Matters Arising:

Speciation in stickleback facilitated by admixture – where is the evidence?

Daniel Berner, Department of Environmental Sciences, Zoology, University of Basel, CH-4051 Basel, daniel.berner@unibas.ch

Abstract

Where genetic variation promoting speciation originates is a crucial question in evolutionary genomics. In a recent article, Marques et al.[1] seek to address this question in lake and stream threespine stickleback fish from the Lake Constance (hereafter LC) basin in Central Europe. Based on population genetic methods, they conclude that incipient speciation between lake and stream stickleback was facilitated by the mixing of genetic variation from old lineages evolved in isolation (i.e., admixture following secondary contact). In this comment, I discuss conceptual and methodological problems and unrecognized conflicts with existing evidence that cast doubt on Marques et al.'s conclusion.

Keywords: evolutionary genomics, Gasterosteus aculeatus, gene flow, hybridization, phylogeny

The origin of stickleback in the LC basin

Marques et al. argue that threespine stickleback populations in the LC basin result from a contact between two deeply separated lineages from Northeastern and Western Europe. In previous work, the authors maintained firmly that stickleback were absent from the LC basin until the late 19th century, when colonization was initiated by human introductions from both lineages[2]. Although the authors adopt a more cautious stance in their present article, they still imply that introductions played an important role in the colonization of the LC basin. To reconsider the evidence of genetically distinct western and northeastern lineages and the origin of stickleback in the LC basin, I generated the currently most detailed nuclear phylogeny for Central European stickleback by combining genome-wide sequence data from 39 freshwater populations.

This phylogeny (Fig. 1, Supplementary Methods) confers two major insights: first, European stickleback populations separate deeply into a Mediterranean and Black Sea lineage on the one hand, and a Central, Eastern, and Northern European lineage on the other

hand (for similar evidence based on ordination see Supplementary Fig. 1). This dichotomy is consistent with a recent phylogeographic investigation[3] establishing that the circum-Mediterranean and Black Sea lineage reflects an ancient southern refugial ancestor, whereas the more northern populations derive from a large-scale postglacial surge in southwestward direction via an ancient Baltic Sea. However, the phylogeny does not support Marques et al.'s claim of the existence of an ancient, genetically distinct Western European stickleback lineage 'evolved in isolation for several thousand generations'[1]: statistical support for the monophyly of the authors' western lineage (indicated by a gray square in Fig. 1) is poor, and the basal branch of this western lineage is not deeper than branches representing populations from other drainages in Northern or Eastern Europe.

The second insight from the phylogeny is that stickleback from the LC basin prove closely related to populations from the Danube drainage (Fig. 1; Supplementary Fig. 1). The broad-scale colonization history of Central Europe[3] in mind, this close genetic relatedness supports the possibility that stickleback in the LC basin may originate from the natural westward colonization by Northeastern European fish via the Danube drainage. The LC basin nowadays drains into the Atlantic via the river Rhine, so its colonization via the Danube drainage may appear counter-intuitive. However, during the retreat of the Pleistocene ice cover, the present-day LC basin drained in an eastward direction via the Danube[4]. Even today, the Danube and the LC drainage remain connected through a sinkhole and a 12 km underground stream system[5] inhabited by fish[6]. Moreover, other fish species within the LC basin, such as European perch (*Perca fluviatilis*), seem to have Danubian ancestors[7]. Given the earlier suggestion based on mitochondrial DNA sequence data that the LC basin may have been colonized via the Danube drainage[8] (albeit not from the Black Sea[3]), it is hard to follow why Marques et al. ignore Danubian stickleback in their analyses; sequence data from Danubian populations would have been available from the same study[3] that Marques et al. used as source for other genomic data.

Evidence of a natural, postglacial colonization of the LC basin also emerges from the demographic analyses presented by the authors themselves: their splitting times among LC populations include estimates of up to 2800 generations (one generation equals approximately one year) before present. Although interpreted as model imprecision or bias, these ancient splitting times – incompatible with an anthropogenic origin – are in line with an

earlier, independent estimate of the splitting age of stream populations in the LC basin (around 2300 generations; Supplementary Fig. 2 in ref. 9).

Taken together, the new fine-grained analysis of population structure among European stickleback presented in this comment questions Marques et al.'s assumption of a Western European stickleback lineage with a long history of evolution isolated from other non-Mediterranean populations. Furthermore, a natural postglacial colonization of the LC basin by stickleback via the Danube drainage appears plausible both phylogeographically and palaeohydrologically. I no way doubt that stickleback may have been transferred within and/or introduced into the LC basin in historical times, nor that a pelagic population within Lake Constance proper may have arisen more recently[9]. But the species may well be native; at least the stream ecotype may have appeared in the LC basin thousands of generations ago.

Inappropriate taxonomy

A second problem in the study by Marques et al. is their view that Western and Northeastern European stickleback coincide with two distinct nominal species, *Gasterosteus gymnurus* and *G. aculeatus,* delimited in the early 19th century[10] based exclusively on the absence vs. presence of body plating (a predator defense trait). It is widely recognized, however, that such a morphological species concept is not meaningful in stickleback[11]. One reason is parallel evolution: low- and completely plated populations have evolved numerous times independently by recurrently recruiting the same, wide-spread ancient genetic variants[12]. Within Europe, for instance, *G gymnurus* would need to include all (low-plated) populations from the circum-Mediterranean lineage (red branch in Fig. 1), but also some low-plated populations nested within the Central and Northeastern European lineage (blue branch in Fig. 1; e.g., the CHA or DOR populations). However, the latter are clearly derived from an ancestor shared with Marques et al.'s *G. aculeatus*, and hence are very distantly related to the Mediterranean lineage. Similarly, the authors' northeastern lineage, equated with completely plated *G. aculeatus*, includes low-plated populations[13]. The two nominal species invoked by Marques et al. are thus clearly polyphyletic, and hence fail to mirror evolutionary history.

Furthermore, numerous stickleback populations worldwide are known to be polymorphic for body plating. That is, they contain both low- and completely plated individuals, and

typically also intermediate, partially plated phenotypes. Some of these populations have been polymorphic for decades and in geographic isolation[14-16], thus challenging both a morphology-based taxonomy and the notion that polymorphism in body plating indicates population admixture. This view is reinforced by low-plated stickleback populations that have evolved[17], or are presently evolving (the SAS population in Fig. 1)[18], toward complete plating in response to habitat changes altering predation risk. Applying the species delimitation of Marques et al. would here lead to the absurd view that stickleback populations transform from one nominal species into another.

A further challenge to the authors' taxonomy is that body plating plays no exceptional role in stickleback reproductive isolation. Within natural populations polymorphic for body plating, for instance, this phenotype has not been observed to influence mate choice[15]. Conversely, within both low-plated and plate-polymorphic stickleback systems, the emergence of strong reproductive isolation without divergence in plating is commonplace. The partially reproductively isolated[19] lake and stream stickleback form the LC basin are a case in point: divergent selection on the genetic polymorphism underlying plate variation is under weak selection between these ecotypes compared to other genomic regions (ref. 9; Supplementary Fig. 1 in ref. 20)

To conclude, it is hard to follow why Marques et al. resurrect a morphological species concept at odds with decades of stickleback research demonstrating extensive parallel evolution in body plating and the rapid adaptive sorting of the underlying standing genetic variation. The populations in Central Europe, as across the northern hemisphere at large, are most profitably considered members of the phenotypically diverse threespine stickleback *G. aculeatus*: 'Given our current state of knowledge about *Gasterosteus*, it would be unwise to apply species labels to any population, lineage, or "species pair"'[11]. I emphasize that adopting a species delimitation lacking conceptual justification, or assigning a species a dubious origin category (native vs. introduced, see previous section), may mislead conservation efforts and fish stock management decisions. I thus feel that clarifying the taxonomic confusion in Marques et al., and emphasizing uncertainty in the origin of stickleback in the LC basin, is more than a semantic subtlety of academic interest.

Overconfidence in population genetic methods

4

Another problem in Marques et al. is that strong conclusions about evolutionary history are derived from population genetic analyses without carefully acknowledging potential violations of the underlying assumptions, and ambiguity in their interpretation. A major issue is that their main methodological tools (demographic modeling, D statistic) assume selectively neutral evolution. However, natural selection imposed by novel local ecological conditions profoundly re-structures genetic variation all across the stickleback genome[21]. That such selection has the potential to bias demographic inference has been indicated in stickleback from the LC basin[9]. For fairness, it must be highlighted that Marques et al. attempt to reduce potential bias due to selection in their demographic analyses by excluding markers located in chromosome regions exhibiting a particularly low recombination rate. While this data manipulation may increase analytical robustness in older organismal systems in which genomic variation is shaped by background selection (a mutation-driven process), it will be ineffective in a young, postglacial system like stickleback strongly influenced by rapid directional selection of standing genetic variation. This skepticism is confirmed directly by a recent genomic analysis of lake-stream stickleback from the LC basin based on whole-genome marker resolution, revealing that signatures of divergent selection are neither less common nor less extensive physically in the chromosome peripheries exhibiting high recombination rates[22] (Fig. 2 and Supplementary Fig. 1 in ref. 20). Likewise, it is easy to imagine that selection can bias D statistics: Marques et al. argue that Northeastern European stickleback are phenotypically closer to ancestral marine stickleback than are Central European populations. However, stickleback within LC are selected for a pelagic life style resembling that of marine fish[8,9], hence greater allele sharing between the LC population and Northeastern European populations is an expected outcome of local adaptation potentially confounding the inference of admixture.

Also, Marques et al. claim to evaluate several possible demographic scenarios for the colonization and subsequent divergence of stickleback in the LC basin, and in particular to refute an 'ecological vicariance' scenario (Fig. 3 in ref. 9). An inherent element of ecological vicariance, however, is population differentiation caused by strong divergent selection, for which there is clear experimental evidence in lake and stream fish from the LC basin[19,20]. Marques et al.'s demographic analysis, assuming the absence of selection, must therefore fail to offer an adequate comparison of relevant evolutionary scenarios.

I further highlight that the population genetic analyses presented by the authors are contingent on the specific populations considered for analysis. Importantly, only a single population sample (the BRO population indicated by a triangle in Fig. 1) is used as representative of the Northeastern European lineage. The phylogeny in Fig. 1, however, casts doubt on this choice; numerous European stickleback populations, in particular those from the Danube drainage or even from the North Sea region, appear genetically closer to stickleback from the LC region. Including samples from these populations may have led to qualitatively different conclusions.

Finally, the estimates of population splitting times among stickleback populations within the LC basin uncover a profound dilemma in the Marques et al. study. As mentioned above, their demographic modeling returns splitting time estimates up to a few thousand generations, estimates appearing plausible when considering natural postglacial colonization. However, the authors seem to favor a much more recent, historical origin of stickleback in the LC basin, implying that some of their demographic models must here miss the truth by a factor of up to nearly twentyfold. Given this bias, why should other demographic parameter estimates in the study be more accurate? Alternatively, the estimated splitting times may indeed be accurate – but this would fundamentally challenge the conceptual framework adopted by the authors in the present and previous studies.

These methodological caveats in mind, Marques et al.'s inference of an admixture history must be regarded speculative and no more plausible than alternative scenarios.

Facilitation of speciation?

A final issue is that even if we assume that admixture between distinct lineages has occurred in the LC basin, the evidence presented by Marques et al. is insufficient for demonstrating that this has promoted reproductive isolation. A first challenge is that within the LC basin, the lake population has adapted relatively recently to the pelagic ecological niche and represents the most derived ecotype in that region, while stream-adapted populations are ancestral[9]. Admixture would thus be expected to promote speciation only if it introduced genetic variation for pelagic adaptation not previously present in the basin already. However, such variation can hardly come from the authors' western lineage, which includes only stream-adapted populations. The proposed admixture therefore appears inconsequential a priori. More

generally, the recent admixture between lineages claimed to have evolved in isolation for several thousand generations leads to the prediction that the populations in the LC basin should be exceptionally genetically diverse. A broad comparison of genetic diversity across European populations, however, reveals that stickleback in the LC basin are no more genetically diverse than populations from other European drainages (Supplementary Fig. 2). This simple observation fundamentally questions an admixture scenario.

Another challenge is that partially reproductively isolated lake and stream stickleback populations are known from numerous watersheds worldwide[23-25], and for many of these incipient species pairs, an admixture origin is implausible biogeographically. Given the ease with which lake-stream divergence emerges, demonstrating admixture as a driver of speciation becomes methodologically demanding and requires i) identifying the specific haplotypes (DNA sequence stretches) holding alleles involved in divergent adaptation, and ii) demonstrating that such haplotypes were initially missing in one or the other original population. However, none of the authors' genetic population samples from outside the LC basin include more than seven individuals, thus precluding robust estimates of haplotype frequencies needed for inferring variational constraints broken by admixture. More fundamentally, the sparse marker resolution of Marques et al. is insufficient for haplotype-level inference in the first place[26]; whole-genome resolution and phased genotype data would be needed. For a few genome regions under divergent lake-stream selection, such haplotype data have been generated in previous studies by targeted sequencing in stickleback from both the LC basin and the authors' western lineage. These regions include the EDA locus underlying variation in body plating (Fig. 5b in ref. 18), and three large inversions (Fig. 7c in ref. 9). These data demonstrate that the genetic variants underlying lake-stream divergence within the LC basin are not only ubiquitous across Europe, but shared among populations on a worldwide scale. The strongest sequence-based evidence currently available thus indicates that lake and stream stickleback within the LC basin have diversified just as stickleback populations do everywhere: by sorting abundant standing genetic variation preexisting in their ancestors. Conclusively assessing a potential contribution of admixture to adaptive diversification and speciation would require a methodological stringency beyond the standard of Marques et al.'s work.

To conclude, in the light of the numerous problems highlighted in this note, Marques et al.'s claim to have 'demonstrated that secondary contact between divergent lineages and the re-assortment of introgressed alleles […] underlie recent ecological speciation' is overconfident and lacks convincing empirical evidence.

# References

1. Marques, D. A., Lucek, K., Sousa, V. C., Excoffier, L. & Seehausen, O. Admixture between old lineages facilitated contemporary ecological speciation in Lake Constance stickleback. *Nat. Commun.* **10**, 4240 (2019).

2. Marques, D. A. *et al.* Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genet.* **12**, e1005887 (2016).

3. Fang, B., Merilä, J., Ribeiro, F., Alexandre, C. M. & Momigliano, P. Worldwide phylogeny of three-spined sticklebacks. *Mol. Phylogenet. Evol.* **127**, 613–625 (2018).

4. Keller, O. & Krayss, E. Die Hydrographie des Bodenseeraums in Vergangenheit und Gegenwart. *Ber. St. Gall. Naturwiss. Ges.* **89**, 39–56 (2000).

5. Hötzl, H. Origin of the Danube-Aach system. *Environ. Geol.* **27**, 87–96 (1996).

6. Behrmann-Godel, J., Nolte, A. W., Kreiselmaier, J., Berka, R. & Freyhof, J. The first European cave fish. *Curr. Biol.* **27**, R257–R258 (2017).

7. Behrmann-Godel, J., Gerlach, G. & Eckmann, R. Postglacial colonization shows evidence for sympatric population splitting of Eurasian perch (*Perca fluviatilis* L.) in Lake Constance. *Mol. Ecol.* **13**, 491–497 (2004).

8. Moser, D., Roesti, M. & Berner, D. Repeated lake-stream divergence in stickleback life history within a Central European lake basin. *PLoS One* **7**, e50620 (2012).

9. Roesti, M., Kueng, B., Moser, D. & Berner, D. The genomics of ecological vicariance in threespine stickleback fish. *Nat. Commun.* **6**, 8767 (2015).

10. Cuvier, G. L. & Valenciennes, M. A. *Histoire naturelle des poissons*. (Levrault, 1829).

11. Mattern, M. Y. Phylogeny, systematics, and taxonomy of sticklebacks. in *Biology of the three-spined stickeback* (eds. Östlund-Nilsson, S., Mayer, I. & Huntingford, F. A.) 1–40 (CRC, 2007).

12. Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).

13. Münzing, J. The evolution of variation and distributional patterns in European populations of the three-spined stickleback, Gasterosteus aculeatus. *Evolution (N. Y).* **17**, 320–332 (1963).

14. Avise, J. C. Genetics of plate morphology in an unusual population of threespine sticklebacks (Gasterosteus aculeatus). *Genet. Res.* **27**, 33–46 (1976).

15. Marchinko, K. B., Matthews, B., Arnegard, M. E., Rogers, S. M. & Schluter, D. Maintenance of a genetic polymorphism with disruptive natural selection in stickleback. *Curr. Biol.* **24**, 1289–1292 (2014).

16. Yamasaki, Y. Y., Mori, S., Kokita, T. & Kitano, J. Armour plate diversity in Japanese freshwater threespine stickleback (Gasterosteus aculeatus). *Evol. Ecol. Res.* **20**, 51–67 (2019).

17. Kitano, J. *et al.* Reverse evolution of armor plates in the threespine stickleback. *Curr. Biol.* **18**, 769–774 (2008).

18. Berner, D., Roesti, M., Hendry, A. P. & Salzburger, W. Constraints on speciation suggested by comparing lake-stream stickleback divergence across two continents. *Mol. Ecol.* **19**, 4963–4978 (2010).

19. Moser, D., Frey, A. & Berner, D. Fitness differences between parapatric lake and stream stickleback revealed by a field transplant. *J. Evol. Biol.* **29**, 711–719 (2016).

20. Laurentino, T. G. *et al.* Genomic release-recapture experiment in the wild reveals within-generation polygenic selection in stickleback fish. *Nat. Commun.* **11**, 1928 (2020).

21. Bassham, S., Catchen, J., Lescak, E., von Hippel, F. A. & Cresko, W. A. Repeated selection of alternatively adapted haplotypes creates sweeping genomic remodeling in stickleback. *Genetics* **209**, 921–939 (2018).

22. Roesti, M., Moser, D. & Berner, D. Recombination in the threespine stickleback genome - patterns and consequences. *Mol. Ecol.* **22**, 3014–3027 (2013).

23. Berner, D., Grandchamp, A.-C. & Hendry, A. P. Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. *Evolution* **63**, 1740–1753 (2009).

24. Deagle, B. E. *et al.* Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proc. R. Soc. B* **279**, 1277–1286 (2012).

25. Ravinet, M., Prodoehl, P. A. & Harrod, C. Parallel and nonparallel ecological, morphological and genetic divergence in lake-stream stickleback from a single catchment. *J. Evol. Biol.* **26**, 186–204 (2013).

26. Lowry, D. B. *et al.* Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol. Ecol. Res.* **17**, 142–152 (2017).

Fig. 1
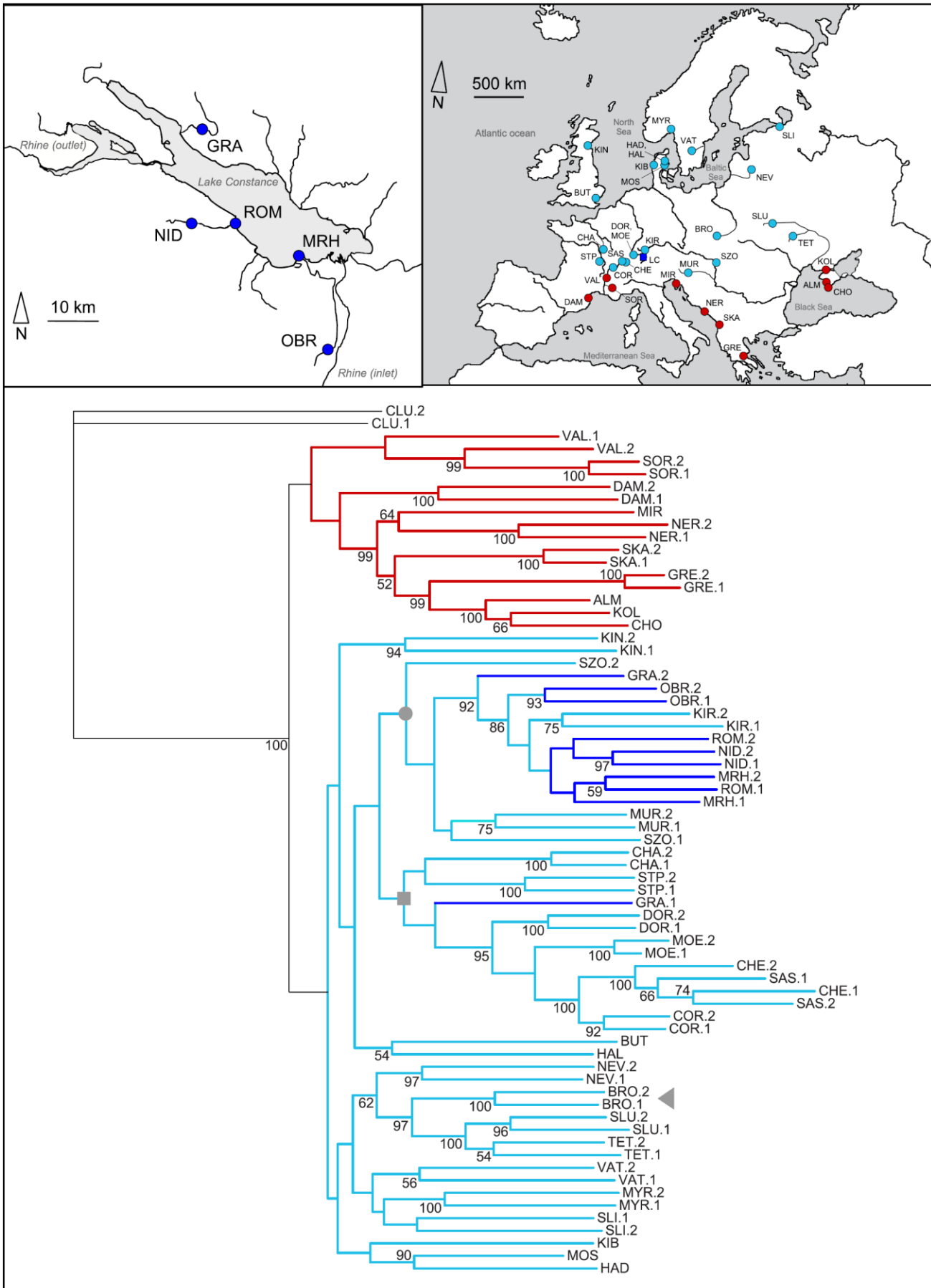
Fig. 1. Phylogeny of European threespine stickleback populations. The phylogram (maximum likelihood tree) is based on DNA sequence data from 69 total stickleback individuals from the 39 freshwater populations indicated in the maps (1-2 individuals per population). The left map represents a close-up of the Lake Constance (LC) region, located by a square in the right map, showing the precise situation of the two lake (ROM, MRH) and three stream (GRA, NID, OBR) sample sites. The color coding separates the populations belonging to the circum-Mediterranean and Black Sea lineage (red) from those belonging to the Central, Northern and Eastern European lineage (blue; populations from the LC basin are labeled in dark blue). The values next to nodes give the strength of monophyly of the corresponding branches based on bootstrapping (500 iterations; shown only for values >= 50%). Note the strong bootstrap support for the reciprocal monophyly of the two major (red, blue) stickleback lineages in Europe. By contrast, the basal nodes within the blue lineage lack bootstrap support, thus challenging Marques et al.'s assumption of an old, genetically distinct Western European lineage (the basal node of this lineage is indicated by a gray square). The branch marked by the gray dot contains exclusively populations from the LC basin and from the Danube river, highlighting their close genetic relatedness. The gray triangle indicates the BRO population chosen by Marques et al. as representative of their Northeastern European lineage. Two individuals derived from Pacific ancestors (CLU) served as outgroup.