 000	3.02×10^6	405 000
scenario3+5	16 800	7 110	4 400	43 000	123 000	32 700	13 900	406 000	5.13×10^6	3.70×10^6	401 000	9.73×10^6	n.a.	n.a.	n.a.	n.a.
Lake Victoria Group (no admixture)																
scenario 3	29 000	12 300	7 360	78 200	210 000	77 200	33 500	473 000	4.47×10^6	378 000	187 000	9.64×10^6	n.a.	n.a.	n.a.	n.a.
scenario 5	47 600	11 200	7 310	85 300	134 000	78 900	17 800	358 000	6.10×10^6	8.97×10^6	1.28×10^6	9.85×10^6	316 000	406 000	2.64×10^6	238 000
scenario3+5	35 500	12 700	7 570	78 600	173 000	74 600	24 600	451 000	5.28×10^6	9.12×10^6	362 000	9.80×10^6	n.a.	n.a.	n.a.	n.a.

SI Figures

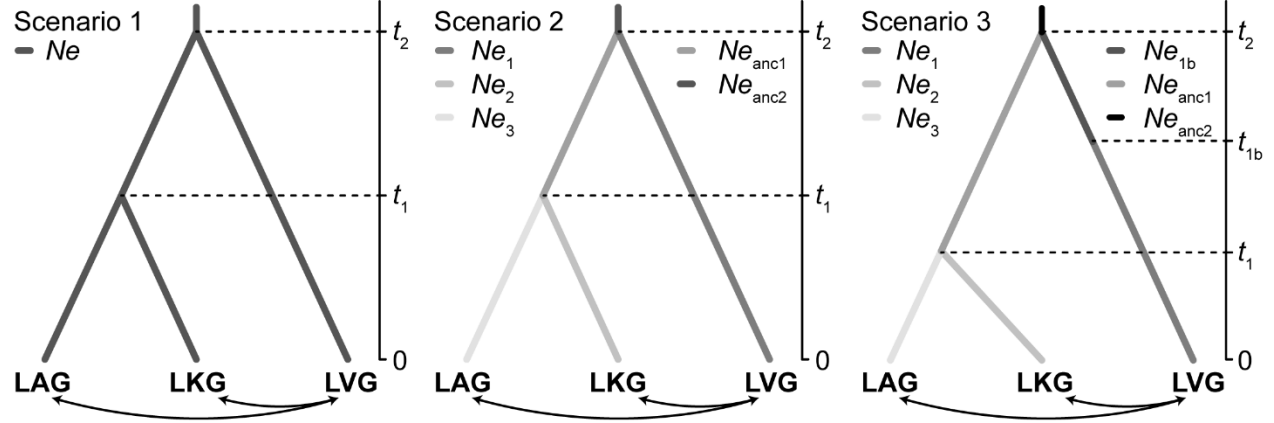


Figure S1. Schematic representation of the 9 examined phylogenetic DIYABC models, including three possible topological relationships between the three molecular groups, i.e. the Lake Albert Group (LAG), the Lake Kyoga Group (LKG), and the Lake Victoria Group (LVG) for each of three scenarios. From scenario 1 to 3 the number of parameters as to effective population size (N_e) increases, allowing with each new parameter new combinations of change (increase or decrease) in N_e .

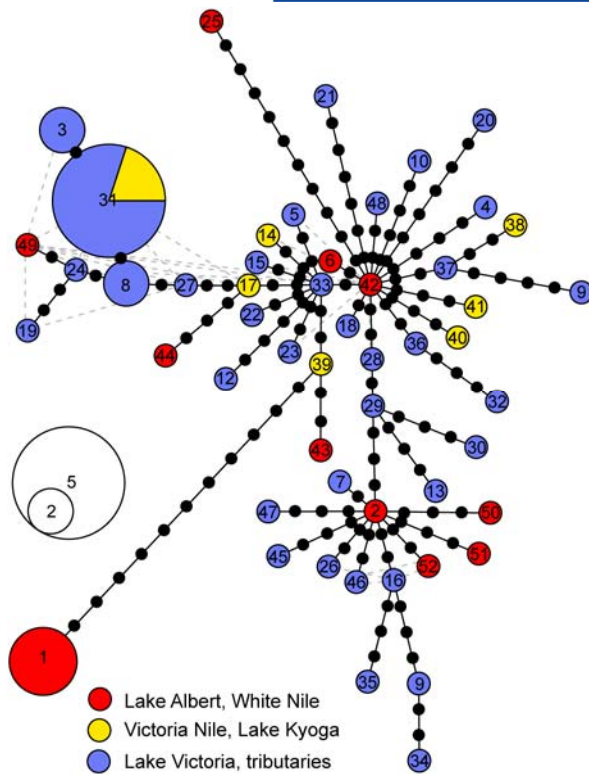


Figure S2. Haplotype network of 52 *Bellamya* haplotypes obtained from 60 sequences of the cytochrome oxidase subunit I (593 bp; see Table S1) belonging to specimens of the Lake Victoria ecoregion, i.e. the 'Victoria clade' of Schultheiß et al. (2014). Dashed light grey edges indicate alternative relationships (mutations, at most 2 mutations/edge, are not indicated for these edges).

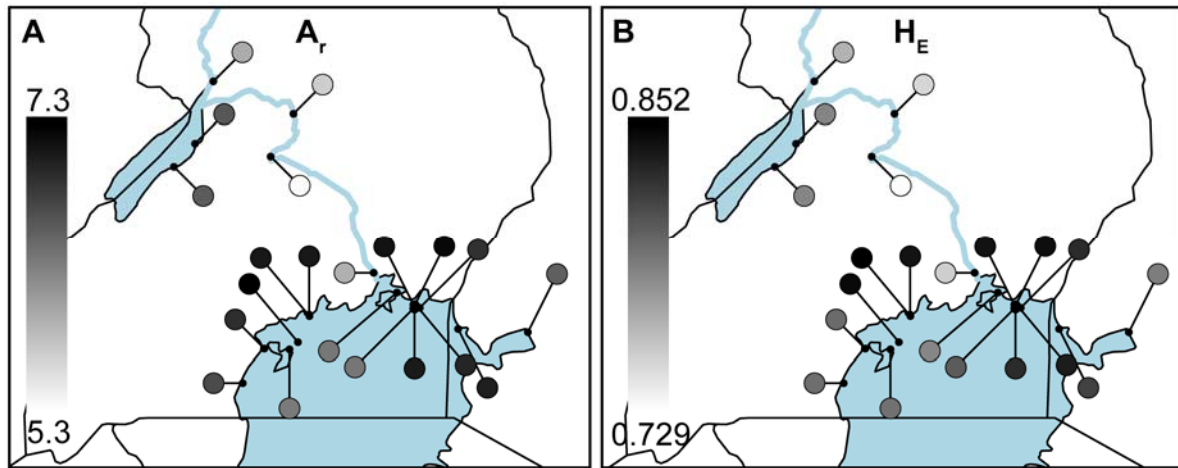


Figure S3. Geographic variation in allelic richness among 14 gene copies (A) and expected heterozygosity (B) in the regrouped dataset of *Bellamya* from the Lake Victoria ecoregion.

MOLECULAR ECOLOGY

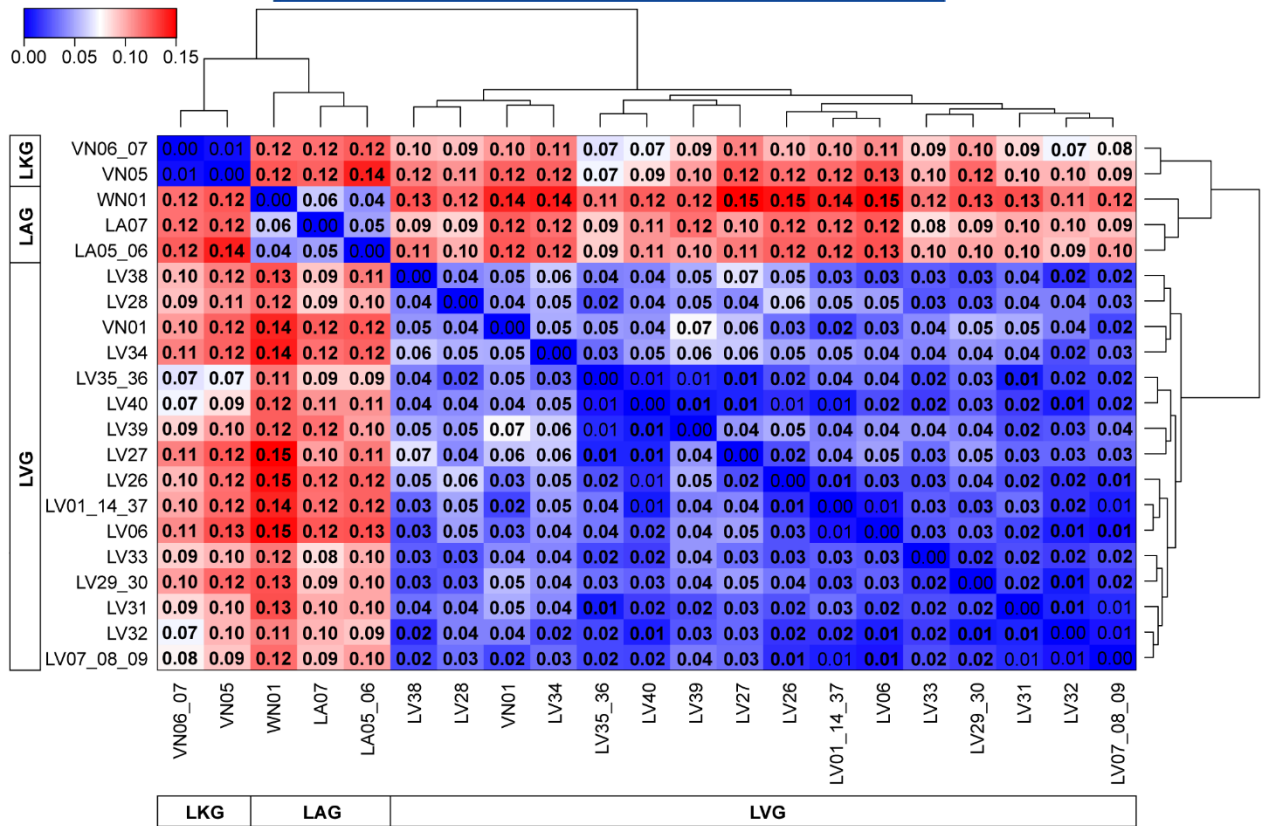


Figure S4. Heatmap of population pairwise F_{ST} values with clustering dendrograms revealing the three molecular groups (= gene pools) of specimens in the Lake Victoria ecoregion. High F_{ST} values are observed between gene pools (red), whereas F_{ST} values within gene pools (blue) are lower.

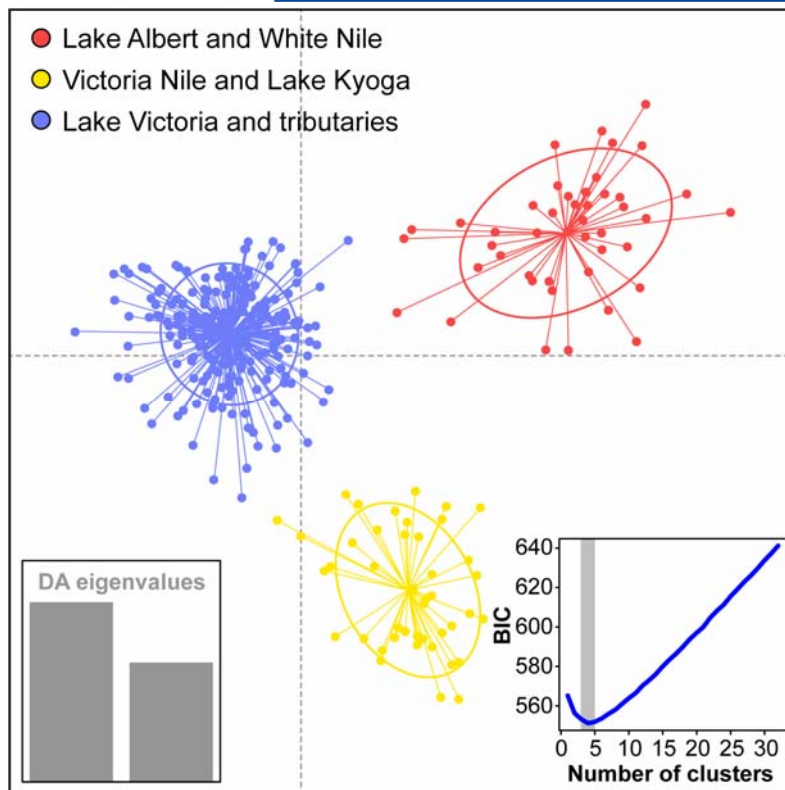


Figure S5. Discriminant analyses of molecular components on the microsatellite dataset (13 loci) of *Bellamyia* from the Lake Victoria ecoregion. The analysis indicates highest support from the Bayesian information criterion (BIC) for four groups, which resulted in poor assignment probabilities for two of these four groups, however. The here depicted three-group solution has still high BIC support and consistently high assignment probabilities. Therefore, it represents a more robust clustering solution. As with the other two clustering methods illustrated in the main text, a few individuals were assigned to geographic clusters that are inconsistent with their sampling locality (not shown). See main text and Fig. 3A, C for more information on this topic.

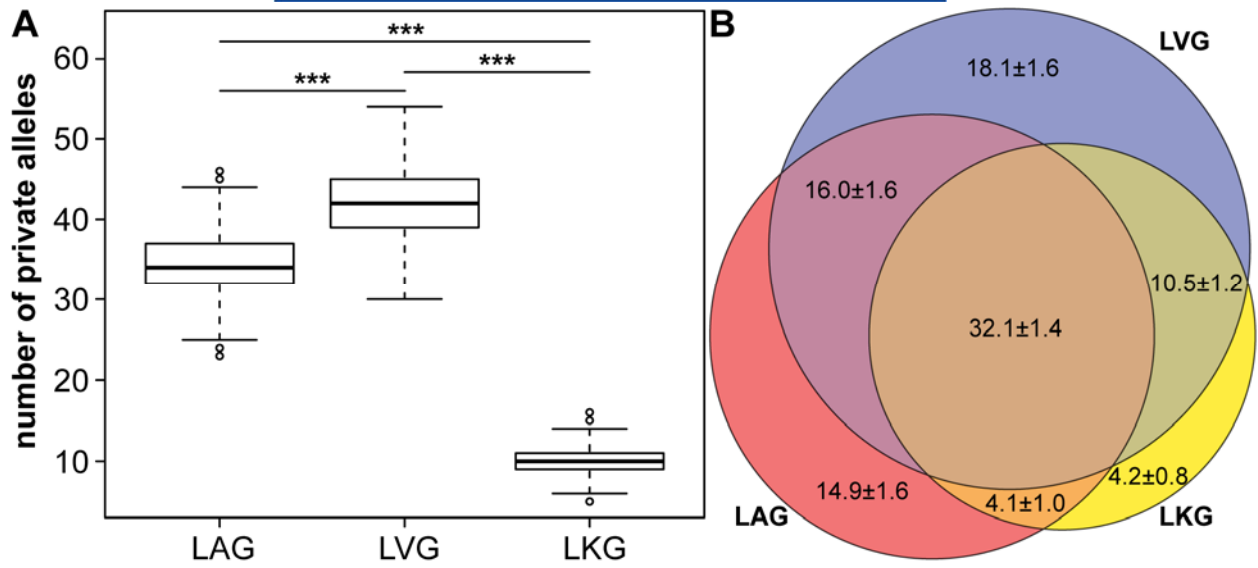


Figure S6. A) Total number of private alleles from 13 microsatellite loci for 1000 random subsamples of 29 individuals (the number of individuals in the regrouped dataset assigned to the smallest genetic cluster, LKG) for each molecular cluster. Differences between these three distributions were tested with FDR-corrected Wilcoxon tests, and are highly significant. B) Relative proportions (in %, mean \pm sd) of private and shared alleles among the three molecular groups following the same subsampling scheme under (A).

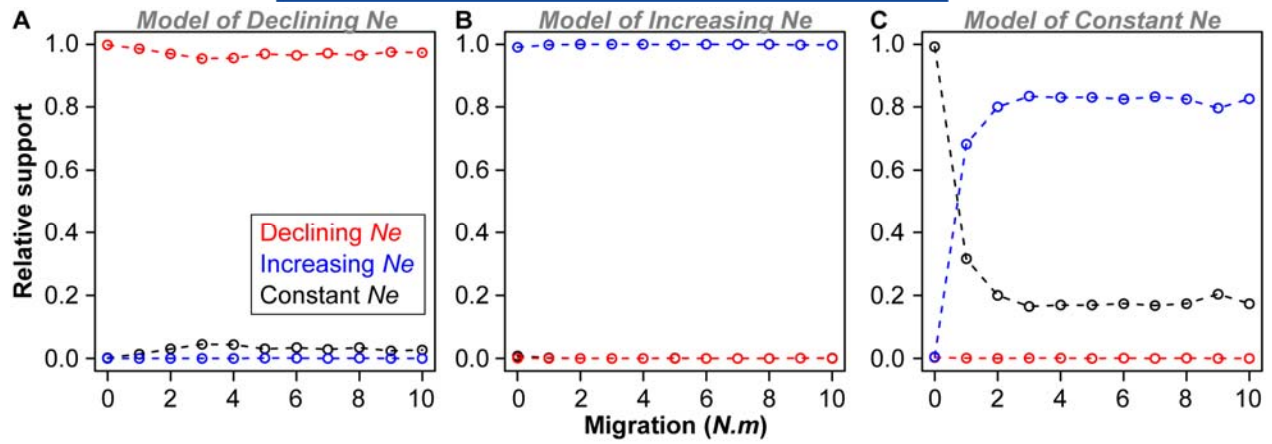


Figure S7. Analysis of how migration into the studied population affects approximate Bayesian computation (ABC) for populations with a decline in N_e (A), an increase in N_e (B) and constant N_e (C) over time. Migration is indicated as the number of immigrants per generation into the studied population ($N.m$). Migration does not bias ABC inferences for populations that are declining or increasing in N_e , however, it is expected to bias conclusions for populations of constant N_e .