JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12767

Does plasticity enhance or dampen phenotypic parallelism? A test with three lake–stream stickleback pairs

K. B. OKE*, M. BUKHARI*, R. KAEUFFER*, G. ROLSHAUSEN*, K. RÄSÄNEN†, D. I. BOLNICK‡, C. L. PEICHEL§ & A. P. HENDRY*

*Redpath Museum and Department of Biology, McGill University, Montreal, QC, Canada
†Department of Aquatic Ecology, Eawag and Institute of Integrative Biology, ETH-Zurich, Duebendorf, Switzerland
‡Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA
§Divisions of Basic Sciences and Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Keywords:

convergent evolution;
parallel evolution;
phenotypic plasticity;
threespine stickleback.

Abstract

Parallel (and convergent) phenotypic variation is most often studied in the wild, where it is difficult to disentangle genetic vs. environmentally induced effects. As a result, the potential contributions of phenotypic plasticity to parallelism (and nonparallelism) are rarely evaluated in a formal sense. Phenotypic parallelism could be *enhanced* by plasticity that causes stronger parallelism across populations in the wild than would be expected from genetic differences alone. Phenotypic parallelism could be *dampened* if site-specific plasticity induced differences between otherwise genetically parallel populations. We used a common-garden study of three independent lake–stream stickleback population pairs to evaluate the extent to which adaptive divergence has a genetic or plastic basis, and to investigate the enhancing vs. dampening effects of plasticity on phenotypic parallelism. We found that lake–stream differences in most traits had a genetic basis, but that several traits also showed contributions from plasticity. Moreover, plasticity was much more prevalent in one watershed than in the other two. In most cases, plasticity enhanced phenotypic parallelism, whereas in a few cases, plasticity had a dampening effect. Genetic and plastic contributions to divergence seem to play a complimentary, likely adaptive, role in phenotypic parallelism of lake–stream stickleback. These findings highlight the value of formally comparing wild-caught and laboratory-reared individuals in the study of phenotypic parallelism.

Introduction

Parallel/convergent phenotypic evolution – the independent and repeated evolution of similar traits in similar environments – provides strong evidence for a deterministic role of natural selection (Langerhans & DeWitt, 2004; Schluter *et al.*, 2004; Arendt & Reznick, 2008; Losos, 2011; Wake *et al.*, 2011). While recognizing that optimal use of the terms 'parallel' vs. 'convergent' is debatable (Arendt & Reznick, 2008), we henceforth use 'parallel' as it is standard for our study system (see below) and because we focus on phenotypic patterns, rather than the underlying genetic/developmental pathways. Parallelism is frequently reported in studies of natural populations, yet a growing number of studies have shown that ostensibly similar habitat divergence is often associated with substantial nonparallel phenotypic divergence. For example, adaptive divergence between contrasting environments can vary dramatically in direction and magnitude among independent evolutionary lineages (e.g. Brinsmead & Fox, 2002; Matos *et al.*, 2002; Langerhans & DeWitt, 2004; Langerhans *et al.*, 2006; Eroukhmanoff *et al.*, 2009; Prunier *et al.*, 2012). These deviations from parallelism are often attributed to site-specific environmental differences within a given 'habitat type' (Brinsmead & Fox, 2002; Landry &

Correspondence: Krista B. Oke, Redpath Museum and Department of Biology, McGill University, 859 Sherbrooke St. W. Montreal, QC, Canada H3A 2K6.
e-mail: krista.oke@mail.mcgill.ca

Bernatchez, 2010; Ravinet *et al.*, 2013), although other factors, including plasticity, also could be important (Teotónio & Rose, 2000; Langerhans & DeWitt, 2004; Nosil & Crespi, 2004; Langerhans *et al.*, 2006; Kaeuffer *et al.*, 2012; Lucek *et al.*, 2014a). Our goal in this study was to explicitly consider the role of contemporary (i.e. current) plasticity in phenotypic parallelism and non-parallelism, which we sometimes refer to jointly as '(non)parallelism'. To determine the relative contributions of plastic and genetic effects to (non)parallelism, we consider the extent to which parallel phenotypic divergence across pairs is retained in a common-garden setting, as compared to patterns in the wild.

Plasticity could interact in several ways with genetic differences to shape phenotypic divergence in nature (Price *et al.*, 2003; Ghalambor *et al.*, 2007, 2015; Paenke *et al.*, 2007; Crispo, 2008; Pfennig *et al.*, 2010; Fitzpatrick, 2012; Wund, 2012). First, plasticity might *hinder* genetic divergence because plastic responses can shield genetic variation from selection (Price *et al.*, 2003; Thibert-Plante & Hendry, 2011). Second, plasticity might *promote* genetic divergence by (1) revealing cryptic variation on which selection can act (West-Eberhard, 2003; Moczek *et al.*, 2011; Schlichting & Wund, 2014), (2) allowing populations to persist in novel habitats until selection can act on genetic variation (Price *et al.*, 2003; Crispo, 2007; Ghalambor *et al.*, 2007; Pfennig & McGee, 2010; Pfennig *et al.*, 2010) or (3) by displacing phenotypes away from adaptive optima (i.e. maladaptive plasticity) and thereby intensifying selection (Grether, 2005; Ghalambor *et al.*, 2015). Third, plasticity can influence the effects of gene flow on adaptive divergence, depending on whether plasticity acts before or after dispersal (Thibert-Plante & Hendry, 2011). Fourth, plasticity might cause non-genetic population divergence if different environments induce plastic differences between genetically similar populations. As a result of these effects, plasticity might increase or decrease phenotypic divergence, a contrast sometimes called cogradient vs. counter-gradient variation (Conover & Schultz, 1995; Conover *et al.*, 2009).

By extension, contemporary plasticity could alter patterns of phenotypic (non)parallelism. On the one hand, plasticity might *enhance* parallelism by making habitat-associated divergence more similar across independent lineages than would be the case in the absence of plasticity. On the other hand, plasticity might *dampen* parallelism by making such divergence less similar than would be expected in its absence. One route to such inferences is to compare phenotypic patterns between in field-collected and laboratory-reared individuals, the latter presumably lacking (or at least minimizing) differential environmental effects on phenotypes. (Of course, genotype-by-environment (GxE) interactions dictate that phenotypic patterns in the laboratory are not, strictly speaking, the patterns that would be seen in nature in the absence of differential plasticity.) Many previous studies have conducted common-garden experiments to infer a genetic basis for phenotypic divergence (e.g. Lavin & McPhail, 1993; O'Steen *et al.*, 2002; Langerhans *et al.*, 2004; Nosil & Crespi, 2004; Herczeg *et al.*, 2009). Although some of these studies involved multiple independent lineages (at the population or species levels) and were thus informative with respect to (non)parallelism, few studies have formally compared (non)parallelism in nature to (non)parallelism for the same populations reared in the laboratory (Torres-Dowdall *et al.*, 2012 provide a rare example). As a result, insights into whether plasticity enhances or dampens phenotypic parallelism are generally lacking. In the present study, we explore this topic through a study of threespine stickleback, *Gasterosteus aculeatus*.

Study system and specific questions

Threespine stickleback are widely distributed across the Northern Hemisphere where they have repeatedly invaded and adapted to diverse habitats, resulting in the replicate parallel evolution of distinct 'ecotypes' (Bell & Foster, 1994; McKinnon & Rundle, 2002; Hendry *et al.*, 2009). In our study system (lake–stream ecotype pairs in British Columbia, Canada), this evolutionary replication arose as different watersheds were independently colonized by marine ancestors following Pleistocene deglaciation (Thompson *et al.*, 1997; Hendry & Taylor, 2004; Berner *et al.*, 2009; Roesti *et al.*, 2012). Lake–stream pairs show considerable parallel trait divergence driven by repeatable natural selection (Moodie, 1972a,b; Reimchen *et al.*, 1985; Lavin & McPhail, 1993; Hendry & Taylor, 2004; Berner *et al.*, 2008, 2009; Kaeuffer *et al.*, 2012; Lucek *et al.*, 2013; Ravinet *et al.*, 2013). For example, lake fish commonly forage on zooplankton in the open water, which favours shallower bodies and numerous gill rakers, whereas stream fish commonly forage on macroinvertebrates in complex benthic habitats. Yet these same traits also exhibit some nonparallelism in that lake–stream divergence is much greater in some watershed than in others. Moreover, other traits such as armour and gill raker length also exhibit nonparallelism in form of different *directions* of divergence in different lake–stream pairs (Moodie, 1972a,b; Reimchen *et al.*, 1985; Lavin & McPhail, 1993; Hendry *et al.*, 2002; Kaeuffer *et al.*, 2012). These deviations from parallel evolution at least partly reflect variation in divergent selection (Berner *et al.*, 2008; Kaeuffer *et al.*, 2012; Hendry *et al.*, 2013a) and gene flow (Hendry & Taylor, 2004; Moore *et al.*, 2007). They could also reflect historical contingency resulting from the particular marine genotypes that happened to colonize a given watershed. Here, we will consider to what extent the above (non)parallelism also might reflect phenotypic plasticity.

Common-garden studies thus far have been performed for only two independent lake–stream lineages: the Misty Lake-Inlet/Outlet pairs on northern Vancouver Island, British Columbia (Lavin & McPhail, 1993; Hendry *et al.*, 2002, 2011; Delcourt *et al.*, 2008; Sharpe *et al.*, 2008; Raeymaekers *et al.*, 2009, 2010; Berner *et al.*, 2011; Baker *et al.*, 2013) and several lake–stream pairs from Lake Constance, Switzerland (Lucek *et al.*, 2014a; Moser *et al.*, 2015). In the Misty watershed, lake and inlet stickleback differ genetically in a wide range of traits, including gill raker length, body shape, swimming ability, nest construction and courtship behaviour (see above references). At the same time, at least some of these traits are known to be influenced by plasticity, especially with respect to differences between Misty Lake and Outlet stickleback (see above references). However, with only a single watershed studied from each of only two continents, generality is currently lacking as to plastic vs. genetic contributions to lake–stream divergence, as well as any (non)parallelism therein.

In this study, we quantify and compare phenotypic divergence for wild-caught and common-garden fish from three Vancouver Island lake–stream pairs. First, considering each of the three watersheds separately, we ask, (1) what lake–stream differences persist in a common garden? and (2) does plasticity increase or decrease lake–stream differences? If lake–stream differences are greater (vs. lesser) in the wild than in the common garden, we conclude that plasticity increases trait differences. If trait differences are greater in the common garden, we conclude that plasticity decreases lake–stream differences. Second, considering all three watersheds together, we ask, (3) to what extent are lake–stream differences parallel in common-garden conditions? We here quantify (non)parallelism of phenotypic *evolutionary divergence* – heritable differences between the ecotypes seen in the common garden – relative to phenotypic *divergence* in the wild that could result from both genetic and plastic effects. Finally, we ask the key question, (4) to what extent does plasticity enhance or dampen parallelism? If lake–stream parallelism across the three watersheds is greater in the wild than in a common garden, we conclude that plasticity enhances phenotypic parallelism. If parallelism is greater in the common garden, we conclude that plasticity dampens phenotypic parallelism.

Materials and methods

We focused on three independent lake–stream pairs from separate watersheds on Vancouver Island that currently contain phenotypically divergent populations of lake and stream stickleback: the Misty, Boot and Roberts pairs (Hendry & Taylor, 2004; Berner *et al.*, 2008; Kaeuffer *et al.*, 2012; Hendry *et al.*, 2013a). Collections for the analysis of morphology in wild-caught (WC) fish were conducted in 2008 and correspond to those described in Kaeuffer *et al.* (2011, 2012). Field collections for common-garden (CG) studies took place in 2011 and 2012 for Boot Lake and outlet stream and Roberts Lake and outlet stream and in 2004 for the Misty Lake and inlet stream (see Fig. S1 for map of sampling sites, Table S1 for coordinates of sampling locations and Table S2 for sample sizes). Thus, the specific stream populations were a mixture of inlets and outlets. Although outlets sometimes show low divergence from upstream lakes, our specific populations all exhibit relatively high phenotypic, ecological and genetic divergence (Table S3; Kaeuffer *et al.*, 2012 estimated F_{ST} of 0.121, 0.178 and 0.045 for Misty Inlet–Lake, Boot Outlet–Lake and Roberts Outlet–Lake, respectively) and low gene flow from their respective lake populations (Hendry & Taylor, 2004; Berner *et al.*, 2009; Kaeuffer *et al.*, 2012; Hendry *et al.*, 2013a). Importantly, all three streams differ from their respective lake populations in environmental characteristics driving lake–stream divergence: the streams are more complex habitats and have fewer limnetic and more benthic prey resources than their respective lakes. The CG Misty collections and data correspond to those described in Sharpe *et al.* (2008), but the fish were redigitized and remeasured so as to match procedures for Boot and Roberts.

For all collections, unbaited minnow traps were used to collect stickleback. For the WC collections, fish were haphazardly selected for preservation in 95% ethanol. For the CG collections, fish collected from the wild were checked for maturity based on gravidity in females and eye colour in males. Gravid females were then crossed by standard methods (Hatfield & Schluter, 1996) with mature males from the same population (4–12 full-sibling families per population). The CG eggs were shipped by air mail to McGill University, where they were raised in glass aquaria according to established methods (Delcourt *et al.*, 2008; Sharpe *et al.*, 2008), separated by family. At the start of exogenous feeding, the fish were fed live brine shrimp nauplii (*Artemia sp*) once daily to satiation. Once the fry had grown large enough (about 20 mm), they also received either frozen blood worms (Chironomid larvae, used in 2004 and 2011), live blackworms (*Lumbriculus* sp., used in 2004) or frozen mysis shrimp (Hikari Sales USA, Inc., Hayward, CA, USA, used in 2012), based on the availability of these food products. Importantly, the different populations did not differ systematically in the rearing and feeding conditions they experienced. Moreover, our previous work has shown that lake–stream differences in a common garden are extremely consistent between experiments in different years and conducted under different conditions – as evidenced by similar results for the Misty system in Lavin & McPhail (1993), Hendry *et al.* (2002), Sharpe *et al.* (2008) and Berner *et al.* (2011).

After 8–9 months of rearing under 'summer' conditions (17 °C, 16 h of light and 8 h of dark), the F1 CG fish were transferred to 'winter' conditions (12 °C, 8 h of light and 16 h of dark), before being transferred back to summer conditions. When mature (either reproductively ready or showing early signs of readiness, such as blue eyes and red pelvic spines in males and swollen abdomens in females), these fish were euthanized with MS-222 (tricaine methanesulphonate), photographed for geometric morphometric and univariate measurements and preserved in 95% ethanol for later measurement of gill rakers. The photographs were taken of the left side of the fish, each including a standard scale, with a digital camera mounted on a tripod. As in Sharpe *et al.* (2008), Berner *et al.* (2009) and Hendry *et al.* (2011, 2013b), fine pins were used to indicate landmarks that are difficult to discern from photographs, as well as to spread the dorsal and anal fins.

Our common-garden experiment was conducted through the first laboratory generation (F1), and maternal effects could thus theoretically persist. However, maternal effects on morphological traits seem to be minimal in threespine stickleback given that F1 and F2 common-garden generations have very similar phenotypes in the Misty lake–stream pair (Berner *et al.*, 2011) and in a benthic-limnetic stickleback pair (Hatfield, 1997). We therefore follow previous work on stickleback in interpreting lake–stream differences in F1 common-garden fish to reflect genetic divergence with minimal (if any) maternal effects.

Morphological measurements

All photographs of WC and CG fish were analysed using tpsDIG2 (life.bio.sunysb.edu/morph/) in haphazard order and blind with respect to sampling site. Fourteen reliable landmarks (Fig. 1) were chosen based on previous studies of lake–stream stickleback (Berner *et al.*, 2008, 2009; Kaeuffer *et al.*, 2012). Consensus shape, centroid size and relative warps were calculated using tpsRelw (life.bio.sunysb.edu/morph/).

Preliminary analyses suggested that our samples showed sexual dimorphism in some traits, and so we analysed males and females separately. Geometric morphometric analyses in tpsRelw were also performed on each sex separately. Inspection of relative warps (RWs) for females revealed that the majority of shape variation was due to reproductive status. For example, the first relative warp (RW1) explained 21.63% of shape variation in females and corresponded to whether the fish was gravid or had recently laid eggs. A linear model performed on RWs that included the reproductive status (gravid, recently spawned or neither) and rearing environment (CG or WC) as fixed factors revealed that all warps except RW4 were influenced by reproductive status. Thus, females were excluded from geometric morphometric analyses.

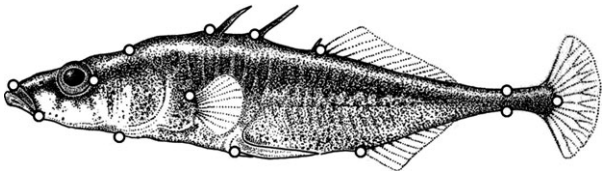

Fig. 1 Position of the landmarks included in geometric morphometric analyses to compare lake–stream shape differences across watersheds and rearing environments. Figure modified from Kaeuffer *et al.* (2012).

Additional morphometric traits not reliably captured by geometric morphometrics were measured from the photographs using IMAGEJ (imagej.nih.gov/ij/). All measurements were again carried out in haphazard order and blind with respect to sampling site. Eleven traits were measured: standard length, jaw length (from the anterior tip of the upper lip to the end of the mouth), eye width (horizontally from the most anterior to most posterior edge of the eye), pectoral fin width (at the base of the fin) and length (along the dorsal edge of the fin), pelvic and dorsal spine lengths (from anterior insertion to tip, both first and second dorsal spines were measured), anal and dorsal fin lengths (at the base of the fin), and caudal peduncle depth (at its narrowest point). A linear model on females similar to that conducted above revealed that several traits were influenced by reproductive status: the lengths of the anal and dorsal fins, the length and width of the pectoral fin and the depth of the caudal peduncle. These traits were excluded from further analyses on females.

Gill raker measurements for WC fish were taken from Kaeuffer *et al.* (2011, 2012). Gill raker measurements for CG fish were performed using the same methods, again in haphazard order and blind with respect to sampling site, on preserved fish examined under a stereomicroscope at 45× magnification. The number of gill rakers on the ventral bone of the first gill arch on the left side of the fish was counted, from the base of the ceratobranchial to the epibranchial–ceratobranchial joint. In addition, the lengths of the second, third and fourth gill rakers (from the epibranchial–ceratobranchial joint on the ceratobranchial) were each measured three times with the stage micrometre. For subsequent analyses, we used the mean (across the three rakers) of the median lengths (of each gill raker) of each fish. We once again used sex-specific analyses for both gill raker length and number.

Statistical analyses

All analyses were performed in R version 3.0.2 (R Development Core Team, 2012), using the packages ggplot2 (Wickham, 2009), car (Fox & Weisberg, 2011),

MASS (Venables & Ripley, 2002), lsmeans (Length, 2015), heplots (Fox *et al.*, 2010), rms (Harrell, 2015) and psych (Revelle, 2015). All models were analysed using type III sums of squares, and inferences were made at $\alpha = 0.05$. Analyses for male geometric morphometric data were performed on relative warps (RWs) extracted from tpsRelw. RW1 (33.62% of the variation) was easily identifiable as being mostly due to variation in body depth. RW2 (16.94%) appeared to be influenced by bending of the body caused by the placement of the fish in the photographs, but was included in analyses because bend could not clearly be separated from potential biological variation (interpretations did not change if RW2 was excluded). RW3 (10.80%), RW4 (8.93%) and RW5 (5.18%) were all influenced by variation in the relative length and shape of the head. All other RWs accounted for <5% of overall shape variation. Due to sexual dimorphism, we analyse and present male and female data separately, except for analyses from which females were excluded (RWs and body depth).

1. Are lake–stream differences present within each watershed in a common garden?

Here, we used three MANCOVAS (one for male RWs, one for male univariate shape traits and one for female univariate shape traits) to analyse the fixed effect of habitat (lake or stream) for the CG fish within each watershed. The logarithm of centroid size (RWs) or standard length (univariate shape traits) was included as a covariate in each model. We also individually analysed several traits of special interest (Hendry & Taylor, 2004; Berner *et al.*, 2008, 2009; Kaeuffer *et al.*, 2012): male body depth (RW1), male and female gill raker number, and male and female gill raker length. All three traits were analysed with AN(C)OVA, using the same model structure as in the above MANCOVAS, except that no size covariate was necessary for analyses on gill raker number.

2. Does plasticity increase or decrease lake–stream differences within watersheds?

Here, we compared WC lake–stream differences to CG lake–stream differences within each watershed. We first analysed WC fish in the manner described in question 1 for CG fish and qualitatively compared results of the two analyses. If phenotypic differences (effect size for the habitat term) are greater in WC than CG fish, then plasticity increases lake–stream differences (or *vice versa*). We then analysed WC and CG fish together using similar models as described above, but with an added fixed factor for rearing environment (wild or lab). A significant habitat-by-rearing environment interaction would provide statistical support for the above assessment of whether divergent phenotypic plasticity (present in wild but not lab) increases or decreases between-ecotype differences.

3. To what extent are lake–stream differences parallel in common-garden conditions?

We here repeated the CG analyses from question 1, but this time including all watersheds together, with watershed added as a fixed factor to allow inferences about specific watersheds and also because watersheds were specifically selected and not a random set. In this analysis, several terms, coupled with inspection of the direction of differences, are important (Langerhans & DeWitt, 2004; Kaeuffer *et al.*, 2012): (i) the habitat term informs parallelism across watersheds in lake–stream CG differences, (ii) the watershed term informs variation across watersheds independent of habitat and (iii) the habitat-by-watershed interaction informs lake–stream differences that are nonparallel across watersheds.

Given our full-sib design for CG fish, family effects would ideally have been included as a random effect in the analyses. However, this was not possible because families were unknown for WC fish and because the generally low and unbalanced number of CG fish per family preclude fitting multivariate mixed models. Instead, as a simple and robust way to determine whether family identity influenced the relative amount of variation explained by the various model terms, we compared ANCOVAS on RW1 and all univariate traits for CG fish from the two largest families (2–11 individuals) for each watershed-habitat combination. These models had the same structures as those described above, but family was either excluded or included as a random effect. In most cases, both models resulted in the same interpretations, and AIC was lower (ΔAIC>2) for models that did not include family. In the few cases with differences, including family did not significantly improve the model fit as judged by AIC and maximum likelihood. Family effects thus had little influence on these data.

4. To what extent does plasticity enhance or dampen parallelism in the wild?

We here repeated the analyses from question 3, but this time including both rearing environments (CG and WC) as well as all watersheds. We first qualitatively compared effect sizes in separate WC and CG models. For a metric of parallelism (relative to nonparallelism), we divided the effect size (partial η^2) for parallelism (the habitat term) by the effect size for nonparallelism (the habitat-by-watershed interaction). If plasticity enhances parallelism, we should see a greater relative effect of parallelism in WC fish than CG fish. These inferences based on effect size were complemented by the examination of the directions of lake–stream

divergence for univariate traits and for multivariate traits along the eigenvector of divergence *d* for the habitat term ($d_{habitat}$, Langerhans, 2009). We next analysed the CG and WC fish from all watersheds together. In this analysis, (1) a habitat-by-rearing environment interaction would inform the effect of plasticity on parallelism, and (2) a significant three-way habitat-by-watershed-by-rearing environment interaction would indicate that nonparallelism differed between WC and CG fish.

Exchangeability analyses

Interpretations of parallelism can be strengthened using the full distribution of observed phenotypes to ask how easily individuals might be exchanged between populations based on their phenotypes (Hendry *et al.*, 2013a). If this 'exchangeability' is high within the same habitat type but not between habitat types, parallelism is indicated, whereas the opposite pattern would indicate nonparallelism. As exchangeability could be trait specific and context specific, we performed separate analysis on WC and CG fish for each trait. We use the dapc function from the adegenet package (Jombart, 2008; Jombart *et al.*, 2010; Jombart & Ahmed, 2011) to perform discriminant analysis on principle components as explained in Jombart *et al.* (2010; see also Hendry *et al.*, 2013a). We select the lowest number of principle components that captured at least 90% of variation (1 for traits investigated individually, 4 for RWs and 5 for univariate shape traits), and all discriminant functions that represented more than about 2% of the variation (1 for traits investigated individually, 4 for RWs and 5 for univariate shape traits). Because DAPC does not readily allow for covariates and we wished to conduct exchangeability analysis independent of size, for this analysis, we used traits allometrically standardized to a common body size (Reist, 1985; Lleonart *et al.*, 2000; see supplemental information for details).

To inform questions 3 and 4, fish were grouped into four categories based on the population into which they were classified: their population of origin (home), the population from the same watershed but the opposite habitat type (parapatric), a population from a different watershed of the same habitat type (allopatric same) or a population from a different watershed of the opposite habitat type (allopatric different). To standardize for random expectations, the number of fish in each category was divided by the number of sites in that category. After comparing the correctly classified and misclassified fish (henceforth referred to as the misclassification analysis), we performed cross-classification by classifying fish to populations excluding the population of origin. Cross-classification is expected to be more powerful than misclassification in informing exchangeability (Hendry *et al.*, 2013a).

Results

Plastic and genetic differences both influenced lake–stream stickleback phenotypes, and their relative effect sizes varied across watersheds and traits. Consistent with previous studies, we found parallelism for some traits but also substantial nonparallelism.

1. Are lake–stream differences present within each watershed in a common garden?

Lake–stream differences were evident in all multivariate analyses on body shape in CG fish in all watersheds, except female univariate shape traits in Misty (Table 1; Fig. 2), indicating that shape differences are genetically based. Unexpectedly, although lake–stream differences were present in univariate shape traits in Roberts (except in WC females), the direction of this difference was opposite to that in Misty and Boot, in both WC and CG fish (Fig. 2). Univariate analyses of other key traits in CG fish generally also showed genetically based differences, but with some nuances. For example, body depth (RW1) was greater in stream fish than in lake fish in Misty and Boot, but not in Roberts (Table 1, Fig. 3a). This suggests that lake–stream differences observed in WC fish from Roberts are plastic. Lake–stream differences in the length of gill rakers were variable in direction and magnitude in CG fish (Figs 3b and S1a): in Boot, gill rakers were longer in stream fish; in Misty, the reverse was true for females but not males; and, in Roberts, no differences were detected. Finally, divergence in gill raker number appears to be genetically based, as lake fish had more rakers than did stream fish in all watersheds (Table 1; Figs 3c and S1b). Thus, lake–stream differences usually had a genetic basis, although not for all traits in all watersheds.

2. Does plasticity increase or decrease lake–stream differences within watersheds?

Based on comparison of lake–stream differences in CG and WC fish, it appears that plasticity usually does not strongly influence the magnitude of divergence within each watershed. Plasticity increased phenotypic differences (more variance explained by the habitat term in WC than CG fish) in two cases (sex/trait/watershed combinations), decreased it in seven cases and had no influence in fifteen cases (Table 1, Table S4). For no trait was the habitat effect uniformly greater or lesser for CG than for WC fish in all three watersheds. Some traits that showed genetically based differences in the CG fish were also significantly increased or decreased by plasticity (habitat-by-rearing environment interaction; Table 1, Table S4). Interestingly, CG lake–stream differences were sometimes not seen in WC fish. This was the case, for instance, for geometric morphometric shape in Roberts males, and gill raker length in both

Table 1 Results for the habitat term from analyses performed CG and WC fish for each watershed to determine lake–stream differences are present in the common garden (question 1) and to compare CG lake–stream differences to those in WC fish (question 2).

		Watershed					
		Misty		Boot		Roberts	
Trait type		WC	CG	WC	CG	WC	CG
Male geometric morphometric shape	Partial η^2	**0.976**	**0.916**	**0.904**	**0.979**	0.890	**0.952**
	F	**$F_{24,11}$ = 19.0**	**$F_{24,14}$ = 6.32**	**$F_{24,19}$ = 7.42**	**$F_{24,4}$ = 7.59**	$F_{24,8}$ = 2.70	**$F_{24,8}$ = 6.59**
	P-value	**<0.0001**	**0.0004**	**<0.0001**	**0.031**	0.074	**0.005**
		Increase		Decrease*		Decrease	
Univariate shape traits – males	Partial η^2	**0.859**	**0.621**	**0.692**	**0.933**	**0.618**	**0.649**
	F	**$F_{9,21}$ = 14.2**	**$F_{9,29}$ = 5.28**	**$F_{9,24}$ = 5.99**	**$F_{9,15}$ = 23.1**	**$F_{9,17}$ = 3.05**	**$F_{9,21}$ = 4.31**
	P-value	**<0.0001**	**0.00028**	**0.00021**	**<0.0001**	**0.023**	**0.0028**
		Increase		Decrease		Decrease	
Univariate shape traits – females	Partial η^2	0.192	0.178	**0.688**	**0.557**	0.246	**0.687**
	F	$F_{4,26}$ = 1.54	$F_{4,33}$ = 1.79	**$F_{4,21}$ = 11.6**	**$F_{4,30}$ = 0.42**	$F_{4,22}$ = 1.79	**$F_{4,19}$ = 10.4**
	P-value	0.219	0.15	**<0.0001**	**<0.0001**	0.17	**0.00012**
		Increase		Increase		Decrease*	
Body depth (RW1)	Partial η^2	**0.814**	**0.486**	**0.4580**	**0.512**	**0.204**	0.019
	F	**$F_{1,34}$ = 149**	**$F_{1,37}$ = 35.1**	**$F_{1,42}$ = 35.5**	**$F_{1,27}$ = 28.4**	**$F_{1,31}$ = 7.94**	$F_{1,31}$ = 0.601
	P-value	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**0.008**	0.44
		Increase		Decrease*		Increase*	
Gill raker length – males	Partial η^2	0.0281	**0.101**	0.0489	**0.409**	0.00674	0.0111
	F	$F_{1,33}$ = 0.954	**$F_{1,37}$=4.15**	$F_{1,44}$ = 2.26	**$F_{1,25}$=17.3**	$F_{1,32}$ = 0.217	$F_{1,32}$ = 0.359
	P-value	0.34	**0.049**	0.14	**0.00033**	0.64	0.55
		Decrease§		Decrease§*		Decrease	
Gill raker length – females	Partial η^2	**0.163**	**0.374**	**0.190**	**0.295**	0.00570	0.0734
	F	**$F_{1,32}$ = 6.25**	**$F_{1,36}$ = 21.5**	**$F_{1,27}$ = 6.35**	**$F_{1,35}$ = 14.7**	$F_{1,32}$ = 0.184	$F_{1,22}$ = 1.74
	P-value	**0.018**	**<0.0001**	**0.018**	**0.00051**	0.67	0.20
		Decrease§*		Decrease§*		Decrease	
Gill raker number – males	Partial η^2	**0.325**	**0.487**	**0.739**	**0.190**	0.089	**0.243**
	F	**$F_{1,30}$ = 14.4**	**$F_{1,37}$ = 35.2**	**$F_{1,45}$ = 128**	**$F_{1,26}$ = 6.11**	$F_{1,34}$ = 3.35	**$F_{1,33}$ = 10.6**
	P-value	**0.0007**	**<0.0001**	**<0.0001**	**0.020**	0.076	**0.003**
		Decrease		Increase*		Decrease	
Gill raker number – females	Partial η^2	**0.251**	**0.279**	**0.525**	**0.369**	0.005	**0.447**
	F	**$F_{1,29}$ = 9.75**	**$F_{1,37}$ = 14.3**	**$F_{1,27}$ = 29.9**	**$F_{1,36}$ = 21.1**	$F_{1,32}$ = 0.162	**$F_{1,23}$ = 18.6**
	P-value	**0.004**	**0.0005**	**<0.0001**	**<0.0001**	0.690	**0.0003**
		Decrease		Increase		Decrease*	

Significant values, in bold, represent genetically based divergence. Traits for which plasticity significantly increases or decreases lake–stream differences are in bold with an asterisk(*) (question 2). The symbol § represents cases in which divergence is in the opposite direction in the lab. Females were excluded from geometric morphometric analyses (see Materials and methods).

sexes in Roberts. For the latter trait, genetic differences were countered by plastic effects in nature, and countergradient variation was evident in Boot and Misty: the ecotype with longer rakers in the wild had shorter rakers in the laboratory (Figs 3b and S1a). Overall, plasticity sometimes increased, sometimes decreased and more often had no influence on lake–stream differences.

3. To what extent are lake–stream differences parallel in common-garden conditions?

Analyses on CG fish provided evidence for both genetically based parallelism (habitat term) and nonparallelism (habitat-by-watershed interaction) (Table 2). In multivariate analyses on shape, parallelism was always greater than nonparallelism, although the difference was not always large (Table 2). For body depth, the effect size was somewhat larger for nonparallelism (partial η^2 = 0.42) than parallelism (partial η^2 = 0.32), primarily because lake–stream differences were evident in CG fish for Misty and Boot but not for Roberts (Fig. 3a). For gill raker length, nonparallelism was stronger than parallelism (Table 2) owing to variation across watersheds in the magnitude and direction of divergence (Figs 3b and S1a). Gill raker number was highly parallel across watersheds (Table 2; Figs 3c and S1b) and was not significantly influenced by nonparallelism. Thus, genetically based lake–stream differences include both parallel and nonparallel components, with

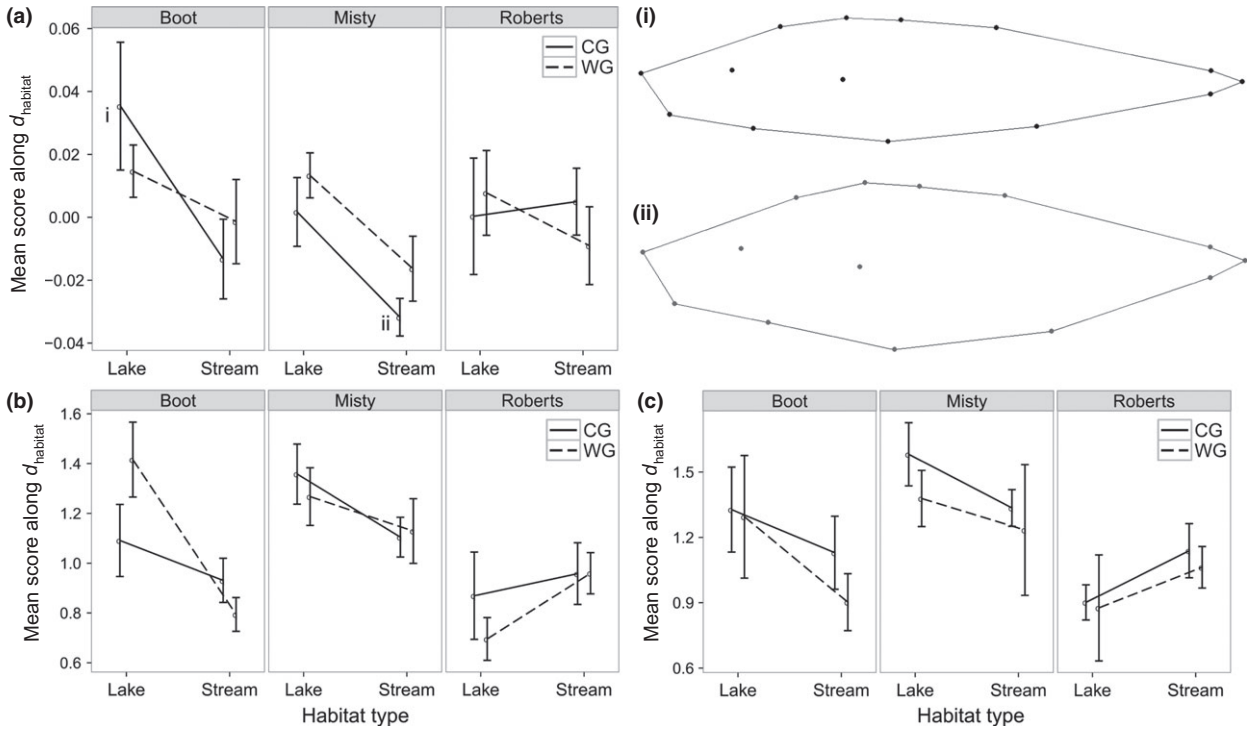

Fig. 2 Lake–stream differences in geometric (a) and univariate shape (b and c) in common-garden (CG) and wild-caught (WC) fish. Similarity in the angle of CG and WC lines within a watershed (within each panel) indicates a genetic basis for shape differences between lake and stream fish. Similarity in the angle of lines within rearing environments (CG or WC) across watersheds (across panels) indicates parallelism. (a) Left: mean scores along the habitat divergence vector ($d_{habitat}$, following Langerhans, 2009) for the main model (question 4) based on geometric morphometric shape for male CG (solid lines) and WC (dashed lines) fish (females were excluded from geometric morphometric analyses). Right: shape deforms primarily in body depth between lake and stream fish. Here, the shape deformations are shown for the most extreme group mean scores along the $d_{habitat}$, the shallowest (top, CG Boot lake, indicated on the $d_{habitat}$ plot by 'i') and deepest (bottom, CG Misty stream, indicated on the $d_{habitat}$ plot by 'ii') bodied groups. (b) Mean $d_{habitat}$ scores based on male univariate shape traits. (c) Mean $d_{habitat}$ scores based on female univariate shape traits. CG and WC $d_{habitat}$ values for a given trait type were estimated from the same models (question 4).

some traits (gill raker length) and some watersheds (Roberts) being especially nonparallel.

Exchangeability analyses supported these inferences. For shape traits, CG fish were overwhelmingly most likely to be classified back to their population of origin (Figs 5 and S3), suggesting that body shape includes substantial components unique to each population. Exclusion of the home population as a classification option for univariate shape traits revealed high 'cross-classification' to the parapatric population of the other habitat type, indicating watershed-specific variation independent of habitat type. Other traits showed higher parallelism. For RWs, cross-classification was much more likely to the allopatric population of the same habitat type. For gill raker number and body depth, fish were again most likely to be classified to their home population, but cross-classified to the allopatric population of the same habitat type. By contrast, cross-classification based on gill raker length was most likely to the parapatric population in the opposing habitat, confirming the lack of genetically based parallel lake–stream differences for this trait.

4. To what extent does plasticity enhance or dampen parallelism in the wild?

Parallelism was qualitatively greater for WC fish than for CG fish for shape traits, including body depth (Fig. 4, Table 2), and this inference was statistically confirmed in almost all cases through a significant habitat-by-watershed-by-rearing environment interaction (Table 3). That is, environmentally induced effects in nature generally enhanced phenotypic parallelism in shape traits relative to that seen for genetic differences manifest in a common environment. However, the opposite was true for gill raker measurements for two keys reasons. First, for gill raker length, plasticity appeared to dampen parallelism – but this effect is misleading because this trait is highly nonparallel and thus even modestly more similar phenotypes seem more parallel despite overall

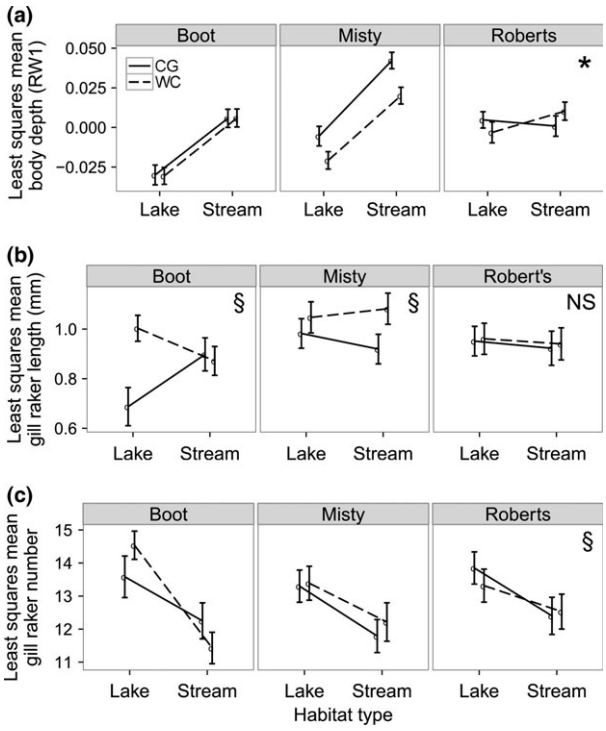

Fig. 3 Lake–stream parallelism in body depth and gill raker traits in males. (a) Stream fish had significantly deeper bodies than lake fish in both common-garden (CG, solid lines) and wild-caught (WC, dashed lines) fish in the Boot and Misty watersheds, but only in WC fish in the Roberts watershed. Thus, body depth divergence has a genetic basis in Boot and Misty, but a plastic basis in Roberts. Here, we show least squares mean body depth (RW1, no units) obtained from the main model including all male fish (see Materials and methods for question 4; females were excluded from geometric morphometric analyses). (b) In males, divergence in gill raker length differed across watersheds. Moreover, divergence reversed direction in the CG fish compared to WC fish in Boot and Misty. Plasticity entirely overwhelmed genetically based divergence in this trait. Here, we show least squares mean gill raker length (mm) obtained from the main model including all fish (see Materials and methods for question 4). (c) Male lake fish had more gill rakers than stream fish in all watersheds and both rearing environments, suggesting that this trait has a genetic basis. Despite differences in the magnitude of divergence in Boot and Roberts, no significant plasticity was detected in this trait. Here, we show least squares mean gill raker number obtained from the main model including all fish (see Materials and methods for question 4). Error bars represent 95% confidence intervals. Panels with * indicate that lake–stream differences are only significant in CG fish, panels with § indicate that lake–stream differences are only significant in WC fish, and those with 'NS' indicate that lake–stream differences were not significant in either CG or WC fish.

nonparallelism across watersheds. Second, for gill raker number, plasticity dampened parallelism – simply because some (minor) nonparallelism was evident for WC fish but none was evident for CG fish. Thus, plasticity usually (although not always) enhances lake–stream phenotypic parallelism in shape in the wild, but not in gill raker traits.

In exchangeability analyses, classification was most common to the population of origin for both WC and CG fish (Figs 5 and S3). However, results differed for 'cross-classification' analyses that excluded the population of origin. Most strikingly, geometric morphometric (and to a lesser extent univariate shape) cross-classification to allopatric populations in similar habitats was much more likely for WC fish than for CG fish. This result confirms the above assertions that plasticity generally enhances lake–stream parallelism. By contrast, cross-classification results were similar for WC and CG fish for body depth and gill raker number (because parallelism was high in both cases) and gill raker length (because parallelism was low in both cases).

Discussion

Our results (summarized in Table 4) indicate that lake–stream phenotypic divergence in threespine stickleback is often genetically based, as is lake–stream parallelism across watersheds. At the same time, phenotypic divergence in nature is also influenced by plasticity, as seen in different patterns for common-garden (CG) fish vs. wild-caught (WC) fish. Although this plasticity had variable effects on within-watershed lake–stream divergence, it generally enhanced phenotypic parallelism in shape in nature. We now consider these outcomes and their implications in more detail.

1. Are lake–stream differences present within each watershed in a common garden?

The genetic vs. plastic basis for lake–stream differences varied among traits, with the importance of genetic effects being greatest for body shape and gill raker number but the importance of plastic effects being greatest for gill raker length (Table 1, Figs 2 and 3). These differences among traits could be shaped by different patterns of inheritance (Berner *et al.*, 2011; Hangartner *et al.*, 2012). For instance, our results are consistent with previous studies on stickleback demonstrating a strong genetic basis for body shape (McPhail, 1992; Cresko *et al.*, 2004; Schluter *et al.*, 2004; Albert *et al.*, 2008; Berner *et al.*, 2011; Leinonen *et al.*, 2011; Rogers *et al.*, 2012; Arnegard *et al.*, 2014) and gill raker number (Hagen, 1973; McPhail, 1992; Peichel *et al.*, 2001; Aguirre *et al.*, 2004; Berner *et al.*, 2011; Arnegard *et al.*, 2014; Glazer *et al.*, 2014). By contrast, although some studies have detected a genetic basis for differences in gill raker length (Lavin & McPhail, 1987; Schluter, 1996; Hatfield, 1997), other studies have reported a strong plastic effect in stickleback (Day *et al.*, 1994; Svanbäck & Schluter, 2012; Wund *et al.*, 2012; Lucek *et al.*, 2014b) and in many other fishes (Robinson & Parsons, 2002). In our study, lake–stream

Table 2 Results for wild-caught (WC) and common-garden (CG) fish analysed separately, with all watersheds included in the analyses to determine whether CG fish show parallelism (question 3) and to compare the extent of parallelism in WC and CG fish (question 4).

			Wild caught		Common garden	
			Males	Females	Males	Females
Geometric morphometrics	Habitat	Partial η^2	0.736		0.718	
		F	$F_{24,86} = 9.98$		$F_{24,74} = 7.85$	
		P	<0.0001		<0.0001	
	Site	Partial η^2	0.623		0.632	
		F	$F_{48,174} = 5.98$		$F_{48,150} = 5.36$	
		P	<0.0001		<0.0001	
	Habitat-by-watershed	Partial η^2	0.561		0.568	
		F	$F_{48,174} = 4.64$		$F_{48,150} = 4.11$	
		P	<0.0001		<0.0001	
Univariate shape traits	Habitat	Partial η^2	0.615	0.421	0.568	0.332
		F	$F_{9,80} = 14.2$	$F_{4,77} = 14.0$	$F_{9,83} = 12.1$	$F_{4,90} = 11.2$
		P	<0.0001	<0.0001	<0.0001	<0.0001
	Watershed	Partial η^2	0.342	0.171	0.414	0.335
		F	$F_{18,162} = 4.68$	$F_{8,156} = 4.02$	$F_{18,168} = 6.58$	$F_{8,182} = 11.5$
		P	<0.0001	0.00023	<0.0001	<0.0001
	Habitat-by-watershed	Partial η^2	0.284	0.147	0.345	0.281
		F	$F_{18,162} = 3.56$	$F_{8,156} = 3.37$	$F_{18,168} = 4.91$	$F_{8,182} = 8.90$
		P	<0.0001	0.0014	<0.0001	<0.0001
Body depth (RW1)	Habitat	Partial η^2	0.423		0.317	
		F	$F_{1,109} = 79.8$		$F_{1,97} = 44.9$	
		P	<0.0001		<0.0001	
	Watershed	Partial η^2	0.244		0.379	
		F	$F_{2,109} = 17.6$		$F_{2,97} = 29.5$	
		P	<0.0001		<0.0001	
	Habitat-by-watershed	Partial η^2	0.220		0.418	
		F	$F_{2,109} = 15.4$		$F_{2,97} = 34.8$	
		P	<0.0001		<0.0001	
Gill raker length	Habitat	Partial η^2	0.0211	0.0433	0.137	0.149
		F	$F_{1,113} = 2.43$	$F_{1,93} = 4.21$	$F_{1,96} = 15.2$	$F_{1,95} = 16.6$
		P	0.12	0.043	0.00018	<0.0001
	Watershed	Partial η^2	0.06700	0.0058	0.291	0.303
		F	$F_{2,113} = 4.09$	$F_{2,93} = 0.269$	$F_{2,96} = 19.7$	$F_{2,95} = 20.6$
		P	0.019	0.76	<0.0001	<0.0001
	Habitat-by-watershed	Partial η^2	–	0.124	0.168	0.289
		F	–	$F_{2,93} = 6.55$	$F_{2,96} = 9.67$	$F_{2,95} = 19.3$
		P	–	0.0022	0.00015	<0.0001
Gill raker number	Habitat	Partial η^2	0.487	0.244	0.285	0.349
		F	$F_{1,109} = 103$	$F_{1,88} = 28.4$	$F_{1,98} = 39.0$	$F_{1,98} = 52.6$
		P	<0.0001	<0.0001	<0.0001	<0.0001
	Watershed	Partial η^2	0.154	0.059	0.0464	0.0352
		F	$F_{2,109} = 9.91$	$F_{2,88} = 2.76$	$F_{2,98} = 2.39$	$F_{2,98} = 1.77$
		P	0.0001	0.07	0.1	0.2
	Habitat-by-watershed	Partial η^2	0.214	0.121	–	–
		F	$F_{2,109} = 14.8$	$F_{2,88} = 6.07$	–	–
		P	<0.0001	0.003	–	–

The habitat term represents parallelism, and the habitat-by-watershed environment interaction represents nonparallelism. Nonsignificant interactions were removed from the models, shown here with a '–'. Females were excluded from geometric morphometric analyses (see Materials and methods).

divergence in gill raker length often differed in direction between WC and CG fish, suggesting that plastic effects act in opposition to genetic effects: that is, counter-gradient variation (*sensu* Conover & Schultz, 1995).

The genetic vs. plastic basis for lake–stream differences also varied among watersheds, with the importance of genetic effects being greatest for Misty and Boot but the importance of plastic effects being greatest for Roberts (Table 1). Several potential explanations can be advanced for these differences. First, divergence in Roberts could be more recent than in the other watersheds, yet all of our watersheds are of

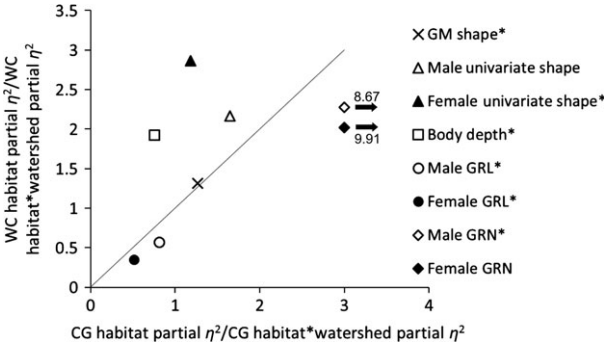

Fig. 4 Parallelism is greater in the wild than in the common garden, thus plasticity enhances parallelism in most shape traits but not gill raker traits. The relative effect of parallelism is shown here as the partial η^2 of the habitat term (parallelism) divided by the partial η^2 of the habitat-by-watershed term (nonparallelism) from models of wild-caught (WC) and common-garden (CG) fish (see Materials and methods, questions 3 and 4). Values falling on the 1:1 line would represent traits that show no influence of plasticity on parallelism (no difference between CG and WC fish). Above the 1:1 line, parallelism is greater in WC fish than CG fish, thus plasticity enhances parallelism. Below it, the opposite is true, and plasticity dampens parallelism. Geometric morphometric is abbreviated to GM. Gill raker length and number are abbreviated to GRL and GRN, respectively. The influence of plasticity was significant in all cases (marked with an asterisk) except male univariate shape traits and female gill raker number. The relative effect of parallelism was much higher for gill raker number in CG fish than any other results, so gill raker number is shown shifted towards zero along the *x*-axis to allow all points to be shown together. The true values for the *x*-axis are shown above (males) or below (female) the arrows.

approximately the same age (Caldera & Bolnick, 2008; Berner *et al.*, 2009). Second, temporal and/or spatial variation could be higher in Roberts than in the other watersheds, which would favour plasticity over genetic divergence (reviewed by Scheiner, 1993; Sultan & Spencer, 2002), yet spatial variation in prey is actually lower in Roberts than in Boot (Berner *et al.*, 2009). Third, lake–stream gene flow could be higher in Roberts than in the other watersheds, which also could promote plasticity over genetic divergence (Sultan & Spencer, 2002; Crispo, 2008; Thibert-Plante & Hendry, 2011). We favour this last explanation because neutral genetic divergence is much lower in Roberts (F_{ST} = 0.045) than in either Misty (F_{ST} = 0.121) or Boot (F_{ST} = 0.178) (Kaeuffer *et al.*, 2012). Moreover, the lower genetic divergence in Roberts than Boot is known to be genomewide (Misty Inlet has yet to be sequenced) and thus likely also applies to genes under selection (Roesti *et al.*, 2012).

2. Does plasticity increase or decrease lake–stream differences within watersheds?

The increasing (cogradient) vs. decreasing (countergradient) effects of plasticity on lake–stream differences varied among watersheds and traits. For example, in Roberts, plasticity strongly increased lake–stream differences in body depth in the wild relative to the common-garden environment, whereas it strongly decreased wild lake–stream differences in gill raker number (Table 1, Figs 3 and S2). This variation has important implications for assumptions about the genetic vs. plastic basis of trait differences in nature because common-garden results obtained from one instance of an environmental contrast might not extrapolate to other instances of the same contrast. Thus, studies attempting general inferences about habitat-associated phenotypic divergence require common-garden studies conducted for multiple independent instances of population divergence between those habitat types.

Our findings further suggest that seeking a general effect of plasticity could be misguided. For instance, many studies find that fish growth shows strong counter-gradient variation, such that phenotypic trends along environmental gradients are much lower in nature than they are in a common-garden environment (Conover *et al.*, 2009). The generality of this pattern for growth might imply a similarly general effect for other traits (e.g. always counter-gradient or cogradient), yet our results show that such an assumption would be incorrect. Instead, trait variation along environmental gradients is likely to show different patterns in different locations. Such variation might be especially likely when testing for parallel evolution because independent lineages likely vary in evolutionary histories, genetic variation and plasticity. Moreover, our finding suggests an interesting analogy to studies on the genetic basis of parallelism: just as one cannot assume that phenotypic parallelism reflects shared genetic changes (Arendt & Reznick, 2008; Chan *et al.*, 2010; Westram *et al.*, 2014), one also cannot assume it reflects shared plastic effects.

3. To what extent are lake–stream differences parallel in common-garden conditions?

Most studies of parallel evolution rely on field-collected samples, perhaps supplemented by common-garden results from only one or a few populations. To confirm this assertion, we scrutinized 75 studies of parallel evolution in fishes and found that only 15 included common-garden experiments (K.B. Oke, C. LeBlond & A.P. Hendry, unpublished). For the lake–stream stickleback system, many studies of phenotypic parallelism have been conducted, yet common-garden experiments are rare (see Introduction). The problem with this state of affairs is that detecting parallel *evolution* requires an assessment of phenotypes while eliminating, or at least minimizing, differential environmental effects that might cause plastic differences.

Table 3 Results for main models (question 4), all fish analysed together. The habitat term represents parallelism, and the habitat-by-watershed-by-rearing environment interaction represents nonparallelism.

Term		Geometric morphometric shape	Univariate shape traits		Body depth (RW1)	Gill raker length		Gill raker number	
		Males	Males	Females	Males	Males	Females	Males	Females
Habitat	Partial η^2	0.517	0.659	0.102	0.246	0.0390	0.0669	0.046	0.225
	F	$F_{24,184} = 8.22$	$F_{9,176} = 37.7$	$F_{4,172} = 4.90$	$F_{1,207} = 67.7$	$F_{1,208} = 8.44$	$F_{1,189} = 13.6$	$F_{1,205} = 10.0$	$F_{1,189} = 54.8$
	P	<0.0001	<0.0001	0.00091	<0.0001	0.004	0.00030	0.002	<0.0001
Watershed	Partial η^2	0.499	0.514	0.299	0.268	0.108	0.179	0.012	0.031
	F	$F_{48,370} = 7.68$	$F_{18,354} = 20.8$	$F_{8,346} = 18.4$	$F_{2,207} = 38.0$	$F_{2,208} = 12.6$	$F_{2,189} = 20.6$	$F_{2,205} = 1.24$	$F_{2,189} = 3.02$
	P	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.291	0.051
Rearing environment	Partial η^2	0.317	0.112	0.022	0.000	0.126	0.0583	0.029	0.006
	F	$F_{24,184} = 3.56$	$F_{9,176} = 2.46$	$F_{4,172} = 0.94$	$F_{1,207} = 0.00$	$F_{1,208} = 30.1$	$F_{1,189} = 11.7$	$F_{1,205} = 6.14$	$F_{1,189} = 1.08$
	P	<0.0001	0.012	0.44	1.0	<0.0001	0.00076	0.014	0.301
Habitat-by-watershed	Partial η^2	0.397	0.272	0.141	0.301	0.0514	0.163	–	0.047
	F	$F_{48,370} = 5.07$	$F_{18,354} = 7.34$	$F_{8,346} = 7.11$	$F_{2,207} = 44.5$	$F_{2,208} = 5.63$	$F_{2,189} = 18.4$	–	$F_{2,189} = 4.67$
	P	<0.0001	<0.0001	<0.0001	<0.0001	0.0041	<0.0001	–	0.010
Habitat-by-Rearing environment	Partial η^2	0.273	–	–	–	0.0558	0.0761	0.052	–
	F	$F_{24,184} = 2.88$	–	–	–	$F_{1,208} = 12.3$	$F_{1,189} = 15.6$	$F_{1,205} = 11.2$	–
	P	<0.0001	–	–	–	0.00055	0.00011	0.001	–
Watershed-by-Rearing environment	Partial η^2	0.263	0.263	0.0848	0.0355	0.0501	0.0863	0.038	–
	F	$F_{48,370} = 2.75$	$F_{18,354} = 7.00$	$F_{8,346} = 4.01$	$F_{2,207} = 3.81$	$F_{2,208} = 5.48$	$F_{2,189} = 8.93$	$F_{2,250} = 4.06$	–
	P	<0.0001	<0.0001	0.00015	0.024	0.0048	0.00020	0.019	–
Habitat-by-Watershed-by-Rearing environment	Partial η^2	0.205	–	0.0576	0.0421	0.0643	0.197	0.114	–
	F	$F_{48,370} = 1.98$	–	$F_{12,522} = 2.66$	$F_{3,207} = 3.03$	$F_{2,208} = 7.15$	$F_{2,189} = 23.2$	$F_{4,205} = 6.62$	–
	P	0.00024	–	0.0018	0.030	0.0010	<0.0001	<0.0001	–

Nonsignificant interactions were removed from the models, shown here with a '–'. Females were excluded from geometric morphometric analyses (see Materials and methods).

Table 4 Summary of results.

	Question 1: Are lake–stream differences present within each watershed in a common garden?	Question 2: Does plasticity increase or decrease lake–stream differences within watersheds?	Question 3: To what extent are lake–stream differences parallel in common-garden conditions?	Question 4: To what extent does plasticity enhance or dampen parallelism in the wild?
Male geometric morphometric shape	Yes (M,B,R)	Decreases (B) or neither (M,R)	Parallel > nonparallel	Enhances
Male univariate shape	Yes (M,B,R)	Neither (M,B,R)	Parallel > nonparallel	Neither
Female univariate shape	Yes (B,R) or no (M)	Decreases (R) or neither (M,B)	Parallel > nonparallel	Enhances
Male body depth (RW1)	Yes (M,B) or no (R)	Increases (R), decreases (B), or neither (M)	Parallel < nonparallel	Enhances
Male gill raker length	Yes (M,B) or no (R)	Decreases (B) or neither (M,R)	Parallel < nonparallel	Dampens
Female gill raker length	Yes (M,B) or no (R)	Decreases (M,B) or neither (R)	Parallel < nonparallel	Dampens
Male gill raker number	Yes (M,B,R)	Increases (B) or neither (M,R)	Parallel > nonparallel	Dampens
Female gill raker number	Yes (M,B,R)	Decreases (R) or neither (M,B)	Parallel > nonparallel	Neither

We asked four questions, 1) are lake–stream differences present within each watershed in a common garden? 2) does plasticity increase or decrease lake–stream differences within watersheds? 3) to what extent are lake–stream differences parallel in common-garden conditions? 4) to what extent does plasticity enhance or dampen parallelism in the wild? We found that 1) for the majority of traits and watersheds, lake–stream differences were maintained in CG fish, indicating that they have a genetic basis, 2) plasticity varied greatly in its increasing or decreasing influence on lake–stream differences within watersheds, 3) for most traits, CG fish are parallel across watersheds, and 4) plasticity more often enhances parallelism, but sometimes also dampens it. Results are provided by watershed (Misty [B], Boot [B] and Roberts [R]) for questions 1 and 2. For question 3, the greater than sign denotes the term of stronger effect size.

Our examination of this problem yielded several insights. In particular, previous studies of lake–stream stickleback were *generally* safe in asserting/assuming that phenotypic parallelism was genetically based. That is, we here found a strong parallel signature for most traits when analysing common-garden fish: that is, the 'habitat' term was highly significant in statistical models. However, this same assumption would not be safe for some *specific* traits/populations. For instance, the genetically based parallelism evident in a common garden for Misty and Boot stickleback was weak or absent for Roberts. However, Robert's stickleback showed low parallelism even in the wild, which suggests a reasonable qualifier: lake–stream parallelism tends to be genetically based *for populations/traits that show strong phenotypic parallelism in nature*. These results highlight the importance of conducting common-garden studies across multiple populations showing varying degrees of (non)parallelism in nature.

4. To what extent does plasticity enhance or dampen parallelism in the wild?

We found that plasticity typically enhanced (as opposed to dampened) phenotypic parallelism. We expect that this enhancement is adaptive because (1) its effects on divergence closely match well-established adaptive interpretations (Lavin & McPhail, 1993; Walker, 1997; Berner *et al.*, 2008; Hendry *et al.*, 2011) and environmental differences (Berner *et al.*, 2008; Kaeuffer *et al.*, 2012; Hendry *et al.*, 2013a), (2) plasticity acting in the same direction as genetically based divergence (and thus selection) implies that natural selection has shaped both (Crispo, 2007; Ghalambor *et al.*, 2007; Handelsman *et al.*, 2013), and (3) adaptive plasticity has been detected in other stickleback systems (Day *et al.*, 1994; Day & McPhail, 1996; Wund *et al.*, 2008, 2012; Svanbäck & Schluter, 2012; Lucek *et al.*, 2014b; Mazzarella *et al.*, 2015). Thus, genetic divergence and plasticity appear to work in concert to shape parallel adaptive phenotypic divergence in lake–stream stickleback, suggesting a deterministic role for selection in shaping both aspects of trait expression.

Although plasticity usually enhanced parallelism, this outcome was not due to a consistent effect of plasticity across watersheds or traits (see the above question). Instead, plastic contributions to trait divergence were stronger in systems where genetic contributions were weaker. Thus, it seems that phenotypic parallelism reflects a complementary interaction between genetic divergence and plasticity, with the two working together to achieve adaptive trait differences. This suggestion echoes and amplifies the common argument that plasticity aids the initial colonization of (and persistence in) novel environments, which can then allow (and provide the initial variation) for subsequent adaptive genetic divergence (Price *et al.*, 2003; Crispo, 2007; Ghalambor *et al.*, 2007; Moczek *et al.*, 2011). In particular, our results suggest that this sort of interaction persists even in well-established populations, such that plasticity continues to contribute to adaptive phenotypic differences, especially when genetic variation is insufficient (or gene flow too strong) to allow complete adaptive difference.

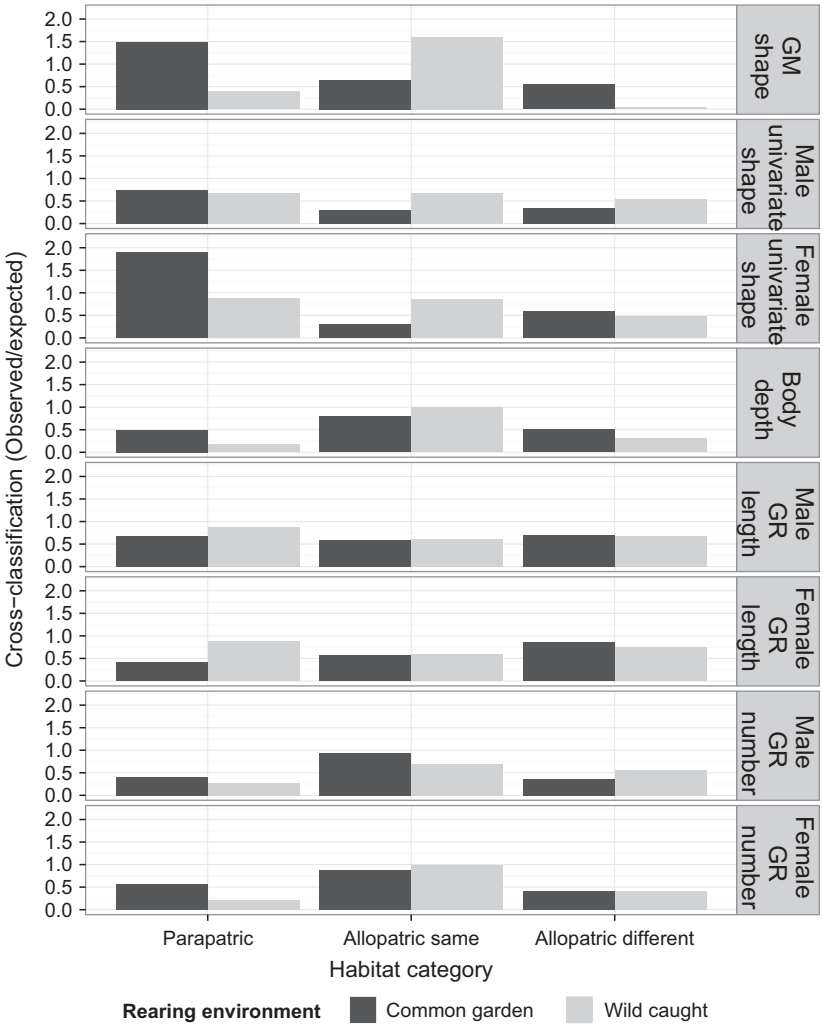

Fig. 5 Cross-classification of common-garden (CG) and wild-caught (WC) fish to populations other than their population of origin shows support for both parallelism and nonparallelism. Here, the number of fish cross-classified into each category is divided by the number of fish that would be expected to be classified to each population in that category at random and by the number of populations in that category. The parapatric category includes the population from the same watershed as their population of origin, but from the other habitat type. The allopatric same category includes populations of the same habitat type from different watersheds, and the allopatric different category includes populations from the different habitat type and different watersheds. GM and GR stand for geometric morphometric and gill raker, respectively.

Caveats and general conclusions

Our study is attended by several caveats. First, we focused on the role of contemporary plasticity; yet, as hinted to above, ancestral plasticity also could have an important role in shaping patterns of trait divergence (West-Eberhard, 2003; Wund *et al.*, 2008; Wund, 2012; Ghalambor *et al.*, 2015; reviewed by Pfennig *et al.*, 2010). Second, we studied only three population pairs, and the variation among them, especially Roberts, implies that additional differences could emerge for other population pairs. Third, GxE dictates that different outcomes might have been obtained under different common-garden conditions, such as in flowing water or with different diets. However, single-population and single-environment common-garden studies have been used to great effect in demonstrating genetically based morphological differences in fishes (e.g. McPhail, 1992; Lavin & McPhail, 1993; Robinson & Parsons, 2002; McGuigan *et al.*, 2003; Langerhans *et al.*, 2004; Burns *et al.*, 2009; Herczeg *et al.*, 2009). Moreover, phenotypic differences in our study were very similar in the laboratory and the wild at least for both Misty and Boot, suggesting that our CG environment revealed at least some biologically relevant patterns. Nevertheless, additional common-garden experiments would be informative for establishing the context dependency (among pairs and environments and for the ancestral state) of trait expression, as well as to explore the potential contributions of maternal effects.

Contributions from phenotypic plasticity are often overlooked in studies of parallel evolution: indeed, most such studies are performed on only wild-caught individuals. Our results indicate that plasticity can contribute in important, and apparently adaptive, ways to parallelism in nature. In particular, plasticity often enhanced parallelism across watersheds and thereby seemingly aided adaptive phenotypic divergence. Importantly, the specific contributions of plasticity differed among traits and watersheds in ways that

appeared to be 'correcting for' variability in genetic divergence, which might arise owing to gene flow, limited genetic variation or other factors. Of course, the generality of our result awaits additional studies comparing parallelism in the wild to that in common-garden conditions. In addition, an important next step would be to determine whether the phenotypic parallelism we see in lake–stream stickleback has a parallel (or convergent) genetic basis, either at the quantitative genetic level or at the level of genes, alleles, developmental pathways, mutations and nucleotides.

Acknowledgments

The authors wish to thank field crews from 2004, 2010 and 2011: Elspeth Kenny, Cecilia Nease, Luke Currin, Erin Van de Water, Prayitno Goenarto, Jason Khoury, Adrian Burrill, Shahin Muttalib, Jean-Sébastien Moore, Angeliina Alopaeus, Martin Turcotte and Friends of the Marble River. Thanks also to Caroline LeBlond, Sebastien Romero and Matthieu Delcourt, as well as student volunteers and students of the McGill Work Studies programme for their help rearing common-garden fish, and Mingsha Zhou for help landmarking photographs. Feedback from Axios review greatly improved the manuscript and discussions with Dean Adams, Michael Collyer, Brian Langerhans and Antigoni Kaliontzopoulou, as well as the suggestions of reviewers improved our analyses. The authors also thank Western Forest Products Inc. for providing accommodations and access to field sites. This work was funded by an NSERC-CGSM and PGSD to KBO, by NSF grants DEB-1144773 to DIB and APH and DEB-1144556 to CLP, and the Howard Hughes Medical Institute (DIB).

References

Aguirre, W.E., Doherty, P.K. & Bell, M.A. 2004. Genetics of lateral plate and gill raker phenotypes in a rapidly evolving population of threespine stickleback. *Behaviour* **141**: 1465–1483.

Albert, A.Y.K., Sawaya, S., Vines, T.H., Knecht, A.K., Miller, C.T., Summers, B.R. *et al.* 2008. The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution* **62**: 76–85.

Arendt, J. & Reznick, D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation?. *Trends Ecol. Evol.* **23**: 26–32.

Arnegard, M.E., McGee, M.D., Matthews, B., Marchinko, K.B., Conte, G.L., Kabir, S. *et al.* 2014. Genetics of ecological divergence during speciation. *Nature* **511**: 307–311.

Baker, J.A., Räsänen, K., Moore, J. & Hendry, A.P. 2013. Genetic and plastic contributions to trait divergence between parapatric habitats: female life-history traits in threespine stickleback within the Misty Lake system. *Evol. Ecol. Res.* **15**: 473–487.

Bell, M.A. & Foster, S.A. eds. 1994. *The Evolutionary Biology of the Threespine Stickleback*. Oxford University Press, Oxford, UK.

Berner, D., Adams, D.C., Grandchamp, A.C. & Hendry, A.P. 2008. Natural selection drives patterns of lake-stream divergence in stickleback foraging morphology. *J. Evol. Biol.* **21**: 1653–1665.

Berner, D., Grandchamp, A.C. & Hendry, A.P. 2009. Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. *Evolution* **63**: 1740–1753.

Berner, D., Kaeuffer, R., Grandchamp, A.C., Raeymaekers, J.A.M., Räsänen, K. & Hendry, A.P. 2011. Quantitative genetic inheritance of morphological divergence in a lake–stream stickleback ecotype pair: implications for reproductive isolation. *J. Evol. Biol.* **24**: 1975–1983.

Brinsmead, J. & Fox, M.G. 2002. Morphological variation between lake- and stream-dwelling rock bass and pumpkinseed populations. *J. Fish Biol.* **61**: 1619–1638.

Burns, J.G., Di Nardo, P. & Rodd, F.H. 2009. The role of predation in variation in body shape in guppies *Poecilia reticulata*: a comparison of field and common garden phenotypes. *J. Fish Biol.* **75**: 1144–1157.

Caldera, E.J. & Bolnick, D.I. 2008. Effects of colonization history and landscape structure on genetic variation within and among threespine stickleback (*Gasterosteus aculeatus*) populations in a single watershed. *Evol. Ecol. Res.* **10**: 575–598.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D. *et al.* 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**: 302–305.

Conover, D.O. & Schultz, E.T. 1995. Phenotypic similarity and the evolutionary significance of countergradient variation. *Trends Ecol. Evol.* **10**: 248–252.

Conover, D.O., Duffy, T.A. & Hice, L.A. 2009. The covariance between genetic and environmental influences across ecological gradients: reassessing the evolutionary significance of countergradient and cogradient variation. *Ann. N. Y. Acad. Sci.* **1168**: 100–129.

Cresko, W.A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P. *et al.* 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc. Natl. Acad. Sci. USA* **101**: 6050–6055.

Crispo, E. 2007. The Baldwin effect and genetic assimilation: revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. *Evolution* **61**: 2469–2479.

Crispo, E. 2008. Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. *J. Evol. Biol.* **21**: 1460–1469.

Day, T. & McPhail, J.D. 1996. The effect of behavioural and morphological plasticity on foraging efficiency in the threespine stickleback (*Gasterosteus sp.*). *Oecologia* **108**: 380–388.

Day, T., Pritchard, J. & Schluter, D. 1994. A comparison of two sticklebacks. *Evolution* **48**: 1723–1734.

Delcourt, M., Räsänen, K. & Hendry, A.P. 2008. Genetic and plastic components of divergent male intersexual behavior in Misty lake/stream stickleback. *Behav. Ecol.* **19**: 1217–1224.

Eroukhmanoff, F., Hargeby, A., Arnberg, N.N., Hellgren, O., Bensch, S. & Svensson, E.I. 2009. Parallelism and historical contingency during rapid ecotype divergence in an isopod. *J. Evol. Biol.* **22**: 1098–1110.

Fitzpatrick, B.M. 2012. Underappreciated consequences of phenotypic plasticity for ecological speciation. *Int. J. Ecol.* **2012**: 1–12.

Fox, J. & Weisberg, S. 2011. *An R Companion to Applied Regression*, 2nd edn. Sage, Thousand Oaks, CA.

Fox, J., Friendly, M. & Monette, G. 2010. Heplots: Visualizing tests in multivariate linear models. R package version 0.9-8. URL http://CRAN.R-project.org/package=heplots

Ghalambor, C.K., McKay, J.K., Carroll, S.P. & Reznick, D.N. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct. Ecol.* **21**: 394–407.

Ghalambor, C.K., Hoke, K.L., Ruell, E.W., Fischer, E.K., Reznick, D.N. & Hughes, K.A. 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* **525**: 372–375.

Glazer, A.M., Cleves, P.A., Erickson, P.A., Lam, A.Y. & Miller, C.T. 2014. Parallel developmental genetic features underlie stickleback gill raker evolution. *Evodevo* **5**: 5–19.

Grether, G.F. 2005. Environmental change, phenotypic plasticity, and genetic compensation. *Am. Nat.* **166**: E115–E123.

Hagen, D.W. 1973. Inheritance of numbers of lateral plates and gill rakers in *Gasterosteus aculeatus*. *Heredity* **30**: 303–312.

Handelsman, C.A., Broder, E.D., Dalton, C.M., Ruell, E.W., Myrick, C.A., Reznick, D.N. *et al.* 2013. Predator-induced phenotypic plasticity in metabolism and rate of growth: rapid adaptation to a novel environment. *Integr. Comp. Biol.* **53**: 975–988.

Hangartner, S., Laurila, A. & Räsänen, K. 2012. The quantitative genetic basis of adaptive divergence in the moor frog (*Rana arvalis*) and its implications for gene flow. *J. Evol. Biol.* **25**: 1587–1599.

Harrell, F.E.J. 2015. rms: Regression Modeling Strategies. R package version 4.3-1. URL http://CRAN.R-project.org/packages=rms

Hatfield, T. 1997. Genetic divergence in adaptive characters between sympatric species of stickleback. *Am. Nat.* **149**: 1009–1029.

Hatfield, T. & Schluter, D. 1996. A test for sexual selection on hybrids of two sympatric sticklebacks. *Evolution* **50**: 2429–2434.

Hendry, A.P. & Taylor, E.B. 2004. How much of the variation in adaptive divergence can be explained by gene flow? An evaluation using lake-stream stickleback pairs. *Evolution* **58**: 2319–2331.

Hendry, A.P., Taylor, E.B. & McPhail, J.D. 2002. Adaptive divergence and the balance between selection and gene flow: lake and stream stickleback in the Misty system. *Evolution* **56**: 1199–1216.

Hendry, A., Bolnick, D., Berner, D. & Peichel, C. 2009. Along the speciation continuum in sticklebacks. *J. Fish Biol.* **75**: 2000–2036.

Hendry, A., Hudson, K., Walker, J., Räsänen, K. & Chapman, L. 2011. Genetic divergence in morphology-performance mapping between Misty Lake and inlet stickleback. *J. Evol. Biol.* **24**: 23–35.

Hendry, A.P., Kaeuffer, R., Crispo, E., Peichel, C.L. & Bolnick, D.I. 2013a. Evolutionary inferences from the analysis of exchangeability. *Evolution* **67**: 3429–3441.

Hendry, A.P., Hendry, A.S. & Hendry, C.A. 2013b. Hendry Vineyard stickleback: testing for contemporary lake – stream divergence. *Evol. Ecol. Res.* **15**: 343–359.

Herczeg, G., Gonda, A. & Merilä, J. 2009. Evolution of gigantism in nine-spined sticklebacks. *Evolution* **63**: 3190–3200.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.

Jombart, T. & Ahmed, I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**: 3070–3071.

Jombart, T., Devillard, S. & Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**: 94.

Kaeuffer, R., Peichel, C., Bolnick, D. & Hendry, A. 2011. Data from: Convergence and non-convergence in ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. Dryad Data Repository http://dx.doi.org/10.5061/dryad.k987h.

Kaeuffer, R., Peichel, C.L., Bolnick, D.I. & Hendry, A.P. 2012. Parallel and nonparallel aspects of ecological phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution* **66**: 402–418.

Landry, L. & Bernatchez, L. 2010. Role of epibenthic resource opportunities in the parallel evolution of lake whitefish species pairs (*Coregonus sp.*). *J. Evol. Biol.* **23**: 2602–2613.

Langerhans, R.B. 2009. Trade-off between steady and unsteady swimming underlies predator-driven divergence in *Gambusia affinis*. *J. Evol. Biol.* **22**: 1057–1075.

Langerhans, R.B. & DeWitt, T.J. 2004. Shared and unique features of evolutionary diversification. *Am. Nat.* **164**: 335–349.

Langerhans, R.B., Layman, C.A., Shokrollahi, A.M. & DeWitt, T.J. 2004. Predator-driven phenotypic diversification in *Gambusia affinis*. *Evolution* **58**: 2305–2318.

Langerhans, R.B., Knouft, J.H. & Losos, J.B. 2006. Shared and unique features of diversification in greater Antillean Anolis ecomorphs. *Evolution* **60**: 362–369.

Lavin, P.A. & McPhail, J.D. 1987. Morphological divergence and the organization of trophic characters among lacustrine populations of the threespine stickleback (*Gasterosteus aculeatus*). *Can. J. Fish Aquat. Sci.* **44**: 1820–1829.

Lavin, P.A. & McPhail, J.D. 1993. Parapatric lake and stream sticklebacks on northern Vancouver Island: disjunct distribution or parallel evolution? *Can. J. Zool.* **71**: 11–17.

Leinonen, T., Cano, J.M. & Merilä, J. 2011. Genetics of body shape and armour variation in threespine sticklebacks. *J. Evol. Biol.* **24**: 206–218.

Length, R.V. 2015. lsmeans: Least-squares means. R package version 2.17. URL http://CRAN.R-project.org/packages=lsmeans

Lleonart, J., Salat, J. & Torres, G.J. 2000. Removing allometric effects of body size in morphological analysis. *J. Theor. Biol.* **205**: 85–93.

Losos, J.B. 2011. Convergence, adaptation, and constraint. *Evolution* **65**: 1827–1840.

Lucek, K., Sivasundar, A., Roy, D. & Seehausen, O. 2013. Repeated and predictable patterns of ecotypic differentiation during a biological invasion: lake-stream divergence in parapatric Swiss stickleback. *J. Evol. Biol.* **26**: 2691–2709.

Lucek, K., Sivasundar, A., Kristjánsson, B.K., Skúlason, S. & Seehausen, O. 2014a. Quick divergence but slow convergence during ecotype formation in lake and stream stickleback pairs of variable age. *J. Evol. Biol.* **27**: 1878–1892.

Lucek, K., Sivasundar, A. & Seehausen, O. 2014b. Disentangling the role of phenotypic plasticity and genetic divergence in contemporary ecotype formation during a biological invasion. *Evolution* **68**: 2619–2632.

Matos, M., Avelar, T. & Rose, M.R. 2002. Variation in the rate of convergent evolution: adaptation to a laboratory environment in *Drosophila subobscura*. *J. Evol. Biol.* **15**: 673–682.

Mazzarella, A.B., Voje, K.L., Hansson, T.H., Taugbøl, A. & Fischer, B. 2015. Strong and parallel salinity-induced phenotypic plasticity in one generation of threespine stickleback. *J. Evol. Biol.* **28**: 667–677.

McGuigan, K., Franklin, C.E., Moritz, C. & Blows, M.W. 2003. Adaptation of rainbow fish to lake and stream habitats. *Evolution* **57**: 104–118.

McKinnon, J.S. & Rundle, H.D. 2002. Speciation in nature: the threespine stickleback model systems. *Trends Ecol. Evol.* **17**: 480–488.

McPhail, J.D. 1992. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): evidence for a species-pair in Paxton Lake, Texada Island, British Columbia. *Can. J. Zool.* **70**: 361–369.

Moczek, A.P., Sultan, S., Foster, S., Ledón-Rettig, C., Dworkin, I., Nijhout, H.F. *et al.* 2011. The role of developmental plasticity in evolutionary innovation. *Proc. Biol. Sci.* **278**: 2705–2713.

Moodie, G.E.E. 1972a. Morphology, life history, and ecology of an unusual stickleback (*Gasterosteus aculeatus*) in the Queen Charlotte Islands, Canada. *Can. J. Zool.* **50**: 721–732.

Moodie, G.E.E. 1972b. Predation, natural selection and adaptation in an unusual threespine stickleback. *Heredity* **28**: 155–167.

Moore, J.-S., Gow, J.L., Taylor, E.B. & Hendry, A.P. 2007. Quantifying the constraining influence of gene flow on adaptive divergence in the lake-stream threespine stickleback system. *Evolution* **61**: 2015–2026.

Moser, D., Kueng, B. & Berner, D. 2015. Lake-stream divergence in stickleback life history: a plastic response to trophic niche differentiation? *Evol. Biol.* **42**: 328–338.

Nosil, P. & Crespi, B.J. 2004. Does gene flow constrain adaptive divergence or vice versa? A test using ecomorphology and sexual isolation in *Timema cristinae* walking-sticks. *Evolution* **58**: 102–112.

O'Steen, S., Cullum, A.J. & Bennett, A.F. 2002. Rapid evolution of escape ability in Trinidadian guppies (*Poecilia reticulata*). *Evolution* **56**: 776–784.

Paenke, I., Sendhoff, B. & Kawecki, T.J. 2007. Influence of plasticity and learning on evolution under directional selection. *Am. Nat.* **170**: E47–E58.

Peichel, C.L., Nereng, K.S., Ohgi, K.A., Cole, B.L.E., Colosimo, P.F., Buerkle, C.A. *et al.* 2001. The genetic architecture of divergence between threespine stickleback species. *Nature* **414**: 901–905.

Pfennig, D.W. & McGee, M. 2010. Resource polyphenism increases species richness: a test of the hypothesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**: 577–591.

Pfennig, D.W., Wund, M.A., Snell-Rood, E.C., Cruickshank, T., Schlichting, C.D. & Moczek, A.P. 2010. Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol. Evol.* **25**: 459–467.

Price, T.D., Qvarnstrom, A. & Irwin, D.E. 2003. The role of phenotypic plasticity in driving genetic evolution. *Proc. R. Soc. B-Biol. Sci.* **270**: 1433–1440.

Prunier, R., Holsinger, K.E. & Carlson, J.E. 2012. The effect of historical legacy on adaptation: do closely related species respond to the environment in the same way? *J. Evol. Biol.* **25**: 1636–1649.

R Development Core Team 2012. *R: A Language and Environment of Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raeymaekers, J.A.M., Delaire, L. & Hendry, A.P. 2009. Genetically based differences in nest characteristics between lake, inlet, and hybrid threespine stickleback from the misty system, British Columbia, Canada. *Evol. Ecol. Res.* **11**: 905–919.

Raeymaekers, J.A.M., Boisjoly, M., Delaire, L., Berner, D., Räsänen, K. & Hendry, A.P. 2010. Testing for mating isolation between ecotypes: laboratory experiments with lake, stream and hybrid stickleback. *J. Evol. Biol.* **23**: 2694–2708.

Ravinet, M., Prodöhl, P.A. & Harrod, C. 2013. Parallel and nonparallel ecological, morphological and genetic divergence in lake-stream stickleback from a single catchment. *J. Evol. Biol.* **26**: 186–204.

Reimchen, T.E., Stinson, E.M. & Nelson, J.S. 1985. Multivariate differentiation of parapatric and allopatric populations of threespine stickleback in the Sangan River watershed, Queen Charlotte Islands. *Can. J. Zool.* **63**: 2944–2951.

Reist, J.D. 1985. An empirical evaluation of several univariate methods that adjust for size variation in morphometric data. *Can. J. Zool.* **63**: 1429–1439.

Revelle, W. 2015. psych: Procedures for Personality and Psychological Research. R package version 1.5.8. URL http://CRAN.R-project.org/package=psych

Robinson, B.W. & Parsons, K.J. 2002. Changing times, spaces, and faces: tests and implications of adaptive morphological plasticity in the fishes of northern postglacial lakes. *Can. J. Fish Aquat. Sci.* **59**: 1819–1833.

Roesti, M., Hendry, A.P., Salzburger, W. & Berner, D. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Mol. Ecol.* **21**: 2852–2862.

Rogers, S.M., Tamkee, P., Summers, B., Balabahadra, S., Marks, M., Kingsley, D.M. *et al.* 2012. Genetic signature of adaptive peak shift in threespine stickleback. *Evolution* **66**: 2439–2450.

Scheiner, S.M. 1993. Genetic and evolution of phenotypic plasticity. *Annu. Rev. Ecol. Syst.* **24**: 35–68.

Schlichting, C.D. & Wund, M.A. 2014. Phenotypic plasticity and epigenetic marking: an assessment of evidence for genetic accommodation. *Evolution* **68**: 656–672.

Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. *Evolution* **50**: 1766–1774.

Schluter, D., Clifford, E.A., Nemethy, M. & McKinnon, J.S. 2004. Parallel evolution and inheritance of quantitative traits. *Am. Nat.* **163**: 809–822.

Sharpe, D.M.T., Räsänen, K., Berner, D. & Hendry, A.P. 2008. Genetic and environmental contributions to the morphology of lake and stream stickleback: implications for gene flow and reproductive isolation. *Evol. Ecol. Res.* **10**: 849–866.

Sultan, S.E. & Spencer, H.G. 2002. Metapopulation structure favors plasticity over local adaptation. *Am. Nat.* **160**: 271–283.

Svanbäck, R. & Schluter, D. 2012. Niche specialization influences adaptive phenotypic plasticity in the threespine stickleback. *Am. Nat.* **180**: 50–59.

Teotónio, H. & Rose, M.R. 2000. Variation in the reversibility of evolution. *Nature* **408**: 463–466.

Thibert-Plante, X. & Hendry, A.P. 2011. The consequences of phenotypic plasticity for ecological speciation. *J. Evol. Biol.* **24**: 326–342.

Thompson, C.E., Taylor, E.B. & Mcphail, J.D. 1997. Parallel evolution of lake-stream pairs of threespine sticklebacks (*Gasterosteus*) inferred from mitochondrial DNA variation. *Evolution* **51**: 1955–1965.

Torres-Dowdall, J., Handelsman, C.A., Reznick, D.N. & Ghalambor, C.K. 2012. Local adaptation and the evolution of phenotypic plasticity in Trinidadian guppies (*Poecilia reticulata*). *Evolution* **66**: 3432–3443.

Venables, W.N. & Ripley, B.D. 2002. *Modern Applied Statistics with S, Fourth*. Springer, New York.

Wake, D.B., Wake, M.H. & Specht, C.D. 2011. Homoplasy: from detecting pattern to determining process and mechanism of evolution. *Science* **331**: 1032–1035.

Walker, J.A. 1997. Ecological morphology of lacustrine threespine (Gasterosteidae) body shape. *Biol. J. Linn. Soc.* **61**: 3–50.

West-Eberhard, M.J. 2003. *Developmental Plasticity and Evolution*. Oxford University Press, Oxford, U.K.

Westram, A., Galindo, J., Rosenblad, M.A., Grahame, J., Panova, M. & Butlin, R. 2014. Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Mol. Ecol.* **23**: 4603–4616.

Wickham, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer, New York.

Wund, M.A. 2012. Assessing the impacts of phenotypic plasticity on evolution. *Integr. Comp. Biol.* **52**: 5–15.

Wund, M.A., Baker, J.A., Clancy, B., Golub, J.L., Fosterk, S.A. & Foster, S.A. 2008. A test of the "Flexible stem" model of evolution: ancestral plasticity, genetic accommodation, and morphological divergence in the threespine stickleback radiation. *Am. Nat.* **172**: 449–462.

Wund, M.A., Valena, S., Wood, S. & Baker, J.A. 2012. Ancestral plasticity and allometry in threespine stickleback reveal phenotypes associated with derived, freshwater ecotypes. *Biol. J. Linn. Soc.* **105**: 573–583.

Supporting information

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

Figure S1 Map of sampling locations (red circles).

Figure S2 Gill raker traits in females (a) No consistent trends were detected in female gill raker length divergence across watersheds, as in males.

Figure S3 Misclassification analyses of common garden (CG) and wild-caught (WC) individuals show support for both parallelism and non-parallelism.

Table S1 Location (latitude and longitude) of sampling sites on Vancouver Island, British Columbia, Canada, and minimum distance from the stream site to the lake measured along the stream.

Table S2 Sample sizes of common garden and wild-caught individuals used in analyses from each system.

Table S3 Mean ecological data for wild-caught (WC) fish from Kaeuffer *et al.* (2011), including isotope data, proportion of limnetic prey, and trophic position.

Table S4 Results table for analyses on common garden (CG) and wild-caught (WC) individuals together (question 4).

Data deposited at Dryad: doi: 10.5061/dryad.1mr07

Received 24 June 2015; revised 18 September 2015; accepted 21 September 2015