Ecology and Evolution Open Access

Signatures of selection in the three-spined stickleback along a small scale brackish water - freshwater transition zone

Journal:	*Ecology and Evolution*
Manuscript ID:	ECE-2015-03-00147.R1
Wiley - Manuscript type:	Original Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Konijnendijk, Nellie; KU Leuven, Laboratory of Aquatic Ecology Shikano, Takahito; University of Helsinki, Department of Biosciences Daneels, Dorien; KU Leuven, Laboratory of Aquatic Ecology Volckaert, Filip; KU Leuven, Laboratory of Aquatic Ecology Raeymaekers, Joost; KU Leuven, Laboratory of Aquatic Ecology
Search Terms:	Evolutionary Ecology, Molecular Ecology, Genomics
Abstract:	Local adaptation is often obvious when gene flow is impeded, such as observed at large spatial scales and across strong ecological contrasts. However, it becomes less certain at small scales such as between adjacent populations or across weak ecological contrasts, when gene flow is strong. While studies on genomic adaptation tend to focus on the former, less is known about the genomic targets of natural selection in the latter situation. In this study we investigate genomic adaptation in populations of the three-spined stickleback Gasterosteus aculeatus L. across a small-scale ecological transition with salinities ranging from brackish to fresh. Adaptation to salinity has been repeatedly demonstrated in this species. A genome scan based on 87 microsatellite markers revealed only few signatures of selection, likely owing to the constraints that homogenizing gene flow puts on adaptive divergence. However, the detected loci appear repeatedly as targets of selection in similar studies of genomic adaptation in three-spined stickleback. We conclude that the signature of genomic selection in the face of strong gene flow is weak, yet detectable. We argue that the range of studies of genomic divergence should be extended to include more systems characterised by limited geographical and ecological isolation, which is often a realistic setting in nature.

Signatures of selection in the three-spined stickleback along a small scale brackish water - freshwater transition zone

Nellie Konijnendijk[1], Takahito Shikano[2], Dorien Daneels[1,*], Filip A.M. Volckaert[1], Joost A. M. Raeymaekers[1]

[1] Laboratory of Biodiversity and Evolutionary Genomics, University of Leuven, Ch. Deberiotstraat, 32, B-3000 Leuven, Belgium
[2] Ecological Genetics Research Unit, Department of Biosciences, University of Helsinki, P.O. Box 65, FI-000 14, Helsinki, Finland

[*] Current address: Centre for Medical Genetics, Reproduction and Genetics; Reproduction, Genetics and Regenerative Medicine, Vrije Universiteit Brussel (VUB), Universitair ziekenhuis Brussel (UZ Brussel), Laarbeeklaan 101, 1090 Brussels, Belgium.

Running header: Adaptation in three-spined stickleback

Address for correspondence:
Joost Raeymaekers, University of Leuven
Laboratory of Biodiversity and Evolutionary Genomics
Ch. Deberiotstraat, 32
B-3000 Leuven, Belgium
e-mail: joost.raeymaekers@bio.kuleuven.be
Phone : + 32 16 32 39 66, Fax : +32 16 32 45 75

ABSTRACT

Local adaptation is often obvious when gene flow is impeded, such as observed at large spatial scales and across strong ecological contrasts. However, it becomes less certain at small scales such as between adjacent populations or across weak ecological contrasts, when gene flow is strong. While studies on genomic adaptation tend to focus on the former, less is known about the genomic targets of natural selection in the latter situation. In this study we investigate genomic adaptation in populations of the three-spined stickleback *Gasterosteus aculeatus* L. across a small-scale ecological transition with salinities ranging from brackish to fresh. Adaptation to salinity has been repeatedly demonstrated in this species. A genome scan based on 87 microsatellite markers revealed only few signatures of selection, likely owing to the constraints that homogenizing gene flow puts on adaptive divergence. However, the detected loci appear repeatedly as targets of selection in similar studies of genomic adaptation in three-spined stickleback. We conclude that the signature of genomic selection in the face of strong gene flow is weak, yet detectable. We argue that the range of studies of genomic divergence should be extended to include more systems characterised by limited geographical and ecological isolation, which is often a realistic setting in nature.

Keywords: adaptation, candidate gene, evolution, fish, genome scan, population genomics, selection

INTRODUCTION

Recent questions on the mechanisms of evolutionary biology revolve around the genomic architecture of species and the effects that processes such as selection, drift, mutation and gene flow have on the genome. Changes in the environment can now be linked to the genetic signatures of these processes and hence promote the understanding of the genetic basis of ecological adaptation. This allows us to understand the various components of the mechanism of adaptation such as which genes are involved, how they are distributed in the genome, and how often the same genes lead to certain adaptations both in the laboratory (Becks *et al.* 2012; see Barrick & Lenski 2013 for other examples) and nature (Mitchell-Olds & Schmitt 2006; Yang *et al.* 2011; Jones *et al.* 2012a; Orsini *et al.* 2012). Additionally we want to understand which processes are essential for adaptation to succeed. One way to find genes that are important for the process of adaptation is to identify genetic signatures of adaptation in natural populations. These "genome scans" identify genes that are under selection across contrasting environments and might thus be involved in the transition from one ecological extreme to the other. When interpreted with care (de Villemereuil *et al.* 2014), this approach has shown to be a promising way to find links between genotype, phenotype and fitness in natural populations (Storz 2005).

Genome scans have so far generated two major insights. First, selection may act on many parts of the genome (Nosil 2009; Hohenlohe *et al.* 2010; Feder *et al.* 2012; Jones *et al.* 2012a; Jones *et al.* 2012b; Strasburg *et al.* 2012; Arnegard *et al.* 2014; Seehausen *et al.* 2014). Nosil (2009) estimated that 5-10 % of the genome is affected by natural selection. Based on a genome-wide analysis of selection, Hohenlohe *et al.* (2012) showed that the genome is much more structured and dynamic than expected from theory. More studies have now contributed to these topics and show that the areas affected by natural selection are typically patchily distributed ("hotspots") across the genome (Voight *et al.* 2006; Papa *et al.* 2008; Hohenlohe *et al.* 2010). Genomic inversions, deletions, repetitive elements and other structural changes facilitate the rise of these islands of divergence, causing a physical barrier to crossing over and thus adding to linkage disequilibrium (Hohenlohe *et al.* 2010). On the other hand, most cases of adaptation in plants involve many genomic areas with a wide distribution across the genome rather than clustered hotspots (Strasburg *et al.* 2012).

A second common finding of genome scans is that organisms may show parallel as well as non-parallel genetic responses to environmental change. For example, Atlantic salmon (*Salmo salar*) populations that undergo a parallel change in environment show non-parallel adaptive divergence at the genomic level (Perrier *et al.* 2013). In cases of parallel phenotypic evolution in the three-spined stickleback (*Gasterosteus aculeatus* L.), common (Hohenlohe *et al.* 2010; Shimada *et al.* 2011; Jones *et al.* 2012b) as well as unique (Mäkinen *et al.* 2008; Jones *et al.* 2012b; Roesti *et al.* 2012) genomic regions are targeted by selection. This has been confirmed in experiments with stick insects (*Timema cristinae*) (Soria-Carrasco *et al.* 2014) and in genome scans of lake white fish (*Coregonus* sp.) (Bernatchez *et al.* 2010). In contrast, parallel phenotypic evolution in other species, such as the nine-spined stickleback (*Pungitius pungitius*) (Shikano *et al.* 2010a) or the rough periwinkle *Littorina saxatilis* (Butlin *et al.* 2014), were mainly characterized by non-parallel genomic signatures of selection.

Adaptive divergence is thought to proceed as a balance between divergent selection and homogenizing gene flow (Levene 1953; Hagen 1967; Endler 1973; Bell 1982) and hence may reach various stages. This has long been recognised at the phenotypic level (Tregenza 2002; Moore *et al.* 2007; Hendry 2009; Schluter 2009). More recently, theoretical and empirical studies have improved our understanding of the genomic architecture at various stages of adaptation as well (Pinho & Hey 2010; Yeaman & Otto 2011; Yeaman & Whitlock 2011). Nevertheless, the focus of genome scans is often on systems where it is reasonable to assume that population divergence has a strong adaptive component. To do so, genome scans often target populations at large spatial scales and across strong ecological contrasts (Hohenlohe *et al.* 2010; Poncet *et al.* 2010; Stapley *et al.* 2010; Zulliger *et al.* 2013). While this has generated great insight in the genomic basis of adaptation and speciation, knowledge might be biased towards stages where gene flow may be largely impeded, and where adaptation is already largely achieved. Gene flow modifies the response to selection by modulating the distribution of the genes that underlie ecologically relevant traits (Slatkin 1987). The study of adaptation with gene flow, especially in study systems where genetic divergence is far from complete such as across a small-scale ecological transitions or between highly connected populations, is therefore crucial to understand how populations diverge to different ecological optima (Hansen *et al.* 2002; Storz 2005; Nielsen *et al.* 2009; Coscia *et al.* 2011; DeFaveri *et al.* 2013; Vandamme *et al.* 2014).

The three-spined stickleback represents an excellent model for the study of adaptive divergence, as phenotypic responses to several ecological changes are frequent and well-documented (McPhail 1994; Foster *et al.* 1998; McKinnon & Rundle 2002; Boughman 2007). A diverse number of populations can be found that occupy various stages of divergence from panmixia to complete and irreversible reproductive isolation (Hendry *et al.* 2009). This provides an excellent framework to investigate population divergence at the phenotypic and genomic level. The genome scans that have been applied to three-spined stickleback have contrasted marine-freshwater (Mäkinen *et al.* 2008; Hohenlohe *et al.* 2010; DeFaveri *et al.* 2011; Shimada *et al.* 2011; Jones *et al.* 2012a; Jones *et al.* 2012b; DeFaveri & Merilä 2013), lake–stream (Deagle *et al.* 2012; Roesti *et al.* 2012), and benthic-limnetic (Olafsdottir & Snorrason 2009; Jones *et al.* 2012a; Lucek *et al.* 2014) population pairs, as well as populations from clean versus polluted water (Lind & Grahn 2011). A common finding is that several genes or gene regions are repeatedly selected across populations and locations, although population-specific regions do appear as well (Hohenlohe *et al.* 2010; DeFaveri *et al.* 2011; Shimada *et al.* 2011; Jones *et al.* 2012b). Others have found that most regions under selection were highly specific to the location under study (Mäkinen *et al.* 2008; Deagle *et al.* 2012; Roesti *et al.* 2012).

In this study we investigate genomic adaptation in three-spined stickleback populations from the Belgian-Dutch lowlands. Populations in this area differ in various morphological traits, which are often correlated with salinity and distance to the coast (Heuts 1947; Raeymaekers *et al.* 2005; Raeymaekers *et al.* 2007; Van Dongen *et al.* 2009; Raeymaekers *et al.* 2012). At the same time, gene flow between these populations is moderate to strong due to high connectivity (Raeymaekers *et al.* 2014). This has the advantage that we can study adaptation with ongoing gene flow, which is often a realistic setting in nature. All sites are part of an interconnected landscape of canals and streams along an ecological transition (i.e. discrete habitats with salinities ranging between brackish and freshwater) at a much smaller

geographical scale than other genome scan studies across salinity gradients in three-spined stickleback (Table 1).

In order to explore the adaptive changes across this transition, we screened two populations from each end of the transition for genomic signatures of selection. We hypothesized that (i) the outcome of small-scale adaptation along ecological transitions may vary at the genomic level and that (ii) adaptation at the genomic level may be influenced by gene flow. Populations were screened for 87 microsatellite markers of which 41 are linked to genes with a range of ecologically relevant functions (Shikano *et al.* 2010b; Shimada *et al.* 2011). We expect that the relatively high gene flow among these populations might constrain local adaptation, despite obvious differences in phenotype across these extremes. We determine outlier loci and compare the results to previous studies in three-spined stickleback across similar ecological contrasts.

MATERIAL AND METHODS

Study area

Three-spined stickleback (*Gasterosteus aculeatus* L.; Gasterosteidae) from the coastal lowlands (polder) of Belgium and the Netherlands (Figure 1) reside in ponds, ditches, streams, estuaries or polder creeks. They have an anadromous or landlocked life style (Heuts 1947; Wootton 1976; Raeymaekers *et al.* 2005; Raeymaekers *et al.* 2007). The polder and surrounding areas contain diked brackish and freshwater habitats of Holocene origin with varying levels of connectivity to adjacent estuaries and the open sea. Populations which live in close proximity to the sea (< 10 km) reside in brackish water of which the salinity is influenced by rainfall and water management. On a scale of less than 50 km further inland, salinity drops to freshwater levels (Raeymaekers *et al.* 2014). Lateral plate number, an important ecological trait, also decreases with distance to the coast, with population averages from 20 to 5 (Heuts 1947; Raeymaekers et al. 2014). Populations bordering the North Sea and the Baltic are typically polymorphic for lateral plate number (Heuts 1947; Raeymaekers *et al*. 2014), so higher or lower population averages are rare. This range is therefore representative for the phenotypic extremes we can find in this part of the stickleback's distribution range.

Field sampling

Field sampling was conducted in spring 2009 in parallel with a multi-year study by Raeymaekers *et al*. (2014), describing the distribution of lateral plate number in the study area. Two brackish water creeks (L01 and L02) and two freshwater ditches (L12 and U01), representing the sites with the most extreme values for salinity, were selected (Table 2). Thirty fish from each population were collected with dip nets, immediately anaesthetized with MS222 (2 $g.l^{-1}$) and frozen on dry ice. In the lab, sticklebacks were thawed on ice, measured (standard length (SL); $\pm$ 0.1 cm), weighed ($\pm$ 0.01 g), photographed and fin clipped for a DNA sample.

Morphological measurements

In line with previous studies on phenotypic divergence between the stickleback populations of the Belgian-Dutch lowlands (Heuts 1947; Raeymaekers *et al.* 2005; Raeymaekers *et al.*

2007; Van Dongen *et al.* 2009; Raeymaekers *et al.* 2012) we investigated variation at six morphological traits. The left side of each fish was photographed from a standard angle, and a ruler was placed in each photograph for scaling. Dorsal spine length, pelvic spine length and pelvic plate length were measured from pictures using the software tpsDig 1.37 (Rohlf 2002). A subsample of the fish was rinsed with water for 72 h, bleached for 4 h (1% KOH bleach solution) and stained with alizarin red (Taylor & Dyke 1985). Stained fish were used to determine the number of lateral plates on the right side of the fish. The right part of the gills was then dissected and the number and length of the large gill rakers were quantified under a dissection scope.

Marker selection

A set of 110 microsatellite markers was selected, including a range of putatively neutral markers, to set a proper neutral F_{ST} background. In our study area there is a salinity cline, but various other factors may co-vary with this cline. Therefore we included 41 markers that are known to be linked to functional genes in a range of ecologically relevant functions such as salinity, growth and immunity (Shikano *et al.* 2010b; Shimada *et al.* 2011). Several of these genes have been found to be under selection in the three-spined stickleback or nine-spined stickleback of other salinity transitions (Hohenlohe *et al.* 2010; Shikano *et al.* 2010a; DeFaveri *et al.* 2011; Shimada *et al.* 2011; Jones *et al.* 2012a; Jones *et al.* 2012b). This allows us to compare these systems to some extent with ours and pinpoint parallel changes across systems. The marker closely linked to the *Eda* gene (*Stn380*), a major effect gene underlying variation in plate number (Colosimo *et al.*, 2005), was included as a reference gene that is often under selection in freshwater-salt water comparisons (Raeymaekers *et al.* 2007; Mäkinen *et al.* 2008; DeFaveri *et al.* 2011; DeFaveri & Merilä 2013).

DNA extraction and genotyping

Genomic DNA was extracted from fin clips using a proteinase K digestion step and the Nucleospin 96 Tissue DNA Extraction kit (Macherey-Nagel). Individuals were genotyped at 110 microsatellite loci that were arranged in 21 multiplexes of 4 to 8 markers at a time, with EST based markers and markers within or near genes with relevant functions included (Table S1). Loci were amplified with the Qiagen® Multiplex PCR Kit (Qiagen, Venlo, the Netherlands). The 10 µl PCR cocktail contained 1-100 ng genomic DNA, 2 pmol each of forward and reverse primers, 1x Qiagen Multiplex PCR master mix, 0.5x Q-solution, and RNase-free water. The reaction consisted of an initial activation step of 15 min at 95 °C, followed by 30 cycles of 30 s at 94 °C, 90 s at 53 °C and 60 s at 72 °C. A final elongation step of 5 min at 60 °C was performed. Allele sizes were determined by means of an internal ET ROX 550 size standard (Amersham Biosciences). Polymerase chain products were visualized using a MegaBace 1000 automated sequencer (Amersham Biosciences). Alleles were scored with the Fragment Profiler v1.2 software (Amersham Biosciences), using visual scoring and manual corrections. Marker Stn380, linked to the *Eda* gene, was scored separately to determine the frequency of the "low plated" allele in each population (Table 2).

Data analysis

Phenotypic differentiation

For each trait, analysis of variance (ANOVA) was performed to test for statistical differences between the four populations. For traits which depend on size, standard length was included as a covariate.

Genetic diversity and genetic differentiation

Genotypes were checked for scoring errors attributable to stutter-products, large allele dropout or null-alleles, using MICRO-CHECKER v2.3 (van Oosterhout *et al.* 2004). Estimates of allelic richness, genetic diversity (H_e, H_o), global and pairwise F_{ST}, with a significance calculated with 1000 bootstraps over loci, were calculated using the GENETIX v4.05.02 software (Belkhir *et al.* 1996).

Genomic signatures of selection

We conducted global outlier tests to find outliers across all populations and used pairwise comparisons to check whether the outliers found by global tests were due to habitat differences. Loci that are under directional selection are expected to have lower intrapopulation variability and larger interpopulation variability than neutral loci. Loci under directional selection can thus be traced by patterns in heterozygosity, differences in F_{ST} values or a combination of the two. Evaluation of several outlier detection methods has shown that these methods differ in number of false positives and false negatives (Narum & Hess 2011). Four methods were therefore compared: LOSITAN (Antao *et al.* 2008), the outlier detection method implemented in ARLEQUIN v3.5.2.3 using hierarchical clustering (Excoffier & Lischer 2010), BAYESCAN v2.01 (Foll & Gaggiotti 2008), and LnRH (Kauer *et al.* 2003). The first three methods were used to determine global outliers. Additionally we did pairwise comparisons of all populations using LOSITAN and LnRH to detect specific signatures of selection in freshwater-brackish water, freshwater-freshwater and brackish water-brackish water populations pairs.

The four methods are based on different underlying assumptions. LOSITAN is based on an island model that uses a coalescent F_{ST}-outlier method based on the distribution of F_{ST} as a function of the heterozygosity. We used the function that first establishes a neutral F_{ST} baseline by removing putative markers under selection outside the 95% interval with 10^5 simulations. The infinite allele model was used with a 95% and 99% confidence interval. We ran 10^5 simulations as recommended by Antao (2008). The outlier detection software implemented in ARLEQUIN uses the same island model, but adds on the option for hierarchical clustering. In the presence of strong hierarchical population structure it reduces false positives by a hierarchical analysis of genetic differentiation (Excoffier *et al.* 2009). We clustered populations according to two scenarios: in two groups (brackish water populations versus freshwater populations), and three groups (brackish water populations versus each freshwater population separately). The first scenario simulates a common descent of the two freshwater populations from the marine population, while the second one simulates a separate split of the two freshwater populations from the marine population. We used the standard settings of 20,000 simulations for each run and 100 demes per group. The method executed by BAYESCAN uses a logistic regression model which explains the observed pattern of diversity by dividing it in a locus- and a population-specific component (Beaumont and Balding 2004). One benefit of this method is that it allows for different migration rates and

different effect sizes and thus can be used for scenarios that deviate from the island model. We conducted 10 pilot runs of 5,000 iterations, followed by an additional 150,000 iterations and a burn-in of 50,000 iterations. Outliers were appointed based on 90%, 95% and 99% posterior probabilities. Finally, the LnRH method is designed especially for microsatellite markers and determines the reduction of heterozygosity. This method is based on the assumption that microsatellites linked to a gene under selection will show reduced levels of diversity between two populations. After standardization of the LnRH estimates with a mean of zero and a standard deviation of one, we determined outliers at 95% and 99% level. We subtracted outliers that were found in pairwise comparisons within the same habitat type, to limit the number of false positives.

All tests based on simulations were executed three times to test for robustness of the results. Only outliers that were detected at least twice with each method were scored as a putative outlier. We detected outlier loci that are under balancing and directional selection. However, interpreting loci under balancing selection is difficult, as there are still limitations for the identification of loci under balancing selection (Hansen *et al.* 2010; Narum & Hess 2011). Therefore we only discuss the loci under directional or positive selection. Loci under balancing selection are provided in Supplementary Table S2.

RESULTS

Phenotypic differentiation

Statistical analysis of variance (ANOVA) confirmed the presence of significant differences between the four populations for standard length, lateral plate number, pelvic plate and spine length, and large gill raker length (Figure 2). In particular, the brackish water populations (L01 and L02) were larger in size than the freshwater populations (L12 and U01; $F_{3,155} = 75.6$; $P < 0.0001$), and had more lateral plates ($F_{3,147} = 10.4$; $P < 0.0001$), longer pelvic spines (especially population L01; $F_{3,93} = 5.51$; $P = 0.0016$), longer pelvic plates ($F_{3,93} = 8.60$; $P < 0.0001$), and larger gill rakers ($F_{3,93} = 19.8$; $P < 0.0001$).

Genetic diversity and genetic differentiation

After genotyping of the individuals with 110 microsatellite markers, we selected 87 markers with good amplification quality to perform the data analysis. A total of 1116 alleles were observed in four populations of 26 to 28 individuals each, with an average of 13 alleles per locus and a range from 2 to 34. Observed heterozygosity ranged from 0.022 to 0.95 across loci. Expected heterozygosity ranged from 0.05 to 0.92 across loci and from 0.57 to 0.69 across populations (Table 2). Allelic richness among populations varied from 6.51 to 9.47, with the freshwater populations being less diverse than the brackish water populations. Eight loci were possibly affected by null alleles or stutters. We therefore performed all final analyses with and without these loci and specify when results differ. Differentiation among populations was moderate with the global F_{ST} value being 0.059. Pairwise F_{ST} values were significant, except between the two brackish water populations (Table 3).

Genomic signatures of selection

Six of the 87 loci were assigned as outliers in at least one of the methods, with the different methods identifying 4, 2, 2, and 4 outliers for Lositan, Bayescan, lnRH and Arlequin respectively (Table 4). The methodologically similar tests performed by Lositan and Arlequin with hierarchical clustering resulted in the same set of outliers. In contrast, only two outliers were shared across methodologically different outlier detection methods; our conservative measure of outliers hence amounts to 2.3 % of the total number of loci. This value is low in comparison with other genome scan studies of three-spined stickleback (Table 1).

The two outliers for directional selection that appear consistently across methodologically different tests are marker *Stn46* (identified by Arlequin/Lositan and Bayescan) and *Ppgm44* (identified by lnRH and Bayescan). *Stn46* has been previously associated with the Rho guanine nucleotide exchange factor 9 (Table S1). It was also assigned as an outlier in one pairwise freshwater versus brackish water comparison. *Ppgm44* is a marker that is linked to the gene *myostatin2*, which is associated with growth (Table S1). It was also assigned as an outlier in all four pairwise freshwater versus brackish water comparisons. Among the four other outliers that were detected (Table 4), one is associated with osmoregulation, namely an alpha subunit of the Na+/K+ ATPase (ATP1A1). Other outliers are either linked to functions such as thermal response (HSPA14) or were assumed to be neutral. Locus *Stn34* was also assigned as an outlier, but since null alleles were detected at this locus, it was excluded from Table 4. The *Eda* gene, with frequencies of the low-plated allele ranging between 0.44-0.51 in the brackish populations and 0.74-0.93 in the freshwater populations (Table 2), did not show up as an outlier in the analyses (Table 4).

DISCUSSION

We found a limited number of loci showing a signature of selection among three-spined stickleback populations inhabiting the coastal Dutch-Belgian lowlands. Six outliers were found when all outlier detection methods were considered and only two of those were shared among methodologically different tests, despite differentiation among populations in ecology and phenotype. Not only the number but also the proportion of outliers was low as compared to other studies (Table 1). The function of one of the outlier loci could be directly related to salinity, the most obvious ecological gradient in our study area. We here discuss possible explanations for the limited amount of outliers, the putative function of the genes linked to outliers, and the consequences for adaptive divergence in the face of gene flow.

The presence of a limited number and proportion of outliers might be attributed to a number of technical aspects. First, we used relatively few markers. However, among the 87 markers selected, 41 markers are linked to ecologically relevant functions, of which several have been shown to be under selection in systems with similar salinity clines (Mäkinen *et al.* 2008; Shimada *et al.* 2011; DeFaveri *et al.* 2013). Second, analysing relatively few populations may have two drawbacks: 1) it might increase the risk of false negatives due to the limited number of individuals sampled, and 2) it might increase the difficulty in separating the signal of selection from the geographical and historical signal (Bierne *et al.* 2011; Bierne *et al.* 2013). The small overlap in outlier loci among detection methods might be a symptom of this. However, in a larger study across 14 populations (2320 individuals) describing the distribution of lateral plate number and the underlying *Eda* gene in our study area (Raeymaekers *et al.* 2014), the signature of selection at the *Eda* gene was neither

significant. This bi-allelic gene is experiencing selection across various other salinity transitions in three-spined stickleback (Raeymaekers et al. 2007; Mäkinen et al. 2008; DeFaveri et al. 2011; DeFaveri & Merilä 2013). This suggests that increasing sample size does not necessarily enhance the detectability of selection. We therefore do not expect that technical issues are a major explanation for our findings.

An alternative explanation for the low proportion of outlier loci is that selection might be weak, owing to weaker environmental contrasts in our study area (i.e., brackish to freshwater) compared to other studies (Table 1). For instance, Raeymaekers *et al.* (2014) found that shifts in *Eda* allele frequencies from one generation to the next were associated with salinity, but that at the landscape level salinity did not correlate with *Eda* allele frequencies. This suggests that selection is still acting, but might be too weak to contribute to local adaptation and leave a signature of selection at this gene. In addition, strong gene flow might confound the effect of selection by mixing adapted and non-adapted alleles in the respective populations. Gene flow being moderate to high, we expect this to be another explanation for why we find so few consistent outlier loci compared to other studies. Accordingly, Raeymaekers *et al.* (2014) found that the spatial distribution of the *Eda* allele frequency correlated with distance to the coast, a proxy for population connectivity. Recurrent contact between freshwater and estuarine or marine populations might lead to the exchange of maladaptive alleles, but at the same time it has been argued that gene flow might cause an opposite effect: efficient flow of advantageous loci (Schluter & Conte 2009; Hohenlohe *et al.* 2012; Bell & Aguirre 2013; Raeymaekers *et al.* 2014), thus facilitating adaptation from standing variation. Whether gene flow in this case fuels or constrains adaptation is hard to say, but the signal of selection might be more difficult to pick up in genome scans due to mixing of the genomic background. We therefore expect that the low number of outlier loci is rather explained by either weak selection, confounding effects of gene flow or a combination of both and are thus inherent to the system we study.

Genes that have been found in previous studies to be under selection relate to biological functions such as bone formation, osmoregulation, growth, thermal response, maturation, pigmentation, scent detection, spiggin production and morphology (Hohenlohe *et al.* 2010; DeFaveri *et al.* 2011; Shimada *et al.* 2011; Jones *et al.* 2012b). Genes that appeared as outliers in this study include markers linked to osmoregulation (*ATP1A1*), thermal response (*HSPA14*) and growth (*myostatin2*). The marker linked to *ATP1A1* has been identified as outlier gene in multiple stickleback saltwater-freshwater transitions worldwide (Jones *et al.* 2006; McCairns & Bernatchez 2010; DeFaveri *et al.* 2011; Shimada *et al.* 2011; Jones *et al.* 2012a; Jones *et al.* 2012b). A major outlier in our study was marker *Stn46*, which is linked to the gene for Rho guanine nucleotide exchange factor 9. It is a member of the gene family coding for rho proteins, a subfamily of the guanine nucleotide exchange factors (GEFs), which are multi-domain proteins involved in the activation of small GTPases (Rossman *et al.* 2005). The Rho family is involved in relaying signals from cell-surface receptors to the actin cytoskeleton and elsewhere (Dvorsky & Ahmadian 2004). Its function can be associated with juvenile growth; in zebrafish it has been specifically linked to angiogenesis (Garnaas *et al.* 2008) and striated muscle and neural development (Raeker *et al.* 2010). Interestingly, locus *Stn46* was also under selection in populations of the nine-spined stickleback (Shikano *et al.* 2010a), indicating that the gene might be involved in local adaptation in multiple species. Another major outlier in our study was the marker linked to *myostatin2*. This gene, a

member of the transforming growth factor-beta (TGF-beta) family, is known to function as a negative regulator of skeletal muscle development and growth in mammals (Walsh & Celeste 2005) and teleost fish (Radaelli *et al.* 2003). Blocking the expression of myostatin in zebrafish has led to the development of a giant phenotype (Acosta *et al.* 2005), but mysostatin is produced in many other tissues than skeletal muscles and is expected to influence many more functions (Radaelli *et al.* 2003). Understanding why these specific genes are selected requires further study due to the broad range of functions these genes might have.

The low number of outlier loci that we find contrasts with the differentiation across populations in several morphological traits. Local differentiation in the number of lateral plates, for instance, has been shown to significantly exceed the level of neutral differentiation in our study area (Raeymaekers *et al.* 2014). Plate number and other morphological traits such as spine length and gill raker length have reasonably high heritability values exceeding 40 % (Schluter 1996; Peichel *et al.* 2001; Berner *et al.* 2014), suggesting that phenotypes are largely determined by genetic rather than by plastic effects. Yet, theory predicts that only functional loci with a relatively large effect size under strong divergent selection will be able to surpass gene flow (Via 2009; Yeaman & Whitlock 2011). The discrepancy between phenotypic and genomic signatures of selection might therefore become particularly strong for traits that involve many genes of small effect. Arnegard *et al*. (2014) have another explanation for the relatively large phenotypic divergence. They show that niche differentiation in sticklebacks, even in early stages of differentiation, can involve many different genes and that gene flow between divergent niche-adapted populations has a bigger impact on the phenotype than just the traits that are directly targeted by selection. It may be attributed to incompatibilities in hybrids that harbor a mix of genes of differentially adapted genotypes. These effects imply that phenotypic changes might not always be adaptive. McCairns & Bernatchez (2010) found indications that freshwater populations might suffer from a loss of plasticity and it might be that epigenetic effects further enhance this discrepancy. For instance, Chaturvedi *et al*. (2014) found that regulation by miRNAs might make a significant contribution to freshwater adaptation in stickleback populations.

Conclusion

We find that weak selection, high levels of gene flow or a combination of both can limit the number of outliers in genome scans. Although genome scans targeting populations across strong environmental contrasts are possibly more effective for pinpointing genes that are involved in adaptation, the genes identified by these studies do not necessarily play an important role at every stage of divergence. Many of the genes involved in salt-freshwater transitions might be site-specific or might not be involved when gene flow is constantly mixing the gene pool. The genes that we do find are likely to be those with a major effect size and thus an underrepresentation of the total number of genes involved. In addition, phenotypic adaptation is not necessarily genetic, but might be facilitated by plastic and epigenetic effects. It remains a challenging task to find which genes and how many are truly involved in local adaptation. We here showed that even with ample gene flow and across weak ecological contrasts, interesting insights on the repeatability of genomic signatures of selection can be obtained.

ACKNOWLEDGEMENTS

We would like to thank A. Chaturvedi, S. Geldof, B. Hellemans, L. Jaquemin, S. Ruyts, N. Thiercelin, I. Verdonck and D. Verheyen for field support and technical assistance. We thank I. Coscia and J. Merilä for insightful and inspiring comments and the latter for collaboration on the whole project. We appreciated the insightful suggestions of two anonymous reviewers. Research was sponsored by the Research Foundation - Flanders (post-doctoral fellowship to J.A.M.R. and research project G.0142.03), by the KU Leuven (GOA/2006/06 "Evolutionary responses to strong selection pressures and their ecological consequences"; PF/2010/07 "Eco- and socio-evolutionary dynamics") and facilitated by the research community W0.037.10N "Eco-evolutionary dynamics in natural and anthropogenic communities" sponsored by the Research Foundation - Flanders. The study received additional support from Academy of Finland grant (129662, 134728 and 218343) to Juha Merilä.

REFERENCES

Acosta J, Carpio Y, Borroto I, González O, Estrada MP (2005) Myostatin gene silenced by RNAi show a zebrafish giant phenotype. *Journal of Biotechnology* **119**, 324-331.

Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* **9**, 323.

Arnegard, M. E., McGee, M. D., Matthews, B., Marchinko, K. B., Conte, G. L., Kabir, S., Bedford, N., Bergek, S., Chan, Y. F., Jones, F. C., Kingsley, D. M., Peichel, C. L. & Schluter, D. 2014. Genetics of ecological divergence during speciation. *Nature* **511**: 307-311.

Barrick JE, Lenski RE (2013) Genome dynamics during experimental evolution. *Nature Reviews: Genetics* **14**, 827-839.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Becks L, Ellner SP, Jones LE, Hairston NG (2012) The functional genomics of an eco-evolutionary feedback loop: linking gene expression, trait evolution, and community dynamics. *Ecology Letters* **15**, 492-501.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations, pp. 1996-2004. Laboratoire génome, populations, interactions, CNRS UMR.

Bell MA (1982) Differentiation of adjacent stream populations of threespine sticklebacks. *Evolution* **36**, 189-199.

Bell MA, Aguirre WE (2013) Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. *Evolutionary Ecology Research* **15**, 377-411.

Bernatchez L, Renaut S, Whiteley AR*, et al.* (2010) On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **365**, 1783-1800.

Berner D, Moser D, Roesti M, Buescher H, Salzburger W (2014) Genetic architecture of skeletal evolution in european lake and stream stickleback. *Evolution* **68**, 1792-1805.

Bierne N, Roze D, Welch JJ (2013) Pervasive selection or is it…? why are FST outliers sometimes so frequent? *Molecular Ecology* **22**, 2061-2064.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**, 2044-2072.

Boughman JW (2007) Speciation in sticklebacks. In: *Biology of the Three-Spined Stickleback* (eds. Östlund-Nilsson S, Mayer I, Huntingford F), pp. 83–126. CRC Press., Boca Raton, FL.

Butlin RK, Saura M, Charrier G*, et al.* (2014) Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution* **68**, 935-949.

Chaturvedi A, Raeymaekers JAM, Volckaert FAM (2014) Computational identification of miRNAs, their targets and functions in three-spined stickleback (*Gasterosteus aculeatus*). *Molecular Ecology Resources* **14**, 768-777.

Colosimo PF, Hosemann KE, Balabhadra S*, et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**, 1928-1933.

Coscia I, Vogiatzi E, Kotoulas G, Tsigenopoulos C, Mariani S (2011) Exploring neutral and adaptive processes in expanding populations of gilthead sea bream, Sparus aurata L., in the North-East Atlantic. *Heredity* **108**, 537-546.

de Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology* **23**, 2006-2019.

Deagle BE, Jones FC, Chan YF*, et al.* (2012) Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **279**, 1277-1286.

DeFaveri J, Jonsson PR, Merilä J (2013) Heterogeneous genomic differentiation in marine threespine sticklebacks: Adaptation along an environmental gradient. *Evolution* **67**, 2530-2546.

DeFaveri J, Merilä J (2013) Evidence for adaptive phenotypic differentiation in Baltic Sea sticklebacks. *Journal of Evolutionary Biology* **26**, 1700-1715.

DeFaveri J, Shikano T, Shimada Y, Goto A, Merilä J (2011) Global analysis of genes involved in freshwater adaptation in threespine stickleback (*Gasterosteus aculeatus*). *Evolution* **65**, 1800-1807.

Dvorsky R, Ahmadian MR (2004) Always look on the bright site of Rho: structural implications for a conserved intermolecular interface. *EMBO reports* **5**, 1130-1136.

Endler JA (1973) Gene flow and population differentiation: Studies of clines suggest that differentiation along environmental gradients may be independent of gene flow. *Science* **179**, 243-250.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285-298.

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

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**, 342-350.

Feulner PGD, Chain FJJ, Panchal M*, et al.* (2013) Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Molecular Ecology* **22**, 635-649.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993.

Foster SA, Scott RJ, Cresko WA (1998) Nested biological variation and speciation. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **353**, 207-218.

Garnaas MK, Moodie KL, Liu M-l*, et al.* (2008) Syx, a RhoA guanine exchange factor, is essential for angiogenesis in vivo. *Circulation Research* **103**, 710-716.

Hagen DW (1967) Isolating mechanisms in threespine sticklebacks (*Gasterosteus*). *Journal of the Fisheries Research Board of Canada* **24**, 1637-&.

Hansen MM, Meier K, Mensberg K-LD (2010) Identifying footprints of selection in stocked brown trout populations: a spatio-temporal approach. *Molecular Ecology* **19**, 1787-1800.

Hansen MM, Ruzzante DE, Nielsen EE, Bekkevold D, Mensberg K-LD (2002) Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (Salmo trutta) populations. *Molecular Ecology* **11**, 2523-2535.

Hendry AP (2009) Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences* **66**, 1383-1398.

Hendry AP, Bolnick DI, Berner D, Peichel CL (2009) Along the speciation continuum in sticklebacks. *Journal of Fish Biology* **75**, 2000-2036.

Heuts MJ (1947) The phenotypical variability of *Gasterosteus aculeatus* L. populations in Belgium. *Verhandelingen van de Koninklijke Vlaamse Academie voor Wetenschappen, Letteren en Schone Kunsten van België* **9**, 1-63.

Hohenlohe PA, Bassham S, Currey M, Cresko WA (2012) Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **367**, 395-408.

Hohenlohe PA, Bassham S, Etter PD*, et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**.

Jones F, Brown C, Pemberton J, Braithwaite V (2006) Reproductive isolation in a threespine stickleback hybrid zone. *Journal of Evolutionary Biology* **19**, 1531-1544.

Jones FC, Chan YF, Schmutz J*, et al.* (2012a) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology* **22**, 83-90.

Jones FC, Grabherr MG, Chan YF*, et al.* (2012b) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61.

Kauer MO, Dieringer D, Schlotterer C (2003) A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of Drosophila melanogaster. *Genetics* **165**, 1137-1148.

Levene H (1953) Genetic equilibrium when more than one ecological niche is available. *American Naturalist* **87**, 331-333.

Lind EE, Grahn M (2011) Directional genetic selection by pulp mill effluent on multiple natural populations of three-spined stickleback (*Gasterosteus aculeatus*). *Ecotoxicology* **20**, 503-512.

Lucek K, Sivasundar A, Kristjánsson BK, Skúlason S, Seehausen O (2014) Quick divergence but slow convergence during ecotype formation in lake and stream stickleback pairs of variable age. *Journal of Evolutionary Biology* **27**, 1878–1892.

Mäkinen HS, Cano JM, Merilä J (2008) Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. *Molecular Ecology* **17**, 3565-3582.

McCairns RJS, Bernatchez L (2010) Adaptive divergence between freshwater and marine sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. *Evolution* **64**, 1029-1047.

McKinnon JS, Rundle HD (2002) Speciation in nature: the threespine stickleback model systems. *Trends in Ecology & Evolution* **17**, 480-488.

McPhail J (1994) Speciation and the evolution of reproductive isolation in the sticklebacks (Gasterosteus) of south-western British Columbia. In: *The evolutionary biology of the threespine stickleback* (eds. Bell AM, Foster S), pp. 399-437. Oxford University Press Inc., New York.

Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature* **441**, 947-952.

Moore J-S, Gow JL, Taylor EB, Hendry AP (2007) Quantifying the constraining influence of gene flow on adaptive divergence in the lake-stream threespine stickleback System. *Evolution* **61**, 2015-2026.

Narum SR, Hess JE (2011) Comparison of F-ST outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, 184-194.

Nielsen E, Hemmer-Hansen J, Poulsen N*, et al.* (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**, 276.

Nosil P (2009) Adaptive population divergence in cryptic color-pattern following a reduction in gene flow. *Evolution* **63**, 1902-1912.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**, 375-402.

Olafsdottir GA, Snorrason SS (2009) Parallels, nonparallels, and plasticity in population differentiation of threespine stickleback within a lake. *Biological Journal of the Linnean Society* **98**, 803-813.

Orsini L, Spanier KI, De Meester LUC (2012) Genomic signature of natural and anthropogenic stress in wild populations of the waterflea Daphnia magna: validation in space, time and experimental evolution. *Molecular Ecology* **21**, 2160-2175.

Papa R, Martin A, Reed RD (2008) Genomic hotspots of adaptation in butterfly wing pattern evolution. *Current Opinion in Genetics & Development* **18**, 559-564.

Peichel CL, Nereng KS, Ohgi KA*, et al.* (2001) The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901-905.

Perrier C, Bourret V, Kent MP, Bernatchez L (2013) Parallel and nonparallel genome-wide divergence among replicate population pairs of freshwater and anadromous Atlantic salmon. *Molecular Ecology* **22**, 5577-5593.

Pinho C, Hey J (2010) Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution and Systematics* **41**, 215-230.

Poncet BN, Herrmann D, Gugerli F*, et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Molecular Ecology* **19**, 2896-2907.

Radaelli G, Rowlerson A, Mascarello F, Patruno M, Funkenstein B (2003) Myostatin precursor is present in several tissues in teleost fish: a comparative immunolocalization study. *Cell and Tissue Research* **311**, 239-250.

Raeker MÖ, Bieniek AN, Ryan AS*, et al.* (2010) Targeted deletion of the zebrafish obscurin A RhoGEF domain affects heart, skeletal muscle and brain development. *Developmental Biology* **337**, 432-443.

Raeymaekers JAM, Konijnendijk N, Larmuseau MHD*, et al.* (2014) A gene with major phenotypic effects as a target for selection vs. homogenizing gene flow. *Molecular Ecology* **23**, 162-181.

Raeymaekers JAM, Lens LUC, Van den Broeck F, Van Dongen S, Volckaert FAM (2012) Quantifying population structure on short timescales. *Molecular Ecology* **21**, 3458-3473.

Raeymaekers JAM, Maes GE, Audenaert E, Volckaert FAM (2005) Detecting Holocene divergence in the anadromous-freshwater three-spined stickleback (*Gasterosteus aculeatus*) system. *Molecular Ecology* **14**, 1001-1014.

Raeymaekers JAM, Van Houdt JKJ, Larmuseau MHD, Geldof S, Volckaert FAM (2007) Divergent selection as revealed by P_{ST} and QTL-based F_{ST} in three-spined stickleback (*Gasterosteus aculeatus*) populations along a coastal-inland gradient. *Molecular Ecology* **16**, 891-905.

Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology* **21**, 2852-2862.

Rohlf FJ (2002) tpsDig, digitize landmarks and outlines. Department of Ecology and Evolution, State University of New York at Stony Brook., New York.

Rossman KL, Der CJ, Sondek J (2005) GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. *Nature Reviews: Molecular Cell Biology* **6**, 167-180.

Schluter D (1996) Adaptive radiation along genetic lines of least resistance. *Evolution*, 1766-1774.

Schluter D (2009) Evidence for Ecological Speciation and Its Alternative. *Science* **323**, 737-741.

Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences, USA* **106**, 9955-9962.

Seehausen O, Butlin RK, Keller I*, et al.* (2014) Genomics and the origin of species. *Nature Reviews: Genetics* **15**, 176-192.

Shikano T, Ramadevi J, Merilä J (2010a) Identification of local-and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *Molecular Biology and Evolution* **27**, 2775-2789.

Shikano T, Ramadevi J, Shimada Y, Merilä J (2010b) Utility of sequenced genomes for microsatellite marker development in non-model organisms: a case study of functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *BMC Genomics* **11**, 334.

Shimada Y, Shikano T, Merilä J (2011) A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Molecular Biology and Evolution* **28**, 181-193.

Slatkin (1987). Gene flow and the geographic structure of natural populations. *Science* **236**,787–792.

693 Soria-Carrasco V, Gompert Z, Comeault AA*, et al.* (2014) Stick insect genomes reveal natural
694 selection's role in parallel speciation. *Science* **344**, 738-742.
695 Stapley J, Reger J, Feulner PGD*, et al.* (2010) Adaptation genomics: the next generation.
696 *Trends in Ecology & Evolution* **25**, 705-712.
697 Storz JF (2005) Invited review: Using genome scans of DNA polymorphism to infer adaptive
698 population divergence. *Molecular Ecology* **14**, 671-688.
699 Strasburg JL, Sherman NA, Wright KM*, et al.* (2012) What can patterns of differentiation
700 across plant genomes tell us about adaptation and speciation? *Philosophical*
701 *Transactions of the Royal Society of London, Series B: Biological Sciences* **367**, 364-
702 373.
703 Taylor WR, Dyke GCV (1985) Revised procedures for staining and clearing small fishes and
704 other vertebrates for bone and cartilage study. *Cybium* **9**, 107-119.
705 Tregenza T (2002) Divergence and reproductive isolation in the early stages of speciation. In:
706 *Genetics of Mate Choice: From Sexual Selection to Sexual Isolation* (eds. Etges WJ,
707 Noor MAF), pp. 291-300. Springer Netherlands.
708 Van Dongen S, Lens L, Pape E, Volckaert FAM, Raeymaekers JAM (2009) Evolutionary history
709 shapes the association between developmental instability and population-level
710 genetic variation in three-spined sticklebacks. *Journal of Evolutionary Biology* **22**,
711 1695-1707.
712 van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for
713 identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology*
714 *notes* **4**, 535-538.
715 Vandamme S, Maes G, Raeymaekers J*, et al.* (2014) Regional environmental pressure
716 influences population differentiation in turbot (Scophthalmus maximus). *Molecular*
717 *Ecology* **23**, 618-636.
718 Via S (2009) Natural selection in action during speciation. *Proceedings of the National*
719 *Academy of Sciences, USA* **106**, 9939-9946.
720 Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the
721 human genome. *Plos Biology* **4**, e72.
722 Walsh F, Celeste A (2005) Myostatin: a modulator of skeletal-muscle stem cells. *Biochemical*
723 *Society Transactions* **33**, 1513-1517.
724 Wootton RJ (1976) *The Biology of the Sticklebacks.* Academic Press, London.
725 Yang L, Jelsbak L, Marvig RL*, et al.* (2011) Evolutionary dynamics of bacteria in a human host
726 environment. *Proceedings of the National Academy of Sciences, USA* **108**, 7481-7486.
727 Yeaman S, Otto SP (2011) Establishment and maintenance of adaptive genetic divergence
728 under migration, selection, and drift. *Evolution* **65**, 2123-2129.
729 Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration–
730 selection balance. *Evolution* **65**, 1897-1911.
731 Zulliger D, Schnyder E, Gugerli F (2013) Are adaptive loci transferable across genomes of
732 related species? Outlier and environmental association analyses in Alpine
733 Brassicaceae species. *Molecular Ecology* **22**, 1626-1639.
734

TABLES

Table 1. Comparison of microsatellite-based genome scan studies in three-spined stickleback across freshwater - brackish/salt water gradients, including spatial scale (from regional to global), percentage of conservative outliers, F_{ST}, F_{ST} at the *Eda* locus, and F_{ST} at the *ATP1A1* locus.

Study	Spatial scale	% outliers	F_{ST}	F_{ST} *Eda*	F_{ST} *ATP1A1*
This study	78 km	2.3	0.059	0.118	0.154
DeFaveri *et al.* 2013	Regional	12.5	0.008	-	0.016
Shimada *et al.* 2011	Continental	8.3	0.107	-	0.225
Mäkinen *et al.* 2008	Continental	3.0	0.166	0.653	-
DeFaveri *et al.* 2011	Global	33	0.119	0.405	-

741
742 **Table 2.** Characteristics of the four three-spined stickleback populations used in this study. DTC:
743 distance to coast; MPN: Mean plate number; *Eda*^{L}: frequency of *Eda* low plated allele; AR: allelic
744 richness; H_e: expected heterozygosity; F_{IS}: inbreeding coefficient, with confidence interval (CI
745 95%).
746

Code	Coordinates (N, E)	Habitat	DTC (km)	Salinity(psu)	MPN	Eda^{L}	AR	H_e	F_{IS} (CI 95%)
L01	51°21'10.66, 3°26'01.83	Creek	3.94	2.04	17.4	0.44	9.47	0.69	0.032 (-0.017 - 0.033)
L02	51°21'56.33, 3°31'11.09	Creek	4.30	1.83	16.8	0.51	9.06	0.69	0.095 (0.048 - 0.096)
L12	51°10'29.03, 3°28'10.45	Stream	22.84	0.32	12.7	0.74	6.80	0.61	0.044 (-0.012 - 0.050)
U01	51°02'48.40, 3°33'01.60	Ditch	36.20	0.52	7.0	0.93	6.51	0.57	0.039 (-0.010 - 0.040)

747 **Table 3.** Pairwise F_{ST} values below the diagonal, and associated P-value above the diagonal. For
748 site codes see Table 2.
749

Code	L01	L02	L12	U01
L01	-	0.600	< 0.010	< 0.010
L02	0.005	-	< 0.010	< 0.010
L12	0.054	0.056	-	< 0.010
U01	0.094	0.097	0.047	-

750

751 **Table 4.** Global outlier loci that are putatively under directional selection detected in four
752 populations of three-spined stickleback. The table includes the locus considered, its F_{ST} -value
753 and either the q value for Bayescan or the P-value for lnRH, LOSITAN and Arlequin. Significance
754 is marked as † = q < 0.1, †† = q < 0.05 and ††† = q < 0.01 or as * = p < 0.05, ** = p < 0.01 and ***
755 = p < 0.001. Marker *Stn380*, linked to the *Eda* gene, was added as a reference.
756

Locus	F_{ST}	Bayescan	LnRH	LOSITAN	Arlequin (2 groups)	Arlequin (3 groups)
ATP1A1	0.149			*	*	*
HSPA14	0.159			*	*	*
Ppgm44	0.120	†	***			
Stn74	0.083		*			
Stn177	0.189			**	**	**
Stn46	0.199	†††		**	**	**
Stn380 (Eda)	0.118					

757

758 **Supplementary Table S1.** Characterisation of microsatellite loci of three-spined stickleback,
759 including name, type of marker, gene assignment and putative function.

Locus	Type of marker	Gene	Putative function
1097PBBE	Random genomic	Novel gene	
1125PBBE	Random genomic	Novel gene	
7033PBBE	Random genomic	Integrin beta-3 precursor	
7080PBBE	Random genomic	Neurensin-1	
ATP1A1	Gene-based	Na+/K+-ATPase alpha-subunit isoform 1	Osmoregulation
ATP4A	Gene-based	H+/K+-ATPase alpha subunit	Osmoregulation
CLCN7	Gene-based	CLC chloride channel isoform 7	Osmoregulation
DCT	Gene-based	L-dopachrome tautomerase precursor	Pigmentation
FGF18	Gene-based	Fibroblast growth factor isoform 18	Growth
FGF8	Gene-based	Fibroblast growth factorisoform 8	Growth
GAest11	Random EST		
GAest16	Random EST	Novel gene	
GAest3	Random EST	Dihydropyrimidinase-related protein 2	
GAest34	Random EST	6-phosphofructokinase type C	
GAest35	Random EST	Rod outer segment membrane protein 1	
GAest41	Random EST	Tumor necrosis factor receptor superfamily member 19 precursor	
GAest47	Random EST		
GAest50	Random EST	Eukaryotic translation initiation factor 3 subunit K	
GAest51	Random EST	Protein kinase C delta-binding protein	
GAest6	Random EST	Ninjurin-2	
GAest67	Random EST	Novel gene	
GAest7	Random EST	Ras-related protein Rab-3A	
GAest82	Random EST	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	
GAest84	Random EST		
GAest87	Random EST		
GS1	Gene-based	Glycogen synthase subtype 1	Growth
HSP70Ab	Gene-based	Heat shock protein 70 kDa subtype 2	Thermal response
HSP70Ad	Gene-based	Heat shock protein 70 kDa subtype 4	Thermal response
IGF-II	Gene-based	Insulin-like growth factor II precursor	Growth
KCNJ4b	Gene-based	Inward rectifier potassium channel subfamily J member 4 subtype 2	Osmoregulation
MYHb	Gene-based	Myosin heavy chain subtype 2	Thermal response
NHE2c	Gene-based	Na+/H+ exchanger subtype 3	Hypoxia
NHE3	Gene-based	Sodium/hydrogen exchanger isoform 3 (SLC9A3)	Osmoregulation
NPY2Rb	Gene-based	Neuropeptide Y receptor Y2 subtype 2	Growth
PER1	Gene-based	Period circadian protein homolog 1	Maturation
Ppgm44	Gene-based	Myostatin subtype 2	Growth
Ppgm17	Gene-based	Type I iodothyronine deiodinase	Morphology

Ppgm16	Gene-based	Cold shock protein 70 kDa subtype 2	Thermal response
Ppgm24	Gene-based	Growth hormone receptor 1	Growth
Ppgm28	Gene-based	General transcription factor IIB	Osmoregulation
Ppgm58	Gene-based	Tristetraprolin	Disease
Ppgm21	Gene-based	Fibroblast growth factor 18	Growth
Ppgm47	Gene-based	Na+/K+/2Cl-cotransporter isoform 1 subtype 2	Osmoregulation
Ppgm35	Gene-based	Heat shock protein 70 kDa isoform beta	Thermal response
Ppgm7	Gene-based	Na+/K+-ATPase alpha-subunit isoform 2	Osmoregulation
Ppgm56	Gene-based	Amine-associated receptor	Taste
Ppgm1	Gene-based	Pituitary adenylate cyclase activating polypeptide receptor sub-type 1	Growth
Ppgm2	Gene-based	Pituitary adenylate cyclase activating polypeptide receptor sub-type 2	Growth
Ppgm49	Gene-based	Neuropeptide Y precursor	Growth
Ppgm30	Gene-based	Heat shock protein 25 kDa	Thermal response
Ppgm33	Gene-based	Heat shock protein 70 kDa isoform alpha subtype 2	Thermal response
Ppgm50	Gene-based	Parvalbumin subtype 2	Thermal response
Ppgm6	Gene-based	Na+/K+-ATPase alpha-subunit isoform 1	Osmoregulation
Ppgm38	Gene-based	Insulin-like growth factor I	Growth
PRL-R	Gene-based	Prolactin receptor a	Osmoregulation
PVALBa	Gene-based	Parvalbumin subtype 1	Growth
RhBG	Gene-based	Rhesus glycoprotein isoform 3	Toxic stress
SLC4A10a	Gene-based	Electroneutral Na-coupled HCO2 transporter isoform 10 subtype 1	Hypoxia
Stn102	Random genomic	EF-hand domain-containing family member A2	
Stn108	Random genomic	Bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 4	
Stn119	Random EST		
Stn122	Random EST		
Stn124	Random EST	Brain-specific angiogenesis inhibitor 2 precursor	
Stn163	Random EST	Novel gene	
Stn164	Random EST		
Stn167	Random genomic		
Stn177	Random genomic		
Stn194	Random genomic	Nuclear-localized factor 1	
Stn20	Random EST		
Stn222	Random genomic	Ras-related protein Rab-18	
Stn223	Random genomic	Zinc finger homeobox protein 4	
Stn235	Random EST		
Stn240	Random EST		
Stn257	Random EST	Solute carrier family 35 member F2	
Stn299	Random genomic		
Stn302	Random genomic		
Stn315	Random genomic	Histone deacetylase 4	
Stn328	Random genomic	RING finger protein 220	
Stn34	Random EST		
Stn380	Gene-based	Eda	plate nr

Stn46	Random genomic	Rho guanine nucleotide exchange factor 9	
Stn49	Random genomic	Dedicator of cytokinesis protein 4	
Stn71	Random EST	Novel gene	
Stn74	Random EST		
Stn79	Random genomic	Reticulon-2	
Stn89	Random EST	Rho GTPase-activating protein SYDE2	
HPX	Gene-based	Hemopexin	Immune response

760

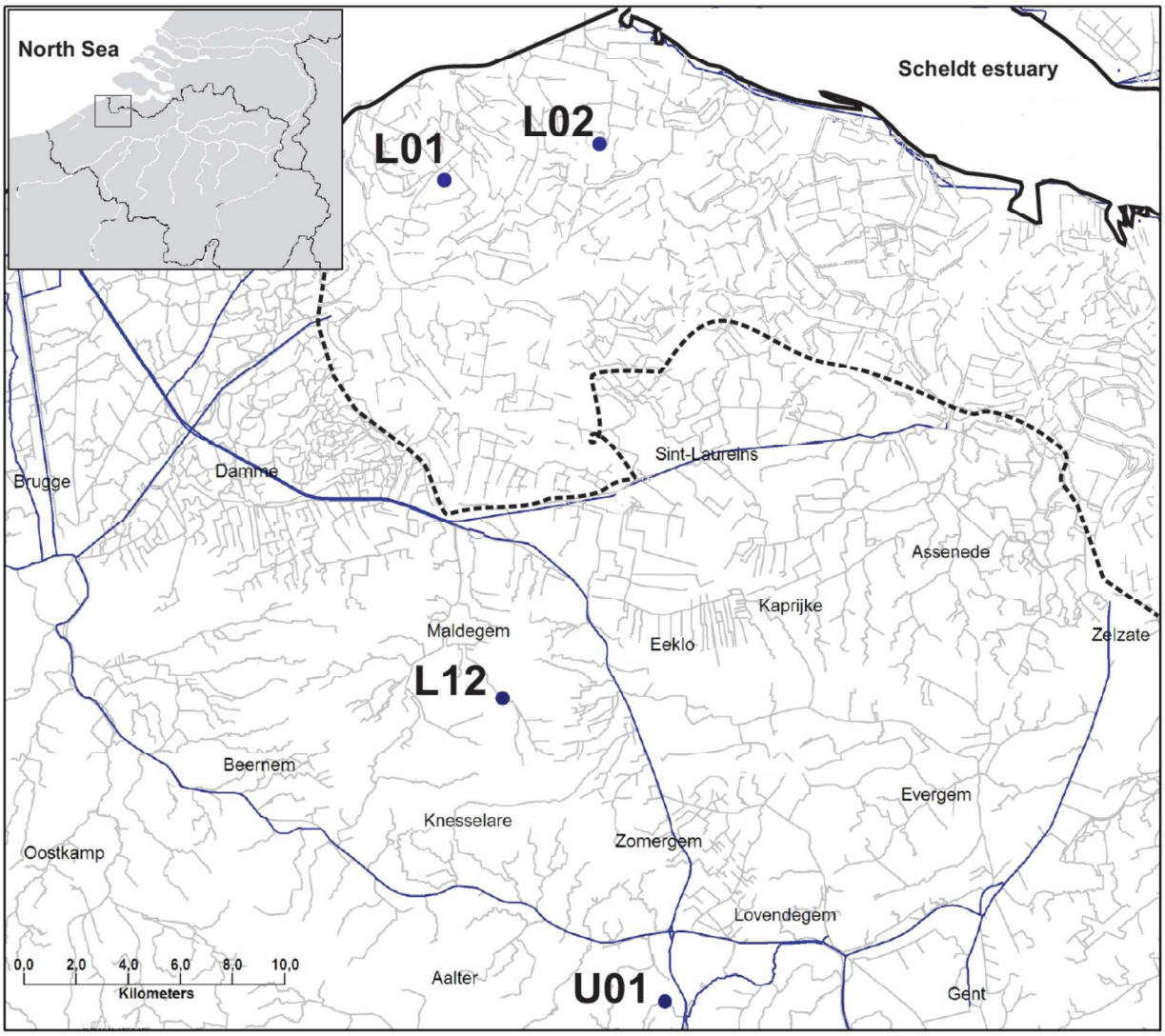
761 **Supplementary Table S2.** Global outlier loci that are putatively under balancing selection as
762 detected with the methods of Bayescan, LOSITAN and Arlequin. Significance is marked as † = q <
763 0.1, †† = q < 0.05 and ††† = q < 0.01 or * = p < 0.05 , ** = p < 0.01 and *** = p < 0.001.

Locus	F_{ST}	Bayescan	LOSITAN	Arlequin (2 groups)	Arlequin (3 groups)
FGF8	0.022		*		
GAest3	0.021		*	*	
GAest34	0.015	††	***	*	*
GAest47	0.009		*	*	*
IGF-II	-0.007		*		
NHE2c	0.027		*	*	
NHE3	0.004		*	*	*
Ppgm16	-0.009		*		
Ppgm2	0.005			**	*
Ppgm30	0.013		*	*	*
Ppgm6	0.005			*	*
Stn122	0.014		***	**	*
Stn20	0.017		*		
Stn223	0.019	†	**	*	*
Stn79	-0.020		***	*	*
Stn89	0.028	†††			

764

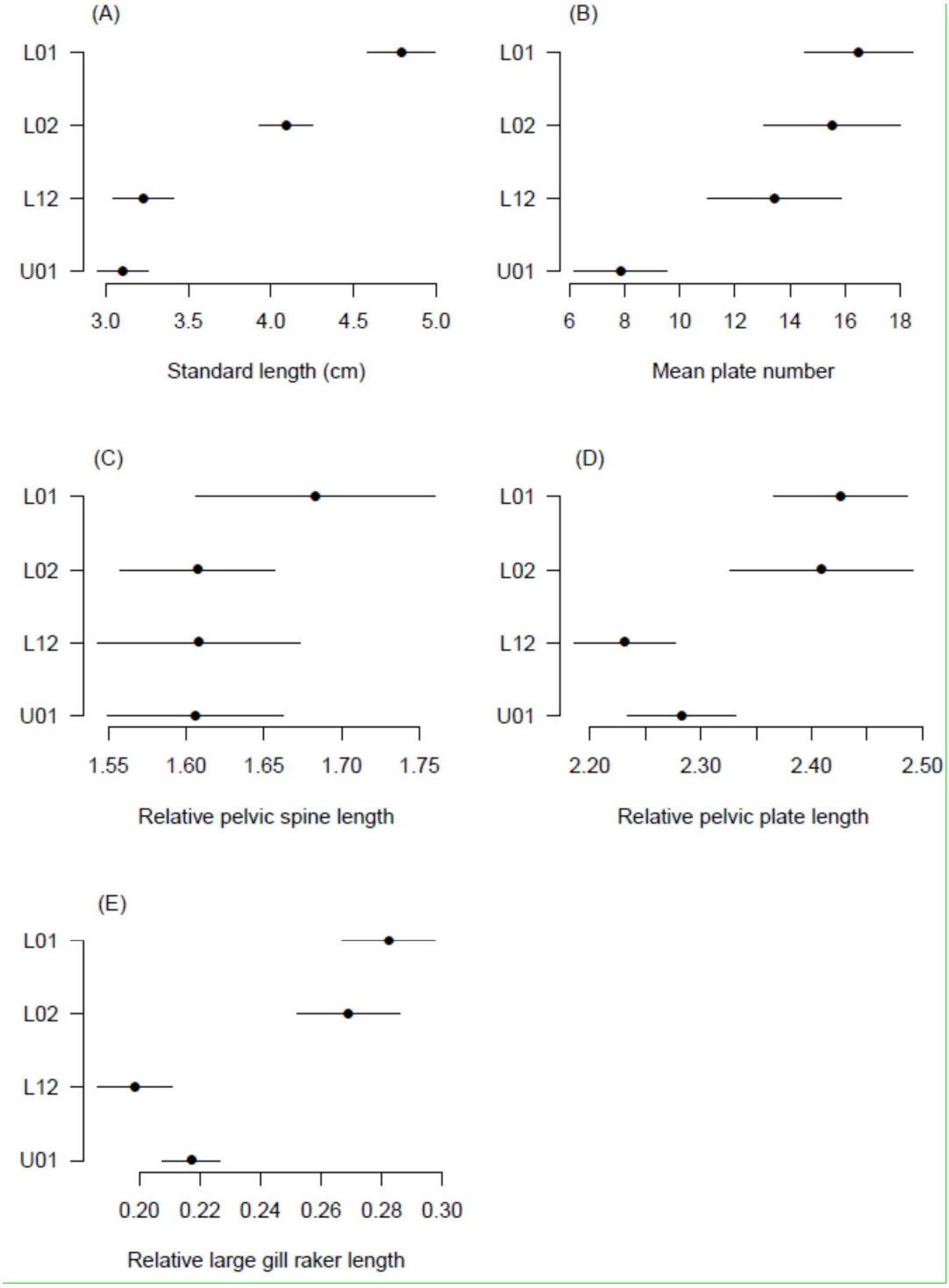
765 **FIGURE LEGENDS**
766
767 **Figure 1.** Map with the four sampling locations of three-spined stickleback in northwestern Belgium and
768 southwestern Netherlands. River network and major towns are mapped.
769
770 **Figure 2.** Average values for A) standard length, B) plate number, C) relative pelvic spine length, D)
771 relative pelvic plate length, and E) relative length of the second large gill raker in two brackish water
772 populations (L01, L02) and two freshwater populations (L12 and U01)) of three-spined stickleback.
773 Horizontal bars represent 95% confidence intervals.
774

775 **FIGURES**
776
777 **Figure 1**
778



779
780
781

Figure 2


(A)
L01
L02
L12
U01
3.0
3.5
4.0
4.5
5.0
Standard length (cm)
(B)
L01
L02
L12
U01
6
8
10
12
14
16
18
Mean plate number
(C)
L01
L02
L12
U01
1.55
1.60
1.65
1.70
1.75
Relative pelvic spine length
(D)
L01
L02
L12
U01
2.20
2.30
2.40
2.50
Relative pelvic plate length
(E)
L01
L02
L12
U01
0.20
0.22
0.24
0.26
0.28
0.30
Relative large gill raker length

Signatures of selection in the three-spined stickleback along a small scale ~~salt~~brackish water~~ and~~ - freshwater transition zone

Nellie Konijnendijk[1], Takahito Shikano[2], Dorien Daneels[1,*], Filip A.M. Volckaert[1], Joost A. M. Raeymaekers[1]

[1] Laboratory of Biodiversity and Evolutionary Genomics, University of Leuven, Ch. Deberiotstraat, 32, B-3000 Leuven, Belgium

[2] Ecological Genetics Research Unit, Department of Biosciences, University of Helsinki, P.O. Box 65, FI-000 14, Helsinki, Finland

[*] Current address: Centre for Medical Genetics, Reproduction and Genetics; Reproduction, Genetics and Regenerative Medicine, Vrije Universiteit Brussel (VUB), Universitair ziekenhuis Brussel (UZ Brussel), Laarbeeklaan 101, 1090 Brussels, Belgium.

Running header: Adaptation in three-spined stickleback

Address for correspondence:
Joost Raeymaekers, University of Leuven
Laboratory of Biodiversity and Evolutionary Genomics
Ch. Deberiotstraat, 32
B-3000 Leuven, Belgium
e-mail: joost.raeymaekers@bio.kuleuven.be
Phone : + 32 16 32 39 66, Fax : +32 16 32 45 75

ABSTRACT

Local adaptation is often obvious when gene flow is impeded, such as observed at large spatial scales and across strong ecological contrasts. However, ~~but~~ it becomes less certain at small scales such as between adjacent populations or across weak ecological contrasts, when gene flow is strong. While studies on ~~Genomic~~ genomic adaptation tend to focus on the former, ~~is often investigated in populations that are either geographically isolated or that occur across strong ecological contrasts. In both cases, gene flow might be strongly impeded, and adaptation might have been largely achieved. L~~less is known about the genomic targets of natural selection ~~at the onset of divergence,~~ in the latter situation. ~~when gene flow is strong.~~ In this study we ~~focus~~ investigate genomic adaptation ~~on~~ in ~~a system characterized by very limited genetic isolation, which is often a realistic setting in nature. We performed a genome scan with 87 microsatellite markers among~~ populations of the three-spined stickleback *Gasterosteus aculeatus* L. across a small-scale ~~along an~~ ecological transition with salinities ranging from brackish to fresh~~water habitats~~. Adaptation to ~~Salinity~~ salinity ~~is known to be an important driver of adaptation~~has been repeatedly demonstrated in this species ~~and has been studied in stark contrasting environments several times. Adaptation to this ecological variable is therefore likely~~. A genome scan based on 87 microsatellite markers ~~As expected, we detected~~revealed only few signatures of selection ~~at this scale of population divergence~~, likely owing to the constraints that homogenizing gene flow puts on adaptive divergence. However, ~~those that we did find~~ the detected loci appear repeatedly as targets of selection in similar studies of genomic adaptation in three-spined sticklebacks~~s~~. We conclude that the signature of genomic selection in the face of strong gene flow is weak, yet detectable. We argue that the range of studies of genomic divergence should be extended to include more ~~the very onset of population divergence~~systems characterised by limited

Formatted: Font: Italic

geographical and ecological isolation, which is often a realistic setting in nature ~~i.e. at the level where it has been hypothesized that adaptation is primarily achieved through non-genomic mechanisms such as phenotypic plasticity~~.

Keywords: adaptation, candidate gene, evolution, fish, genome scan, population genomics, selection

INTRODUCTION

Recent questions on the mechanisms of evolutionary biology revolve around the genomic architecture of species and the effects that processes such as selection, drift, mutation and gene flow have on the genome. Changes in the environment can now be linked to the genetic signatures of these processes and hence promote the understanding of the genetic basis of ecological adaptation. This allows us to understand the various components of the mechanism of adaptation such as which genes are involved, how they are distributed in the genome, and how often the same genes lead to certain adaptations both in the laboratory (Becks *et al.* 2012; ~~and~~ see Barrick & Lenski 2013 for other examples) and nature (Mitchell-Olds & Schmitt 2006; Yang *et al.* 2011; Jones *et al.* 2012a; Orsini *et al.* 2012). Additionally we want to understand which processes are essential for adaptation to succeed. One way to find genes that are important for the process of adaptation is to identify genetic signatures of adaptation in natural populations. These "genome scans" identify genes that are under selection across contrasting environments and might thus be involved in the transition from one ecological extreme to the other. When interpreted with care (de Villemereuil *et al.* 2014), this approach has shown to be a promising way to find links between genotype, phenotype and fitness in natural populations (Storz 2005).

Genome scans have so far generated two major insights. First, selection may act on many parts of the genome (Nosil 2009; Hohenlohe *et al.* 2010; Feder *et al.* 2012; Jones *et al.* 2012a; Jones *et al.* 2012b; Strasburg *et al.* 2012; Arnegard *et al.* 2014; Seehausen *et al.* 2014). Nosil (2009) estimated that 5-10 % of the genome is affected by natural selection ~~(although such estimates should be interpreted cautiously as the authors suggested, see also~~

Stapley *et al.* (2010) and Excoffier *et al.* (2009)). Based on a genome-wide analysis of selection, Hohenlohe *et al.* (2012) showed that the genome is much more structured and dynamic than expected from theory. More studies have now contributed to these topics and show that the areas affected by natural selection are typically patchily distributed ("hotspots") across the genome (Voight *et al.* 2006; Papa *et al.* 2008; Hohenlohe *et al.* 2010). Hohenlohe *et al.* (2010) suggest thatG genomic inversions, deletions, repetitive elements and other structural changes facilitate the rise of these islands of divergence, causing a physical barrier to crossing over and thus adding to linkage disequilibrium (Hohenlohe *et al.* (2010). On the other hand, a review dedicated to genome scans in plants (Strasburg *et al.* 2012) describes that most cases of adaptation across environmentsin plants involve many genomic areas with a wide distribution across the genome rather than clustered hotspots (Strasburg *et al.* 2012).

A second common finding of genome scans is that organisms may show parallel as well as non-parallel there is large variation across studies in the degree of parallelism for genetic responses to environmental change. For example, Atlantic salmon (*Salmo salar*) populations that undergo a parallel change in environment show non-parallel adaptive divergence within the genomeat the genomic level (Perrier *et al.* 2013). In cases of parallel phenotypic evolution in the three-spined stickleback (*Gasterosteus aculeatus* L.), common (Hohenlohe *et al.* 2010; Shimada *et al.* 2011; Jones *et al.* 2012b) as well as unique (Mäkinen *et al.* 2008; Jones *et al.* 2012b; Roesti *et al.* 2012) genomic regions are targeted by selection. This has been confirmed in experiments with stick insects (*Timema cristinae*) (Soria-Carrasco *et al.* 2014) and in genome scans of lake white fish (*Coregonus* sp.) (Bernatchez *et al.* 2010). In contrast, parallel phenotypic evolution in a other species, like such as the nine-spined

stickleback (*Pungitius pungitius*) (Shikano *et al.* 2010a) or the rough periwinkle *Littorina saxatilis* (Butlin *et al.* 2014), were mainly characterized by non-parallel genomic signatures of selection.

Adaptive divergence is thought to proceed as a balance between divergent selection and homogenizing gene flow (Levene 1953; Hagen 1967; Endler 1973; Bell 1982) and hence may reach various stages. This has long been recognised at the phenotypic level (Tregenza 2002; Moore *et al.* 2007; Hendry 2009; Schluter 2009). More recently, theoretical and empirical studies have improved our understanding of the genomic architecture at various stages of adaptation as well (Pinho & Hey 2010; Yeaman & Otto 2011; Yeaman & Whitlock 2011). Nevertheless, the focus of genome scans is often on systems where it is reasonable to assume that population divergence has a strong adaptive component. To do so, genome scans often target populations ~~from highly contrasting environments~~at large spatial scales and across strong ecological contrasts (Hohenlohe *et al.* 2010; Poncet *et al.* 2010; Stapley *et al.* 2010; Zulliger *et al.* 2013). While this has generated great insight in the genomic basis of adaptation and speciation, knowledge might be biased towards stages ~~of population divergence with very little gene flow~~where gene flow may be largely impeded, and where adaptation is already largely achieved. Gene flow modifies the response to selection by modulating the distribution of the genes that underlie ecologically relevant traits (Slatkin 1987). The study of adaptation with gene flow, especially in study systems where genetic divergence is far from complete such as across a small-scale ecological transitions or between highly connected populations, is therefore crucial to understand how populations diverge to different ecological optima~~Less is known about the genomic targets of natural selection at the onset of divergence, when gene flow is strong (Feder *et al.* 2012). To~~

~~understand how populations diverge to different ecological optima it is crucial to study systems where isolation is far from complete. Populations with recurrent gene flow can teach us about the generality and constraints of adaptive divergence, and the importance of genomic processes at the onset of population divergence~~ (Hansen *et al.* 2002; Storz 2005; Nielsen *et al.* 2009~~a~~; Coscia *et al.* 2011; DeFaveri *et al.* 2013; Vandamme *et al.* 2014).

The three-spined stickleback represents an excellent model for the study of adaptive divergence, as phenotypic responses to several ecological changes are frequent and well-documented (McPhail 1994; Foster *et al.* 1998; McKinnon & Rundle 2002; Boughman 2007). A diverse number of populations can be found that occupy ~~different positions along a "speciation continuum" from state 1 where there is continuous gene flow in a panmictic population to state 4 where there is~~various stages of divergence from panmixia to complete and irreversible reproductive isolation (Hendry *et al.* 2009~~; Feder *et al.* 2012~~). This provides an excellent framework to investigate ~~progress towards ecological speciation~~population divergence at the phenotypic and genomic level. The genome scans that have been applied to three-spined stickleback ~~populations cover a broad diversity of habitat pairs across a wide geographical range. The habitat pairs have been divided in~~have contrasted marine-fresh water (Mäkinen *et al.* 2008; Hohenlohe *et al.* 2010; DeFaveri *et al.* 2011; Shimada *et al.* 2011; Jones *et al.* 2012a; Jones *et al.* 2012b; DeFaveri & Merilä 2013), lake–stream (Deagle *et al.* 2012; Roesti *et al.* 2012), and benthic-limnetic (Olafsdottir & Snorrason 2009; Jones *et al.* 2012a; Lucek *et al.* 2014) population pairs, as well as populations from ~~and~~ clean versus polluted water (Lind & Grahn 2011). ~~In fresh–salt water comparisons, the studies only cover a limited part of the speciation continuum. They have largely been carried out on populations in stark contrasting environments with limited gene flow (Table 1), although~~

~~some global studies have added some populations with recurrent gene flow, (Jones *et al.* 2012a), like the Scottish river system that harbours a very narrow range of hybrid zone (Jones *et al.* 2006) or with a limited set of markers in the Baltic Sea (DeFaveri *et al.* 2013). Several studies cover a global or regional scale (Mäkinen *et al.* 2008; DeFaveri *et al.* 2011; Jones *et al.* 2012a; Jones *et al.* 2012b; DeFaveri *et al.* 2013) and a larger area than what would be considered the dispersal rates typical of any of these populations of three-spined sticklebacks.~~ A common finding is that several genes or gene regions are repeatedly selected across populations and locations, although population-specific regions do appear as well (Hohenlohe *et al.* 2010; DeFaveri *et al.* 2011; Shimada *et al.* 2011; Jones *et al.* 2012b). Others have found that most regions under selection were highly specific to the location under study (Mäkinen *et al.* 2008; Deagle *et al.* 2012; Roesti *et al.* 2012). ~~Whether any of these patterns can be attributed to the stage in the speciation continuum was not considered.~~

In this study we investigate genomic adaptation in three-spined stickleback populations from the Belgian-Dutch lowlands. Populations in this area differ in various morphological traits, which are often correlated with salinity and distance to the coast (Heuts 1947; Raeymaekers *et al.* 2005; Raeymaekers *et al.* 2007; Van Dongen *et al.* 2009; Raeymaekers *et al.* 2012). At the same time, gene flow between these populations is ~~recurrent~~ moderate to strong due to high connectivity (Raeymaekers *et al.* 2014). This has the advantage that we can study adaptation with ongoing gene flow, which is often a realistic setting in nature. All sites are part of an interconnected landscape of canals and streams along an ecological ~~gradient~~ transition (i.e. discrete habitats with salinities ranging between brackish and freshwater) at a much ~~comparatively~~ smaller geographical scale than other genome scan studies across salinity gradients in three-spined stickleback (Table 1).

Formatted: Indent: First line: 0"

~~We can thus use this system to deduce which genes are the targets of selection at the very onset of population divergence.~~

In order to explore the adaptive changes across this transition, we screened two populations from each end of the transition for genomic signatures of selection. We hypothesized~~We performed a genome scan along this ecological transition hypothesising~~ that (i) the outcome of small-scale adaptation along ecological transitions may vary at the genomic level and that (ii) adaptation at the genomic level may be influenced by gene flow. ~~Four populations were selected along a salinity cline and screened with~~Populations were screened for 87 microsatellite markers of which 41 are linked to genes with a range of ecologically relevant functions (Shikano *et al.* 2010b; Shimada *et al.* 2011). We expect that the relatively high gene flow among these populations might constrain local adaptation, despite obvious differences in phenotype across these extremes. We determine outlier loci and compare the results to previous studies in three-spined stickleback across similar ecological contrasts.

Formatted: Indent: First line: 0.49"

Formatted: Indent: First line: 0"

MATERIAL AND METHODS

Study area

Three-spined stickleback (*Gasterosteus aculeatus* L.; Gasterosteidae) from the coastal lowlands (polder) of Belgium and the Netherlands (Figure 1) reside in ponds, ditches, streams, estuaries or polder creeks. They have an anadromous or landlocked life style (Heuts 1947; Wootton 1976; Raeymaekers *et al.* 2005; Raeymaekers *et al.* 2007). The polder and surrounding areas contain diked brackish and freshwater habitats of Holocene origin with

varying levels of connectivity to adjacent estuaries and the open sea. Populations which live in close proximity to the sea (< 10 km) reside in brackish water of which the salinity is influenced by rainfall and water management. On a scale of less than 50 km further inland, salinity ~~and associated environmental variables change with increasing~~drops ~~distance to the coast~~to freshwater levels (Raeymaekers *et al.* 2014). ~~Populations which live in close proximity to the sea (< 10 km) reside in brackish water of which the salinity is influenced by rainfall and water management~~Lateral plate number, an important ecological trait, also decreases with distance to the coast, with population averages from 20 to 5 (Heuts 1947; Raeymaekers et al. 2014). Populations bordering the North Sea and the Baltic are typically polymorphic for lateral plate number (Heuts 1947; Raeymaekers *et al*. 2014), so higher or lower population averages are rare. This range is therefore representative for the phenotypic extremes we can find in this part of the stickleback's distribution range.

Field sampling

Field sampling was conducted in spring 2009 in parallel with a multi-year study by Raeymaekers *et al*. (2014), describing the distribution of lateral plate number in the study area. ~~Fifteen sites inhabited by the three spined stickleback were monitored on a regular basis during two years for various environmental variables including salinity, pH and water depth (Raeymaekers *et al.* 2014).~~ Two brackish water creeks (L01 and L02) and two freshwater ditches (L12 and U01), representing the sites with the most extreme values for salinity, were selected (Table 2). Thirty fish from each population were collected with dip nets, immediately anaesthetized with MS222 (2 $g.l^{-1}$) and frozen on dry ice. In the lab, sticklebacks were thawed on ice, measured (standard length (SL); ± 0.1 cm), weighed (± 0.01 g), photographed and fin clipped for a DNA sample. ~~Field measurements included~~

monitoring the water for pH and conductivity (µS.cm^{-1}; converted to salinity in psu for analysis) using a Hach probe (Hach, Loveland, Co, USA). An index for the turbidity of the water was obtained with a Sneller tube and water depth was calculated as the median depth (cm) of five equidistant points along a stretch of 100 m of the water body. All macro-invertebrate predators of three-spined stickleback, including backswimmers (*Notonecta glauca*), dragonfly larvae (*Anax* sp. and *Aeschna* sp.) and great diving beetles (*Dytiscus marginalis*) obtained in the nets were counted.

Morphological measurements

In line with previous studies on phenotypic divergence between the stickleback populations of the Belgian-Dutch lowlands (Heuts 1947; Raeymaekers *et al.* 2005; Raeymaekers *et al.* 2007; Van Dongen *et al.* 2009; Raeymaekers *et al.* 2012) we investigated variation at six morphological traits. The left side of each fish was photographed from a standard angle, and a ruler was placed in each photograph for scaling. Dorsal spine length, pelvic spine length and pelvic plate length were measured from pictures using the software tpsDig 1.37 (Rohlf 2002). A subsample of the fish was rinsed with water for 72 h, bleached for 4 h (1% KOH bleach solution) and stained with alizarin red (Taylor & Dyke 1985). Stained fish were used to determine the number of lateral plates on the right side of the fish. The right part of the gills was then dissected and the number and length of the large gill rakers were quantified under a dissection scope. Statistical analysis of variance (ANOVA) confirmed the presence of significant differences between the four populations for standard length, lateral plate number, pelvic plate and spine length and large gill raker length (Figure 2). The most studied trait in these populations is the plate morphology which can either be described as low plated or high plated. These terms refer to the number of lateral plates the individual has

~~with high plated individuals having 20 or more plates, while low plated individual have less than 10. It is well established that this trait correlates with the salinity of the water within the habitat and that this trait is regulated by a major effect gene called *Eda* and several minor effect genes (Colosimo *et al.* 2005). All populations in this study system are polymorphic for this trait, but the frequency of the different morphs change considerably (Raeymaekers *et al.* 2014).~~

Marker selection

A set of 110 microsatellite markers was selected, including a range of putatively neutral markers, to set a proper neutral F_{ST} background. In our study area there is a salinity cline, but ~~numerous~~ various other factors may co-vary with this cline. Therefore we included 41 markers that are known to be linked to functional genes in a range of ecologically relevant functions such as salinity, growth and immunity (Shikano *et al.* 2010b; Shimada *et al.* 2011). Several of these genes have been found to be under selection in the three-spined stickleback or nine-spined stickleback of other ~~fresh-salt water~~salinity ~~systems~~ transitions (Hohenlohe *et al.* 2010; Shikano *et al.* 2010a; DeFaveri *et al.* 2011; Shimada *et al.* 2011; Jones *et al.* 2012a; Jones *et al.* 2012b). This allows us to compare these systems to some extent with ours and pinpoint parallel changes across systems. The marker closely linked to the *Eda* gene (*Stn380*), a major effect gene underlying variation in plate number (Colosimo *et al.*, 2005), was included as a reference gene that is often under selection in freshwater-salt water comparisons (Raeymaekers *et al.* 2007; Mäkinen *et al.* 2008; ~~Barrett *et al.* 2009; Van Dongen *et al.* 2009;~~ DeFaveri *et al.* 2011; DeFaveri & Merilä 2013).

Formatted: Dutch (Belgium)

DNA extraction and genotyping

Genomic DNA was extracted from fin clips using a proteinase K digestion step and the Nucleospin 96 Tissue DNA Extraction kit (Macherey-Nagel). Individuals were genotyped at 110 microsatellite loci that were arranged in 21 multiplexes of 4 to 8 markers at a time, with EST based markers and markers within or near genes with relevant functions included (Table S1). Loci were amplified with the Qiagen® Multiplex PCR Kit (Qiagen, Venlo, the Netherlands). The 10 µl PCR cocktail contained 1-100 ng genomic DNA, 2 pmol each of forward and reverse primers, 1x Qiagen Multiplex PCR master mix, 0.5x Q-solution, and RNase-free water. The reaction consisted of an initial activation step of 15 min at 95 °C, followed by 30 cycles of 30 s at 94 °C, 90 s at 53 °C and 60 s at 72 °C. A final elongation step of 5 min at 60 °C was performed. Allele sizes were determined by means of an internal ET ROX 550 size standard (Amersham Biosciences). Polymerase chain products were visualized using a MegaBace 1000 automated sequencer (Amersham Biosciences). Alleles were scored with the Fragment Profiler v1.2 software (Amersham Biosciences), using visual scoring and manual corrections. Marker Stn380, linked to the *Eda* gene, was scored separately to determine the frequency of the "low plated" allele in each population (Table 2).

Data analysis

Phenotypic differentiation

For each trait, analysis of variance (ANOVA) was performed to test for statistical differences between the four populations. For traits which depend on size, standard length was included as a covariate.

Genetic diversity and genetic differentiation

Genotypes were checked for scoring errors attributable to stutter-products, large allele dropout or null-alleles, using MICRO-CHECKER v2.3 (van Oosterhout *et al.* 2004). Estimates of allelic richness, genetic diversity (H_e, H_o), global and pairwise F_{ST}, with a significance calculated with 1000 bootstraps over loci, were calculated using the GENETIX v4.05.02 software (Belkhir *et al.* 1996).

Genomic signatures of selection

We conducted global outlier tests ~~using our entire dataset~~ to find outliers across all populations and used pairwise comparisons to check whether the outliers found by global tests ~~could be confirmed and~~ were due to habitat differences. Loci that are under directional selection are expected to have lower intrapopulation variability and larger interpopulation variability than neutral loci. Loci under directional selection can thus be traced by patterns in heterozygosity, differences in F_{ST} values or a combination of the two. ~~We used several outlier detection approaches as e~~Evaluation of several outlier detection methods has shown that these methods differ in number of false positives and false negatives (Narum & Hess 2011). ~~Therefore we implemented and compared the results from four different outlier detection tests: LOSITAN (Antao *et al.* 2008), the outlier detection method within ARLEQUIN v3.5.2.3 (Excoffier & Lischer 2010), hierarchical clustering and BAYESCAN v2.01 (Foll & Gaggiotti 2008) LnRH (Kauer *et al.* 2003).~~ Four methods were therefore compared: LOSITAN (Antao *et al.* 2008), the outlier detection method implemented in ARLEQUIN v3.5.2.3 using hierarchical clustering (Excoffier & Lischer 2010), BAYESCAN v2.01 (Foll & Gaggiotti 2008), and LnRH (Kauer *et al.* 2003). The first three methods ~~were used~~ were used to determine global outliers. Additionally we did pairwise comparisons of all populations using LOSITAN and LnRH to detect specific ~~outliers~~ signatures of selection in freshwater ~~salt~~ brackish water

~~population pairs~~, fresh~~water~~-freshwater and ~~salt~~brackish water-~~salt~~ brackish water populations pairs.

The four methods are based on different underlying assumptions. LOSITAN is based on an island model that uses a coalescent F_{ST}-outlier method based on the distribution of F_{ST} as a function of the heterozygosity. We used the function that first establishes a neutral F_{ST} baseline by removing putative markers under selection outside the 95% interval with 10^5 simulations. The infinite allele model was used with a 95% and 99% confidence interval. We ran 10^5 simulations as recommended by Antao (2008). The outlier detection software implemented in ARLEQUIN uses the same island model, but adds on the option for hierarchical clustering. In the presence of strong hierarchical population structure it reduces false positives by a hierarchical analysis of genetic differentiation (Excoffier *et al.* 2009). We clustered populations ~~for~~ according to two ~~probable~~ scenarios: in two groups (brackish water populations versus freshwater populations), and three groups (brackish water populations versus each freshwater population separately). The first scenario simulates a common descent of the two freshwater populations from the marine population, while the second one simulates a separate split of the two freshwater populations from the marine population. We used the standard settings of 20,000 simulations for each run and 100 demes per group. The method executed by BAYESCAN uses a logistic regression model ~~described by Beaumont and Balding (2004),~~ which explains the observed pattern of diversity by dividing it in a locus- and a population-specific component~~.~~ (Beaumont and Balding 2004). One benefit of this method is that it allows for different migration rates and different effect sizes and thus can be used for scenarios that deviate from the island model. We conducted 10 pilot runs of 5,000 iterations, followed by an additional 150,000 iterations and a burn-in of 50,000 iterations. Outliers were appointed based on 90%, 95% and 99%

posterior probabilities. Finally, ~~The~~ the LnRH method is designed especially for microsatellite markers and determines the reduction of heterozygosity. This method is based on the assumption that microsatellites linked to a gene under selection will show reduced levels of diversity between two populations. After standardization of the LnRH estimates with a mean of zero and a standard deviation of one, we determined outliers at 95% and 99% level. We subtracted outliers that were found in pairwise comparisons within the same habitat type, to limit the number of false positives.

All tests based on simulations were executed three times to test for robustness of the results. Only outliers that were detected ~~in two out of three trials within each method~~ at least twice with each method were scored as a putative outlier. We detected outlier loci that are under balancing and directional selection. However, interpreting loci under balancing selection is difficult, as there are still limitations for the identification of loci under balancing selection (Hansen *et al.* 2010; Narum & Hess 2011). Therefore we only discuss ~~only~~ the loci under directional or positive selection. Loci under balancing selection are provided in Supplementary Table S2.

RESULTS

Phenotypic differentiation

Statistical analysis of variance (ANOVA) confirmed the presence of significant differences between the four populations for standard length, lateral plate number, pelvic plate and spine length, and large gill raker length (Figure 2). In particular, the brackish water populations (L01 and L02) were larger in size than the freshwater populations (L12 and U01; $F_{3,155}$ = 75.6; P < 0.0001), and had more lateral plates ($F_{3,147}$ = 10.4; P < 0.0001), longer pelvic

spines (especially population L01; $F_{3,93}$ = 5.51; P = 0.0016), longer pelvic plates ($F_{3,93}$ = 8.60; P < 0.0001), and larger gill rakers ($F_{3,93}$ = 19.8; P < 0.0001).

Genetic diversity and genetic differentiation

After genotyping of the individuals with 110 microsatellite markers, we selected 87 markers with good amplification quality to perform the data analysis. A total of 1116 alleles were observed in four populations of 26 to 28 individuals each, with an average of 13 alleles per locus and a range from 2 to 34. Observed heterozygosity ranged from 0.022 to 0.95 across loci. ~~and e~~Expected heterozygosity ranged from 0.05 to 0.92 across loci and from 0.57 to 0.69 across populations (Table 2). Allelic richness among populations varied from 6.51 to 9.47, with the freshwater populations being less diverse than the brackish water populations. Eight loci were possibly affected by null alleles or stutters. We therefore performed all final analyses with and without these loci and specify when results differ. Differentiation among populations was moderate with the global F_{ST} value being 0.059. Pairwise F_{ST} values were significant, except between the two brackish water populations (Table 3).

Genomic signatures of selection

Six of the 87 loci were assigned as outliers in at least one of the methods, with the different methods identifying 4, 2, 2, and 4 outliers for Lositan, Bayescan, lnRH and Arlequin ~~(with hierarchical nesting)~~ respectively (Table 4). The methodologically similar tests performed by Lositan and Arlequin with hierarchical ~~grouping~~ clustering resulted in the same set of outliers. In contrast, only two outliers were shared across methodologically different outlier detection methods; our conservative measure of outliers hence amounts to 2.~~2~~ 3 % of the

total number of loci. This value is low in comparison with other ~~outlier~~ genome scan studies ~~of~~ of three-spined stickleback ~~populations~~ (Table 1).

The two outliers for directional selection that appear consistently across methodologically different tests are marker *Stn46* (identified by Arlequin/Lositan and Bayescan) and *Ppgm44* (identified by lnRH and Bayescan). *Stn46* has been previously associated with the Rho guanine nucleotide exchange factor 9 (Table S1). I~~n pairwise comparisons i~~t ~~is~~ was also assigned as an outlier in ~~only~~ one ~~of the~~ pairwise freshwater versus ~~salt~~ brackish water comparison~~s~~. ~~*In lnRH it is not assigned as an outlier.*~~ *Ppgm44* is a marker that is linked to the gene *myostatin2~~MSTNb~~*, which is associated with growth (Table S1). I~~n pairwise comparisons of LnRH, PPGM44 is found in all the~~t was also assigned as an outlier in all four pairwise fresh~~—~~water versus ~~salt~~ brackish water comparisons~~. In LOSITAN the same was found in pairwise comparisons, yet PPGM44 was not assigned as outlier in the global analysis~~. ~~Of~~ Among the four other ~~markers~~ outliers that ~~are~~ were detected (Table 4), one is associated with osmoregulation, namely an alpha subunit of the Na+/K+ ATPase (ATP1A1). Other outliers are either linked to functions ~~like~~ such as thermal response (HSPA14) or ~~are~~ were assumed to be neutral. ~~The outlier~~Locus *Stn34* was also assigned as an outlier, but ~~this~~ since ~~might be caused by~~ null alleles were detected at this locus, it was~~and was deleted~~ excluded from Table 4. The *Eda* gene, with frequencies of the low-plated allele ranging between 0.44-0.51 in the brackish populations and 0.74-0.93 in the freshwater populations (Table 2), did not show up as an outlier in the analyses (Table 4).

DISCUSSION

We found ~~that~~ a limited ~~proportion~~ number of loci ~~show~~ showing a signature of selection among three-spined stickleback populations inhabiting the coastal Dutch-Belgian lowlands ~~(Table 4)~~. Six outliers were found when all outlier detection methods were considered and only two ~~out of those six~~of those were shared among methodologically different tests, despite differentiation among populations in ecology and phenotype. Not only the number but also the proportion of outliers was low as compared to other studies (Table 1). The function of one of the outlier loci ~~can be~~could be directly related to salinity, the most obvious ecological gradient in our study area. We here discuss possible explanations for the limited amount of outliers, the putative function of the genes linked to outliers, and ~~what this tells us about~~ the consequences for ~~genetic differentiation in populations with concurrent~~ adaptive divergence in the face of gene flow.

The presence of a limited ~~amount~~ number and proportion of outliers might be attributed to ~~several~~ a number of technical aspects. First, we used relatively few markers. However, among the 87 markers selected, 41 markers are linked to ecologically relevant functions, of which several have been shown to be under selection ~~in other fresh-salt water~~ in systems with similar salinity clines (Mäkinen *et al.* 2008; Shimada *et al.* 2011; DeFaveri *et al.* 2013). ~~Additionally, not only the number, but also the proportion of loci under selection was low compared to other studies, indicating that the number of markers did not cause this limited amount of outliers.~~ Second, analysing relatively few populations may have two drawbacks: 1) it might increase the risk of ~~not detecting all possible genes under selection~~ false negatives due to the limited number of individuals sampled, and 2) it might increase the difficulty in separating the signal of selection from the geographical and historical signal

(Bierne *et al.* 2011; Bierne *et al.* 2013) ~~as we miss the possibility to confirm outliers across more than two population pairs~~. The small overlap in outlier loci among detection methods might be a symptom of this. However, in a larger ~~set~~ study ~~of~~ across 14 populations ~~(and~~ 2320 individuals), describing the distribution of lateral plate number and the underlying *Eda* gene in ~~from~~ our study area (Raeymaekers *et al.* 2014), the signature of selection at the *Eda* gene was ~~equally weak~~neither~~ot~~ significant ~~either~~. This bi-allelic gene is experiencing selection across various other salinity transitions in three-spined stickleback (Raeymaekers et al. 2007; Mäkinen et al. 2008; DeFaveri et al. 2011; DeFaveri & Merilä 2013) ~~(Raeymaekers *et al.* 2014)~~. ~~This~~ ~~It shows~~This suggests that increasing sample size does not necessarily enhance ~~detecting the signatures~~the detectability of selection. We therefore do not expect that technical issues are a major explanation for our findings.

~~There might be several other reasons~~An alternative explanation for the low proportion of outlier loci~~, despite the established knowledge from other salt-freshwater comparisons in the three-spined stickleback that several genes can be under natural selection in these contrasting environments (Hohenlohe *et al.* 2010; DeFaveri *et al.* 2011; Shimada *et al.* 2011; Jones *et al.* 2012a; Jones *et al.* 2012b). S~~ is that selection might be weak, owing to ~~the relatively weak~~weaker environmental contrasts in our study area (i.e., brackish to freshwater) compared to other studies (Table 1). For instance, Raeymaekers *et al.* (2014) found that ~~Raeymaekers et al. (2014) identified only a weak signature of selection in our study area on *Eda*, which is a bi-allelic gene underlying variation in lateral plate number and experiencing selection in many other systems. In line with this we find no evidence for selection on the marker that is linked to *Eda*. Nevertheless, the frequency of the *Eda* allele accounting for low plate number showed a significant increase with distance from the coast,~~

~~and~~ shifts in *Eda* allele frequencies from one generation to the next were ~~linked~~ associated with salinity ~~(Raeymaekers *et al.* 2014)~~, but that at the landscape level salinity did not correlate with *Eda* allele frequencies. This suggests that selection is still acting, but might be too weak to contribute to local adaptation and leave a signature of selection at this gene. ~~Additionally~~In addition, strong gene flow might confound the effect of selection by mixing adapted and non-adapted alleles in the respective populations. Gene flow being moderate to high, we expect this to be ~~one~~ another explanation for why we find so few consistent outlier loci compared to other studies ~~(see Table 1)~~. Accordingly, Raeymaekers *et al.* (2014) found that the spatial distribution of the *Eda* allele frequency correlated with distance to the coast, a proxy for population connectivity. Recurrent contact between freshwater and estuarine or marine populations might lead to the exchange of maladaptive alleles, but at the same time it has been argued that gene flow might cause an opposite effect: efficient flow of advantageous loci (Schluter & Conte 2009; Hohenlohe *et al.* 2012; Bell & Aguirre 2013; Raeymaekers *et al.* 2014), thus facilitating adaptation from standing variation. Whether gene flow in this case fuels or constrains adaptation is hard to say, but the signal of selection might be more difficult to pick up in genome scans due to mixing of the genomic background. ~~Finally, population differentiation among the brackish water populations was very low, which could arguably lead to the conclusion that these are in fact one population. However, this is a feature of many marine populations where connectivity is high and populations structure thus low, and has been shown neither to prevent the detection of genomic regions under selection (Nielsen *et al.* 2009b; DeFaveri *et al.* 2013; Feulner *et al.* 2013; Milano *et al.* 2014; Vandamme *et al.* 2014) nor can it be assumed *a priori* that these populations have not diverged in adaptive loci (Pinho & Hey 2010).~~ We therefore expect that

the low number of outlier loci is rather explained by either weak selection, confounding effects of gene flow or a combination of both and are thus inherent to the system we study.

Genes that have been found in previous studies to be under selection relate to biological functions such as bone formation, osmoregulation, growth, thermal response, maturation, pigmentation, scent detection, spiggin production and morphology (Hohenlohe *et al.* 2010; DeFaveri *et al.* 2011; Shimada *et al.* 2011; Jones *et al.* 2012b). Genes that appeared as outliers in this study include markers linked to osmoregulation (*ATP1A1*), thermal response (*HSPA14*) and growth (*myostatin2*~~*MSTNb*~~). The marker linked to *ATP1A1* has been identified as outlier gene in multiple stickleback saltwater-freshwater transitions worldwide (Jones *et al.* 2006; McCairns & Bernatchez 2010; DeFaveri *et al.* 2011; Shimada *et al.* 2011; Jones *et al.* 2012a; Jones *et al.* 2012b). ~~One of the two markers pinpointed as outliers in more than one detection method~~A major outlier in our study was marker *Stn46*, which is linked to the gene for Rho guanine nucleotide exchange factor 9. It is a member of the gene family coding for rho proteins, a subfamily of the guanine nucleotide exchange factors (GEFs), which are multi-domain proteins involved in the activation of small GTPases (Rossman *et al.* 2005). The Rho family is involved in relaying signals from cell-surface receptors to the actin cytoskeleton and elsewhere (Dvorsky & Ahmadian 2004). Its function can be associated with juvenile growth; in zebrafish it has been specifically linked to angiogenesis (Garnaas *et al.* 2008) and striated muscle and neural development (Raeker *et al.* 2010). Interestingly, locus ~~*marker*~~ *Stn46* was also under selection in populations of the nine-spined stickleback (Shikano *et al.* 2010a), indicating that the gene might be involved in local adaptation in multiple species. ~~The other outlier detected in more than one method is~~ Another major outlier in our study was the marker linked to ~~the gene~~ *myostatin2*~~*MSTNb*~~. ~~Myostatin b,~~This gene, a member of

the transforming growth factor-beta (TGF-beta) family, is known to function as a negative regulator of skeletal muscle development and growth in mammals (Walsh & Celeste 2005) and teleost fish (Radaelli *et al.* 2003). Blocking the expression of myostatin in zebrafish has led to the development of a giant phenotype (Acosta *et al.* 2005), but mysostatin is produced in many other tissues than skeletal muscles and is expected to influence many more functions (Radaelli *et al.* 2003). Understanding why these specific genes are selected requires further study due to the broad range of functions these genes might have. ~~Interestingly, marker Stn46 was also under selection in populations of the nine-spined stickleback (Shikano *et al.* 2010a), indicating that the gene might be involved in local adaptation in multiple species.~~

The low number of outlier loci that we find contrasts with the differentiation across populations in several morphological traits. Local differentiation in the number of lateral plates, for instance, has been shown to significantly exceed the level of neutral differentiation in our study area (Raeymaekers *et al.* 2014). Plate number and other morphological traits such as spine length and gill raker length have reasonably high heritability values exceeding 40 % (Schluter 1996; Peichel *et al.* 2001; Berner *et al.* 2014), suggesting that phenotypes are largely determined by genetic rather than by plastic effects. Yet, theory predicts that only functional loci with a relatively large effect size under strong divergent selection will be able to surpass gene flow (Via 2009; Yeaman & Whitlock 2011). The discrepancy between phenotypic and genomic signatures of selection might therefore become particularly strong for traits that involve many genes of small effect. ~~That the signature of selection on a major effect locus such as the *Eda* gene is not detected in our study is unexpected, but might be partially due to the limited power of outlier detection~~

tests. Arnegard *et al*. (2014) have another possible explanation for the relatively large phenotypic divergence. They show that niche differentiation in sticklebacks, even in early stages of differentiation, can involve many different genes and that gene flow between divergent niche-adapted populations has a bigger impact on the phenotype than just the traits that are directly targeted by selection. It may be attributed to incompatibilities in hybrids that harbor a mix of genes of differentially adapted genotypes. These effects imply that phenotypic changes might not always be adaptive. McCairns & Bernatchez (2010) found indications that freshwater populations might suffer from a loss of plasticity and it might be that epigenetic effects further enhance this discrepancy. For instance, Chaturvedi *et al*. (2014) found that regulation by miRNAs might have make a significant contribution toan important function in the salt to freshwater adaptation in stickleback populations.

When taking the different states of ecological speciation into account, we expect a large range of possible outcomes and degrees of differentiation in both strength of genetic and morphological divergence and associated genes. In the three-spined stickleback, this has already been shown in lake-river habitats (Berner *et al.* 2009; Berner *et al.* 2010). Seehausen et al. (2014) pointed out that studying population differentiation and estimating the effect of gene flow will teach us a lot about the order in which reproductive barriers emerge. Furthermore, patterns of divergence might be substantially different from study systems that experience recurrent gene flow. Other genes might be selected due to divergence hitchhiking (Via *et al.* 2012), resulting in a mosaic pattern of genomic divergence. Genome scans might seem useful to discover these genes, but our study shows that it is challenging to detect the outliers in the first stages of divergence due to the inherent nature of these

~~systems. Yet, others have shown it is possible (Rogers & Bernatchez 2007; Egan *et al.* 2008; Michel *et al.* 2010) and the relevance for understanding the mechanism of ecological speciation is high.~~

Conclusion

~~As expected, w~~We find that weak selection, high levels of gene flow or a combination of both can limit the number of outliers in genome scans. Although genome scans targeting populations across ~~starker~~ strong environmental contrasts are possibly more effective for pinpointing genes that are involved in adaptation, the genes identified by these studies do not necessarily play an important role at every stage of divergence. Many of the genes involved in salt-freshwater transitions might be site-specific or might not be involved ~~in early stages of divergence~~ when gene flow is constantly mixing the gene pool. The genes that we do find are likely to be those with a major effect size and thus an underrepresentation of the total number of genes involved. In addition, phenotypic adaptation is not necessarily genetic, but might be facilitated by plastic and epigenetic effects. It remains a challenging task to find which genes and how many are truly involved ~~at the onset of population divergence~~in local adaptation. We here showed that even with ample gene flow and across weak ecological contrasts, ~~, but we show that these systems reveal~~ interesting insights on the repeatability ~~in the~~ of genomic signatures of ~~adaptation~~ selection ~~across systems, even at the very onset of divergence~~can be obtained.

ACKNOWLEDGEMENTS

We would like to thank A. Chaturvedi, S. Geldof, B. Hellemans, L. Jaquemin, S. Ruyts, N. Thiercelin, I. Verdonck and D. Verheyen for field support and technical assistance. We thank I. Coscia and J. Merilä for insightful and inspiring comments and the latter for collaboration on the whole project. We appreciated the insightful suggestions of two anonymous reviewers. Research was sponsored by the Research Foundation - Flanders (post-doctoral fellowship to J.A.M.R. and research project G.0142.03), by the KU Leuven (GOA/2006/06 "Evolutionary responses to strong selection pressures and their ecological consequences"; PF/2010/07 "Eco- and socio-evolutionary dynamics") and facilitated by the research community W0.037.10N "Eco-evolutionary dynamics in natural and anthropogenic communities" sponsored by the Research Foundation - Flanders. The study received additional support from Academy of Finland grant (129662, 134728 and 218343) to Juha Merilä.

589 **TABLES**

590 **Table 1.** Comparison of microsatellite-based genome scan studies in three-spined stickleback
591 across freshwater - brackish/salt water gradients, including spatial scale (from regional to
592 global), percentage of conservative outliers, F_{ST}, F_{ST} at the *Eda* locus, and F_{ST} at the *ATP1A1*
593 locus.
594

Study	Spatial scale	% outliers	F_{ST}	F_{ST} *Eda*	F_{ST} *ATP1A1*
This study	78 km	2.3	0.059	0.118	0.154
DeFaveri *et al.* 2013	Regional	12.5	0.008	-	0.016
Shimada *et al.* 2011	Continental	8.3	0.107	-	0.225
Mäkinen *et al.* 2008	Continental	3.0	0.166	0.653	-
DeFaveri *et al.* 2011	Global	33	0.119	0.405	-

595

596

597 **Table 2.** Characteristics of the four three-spined stickleback populations used in this study. DTC: distance to coast; MPN: Mean plate
598 number; *Eda*^{L}: frequency of *Eda* low plated allele; AR: allelic richness; H_e: expected heterozygosity; F_{IS}: inbreeding coefficient, with
599 confidence interval (CI 95%).
600

Code	Coordinates (N, E)	Habitat	DTC (km)	Salinity (psu)	MPN	Eda^{L}	AR	H_e	F_{IS} (CI 95%)
L01	51°21'10.66, 3°26'01.83	Creek	3.94	2.04	17.4	0.44	9.47	0.69	0.032 (-0.017 - 0.033)
L02	51°21'56.33, 3°31'11.09	Creek	4.30	1.83	16.8	0.51	9.06	0.69	0.095 (0.048 - 0.096)
L12	51°10'29.03, 3°28'10.45	Stream	22.84	0.32	12.7	0.74	6.80	0.61	0.044 (-0.012 - 0.050)
U01	51°02'48.40, 3°33'01.60	Ditch	36.20	0.52	7.0	0.93	6.51	0.57	0.039 (-0.010 - 0.040)

601 **Table 3.** Pairwise F_{ST} values below the diagonal, and associated P-value above the diagonal. For

602 site codes see Table 2.

Code	L01	L02	L12	U01
L01	-	0.600	< 0.010	< 0.010
L02	0.005	-	< 0.010	< 0.010
L12	0.054	0.056	-	< 0.010
U01	0.094	0.097	0.047	-

603

Table 4. Global outlier loci that are putatively under directional selection detected in four populations of three-spined stickleback. The table includes the locus considered, its F_{ST} -value and either the q value for Bayescan or the P-value for lnRH, LOSITAN and Arlequin. Significance is marked as † = q < 0.1, †† = q < 0.05 and ††† = q < 0.01 or as * = p < 0.05, ** = p < 0.01 and *** = p < 0.001. Marker *Stn380*, linked to the Eda gene, was added as a reference.

Locus	F_{ST}	Bayescan	LnRH	LOSITAN	Arlequin (2 groups)	Arlequin (3 groups)
ATP1A1	0.149			*	*	*
HSPA14	0.159			*	*	*
Ppgm44	0.120	†	***			
Stn74	0.083		*			
Stn177	0.189			**	**	**
Stn46	0.199	†††		**	**	**
Stn380 (Eda)	0.118					

610 **Supplementary Table S1.** Characterisation of microsatellite loci of three-spined stickleback,
611 including name, type of marker, gene assignment and putative function.

Locus	Type of marker	Gene	Putative function
1097PBBE	Random genomic	Novel gene	
1125PBBE	Random genomic	Novel gene	
7033PBBE	Random genomic	Integrin beta-3 precursor	
7080PBBE	Random genomic	Neurensin-1	
ATP1A1	Gene-based	Na+/K+-ATPase alpha-subunit isoform 1	Osmoregulation
ATP4A	Gene-based	H+/K+-ATPase alpha subunit	Osmoregulation
CLCN7	Gene-based	CLC chloride channel isoform 7	Osmoregulation
DCT	Gene-based	L-dopachrome tautomerase precursor	Pigmentation
FGF18	Gene-based	Fibroblast growth factor isoform 18	Growth
FGF8	Gene-based	Fibroblast growth factorisoform 8	Growth
GAest11	Random EST		
GAest16	Random EST	Novel gene	
GAest3	Random EST	Dihydropyrimidinase-related protein 2	
GAest34	Random EST	6-phosphofructokinase type C	
GAest35	Random EST	Rod outer segment membrane protein 1	
GAest41	Random EST	Tumor necrosis factor receptor superfamily member 19 precursor	
GAest47	Random EST		
GAest50	Random EST	Eukaryotic translation initiation factor 3 subunit K	
GAest51	Random EST	Protein kinase C delta-binding protein	
GAest6	Random EST	Ninjurin-2	
GAest67	Random EST	Novel gene	
GAest7	Random EST	Ras-related protein Rab-3A	
GAest82	Random EST	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	
GAest84	Random EST		
GAest87	Random EST		
GS1	Gene-based	Glycogen synthase subtype 1	Growth
HSP70Ab	Gene-based	Heat shock protein 70 kDa subtype 2	Thermal response
HSP70Ad	Gene-based	Heat shock protein 70 kDa subtype 4	Thermal response

IGF-II	Gene-based	Insulin-like growth factor II precursor	Growth
KCNJ4b	Gene-based	Inward rectifier potassium channel subfamily J member 4 subtype 2	Osmoregulation
MYHb	Gene-based	Myosin heavy chain subtype 2	Thermal response
NHE2c	Gene-based	Na+/H+ exchanger subtype 3	Hypoxia
NHE3	Gene-based	Sodium/hydrogen exchanger isoform 3 (SLC9A3)	Osmoregulation
NPY2Rb	Gene-based	Neuropeptide Y receptor Y2 subtype 2	Growth
PER1	Gene-based	Period circadian protein homolog 1	Maturation
Ppgm44	Gene-based	Myostatin subtype 2	Growth
Ppgm17	Gene-based	Type I iodothyronine deiodinase	Morphology
Ppgm16	Gene-based	Cold shock protein 70 kDa subtype 2	Thermal response
Ppgm24	Gene-based	Growth hormone receptor 1	Growth
Ppgm28	Gene-based	General transcription factor IIB	Osmoregulation
Ppgm58	Gene-based	Tristetraprolin	Disease
Ppgm21	Gene-based	Fibroblast growth factor 18	Growth
Ppgm47	Gene-based	Na+/K+/2Cl-cotransporter isoform 1 subtype 2	Osmoregulation
Ppgm35	Gene-based	Heat shock protein 70 kDa isoform beta	Thermal response
Ppgm7	Gene-based	Na+/K+-ATPase alpha-subunit isoform 2	Osmoregulation
Ppgm56	Gene-based	Amine-associated receptor	Taste
Ppgm1	Gene-based	Pituitary adenylate cyclase activating polypeptide receptor sub-type 1	Growth
Ppgm2	Gene-based	Pituitary adenylate cyclase activating polypeptide receptor sub-type 2	Growth
Ppgm49	Gene-based	Neuropeptide Y precursor	Growth
Ppgm30	Gene-based	Heat shock protein 25 kDa	Thermal response
Ppgm33	Gene-based	Heat shock protein 70 kDa isoform alpha subtype 2	Thermal response
Ppgm50	Gene-based	Parvalbumin subtype 2	Thermal response
Ppgm6	Gene-based	Na+/K+-ATPase alpha-subunit isoform 1	Osmoregulation
Ppgm38	Gene-based	Insulin-like growth factor I	Growth
PRL-R	Gene-based	Prolactin receptor a	Osmoregulation
PVALBa	Gene-based	Parvalbumin subtype 1	Growth
RhBG	Gene-based	Rhesus glycoprotein isoform 3	Toxic stress
SLC4A10a	Gene-based	Electroneutral Na-coupled HCO2 transporter isoform 10 subtype 1	Hypoxia
Stn102	Random genomic	EF-hand domain-containing family member A2	
Stn108	Random genomic	Bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 4	

Stn119	Random EST		
Stn122	Random EST		
Stn124	Random EST	Brain-specific angiogenesis inhibitor 2 precursor	
Stn163	Random EST	Novel gene	
Stn164	Random EST		
Stn167	Random genomic		
Stn177	Random genomic		
Stn194	Random genomic	Nuclear-localized factor 1	
Stn20	Random EST		
Stn222	Random genomic	Ras-related protein Rab-18	
Stn223	Random genomic	Zinc finger homeobox protein 4	
Stn235	Random EST		
Stn240	Random EST		
Stn257	Random EST	Solute carrier family 35 member F2	
Stn299	Random genomic		
Stn302	Random genomic		
Stn315	Random genomic	Histone deacetylase 4	
Stn328	Random genomic	RING finger protein 220	
Stn34	Random EST		
Stn380	Gene-based	Eda	plate nr
Stn46	Random genomic	Rho guanine nucleotide exchange factor 9	
Stn49	Random genomic	Dedicator of cytokinesis protein 4	
Stn71	Random EST	Novel gene	
Stn74	Random EST		
Stn79	Random genomic	Reticulon-2	
Stn89	Random EST	Rho GTPase-activating protein SYDE2	
HPX	Gene-based	Hemopexin	Immune response

612

613

614 **Supplementary Table S2.** Global outlier loci that are putatively under balancing selection as
615 detected with the methods of Bayescan, LOSITAN and Arlequin. Significance is marked as † = q <
616 0.1, †† = q < 0.05 and ††† = q < 0.01 or * = p < 0.05 , ** = p < 0.01 and *** = p < 0.001.

Locus	F_{ST}	Bayescan	LOSITAN	Arlequin (2 groups)	Arlequin (3 groups)
FGF8	0.022		*		
GAest3	0.021		*	*	
GAest34	0.015	††	***	*	*
GAest47	0.009		*	*	*
IGF-II	-0.007		*		
NHE2c	0.027		*	*	
NHE3	0.004		*	*	*
Ppgm16	-0.009		*		
Ppgm2	0.005			**	*
Ppgm30	0.013		*	*	*
Ppgm6	0.005			*	*
Stn122	0.014		***	**	*
Stn20	0.017		*		
Stn223	0.019	†	**	*	*
Stn79	-0.020		***	*	*
Stn89	0.028	†††			

617

REFERENCES

Acosta J, Carpio Y, Borroto I, González O, Estrada MP (2005) Myostatin gene silenced by RNAi show a zebrafish giant phenotype. *Journal of Biotechnology* **119**, 324-331.

Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* **9**, 323.

Arnegard, M. E., McGee, M. D., Matthews, B., Marchinko, K. B., Conte, G. L., Kabir, S., Bedford, N., Bergek, S., Chan, Y. F., Jones, F. C., Kingsley, D. M., Peichel, C. L. & Schluter, D. 2014. Genetics of ecological divergence during speciation. *Nature* **511**: 307-311.

~~Barrett, R. D. H., Rogers, S. M. & Schluter, D. 2009. Environment specific pleiotropy facilitates divergence at the Ectodysplasin locus in threespine stickleback. *Evolution* **63**: 2831-2837.~~

Barrick JE, Lenski RE (2013) Genome dynamics during experimental evolution. *Nature Reviews: Genetics* **14**, 827-839.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Becks L, Ellner SP, Jones LE, Hairston NG (2012) The functional genomics of an eco-evolutionary feedback loop: linking gene expression, trait evolution, and community dynamics. *Ecology Letters* **15**, 492-501.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations, pp. 1996-2004. Laboratoire génome, populations, interactions, CNRS UMR.

Bell MA (1982) Differentiation of adjacent stream populations of threespine sticklebacks. *Evolution* **36**, 189-199.

Bell MA, Aguirre WE (2013) Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. *Evolutionary Ecology Research* **15**, 377-411.

Bernatchez L, Renaut S, Whiteley AR*, et al.* (2010) On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **365**, 1783-1800.

~~Berner D, Grandchamp A-C, Hendry AP (2009) Variable Progress toward Ecological Speciation in Parapatry: Stickleback across Eight Lake-Stream Transitions. *Evolution* **63**, 1740-1753.~~

Berner D, Moser D, Roesti M, Buescher H, Salzburger W (2014) Genetic architecture of skeletal evolution in european lake and stream stickleback. *Evolution* **68**, 1792-1805.

~~Berner D, Roesti M, Hendry AP, Salzburger W (2010) Constraints on speciation suggested by comparing lake-stream stickleback divergence across two continents. *Molecular Ecology* **19**, 4963-4978.~~

Bierne N, Roze D, Welch JJ (2013) Pervasive selection or is it…? why are FST outliers sometimes so frequent? *Molecular Ecology* **22**, 2061-2064.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**, 2044-2072.

Boughman JW (2007) Speciation in sticklebacks. In: *Biology of the Three-Spined Stickleback* (eds. Östlund-Nilsson S, Mayer I, Huntingford F), pp. 83–126. CRC Press., Boca Raton, FL.

Butlin RK, Saura M, Charrier G*, et al.* (2014) Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution* **68**, 935-949.

Chaturvedi A, Raeymaekers JAM, Volckaert FAM (2014) Computational identification of miRNAs, their targets and functions in three-spined stickleback (*Gasterosteus aculeatus*). *Molecular Ecology Resources* **14**, 768-777.

Colosimo PF, Hosemann KE, Balabhadra S*, et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**, 1928-1933.

Coscia I, Vogiatzi E, Kotoulas G, Tsigenopoulos C, Mariani S (2011) Exploring neutral and adaptive processes in expanding populations of gilthead sea bream, Sparus aurata L., in the North-East Atlantic. *Heredity* **108**, 537-546.

de Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology* **23**, 2006-2019.

Deagle BE, Jones FC, Chan YF*, et al.* (2012) Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **279**, 1277-1286.

DeFaveri J, Jonsson PR, Merilä J (2013) Heterogeneous genomic differentiation in marine threespine sticklebacks: Adaptation along an environmental gradient. *Evolution* **67**, 2530-2546.

DeFaveri J, Merilä J (2013) Evidence for adaptive phenotypic differentiation in Baltic Sea sticklebacks. *Journal of Evolutionary Biology* **26**, 1700-1715.

DeFaveri J, Shikano T, Shimada Y, Goto A, Merilä J (2011) Global analysis of genes involved in freshwater adaptation in threespine stickleback (*Gasterosteus aculeatus*). *Evolution* **65**, 1800-1807.

Dvorsky R, Ahmadian MR (2004) Always look on the bright site of Rho: structural implications for a conserved intermolecular interface. *EMBO reports* **5**, 1130-1136.

~~Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of neochlamisus bebbianae leaf beetles. *Evolution* **62**, 1162-1181.~~

Endler JA (1973) Gene flow and population differentiation: Studies of clines suggest that differentiation along environmental gradients may be independent of gene flow. *Science* **179**, 243-250.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285-298.

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

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**, 342-350.

Feulner PGD, Chain FJJ, Panchal M*, et al.* (2013) Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Molecular Ecology* **22**, 635-649.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993.

Foster SA, Scott RJ, Cresko WA (1998) Nested biological variation and speciation. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **353**, 207-218.

Garnaas MK, Moodie KL, Liu M-l*, et al.* (2008) Syx, a RhoA guanine exchange factor, is essential for angiogenesis in vivo. *Circulation Research* **103**, 710-716.

Hagen DW (1967) Isolating mechanisms in threespine sticklebacks (*Gasterosteus*). *Journal of the Fisheries Research Board of Canada* **24**, 1637-&.

Hansen MM, Meier K, Mensberg K-LD (2010) Identifying footprints of selection in stocked brown trout populations: a spatio-temporal approach. *Molecular Ecology* **19**, 1787-1800.

Hansen MM, Ruzzante DE, Nielsen EE, Bekkevold D, Mensberg K-LD (2002) Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (Salmo trutta) populations. *Molecular Ecology* **11**, 2523-2535.

Hendry AP (2009) Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences* **66**, 1383-1398.

Hendry AP, Bolnick DI, Berner D, Peichel CL (2009) Along the speciation continuum in sticklebacks. *Journal of Fish Biology* **75**, 2000-2036.

Heuts MJ (1947) The phenotypical variability of *Gasterosteus aculeatus* L. populations in Belgium. *Verhandelingen van de Koninklijke Vlaamse Academie voor Wetenschappen, Letteren en Schone Kunsten van België* **9**, 1-63.

Hohenlohe PA, Bassham S, Currey M, Cresko WA (2012) Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **367**, 395-408.

Hohenlohe PA, Bassham S, Etter PD*, et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**.

Jones F, Brown C, Pemberton J, Braithwaite V (2006) Reproductive isolation in a threespine stickleback hybrid zone. *Journal of Evolutionary Biology* **19**, 1531-1544.

Jones FC, Chan YF, Schmutz J*, et al.* (2012a) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology* **22**, 83-90.

Jones FC, Grabherr MG, Chan YF*, et al.* (2012b) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61.

Kauer MO, Dieringer D, Schlotterer C (2003) A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of Drosophila melanogaster. *Genetics* **165**, 1137-1148.

Levene H (1953) Genetic equilibrium when more than one ecological niche is available. *American Naturalist* **87**, 331-333.

Lind EE, Grahn M (2011) Directional genetic selection by pulp mill effluent on multiple natural populations of three-spined stickleback (*Gasterosteus aculeatus*). *Ecotoxicology* **20**, 503-512.

Lucek K, Sivasundar A, Kristjánsson BK, Skúlason S, Seehausen O (2014) Quick divergence but slow convergence during ecotype formation in lake and stream stickleback pairs of variable age. *Journal of Evolutionary Biology* **27**, 1878–1892.

Mäkinen HS, Cano JM, Merilä J (2008) Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. *Molecular Ecology* **17**, 3565-3582.

McCairns RJS, Bernatchez L (2010) Adaptive divergence between freshwater and marine sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. *Evolution* **64**, 1029-1047.

McKinnon JS, Rundle HD (2002) Speciation in nature: the threespine stickleback model systems. *Trends in Ecology & Evolution* **17**, 480-488.

McPhail J (1994) Speciation and the evolution of reproductive isolation in the sticklebacks (Gasterosteus) of south-western British Columbia. In: *The evolutionary biology of the threespine stickleback* (eds. Bell AM, Foster S), pp. 399-437. Oxford University Press Inc., New York.

~~Michel AP, Sim S, Powell THQ, *et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **107**, 9724-9729.~~

~~Milano I, Babbucci M, Cariani A, *et al.* (2014) Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (Merluccius merluccius). *Molecular Ecology* **23**, 118-135.~~

Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature* **441**, 947-952.

Moore J-S, Gow JL, Taylor EB, Hendry AP (2007) Quantifying the constraining influence of gene flow on adaptive divergence in the lake-stream threespine stickleback System. *Evolution* **61**, 2015-2026.

Narum SR, Hess JE (2011) Comparison of F-ST outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, 184-194.

Nielsen E, Hemmer-Hansen J, Poulsen N*, et al.* (2009~~a~~) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**, 276.

~~Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009b) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology* **18**, 3128-3150.~~

Nosil P (2009) Adaptive population divergence in cryptic color-pattern following a reduction in gene flow. *Evolution* **63**, 1902-1912.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**, 375-402.

Olafsdottir GA, Snorrason SS (2009) Parallels, nonparallels, and plasticity in population differentiation of threespine stickleback within a lake. *Biological Journal of the Linnean Society* **98**, 803-813.

Orsini L, Spanier KI, De Meester LUC (2012) Genomic signature of natural and anthropogenic stress in wild populations of the waterflea Daphnia magna: validation in space, time and experimental evolution. *Molecular Ecology* **21**, 2160-2175.

Papa R, Martin A, Reed RD (2008) Genomic hotspots of adaptation in butterfly wing pattern evolution. *Current Opinion in Genetics & Development* **18**, 559-564.

Peichel CL, Nereng KS, Ohgi KA*, et al.* (2001) The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901-905.

Perrier C, Bourret V, Kent MP, Bernatchez L (2013) Parallel and nonparallel genome-wide divergence among replicate population pairs of freshwater and anadromous Atlantic salmon. *Molecular Ecology* **22**, 5577-5593.

Pinho C, Hey J (2010) Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution and Systematics* **41**, 215-230.

Poncet BN, Herrmann D, Gugerli F*, et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Molecular Ecology* **19**, 2896-2907.

Radaelli G, Rowlerson A, Mascarello F, Patruno M, Funkenstein B (2003) Myostatin precursor is present in several tissues in teleost fish: a comparative immunolocalization study. *Cell and Tissue Research* **311**, 239-250.

Raeker MÖ, Bieniek AN, Ryan AS*, et al.* (2010) Targeted deletion of the zebrafish obscurin A RhoGEF domain affects heart, skeletal muscle and brain development. *Developmental Biology* **337**, 432-443.

Raeymaekers JAM, Konijnendijk N, Larmuseau MHD*, et al.* (2014) A gene with major phenotypic effects as a target for selection vs. homogenizing gene flow. *Molecular Ecology* **23**, 162-181.

Raeymaekers JAM, Lens LUC, Van den Broeck F, Van Dongen S, Volckaert FAM (2012) Quantifying population structure on short timescales. *Molecular Ecology* **21**, 3458-3473.

Raeymaekers JAM, Maes GE, Audenaert E, Volckaert FAM (2005) Detecting Holocene divergence in the anadromous-freshwater three-spined stickleback (*Gasterosteus aculeatus*) system. *Molecular Ecology* **14**, 1001-1014.

Raeymaekers JAM, Van Houdt JKJ, Larmuseau MHD, Geldof S, Volckaert FAM (2007) Divergent selection as revealed by P_{ST} and QTL-based F_{ST} in three-spined stickleback (*Gasterosteus aculeatus*) populations along a coastal-inland gradient. *Molecular Ecology* **16**, 891-905.

Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology* **21**, 2852-2862.

~~Rogers S, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (Coregonus sp. Salmonidae) species pairs. *Molecular Biology and Evolution* **24**, 1423-1438.~~

Rohlf FJ (2002) tpsDig, digitize landmarks and outlines. Department of Ecology and Evolution, State University of New York at Stony Brook., New York.

Rossman KL, Der CJ, Sondek J (2005) GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. *Nature Reviews: Molecular Cell Biology* **6**, 167-180.
Schluter D (1996) Adaptive radiation along genetic lines of least resistance. *Evolution*, 1766-1774.
Schluter D (2009) Evidence for Ecological Speciation and Its Alternative. *Science* **323**, 737-741.
Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences, USA* **106**, 9955-9962.
Seehausen O, Butlin RK, Keller I*, et al.* (2014) Genomics and the origin of species. *Nature Reviews: Genetics* **15**, 176-192.
Shikano T, Ramadevi J, Merilä J (2010a) Identification of local-and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *Molecular Biology and Evolution* **27**, 2775-2789.
Shikano T, Ramadevi J, Shimada Y, Merilä J (2010b) Utility of sequenced genomes for microsatellite marker development in non-model organisms: a case study of functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *BMC Genomics* **11**, 334.
Shimada Y, Shikano T, Merilä J (2011) A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Molecular Biology and Evolution* **28**, 181-193.
Slatkin (1987). Gene flow and the geographic structure of natural populations. *Science* **236**,787–792.
Soria-Carrasco V, Gompert Z, Comeault AA*, et al.* (2014) Stick insect genomes reveal natural selection's role in parallel speciation. *Science* **344**, 738-742.
Stapley J, Reger J, Feulner PGD*, et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution* **25**, 705-712.
Storz JF (2005) Invited review: Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.
Strasburg JL, Sherman NA, Wright KM*, et al.* (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **367**, 364-373.
Taylor WR, Dyke GCV (1985) Revised procedures for staining and clearing small fishes and other vertebrates for bone and cartilage study. *Cybium* **9**, 107-119.
Tregenza T (2002) Divergence and reproductive isolation in the early stages of speciation. In: *Genetics of Mate Choice: From Sexual Selection to Sexual Isolation* (eds. Etges WJ, Noor MAF), pp. 291-300. Springer Netherlands.
Van Dongen S, Lens L, Pape E, Volckaert FAM, Raeymaekers JAM (2009) Evolutionary history shapes the association between developmental instability and population-level genetic variation in three-spined sticklebacks. *Journal of Evolutionary Biology* **22**, 1695-1707.
van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology notes* **4**, 535-538.
Vandamme S, Maes G, Raeymaekers J*, et al.* (2014) Regional environmental pressure influences population differentiation in turbot (Scophthalmus maximus). *Molecular Ecology* **23**, 618-636.
Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences, USA* **106**, 9939-9946.
~~Via S, Conte G, Mason-Foley C, Mills K (2012) Localizing FST outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation with gene flow. *Molecular Ecology* **21**, 5546-5560.~~
Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *Plos Biology* **4**, e72.
Walsh F, Celeste A (2005) Myostatin: a modulator of skeletal-muscle stem cells. *Biochemical Society Transactions* **33**, 1513-1517.

853 Wootton RJ (1976) *The Biology of the Sticklebacks.* Academic Press, London.
854 Yang L, Jelsbak L, Marvig RL*, et al.* (2011) Evolutionary dynamics of bacteria in a human host
855 environment. *Proceedings of the National Academy of Sciences, USA* **108**, 7481-7486.
856 Yeaman S, Otto SP (2011) Establishment and maintenance of adaptive genetic divergence under
857 migration, selection, and drift. *Evolution* **65**, 2123-2129.
858 Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration–selection
859 balance. *Evolution* **65**, 1897-1911.
860 Zulliger D, Schnyder E, Gugerli F (2013) Are adaptive loci transferable across genomes of related species?
861 Outlier and environmental association analyses in Alpine Brassicaceae species. *Molecular*
862 *Ecology* **22**, 1626-1639.

863 **FIGURE LEGENDS**

864

865 **Figure 1.** Map with the four sampling locations of three-spined stickleback in northwestern Belgium and

866 southwestern Netherlands. River network and major towns are mapped.

867

868 **Figure 2.** Average values for A) standard length, B) plate number, C) relative pelvic spine length, D)

869 relative pelvic plate length, and E) relative length of the second large gill raker in two brackish water

870 populations (L01, L02) and two freshwater populations (L12 and U01)) of three-spined stickleback.

871 Horizontal bars represent 95% confidence intervals.

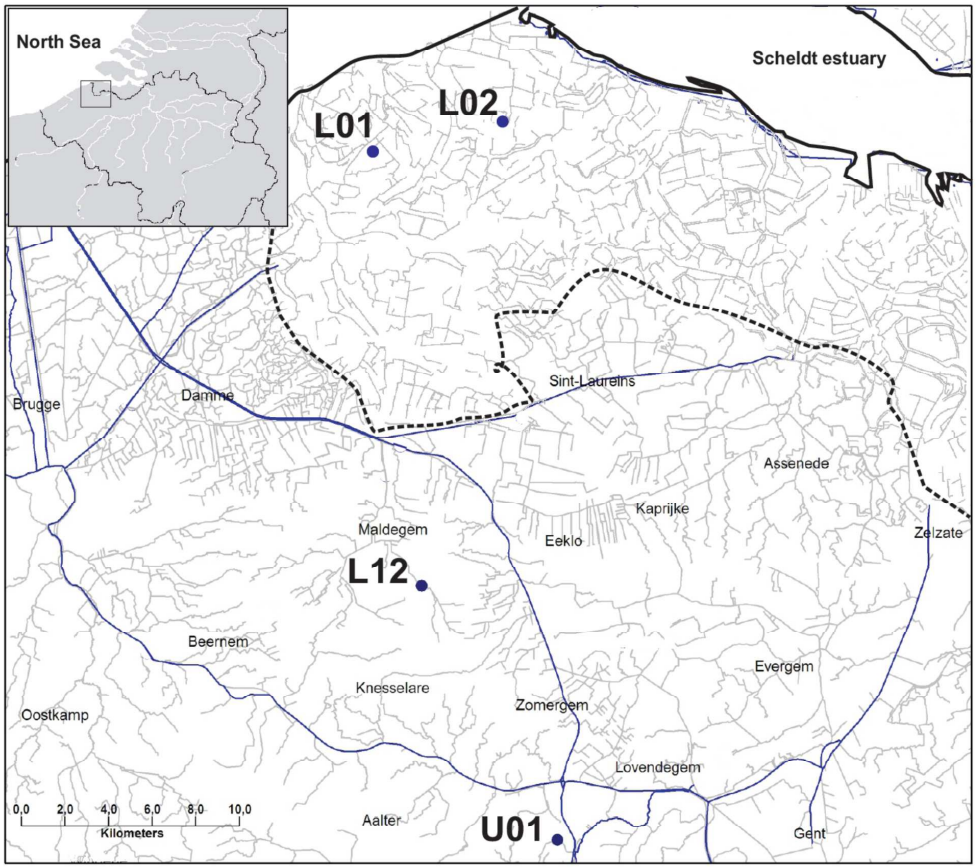
872

873 **FIGURES**

874

875 **Figure 1**

876

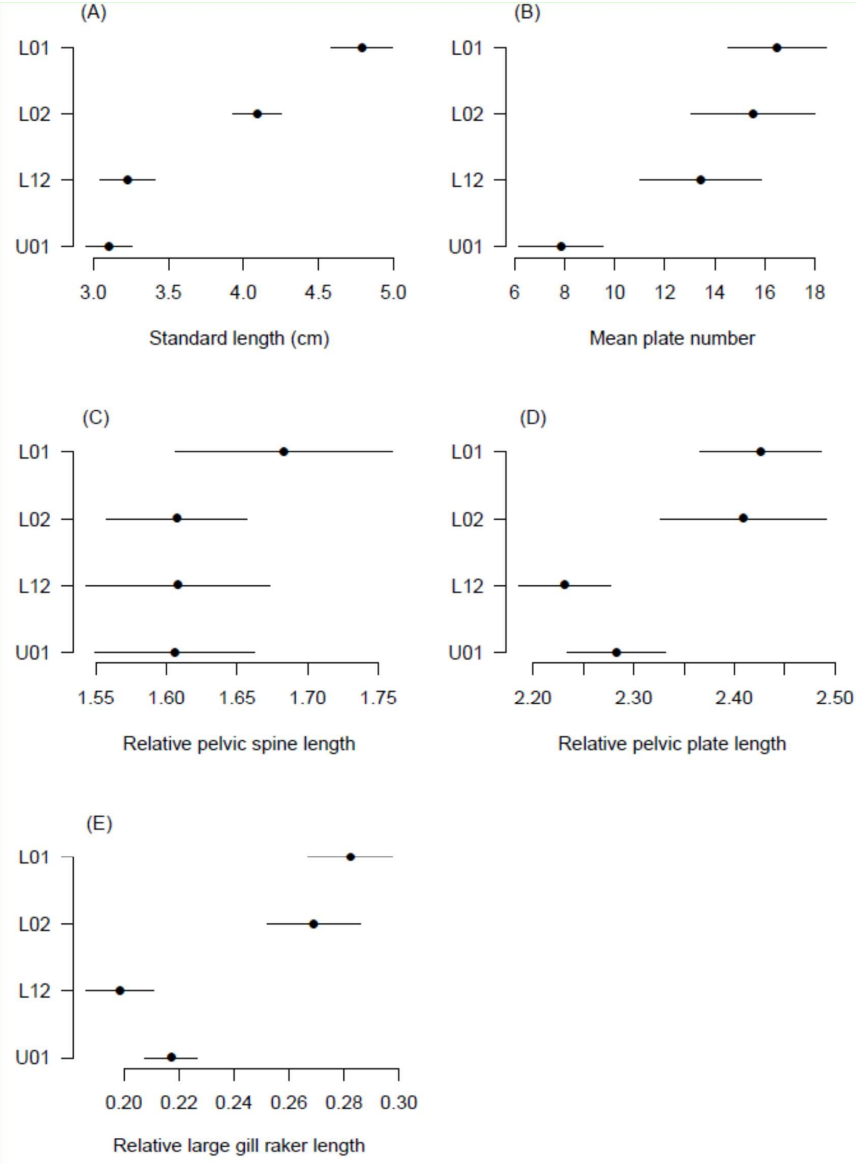


877

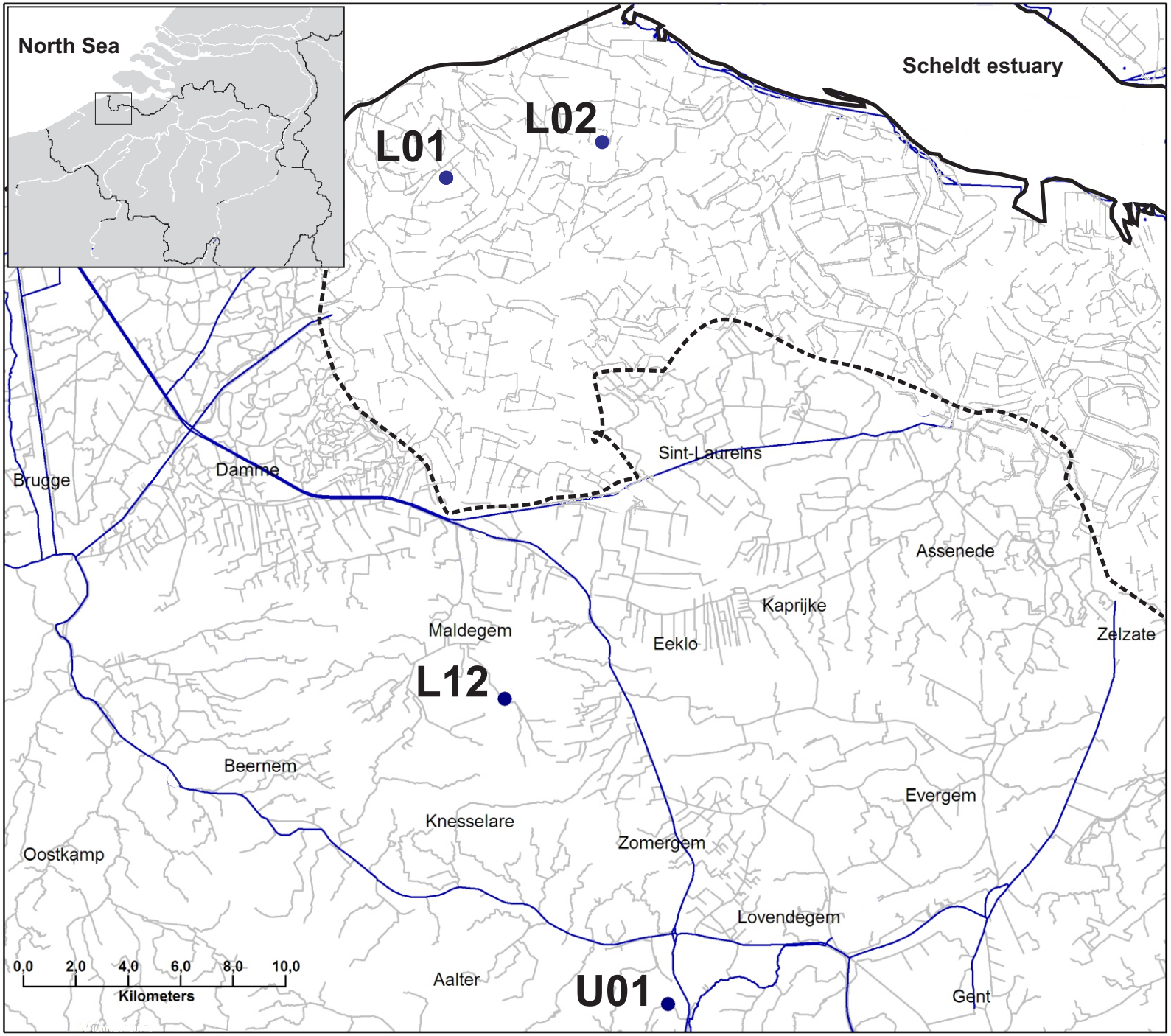
878

879

880 **Figure 2**
881



882
883
884
885


North Sea
Scheldt estuary
L01
L02
L12
U01
Brugge
Damme
Sint-Laureins
Assenede
Kaprijke
Eeklo
Maldegem
Zelzate
Beernem
Knesselare
Evergem
Zomergem
Oostkamp
Lovendegem
Aalter
Gent
0,0 2,0 4,0 6,0 8,0 10,0
Kilometers