

Secondary contact between two divergent lineages of grayling *Thymallus* in the lower Enisey basin and its taxonomic implications

S. WEISS*†, I. KNIZHIN‡, V. ROMANOV§ AND T. KOPUN*

*Karl-Franzens Universität Graz, Institut für Zoologie, Universitätsplatz 2, A-8010 Graz, Austria, ‡Irkutsk State University, Department of Vertebrate Zoology, 664003, Irkutsk, Karla Marksa 1, Russia and §Tomsk State University, Department of Ichthyology and Hydrobiology, 634050, Tomsk, Lenina, 36, Russia

(Received 23 November 2006, Accepted 30 July 2007)

A contact zone between two widespread divergent lineages of grayling *Thymallus* sp. has been identified in the lower Enisey basin (Khantaiskoye Lake) in north central Siberia (Russia). Sequences of the mitochondrial DNA control region revealed two lineages with 2.6% divergence. Haplotypes from these two lineages clearly fall within two previously described clades, one presumably corresponding to Arctic grayling *Thymallus arcticus* found throughout the Siberian coastal zone as well as North America, and the other corresponding to grayling found elsewhere in the Enisey basin, including Lake Baikal and its tributaries. Allelic variation at seven microsatellite loci strongly suggests reproductive isolation in sympatry between these two lineages in Khantaiskoye Lake. Integration of these new data with existing knowledge on *Thymallus* throughout its Siberian range supports a taxonomic revision. The currently recognized sub-specific lineages of Arctic grayling along the Siberian Arctic coast as well as grayling from North America should probably be regarded as *T. arcticus*. Arctic grayling, however, are limited to the coastal zone and lower reaches of the Enisey. Most of the Enisey basin, including Lake Baikal and its tributaries are occupied by another taxon, which should be recognized as *Thymallus baicalensis*. This perspective strictly follows evolutionary principles, eliminates the use of four sub-specific names and circumvents the illogical mosaic distribution of *T. arcticus arcticus*, which nevertheless does not represent a monophyletic lineage.

© 2007 The Authors

Journal compilation © 2007 The Fisheries Society of the British Isles

Key words: Arctic grayling; Lake Khantaiskoye; *Thymallus arcticus*; *Thymallus baicalensis*.

INTRODUCTION

The genus *Thymallus* Cuvier is widely distributed in Eurasia, with taxa found in all four major Siberian rivers. These systems, the Ob, Enisey, Lena and Amur are among the 10 largest freshwater rivers of the world and each contains large areas that are difficult to access and have thus received limited research attention. Despite this sampling challenge, a great deal of phenotypic diversity has been described for Siberian grayling within and among these basins, and this

†Author to whom correspondence should be addressed. Tel.: +43 316 380 5599; fax: +43 316 380 9875; email: steven.weiss@uni-graz.at

has led to various controversial schemes of systematics and taxonomy. Excluding the DNA \times DNA hybridization work of Skurikhina (1984), the first broad-scale molecular genetic work on Siberian *Thymallus* revealed strong phylogeographic structure, with major drainages harbouring distinct mitochondrial (mt) DNA lineages (Koskinen *et al.*, 2002). More recent molecular research, however, coupled with systematic morphological analysis, supports the existence of multiple lineages within major river systems, notably in the Amur (Froufe *et al.*, 2003; Antonov, 2004; Knizhin *et al.*, 2004) and the Lena (Knizhin *et al.*, 2006a; Weiss *et al.*, 2006). Except for a few relatively small regions in the Amur basin, these lineages were all found in distinct geographic areas with no zones of overlap.

The genetic analysis in Weiss *et al.* (2006) demonstrated that mtDNA haplotypes throughout the Arctic coastal region in Siberia, along with those reported in Stamford & Taylor (2004) in North America, all belong to one diverse clade of Arctic grayling *Thymallus arcticus* (Pallas). Weiss *et al.* (2006) also showed that, in the Lena system, this clade was limited to the Delta region, whereas the rest of the Lena catchment contained another lineage distinguishable with both genetic and phenotypic characters and thus represented a distinct taxon. No sample from the Enisey River basin was included in that analysis. Romanov (2005) reported two distinct phenotypes of Arctic grayling from a region in the lower Enisey basin and assigned one to the sub-specific taxon *T. arcticus pallasii* Valenciennes and the other to the nominal sub-species *T. arcticus arcticus* (Pallas). These observations were primarily made in and around Lake Khantaiskoye, a 822 km² body of water located 174 km east of the main channel of the Enisey.

Whereas North American biologists currently group all populations of Arctic grayling under one taxon (*T. arcticus*), the recognition of several sub-species (e.g. *T. a. arcticus*, *T. a. pallasii*, *T. a. mertensii*, *T. a. signifer*, *T. a. nigrescens*, *T. a. baicalensis* and *T. a. grubii*) is common practice among Russian ichthyologists and field biologists. Most populations along the Arctic coast are recognized as *T. a. pallasii*, first described from the Kolyma River (Cuvier, Valenciennes, 1848), but ranging from the Taimyr Peninsula in the west (just east of the Enisey Delta) to some tributaries of the Bering and Okhotsk seas (Chereshnev *et al.*, 2002). The nominal sub-species *T. a. arcticus* has a large and mosaic distribution existing throughout Siberia including areas both upstream (the Selenga basin in Mongolia) and downstream (Enisey) of Lake Baikal (Svetovidov, 1936; Berg, 1948; Shatunovskiy, 1983; Dorofeeva, 2002). *Thymallus arcticus baicalensis* is limited to Lake Baikal and its tributaries (Svetovidov, 1936; Shatunovskiy, 1983; Knizhin *et al.*, 2006b). *Thymallus a. mertensii* is recognized from some populations of the Okhotsk Sea and Kamchatka, and *T. a. signifer* only for several tributaries of the Bering Sea (Chereshnev *et al.*, 2002). Regardless of one's philosophy concerning the use of sub-specific nomenclature and the lack of clarity concerning these distributions, the lack of diagnostic characters as well as the mosaic distribution of *T. a. arcticus* strongly suggests that this taxonomic scheme is not concordant with evolutionary principles.

This work aims to (1) evaluate the hypothesis of two sympatric grayling taxa in the Khantaiskoye Lake region of the lower Enisey basin, (2) integrate these

new data with existing data to assign these lineages systematically within the genus and (3) report the taxonomic implications of these results.

MATERIALS AND METHODS

A total of 104 grayling were sampled from 24 locations by angling or gill netting (Table I and Fig. 1). A small piece of fin tissue was preserved in 96% ethanol, and whole genomic DNA was isolated using a high salt extraction protocol, modified from Miller *et al.* (1988). Samples from the hypothesized area of sympatry (Khantaiskoye Lake region, $n = 67$) contained two phenotypes, which local biologists claim can be distinguished primarily based on the pigmentation pattern of the dorsal fin. These samples are from three locations within the lake basin, one of which (Gogochenda) revealed both phenotypes in sympatry (Table I). Romanov (2005) considers these phenotypes to represent two sub-species of Arctic grayling, *T. a. arcticus* and *T. a. pallasii*.

The complete mtDNA control region was amplified using the LRBT-25 and LRBT-1195 primers described in Uiblein *et al.* (2001). Polymerase chain reaction (PCR) conditions (25 μ l reactions) were as follows: each reaction contained 19 μ l H₂O, 2.5 μ l $\times 10$ Promega Buffer B, 0.5 μ l (10 mM) of each primer, 1.5 μ l (25 mM) MgCl₂, 0.5 μ l (10 mM) dNTP's, 0.1 μ l (25 mM) Taq-polymerase (Peq Lab, Erlangen, Germany) DNA polymerase, 0.5 μ l (100 ng μ l⁻¹) DNA template. The cycle parameters were as follows: initial denaturation at 94° C for 3 min, denaturation at 94° C (40 s), annealing at 55° C (40 s) and extension at 72° C (40 s) repeated for 30 cycles. Amplified DNA templates were purified using ExoSAP-IT (Amersham Biosciences, Baden-Württemberg, Germany) and *c.* 100 ng of purified PCR product was used in cycle sequencing reactions following ABI PRISM BigDye Terminator protocols. Sequences were visualized on an ABI-3100 genotype analyser. New mtDNA sequences generated in this study were deposited in GenBank (accession numbers EU-168909-EU168943).

Sequences (1106 bp) produced in this study were aligned by eye, together with published haplotypes from representative populations from the Enisey and Selenga/Baikal basins, the Arctic coast from the Urals to Alaska and several interior Siberian populations. Two haplotypes of European grayling *Thymallus thymallus* (Linnaeus, 1758) serve as an outgroup (Table I). Sequence data were imported into PAUP Version 4.0b10 (Swofford, 2001) for phylogenetic analysis. Maximum parsimony (MP) and neighbour-joining (NJ) (with Kimura two-parameter distances) were used for phylogenetic reconstruction. A heuristic search (10 replicates) was used to estimate the most probable topology for MP, starting with a stepwise addition of trees and coding indels as a fifth character. Node support was estimated with 1000 bootstrap replicates. Between-group variation (corrected for within-group variation) was calculated using the net nucleotide divergence (D_n) in MEGA 3.1 (Kumar *et al.*, 2004).

Data on mtDNA variation were supplemented with data from bi-parentally inherited genetic markers to evaluate the potential for reproductive isolation for samples in the Khantaiskoye Lake region as well as the two nearest locations within the Enisey basin ($n = 67$). For this analysis, seven microsatellite loci, previously used in Arctic grayling populations (Koskinen *et al.*, 2002) were screened through the use of one single and three duplex PCR reactions. Each 10 μ l multiplex PCR was composed of 1 μ l of $\times 10$ reaction buffer (10 mM), 0.6 μ l MgCl₂, between 0.25 μ l and 0.50 μ l (10 mM) of each primer, 0.25 μ l (10 mM) dNTPs, 0.1 μ l (25 mM) Taq-Polymerase (Peq Lab), 1 μ l of template DNA (*c.* 100 ng μ l⁻¹) and a variable amount of H₂O. Each reaction consisted of 30 cycles of initial denaturation at 94° C (45 s), annealing at 57° C (20 s), extension at 72° C (20 s) and a final extension of 72° C for (60 s). Fluorescently labelled PCR products were loaded onto 96 well plates and dried by heating at 60° C for 20 min. A loading solution of 9.6 μ l formamide and 0.4 μ l GeneScan-350 (ROX) size standard was added to each sample. Samples were denatured at 60° C for 1 min and put on ice before loading onto a 16 capillary ABI PRISM® 3100 Genetic Analyser. Results were viewed and genotyping of all loci was performed using the programme GeneScan 3.7 (Applied Biosystems, Foster City, CA, U.S.A.).

TABLE 1. Sample locations including drainage, geographical coordinates, number of individuals screened and GenBank accession numbers. For Siberian populations, a three-letter code and number corresponds to site locations in Fig. 1. The commonly used sub-specific name, if available, along with the full species name proposed in this manuscript

Sample site	Code	Drainage	Prior sub-specific designation	Suggested species taxon	Latitude (N)	Longitude (E)	Number of individuals		GenBank accession numbers
							mtDNA	msats	
Edyngde Lake	Edy (1)	Khantaiskoye Lake, Enisey	<i>pallasii</i>	<i>T. arcticus</i>	68°16'	91°16'	5	5	—
Gogochenda River	Gog (2)	Khantaiskoye Lake, Enisey	<i>pallasii</i>	<i>T. arcticus</i>	68°17'	91°07'	9	9	—
Gogochenda River	Gog (2)	Khantaiskoye Lake, Enisey	<i>arcticus</i>	<i>T. baicalensis</i>	68°17'	91°07'	11	11	—
Nerotkar River	Ner (3)	Khantaiskoye Lake, Enisey	<i>arcticus</i>	<i>T. baicalensis</i>	68°25'	91°13'	5	5	—
N. Tunguska River*	Ntr (4)	Enisey	<i>arcticus</i>	<i>T. baicalensis</i>	60°55'	107°56'	2	—	AY168366; AY168367
Varlamovka River	Var (5)	Enisey	<i>arcticus</i>	<i>T. baicalensis</i>	62°23'	89°25'	5	9	—
Urunge Nur	Uru (6)	Oka, Angara, Enisey	<i>arcticus</i>	<i>T. baicalensis</i>	51°57'	100°23'	5	28	—
Mongosha Lake	Mon (7)	Oka, Angara, Enisey	<i>arcticus</i>	<i>T. baicalensis</i>	51°59'	100°24'	1	—	—
Taltzinka River	Irt (8)	Irkutsk Reservoir, Angara, Enisey	<i>arcticus</i>	<i>T. baicalensis</i>	51°59'	104°36'	6	—	AY168358; AY168365
Angara River	Ang (9)	Enisey	<i>arcticus</i>	<i>T. baicalensis</i>	52°26'	104°08'	4	—	—
Dagary Bay, Baikal*	Dbb (10)	Baikal, Angara, Enisey	<i>baicalensis</i>	<i>T. baicalensis</i>	55°41'	109°53'	1	—	AY168351
Zavorotnaya, Baikal*	Zbb (11)	Baikal, Angara, Enisey	<i>baicalensis</i>	<i>T. baicalensis</i>	54°18'	108°29'	1	—	AY168376
Shishkid Gol	Shi (12)	Enisey	<i>none</i>	<i>T. sp?</i>	51°28'	99°01'	3	—	AY168403; AY168405

TABLE I. Continued

Sample site	Code	Drainage	Prior sub-specific designation	Suggested species taxon	Latitude (N)	Longitude (E)	Number of individuals		GenBank accession numbers
							mtDNA	msats	
Anabar River*	Ana (13)	Arctic Ocean	<i>pallasii</i>	<i>T. arcticus</i>	70°48'	113°13'	1	—	DQ683684
Olenek River*	Ole (14)	Arctic Ocean	<i>pallasii</i>	<i>T. arcticus</i>	68°31'	112°00'	2	—	DQ683718; DQ683719
Lena Delta*	Del (15)	Arctic Ocean	<i>pallasii</i>	<i>T. arcticus</i>	71°08'	127°08'	2	—	AY779015; AY779016
Kutima River*	Len (16)	Kirenga, Lena, Arctic Ocean	<i>lenensis</i> †	<i>T. sp?</i>	57°02'	108°36'	2	—	AY168379; AY168380
Pilka River*	Pil (17)	Lena, Arctic Ocean	<i>lenensis</i> †	<i>T. sp?</i>	60°00'	113°00'	2	—	DQ683696; DQ683697
Inderkey River*	Ind (18)	Yana, Arctic Ocean	<i>pallasii</i>	<i>T. arcticus</i>	64°43'	132°33'	2	—	DQ683711; DQ683723
Turakh Lake*	Tur (19)	Indigirka, Arctic Ocean	<i>pallasii</i>	<i>T. arcticus</i>	64°15'	145°15'	2	—	DQ683716; DQ683717
Kolyma River*	Kol (20)	Arctic Ocean	<i>pallasii</i>	<i>T. arcticus</i>	66°10'	151°03'	2	—	DQ683721; DQ683720
Sjamzhenga River*	Sja (21)	S. Dvina River, White Sea	<i>pallasii</i>	<i>T. arcticus</i>	63°70'	46°26'	1	—	AY779014
Burnt River*	Pac	British Columbia, Pacific (N. America)	—	<i>T. arcticus</i>	122°20'	55°25'	1	—	AY168402
Soča River*	Ad	Soča, Isonzo, Adriatic Sea (Europe)	—	<i>T. thymallus</i>	47°50'	13°20'	1	—	AF522419
Loire River*	At	Atlantic Ocean (Europe)	—	<i>T. thymallus</i>	44°51'	03°55'	1	—	AF522425

*Samples screened in a previous investigation (Koskinen *et al.*, 2002; Froufe *et al.*, 2003; Weiss *et al.*, 2006).

†Sub-specific name *T. a. lenensis* was used in Weiss *et al.* (2006).

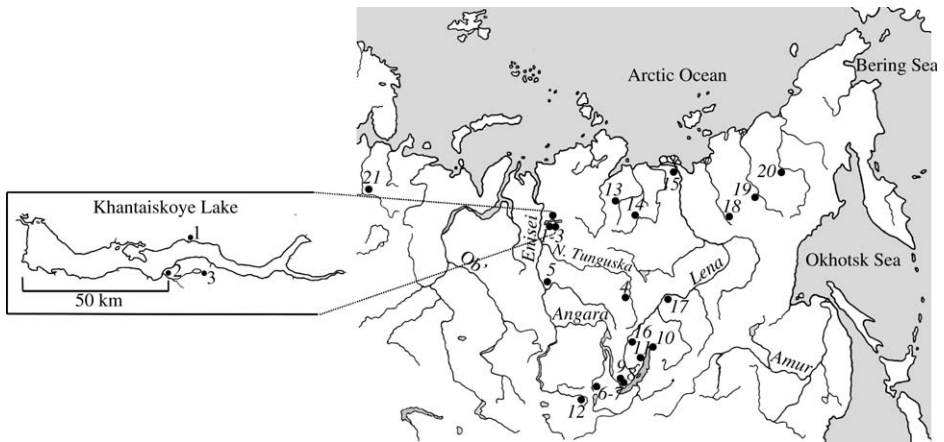


FIG. 1. Map of Siberia showing 21 sample locations (numbering as in Table I). Enlargement shows sample sites 1–3 in Khantaiskoye Lake.

The number of alleles per locus, allele frequencies, heterozygosity, allelic richness and deviations from Hardy–Weinberg expectations and genotypic linkage equilibrium were calculated with GENETIX 4.03 (Belkhir & Borsa, 1998) and FSTAT 2.9.3.2 (Goudet, 2002). To test the hypothesis of two reproductive units in Khantaiskoye Lake and considering the limited sample sizes for this region, three independent approaches were taken all of which assume no *a priori* membership of an individual to a taxon or sample site. Data were first analysed with the Bayesian clustering method in STRUCTURE 2.0 (Pritchard *et al.*, 2000). The posterior probabilities of K (number of populations) were estimated assuming uniform prior values of K between 1 and 6 with five replicates of 200 000 iterations following a burn-in of 20 000 iterations. An admixture model and independent allele frequencies were assumed. Second, a general relationship among all individuals in the dataset was assessed through the use of a factorial correspondence analysis (FCA) as computed in GENETIX. Third, pair-wise distances among individuals were calculated with an allele-sharing distance (D_{AS}) (Chakraborty & Jin, 1993) and a NJ tree of individuals was constructed using the programme POPULATIONS 1.2.28 (Langella, 2002). Based on the groups resolved with these analyses, summary statistics (e.g. allelic richness) and pair-wise F_{ST} and R_{ST} distances among groups are also provided using ARLEQUIN 2.0 (Schneider *et al.*, 2000).

RESULTS

MTDNA

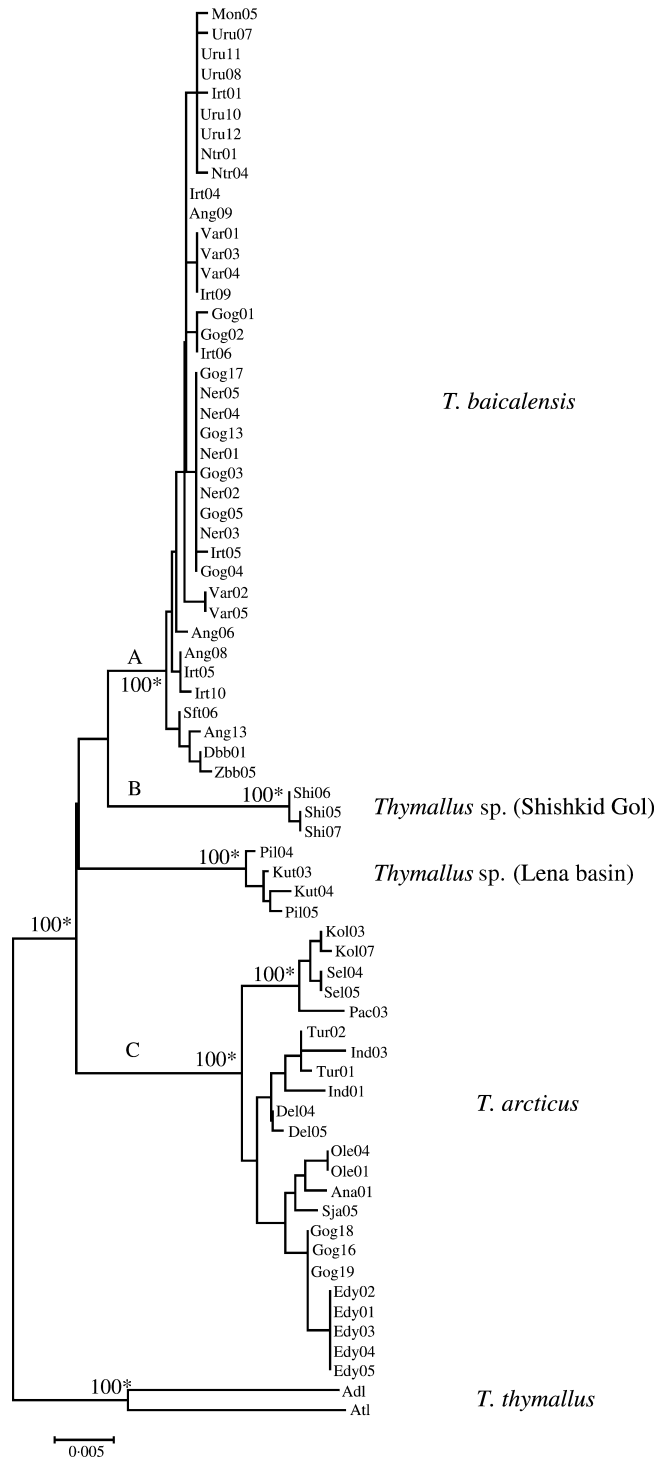
The alignment of 77 CR sequences included the entire control region (1006 bp) and partial segments of the tRNA proline (68 bp) and tRNA phenylalanine (15 bp) genes. There were 127 variable sites, of which 91 were parsimony informative and 20 indels. The NJ tree depicts three distinct clades (A–C), each with 100% bootstrap support, in samples from the Enisey basin (Fig. 2). The shortest MP tree was 199 steps and was nearly identical to the NJ tree and thus is represented by corresponding bootstrap values only (Fig. 2). All three lineages correspond to previously described clades first presented in Koskinen *et al.* (2002). Clade A corresponds to the shallow but diverse lineage representing

populations from Lake Baikal, as well as much of the Enisey basin sampled in this study, including samples from the Khantaiskoye Lake region. Populations from this region are traditionally assigned to either *T. a. arcticus* for fish in the Enisey basin or *T. a. baicalensis* for fish in Lake Baikal. Clade B represents a few samples from the Shishkid Gol, a river in the uppermost Enisey basin in Mongolia. Clade C corresponds to a lineage discussed in detail in Weiss *et al.* (2006) representing populations from the Urals in the west, to North America in the east, but herein including samples from the Khantaiskoye Lake region. Most populations of the Arctic coastal region of Siberia are traditionally assigned to *T. a. pallasii*. Thus, individuals of both clade A (sub-species *arcticus* and *baicalensis*) and C (sub-species *pallasii*) are found in the Khantaiskoye Lake basin, including the same site (Gogochenda River). This is the first location in Siberia where individuals assigned to these two lineages (A and C) are found in sympatry. An additional 10 individuals from the Khantaiskoye Lake region were assigned to these clades (data not shown) based on the sequencing of c. 550 bp of the 5' end of the CR; each of these partial sequences showed 100% identity with the 5' end of a haplotype in Fig. 2. Thus, all 16 individuals from Khantaiskoye Lake identified *a priori* by dorsal fin colour pattern as being *T. a. arcticus* were assigned to clade A and 12 of 14 individuals identified *a priori* as *T. a. pallasii* belonged to clade C. The remaining two individuals identified as *T. a. pallasii* revealed a haplotype grouping with clade A.

For the two major lineages (A and C), there were 22 haplotypes among 39 individuals for clade A (Baikal–Enisey) and 16 haplotypes among 23 individuals for clade C (Arctic coast, North America). The net mean divergence between these clades was 2.6%, with a total of 19 fixed differences between them, and the maximum pair-wise divergence reached 3.5%. Clade C exhibited more nucleotide diversity ($\pi = 0.0087$) than clade A ($\pi = 0.0028$). Clade B, for which no taxonomic assignment is given, exhibited a net divergence of 2.2% from clade A and 4.0% from clade C, with a maximum pair-wise divergence of 4.8%.

MICROSATELLITES

All seven microsatellites could be unambiguously genotyped among samples. The results of STRUCTURE revealed the largest change in the posterior probabilities of K (ΔK) between $K = 1$ and $K = 2$ (Ln Prob -1767 to -1390) (Evanno *et al.*, 2005). At $K = 2$ all samples from the Lake Khantaiskoye region exhibiting an mtDNA haplotype from clade C formed one group, which included the two individuals phenotypically identified as *T. a. pallasii* but carrying an mtDNA haplotype from clade A. The second group was formed by all individuals from Lake Khantaiskoye exhibiting a haplotype from clade A as well as samples from Varlamovka and Urunge Nur. Assignment probabilities (Q-values) were above 98% for all individuals at $K = 2$. Further iterations at $K = 3$ ($\Delta K = 231$) and $K = 4$ ($\Delta K = 59$) revealed a division in the clade A group, resolving Varlamovka and Urunge Nur populations. Assignment probabilities at $K = 4$ remained above 98% for all individuals in the Khantaiskoye Lake region, with no evidence of admixture between genotypes in the Khantaiskoye Lake region assigned to the two phenotypes. The only possibility of admixed genotypes (though with low Q-values) was seen for a few individuals



between Varlamovka and Urunge Nur sampling sites. At higher levels of K , ΔK was <20 , and no further structure was revealed.

Based on this structure (*i.e.* $K = 4$), representing the two mid to upper Enisey sample areas, and the two phenotypes within the Khantaiskoye Lake region, some summary analyses are provided. Mean allelic richness varied from 2.46 for *T. a. arcticus* from Khantaiskoye Lake and 5.62 for *T. a. arcticus* from the Varlamovka River (Table II). Eight of the 20 alleles (40%) found in *T. a. arcticus* from Khantaiskoye Lake were private to this population, some of which were either fixed or found at high frequencies (Table II). This resulted in highly significant differentiation estimates between *T. a. arcticus* from Khantaiskoye Lake and all other samples, including those found in sympatry (F_{ST} : 0.478–0.546; R_{ST} : 0.463–0.668). This divergence is clearly seen in the FCA diagram whereby all 14 individuals cluster together, highly distinct from all others, along the first and most explanatory axis (Fig. 3). There were low to moderate levels of pair-wise population differentiation (F_{ST} : 0.154–0.264; R_{ST} : 0.008–0.119) among the three other sample units. Only the R_{ST} values between Urunge and Varlamovka were not statistically significant. The robustness of this differentiation is also supported by the 100% clustering of individuals from each population in the tree of individuals as well as the longer branch length for the *T. a. arcticus* individuals from Lake Khantaiskoye (Fig. 4).

Thus, *a priori* assignment of all 30 individuals from the zone of sympatry (Lake Khantaiskoye) to two distinct phenotypes was fully concordant with the microsatellite data in presenting two genetic entities with no sign of interbreeding, despite the presence of two individuals showing discordance in their mtDNA.

DISCUSSION

Despite limited sample sizes in the zone of sympatry, the disparate allelic profiles at seven microsatellite loci (F_{ST} and $R_{ST} >0.5$) strongly suggest reproductive isolation of two lineages of grayling in the lower reaches of the Enisey basin. Moreover, the mtDNA clades are highly divergent and are well described over vast regions of Siberia (Koskinen *et al.*, 2002; Knizhin *et al.*, 2006b; Weiss *et al.*, 2006). Clade C is apparently fixed throughout the North American range of *Thymallus* (Stamford & Taylor, 2004; Weiss *et al.*, 2006). The two individuals identified as *T. a. pallasii* by phenotype and grouped with *T. a. pallasii* based on microsatellite loci but carried an mtDNA from clade A must be either the product of ancient introgression or more recent hybridization coupled with complete nuclear replacement *via* unidirectional backcrossing with males of *T. a. pallasii* (*i.e.* mtDNA capture), a common phenomena in fishes (Sušnik *et al.*, 2007 and references therein). The small number of sampling locations limits inferences about the level of reproductive isolation at other sites in the Khantaiskoye Lake catchment or in other areas of

FIG. 2. A neighbour-joining (NJ) tree of complete mtDNA control region haplotypes for *Thymallus* sp. based on Kimura two-parameter distances. Nodes with 100% bootstrap support (1000 replicates) from both NJ and maximum parsimony approaches are labelled '100*'. Clades A, B and C contain individuals captured in the Enisey basin. New haplotypes from present study are shown in bold. The tree is rooted with two haplotypes from European grayling *Thymallus thymallus*.

TABLE II. Sample sizes (n), allelic richness (A_R), expected (H_{exp}) and observed (H_{obs}) heterozygosity, allele frequencies (A_N) and the number of alleles per locus (A_N) for each of the four sample clusters, as identified in STRUCTURE. Lake Khantaiskoye samples are separated into the two groups proposed to represent two separate species. Private alleles of *T. arcticus* are given in bold

	Lake Khantaiskoye		Urunge Nur	Varlamovka River
	<i>T. arcticus</i>	<i>T. baicalensis</i>	<i>T. baicalensis</i>	<i>T. baicalensis</i>
n	14	16	28	9
A_R	2.46	4.69	2.84	5.62
H_{exp}	0.318	0.681	0.450	0.705
H_{obs}	0.265	0.598	0.428	0.663
Locus/allele	Allele frequencies			
Str85				
171	0	0	0.02	0
173	0	0.13	0.24	0.07
175	0	0	0	0.29
177	0	0	0	0.07
179	0	0.72	0.72	0.36
181	0	0	0.02	0
193	0.89	0	0	0
195	0.11	0	0	0.14
207	0	0	0	0.07
211	0	0.16	0	0
A_N	2.00	3.00	4.00	6.00
One2				
185	0	0	0	0.06
187	0	0	0	0.06
199	0	0.09	0	0
201	0	0.09	0	0
203	0	0	0.54	0.39
205	0.04	0.06	0	0.11
207	0.96	0.25	0.16	0.11
209	0	0.09	0	0.11
211	0	0.19	0.27	0
213	0	0	0	0.11
217	0	0	0.04	0
231	0	0.03	0	0
233	0	0.19	0	0.06
A_N	2.00	8.00	4.00	8.00
Thy18				
231	0	0	0.02	0
243	0	0.22	0	0
245	0.11	0.47	0.11	0
246	0	0	0.88	0.83
247	0.04	0.16	0	0
248	0	0	0	0.06
250	0	0	0	0.11

TABLE II. Continued

	Lake Khantaiskoye		Urunge Nur	Varlamovka River
	<i>T. arcticus</i>	<i>T. baicalensis</i>	<i>T. baicalensis</i>	<i>T. baicalensis</i>
263	0.25	0.06	0	0
265	0.61	0.09	0	0
A_N	4.00	5.00	3.00	3.00
Coc123				
195	0.36	0	0	0
233	0.07	0	0	0
235	0	0.03	0	0
239	0.14	0	0	0
241	0.29	0	0	0
242	0	0	0	0.13
244	0	0	0	0.13
245	0	0.09	0	0
246	0	0	0	0.13
247	0.07	0.38	0.06	0
249	0.04	0.28	0.67	0.06
250	0	0	0	0.25
252	0	0	0	0.06
254	0	0	0.26	0
256	0	0	0	0.19
259	0	0.13	0.02	0
261	0	0.03	0	0
265	0.04	0.06	0	0
268	0	0	0	0.06
A_N	7	7	4	8
Thy72				
131	1.00	0	0	0
142	0	0.41	0.04	0.31
143	0	0.53	0.96	0.31
145	0	0.06	0	0.13
146	0	0	0	0.19
148	0	0	0	0.06
A_N	1.00	3.00	2.00	5.00
Thy2/2				
155	0	0	0.02	0
159	0.25	0.03	0	0
163	0	0.06	0	0.17
165	0	0	0.38	0
167	0	0.19	0	0
169	0	0.03	0	0.17
171	0	0.13	0	0.11
173	0.43	0.13	0.46	0.22
175	0	0.38	0	0.33
183	0	0.03	0.14	0
185	0	0.03	0	0

TABLE II. Continued

	Lake Khantaiskoye		Urunge Nur	Varlamovka River
	<i>T. arcticus</i>	<i>T. baicalensis</i>	<i>T. baicalensis</i>	<i>T. baicalensis</i>
191	0·32	0	0	0
A_N	3·00	9·00	4·00	5·00
Thy54				
173	0	0·31	0·13	0·07
175	0	0·41	0	0·43
177	0	0·13	0·30	0
179	0	0	0	0·14
181	1·00	0·03	0·41	0·07
182	0	0	0·11	0
183	0	0	0·04	0·07
184	0	0	0·02	0
189	0	0	0	0·21
191	0	0·03	0	0
193	0	0·09	0	0
A_N	1·00	6·00	6·00	6·00

sympatry. Local biologists report that the two taxa exist in other tributaries of Lake Khantaiskoye and in at least one other nearby lake system (Kutaramakam), a few kilometres to the north-east. Elsewhere in the Enisey basin, from Varlamovka River (site 5; Fig. 1) upstream up to and including Lake Baikal and its tributaries (sites 6–11), no individual carrying a clade C haplotype has been found. This assessment includes a large dataset for Lake Baikal (S. Weiss, unpubl. data). Outside the Enisey, in all other Arctic draining rivers sampled (sites 13–15 and 18–21), no individual from clade A has been found (Weiss *et al.*, 2006). As the area around Khantaiskoye and Kutaramakan lakes is remote as well as the several hundred kilometres of basin downstream to the Enisey

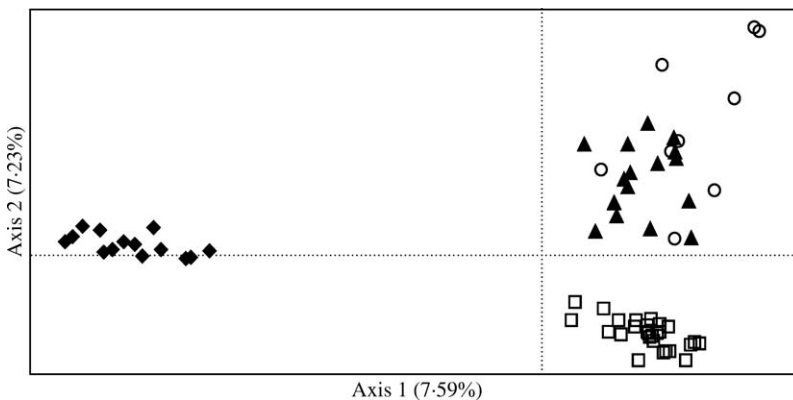


FIG. 3. Bi-variate plot of the first two factors of a factorial correspondence analysis of seven microsatellite allele frequencies scored in 67 individuals from the Enisey basin. Symbols [◆, Khantaiskoye (*T. arcticus*); ▲, Khantaiskoye (*T. baicalensis*); □, Urunge Nur; ○, Varlamovka] denote individuals from the four clusters identified with STRUCTURE.

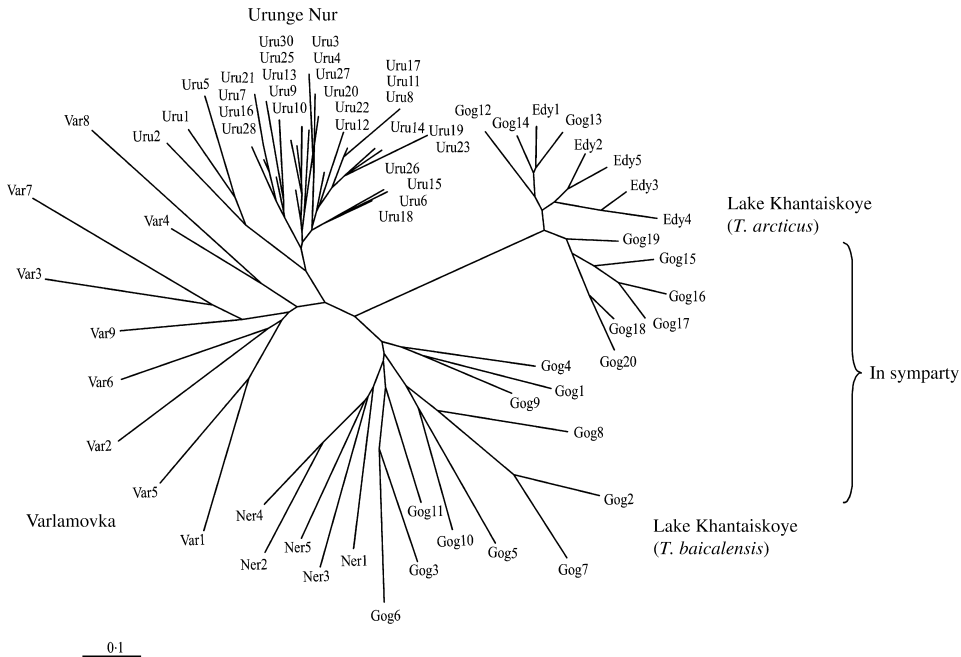


FIG. 4. Neighbour-joining tree of individuals ($n = 67$) based on pair-wise D_{AS} distances (shared proportion of alleles) across seven microsatellite loci. Three-letter samples codes as in Table I, whereby Ner, Gog and Edy are separated by only a few kilometres in Lake Khantaiskoye.

delta, other unrecorded sympatric populations may exist. The cyto-nuclear and cyto-phenotypic discordance seen in two individuals presumably stems from historical introgression and capture, a common phenomenon in fishes (Sušnik *et al.*, 2007 and references therein).

The stark divergence seen at microsatellite loci, best expressed in that 40% of the alleles scored in *T. a. pallasii* were private, many at high frequencies, underscores the robustness of this conclusion. The analytical methods applied to test this hypothesis (FCA, tree of individuals and mtDNA tree reconstruction) were all blind to this *a priori* labelling, and thus the labelling serves only to demonstrate that the taxa were phenotypically recognizable prior to genetic analysis. The use of dorsal fin colour patterns for the recognition of distinct lineages of Arctic grayling in Siberia has been promoted by several Russian ichthyologists (Makoedov, 1999; Chereshev *et al.*, 2002; Knizhin *et al.*, 2006a) and supports the speculation that mate choice based on recognition of this character may be an isolating or re-enforcing mechanism promoting speciation in *Thymallus*. These colour patterns, however, are difficult to describe and await future quantification. As mtDNA haplotypes from these two lineages can be assigned to two relatively widespread and highly divergent clades in Siberia, the lower Enisey appears to be a zone of secondary contact, rather than *in situ* speciation. Reproductively isolated lineages in sympatry cannot be recognized as two sub-species, notwithstanding the highly subjective and mosaic distribution of *T. a. arcticus*, and thus this finding allows us to offer a clear and much needed taxonomic revision.

The distribution of clade C (Fig. 2) is reported to extend from the Urals in the west, throughout the Siberian Arctic coastal region and includes perhaps all grayling populations in North America (Stamford & Taylor, 2004; Weiss *et al.*, 2006). Based on the present results, as well as those of Koskinen *et al.* (2002), Froufe *et al.* (2005) and Weiss *et al.* (2006), the following conclusions are drawn: (1) the sub-specific designations of Arctic grayling along the Siberian Arctic coast should be abandoned in favour of the name *T. arcticus* and (2) the distribution of *T. arcticus*, though still large, is excluded from much of interior Siberia, most probably due to the existence of other taxa.

For the Enisey basin, this implies that *T. arcticus* is excluded from all but the lowermost reaches by a taxon represented genetically herein by clade A (Fig. 2). This lineage, based on the geographic distribution of mtDNA haplotypes as well as dorsal fin patterns is distributed throughout most of the Enisey basin including Lake Baikal and its major tributary, the Selenga River (Knizhin *et al.*, 2006b). While populations in Lake Baikal are currently recognized by some researchers as *T. a. baicalensis* (Dorofeeva, 2002; Knizhin *et al.*, 2006c, d), they have been previously recognized as a full species *Thymallus baicalensis* Dybowski, 1874 by Grazianov (1902) and Pivnička & Hensel (1978). A recent species list of Russian fishes (Bogutskaya & Naseka, 2004) includes not only *T. baicalensis* but also a second taxon *Thymallus brevipinnis* Svetovidov, 1931 for Lake Baikal. Presently, neither genetic nor phenotypic support the existence of a second taxon of grayling in Lake Baikal or its main tributary Selenga River (Koskinen *et al.*, 2002; Knizhin *et al.*, 2006b; Weiss *et al.*, unpubl. data). The recognition of *T. baicalensis* for grayling in Lake Baikal, populations in the Selenga River basin of Mongolia and in most of the Enisey basin, resolves the problem of the mosaic distribution of a sub-species throughout central Siberia and strictly follows an evolutionary perspective as all these populations appear to belong to one large shallow clade, all possess some common phenotypic features and the lineage shows no signs of introgression with other proposed taxa in zones of contact (Koskinen *et al.*, 2002; Froufe *et al.*, 2005; Knizhin *et al.*, 2006c, d; Weiss *et al.*, unpubl. data and data herein).

Haplotypes from the Shishkid Gol River in Mongolia (clade B) cannot presently be assigned to a taxon with any degree of confidence. Existing phenotypic data (Pivnička & Hensel, 1978; Shatunovskiy, 1983) are not comparable with the authors' data, and phylogenetic analyses place this clade as a sister group to Mongolian grayling *T. brevirostris* (Koskinen *et al.*, 2002; Froufe *et al.*, 2005). Thus, the remote areas of the uppermost Enisey basin may indeed contain a third taxon awaiting formal description.

The integration of morphological and genetic approaches to the description and understanding of diversity in Eurasian *Thymallus* has repeatedly shown that the current systematics is inadequate or wholly misleading. This small but key report identifies a contact zone for two putative sub-species of Arctic grayling (*T. a. pallasii* and *T. a. arcticus*) and strongly suggests that they are reproductively isolated as samples in sympatry display disparate microsatellite allele profiles. Furthermore, as *T. a. arcticus* of the Enisey and grayling found in Lake Baikal (*i.e.* *T. a. baicalensis*) appear to be synonymous, a systematic revision following evolutionary principles is permitted.

This study was funded, in part, by the Russian Foundation of Basic Research, 06-04-90611-BNTS-9.

References

- Antonov, A. L. (2004). A new species of grayling *Thymallus burejensis* sp. nov. (Thymallidae) from the Amur Basin. *Journal of Ichthyology* **44**, 401–411.
- Berg, L. S. (1948). *Freshwater Fishes of the USSR and Adjacent Countries*, 4th edn, Vol. 1. Moscow: Akad. Nauk SSSR (in Russian).
- Bogutskaya, N. G. & Naseka, A. M. (2004). *Catalog of Agnathans and Fishes of Fresh and Brackish Waters of Russia with Comments on Nomenclature and Taxonomy*. Moscow: T-vo Nauch. Izd. KMK (in Russian).
- Chakraborty, R. & Jin, L. (1993). Estimation of genetic distance and coefficient of gene diversity from single-probe multilocus DNA fingerprinting data. *Molecular Biology and Evolution* **11**, 120–127.
- Chereshnev, I. A., Volobuyev, V. V., Shestakov, A. V. & Frolov, S. V. (2002). *Lososevidniye ribi severo-vostoka Rossii*. Vladivostok: Dalnauka Press (in Russian).
- Cuvier, G. & Valenciennes, A. (1848). *Histoire naturelle des poissons. Tome vingt et unième. Suite du livre vingt et unième et des Clupéoides. Livre vingt-deuxième. De la famille des Salmonoides*. Paris: Bertrand.
- Dorofeeva, E. A. (2002). "The Genus *Thymallus*". In *Atlas of Freshwater Fish of Russia* (Reshetnikov, Y. S., ed.), pp. 163–169. Moscow: Nauka (In Russian).
- Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611–2620.
- Froufe, E., Knizhin, I., Koskinen, M., Primmer, C. R. & Weiss, S. (2003). Identification of reproductively isolation lineages of Amur grayling (*Thymallus grubii* Dybowski 1869): concordance between phenotypic and genotypic variation. *Molecular Ecology* **12**, 2345–2355.
- Froufe, E., Knizhin, I. & Weiss, S. (2005). Phylogenetic analysis of the genus *Thymallus* (grayling) based on mtDNA control region and ATPase 6 genes, with inferences on control region constraints and broad-scale Eurasian phylogeography. *Molecular Phylogenetics and Evolution* **34**, 106–117.
- Grazianov, V. O. (1902). Ikhtiofauna Baikala. *Izvestiya obshestva lyubitelei estestvoznaniya, antropologii i etnographii* **97**, 18–61 (in Russian).
- Knizhin, I., Weiss, S., Antonov, A. L. & Froufe, E. (2004). Morphological and genetic diversity of Amur graylings (*Thymallus*; Thymallidae). *Journal of Ichthyology* **44**, 59–76.
- Knizhin, I. B., Kirillov, A. F. & Weiss, S. J. (2006a). On the diversity and taxonomic status of graylings (*Thymallus*; Thymallidae) in the Lena River. *Journal of Ichthyology* **46**, 234–246.
- Knizhin, I. B., Weiss, S. J. & Sušnik, S. (2006b). Graylings of Baikal Lake Basin (*Thymallus*, Thymallidae): diversity of forms and their taxonomic status. *Journal of Ichthyology* **46**, 418–435.
- Knizhin, I. B., Weiss, S. J., Bogdanov, B. E., Samarina, S. S. & Froufe, E. (2006c). Finding a new form of the grayling *T. arcticus* (Thymallidae) in the basin of Lake Baikal. *Journal of Ichthyology* **46**, 34–43.
- Knizhin, I. B., Bogdanov, B. E. & Vasileva, E. A. (2006d). Biological and morphological characteristics of the Arctic grayling *Thymallus arcticus* (Thymallidae) from alpine lakes of the upper reaches of the Angara River. *Journal of Ichthyology* **46**, 709–721.
- Koskinen, M. T., Knizhin, I., Primmer, C. R., Schlötterer, C. & Weiss, S. (2002). Mitochondrial and nuclear DNA phylogeography of *Thymallus* spp. (grayling) provides evidence of ice-age mediated environmental perturbations in the world's oldest body of freshwater, Lake Baikal. *Molecular Ecology* **11**, 2599–2611.
- Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150–163.

- Makoedov, A. N. (1999). *Relations of Graylings of Siberia and Far East*. Moscow: UMK Psikhologiya (in Russian).
- Miller, S. A., Dykes, D. D. & Polesky, H. F. (1988). A simple salting out procedure from human nucleated cells. *Nucleic Acids Research* **16**, 1215.
- Pivnička, K. & Hensel, K. (1978). Morphological variation in the genus *Thymallus* Cuvier, 1829 and recognition of the species and subspecies. *Acta Universitatis Carolinae-Biologica* 1975–1976 **4**, 37–67.
- Pritchard, J. K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
- Romanov, V. I. (2005). Фауна, систематика и биология рыб в условиях озерно-реки гидросистем южного Тaimira. Автореф. дисс. докт. биол. наук. Томск.: Томский госуд. Университет (Dissertation, in Russian).
- Shatunovskiy, M. I. (Ed.) (1983). *Fishes of Mongolian People's Republic*. Moscow: Nauka Press (in Russian).
- Skurikhina, L. A. (1984). Geneticheskaya divergentsiya khariusov (rod *Thymallus* Cuvier, 1829) Evrazii v svete dannikh molekulyarnoi gibrizatsii DNKxDNK. Candidate's Dissertation in Biology, Mosk. Gos. Univ., MSU, Moscow (in Russian).
- Stamford, M. D. & Taylor, E. B. (2004). Phylogeographic lineages of Arctic grayling (*Thymallus arcticus*) in North America: divergence, origins and affinities with Eurasian *Thymallus*. *Molecular Ecology* **13**, 1533–1549.
- Sušnik, S., Weiss, S., Odak, T., Dellling, B., Treer, T. & Snoj, A. (2007). Reticulate evolution: ancient introgression of the Adriatic brown trout mtDNA in softmouth trout *Salmo obtusirostris* (Teleostei: Salmonidae). *Biological Journal of the Linnean Society* **90**, 139–152.
- Svetovidov, A. N. (1936). European-Asiatic Graylings (the Genus *Thymallus* Cuvier). *Trudi Zoologicheskogo Instituta Akademia Nauk SSSR* **3**, 183–301 (in Russian).
- Swofford, D. L. (2001). *PAUP* ver 4.0.b3a. Phylogenetic Analysis Using Parsimony and Other Methods*. Sunderland, MA: Sinauer Associates.
- Uiblein, F., Jagsch, A., Honsig-Erlenburg, W. & Weiss, S. (2001). Status, habitat use, and vulnerability of the European grayling (*Thymallus thymallus*) in Austrian waters. *Journal of Fish Biology* **59**, 223–247.
- Weiss, S., Knizhin, I., Kirillov, A. & Froufe, E. (2006). Phenotypic and genetic differentiation of two major phylogeographic lineages of arctic grayling *Thymallus arcticus* in the Lena River, and surrounding Arctic draining basins. *Biological Journal of the Linnean Society* **88**, 511–525.

Electronic References

- Belkhir, K. & Borsa, P. (1998). *GENETIX, logiciels sous Windows™ pour la génétique des populations*. Laboratoire Génome et Populations, CNRS UPR 9060, Université Montpellier II, Montpellier, France. Available at <http://univ-montp2.fr/~genetix/genetix.htm>
- Goudet, J. (2002). *FSTAT, Version 2.9.3.2: a Program to Estimate and Test 783 Gene Diversities and Fixation Indices*. Available at <http://www2.unil.ch/popgen/software/fstat.htm>
- Langella, O. (2002). *POPULATIONS Version 1.2.26*. Available at <http://www.pge.cnrs-gif.fr/bioinfo/populations>
- Schneider, S., Roessli, D. & Excoffier, L. (2000). *Arlequin ver.2.0. A Software for Population Genetics Analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland. Available at <http://lgb.unige.ch/arlequin>