

RESOURCE ARTICLE

A population genetic assessment of taxonomic species: The case of Lake Malawi cichlid fishes

Catarina Pinho¹  | Vera Cardoso¹ | Jody Hey^{2,3} 

¹CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Vairão, Portugal

²Rutgers, the State University of New Jersey, Piscataway, New Jersey

³CCGG, Center for Computational Genetics and Genomics, Department of Biology, Temple University, Philadelphia, Pennsylvania

Correspondence

Jody Hey, CCGG, Department of Biology, Temple University, Philadelphia, PA.
Email: hey@temple.edu

Funding information

National Institutes of Health, Grant/Award Number: R01GM078204; Fundação para a Ciência e a Tecnologia, Grant/Award Number: PTDC/BIA-BDE/66210/2006 and IF/01597/2014; European Social Fund; Ministério da Educação e Ciência

Abstract

Organisms sampled for population-level research are typically assigned to species by morphological criteria. However, if those criteria are limited to one sex or life stage, or the organisms come from a complex of closely related forms, the species assignments may misdirect analyses. The impact of such sampling can be assessed from the correspondence of genetic clusters, identified only from patterns of genetic variation, to the species identified using only phenotypic criteria. We undertook this protocol with the rock-dwelling mbuna cichlids of Lake Malawi, for which species within genera are usually identified using adult male coloration patterns. Given high local endemism of male colour patterns, and considerable allele sharing among species, there persists considerable taxonomic uncertainty in these fishes. Over 700 individuals from a single transect were photographed, genotyped and separately assigned: (a) to morphospecies using photographs; and (b) to genetic clusters using five widely used methods. Overall, the correspondence between clustering methods was strong for larger clusters, but methods varied widely in estimated number of clusters. The correspondence between morphospecies and genetic clusters was also strong for larger clusters, as well as some smaller clusters for some methods. These analyses generally affirm (a) adult male-limited sampling and (b) the taxonomic status of Lake Malawi mbuna, as the species in our study largely appear to be well-demarcated genetic entities. More generally, our analyses highlight the challenges for clustering methods when the number of populations is unknown, especially in cases of highly uneven sample sizes.

KEYWORDS

cichlids, hybridization, population genetics, speciation

1 | INTRODUCTION

The processes of discovering new species and of assigning individuals to previously described species sometimes come into conflict with the concept of a species as a sexual evolving lineage. For example, species diagnosis and species identification, by human investigators,

frequently depend upon the use of diagnostic characters (Cracraft, 1983), but diagnostic characters per se play no role in the theory of how species come into being or how they persist. Species do of course evolve derived fixed traits that sometimes serve well as diagnostic characters, but cryptic species are also common (Bickford et al., 2007). For species that are highly visual, and for which coloration

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2019 The Authors. *Molecular Ecology Resources* Published by John Wiley & Sons Ltd.

patterns are cues, both for reproduction and for species diagnosis by human investigators, we might expect that species diagnosis by human investigators will accurately capture real species.

We undertook a close examination of the process of species identification in a classic species group where male colour variation plays a large role in both reproduction and species identification by investigators. Specifically, we asked how well species assignment based on morphological criteria (primarily coloration) corresponded to species as genetic entities, as estimated using clustering methods that minimize linkage disequilibrium (LD) and inbreeding disequilibrium among multiple polymorphic loci.

The cichlid fishes of the great African lakes represent arguably the single most striking known radiation of vertebrate species, with each of the lakes harbouring hundreds of species, and with some parts of these radiations having occurred very recently, and thus exceedingly rapidly (Kocher, 2004; Seehausen, 2006). This system constitutes one of the most important natural laboratories for studies in speciation, particularly because of the complex interplay between ecological opportunity and sexual selection, which presumably drove the formation of multiple species (Salzburger, 2009; Wagner, Harmon, & Seehausen, 2012). However, the system also presents significant challenges in species diagnosis and identification (Turner, Seehausen, Knight, Allender, & Robinson, 2001).

The systematics of African cichlids has traditionally relied on features such as body shape and morphology of the pharyngeal and oral jaw apparatus (Rüber, Verheyen, & Meyer, 1999; Trewavas, 1984). More recently, male secondary sexual characters such as bower shape (Kidd, Kidd, & Kocher, 2006) and especially male nuptial coloration have also been used to describe species (Genner et al., 2004). A reliance on male breeding coloration to delimit species necessarily overlooks the identification of females and juveniles, which are comparatively drab in coloration (Kornfield & Smith, 2000). Because of this difficulty, sampling schemes in evolutionary or ecological studies usually rely on typical males that can indeed be readily identified, while overlooking other individuals, including less typical males, females, juveniles and possibly intermediate phenotypes.

When species are studied by sampling only males in breeding coloration there are implicit and possibly erroneous assumptions about the nature of those species. One assumption is that such males are representative of the species, which will be in error if they somehow differ genetically from other life stages or from females. Secondly, a reliance on typical male individuals might skew investigators' perception of the magnitude of species colour differences and cause individuals that are intermediate between investigators' diagnostic patterns for related species to be overlooked. Species hybrids or backcrossed individuals are likely to be overlooked by this type of sampling, and sampled stereotypical males may thereby mislead on the level of gene exchange between species. If investigators are focused strongly on males that show diagnostic patterns, and thereby overlook individuals that show a more continuous gradation between species, then they may come to erroneous conclusions regarding such things as the action of disruptive natural or sexual selection as engines of diversification. A second more inclusive

assumption underlying this type of sampling is that male nuptial coloration and other distinctive morphological features are indeed good characters to delimit species. There are several reasons why this might not be generally so. On the one hand, the diversity of colour could be driven in part by phenotypic plasticity. Also, intraspecific colour polymorphism in sympatry has been well documented in several species (Sinervo & Lively, 1996), including cichlids (Maan & Sefc, 2013, and references therein).

It is also possible that sampling on the basis of one set of morphological criteria can fail to reveal the presence of cryptic species that are similar by those criteria. Cryptic species are common in many biological systems (Bickford et al., 2007), and they have been reported among Lake Malawi cichlids (Kidd et al., 2006). If cryptic species are frequent, our view on cichlid diversity might be significantly underestimated.

In this study, we asked the following question: how do the species identifications of individuals, based only on traditional morphological criteria, correspond to genetic entities (i.e., gene pools) discovered using a population genetics approach? To be clear when referring to instances of the two kinds of groups, we refer to the species to which individuals were assigned using morphological criteria, as "morphospecies" and to the groups identified only by patterns of genetic variation as "genetic clusters" or "genetic populations".

We focused mainly on a community of cichlids, the rock-dwelling "mbuna" species of Lake Malawi that live primarily in the shallow waters of the rocky portions of the coastline. Individuals were collected without regard to coloration by sweep netting a transect along a rocky shore area of Lake Malawi. A large collection of juveniles and adults of both sexes were then photographed and identified to morphospecies, and then later genotyped at a set of highly polymorphic loci to identify genetic clusters. With these two fundamentally different types of identifications in hand we could then ask how well the two different groupings corresponded to each other, and we could potentially identify hybrid individuals. Finally, we could also ask: had we sampled only males in breeding coloration, would we have a biased or misleading sample of the genetic populations?

2 | MATERIALS AND METHODS

2.1 | Sampling and photographic record of cichlids

Samples were collected off the rocky shore of the western side of Domwe Island, in southern Lake Malawi. Our choice of sampling habitat reflects the intention of collecting especially the brightly coloured rock-dwelling mbuna, which are popular objects of both research and commercial trade for aquarium hobbyists. The mbuna are an informal assemblage of 13 speciose genera (Oliver & Arnegard, 2010). Prior to sampling, we set up a transect 5 m wide and 50 m long, extending from the shoreline and reaching a maximum depth of 20 m. The transect was divided into quadrants which were sampled individually using SCUBA and with the aid of gill nets that were 2 m in height and 5 m in length. Nets were swept repeatedly through each quadrant with the bottom edge of the net just above the rocky

bottom. In each quadrant, we collected the majority of individuals that could be caught, without regard to species, sex, colour or life stage of each individual. Sampling was not uniform with respect to size and elusiveness, as smaller fish, particularly juveniles, and fish that are more elusive were more likely to escape the net or to escape while transferring from the net. Although the focus of our sample were the mbuna, we included all fish that were collected in the nets, including some that were later assigned to nonmbuna genera.

After their collection and prior to handling, fish were kept in underwater bags that were brought to the surface in a slow and gradual fashion to allow acclimation to lower water pressures. To facilitate handling and minimize stress, all fish were individually anaesthetized using a solution of clove oil (10 ml of 1:10 solution in ethanol to ~10 L of water). Samples consisted of a portion of the dorsal and caudal fins, preserved in a salt-saturated buffer containing 20% dimethyl sulfoxide (DMSO). Before their release back to the lake, all fish were photographed underwater as described in Joo et al. (2013).

Samples from a total of 710 individuals were collected and photographed. We asked two local divers, experienced with cichlid identification—Richard Zatha and James W. Maluza—to perform an a posteriori identification of the species, sex and life stage of each fish from the photographs obtained. These two experts performed the identification together, based on publications exhibiting photos of the different species found in the area (Konings, 1990, 2007). Based on this assignment, our sample included both mbuna (663 specimens from the genera *Metriaclima*, *Tropheops*, *Labeotropheus*, *Genyochromis*, *Petrotilapia* and *Melanochromis*) and nonmbuna (47 specimens from a diverse set of genera) resulting in a minimum of 35 species (some specimens were only assigned to the generic level), two of which are represented by more than 200 individuals, *Metriaclima zebra* and *Tropheops gracilior*, and 26 by fewer than five individuals.

Sampling took place under appropriate permits from the Lake Malawi National Park and samples were transported to Portugal under a licence from the Instituto da Conservação da Natureza e das Florestas (ICNF).

2.2 | Laboratory procedures and genetic data collection

DNA was extracted from fin tissue (Sambrook, Fritsch, & Maniatis, 1989). For genetic analyses, we chose a total of 26 dinucleotide short tandem repeat (STR) loci. These STRs were mainly chosen from the set developed for Lake Malawi cichlids by Albertson, Streebman, and Kocher (2003) but also from other works (Booton, Kaufman, Chandler, & Fuerst, 1996; Kellogg, Markert, Stauffer, & Kocher, 1995; Van Oppen, Rico, Deutsch, Turner, & Hewitt, 1997). Whenever possible we preferred markers from different linkage groups and avoided STRs linked to known phenotypes to minimize the possibility of using markers under strong selection.

After preliminary assays to evaluate amplification conditions and observe STR length range, the 26 markers were arranged into six different multiplex PCRs. These were carried out with three primers

per locus, following the M13-tailed primer method (Oetting et al., 1995). In brief, this method consists of adding four different universal M13 primers labelled with fluorescent dyes (6-FAM, VIC, NED and PET) which anneal to tails incorporated 5' into the forward primer sequence. The complete list of the markers used, as well as their organization into multiplexes and composition of each multiplex primer mix is given in Table S1. Multiplex PCRs were carried 10- μ l volumes containing 5 μ l of Master Mix (Qiagen), 3 μ l of H₂O, 1 μ l of preprepared primer mix and 1 μ l of DNA. Amplification conditions were as follows: an initial denaturing step of 15 min at 95°C, followed by 42 cycles of 95°C for 30 s, 1 min of annealing and 45 s at 72°C. This was followed by a final stage of 30 min at 60°C. For multiplexes 1, 4, 5 and 6 annealing temperatures started at 60°C and decreased 0.5°C per cycle during the first 12 cycles; this was followed by 22 cycles of stable annealing at 54°C. For the final eight cycles annealing temperature was 53°C. For multiplexes 2 and 3 the initial annealing temperature was 62°C and was lowered at a rate of 0.5°C per cycle during the first 12 cycles; 22 cycles with annealing at 56°C followed and the final eight cycles were conducted with annealing at 53°C.

PCR products were separated on an ABI Prism 3130xl Genetic Analyzer (Applied Biosystems) along with an internal size standard (Genescan-500 LIZ; ABI). Alleles were scored based on sizing bin windows using GeneMapper software version 3.1.2 (Applied Biosystems).

To identify linked loci, or loci with null alleles or typing errors, we faced a conflict between the need to screen loci by assessing departures from Linkage and Hardy-Weinberg equilibrium (LE and HWE, respectively), which requires taxon or population assignment, and our goal to identify populations using genetic data. We settled on screening loci using preliminary identification of eight morphospecies (as identified by the two experts) for which we sampled more than 10 individuals (*Genyochromis mento*, *Labeotropheus fuelleborni*, *Metriaclima zebra*, *Protomelas fenestratus*, *Petrotilapia genalutea*, *Tropheops gracilior*, *T. microstoma* and *T. "orange chest"*), and we considered a locus as problematic when there was consistent evidence for problems in at least three out of the eight species. Loci were screened for LE using Genepop on the Web (Raymond & Rousset, 1995; Rousset, 2008), and for HWE using FSTAT 2.9.3 (Goudet, 2001). Additionally, we performed more sensitive analyses aimed at detecting microsatellite typing errors using MICROCHECKER 2.2.3 (Van Oosterhout, Hutchinson, Wills, & Shipley, 2004), with 10,000 replicates per sample.

2.3 | Analysis of population structure

We used four methods for identifying populations using genetic data: the model-based Bayesian methods implemented in BAPS 6 (Corander & Marttinen, 2006; Corander, Marttinen, Sirén, & Tang, 2008), STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) and SNAPCLUST (Beugin, Gayet, Pontier, Devillard, & Jombart, 2018); and the model-free discriminant analysis of principal components, DAPC (Jombart, Devillard, & Balloux, 2010). BAPS, STRUCTURE and SNAPCLUST

are similar in that they identify clusters that maximize HWE and LE, and all can be used to assess admixture; however, they differ in other respects. DAPC, on the other hand, is a multivariate genetic clustering method that does not make use of LE or HWE, or indeed any genetic model per se.

For estimating both K and admixture *BAPS* relies on a two-step analysis: first, determination of the number of clusters and individual assignment using a Bayesian stochastic partition approach and a subsequent study of the proportion of each individual's genome originating in each of the inferred ancestral populations ("admixture analysis"). As part of the admixture analyses, *BAPS* performs a simulation test allowing the evaluation of the statistical significance of a deviation from the null expectation of no admixture, which prevents spurious inference of admixture between similar clusters. *BAPS* requires the input of one or multiple priors on a maximum bound on the number of clusters. As recommended by the authors, we provided the program with a vector including a wide range of values (up to $K = 200$, each repeated several times) to ensure sufficient independent runs. We also repeated this procedure multiple times to evaluate the consistency of the results. The final analyses included two sets of ~1,100 runs. For the subsequent admixture analysis, we ran the program based only on the most likely partition including all the clusters regardless of their sample size. For these analyses, 200 iterations were performed to estimate admixture coefficients. Two-hundred reference individuals from each population were simulated and their admixture coefficient was estimated based on 50 iterations to assess the p -value of admixture coefficients.

STRUCTURE uses a Markov chain Monte Carlo approach to estimate each individual's admixture proportions for a predefined value of K .

STRUCTURE includes two models relative to prior expectations on differences between allele frequencies in inferred clusters. The independent allele frequency (IAF) model assumes that different clusters are not connected by recent divergence and that their genetic composition can be assumed to be independent. The correlated allele frequency (CAF) model (Falush, Stephens, & Pritchard, 2003) assumes that clusters derived recently from an ancestral population and that as such their allele frequencies are not independent from one another. The latter model can be more useful to detect very subtle differentiation. Because we had no prior expectation on which model would be more suitable for the biological problem in hand, we implemented both models. This was accomplished using the same settings in both cases. We ran analysis in *STRUCTURE* for 200,000 steps, of which the first 50,000 were discarded as burn-in. We considered a range from 1 to 30 for the number of populations (K) and performed 10 replicates for each. We allowed the parameter α to vary across populations, and used $1/K$ as the initial value for α , as suggested by Wang (2017) for putatively large K values and highly unbalanced sample sizes (as we predict to occur in the present study). We considered two methods for estimating K : (a) using Bayes Rule with a uniform prior on K and evaluation of the estimated log likelihood, $\ln(\text{Pr}(\text{Data}|K))$, as suggested by the authors in the program manual (Pritchard, Wen, & Falush, 2010); (b) the method outlined by Evanno, Regnaut, and Goudet (2005), which searches for a mode in

the distribution of ΔK , a quantity related to the second order rate of change of $\ln(\text{Pr}(\text{Data}|K))$. This latter analysis was conducted using the standalone version of *STRUCTUREHARVESTER* (Earl, 2012).

DAPC performs a discriminant analysis (DA) on genetic data previously transformed using principal component analyses (PCAs); therefore, this procedure combines the advantages of a DA (maximizing between-group differences while overlooking within-group variation) with those of a PCA (summarizing the differences in a large set of allele frequencies into a discrete set of informative and uncorrelated variables), while imposing no model with respect to allele frequency variation among and within groups. The definition of clusters, a prior step to DAPC, is based on K -means clustering of principal components; this procedure divides variation into a between- and a within-group component and describes groups that minimize the latter for each number of populations (K). For each value of K there is an associated likelihood that can be used for model choice using, for example, the Bayesian information criterion (BIC). DAPC is implemented in the R package *ADEGENET* 2.1.1 (Jombart, 2008), which we ran in R version 3.5.1. We began by using the *find.clusters* function to define the number and nature of populations to consider. We used 10^6 iterations (*n.iter*) and 500 starting points (*n.start*) in each run of the K -means algorithm. We used the BIC, as suggested by the authors. For this step we retained all of the PCA axes by defining *n.pca* = 800. We used visual inspection of a plot relating the BIC for each K to select the optimal number of populations for the data set. Using these groups, we proceeded with a DAPC on the data. The difficult step in DAPC is the choice of the number of principal components to retain, because a large number can cause model overfitting whereas a small number may lead to low discrimination power. To assist in this choice, *ADEGENET* includes a cross-validation function named *xvalDapc*, which splits the data into training and validation sets (using stratified random sampling across inferred groups) and performs DAPC on the training set with different numbers of retained components (PCs). By evaluating the ability of each analysis to determine the correct group membership for each individual in the validation set, the output of this function allows us to determine the minimum number of PCs that are sufficient for accurate assignment. We used this function with the default proportion of 90% of individuals in the training set and each analysis was replicated 500 times. We found the optimal number of retained PCs to be 28 for our data and used this value in our subsequent DAPC analyses. Finally, we recorded the posterior group assignment of each individual as a measure of the admixture proportion originating in each cluster.

SNAPCLUST uses the same model-based approach as *STRUCTURE* and *BAPS* to identify groups that minimize disequilibria, yet relies on a faster algorithm, a combination of geometric approaches like those used by DAPC and expectation-maximization (EM) likelihood optimization, to achieve the maximum-likelihood solution for the specific clustering problem. This method is also implemented in the R package *ADEGENET*. We began by using function *snapclust.choose.k* and by considering the maximum possible K as 50. We used Akaike's information criterion (AIC) for all inferences using *SNAPCLUST* based on trial runs. After this prior step of choosing the more adequate number of

populations, we proceeded with the clustering analysis itself using the function *snapclust*. We used the default option of using the Ward algorithm to define the initial groups, and we ran the EM algorithm 50 times. The maximal number of iterations was set at 500. The output from the *snapclust* function includes information on whether the run converged, and this was checked before recording the result.

In the case of BAPS, unlike the other methods, the results include an estimate of whether an individual is admixed. Among those individuals identified as admixed by BAPS, we noticed that they showed a maximum probability of assignment of 74% or lower to a single cluster. However, this threshold may not be easily applied to the case of other methods, because the assignment probabilities may be differently impacted by noise in the estimation procedure. Therefore, in order to use a threshold (for identifying admixed individuals) that is appropriate for each method, and yet comparable across methods, we used a relative threshold of $0.75/(\text{mean } Q_{\max})$, where Q_{\max} is the maximum probability of assignment to a single cluster for each individual in that particular method.

Additionally, we evaluated whether the adoption of an iterative strategy to the clustering problem (performing a second set of analysis in a partial data set from where previously detected clusters were removed) could increase the detectability of under-sampled clusters in methods which recovered few clusters. In order to do so, for the three methods with fewer clusters detected (STRUCTURE with IAF, DAPC and SNAPCLUST) we removed from the data set individuals that had been classified in the four largest clusters, which were recovered in common by different methods, and applied exactly the same approach described above to the resulting data sets.

2.4 | Bootstrapping to assess cluster support

Although all of the methods described above are potentially able to detect clusters within a sample, all of them lack a measure of the support for each group. Not all groups may be equally supported by the data. We thus implemented a bootstrap methodology to assess support for each of the groups detected by the different methods.

To assess how strongly the data support the groups that were detected, we generated 100 data sets of the same size as the actual data set. For each bootstrapped data set of 22 loci, each locus was drawn at random from the full data set of 22 loci, with replacement. Each data set was then analysed by BAPS, STRUCTURE, DAPC and SNAPCLUST using faster versions of the implementations described above for the main analyses. In BAPS each of the bootstrap replicates was analysed 30 times for the mixture analyses, using a range of different priors for the maximum number of populations; admixture analyses were performed based on each output using 50 iterations, 50 reference individuals and 10 iterations for reference individuals. Each bootstrap replicate was run in STRUCTURE for K ranging from 1 to 30, with five iterations for each K ; the runs consisted of 10,000 burn-in and 20,000 post-burn-in steps and the same model settings chosen for the main analyses; for each bootstrapped data set we then chose the run with the highest $\text{Ln}(\text{Pr}(X|K))$. For DAPC we reduced the number of starting points to 50 and the number of iterations to 1×10^5

unless convergence problems were detected (in these data sets we raised the number of iterations until no more convergence warnings were raised by the program). Also, for simplicity, the choice of the number of populations after running *find.clusters* was conducted in an automated fashion and not by visually inspecting each BIC curve. ADEGENET provides different algorithms for this choice. We opted for using the “min” algorithm (choosing the model with the lowest value of BIC), because it selected K values closer to those chosen by visual inspection in experiments using a small subset of replicates than any of the other available algorithms. *xvalDapc* prior to DAPC was also performed but the number of replicates was reduced to 25. Since SNAPCLUST is extremely fast we did not perform any adjustment to the methodology described above to carry on the bootstrap analyses.

After obtaining the admixture proportions per individual based on all methods for all the data sets, the process of calculating the support for each group was identical across methods. First, we used the maximum admixture proportion to assign individuals to a group, without taking into account whether an individual might be admixed or not. Second, we defined the reference group for each individual as the group it was assigned to in the main analysis. Then, for each replicate, we compared each individual's assignment to the assignments of its reference group elements by: (a) defining the modal cluster(s) of its reference group, that is, the cluster(s) receiving the higher number of assignments in that set of individuals for that particular replicate; (b) comparing each individual's assignment with the modal cluster(s): if the individual was assigned to a modal cluster of its reference group it received a score of $1/(\text{number of modal clusters})$, otherwise it received a score of 0. Next, we evaluated a consistency index for each individual as the sum of its scores across the 100 replicates. Finally, the bootstrap support of each group was then calculated as the arithmetic mean of the individual consistency indices for members of the group.

2.5 | Analyses of simulated data sets

For our real data set we do not know the true number of genetic populations. While we can compare the results of BAPS, STRUCTURE, SNAPCLUST and DAPC to each other, and compare these outcomes with morphospecies identifications, our agnostic approach cannot directly assess which if any partition is correct. To gain some perspective on the relative performance of BAPS, STRUCTURE, SNAPCLUST and DAPC under cases where we do know the number of true genetic populations, we evaluated the performance of each of the methods using simulated data sets that resembled our actual data in terms of sample sizes and numbers of morphospecies.

We simulated data sets using EASYPOP 2.01 (Balloux, 2001). Although we do not know the true number of genetic populations in our real data, we nevertheless tried to reflect some features of the real data in the simulated data. In particular, given the highly skewed distribution of morphospecies in our data, we used a highly skewed distribution for each of the simulated data sets. For each data set we simulated 25 populations of equal sizes, from which we randomly sampled individuals following a skewed distribution of sample sizes:

two populations with 200 individuals, one with 100, one with 50, three with 25, four with 10, eight with five, three with two, and four groups with just one individual. After initial trials to evaluate which parameter combinations would give appropriate results, we ran *EASY-POP* with different parameter combinations for mutation rates (from 0.00045 to 0.00055), migration rates (0.002 to 0.004), maximum number of alleles per locus (80 or 100) and number of generations (3,500–8,000), although the 100 simulations that were retained in the end (see below) were all performed with migration rates of 0.0025 and 4,000 generations. In all simulations we chose the island model for migration, the SMM-allele model for mutation and a maximal variability of the initial population (as opposed to all populations starting with the same allele). We conducted simulations using two different data set sizes: 22 loci, as our own data set, and 100 loci, to evaluate whether differences between methods were consistent using a larger data set. To ensure that these data sets were comparable to ours in other key aspects related to the methods' performance, we retained only the 100 simulated data sets that were most similar to the real data in terms of: (a) the total number of alleles, "TA" (in order to monitor overall diversity), controlled for the number of loci; (b) the total number of alleles private to a single individual, "TP" (in order to control for the possibility that individuals with private alleles were more likely to be placed in a different cluster), again controlled for the number of simulated loci; and (iii) Weir and Cockerham's (1984) estimator of F_{ST} , Θ_{ST} (as estimated with the R package *HIERFSTAT* (Goudet, 2005)). We considered as the guide value for $\Theta_{ST} = 0.1$, which is close to the mean value of those estimated from the population structures inferred by the various methods of assignment in our original data set (0.095), particularly to those that appear to be able to recover a number of populations more similar to the real value (see Section 3). The approximation of these measures to those observed was taken as the percentage difference from observed to expected, taken cumulatively. The final set of simulated data sets with 22 loci had a median TA of 954 (vs. 938 in the original data set), median TP of 108 (109 in the original data set) and median Θ_{ST} of 0.0994. The simulated data sets with 100 loci had a median Θ_{ST} of 0.1004, a median TA of 4,331 (vs. an expected value of 4,264 for 100 loci) and a median TP of 489 (vs. 495 for 100 loci). Using these data sets, the different assignment methods were run using the same conditions described for the bootstrap procedures (see above).

We recorded the value of K obtained across all simulations by the different methods (and using the two alternative estimators of K in *STRUCTURE*). Furthermore, we calculated the Rand Index (RI) (Rand, 1971) to compare the inferred clusterings with the simulated partition. RI was calculated according to the formula $RI = (a + b) / (a + b + c + d)$, in which a refers to the number of pairs of individuals which were placed in the same cluster in both the inferred and the real (simulated) clustering, b is the number of pairs which were placed in different clusters in the two clusterings, c is the number of pairs that were in the same group in the real clustering but that were separated in the inferred clustering and d is the number of pairs that had individuals from two real clusters but that were placed

together in the inferred clustering. To compute these values, admixture proportions do not need to be taken into account because the calculation of RI requires clusters to be fully discrete. Therefore, as for bootstrap analysis, we assigned individuals to the populations where they had allocated the largest proportion of their genome irrespectively of admixture being present or not. As a measure of the quality of the assignment, we also computed the mean Q_{max} for each simulation.

2.6 | Effects of male-biased sampling schemes

One concern when studying species complexes with strongly sex-limited patterns of colour variation is whether samples consisting of typically coloured individuals of just one sex could provide biased perspectives on the discontinuity between species. To address this issue, we repeated the analyses of population structure described above using a data set comprised only of adult male individuals. For these to be comparable to data sets with both sexes we took 100 random samples of 292 individuals (the number of adult males) from the full data set and analysed them using the clustering methods. We then asked whether the male-only sample differed from random samples of the same size for (a) overall assignment probabilities; (b) the number of individuals recovered as admixed; and (c) the number of inferred clusters.

Finally, we investigated the common practice of sampling only adult males by asking the question: are females and juveniles more prone to be erroneously identified in the field? To do so, we focused on morphospecies with more than five individuals, and evaluated whether the genetic assignment of each individual was consistent with that of the majority of individuals of the same morphospecies (excluding putatively admixed individuals). For each individual we determined whether it was a genetic outlier from its morphospecies, and then we compared the number of wrongly assigned female and juvenile individuals with the number of wrongly assigned adult male individuals. We also conducted a second analysis in which we focused on the largest *BAPS* clusters and asked whether males and nonmales were assigned to the same morphospecies as the majority of individuals in those clusters.

In all the above procedures, format conversions, complex calculations, and handling of data sets and outputs were performed using a series of scripts written in the *PYTHON* and *R* programming languages.

3 | RESULTS

3.1 | Microsatellite variability

Twenty-six dinucleotide microsatellites were genotyped in the 710 individuals collected from the sampled transect. The average missing data per individual was 0.88 loci (3.4%). The number of alleles per locus ranged from 12 (in locus UNH2168) to 80 (in locus UNH2059), with an average of 40.42 alleles per locus. We used eight morphospecies to identify problematic loci, and then, to guard against the misapplication of morphospecies, we rejected loci if they failed in

three or more of these. Analyses of HWE using *FSTAT* revealed multiple cases of significant deviations from HW expectations (Table S2). Application of the *MICROCHECKER* program gave similar results. Although wrongly identified morphospecies may contribute to a lack of HWE, disequilibrium can also be caused by confounding population structure within morphospecies. For example, 14 and 11 out of 26 loci show deviations in individuals identified as *Metriaclima zebra* and *Tropheops gracilior*, respectively. However, there are also cases suggesting problems in genotyping. The most striking case is locus UNH2203, which exhibits deviations from HW and presence of null alleles in seven out of the eight "species." Other loci showed some evidence of missing alleles. Based on these different tests, we excluded UNH2168, UNH2203, UNH2035 and UNH2065 from population structure analyses.

Analyses of LD revealed several dozen significant deviations from independent segregation, of which 42 remained significant after Bonferroni correction. As with departure from HWE, LD can be caused by population structure, and indeed only one pair out of the 42 (UNH2127 and UNH2128) was found to be in disequilibrium in more than one "species" (*Tropheops gracilior* and *T. microstoma*). It is noteworthy to mention that four pairs of markers are known to belong to the same linkage group (Albertson et al., 2003) and that none of these showed significant evidence for LD in any of the eight species used for identifying problematic loci. Overall, we interpreted these linkage analysis results as consistent with loci being effectively unlinked.

3.2 | Numbers of genetic clusters, and correspondence to morphospecies

Table 1 summarizes the partition of individuals into genetic clusters and the correspondence between those clusters and morphospecies for clusters with more than one individual for the various methods applied in this study. Overall, four clusters, roughly corresponding to *Metriaclima zebra*, *Tropheops gracilior*, *Tropheops* "orange chest" and *Tropheops microstoma*, are concordantly recovered by all methods. With respect to the remaining population structure the methods differ to a great extent. Mixture analyses in *BAPS* suggested that the most likely number of clusters in the sample was 49; however, partitions with 48 and 50 clusters were recovered with similar probability (0.381, 0.269 and 0.346, respectively). For $K = 49$, *BAPS* recovers several clusters that correspond well to morphospecies, including (in addition to Clusters 1–4 mentioned above), Clusters 5 (corresponding to *Petrotilapia genalutea*), 6 (*Genyochromis mento*), 8 (*Protomelas fenestratus*), 9 (*Labeotropheus fuelleborni*), 10 (*Melanochromis vermicorus*), 12 (*Melanochromis auratus*), 12 (*Ctenopharynx pictus*), 15 (*Otopharynx heterodon*), 16 (*Aulonocara jacobfreibergi*) and 17 (*Tyrannochromis nigriventer*). For all of these genetic clusters, a single morphospecies accounts for 67% or more of the individuals. And of these modal species that make up a majority of a single genetic cluster, the percentage of individuals that do not fall within that cluster ranges from 0% to 33%. There are also some genetic clusters recovered by *BAPS* (Clusters 7, 11, 13 and 18 in Table 1) that include individuals from multiple morphospecies and for which the

corresponding morphospecies is unclear. Based on admixture analyses, a total of 21 individuals were not assigned to a single cluster with over 0.74 posterior probability, the threshold defined for this method. Remarkably, 29 out of the 49 clusters recovered by *BAPS* (59%) consist of a single individual ("singleton cluster").

Analyses using *STRUCTURE* and the CAF model recovered 22 groups when choosing the run with the highest $\ln(\Pr(X|K))$ and just two groups using the method of Evanno et al. (2005). Because the latter seemed biologically more improbable than the former, we chose to analyse results only for $K = 22$. Although the assignment of some individuals differed between the two methods (see Figure 1), in general there is a high degree of correspondence between the groups defined by *STRUCTURE* CAF and those recovered by *BAPS*. In general *STRUCTURE* fails to classify more individuals (78 were classified as "admixed" based on the threshold defined), which translates into a lower degree of correspondence between genetic clusters and morphospecies. From the morphospecies picked up as distinct by *BAPS*, only *O. heterodon*, *A. jacobfreibergi* and *T. nigriventer* (those that are recovered with two or three individuals) were not recovered as independent groups by *STRUCTURE*. Also, some *BAPS* clusters without obvious morphological correspondence are also recovered by *STRUCTURE* (e.g., Clusters 11 in both methods or Cluster 7 in *BAPS*/8 in *STRUCTURE* or Cluster 13 in *BAPS*/14 in *STRUCTURE*). Bootstrap support was generally lower than in the case of *BAPS* groups, indicating somewhat less robust assessments. Two clusters included a single individual (both morphologically assigned to *M. zebra*) and four of the 22 clusters included only individuals with posterior probability of belonging to a single clade below the threshold we defined for this method (0.61). In three of the clusters a few individuals were found with admixture proportions between 0.5 and 0.61, but cluster 22 did not include a single individual with an admixture proportion higher than 0.5.

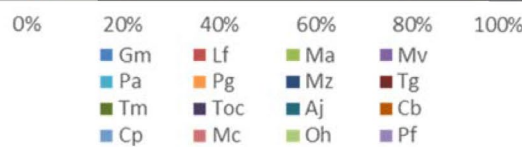
The CAF results were quite different from those obtained from *STRUCTURE* with the IAF model. In this case the program recovered just seven different groups using the run with the highest $\ln(\Pr(X|K))$ —and again two groups using the Evanno et al. (2005) method—far fewer than the number of morphospecies in the sample. From Table 1, Clusters 1–3 and 5 are recovered in common with *BAPS* and *STRUCTURE* assuming the CAF model. The remaining groups are amalgamations of morphospecies: Group 4 includes all nonmbuna forms, Group 6 includes mostly *P. genalutea* and *M. vermicorus* and Group 7 includes *G. mento*, *T. gracilior*, *M. auratus* and *M. zebra*. Thirty-seven individuals, including the majority of well-sampled species such as *L. fuelleborni*, were not placed in a single cluster.

DAPC recovered five genetic clusters, four of which roughly correspond to the four major groups identified using both *BAPS* and *STRUCTURE*. All remaining individuals are placed into a single group, except one individual that could not be assigned to a single cluster with a group membership score of over 0.75.

The results obtained with *SNAPCLUST* are very similar to those obtained with DAPC (five clusters with the same rough composition), with differences in the assignment of only six individuals. No individual was classified as admixed using the threshold defined for this method.

TABLE 1 Non-singleton clusters detected using the various software [Colour figure can be viewed at wileyonlinelibrary.com]

Cluster	N	Bootstrap	Distribution of morphospecies in cluster	Modal morpho-species	% in cluster ¹
BAPS					
1	246	100		<i>M. zebra</i> (Mz)	94
2	194	100		<i>T. gracilior</i> (Tg)	86
3	69	96		<i>T. "orange chest"</i> (Toc)	81
4	38	99		<i>T. microstoma</i> (Tm)	89
5	29	90		<i>P. genalutea</i> (Pg)	93
6	14	100		<i>G. mento</i> (Gm)	82
7	13	83		<i>T. gracilior</i> (Tg)	5
8	10	52		<i>P. fenestratus</i> (Pf)	56
9	9	85		<i>L. fuelleborni</i> (Lf)	90
10	7	99		<i>M. vermivorus</i> (Mv)	100
11	6	95		Gm + Mz + Tg	18/0.8/0.9
12	4	72		<i>C. pictus</i> (Cp)	80
13	4	95		<i>P. aurora</i> (Pa)	50
14	4	94		<i>M. auratus</i> (Ma)	100
15	3	77		<i>O. heterodon</i> (Oh)	100
16	2	69		<i>A. jacobfreibergeri</i> (Aj)	100
17	2	85		<i>T. nigriventer</i> (Tn)	100
18	2	73		Mz + Pa	0.4/25
19	2	74		Pf + <i>T. placodon</i>	6/33
20	2	88		Pf + Cp	6/20
STRUCTURE CAF model					
1	216	32		Mz	83
2	192	54		Tg	86
3	72	60		Toc	87
4	38	98		Tm	89
5	27	89		Pg	83
6	18	76		Pf	83
7	13	99		Gm	82
8	13	70		Tg	5
9	8	89		Lf	80
10	8	97		Mv	100
11	6	88		Gm + Mz + Tg	18/0.8/0.9
12	4	82		Cp	80
13	4	95		Ma	100
14	3	58		Pa + Tg	50/0.4
15	3	85		Tg	1.4
16	2	42		Cb + Mc	25/50
STRUCTURE IAF model					
1	252	95		Mz	96
2	199	99		Tg	88
3	83	93		Toc	93
4	42	62		Pf, other non-mbuna	100
5	38	97		Tm	89
6	36	82		Pg + Mv	93/100
7	30	57		Gm + Ma + Mz + Tg	100/100/1.6/4.5
SNAPCLUST					
1	252	94		Mz	96
2	199	99		Tg	88
3	136	70		All other	-
4	85	94		Toc	93
5	38	99		Tm	89
DAPC					
1	252	52		Mz	96
2	199	72		Tg	88
3	136	61		All other	-
4	82	87		Toc	93
5	40	94		Tm	89



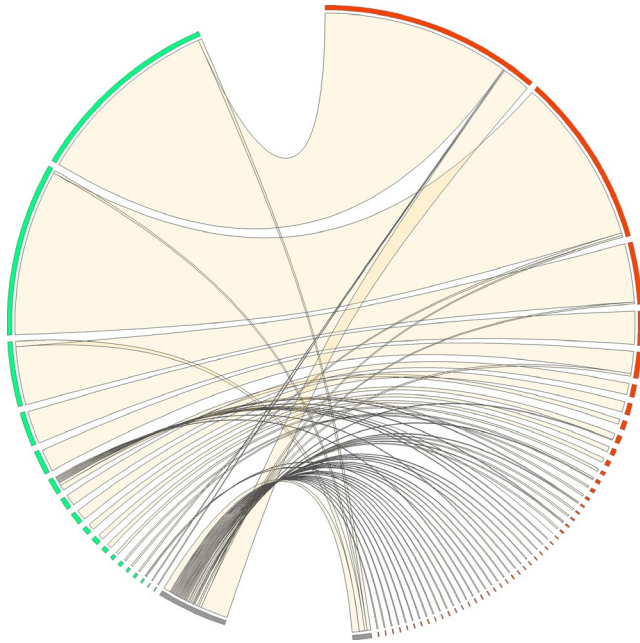


FIGURE 1 Correspondence between the clusters revealed by BAPS (in red) and STRUCTURE using the correlated allele frequency model (in green). Clusters are ordered by their number (with larger clusters ranked first; see Tables 1 and the main text for a description of each group). Individuals not assigned to any of the clusters (“admixed” individuals) are shown in grey as a last group of individuals for both methods. Correspondence between individuals is shown by means of ribbons linking clusters from the two methods [Colour figure can be viewed at wileyonlinelibrary.com]

As a measure of differentiation of these different clusters, we calculated F_{ST} based on the estimated clusters using the R package *hierfstat*. For BAPS, F_{ST} was 0.111 and for STRUCTURE with the CAF model F_{ST} was 0.106. For the other methods, all of which gave many fewer clusters, F_{ST} estimates are lower: STRUCTURE with the IAF, $F_{ST} = 0.090$; SNAPCLUST, $F_{ST} = 0.084$; and DAPC, $F_{ST} = 0.084$. These lower values probably reflect the high within-group variation of some of the clusters. Overall F_{ST} among morphospecies was 0.0927.

These clustering analyses also uncovered an important omission in our morphology-based identification of morphospecies in that we failed to include the taxon *Pseudotropheus elongatus* for consideration. From previous experience and taxonomic descriptions adult male *P. elongatus*, which resemble those of *T. gracilior*, had been reported near the collecting area. Upon realizing that BAPS and STRUCTURE had identified two large separate clusters made up primarily by individuals identified as *T. gracilior*, we returned to the photographs and the taxon list and realized that many male individuals that had been identified as *T. gracilior* in BAPS group 7 and in STRUCTURE group 8 (Table 1) were more consistent with the appearance of adult male *P. elongatus*. In the case of STRUCTURE with independent allele frequencies, these males that should have been identified as *P. elongatus*, but were identified as *T. gracilior*, fell into group 7 and in the case of DAPC and SNAPCLUST they were clustered in group 3 (Table 1).

Correspondence between the assignment obtained in BAPS and STRUCTURE with the CAF model is depicted in Figure 1, which was

drawn using software CIRCOS (Krzywinski et al., 2009). For the correspondence between other pairs of methods, figures are presented as Supporting information.

Given these important differences in the number of populations suggested by different methods, which probably reflect different abilities in coping with highly unbalanced sample sizes, for methods that probably underestimated K (STRUCTURE IAF, DAPC and SNAPCLUST) we removed from the data sets individuals allocated to the four larger clusters—i.e., those that were recovered in common by all methods. In the case of DAPC and SNAPCLUST we removed all individuals except those allocated to cluster 3, which corresponds to an amalgamation of other clusters detected by the other methods. STRUCTURE IAF clearly detects more of the remaining clusters; in a set of individuals where it had previously inferred only three clusters, the optimal K for these runs was $K = 7$. Clusters roughly corresponding to *P. genalutea*, *G. mento*, *M. vermivorus*, *M. auratus*, *L. fuelleborni* and *P. elongatus* can now be recovered, as well as a cluster including *P. fenestratus* and other nonmbuna. Removing the four well-defined clusters, DAPC now recovered three clusters within former Cluster 3: one includes only *G. mento*, another includes *L. fuelleborni*, *M. auratus* and roughly all *P. fenestratus* and other nonmbuna, and a third cluster includes the remaining individuals. SNAPCLUST, on the other hand, recovered two clusters: one includes *L. fuelleborni*, *M. auratus* and all nonmbuna, while the remaining individuals are placed in a separate cluster. Overall these analyses suggest that these three methods may recover additional population structure when dominant clusters are removed, but that even so they fail to recover clusters that other methods identify without correcting for biased sample sizes.

3.3 | Simulations

To assess the relative performance of the methods for cases where we know the true number of genetic clusters, we simulated 100 data sets that resemble the actual data and 100 additional data sets extending the number of loci to 100. Although we do not know the true number of clusters in our actual data, it seems probable that size distribution is highly skewed, and we included this kind of distribution in our simulations of 25 populations.

As shown in Table 2, SNAPCLUST, DAPC and to a lesser extent STRUCTURE with independent allele frequencies tend to severely underestimate the real number of clusters, and this is true both for data sets with 22 loci and data sets with 100 loci. DAPC recovers on average 5.57 and 5.13 clusters with 22 and 100 loci, respectively, and SNAPCLUST's performance also does not seem to be affected by the number of loci, recovering 7.16 and 7.24 clusters on average based on 22 and 100 loci, respectively. STRUCTURE using independent allele frequencies, on the other hand, improves its performance in data sets with 100 loci, suggesting it might even recover more complex structures in larger sized data sets.

Both BAPS and STRUCTURE with the CAF model produce clusterings with Rand Indexes close to 1, suggesting that these two methods perform well in assigning simulated individuals to their correct populations of origin. However, their performance is not

TABLE 2 Recovered population structure based on simulated data sets of 22 and 100 loci with $K = 25$ and unbalanced sample sizes

Method	Data set	Mean K (min–max)	Mean RI (min–max)	Mean Q_{\max} (min–max)
DAPC	22 loci	5.57 (5–7)	0.958 (0.914–0.982)	1.000 (0.999–1.000)
	100 loci	5.13 (5–6)	0.956 (0.953–0.967)	1.000 (1.000–1.000)
SNAPCLUST	22 loci	7.16 (6–8)	0.977 (0.962–0.986)	1.000 (0.999–1.000)
	100 loci	7.24 (6–8)	0.980 (0.967–0.987)	1.000 (1.000–1.000)
STRUCTURE	22 loci	8.9 (7–11)	0.990 (0.983–0.996)	0.950 (0.937–0.964)
IAF model	100 loci	16.13 (10–24)	0.999 (0.991–1.000)	0.972 (0.956–0.980)
STRUCTURE	22 loci	21.88 (14–30)	0.997 (0.959–0.999)	0.886 (0.850–0.918)
CAF model	100 loci	23.13 (15–29)	1.000 (0.999–1.000)	0.942 (0.883–0.969)
BAPS	22 loci	32.11 (27–39)	0.998 (0.993–1.000)	0.994 (0.990–0.998)
	100 loci	27.31 (25–31)	1.000 (0.999–1.000)	0.997 (0.995–0.999)

Note: K , number of populations; RI, Rand (1971) index; Q_{\max} , mean maximum individual probability of assignment to a cluster.

completely similar: despite the difference in mean Rand Index is small for data sets with 22 loci (0.998 for BAPS versus 0.997 for STRUCTURE), the difference in this assignment success measure is significant ($p = 9.41 \times 10^{-6}$, paired Wilcoxon signed ranks test). In practice, BAPS tends to overestimate the number of clusters (the mean value of K inferred for the data set with 22 loci was 32.11 with a real K of 25). The high number of singleton clusters recovered by BAPS in our data and the observed tendency for this software to overestimate the number of clusters led us to further investigate whether this overestimation results from the spurious estimation of singleton clusters. In all 100 simulated data sets, BAPS recovered spurious (i.e., nonexistent) singleton clusters (anywhere from two to 17, with an average of 7.99). Removing these spurious singleton clusters, the number of clusters recovered by BAPS would be between 21 and 28, with an average of 24.12 (i.e., much closer to the real simulated data). This suggests that the overestimation of K observed in BAPS is largely due to the spurious allocation of individuals to singleton clusters. It should be noted that our simulated data sets included four real singleton clusters which, in general, BAPS recovers rather well (an average of 3.84 of the four real singleton clusters are indeed recovered as such). Increasing the number of loci to 100, BAPS' performance improves, in terms of both Rand Index and the lower tendency to allocate individuals to nonexistent singleton clusters. This problem still persists but on average BAPS now recovers 4.13 spurious singleton clusters, which is reflected in a more accurate overall estimate of K . Using this set BAPS recovers the four real singleton clusters as such in 100% of the simulations.

In contrast to BAPS, STRUCTURE with the CAF model tends to slightly underestimate K (the mean value for K using the 22-loci data set was 21.88, ranging between 14 and 30). Intriguingly, some of these clusters do not include any individual assigned with a high posterior probability, a feature that mirrors the findings in our empirical data set. On average, STRUCTURE recovered 2.57 empty clusters (when considering a threshold for the posterior probability of 0.50), which suggests that the number of actual clusters recovered by this method is in practice lower than the number of reported clusters.

Correcting for empty clusters, the detected K value is on average 19.34. Increasing the number of loci to 100, STRUCTURE also improves its estimates slightly, recovering fewer empty clusters (1.95 on average), increasing the Rand Index and recovering a more accurate average K (23.13 on average, 21.17 when correcting for empty clusters). Another important difference between STRUCTURE and the other methods is a much noisier assignment; individuals were generally assigned with less confidence than using BAPS, both in the 22- and in the 100-loci data sets (the mean maximum assignment proportion was 0.886 compared to 0.994 using BAPS for the 22-loci set, and an attenuated difference from 0.942 of STRUCTURE to 0.997 for BAPS using the 100-loci set). Unlike BAPS, STRUCTURE did not recover real singleton clusters (it recovered on average 0.25 and 0.61 of the four simulated singleton clusters using the 22- and the 100-loci data set, respectively).

In summary, both BAPS and STRUCTURE with the CAF model perform much better than the alternatives, but with problems specific to each method—the tendency to overestimate K by assuming nonexistent singleton clusters in the case of BAPS, and both the tendency to have empty clusters and an increase in the overall noise (lowering the ability to detect hybridization events, for example) in the case of STRUCTURE.

All inferences using STRUCTURE mentioned above were performed using the run with the highest value of $\ln(\Pr(X|K))$. We also considered the method of Evanno et al. (2005) for selecting K but our simulations showed that this method clearly underestimated K in all STRUCTURE runs (the average K was always below 5 with or without correlated allele frequencies or with 22 or 100 loci).

3.4 | Evidence of hybridization between species

Given the problems exhibited by the various methods in detecting the real underlying structure, detecting actual hybrids is a difficult task. This is further complicated by highly discordant sets of individuals highlighted as putatively admixed based on their probability of assignment to more than a single cluster by the different methodologies.

The number of individuals not assigned to a single cluster based on the thresholds defined for each method varied considerably depending on the method: 78 in STRUCTURE (CAF model), 30 in STRUCTURE (IAF model), 21 in BAPS, one in DAPC and zero in SNAPCLUST. In total, 103 individuals were classified as admixed based on the thresholds defined at least by one of the methods. There was little overlap in the individuals based on different methods: one individual appears as admixed according to three different methods (DAPC and the two implementations of STRUCTURE), eight individuals are inferred as admixed according to both BAPS and STRUCTURE with the CAF model, and 18 individuals (none in common with BAPS) are not assigned to a single cluster based on both implementations of STRUCTURE.

Unlike the other methods, BAPS incorporates a test of whether an individual is a hybrid based on the simulation of pure genotypes. Of the 21 identified as admixed by BAPS, 12 were indicated as hybrids by BAPS. Also in the case of BAPS, only four of the 21 putatively admixed individuals had more than 10% (and the highest of those had no more than 13%) of assignment probability to a second cluster; and in general the assignment probabilities were scattered throughout multiple clusters without a pattern suggestive of hybridization.

The analysis of patterns of possible hybridization is more complicated in STRUCTURE with the CAF model because of the overall noisier pattern (the maximum average Q was 0.81, considerably lower than other methods), making it harder to disentangle lack of power to assign individuals to clusters from real admixture. Of the 78 individuals with putative admixture, 52 had a pattern that might fit with hybridization between two clusters (i.e., with large contributions from two major clusters).

3.5 | Do male-biased sample schemes affect the inference of population structure?

We used a reduced sample of 292 adult males to test three hypotheses: (a) that a sample of only adult males will have higher mean maximum assignment proportions to a single cluster (Q_{\max}); (b) that a sample of

only adult males will have fewer admixed individuals; and (c) that there will be fewer inferred clusters when only adult males are used. For all three hypotheses, the null hypothesis is that there will be no difference between a sample of only adult males and a random sample of the same size that includes males, females and juveniles. The results of these analyses are shown in Table 3. Only the last test, of the hypothesis of fewer clusters inferred for an adult male-only sample, was supported by the analyses, and in that case only when using BAPS. In this case none of the 100 simulated data sets had an equal or fewer number of clusters than the male-only data set.

We also tried to evaluate whether such sampling schemes are even justified by supposed difficulties in identifying females and juveniles. When asking "are females and juveniles more prone to be misidentified?" the answer from our data is negative: the number of males misidentified according to both the two best-performing methods (18 out of 249 sexed individuals from well-defined species) and the number of females or juveniles misidentified (22 out of 340) are not significantly different ($\chi^2 = 0.038922$, 1 *df*, $p = 0.8849$).

Finally we inquired in a different way regarding male-based taxon identification by asking whether identification to morphospecies was independent of whether an individual was identified as a male. For the largest five BAPS clusters there were 242 male individuals identified to the predominant species in that group, and 312 females and juveniles identified to the predominant species in that group. However, among individuals that were not identified to the predominant species there were fewer males (four males and a total of 16 females and juveniles). A test of independence of male status and predominant species taxon assignment is rejected ($\chi^2 = 4.4207$, 1 *df*, $p = 0.03551$). If we consider the same kind of analysis using groups recovered by STRUCTURE with the CAF model the difference is even more pronounced (three males misidentified out of 230 and 15 out of 295 females, $\chi^2 = 5.2339$, 1 *df*, $p = 0.02215$). This kind of result is consistent with male individuals being assigned to the correct morphospecies more easily than females and juveniles, although the overall interaction effect is quite modest.

	BAPS	STRUCTURE (CAF model)	STRUCTURE (IAF model)	SNAPCLUST	DAPC
Mean maximum assignment proportion					
Males only	0.99	0.91	0.95	1.00	1.00
Mean resampled data sets	1.00	0.89	0.95	1.00	1.00
p (resampled >only males)	0.65	0.36	0.15	0.11	0.14
Number of admixed individuals					
Males only	1	5	17	0	0
Mean resampled data sets	2.1	11.28	13.84	0.09	0.11
p (resampled ≤only males)	0.39	0.09	0.84	0.91	0.89
Number of inferred clusters					
Males only	25	9	6	3	3
Mean resampled data sets	37.64	10.55	6.4	3.09	3.25
p (resampled >only males)	0	0.59	0.4	0.09	0.25

Bold values indicates cases where resampled data sets differ from the "only males" data set, $p < 0.05$.

TABLE 3 Comparison between data sets comprising only males and resampled data sets containing the same number of individuals ($N = 292$)

4 | DISCUSSION

We assessed the correspondence between population genetic entities and morphospecies in the highly diverse mbuna cichlid complex of Lake Malawi. Previous studies that have assessed the correspondence between genetic clusters and morphospecies have typically found a high level of correspondence (Altermann, Leavitt, Goward, Nelsen, & Lumsch, 2014; Leray et al., 2010; Pinho, Harris, & Ferrand, 2007; Sekino & Hara, 2007). One outlier to this pattern comes from oaks where the levels of hybridization are quite high (Lepais et al., 2009).

4.1 | Diverse methods for identifying genetic populations give widely different results

In this study we used various different approaches to estimate genetic discontinuities within a sympatric community of cichlid species. Each of the five different approaches yielded a very different partition of specimens into genetic clusters, recovering from five to 49 different clusters depending on the method used and suggesting widely differing degrees of admixture between clusters.

We also found qualitatively similar (i.e., large) differences among methods when using simulated data sets, with BAPS consistently overestimating the true number of clusters (K), STRUCTURE with the CAF model slightly underestimating it and STRUCTURE assuming IAF, SNAPCLUST and DAPC greatly underestimating K . That these patterns were found in the simulated data suggests that the true number of genetic populations in the real data lies between the higher BAPS count and the lower counts for the other methods. However, we cannot rule out that our simulated data are dissimilar from the actual data in important ways that bear on this interpretation and that we cannot discern.

While it is possible that more genetic markers would greatly improve the correspondence of the methods, this seems unlikely given the high bootstrap support found for most clusters with each method and simulations showing that some of the methods do not improve their performance with larger data sets.

Unlike BAPS, STRUCTURE and SNAPCLUST, DAPC does not rely upon genetic disequilibria. Given the model-free aspect of DAPC it is perhaps not surprising that the returned estimates of K for simulated data were far from the true value; the simulated data were very much the result of a specific population genetic model (i.e., as implemented in the EASY-POP software; Balloux, 2001). However, this does not explain the overall similar results obtained for SNAPCLUST, which does rely on a model similar to that of BAPS and STRUCTURE.

The genetic model underlying BAPS, STRUCTURE with independent allele frequencies and SNAPCLUST is a very simple one that assumes only that individuals are made up of independent draws of alleles from a set of allele frequencies. These two clustering methods, and a number of other related methods (Alexander, Novembre, & Lange, 2009; Cheng, Mailund, & Nielsen, 2017; Huelsenbeck & Andolfatto, 2007; Raj, Stephens, & Pritchard, 2014; Tang, Peng, Wang, & Risch, 2005), make no other population genetic assumptions,

and they are inherently ahistorical, as they make no reference to ancestors or descendants or to any phylogenetic component of history (Lawson, Dorp, & Falush, 2018; Liang & Nielsen, 2014; Novembre, 2014; Wang, 2017). The lack of a historical phylogenetic component is a considerable shortcoming, as real data sets are sampled from populations for which the allele frequencies have strong correlations as a result of the phylogenetic history of the populations. To help account for this, methods such as the correlated allele frequency model in STRUCTURE (Falush et al., 2003) implement allele frequency priors that introduce correlations in allele frequencies. This method assumes that populations have had a star phylogeny, but with varying amounts of genetic drift. This method performed slightly better than BAPS, for example, in estimating K ; however, this does not translate into an increased performance in clustering success, for example, and it suffers from other problems that do not affect other methods, such as a lower resolution in the assignment proportions which hampers admixture analysis or the presence of clusters with no individuals fully assigned to them.

Our cichlid data probably present a highly challenging problem in that (a) they have a recent phylogenetic history (Salzburger, 2018) and (b) the sample sizes vary greatly, with some apparent groups having just a single individual. In particular, the lack of a phylogenetic component in their respective models probably underlies the generally poor performance of these methods for estimating K . The consistent finding from simulation studies is that when sample sizes are unequal or the data have a substantial phylogenetic component, estimates of K are poor (Kalinowski, 2011; Neophytou, 2014; Puechmaile, 2016; Wang, 2017). Those simulation studies that do find a good correspondence between true and estimated K values have used an island model of divergence that inherently lacks phylogenetic structure (Evanno et al., 2005; Latch, Dharmarajan, Glaubitz, & Rhodes, 2006).

4.2 | Cichlid morphospecies reflect population genetic discontinuities

Although cichlid species are among the best known model systems for the study of speciation (Brawand et al., 2014), few studies have actually addressed *in loco* species differentiation and delimitation based on molecular markers. In general, species of Malawi cichlids are assumed to be valid primarily because they have the ability to maintain phenotypic distinction in sympatry. Our results show that, notwithstanding the difficulties pointed out in the section above, the clusters obtained by the different methods correspond roughly to named species. This is true especially for the best sampled and easier to identify species of mbuna, which correspond rather well to genetic discontinuities across all three methods, but it is also true for less well sampled species, at least using BAPS, the method that, if we ignore the tendency for recovering spurious singleton clusters, seems to be performing better in recovering clusters with low sample sizes. Fifteen of the 20 non-singleton clusters recovered by BAPS correspond rather well to named species. Using this method, and focusing on morphological species sampled with $N > 5$ and discarding singleton cluster allocations, we had a misidentification rate based on genetic assignment of ~7.5%,

which is not high considering the diversity of species, some probably cryptic, and the fact that photographs were used for the identification, not the actual individuals. Plus, although this concordance should be taken with caution given the results of our simulation study, of the 11 morphospecies with sample sizes of 1 in our data set, nine are indeed recovered as singleton species by BAPS.

This overall correspondence between morphological species and genetic clusters is reflected in the lack of evidence for both (a) intra-specific phenotypic morphs and (b) cryptic species. In particular in the case of BAPS and STRUCTURE with the CAF model we did not find evidence for genetic clusters clearly consisting of individuals of two or more assigned morphospecies (i.e., based on photographs), except Cluster 11 in both methods. We also did not find examples of genetic clusters that include just a subset of a morphospecies as would be expected in the presence of cryptic species (the only case would be STRUCTURE Cluster 15, which we do not find in BAPS). A few of the groups recovered do not correspond to a morphospecies, but in these cases they are not composed mostly of members of a morphospecies that also includes individuals that constitute the majority of an additional cluster. The one case that at first suggested a cryptic morphospecies (BAPS group 7 and STRUCTURE with CAF model group 8 in Table 1) was found to be largely explained by a failure to include *P. elongatus* in our list of morphospecies to be considered (see Section 3).

Overall the picture is one in which most of the species-level diversity within cichlids can be identified by the use of morphology and colour (as represented in the photographs used in this study). This support for morphology, and particularly colour in the case of adult males, reinforces the idea of morphological adaptation and sexual selection on male nuptial coloration as major drivers of speciation in this system (Albertson et al., 2003; Allender, Seehausen, Knight, Turner, & Maclean, 2003). This result does not match the high prevalence of cryptic species reported by Kidd et al. (2006) in bower-building cichlids of Lake Malawi, suggesting that the utility of morphological characters for species diagnosis—and ultimately the major forces driving speciation—might differ in different groups of cichlids.

A clear correspondence between phenotypic and genetic clusters was also reported by Mims, Darrin Hulsey, Fitzpatrick, and Todd Strelman (2010) in a study of two mbuna species (*Labeotropheus fuelleborni* and *Metriaclima zebra*). Overall, these two species can be clearly differentiated using genetic markers, although various aspects of the study point to the existence of gene flow between them.

4.3 | Absence of evidence for significant recent hybridization

Cichlids of the African rift valley are clearly capable of exchanging genes, as evidenced by widespread interfertility (Albertson et al., 2003; McElroy & Kornfield, 1990; Turner et al., 2001) and some cases of hybrids in nature (Smith, Konings, & Kornfield, 2003; Stauffer, Bowers, Kocher, & McKaye, 1996; Strelman et al., 2004). More recently, genomic studies show high levels of allele sharing between species and genera (Loh et al., 2013; Malinsky et al., 2017), consistent with the low F_{ST} levels that are typically reported

between populations and species (Arnegard et al., 1999; Danley, Markert, Arnegard, & Kocher, 2000; Rico et al., 2003).

However, we found no clear evidence of recent hybridization. Certainly we found some individuals that did not cluster well using one of our methods. However, the overlap among methods of candidate “admixed” individuals was very low, and a detailed analysis of these candidates did not suggest that the apparent pattern of admixture was consistent with a recent hybridization event for most of them. Hybridization thus seems to be a very restricted phenomenon in this system.

4.4 | A critical evaluation of male-biased sampling schemes

While the diagnostic criteria of mbuna genera are primarily skeletal and dental (Trewavas, 1935), most morphospecies within mbuna genera have been identified on the basis of nuptial coloration (Kocher, 2004; Konings, 2001; Ribbink, Marsh, Marsh, & Sharp, 1983). This focus on coloration at the species level is also supported by research suggesting that rapid evolution of male coloration is a primary contributor to reproductive isolation (Kocher, 2004; Kornfield & Smith, 2000). The evolutionary focus on male coloration, as well as the ease of species identification in the field, necessarily leads investigators seeking samples from particular species to sample primarily males showing representative coloration. By investigating genetic populations without regard to male coloration we are able to assess whether such a male-focused sampling scheme has a downside, i.e., by undersampling individuals that might otherwise show important patterns of genetic variation. Sampling individuals that are uniform in appearance not only might reduce our notion of intraspecific morphological variation, but also could bias a study towards individuals that are also less admixed genetically, and hence provide an incomplete picture on how distinct species really are.

Our results overall, using the subset of adult males in comparison with randomized subsets that include all kinds of individuals, revealed little evidence that a male-limited sample leads to different conclusions than one that would be found with a broader sampling design. The male-only samples did not show lower assignment proportions, and they did not show more cases of apparent admixture. However, in the case of BAPS analyses we did see a lower number of estimated genetic clusters for the male-only sample. This puzzling finding suggests the presence of groups for which males went unrecognized in our protocol.

We also observed that nonmale individuals were largely placed in the same taxa as males. We did not observe a difference between males and nonmales in their assignment to genetic clusters; individuals of both groups saw very similar levels of assignment to the modal cluster for their respective species taxon ID. However, a different test of independence of male/nonmale status and assignment found a small yet significant effect. Nonmale individuals were identified to species taxa other than the dominant taxon for their respective genetic cluster at higher rates, although the effect was modest. On balance species taxa were identified with remarkable independence

of whether an individual was male, which is all the more surprising considering that these identifications were done using photographs.

5 | CONCLUDING REMARKS

The overall impression of the population genetic and taxonomic status of Lake Malawi mbuna is one in which apparent species (i.e., morphospecies based on nuptial coloration) are well-demarcated genetic entities, and that investigators are able to clearly observe and identify these discontinuities within a sample, relying or not on typical (i.e., male) individuals. This common type of sampling scheme, although possibly problematic for some questions, does not appear to affect our perception of genetic discontinuities between species of Lake Malawi mbuna cichlids.

Although we did not find clear evidence of recent hybridization, we did observe high levels of allele sharing and a number of sampled individuals presented difficulties for different clustering methods. As only a modest amount of gene exchange can cause high levels of allele sharing, our results are not inconsistent with occasional hybridization between species. Thus, our findings do not rule out a significant role for gene exchange in the evolution of these species, and they are not inconsistent with these species conforming poorly to strictly bifurcating phylogenetic models.

Our results demonstrate that widely used clustering methods have important limitations, particularly in settings with high *K* and highly unbalanced sample sizes. Importantly these findings also hold for the well-studied STRUCTURE program, even when using mitigating strategies such as those suggested by Wang (2017).

Cichlids from the African great lakes are a unique system in the sense that they represent a large assemblage of closely related, often sympatric species evolving under nonnegligible levels of gene flow, which poses well-known difficulties for species identification and for the assessment of their evolutionary history (Loh et al., 2013; Malinsky et al., 2017; Salzburger, 2018). However, the results of our study are potentially relevant for other biological systems where the number of real clusters is unknown and/or the sample sizes are naturally unbalanced. Examples of this situation are other cichlid fish assemblages (e.g., in crater lakes; Elmer et al., 2014; Martin et al., 2015) or other recent sympatric species complexes (Levitan et al., 2004; Lin & Wood, 2002; Suatoni, Vicario, Rice, Snell, & Caccone, 2006). The limitations of the methods are also not restricted to species-level differentiation; studies on assortative mating between host races, colour morphs or ecotypes (e.g., Ólafsdóttir, Ritchie, & Snorrason, 2006; Reynolds & Fitzpatrick, 2007) or other kinds of intraspecific discontinuities may be similarly affected by these challenges.

ACKNOWLEDGEMENTS

We thank Richard Zatha, James W. Maluza, Stella Capoccia and Yong-Jin Won for assistance during the sampling campaign. We are further extremely grateful to Richard and James for the long hours spent

identifying fish from photographic material. We thank Susana Lopes for help in laboratory work. We would also like to thank Jorge Fernandes for computer assistance. We further thank Jukka Corander for helpful discussion related to BAPS software, and to Rui Faria and Xana Sá Pinto for valuable suggestions. We also thank three anonymous reviewers. This work was partially supported by FEDER through the COMPETE programme and Portuguese national funds through the FCT (Fundação para a Ciência e a Tecnologia) PTDC/BIA-BDE/66210/2006 research project and by Programa Operacional Potencial Humano—Quadro de Referência Estratégico Nacional funds from the European Social Fund and Portuguese Ministério da Educação e Ciência through an IF contract (IF/01597/2014) to C.P., and by NIH Grant R01GM078204 to J.H.

AUTHOR CONTRIBUTIONS

This study was designed by J.H. and C.P. Both also participated in sample collection. V.C. produced genetic data from the samples. C.P. analysed the data with input from J.H. C.P. and J.H. wrote the article.

DATA ACCESSIBILITY

Data summaries, scripts and simulation results are available at <https://github.com/catpinho>. The data are available at <https://doi.org/10.5061/dryad.258nm86>.

ORCID

Catarina Pinho  <https://orcid.org/0000-0001-5150-2850>

Jody Hey  <https://orcid.org/0000-0001-5358-6488>

REFERENCES

- Albertson, R. C., Streelman, J. T., & Kocher, T. D. (2003). Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences USA*, *100*, 5252–5257. <https://doi.org/10.1073/pnas.0930235100>
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Allender, C. J., Seehausen, O., Knight, M. E., Turner, G. F., & Maclean, N. (2003). Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 14074–14079. <https://doi.org/10.1073/pnas.2332665100>
- Altermann, S., Leavitt, S. D., Goward, T., Nelsen, M. P., & Lumbsch, H. T. (2014). How do you solve a problem like *Letharia*? A new look at cryptic species in lichen-forming fungi using Bayesian clustering and SNPs from multilocus sequence data. *PLoS ONE*, *9*, e97556. <https://doi.org/10.1371/journal.pone.0097556>
- Arnegard, M. E., Markert, J. A., Danley, P. D., Stauffer, J. R., Ambali, A. J., & Kocher, T. D. (1999). Population structure and colour variation in the cichlid fish *Labeotropheus fuelleborni* Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. *Proceedings of the Royal Society, London, Series B*, *266*, 119–130.
- Balloux, F. (2001). EASYPOP (version 1.7): A computer program for population genetics simulations. *Journal of Heredity*, *92*, 301–302. <https://doi.org/10.1093/jhered/92.3.301>

- Beugin, M. P., Gayet, T., Pontier, D., Devillard, S., & Jombart, T. (2018). A fast likelihood solution to the genetic clustering problem. *Methods in Ecology and Evolution*, 9, 1006–1016. <https://doi.org/10.1111/2041-210X.12968>
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., ... Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22, 148–155. <https://doi.org/10.1016/j.tree.2006.11.004>
- Bootton, G. C., Kaufman, L., Chandler, M., & Fuerst, P. A. (1996). Use of DNA microsatellite loci to identify populations and species of Lake Victoria haplochromine cichlids. In E. M. Donaldson, & D. D. MacKinlay (Eds.), *Symposium Proceedings, International Congress on the Biology of Fishes* (pp. 105–113). Vancouver, CA: American Fisheries Society, Physiology Section.
- Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., ... Di Palma, F. (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513(7518), 375–381. <https://doi.org/10.1038/nature13726>
- Cheng, J. Y., Mailund, T., & Nielsen, R. (2017). Fast admixture analysis and population tree estimation for SNP and NGS data. *Bioinformatics*, 33, 2148–2155. <https://doi.org/10.1093/bioinformatics/btx098>
- Corander, J., & Marttinen, P. (2006). Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology*, 15, 2833–2843. <https://doi.org/10.1111/j.1365-294X.2006.02994.x>
- Corander, J., Marttinen, P., Sirén, J., & Tang, J. (2008). Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, 9, 539. <https://doi.org/10.1186/1471-2105-9-539>
- Cracraft, J. (1983). Species concepts and speciation analysis. *Current Ornithology*, 1, 159–187.
- Danley, P., Markert, J., Arnegard, M., & Kocher, T. (2000). Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution*, 54, 1725–1737. <https://doi.org/10.1111/j.0014-3820.2000.tb00716.x>
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Elmer, K. R., Fan, S., Kusche, H., Luise Spreitzer, M., Kautt, A. F., Franchini, P., & Meyer, A. (2014). Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nature Communications*, 5, 5168. <https://doi.org/10.1038/ncomms6168>
- 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. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of Population Structure Using Multilocus Genotype Data. Linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.
- Genner, M. J., Seehausen, O., Cleary, D. F. R., Knight, M. E., Michel, E., & Turner, G. F. (2004). How does the taxonomic status of allopatric populations influence species richness within African cichlid fish assemblages? *Journal of Biogeography*, 31, 93–102. <https://doi.org/10.1046/j.0305-0270.2003.00986.x>
- Goudet, J. (2001). *FSTAT, a program to estimate and test gene diversities and fixation indices* (version 2.9. 3). 2001. <http://www2.unil.ch/popgen/softwares/fstat.htm>
- Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Resources*, 5, 184–186.
- Huelsenbeck, J. P., & Andolfatto, P. (2007). Inference of population structure under a Dirichlet process model. *Genetics*, 175, 1787–1802. <https://doi.org/10.1534/genetics.106.061317>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94. <https://doi.org/10.1186/1471-2156-11-94>
- Joo, D., Kwan, Y.-S., Song, J., Pinho, C., Hey, J., & Won, Y.-J. (2013). Identification of cichlid fishes from Lake Malawi using computer vision. *PLoS ONE*, 8, e77686. <https://doi.org/10.1371/journal.pone.0077686>
- Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: Simulations and implications for human population structure. *Heredity*, 106, 625. <https://doi.org/10.1038/hdy.2010.95>
- Kellogg, K. A., Markert, J. A., & J. R. Stauffer J, Kocher TD., (1995). Microsatellite variation demonstrates multiple paternity in lekking cichlid fishes from Lake Malawi, Africa. *Proceedings of the Royal Society, London, Series B*, 260, 79–84.
- Kidd, M. R., Kidd, C. E., & Kocher, T. D. (2006). Axes of differentiation in the bower-building cichlids of Lake Malawi. *Molecular Ecology*, 15, 459–478. <https://doi.org/10.1111/j.1365-294X.2005.02787.x>
- Kocher, T. D. (2004). Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Reviews Genetics*, 5, 288. <https://doi.org/10.1038/nrg1316>
- Konings, A. (1990). *Ad Koning's book of Cichlids and all the other Fishes of Lake Malawi*. Neptune City, New Jersey: TFH Publications.
- Konings, A. (2001). *Malawi Cichlids in their Natural Habitat*, 3rd ed. El Paso, TX: Cichlid Press.
- Konings, A. (2007). *Malawi Cichlids in Their Natural Habitat*, 4th ed. El Paso, TX: Cichlid Press.
- Kornfield, I., & Smith, P. F. (2000). African cichlid fishes: Model systems for evolutionary biology. *Annual Review of Ecology and Systematics*, 31, 163–196. <https://doi.org/10.1146/annurev.ecolsys.31.1.163>
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, 19, 1639–1645. <https://doi.org/10.1101/gr.092759.109>
- Latch, E. K., Dharmarajan, G., Glaubitz, J. C., & Rhodes, O. E. (2006). Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, 7, 295–302. <https://doi.org/10.1007/s10592-005-9098-1>
- Lawson, D. J., Dorp, L. van, & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9(1), 3258. <https://doi.org/10.1038/s41467-018-05257-7>
- Lepais, O., Petit, R. J., Guichoux, E., Lavabre, J. E., Alberto, F., Kremer, A., & Gerber, S. (2009). Species relative abundance and direction of introgression in oaks. *Molecular Ecology*, 18, 2228–2242. <https://doi.org/10.1111/j.1365-294X.2009.04137.x>
- Leray, M., Beldade, R., Holbrook, S. J., Schmitt, R. J., Planes, S., & Bernardi, G. (2010). Allopatric divergence and speciation in coral reef fish: The three-spot dascyllus, *dascyllus trimaculatus*, species complex. *Evolution*, 64, 1218–1230.
- Levitán, D. R., Fukami, H., Jara, J., Kline, D., McGovern, T. M., McGhee, K. E., ... Knowlton, N. (2004). Mechanisms of reproductive isolation among sympatric broadcast-spawning corals of the *Montastraea annularis* species complex. *Evolution*, 58, 308–323. <https://doi.org/10.1111/j.0014-3820.2004.tb01647.x>
- Liang, M., & Nielsen, R. (2014). The lengths of admixture tracts. *Genetics*, 197, 953–967. <https://doi.org/10.1534/genetics.114.162362>
- Lin, C. P., & Wood, T. K. (2002). Molecular phylogeny of the North American *Echenopa binotata* (Homoptera: Membracidae) species complex. *Annals of the Entomological Society of America*, 95, 162–171.
- Loh, Y.-H., Bezault, E., Muenzel, F. M., Roberts, R. B., Swofford, R., Barluenga, M., ... Streelman, J. T. (2013). Origins of shared genetic variation in African cichlids. *Molecular Biology and Evolution*, 30, 906–917. <https://doi.org/10.1093/molbev/mss326>
- Maan, M. E., & Sefc, K. M. (2013). Colour variation in cichlid fish: Developmental mechanisms, selective pressures and evolutionary consequences. *Seminars in Cell & Developmental Biology*, 24, 516–528. <https://doi.org/10.1016/j.semcdb.2013.05.003>

- Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2017). Whole genome sequences of malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*. Retrieved from <https://www.nature.com/articles/s41559-018-0717-x>
- Martin, C. H., Cutler, J. S., Friel, J. P., Dening Touokong, C., Coop, G., & Wainwright, P. C. (2015). Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution*, *69*, 1406–1422. <https://doi.org/10.1111/evo.12674>
- McElroy, D. M., & Kornfield, I. (1990). Sexual selection, reproductive behavior, and speciation in the mbuna species flock of Lake Malawi (Pisces: Cichlidae). *Environmental Biology of Fishes*, *28*, 273–284. <https://doi.org/10.1007/BF00751041>
- Mims, M. C., Darrin Hulsey, C., Fitzpatrick, B. M., & Todd Strelman, J. (2010). Geography disentangles introgression from ancestral polymorphism in Lake Malawi cichlids. *Molecular Ecology*, *19*, 940–951. <https://doi.org/10.1111/j.1365-294X.2010.04529.x>
- Neophytou, C. (2014). Bayesian clustering analyses for genetic assignment and study of hybridization in oaks: Effects of asymmetric phylogenies and asymmetric sampling schemes. *Tree Genetics & Genomes*, *10*, 273–285. <https://doi.org/10.1007/s11295-013-0680-2>
- Novembre, J. (2014). Variations on a common STRUCTURE: New algorithms for a valuable model. *Genetics*, *197*, 809–811. <https://doi.org/10.1534/genetics.114.166264>
- Oetting, W. S., Lee, H. K., Flanders, D. J., Wiesner, G. L., Sellers, T. A., & King, R. A. (1995). Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics*, *30*, 450–458. <https://doi.org/10.1006/geno.1995.1264>
- Ólafsdóttir, G. Á., Ritchie, M. G., & Snorrason, S. S. (2006). Positive assortative mating between recently described sympatric morphs of Icelandic sticklebacks. *Biology Letters*, *2*, 250–252. <https://doi.org/10.1098/rsbl.2006.0456>
- Oliver, M. K., & Arnegard, M. E. (2010). A new genus for *Melanochromis labrosus*, a problematic Lake Malawi cichlid with hypertrophied lips (Teleostei: Cichlidae). *Ichthyological Exploration of Freshwaters*, *21*(3), 209–232.
- Pinho, C., Harris, D. J., & Ferrand, N. (2007). Comparing patterns of nuclear and mitochondrial divergence in a cryptic species complex: The case of Iberian and North African wall lizards (Podarcis, Lacertidae). *Biological Journal of the Linnean Society*, *91*, 121–133. <https://doi.org/10.1111/j.1095-8312.2007.00774.x>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.
- Pritchard, J., & Wen, X., Falush, D. (2010). *Documentation for STRUCTURE software, version 2.3*. Chicago, IL: University of Chicago.
- Puechmaile, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*, 608–627.
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*, 573–589. <https://doi.org/10.1534/genetics.114.164350>
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*, 846–850. <https://doi.org/10.1080/01621459.1971.10482356>
- Raymond, M., & Rousset, F. (1995). GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, *86*, 248–249. <https://doi.org/10.1093/oxfordjournals.jhered.a111573>
- Reynolds, R. G., & Fitzpatrick, B. M. (2007). Assortative mating in poison-dart frogs based on an ecologically important trait. *Evolution*, *61*, 2253–2259. <https://doi.org/10.1111/j.1558-5646.2007.00174.x>
- Ribbink, A. J., Marsh, A. C., Marsh, B. A., & Sharp, B. J. (1983). A preliminary survey of the cichlid fishes of the rocky habitats in Lake Malawi. *South African Journal of Zoology*, *18*, 149–310.
- Rico, C., Bouteillon, P., Van Oppen, M. J. H., Knight, M. E., Hewitt, G. M., & Turner, G. F. (2003). No evidence for parallel sympatric speciation in cichlid species of the genus *Pseudotropheus* from northwestern Lake Malawi. *Journal of Evolutionary Biology*, *16*, 37–46. <https://doi.org/10.1046/j.1420-9101.2003.00477.x>
- Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*, 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Rüber, L., Verheyen, E., & Meyer, A. (1999). Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. *Proceedings of the National Academy of Sciences*, *96*, 10230–10235. <https://doi.org/10.1073/pnas.96.18.10230>
- Salzburger, W. (2009). The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology*, *18*, 169–185. <https://doi.org/10.1111/j.1365-294X.2008.03981.x>
- Salzburger, W. (2018). Understanding explosive diversification through cichlid fish genomics. *Nature Reviews Genetics*, *19*, 705–717. <https://doi.org/10.1038/s41576-018-0043-9>
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Seehausen, O. (2006). African cichlid fish: A model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 1987–1998. <https://doi.org/10.1098/rspb.2006.3539>
- Sekino, M., & Hara, M. (2007). Individual assignment tests proved genetic boundaries in a species complex of Pacific abalone (genus *Haliotis*). *Conservation Genetics*, *8*, 823–841. <https://doi.org/10.1007/s10592-006-9229-3>
- Sinervo, B., & Lively, C. M. (1996). The rock–paper–scissors game and the evolution of alternative male strategies. *Nature*, *380*, 240. <https://doi.org/10.1038/380240a0>
- Smith, P. F., Konings, A., & Kornfield, I. (2003). Hybrid origin of a cichlid population in Lake Malawi: Implications for genetic variation and species diversity. *Molecular Ecology*, *12*, 2497–2504. <https://doi.org/10.1046/j.1365-294X.2003.01905.x>
- Stauffer, J. R., Bowers, N. J., Kocher, T. D., & McKaye, K. R. (1996). Evidence of hybridization between *Cynotilapia afra* and *Pseudotropheus zebra* (Teleostei: Cichlidae) following an intralacustrine translocation in Lake Malawi. *Copeia*, *1996*, 203–208.
- Strelman, J. T., Gmyrek, S. L., Kidd, M. R., Kidd, C., Robinson, R. L., Hert, E., ... Kocher, T. D. (2004). Hybridization and contemporary evolution in an introduced cichlid fish from Lake Malawi National Park. *Molecular Ecology*, *13*, 2471–2479. <https://doi.org/10.1111/j.1365-294X.2004.02240.x>
- Suatoni, E., Vicario, S., Rice, S., Snell, T., & Caccione, A. (2006). An analysis of species boundaries and biogeographic patterns in a cryptic species complex: The rotifer–*Brachionus plicatilis*. *Molecular Phylogenetics and Evolution*, *41*, 86–98. <https://doi.org/10.1016/j.ympev.2006.04.025>
- Tang, H., Peng, J., Wang, P., & Risch, N. J. (2005). Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology*, *28*, 289–301. <https://doi.org/10.1002/gepi.20064>
- Trewavas, E. (1935). A synopsis of the cichlid fishes of Lake Nyasa. *Annals and Magazine of Natural History*, *16*, 65–118.
- Trewavas, E. (1984). The pharyngeal apophysis on the base of the skull in cichlid fishes. *Netherlands Journal of Zoology*, *35*, 716–719. <https://doi.org/10.1163/002829685X00262>
- Turner, G. F., Seehausen, O., Knight, M. E., Allender, C. J., & Robinson, R. L. (2001). How many species of cichlid fishes are there in African lakes? *Molecular Ecology*, *10*, 793–806. <https://doi.org/10.1046/j.1365-294x.2001.01200.x>

- Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Resources*, 4, 535–538.
- Van Oppen, M. J. H., Rico, C., Deutsch, J. C., Turner, G. F., & Hewitt, G. M. (1997). Isolation and characterization of microsatellite loci in the cichlid fish *Pseudotropheus zebra*. *Molecular Ecology*, 6, 387–388.
- Wagner, C. E., Harmon, L. J., & Seehausen, O. (2012). Ecological opportunity and sexual selection together predict adaptive radiation. *Nature*, 487, 366. <https://doi.org/10.1038/nature11144>
- Wang, J. (2017). The computer program structure for assigning individuals to populations: Easy to use but easier to misuse. *Molecular Ecology Resources*, 17, 981–990.
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Pinho C, Cardoso V, Hey J. A population genetic assessment of taxonomic species: The case of Lake Malawi cichlid fishes. *Mol Ecol Resour*. 2019;19:1164–1180. <https://doi.org/10.1111/1755-0998.13027>