Phylogeography of the prickly sculpin (Cottus asper) in north-western North America reveals parallel phenotypic evolution across multiple coastal–inland colonizations

Stefan Dennenmoser1,2*, Arne W. Nolte2, Steven M. Vamosi1 and Sean M. Rogers1

1Department of Biological Sciences, University of Calgary, Calgary AB T2N 1N4, Canada, 2Max-Planck Institute for Evolutionary Biology, 24306 Plön, Germany

ABSTRACT

Aim Glacial cycles during the Pleistocene may have frequently contributed to parallel evolution of phenotypes across independently evolving genetic lineages associated with separate glacial refugia. Previous studies based on morphology suggested that the prickly sculpin (Cottus asper) survived the Last Glacial Maximum (LGM) in southern coastal and inland refugia, favouring allopatric divergence between coastal and inland prickling phenotypes, which vary in the degree to which spine-like scales cover the body of the fish. Herein, we aimed to test whether parallel evolution across multiple genetic lineages rather than a single-lineage origin of highly prickled inland sculpins could serve as an explanation for the biogeographical distribution of prickling phenotypes.

Location North-western North America, Southeast Alaska and Canada (British Columbia).

Methods We used data from mitochondrial haplotypes and 19 microsatellite loci to identify distinct genetic lineages as a basis to interpret patterns of phenotypic evolution.

Results The occurrence of multiple mtDNA groups suggests that highly prickled inland phenotypes comprise more than one genetic lineage. Both mtDNA and microsatellite data are consistent with post-glacial dispersal along the coast and repeated coastal to inland colonization events, as opposed to inland dispersal of a single lineage from a southern refugium to northern regions.

Main conclusions Our results suggest that highly prickled inland phenotypes evolved repeatedly following multiple inland colonization events, probably via coastal rivers. The prickly sculpin therefore provides an example of recent (post-glacial) parallel evolution, potentially facilitated by standing genetic variation already present in the ancestral coastal populations.

Keywords Cottus asper, fish, morphometrics, multiple colonization events, North America, Pacific Northwest, parallel evolution, phylogeography, Pleistocene.

INTRODUCTION

The north-west of North America is characterized by a complex geomorphological history associated with multifaceted phylogeographical patterns largely shaped by repeated glaciations during the Pleistocene (c. 2.5 Ma–12,000 years ago; reviewed in Shafer et al., 2010). The increasing knowledge of the post-glacial history of numerous taxa in this region creates unique opportunities to investigate evolutionary consequences of allopatric separation in glacial refugia, rapid range expansions into recently deglaciated areas, and mixing of genetic lineages in zones of secondary contact (Shafer et al., 2010). Such investigations elucidate the relative importance of historical and ecological effects on intraspecific geographical variation (Avise et al., 1987; Shikano et al., 2010), and are therefore integral towards testing predictions about
Phylogeography reveals parallel evolution

The prickly sculpin, Cottus asper Richardson, 1836, is a euryhaline fish species characterized by both amphidromous and purely freshwater populations. Consistent phenotypic variants have been observed across a wide distributional range between southern Alaska and California, representing potentially diverse historical backgrounds (Krejsa, 1965; McPhail, 2007). Post-glacial colonization of inland regions has resulted in the predominance of a highly prickled phenotype characterized by a dense lateral coverage with spiny, scale-like epidermal structures (Krejsa, 1965). A heritable genetic basis of prickling probably exists, because a major quantitative trait locus (QTL) that affects prickling has been found in the European sculpin, Cottus perifretum (Cheng, 2013). Geographically disjunct distributions in prickling intensity have been found in additional Cottus species in Europe and North America (Koli, 1969; Freyhof et al., 2005; McPhail, 2007), but we are not aware of any apparent phenotype-environment associations that would indicate a common adaptive function of prickling.

Cottus asper is thought to have survived the Last Glacial Maximum (LGM) in two glacial refugia south of the ice sheets: a coastal refugium located in the Chehalis River region (south of Puget Sound, Washington State, USA) and an inland refugium in the Columbia River system (British Columbia, Canada, and Washington State, USA) (McAllister & Lindsey, 1961; Krejsa, 1965). Accordingly, two northward post-glacial colonization routes have been proposed: (1) coastal along-shore dispersal through salt-tolerant planktonic larvae; and (2) inland through proglacial lakes creating ephemeral connections between river systems (Krejsa, 1965).

Phenotypic divergence between highly prickled inland sculpins and weakly prickled coastal sculpins has therefore been interpreted as a consequence of allopatric divergence between coastal and inland glacial refugia, followed by northward dispersal during deglaciation (Krejsa, 1965; McPhail, 2007). However, assumptions about post-glacial dispersal routes and phenotypic divergence of C. asper are based on the traditional view that species persisted solely in glacial refugia south of the ice sheets (Hewitt, 2004). Meanwhile, phylogeographical studies have increasingly revealed the existence of more northern, cryptic glacial refugia (reviewed in Shafer et al., 2010) that may have provided additional sources for allopatric divergence and post-glacial range expansions. Thus, highly prickled inland populations may have evolved in parallel across multiple glacial lineages originating from both southern and northern refugia.

Here, we use multi-locus data (mitochondrial DNA and microsatellites), as well as phenotypic data (prickling categories, morphometrics), to elucidate the phylogeographical context for the disjunct inland distribution of highly prickled phenotypes in the prickly sculpin. We aimed to distinguish between two main hypotheses.

1. Our first hypothesis is that highly prickled inland populations originated from a single southern inland refugium, followed by post-glacial colonization of inland areas through proglacial lakes temporarily connecting neighbouring watersheds. Accordingly, a long glacial phase of allopatric divergence between coastal and inland populations should result in pronounced coastal–inland differentiation in both genetic and morphological data. A decrease of genetic diversity towards northern latitudes would be expected if founder events characterized post-glacial colonization.

2. Our second hypothesis is that highly prickled inland sculpins evolved following colonization from multiple inland and coastal refugia. In this case, highly prickled inland phenotypes should be associated with multiple genetic lineages rather than a southern inland lineage. If inland groups do not share a common glacial history of long-time coastal–inland separation, morphological divergence between coastal and inland groups is not expected to be strong. Under this scenario, decreases in genetic diversity towards northern regions are not necessarily expected, because of the presence of northern glacial refugia and admixture zones.

**MATERIALS AND METHODS**

**Sample collection and DNA sequence amplification**

Prickly sculpins were sampled with baited minnow traps during spring and summer 2009 and 2010 from 29 sites in British Columbia and one site in Alberta (Fig. 1). A total of 1053 sculpins were collected, euthanized with an overdose of clove oil, and preserved in 95% ethanol. Additionally, 14 fin clips from Auke Creek (Juneau, SE Alaska) were generously provided by David Tallmon (University of Alaska Southeast). Extractions of genomic DNA were obtained from...
either abdominal muscle or fin-clip tissue using a standard phenol–chloroform technique. Polymerase chain reactions (PCRs) were used to amplify a total of 1585 base pairs (bp) of mitochondrial DNA (mtDNA) sequence for 141 sculpins from 12 sites: Alaska (Auke Creek, \(n = 14\)), Meziadin Lake (\(n = 11\)), Mosquito Lake (\(n = 11\)), Bella Coola River (\(n = 12\)), McLeod Lake (\(n = 13\)), Peace River in British Columbia (\(n = 11\)), Peace River in Alberta (\(n = 10\)), Falls Creek (\(n = 10\)), Little Campbell River (\(n = 16\)), Harrison Lake (\(n = 10\)), and Okanagan Lake (\(n = 13\)) (Fig. 1, Table 1; for primer sequences and PCR protocols see Appendix S1 in Supporting Information). Sequence proofreading was conducted with 4peaks 1.7.2 (Griekspoor & Groothuis, 2006), and alignments were carried out with ClustalX 2.1 (Larkin et al., 2007). Sequences of mtDNA haplotypes are deposited in GenBank (accession numbers KR024578–KR024637).

Morphology

Sculpins were bleached for 2–10 days (3 parts of 0.5% potassium hydroxide to 1 part glycerol, containing 50 μL of 50% hydrogen peroxide per 10 mL KOH:glycerol solution), and stained for 48 h in a 1% potassium hydroxide and Alizarin red solution. We assessed prickling pattern and body shape variation among coastal and inland populations. Prickling intensity was characterized as ‘high’, ‘medium’ or ‘low’ for each individual, which corresponds to prickling categories 2–4 as described by Nolte et al. (2005b) and reflects the coverage of the lateral body side with prickle-shaped, scale-like bony structures (Fig. 2). A Pearson’s chi-square test was conducted to test whether coastal and inland sculpin populations differ in frequencies of prickling categories.

We analysed body shape using a geometric morphometrics approach to assess morphological divergence among coastal and inland sculpins. We applied 23 homologous landmarks digitized on the left lateral body side of photographed sculpins using tpsDIG 2.14 and included a ruler in the image for scaling (Rohlf, 2009a) (see Appendix S2). Because preservation caused pronounced arching of the fish bodies, we included only samples showing minimal bending in analyses (coastal: \(n = 548\); inland: \(n = 183\)). To remove remaining effects of bending from the landmark configuration, we used the ‘unbend’ function implemented in tpsUtil 1.44 (Rohlf, 2009b). For this purpose, landmarks 8 and 21–23 were chosen to represent a straight line (Appendix S2). A quadratic curve was then fitted along these landmarks and used to correct for the arch in the complete landmark set. After removing landmarks 21–23, we used MorphiJ 1.05d (Klingenberg, 2011) to conduct a generalized least squares Procrustes superimposition to remove the effects of rotation, scale and non-allometric effects of size. The resulting partial warp scores and uniform components were used as shape variables in subsequent analyses. To test for allometric effects within coastal and inland groups, we used a multivariate analysis of covariance (MANCOVA) with (log) centroid size as a covariate, and tested for equal slopes among coastal and inland groups using a MANCOVA as implemented in tpsRegr 1.38 (Rohlf, 2011). We obtained size-corrected shape variables by performing a MANCOVA of Procrustes coordinates on (log) centroid size using a permutation test for each subgroup (10,000 permutations), and used the regression residuals for
a size-corrected analysis (Berner, 2011). We used a jackknifed discriminant function analysis (DFA) on both size-corrected and uncorrected data to determine whether a priori defined coastal and inland populations could be distinguished by body shape. A principal components analysis (PCA) on the variance-covariance matrix of shape coordinates was conducted to visualize patterns of shape variability. Landmark data are deposited under DRYAD (doi: 10.5061/dryad.8ht04).

**Genetic diversity and phylogenetic analysis of sequence data**

We used ARLEQUIN 3.5.1.2 (Excoffier et al., 2005) for calculating population pairwise $F_{ST}$ values for mtDNA sequences (θ; Weir & Cockerham, 1984) and genetic diversity estimates. We performed a spatial analysis of molecular variance (SAMOVA) using the software SAMOVA 1.0 (Dupanloup et al., 2002), to test for geographical groupings that are genetically homogeneous and maximally differentiated from each other, without making a priori assumptions about group assignments. Genetic diversity estimates were calculated for each putative population and geographical mtDNA group, which were derived from the SAMOVA analyses.

In order to characterize genetic lineages and to estimate phylogenetic relationships among them, phylogenetic trees were built from concatenated mtDNA sequences using a maximum-likelihood (ML) approach implemented in RAxML 7.3.2 (Silvestro & Michalak, 2012) with 1000 bootstrap runs, and a Bayesian Inference (BI) approach in MrBayes 3.2 (Ronquist et al., 2012). The best-fit nucleotide substitution model was determined in ModelTest 0.1.1 (Posada, 2008), based on the corrected Akaike information criterion (AICc). A Bayes Factor test implemented in TRACER 1.5 (Rambaut & Drummond, 2009) indicated a better fit of a codon-partitioned model (by gene and 3rd codon) as opposed to an unpartitioned model. Analyses in MrBayes consisted of four Markov chain Monte Carlo (MCMC) chains running for 30 million generations (average standard deviation of split frequencies < 0.01). Trees were sampled...
every 1000th generation with a default burn-in of 25% of samples. A 50% majority-rule consensus tree was built from the remaining trees and used to calculate posterior probabilities of tree branches. Convergence and stationarity of likelihood scores was confirmed in Tracer. Average uncorrected pairwise genetic distances (p-distances) within and between mtDNA tree groups were calculated in mega 5.05 (Tamura et al., 2011).

To better distinguish alternative colonization scenarios of either single origin or multiple coastal–inland colonizations of highly prickled inland populations, we used a constrained tree analysis as implemented in paup* 4.0b1 (Swofford, 1998). After determining the most appropriate substitution model in jModelTest, a ML search under the HKY+I model in paup* was performed for each of three tree topologies: (1) unconstrained tree; (2) constrained tree with monophyly of highly prickled inland populations (Okanagan Lake, Harrison Lake, Nimpo Lake, McLeod Lake); and (3) constrained tree with independent colonization histories reflected by monophyly of Okanagan Lake assuming an independent colonization history via the Columbia River watershed. Resulting tree topologies were compared using a one-tailed Shimodaira–Hasegawa likelihood ratio test with full likelihood parameter estimation using a RELL approximation with 1000 bootstrap replicates (Shimodaira & Hasegawa, 1999).

A haplotype network under statistical parsimony was constructed with tcs 1.13 (Clement et al., 2000) to visualize phylogenetic relationships between haplotypes without enforcing bifurcating tree branching that may be inappropriate for describing intraspecific patterns (Posada & Crandall, 2001). This method collapses sequences into haplotypes and calculates pairwise distances to construct connections among haplotypes where the probability of parsimony exceeds 95% (Templeton et al., 1992).

**Microsatellite amplification and analyses**

We amplified a total of 19 microsatellite loci, including 18 microsatellite loci from Nolte et al. (2005a) and one newly designed marker ‘EDA1’. This locus is linked to the Ectodysplasin (EDA) region in Cottus perifretum and was potentially associated with prickling in C. asper, but showed no deviations from neutrality in a FST outlier test or gametic phase disequilibrium with the prickling phenotype (S.D., unpublished data). We genotyped 273 individuals from the 12 populations used for the sequence analyses (including the same individuals), and 132 individuals from four additional sites (Fig. 1, Table 1): Lakelse Lake (n = 30), Martins Lake on Campbell Island (n = 51), Martins Lake outlet (n = 31), and Tlell River (n = 20). Alleles were separated and genotyped via electrophoresis on a ABI 3500xl sequencer (Applied Biosystems, Carlsbad, CA, USA), with allelic sizes (in base pairs) determined by reference to an internal sizing standard in the software GeneMapper 4.1 (Applied Biosystems). For primer sequences and PCR conditions, see Appendix S1.

We used an extended version of Fisher’s exact test for testing microsatellite loci for deviations from Hardy–Weinberg
equilibrium (HWE, 10^5 MCMC chain steps, 10^4 dememorization steps) and linkage disequilibrium (LDE, 10^5 MCMC chain steps, 10^4 dememorization steps) with ARLEQUIN 3.5.1.2 (Excoffier et al., 2005). Sequential Bonferroni correction was applied to the P-values to correct for multiple comparisons in the same dataset. Pairwise FST values, as well as observed (Hs) and expected (He) heterozygosities were calculated in ARLEQUIN. Genetic diversity and allelic richness (mean number of alleles, corrected for smallest sample number) was calculated in FSTAT 2.9.3.2 (Goudet, 2001). We used a permutation test (15,000 permutations) in FSTAT to compare allelic richness and genetic diversity between coastal and inland populations. The possibility of null alleles, large allele drop-out and scoring errors was evaluated using MICRO-CHECKER 2.2.3 (Van Oosterhout et al., 2004).

An individual-based assignment approach implemented in STRUCTURE 2.3.3 (Pritchard et al., 2000) was used to visualize geographical clustering of microsatellite data and infer genetic population structure. Following initial test runs, we used an admixture model to test for number of populations (K) from 4 to 16. We applied a 100,000-iteration burn-in followed by 900,000 iterations, and repeated each K four times to ensure stability. The most likely number of genetic clusters was determined by using the ΔK method (Evanno et al., 2005) implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2012). Additionally, individual-based chord distance measures (DG; Cavalli-Sforza & Edwards, 1967) were calculated and used for the construction of a 1000 times bootstrapped neighbour-joining (NJ) tree in POPULATIONS 1.2.32 (Langella, 1999). Concordance of genetic variability between microsatellite data and mtDNA groups was assessed with an analysis of molecular variance (AMOVA) using the three mtDNA groups as a grouping variable, and excluding genetically admixed populations [Auke Creek (Alaska) and Bella Coola River, see Results]. For comparison, a second AMOVA was conducted using the clusters inferred from the STRUCTURE analysis as a grouping variable. Genotype data are deposited under DRYAD (doi: 10.5061/dryad.8ht04).

RESULTS

Morphology

Consistent with Krejsa’s (1965) findings, prickling intensity differed significantly between inland and coastal sculpins (χ^2 = 208.2, d.f. = 2, P < 0.001). Inland populations were dominated by highly prickled sculpins, whereas coastal sculpins exhibited high frequencies of medium and low prickling categories (Fig 2c, Appendix S2). Allometric effects of centroid size on body shape variation were significant, but explained only 3.91% (coastal: MANCOVA, 10,000 permutations, P < 0.001) and 3.88% (inland: MANCOVA, 10,000 permutations, P < 0.001) of shape variation, respectively. A discriminant function analysis (DFA) on size-corrected and uncorrected data gave similar results and indicated a small, but significant difference between coastal and inland groups (Procrustes distance: 0.0088, P < 0.001). Classification accuracy in the cross-validated DFA was 81.9% for coastal and 77% for inland sculpins for size-corrected data. Inland sculpins showed a subtle tendency towards having a shorter and higher caudal peduncle, a smaller eye and a slightly stouter snout than coastal sculpins (Appendix S2). These differences among coastal and inland groups were also visible when performing a DFA on highly prickled (category 4) sculpins only (not shown).

Mitochondrial DNA

The geographical distribution of mtDNA haplotypes did not support monophyly of highly prickled inland populations. We detected 20 mitochondrial haplotypes (20 variable positions, with five being parsimony informative) in 141 individuals from 12 populations. The substitution model HKY + I was chosen as the most appropriate model, as determined by the lowest AIC in jMODELTEST.

The SAMOVA supported five distinct geographical groups: (1) Alaska (Auke Creek), (2) Bella Coola River, (3) Okanagan Lake, (4) a central British Columbian group (Meziadin Lake, McLeod Lake, Mosquito Lake, Nimpo Lake, Peace River), and (5) a southern coastal group (Falls Creek, Harrison Lake, Little Campbell River) (FCT = 0.638, P < 0.001). In total, 63.88% of the variance was explained by differences among geographical regions (Table 2). Two of the five SAMOVA groups [Alaska (Auke Creek) and Bella Coola River] exhibited pronounced admixture of haplotypes from different mtDNA haplotype clusters (Fig. 2a,b). Consistently, mtDNA haplotype and nucleotide diversity was higher in Auke Creek and Bella Coola River than in the three main groups (Table 1), supporting their potential role as admixture zones.

Highly prickled inland populations were present in all three major mtDNA groups, which was also evident in the mtDNA haplotype network (Fig. 2a,b). Here, major groups are visible as star-like clusters around a frequent haplotype and denoted as lineages A, B and C. Importantly, lineages ‘A’ (central BC) and ‘B’ (southern BC) include highly prickled inland populations as well as coastal populations that are weakly to moderately prickled (Fig. 2c).

The same genetic groups could be resolved in a phylogenetic tree, and exhibited identical topologies and similar posterior probability support for both maximum likelihood (ML) and Bayesian inference (BI) constructions (see Appendix S3). Uncorrected average pairwise distances within and between major tree clusters were low, ranging from 0.09% to 0.21% between clusters, and from 0.02% to 0.03% within each cluster. The constrained tree analysis supported an unconstrained tree (negative log-likelihood, −ln L = 2363.49) over monophyly of highly prickled inland populations (−ln L = 2405.18, Shimodaira–Hasegawa (SH) test, P = 0.023). In contrast, a constrained tree topology mirroring multiple independent coastal–inland colonization in northern (central BC) and southern populations was not
significantly different from an unconstrained ML tree (−ln L = 2363.6, SH-test, P = 0.66).

**Microsatellite loci**

In 405 individuals, the degree of polymorphism in the 19 microsatellite loci was highly variable, with the total number of alleles ranging from 2 to 31. Allelic richness (standardized to a minimum sample size of seven) of polymorphic loci ranged from 1.8 to 6.3, and gene diversity ranged from 0.15 to 0.77 (Table 1). Significant departures from HWE were found for locus Cott153 in Okanagan Lake and Martins Lake, and for locus CottE30 in Martins Lake outlet. Evidence for linkage disequilibrium was found in 38 (out of 2736) comparisons, and potential null alleles occurred in nine locus-pair combinations, and potential null alleles occurred in nine locus-pair combinations. We did not exclude any loci from the analyses because no consistent deviations across the majority of sites were found. Coastal populations showed overall higher genetic diversity estimates than inland populations, with an average allelic richness of 5.04 and 3.23 (F-stat permutation test, two-sided, P = 0.004), respectively, and a gene diversity of 0.665 and 0.422 (P = 0.001), respectively. Significant pairwise FST values ranged from 0.038 (Falls Creek–Martins Lake outlet) to 0.570 (McLeod Lake–Okanagan Lake).

Genetic population structure was more pronounced under microsatellites compared to mitochondrial DNA, yet largely concordant with the inferred mtDNA groups. Clustering results in STRUCTURE and a highest ΔK value (Evanno et al., 2005) supported 11 clusters by grouping Peace River, McLeod Lake and Nimpo Lake in one cluster, Martins Lake and outlet in another cluster, and forming a third cluster including Little Campbell River and Falls Creek (Fig. 3a,b). These clusters were consistent with a NJ-tree of ΔC distances from microsatellites (Fig. 3c). The AMOVA of microsatellite genetic variation revealed that most of the variation existed within populations (61.3%), while grouping according to the three mtDNA lineages explained 20.25% of the variation, overall indicating a high degree of concordance between data sets (Table 3). Similarly, the AMOVA using STRUCTURE K clusters as grouping variable also revealed elevated within-population variability (71%), with genetic clusters explaining 26.25% of the variation (Table 3).

**DISCUSSION**

In this study, we presented phenotypic and multi-locus genetic data to evaluate the phylogeographical history of the prickly sculpin (*Cottus asper*) and the evolutionary context of a geographical discontinuity in prickling intensity described by Krejsa (1965). The pronounced coastal–inland split of prickling phenotypes was accompanied by weak, yet significant differentiation in body shape, which supports divergent evolutionary trajectories in alternative environments. Our finding that highly prickled inland phenotypes are genetically distinct from coastal populations but associated with multiple genetic lineages implies parallel evolution subsequent to wide coastal dispersal and repeated coastal to inland colonizations, rather than a single origin from a southern inland refugium as has been assumed since Krejsa (1965).

**Post-glacial colonization**

The widespread occurrence of northern and southern mitochondrial haplotypes along the coast underlines the importance of coastal colonization routes. Inland routes through ephemeral connections between river systems may explain the occurrence of shared coastal and inland haplotypes in

<table>
<thead>
<tr>
<th>Grouping schemes</th>
<th>d.f.</th>
<th>Sum of squares (among groups)</th>
<th>Variance components (among groups)</th>
<th>Percentage of variation (among groups)</th>
<th>FCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 groups</td>
<td>1</td>
<td>46.23</td>
<td>0.81</td>
<td>59.04</td>
<td>0.59*</td>
</tr>
<tr>
<td>3 groups</td>
<td>2</td>
<td>61.18</td>
<td>0.73</td>
<td>62</td>
<td>0.62*</td>
</tr>
<tr>
<td>4 groups</td>
<td>3</td>
<td>66.17</td>
<td>0.68</td>
<td>62.92</td>
<td>0.634*</td>
</tr>
<tr>
<td>5 groups</td>
<td>4</td>
<td>67.83</td>
<td>0.69</td>
<td>63.88</td>
<td>0.638*</td>
</tr>
<tr>
<td>6 groups</td>
<td>5</td>
<td>68.47</td>
<td>0.67</td>
<td>63.46</td>
<td>0.634*</td>
</tr>
<tr>
<td>7 groups</td>
<td>6</td>
<td>69.6</td>
<td>0.63</td>
<td>63.01</td>
<td>0.63*</td>
</tr>
<tr>
<td>8 groups</td>
<td>7</td>
<td>69.76</td>
<td>0.62</td>
<td>62.62</td>
<td>0.626*</td>
</tr>
<tr>
<td>9 groups</td>
<td>8</td>
<td>69.5</td>
<td>0.59</td>
<td>61.36</td>
<td>0.614*</td>
</tr>
</tbody>
</table>

*SAMOVA grouping schemes were as follows: 2 groups: (Falls Creek, Harrison Lake, Little Campbell River), (Alaska, Bella Coola, Meziadin, McLeod, Mosquito, Nimpo, Okanagan, Peace River); 3 groups: (Falls Creek, Harrison Lake, Little Campbell River), (Bella Coola, Okanagan), (Alaska, Meziadin, McLeod, Mosquito, Nimpo, Peace); 4 groups: (Alaska, Bella Coola, Okanagan), (Falls Creek, Harrison Lake, Little Campbell River), (Meziadin, McLeod, Mosquito, Nimpo, Peace); 5 groups: (Alaska), (Bella Coola), (Okanagan), (Falls Creek, Harrison Lake, Little Campbell River), (Meziadin, McLeod, Mosquito, Nimpo, Peace); 6 groups: (Alaska), (Bella Coola), (Okanagan), (Harrison), (Falls Creek, Little Campbell River), (Meziadin, McLeod, Mosquito, Nimpo, Peace); 7 groups: (Alaska), (Bella Coola), (Okanagan), (Harrison), (Falls Creek, Little Campbell River), (Meziadin), (McLeod, Mosquito, Nimpo, Peace); 8 groups: (Alaska), (Bella Coola), (Okanagan), (Harrison), (Falls Creek), (Little Campbell River), (Meziadin), (McLeod, Mosquito, Nimpo, Peace); 9 groups: (Alaska), (Bella Coola), (Okanagan), (Harrison), (Falls Creek), (Little Campbell River), (Meziadin), (Mosquito), (McLeod, Nimpo, Peace).
now disconnected inland watersheds such as Peace River. However, the lack of shared haplotypes among southern and northern inland regions does not support far northward dispersal of southern inland haplotypes. Gene flow among coastal populations of *C. asper* is facilitated by a high dispersal capability of salt-tolerant, planktonic larval stages (Krejza, 1965; McPhail, 2007), which is well known from many amphidromous species including sculpins (Whiteley et al., 2009; Dennenmoser et al., 2014). Accordingly, secondary contact between coastal northern and southern genetic lineages may have contributed to the occurrence of elevated genetic diversities and the admixture of haplotypes in Auke Creek (Southeast Alaska) and Bella Coola River (central BC). Genetic diversity did not decrease towards northern latitudes, as would be expected if founder events accompanied a long period of northward colonization. Instead, genetic diversities were lower in inland populations, which is presumably a consequence of founder effects associated with coastal to inland colonizations, as well as larger effective population sizes and more gene flow among coastal populations (Teacher et al., 2011).

Inland dispersal was facilitated during deglaciation, when drainage of large proglacial lakes allowed crossing watersheds through temporal connections between river systems such as Columbia/ Fraser, Fraser/Peace, Nass/Skeena or Skeena/Peace rivers (McPhail & Carveth, 1993). Similar to other species (e.g. pygmy whitefish, *Prosopium coulterii*; Witt et al., 2011), the prickly sculpin may have accessed the Peace River watershed via the Fraser and Skeena rivers (McPhail, 2007). Far inland dispersal and an exclusively southern inland origin of all highly prickled inland sculpins via multiple inland watershed crossings is unlikely, given the high genetic diversity in southern regions.

**Figure 3** Genetic structuring of the prickly sculpin (*Cottus asper*) in north-western North America. (a) Log-likelihood plot supporting $K = 11$. (b) STRUCTURE plot showing assignment of 405 individuals of *Cottus asper* from 16 sampling sites into 11 clusters. (c) Neighbour-joining tree based on $D_{st}$ distances among populations based on 19 microsatellite loci (numbers indicating 1000× bootstrap support).
Table 3 Analyses of molecular variance (AMOVAs) for microsatellite data of the prickly sculpin (Cottus asper) sampled in northwestern North America (*P < 0.001, 10,100 permutations).

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Sum of squares</th>
<th>Variance component</th>
<th>Percentage of total variation</th>
<th>Fixation indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 groups (mtDNA lineages)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among groups</td>
<td>2</td>
<td>326.2</td>
<td>0.76</td>
<td>20.25</td>
<td>F&lt;sub&gt;CT&lt;/sub&gt; = 0.2*</td>
</tr>
<tr>
<td>Among populations, within groups</td>
<td>10</td>
<td>387.4</td>
<td>0.74</td>
<td>19.71</td>
<td>F&lt;sub&gt;SC&lt;/sub&gt; = 0.25*</td>
</tr>
<tr>
<td>Among individuals, within populations</td>
<td>302</td>
<td>667.97</td>
<td>0</td>
<td>0</td>
<td>F&lt;sub&gt;IS&lt;/sub&gt; = 0</td>
</tr>
<tr>
<td>Within individuals</td>
<td>315</td>
<td>726.5</td>
<td>2.3</td>
<td>61.3</td>
<td>F&lt;sub&gt;RT&lt;/sub&gt; = 0.39*</td>
</tr>
<tr>
<td>11 groups (structure clusters)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among groups</td>
<td>10</td>
<td>887.94</td>
<td>1.0</td>
<td>26.25</td>
<td>F&lt;sub&gt;CT&lt;/sub&gt; = 0.26*</td>
</tr>
<tr>
<td>Among populations, within groups</td>
<td>7</td>
<td>65.5</td>
<td>0.13</td>
<td>3.45</td>
<td>F&lt;sub&gt;SC&lt;/sub&gt; = 0.05*</td>
</tr>
<tr>
<td>Among individuals, within populations</td>
<td>442</td>
<td>1177.5</td>
<td>0</td>
<td>0</td>
<td>F&lt;sub&gt;IS&lt;/sub&gt; = 0</td>
</tr>
<tr>
<td>Within individuals</td>
<td>460</td>
<td>1250</td>
<td>2.72</td>
<td>71</td>
<td>F&lt;sub&gt;RT&lt;/sub&gt; = 0.29*</td>
</tr>
</tbody>
</table>

differentiation among northern and southern inland areas. Overall, our data highlight the importance of large coastal rivers such as the Skeena and Fraser rivers for the colonization of inland freshwater habitats, such as the highly prickled populations present in Lakelse Lake or Harrison Lake. This suggests that post-glacial colonization of inland populations has most likely advanced by separate coastal–inland colonizations across different genetic lineages, resulting in parallel evolution of highly prickled inland populations.

This interpretation contradicts the previous assumption of monophyly of highly prickled inland sculpins, associated with only southern glacial refugia and post-glacial dispersal from southern towards northern areas. Instead, our results show a pattern of mitochondrial haplotypes forming three genetic clusters, albeit weakly differentiated but consistent with divergence among three glacial lineages. These patterns could suggest the presence of a northern coastal refugium, as has been suggested for other species in the Queen Charlotte Islands and the Alexander Archipelago area, such as threespine stickleback (Gasterosteus aculeatus), rainbow trout (Oncorhynchus mykiss) and salmon (Oncorhynchus nerka, Oncorhynchus kisutch) (O’Reilly et al., 1993; Wood et al., 1994; McCusker et al., 2000; Smith et al., 2001). While our limited sampling scheme and possibly confounding effects of post-glacial admixture does not allow pinpointing the exact location of a northern coastal refugium, the microsatellite genetic structure among northern coastal populations could indicate the presence of a mosaic of separate, partly disconnected coastal refugia (‘refugia within refugia’; Shafer et al., 2010). Regarding southern refugia, a strong genetic break between southern coastal (Fraser Valley) and inland (Okanagan Lake) populations supports the previous assumption of separate coastal and inland refugia in this region (Krejsa, 1965; McPhail, 2007). A similar coastal–inland break has frequently been attributed to refugia areas located in the Chehalis River and lower Columbia River valleys (e.g. McPhail & Carveth, 1993; Taylor et al., 1999), including an ‘Okanagan group’ in the Okanagan Valley that connects to a glacial refugium in the Columbia River (Larson et al., 2012).

Comparison of mitochondrial and nuclear genetic data

Analyses of microsatellite loci revealed pronounced population structuring, which was overall consistent with the characterization of the three mtDNA lineages and two mtDNA admixture zones. Genetic divergence among mtDNA lineages has not been eroded by high migration rates during more recent times, as has been suggested for other species (e.g. Atlantic herring, Clupea harengus; Gaggiotti et al., 2009). Moreover, the structure in the microsatellite data indicates that male-biased dispersal or female-biased philopatry are not likely major drivers of mtDNA structuring. The finding of additional, more fine-scaled genetic structure revealed by microsatellites is expected in the absence of high migration rates and promoted by the relatively higher microsatellite mutation rates (Koskinen et al., 2002). While elevated mutation rates and homoplasy render microsatellite markers less informative for more ancient divergences, and have frequently failed to find structure in the presence of mitochondrial genetic structuring (Zink, 2010), they still may be useful to support phylogeographical patterns that reflect more recent (late Pleistocene) divergences in some taxa (Koskinen et al., 2002; Hänfling et al., 2002).

An unexpected discordance between nuclear and mitochondrial DNA was found in a pronounced genetic differentiation between northern coastal and inland populations in the microsatellite loci but not the mtDNA, which is dominated by the most common northern haplotype (lineage ‘A’). This could suggest mitochondrial introgression following secondary contact between glacial lineages, as has been reported in other studies (Nolte et al., 2005b; Redenbach & Taylor, 2003; Toews & Brelsford, 2012). A possible double-colonization scenario allowing mitochondrial introgression
upon secondary contact could be post-glacial dispersal of the southern-coastal lineage ‘B’ into the Peace River watershed via the Fraser River, followed by the introgressive invasion of the northern mtDNA lineage ‘A’ via the Skeena River. We recognize the speculative nature of such a double-colonization scenario, and propose that future studies should test for introgression of a northern mitochondrial haplotype that might have replaced a southern haplotype (Glémet et al., 1998). In general, selection on the mitochondrial genome may be common (Dowling et al., 2008), and selective sweeps could play an important yet frequently overlooked role in shaping phylogeographical patterns (Rato et al., 2011).

**Parallel evolution**

Our results suggest that multiple genetic lineages have contributed to the occurrence of highly prickled inland populations following colonizations via coastal rivers, which indicates parallel phenotypic evolution (Elmer & Meyer, 2011). The high variability in prickling intensity in the presumed ancestral coastal populations may indicate a role for standing genetic variation rather than new mutations as a source of parallel evolution. This distinction is important, because consequences for genomic architecture and speed of adaptive divergence can differ dramatically (Elmer & Meyer, 2011; Rogers et al., 2013). For example, the availability of standing genetic variation allows faster fixation of alleles compared to low-frequency new mutations, which may increase the probability of finding parallel phenotypes in recently deglaciated, post-glacial environments such as those found in north-western North America (e.g. threespine stickleback; Colosimo et al., 2004). If high gene flow promotes the wide dispersal of pre-existing adaptive alleles, parallel evolution can rapidly occur in derived populations across a large geographical area as has been shown for the parallel evolution of freshwater ecotypes in threespine stickleback (Schluter & Conte, 2009). Similarly, our study suggests parallel evolution of highly prickled inland groups following colonization through coastal rivers, which is based on standing genetic variation and wide dispersal among the ancestral coastal populations. The parallel fixation of a highly prickled phenotype across multiple genetic lineages suggests a heritable genetic basis of prickling. This seems plausible given the finding of a major prickling QTL in Cottus perifretum (Cheng, 2013). While we currently do not know whether prickling is adaptive, the finding of weak differentiation in body shape among coastal and inland groups besides prickling intensity also suggests that inland freshwater habitats represent a different selective environment. Further studies are needed to test for possible functions of prickling, and to reveal whether parallel evolution of prickling phenotypes in C. asper is based on a single mutation in the EDA pathway as suggested by Cheng (2013) for European Cottus, or whether historical differentiation has diversified genetic variation underlying highly prickled phenotypes.

**CONCLUSIONS**

In summary, we provide an example of a parallel phenotype that has evolved across multiple genetic lineages, and repeatedly became the predominant phenotype after multiple inland colonizations via coastal rivers. Future studies should test whether mitochondrial introgression between two lineages may have facilitated colonization of northern inland regions, and whether standing genetic variation of the highly admixed ancestral coastal populations could have resulted in lineage-specific genetic polymorphisms underlying highly prickled phenotypes. While the adaptive function of parallel phenotypes often remains to be experimentally tested, the possibility of alternative parallel evolutionary pathways (genetic constraints, ancient selective regimes) deserves more attention.

**ACKNOWLEDGEMENTS**

We are grateful to David Tallmon (University of Alaska Southeast) for providing fin-clip samples and to Susan Cousineau, Andrew Rezansoff, Stevi Vanderzwan and Lisa Duke for field assistance. Comments from Daniel Ruzzante and two anonymous referees improved the manuscript. We thank the Alberta Innovates Technology Futures for a graduate scholarship to S.D. This research was supported by NSERC Discovery Grants (S.M.R. and S.M.V.), a European Research Council starting grant (A.W.N.) and an Alberta Innovates Technology Futures New Faculty Award (S.M.R.).

**REFERENCES**


Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. Proceedings of the Royal Society B: Biological Sciences, 272, 2379–2387.


**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article:

- **Appendix S1** Primer sequences and PCR protocols.
- **Appendix S2** Prickling and body shape differences between inland and coastal sculpins.
- **Appendix S3** Fifty per cent majority-rule consensus tree of mtDNA sequences.

**DATA ACCESSIBILITY**

Landmark and genotype data are deposited under DRYAD (doi: 10.5061/dryad.8ht04), and mitochondrial haplotypes are deposited under GenBank (accession numbers KR024578-KR024637).

**BIOSKETCH**

**Stefan Dennenmoser** is broadly interested in the intersection of biogeography, natural history and local adaptation and its implications for the genetic basis of adaptive change. This study was part of his PhD research on the phylogeography of *Cottus asper* in north-western North America.

Author contributions: S.D., S.M.V., A.W.N. and S.M.R. designed the study. S.D. performed the research and analysed the data, and all authors contributed to writing the paper.

Editors: John Stewart and Robert Whittaker