The impact of estimator choice: Disagreement in clustering solutions across K estimators for Bayesian analysis of population genetic structure across a wide range of empirical datasets

Kathryn Stankiewicz, Kate Vasquez Kuntz, Jean-Baptiste Ledoux, D. Aurelle, Joaquim Garrabou, Yuichi Nakajima, Mikael Dahl, Yuna Zayasu, Sabri Jaziri, Federica Costantini, et al.

Title: The impact of estimator choice: Disagreement in clustering solutions across *K* estimators for Bayesian analysis of population genetic structure across a wide range of empirical datasets

Authors: Kathryn H. Stankiewicz[1]*, Kate L. Vasquez Kuntz[1]*,

Coral Microsatellite Group**, Iliana B. Baums[1]

* Kathryn H. Stankiewicz and Kate L. Vasquez Kuntz are co-first authors

** Jean-Baptiste Ledoux[2,3], Didier Aurelle[4], Joaquim Garrabou[3,4], Yuichi Nakajima[5], Mikael Dahl[6], Yuna Zayasu[7], Sabri Jaziri[8], Federica Costantini[9]

Affiliations

1. Pennsylvania State University, Department of Biology, University Park, PA, 16802
2. CIIMAR/CIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Porto, Portugal.
3. Institut de Ciències del Mar, CSIC, Passeig Marítim de la Barceloneta, 37-49, 08003, Barcelona, Spain
4. Aix Marseille Univ., Université de Toulon, CNRS, IRD, MIO UM 110, 13288, Marseille, France; Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, 57 rue Cuvier, 75005 Paris, France
5. Marine Biophysics Unit, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna, Okinawa 904-0495, Japan
6. UiT The Arctic University of Norway · Department of Medical Biology (IMB), Hansine Hansens veg 18, 9019 Tromsø, Norway
7. Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

8. Institut National des Sciences et Technologies de la Mer, 28 rue du 2 mars 1934 - 2025 Salammbô, Tunisia

9. Università di Bologna, Dipartimento di Scienze Biologiche, Geologiche e Ambientali (BiGeA), UOS Ravenna, Ravenna, Italy

Corresponding Author: Iliana B. Baums

Email: baums@psu.edu

Keywords: Δ*K* method, STRUCTURE, population structure, microsatellites, Bayesian clustering methods, corals, gene flow, genotype, sampling effort, Puechmaille statistics

Running Title: Impact of *K* estimator choice in empirical data

Abstract

The software program STRUCTURE is one of the most cited tools for determining population structure. To infer the optimal number of clusters from STRUCTURE output, the Δ*K* method is often applied. However, a recent study relying on simulated microsatellite data suggested that this method has a downward bias in its estimation of *K* and is sensitive to uneven sampling. If this finding holds for empirical datasets, conclusions about the scale of gene flow may have to be revised for a large number of studies. To determine the impact of method choice, we applied recently described estimators of *K* to re-estimate genetic structure in 41 empirical microsatellite datasets; 15 from a broad range of taxa and 26 focused on a diverse phylogenetic group, coral. We compared alternative estimates of *K* (Puechmaille statistics) with traditional (Δ*K* and posterior probability) estimates and found widespread disagreement of estimators across datasets. Thus, one estimator alone is insufficient for determining the optimal number of clusters regardless of study organism or evenness of sampling scheme. Subsequent analysis of molecular variance (AMOVA) between clustering solutions did not necessarily clarify which solution was best. To better infer population structure, we suggest a combination of visual inspection of STRUCTURE plots and calculation of the alternative estimators at various thresholds in addition to Δ*K*. Differences between estimators could reveal patterns with important biological implications, such as the potential for more population structure than previously estimated, as was the case for many studies reanalyzed here.

Introduction

To date, one of the most cited tools to determine genetic population structure is the software program STRUCTURE (Pritchard, Stephens, & Donnelly, 2000). STRUCTURE is a free software package that uses multi-locus genotype data and a Bayesian clustering approach

relying on a Monte Carlo Markov Chain (MCMC) algorithm to infer population structure and assign individuals to populations based on their genotypes. The specification of models and the use of a random walk approach allows users to more easily incorporate prior information and account for uncertainty when clustering. In addition, STRUCTURE accepts common genetic marker types as input such as amplified fragment length polymorphisms (AFLPs), restriction fragment length polymorphisms (RFLPs), single nucleotide polymorphisms (SNPs), and microsatellites. In 2003, Falush et al. built upon STRUCTURE by developing models that allow inference of population structure with linked loci and correlated allele frequencies. Using the correlated allele frequencies method quickly became the gold standard for parsing samples into population clusters, because it assumes a level of non-independence. This model could uncover previously undetected correlation without impacting the results if the correlation did not exist (Falush, Stephens, & Pritchard, 2003; Porras-Hurtado et al., 2013).

Important to the function of STRUCTURE is the identification of clusters, which represent the main genetic divisions or 'subpopulations' within a species (Kalinowski, 2011; Puechmaille, 2016). A common problem for clustering algorithms is to determine which clustering solution is the best (Hoban, Bertorelle, & Gaggiotti, 2012; Novembre, 2016). The *K* estimation method implemented in STRUCTURE is the posterior probability of the data for a given *K* (ln Pr(X|*K*) and it has been widely used for determining the optimal number of clusters and assigning individuals to clusters. However, determining the maximal value from the posterior probability distribution is difficult, as peaks are not always clear (Evanno, Regnaut, & Goudet, 2005; Pritchard et al., 2000). To complicate matters further, in cases in which STRUCTURE model assumptions are violated, such as the presence of hierarchical population

structure, clustering solutions may be affected and subject to over-interpretation (Lawson, van Dorp, & Falush, 2018).

To solve this issue, Evanno *et al.* (2005) developed the Δ*K* statistic which is an ad hoc quantity related to the second order rate change of the log probability of data with respect to the number of clusters (Evanno et al., 2005). The Δ*K* statistic has since been a popular method for determining the number of clusters and has been cited over 12,000 times. Evanno *et al*. (2005) state that when the Δ*K* method was used on their simulated data, Δ*K* accurately estimated the true *K*, with the reservation that partial or uneven sampling could compromise the statistic from revealing the true number of clusters.

In addition, the Δ*K* method makes some biologically simplistic assumptions, which may not hold with real populations and their complex relationships. Specifically, Evanno *et al*. (2005) used a hierarchical island model of gene flow which assumed that all groups of populations were *equally* different from each other (Kalinowski, 2011). Overlying complex biological relationships, and uneven sampling appears to affect the accuracy of the Δ*K* method, as well as the program STRUCTURE itself (Puechmaille, 2016; Toyama, Crochet, & Leblois, 2020). For instance, Kalinowski (2011) states that in some cases, STRUCTURE simply put all the individuals from the largest population sample in the same cluster. To remedy the uneven sampling problem, four alternative best *K* estimators, commonly referred to as Puechmaille statistics, were created (Puechmaille, 2016).

Puechmaille (2016) tested the robustness of Δ*K* when hierarchical levels of population structure were detected in simulated and empirical datasets and found that Δ*K* did not compensate for STRUCTURE's inability to cluster subpopulations correctly, and thus Δ*K* could not reliably recover the true number of clusters. This is crucial because many empirical datasets

display hierarchical population structure and using the Δ*K* method without a proper hierarchical analysis could lead to a faulty conclusion of the number of clusters. In a meta-literature review of 1,264 studies that used Δ*K*, the authors found that very few studies performed the hierarchical analysis recommended by the authors of both Δ*K* and STRUCTURE to fully explore population subdivision (Janes et al., 2017). Janes *et al*. (2017) also found that over half of the studies that used Δ*K* concluded that the best *K* was 2. Further investigation of this issue revealed that Δ*K* was biased towards 2 due to either the presence of hierarchical populations structure, or when structure is limited (*K* = 1) (Cullingham et al., 2020). This echoes previous work on best practices for running STRUCTURE in which authors advise paying special attention to cases of *K* = 1 due to the inability of the Δ*K* method to detect such a case (Gilbert et al., 2012).

Puechmaille (2016) tested the alternative *K* estimators using almost exclusively simulated data modeled on microsatellite markers. Yet, simulated data may not reflect the complexities of empirical data, particularly in organisms with complex population structure due to life cycles or historical factors. Thus, with many available *K* estimation tools, a large-scale meta-analysis of empirical data comparing the functional outcome of estimator choice could assist researchers in methodology decisions. Previous work has evaluated the impact of different STRUCTURE parameters on determining the optimal *K* in empirical data (Funk et al., 2020), however, to date no study has evaluated the impact of choice of *K* estimator across a wide range of empirical datasets. If estimators largely disagree, greater emphasis on methodology decisions is needed and a large number of population genetic studies may need to be revised. To provide a comprehensive analysis of the choice of method to determine the optimal *K* on the outcome of population genetic studies, we re-estimated genetic structure patterns based on a total of 41 microsatellite datasets; 26 derived from corals which represent taxa that have diverse life

histories and 15 from a broad range of taxa. We tested Puechmaille's (2016) alternative *K* estimators and compared the results to the outcomes of using traditional best *K* estimation methods (Δ*K* and posterior probability). Our objectives were: (1) determine the degree of disagreement between alternative *K* estimators and traditional *K* estimators in empirical datasets (Δ*K* and posterior probability), (2) analyze potential causes of any disagreement between *K* estimation methods across datasets (sampling scheme and study organism), and (3) determine the best way to reconcile traditional *K* estimation methods with newer methods.

Methods

Dataset selection

To determine whether study organism impacts disagreement between *K* estimation methods, two dataset collections were compiled ('focused' and 'broad'). The 'focused' category was comprised of microsatellite studies on corals known to have complex population structures influenced by ocean currents. To test if findings in the 'focused' group are extendable to other systems, this was complemented by the 'broad' category of microsatellite studies on a wide range of other terrestrial, freshwater, and marine taxa. To compare the four alternative *K* estimators (Puechmaille 2016) to traditional methods (Δ*K* and ln Pr(X|*K*)), we first conducted a literature review of coral population genetics studies by searching the Web of Science using keyword combinations "coral population genetics" and "coral AND population genetics". From these searches we assembled a database of coral microsatellite datasets to represent our focused study system. To assemble a database of broad representation of taxa, we performed a search on The Dryad Digital Repository using the keywords "microsatellite population genetic structure". Studies based on single nucleotide polymorphism (SNP) data were excluded, as Puechmaille's

(2016) tested the alternative estimators using only microsatellite data. Puechmaille (2016) states that further testing is necessary to confirm whether conclusions about the alternative estimators can be extended to SNP datasets. Further, since Puechmaille (2016) created these estimators to analyze output from the software program STRUCTURE (Falush et al., 2003), datasets were selected if they had been analyzed using STRUCTURE. Additionally, we selected datasets that met two criteria: loci were not found to be under selection and population structure was analyzed using a minimum of five microsatellite loci.

Broad Datasets

The 'broad' category included 15 studies, each targeting a different species from a wide range of taxonomic groups including plants and animals of marine, freshwater, and terrestrial habitats. The sample size across these datasets ranged from 73 to 913 individuals, and thus, sampling effort differed among studies (See Supplementary Table 1). This group serves to provide a benchmark against which to compare the datasets focused on one phylogenetic group outlined below.

Focused Datasets

The 'focused' category included 26 datasets targeting 20 coral species. The sample size of datasets in the 'focused' category also varied (64 to 2,014 individuals; Supplementary Table 1). Corals were specifically chosen to represent the 'focused' category of datasets for the reasons outlined below.

STRUCTURE and the ΔK method have been widely applied to the detection of population genetic structure in marine organisms with planktonic dispersal and complex life histories (Palumbi, 2003). Corals are chief among them (Baums, Boulay, Polato, & Hellberg, 2012; Ledoux et al., 2015; Nakajima et al., 2017; Ruiz-Ramos, Saunders, Fisher, & Baums,

2015). Corals' diverse life histories include asexual and sexual reproductive modes for some species (Baird, Guest, & Willis, 2009). STRUCTURE plots often show complex patterns and determination of the best *K* results can be problematic in such cases (Lukoschek, Riginos, & van Oppen, 2016; Warner, van Oppen, & Willis, 2015). It is unclear, however, whether the complex patterns are the result of biological phenomena such as unidentified cryptic species (Boulay, Hellberg, Cortés, & Baums, 2013), violations of the corresponding model assumptions such as non-overlapping generation times (Potts, 1984), extensive inbreeding (Richards & Oppen, 2012), isolation by distance (Aurelle & Ledoux, 2013), lack of strong differentiation, or poorly performing genetic markers (i.e. null alleles) (Dubé, Planes, Zhou, Berteaux-Lecellier, & Boissin, 2017).

Focusing on one phylogenetic group containing diverse life histories allows for testing across a wide range of traits, while still preserving comparability due to shared evolutionary history. The complexity and diversity of corals makes for an excellent focused taxonomic group with which to test the performance of best *K* estimators under less simplistic study systems than those often represented by simulated data. In addition to a more general testing of a broad range of taxa, we included a separate analysis of this particularly complex study system to tease apart the nuances of how each *K* estimator may be impacted by biological intricacies found in empirical data.

Population structure analysis

To assess the performance of each estimator on empirical data, we analyzed each microsatellite dataset using *ParallelStructure* (Besnier & Glover, 2013). To ensure comparability of the results, we ran our analysis with the STRUCTURE parameters described in

197 the corresponding study. All studies considered each repeated multi-locus genotype only once
198 before running STRUCTURE. In the 'focused' category, all 26 studies ran STRUCTURE under
199 the admixture model and 24 studies used the correlated allele frequencies model. Seventeen of
200 the studies used a location prior (Hubisz, Falush, Stephens, & Pritchard, 2009) to assist with
201 clustering. In the 'broad' category, all 15 studies used the admixture model, 14 studies used the
202 correlated allele frequencies model, and three were run using a location prior. Complete details
203 for the parameter settings of each dataset can be found in Supplementary materials on Dryad
204 (DOI pending).
205 First, we calculated the Δ*K* and the posterior probability (which relies on ln Pr(X|*K*))
206 estimate using Puechmaille's (2016) R script *Kestimator* V-1.13. Then, we estimated the best *K*
207 according to Puechmaille's four alternative *K* estimators using the same R script (Puechmaille,
208 2016): the MaxMedK (the maximum of medians), the MaxMeaK (the maximum of means), the
209 MedMedK (the median of medians), and the MedMeaK (the median of means). Each of the four
210 alternative estimators were calculated at four membership coefficient thresholds (0.5, 0.6, 0.7,
211 0.8) according to the recommended default settings of the script. These threshold values are
212 based upon the finding from Guillot, Estoup, Mortier, and Cosson (2005) which defined a
213 spurious cluster as one in which no individuals have a membership coefficient >0.5. However,
214 Puechmaille (2016) extended this membership threshold by increasing the stringency in steps of
215 0.1 until reaching a threshold of 0.8. The proportion of cases in which each alternative *K*
216 estimate agreed with the Δ*K* estimate was calculated (See Supplementary Table 1). An ANOVA
217 was performed on a linear mixed model fit by residual maximum likelihood (REML) to
218 determine the effect of threshold on disagreement with Δ*K*. Following the same method, each
219 Puechmaille statistic was compared to the posterior probability estimate based on ln Pr(X|*K*)

described in (Pritchard et al., 2000). CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015) was used to visualize STRUCTURE plots.

In addition, to assess support for clustering solutions, analysis of molecular variance (AMOVA) (Excoffier, Smouse, & Quattro, 1992) was conducted using *Poppr* v2.9.1 (Kamvar, Tabima, & Grünwald, 2014) with 999 permutations for a randomly selected subset of two datasets from each category ('focused' and 'broad') in which all alternative estimators disagreed with both the ΔK and the ln Pr(X|K). For each of the four datasets, individuals were assigned by majority rule according to their membership coefficients into the number of clusters identified by the different K estimation methods (the alternative estimators, the ΔK, and the ln Pr(X|K)). AMOVA was run on each clustering solution for each dataset, with only two exceptions. For the dataset baums_et_al_2010_1 (Baums, Johnson, Devlin-Durante, & Miller, 2010), all alternative estimators found K =1. AMOVA requires >1 group in order to compare variation between groups, thus, it was not run on a clustering of individuals into one singular population. For the dataset perez_et_al_2014 (Perez et al., 2014), majority rule assigned individuals to only 11 clusters, with no single individual having a membership coefficient high enough for assignment to a twelfth cluster. Thus, K = 12 as identified by the posterior probability method was excluded from AMOVA.

Assessment of sampling strategy

The program STRUCTURE may not reliably estimate the true number of clusters when sampling is uneven (Puechmaille, 2016). Consequently, we calculated sampling evenness scores for each dataset to test whether the alternative estimators perform differently than traditional methods in situations of uneven sampling. We calculated the evenness score, E, for each data set

using Shannon's Diversity Index (Equation 1). The number of multi-locus genotypes (MLGs) per sampling site was used to calculate the evenness of each dataset with respect to sampling scheme.

The result of the equation below yields a score from 0 to 1 where higher scores indicate a more even sampling scheme (See Supplementary Table 4 for calculations).

Equation 1

$$E = -1 * \sum \frac{\left(\frac{N_{isite}}{N_{itotal}} * \ln \frac{N_{isite}}{N_{itotal}} \right)}{\ln N_{itotal}}$$

Where E = eveneness, N_{isite} = number of MLGs at site, and N_{itotal} = total number of MLGs.

Next, we tested if there was a relationship between the proportion of the new *K* estimators that were congruent with each traditional method (ΔK and ln Pr(X|*K*)) and the evenness of the sampling strategy. To do so, we performed a linear regression with sampling evenness as a predictor for proportion agreement.

Results

Comparison of K estimator performances: Focused category

For each dataset in the 'focused' category, 16 estimates of *K* were calculated from the R script *Kestimator* V-1.1 (the four alternative *K* estimators, each at four membership thresholds). The script also calculated the traditional ΔK and the posterior probability estimates. The proportion of these 16 alternative *K* estimators that were congruent with the ΔK method varied across studies. The relative frequency of coral studies in which less than 20% of the 16 new *K*

estimators agreed with the Δ*K* estimate was 50% (Fig. 1A). Additionally, most (62%) of the studies had less than 50% agreement with Δ*K*.

The alternative *K* estimators tended to return higher values of *K* than the Δ*K* method, with only two exceptions. On average, the MaxMeaK and the MedMeaK estimates, each at a membership threshold of 0.8, returned lower values of *K* than the Δ*K* method (Fig. 2A). In the empirical 'focused' category datasets we analyzed here, lower membership coefficient thresholds led to a higher magnitude of disagreement from Δ*K* across all four new estimators (Fig. 3). The effect of threshold was significant on the disagreement from Δ*K* according to ANOVA on a linear mixed model fit by REML (f-value = 4.348; $p = 0.005$). The effect of estimator, however, was not significant (f-value = 0.0998; $p = 0.96$). The combined effect of threshold and estimator was also not significant (f-value = 0.2244; $p = 0.991$). Notably, estimators based on the median (the MaxMedK and the MedMedK) tended to disagree with Δ*K* by more than those based on the mean (the MaxMeaK and the MedMeaK, Fig. 2A). Unsurprisingly, the estimators that use the maximum number of clusters in their calculations of the best *K* (the MaxMeaK and the MaxMedK), as opposed to the median, tended to disagree with Δ*K* by a higher magnitude (Fig. 2A).

In fourteen of the 26 coral datasets, less than 20% of the alternative estimates of *K* agreed with the posterior probability estimate (Fig. 1B). However, on average, the alternative Puechmaille statistics returned lower estimates of *K* than the posterior probability method (Fig. 2B) in the 'focused' category datasets. This was not the case with Δ*K*.

Comparison of K estimator performances: Broad category

Nearly all of the patterns present in the 'focused' dataset analysis were mirrored in the 'broad' dataset category analysis. Nine of the datasets in the 'broad' category had lower than

20% proportion agreement between the alternative Puechmaille statistics and the Δ*K* estimate (Fig. 1C). Additionally, on average, all alternative *K* estimators were higher than the Δ*K* estimate regardless of threshold (Fig 2C).

Proportion agreement between the posterior probability estimate and the alternative statistics was similarly low, with 11 out of the 15 'broad' category datasets showing less than 20% proportion agreement (Fig. 1D). In comparison to the posterior probability estimate, the Puechmaille statistics resulted in lower estimates of *K* on average (Fig. 2D)—again, consistent with the trend present in the 'focused' category of datasets (Fig. 2B). As in the 'focused' datasets, lower membership coefficient thresholds led to a higher magnitude of disagreement from Δ*K* (Fig. 3).

Influence of sampling effort on K estimates: Focused category

In the coral datasets, we found no significant relationship between sampling evenness and the proportion of alternative *K* estimators that agree with the Δ*K* estimator (Fig. 4A) or the posterior probability (Fig. 4B). We compared sampling evenness and proportion agreement with Δ*K* using a linear and a polynomial model. However, neither resulted in a significant best fit (linear: $R^2 = 0.025$, $p = 0.444$; polynomial: $R^2 = 0.137$, $p = 0.070$). To account for differences in sample size, we weighted each point in the plot accordingly, but the relationship remained insignificant (See Supplementary Figure 1). Similarly, under a linear model, there was no significant relationship between proportion agreement of the alternative estimators and the posterior probability (Fig. 4B; $R^2 = 0.116$, $p = 0.088$).

Influence of sampling effort on K estimates: Broad category

Echoing the trends found in the 'focused' category, the 15 datasets included in the 'broad' category also returned no significant relationship between sampling evenness and proportion agreement for either the ΔK estimator (Fig. 4C; $R^2 = 0.231$, $p = 0.070$) or the posterior probability (Fig. 4D; $R^2 = 0.10$, $p = 0.258$) under a linear model. Each dataset was weighted by sample size for all tests.

Additionally, we visualized the STRUCTURE plots for the Perez et al. (2014) *Testudo hermanni* dataset as an example with a relatively low evenness score (E=0.84, See Supplementary Table 1). The reanalysis yielded complete agreement between the Puechmaille statistics that contrasted with published findings using traditional methods (ΔK and ln Pr(X|K)). The authors reported a $K = 5$ (Fig. 5A), however, alternative estimators reported a $K = 7$ (Fig. 5B).

Precision of Puechmaille estimates

Across all 41 datasets, the 16 Puechmaille estimates most commonly offered two (13/41 datasets) or one (11/41 datasets) K estimate(s) (See Supplementary Table 1). In 75.6% of cases, the range of solutions offered by the Puechmaille estimators was ≤ 3. The largest range of K estimates provided by the Puechmaille statistics was 6 (1/41 datasets) and was found in only one dataset, Kurita_et_al_2014.

Analysis of molecular variance

From the 'broad' category, the datasets kim_et_al_2017 (Kim et al., 2017) and perez_et_al_2013 (Perez et al., 2014) were randomly selected. From the 'focused' category, datasets baums_et_al_2010_1 (Baums et al., 2010) and rippe_et_al_2017 (Rippe et al., 2017) were

randomly selected. *K* estimation for each dataset included a range of *K* values each supported by different *K* estimation methods (Table 1). Across all datasets and all clustering solutions, the majority of the variation was explained by differences within samples (Table 2). Additionally, the proportion of variation across all strata levels (between clusters, between samples within clusters, and within samples) were significant ($p < 0.01$ in all cases; Supplementary Table 5) across all datasets and clustering solutions. The magnitude of the proportion of variation explained by differences between clusters varied by dataset (Table 2). However, a notable trend found in all clustering solutions across all datasets, was the slight increase in the magnitude of the proportion of variation attributed to differences between clusters with an increase in *K* (Table 2).

Discussion

Accurate characterizations of population genetic structure are at the core of eco-evolutionary studies. Knowledge of population genetic structure enables inferences about the ecological and evolutionary dynamics of a species such as the scale of dispersal, the breeding system, and demographic history (Bohonak, 1999; Dillane et al., 2008; Les, 1988). The development of cost-effective molecular markers for non-model organisms combined with the adoption of Bayesian methods to detect even weak signals of population genetic structure has propelled the field forward (Baums, Miller, & Hellberg, 2005; Garris, Tai, Coburn, Kresovich, & McCouch, 2005; Latch, Dharmarajan, Glaubitz, & Rhodes, 2006). Yet, especially in non-model organisms, the determination of the best solution among the tested number of clusters in a Bayesian model can be difficult. Here, we report that the more recently developed best *K* estimators (Puechmaille, 2016) suggest more population genetic structure across the majority of empirical 'focused' coral microsatellite datasets tested compared to the most popular *K* estimation method, Δ*K* estimator.

In contrast, the alternative estimators suggested less genetic structure than the posterior probability (ln Pr(X|K)). These patterns hold when extended to a broad group of taxa, and results agree with a previous study using simulated datasets (Puechmaille, 2016). Even sampling effort among populations is expected to lead to more accurate determination of best *K* and yet we found no significant relationship between sampling evenness and proportion agreement in the empirical data.

Comparison to Δ*K*

Because we used the same parameters for STRUCTURE modeling that were used in the original studies, if there was hierarchy among clusters present, it remained intact. In other words, genotypes in the original and in our reanalysis were always split between the first two clusters in the same way, and then were assigned to the next cluster in the same way, and so forth for each higher number of *K*. This design allowed us to compare the solution suggested by the alternative *K* estimators to the results of the original studies. Alternative estimators agreed with the Δ*K* method across thresholds only when the best *K* was one or two ('focused' category: five out of 20 species, 'broad' category: one out of 15 species, Supplementary Table 1). In most other cases, alternative *K* estimators suggested that species may have more pronounced population structure than previously thought. In the 'focused' dataset category, 11 out of 20 species of varying habitat type and study design displayed this phenomenon. In the 'broad' category, in ten out of 15 studies alternative *K* estimators returned higher *K* solutions. Thus, across a wide range of taxa, the alternative *K* estimators indicated more population genetic structure than the Δ*K* method.

One notable case where we found evidence for more pronounced population structure was in the 'focused' category dataset corresponding to the coral *Porites lobata* (Baums et al., 2012).

375 Initially, the Δ*K* method returned a best *K* = 5. *Porites lobata* from the Eastern Tropical Pacific
376 were distinct from colonies from the central Pacific and Hawaii (Baums et al., 2012). Within
377 Hawaii, there existed three co-occurring clusters that remained distinct from the remainder of
378 central Pacific clusters. Another clustering algorithm, GENELAND (Guillot et al., 2005),
379 returned a best *K* of seven with the possibility of an additional cluster being split in two, yielding
380 nine clusters in total (Baums et al., 2012). Upon reanalysis with the alternative *K* estimators, the
381 clear majority (14/16 estimators) pointed to a best *K* between seven and nine. One estimator
382 agreed with Δ*K* and another reported a lower estimate of *K* = 4. The study's main conclusion that
383 there is a lack of geneflow across the eastern pacific barrier was upheld (see also (Wood et al.,
384 2016)), but our reanalysis suggested additional population structure in the central Pacific with
385 putative conservation implications at the regional scale.
386 Conversely, in some select cases the Δ*K* estimate was higher than the alternative estimates.
387 One such case in the 'focused' category was the dataset corresponding to the coral *Acropora*
388 *digitifera* (Nakajima, Nishikawa, Iguchi, & Sakai, 2012). Though the Δ*K* estimate returned a best
389 *K* of 2, the authors found evidence to suggest there was only one population. Δ*K* is known to be
390 unable to report when the best *K* is 1 and instead most often reports *K* = 2 (Cullingham et al.,
391 2020). However, the alternative Puechmaille statistics identified the best *K* = 1, except for those
392 at the lowest (0.5) threshold. This same phenomenon can be found in the 'broad' category of
393 datasets in the New Zealand Sea Lion, *Phocarctos hookeri* (Osborne et al., 2016). Again, here
394 the Δ*K* estimate suggested two populations. However, Osborne et al. (2016) found weak
395 population differentiation with F_{ST} values low enough to suggest no population structure and
396 concluded that the result was more consistent with one population of individuals living in
397 familial clusters. All of the alternative Puechmaille statistics again identified a best *K* of 1,

except those at the lowest threshold. This highlights the benefit to calculating these alternative statistics, while considering a range of thresholds. In adding to the recommendations by Cullingham et al., 2020 for determining when $K = 2$, we propose using the alternative estimators to help determine the level of support.

In four cases within the 'focused' category, the alternative estimators showed little agreement amongst themselves and with ΔK in their best K solutions (Supplementary Table 1). We suggest that difficulties in determining the best K can arise from hidden genetic diversity in the investigated species (Hajibabaei, Singer, Hebert, & Hickey, 2007; Hebert, Penton, Burns, Janzen, & Hallwachs, 2004). *Seriatopora hystrix* had a wide spread of best K estimates with ΔK suggesting $K = 3$. The authors conducted a hierarchical analysis investigating all three clusters further. Clusters were grouped based on regional scales of clustering and five major genetic clusters were distinguished. However, reanalysis with new estimators suggested a minimum $K = 4$ and a maximum $K = 9$ (Supplementary Table 1). The authors mention that cryptic species may have masked the true population connectivity signals, further investigation of which may be warranted based on our reanalysis of population structure. Corals hybridize frequently and the speciation process in this group may follow a pattern of reticulate evolution and thus cryptic lineages at all taxonomic levels are expected to be common (Kenyon, 1997; Veron, 1995; Vollmer & Palumbi, 2002; Willis, van Oppen, Miller, Vollmer, & Ayre, 2006).

In the 'broad' category, one case in which all alternative estimators agreed with one another, but disagreed with ΔK occurred in a dataset for *Testudo hermanni*, an endangered tortoise species in Mediterranean (Perez et al., 2014). All alternative Puechmaille statistics indicated $K = 7$, while the posterior probability indicated $K = 12$. However, the ΔK estimate found the best $K = 2$. Perez et al. (2014) used STRUCTURE and GENELAND (Guillot et al., 2005) to draw their

conclusions about population structure in this study. Using STRUCTURE, the authors found evidence for $K = 2$ and $K = 5$ by employing several traditional K estimation methods (ΔK and ln Pr(X|K)). However, GENELAND analysis indicated a best $K = 6$. Curiously, under $K = 5$, samples from geographically distant localities (Spain, Sicily and Corsica) clustered together according to STRUCTURE (Fig. 5A). The authors assert that massive translocations between Spain, Sicily, and Corsica are unlikely for this sedentary species of tortoise and instead suggest that prehistoric events could explain the admixture (Perez et al., 2014). However, the alternative estimators suggest $K = 7$ (Fig. 5B). At $K = 7$, Spain, Sicily and Corsica contain distinct population clusters, as does the region of Macedonia (MAC). Previously, MAC clustered together with the Bosco Mesola population (BM) in the $K = 5$ solution (Fig. 5A) and Perez et al. (2014) report that it clustered with Greece (GR) in the GENELAND analysis. Though the true K can't be known, inclusion of the alternative estimators may have provided helpful insight in parsing the different solutions between GENELAND and STRUCTURE in this study.

Comparison to Posterior Probability

In both the 'focused' and the 'broad' categories, the posterior probability method yielded higher estimates of K than the alternative estimators and the ΔK method. This result could be due to the fact that ln Pr(X|K), the basis for calculating the posterior probability according to Bayes rule, is known to be sensitive to the STRUCTURE model which allows for allele frequencies to be correlated between subpopulations (Falush et al., 2003). The STRUCTURE manual recommends that default settings should include allowing for correlated allele frequencies, and indeed most (38/41) datasets re-analyzed here, regardless of category, followed this recommendation. However, Falush et al. (2003) find that this could result in a higher risk of

overestimating *K* compared to the independent allele frequencies model. Since the alternative estimates of *K* are lower, on average, than the estimates calculated by the posterior probability method, it is possible that the Puechmaille statistics are less sensitive to such deviations in model assumptions. This corroborates Puechmaille's (2016) simulation study, which exclusively used the correlated allele frequencies model, showing that the posterior probability method overestimated the true *K*.

Analysis of molecular variance

As the true *K* cannot be known in empirical data, we applied analysis of molecular variance (AMOVA) to a subset of datasets to evaluate its use as a method for determining which clustering solution was most supported. Datasets in which there was full disagreement between the Puechmaille statistics and both traditional *K* estimation methods were selected, as these cases are the most difficult to interpret and additional analysis is warranted to determine the best clustering solution. Previous work has pointed out that it may be inappropriate to test the significance of AMOVA results on STRUCTURE clustering solutions (despite this being a common practice) (Meirmans, 2015). However, Meirmans (2015) indicate that reporting F_{ST} values is perfectly acceptable. With the expectation that the magnitude of significant variance explained by differences between clusters should be maximized in the best solution, we compared AMOVA results across clustering solutions for each dataset. Perhaps not unexpectedly, we noted that across datasets, the proportion of variance explained by differences between clusters increased slightly with increasing number of clusters, *K* (Table 2). This finding is similar to the results of a recent simulation-based study which found that the magnitude of ΔK was correlated with F_{ST}, with higher values of ΔK having more supported population structure

(Cullingham et al., 2020). It may well be that the magnitude of variance explained is simply always maximized at the highest value of *K*. This could be the case if increasing the number of model parameters by adding more clusters increases the distance between clusters. Thus, a simulation study is necessary to assess whether AMOVA can assist with identifying the best clustering solution.

Evenness Assessment

Since STRUCTURE's inception, Evanno (2005) and others have warned users that uneven sampling across strata may influence the accuracy of determining the best *K* (Evanno et al., 2005; Kalinowski, 2011; Puechmaille, 2016). In fact, previous work has recommended modifying alpha values when running STRUCTURE to address this (Wang, 2017). Because STRUCTURE can detect weak population signals (Latch et al., 2006), Puechmaille (2016) theorized that uneven sampling was the main contributor to Δ*K*'s inability to identify the correct *K*. Further, previous work has found that uneven sampling design in a multi-species empirical dataset did impact STRUCTURE results (Meirmans, 2019). Thus, we initially hypothesized that the discrepancy between Puechmaille's estimators and Δ*K* was due to uneven sampling across clusters. Δ*K* is affected by uneven sampling because STRUCTURE tends to place individuals from an oversampled subpopulation into their own cluster while putting a sparsely sampled subpopulation into its own cluster, regardless of the true evolutionary history. Puechmaille's new estimators avoid this by implementing a range of cluster membership coefficients (from least stringent, 0.5 to most, 0.8) and accounting for maximum population subdivision via the estimators MaxMeaK and MaxMedK, thus correcting for STRUCTURE's downward biased estimates of *K*.

To test how sampling evenness affects best *K* estimates, we calculated evenness scores for each study (Fig. 4) and correlated these scores with the proportion of estimators that agreed among all best *K* estimators. Surprisingly, we found no significant relationship between sampling evenness and proportion agreement among best *K* estimators for both the 'focused' and 'broad' category datasets. In fact, a subset of studies at all levels of sampling evenness had high proportion agreement scores. Unexpectedly, the study that was the least evenly sampled, had one of the highest proportion agreement scores.

The unexpected poor power of sampling evenness to predict the ease of which the best *K* could be determined may stem from overestimating evenness. In human studies, populations are typically grouped based on linguistic, cultural, or physical characters and then sampled as evenly as possible (Pritchard et al., 2000). However, *a priori* stratification of many non-model organisms into sampling groups is often not possible due to a lack of obvious phenotypes and poor understanding of metapopulation structure. In fact, the latter is often a motivation to conduct a STRUCTURE analysis. Yet, to have confidence in STRUCTURE results, even sampling is required, thus the paradox arises. Per design, the sites in each study might have been sampled evenly, which yielded high evenness scores ($E > 80$). However, sampling sites do not equate to populations and thus, some populations were unintentionally oversampled while others were under-sampled. Therefore, evenness scores as calculated here for a given study might be high and yet do not reflect even sampling of populations. Additionally, even sampling of populations across a species' range is logistically challenging. Oversampling may occur at the center of a species' range because there are more individuals per unit area making sampling easier. Likewise, under-sampling may occur at the margins of the range because, by definition, organisms occur at lower density requiring higher sampling effort.

Regardless of the reason why there is a lack of correlation between evenness and ease of determining the best K in this meta-analysis, it is very difficult to achieve even sampling across populations in practice even if it is desirable. It thus behooves us to use population genetics tools that can deal with reality by correcting for sampling unevenness *ex post facto*, as the alternative estimators do. We recommend using ΔK and the posterior probability to get a basic cluster estimation, followed by an analysis that uses all alternative K estimators at a range of thresholds. Since each estimator has different sensitivities and choice of threshold has a significant effect on result, comparing each to ΔK and the posterior probability during analysis offers the most robust procedure for estimating K in the case of potentially ambiguous sampling evenness. We additionally recommend inspecting STRUCTURE plots to tease out the best estimation of K in case new estimators give an ambiguous result (rare). Combining all four strategies—the ΔK, the posterior probability, the alternative K estimators, and examination of STRUCTURE plots—ensures the most robust estimation of K and will allow researchers to detect biological subtleties that may not be recognizable using the ΔK estimate alone.

Final Thoughts

Our comprehensive re-analysis of population genetic structure across both a focused group of taxa (corals) and a broad group of taxa from across the Tree of Life indicates that population genetic structure may be more pronounced than previously described. The alternative K estimators typically agreed with each other across thresholds and ΔK when there was clear population structure across space. However, there were cases showing disagreement amongst estimators when population structure was more complicated, for example when sympatric samples were assigned with high probability to different clusters. Since the new estimators more

accurately predicted *K* than Δ*K*'s and the posterior probability's predictions in studies where the best *K* was known (i.e., simulated data; Puechmaille, 2016) and there were substantially more empirical studies whose alternative *K* estimates differed drastically from traditional *K* estimation predictions than agreed with it (See Supplementary Table 2), we recommend the incorporation of the alternative estimators to determine the best *K*.

Our finding of little agreement between *K* estimation methods across a wide range of datasets indicate that choice of estimator has a substantial impact on the results in empirical data. Further, we found that this is not restricted to a particularly complex taxonomic group (i.e., corals), nor to studies with obviously uneven sampling schemes. Thus, our recommendations for careful consideration in methodology applies to a wide range of studies. We find here that due to large scale disagreement between estimator solutions across datasets, a multi-estimator approach is always required, regardless of study species or sampling approach. Additionally, broader re-analysis of existing microsatellite datasets may be warranted and has the added benefit of preserving these datasets for future use as many of these datasets were published before the advent of online repositories.

Acknowledgements

We thank all authors who made their microsatellite data publicly available.

This work was made possible by NSF grant OCE-1537959 to IBB, NIH grant T32: Computation, Bioinformatics, and Statistics (CBIOS) Training Program to KHS, a Bunton-Waller fellowship to KLVK, the strategic Funding UIDB/04423/2020 and UIDP/04423/2020 to JBL, and the Pennsylvania State University Biology Department. The project leading to this publication has received funding from European FEDER Fund under project 1166-39417 to DA. We

acknowledge the funding of the Spanish government through the ‘Severo Ochoa Centre of Excellence’ accreditation (CEX2019-000928-S). Special thanks to that Pennsylvania State University’s Statistical Consulting Center.

Data Accessibility Statement

All Supplementary figures and their corresponding raw data can be accessed on Dryad (DOI pending).

Author Contributions

KLVK Designed research, assembled microsatellite database, analyzed data, and wrote the paper. KHS Designed research, assembled microsatellite database, analyzed data, and wrote the paper. Coral Microsatellite Group contributed data and edited the paper. Key contributions from the Coral Microsatellite Group were made by DA, JBL, FC, and JG. IBB conceived the project and wrote the paper.

References

Aurelle, D., & Ledoux, J. B. (2013). Interplay between isolation by distance and genetic clusters in the red coral Corallium rubrum: insights from simulated and empirical data. *Conservation Genetics, 14*(3), 705-716. doi:10.1007/s10592-013-0464-0

Baird, A. H., Guest, J. R., & Willis, B. L. (2009). Systematic and Biogeographical Patterns in the Reproductive Biology of Scleractinian Corals. *Annual Review of Ecology, Evolution, and Systematics, 40*(1), 551-571. doi:10.1146/annurev.ecolsys.110308.120220

Baums, I., Johnson, M., Devlin-Durante, M., & Miller, M. (2010). Host population genetic structure and zooxanthellae diversity of two reef-building coral species along the Florida Reef Tract and wider Caribbean. *Coral Reefs, 29*, 835-842.

Baums, I. B., Boulay, J. N., Polato, N. R., & Hellberg, M. E. (2012). No gene flow across the Eastern Pacific Barrier in the reef-building coral Porites lobata. *Mol Ecol, 21*(22), 5418-5433. doi:10.1111/j.1365-294X.2012.05733.x

Baums, I. B., Miller, M. W., & Hellberg, M. E. (2005). Regionally isolated populations of an imperiled Caribbean coral, *Acropora palmata*. *Mol Ecol, 14*(5), 1377-1390. doi:10.1111/j.1365-294X.2005.02489.x

Besnier, F., & Glover, K. A. (2013). ParallelStructure: a R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. *PLoS One, 8*(7), e70651. doi:10.1371/journal.pone.0070651

Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *Q Rev Biol, 74*(1), 21-45.

Boulay, J. N., Hellberg, M. E., Cortés, J., & Baums, I. B. (2013). Unrecognized coral species diversity masks differences in functional ecology. *Proceedings. Biological sciences, 281*(1776), 20131580-20131580. doi:10.1098/rspb.2013.1580

Cullingham, C. I., Miller, J. M., Peery, R. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., & Janes, J. K. (2020). Confidently identifying the correct K value using the ΔK method: When does K = 2? *Molecular Ecology, 29*(5), 862-869. doi:https://doi.org/10.1111/mec.15374

Dillane, E., McGinnity, P., Coughlan, J. P., Cross, M. C., De Eyto, E., Kenchington, E., . . . Cross, T. F. (2008). Demographics and landscape features determine intrariver population structure in Atlantic salmon (Salmo salar L.): the case of the River Moy in Ireland. *Mol Ecol, 17*(22), 4786-4800. doi:10.1111/j.1365-294X.2008.03939.x

Dubé, C. E., Planes, S., Zhou, Y., Berteaux-Lecellier, V., & Boissin, E. (2017). Genetic diversity and differentiation in reef-building Millepora species, as revealed by cross-species amplification of fifteen novel microsatellite loci. *PeerJ, 5*, e2936-e2936. doi:10.7717/peerj.2936

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol, 14*(8), 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics, 131*(2), 479-491.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics, 164*(4), 1567-1587.

Funk, S. M., Guedaoura, S., Juras, R., Raziq, A., Landolsi, F., Luís, C., . . . Cothran, E. G. (2020). Major inconsistencies of inferred population genetic structure estimated in a large set of domestic horse breeds using microsatellites. *Ecology and Evolution, 10*(10), 4261-4279. doi:https://doi.org/10.1002/ece3.6195

Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S., & McCouch, S. (2005). Genetic structure and diversity in Oryza sativa L. *Genetics, 169*(3), 1631-1638. doi:10.1534/genetics.104.035642

Gilbert, K. J., Andrew, R. L., Bock, D. G., Franklin, M. T., Kane, N. C., Moore, J.-S., . . . Vines, T. H. (2012). Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure. *Molecular Ecology, 21*(20), 4925-4930. doi:https://doi.org/10.1111/j.1365-294X.2012.05754.x

Guillot, G., Estoup, A., Mortier, F., & Cosson, J. F. (2005). A spatial statistical model for landscape genetics. *Genetics, 170*(3), 1261-1280. doi:10.1534/genetics.104.033803

Hajibabaei, M., Singer, G. A., Hebert, P. D., & Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet, 23*(4), 167-172. doi:10.1016/j.tig.2007.02.001

Hebert, P. D., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proc Natl Acad Sci U S A, 101*(41), 14812-14817. doi:10.1073/pnas.0406166101

Hoban, S., Bertorelle, G., & Gaggiotti, O. E. (2012). Computer simulations: tools for population and evolutionary genetics. *Nat Rev Genet, 13*(2), 110-122. doi:10.1038/nrg3130

Hubisz, M., Falush, D., Stephens, M., & Pritchard, J. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources, 9*(5), 1322-1332.

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K = 2 conundrum. *Mol Ecol, 26*(14), 3594-3602. doi:10.1111/mec.14187

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 (Edinb), 106*(4), 625-632. doi:10.1038/hdy.2010.95

Kamvar, Z., Tabima, J., & Grünwald, N. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ 2:e281*. doi:https://doi.org/10.7717/peerj.281

Kenyon, J. C. (1997). Models of Reticulate Evolution in the Coral Genus Acropora Based on Chromosome Numbers: Parallels with Plants. *Evolution, 51*(3), 756-767. doi:10.1111/j.1558-5646.1997.tb03659.x

Kim, J. H., Kang, J. H., Jang, J. E., Choi, S. K., Kim, M. J., Park, S. R., & Lee, H. J. (2017). Population genetic structure of eelgrass (Zostera marina) on the Korean coast: Current status and conservation implications for future management. *PLoS One, 12*(3), e0174105. doi:10.1371/journal.pone.0174105

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour, 15*(5), 1179-1191. doi:10.1111/1755-0998.12387

Latch, E. K., Dharmarajan, G., Glaubitz, J. C., & Rhodes, O. E. (2006). Relative performance of Bayesian clustering software for inferringpopulation substructure and individual assignment at low levels of population differentiation. *Conservation Genetics, 7*(2), 295-302. doi:10.1007/s10592-005-9098-1

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat Commun, 9*(1), 3258. doi:10.1038/s41467-018-05257-7

Ledoux, J. B., Aurelle, D., Bensoussan, N., Marschal, C., Feral, J. P., & Garrabou, J. (2015). Potential for adaptive evolution at species range margins: contrasting interactions between red coral populations and their environment in a changing ocean. *Ecol Evol, 5*(6), 1178-1192. doi:10.1002/ece3.1324

Les, D. H. (1988). Breeding Systems, Population Structure, and Evolution in Hydrophilous Angiosperms. *Annals of the Missouri Botanical Garden, 75*(3), 819-835. doi:10.2307/2399370

Lukoschek, V., Riginos, C., & van Oppen, M. J. (2016). Congruent patterns of connectivity can inform management for broadcast spawning corals on the Great Barrier Reef. *Mol Ecol, 25*(13), 3065-3080. doi:10.1111/mec.13649

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology, 24*(13), 3223-3231. doi:https://doi.org/10.1111/mec.13243

Meirmans, P. G. (2019). Subsampling reveals that unbalanced sampling affects Structure results in a multi-species dataset. *Heredity, 122*(3), 276-287. doi:10.1038/s41437-018-0124-8

Nakajima, Y., Nishikawa, A., Iguchi, A., & Sakai, K. (2012). Regional genetic differentiation among northern high-latitude island populations of a broadcast-spawning coral. *Coral Reefs, 31*(4), 1125-1133. doi:10.1007/s00338-012-0932-x

Nakajima, Y., Wepfer, P. H., Suzuki, S., Zayasu, Y., Shinzato, C., Satoh, N., & Mitarai, S. (2017). Microsatellite markers for multiple Pocillopora genetic lineages offer new insights about coral populations. *Sci Rep, 7*(1), 6729. doi:10.1038/s41598-017-06776-x

Novembre, J. (2016). Pritchard, Stephens, and Donnelly on Population Structure. *Genetics, 204*(2), 391-393. doi:10.1534/genetics.116.195164

Osborne, A. J., Negro, S. S., Chilvers, B. L., Robertson, B. C., Kennedy, M. A., & Gemmell, N. J. (2016). Genetic Evidence of a Population Bottleneck and Inbreeding in the Endangered New Zealand Sea Lion, Phocarctos hookeri. *Journal of Heredity, 107*(5), 392-402. doi:10.1093/jhered/esw015

Palumbi, S. R. (2003). POPULATION GENETICS, DEMOGRAPHIC CONNECTIVITY, AND THE DESIGN OF MARINE RESERVES. *Ecological Applications, 13*(sp1), 146-158. doi:10.1890/1051-0761(2003)013[0146:Pgdcat]2.0.Co;2

Perez, M., Livoreil, B., Mantovani, S., Boisselier, M. C., Crestanello, B., Abdelkrim, J., . . . Bertorelle, G. (2014). Genetic variation and population structure in the endangered Hermann's tortoise: the roles of geography and human-mediated processes. *J Hered, 105*(1), 70-81. doi:10.1093/jhered/est071

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, A., & Lareu, M. V. (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. *Front Genet, 4*, 98. doi:10.3389/fgene.2013.00098

Potts, D. C. (1984). Generation Times and the Quaternary Evolution of Reef-Building Corals. *Paleobiology, 10*(1), 48-58.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Puechmaille, 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. *Mol Ecol Resour, 16*(3), 608-627. doi:10.1111/1755-0998.12512

Richards, Z. T., & Oppen, M. J. H. (2012). Rarity and genetic diversity in Indo-Pacific Acropora corals. *Ecology and Evolution, 2*(8), 1867-1888. doi:10.1002/ece3.304

Rippe, J. P., Matz, M. V., Green, E. A., Medina, M., Khawaja, N. Z., Pongwarin, T., . . . Davies, S. W. (2017). Population structure and connectivity of the mountainous star coral, Orbicella faveolata, throughout the wider Caribbean region. *Ecology and Evolution, 7*(22), 9234-9246. doi:https://doi.org/10.1002/ece3.3448

Ruiz-Ramos, D. V., Saunders, M., Fisher, C. R., & Baums, I. B. (2015). Home Bodies and Wanderers: Sympatric Lineages of the Deep-Sea Black Coral Leiopathes glaberrima. *PLoS One, 10*(10), e0138989. doi:10.1371/journal.pone.0138989

Toyama, K. S., Crochet, P.-A., & Leblois, R. (2020). Sampling schemes and drift can bias admixture proportions inferred by structure. *Molecular Ecology Resources, 20*(6), 1769-1785. doi:https://doi.org/10.1111/1755-0998.13234

Veron, J. E. N. (1995). *Corals in space and time : the biogeography and evolution of the Scleractinia / J.E.N. Veron*. Sydney: University of New South Wales Press.

Vollmer, S. V., & Palumbi, S. R. (2002). Hybridization and the evolution of reef coral diversity. *Science, 296*(5575), 2023-2025. doi:10.1126/science.1069524

Wang, J. (2017). The computer program structure for assigning individuals to populations: easy to use but easier to misuse. *Molecular Ecology Resources, 17*(5), 981-990. doi:10.1111/1755-0998.12650

Warner, P. A., van Oppen, M. J., & Willis, B. L. (2015). Unexpected cryptic species diversity in the widespread coral Seriatopora hystrix masks spatial-genetic patterns of connectivity. *Mol Ecol, 24*(12), 2993-3008. doi:10.1111/mec.13225

Willis, B. L., van Oppen, M. J. H., Miller, D. J., Vollmer, S. V., & Ayre, D. J. (2006). The Role of Hybridization in the Evolution of Reef Corals. *Annual Review of Ecology, Evolution, and Systematics, 37*(1), 489-517. doi:10.1146/annurev.ecolsys.37.091305.110136

Wood, S., Baums, I. B., Paris, C. B., Ridgwell, A., Kessler, W. S., & Hendy, E. J. (2016). El Nino and coral larval dispersal across the eastern Pacific marine barrier. *Nat Commun, 7*, 12571. doi:10.1038/ncomms12571

Main Text Figures 1-5, and Tables 1 & 2

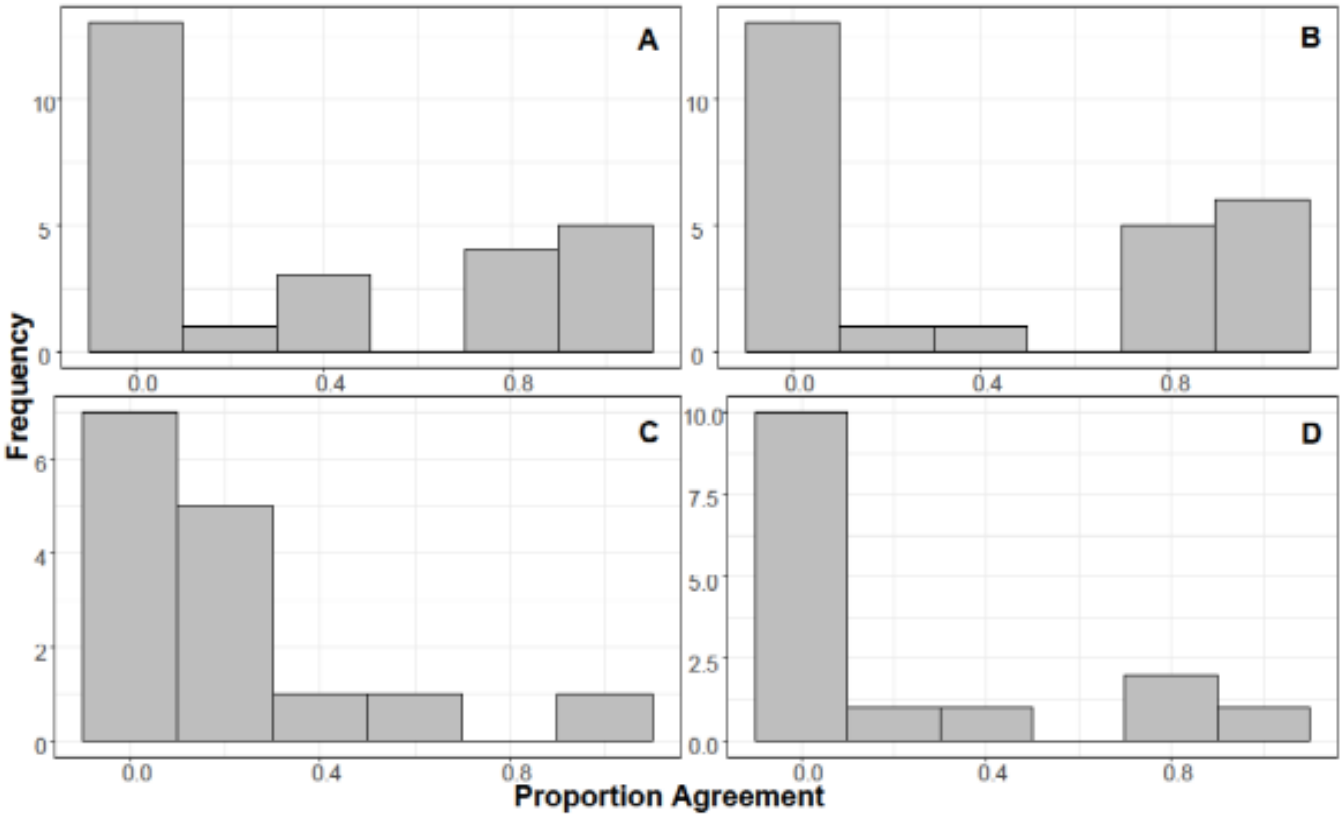


Fig. 1 Histogram of Proportion Agreement. Focused category microsatellite datasets (*n*=26) were binned according to the proportion of the 16 alternative estimators (Puechmaille 2016) which agree with the **(A)** Δ*K* estimate and **(B)** the posterior probability estimate. Broad category datasets (*n*=15) were similarly binned according to the proportion agreement between the 16 alternative estimators and **(C)** Δ*K*, and **(D)** the posterior probability estimate.

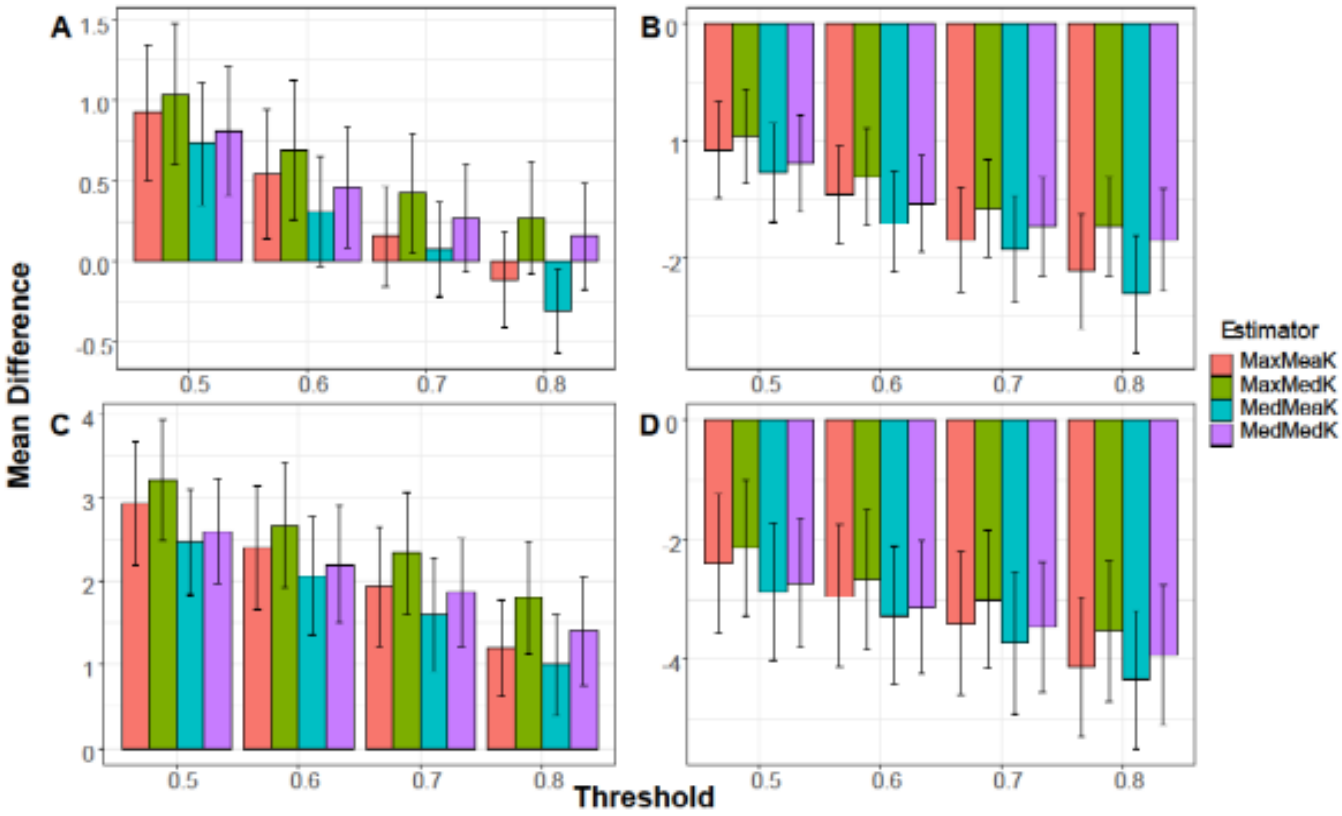


Fig. 2 Average Discrepancy from Δ*K* and Posterior Probability (+/- SEM). The alternative estimators include MaxMeaK (maximum of means), the MaxMedK (maximum of medians), the MedMeaK (median of means), and the MedMedK (median of medians). Each estimator was calculated at four membership coefficient thresholds (0.5, 0.6, 0.7, 0.8) which are shown on the x-axis. For the 'focused' category (*n*=26), the difference from each of the 16 alternative estimators to (**A**) the Δ*K* estimate and to (**B**) the Posterior Probability was calculated and averaged across all 26 studies for each estimator (shown on the y-axis). For the 'broad' category

(*n*=15), the difference from the alternative estimators to (**C**) the Δ*K* estimate and to (**D**) the Posterior Probability is also shown. SEM = Standard Error of the Mean.

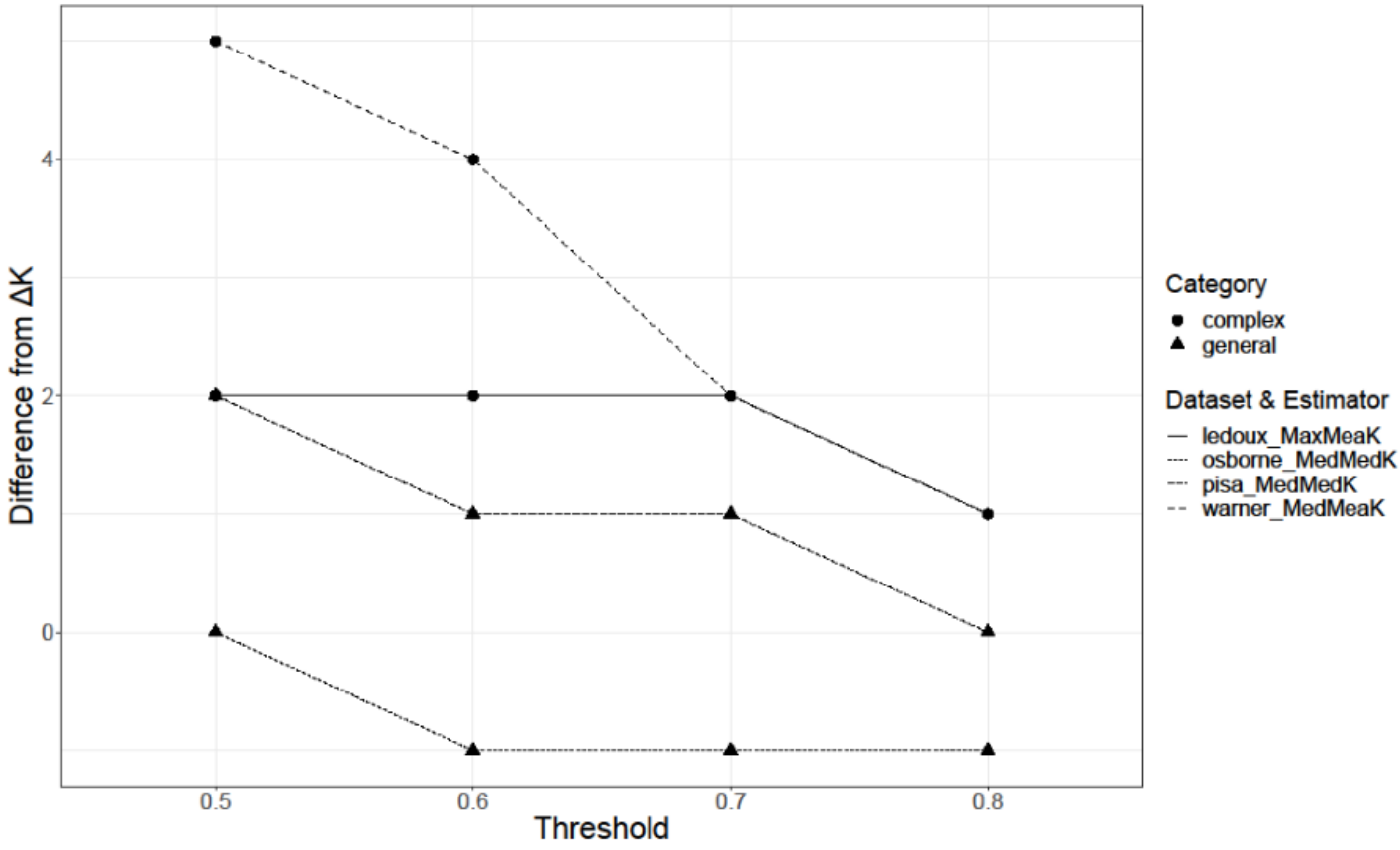
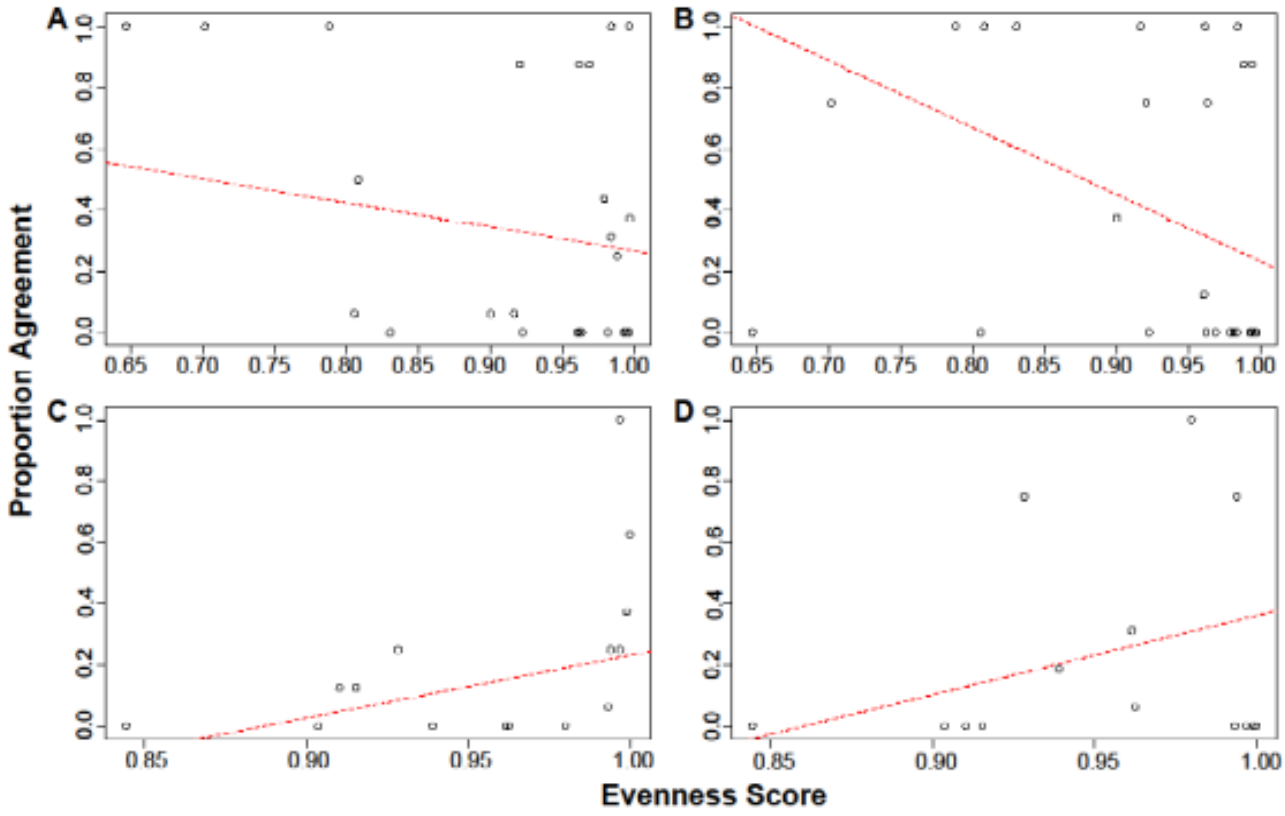


Fig. 3 Difference from Δ*K* by threshold. A randomly selected subset of the alternative *K* estimators from a randomly selected subset of datasets from both the 'focused' and 'broad' categories is shown here to illustrate the effect of threshold for the alternative estimators (Puechmaille 2016) on the magnitude of deviation from Δ*K*.



768 **Fig. 4 Sampling Scheme and Estimator Precision.** For each dataset, an evenness score was
769 calculated using the Shannon Diversity Index (plotted on the x-axis). Studies which had a more
770 even number of samples taken from each site had a higher score (between 0 and 1). The
771 proportion of alternative estimators that agreed with (**A**) the Δ*K* estimate and (**B**) the Posterior
772 Probability was calculated for each dataset in the 'focused' category (*n*=26) and plotted on the y-
773 axis. A linear regression was plotted (red-dotted line) to show the relationship between evenness
774 and proportion agreement with the (**A**) Δ*K* estimate (Adj. R^2 = -0.016; Intercept = 1.0487; Slope
775 = -0.7810; *p*-value = 0.4438) and (**B**) the Posterior Probability (Adj. R^2 = 0.07939; Intercept =

776 2.415; Slope = -2.182; *p*-value = 0.08833). Similarly, for datasets in the 'broad' category (*n*=15)
777 a linear regression between evenness and proportion agreement with (**C**) Δ*K* estimate (Adj. R^2 =
778 0.1722; Intercept = -1.8212; Slope = 2.0533; *p*-value = 0.06951) and (**D**) the Posterior
779 Probability (Adj. R^2 = 0.02763; Intercept = -2.222; Slope = 2.583; *p*-value =0.2583) is shown. In
780 all cases, each point is weighted according to sample size.

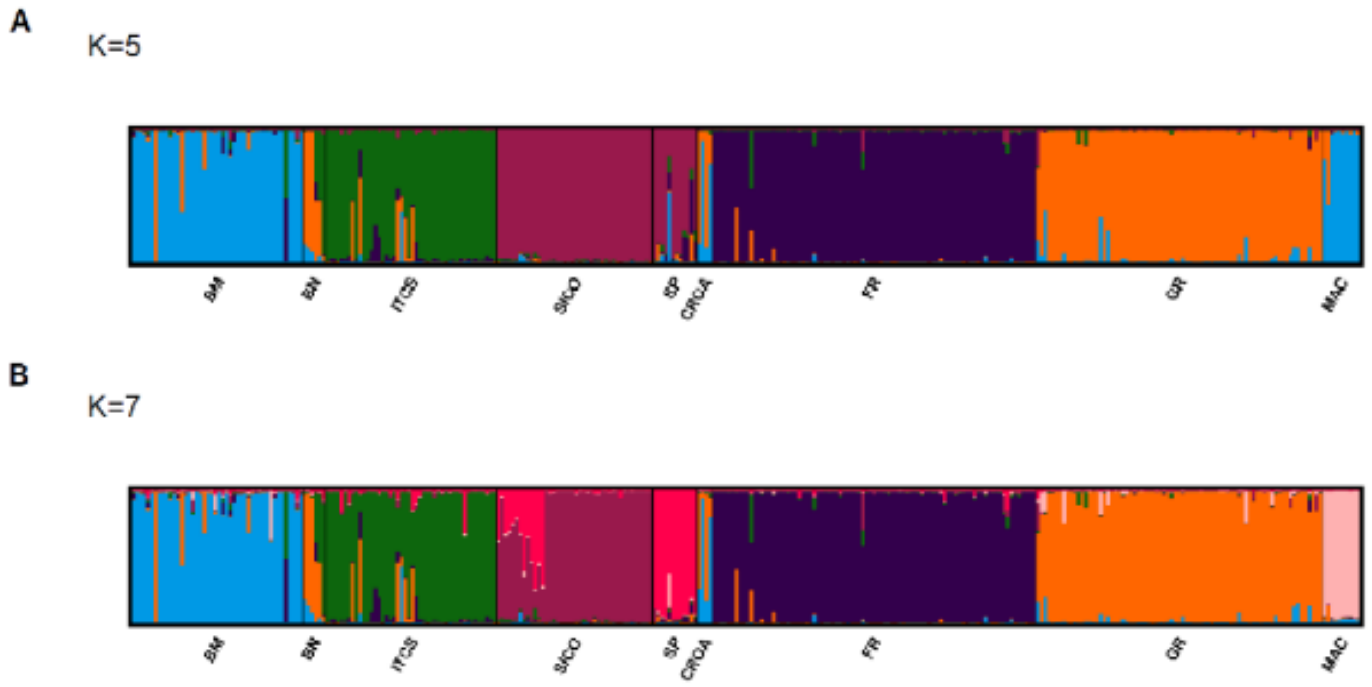


Fig. 5 Membership plots for *Testudo hermanni*. Membership plots for STRUCTURE runs when (**A**) *K*=5 and (**B**) *K*=7 for the 'broad' category dataset reanalyzing Perez et al. (2014) data for Hermann's Tortoise (*Testudo hermanni*). According to Perez et al. (2014) BM = Bosco Mesola population (Italy), BN = Bosco Nordio population (Italy), ITCS = Central and Southern Italian population (Italy), SICO = Sicilian and Corsican population (Italy, France), SP = Spain Population (Spain), CROA = Croatian population (Croatia), FR = French population (France), GR = Greek population (Greece), and MAC = Macedonian population (Macedonia).

Table 1. *K* values according to each estimator for datasets randomly selected for AMOVA from the 'broad' category (a) and the 'focused' category (b).

Dataset ID	*K*	Estimator support
(a) kim_et_al_2017	2	Δ*K*
	5	MaxMeaK0.8, MedMeaK0.8
	6	MedMeaK0.5, MedMeaK0.6, MedMeaK0.7, MedMedK0.5, MedMedK0.6, MedMedK0.7, MedMedK0.8
	7	MaxMeaK0.5, MaxMeaK0.6, MaxMeaK0.7, MaxMedK0.5, MaxMedK0.6, MaxMedK0.7, MaxMedK0.8
	8	PPK
(a) perez_et_al_2014	2	Δ*K*
	7	All Puechmaille estimators
	12	PPK
(b) baums_et_al_2010_1	1	All Puechmaille estimators
	2	Δ*K*
	3	PPK
(b) rippe_et_al_2017	2	MaxMeaK0.8, MedMeaK0.8
	3	MaxMeaK0.6, MaxMeaK0.7, MaxMedK0.6, MaxMedK0.7, MaxMedK0.8, MedMeaK0.6, MedMeaK0.7, MedMedK0.5, MedMedK0.6, MedMedK0.7, MedMedK0.8
	4	MaxMeaK0.5, MaxMedK0.5, MedMeaK0.5
	5	Δ*K*
	10	PPK

Table 2. Analysis of molecular variance (AMOVA) across clustering solutions for randomly selected datasets in the 'broad' category (a) and the 'focused category (b).

Dataset ID	*K*	Partitioning	*df*	Sum of squares	Variance	% Variation
(a) kim_et_al_2017						
	2					
		Between clusters	1	349.275	1.470	24.951
		Between samples within clusters	318	1716.631	0.976	16.561
		Within samples	32	1102.861	3.446	58.488

		Total	639	3168.767	5.893	100.000
	5					
		Between clusters	4	697.017	1.333	25.490
		Between samples within clusters	315	1368.889	0.450	8.599
		Within samples	320	1102.861	3.446	65.911
		Total	639	3168.767	5.229	100.000
	6					
		Between clusters	5	722.428	1.340	25.753
		Between samples within clusters	314	1343.478	0.416	7.998
		Within samples	320	1102.861	3.446	66.249
		Total	639	3168.767	5.202	100.000
	7					
		Between clusters	6	758.122	1.360	26.299
		Between samples within clusters	313	1307.783	0.366	7.074
		Within samples	320	1102.861	3.446	66.628
		Total	639	3168.767	5.173	100.000
	8					
		Between clusters	7	816.925	1.420	27.603
		Between samples within clusters	312	1248.980	0.278	5.410
		Within samples	320	1102.861	3.446	66.987
		Total	639	3168.767	5.145	100.000
(a) perez_et_al_2014						
	2					
		Between clusters	1	789.683	2.424	36.605
		Between samples within clusters	328	1791.112	1.263	19.082
		Within samples	330	968.184	2.934	44.313
		Total	659	3548.979	6.621	100.000
	7					
		Between clusters	6	1320.461	2.438	41.637
		Between samples within	32	1260.334	0.484	8.265

		clusters	3			
		Within samples	330	968.184	2.934	50.098
		Total	659	3548.979	5.856	100.000
(b) baums_et_al_2010_1						
	2					
		Between clusters	1	52.898	0.692	16.837
		Between samples within clusters	180	652.745	0.209	5.081
		Within samples	182	584.000	3.209	78.082
		Total	363	1289.643	4.110	100.000
	3					
		Between clusters	2	66.981	0.821	19.512
		Between samples within clusters	179	638.662	0.180	4.266
		Within samples	182	584.000	3.209	76.222
		Total	363	1289.643	4.210	100.000
(b) rippe_et_al_2017						
	2					
		Between clusters	1	104.173	0.305	5.400
		Between samples within clusters	367	2483.876	1.429	25.322
		Within samples	369	1442.744	3.910	69.278
		Total	737	4030.792	5.644	100.000
	3					
		Between clusters	2	160.059	0.299	5.362
		Between samples within clusters	366	2427.990	1.362	24.450
		Within samples	369	1442.744	3.910	70.188
		Total	737	4030.792	5.571	100.000
	4					
		Between clusters	3	218.816	0.364	6.550
		Between samples within clusters	365	2369.233	1.291	23.191
		Within samples	369	1442.744	3.910	70.259
		Total	737	4030.792	5.565	100.000

			7		
	5				
		Between clusters	4	263.692	0.416	7.480
		Between samples within clusters	36 4	2324.357	1.238	22.248
		Within samples	36 9	1442.744	3.910	70.272
		Total	73 7	4030.792	5.564	100.000
	10				
		Between clusters	9	386.415	0.525	9.470
		Between samples within clusters	35 9	2201.634	1.111	20.038
		Within samples	36 9	1442.744	3.910	70.492
		Total	73 7	4030.792	5.547	100.000

Note

df = degrees of freedom

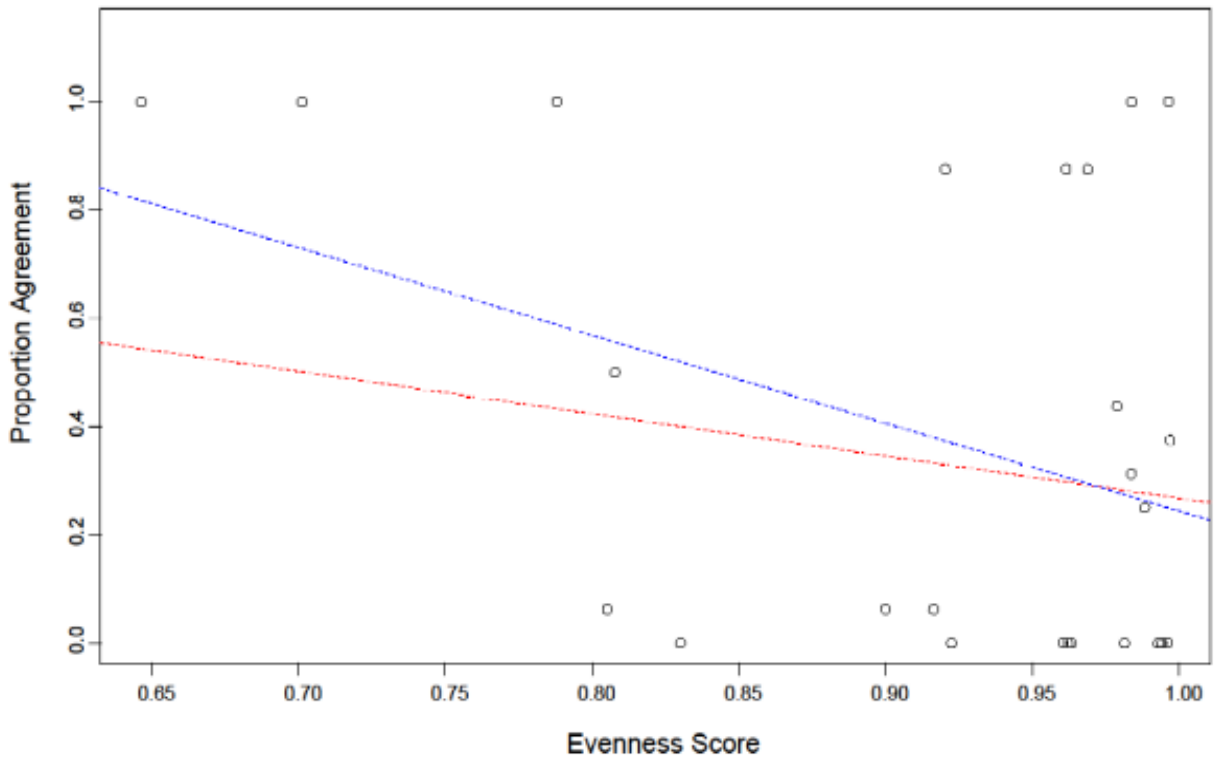


Fig. S1 Weighted versus unweighted linear regression: Focused category. Each point in Fig. 5A was weighted by sample size (blue line) and compared to the results of the unweighted

785 regression (red line). For the unweighted regression: Adj. R^2 = 0.1051; Intercept = 1.8666; Slope
786 = -1.6223; p-value = 0.0588. For the weighted regression: Adj. R^2 = -0.016; Intercept = 1.0487;
787 Slope = -0.7810; p-value = 0.4438.

788