Timing of the population dynamics of bullhead *Cottus gobio* (Teleostei: Cottidae) during the Pleistocene

F. A. M. VOLCKAERT*, B. HÄNFLING†, B. HELLEMANS* & G. R. CARVALHO†

Laboratory of Aquatic Ecology, Katholieke Universiteit Leuven, Leuven, Belgium
†Molecular Ecology and Fisheries Genetics Laboratory, Department of Biological Sciences, University of Hull, Hull, UK

Abstract

Pleistocene genetic structure of the bullhead, *Cottus gobio,* was evaluated across the western Palearctic using a 771-bp long fragment of the mitochondrial control region in 123 individuals collected at 35 sites (data set I). In total, 59 haplotypes that differed at 73 positions (9.3%) were detected. Data analysis also included sequences from Englbrecht *et al.* (2000), thus increasing the sampling to a more comprehensive data set of 529 fish and 63 control region sequences of 482 bp (data set II). A minimum spanning and phylogenetic tree identified a seventh clade (Brittany–Loire) in addition to the previously identified six clades. The geographical range of the North Sea and Lower Rhine clades was considerably larger than thought previously. Haplotype diversity was generally low, and the total fixation index high ($F_{ST} = 0.49$). Among-group differentiation accounted for 41.7% (data set I) of the variation. Contiguous range expansions and restricted gene flow combined with isolation by distance, interspersed with past fragmentation characterize bullhead across its range. New is the knowledge that dated interglacial periods correlated with population expansions; river captures, proglacial lake systems and sea level played a significant role in the dispersal and expansion either in northern or southern direction. Hence it became possible to identify and date the colonization routes and putative palaeorefugia, most of which were located in Central and North-west Europe. Glacial periods resulted in distinct fragmentation events and lineage sorting.

Introduction

The evolutionary legacy of the Cenozoicum, incorporating such events as vicariance and palaeoclimate, can be traced back in the genome of a wide variety of terrestrial and aquatic organisms (Avise, 2000). It resulted in an identifiable impact on species richness, abundance, genetic diversity and structure. The phylogeographical approach (Avise *et al.*, 1987) provides insights into the determinants of current distribution and abundance, and a powerful method for investigating the nature and efficacy of evolutionary processes. For example, phylogeographical analyses have shown that terrestrial animals responded across Europe in three general ways to the glacial recession over the past 20 000 years (Hewitt, 2000). Species belonging to the 'hedgehog' pattern colonized from three refugia (Balkan, Iberic and Italian peninsula) to Central and Western Europe, those following the 'bear' pattern did so from the Iberic and Caucasian refugia, and those following the 'grasshopper' pattern colonized from a single refugium in the Balkans. In addition to the vast amount of information provided on Europe's terrestrial animals from phylogeographical studies, there is a steadily increasing knowledge on the Pleistocene history of European freshwater ichthyofauna from genetic studies (Nesbø *et al.*, 1999; Bernatchez, 2001; Brunner *et al.*, 2001). The dispersal of aquatic

Correspondence: Dr Filip Volckaert, Katholieke Universiteit Leuven, Laboratory of Aquatic Ecology, Ch. de Bériotstraat 32, B-3000 Leuven, Belgium.
Tel.: +32 16 32 39 66; fax: +32 16 32 45 75;
e-mail: filip.volckaert@bio.kuleuven.ac.be

organisms is constrained by the availability of suitable river and lake habitats. Hence aquatic organisms might show a phylogeographical pattern, which is – or which is not – distinct from their terrestrial counterparts. River captures and connections throughout geological time might provide both the necessary migration routes for colonization and for retreat during deteriorating environmental conditions. Most fish, for example, were well established in Europe by the Late Miocene (Banarescu, 1990, 1992), including bullhead (Heizmann, 1992). Their distribution, however, became subject to the continuous remodelling of the river drainage (Gibbard, 1988) and to climatic oscillations, of which the glacial⁄interglacials of the Late Pleistocene played a crucial role (Banarescu, 1990, 1992). Periods of above or below average temperatures and precipitation levels, as well as periods with strong climate fluctuations, have affected the viability of populations as well as their chances for expansion or reduction.

Here, the population structure and Pleistocene history of a small secondary freshwater fish living in the western Palearctic is analysed. Despite a benthic life style with paternal brood care (Marconato & Bisazza, 1988) and limited individual dispersal ability (Downhower & Brown, 1979; Waterstraat, 1992), the colonization potential of the bullhead, *Cottus gobio* L. (Cottidae, Teleostei), is considerable as evidenced by its wide distribution. Bullhead have, for example, recolonized Scandinavia across an approximately 1500 km range within a 10 000-year period (Kontula & Väinölä, 2001). Its current range fits between the maximum July isotherm of 20 °C and the minimum January isotherm of −20 °C. In addition, it inhabits rheophilic well-oxygenated upland rivers and the surf zone of lakes of the western Palaearctic (Lelek, 1980; Wanzenböck *et al.*, 2000).

Previous allozyme studies have shown that *C. gobio* has experienced considerable population fragmentation during the Pleistocene (Riffel & Schreiber, 1995; Hänfling & Brandl, 1998a,b; Eppe *et al.*, 1999). In a recent study based on mtDNA sequences, Englbrecht *et al.* (2000) identified six major clades or haplotype groups across the European range of *C. gobio*. They proposed that the founder population of *C. gobio* inhabited the Parathetys Sea some 10 ma BP, invaded Europe during the Late Pliocene, and colonized Central and Northern Europe along the Danubian – Upper Rhenian – Elbe – Baltic axis (see also Kontula & Väinölä, 2001). During the Late Pliocene a Central European lineage must have reached the north-eastern North Sea along a northern route (Dnjepr and Baltic plain) and has colonized the East-Atlantic region south-west to northern Spain. However, because of the restricted spatial sampling of the previous studies, it remained unclear whether all bullhead clades had been identified to the west of the rivers Vistula and Danube, and what the full distribution was of the clades. Additional contentious issues included

the timing of the population expansions and reductions caused by the uncertain calibration of the molecular clock, and the identification of the various colonization routes in the western Palearctic during the Late Pleistocene.

A phylogeographical analysis (as measured at a 771-bp and a smaller 482 bp fragment of the mitochondrial control region) was performed here on a regional scale to address the colonization, radiation and extinction of *C. gobio* using inferences in a multidisciplinary historic perspective.

Materials and methods

Collection of samples

Juvenile and adult bullhead were collected across Western Europe (UK, Belgium, France and the Netherlands) and Central Europe (Slovakia and Germany) at representative sites with electrofishing (Table 1 and Fig. 1). Previous personal research had indicated these areas to harbour phylogeographically informative populations. In total, 35 (labelled V1–V52) rivers and lakes were sampled between 4°W and 30°E and 47–62°N (data set I). Biopsies (fin clips) were collected and preserved in either 100% ethanol or a solution of salt (NaCl) saturated dimethylsulphoxide.

DNA extractions

The tissue was incubated overnight in cetyltrimethylammonium bromide (CTAB)-buffer [2% CTAB, 1.4 M NaCl, 0.2% β-mercaptoethanol, 20 mM ethylenediaminetetraacetic acid (EDTA) and 100 mM Tris pH 8.0] at 55 °C (Winnepenninckx *et al.*, 1993). DNA, extracted with phenol-chloroform, was precipitated with ethanol, followed by resuspension in 50 μL distilled water.

Sequencing of the control region

An mtDNA control region fragment of 1.0 kb was amplified using the HN20 (Bernatchez & Danzmann, 1993) and the Pro19-primers (Bernatchez *et al.*, 1995), located in the phenylalanine tRNA-gene and the proline tRNA-gene, respectively. The reaction product was purified with the 'GFX PCR DNA and Gel Band Purification' kit (Amersham Biosciences, Little Chalfont, UK) and eluted in 50 μL 10 mM Tris-Cl (pH 8.0). Samples were cloned into the PCR2.1-TOPO vector (Invitrogen, Paisley, UK) and sequenced with standard M13 primers using the 'SequiTherm Excell II' kit (Epicentres Technologies, Madison, OH, USA) according to the protocol of the producer. The reaction products (771 bp) were analysed on an automatic sequencer 'LI-COR GeneReadIR DNA system' with the AlignIR software (LI-COR).

Table 1 Sample code, catchment, river, geographical location of sampling sites and control region (771 bp) haplotypes as identified in each population of *Cottus gobio* (data set I). Between brackets are given the number of individuals sharing a given haplotype.

Code	Catchment	River	Longitude	Latitude	Haplotype
V1	Danube	Ulicka	22°27′E	48°57′N	HV3 (2), HV4 (1)
V2	Danube	Turiec	18°51′E	48°54′N	HV1 (1), HV2 (1)
V4	Danube	Otterbach	12°15′E	49°01′N	HV5 (3)
V12	Scorff	Scorff	3°24′W	47°57′N	HV6 (2), HV7 (1)
V13	Elbe	Enziusbach	11°44′E	50°12′N	HV8 (1), HV9 (1), HV10 (1), HV11 (1)
V14	Nybroan	Kullean	13°52′E	55°31′N	HV12 (1), HV13 (1)
V15	Rhine	Thiemitz	11°32′E	50°04′N	HV14 (2), HV15 (1)
V17	Rhine	Aufsess	11°15′E	49°50′N	HV16 (1), HV17 (1)
V20	Rhine	Ketelmeer	5°44′E	52°37′N	HV18 (3), HV21 (1), HV22 (1), HV23 (1), HV24 (1), HV25 (1), HV26 (1), HV27 (1)
V21	Rhine	IJsselmeer	5°25′E	52°43′N	HV18 (3), HV19 (1), HV20 (1), HV26 (1), HV56 (1)
V22	Maas	Perlbach	6°14′E	50°33′N	HV28 (1)
V23	Maas	Veurs	5°49′E	50°45′N	HV29 (7)
V24	Maas	Aisne	5°32′E	50°23′N	HV30 (2)
V25	Maas	Lesse	5°10′E	49°60′N	HV31 (3)
V26	Maas	Solre	4°05′E	50°12′N	HV31 (2)
V27	Scheldt	Witte Nete	5°06′E	51°14′N	HV21 (2), HV32 (2)
V28	Scheldt	Molenbeek-Zenne	4°18′E	50°44′N	HV34 (8)
V29	Scheldt	Steenputbeek	4°17′E	50°43′N	HV33 (7)
V30	Scheldt	Molenbeek-Zwalm	3°44′E	50°47.5′N	HV35 (2)
V31	Scheldt	Molenbeek-Maarkebeek	3°41′E	50°48′N	HV41 (2)
V32	Scheldt	Piétrebais	4°44′E	50°45′N	HV36 (3)
V33	Scheldt	Trouille	4°00′E	50°24′N	HV37 (1), HV38 (2)
V34	Scheldt	St-Georges	3°31′E	50°15′N	HV39 (1), HV40 (1), HV41 (1), HV42 (1), HV43 (1), HV44 (1)
V35	Aa	Aa	2°03′E	50°34′N	HV45 (3)
V36	Humber	Rye	1°03′W	54°14′N	HV46 (2)
V37	Ribble	Ribble	2°21′W	53°54′N	HV41 (1), HV47 (1), HV50 (1)
V38	Ribble	Hodder	2°31′W	53°55′N	HV41 (1), HV48 (1), HV49 (1)
V39	Dee	Alwen	3°27′W	53°00′N	HV41 (1), HV50 (4), HV51 (1)
V40	Dee	Terrig	3°07′W	53°07′N	HV52 (3)
V41	Teifi	Dulas	4°04′W	52°07′N	HV52 (1), HV53 (1), HV54 (1)
V42	Towy	Cennen	3°59′W	51°52′N	HV50 (2)
V43	Neet	Neet	4°30′ W	50°47′N	HV55 (1), HV56 (2)
V45	Caerhayes	Caerhayes	4°50′W	50°17′N	HV57 (3)
V51	Puruvesi	Lake Saimaa	29°40′E	61°55′N	HV58 (2)
V52	Ylä-Keitele	Keihärinkoski	25°35′E	63°05′N	HV59 (1)

Analysis of haplotypes

Alignment of sequences was undertaken using the CLU-STALW, version 1.7 software (Thompson *et al.*, 1994). Two data sets were analysed: a full length control region sequence of 771 bp from the newly collected samples (data set I – labelled HV) and a shorter control region sequence (482 bp) from a comprehensive data set whose nucleotides are shared between the published data of Englbrecht *et al.* (2000) (523 bp – labelled HE) and this study (data set II). The haplotypes of the short fragments (482 bp – data set II) were relabelled and named according to the most common haplotype found by Englbrecht *et al.* (2000) or in this study (in total 63 HE and HV haplotypes). Mitochondrial DNA polymorphism was estimated as nucleotide (π; Nei, 1987) and haplotype diversity (h; Nei & Tajima, 1981) and within drainage and among *post hoc* defined haplotype groups

using the program ARLEQUIN, version 2.0 (Schneider *et al.*, 2000).

In a first step we used Modeltest 1.05 to select the model of DNA evolution that best fits data set I (771 bp fragment including as outgroup *C. cognatus*) based on log likelihood scores (Posada & Crandall, 1998). The Hase-gawa–Kishino–Yano model (Hasegawa *et al.*, 1985), with a gamma distribution shape parameter of 0.816, a proportion of invariable sites of 0.745 and base frequencies of A: 0.313, C: 0.217, G: 0.179 and T: 0.291 proved to give the best fit. We also compared the base composition for all sequences using a 5% χ^2-test on the average composition (PUZZLE; Strimmer & von Haeseler, 1996). The molecular-clock hypothesis was tested assuming the TrN model (Tamura & Nei, 1993) with γ-distributed rates across sites, with the likelihood ratio test for the clock hypothesis implemented in PUZZLE. Gaps were treated as missing data.
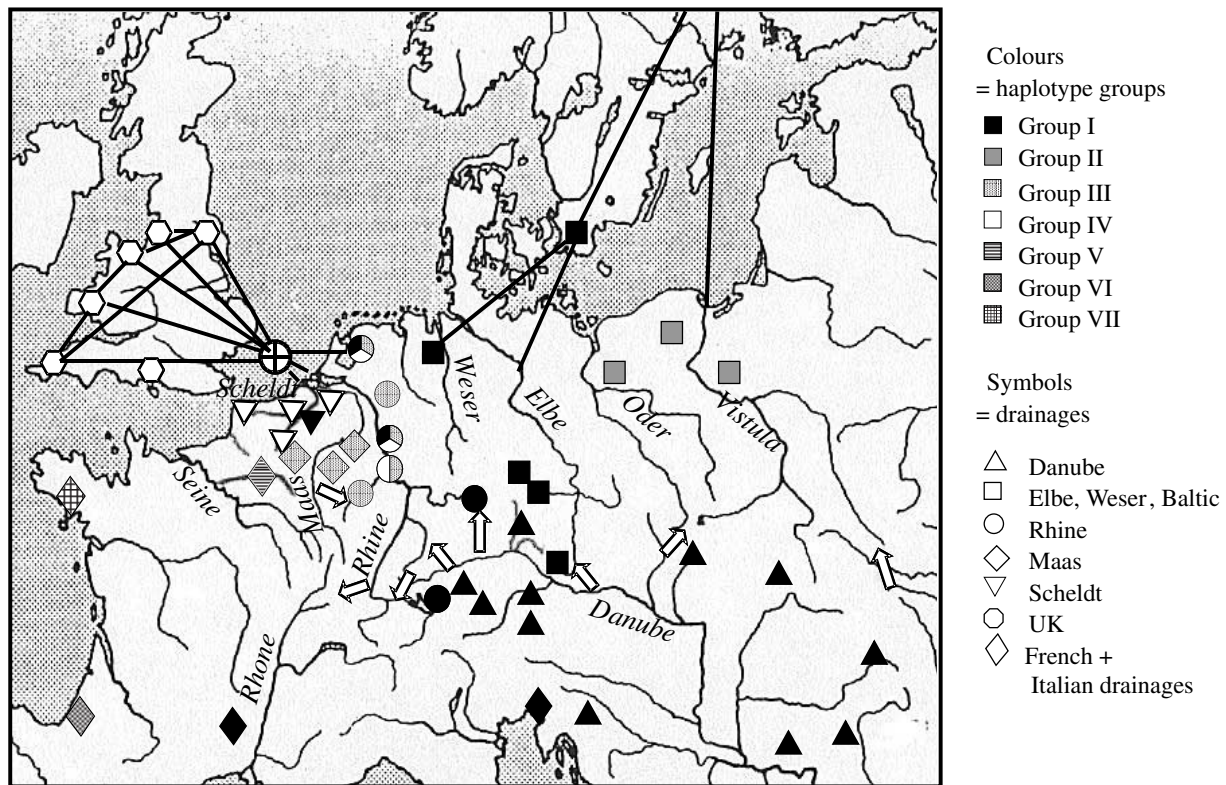
Fig. 1 Geographical distribution of sampling sites and major haplotype groups. Drainage or geographical areas are coded as symbols, major haplotype groups are coded as grey shades (see legend). Note that marks may represent a group of sample sites. The geographical locations of all sample sites can be found in Table 1. The map also shows the connections used for calculating the geographical distances for the nested clade analysis (NCA). Within drainage river distances were used, among drainage information on historic connections (arrows) or Euclidean distances (straight lines).

To infer a phylogeny of the 771 bp fragment, we used maximum parsimony (MP), maximum likelihood (ML) and distance-based methods. With MP the following model was used: weighted parsimony with all characters unordered, gaps treated as missing data, transition : transversion (ti / tv) ratios of 1.84 : 1.00 and transversion parsimony (ti / tv weight 0 : 1). The ML analysis was performed using the parameters estimated under the best-fit model (quartet method). With MP and ML we conducted the exhaustive search method and bootstrapped (n = 1000 or 100, respectively) with the branch and bound, and heuristic algorithm, respectively. A neighbour joining (NJ) phenogram of pair-wise distances (Kimura-2 parameter) between haplotypes was also prepared. Analyses were performed with PAUP vs. 4.0b4a (Swofford, 1998), PUZZLE vs. 4.0.2. (Strimmer & von Haeseler, 1996) and MEGA vs. 2.1 (Kumar *et al.*, 2001). The slimy sculpin, *C. cognatus*, was employed for outgroup comparison (0.049–0.065 p-distance) (Kiril'chik & Slobodyanyuk 1997), using bootstrap values (1000 replicates) to assess the node support. In order to test for the presence of saturation in the sequences, we compared the saturation index expected when assuming full saturations with the observed saturation index (DAMBE 4.0.75; Xia & Xie, 2001). A *t*-test with infinite degrees of freedom was used to assess statistical significance.

As the phylogenetic analysis detailed only some evolutionary aspects, an analysis of hierarchical genetic structure as adapted to molecular data (AMOVA as implemented in ARLEQUIN vs. 2.0; Schneider *et al.*, 2000) was undertaken. The significance of variance components and *F*-statistic analogues were tested by multiple permutation of the original data set. The AMOVA was run using the full 771 bp and the shorter 482 bp control region fragment, which integrated the present data with that from Englbrecht *et al.* (2000).

We also tested whether *C. gobio* had experienced population expansions and contractions in two ways. (1) A frequency distribution of the pair-wise number of mutational differences for the whole data set and subsets (clades I, III and IV) was constructed as implemented in DNASP vs. 3.53 (Rozas & Rozas, 1999). A population, which has recently experienced a population growth and decline, shows a smooth, unimodal and Poisson-like

distribution; significance was tested by coalescence. The age expansion number ($\tau = \mu t$ with μ the mutation rate and t the expansion time in generations) was used to compare the timing of population expansions among bullhead clades. (2) A nested clade analysis (NCA) (Templeton, 1998) was performed on an unrooted minimum spanning network (MSP) drawn by hand and using the MSP output of ARLEQUIN vs. 2.0 (Schneider *et al.*, 2000) with probabilities assessed as implemented in the software program PARSPROB vs. 1.1 (available at http://bioag.byu.edu/zoology/crandall_lab/programs.htm). The null hypothesis was tested that there were no associations between the haplotypes and geographical locations at different genealogical levels. The nesting design was constructed following the rules described in Templeton *et al.* (1992). The program GEODIS vs. 2.0 (Posada *et al.*, 2000) was used for implementing the calculations of the distance measures and their statistical significance. The distance matrix consisted of the shortest river distances as measured manually on maps. Within drainage, distances between populations were measured along present river courses. Distances between populations from different drainages were measured using the information on historic connections such as river captures and common ancestral drainage systems. If more than one connection between two populations was possible, the shortest distance was incorporated. The historic connections used are indicated in Fig. 1. Problems arose regarding the historic interpretation of the connections among some river basins. In such cases the Euclidean distance was used (straight lines in Fig. 1). All UK samples were measured relative to a central point in the North Sea (51°52′ 00′ N; 2°37′ 00′ E), which is based on the assumption that the Thames was connected to the Scheldt, Maas and Rhine during periods of low sea level (Gibbard, 1988). As the matrix of river distances was constructed by hand, errors were checked by conducting a cluster analysis of the matrix. The hydrographical distances estimated from the data were $D_{c(x)}$, $D_{n(x)}$, I–$T_{c(x)}$ and I–$T_{n(x)}$ (Templeton, 1998). $D_{c(x)}$ is the average distance of all individuals of clade x from their geographical centre, indicating how widespread a clade is. $D_{n(x)}$ is the average distance of all members of clade x from the geographical centre of nesting clade y, estimating how far individuals of clade x haplotypes are from individuals bearing clade y haplotypes. I–$T_{c(x)}$ and I–$T_{n(x)}$ are the average D_c and D_n values of all the interior clades within the nesting clade minus the average D_c and D_n values for all interior clades within nesting clade y. They give an estimate of the distribution of old vs. young clades. The distribution of all distance measures was determined by recalculating all distances over 10 000 random permutations of clades against sampling locality. Interpretation of the contingency tests followed the updated (24.10.01) version of the inference key of Templeton (1998) (http://bioag.byu.edu/zoology/crand-all_lab/geodis.htm).

Results

Haplotype diversity

Alignment of sequences (GENBANK accession numbers AF381189-AF381217) revealed a total of 59 haplotypes among the 123 bullheads from the 35 sites analysed at the 771 bp control region sequence (data set I) (Appendix 1). Seventy-two positions (9.3%) were polymorphic, of which 45 positions (5.9%) were parsimony informative, including three indels. No evidence for the presence of nuclear copies was found (e.g. intraindividual allelic variation). When reducing the data to the 482 bp control region sequence, which is the fragment in common with the data of Englbrecht *et al.* (2000), 40 haplotypes were found (Appendix 1). All haplotypes were drainage-specific (Danube, Rhine, Elbe, North Sea and Gulf of Biscay). The number of haplotypes and nucleotide diversity (π) (based on the 482 bp fragment and including the data of Englbrecht *et al.* (2000) within drainage) was lowest in the Elbe, Upper Danube, Upper Rhine, Poland, Seine, Adour and Brittany (Table 2), higher in the Baltic ($\pi = 0.00644$), Lower Danube (0.00700), Lower Rhine (0.00599), and highest in the Netherlands (0.015535) because of the presence of haplotypes of lineage I, III and IV. Although sample sizes were often small, most catchments indeed harboured very few haplotypes [E06 (Englbrecht *et al.*, 2000), V23, V28, V29 (Table 1) and sites mentioned in Knapen *et al.*, 2002]; but see E05, E43 (Englbrecht *et al.*, 2000), V20, V21 and V34]. Additional individuals were genotyped at those sites where we noticed considerable diversity on a small number of fish.

Phylogeny

The base composition for all sequences was not significantly different (5% χ^2-test on the average composition) (PUZZLE). The molecular-clock was not enforced since PUZZLE showed rejection of the molecular clock hypothesis. There was no indication of saturation, either in the full data set of 771 bp, or clade I, III or IV (DAMBE). Between clade sequence differences of the control region (data set II) point to weak differences between clades I and II, and strong differences between clades I and IV, II and IV–V–VII (Table 3).

Maximum likelihood, MP and distance-based phylogenetic analysis based on neighbour joining of Kimura-2 parameter distances were performed on the 771 bp fragment (Figs 1 and 2). The five major clades were congruent among the three methods, although on average bootstrap support was low because of the small number of phylogenetically informative sites. The root of the bullhead haplotypes was positioned within clade I, indicating that this clade contains the most ancestral haplotypes. Clade IV was located at the tip of the phenogram. Clade I showed a deeper branching profile

Table 2 Number of fish collected (n), haplotype number (n_h) and type, and nucleotide diversity within drainage (data set II, 482 bp) of the control region of bullhead.

Drainage (clade)	n	Haplotype n_h	Haplotypes	Nucleotide diversity (π)
Elbe (I)	45	4	HE12, HE14, HV9, HV10	0.00103 ± 0.00100
Baltic (I)	8	2	HE23, HE24	0.00644 ± 0.00426
Upper Danube (I)	121	9	HE1, HE2, HE3, HE4, HE5, HE6, HV1, HV2, HV4	0.00147 ± 0.00124
Lower Danube (I)	26	5	HE1, HE7, HE8, HE10, HE11	0.00700 ± 0.00411
Main (I)	62	6	HE8, HE12, HE18, H20, HE21, HV15	0.00301 ± 0.00206
Upper Rhine (I)	12	1	HE1	0.00000 ± 0.00000
Poland (II)	43	2	HE28, HE30	0.00010 ± 0.00027
Lower Rhine – Maas (III)	25	1	HE34, HE35, HE36, HE37, HE38, HE39,	0.00214 ± 0.00161
		1	HE40, HV19, HV22, HV28, HV29	
Lower Rhine – Maas (III and IV)	111	1	HE34, HE35, HE36, HE37, HE38, HE39, HE40, HE41, HE42, HV19,	0.00599 ± 0.00353
		4	HV22, HV25, HV28, HV29	
Netherlands (III and IV)	17	8	HE34, HE39, HE40, HE41, HE42, HV19, HV22, HV25	0.01554 ± 0.00855
North Sea (IV)	57	1	HE41, HE42, HV35, HV36, HV37, HV39, HV42, HV44, HV46, HV47,	0.00387 ± 0.00250
		6	HV48, HV49, HV51, HV53, HV55, HV57	
Seine (V)	7	2	HE43, HE44	0.00060 ± 0.00082
Adour (VI)	16	2	HE45, HE46	0.00192 ± 0.00157
Brittany–Loire (VII)	3	1	HV6	0.00000 ± 0.00000

Table 3 Matrix of between clade sequence difference (p-distances including standard error) of the mitochondrial control region of *Cottus gobio* (data set II, 482 bp).

Group	I	II	III	IV	V	VI
II	0.006 ± 0.003					
III	0.010 ± 0.004	0.018 ± 0.006				
IV	0.031 ± 0.007	0.039 ± 0.009	0.021 ± 0.006			
V	0.028 ± 0.007	0.031 ± 0.008	0.023 ± 0.007	0.014 ± 0.005		
VI	0.023 ± 0.006	0.023 ± 0.007	0.024 ± 0.006	0.028 ± 0.007	0.023 ± 0.007	
VII	0.027 ± 0.007	0.031 ± 0.008	0.022 ± 0.006	0.020 ± 0.006	0.021 ± 0.007	0.023 ± 0.007

than clades III and IV; haplotypes from the Danube – Tatra mountains (HV1 and HV2) and the Elbe (HV8 to HV11) are distinct. Unexpectedly, the haplotypes from the Scheldt – V28 and V29 (HV33 and HV34) clustered within the Central European clade I.

A minimum spanning tree was constructed from the squared distance matrix among haplotypes of the 482 bp fragment (data set II) to understand the phylogenetic relationships between the haplotypes and to infer the evolutionary history (Fig. 3). It yielded the major groupings (clades) as defined by Englbrecht *et al.* (2000), with the addition of a new clade (Scorff, VII). The tree was used to assign clade levels for a NCA (see below). The parsimony criterion was set at nine mutations, and excluded the Adour samples from NCA.

The pair-wise number of mutational differences in haplotype sequence (771 bp – data set I) deviated significantly from the expected distribution under a population growth–decline model (Fig. 4). Frequency peaks were observed at 2, 13, 18 and 24–27 mutations, the latter peak representing differences between the major clades (except clade II). Mismatch distributions differed considerably among clades with the mean number of differences being the lowest in clade III (average number of mutational differences 'standard deviation: 1.8 '1.3) and IV (2.7 '1.6), and the highest in clade I (14.5 '9.3). In the case of clade I three frequency peaks were observed. The mismatch distribution within clade IV fitted well the model of population growth–decline. The observed values of the age expansion parameter (τ) differed considerably among clades (I: 3392, III: 5000 and IV: 1787 generations).

Population subdivision

A population-based analysis was undertaken using a hierarchical analysis of variance (ANOVA) (Schneider *et al.*, 2000) and an NCA (Templeton *et al.*, 1992). The first method tracks population dynamics by taking into
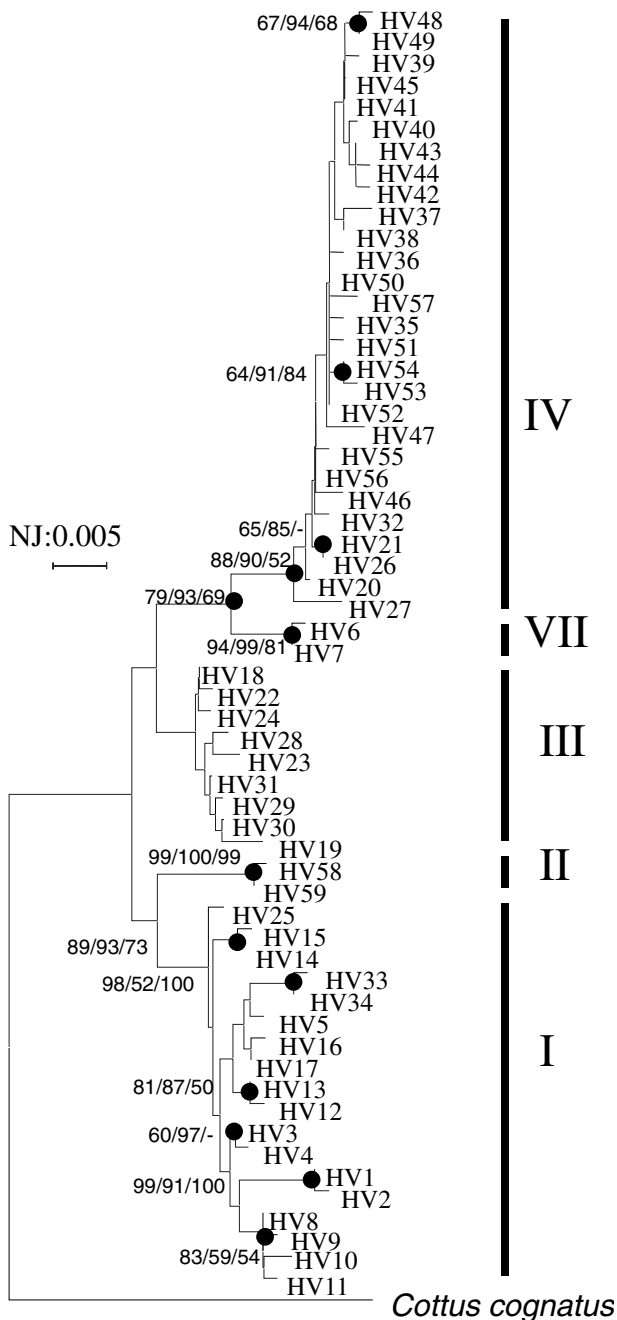
Fig. 2 Neighbour joining illustration of haplotypes (771 bp – data set I) of *Cottus gobio* using *C. cognatus* as outgroup. Bootstrap support for branching is shown for neighbour joining of Kimura-2-P distances (NJ), maximum likelihood (ML) and maximum parsimony (MP) (NJ / ML / MP) when NJ values are superior to 60%. Haplotype clusters are grouped by claded I (Danube, Upper Rhine and Elbe), II (eastern Baltic), III (Lower Rhine and Maas), IV (Lower Rhine, Scheldt and UK) and VII (Brittany).

account the sequence information and haplotype frequency at various arbitrarily defined hierarchical levels. The second test takes into account haplotype

frequencies, their phylogenetic hierarchy and geographical distances.

The hierarchical AMOVA, based on eight *post hoc* defined groups (UK, Scheldt, Maas, the Netherlands, Brittany, Main, Elbe and Danube), showed a high variance among groups and among populations within groups in both the 771 bp (data set I) and 482 bp fragment data (data set II) (Table 4). Variance within populations was small with low diversity. Thus, differences between the eight groups and between all 33 samples are higher than any variance observed within each population. Indeed, only the Dutch samples showed high haplotype diversity, whilst most samples had only one or two haplotypes.

The second approach, NCA, differentiated between historic and recurrent events by incorporating distance information, frequency and mutational information. Of the 40 contingency tests at six hierarchical levels, 15 hypotheses could not be assessed, eight represented restricted gene flow with isolation by distance, seven represented a contiguous range expansion, four past fragmentation events and six were inconclusive (Appendix 2). When adjusting for spurious results (Bonferroni correction) the hypotheses of clades 2–14, 3–4 and 4–2 couldn't be tested any more; the hypothesis of clade 1–30 changed from restricted gene flow to a contiguous range expansion. At the highest level (6-level clade) the hypothesis could not be tested. Of the 5-level clades, only clade 5–2 (clades III to VII) differentiated significantly by contiguous range expansion. Of the 4-level clades, 4–1 was not significant, 4–2 and 4–3 had an inconclusive outcome and 4–4 a contiguous range expansion. Contiguous range expansions and fragmentation events dominate the eastern range (sample V1 – Turiec River; sample E15 (HE11) – Kolpa River; samples E36–40 – Vistula River) whereas fragmentation (sample E53 – Seine River; samples E54–56 – Adour River; sample V12 – Scorff River) and restricted gene flow dominate the western range (Appendix 2). Interesting in clade IV is the gene flow between the Scheldt and the Neet basin (1–30) and the range expansion to the Humber (2–14). Thus, the texture of the pattern differs between the western and eastern group; the former is dominated by a single clade which shows smaller internal expansions and fragmentations, whereas the latter includes a set of clades which are the result of successive expansions and fragmentations (Fig. 5).

Discussion

Spatial differentiation west of the rivers Vistula and Danube

Most haplotypes found in this study were assigned to one of six clades defined by Englbrecht *et al.* (2000) and confirmed in Northern Europe by Kontula & Väinölä (2001), excluding the regions of clades V (Seine basin)
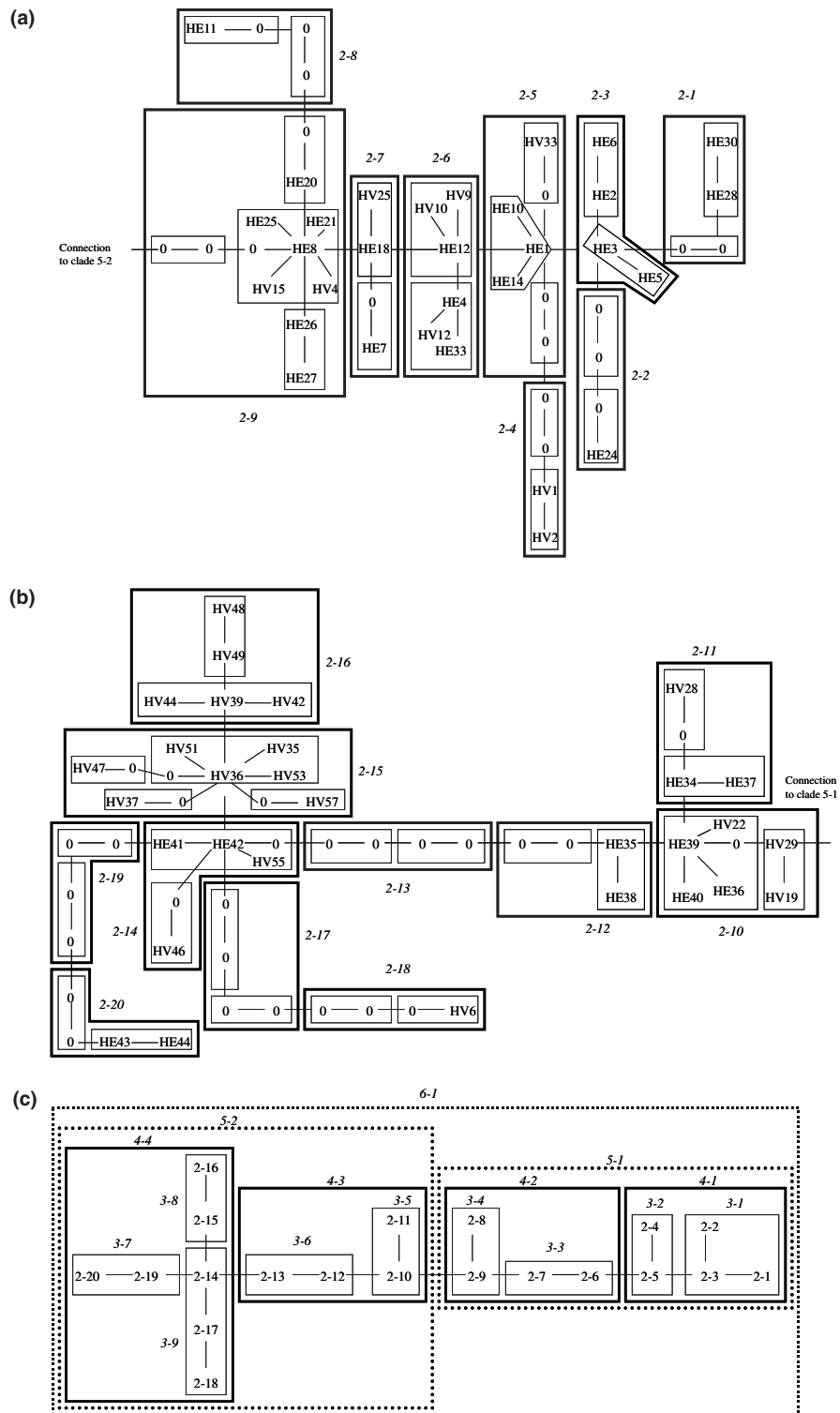
Fig. 3 Minimum spanning network and hierarchical grouping (six levels) of the 59 control region haplotypes of bullhead from this study (771 bp – HV) (data set I), as well as the 46 haplotypes detected by Englbrecht *et al.* (2000) (523 bp – HE), but reduced to 40 haplotypes of a 482-bp fragment (data set II). Level 1 clades are not numbered (See Appendix 1). 3a. Clade 5–1; 3b. Clade 5–2; 3c. Clade 6–1.
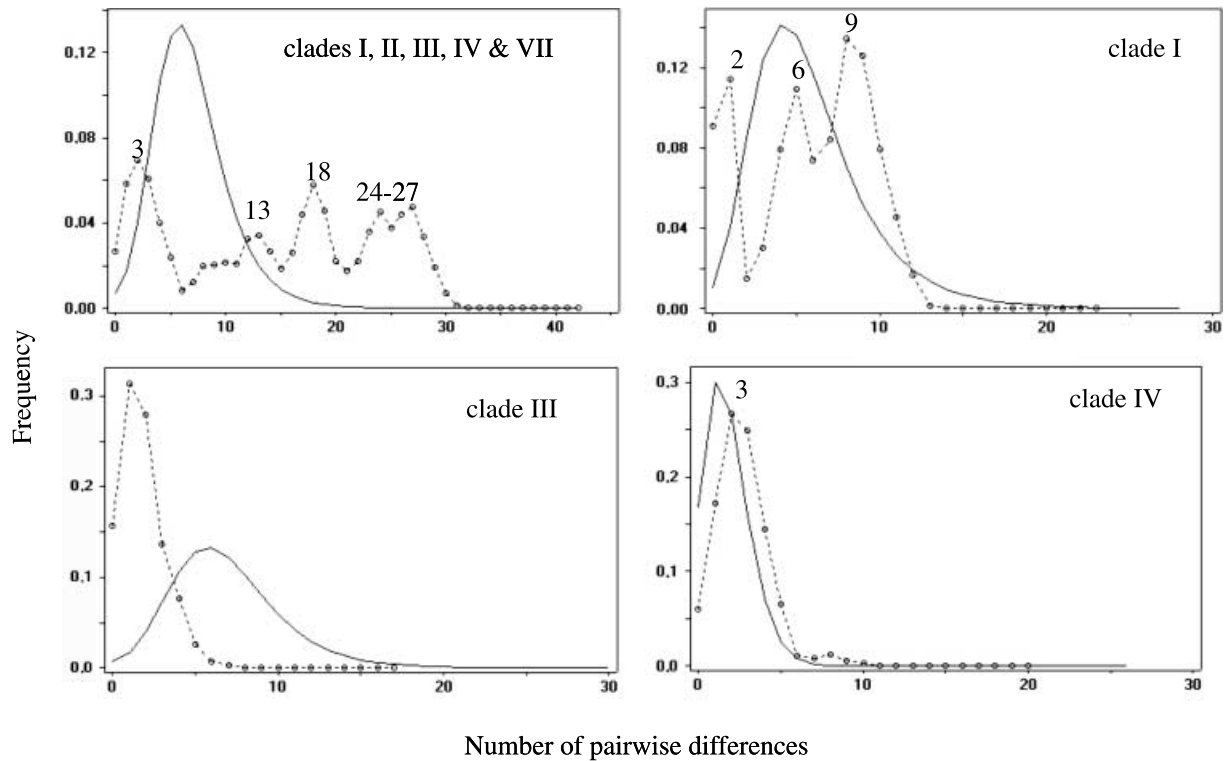
Fig. 4 Frequency distribution of pair-wise number of mutational differences in a 771-bp control region fragment between bullhead individuals observed in all lineages combined ($n = 123$; upper left), in clade I ($n = 35$; upper right panel), in clade III ($n = 25$; lower left panel) and in clade IV ($n = 57$; lower right panel). The dashed line represents the observed data, the full line the model of population growth – decline fitted to the data. The numbers at the peak frequencies represent the number of pair-wise differences.

Table 4 Hierarchical analysis of genetic structure analysis of molecular variance (AMOVA) in bullhead for the eight partitions of samples at the full 771 bp control region sequence and a shortened 482 bp control region sequence of bullhead at 33 sites (data set I). The eight groups have been defined *post hoc* (see text).

	D-loop (771 bp)		D-loop (482 bp)	
Source of variation	d.f.	% var.	d.f.	% var.
Among groups	7	41.70*	7	44.23***
Among populations / within groups	25	44.00***	25	41.42***
Within populations	87	14.30***	87	14.34***
Total	119		119	

*Statistical significance is shown at the 1% level;
***and the 0.01% level;
Samples included in the analysis are the same as in Table 1 and Fig. 1; d.f., degrees of freedom.

and VI (Adour basin) that were not sampled. The Breton population was identified as a new phylogenetically distinct clade (clade VII). Bullhead from the Loire basin is distinct (Eppe *et al.*, 1999) and might either belong to the Breton clade (which is likely given the local geography), or represent an eighth clade. Similar uncertainty exists for the Garonne basin (Eppe *et al.*, 1999), which possibly

encompasses the Adour clade, or may represent a new clade. The sampling range is now sufficiently comprehensive to describe in detail the geographical extent of clades I, III and IV (Fig. 5). The range of clade I extends over the catchments of Danube and Rhine (Oberrhein and Hochrhein, but in two cases also in the Lower Rhine), Weser, Elbe and the Baltic region (Kontula &
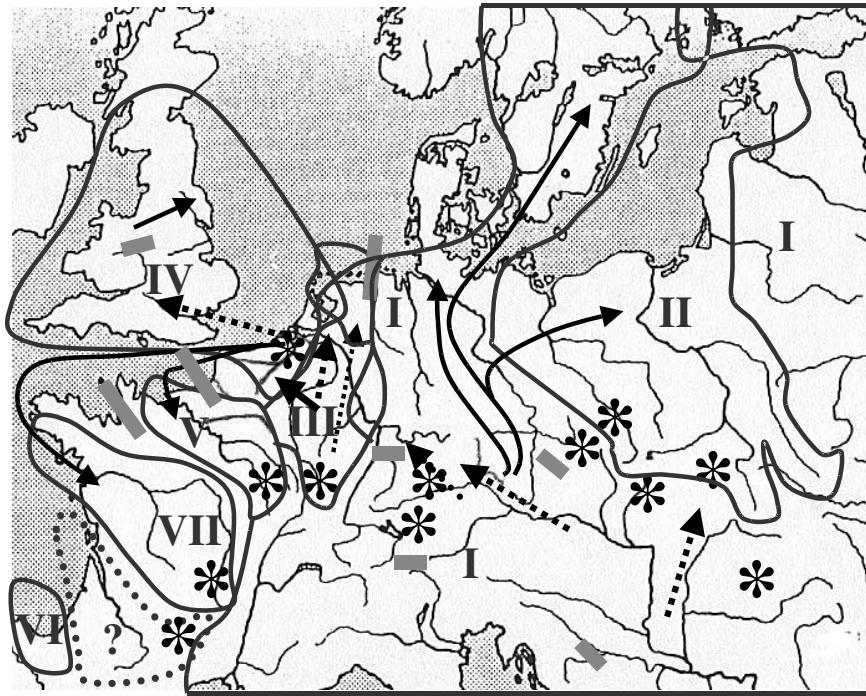
Fig. 5 Geographical extent of the seven clades of *Cottus gobio* with indication of geographical contiguous range expansions ($\rightarrow$), restricted gene flow with isolation by distance ($--\blacktriangleright$), fragmentations (|) and palaeorefugia (*).

Väinölä, 2001). Hybridization between an eastern group and a western group is evident in northern Scandinavia (Kontula & Väinölä, 2001). Unexpectedly, one sample in central Belgium belongs to clade I, which may represent the first documented introduction of bullhead, a species considered less susceptible to anthropogenic transfers. Clade II extends from Poland into the eastern Baltic (Kontula & Väinölä, 2001). Clade III haplotypes were observed in the Lower Rhine and the Maas drainage, one population of the Scheldt and in the Middle Rhine. The range of clade IV haplotypes extends from the Lower Rhine into the Middle Rhine, the Scheldt and the United Kingdom. The range of bullhead to the east (including the Dnjepr and Volga catchments) includes clade I and possibly another clade. Clades V (Seine), VI (Adour), VII (Brittany – Loire?), and may be an eighth clade (Garonne) stretch along the French Atlantic coast.

Temporal differentiation of the clades

An important but contentious issue is the timing of the events, which led to the differentiation of the various distinct clades. Three sources of information are considered here. First, although molecular clock rates are not necessarily consistent across taxa and loci, there are some widely accepted generalizations for applying molecular clocks. Noncoding regions, such as the mitochondrial control region, are expected to evolve faster than coding regions because of the lack of mutational constraints. This has been demonstrated for many groups of animals including cichlid fishes (Meyer *et al.*, 1990; 6–7% divergence per million years), Arctic charr (Brunner *et al.*, 2001; 5–10% per million years) and butterfly fishes (McMillan *et al.*, 1999) with as much as 33–100% per million years at selected hotspots. As Cichlidae and Cottidae belong to the order of the Percomorpha, one option is to adopt the calibration from cichlid fishes (6%) to estimate divergence times in *C. gobio* (Table 5).

Secondly, molecular clock estimates of the control region of bullhead itself have been made by Englbrecht *et al.* (2000) (1% per million years) and Kontula & Väinölä (2001) (2.4–8.4% per million years). Two values (1 and 6% per million years) were used to estimate the divergence times among the major clades of haplotypes (Table 5).

Thirdly, it might be possible to identify an association between major mutation events and postulated population expansions (related to glaciations) by incorporating information from the pair-wise haplotype divergence estimates at the control region. Thienemann (1950) identified in Central Europe a faunal group of 'southern ice cap marginal species' (including bullhead) in response to glaciation. They inhabited the Alps and Central Europe during preglacial times and moved to the south of the unglaciated belt of Southern Europe during the Ice Ages. Rapid expansions of bullhead can be envisaged at two

Table 5 Calibration of the control region of bullhead according to three rates by means of pair-wise comparisons among clades. Ages are given in million years (ma).

Pair-wise clade	I–II	I–III	I–IV	I–VII	III–IV	III–VI	IV–V	IV–VII
Sequence divergence (%)	0.006	0.010	0.031	0.027	0.021	0.024	0.014	0.020
1% (ma)$^{-1}$ (ref. 1)	0.6	1.0	3.1	2.7	2.1	2.4	1.4	2.0
6% (ma)$^{-1}$ (ref. 2; ref. 3)	0.1	0.17	0.5	0.45	0.35	0.4	0.23	0.33
9% (ma)$^{-1}$ (this study)	0.07	0.11	0.34	0.30	0.23	0.27	0.16	0.22

Ref. 1: Englbrecht *et al.* (2000); Ref. 2: Meyer *et al.* (1990); Ref. 3: Kontula & Väinölä (2001).

times during the glacial/interglacial phases. At the beginning of the interglacials, when the climate warms, fish either follow the receding ice margin northwards as vast stretches of habitat become available, or move upstream into mountain refugia, which induces major discontinuities or even southerly extinctions. At the end of the interglacials, when rapid climatic cooling forces the (sub) boreal zone southwards, bullhead descend from the refugia and expand southwards. The northern range becomes covered with ice sheets, and thus inaccessible for bullhead.

The four major phases of mutations appear to coincide with the population expansions at the end of the four interglacials southwards and within lowlands, and at the beginning of the interglacials, northwards. By using the assumptions for a molecular clock of the control region, the four peaks in pair-wise sequence differences (0.4, 1.7, 2.6 and 3.5%) would correspond to approximately 0–11, 110–130, 210–230 and 310–330 ka BP, respectively, which translates into a molecular clock of about 9% per million years (Table 5).

Molecular clock rates should be interpreted with caution as lineage-specific rate variation has been demonstrated in many groups (Rambaut & Bromham, 1998). We have assumed a constant rate, although a molecular clock likelihood ratio test rejects the clocklike behaviour. Small and isolated populations of bullhead might suffer from drift, whereas large populations are more likely to be under selective pressure. In addition, local extinction of intermediary lineages seems to be a recurrent feature in bullhead.

A scenario for the colonization history of bullhead

Thus, during warm interglacial epochs, such as today, the bullhead must have survived in refugia at higher altitudes; the distribution was probably patchy and not unlike that of today. The putative palaeorefugia of clade I are rather complex, as the Danube catchment is bordered by numerous high altitude regions: Alps (4000 m), Tatra mountains (2600 m), Carpathians (2300 m) and Dinarides (2400 m) (Fig. 5). Notice also that the populations from Borovo (Carpathians), Kolpa (Dinarides), Nera (Carpathians) (Englbrecht *et al.*, 2000) and Turiec (Tatra) in the Middle and Lower Danube are well differentiated, and thus have been isolated for some time. Clade II

bullhead probably retreated into the Tatra Mountains (2600 m). The remaining putative refugia are located along the south-western Palearctic. Clade III most recently (Würm) retreated in the Vosges (1000 m) during a warm period or maybe in a limestone area with a milder microclimate during a cold period (Stewart & Lister, 2001). A more ancient (during the Günz and Mindel) northern location cannot be excluded. Clade IV probably survived in the Ardennes (B–F) (700 m) and possibly during the last Glacial in the limestone region of Dartmoor (600 m) (Stewart & Lister, 2001). The Langres/Morvan region (900 m) (clade V), Central Massif (1800 m) (clade VII) and Western Pyrenees (3300 m) (clade VI) complete the string of western refugia. The upper temperature limit to maintain a stable population of bullhead probably approaches 20 °C in summer as judged from its current distribution range, which overlaps largely with brown trout *Salmo trutta* (Lelek, 1980).

From the molecular clock data and the location of the palaeorefugia, the colonization route and timing of the colonizations of the western Palearctic by bullhead can be completed as follows (Fig. 6). From the early Pleistocene onwards, bullhead occupied the Lower Danube as temperatures remained mild across the western Palearctic (Lowe & Walker, 1997). Although a dispersal wave before the first glacial is not visible in the sequence variation, at the beginning of the first glacial (Günz about 430 kA BP) warming resulted in a first north-western dispersal wave along a route following most likely the Dnjepr River, Pritpjat Swamps, Vistula River and Baltic Plain catchments (Englbrecht *et al.*, 2000; Nagel, 2000). Most populations of this old Günz colonization have disappeared because of more recent glacials (Bernatchez & Wilson, 1998), but clade III survived at the western edge of the Baltic Plain. At the end of the Günz glacial (325 ka BP) bullhead expanded from clade III further west (clade IV). The Mindel interglacial resulted in a contraction of clades III and IV, and expansion southward of clade VI and VII. At the beginning of the Riss interglacial, the clades III, IV, VI and VII retreated into refugia, and group I expanded northwards (subclades I east and I west). At the end of the Riss interglacial, group IV expanded along the English Channel into the Seine Basin (clade V), and several Danubian subclades developed in the Dinarides, Carpathians and Tatra mountains.
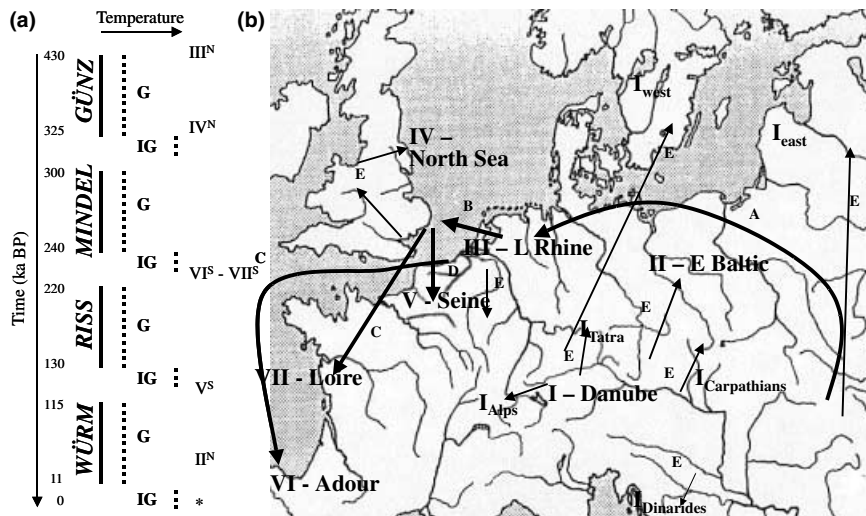
Fig. 6 Model for the colonization history of bullhead. (a) The upper panel refers to the timing of cold and warm periods during the Late-Pleistocene and Holocene, as well as the formation of the clades (IG: interglacial period; G: glacial period). The roman numbers refer to the seven clades, the superscript whether an expansion occurred in northern (N) or southern (S) direction, and the * refers to a Holocene expansion. (b) The lower panel depicts the spatial evolution of each clade with special reference to the timing. The letter associated with each arrow refers to the glacials / interglacials; A: preGünz glacial; B: Günz; C: Mindel; D: Riss and E: Würm.

Clade II expanded from clade I during the Würm glacial at about 70 ka BP, just before a major glacial cooling event (Lowe & Walker, 1997) and possibly because of river capture in the Tatra mountains. During the current Würm interglacial, clades I and II have expanded north, whereas clades III, IV and probably V, VI and VII have locally expanded after retreating from their respective interglacial refugia.

The clades I, II, III and IV show the impact of a recent population expansion. Phylogeny of the control region suggests a simultaneous radiation from a few ancestral haplotypes within each clade (clades III and IV). Control region and microsatellite population data (Hänfling *et al.*, 2002) confirm the discreteness of most populations within each drainage basin: bullheads are slow dispersers on a short time scale. The timing of the most recent population expansion is probably the beginning of the last interglacial (20 ka BP) and might harbour elements from the before last interglacial (150 ka BP) (see above). For example, British bullhead (clade IV) seems to occur in three groups (Fig. 2): Cornwall (the oldest), Humber (most recently colonized from the Upper Scheldt) and Humber, Wales and Cornwall (colonized from the Middle Scheldt). As far south, 50°N of the UK, was covered by ice during the Riss and Würm glacials, the colonization source(s) must have been to the south-east. Hence, unlike the conclusions of Englbrecht *et al.* (2000), our analyses point towards the strong impact of glacial periodicity on bullhead population dynamics.

Morphometric, allozyme, microsatellite and mitochondrial information point to the discreteness of most populations (Riffel & Schreiber, 1995; Hänfling & Brandl,

1998a,b; Eppe *et al.*, 1999; Knapen *et al.*, 2002). Limited gene flow for a certain amount of time was sufficient to cause drift among semi-isolated populations. However, there are at least two regions where this is not the case with bullhead. An interpretation within a coalescent theory framework would mean that the populations of the Lower Rhine represent a recent secondary contact. Many haplotypes belonging to clades III and IV are found in the same drainage. There is even a haplotype of clade I present in the Ketelmeer (NL) and several haplotypes in the Middle Rhine (Englbrecht *et al.*, 2000; Nolte, pers. comm.). Such features explain the suggestion of de Nie (1996) that a 'lowland' bullhead population inhabiting the lakes and slow flowing major rivers (the mixed clade I, III and IV identified in this study) and an 'upland' (clade III) bullhead population inhabiting the small fast flowing mid-altitude rivers are found in the Netherlands. A similar suture zone is observed in the northern Baltic where two morphs of clade I hybridize (Kontula & Väinölä, 2001). Initially both groups were morphologically and ecologically identified as two species. Again the regional habitat has only recently become vacant based on the timing of isostatic changes in northern Scandinavia during the Holocene.

The Late Pleistocene structure of most (sub)boreal and temperate zone freshwater fishes across Europe [Atlantic salmon (Nilsson *et al.*, 2001), brown trout (Bernatchez, 2001), Arctic charr (Brunner *et al.*, 2001), chub (Durand *et al.*, 1999), grayling (Koskinen *et al.*, 2000), perch (Nesbø *et al.*, 1999) and whitefish (Bernatchez & Dodson, 1994; Hansen *et al.*, 1999)] typically shows a Danubian clade, an Atlantic (western) clade and an

eastern clade. There is no other species than bullhead with such high and ancient clade diversity at the western edge of the Palearctic. In addition, many fish taxa show a number of regional clades in the northern Mediterranean which, however, rarely have given rise to widespread populations as in terrestrial animals (Hewitt, 2000). Palaeorefugia of freshwater fish are species-specific, as their individual life-style, ecophysiology and life-history traits determine their location. Postglacial colonization by most fishes seems initially to have happened along an east–west axis along the Danube catchment, whereas more recent population expansions have occurred on a north–south axis. This fits the arguments of Bernatchez & Wilson (1998) who showed that latitude correlates with a breakpoint in nucleotide diversity and tree branching between nonglaciated and glaciated areas (at 46°N), with evolutionary rates of dispersal of northern species being 10 times larger than southern species.

General conclusion

Phylogenetic, nested clade and hierarchical analysis of molecular variance show that *C. gobio* has one of the most diverse and fine grained genetic patterns of the terrestrial and aquatic animals analysed so far in the western Palearctic (Bernatchez & Wilson, 1998; Hewitt, 2000). Colonization happened along two major routes and in several waves, which can be dated with great precision. The distinction between clades is unambiguous and points to periods of colonization caused by habitat expansion followed by population fragmentation caused by habitat loss, and again followed by population expansion. The species has thus responded actively and regularly to the fluctuations in its natural range. An issue remaining is the population structure in the Russian Plain, where some of the most ancestral populations might be found. A study in this area has therefore priority.

Supplementary material

The following material is available from http://www.blackwell-science.com/products/journals/supp-mat/JEB/Jeb469/Jeb469sm.htm
Appendix 1
Appendix 2

References

Avise, J.C. 2000. *Phylogeography*. Harvard University Press, Cambridge.

Avise, J.C., Arnold, J., Ball, R.M. *et al.* 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.* **18**: 489–522.

Banarescu, P. 1990. *Zoogeography of Freshwaters*, Vol. 1. *General Distribution and Dispersal of Freshwater Animals*. Aula-Verlag, Wiesbaden.

Banarescu, P. 1992. *Zoogeography of Freshwaters*, Vol. 2. *General Distribution and Dispersal of Freshwater Animals in North America and Eurasia*. Aula-Verlag, Wiesbaden.

Bernatchez, L. 2001. The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution* **55**: 351–379.

Bernatchez, L. & Danzmann, R.G. 1993. Congruence in control-region sequence and restriction-site variation in mitochondrial DNA of brook charr (*Salvelinus fontinalis* Mitchell). *Mol. Biol. Evol.* **10**: 1002–1014.

Bernatchez, L. & Dodson, J.J. 1994. Phylogenetic relationships among Palearctic and Nearctic whitefish (*Coregonus* sp.) populations as revealed by mitochondrial DNA variation. *Can. J. Fish. Aquat. Sci.* **51** (Suppl.): 240–251.

Bernatchez, L., Glémet, H., Wilson, C.S. & Danzmann, R.G. 1995. Introgression and fixation of Arctic charr (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*). *Can. J. Fish. Aquat. Sci.* **52**: 179–185.

Bernatchez, L. & Wilson, C.C. 1998. Comparative phylogeography of Nearctic and Palearctic fishes. *Mol. Ecol.* **7**: 431–452.

Brunner, P.C., Douglas, M.R., Osinov, A., Wilson, C.C. & Bernatchez, L. 2001. Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial dNA sequences. *Evolution* **55**: 573–586.

Downhower, J. & Brown, L. 1979. Seasonal changes in the social structure of a mottled sculpin (*Cottus bairdi*) population. *Anim. Beh.* **27**: 451–458.

Durand, J., Persat, H. & Bouvet, Y. 1999. Phylogeography and postglacial dispersal of the chub *Leuciscus cephalus* in Europe. *Mol. Ecol.* **8**: 989–997.

Englbrecht, C.C., Freyhof, J., Nolte, A., Rassmann, K., Schliewen, U. & Tautz, D. 2000. Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggests a pre-Pleistocene origin of the major central European populations. *Mol. Ecol.* **9**: 709–722.

Eppe, R., Persat, H., Beaudou, D. & Berrebi, P. 1999. Genetic variability in sculpin (genus *Cottus*) from southern France, with reference to the taxonomic status of an endemic species, *C. petiti*. *Heredity* **83**: 533–540.

Gibbard, P.L. 1988. The history of the great northwest European rivers during the past three million years. *Phil. Trans. R. Soc. London B* **318**: 559–602.

Hänfling, B. & Brandl, R. 1998a. Genetic variability, population size and isolation of distinct populations in the freshwater fish *Cottus gobio* L. *Mol. Ecol.* **7**: 1625–1632.

Hänfling, B. & Brandl, R. 1998b. Genetic differentiation of the bullhead *Cottus gobio* L. across watersheds in Central Europe: evidence from two taxa. *Heredity* **80**: 110–117.

Hänfling, B., Hellemans, B., Volckaert, F.A.M. & Carvalho, G.R. 2002. Late glacial history of the cold-adapted freshwater fish *Cottus gobio*, revealed by microsatellites. *Mol. Ecol.* **11**: 1717–1729.

Hansen, M.M., Mensberg, K.-L.D. & Berg, S. 1999. Postglacial recolonisation patterns and genetic relationships among whitefish (*Coregonus* sp.) populations in Denmark, inferred from mitochondrial DNA and microsatellite markers. *Mol. Ecol.* **8**: 239–252.

Hasegawa, M., Kishino, H. & Yano, T.-A. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**: 160–174.

Heizmann, E.P.J. 1992. Das Tertiär in Südwestdeutschland. *Stutt. Beitr. Naturkunde* **33**: 1–61.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.

Kiril'chik, S.V. & Slobodyanyuk, S.Ya. 1997. Evolution of *cytochrome b* gene fragment from mitochrondrial DNA in some Baikalian and non-Baikalian Cottoidei fishes. *Mol. Biol. (Moscou)* **31**: 141–148.

Knapen, D., Knaepkens, G., Bervoets, L., Taylor, M.I., Eens, M. & Verheyen, E. 2002. Conservation units based on mitochondrial and nuclear DNA variation among European bullhead populations (*Cottus gobio* L 1758) from Flanders (Belgium). *Conserv. Genet.* **3**: in press.

Kontula, T. & Väinölä, R. 2001. Postglacial colonisation of Northern Europe by distinct phylogeographic lineages of the bullhead, *Cottus gobio*. *Mol. Ecol.* **10**: 1893–2002.

Koskinen, M.T., Ranta, E., Piironen, J., Veselov, A., Titov, S., Haugen, T.O., Nilsson, J., Carlstein, M. & Primmer, C.R. 2000. Genetic lineages and postglacial colonisation of grayling (*Thymallus thymallus*, Salmonidae) in Europe, as revealed by mitochondrial DNA analyses. *Mol. Ecol.* **9**: 1609–1624.

Kumar, S., Tamura, K., Jakobsen, I.B. & Nei, M. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* **16**: 140–151.

Lelek, A. 1980. *The Freshwater Fishes of Europe*, Vol. 9. *Threatened Fishes of Europe*. Aula-Verlag, Wiesbaden.

Lowe, J.J. & Walker, M.J.C. 1997. *Reconstructing Quaternary Environments*. Addison-Wesley-Longman, London.

Marconato, A. & Bisazza, A. 1988. Mate choice, egg cannibalism and reproductive success in the river bullhead, *Cottus gobio* L. *Behav. Biol.* **33**: 905–916.

McMillan, W.O., Weigt, L.A. & Palumbi, S.R. 1999. Color pattern evolution, assortative mating, and genetic differentiation in brightly colored butterflyfishes (Chaetodontidae). *Evolution* **53**: 247–260.

Meyer, A., Kocher, T.D., Basasibwaki, P. & Wilson, A.C. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* **247**: 550–553.

Nagel, K.O. 2000. Testing hypotheses on the dispersal and evolutionary history of freshwater mussels (Mollusca: Bivalvia: Unionidae). *J. Evol. Biol.* **13**: 854–865.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M. & Tajima, F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* **105**: 207–217.

Nesbø, C.L., Fossheim, T.V., Øllestad, L.A. & Jakobsen, K.S. 1999. Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. *Mol. Ecol.* **8**: 1387–1404.

de Nie, H.W. 1996. *Atlas Van de Nederlandse Zoetwatervissen*. Media Publishing, Doetinchem.

Nilsson, J., Gross, R., Asplund, T., Dove, O., Jansson, H., Kelloniemi, J., Kohlmann, K., Loytynoja, A., Nielsen, E.E., Paaver, T., Primmer, C.R., Titov, S., Vasemagi, A., Veselov, A., Ost, T. & Lumme, J. 2001. Matrilinear phylogeography of Atlantic salmon (*Salmo salar* L.) in Europe and postglacial colonization of the Baltic Sea area. *Mol. Ecol.* **10**: 89–102.

Posada, D. & Crandall, K.A. 1998. Modeltest: testing the model of DNA substitution. *Bioinform.* **14**: 817–818.

Posada, D., Crandall, K.A. & Templeton, A.R. 2000. Geodis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* **9**: 487–488.

Rambaut, A. & Bromham, L. 1998. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15**: 442–448.

Riffel, M. & Schreiber, A. 1995. Coarse-grained population structure in Central European sculpin (*Cottus gobio* L.): secondary contact or ongoing genetic drift? *Z. Zoo. Syst. Evol. Research* **33**: 173–184.

Rozas, J. & Rozas, R. 1999. DnaSP, Version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinform.* **15**: 174–175.

Schneider, S., Roessli, D. & Excoffier, L. 2000. *Arlequin. A Software for Population Genetics Data Analysis*, Version 2.00. University of Geneva, Geneva.

Stewart, J.R. & Lister, A.M. 2001. Cryptic northern refugia and the origins of the modern biota. *Trends Ecol. Evol.* **16**: 608–613.

Strimmer, K. & von Haeseler, A. 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree t (opologies. *Mol. Biol. Evol.* **13**: 964–969.

Swofford, D. 1998. *Paup*. Phylogenetic Analysis Using Parsimony (*and other methods)*, Version 4. Sinauer Associates, Sunderland, MA.

Tamura, K. & Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**: 512–526.

Templeton, A.R. 1998. Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* **7**: 381–397.

Templeton, A.R., Crandall, K.A. & Sing, C.F. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. *Genetics* **132**: 619–633.

Thienemann, A. 1950. Verbreitungsgeschichte der Süsswassertierwelt Europas. *Die Binnengewässer* **18**: 1–809.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. 1994. Clustal-W – Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**: 4673–4680.

Wanzenböck, J., Lahnsteiner, B. & Maier, K. 2000. Pelagic early life phase of the bullhead in a freshwater lake. *J. Fish Biol.* **56**: 1553–1557.

Waterstraat, A. 1992. Investigations on the ecology of *Cottus gobio* L. and other fish species from two lowland streams of Northern Germany. *Limnologica* **22**: 137–149.

Winnepenninckx, B., Backeljau, T. & De Wachter, R. 1993. Extraction of high molecular weight DNA from molluscs. *Trends Genet* **9**: 407.

Xia, X. & Xie. Z. 2001. DAMBE: data analysis in molecular biology and evolution. *J. Heredity* **92**: 371–373.