Mitochondrial and nuclear DNA phylogeography of *Thymallus* spp. (grayling) provides evidence of ice-age mediated environmental perturbations in the world's oldest body of fresh water, Lake Baikal

MIKKO T. KOSKINEN,* IGOR KNIZHIN,† CRAIG R. PRIMMER,* CHRISTIAN SCHLÖTTERER‡ and STEVEN WEISS§¶

*Department of Ecology and Systematics, Division of Population Biology, PO Box 65, FIN-00014 University of Helsinki, Finland, †Irkutsk State University, Department of Vertebrate Zoology, 664011, Irkutsk, Sukhe-Bator 5, Russia, ‡Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Veterinärplatz 1, 1210 Vienna, Austria, §Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), Campus Agrário de Vairão, R.Monte-Crasto, 4485–661, Vairão, Portugal

Abstract

Theories on the hydrological history of Lake Baikal, the world's oldest and deepest body of freshwater, and its surrounding great rivers, are currently based solely on geological evidence and are conflicting. Baikal is inhabited by numerous zoogeographical enigmas but their high level of endemism has hindered phylogeographic inferences. We provide a biological perspective of the region's palaeo-hydrological development based on the demographic and genealogical history of the widespread *Thymallus* spp. (grayling). Phylogenetic reconstruction reveals that old lineages of grayling (pre-Pleistocene) currently inhabit the Enisey, Lena and Amur River basins. For Lake Baikal however, we conclude that a mid-Pleistocene colonization (110 000–450 000 years ago) of an unoccupied niche has occurred. Population genetic inferences support an Enisey–Angara river route of colonization into Baikal, corresponding to the cataclysmic palaeo-hydrological event that led to the formation of the lake's only contemporary outlet, and a subsequent range expansion several thousand kilometres into the uppermost reaches of the Selenga River basin. The evolutionary history of Lake Baikal grayling is congruent with the controversial hypothesis of repeated glaciation. However, considering the extraordinary levels of endemism and proposed Miocene or Oligocene coalescence of other faunal lineages, a less profound but equally consequential cycle of environmental perturbations may have taken place. Bi-parentally inherited microsatellite DNA loci supported the phylogenetic relationships of *Thymallus* spp. and the geographical expansion of Baikal grayling strikingly well. A Markov Chain Monte Carlo modelling approach suggested severe contemporary population decline during the last century, possibly reflecting the influence of an uncontrolled fishery on this treasured ecosystem. These complementary pictures of the demographic history of grayling underscore the breadth of historical inquiry that can be entertained through the modelling of sufficient molecular data, and may significantly alter the zoogeographical and limnological perspectives of Baikal's history.

Keywords: glaciation, Lake Baikal, palaeohydrology, phylogeography, *Thymallus*

Received 9 May 2002; revision received 19 August 2002; accepted 19 August 2002

Introduction

Lake Baikal, the world's oldest and deepest body of fresh water, lies near the headwaters of three great rivers, the

Correspondence: Steven Weiss. ¶Present address: Karl-Franzens Universität Graz, Institut für Zoologie, Universitätsplatz 2, A-8010 Graz, Austria. E-mail: steven_weiss@uni-graz.at

Enisey, Lena and Amur. Each of these rivers is thought to have been connected to Baikal in the past, providing colonization corridors for the ancestors of such zoogeographical enigmas as freshwater seals (*Phoca sibirica*), endemic families of freshwater sponges (Lubomirskiidae) and caddisflies (Limnephilidae), several endemic genera of leeches, and a monotypic family of turbellarians (Baicalarctidae)

(Bânârescu 1991). However, zoogeographers have thus far been unable to use such taxa for reconstructing the region's palaeo-hydrological development. Either these taxa are too unique, with no close living ancestors, or, geographically comprehensive sampling combined with molecular approaches has not been attempted. Thus, current theories on this ancient ecosystem's history are based primarily on geological evidence and are in conflict (Grosswald 1998; Kuz'min 2001). While some state that Lake Baikal has never been glaciated (Kuz'min 2001), others speculate on cyclic patterns of glaciation paralleling Quaternary fluctuations elsewhere in Eurasia (Grosswald 1998).

In contrast to a rich invertebrate community, Baikal's ichthyofauna is not diverse, aside from a species flock of mostly benthic cottoid fish, of which the most recent common ancestor existed as recently as 2.5 million years ago (Ma) (Kiril'chik & Slobodyanyuk 1997; Sherbakov *et al.* 1998)—a puzzling suggestion for a lake that is at least 25 million years old (Bânârescu 1991). However, the salmonid genus *Thymallus* (grayling) may be more ideal for reconstructing hydrological history as it occurs throughout the region's river systems, as well as in isolated lakes of Western Mongolia. The genus is represented by four species: European grayling, *T. thymallus*; Mongolian grayling, *T. brevirostris*; the Amur grayling, *T. grubei*; and Arctic grayling, *T. arcticus*; characterized by numerous subspecific designations throughout its range.

Cataclysmic events involving Lake Baikal and its surrounding regions are expected to leave an imprint on the genetic architecture of organisms (Hewitt 2000). The widespread populations of grayling should harbour a rich genetic impression reflecting palaeo-change and disturbance and the phylogeographic patterns of *Thymallus* may shed further light on Baikal's history. A rudimentary understanding of the lake's environmental fluctuations is paramount to the critical study of any of Baikal's evolutionary radiations. Thus, we attempt to shed light on both the pattern and dynamics of Lake Baikal's more recent physical past through a comprehensive phylogeographic analysis of one widespread genus of fish (*Thymallus* spp.). In addition to the genealogical reconstruction, emphasis is placed on evaluating historical demography to gain an objective reflection of the region's past perturbations. To our knowledge this is the first study to provide a biological perspective, based on molecular genetic techniques, of the region's palaeohydrological development.

Materials and methods

Mitochondrial DNA sequencing

Grayling (*n* = 142) were collected by electrofishing, angling, and gill nets in 1998 and 1999 from 17 Asian populations and six locations in Europe and North America (Table 1,

Fig. 1a). The complete mitochondrial DNA (mtDNA) control region (≈ 937 base pairs) and 228 base pairs of flanking tRNA gene regions were sequenced using standard procedures with an ABI-377 or ABI-310 automated sequencer as described in Uiblein *et al.* (2001) and Weiss *et al.* (2002). The mtDNA sequences of this study have been deposited into GenBank under accession numbers AY168348–168405.

Phylogenetic reconstruction

A sequence evolution model was chosen using a nested series of likelihood ratio tests (Huelsenbeck & Crandall 1997) in the program MODELTEST (Posada & Crandall 1998) incorporated into PAUP* 4.0b3a (Swofford 2000). Using the most likely model for the data, a neighbour-joining tree was constructed to obtain clusters of well-supported (97–100%) terminal taxa (haplotypes). These clusters were then constrained to allow maximum likelihood estimation of the major node topology without wasting computational effort on minor rearrangements of terminal taxa for which little confidence can be placed. Support for these nodes was calculated with bootstrap replicates using maximum likelihood (100), maximum parsimony (1000) and neighbour-joining (1000) search methodologies.

Nested clade analysis

Because intraspecific, dichotomously branching haplotype trees do not necessarily reflect population history and may indeed be misleading for shallow genealogies (Goldstein *et al.* 2000; Posada & Crandall 2001), we also evaluated the haplotype genealogy of sample sites from Lake Baikal and its Enisey basin outflow with an unrooted network using a statistical parsimony criterion (Templeton *et al.* 1992) in the program TCS 1.13 (Clemens *et al.* 2000). This network was used to apply nested clade analysis (Templeton 1998), a standard tool for phylogeographic analyses (Templeton 2002). Nested clade analysis, as implemented in the program GEODIS 2.0 (Posada *et al.* 2000), provides a model-free statistical framework to test for phylogeographic signal. When the null-hypothesis of no correlation between genealogy and geography is rejected, biological inferences are drawn using *a priori* criteria thus preventing *ad hoc* or subjective interpretations. Although not yet subjected to extensive computer simulation studies, nested clade analysis is nevertheless considered a conservative technique and not prone to false-positive conclusions, i.e. type I error (Templeton 1998; Templeton 2002).

After transforming the network into a nested series of clades (Templeton *et al.* 1987; Templeton & Sing 1993), association between these clades and geographical distances was tested with a permutational contingency analysis using the program GEODIS 2.0. Inferences on the population-level processes resulting in the observed pattern

Table 1 Sample locations including major river drainage, basin, geographical coordinates and the number of individuals screened for both mtDNA and microsatellite (micro.) variation

Drainage	Population	Drain. code	Pop. code	Species	Lat.(N)	Long.(E)	*n* (mtDNA)	*n* (micro.)
Lake Baikal (North)	Dagary Bay	Bai	DBB	*T. arcticus*	55°41′	109°53′	8	9
Lake Baikal (North)	Frolikha Bay	Bai	FBB	*T. arcticus*	55°31′	109°51′	7	19
Lake Baikal (West)	Zavarotnaya Bay	Bai	ZBB	*T. arcticus*	54°18′	108°29′	10	33
Lake Baikal (Central)	Ushkaniy Islands	Bai	UIB	*T. arcticus*	53°40′	108°37′	16	10
Lake Baikal (South)	Selenga River Delta	Bai	SRD	*T. arcticus*	52°08′	106°34′	2	—
Snezhnaya River–Lake Baikal	Sobolinoye Lake	Bai	SOB	*T. arcticus*	51°17′	104°41′	9	29
Selenga River–Lake Baikal	South Fork Terkhyn	Bai	SFT	*T. arcticus*	47°45′	99°20′	10	34
Selenga River–Lake Baikal	Lake Chovsgul	Bai	CHV	*T. arcticus*	51°27′	100°39′	4	34
Irkutsk Reservoir–Angara-Enisey	Taltzinka River	Eni	IRT	*T. arcticus*	51°59′	104°36′	10	31
Enisey River	Nizhnyaya Tunguska River	Eni	NTR	*T. arcticus*	60°55′	107°56′	8	22
Shilka–Amur River	Onon River	Amu	ONG	*T. grubei*	48°75′	110°25′	10	38
Olengui–Ingoda-Amur River	Sypchergurka River	Amu	SYP	*T. grubei*	51°20′	113°26′	8	—
Central Asia	Khökh Nuur	Cas	KN	*T. brevirostris*	47°31′	98°27′	7	28
Kirenga River–Lena	Kutima River	Len	KTR	*T. arcticus*	57°02′	108°36′	2	—
Kuanda River–Lena	Leprindokan Lake	Len	LLL	*T. arcticus*	56°33′	117°31′	8	—
Olekma River–Lena	Lesha Lake	Len	LLO	*T. arcticus*	57°09′	119°37′	15	—
Enisey River	Shishkhid River	Eni	SHG	*T. arcticus*	51°30′	98°30′	2	—
Pacific, Alaska	Bristol Bay	Pac	Pac	*T. arcticus*	59°00′	155°00′	1	—
Blue River; British Columbia	Frank Lake	Pac	Pac	*T. arcticus*	59°20′	130°00′	1	—
Pacific, Yukon Territory	Pierce Lake	Pac	Pac	*T. arcticus*	—	—	1	—
Pacific, British Columbia	Burnt River	Pac	Pac	*T. arcticus*	122°20′	55°25′	—	11
Poe River–Adriatic, Italy	Gar Maggia	Ad	Ad	*T. thymallus*	46°15′	08°44′	1	—
Danube River–Black Sea, Austria	Füschler Ache	Da	Da	*T. thymallus*	47°50′	13°20′	1	—
Loire River–Atlantic, France	Loire River	At1	At1	*T. thymallus*	44°51′	03°55′	1	—
Lake Saimaa, eastern Finland	Lake Saimaa	Esa	Esa	*T. thymallus*	61°20′	28°20′	—	48
Total							142	346

Drain. code, drainage code; Pop. code, population code; Lat., latitude; Long., longitude; micro., microsatellites.

of clade dispersal followed the February 2001 update of an inference key, available at http://bioag.byu/zoology/crandall_lab/geodis.htm.

Pairwise mismatch analysis

We investigated the demographic signature of Lake Baikal and Enisey outflow mtDNA haplotypes based on coalescence theory. We used a pairwise mismatch distribution to test for population expansion (Rogers & Harpending 1992). The parameter of demographic expansion (τ) was estimated with a generalized nonlinear least squares approach and approximate confidence intervals were obtained with 1000 parametric bootstrap replicates in the program ARLEQUIN (Schneider *et al.* 2000). The goodness-of-fit of the observed data to a simulated model of expansion was tested with the sum of squared deviations and Harpending's raggedness index (Harpending 1994), as implemented in ARLEQUIN. The age of expansion was estimated with the formula $\tau = 2\mu t$, where μ equals the aggregate mutation rate across all nucleotides per generation (set at 5 years; Tugarina

1981) and t is the expansion time in generations. Estimates were plotted across a range of mutation rates reported for salmonid mtDNA. For graphical representation, the mismatch distribution was also calculated with Roger's method of moments (Rogers 1995), using the program DNASP version 3.51 (Rozas & Rozas 1999). A Kolmogorov–Smirnov two-sample test was used to test the distribution of observed values against the null hypothesis of a stable population.

Microsatellite analysis

A total of 346 individuals were sampled from 11 populations of *Thymallus* in Asia, one from Europe, and one from North America (Table 1, Fig. 1a). Individuals from the sampling sites DBB, FBB, UIB and ZBB (Table 1; Fig. 1a) were pooled to form the within-Baikal population (Bai) for all further analyses. Four randomly selected individuals from each population were screened for 17 microsatellite loci using previously described genotyping procedures (Koskinen & Primmer 2001). The following loci
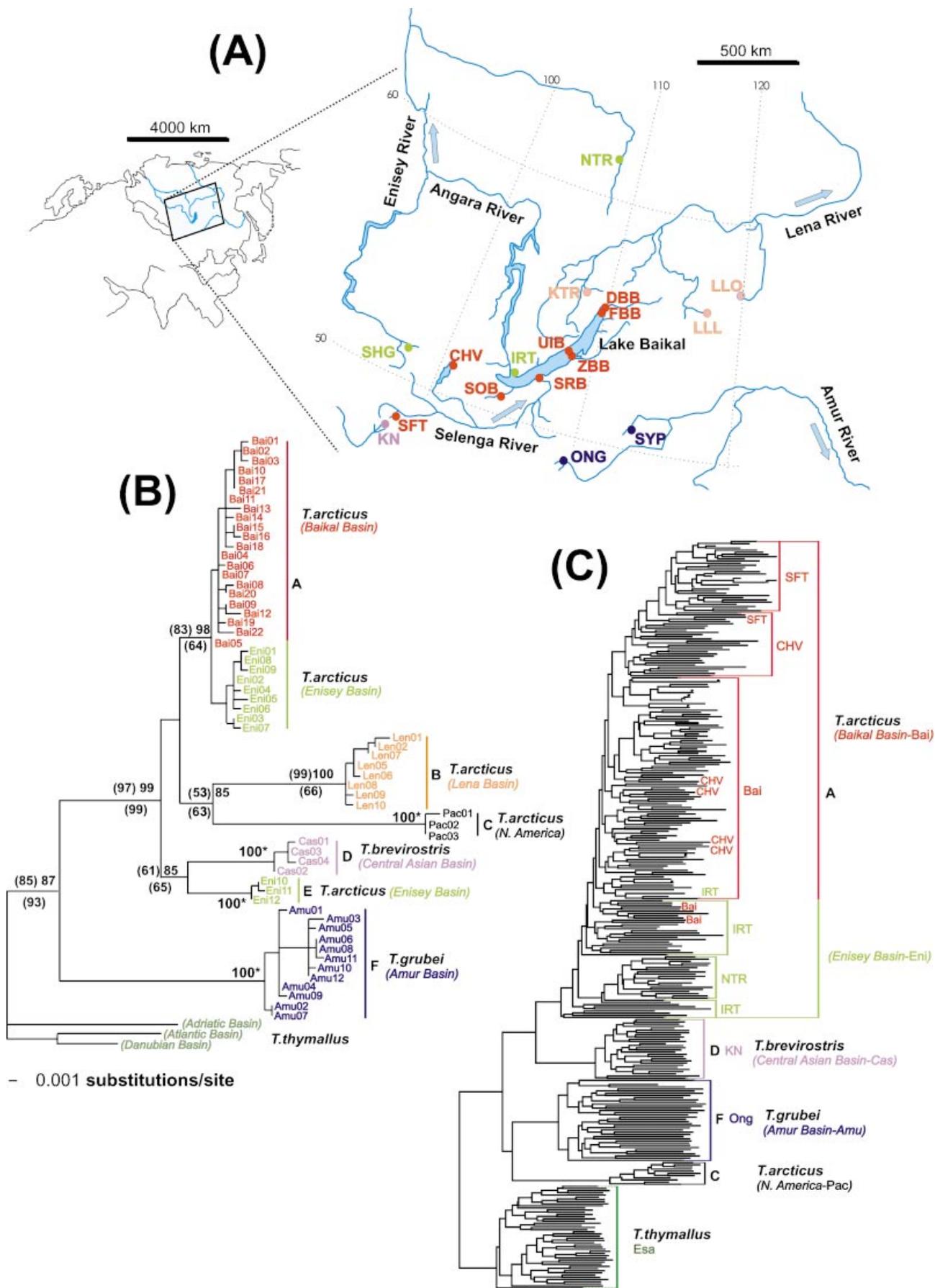
exhibited unambiguously scorable amplification products in all Asian grayling and were thus selected for further analyses: *BFRO004*, *BFRO009*, *BFRO010*, *BFRO013*, *BFRO015*, *BFRO018*, *Cocl23*, *One2*, *Str73INRA* and *Str85INRA*. Six of the microsatellites (*BFRO004–018*) have been isolated from *T. thymallus*, while the remaining four loci were derived from other salmonids and cross-amplify polymorphic alleles in *T. thymallus* (Koskinen & Primmer 1999). Original references and details of polymerase chain reaction and electrophoresis (on ABI-377) methods of the microsatellites are outlined elsewhere (Koskinen & Primmer 2001).

The program GENEPOP 3.2a (Raymond & Rousset 1995) was used to conduct exact probability tests for deviations from Hardy–Weinberg equilibrium, exact tests for deviations from genotypic linkage equilibrium, and to test for genic differentiation between populations. Corrections for multiple significance tests were performed by applying Fisher's method and a sequential Bonferroni-type correction (Rice 1989). GENEPOP 3.2a was further used to estimate among-population variance in allele frequencies (F_{ST}; Weir & Cockerham 1984) and variance in allele sizes (ρ_{ST}; Michalakis & Excoffier 1996).

Interpopulation dendrograms were derived with infinite allele model-based D_A distances (Nei *et al.* 1983), stepwise mutation model-based D_{SW} distances (Shriver *et al.* 1995) and D_{AS} distances that do not make explicit evolutionary assumptions (Bowcock *et al.* 1994). Statistical support estimates for dendrogram nodes were obtained by resampling loci across 2000 bootstrap replicates. Additionally, an individual based D_{AS} phenogram (Bowcock *et al.* 1994) was constructed. The D_A and D_{AS} distances and the corresponding trees were obtained using a program kindly provided by Jean-Marie Cornuet and Sylvain Piry (INRA, Montpellier, France). The D_{SW} distances were calculated using MICROSAT (Minch *et al.* 1995) and the corresponding dendrogram was constructed using PHYLIP 3.57c (Felsenstein 1995).

Microsatellite data were applied to test for variation in past population size within Lake Baikal (i.e. the pooled population Bai), using Markov Chain Monte Carlo simulations to recover the posterior probability distributions of several geneological and demographic parameters as implemented in the program MSVAR 0.4.1b (Beaumont 1999). A demographic history was assumed whereby a stable population of size N_1 chromosomes at a time *ta* ago, exponentially changed to its contemporary size of N_0. Microsatellite loci were assumed to follow a stepwise mutation model with a mutation rate μ (equal rates were assumed for all loci). Inter-relations of N_1, N_0 and μ were summarized with parameters *r*, *tf* and θ: $r = N_0/N_1$; $tf = ta/N_0$; and $\theta = 2N_0\mu$ (Beaumont 1999). The *a priori* lower and upper bounds on $\log_{10}(r)$, $\log_{10}(tf)$ and $\log_{10}(\theta)$ were set at: −1 and 1; −5 and 0; and −2 and 4, respectively. In order to confirm that the Markov Chain had converged, we used 10 000 iterations and replicated the simulations three times. The density plots of the parameters of interest were always similar across the replicates and unimodal, providing a strong indication that the Markov Chain had converged (Beaumont 1999). For estimating *ta*, a 5-year generation interval for Lake Baikal grayling (Tugarina 1981) and two different microsatellite mutation rates were assumed: (i) 2×10^{-4}, a germline mutation rate reported across 2000 zebrafish loci (Shimoda *et al.* 1999); and (ii) 6×10^{-4}, a germline dinucleotide mutation rate observed and often applied in humans (Weber & Wong 1993).

Results

mtDNA sequence diversity and inferred phylogeny

A total of 150 variable sites (110 parsimony informative) were derived from 142 sequences yielding 64 haplotypes. Phylogenetic reconstruction revealed six matrilineal clades (labelled A–F) with neighbour-joining bootstrap support ≥ 98% (Fig. 1b). Lineage A contained all Baikal basin haplotypes as well as those from two populations (NTR and IRT) in the Enisey (Baikal outflow) drainage (Fig. 1b). The most basal clade (F) represents *Thymallus grubei* from

Fig. 1 (A) Map indicating the origin of samples. Baikal basin (Bai): SFT = South Fork Terkhyn Gol, Mongolia; CHV = Lake Chovsgul, Mongolia; SOB = Sobolinoye Lake; SRD = Selenga River Delta; ZBB = Zavorotnaya Bay, Baikal; UIB = Ushkaniy Islands, Baikal; FBB = Frolikha Bay, Baikal; DBB = Dagary Bay, Baikal. Enisey basin (Eni): SHG = Shishkhed Gol, Mongolia; NTR = Nizhnaya Tunguska River; IRT = Irkutsk Reservoir Taltzinka. Lena basin (Len): KTR = Kutima River; LLL = Lesha Lake; LLO = Leprinidokan Lake. Amur basin (Amu): ONG = Onon Gol, Mongolia; SYP = Sypchegurka River. Central Asian basin (Cas): Khökh Nuur, Mongolia. (B) The inferred mtDNA phylogeny, generated with maximum likelihood, an Hasegassa-Kishino-Yana (HKY) model of evolution (γ = 0.005) and 10 replicate heuristic searches. Node support is shown as the percentage of bootstrap replicates for maximum likelihood (100 replicates; below node) and maximum parsimony (1000; above left) criteria, and neighbour-joining (1000; above right) algorithm. 100* designates 100% support for all three search methodologies. Six well-supported clades are labelled A–F. Haplotypes are colour-coded to match sample points on the map. (C) Neighbour-joining tree of individuals based on D_{AS} distances and 10 microsatellite loci. Vertical bars indicate population clusters. Individuals that do not group within the correct clusters are marked with their population code. Population codes and colours correspond to those on the mtDNA tree and map. The mtDNA haplotypes have been named according to the drainage basins, whereas for the microsatellite tree population codes are indicated beside the population clusters and drainage basins below the species designations. The within-Baikal populations DBB, FBB, UIB and ZBB are all marked as Bai in the microsatellite tree.
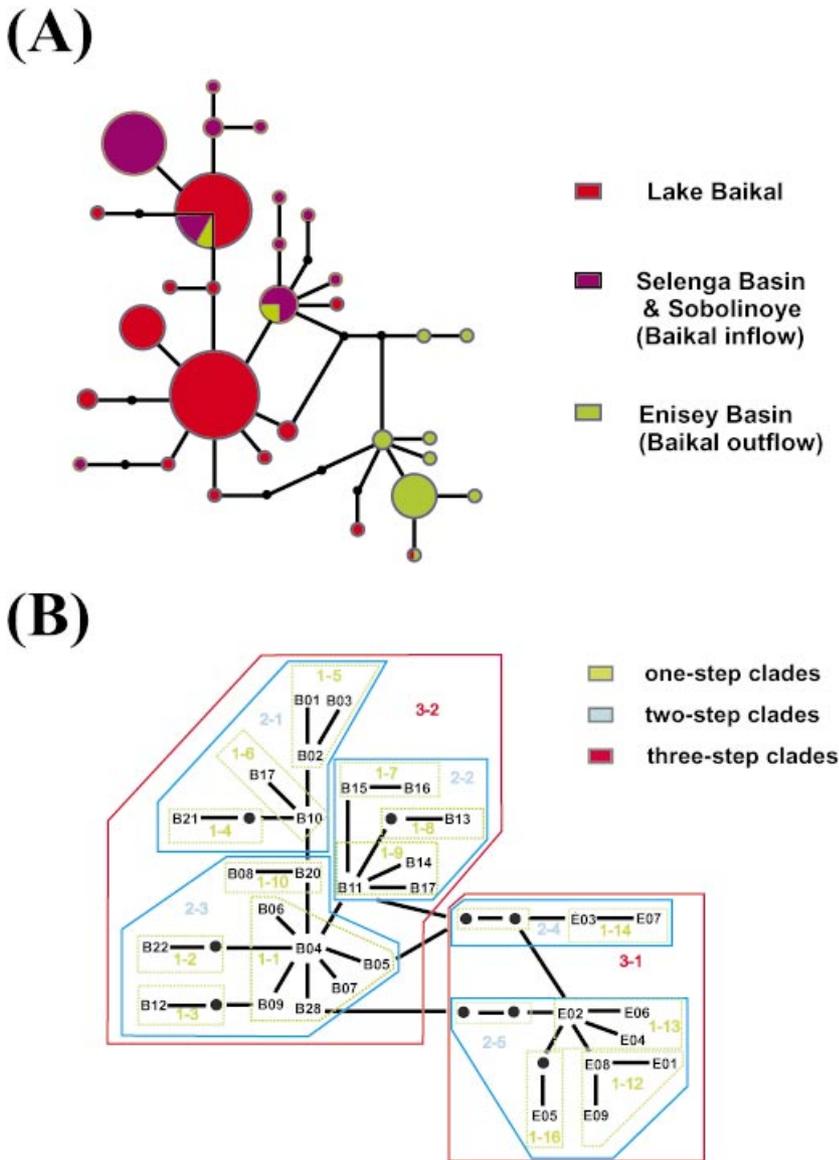
Fig. 2 (A) Haplotype cladogram for Baikalian drainage populations. Circle size reflects the frequency of each haplotype and solid lines connecting each haplotype (regardless of their length) represent a single mutational event. Small black circles represent missing or theoretical haplotypes. (B) The series of nested clades used in nested clade analysis.

the Amur basin (ONG and SYP; Fig. 1a,b). The morphologically divergent Mongolian grayling, *T. brevirostris* (KN; Clade D) appears as a sister clade to a sample of *T. arcticus* captured in an extreme headwater region of the Enisey drainage in northwestern Mongolia (SHG; Clade E) (Fig. 1b). Lena basin haplotypes (LLL, LLO and KTR; Clade B) appear as a sister clade to those from North America (Clade C). Clade B + C is a sister group to the Baikalian clade (A) under maximum likelihood (Fig. 1b) whereas D + E is basal. Neighbour joining and maximum parsimony support a minor alteration to this topology with the Mongolian clades (D + E) forming the sister group to the Baikalian clade, and Clade B + C as basal. Either topology results in a paraphyletic status for *T. arcticus*.

Nested clade analysis

The 95% parsimony network of Baikal basin and outflow (IRT and NTR) haplotypes spanned 11 mutations, revealing several star-like patterns of divergence (Fig. 2a). The nested design included 23 clades across four nested levels, whereby five clades revealed a significant association between genetic and geographical variation (Fig. 2b). At the first nesting level, reflecting events of the most recent past, past fragmentation was inferred for clade 1–6, containing haplotype (B17) fixed in Sobolinoye Lake (SOB), which belongs to a southern tributary system of Lake Baikal (Fig. 1a). At the next nesting level, the only inference drawn was that of range expansion for clade 2–3,
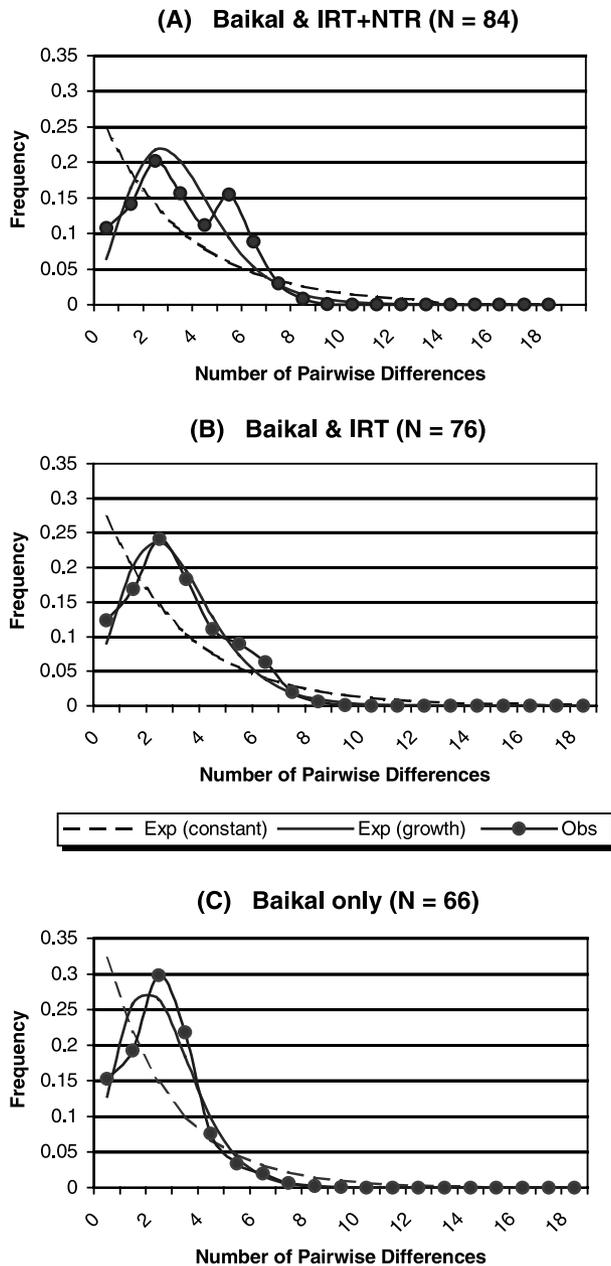
Fig. 3 The pairwise mismatch distributions for: (A) Baikal basin and the two Enisey basin outflow populations (IRT and NTR); (B) Baikal basin and the nearby Enisey basin (IRT) population; and (C) Baikal basin population.

which contained frequent Bakalian haplotypes along with tip haplotypes (B12 and B22) occurring in an upper Selenga River population in Mongolia (SFT; Fig. 2b). While clade 2–5 revealed a highly significant genetic/geographical association, sampling was inadequate to discriminate between past fragmentation, range expansion, and isolation by distance. At the third level, range expansion was inferred again, for clade 3–1, which contained tip haplotypes occurring in the Selenga basin, including 3

haplotypes private to Lake Chovsgul in Mongolia (B01, B02 and B03; CHV; Fig. 2b). A highly significant association was also seen for the total cladogram, whereby past fragmentation was inferred between clades within 3–1 (representing haplotypes from Baikal and its tributaries) and 3–2 (outflow populations IRT and NTR; Fig. 2b).

Pairwise mismatch analysis

The pairwise mismatch distribution of Baikal and Enisey (IRT and NTR) basin haplotypes was strongly bi-modal (Fig. 3a). However, when the distant NTR population was removed, the second mode nearly disappeared (Fig. 3b) implying that it represented population subdivision (Marjoram & Donnelly 1994). When both Enisey populations were removed, the leading wave completely disappeared, and the remaining uni-modal distribution reflected a model of sudden expansion (Fig. 3c). The observed pairwise mismatch distribution was significantly different from that expected under a stable population size as calculated in DNASP (K–S-test, $Z = 1.697$, $P = 0.006$). Under the least-squares approach, tests for the goodness-of-fit of the observed data to the expansion model revealed no significant differences (sum of squared deviations, $P = 0.223$; Harpending's raggedness index, $P = 0.488$). Using the expansion parameter ($\tau = 2.3$; 95% CI 0.37–4.24) and a divergence rate ranging from 0.5 to 2.0% per million years, the estimated mean age of expansion ranged from approximately 110 000 to 450 000 years (Fig. 4).

Microsatellite results

Unambiguously scorable amplification products were obtained for all 10 microsatellites across all populations,
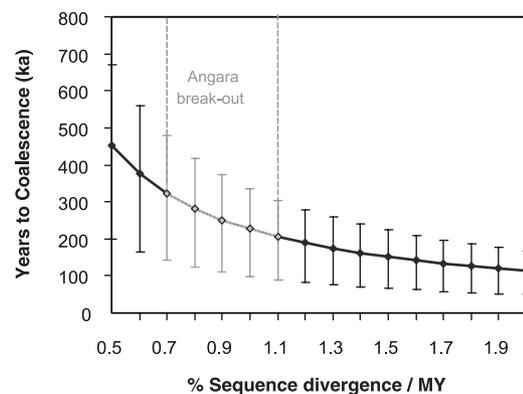

Fig. 4 Coalescence estimates based on τ and its 95% CI shown for the full range (0.5–2.0%) of divergence rates reported for salmonid mtDNA. Rates of around 1% however, are most likely for the mtDNA region sequenced (see text). The dotted lines indicate the divergence estimates corresponding to the putative time window when the Angara River broke out from Lake Baikal.
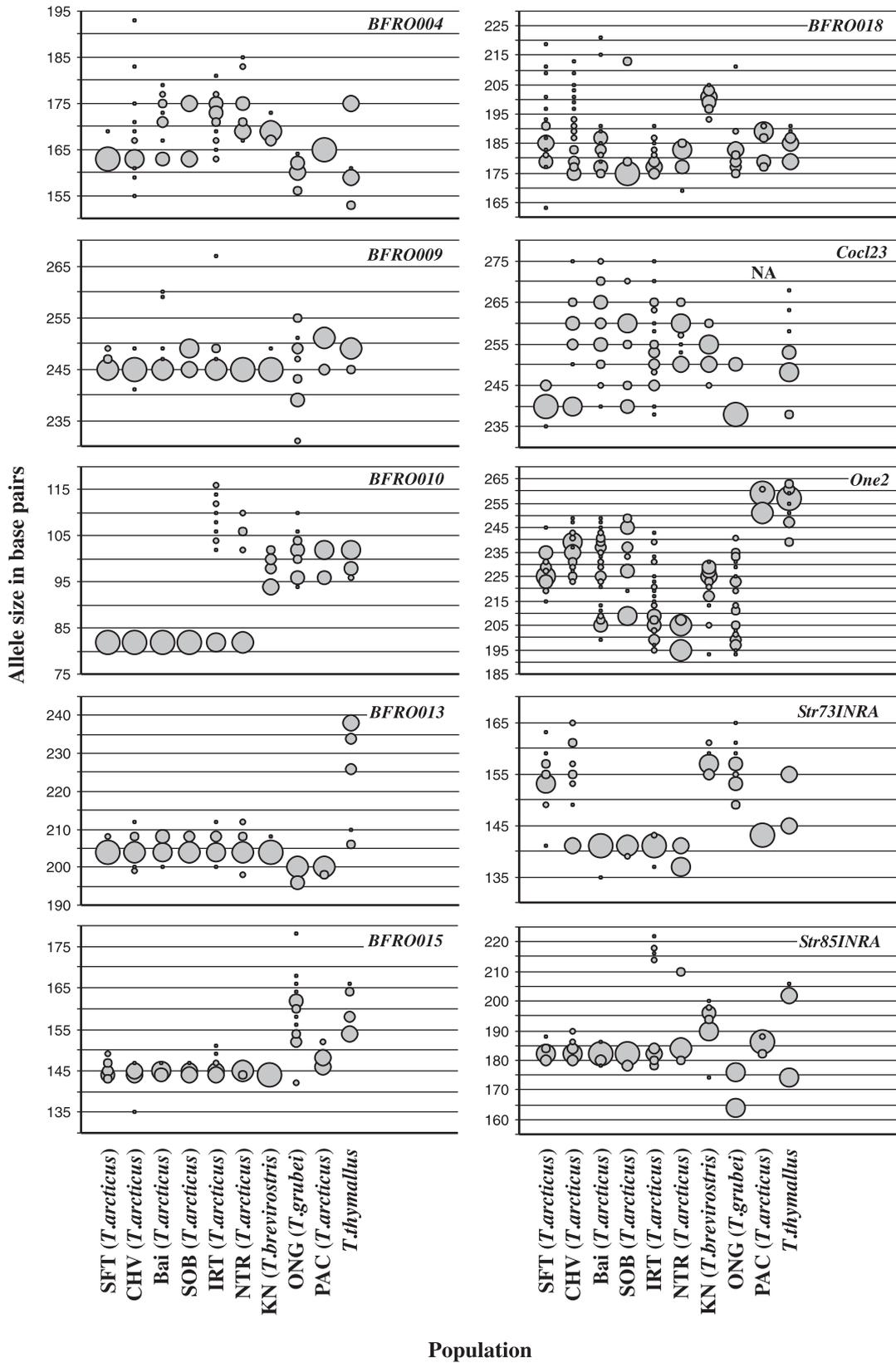
Fig. 5 Allele frequencies and size distributions of 10 microsatellite loci in *Thymallus* spp. (see Fig. 1 for the sampling locations of the populations). Areas of the bubbles correspond to frequencies of the respective alleles in given populations. NA indicates that, despite numerous attempts, the locus could not be amplified in the given population.

Table 2 Genetic differentiation among Asian *Thymallus* populations based on variance in allele frequencies (F_{ST}; above diagonal) and variance in allele sizes (ρ_{ST}; below diagonal)

	ONG	KN	NTR	IRT	CHV	SFT	SOB	Bai
ONG	—	0.403	0.401	0.335	0.386	0.417	0.428	0.420
KN	0.698	—	0.397	0.326	0.346	0.412	0.463	0.402
NTR	0.621	0.810	—	0.174	0.252	0.378	0.366	0.242
IRT	0.499	0.594	0.100	—	0.124	0.271	0.163	0.072
CHV	0.640	0.682	0.729	0.528	—	0.417	0.166	0.097
SFT	0.600	0.724	0.752	0.521	0.214	—	0.310	0.262
SOB	0.562	0.654	0.518	0.299	0.176	0.307	—	0.160
Bai	0.576	0.595	0.474	0.295	0.234	0.301	0.029	—

except for the locus *Cocl23* which, despite numerous attempts, could not be amplified in the North American *T. arcticus* population (Fig. 5). The mean number of alleles (across microsatellites) was 5.6 ($\pm$ 2.0) within the Asian *Thymallus* spp. populations. The level of polymorphism of the microsatellites across the Asian populations was much higher, averaging 17.5 ($\pm$ 6.7) (Fig. 5). All of the loci (across populations) and all of the populations (across loci) displayed genotypes in Hardy–Weinberg equilibrium proportions after correcting for multiple tests. Four significant ($P \leq 0.05$) departures from linkage equilibrium were observed between pairs of loci after correcting for multiple tests, however, no deviation occurred in more than one population, so there was no compelling evidence for linkages between the microsatellites.

Many alleles were geographically confined in the populations investigated (Fig. 5). Visualization of the allele-size distributions suggested that the *T. brevirostris* (KN) and *T. grubei* (ONG) populations were very diverged from the other Asian *Thymallus* spp. (Fig. 5). For instance, at the locus *Str85INRA*, the *T. brevirostris* and *T. grubei* specimens exhibited only alleles that were not found in any other population (Fig. 5). Accordingly, highly significant ($P < 0.001$) genic differentiation was detected between all pairwise population comparisons, and the level of genetic differentiation was often very high [$F_{ST} = 0.072–0.463$ (mean = 0.307); $\rho_{ST} = 0.029–0.810$ (mean = 0.490); Table 2].

The individual-based D_{AS} distances and neighbour-joining phenogram provided strikingly clear resolution of *Thymallus* evolutionary relationships (Fig. 1c). As already implied by the allele size distributions (Fig. 5), the *T. grubei* and *T. brevirostris* specimens formed clusters distinct from all *T. arcticus* from Baikal and Enisey basins (Fig. 1c). The Enisey basin *T. arcticus* (IRT and NTR) formed a relatively clear subcluster among Baikal basin samples (Fig. 1c). Similarly, most individuals from Lake Chovsgul (CHV) and Selenga River (SFT) grouped distinct from grayling sampled within Lake Baikal (Fig. 1c). At the population level of resolution, all three distance metrics grouped *T.*

grubei distinct from other Asian *Thymallus* with bootstrap support estimates of 91–98%. Furthermore, as revealed by D_A and D_{AS} distances, the Lake Baikal population and all populations physically connected to Baikal formed a well-supported group (70–82% bootstrap support). These metrics also indicated grouping of Lake Chovsgul (CHV) and Selenga River (SFT) populations with bootstrap support of 60–66%. The D_{SW} distance supported these conclusions, however, bootstrap estimates for separating Lake Baikal populations from *T. brevirostris* and grouping the SFT and CHV populations were lower (49 and 45%, respectively).

Analyses of population size variation provided a signal of population decline within Lake Baikal. When the simulations were run across loci, all sampled points of $\log_{10}(r)$ were substantially below zero, with a 95% posterior probability interval of −1.60 to −0.85 (mean = −1.21; Fig. 6a). Hence, from $r = N_0/N_1$ (Beaumont 1999) we estimate that the contemporary Baikal *Thymallus* population size is roughly only 2–14% of its historical size. Notably, when the polymorphic microsatellites were analysed individually, the upper bounds of the 95% probability intervals of $\log_{10}(r)$ were always clearly below zero, indicating that the signal of population size decline was not due to just a proportion of the loci, but a general trend across all polymorphic microsatellites. The 95% probability intervals for $\log_{10}(tf)$ and $\log_{10}(\theta)$ were −0.60–0.29 (mean = −0.18; Fig. 6a) and −1.88–0.18 (mean = −1.02), respectively. The most probable estimates of *ta*, the time since the commencement of decline, ranged from 20 to 50 years ago, depending on the microsatellite mutation rates applied in the calculations (Fig. 6b).

Discussion

Phylogenetic analysis

The mtDNA-based phylogenetic reconstruction was congruent with microsatellite-based dendrograms, revealing grayling lineages delimited by their occurrence in disjunct hydrological basins. The only exception to this pattern was the deep divergence between *Thymallus arcticus* in Northwestern Mongolia (SHG; Clade E) and *T. arcticus* found in the remaining Enisey basin populations (IRT and NTR). This incongruity, and the sister group status of Clade E with Mongolian grayling (Clade D), thought to represent the primitive member of the genus (Svetovidov 1936; Tugarina & Dashidorzhi 1972), suggests that the present systematic scheme of *Thymallus* needs revision. Mongolian grayling are a large-headed, lake-dwelling primarily piscivorous taxon confined to the Central Asian basin, whereas Clade E represents individuals displaying a typical *T. arcticus* phenotype, captured in the headwaters of the arctic draining Enisey (SHG; Fig. 1a). The SHG individuals may represent a refugial population with its
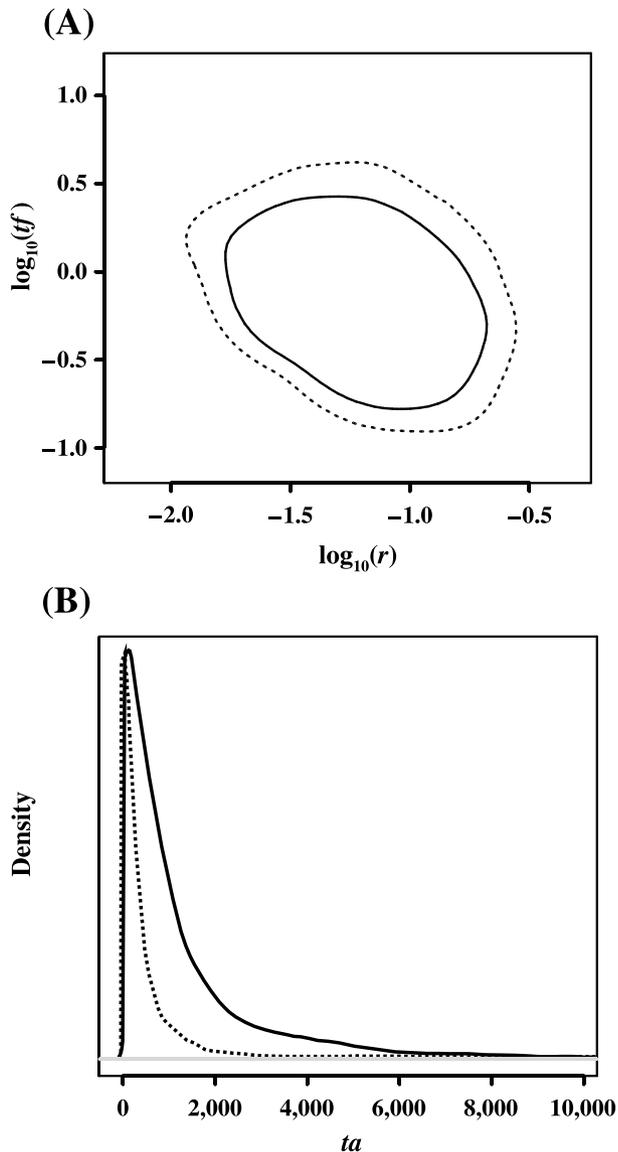
(A)

(B)

Fig. 6 (A) Plot of the marginal posterior distribution of $\log_{10}(r)$ and $\log_{10}(tf)$ obtained from Markov Chain Monte Carlo modelling approach applied to estimate demographic trends for the Lake Baikal grayling microsatellite data. The inner (solid line) and outer (dashed line) circles indicate 95% and 99% probability intervals, respectively. (B) Density curves generated from the posterior distribution of sampled points presented in (A) for *ta*, the estimated time since the commencement of the population size decline within Lake Baikal. The solid and dashed curves represent *ta* assuming mutation rates of 2×10^{-4} and 6×10^{-4}, respectively.

age predating the split between Mongolian and Arctic grayling. Furthermore, rather than representing a primitive taxon, *T. brevirostris* appears to have evolved relatively recently within the genus. The apparently more rapid morphological evolution that has occurred in Mongolian

grayling may result from the fact that this taxon has become isolated in the species-poor lakes of the Central Asian basin, where it has adapted to a primarily piscivorous feeding behaviour.

Contrasting the derived position of all Baikal basin haplotypes (Clade A), more ancestral lineages occur in all major basins (Amur, Clade F; Lena, Clade B; and Enisey, Clade E). Baikal's previous outlet was to the Lena drainage through the ancient Manzurka River valley, existing through the Pliocene, until tectonic uplift resulted in the lake's isolation, between 300 000 and 800 000 years ago (Florensov 1978). Consequently, grayling should have had access to Baikal prior to its isolation from the Lena system (i.e. throughout the Pliocene). However, pairwise divergence between Lena and Baikal basin haplotypes ranges from 2.3 to 2.8%, corresponding to an early Pleistocene to Pliocene/Pleistocene split. Thus, if ancient Lena basin grayling were abundant in Baikal, they have either gone through an extreme bottleneck resulting in the loss of older haplotypes, or have been extirpated sometime before the present inhabitants re-colonized the lake. The most recent major cooling periods (1.8–1.6 Ma and 2.8–2.6 Ma), calibrated with biogenic silica modelling (Williams *et al.* 1997), occurred in the early Pleistocene. It is therefore possible that grayling were extirpated from Baikal during one of these palaeoclimatic benchmarks. However, significant temperature fluctuations have occurred throughout the last 800 000 years (Williams *et al.* 1997); consequently, the effects of more recent climatic events cannot be excluded. For example, water-level fluctuations may have been sufficient to isolate Baikal's tributaries, where grayling spawn, resulting in complete extirpation of *Thymallus* from the lake, without incidence of glaciation. An alternative scenario would be that *Thymallus* did not occupy Lake Baikal until the most recent mid-to late Pleistocene colonization. While this scenario may appear more parsimonious concerning Baikal itself, it is difficult to envisage the degree of the lake's isolation from the surrounding river systems, that would have had to exist from the Pliocene through the early Pleistocene, given that diverse lineages of *Thymallus* are widespread in all Siberian drainages.

Nested clade analysis reveals patterns of colonization and expansion

The relation of clade 3–1 to 3–2, joined by mutational pathways with missing haplotypes (Fig. 2), gave both statistical and intuitive support for past fragmentation between Baikal and its present outflow to the Enisey. For Baikal's primary tributary Selenga, however, range expansion was inferred at two nesting levels. These inferences provided complementary support for a colonization–expansion pattern into Lake Baikal, and out into the Selenga River drainage. Expansion into the Selenga was

further supported by microsatellite-based dendrograms, as they grouped CHV and SFT individuals (and populations) together in a derived position, with respect to Lake Baikal. The overall pattern of range expansion implies that it occurred from a single-source (Lake Baikal) and proceeded relatively rapidly up to the extreme headwaters of the Selenga basin, including Baikal's ancient sister lake, Lake Chovsgul. The ancestral source of the Baikal lineage lies either in the Enisey basin (as implied by the phylogenetic analysis), or in some yet-to-be located refugium, and has potentially been associated with connections between the Lena and Enisey nearby Lake Baikal. An interesting aspect of this expansion is that it apparently did not include derived Baikalian haplotypes, re-invading the presumably occupied habitat in the Angara River. In other words, colonization was able to take place into an unoccupied habitat, but extensive gene flow did not occur back into the Angara system.

A mid-Pleistocene explosion

A monotonic, sharply peaked distribution of pairwise mtDNA differences (Fig. 3c) can have two potential causes. Either a sudden or exponential expansion, as our significant fit to the model suggests, or selection acting on mtDNA. The null hypothesis of neutrality could not be rejected with our data (Tajima's D, -1.609, $P > 0.05$). It is also worth mentioning that the distribution of pairwise differences (Fig. 3c) was drawn from populations occupying diverse habitats separated by over 2000 km of waterway, a difficult landscape to accept a selection/replacement event at the mtDNA locus. Additional support arguing against selection stems from the selectively neutral nuclear microsatellite loci, which were largely concordant with mtDNA in reflecting overall patterns of interbasin divergence.

The estimated age of expansion is plotted in terms of a broad range of mtDNA divergence rates, whereby a salmonid-specific calibration is 1% per million years (Smith 1992). This rate appears appropriate for the mtDNA sequence used in this study, which has been shown to evolve slower than NADH subunits 1, 2, 5 and 6 (Hansen & Loeschcke 1996; Apostolidis *et al.* 1997; Chirikov *et al.* 2001), but faster than the remaining coding segments of the molecule (Chirikov *et al.* 2001). Thus, a more realistic mean estimate of the age of expansion is between approximately 225 000 and 325 000 years ago.

This estimated age of expansion is fully concordant with a hypothesis regarding the formation of Baikal's contemporary outlet, the Angara River. Namely, subsequent to Baikal's isolation from the Lena drainage, its water levels are presumed to have risen until they broke through the Primorskiy mountain range approximately 200 000–300 000 years ago (Kozhov 1972), forming an outlet to the Enisey basin. The sudden outbreak of the Angara River

provided a new hydrological corridor between the Enisey and Lake Baikal, which we presume was associated with significant ecological changes in Baikal that provided an ideal setting for rapid population expansion. The dating of this palaeo-hydrological event results in a mean mtDNA divergence estimate of 0.75–1.10% per million years, straddling the calibrated estimate of 1%.

This scenario is also concordant with a molecular based divergence estimate (380 000 years) between the Baikal seal (*Phoca sibirica*) and its presumed ancestor, the Ringed seal (*Phoca hispida*) (Sasaki & Numachi 1997), assuming an Enisey River colonization route. Additional biological evidence for Pleistocene perturbations stems from molecular analyses of the diverse amphipod fauna, which reveals both old and new lineages, as well as vicariant patterns of divergence (Sherbakov *et al.* 1999; Väinölä & Kamaltynov 1999). It is hypothesized that at least some of this diversification developed during the Pleistocene, through fragmentation events involving one or more of the Baikal's three major isobaths.

Holocene declines

Despite the ancient expansion supported by mtDNA, microsatellite loci provided a clear signal of recent population size decline in Lake Baikal. Such contemporary declines, occurring during the last century (Fig. 6b), will have little or no effect on a uni-modal pairwise mismatch distribution of mtDNA haplotypes resulting from a more ancient expansion event, i.e. the two marker types provided complementary, not contrasting, resolution (Rogers 1996). Currently, there is a thriving sport and commercial (gillnet) fishery on Baikal and its Angara River outflow, and it has been suggested that overfishing resulted in declines of Baikal *T. arcticus* in the 1960s (Kozhova & Izmest'eva 1998). The microsatellite data appear to reflect the severity and timing of this anthropogenic influence strikingly well.

Conclusions

As molecular-based phylogeographic data sets are often used in connection with *a priori*, or *post hoc* palaeo-climatic scenarios, it appears important that multiple marker systems are used, and that care is taken to consider the variable temporal depth of demographic history that specific markers may reflect. For this study, involving one of the world's most enigmatic hydrological systems, mtDNA and microsatellite loci were largely congruent in reflecting several hypothesized palaeo-geological events, such as the break-out of the Angara river, and the isolation of Baikal from the Lena basin. While the coalescence of several invertebrate lineages in Baikal is thought to reach into the Miocene or Oligocene (Sherbakov *et al.* 1999), organisms dependant on an ice-free environment, such as

fish, would seem to have a relatively recent ancestry in the lake (Kiril'chik & Slobodyanyuk 1997; Sherbakov *et al.* 1998, 1999). While complete glacial coverage of Lake Baikal (Grosswald 1998; references therein) is difficult to envisage considering the diverse, ancient and solar-dependent invertebrate community that is present, our results demonstrate that environmental perturbations have had a prevailing influence on the ability of *Thymallus* to survive in Lake Baikal. Thus, a more astute consideration of the biotic effects of ice-age phenomena in this region may significantly alter the zoogeographic and limnological perpectives of Baikal's history.

Acknowledgements

References

Apostolidis AP, Triantaphyllidis C, Kouvatsi A, Economidis PS (1997) Mitochondrial DNA sequence variation and phylogeography among Salmo trutta L. (Greek brown trout) populations. *Molecular Ecology*, **6**, 531–542.

Bânârescu P (1991) *Zoogeography of Fresh Waters, 2: Distribution and Dispersal of Freshwater Animals in North America and Eurasia*. AULA-Verlag, Wiesbaden.

Beaumont M (1999) Detecting population expansion and decline using microsatellites. *Genetics*, **153**, 2013–2029.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza L-L (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368**, 455–457.

Chirikov D, Matsuoka M, Luan X, Gray AK, Brykov VLA, Gharrett AJ (2001) Assessment of concordance among genealogical reconstructions from various mtDNA segments in three species of Pacifid salmon (genus *Oncorhynchus*). *Molecular Ecology*, **10**, 2329–2339.

Clemens M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Felsenstein J (1995) PHYLIP *(Phylogeny Inference Package)*, Version 3.57c. Department of Genetics, University of Washington, Seattle, WA.

Florensov NA (1978) *Istoriya Ozera/Problemi Baikala*, pp. 9–17. Nauka Press, Novosibirsk (in Russian).

Goldstein PZ, DeSalle R, Amato G, Vogler AP (2000) Conservation genetics at the species boundary. *Conservation Biology*, **14**, 120–131.

Grosswald MG (1998) New approach to the ice age paleohydrology of northern Eurasia. In: *Paleohydrology and Environmental Change* (eds Benito G, BakerVR, Gregory KJ), pp. 199–214. Wiley and Sons, Chichester.

Hansen MM, Loeschcke V (1996) Genetic differentiation among Danish brown trout populations, as detected by RFLP analysis of PCR amplitied mitochondrial DNA segments. *Journal of Fish Biology*, **48**, 422–436.

Harpending H (1994) Signature of ancient population growth in a low resolution mitochondrial DNA mismatch distribution. *Human Biology*, **66**, 591–600.

Hewitt G (2000) The genetic legacy of quaternary ice ages. *Nature*, **405**, 907–913.

Huelsenbeck JP, Crandall KA (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics*, **28**, 437–466.

Kiril'chik SV, Slobodyanyuk SY (1997) Evolution of a fragment of mitochondrial gene *cyt b* in some Baikalian and non-Bakalian species of cottoid fishes. *Molecular Biology*, **31**, 168–175.

Koskinen MT, Primmer CR (1999) Cross-species amplification of salmonid microsatellites which reveal polymorphism in European and arctic grayling, Salmonidae: *Thymallus* spp. *Hereditas*, **131**, 171–176.

Koskinen MT, Primmer CR (2001) High throughput analysis of 17 microsatellite loci in grayling (*Thymallus* spp. Salmonidae). *Conservation Genetics*, **2**, 173–177.

Kozhov MM (1972) *Ocherki Po Baikalovedeniyu*. East-Siberian Press, Irkutsk (in Russian).

Kozhova OM, Izmest'eva LR (1998) *Lake Baikal. Evolution and Biodiversity*. Backhyus Publishers, Leiden.

Kuz'min MI (2001) *Vo l'dakh Baikala*, p. 140. SB RAS Publishers, 'Geo' branch, Novosibirsk (in Russian).

Marjoram P, Donnelly P (1994) Pairwise comparisons of mitochondrial DNA sequences in subdivided popuations and implications for early human evolution. *Genetics*, **136**, 673–683.

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special interest to microsatellite loci. *Genetics*, **142**, 1061–1064.

Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza L-L (1995) *Microsat (Version 1.5): A Computer Program for Calculating Various Statistics on Microsatellite Allele Data*. Stanford University Medical Center, Stanford, CA.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, **19**, 153–170.

Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, **16**, 37–45.

Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rice WR (1989) Analysing tables of statistical tests. *Evolution*, **43**, 223–225.

Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. *Evolution*, **49**, 608–615.

Rogers AR (1997) Population structure and modern human origins. In: *Progress in Population Genetics and Human Evolution* (eds Donnelly PJ, Tavare S), pp. 55–79. Springer Verlag, New York.

Rogers AR, Harpending HC (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Rozas J, Rozas R (1999) DnaSP, Version 3.1 An integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174–175.

Sasaki H, Numachi K (1997) The origin and genetic relationships of Baikal seal elucidated by restriction analysis of mitochondrial DNA. In: *Animal Community, Environment and Phylogeny in Lake Baikal* (ed. Miyazaki N), pp. 91–105. Otsuchi Marine Research Center, Tokyo.

Schneider S, Roessli D, Excoffier L (2000) *Arlequin*, Version 2. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva.

Sherbakov DYu, Kamaltynov RM, Ogarkov OB, Verheyen E (1998) Patterns of evolutionary change in Baikalian gammarids inferred from DNA sequences (Crustacea, Amphipoda). *Molecular Phylogenetics and Evolution*, **10**, 160–167.

Sherbakov DYu, Kamaltynov RM, Ogarkov OB, Väinölä R, Vainio JK, Verheyen E (1999) On the phylogeny of Lake Baikal amphipods in the light of mitochondrial and nuclear DNA sequence data. *Crustaceana*, **72**, 911–919.

Shimoda N, Knapik EW, Ziniti J *et al.* (1999) Zebrafish genetic map with 2000 microsatellite markers. *Genomics*, **58**, 219–232.

Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrel RE, Chakraborty R (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Molecular Biology and Evolution*, **12**, 914–920.

Smith GR (1992) Introgression in fishes: significance for paleontology, cladistics, and evolutionary rates. *Systematic Biology*, **41**, 41–57.

Svetovidov AN (1936) Graylings (genus *Thymallus*, Cuvier) of Europe and Asia. *Proceedings of Institute of Zoology Academy of Sciences USSR*, **III**, 183–301 (in Russian).

Swofford DL (2000) PAUP*, Version 4.0.b3a. *Phylogenetic Analysis Using Parsimony and Other Methods*. Sinauer Associates, Sunderland, MA.

Templeton AR (1998) Nested clade analysis of phylogeographic data: Testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–398.

Templeton AR (2002) Out of Africa again and again. *Nature*, **416**, 45–51.

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*, **134**, 659–669.

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred fronm restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, **117**, 343–351.

Templeton AR, Crandall KA, Sing CR (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–623.

Tugarina PYa, (1981) *Khariusi Baikala (Grayling of the Lake Baikal)*. Nauka Press, Novosibirsk (in Russian).

Tugarina PYa, Dashidorzhi A (1972) Mongolian grayling *Thymallus brevirostris* (Kessler) of Dzabakhan River Basin. *Journal of Ichthyology* **12 N 5**, 843–856 (in Russian).

Uiblein F, Jagsch A, Honsig-Erlenburg W, Weiss S (2001) Status, habitat use, and vulnerability of the European grayling in Austrian water. *Journal of Fish Biology*, **59**, 223–247.

Väinölä R, Kamaltynov RM (1999) Species diversity and speciation in the endemic amphipods of Lake Baikal: Molecular evidence. *Crustaceana*, **72**, 945–956.

Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics*, **2**, 1123–1128.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Weiss S, Persat H, Eppe R, Schlötterer C, Uiblein F (2002) Complex patterns of colonization and refugia revealed for European grayling *Thymallus thymallus*, based on complete sequencing of the mtDNA control region. *Molecular Ecology*, **11**, 1393–1408.

Williams DF, Peck J, Karabanov EB *et al.* (1997) Lake Baikal record of continental climate response to orbital insolation during the past 5 million years. *Science*, **278**, 1114–1117.