Biological Journal of the Linnean Society, 2018, **125**, 115–125. With 4 figures.

Contrasted patterns of divergence and gene flow among five fish species in a Mongolian rift lake following glaciation

IVANA ROMAN[1†], YANN BOURGEOIS[2†], JACOBO REYES-VELASCO[2], OLAF P. JENSEN[3], JOHN WALDMAN[1*] and STEPHANE BOISSINOT[2*]

[1]*Department of Biology, Queens College, the City University of New York, Flushing, NY, USA*
[2]*New York University Abu Dhabi, Saadiyat Island Campus, Abu Dhabi, United Arab Emirates*
[3]*Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA*

Received 2 April 2018; revised 2 June 2018; accepted for publication 2 June 2018

Lakes and rivers in the Baikal ecoregion of central Asia provide particularly interesting models to assess the emergence and maintenance of species diversity, as they display strong contrasts in the composition of their aquatic vertebrate communities. Here, we used RAD-seq to study the recent evolutionary history of five fish species from Lake Hovsgol in Mongolia. This ancient lake was probably recolonized from its outlet following the Last Glacial Maximum, and harbours only ten native fish species. We detected substantial genetic differentiation between the lake and the putative ancestral refuge river in five species but we also found that these species have experienced different population dynamics. Some species appear to have colonized the lake after the Last Glacial Maximum, while others harbour levels of genetic differentiation consistent with the presence of refugia in the lake that could have persisted during glacial maxima or suggestive of colonization from a distinct source. We further demonstrated that fish species have experienced different levels of gene flow following colonization of the lake, suggesting that ecology and habitat use have had a substantial impact on the differentiation of lake populations.

ADDITIONAL KEYWORDS: Baikal ecoregion – fish communities – Lake Hovsgol – population genomics – RAD-seq.

INTRODUCTION

The study of island biota has been crucial to our understanding of how colonization and subsequent competition among colonist species shape local communities (Warren *et al.*, 2015). Lakes are similar to islands in the way that they can be colonized and exchange migrants with nearby populations. Isolated islands and lakes are expected to display species communities that are divergent from their continental counterparts, with opportunities for *in situ* diversification (Keller *et al.*, 2013). This process can occur quickly in fish populations, with multiple well-known examples provided by cichlid fishes in several lakes of Africa and Central America (Elmer *et al.*, 2010). At the same time, continuous immigration from a continental (or equivalent) source can prevent local adaptation (Kawecki & Ebert, 2004). At the scale of communities, comparative population genetics can quantify the extent to which multiple species share the same recent history and can help to identify processes that promote or prevent local differentiation, such as time of colonization, local adaptation, behaviour, dispersal abilities or generation times (Vellend, 2005). It also provides much needed insights on the relative roles of drift and selection in shaping divergence between populations.

We investigated the early stage of differentiation in fish species that have colonized Lake Hovsgol since the Last Glacial Maximum (LGM). Lake Hovsgol (Khövsgöl Nuur) is a rift lake located in northern Mongolia and is one of the world's oldest, largest and most unusual freshwater bodies (Goulden *et al.*, 2006). Lake Hovsgol belongs to the same watershed as Lake Baikal. Its outlet is the origin of the lengthy Eg River, a tributary of the Selenga River, which enters Lake Baikal, situated about 200 km east of Hovsgol. Despite its age and size, Hovsgol harbours fewer than 400 species and only ~20

*Corresponding authors. E-mails: john.waldman@qc.cuy.edu and stephane.boissinot@nyu.edu.
†These two authors contributed equally to the research.

endemics (Karabanov *et al.*, 2004). This low level of diversity is in contrast to that of Lake Baikal, one of the world's most biodiverse lakes, with more than half of its 2500 animal species and 30% of its plant species being endemic to it (Moore *et al.*, 2009). Although Lake Hovsgol is estimated to be between 2 and 5 Myr old, its biota is thought to be of recent origin due to the glacial history of the region. Sediment analyses from a single core in Lake Hovsgol suggest that during the LGM, diatoms, chrysophyte cysts, sponge spicules and zooplankton remains did not accumulate (Karabanov *et al.*, 2004). Karabanov *et al.* interpreted the absence of these remains as an indication of a drastic lowering of photosynthetic production and, perhaps, the collapse of the lake's ecosystem. Moreover, lake levels probably fell as much as 200 m during the LGM (Fedotov *et al.*, 2002). Starting from ~12 000 to 13 000 years BP, the planktonic communities redeveloped and primary production increased, as shown by the composition of Holocene sediments (Karabanov *et al.*, 2004). From this perspective, Lake Hovsgol can be considered equivalent to a recently formed island colonized from a reservoir of diversity (in this case, the Eg River), with opportunities for continuous genetic exchange between source and sink populations.

Only ten native fish species are known from Lake Hovsgol (Sideleva, 2006; Ahrenstorff *et al.*, 2012; Young *et al.*, 2015) while one species has been introduced. An additional 12 fish species occur in the nearby Selenge River, but are not known from Lake Hovsgol. Despite the low diversity of fish species in the lake, the native species are distributed across eight genera and seven families. This shows that there is considerable phylogenetic diversity, and associated morphological, ecological and life history variation within its ichthyofauna, which makes this a functionally diverse fish community. Most of the current fish community of Hovsgol probably arrived after the deglaciation and subsequent warming of the lake. The post-glacial lake has only minor tributaries that were probably too small and ice covered to have served as refugia, and thus it is believed that the lake was recolonized from its outlet waters (Karabanov *et al.*, 2004). However, it remains possible that isolated bays in the lake served as refugia, where divergence from other lake populations could have occurred.

Given these uncertainties, we decided to test whether the level of differentiation between fish populations from the lake and the river was consistent with a single post-glacial episode of re-colonization or if the lake acted as a refugium for some species. We collected samples of five fish species that occur both in the Eg River and one of its main tributaries, the Uur River, as well as in Lake Hovsgol. We used thousands of single nucleotide polymorphisms (SNPs) obtained from double digest restriction site associated DNA sequencing (ddRAD-seq) to infer the genetic structure of those species from both river and lake populations. We found that the amount of gene flow between lake and river as well as the timing of re-colonization of the lake differs substantially among taxa, suggesting that the apparently simple fish community of Lake Hovsgol is the result of heterogeneous colonization processes.

MATERIALS AND METHODS

Sampling

We collected tissue samples of five fish species from nine localities in Lake Hovsgol and the Eg and Uur Rivers in July 2011 (Fig. 1 and Supporting Information, Table S1). The other five species represented in the lake were not analysed because we failed to collect sufficient samples from both the lake and the river. The five species are representative of four families: a Cyprinidae, the Eurasian minnow (*Phoxinus phoxinus*), a Nemacheilidae, the Siberian stone loach (*Barbatula toni*), a Lotidae, the burbot (*Lota lota*), and two Salmonidae, the lenok (*Brachymystax lenok*) and the grayling (*Thymallus arcticus nigrescens*). Graylings from Lake Hovsgol are thought to be a different subspecies (*Thymallus arcticus nigrescens*) from those in the rivers (*T. arcticus baicalensis*) and are morphologically distinct (Knizhin *et al.*, 2006). We used a combination of gill netting, seining, electrofishing and angling to collect fish samples. When possible, at least eight individuals were caught per species and location. Pectoral fin clippings were obtained immediately after capture and preserved in 95% ethanol. Fish collections were conducted under an approved Institutional Animal Care and Use Committee (IACUC) protocol from Rutgers University (11-005) and permits from the Mongolian Ministry of Nature, Environment, and Tourism.

DNA extraction, ddRAD-seq library preparation and sequencing

We initially digested the tissue samples in 500 µL of cell lysis buffer, proteinase K and RNase A for 12 h at 56 °C. We then extracted genomic DNA from the digested tissue samples with the use of Serapure beads (Rohland & Reich, 2012) at a 1:1 volume. After extraction, we measured DNA concentration with a High Sensitivity Assay kit on a Qubit fluorometer (Life Technologies).

We obtained SNPs from across the genome for each sample with the use of ddRAD-seq (Peterson *et al.*, 2012). We digested DNA with the enzymes *Sbf*I and *Msp*I for 7 h at 37 °C with the 10× CutSmart buffer (Thermo Scientific). We then used Serapure beads to purify the DNA fragments before ligation of the barcoded Illumina adaptors (Table S2). We multiplexed all samples into a single sequencing lane with the use of

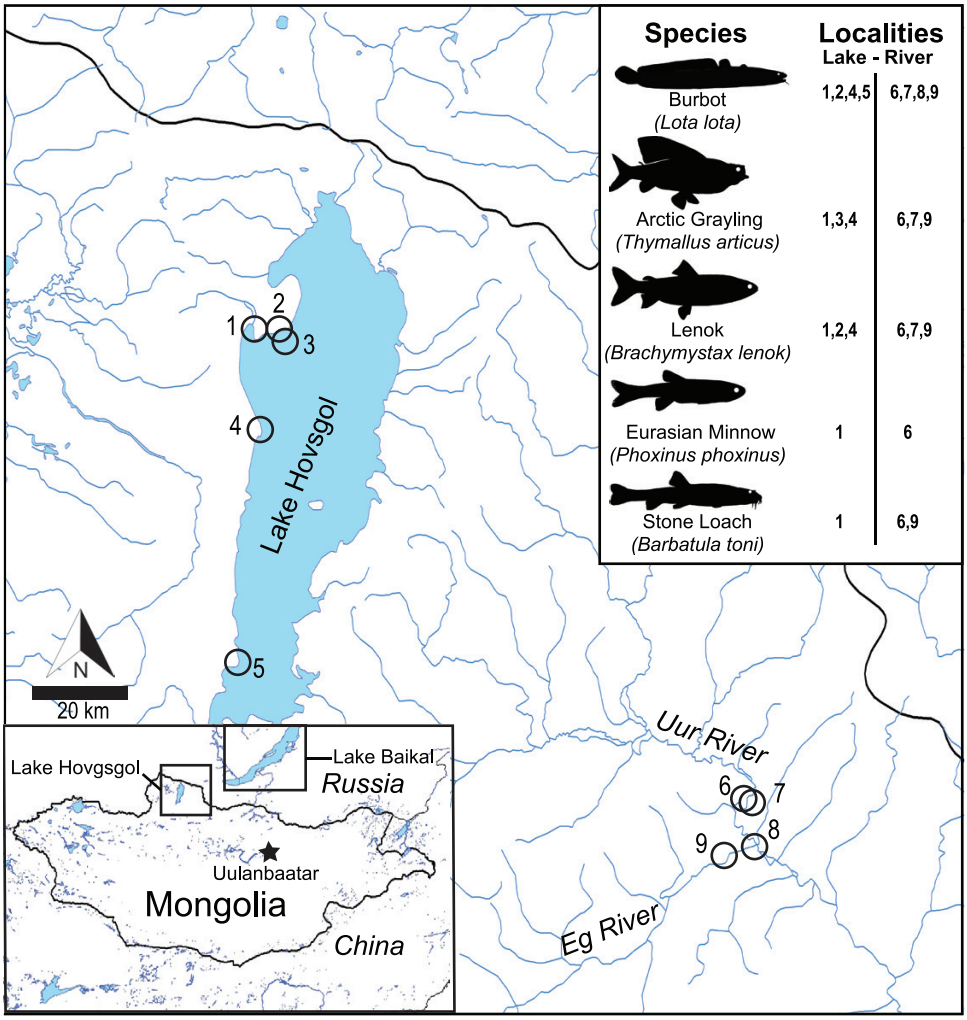

Figure 1. Map of collecting localities in Lake Hovsgol and rivers, with an inset showing the lake relative to Mongolia and Lake Baikal.

eight barcodes and 12 Illumina indexes. We pooled each column of barcoded samples and performed size selection on a Pippin Prep system (Sage Science), selecting for DNA fragments between 450 and 550 bp. We then amplified the pooled samples and attached Illumina's indexes by performing a PCR with Illumina's indexed primers. Before sequencing we used a Bioanalyzer 7500 high sensitivity DNA chip (Agilent) to determine the distribution of the fragment sizes amplified as well as their concentration. We pooled all libraries and sequenced them on an Illumina HiSeq 2500 with paired-end reads at the Genome Core Facility of the New York University Abu Dhabi, United Arab Emirates.

RAD-SEQ ASSEMBLY AND SNP CALLING

We assembled paired reads into contigs for each species using the Stacks (v1.40) pipeline (Catchen *et al.*, 2011). We removed reads with more than 15 consecutive nucleotides with low-quality calls (Phred score < 10). Paired reads were then concatenated and used for RAD-seq assembly, using the default setting. We then obtained VCF files using the *populations* module in Stacks, requiring no more than 70% missing data in each population. The resulting VCF files were further filtered with the use of VCFTOOLS (Danecek *et al.*, 2011) to only include SNPs with at least 60% of covered individuals and a minimum sequencing depth of 6× for each genotype. Prior to filtering with VCFTOOLS, we excluded individuals of two species (burbot and stone loach) with high levels of missing data (more than 80%). We then converted the resulting VCF files into STRUCTURE and Arlequin formats using PGDSpider (Lischer & Excoffier, 2012). All Illumina raw sequence data are available at the NCBI Sequence Read Archive accessioned under Bioproject ID RJNA476399 and SRA ID SRP150668.

Population structure

We estimated relatedness between samples using the network-based approach implemented in Splitstree4 (Huson & Bryant, 2006). This approach is based on reticulation and is therefore more suited to characterize structure and evolutionary relationships when gene flow or incomplete lineage sorting is high. We confirmed the existence of population structure with the software STRUCTURE (Pritchard *et al.*, 2000). This method clusters individuals into a predefined number of populations without any a priori assumption about assignment. Our goal was to confirm the existence of a detectable genetic differentiation between the two environments (river and lake) without any information about the origin of individuals. We therefore tested a maximum number of three clusters (K = 1–3). For each species we tested the likelihood for individuals to be assigned to K genetic clusters using the admixture model and correlated allele frequencies in STRUCTURE. We ran ten replicates for each value of K.

Descriptive statistics

Multiple factors can impact the sharing of alleles between populations, including time since divergence, changes in population size and migration rates. We therefore explored how genetic diversity was partitioned between the two sampled localities for each species. We computed measures of differentiation (F_{ST}), nucleotide diversity per segregating site, and the relative proportion of fixed and shared polymorphisms in Arlequin (Excoffier & Lischer, 2010). We then assessed significance for F_{ST} in Arlequin by comparing observed values to those obtained from 10 000 permutations. We also computed Tajima's D (Tajima, 1989) over all polymorphic sites in VCFTOOLS (Danecek *et al.*, 2011). Tajima's D is sensitive to changes in population size, with negative values as a result of a recent expansion, while positive values can be expected after a bottleneck in population size.

Coalescent simulations

Genetic structure between populations is impacted by gene flow and incomplete lineage sorting, both of which increase the number of shared alleles. Recent methodological improvements have made it possible to estimate at genome-wide scales the most likely combination of gene flow and demographic events that lead to a given allele frequency spectrum (Gutenkunst *et al.*, 2009; Wegmann *et al.*, 2010; Excoffier *et al.*, 2013). To further characterize the extent of isolation between river and lake populations, we estimated parameters for an isolation-with-migration model for each of the five species. We used the likelihood framework implemented in fastsimcoal2.5 (Excoffier & Foll, 2011; Excoffier *et al.*, 2013), which is based on the allele frequency spectrum (AFS). The method estimates the most likely set of parameters explaining an observed AFS by performing coalescent simulations. The model included six parameters, including the splitting time between populations, two migration rates, and effective population sizes for current and ancestral populations. Instead of considering only loci covered in all individuals, we projected the AFS down in each population to increase the number of segregating sites, with the use of a custom python script (available at https://github.com/isaacovercast/easySFS, accessed on 2 July 2018). We obtained parameters with the highest likelihood after 40 cycles of the algorithm, starting with 50 000 coalescent simulations per cycle, and ending with 100 000 simulations. We replicated this procedure 50 times, and retained the set of parameters with the highest final likelihood as the best point estimate. We estimated 95% confidence intervals (CIs) using a parametric bootstrap procedure, creating 100 pseudo-observed AFS using the set of parameters estimated from the actual dataset and repeating the estimation procedure on these datasets.

RESULTS

We performed ddRAD-seq on five species of fish from Lake Hovsgol and its tributary Eg and Uur rivers. The number of usable read pairs per individual ranged from 869 to 3 566 346 with an average of 1 117 824 (SD = 792 682). After filtering, the mean depth of coverage for the remaining SNPs was 44× per individual (SD = 20). The final dataset included between 497 and 7569 SNPs for each species, with a number of assembled polymorphic loci ranging from 277 to 5285 (Table S3).

Population structure, diversity and differentiation

Visual examination of the Splitstree networks (Fig. 2) clearly shows that individuals from the lake and from the river group according to their habitat. The networks also suggest that the level of differentiation between lake and river populations was higher for grayling than for the other species, which is consistent with the morphological differences previously reported (Knizhin *et al.*, 2006). STRUCTURE analyses assigned individuals to clusters that matched sampling location for four of the five species that we surveyed, with the exception of lenok (Fig. 3). For stone loach and Eurasian minnow, the favoured number of genetic clusters was 3, due to the high relatedness between two individuals. Removing one individual of the pair gave a clustering of two genetic clusters, consistent with the localities of origin. We confirmed the differences in genetic structure among species by calculating F_{ST} (Table 1). The

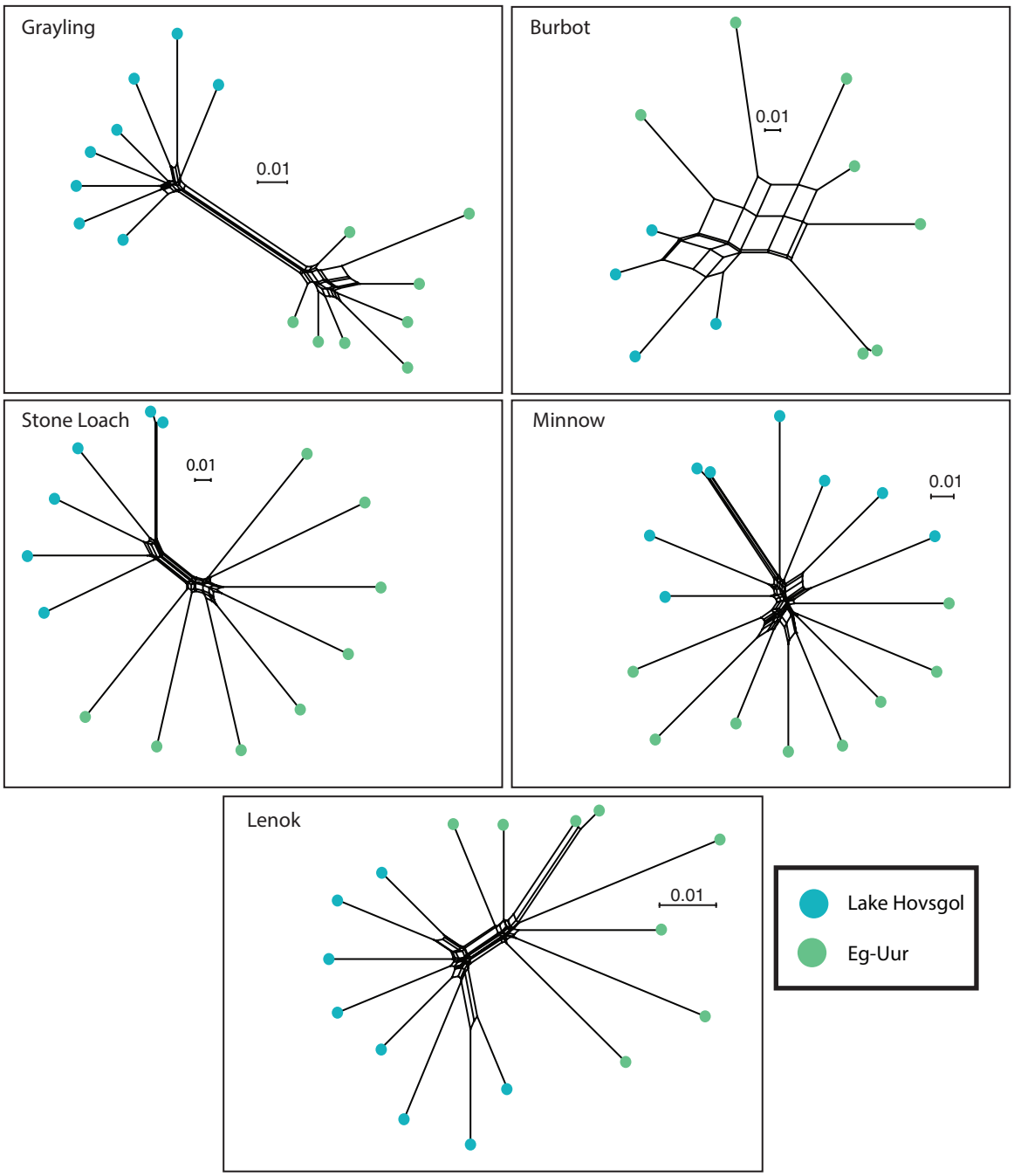

Figure 2. Network analysis obtained from Splitstree for the five species. The scale represents the average number of nucleotide differences per site.

highest level of differentiation was found in grayling, while genetic structure was less pronounced for burbot and stone loach, while the lowest, yet significant, level of differentiation was found in Eurasian minnow and the lenok. These analyses clearly demonstrate a substantial level of differentiation between lake and river for all five species, but reveal important differences among them.

Nucleotide diversity and number of private segregating sites for populations in Lake Hovsgol were lower than in populations of the Eg/Uur Rivers in four out of the five species (Table 1). This suggests a lower effective population size for those four species in the lake compared to the rivers. The only exception to this pattern was populations of grayling, which displayed a higher diversity in the lake than in the rivers. In

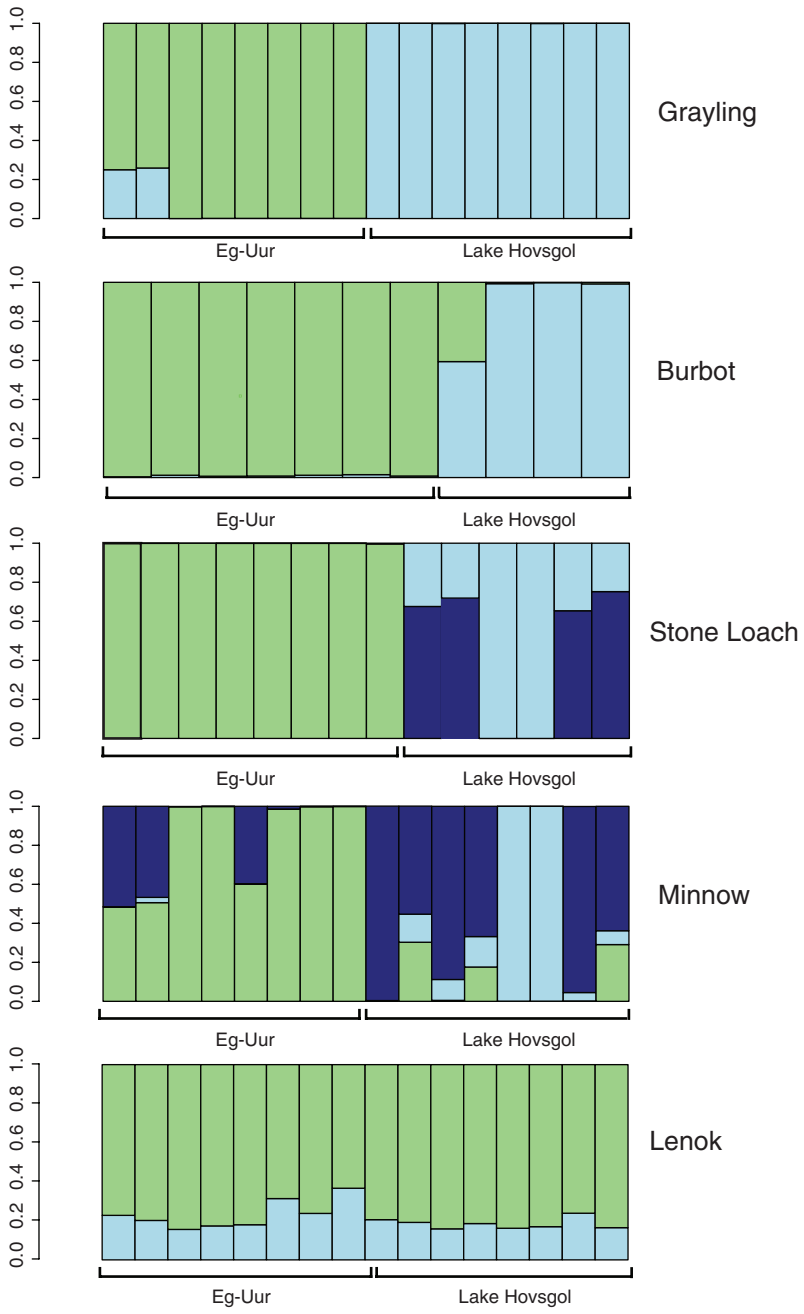

Figure 3. STRUCTURE analyses for the five species at $K = 2$ or 3. Vertical bars correspond to one individual and proportion of colour represents inferred probability of belonging to a cluster.

all species and populations Tajima's D was close to 0, which is consistent with an absence of strong bottlenecks or population expansion in recent times.

Gene flow and time since divergence

To assess whether the generally low genetic differentiation between lake and river populations across species was due to a recent split between lake and river or to an older split followed by high gene flow, we performed coalescence analyses using the likelihood framework implemented in fastsimcoal2.5. The results (Fig. 4) were consistent with the descriptive analyses above, with grayling, burbot and stone loach displaying low gene flow ($2N_em < 0.5$) between the rivers and lake populations. We observed the opposite pattern for lenok and Eurasian minnow, which displayed a low F_{ST}. For all except one species (grayling), effective

Table 1. Summary statistics for the two populations and the five species studied

	Grayling	Burbot	Stone loach	Minnow	Lenok
F_{ST}	0.192***	0.130***	0.123***	0.057***	0.029***
Nucleotide diversity (H)	0.133	0.096	0.154	0.112	0.118
Nucleotide diversity (EU)	0.090	0.180	0.167	0.119	0.142
Private polymorphism (H)	2780	61	658	1408	1950
Private polymorphism (EU)	1338	293	1856	1955	2914
Shared polymorphisms	1410	134	1817	1529	2631
Tajima's *D* (H)	−0.550	0.627	0.753	0.068	−0.074
Tajima's *D* (EU)	0.444	0.548	0.444	−0.212	−0.412

Significance levels were assessed through 10 000 permutations; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. H, Lake Hovsgol; EU, Eg/Uur rivers.

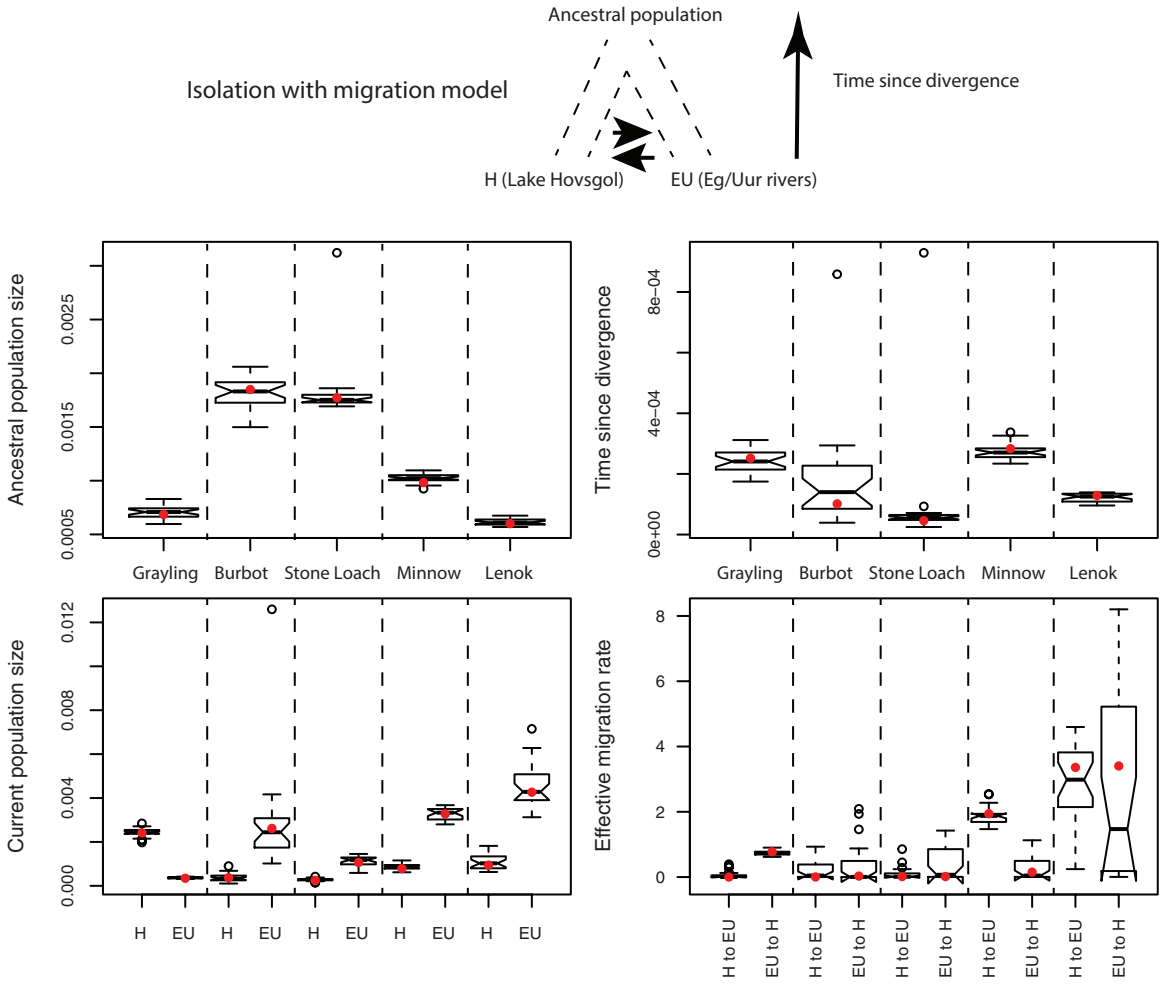

Figure 4. Parametric bootstrap replicates for parameters estimated under an isolation-with-migration model for the five species. Point estimates obtained from the observed dataset are indicated by a red dot. Parameters are scaled by mutation rate: population sizes are in units of $4N_e\mu$, with N_e the diploid population size and μ the mutation rate per nucleotide per generation. Effective migration rates are given in number of gene copies exchanged per generation ($2N_e m$, with m the migration rate per generation). H, Lake Hovsgol; EU, Eg/Uur rivers.

population sizes were lower in the lake than in the rivers, as suggested by their lower nucleotide diversity and number of private alleles (Table 1).

To determine whether divergence times between populations were consistent with a re-colonization following the LGM, we used previously estimated mutation

Table 2. Time since divergence estimated in years; mutation rates were obtained from previous studies on related species from the same family

Species	Time estimate (calibrated)	Lower bound (95% CI)	Upper bound (95% CI)	Mutation rate (per year)	Generation time (years)	Time (years)	Lower bound (95% CI)	Upper bound (95% CI)	References (mutation rate)	References (generation time and fecundity)
Grayling	0.000251994	0.000191812	0.000303956	1.21E-09	4.5	46191.55	35159.98	55716.33	(Crête-Lafrenière *et al.*, 2012)	(Dulmaa 1999; Sideleva *et al.*, 2006; Tsogtsaikhan *et al.*, 2017)
Burbot	0.000101282	4.05872E-05	0.00028312	5.00E-09	6.4	3165.06	1268.35	8847.49	(Malmstrøm *et al.*, 2016)	(Podlesnyi, 1958; Dulmaa, 1999)
Stone loach	0.000045757	3.24508E-05	7.05E-05	3.51E-09	1	13036.18	9245.24	20075.44	(Xu *et al.*, 2014)	(Breder & Rosen, 1966; Barluenga & Meyer, 2005)
Minnow	0.00028348	0.000236097	0.00033068	3.51E-09	2.6	31062.90	25870.85	36234.93	(Xu *et al.*, 2014)	(Mills & Eloranta, 1985; Sideleva *et al.*, 2006; Kottelat & Freyhof, 2007)
Lenok	0.000128681	9.97352E-05	0.000139049	1.21E-09	8	13268.12	10283.56	14337.19	(Crête-Lafrenière *et al.*, 2012)	(Sideleva *et al.*, 2006; Tsogtsaikhan *et al.*, 2017)

rates and generation times to translate parameters into demographic units (Table 2). These are probably overestimates because they are based either on divergence at slowly evolving exonic markers (Xu *et al.*, 2014; Malmstrom *et al.*, 2016) or on deep phylogenies for which mutation rate estimates tend to be biased downwards (Crete-Lafreniere *et al.*, 2012). Nevertheless, point estimates of divergence times differ considerably among species, and the absence of overlap in confidence intervals suggests that fish species have colonized the lake at different times. For three species (stone loach, burbot and lenok), the divergence time estimates between river and lake are consistent with colonization of the lake after the LGM. However, for the European minnow and the grayling, the divergence time pre-dates the LGM by approximately 15–30 kyr, and the confidence intervals do not overlap with the LGM.

DISCUSSION

Genetic analyses of five of the ten native fishes of Lake Hovsgol provide insight into the history of post-Pleistocene colonization of the lake. Our results can be summarized as follows: (1) there is significant genetic differentiation between lake and river populations in all species; (2) colonization of burbot and stone loach in the lake probably occurred after the LGM (3.1 kyr for burbot and 13 kyr for stone loach) and was followed by low gene flow between river and lake populations; (3) in lenok, the colonization also occurred after the LGM (13 kyr), but high levels of gene flow between river and lake populations resulted in the lowest level of genetic differentiation among the species analysed; and (4) in grayling and possibly minnow, the differentiation between lake and river is not consistent with a post-glacial colonization of the lake from the Eg/Uur rivers, as the coalescent estimates are older than the LGM (31 and 46 kyr, respectively). Altogether, our results suggest two main differences among fish species in Lake Hovsgol: in the time of divergence between lake and river populations and in the amount of gene flow between populations from the two habitats.

HETEROGENEOUS DIVERGENCE AMONG SPECIES FROM THE SAME FISH COMMUNITY

Our results suggest two pulses of differentiation: one that is consistent with a colonization of the lake after the LGM (in burbot, stone loach and lenok) and one that pre-dates the LGM by 15–30 kyr (in minnow and grayling). Assuming our mutation rates and generation times are accurate, two distinct models, a two-steps model and a two-sources model, can explain this pattern. In the two-steps model, two waves of colonization of Lake Hovsgol occurred, a first one more than 30 kya and a second that began ~13 kya. This model suggests that

the species that first colonized the lake (minnow and grayling) remained isolated in the lake during the LGM in refugia. This model is identical to the one proposed to explain the persistence of a high level of endemism in Lake Baikal despite the instability of the lake habitat during the glacial and inter-glacial episodes of the Pleistocene (Kamaltynov, 1999; Karabanov *et al.*, 2001). However, the ecosystem of Lake Hovsgol seems to have been much more impacted by glacial episodes than that of Lake Baikal (Karabanov *et al.*, 2004). It is believed that the lacustrine habitat of Lake Hovsgol collapsed completely during the LGM, as suggested by analysis of sedimentary sections. The persistence of refugia in the lake does not seem likely, but additional analyses may be necessary to confirm the complete collapse of the lake habitat proposed by Karabanov *et al.* (2004).

The two-sources model posits that the lake was colonized after the LGM but from two distinct sources. The recent divergence time between lake and river populations of burbot, lenok and stone loach is consistent with a colonization of the lake from the Eg/Uur Rivers after the LGM. It is interesting to note that the colonization of the lake by two species (the loach and lenok) seems to be concomitant with the development of planktonic communities and the increase in primary production ~13 kya (Karabanov *et al.*, 2004). We propose that the colonization of the lake by the minnow and the grayling occurred from other source populations than those in the Eg/Uur Rivers. Under this scenario, we hypothesize the existence of minnow and grayling populations that were isolated from the Eg/Uur populations for 30 to 45 kyr. Although the Eg/Uur watershed is one of the major feeders of the lake, other rivers, as well as small neighbouring lakes, could have contributed to the colonization of the lake 13 kya. This hypothesis will need to be tested on a geographically wider sampling of the fish species analysed here.

Life history traits and gene flow

The different fish species studied exhibit striking differences in the level of gene flow between lake and river. The minnow and lenok show a high level of gene flow while the other three species seem to exchange very few migrants. We propose that these differences are related to the ecology of the species and in particular to the distribution of species at the level of microhabitats (Dieckmann *et al.*, 1999). Generalist species that are distributed continuously in the river should exhibit a higher gene flow between river and lake, compared to specialist species restricted to some microhabitats in the river. Much of the Eg River is pool and riffle habitat and the five fish species are not dispersed similarly along the length of the river. For instance, burbot is restricted to deep, slow moving pools in the river and hence has a punctuated distribution (Froese & Pauly, 2018). Not surprisingly, the estimated level of gene flow for this species is relatively low (Fig. 4). Similarly, the stone loach, which inhabits shallow riverbeds where the fish can partially burrow in the gravel, may have a discontinuous distribution, thus limiting their dispersal and subsequent gene flow. In contrast, the lenok prefers deeper, fast flowing water, typically at the heads or tailouts of pools, which could facilitate gene flow. The Eurasian minnow has a continuous distribution along the rivers, which is consistent with the high gene flow detected here. This species is an ecological generalist, living in pools and shallow eddies along faster flowing waters.

The grayling seems to contradict this model as it shares the same habitat preference as the lenok in the river, yet exhibits limited levels of gene flow. The grayling was the only one to show higher genetic diversity in Lake Hovsgol than in the rivers. This is consistent with a relatively larger effective population size in the lake, the opposite of what was seen in the other species. Moreover, grayling have shown differentiation within the lake, evolving a pelagic form with high gill raker counts (used for suspension feeding), and a more omnivorous littoral form with fewer gill rakers (K. Olson *et al.*, in preparation). Additionally, effective migration rates for grayling between the lake and the river were asymmetrical, with higher river-to-lake than lake-to-river migration, as might be expected given life history specialization in the lake. This finding is supported by observations of the riverine subspecies (*T. arcticus baicalensis*) captured within Lake Hovsgol near its outlet, but no observations of the lake subspecies (*T. arcticus nigrescens*) captured within the river. This suggests the possibility that a river-dwelling grayling would experience a similar trophic niche in the littoral zone of the lake as to that in the river, while pelagic grayling of the lake would be at a disadvantage when migrating into the river. The lake may therefore have acted as a sink for gene flow and possibly provides better conditions to sustain a large population, facilitating the accumulation of genetic diversity compared to nearby rivers. Our results also suggest incipient niche specialization in the lake that may be facilitated by low gene flow. The high estimate for time since divergence for the grayling may indicate that differentiation took place before the recolonization of the lake, and calls for more detailed studies at the scale of the range of this species to identify potential refugia and sources for this population. In that case, river and lake populations would be in secondary contact, making this species an interesting model to assess how ecology and reproductive isolation facilitate speciation.

Our genetic comparison is limited by design to species that occur in both Lake Hovsgol and its outlet. An interesting related ecological question not amenable to genetic analysis is why some fish species found in the river have not colonized the lake, given the success of other species. For instance, northern pike (*Esox lucius*) is found throughout the Eg and Selenga Rivers and also

in Lake Baikal, but not in Hovsgol (Sideleva, 2006). This reinforces the importance of the ecological suitability of the potentially colonizable waters for individual species, a factor that probably shifts with environmental conditions across millennia. In the northern pike, it may be that the food web of this ultra-oligotrophic lake is presently insufficient to support such a large apex predator.

CONCLUSIONS

Lake Hovsgol, although geologically old, is a newly available habitat whose fish communities have not yet reached equilibrium. Our results show that a community of species that differ in their ability to migrate and diversify colonized this new 'island' in recent times. By contrasting the demographic history of the five fish species, we identified possible historical and ecological factors affecting the level of genetic differentiation between lake and river populations. Both our results and previous morphological studies suggest that, at least for graylings, isolation-by-ecology may have already taken place. Future genome-wide studies should focus on additional localities at a broader spatial scale to clearly assess the origin of colonization and quantify population structure along neighbouring rivers and lakes. Ultimately, this should lead to the quantification of biotic and abiotic mechanisms that drive local adaptation.

ACKNOWLEDGEMENTS

This project was funded by a National Science Foundation grant (NSF OISE: 1064843) to O. Jensen and by New York Abu Dhabi Research Funds (AD180) to S. Boissinot. We are grateful to our Mongolian-American research team for assistance in the field, including M. Amaraa, Boidkhand, Boloroo, H. Chantuu, L. Chuluunchimeg, B. Ganzorig, T. Hrabik, M. Hickman, G. Jargalmaa, M. Kessler, B. Mendsaikhan, N. McAuliffe, G. Natsag, U. Nemekhbayar, K. Olson, M. Provost and T. Zimmerman. We thank Marc Arnoux from the Genome Core Facility at NYUAD for sequencing RAD libraries. NYUAD Sequencing and Bioinformatics Cores are supported by NYUAD Research Institute grant G1205-1205A to the Center for Genomics and Systems Biology at NYUAD. We thank two anonymous reviewers for their comments on an earlier version of the manuscript.

REFERENCES

Ahrenstorff TD, Jensen OP, Weidel BC, Mendsaikhan B, Hrabik TR. 2012. Abundance, spatial distribution, and diet of endangered Hovsgol grayling (*Thymallus nigrescens*). *Environmental Biology of Fishes* **94:** 465–476.

Barluenga M, Meyer A. 2005. Old fish in a young lake: stone loach (Pisces: *Barbatula barbatula*) populations in Lake Constance are genetically isolated by distance. *Molecular Ecology* **14:** 1229–1239.

Breder CM, Rosen DE. 1966. *Modes of reproduction in fishes*. Neptune City: T.F.H. Publications.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes Genomes Genetics* **1:** 171–182.

Crête-Lafrenière A, Weir LK, Bernatchez L. 2012. Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. *PLoS One* **7:** e46662.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G. 2011. The variant call format and VCFtools. *Bioinformatics* **27:** 2156–2158.

Dieckmann U, O'Hara B, Weisser W. 1999. The evolutionary ecology of dispersal. *Trends in Ecology and Evolution* **14:** 88–90.

Dulmaa A. 1999. Fish and fisheries in Mongolia. In: Petr T, ed. *Fish and fisheries at higher altitudes: Asia*. Rome: FAO Fisheries Technical Paper 385.

Elmer KR, Lehtonen TK, Kautt AF, Harrod C, Meyer A. 2010. Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. *BMC Biology* **8:** 60.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genetics* **9:** e1003905.

Excoffier L, Foll M. 2011. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27:** 1332–1334.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564–567.

Fedotov A, De Batist M, Shapron E, De Reiker K, Paule T, Grachev M. 2002. Seismic survey of sediments of Lake Hovsgol. *Doklady of the Russian Academy of Sciences* **382:** 261–263.

Froese R, Pauly D. 2018. *FishBase*. Available at: http://www.fishbase.org

Goulden CE, Sitnikova T, Gelhaus J, Boldgiv B. 2006. *The geology, biodiversity and ecology of Lake Hövsgöl (Mongolia)*. Leiden: Backhuys Publishers BV.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* **5:** e1000695.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23:** 254–267.

Kamaltynov RM. 1999. On the evolution of Lake Baikal amphipods. *Crustaceana* **72:** 921–931.

Karabanov E, Williams D, Kuzmin M, Sideleva V, Khursevich G, Prokopenko A, Solotchina E, Tkachenko

L, Fedenya S, Kerber E, Gvozdkov A, Khlustov O, Bezrukova E, Letunova P, Krapivina S. 2004. Ecological collapse of Lake Baikal and Lake Hovsgol ecosystems during the Last Glacial and consequences for aquatic species diversity. *Palaeogeography, Palaeoclimatology, Palaeoecology* **209:** 227–243.

Karabanov EB, Williams DF, Sideleva VG, Prokopenko AA, Khursevich GK, Kuzmin MI, Gvozdkov AN, Soshina NM, Afonina TE. 2001. Glacial ecological catastrophes of Lake Baikal (Siberia) and Lake Hovsgol (Mongolia) during the Pleistocene and Late Pliocene. In: Jacoby GJ, Pederson N, eds. *Proceedings of the conference on Mongolian paleoclimatology and environmental research*. Palisades: Lamont-Doherty Earth Observatory, 21–32.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecology Letters* **7:** 1225–1241.

Keller I, Wagner CE, Greuter L, Mwaiko S, Selz OM, Sivasundar A, Wittwer S, Seehausen O. 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology* **22:** 2848–2863.

Knizhin IB, Antonov AL, Weiss SJ. 2006. A new subspecies of the Amur grayling *Thymallus grubii flavomaculatus* ssp. *nova* (Thymallidae). *Journal of Ichthyology* **46:** 555–562.

Kottelat M, Freyhof J. 2007. *Handbook of European fishes*. Berlin: Kottelat, Cornol, Switzerland and Freyhof.

Lischer HE, Excoffier L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28:** 298–299.

Malmstrøm M, Matschiner M, Tørresen OK, Star B, Snipen LG, Hansen TF, Baalsrud HT, Nederbragt AJ, Hanel R, Salzburger W, Stenseth NC, Jakobsen KS, Jentoft S. 2016. Evolution of the immune system influences speciation rates in teleost fishes. *Nature Genetics* **48:** 1204–1210.

Mills CA, Eloranta A. 1985. The biology of *Phoxinus phoxinus* (L.) and other littoral fishes in Lake Konnevesi, Central Finland. *Annales Zoologici Fennici* **22:** 1–12.

Moore MV, Hampton SE, Izmest'eva LR, Silow EA, Peshkova EV, Pavlov BK. 2009. Climate change and the world's "Sacred Sea" - Lake Baikal, Siberia. *Bioscience* **59:** 405–417.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7:** e37135.

Podlesnyi AV. 1958. Fishes of Enisey, their environments and use. In: Pravdin IF, Pirozhnikov PL, eds. *Commercial fishes of Ob and Enisey and their use*. Moscow: Pishchepromizdat: Izv. VNIORH 44.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Rohland N, Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* **22:** 939–946.

Sideleva V. 2006. Fish fauna of Lake Hovsgol and Selenga River in comparison with ichthyofauna of Lake Baikal. In: Goulden CE, Sitnikova T, Gelhaus J, Boldgiv B, eds. *The geology, biodiversity and ecology of Lake Hövsgöl (Mongolia)*. Leiden: Backhuys Publishers BV, 357–378.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:** 585–595.

Tsogtsaikhan P, Mendsaikhan B, Jargalmaa G, Ganzorig B, Weidel BC, Filosa C, Free C, Young T, Jensen OP. 2017. Age and growth comparisons of Hovsgol grayling (*Thymallus nigrescens* Dorogostaisky, 1923), Baikal grayling (*T. baicalensis* Dybowski, 1874), and lenok (*Brachymystax lenok* Pallas, 1773) in lentic and lotic habitats of Northern Mongolia. *Journal of Applied Ichthyology* **33:** 108–115.

Vellend M. 2005. Species diversity and genetic diversity: parallel processes and correlated patterns. *The American Naturalist* **166:** 199–215.

Warren BH, Simberloff D, Ricklefs RE, Aguilée R, Condamine FL, Gravel D, Morlon H, Mouquet N, Rosindell J, Casquet J, Conti E, Cornuault J, Fernández-Palacios JM, Hengl T, Norder SJ, Rijsdijk KF, Sanmartín I, Strasberg D, Triantis KA, Valente LM, Whittaker RJ, Gillespie RG, Emerson BC, Thébaud C. 2015. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecology Letters* **18:** 200–217.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L. 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* **11:** 116.

Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, Xu J, Zheng X, Ren L, Wang G, Zhang Y, Huo L, Zhao Z, Cao D, Lu C, Li C, Zhou Y, Liu Z, Fan Z, Shan G, Li X, Wu S, Song L, Hou G, Jiang Y, Jeney Z, Yu D, Wang L, Shao C, Song L, Sun J, Ji P, Wang J, Li Q, Xu L, Sun F, Feng J, Wang C, Wang S, Wang B, Li Y, Zhu Y, Xue W, Zhao L, Wang J, Gu Y, Lv W, Wu K, Xiao J, Wu J, Zhang Z, Yu J, Sun X. 2014. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nature Genetics* **46:** 1212–1219.

Young T, Jensen OP, Weidel BC, Chandra S. 2015. Natural trophic variability in a large, oligotrophic near-pristine lake. *Journal of Great Lakes Research* **41:** 463–472.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Table S1. Origin of samples used in this study.
Table S2. Barcodes and adaptors used in this study.
Table S3. Number of loci and SNPs recovered in this study.