Targeted resequencing reveals geographical patterns of differentiation for loci implicated in parallel evolution

ANJA M. WESTRAM,* MARINA PANOVA,† JUAN GALINDO‡ § and ROGER K. BUTLIN*†
*Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK, †Department of Marine Sciences – Tjärnö, University of Gothenburg, SE-452 96 Strömstad, Sweden, ‡Departamento de Bioquímica, Genética e Inmunología, Facultad de Biología, Universidade de Vigo, 36310 Vigo, Spain, §ECIMAT, Estación de Ciencias Mariñas de Toralla, Universidade de Vigo, Illa de Toralla, 36331 Vigo, Spain

Abstract

Parallel divergence and speciation provide evidence for the role of divergent selection in generating biological diversity. Recent studies indicate that parallel phenotypic divergence may not have the same genetic basis in different geographical locations – 'outlier loci' (loci potentially affected by divergent selection) are often not shared among parallel instances of phenotypic divergence. However, limited sharing may be due, in part, to technical issues if false-positive outliers occur. Here, we test this idea in the marine snail *Littorina saxatilis*, which has evolved two partly isolated ecotypes (adapted to crab predation vs. wave action) in multiple locations independently. We argue that if the low extent of sharing observed in earlier studies in this system is due to sampling effects, we expect outliers not to show elevated F_{ST} when sequenced in new samples from the original locations and also not to follow predictable geographical patterns of elevated F_{ST}. Following a hierarchical sampling design (within vs. between country), we applied capture sequencing, targeting outliers from earlier studies and control loci. We found that outliers again showed elevated levels of F_{ST} in their original location, suggesting they were not generated by sampling effects. Outliers were also likely to show increased F_{ST} in geographically close locations, which may be explained by higher levels of gene flow or shared ancestral genetic variation compared with more distant locations. However, in contrast to earlier findings, we also found some outlier types to show elevated F_{ST} in geographically distant locations. We discuss possible explanations for this unexpected result.

Keywords: adaptive divergence, capture sequencing, genome scan, parallel evolution, speciation

Received 9 July 2015; revision received 8 March 2016; accepted 23 March 2016

Introduction

Parallel phenotypic divergence occurs where the same traits diverge repeatedly in different geographical locations, often associated with similar environmental transitions. Parallel divergence coupled with the repeated evolution of reproductive isolation provides fascinating evidence for the role of selection as a driving force in speciation (Schluter & Nagel 1995; Elmer & Meyer 2011): When similar patterns of divergence occur repeatedly, this is more likely to be driven by a directional process than by chance.

With the availability of genomic tools, researchers have now begun to focus on the genetic basis of phenotypic divergence and speciation (Seehausen *et al.* 2014). For example, genome scans can identify genomic regions potentially affected by divergent selection (directly or via linkage disequilibrium) (Luikart *et al.* 2003; Butlin 2010). These methods utilize data from large numbers of genomic markers to identify 'outlier' genomic regions showing increased differentiation (typically measured as F_{ST}) compared with the rest of the genome (Lewontin & Krakauer 1973; Beaumont

Correspondence: Anja M. Westram, Fax: +44 (114) 222 0002; E-mail: a.westram@sheffield.ac.uk

2005). Outliers have been identified in multiple different study systems, including some where parallel phenotypic evolution has been described (e.g. Nosil *et al.* 2009). These systems are particularly interesting because they allow us to ask whether the same alleles, genes or genetic pathways underlie parallel phenotypic patterns of divergence. Answering this question will help in understanding the repeatability of adaptation at the molecular level (Elmer & Meyer 2011; Conte *et al.* 2012). The extent of 'sharing' of outliers among instances of parallel evolution should be indicative of the extent to which the genetic basis of divergence is the same. Studies using candidate genes or QTL mapping approaches often find reuse of the same genes in adaptation to similar selection pressures, especially among closely related species (Conte *et al.* 2012). However, these analyses may be biased towards genes of large effect (Rockman 2012) and candidate genes are generally more likely to appear as shared (Conte *et al.* 2012). In contrast to these results, recent genome scans focusing on populations within species have shown that parallel phenotypic divergence may not necessarily be accompanied by similar genetic patterns: that is, outlier loci may not be the same across different locations, even if they are geographically close (Deagle *et al.* 2012; Kautt *et al.* 2012; Gagnaire *et al.* 2013; Perrier *et al.* 2013; Roda *et al.* 2013; Soria-Carrasco *et al.* 2014; Westram *et al.* 2014).

The extent of sharing detected in genome scans may be influenced by both biological/evolutionary explanations and technical artefacts (Narum & Hess 2011). To understand the genetic basis of (parallel) divergence, we need to tease apart these alternatives and improve the technical aspects of the methodology to avoid false positives.

A similar genetic basis of divergence is expected especially when there is a strong connection by gene flow that may transport adaptive alleles (e.g. evolution in concert; Johannesson *et al.* 2010), or when there is a lot of shared standing genetic variation among instances of parallel divergence (e.g. Hohenlohe *et al.* 2010). In these cases, shared outliers should be identifiable in genome scans, although this method may sometimes fail when patterns of linkage disequilibrium between marker loci and the actual targets of selection differ among populations. On the other hand, low outlier sharing may be expected when the genetic variation available for adaptation is substantially different among locations. In this scenario, solutions for the same problem evolve independently and might therefore be based on different novel mutations, or on different components of old variation that have been maintained only locally. Such a pattern may be more likely for polygenic traits, which allow for a larger number of potential 'genetic solutions'. In this case, the phenotypic outcome

(i.e. adaptation) may be similar, and outliers might still belong to the same metabolic pathways (Roda *et al.* 2013), but the markers appearing as outliers may be different. Such independent evolution is expected when instances of divergence are separated by long evolutionary distances (reducing the amount of shared standing genetic variation) and/or geographical distances (less connected by gene flow) (e.g. Conte *et al.* 2012). If the limited extent of outlier sharing often observed in recent studies is mainly due to these evolutionary reasons (rather than technical explanations), we would therefore expect sharing to increase when sampling geographically closer locations (unless adaptation is completely based on local novel mutation even on small geographical scales) (e.g. Roda *et al.* 2013).

As an alternative (but not mutually exclusive) explanation, various technical issues may affect outlier scans and therefore the observed extent of sharing of outliers. Whether a marker appears as an outlier reflects not only the effect of divergent selection, but also potentially several other factors. Regions of low intrapopulation diversity may be identified as outliers even if they do not show increased divergence between populations (Noor & Bennett 2009; Cruickshank & Hahn 2014). Features of the genome may also have an influence on F_{ST}. For example, genomic regions of low recombination rate (e.g. near centromeres) tend to show increased F_{ST} estimates (e.g. Roesti *et al.* 2012; Renaut *et al.* 2013), because in these regions the effects of both divergent and purifying selection spread further along the chromosome. If local recombination rate is a conserved feature, this may lead to sharing of outlier loci that is not necessarily caused by divergent selection alone.

Other factors may generate spurious outliers which are not expected to be shared among populations, and which could therefore decrease the observed amount of sharing and falsely indicate large differences in the genetic basis of parallel divergence. A proportion of outliers will often represent false positives, while some loci influenced by divergent selection will be missed. False positives may occur due to violations of the neutral model applied in outlier scans, for example deviations from the assumed population structure or demographic history (Foll & Gaggiotti 2008; Excoffier *et al.* 2009). Additionally, false positives may simply occur due to sampling noise (i.e. allele frequencies in the sampled individuals do not reflect population allele frequencies). The probability of misrepresentation of population allele frequencies increases with smaller sample sizes. Effects of sampling noise may also be increased when intrapopulation genetic diversity is spatially structured and sampling is limited to a few spots (as may often be the case for practical reasons). Furthermore, some studies pool DNA/RNA from multiple

individuals of the same population (e.g. Roda *et al.* 2013; Westram *et al.* 2014), further reducing the confidence in the data. A large number of false positives represent a possible explanation for a low extent of observed outlier sharing because such outliers may not be expected to be the same in different locations.

Following from the above considerations, we predict the following patterns: (i) if a low observed amount of outlier sharing among distant locations is mainly due to the biological explanations outlined above, outliers may still show elevated levels of differentiation in geographically close populations (where there is more potential for gene flow and a shared evolutionary history). (ii) If low sharing is mainly due to false positives caused by sampling noise, outliers should not be repeatable (i.e. when taking new samples from the same locations, outliers should often not show elevated levels of differentiation again).

Here, we test for these patterns in a system where parallel divergence is well studied, the intertidal snail *Littorina saxatilis.* This ovoviviparous periwinkle has evolved morphologically distinct ecotypes, partly separated by reproductive barriers, in multiple geographical locations including Spain, Sweden and the UK (Rolán-Alvarez 2007; Johannesson *et al.* 2010; Butlin *et al.* 2014). A model of repeated divergence, within and between countries, fits genetic data better than a single origin of the ecotypes (Butlin *et al.* 2014). One ecotype ('crab ecotype') is adapted to predation and occurs in crab-occupied habitats (e.g. boulder fields), while the other one occurs in crab-free environments (e.g. cliffs) and is adapted to exposure to wave action ('wave ecotype'). Parallel phenotypic divergence involves multiple traits, including body size (crab ecotype larger), shell thickness (thicker shell in crab ecotype) and behaviour (crab ecotype more wary) (reviewed in Johannesson *et al.* 2010). However, there are also nonparallel patterns of divergence, for example with regard to ridging of the shell (Johannesson *et al.* 2010). In addition to divergent natural selection, habitat choice and assortative mating contribute to reproductive isolation between the ecotypes (Johannesson *et al.* 1995, 2010; Rolán-Alvarez 2007; Webster *et al.* 2012). Multiple studies have analysed genetic divergence between ecotypes in this system, finding that gene flow between ecotypes is restricted across the genome (Grahame *et al.* 2006; Panova *et al.* 2006; Galindo *et al.* 2009) and detecting outliers with differentiation increased above a low background level of differentiation (Wilding *et al.* 2001; Galindo *et al.* 2010; Westram *et al.* 2014; Ravinet *et al.* 2016).

Because population sizes are large, and some populations share a recent postglacial history (Swedish and UK populations) (Butlin *et al.* 2014), it seems feasible

that parallel adaptation involves a partly shared genetic basis, as observed in other systems (e.g. sticklebacks; Hohenlohe *et al.* 2010; Jones *et al.* 2012). Most outlier studies so far have been limited to specific geographical regions (Spain, Galindo *et al.* 2009; Sweden, Hollander *et al.* 2015; UK, Wilding *et al.* 2001; Grahame *et al.* 2006; Galindo *et al.* 2010). However, a recent transcriptome scan comparing one location from each of Sweden, Spain and the UK found relatively low sharing of outliers – approximately 20% of outliers were shared among pairs of countries (Westram *et al.* 2014). An outlier scan based on AFLP (amplified fragment length polymorphism) markers found <10% sharing on a similar geographical scale (Butlin *et al.* 2014). A recent study using RAD sequencing suggests that sharing among islands within the same archipelago in Sweden may be similarly low (Ravinet *et al.* 2016).

In this study, we aim to answer two main questions: (i) Do outlier loci still show increased differentiation when they are sequenced in new samples from the locations of original outlier detection? In order to gain more reliable data, we perform more extensive sequencing and obtain individual sequencing data (rather than pooled, as in Galindo *et al.* 2010; Westram *et al.* 2014). (ii) When using a hierarchical sampling design, including geographically close (within-country) and distant (between-country) samples, do outliers show elevated F_{ST} levels in geographically closer locations?

We use a novel approach relying on resequencing of outliers detected in earlier work (see above) as well as control loci, rather than performing a *de novo* genome scan (which requires the sequencing of a large number of random loci across the genome). Thus, instead of detecting novel outliers, we ask whether known outliers show elevated F_{ST} in various geographical locations. All loci were sequenced in snails from six locations, which contained (a) locations where the outliers were first detected (original locations), (b) a second (new) location in each country where outliers were first detected and (c) locations in other countries. This sampling design allowed us to answer both of our main questions.

As one specific objective was to test the confidence in previously identified outliers, we aimed at increasing the amount of sequence information per locus, rather than including a large number of loci. Towards that aim, we used a capture sequencing approach (e.g. Nadeau *et al.* 2012; Smadja *et al.* 2012; Hebert *et al.* 2013), which allows for the generation of longer sequences for regions for which some sequence information is already available. In this approach, genomic DNA libraries are generated and hybridized with probes targeting regions of interest, and only targeted fragments are retained and sequenced. Because the library fragments are typically longer than the probes,

sequence information upstream and downstream of the probe region can be obtained.

Methods

Selection of outlier and control loci from previous studies

Various previous studies have identified outliers between ecotypes, using different methods. We included outliers from an RNAseq transcriptome scan (Westram *et al.* 2014), a 454 transcriptome scan (Galindo *et al.* 2010) and a RADseq genome scan (Ravinet *et al.* 2016) in the current work. These outlier types are referred to as 'RNAseq', '454' and 'RAD' outliers in the following (Table 1; Fig. 1). We also included some loci that showed gene expression differences in a microarray study but for which we did not have information about sequence divergence (M. Panova, unpublished). While we hypothesized that most transcriptome and genome scan outliers will show elevated F_{ST} in their original location if outlier detection is reliable, this hypothesis does not apply for the genes showing expression differences. If gene expression is controlled by genomic regions distant from the gene itself, we would not expect these loci to show elevated F_{ST} levels here.

Most of the previous studies did not cover all three countries studied here (Spain – SP, Sweden – SW, United Kingdom – UK), but used samples from multiple locations within countries (Table 1; Fig. 1). In some cases, they identified local outliers as well as outliers shared among multiple locations.

Depending on how the original study was conducted, outlier data sets consisted of contigs from the *Littorina saxatilis* reference genome (from one 'crab' ecotype snail from Sweden; The IMAGO Marine Genome project, http://www.cemeb.science.gu.se/research/imago-marine-genome-projects/; project coordinated by Anders Blomberg and Kerstin Johannesson), contigs from EST or transcriptome assemblies (Galindo *et al.* 2010; Canbäck *et al.* 2011), or RADtags (Ravinet *et al.* 2016).

Table 1 Numbers of outlier and control loci analysed in this study. The first column indicates the original outlier scan studies. The following columns show in which locations outliers were identified (see Fig. 1 for location codes). Note that outliers shared across multiple locations were defined in varying ways. They were either detected in multiple locations independently using multiple, local outlier scans (Ravinet *et al.* 2016; Westram *et al.* 2014), or they were identified in pooled samples that contained individuals from two locations (Galindo *et al.* 2010). In the gene expression study (M Panova, unpublished) country-wise expression outliers were detected as loci showing an overall ecotype effect in a country-wise ANOVA. Outliers shared between Sweden and the UK were then identified as those that were significant in both countries independently. The last column shows the number of outlier/control loci followed up in this study, as well as the total number of probes designed for these loci (in brackets). The numbers of outlier loci for which we could actually obtain data in the current study (i.e. which were not discarded due to missing data) are shown in Figs S7–S10 (Supporting information) for each of the studied locations

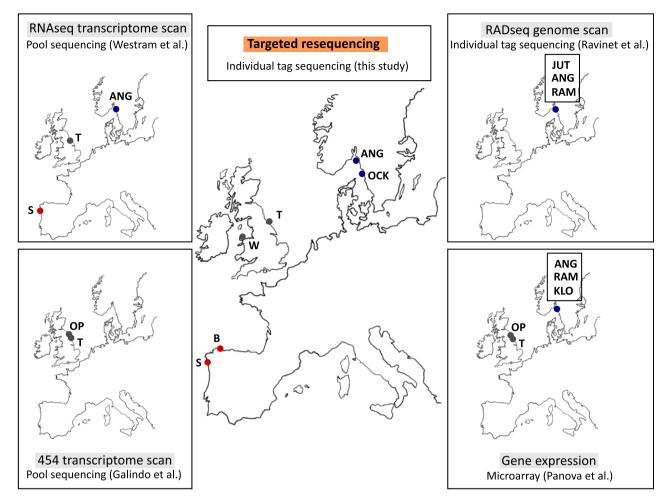| | Outlier location | | | | | | | |
Original study	Spain S	Sweden ANG	RAM	KLO	UK T	OP	None (i.e. control)	Number of loci followed up (number of probes)
RNAseq (transcriptome scan)	x							15 (94)
Westram *et al.* (2014)		x						15 (80)
					x			16 (114)
	x	x						21 (132)
	x				x			16 (183)
		x			x			25 (179)
454 (transcriptome scan) Galindo *et al.* (2010))					x	x		16 (226)
RAD (genome scan) Ravinet *et al.* (2015)		x						10 (57)
Gene expression (microarray)		x	x	x				7 (48)
(Panova *et al.* unpublished)					x	x		6 (55)
		x	x	x	x	x		16 (117)
Controls: random RAD loci Ravinet *et al.* (2015)							x	28 (234)
Controls: random *Littorina* sequence database (transcriptome) loci Canbäck *et al.* (2011)							x	80 (794)

Figure 1 Maps of sampling locations used for detection of outlier loci between ecotypes in previous studies (smaller plots) and the present work (large, central plot). Note that the present study comprises previously studied locations as well as new sites. Coordinates for sites sampled for the current study: B (Burela): 43.6763°,−7.3679°; S (Silleiro): 42.1012°, −8.8972°; ANG (Ängklåvebukten): 58.8710°, 11.1203°; OCK (Öckerö): 57.7181, 11.6290; T (Thornwick Bay): 54.1328°, −0.1129°; W (Anglesey): 53.2999°, −4.6795°. For the current study, 12 female individuals were sampled per population and ecotype.

Detailed methods for outlier selection and probe design are described in Data S1 (Supporting information). Briefly, we used reference genome contigs as the basis for probe design, as these are often longer than the original outlier sequences and contain introns as well as exons. These genome contigs were identified using a BLAST search of the outliers against the reference genome. The strongest outliers (sorted by P-value or F_{ST} estimate) which had a BLAST result in the genome were selected for probe design. We included local outliers (i.e. those appearing in a single location only) as well as outliers shared among different locations (Table 1), but analysed them separately (see below).

We also included control loci that should reflect the extent of neutral divergence between ecotypes (Table 1; Data S1, Supporting information). To obtain control loci comparable to transcriptome outliers, we obtained random sequences from the *Littorina* sequence database (Canbäck *et al.* 2011), which contains expressed sequences. For control loci comparable to RAD outliers, we randomly selected nonoutlier RAD loci from the study by Ravinet *et al.* (2016). These control loci were then processed in the same way as the outlier loci.

In the following, loci obtained from original studies and databases (i.e. both outliers and control loci) will simply be referred to as 'loci'. Note that each locus can potentially be represented by multiple reference genome contigs (e.g. due to splicing and the shortness of some of the reference contigs, see Data S1, Supporting information).

Probe design for capture sequencing

Probe design was performed by RAPiD Genomics (FL, USA) for the reference contigs selected as described

above. Probes were designed to map only to a single position in the reference genome to avoid paralogs and repetitive regions and were further selected based on GC content. This resulted in a total of 2313 120-bp capture probes, representing 271 loci (163 outlier loci and 108 control loci). To obtain long sequences, there were usually multiple nonoverlapping probes per locus, separated by short gaps. Numbers of loci and probes are shown in Table 1.

Sampling of Littorina saxatilis

Littorina saxatilis snails were sampled from two locations (100–300 km apart) in each of three countries, Spain, Sweden and the UK (Fig. 1, central plot; coordinates in figure legend). To test whether outliers again show elevated F_{ST} levels, we included the original locations where outliers were first detected (SP: S; SW: ANG; UK: T); but we also sampled a second location per country where outliers had not been tested before (SP: B; SW: OCK; UK: W), to analyse the extent to which outliers showed elevated F_{ST} in other locations in the same country. Twelve female individuals of each ecotype were sampled, avoiding any area of habitat overlap or unclear morphological classification. This resulted in a total of 144 *L. saxatilis* individuals from 12 populations (six locations × two ecotypes). Snails were stored in 100% ethanol.

DNA extraction and sequencing

DNA was extracted from a piece of foot tissue from each individual using a protocol modified from Winnepenninckx *et al.* (1993). Individually barcoded libraries were prepared and fragments were captured, at RAPiD Genomics. Fragments were then sequenced on an Illumina HiSeq 2000 machine (125-bp single-end reads), generating on average 2.5 M reads per individual sample (after excluding three samples that failed).

Read mapping, variant calling and consensus generation

Reads containing adapter contamination were removed, and quality trimming (-q 20) was performed using the software CUTADAPT (Martin 2011). Sequences shorter than 40 bp after trimming were discarded.

Four individuals were excluded because of low read numbers. The remaining reads were mapped to the *Littorina saxatilis* genome using the program bwa-mem (Li 2013). Only primary hits (i.e. excluding shorter, split hits) mapping to the targeted reference contigs and with a mapping quality of at least 30 were retained for

further analyses (these were on average 76% of the reads per sample). PicardTools (http://broadinstitute.github.io/picard/) MarkDuplicates was used to remove reads potentially originating from PCR duplicates based on identical 5′ mapping positions.

For each individual and reference contig, a consensus sequence was generated, using the recommended pipeline in samtools (Li *et al.* 2009, http://www.htslib.org/doc/samtools-1.1.html). For that, first, an mpileup file (containing the mapping data for each position) was generated, and variants were called with the bcftools 'call-c' command, using bases with a quality of at least 20. From this, a fastq file was generated using the VCFUTILS 'VCF2FQ' program, keeping only positions with a minimum read depth of 20, and then converted into a fasta file using seqtk (https://github.com/lh3/seqtk). Regions of low quality or low sequencing depth, as well as regions containing indels, were replaced by N. Working with sequences (rather than extracting SNPs) allowed us to obtain a more reliable F_{ST} estimate, as information was integrated over multiple SNPs.

The individual consensus sequences contained ambiguous bases for heterozygous sites. Two 'pseudo-haplotypes' were produced by generating two copies of the consensus sequence and replacing ambiguous bases by the two component bases: that is, the base composition of the sequences was correct, while the phase (which was irrelevant for downstream analyses) was arbitrary. Sequences shorter than 100 bp were discarded, and the remaining sequences were stored in a single fasta file per reference genome contig. Finally, for each locus for which we targeted multiple reference genome contigs (see Data S1, Supporting information), the contig-wise fasta files were concatenated.

Filtering based on sequencing depth, Hardy–Weinberg equilibrium and number of SNPs

Visual inspection of the fasta files indicated that some SNPs were heterozygous in almost all individuals in one or more populations. We suspected that this might be due to the presence of duplicated loci/repetitive sequences in our data. Potentially duplicated sequences were filtered using two approaches. First, we discarded loci with unusually high sequencing depth as follows. For each individual and reference genome contig, we averaged the sequencing depth across all covered bases. We then calculated the average sequencing depth across all contigs for each individual. Then, for each individual, the value of each contig was divided by this individual average. In this way, we obtained a contig-wise value for each individual that was corrected for general

differences in sequencing depth between individuals. To obtain a sequencing depth estimate for each population and contig, these individual-based values were averaged for each reference contig and population (ecotype within site). An average value of two for a given contig and population indicates that the average coverage depth for this contig is approximately twice as high as expected, which is high even when allowing for some variation between contigs. Therefore, loci associated with one or more reference contigs with an average coverage depth >2 in any of the 12 populations were discarded.

Second, SNPs with frequency differences between paralogs are expected to show negative F_{IS} values indicating deviations from Hardy–Weinberg equilibrium (HWE). We therefore used the following procedure to filter loci with evidence for departure from HWE. SNPs were extracted from alignment files for each locus. For each SNP with data for more than five individuals in a population, the number of homozygotes and heterozygotes was determined. SNP-wise F_{IS} was then calculated as $(H_e–H_o)/H_e$, where H_e is the expected heterozygosity based on allele frequencies, and H_o is the observed heterozygosity. Deviations of genotype frequencies from HWE were tested for each SNP and population using the approach of Wigginton et al. (2005). The fraction of SNPs with significantly negative F_{IS} was determined for each locus and population. In addition, the average F_{IS} per locus was calculated, using the average H_e and H_o estimates across SNPs in the equation above. Loci were discarded in a population-wise manner if their average F_{IS} was below −0.4 or above 0.4, and/or if more than 10% of SNPs showed significantly negative F_{IS} estimates. Loci with fewer than two SNPs across individuals were also discarded as the confidence in F_{ST} estimates should increase with the number of SNPs.

Nucleotide diversity (π)

Several population genetic statistics were calculated for each locus. The nucleotide diversity (π) is the average number of sequence differences per nucleotide position between sequences sampled from the same population. To calculate π, the number of sequence differences was counted for each possible pair of sequences from the focal population. These counts were summed over all sequence pairs and divided by the total number of nucleotide positions compared. Calculations were performed only if there were sequence data for at least six individuals in the focal population. We used the PYTHON EGGLIB library (De Mita & Siol 2012) and custom PYTHON scripts based on scripts developed by Martin et al. (2015).

d_{xy}

The measure d_{xy} (Nei 1987) corresponds to the average number of sequence differences per nucleotide position between sequences sampled from two different populations (here, two ecotypes from the same geographical location). We needed a measure of d_{xy} to calculate F_{ST} (see below).

The number of sequence differences between each possible pair of sequences from the two ecotypes was counted. These counts were summed over all possible pairs and divided by the total number of sites compared. Calculations were performed only if there were sequence data for at least six individuals in both focal populations. We used the PYTHON EGGLIB library (De Mita & Siol 2012) and custom PYTHON scripts based on scripts developed by Martin et al. (2015).

F_{ST}

For each location and locus, we calculated F_{ST} between the two ecotypes. Following the suggestion made by Bhatia et al. (2013), we used Hudson's estimator of F_{ST} (Hudson et al. 1992), which is less sensitive to sample size issues than other estimators. F_{ST} is defined as 1-(Hw/Hb), where Hw is the average number of sequence differences within populations, and Hb is the average number of sequence differences between populations. Here, this measure is equivalent to 1-(mean π/d_{xy}), where mean π is the average π across the two populations, if π and d_{xy} are calculated as explained above. We therefore simply used these values to calculate F_{ST}.

We also calculated F_{ST} between locations (within ecotypes), using the same approach.

F_{ST} values of outliers and patterns within and between countries

Our approach differed from a classical genome scan as we sequenced previously detected outlier loci as well as a set of control loci, rather than sequencing loci randomly picked from the genome. Therefore, we did not aim at identifying outliers in the current work and could not apply standard genome scan software (e.g. BAYESCAN, FDIST). Instead, we used the empirical F_{ST} distribution to ask whether outliers showed elevated F_{ST} estimates. When we mention 'outliers' in the following, we always refer to loci detected as outliers in previous work, and not to inferences from the F_{ST} estimates observed in the current study.

For each location, we first used the control loci, pooling those reflecting RAD loci and transcriptome loci, to obtain an expected distribution of F_{ST} between ecotypes. We then used the 80% quantile of this distribution as a

threshold to identify those outliers that showed a higher-than-expected F_{ST} estimate. We note that this threshold is arbitrary. We picked the relatively low threshold of 80% because the number of control loci with sufficient data was relatively small (the number and identity of control loci with insufficient data differed between locations, leaving between 58 and 84 loci per location; that is, each control distribution was based on a different set of loci.). With a higher quantile, the position of the threshold is driven by a very small number of loci and may therefore be less comparable among populations. However, for completeness we report the results for the 95% quantile in Table S2 (Supplementary Information).

Each outlier was coded in each location according to whether it fell above the 80% quantile of the respective control distribution or not. For each outlier type (RNA-seq, 454, RAD, expression) separately, we then estimated the probability that outliers show elevated F_{ST} (a) in their original location(s), (b) in a second location in the original country/countries and (c) in different countries. We set up a likelihood function assuming sampling from a binomial distribution, where n is the number of loci and P is the probability of 'success' (i.e. of obtaining a locus with $F_{ST} > 80\%$ quantile of the control distribution). Using maximum-likelihood estimation, we determined P and obtained the log-likelihood (mle2 function in the BBMLE package in R, using the 'L-BFGS-B' method; limited-memory Broyden–Fletcher–Goldfarb–Shanno). The probability, P, equals the observed proportion of loci above the 80% quantile; the log-likelihood estimation allowed us to compare different models.

To test whether the probabilities for (a), (b) and (c) were significantly different, we used a model selection approach to compare a total of eight models for each outlier type. The simplest model assumed a single probability of 0.2 in all three cases (i.e. no parameter estimation, 'fixed' in Table 2). A value of 0.2 is equal to the chance expectation because 20% of loci are expected to fall above the 80% quantile of the control distribution by chance. The second model still forced the probability to be the same for (a), (b) and (c), but allowed for the parameter to be estimated. Additional, more complex models allowed separate probabilities for the three different cases (details see Table 2). For each model, the AIC (Akaike information criterion) was calculated and the model(s) with the lowest AIC was selected as the best fit for the data.

As mentioned above, the control distribution against which outliers were compared was generated by pooling RAD and transcriptome control loci (because the number of loci in each category was relatively small; Table 1). However, we note that the F_{ST} distributions of these loci might differ, for example if RAD loci are less likely to be located in genic regions. We therefore compared the two F_{ST} distributions for each location using a Kolmogorov–Smirnov test. In two cases, it was significant (OCK: $P < 0.01$; T: $P = 0.03$), with the average F_{ST} for RAD loci being higher. To ensure that our results were not affected by this, we repeated the maximum-likelihood analysis described above, after separating RAD and transcriptome loci, testing for elevated F_{ST} (using the quantile approach described above) when using only the same locus type to generate the control distribution. The results were qualitatively and quantitatively very similar to the ones obtained with the pooled distribution and are therefore not shown.

Results

Data set after filtering

A total of 253 loci were retained after filtering for number of SNPs, sequencing depth and Hardy–Weinberg equilibrium (see also Fig. S2, Supporting information). On average, a locus had data in ~10 (of 12) populations and an average sequence length of 1165 base pairs (Fig. S2, Supporting information). Per location, between 133 and 205 loci passed filters and had data in both ecotypes, so that F_{ST} could be calculated.

F_{ST} estimates between ecotypes within locations, and between locations (using control loci only) are shown in Table S1, Supporting information. As expected, F_{ST} between ecotypes was generally lower than between locations, and F_{ST} between locations within countries was lower than F_{ST} between locations from different countries. Differentiation between ecotypes in the new UK location (W, west coast) could not be detected ($F_{ST} = 0$, compared to 0.04 at the other UK location) despite clear phenotypic differentiation.

F_{ST} values for outliers in their original location

One of our main questions was whether loci identified as F_{ST} outliers in earlier studies still showed elevated F_{ST} when studied in a new sample from the same location. Table 2 shows the probability that outliers fall above the 80% quantile obtained from the F_{ST} distributions of control loci, estimated using maximum likelihood. The best models (those with the lowest AIC for each outlier type) are also shown in Fig. 2 (see Figs S3–S6, Supporting information, for a direct comparison of outliers with the control distribution; and Figs S7–S10, Supporting information for the numbers of loci falling above the 80% quantile for each outlier type). The 'original locations' entries reflect these probabilities for the locations where

Table 2 Probabilities that outliers resequenced in samples from original and new locations fall above the 80% quantile of the control F_{ST} distribution. Eight different models estimated the probability for outliers in the location where they were first detected (original location), a second location in the same country (new location, original country), and other countries. Models varied in whether probabilities were estimated separately for these different categories or not. Where a single probability was estimated across multiple categories, the respective cells in the table are merged. 'Fixed' indicates that probabilities were not estimated by the model, but fixed at the value expected if outliers are as likely to have elevated F_{ST} estimates as control loci (0.2). The model with the lowest AIC for each outlier type is highlighted in bold. Any model <2 AIC units different from the model with the lowest AIC, that is with ΔAIC < 2, is highlighted as well as these models cannot clearly be distinguished. Numbers in brackets reflect the number of observations on which the estimation is based, that is the number of locus*population combinations for which an F_{ST} estimate was available. (a) Outliers detected in a single location in the original study. In the case of RNAseq outliers and loci showing expression differences, this means that F_{ST} was elevated in only one of the three/two countries tested in the original work, while 454 and RAD outliers were only tested in a single country and no information about the others was available. (b) Outliers shared between two countries in the original work. The 'original location' column shows the probability as estimated for both locations simultaneously

a)

outlier type	model	original location	new location in original country	other countries	-2*LL	AIC	ΔAIC
RNAseq	1 (null model)	0.2 (fixed) (174)			338.28	338.28	152.03
	2	0.54 (174)			240.09	242.09	55.84
	3	1.00 (26)	0.2 (fixed) (148)		254.59	256.59	70.34
	4	1.00 (26)	0.46 (148)		204.20	208.20	21.95
	5	0.89 (62)		0.2 (fixed) (112)	201.83	203.83	17.58
	6	0.89 (62)		0.35 (112)	188.49	192.49	6.24
	7	1.00 (26)	0.81 (36)	0.2 (fixed) (112)	193.58	197.58	11.33
	8	**1.00 (26)**	**0.81 (36)**	**0.35 (112)**	**180.25**	**186.25**	**0**
454	1 (null model)	0.2 (fixed) (52)			86.98	86.98	14.94
	2	**0.44 (52)**			**71.39**	**73.39**	**1.35**
	3	0.70 (10)	0.2 (fixed) (42)		75.32	77.32	5.28
	4	**0.70 (10)**	**0.38 (42)**		**68.04**	**72.04**	**0**
	5	0.52 (21)		0.2 (fixed) (31)	76.17	78.17	6.13
	6	0.52 (21)		0.39 (31)	70.45	74.45	2.41
	7	0.70 (10)	0.36 (11)	0.2 (fixed) (31)	73.74	77.74	5.7
	8	**0.70 (10)**	**0.36 (11)**	**0.39 (31)**	**68.02**	**74.02**	**1.98**
RAD	1 (null model)	0.2 (fixed) (48)			118.46	118.46	64.01
	2	0.73 (48)			56.07	58.07	3.62
	3	1.00 (8)	0.2 (fixed) (40)		92.71	94.71	40.26
	4	**1.00 (8)**	**0.68 (40)**		**50.45**	**54.45**	**0**
	5	0.87 (16)		0.2 (fixed) (32)	84.56	86.56	32.11
	6	0.87 (16)		0.66 (32)	53.24	57.24	2.79
	7	1.00 (8)	0.75 (8)	0.2 (fixed) (32)	81.50	85.50	31.05
	8	**1.00 (8)**	**0.75 (8)**	**0.66 (32)**	**50.18**	**56.18**	**1.73**
expression	1 (null model)	0.2 (fixed) (57)			91.98	91.98	15.57
	2	0.42 (57)			77.59	79.59	3.18
	3	0.60 (10)	0.2 (fixed) (47)		84.34	86.34	9.93
	4	0.60 (10)	0.38 (47)		76.02	80.02	3.61
	5	**0.63 (19)**		**0.2 (fixed) (38)**	**75.24**	**77.24**	**0.83**
	6	**0.63 (19)**		**0.32 (38)**	**72.41**	**76.41**	**0**
	7	0.60 (10)	0.67 (9)	0.2 (fixed) (38)	75.15	79.15	2.74
	8	**0.60 (10)**	**0.67 (9)**	**0.32 (38)**	**72.32**	**78.32**	**1.91**

Table 2 *Continued*

b)

outlier type	model	original locations	new locations in original countries	other countries	-2*LL	AIC	ΔAIC
RNAseq	1 (null model)	0.2 (fixed) (252)			592.12	592.12	358.77
	2	0.69 (252)			313.42	315.42	82.07
	3	0.93 (89)	0.2 (fixed) (163)		366.23	368.23	134.88
	4	0.93 (89)	0.55 (163)		268.14	272.14	38.79
	5	0.85 (180)		0.2 (fixed) (72)	239.76	241.76	8.41
	6	0.85 (180)		0.28 (72)	237.26	241.26	7.91
	7	**0.93 (89)**	**0.77 (91)**	**0.2 (fixed) (72)**	**229.85**	**233.85**	**0.5**
	8	**0.93 (89)**	**0.77 (91)**	**0.28 (72)**	**227.35**	**233.35**	**0**
expression	1 (null model)	0.2 (fixed) (59)			**70.69**	**70.69**	**0.53**
	2	0.27 (59)			**68.96**	**70.96**	**0.8**
	3	0.30 (20)	0.2 (fixed) (39)		**69.57**	**71.57**	**1.41**
	4	0.30 (20)	0.26 (39)		68.84	72.84	2.68
	5	**0.30 (43)**		**0.2 (fixed) (16)**	**68.16**	**70.16**	**0**
	6	**0.30 (43)**		**0.19 (16)**	**68.15**	**72.15**	**1.99**
	7	**0.30 (20)**	**0.30 (23)**	**0.2 (fixed) (16)**	**68.16**	**72.16**	**2**
	8	0.30 (20)	0.30 (23)	0.19 (16)	68.14	74.14	3.98

outliers were first detected. These probabilities were very high (>0.9) for the RNAseq and RAD outliers. The probabilities for the 454 outliers and the non-shared loci showing gene expression differences were lower, but in both cases still clearly above the chance expectation of 0.2 (horizontal line in Fig. 2). In cases where multiple models obtained similarly low AIC values, these results were supported by all such models (Table 2). High probabilities were still observed when the threshold for identification of loci with elevated F_{ST} levels was moved from the 80% quantile to the 95% quantile of the control distribution (Table S2, Supporting information).

F_{ST} of outliers in geographically close locations

We asked whether outliers from a given location are also likely to show elevated F_{ST} in another location within the same country. In Fig. 2, this corresponds to the 'new locations, original countries' bars. In all cases except for the shared loci showing expression differences, the probability was higher than the 0.2 expected by chance, strongly so for the RNAseq and RAD outliers and the nonshared loci showing expression differences. These results indicate that F_{ST} of outliers is elevated not only in the location where they were originally detected, but also in other locations in the same country. This conclusion was also supported by other models with low AIC values similar to the 'best' models shown in Fig. 2 (Table 2). Furthermore, the same conclusion emerged when the 95% rather than the 80% quantile of the control distribution was used in the analysis (Table S2, Supporting information).

Elevated F_{ST} in geographically close locations was also reflected by the fact that F_{ST} values (across all loci) were strongly correlated between locations within countries, particularly in Spain and Sweden (Pearson's correlation coefficient: Spain – 0.78; Sweden – 0.71; UK – 0.48; Fig. S11, Supporting information).

F_{ST} of outliers in geographically distant locations

The 'other countries' bars in Fig. 2 indicate the probability that outliers fall above the 80% quantile in countries other than where they were first detected. For all nonshared outliers, this probability was larger than 0.2. For the RNAseq outliers, there was only a slight increase above 0.2, indicating that most outliers are country specific. In contrast, for the RAD outliers the estimated probability in a different country was the same as that for a new location in the original country (0.68), indicating that F_{ST} of outliers may be elevated on a large geographical scale. Clearly elevated F_{ST} in a
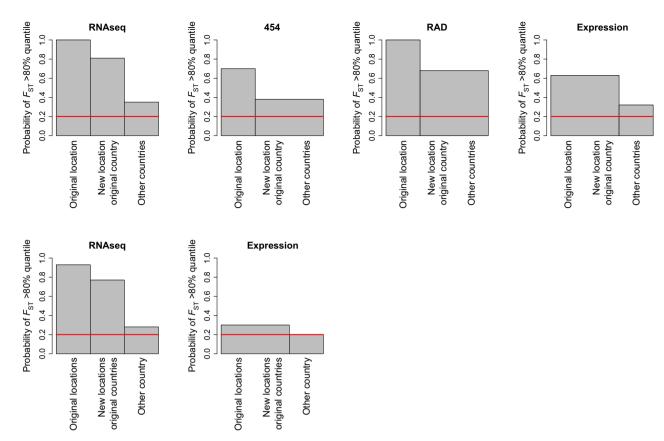
Figure 2 Probabilities that outliers sequenced in original and new locations fall above the 80% quantile of the control F_{ST} distribution. Subplots show different types of outliers (454: Galindo *et al.* 2010; RNAseq: Westram *et al.* 2014; RAD: Ravinet *et al.* 2016; expression: Panova *et al.* unpublished). Probabilities are estimates from the best-fitting model (see main text and Table 2). The models estimated the probability for outliers in the location where they were first detected (original location), a second location in the same country (new location, original country), and other countries. Models varied in whether probabilities were estimated separately for the different categories or not. If the best model did not distinguish between probabilities for different categories, the bars are merged here. For example, for the 454 outliers, the best model estimated a single probability of 0.38 of loci falling above the 80% quantile for both a new location in the original country and locations in other countries. The horizontal lines at a probability of 0.2 indicate the chance expectation. (a) Probabilities for outliers that were not shared among countries in the original study. (b) Probabilities for outliers shared among two countries in the original study. All 454 outliers were only tested in the UK in the original study; therefore, there are no shared 454 outliers. Similarly, all RAD outliers were only tested in Sweden in the original study.

different country for the RAD data was also supported by alternative models with similarly low AIC values (Table 2). When the 95% rather than the 80% quantile of the control distribution was used to identify outliers showing elevated F_{ST} (Table S2, Supporting information), the probabilities were lower but still clearly above the chance expectation (0.05).

Discussion

Several recent studies have found that the proportion of outliers shared among instances of parallel divergence may be small, but the relative roles of technical issues and evolutionary explanations causing this pattern are often not clear. Here, we have tried to disentangle these issues. First, we asked whether outliers detected in

earlier studies show elevated F_{ST} when sequenced in new samples; second, we asked whether outliers show a pattern of decreasing outlier sharing with increasing geographical distance, as expected if the observed level of sharing reflects the true extent of a common genetic basis of adaptation. We found that most outliers did again show high levels of differentiation (F_{ST}), but the few outliers with low F_{ST} may indicate false-positive outliers that could have decreased observed sharing among locations. However, we found clear evidence for elevated F_{ST} of outliers in geographically close and sometimes even geographically distant locations, reflecting both recent common ancestry (Panova *et al.* 2011) and the potential for gene flow.

We used a capture sequencing approach to resequence outliers from previous studies as well as

surrounding genomic regions. Capture sequencing is a powerful approach when the number of loci of interest is too large for PCR-based approaches and some sequence information is already available (Jones & Good 2016). While the latter represents a limitation in a nonmodel organism without a reference genome, we show here that a draft genome consisting of short contigs (N50: ~950 bp) is sufficient for probe design and generating relatively long stretches of sequence data. We obtained sequence information for almost all of the targeted loci (although we had to filter some out due to potential duplication, see below). The advantage of capture sequencing is that data for large genomic regions (rather than short markers; e.g. RAD sequencing) can be obtained without the need to resequence the whole genome. Longer sequences provide more information for a given locus because they often contain multiple SNPs (on average 24 in this study; average sequence length 1165 base pairs), making F_{ST} and other estimates more reliable. The additional advantage of capture sequencing exploited in this study is that specific genomic regions of interest can be targeted, rather than, or in addition to, random markers across the genome.

While capture sequencing yielded information for almost all the targeted loci, some loci had to be removed (at least locally) because they showed unusually high levels of heterozygosity. This is probably caused by repetitive regions/paralogs that are sufficiently similar to be captured by the same probe (and that were assembled into the same genome contig). In studies where pooled samples are used (or where deviation from Hardy–Weinberg equilibrium is not tested), such loci may affect the neutral F_{ST} distribution because they introduce spurious SNPs with low differentiation. This observation emphasizes the need to confirm outliers detected in pool-seq studies using individual barcoding. In general, these loci represent an interesting phenomenon as they seem to be more prevalent in some populations (e.g. crab ecotype, ANG, Sweden; and wave ecotype, S, Spain; Fig. S1, Supporting information). This might be the result of recent duplications of larger genomic regions. There is evidence of common gene duplications and variation in copy numbers both between the geographic populations and the ecotypes in *Littorina* (Panova *et al.* 2014). These loci will be an interesting target for future studies especially given the potential role of chromosomal rearrangements and gene duplications in divergence and speciation (Lynch & Force 2000; Faria & Navarro 2010).

Using the filtered data set, we found that general patterns of differentiation between populations in control loci are in line with results from earlier studies. Differentiation between countries was relatively high, especially between Spain and the other two countries (Table S1, Supporting information), where populations probably have a more recent shared origin (Doellman *et al.* 2011; Panova *et al.* 2011). Genetic differentiation between locations within the same country was lower. Differentiation between ecotypes was generally lower than between locations and showed similar patterns to earlier studies, with ecotypes being most differentiated in Spain (Butlin *et al.* 2014; Westram *et al.* 2014).

Regarding our first main question, we show that the probability that outliers exhibit elevated F_{ST} in their original location is not always 100% (Fig. 2), which is not surprising. First, methods of outlier detection differed among studies, two of them (Galindo *et al.* 2010; Ravinet *et al.* 2016) using model-based approaches – the softwares WINKLES (Wilding *et al.* 2001) and FDIST (Beaumont & Nichols 1996) – and one using the empirical F_{ST} distribution (Westram *et al.* 2014). These methods differ in the underlying assumptions and sensitivity (e.g. Storz 2005). Maybe most importantly, observed allele frequencies (both in previous work and in the current study) were simply subject to sampling noise due to small sample size, which affects the reliability of outlier scans (Vilas *et al.* 2012). This may be even more true for studies where pooled samples were used, and individuals may contribute differently to the pools (Galindo *et al.* 2010; Westram *et al.* 2014). Similarly, two of the earlier studies (Galindo *et al.* 2010; Westram *et al.* 2014) were based on transcriptome data, with noise introduced by expression variation among individuals, and potentially between different alleles at the same locus. The capture sequencing and analytical approaches used here were expected to provide more robust estimates of differentiation than previous studies and yet outlier repeatability was high. Outliers that did not show high differentiation in the current work may therefore represent false positives. These are not expected to be shared and as a result may reduce the observed amount of outlier sharing.

We suggest that replication can be an important means for estimating the reliability of outliers. Many studies use replicates in the form of pairwise samples from several different geographical locations and define outliers not shared among comparisons as false positives (e.g. Bonin *et al.* 2006). We believe that this method is suitable to increase confidence in outliers shared among locations and may be appropriate especially in systems where replicates are geographically close and where high sharing among locations is expected. However, it will not be appropriate for confirming location-specific outliers, or for understanding the relative contribution of shared vs. local outliers to the overall amount of divergence. These points can only be targeted using replicates from the same location.

Despite some outliers with low F_{ST}, overall the RNA-seq and RAD outliers show probabilities of >0.9, and the

454 outliers of 0.7, to have elevated F_{ST} in the original location (i.e. to fall above the 80% quantile of the control distribution; Fig. 2). These results indicate that genuine outliers could be detected quite reliably in the original work despite limited sample sizes, pooled sequencing and the use of transcriptome data in some of the studies. Additionally, patterns of differentiation in short markers in the original studies (e.g. RAD tags; Ravinet *et al.* 2016) are recovered when studying longer stretches of sequence. The repeatability of high differentiation patterns represents evidence that at least in the *L. saxatilis* system, the low sharing of outliers between countries is unlikely to be solely caused by a large number of false positives caused by sampling effects. It also indicates that F_{ST} patterns are temporally stable over several years.

However, it is important to note that loci not directly relevant for ecotype divergence could also show repeatable patterns of high divergence. For example, outliers mainly caused by low recombination rates in combination with selection on distant linked loci are probably repeatable. Further studies of demographic history and variation of recombination rate across the genome will be necessary to rule out these types of effects. Ultimately, confidence in outliers can only be gained if independent signatures of selection can be detected, they can be associated with divergent traits (e.g. via QTL mapping or, ideally, via functional manipulation), or if selection acting on them can be tested more directly (e.g. experimentally; Soria-Carrasco *et al.* 2014).

In addition to outliers from genome scans, we also followed up loci showing expression differences between ecotypes in Sweden and the UK (Panova *et al.* unpublished). The nonshared loci often had high F_{ST} in the original location, similar to the genome scan outliers (Table 2, Fig. 2). This may indicate that in some cases regulatory elements may be closely linked with the coding sequence. However, this does not seem to be a general pattern as the shared loci did not have high F_{ST} estimates (Table 2, Fig. 2). In these cases, gene expression differences may be caused by trans-regulatory elements remote from the expressed sequence and therefore probably not included in our targets. A similar pattern has been found, for example, for the divergence between dwarf and normal whitefish, where the extent of expression differences is not correlated with the amount of sequence divergence at the same locus (Jeukens *et al.* 2010).

Our second main question was whether outliers showed elevated F_{ST} in geographically close and distant locations. Because the RNAseq, 454 and RAD outliers were identified in different ways in the original work and produced different results here, we consider them separately in the following discussion.

The RNAseq outliers were the group of loci for which we could perform the most powerful analysis because the number of observations was much larger than for the other marker types (see Table 2, Figs S7–S10, Supporting information). The study identifying the RNAseq outliers (Westram *et al.* 2014) differed from the other two outlier scan studies in that loci were tested in all three countries (Spain, Sweden and UK; see Table 1). Therefore, there were clear expectations: Nonshared outliers should show low F_{ST} estimates in both other countries, and outliers shared between two countries should show low F_{ST} in the third country. In accordance with these expectations, we found that the RNAseq outliers, while having elevated F_{ST} values in the country/countries where they were first detected, exhibited only slightly increased probabilities for elevated F_{ST} in the other country/countries (Fig. 2). In contrast, RNAseq outliers had a probability of about 0.8 of showing elevated F_{ST} in locations within the same country, where they had not been tested before (Fig. 2). These results are consistent with a similar genomic basis of divergence on a small (but not a large) geographical scale. They also suggest that the low levels of outlier sharing among countries observed earlier (Westram *et al.* 2014), and supported in the current study, cannot be explained by a large number of false positives due to sampling effects. If such effects played a large role, we would not expect to find any geographical pattern; instead, outliers should rarely show elevated F_{ST} in any other location.

However, for the 454 outliers, which should mostly be located within expressed genes similar to the RNAseq outliers, we obtained a different result. The probability of showing elevated F_{ST} in a second location in the original country was relatively low (0.38), and the same as that estimated for other countries (Fig. 2). As we expected elevated F_{ST} within the same country, but not in other countries (as observed for the RNAseq outliers), this result is slightly surprising. However, we note that the number of loci in this analysis was much smaller than for the RNAseq outliers. It is possible that because of this, by chance, we included a relatively high proportion of loci that are location specific.

The RAD outliers showed an elevation of F_{ST} not only in geographically close but also distant locations (Table 2, Fig. 2). This result is surprising especially in so far as most of these outliers were not shared among Swedish locations in the original study, and so were not expected to show elevated F_{ST} on an even larger geographical scale. However, here we used a relatively relaxed criterion to identify loci with 'elevated F_{ST}' (loci that fall above the 80% quantile of the neutral F_{ST} distribution) compared with the original study, which may explain why loci can show elevated F_{ST} across large

geographical scales without appearing as shared outliers in the original study (Ravinet *et al.* 2016) (also see discussion of this point below). The result of high F_{ST} even in other countries might also partly be explained by the fact that these outliers are less focussed on coding regions. Markers in regions with low gene densities might also be in regions with lower recombination rates (Flowers *et al.* 2012), making elevation of F_{ST} on large geographical scales more likely because associations with selected alleles are more likely to be maintained.

Taken together, the results from the different outlier types suggest that outliers frequently show elevated F_{ST} on small, and sometimes also on large geographical scales. This contrasts with work in other systems where outliers were often not shared among instances of parallel evolution, for example in stick insects (Soria-Carrasco *et al.* 2014) and groundsels (Roda *et al.* 2013). While these differences between study systems may be explained by taxon-specific patterns of gene flow and evolutionary history, or by differences in the geographical sampling scheme, it is more surprising that our current results seem to contradict earlier work on *L. saxatilis*, where outlier sharing was low both on small (within archipelago; Ravinet *et al.* 2016) and large (between country; Westram *et al.* 2014) geographical scales. The discrepancies between studies may partly be explained by the fact that here we did not try to identify outliers using a stringent threshold, but rather asked whether known outliers showed elevated F_{ST} in other geographical locations. In the earlier studies, outlier loci may frequently have shown high differentiation in a second location but not met the threshold for outlier detection. One reason for such patterns could be the position of these loci in regions of low recombination, where they may be affected by selection on multiple linked loci (Roesti *et al.* 2012), but only appear as strong outliers locally. Alternatively, if divergent traits are polygenic and a lot of loci are affected by divergent selection, not all of them may appear as statistically significant outliers at any given point in time, which might lead to differences among locations if a stringent criterion for outlier detection is used. Many of the traits responding to divergent selection in *L. saxatilis* are probably highly polygenic (e.g. body size and shape, which show continuous variation across the hybrid zones between ecotypes; e.g. Hollander *et al.* 2015).

Our results indicate that, even though there is evidence that ecotypes have evolved repeatedly within countries and within regions (Butlin *et al.* 2014), divergence may partly be based on the same genomic regions. This may be due to shared standing genetic variation potentially predating the colonization of the different areas, as well as gene flow among locations. Sharing of outliers, sometimes even on large geographical scales, has been observed in other system, for example for the divergence between freshwater and marine sticklebacks, where the same genomic regions are involved in divergence repeatedly (Hohenlohe *et al.* 2010; Jones *et al.* 2012). Even though gene flow between locations in *L. saxatilis* is probably low, especially given that this species is ovoviviparous and has no pelagic larval stage (Reid 1996), small amounts of gene flow may be sufficient to transport strongly adaptive alleles among locations (Morjan & Rieseberg 2004). Additionally, *L. saxatilis* populations are typically very large; therefore, the maintenance of a large amount of ancestral genetic variation seems conceivable. We did not find a pattern of more negative Tajima's D (Tajima 1989) for outliers compared with control loci (Figs S12–S15, Supporting information), which is consistent with adaptation from standing genetic variation as well as variation introduced by past gene flow, but not with hard sweeps of recent novel mutations or alleles introduced by recent gene flow (Nielsen 2005; Barrett & Schluter 2008).

Both shared old genetic variation and gene flow are expected to be higher among geographically close populations, likely explaining the differences between close and distant populations that we observe mainly for the RNAseq outliers. Additionally, or alternatively, selection pressures could also be more similar among geographically close populations, which should increase the proportion of shared outliers. For example, in sunflowers, species pairs diverging across a similar environmental gradient show more similar patterns of divergence than species pairs diverging across different gradients (Renaut *et al.* 2014). In *L. saxatilis* sampling sites in the UK, the crab ecotype is located in the mid shore and the wave ecotype in the high shore. In Sweden, such a tidal pattern is absent. Therefore, any outliers generated by selection along this axis (e.g. those conferring adaptation to desiccation resistance) could potentially be shared among UK locations, but are not expected to be shared with Swedish locations (Westram *et al.* 2014).

In addition to the geographical patterns of outliers, their biological role is of interest. In the future, the availability of an improved version of the *L. saxatilis* reference genome and of more information for related mollusc species will allow for a detailed investigation of the molecular functions of outlier loci. This will help in answering, for example, the question whether similar functions are fulfilled by different loci in different geographical locations. This study and previous work already allow for some insights. A previous transcriptome scan identified genes involved in shell calcification, foot muscle operation and energetic metabolism (Galindo *et al.* 2010). Here, we annotated the outlier loci again showing very high levels of F_{ST} (above the 95% quantile of the control

distribution) in the current study (methods described in Data S1, Supporting information). These loci may be particularly good candidates for further in-depth study, and their annotations are shown in Table S3 (Supporting information). For example, one of the outlier loci showed similarity with proteins of the family of alpha carbonic anhydrases. These enzymes are known for their role in the biomineralization of calcium carbonate structures (Moya *et al.* 2008; Le Roy *et al.* 2014) like mollusc shells. Carbonic anhydrase secreted by the mantle tissue is responsible for the hydration of CO_2 and the deposition of calcium carbonate on the shell (Le Roy *et al.* 2012). This function could be under divergent selection between the crab and wave ecotypes in Spain (Table S3, Supporting information), promoting differences in shell structure. Other outliers are more difficult to assign to specific functions in the barriers to gene flow between these ecotypes, although some functions overlap between outliers (i.e. translation initiation factors, serine–threonine phosphatases) and therefore should be investigated in future studies.

Future work will also link the shared and nonshared outlier regions with phenotypes to understand their role in divergence. Furthermore, with improved genomic resources, we will be able to narrow down the actual targets of selection and study outlier loci within their genomic context.

Acknowledgements

References

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.

Beaumont MA (2005) Adaptation and speciation: what can Fst tell us? *Trends in Ecology & Evolution*, **20**, 435–440.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **263**, 1619–1626.

Bhatia G, Patterson N, Sankararaman S, Price AL (2013) Estimating and interpreting FST: the impact of rare variants. *Genome Research*, **23**, 1514–1521.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773–783.

Butlin RK (2010) Population genomics and speciation. *Genetica*, **138**, 409–418.

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.

Canbäck B, André C, Galindo J *et al.* (2011) The *Littorina* sequence database (LSD) – an online resource for genomic data. *Molecular ecology resources*, **12**, 142–148.

Conte GL, Arnegard ME, Peichel CL, Schluter D (2012) The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 5039–5047.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

De Mita S, Siol M (2012) EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics*, **13**, 27.

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 B: Biological Sciences*, **279**, 1277–1286.

Doellman MM, Trussell GC, Grahame JW, Vollmer SV (2011) Phylogeographic analysis reveals a deep lineage split within North Atlantic *Littorina saxatilis*. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 3175–3183.

Elmer KR, Meyer A (2011) Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution*, **26**, 298–306.

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

Faria R, Navarro A (2010) Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends in Ecology & Evolution*, **25**, 660–669.

Flowers JM, Molina J, Rubinstein S *et al.* (2012) Natural selection in gene-dense regions shapes the genomic pattern of polymorphism in wild and domesticated rice. *Molecular Biology and Evolution*, **29**, 675–687.

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.

Gagnaire P-A, Pavey SA, Normandeau E, Bernatchez L (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, **67**, 2483–2497.

Galindo J, Morán P, Rolán-Alvarez E (2009) Comparing geographical genetic differentiation between candidate and non-candidate loci for adaptation strengthens support for parallel

ecological divergence in the marine snail *Littorina saxatilis*. *Molecular Ecology*, **18**, 919–930.

Galindo J, Grahame JW, Butlin RK (2010) An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. *Journal of evolutionary biology*, **23**, 2004–2016.

Grahame JW, Wilding CS, Butlin RK (2006) Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis*. *Evolution*, **60**, 268–278.

Hebert FO, Renaut S, Bernatchez L (2013) Targeted sequence capture and resequencing implies a predominant role of regulatory regions in the divergence of a sympatric lake whitefish species pair (*Coregonus clupeaformi*). *Molecular Ecology*, **22**, 4896–4914.

Hohenlohe PA, Bassham S, Etter PD, *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Hollander J, Galindo J, Butlin RK (2015) Selection on outlier loci and their association with adaptive phenotypes in Littorina saxatilis contact zones. *Journal of Evolutionary Biology*, **28**, 328–337.

Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, **9**, 138–151.

Jeukens J, Renaut S, St-Cyr J, Nolte AW, Bernatchez L (2010) The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Molecular Ecology*, **19**, 5389–5403.

Johannesson K, Rolan-Alvarez E, Ekendahl A (1995) Incipient reproductive isolation between two sympatric morphs of the intertidal snail *Littorina saxatilis*. *Evolution*, **49**, 1180.

Johannesson K, Panova M, Kemppainen P *et al.* (2010) Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1735–1747.

Jones MR, Good JM (2016) Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, **25**, 185–202.

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

Kautt AF, Elmer KR, Meyer A (2012) Genomic signatures of divergent selection and speciation patterns in a "natural experiment", the young parallel radiations of Nicaraguan crater lake cichlid fishes. *Molecular Ecology*, **21**, 4770–4786.

Le Roy N, Marie B, Gaume B *et al.* (2012) Identification of two carbonic anhydrases in the mantle of the European abalone *Haliotis tuberculata* (Gastropoda, Haliotidae): Phylogenetic implications. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **318**, 353–367.

Le Roy N, Jackson DJ, Marie B, Ramos-Silva P, Marin F (2014) The evolution of metazoan α-carbonic anhydrases and their roles in calcium carbonate biomineralization. *Frontiers in Zoology*, **11**, 75.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio].

Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Lynch M, Force AG (2000) The origin of interspecific genomic incompatibility via gene duplication. *The American Naturalist*, **156**, 590–605.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, **17**, 10–12.

Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, **32**, 244–257.

Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, **13**, 1341–1356.

Moya A, Tambutté S, Bertucci A *et al.* (2008) Carbonic anhydrase in the scleractinian coral stylophora pistillata characterization, localization, and role in biomineralization. *Journal of Biological Chemistry*, **283**, 25475–25484.

Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 343–353.

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

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

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species *Heredity*, **103**, 439–444.

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

Panova M, Hollander J, Johannesson K (2006) Site-specific genetic divergence in parallel hybrid zones suggests nonallopatric evolution of reproductive barriers. *Molecular Ecology*, **15**, 4021–4031.

Panova M, Blakeslee AMH, Miller AW *et al.* (2011) Glacial history of the north atlantic marine snail, *Littorina saxatilis*, inferred from distribution of mitochondrial DNA lineages. *PLoS One*, **6**, e17511.

Panova M, Johansson T, Canbäck B *et al.* (2014) Species and gene divergence in *Littorina* snails detected by array comparative genomic hybridization. *BMC Genomics*, **15**, 687.

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.

Ravinet M, Westram A, Johannesson K *et al.* (2016) Shared and non-shared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Molecular Ecology*, **25**, 287–305.

Reid DG (1996) *Systematics and Evolution of Littorina*. The Ray Society, London.

Renaut S, Grassa CJ, Yeaman S *et al.* (2013) Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, **4**, 1827.

Renaut S, Owens GL, Rieseberg LH (2014) Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology*, **23**, 311–324.

Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, **66**, 1–17.

Roda F, Liu H, Wilkinson MJ *et al.* (2013) Convergence and divergence during the adaptation to similar environments by an Australian groundsel. *Evolution*, **67**, 2515–2529.

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.

Rolán-Alvarez E (2007) Sympatric speciation as a by-product of ecological adaptation in the Galician *Littorina saxatilis* hybrid zone. *Journal of Molluscan Studies*, **73**, 1–10.

Schluter D, Nagel LM (1995) Parallel speciation by natural selection. *The American Naturalist*, **146**, 292–301.

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

Smadja CM, Canbäck B, Vitalis R *et al.* (2012) Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution*, **66**, 2723–2738.

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.

Storz JF (2005) INVITED REVIEW: using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Vilas A, Pérez-Figueroa A, Caballero A (2012) A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *Journal of Evolutionary Biology*, **25**, 1364–1376.

Webster SE, Galindo J, Grahame JW, Butlin RK (2012) Habitat choice and speciation. *International Journal of Ecology*, **2012**, 1–12.

Westram AM, Galindo J, Alm Rosenblad M *et al.* (2014) Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Molecular Ecology*, **23**, 4603–4616.

Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *The American Journal of Human Genetics*, **76**, 887–893.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**, 611–619.

Winnepenninckx B, Backeljau T, De Wachter R (1993) Extraction of high-molecular-weight DNA from molluscs. *Trends in Genetics*, **9**, 407.

J.G. and R.K.B. developed the idea and A.M.W. and R.K.B. designed the study. A.M.W., J.G. and R.K.B. collected samples. A.M.W., M.P. and R.K.B. analysed the data. A.M.W., J.G., M.P. and R.K.B. wrote the manuscript.

Data accessibility

Supporting information

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

Data S1. Selection of outlier loci.

Table S1. Average F_{ST} estimates across control loci ($\pm$SD) between sites and between ecotypes.

Table S2. Probabilities that outliers sequenced in samples from original and new locations fall above the 95% quantile of the control F_{ST} distribution.

Table S3. Outliers from earlier studies that showed highly elevated F_{ST} (above the 95% quantile of the control distribution) in the current study, and that could be functionally annotated.

Fig. S1. F_{IS} estimates per locus for both ecotypes (C = crab; W = wave) at the six locations sampled for this study.

Fig. S2. Histograms showing information about the filtered dataset used for downstream analyses.

Fig. S3. Distribution of F_{ST} estimates between ecotypes for each of the six locations sampled for this study, using control loci only (grey bars).

Fig. S4. Distribution of F_{ST} estimates between ecotypes for each of the six locations sampled here.

Fig. S5. Distribution of F_{ST} estimates between ecotypes for each of the six locations sampled here.

Fig. S6. Distribution of F_{ST} estimates between ecotypes for each of the six locations sampled here.

Fig. S7. Proportions of RNAseq outliers (Westram *et al.* 2014) falling above the 80% quantile of the control F_{ST} distribution in the current study.

Fig. S8. Proportions of transcriptome outliers from (Galindo *et al.* 2010) falling above the 80% quantile of the control F_{ST} distribution in the current study.

Fig. S9. Proportions of RADseq outliers (Ravinet *et al.* 2016) falling above the 80% quantile of the F_{ST} distribution for control loci in the current study.

Fig. S10. Proportions of loci showing differential gene expression between ecotypes in Panova *et al.* (unpublished) falling above the 80% quantile of the F_{ST} distribution for control loci in the current study.

Fig. S11. Correlation of per-locus F_{ST} estimates between locations within countries (left: Spain; middle: Sweden; right: UK) for control and outlier loci.

Fig. S12. Boxplots of Tajima's D for outliers from an RNAseq outlier scan (Westram *et al.* 2014).

Fig. S13. Boxplots of Tajima's D for outliers from the 454 transcriptome scan (Galindo *et al.* 2010).

Fig. S14. Boxplots of Tajima's D for RAD outliers (Ravinet *et al.* 2016).

Fig. S15. Boxplots of Tajima's D for loci showing differential gene expression between ecotypes in Panova *et al.* (unpublished).