Benchmarking the performance of time-dependent density functional theory methods on biochromophores

Shao, Yihan

2020-01-14

Shao , Y , Mei , Y , Sundholm , D & Kaila , V R I 2020 , ' Benchmarking the performance of time-dependent density functional theory methods on biochromophores ' , Journal of Chemical Theory and Computation , vol. 16 , no. 1 , pp. 587-600 . https://doi.org/10.1021/acs.jctc.9b00823

http://hdl.handle.net/10138/309566
https://doi.org/10.1021/acs.jctc.9b00823

acceptedVersion

Phylogeography and historical introgression in smoothtail nine-spined sticklebacks, *Pungitius laevis* (Gasterosteiformes: Gasterosteidae)

Cui Wang[1]*, **Takahito Shikano**[1], **Henri Persat**[2], **Juha Merilä**[1]

[1]*Ecological Genetics Research Unit, Department of Biosciences, University of Helsinki, FI-00014 Helsinki, Finland*

[2]*UMR 5023 - LEHNA, Université Claude Bernard Lyon 1, 69622 Villeurbanne, France*

Running head: Historical introgression in *Pungitius laevis*

*Corresponding author: Cui Wang, Ecological Genetics Research Unit, Department of Biosciences, University of Helsinki, FI-00014 Helsinki, Finland.

E-mail: cui.wang@helsinki.fi

Abstract

Pleistocene glaciations have strongly affected the biogeography of many species residing in periglacial and previously glaciated regions. Smoothtail nine-spined sticklebacks (*Pungitius laevis*) have three highly divergent mitochondrial lineages in France, one of which shares the same mitochondrial cluster with a congener *P. pungitius*. To understand if interspecific introgression has happened between the two species, we carried out phylogeographic and population genetic analyses using mitochondrial and nuclear gene sequences. Our results indicated asymmetric mitochondrial introgression from *P. pungitius* to *P. laevis* and genetic admixture of these species in one of the *P. laevis* lineages, suggesting historical hybridization. Deep intraspecific mitochondrial divergence within *P. laevis* in central and southern France mostly coinciding with major drainages suggests that these areas were important glacial refugia for the species explaining the observed intraspecific divergence. The historical hybridization between *P. laevis* and *P. pungitius* likely occurred in a refugium at central France, and the newly formed *P. laevis* lineage spread northward during postglacial recolonization. The study adds to the long list of species showing complete mitogenome capture owing to historical hybridizations, and highlights the reticulate nature of population differentiation in taxa subject to postglacial range-expansions.

Keywords: genetic admixture, hybridization, introgression, phylogeography, *Pungitius*, stickleback

Introduction

Pleistocene glaciation cycles caused dramatic climate oscillations, and significantly affected the diversification of organisms distributed in boreal and temperate regions (Hewitt, 2000, 2004; Wiens & Donoghue, 2004). During repeated ice sheet expansions, the distribution ranges of species became restricted to glacial refugia in the south (Taberlet *et al.*, 1998; Hewitt, 1999). Comparative analyses of many terrestrial and aquatic organisms have revealed that prolonged isolation in different refugia has led to extensive intraspecific genetic differentiation (Hewitt, 1999; Pamilo & Savolainen, 1999; Schmitt, 2007), and even promoted speciation (Avise & Walker, 1998; Stewart & Lister, 2001). Since a limited number of founding individuals often seeded postglacial recolonizations, northern populations also tend to have lower genetic diversity than populations in the south (Hewitt, 1996; Bernatchez & Wilson, 1998; Hawkins & Porter, 2003). In addition, as distinct evolutionary lineages or species that inhabited in different refugia often used multiple routes for postglacial recolonization, hybridization has frequently occurred in their secondary contact areas (Remington, 1968; Hewitt, 2001, 2011).

Hybridization between distinct lineages or species often results in genetic introgression from one group to another, and backcrossing of hybrids with a parental group can give rise to hybrid swarms (Rhymer & Simberloff, 1996). Incongruent phylogeographic patterns between mitochondrial and nuclear markers are found in various organisms (Toews & Brelsford, 2012). For instance, although different taxonomic groups can be readily distinguished using nuclear genetic markers, they can become assigned into a single mitochondrial group if mitochondrial introgression has occurred (Irwin, Rubtsov & Panov, 2009; Wiens, Kuczynski & Stephens,

2010; Boratyński *et al*., 2011; Darras & Aron, 2015). Postglacial introgression has often been observed in both North American and Eurasian organisms, typically in specific areas where diverged lineages from different refugia have come into secondary contact after northward expansions (Zink, 1994; Hewitt, 1999; Spellman, Riddle & Klicka, 2007). While such admixture can complicate phylogeographic reconstructions, it is becoming clear that hybridization can be also involved in generating new species or lineages (Mallet *et al*., 2007; Schumer, Rosenthal & Andolfatto, 2014).

Evolutionary history and genetic structuring of freshwater taxa are strongly influenced by the historical and contemporary connectivity of the waterways (McGlashan & Hughes, 2001; Smith & Dowling, 2008). As fluvial conditions are strongly affected by climate, local geomorphological events, river captures and sea level changes, glaciation cycles probably resulted in numerous drainage re-connections and disconnections (Blum & Törnqvist, 2000; Whitfield & Harvey, 2012). However, as evidenced by large body of research, fairly detailed reconstructions of historical biogeography of freshwater fauna in formerly glaciated areas are possible with the aid of genetic markers (Bernatchez & Wilson, 1998; Makhrov & Bolotov, 2006).

Stickleback fishes of the family Gasterosteidae are widely distributed in the northern hemisphere (Wootton, 1976), and their diversification has been strongly impacted by glaciation events (Mäkinen & Merilä, 2008; Münzing, 1969; Orti *et al.*, 1994; Takahashi & Goto, 2001; Wang *et al*., 2015). The smoothtail nine-spined stickleback (*Pungitius laevis*) is a small freshwater fish found in coastal and inland areas of western Europe (Kottelat & Freyhof, 2007).

It is morphologically very similar to the nine-spined stickleback (*P. pungitius*), although they can be distinguished on the basis of the absence or presence of lateral scutes and keels (Keivany & Nelson, 2000; Kottelat & Freyhof, 2007). Because of their morphological similarities, *P. laevis* has often been taxonomically considered as a subspecies of *P. pungitius* (Münzing, 1969; Gross, 1979; Paepke, 1996). However, a previous phylogenetic study showed high degree of mitochondrial divergence between *P. laevis* and *P. pungitius*, additionally demonstrating the presence of three highly divergent *P. laevis* lineages in France (Wang *et al*., 2015). The divergences of these lineages were estimated to have occurred around 1.95 to 1.38 Mya in Pleistocene, which are much older divergences than those estimated for globally distributed *P. pungitius* lineages (Wang *et al*., 2015). Given that all divergent *P. laevis* lineages, as well as *P. pungitius* are found in the central and southern parts of France, these areas have been considered as important glacial refugia for ancestral European *Pungitius* fishes (Wang *et al*., 2015). It is noteworthy that one of the three divergent mitochondrial lineages in *P. laevis* is phylogenetically positioned in the western European clade of *P. pungitius* (Wang *et al*., 2015). Thus, it appears that this lineage has a unique evolutionary history that differs from those of the other *P. laevis* lineages. For instance, this lineage might have experienced hybridization and introgression with *P. pungitius* and expanded its geographic range to the northern parts of France where this lineage is currently distributed (Wang *et al*., 2015). It is also possible that this lineage represents a convergent form of *P. pungitius* that has lost its lateral scutes and keels resulting in morphological similarity to *P. laevis*.

The aim of this study was to investigate the evolutionary history and processes underlying

the divergence between different lineages of *P. laevis*. In particular, we were interested in elucidating the hypothesis that *P. laevis* lineage III was formed as a result of hybridization and introgression between *P. laevis* and *P. pungitius*, rather than being a morphologically distinct form of *P. pungitius*. To address these issues, we conducted a fine scale phylogeographic analyses of samples collected from 30 sites in France using both mitochondrial and nuclear gene sequences.

Materials and methods

Samples

We collected 114 individuals of *P. laevis* from 25 sites and 22 individuals of *P. pungitius* from five sites in France (Fig. 1 and Table S1, Supporting information). *P. laevis* and *P. pungitius* were distinguished based on the absence and presence of keels at caudal regions, respectively, which is a diagnostic morphological and taxonomic trait characterizing these species (Kottelat & Freyhof, 2007). The sampling sites covered most parts of the species distribution ranges in France (Wootton, 1976; Paepke, 2001; Kottelat & Freyhof, 2007), including seven main drainage basins (*viz*. Seine, Loire, Dordogne, Charente, Meuse, Rhine and Rhône basins). The *P. laevis* individuals were sampled from three sites in the Dordogne River tributaries (Dordogne basin), two sites in the Charente River (Charente basin), eight sites in the Loire River (Loire basin), seven sites in the Seine River tributaries (Seine basin), four sites in the Meuse River (Meuse basin) and one site in the Mosel River (Rhine basin; Fig. 1). The *P. pungitius* individuals were collected from five sites in the Saône River (Rhône basin; Fig. 1). Although

all the individuals were included in the mitochondrial gene analyses, 82 *P. laevis* individuals from 18 sites and 20 *P. pungitius* individuals from four sites were used for nuclear gene analysis due to small sample sizes in some sites (Table S1, Supporting information). Fin clips were collected and preserved in ethanol for DNA extraction. *P. platygaster* collected from Greece (40°50´N, 22°18´E) was used as an outgroup in nuclear phylogenetic analyses. Mitochondrial data for *P. platygaster* were adopted from Wang *et al*. (2015).

DNA extraction and sequencing

Whole genomic DNA was extracted using the silica-based method (Elphinstone *et al*., 2003) or DNeasy Tissue Kit (QIAGEN). Phylogenetic analyses were conducted with one mitochondrial gene (cytochrome *b*) and eight nuclear gene fragments, including four exon primed intron crossing (EPIC) markers (04174E20, 19231E4, 36298E1 and 55305E1) and four conserved coding regions (myh6, plagl2, SH3PX3 and sreb2; Table S2, Supporting information). A total length of 1104 bp of cytochrome *b* gene was obtained using two primer pairs (Kocher *et al.*, 1989; Palumbi, 1996; Shikano *et al*., 2010; Table S2, Supporting information). Each nuclear gene was amplified and sequenced using the primers reported by earlier studies (Li *et al*., 2007; Li, Riethoven & Ma, 2010; Table S2, Supporting information), resulting in 274 to 853 bp length for each gene with a total alignment length of 4919 bp (Table S2, Supporting information). Polymerase chain reactions (PCRs) for cytochrome *b* and nuclear genes were performed following Shikano *et al*. (2010) with slight modifications on annealing temperature for each gene (Table S2, Supporting information). PCR procedures for the four coding genes included

the second PCR to avoid nonspecific amplification (Li *et al*., 2007). Direct sequencing of PCR products was conducted following Shikano *et al*. (2010) with MegaBACE 1000 (Amersham Biosciences) and ABI 3730XL (Applied Biosystems) for mitochondrial and nuclear genes, respectively. Cytochrome *b* sequences for 45 individuals from 13 sites were obtained from Wang *et al*. (2015).

DNA sequences were aligned using MEGA6 (Tamura *et al*., 2013). To minimize the effects of sequencing error in nuclear genes, only SNPs observed in at least two individuals were considered as polymorphic sites according to Hey & Wakeley (1997). Phylogenetic tree reconstruction and pairwise nucleotide difference estimation (see below) were performed with IUPAC codes for heterozygous sites, and other analyses were conducted using genotypic data transformed with PGDSpider (Lischer & Excoffier, 2012). The mitochondrial and nuclear datasets were analyzed separately, since possible mitochondrial introgression from *P. pungitius* to *P. laevis* was indicated by an earlier study (Wang *et al*., 2015). Novel mitochondrial and nuclear gene sequences were deposited in GenBank (accession numbers: KX384688–KX384725, KX758649–KX758992).

Genetic diversity

For the mitochondrial data, nucleotide diversity (pi), haplotype diversity (Hd) and number of polymorphic sites (S) were calculated using DnaSP 5.10.1 (Librado & Rozas, 2009). For the nuclear data, the number of alleles (Na), expected heterozygosity (H_E) and heterozygosity deficiency (F_{IS}) at polymorphic sites were calculated using GenAlEx 6.5 (Peakall & Smouse,

2012). Tests for linkage disequilibrium and Hardy-Weinberg equilibrium (HWE) were conducted using Genepop 4.2 (Raymond & Rousset, 1995; Rousset, 2008) with Bonferroni correction (Bonferroni, 1936). Given that genetic population structure is highly heterogeneous even within the same phylogenetic lineage (see results), these tests were performed for each site. Thus, it should be noted that the results of these tests can be conservative due to a relatively small sample size in each site. Statistical significance in the level of genetic diversity among four different taxonomic or phylogenetic groups (i.e. *P. laevis* lineage I, II and III and *P. pungitius*; see results) was examined using ANOVA followed by Fisher's LSD post-hoc test. The analyses were performed on nucleotide diversity (pi) and haplotype diversity (Hd) in the mitochondrial data, and for allele number (Na) and expected heterozygosity (H_E) in the nuclear data. The hierarchical analysis of molecular variance (AMOVA) was conducted to evaluate the distribution of genetic variation within populations, among populations and among the four phylogenetic groups (see above) using Arlequin v3.5 (Excoffier & Lischer, 2010).

Phylogeny and population structuring

Bayesian inference phylogenetic analysis was conducted using MrBayes 3.2 (Ronquist *et al*., 2012). The best-fit substitution model was determined based on BIC criteria with Kakusan 4 (Tanabe, 2007). For the mitochondrial data, K80 + Gamma, HKY85 and GTR + Gamma were used for the first, second and third codon positions, respectively. The phylogenetic analysis for the nuclear data was performed with GTR + Gamma for 04174E20 and 55305E1, HKY85 + Gamma for 36298E1, plagl2, SH3PX3 and sreb2, JC69 + Gamma for 19231E4 and K80 +

Gamma for myh6. The MCMC chains were run for 10 000 000 generations (2500 trees were used as burn-in and every 1000 generations were sampled), at which the average standard deviation of split frequencies reached less than 0.01. Tree topology, as well as posterior probabilities, were viewed using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). The incongruence between the mitochondrial and nuclear trees was tested using Congruence Among Distance Matrices (CADM) with 9999 permutations (Campbell, Legendre & Lapointe, 2011). In this analysis, the average number of nucleotide differences (pairwise nucleotide differences) between pairs of the four phylogenetic groups (*P. laevis* lineage I, II and III and *P. pungitius*) were calculated for the mitochondrial and nuclear data separately. The null hypothesis was set to incongruence of phylogenetic patterns in the two data sets.

For the nuclear data, phylogenetic relationships were also inferred based on principal component analysis (PCA) using Eigensoft (Patterson, Price & Reich, 2006). Perl script SmartPCA was used to calculate principal components and determine the statistical significance of each component. Graphical plotting was conducted using R v3.2.3 (R Development Core Team 2008). In addition, population admixture analysis was carried out for the nuclear data using STRUCTURE 2.3.4 (Pritchard, Stephens & Donnelly, 2000). To investigate different genetic groups (K) of *P. laevis* and *P. pungitius*, the analysis was conducted for K = 1 to 9 under the admixture and independent allele frequency models. Each K was run independently with 10 replicates. The burn-in period was set to be 500 000 iterations, and the running period after burn-in was 1 000 000 iterations. The highest hierarchical level of genetic groups in the data was inferred from delta K (Evanno, Regnaut & Goudet, 2005) using Structure Harvester

208 (Earl 2012).

209

210 *Colonization history and recent gene flow*

211 The evolutionary history of *P. laevis* lineage III was assessed based on the geographic trends
212 in genetic diversity within the lineage and the patterns of genetic divergence between
213 populations in this and other lineages. Since *P. laevis* lineage III is phylogenetically distinct
214 from *P. pungitius* based on the analysis of nuclear genes (see results), it appears that
215 hybridization and introgression occurred between *P. laevis* and *P. pungitius* in the past before
216 *P. laevis* lineage III expanded its distribution in the northern parts of France including the Seine,
217 Meuse and Mosel River basins. Given that divergence between *P. laevis* and *P. pungitius* traces
218 back to Pleistocene glaciations (Wang *et al*., 2015), hybridization could have occurred in glacial
219 refugia in the south when these species retreated southward during glaciations. Since postglacial
220 recolonization to the north is expected to lead to the northward reduction of genetic diversity
221 due to founder effects and population bottlenecks associated with range expansion (Hewitt,
222 2000, 2004), we investigated correlation between mitochondrial DNA nucleotide diversity
223 (sensitive to population size reductions) and latitudinal location to infer the colonization history
224 of *P. laevis* lineage III.

225 In addition, we tested possible recent gene flow between *P. laevis* lineage III and *P.*
226 *pungitius*, as well as between *P. laevis* lineage III and lineage I, which have allopatric
227 distribution patterns. Under the null hypothesis that there is no recent gene flow between the
228 different phylogenetic groups, we expected that the level of genetic differentiation is

independent of geographic distance between the sites within each of these groups. In contrast, if there is recent gene flow between the groups, we expected to see lower degree of genetic differentiation between the sites where these groups are geographically more closely located. Linear regressions of genetic differentiation (F_{ST}) at the nuclear genes against geographic distance were performed by Real Statistics Resource Pack (http://www.real-statistics.com/) in Excel 2016 to assess if genetic differentiation within each of the lineages is independent of geographic distance. The same analysis was also conducted with the average pairwise nucleotide differences between pairs of the populations. Since the distance data violates assumption of independence among data points, the significance testing was performed using randomization with 2000 permutations.

Results

Mitochondrial phylogeny

The 1104 bp mitochondrial sequence contained 128 segregating sites defining 61 haplotypes among 114 *P. laevis* and 22 *P. pungitius* individuals (Tables S1, S3, Supporting information). Nucleotide diversity and haplotype diversity were 0.0342 and 0.973 in *P. laevis*, and 0.0029 and 0.887 in *P. pungitius*, respectively (Table S1, Supporting information). Nucleotide diversity differed significantly among the four phylogenetic groups (i.e. *P. laevis* lineage I, II and III and *P. pungitius*; ANOVA, $F_{3,25} = 5.44$, $P = 0.0051$; Fig. 2), although there was no significant difference in haplotype diversity among them ($F_{3,25} = 1.10$, $P = 0.37$). While nucleotide diversity did not significantly differ between *P. laevis* lineage III (0.00069) and *P. pungitius*

(0.00067; Fisher's LSD, P = 0.97; Fig. 2), each of these lineages showed lower nucleotide diversity than *P. laevis* lineage II (0.00254; Fisher's LSD, $P < 0.01$; Fig. 2).

In the Bayesian phylogenetic tree, three major mitochondrial clades were found. *P. laevis* individuals were divided into three highly divergent lineages (i.e. lineage I, II and III) with high posterior probabilities (>0.99; Fig. 3A). All *P. pungitius* individuals clustered with *P. laevis* lineage III (Fig. 3A). *P. laevis* lineage I included individuals from the Loire River drainage (LI_FON, LI_ERD, LI_FIL, LI_RID, LI_VRI and LI_NOH) and a close Loire bordering area of the Seine drainage (LI_LOI and LI_OUA), which is connected to the Loire River through the Canal de Briare. This lineage was further divided into two subclades, with one composed of individuals from LI_FON, LI_ERD, LI_LOI and LI_OUA, and the other composed of individuals from LI_FIL, LI_RID, LI_VRI and LI_NOH (Fig. 3A). *P. laevis* lineage II consisted of individuals from southwestern France, including the Dordogne River and its tributaries (LII_MED, LII_LAR and LII_BLA) and the Charente River and its vicinity area (LII_ANT, LII_TOU, LII_PUY and LII_PAY). This lineage was also divided into two subclades (Fig. 3A). One subclade was composed of individuals from four sites close to the Dordogne estuary (LII_ANT, LII_TOU, LII_MED and LII_LAR), and the other subclade included those from four sites farer from the coastline (LII_PUY, LII_PAY and LII_BLA; Fig. 3A). *P. laevis* lineage III was composed of *P. laevis* individuals from the Seine drainage (LIII_ARO, LIII_EUR, LIII_CHA, LIII_YON and LIII_DRU), the Meuse River (LIII_TRO, LIII_BAR, LIII_MAZ and LIII_MOU) and the Mosel River (LIII_ORN), as well as *P. pungitius* individuals from the Saône drainage (PP_ORA, PP_OGN, PP_MON, PP_MEN and

PP_VEY). No subdivision was found between the *P. laevis* and *P. pungitius* individuals (Fig. 3A).

The AMOVA revealed that the variance among phylogenetic groups accounted for majority (81.3%) of the total variance in the data (Table 1). Variation within phylogenetic groups and within populations only accounted for 15.3% and 3.4% of the total variance, respectively. Genetic differentiation among groups (F_{CT}), within populations (F_{SC}), and among populations (F_{ST}) were 0.813, 0.819 and 0.966 respectively (Table 1).

Nuclear phylogeny

In the total 4905 bp sequence of the eight nuclear gene fragments (Table S2, S4, Supporting information), 73 SNPs were identified in the 82 *P. laevis* and 20 *P. pungitius* individuals. None of the SNPs showed significant linkage disequilibrium or departure from HWE after Bonferroni correction. The mean values of the number of alleles (Na) and expected heterozygosity (H_E) in the study sites were 1.116 and 0.043 for *P. laevis* and 1.076 and 0.020 for *P. pungitius*, respectively (Table S1, Supporting information). The number of alleles (Na) and expected heterozygosity (H_E) were significantly different among the four phylogenetic groups (ANOVA, $F_{3,18}$ = 5.93, P = 0.005 for Na; $F_{3,18}$ = 5.28, P = 0.009 for H_E; Fig. 2). In contrast to the mitochondrial data, *P. laevis* lineage III showed significantly higher Na (1.20) and H_E (0.070) than all of the other groups (Na = 1.05–1.08, H_E = 0.019–0.033; Fisher's LSD, $P < 0.05$ or $P < 0.01$: Fig. 2).

The nuclear phylogenetic tree revealed two main clusters corresponding to *P. laevis* and

292 *P. pungitius* with high posterior probabilities (1.00 and 1.00, respectively; Fig. 3B). Although
293 the mitochondrial phylogenetic tree indicated a single cluster for the individuals of *P. laevis*
294 lineage III and *P. pungitius*, these were not clustered together in the nuclear phylogenetic tree
295 (Fig. 3B). In the *P. laevis* cluster, individuals belonging to the lineage I and II formed a
296 subcluster with a high posterior probability (0.98), and those of the lineage II were indicated as
297 a monophyletic group (Fig. 3B). In contrast, the individuals of *P. laevis* lineage III did not form
298 a subcluster (Fig. 3B). The CADM test indicated that the phylogenetic relationships of the four
299 phylogenetic groups (*P. laevis* lineage I, lineage II, lineage III and *P. pungitius*) are incongruent
300 with those obtained from the mitochondrial data ($P = 0.58$).

301 The AMOVA revealed that 58.6% of the genetic variation was explained by phylogenetic
302 groups, whereas variance within phylogenetic groups and within populations accounted for 26.4%
303 and 15.0% of variance, respectively (Table 1). The F_{CT}, F_{SC} and F_{ST} values were 0.586, 0.637
304 and 0.850, respectively (Table 1). In the PCA, three principal components were identified to be
305 significant with inertia values of 29.3, 14.1 and 8.9 (Fig. 4). All *P. laevis* individuals formed a
306 single cluster distinct from *P. pungitius* individuals (Fig. 4). Within the *P. laevis* cluster, the
307 individuals from the same mitochondrial lineages tended to cluster together, but the individuals
308 of *P. laevis* lineage II were further separated into two subgroups (Fig. 4). Notably, the
309 individuals of *P. laevis* lineage III clustered in between *P. pungitius* and other *P. laevis* lineages
310 showing large spread along the first principal component axis (Fig. 4).

311 In the Bayesian admixture analysis with STRUCTURE, the delta K showed a clear peak
312 at K = 2, indicating that population structure was best explained by two genetic clusters (Fig.

S1, Supporting information). At K = 2, one genetic cluster was found for the individuals of *P. laevis* lineages I and II, and another cluster was observed for those of *P. pungitius* (Fig. 5). However, the individuals of *P. laevis* lineage III showed a pattern of admixture between these clusters (Fig. 5). At K = 3, *P. laevis* lineage I and II were separated into two different clusters, and *P. laevis* lineage III was indicated to be an admixture of *P. laevis* lineage I and *P. pungitius* (Fig. 5). At K = 4, *P. laevis* lineage II was divided into two subgroups, and at K = 5, *P. laevis* lineage III was indicated as an independent cluster, although admixture from *P. laevis* lineage I and/or *P. pungitius* were found in some individuals (Fig. 5).

Colonization history and recent gene flow

Both nucleotide diversity (pi) and haplotype diversity (Hd) in the mitochondrial data decreased significantly with increasing latitude in *P. laevis* lineage III (pi: $r^2 = 0.442$, N = 10, $P = 0.036$; Hd: $r^2 = 0.540$, N = 10, $P = 0.016$; Fig. 6). However, no such a trend was found in *P. pungitius* (pi: $r^2 = 0.289$, N = 5, $P = 0.084$; Hd: $r^2 = 0.033$, N = 5, $P = 0.77$). Hence, while *P. laevis* lineage III and *P. pungitius* belong to the same mitochondrial clade, they show different geographic patterns of mitochondrial diversity.

In the tests for recent gene flow with the nuclear genes, no significant correlation was found between genetic (F_{ST}) and geographic distance across *P. laevis* lineage III and *P. pungitius* sites ($r^2 = 0.006$, N = 28, $P = 0.52$), or in between the *P. laevis* lineage III and lineage I sites ($r^2 = 0.012$, N = 42, $P = 0.35$; Fig. S2, Supporting information). Likewise, no significant correlation was found between pairwise nucleotide difference and geographic distance across

P. laevis lineage III and *P. pungitius* sites (r^2 = 0.134, N = 28, *P* = 0.06), or in between the *P. laevis* lineage III and lineage I sites (r^2 = 0.023, N = 42, *P* = 0.34; Fig. S3, Supporting information). Thus, the null hypothesis of the presence of recent gene flow was rejected.

Discussion

Our results provide a basis to reject the hypothesis that *P. laevis* lineage III would be a phenotypically convergent form of *P. pungitius* which has lost its keel plates and become morphologically indistinguishable from *P. laevis*. Instead, the results provide evidence for historical interspecific introgression between *P. pungitius* and *P. laevis*, resulting in a formation of a new evolutionary lineage which appears to be morphologically indistinguishable from pure *P. laevis*, but carries traces of genomic admixture between the two parental species. Most notably, this introgression led to capture of *P. pungitius* mitogenome to *P. laevis* lineage III, but traces of nuclear introgression are clearly visible. The lack of evidence for recent gene flow between the species indicates that this secondary contact leading to the observed introgression took place historically.

Hybridization and mitochondrial introgression

While *P. laevis* lineage III clustered together with *P. pungitius* in the mitochondrial analysis, nuclear phylogenetic tree identified that all *P. laevis* individuals formed a monophyletic group distinct from *P. pungitius*. The Structure analyses indicated that *P. laevis* lineage III individuals are a genetic admixture between *P. laevis* and *P. pungitius*, suggesting that *P. laevis* lineage III

was formed in an asymmetric introgression between *P. pungitius* and *P. laevis*, which is also consistent with the PCA results. As hybridization between different species leads to transfer of alleles from one species to another, introgressed populations are generally expected to have higher genetic variability than either of the parental species (Katoh & Ribi, 1996). In fact, *P. laevis* lineage III was found to exhibit a higher level of genetic variation at nuclear genes than the other *P. laevis* lineages (and *P. pungitius*), further supporting the admixed origin of the *P. laevis* lineage III individuals. Since our data do not indicate ongoing gene flow either between *P. laevis* lineages I and III or between *P. laevis* lineage III and *P. pungitius*, hybridization likely occurred historically. Genetic introgression is often observed in stickleback fishes both in Eurasia and North America (Takahashi & Takata, 2000; Takahashi *et al*., 2016, Taylor & McPhail, 1999). Takahashi *et al*. (2016) reported extensive genetic introgression among several *Pungitius* species in East Asia, including that from *P. pungitius* to *P. sinensis*, as well as from *P. sinensis* to *P. tymensis* and *P. kaibarae*. The frequent occurrence of introgression in genus *Pungitius* could be due to relatively low degree of reproductive isolation among *Pungitius* species.

Given that *P. laevis* lineage III is widely spread over the Seine, Meuse and Mosel Rivers, a problem to be solved is how such a vast area became colonized by this lineage. Mitochondrial genetic diversity in *P. laevis* lineage III showed clear decrease with the increasing of latitude, indicating that the lineage may have gone through northward population expansion after glaciations. Given the northward latitudinal reduction of mitochondrial diversity in lineage III, hybridization might have occurred at a southern refugium when *P. pungitius* and *P. laevis*

retreat during glaciations and the newly formed lineage spread to the current distribution area during postglacial recolonization. Sediments at the upstream of the Seine and Aube Rivers were deposited during the last glacial period and filled incised valleys forming alluvial plains (Bendjoudi *et al*., 2002). This might have facilitated water connections in the Seine drainage, and provided a passage for the lineage to spread in it. Of course, given that the river networks in France have been strongly influenced by human activities including artificial canals (Persat & Keith, 2011), it is also possible that the spread of lineage III was assisted by humans.

Although *P. laevis* lineage III is genetically admixed by *P. pungitius* and *P. laevis* in the analyses of nuclear DNA, all the individuals carried the *P. pungitius* mitochondria. This type of asymmetric introgression, in which the mitochondria of one species is replaced by that of another (i.e. mitogenome capture), is rather common in fish and other taxonomic groups (Sousa-Santos *et al*., 2014; Carson & Dowling, 2006; Nevado *et al*., 2009; Toews & Brelsford, 2012). Asymmetric introgression can come about in various different ways. For example, asymmetric reproductive isolation, differences in generation length, selective sweeps and different dispersal distances between species can all cause asymmetric gene flow from one species to another (Chan & Levin, 2005; Crespin, Berrebi & Lebreton, 1999; Harrison & Larson, 2014). In East Asian *Pungitius* fishes, postzygotic reproductive isolation was found between freshwater and brackish-water types, in which mitochondrial introgression has occurred relatively recently (Takahashi, Tsuruta & Goto, 2003; Wiens, 2004), indicating that F_1 hybrid males are sterile but females are fertile (Takahashi, Nagai & Goto, 2005). Reproductive isolation has been proven also between landlocked and marine forms of the three-spined stickleback (*Gasterosteus*

aculeatus), in which mitochondrial introgression has occurred (Yamada, Higuchi & Goto, 2001). In this case, all the F_1 hybrid females were sterile in both pairing directions, and only male F_1 hybrids generated from female landlocked and male marine forms were fertile (Honma & Tamura, 1984). The asymmetric introgression from *P. pungitius* to *P. laevis* might have occurred due to such postzygotic reproductive isolation, although no information is currently available to evaluate this possibility. However, also differences in population size are considered as a possible reason for the mitochondrial introgression between the freshwater and brackish-water types and between the Pacific and Japan Sea three-spined stickleback forms (Yamada *et al*., 2001; Takahashi *et al.*, 2003). Similarly, the asymmetric introgression of *P. pungitius* mitochondrial DNA into *P. laevis* lineage III could also be due to other causes, such as possible selective advantage of the *P. pungitius* mitochondrial DNA on *P. laevis* genetic background in the environments inhabited by lineage III.

Geographic distribution of different lineages and species

The three *P. laevis* lineages and *P. pungitius* were found to be geographically clearly isolated from each other and distributed mostly in different drainage systems. *P. laevis* lineage II occurred in southwestern France including the Charente and Dordogne Rivers and their tributaries or vicinities, whereas *P. laevis* lineage I occurred in central France in the Loire River drainage and in upstream parts of the Loing River. *P. laevis* lineage III was confined at the Seine, Meuse and Mosel River basins, whereas *P. pungitius* was confined in the Saône basin. The divergence between *P. laevis* lineages I and II traces back to the Pleistocene glaciations

(1.95 Ma; Wang *et al*., 2015), which strongly affected the biogeography of many species in France (Gouin *et al*., 2006). Historically, the southernmost range of ice sheets reached the northern France (Hewitt, 2004; Buoncristiani & Campy, 2004), and while regions from the Seine-Normandie basin to the northern part of the Adour-Garonne basin, which contains the Charente and Dordogne Rivers, were not covered by ice, they experienced continuous or discontinuous permafrost during Pleistocene glaciations (Bertran *et al*., 2014). However, the southern part of the Adour-Garonne basin was nearly free of ice during the Pleistocene, which may suggest a refugium for the *P. laevis* lineage II (Bertran *et al*., 2014). Several refugial areas have been identified in southern France, and these refugia are frequently associated with the divergence between different lineages of freshwater species. For example, freshwater crayfish (*Austropotamobius pallipes*) was found to have diverged into three deep lineages ($\Phi st = 0.731$) distributed in river basins in southern, northwestern and eastern France, respectively. The intraspecific divergence in this species was inferred to have risen from retreat to different refugia during glaciations. One refugium was possibly located in south-western France when the Adour-Garonne basin was free of ice, while the others were suggested to be located at the Mediterranean coast and Rhine basin (Gouin *et al*., 2001). Similarly, brown trout (*Salmo trutta* L.) diverged into five lineages in Eurasia during Pleistocene glaciations (0.5–2.0 Mya; Bernatchez, 2001). Accordingly, five glacial refugia are suggested for this species, two of which were located in southern France with corresponding lineages currently distributed to eastern and northwestern France (Bernatchez, 2001). For graylings (*Thymallus thymallus*) and dace (*Leuciscus leuciscus*), the Loire region has been suggested to be a refugium, although it is

located close to periglacial area (Weiss *et al*., 2002; Costedoat & Gilles, 2009). By inference, as both central and southern France have been identified as refugia for some species, it is possible that the ancestors of *P. laevis* lineage I have inhabited the Loire River region as a refugium, and become isolated from those of the lineage II, which currently inhabits the Charente and Dordogne Rivers. The Loire River has been mostly independent from the Adour-Garonne basin since 4.5 Mya (Persat & Keith, 2011), and the separation of these drainages may have facilitated the divergence between the lineages I and II. Moreover, *P. pungitius* in the Saône catchment represents the southernmost populations of this species in Europe. These *P. pungitius* populations are isolated from other western European *Pungitius* populations. Thus, it seems likely that the Saône basin was a refugium to *P. pungitius*, and the population in this catchment persisted there during and after glaciations. A possible scenario for the origin of *P. laevis* lineage III is that *P. laevis* lineage I retreated to the upstream of the Seine River during the glaciations, and met *P. pungitius* that formerly resided in the area or colonized from the Saône watershed through periodic waterway connection caused by ancient river capture. This contact between the two species may have resulted in the formation of *P. laevis* lineage III.

Comparative phylogeographic studies of various species have discovered that genetic diversity of species or lineages currently inhabiting formerly glaciated regions have lower genetic diversity than species of lineages occurring in south (Hewitt, 1996; Bernatchez & Wilson, 1998). By inference, the higher genetic diversity of *P. laevis* lineage II than the two other *P. laevis* lineages and *P. pungitius* implies that *P. laevis* lineage II might have been less affected by glaciations than the others. Since the *P. laevis* lineage II was inferred to have

occupied the southernmost refugium, the results concur with general patterns seen in other studies. However, in the mitochondrial phylogenetic tree of *P. laevis* lineage II, populations from the Charente and Dordogne Rivers form two distinct clusters with haplotypes from both rivers. This clustering pattern could be explained by admixture owing to historical sea level fluctuations, which occurred during Quaternary glaciations in European regions (Lericolais *et al.*, 2007; Mellett *et al*., 2013). It is widely recognized that freshwater fishes are forced to retreat to inlands during sea level upraises, and recolonization towards the coastline occurs when sea level drops (De Bruyn & Mather, 2007; Swartz *et al.*, 2014). However, an alternative explanation to these patterns is provided by artificial transfers by humans.

Conclusion

The results identified and mapped the occurrence and phylogeographic distributions of three distinct *P. laevis* lineages in France, as well provided evidence for interspecific hybridization between *P. laevis* and *P. pungitius* being behind the formation of *P. laevis* lineage III. The two other deep lineages (I and II) of *P. laevis* have likely diverged from each other in different refugia during glaciations. Although future work is needed to test the extent of reproductive isolation between *P. pungitius* and *P. laevis*, as well as between the different *P. laevis* lineages, the results add to the evidence that interspecific hybridization between closely related fish species has been probably more common than previously thought.

Acknowledgements

We thank Scott McCairns and Alexandre Budria for their help with translating French references, and Jacquelin De Faveri for linguistic corrections for an earlier draft of this manuscript. We thank the two anonymous reviewers for comments that improved this manuscript. Our research was supported by the Academy of Finland (AF, grant #s: 129662, 134728 and 218343; JM), Chinese Scholarship Council (CW) and Researcher Exchange Program between JSPS and AF (TS).

References

April J, Hanner RH, Dion-Côté AM, Bernatchez L. 2013. Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. *Molecular Ecology* **22**: 409-422.

Avise JC, Walker D. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. *Proceedings of the Royal Society B: Biological Sciences* **265**: 457–463.

Bendjoudi H, Weng P, Guerin R, Pastre JF. 2002. Riparian wetlands of the middle reach of the Seine river (France): historical development, investigation and present hydrologic functioning. A case study. *Journal of Hydrology* **263**: 131-155.

Bernatchez L, Wilson CC. 1998. Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology* **7**: 431–452.

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.

Bertran P, Andrieux E, Antoine P, Coutard S, Deschodt L, Gardère P, Hernandez M, Legentil C, Lenoble A, Liard M, Mercier N, Moine O, Sitzia L, Vliet-Lanoë BV. 2014. Distribution and chronology of Pleistocene permafrost features in France: database and first results. *Borea*s, **43**, 699-711.

Blum MD, Törnqvist TE. 2000. Fluvial responses to climate and sea‐level change: a review and look forward. *Sedimentology* **47**: 2-48.

Bonferroni CE. 1936. Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze **8**: 3–62.

Boratyński Z, Alves PC, Berto S, Koskela,E, Mappes T, Melo-Ferreira J. 2011. Introgression of mitochondrial DNA among Myodes voles: consequences for energetcuiwangics? *BMC Evolutionary Biology* **11**: 1.

Buoncristiani JF, Campy M. 2004. Palaeogeography of the last two glacial episodes in the

Massif Central, France. *Developments in Quaternary Sciences* **2**: 111-112.

Campbell V, Legendre P, Lapointe FJ. 2011. The performance of the Congruence Among Distance Matrices (CADM) test in phylogenetic analysis. *BMC Evolutionary Biology* **11:** 1.

Carson EW, Dowling TE. 2006. Influence of hydrogeographic history and hybridization on the distribution of genetic variation in the pupfishes *Cyprinodon atrorus* and *C. bifasciatus*. *Molecular Ecology* **15**: 667-679.

Costedoat C, Gilles A. 2009. Quaternary pattern of freshwater fishes in Europe: comparative phylogeography and conservation perspective. *The Open Conservation Biology Journal* **3**: 36-48.

Chan KM, Levin SA. 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution* **59**: 720-729.

Crespin L, Berrebi P, Lebreton JD. 1999. Asymmetrical introgression in a freshwater fish hybrid zone as revealed by a morphological index of hybridization. *Biological Journal of the Linnean Society* **67**: 57-72.

Darras H, Aron S. 2015. Introgression of mitochondrial DNA among lineages in a hybridogenetic ant. *Biology Letters* **11**: 20140971.

De Bruyn M, Mather PB. 2007. Molecular signatures of Pleistocene sea-level changes that affected connectivity among freshwater shrimp in Indo-Australian waters. *Molecular Ecology* **16**: 4295-4307.

Earl DA. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359-361.

Elphinstone MS, Hinten GN, Anderson MJ, Nock CJ. 2003. An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. *Molecular Ecology Notes* **3**: 317–320.

Gouin N, Grandjean F, Bouchon D, Reynolds JD, Souty-Grosset C. 2001. Population genetic structure of the endangered freshwater crayfish *Austropotamobius pallipes*, assessed using RAPD markers. *Heredity* **87**: 80-87.

Gross HP. 1979. Geographic variation in European ninespine sticklebacks, *Pungitius pungitius*. *Copeia* **1979**: 405–412.

Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity* **105**: 795-809.

Hawkins BA, Porter EE. 2003. Relative influences of current and historical factors on mammal and bird diversity patterns in deglaciated North America. *Global Ecology and Biogeography* **12**: 475-481.

Hewitt GM. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**: 247-276.

Hewitt GM. 1999. Postglacial recolonization of European biota. *Biological Journal of the Linnean Society* **68**: 87-112.

Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907-913.

Hewitt GM. 2001. Speciation, hybrid zones and phylogeography—or seeing genes in space and time. *Molecular Ecology* **10**: 537-549.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**: 183-195.

Hewitt GM. 2011. Quaternary phylogeography: the roots of hybrid zones. *Genetica* **139**: 617-638.

Hey J, Wakeley J. 1997. A coalescent estimator of the population recombination rate. *Genetics* **145**: 833-846.

Honma Y, Tamura E. 1984. Anatomical and behavioral difference among threespine sticklebacks: the marine form, the landlocked form and their hybrids. *Acta Zoologica* **65**: 79–87

Irwin DE, Rubtsov AS, Panov EN. 2009. Mitochondrial introgression and replacement between yellowhammers (*Emberiza citrinella*) and pine buntings (*Emberiza leucocephalos*) (Aves: Passeriformes). *Biological Journal of the Linnean Society* **98**: 422-438.

Katoh M, Ribi G. 1996. Genetic evidence for natural hybridization and apparent introgression between freshwater snail species (*Viviparus ater* and *V. contectus*). *Journal of Evolutionary Biology* **9**: 67-82.

Keivany Y, Nelson JS. 2000. Taxonomic review of the genus *Pungitius*, ninespine sticklebacks (Gasterosteidae). *Cybium: international journal of ichthyology* **24**: 107-122.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 6196-6200.

Kottelat M, Freyhof J. 2007. Handbook of European freshwater fishes. Kottelat, Cornol, Switzerland and Freyhof, Berlin, Germany. pp 267–299.

Lericolais G, Popescu I, Guichard F, Popescu SM, Manolakakis L. 2007. Water-level fluctuations in the Black Sea since the Last Glacial Maximum. In: Yanko-Hombach V, Gilbert AS, Panin N, Dolukhanov PM (eds) *The Black Sea Flood Question: Changes in Coastline, Climate, and Human Settlement*, pp 437-452. Springer, Netherlands.

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

Li C, Ortí G, Zhang G, Lu G. 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evolutionary Biology* **7**: 44.

Li C, Riethoven JJM, Ma L. 2010. Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. *BMC Evolutionary Biology* **10**: 90.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451-1452.

Makhrov AA, Bolotov IN. 2006. Dispersal routes and species identification of freshwater animals in Northern Europe: a review of molecular evidence. *Russian Journal of Genetics* **42**: 1101-1115.

Mallet J, Beltrán M, Neukirchen W, Linares M. 2007. Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology* **7**: 28.

McGlashan DJ, Hughes JM. 2001. Low levels of genetic differentiation among populations of the freshwater fish *Hypseleotris compress*a (Gobiidae: Eleotridinae): implications for

its biology, population connectivity and history. *Heredity* **86**: 222-233.

Melletta CL, Hodgsonc DM, Platera AJ, Mauza B, Selbyb I, Langa A. 2013. Denudation of the continental shelf between Britain and France at the glacial–interglacial timescale. *Geomorphology* **203**: 79-96.

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

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.

Mäkinen H, Merilä J. 2008. Mitochondrial DNA phylogeography of the three-spined stickleback (*Gasterosteus aculeatus*) in Europe—evidence for multiple glacial refugia. *Molecular phylogenetics and evolution* 46: 167-182.

Münzing J. 1969. Variabilität, Verbreitung und Systematik der Arten und Unterarten in der Gattung Pungitius Coste, 1848 (Pisces, Gasterosteidae). *Journal of Zoological Systematics and Evolutionary Research* **7**: 208-233.

Nevado B, Koblmüller S, Sturmbauer C, Snoeks J, Usano-Alemany J, Verheyen E. 2009. Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish. *Molecular Ecology* **18**: 4240-4255.

Orti G, Bell MA, Reimchen TE, Meyer A. 1994. Global survey of mitochondrial DNA sequences in the threespine stickleback: evidence for recent migrations. *Evolution* **48**: 608-622.

Paepke HJ. 1996. Die Stichlinge: Gasterosteidae.-2. überarbeitete und ergänzte Auflage.-Die Neue Brehmbücherei 10, Westarp Verlag Magdeburg. pp.175

Paepke HJ. 2001. *Pungitius pungitius* (Linnaeus, 1758). *The freshwater fishes of Europe, vol. 5* (ed. by P. Bănărescu and H.J. Paepke), pp. 277–299. Aula-Verlag, Wiesbaden.

Palumbi SR. 1996. Nucleic acids. II. The polymerase chain reaction. *Molecular systematics, 2nd ed*. (ed. by D.M. Hillis, C. Moritz and B.K. Mable), pp. 205–248. Sinauer Associates, Sunderland, MA.

Pamilo P, Savolainen O. 1999. Post-glacial colonization, drift, local selection and conservation value of populations: A northern perspective. *Hereditas* **130**: 229-238.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genetics* **2**: e190.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**: 2537-2539.

Persat, H, Keith P. 2011. Biogéographie et historique de la mise en place des peuplements ichtyologiques de France métropolitaine. In: Keith P, Persat H, Feunteun E, Allardy J (eds) *Les poissons d'eau douce de France,* pp 36-88. Editions Biotope-Mèze and MNHN-Paris.

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

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

Remington CL. 1968. Suture-zones of hybrid interaction between recently joined biotas.

644 *Evolutionary Biology* **2**: 321-428.
645 **Rhymer JM, Simberloff D. 1996.** Extinction by hybridization and introgression. *Annual*
646 *Review of Ecology and Systematics* **27**: 83-109.
647 **Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu**
648 **L, Suchard MA, Huelsenbeck JP. 2012.** MrBayes 3.2: efficient Bayesian phylogenetic
649 inference and model choice across a large model space. *Systematic Biology* **61**: 539-542.
650 **Rousset F. 2008.** GENEPOP'007: a complete re-implementation of the genepop software for
651 Windows and Linux. *Molecular Ecology Resources* **8**: 103-106.**Schmitt, T. 2007.**
652 Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers*
653 *in Zoology* **4**: 11.
654 **Schumer M, Rosenthal GG, Andolfatto P. 2014.** How common is homoploid hybrid
655 speciation? *Evolution* **68**: 1553-1560.
656 **Shikano T, Shimada Y, Herczeg G, Merilä J. 2010.** History vs. habitat type: explaining the
657 genetic structure of European nine-spined stickleback (*Pungitius pungitius*) populations.
658 *Molecular Ecology* **19**: 1147-1161.
659 **Slatkin M. 1987.** Gene flow and the geographic structure of natural populations. *Science* **236**:
660 787-792.
661 **Smith GR, Dowling TE. 2008.** Correlating hydrographic events and divergence times of
662 speckled dace (Rhinichthys: Teleostei: Cyprinidae) in the Colorado River drainage.
663 *Geological Society of America Special Papers* **439**: 301-317.
664 **Sousa-Santos C, Gante HF, Robalo J, Cunha PP, Martins A, Arruda M, Alves MJ,**
665 **Almada V. 2014.** Evolutionary history and population genetics of a cyprinid fish
666 (*Iberochondrostoma olisiponensis*) endangered by introgression from a more abundant
667 relative. *Conservation Genetics* **15**: 665-677.
668 **Spellman GM, Riddle B, Klicka J. 2007.** Phylogeography of the mountain chickadee (*Poecile*
669 *gambeli*): diversification, introgression, and expansion in response to Quaternary climate
670 change. *Molecular Ecology* **16**: 1055-1068.
671 **Stewart JR, Lister AM. 2001.** Cryptic northern refugia and the origins of the modern biota.
672 *Trends in Ecology & Evolution* **16**: 608–613.
673 **Swartz ER, Chakona A, Skelton PH, Bloomer P. 2014.** The genetic legacy of lower sea
674 levels: does the confluence of rivers during the last glacial maximum explain the
675 contemporary distribution of a primary freshwater fish (*Pseudobarbus burchelli,*
676 *Cyprinidae*) across isolated river systems? *Hydrobiologia* **726**: 109-121.
677 **Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F. 1998.** Comparative phylogeography
678 and postglacial colonization routes in Europe. *Molecular Ecology* **7**: 453-464.
679 **Takahashi H, Goto A. 2001.** Evolution of East Asian ninespine sticklebacks as shown by
680 mitochondrial DNA control region sequences. *Molecular Phylogenetics and Evolution*
681 **21**: 135-155.
682 **Takahashi H, Nagai T, Goto A. 2005.** Hybrid male sterility between the fresh-and brackish-
683 water types of ninespine stickleback *Pungitius pungitius* (Pisces, Gasterosteidae).
684 *Zoological Science* **22**: 35-40.
685 **Takahashi H, Møller PR, Shedko VS, Ramatulla T, Joen SR, Zhang CG, Sideleva VG,**
686 **Takata K, Sakai H, Goto A, Nishida M. 2016.** Species phylogeny and diversification

process of Northeast Asian Pungitius revealed by AFLP and mtDNA markers. *Molecular Phylogenetics and Evolution* **99**: 44-52.

Takahashi H, Takata K. 2000. Multiple lineages of the mitochondrial DNA introgression from *Pungitius pungitius* (L.) to *Pungitius tymensis* (Nikolsky). *Canadian Journal of Fisheries and Aquatic Sciences* **57**: 1814-1833.

Takahashi H, Tsuruta T, Goto A. 2003. Population structure of two ecologically distinct forms of ninespine stickleback, *Pungitius pungitius*: gene flow regimes and genetic diversity based on mtDNA sequence variations. *Canadian Journal of Fisheries and Aquatic Sciences* **60**: 421-432.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725-2729.

Tanabe AS. 2007. Kakusan: a computer program to automate the selection of a nucleotide substitution model and the configuration of a mixed model on multilocus data. *Molecular Ecology Notes* **7**: 962-964.

Taylor EB, McPhail J. 1999. Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (*Gasterosteus*): insights from mitochondrial DNA. *Biological Journal of the Linnean Society* **66**: 271-291.

Toews DP, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* **21**: 3907-3930.

Wang C, Shikano T, Persat H, Merilä J. 2015. Mitochondrial phylogeography and cryptic divergence in the stickleback genus Pungitius. *Journal of Biogeography* **42**: 2334-2348.

Weiss S, Persat H, Eppe R, Schlötterer C, Uiblein F. 2002. Complex patterns of colonization and refugia revealed for European grayling *Thymallus thymallus*, based on complete sequencing of the mitochondrial DNA control region. *Molecular Ecology* **11**: 1393-1407.

Whitfield (née Maher) E, Harvey AM. 2012. Interaction between the controls on fluvial system development: tectonics, climate, base level and river capture-Rio Alias, Southeast Spain. *Earth Surface Processes and Landforms* **37**: 1387-1397.

Wiens JJ, Donoghue MJ. 2004. Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution* **19**: 639-644.

Wiens JJ, Kuczynski CA, Stephens PR. 2010. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. *Biological Journal of the Linnean Society* **99**: 445-461.

Wootton RJ. 1976. *The biology of the sticklebacks*. Academic Press, London.

Yamada M, Higuchi M, Goto A. 2001. Extensive introgression of mitochondrial DNA found between two genetically divergent forms of threespine stickleback, *Gasterosteus aculeatus*, around Japan. *Environmental Biology of Fishes* **61**: 269-284.

Zink RM. 1994. The geography of mitochondrial DNA variation, population structure, hybridization, and species limits in the fox sparrow (*Passerella iliaca*). *Evolution* **48**: 96-111.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1. Genetic diversity at mitochondrial cytochrome *b* gene and eight nuclear genes in 25 *P. laevis* and five *P. pungitius* sites.

Table S2. Primers and PCR conditions used for mitochondrial and nuclear genes.

Table S3. Mitochondrial haplotype information in 25 *P. laevis* and five *P. pungitius* sites.

Table S4. Sequence information of nuclear genes in 18 *P. laevis* and four *P. pungitius* sites.

Fig. S1. Delta K values at K = 2 to K = 8 in the STUCTURE analyses based on nuclear genetic data.

Fig. S2. Relationships between nuclear genetic distance (F_{ST}) and geographic distance in the data between (A) the *P. laevis* lineage III and *P. pungitius* sites and between (B) the *P. laevis* lineage III and I sites.

Fig. S3 Relationships between nuclear genetic distance (pairwise nucleotide difference) and geographic distance in the data between (A) the *P. laevis* lineage III and *P. pungitius* sites and between (B) the *P. laevis* lineage III and I sites.

Legends to figures

Fig. 1 Sampling sites of *P. laevis* and *P. pungitius* used in this study. The four different colours of site IDs and symbols indicate different phylogenetic groups identified by mitochondrial and nuclear DNA data. Sampling sites with circle symbols were used in both mitochondrial and nuclear analyses, whereas those with square symbols were used only in mitochondrial analyses. The seven sampled main rivers are indicated in shaded colors. The dashed circle and solid circle show the distribution areas of *P. laevis* and *P. pungitius*, respectively.

Fig. 2 Comparisons of genetic diversity among four different phylogenetic groups based on (A) mitochondrial nucleotide diversity and (B) haplotype diversity, as well as on (C) expected heterozygosity and (D) the number of alleles at nuclear loci. The vertical bars represent the standard errors of the mean. Statistically significant comparisons are indicated with asterisks (*$P < 0.05$ and **$P < 0.01$).

Fig. 3 Bayesian phylogenetic trees of *P. laevis* and *P. pungitius* individuals based on (A) mitochondrial and (B) nuclear genetic data. Different phylogenetic groups are indicated with differently coloured bars. The numbers represent the posterior probability (>0.85) of each node. Numbers in brackets indicate the number of individuals in the same population.

Fig. 4 Scatter plot of *P. laevis* and *P. pungitius* individuals based on three principal components (PC1, PC2 and PC3) of nuclear data. The symbols in blue, yellow, red and green indicate the individuals of *P. laevis* lineage I, lineage II and lineage III and *P. pungitius*, respectively.

Fig. 5 Bayesian clustering of *P. laevis* and *P. pungitius* individuals at K = 2 to K = 5 based on nuclear data. Each vertical bar represents each individual. The results based on K = 2 was indicated to be the optimal number of clusters in the data (Fig. S1, Supporting information).

Fig. 6 Mitochondrial (A) nucleotide and (B) haplotype diversity as a function of latitude in the *P. laevis* lineage III sites.

Tables

Table 1. AMOVA statistics in mitochondrial and nuclear data

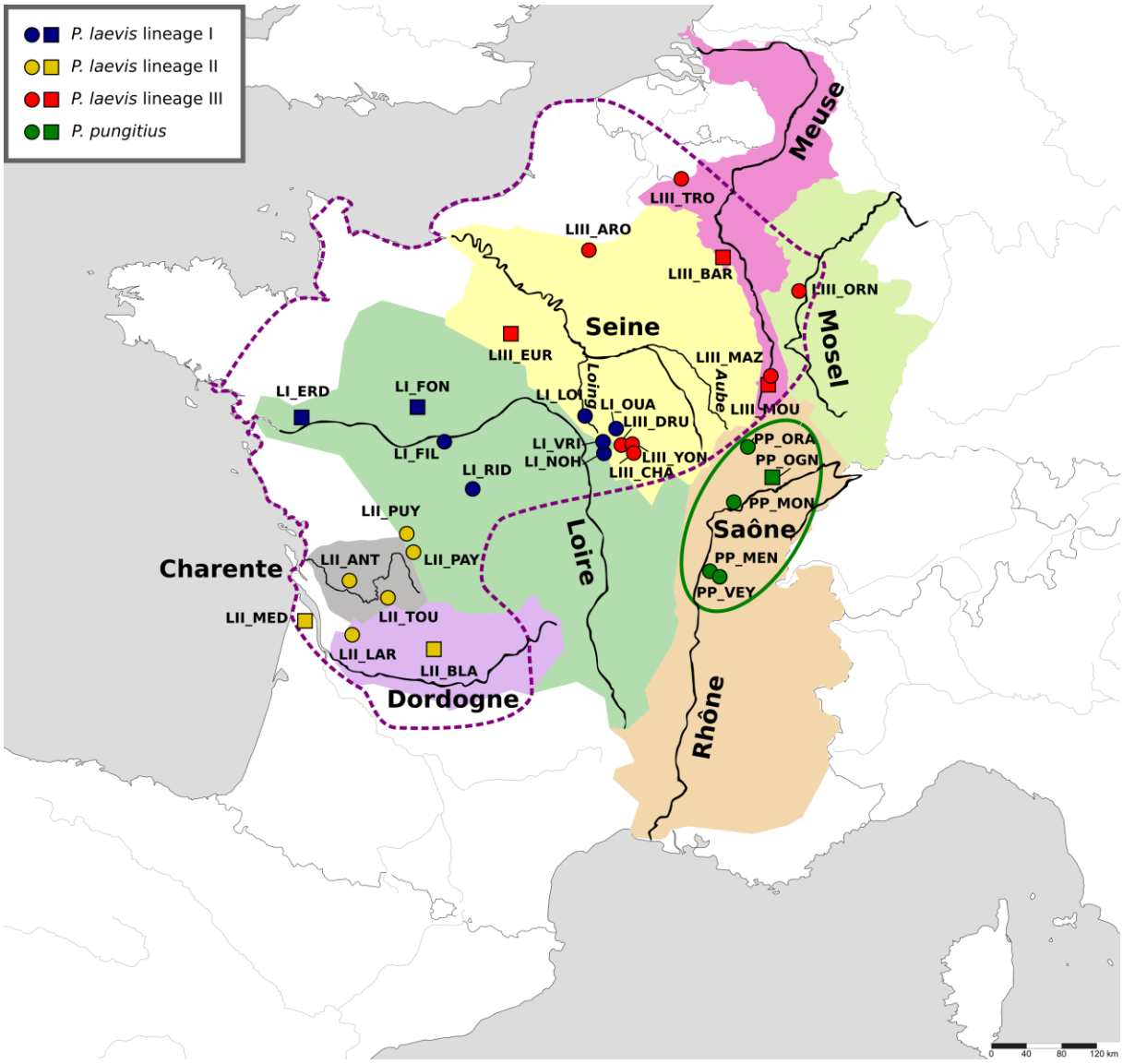

Fig. 1 Sampling sites of *P. laevis* and *P. pungitius* used in this study. The four different colours of site IDs and symbols indicate different phylogenetic groups identified by mitochondrial and nuclear DNA data. Sampling sites with circle symbols were used in both mitochondrial and nuclear analyses, whereas those with square symbols were used only in mitochondrial analyses. The seven sampled main rivers are indicated in shaded colors. The dashed circle and solid circle show the distribution areas of *P. laevis* and *P. pungitius*, respectively.

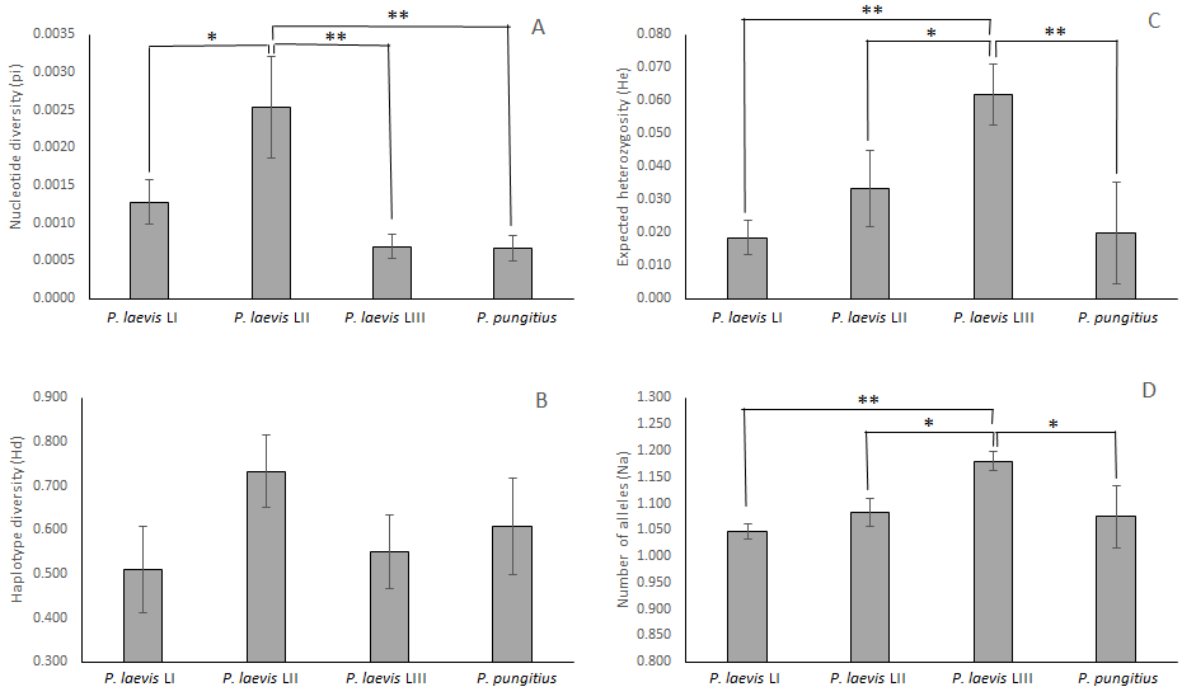

Fig. 2 Comparisons of genetic diversity among four different phylogenetic groups based on (A) mitochondrial nucleotide diversity and (B) haplotype diversity, as well as on (C) expected heterozygosity and (D) the number of alleles at nuclear loci. The vertical bars represent the standard errors of the mean. Statistically significant comparisons are indicated with asterisks (*P < 0.05 and **P < 0.01).

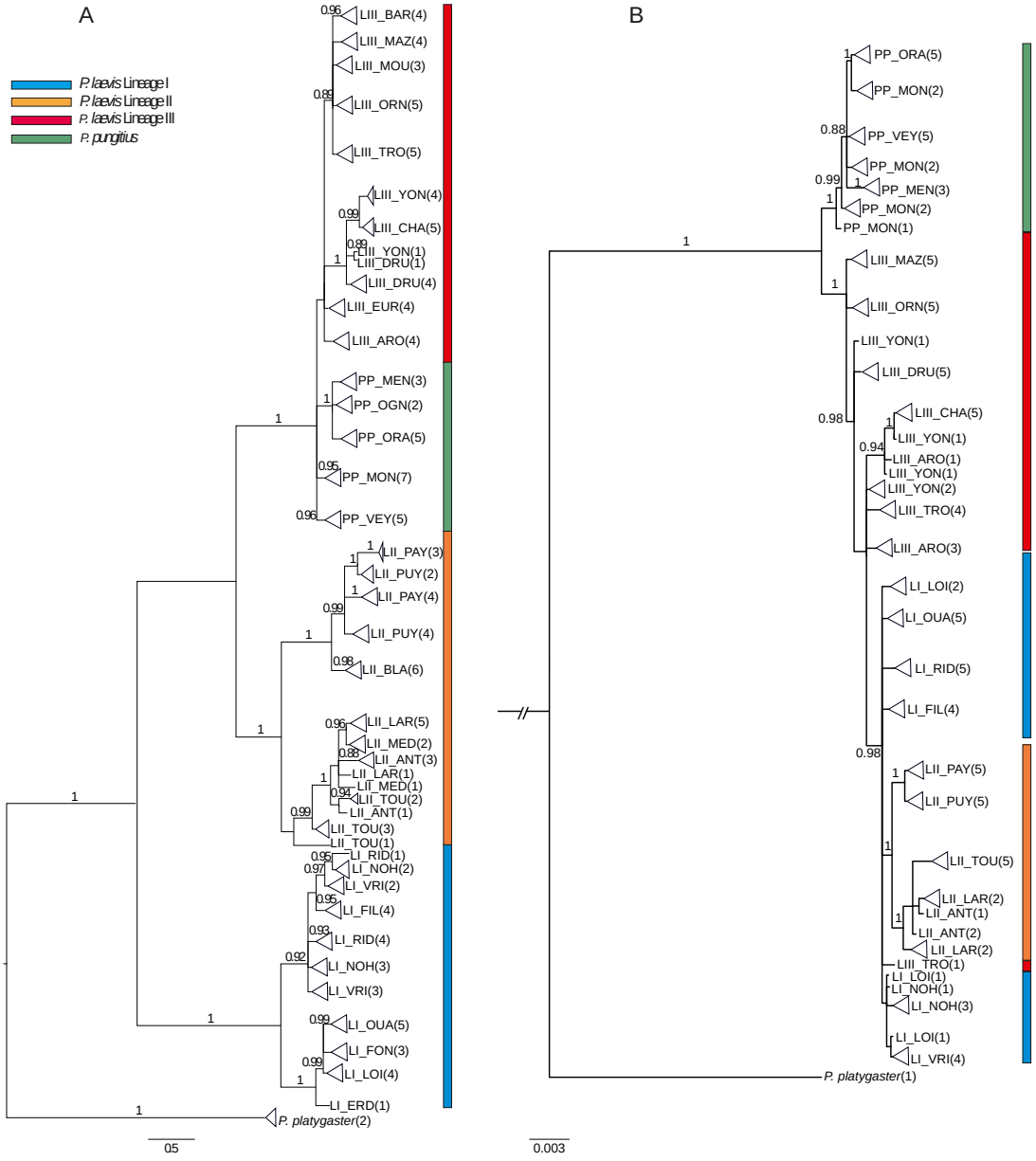

Fig. 3 Bayesian phylogenetic trees of *P. laevis* and *P. pungitius* individuals based on (A) mitochondrial and (B) nuclear genetic data. Different phylogenetic groups are indicated with differently coloured bars. The numbers represent the posterior probability (>0.85) of each node. Numbers in brackets indicate the number of individuals in the same population.

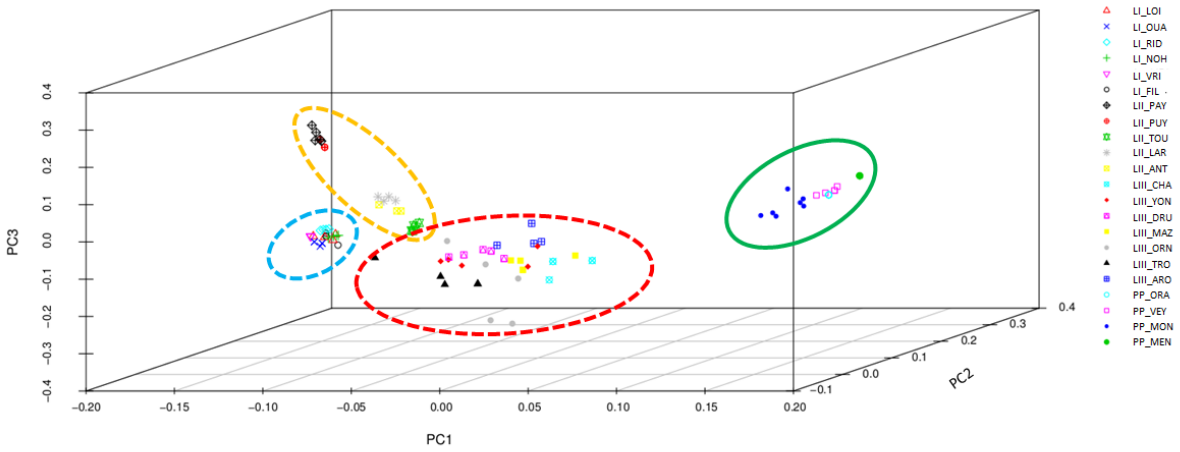

Fig. 4 Scatter plot of *P. laevis* and *P. pungitius* individuals based on three principal components (PC1, PC2 and PC3) of nuclear data. The symbols in blue, yellow, red and green indicate the individuals of *P. laevis* lineage I, lineage II and lineage III and *P. pungitius*, respectively.

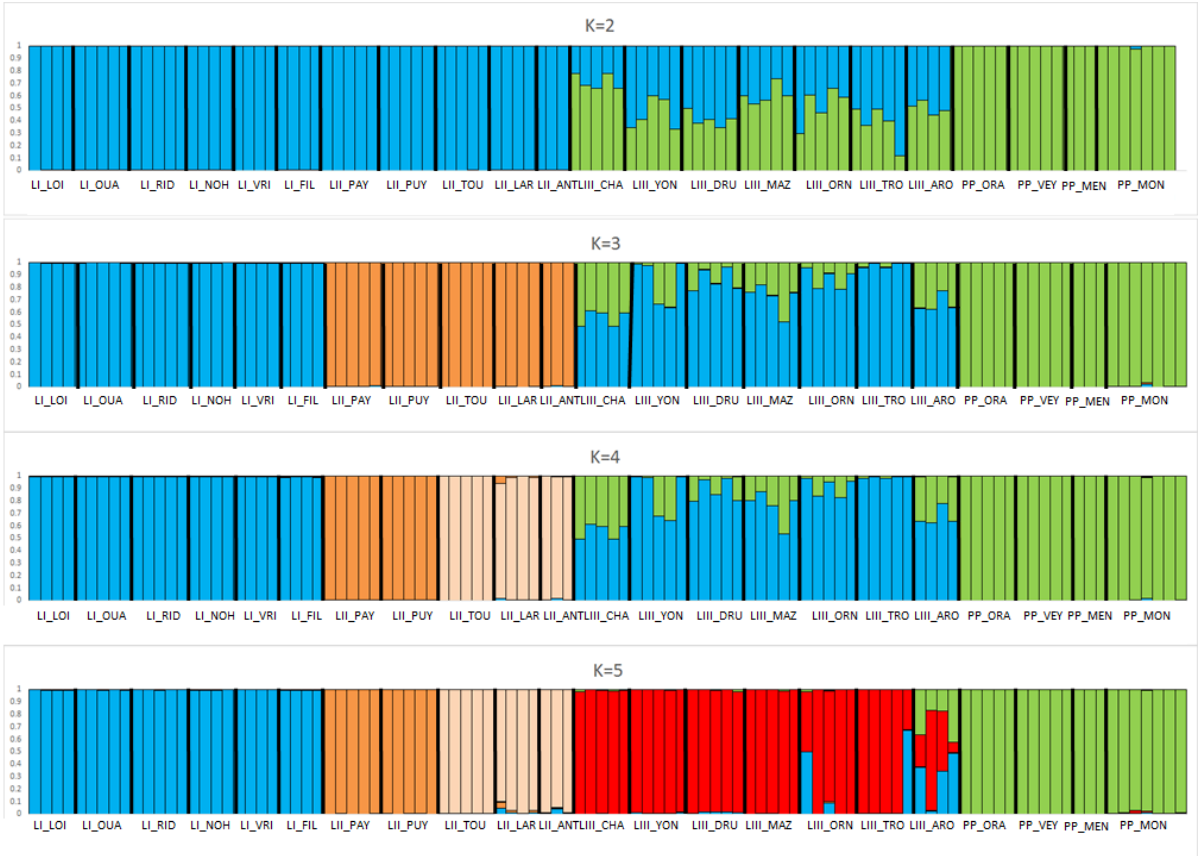

Fig. 5 Bayesian clustering of *P. laevis* and *P. pungitius* individuals at K = 2 to K = 5 based on nuclear data. Each vertical bar represents each individual. The results based on K = 2 was indicated to be the optimal number of clusters in the data (Fig. S1, Supporting information).

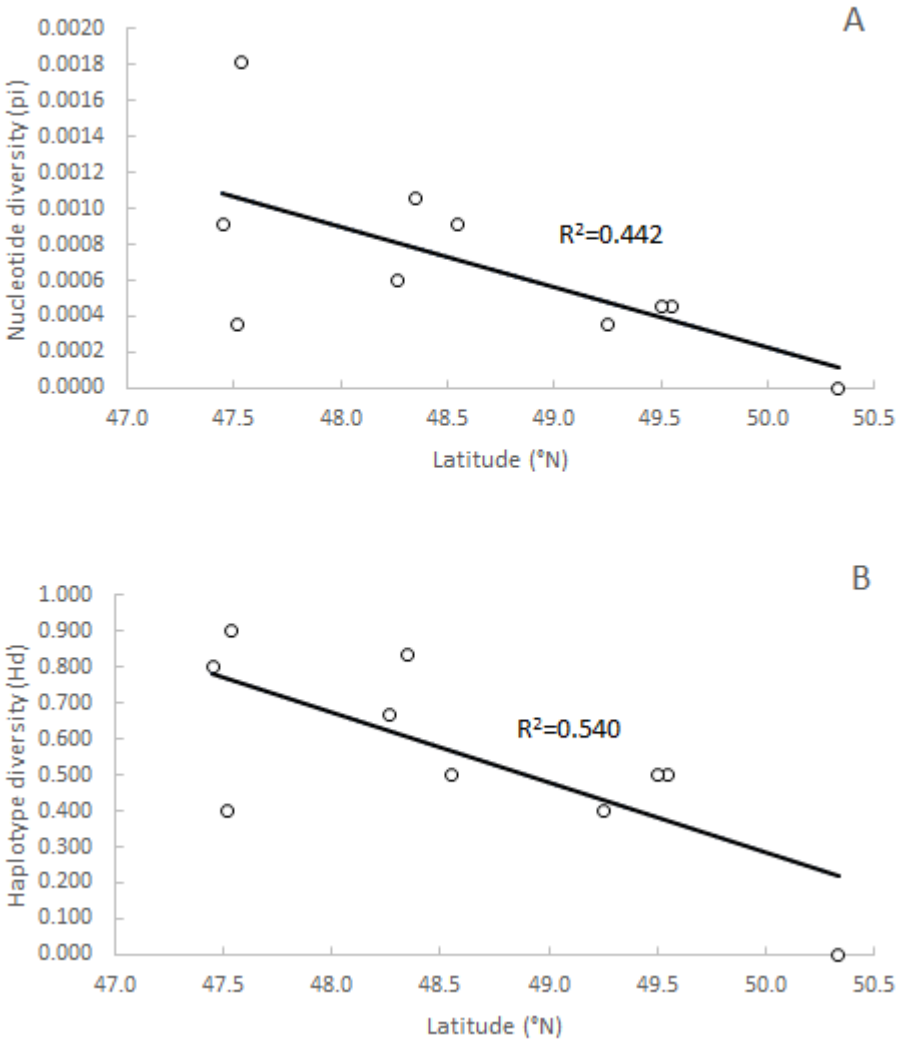

Fig. 6 Mitochondrial (A) nucleotide and (B) haplotype diversity as a function of latitude in the *P. laevis* lineage III sites.

Table 1. AMOVA statistics in mitochondrial and nuclear data

Source of variation	Mitochondrial data				Nuclear data			
	d.f.	Sum of squares	Variance component	Percentage of variation	d.f.	Sum of squares	Variance component	Percentage of variation
Among groups	3	1880.2	18.3	81.3	3	1025.0	6.2	58.6
Among populations within group	26	418.7	3.4	15.3	18	492.8	2.8	26.4
Within populations	106	80.2	0.8	3.4	182	290.5	1.6	15.0

Mitochondrial data: F_{CT} = 0.813, F_{ST} = 0.966 and F_{SC} = 0.819; nuclear data: F_{CT} = 0.586, F_{ST} = 0.850 and F_{SC} = 0.637.