

SCIENTIFIC REPORTS



OPEN

Population genomics of parallel evolution in gene expression and gene sequence during ecological adaptation

María José Rivas¹, María Saura¹, Andrés Pérez-Figueroa¹ , Marina Panova², Tomas Johansson³, Carl André², Armando Caballero¹, Emilio Rolán-Alvarez¹, Kerstin Johannesson² & Humberto Quesada¹

Natural selection often produces parallel phenotypic changes in response to a similar adaptive challenge. However, the extent to which parallel gene expression differences and genomic divergence underlie parallel phenotypic traits and whether they are decoupled or not remains largely unexplored. We performed a population genomic study of parallel ecological adaptation among replicate ecotype pairs of the rough periwinkle (*Littorina saxatilis*) at a regional geographical scale (NW Spain). We show that genomic changes underlying parallel phenotypic divergence followed a complex pattern of both repeatable differences and of differences unique to specific ecotype pairs, in which parallel changes in expression or sequence are restricted to a limited set of genes. Yet, the majority of divergent genes were divergent either for gene expression or coding sequence, but not for both simultaneously. Overall, our findings suggest that divergent selection significantly contributed to the process of parallel molecular differentiation among ecotype pairs, and that changes in expression and gene sequence underlying phenotypic divergence could, at least to a certain extent, be considered decoupled processes.

The importance of natural selection on population divergence and the genesis of new species remains poorly understood. One reason for this limited knowledge is the stochasticity linked to the somewhat unique history of each population and species, which can overwhelm the fingerprint of adaptive divergence¹. Instances of repeated, parallel phenotypic evolution in response to similar environmental pressures provide strong evidence of evolution by natural selection, as genetic drift is unlikely to generate a concerted change in multiple, independent lineages^{2,3}. However, the underlying genetic basis of this process is unclear. In particular, we know very little as to whether selection acts upon the same genetic machineries to generate repeated phenotypes, or if its action follows alternative genetic routes^{4–6}. Similarly, it remains unknown to what extent constraints faced by organic evolution might facilitate the repeated use of the same genes during independent phenotypic evolution^{7,8}.

Evidence for parallel evolution has been found in many taxa^{4,9}. However, one limitation of our view that parallel evolution is rather abundant comes from the fact that many studies are based on targeted candidate gene surveys that suffer from an inevitable ascertainment bias, as they do not allow answering whether repeated genetic changes are ubiquitous across the genome or more frequent than the neutral expectation³. Recent studies using a genome-wide approach have provided some unbiased insights into our understanding of the level of genome-wide repeatability linked to parallel evolution. For instance, molecular footprints of selection underlying parallel phenotypic evolution in cichlid fishes¹⁰, Australian groundsel¹¹ and lake trout¹² involve replicated evolution on a rather restricted subset of genes and, more frequently, divergence events that are unique to each population. Similarly, the early stages of parallel speciation in the stick insect *Tinema cristinae* involve mostly nonparallel divergence despite evidence of the importance of repeated selection on the same genes¹³. However, the repeatability of evolution through the reuse of the same genes may be substantial amongst recently diverged lineages^{9,14}. Overall, these and other studies^{15–18} suggest that the genomic architecture underlying parallel

¹Departamento de Bioquímica, Genética e Inmunología, Universidad de Vigo, 36310, Vigo, Spain. ²Department of Marine Sciences, Tjärnö, University of Gothenburg, SE-452 96, Strömstad, Sweden. ³Department of Biology, University of Lund, SE-223 62, Lund, Sweden. Correspondence and requests for materials should be addressed to H.Q. (email: hquesada@uvigo.es)

phenotypic evolution frequently follows complex genetic trajectories, affecting multiple loci that show a mosaic pattern of both repeatable and idiosyncratic divergence, and where ancestral standing variation is frequently an important source of adaptive variation.

Very few studies have attempted to address the extent to which parallel gene expression differences and genomic divergence underlie parallel phenotypic traits^{19–22}. This question is of central importance, because adaptive variation is likely to be underpinned by changes in both regulatory and coding sequences²³. If evolution in coding sequences and regulatory regions are two highly related phenomena, then patterns of differentiation in coding sequence and gene expression should be markedly similar, i.e. they should be coupled. However, previous attempts to test the coupling between coding sequences and gene expression in multicellular organisms have given conflicting results, with markedly similar patterns of differentiation found in some datasets^{24–27}, but very dissimilar in others^{17,28,29}. To add further uncertainty, the specific mechanism underlying these observations remains elusive. Thus, adaptive selection driving rapid evolution of both gene expression and coding sequence may account for the coupling²⁴, but also variation in functional constraints, in which genes less constrained in coding sequence would also be less constrained in expression^{26–30}. Alternatively, markedly dissimilar patterns of differentiation would point towards the possibility that changes in coding sequence and gene expression underlying phenotypic evolution play different roles during evolution and could, at least to a certain extent, be considered decoupled processes^{31,32}. Unveiling the degree of convergence at different levels of genomic organization will help to establish to what extent natural selection, genetic constraints, or independent modes of evolution, determine whether patterns of genetic differentiation associated with adaptation are predictable.

The marine snail *Littorina saxatilis* provides an excellent opportunity for testing these aspects of evolutionary repeatability. In this ovoviviparous species, pairs of ecotypes adapted to distinct parts of the same shores have repeatedly evolved in different geographical regions from Europe^{33–35}. Parallel phenotypic divergence involves many traits, including body size, shell shape, shell thickness, and behavior³⁶. Hybridization occurs in a relatively narrow zone, but gene flow among ecotypes is restricted due to assortative mating, immigrant inviability, and habitat choice^{37–39}. Consistent with the prediction of parallel evolution, pairs of sympatric ecotypes cluster in phylogenetic trees by geographic origin but not by ecotype⁴⁰. Computer simulations assessing the confounding effect of gene flow on phylogenetic inference confirm this result, demonstrating that the time elapsed since the emergence of ecotypes would not be enough to erode the distinctive phylogenetic signal linked to a parallel or a non-parallel (allopatric) origin of ecotypes⁴¹. Moreover, the comparison between alternative evolutionary models further supports that data better fit a scenario in which the separation of pairs of ecotypes occurred in parallel at both regional and local scales³⁵. Recent genomic studies comparing populations from three geographically distant regions (Spain, Sweden and United Kingdom) suggest that footprints of selection are frequently region-specific^{42,43}, or even site-specific at a very local scale⁴⁴. Still, no study in *Littorina* has so far investigated the extent of parallelism in gene expression nor the relation between variation in gene expression and divergence in coding sequences.

Despite the ongoing development of next-generation sequencing technologies for genome-wide evolutionary analyses, it remains technically and financially unapproachable for many laboratories to sequence whole genomes or transcriptomes. Microarrays remain a simple and inexpensive alternative for genotype-related purposes and gene expression analyses⁴⁵. Array-based comparative genomic hybridization can be accurately used as a proxy to estimate genome-wide divergence by comparing hybridization intensities of individuals on the microarray^{46,47}. Modifications of this method have been successfully used to identify SNPs or copy number variants without the need of allele-specific probes, thanks to a linear relationship between hybridization signal and sequence divergence⁴⁷. Similarly, microarrays remain widely used for gene expression profiling, as correlation between microarray data and other platforms such as RNA-seq is usually pretty good^{48,49}. Although microarrays may be problematic for the study of low expressed genes, RNA-seq shows a greater degree of intensity-dependent variation than do microarrays⁵⁰, which has led some authors to recommend the use of RNA-seq^{51,52}, and others challenging that conclusion^{53,54}.

Here we combine genome-wide evolutionary analyses of coding sequences and gene expression data using microarrays for investigating the molecular basis of adaptive divergence, employing *L. saxatilis* ecotypes from NW Spain as a model system. In this region, a large “crab ecotype” and a smaller “wave ecotype” have evolved repeatedly in response to crab predation and wave exposure respectively^{33,35,40,55}. The recent origin of these ecotypes (<10,000 years)³⁵ is expected to be associated with high levels of shared genetic constraints and standing variation that would facilitate a rapid and more pervasive repeated evolution. Our objectives were i) assess to what extent expression and sequence differences between ecotypes affect the same genes, ii) determine the level of correspondence between gene expression divergence and coding sequence divergence, and iii) quantify how natural selection may affect repeatability. Our results stress the important contribution that both gene regulation and coding regions can make to rapid phenotypic evolution and adaptation.

Materials and Methods

Sampling. Adult snails were collected in August 2010 from three Galician (NW Spain) localities: Burela (N 43°40'54.8"; W 7°22'1.8"), Roncudo (N 43°16.31'32"; W 8°59.23'93"), and Silleiro (N 42°6'17.20"; W 8°53'56.59") (Fig. 1). At each locality, specimens from the “crab” and the “wave” ecotypes were obtained from the upper and lower shore level respectively to avoid collecting intermediate forms (i.e. hybrids). Snails were immediately transported to our facilities in the ECIMAT Marine Station, and sexed according to the presence or absence of penis. Specimens used for DNA extractions were stored at –80 °C until processed. Specimens targeted for expression analysis were maintained alive in an aquarium under identical environmental conditions for two weeks using a continuous sea water flow (16 °C, 36.1‰ salinity, 6.8 mg/L oxygen level). This step aimed to minimize the impact of environmental variance on gene expression patterns by ensuring that all individuals shared

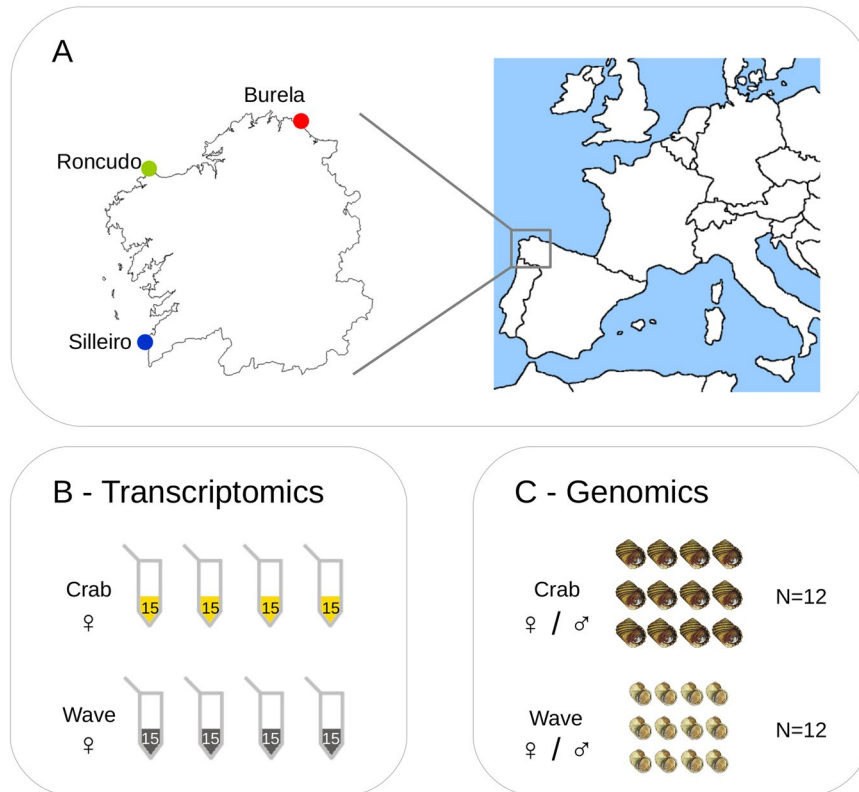


Figure 1. Experimental design. **(A)** Geographical situation of the three localities sampled in this study. **(B)** Transcriptomic scan: 4 pools were assayed for each ecotype and locality (24 pools in total); each pool consisted of 15 females with an equimolecular contribution to the pool. **(C)** Genomic sequence scan: 12 individuals were assayed for each ecotype and locality (72 individuals in total); each single individual was separately hybridized against the array.

the same environmental conditions prior to expression analysis. After this period, snails were stored at -20°C in RNAlater solution (Ambion) until RNA extraction.

RNA and DNA extraction. Total RNA was isolated from the foot muscle tissue of single females using TRIZOL reagent (Invitrogen) according to the manufacturer's instructions. Females have the advantage of providing larger RNA yields than males given their bigger size, while displaying expression patterns similar to those from males across the different ontogenetic stages of each ecotype⁵⁶. The Turbo DNA-freeTM kit (Ambion) was used to remove any remaining DNA from RNA extractions. Genomic DNA was isolated from the foot muscle tissue of single males and females using a CTAB extraction method⁵⁷ modified to include RNase treatment. DNA samples were further cleaned with NucleoSpin columns (Macherey-Nagel) following manufacturer's instructions. RNA and DNA purity was assessed using a NanoDrop spectrophotometer (NanoDrop Tech. Inc., Wilmington, DE). All extractions were standardized to 100 ng/ μL after checking their integrity in agarose gels.

NimbleGen *L. saxatilis* microarray. The *L. saxatilis* oligonucleotide microarray⁵⁸ was developed by NimbleGen Roche (090824_L_saxatilis_expr_HX12, $12 \times 135\text{K}$ array format) on the basis of draft or versioned assemblies from the *Littorina saxatilis* EST database⁵⁹ and the GenBank database. The microarray contained sequence information based on 25,205 partial transcripts, hereafter referred to as "genes" for simplicity, and which represent the coding part of the genome. These transcripts were obtained mainly by 454 sequencing of cDNA libraries from both the "crab" and "wave" ecotypes⁵⁹. Each gene was usually represented on the array by five non-overlapping 60-nt probes. Each NimbleGen slide contained 12 identical subarrays. We used this microarray to assess variation in gene expression and also in genomic sequence using, for the latter, a comparative genomic hybridization (CGH) approach, which is based on hybridization of labeled DNA fragments to a microarray⁴⁶. This allowed us to compare variation in expression and nucleotide sequence for the same subset of the *L. saxatilis* genome.

Gene expression profiling. For this analysis, pools of total RNAs were retrotranscribed to cDNAs representing the coding part of the transcriptome, which were then compared to establish patterns of over- and under-expressed genes. We prepared four replicate samples from each ecotype and locality (24 samples in total), each including 15 pooled female specimens with an equimolar RNA contribution of each specimen to the pool. The RNA from each pool was retrotranscribed with the SuperScriptTM Double-Stranded cDNA Synthesis Kit (Invitrogen) following the conditions recommended by the manufacturer. Cy3 labeling was performed with the

NimbleGen One-Color DNA labeling kit (Roche) using a starting amount of 1 µg of cDNA per pool. Then, for each pool, 4 µg of Cy3 labeled cDNA was resuspended in 12 µL of hybridization solution, of which 6 µL was applied onto a subarray. Pools were randomly distributed in the subarrays. Hybridization was carried out at 42 °C for 19 h on a NimbleGen Hybridization System with continuous mixing. After hybridization, the arrays were washed in buffers of various stringencies using the NimbleGen Washing kit. Arrays were scanned using an Agilent G2565AA microarray scanner (Agilent Technologies) with a resolution of 2 µm. The quality of the images was assessed using the NimbleScan v.2.5. software (NimbleGen/Roche), discarding those images with signal intensity or other metrics outside the range recommended by the manufacturer.

Data were extracted using NimbleScan v.2.5 and analyzed in the R/Bioconductor statistical environment. The `pdfInfoBuilder` and `oligo`⁶⁰ packages were used for data handling and pre-processing, with the robust multichip average (RMA) method⁶¹ used for background correction, quantile normalization and probe-level summarization of the microarray samples. A single value was obtained for each gene, resulting from each summarization of probe-level data. To decide whether a gene was expressed, a threshold level representing the “background signal” was calculated based on the average hybridization signal of the empty spots present in the array. Genes for which more than 20% of the probes had an average hybridization signal lower than the “background signal” were disregarded⁶².

Genomic divergence profiling. For the analysis of variation in genomic sequence, each subarray hosted the genomic DNA of one single individual and the genomic DNA of a common reference sample. Thus, in this experiment, genomic DNA was hybridized against the coding portion of the *L. saxatilis* genome represented in the microarray. The reference sample was composed of a DNA pool of 100 “crab” and “wave” snails from two British *L. saxatilis* locations (Dunvar and Thornwick, the latter used in the array design⁵⁸), to ensure consistent and non-zero hybridization signals for the reference sample in all the probes from the array. Therefore, “crab” and “wave” Galician ecotypes should be equally diverged from both the array and the reference, as the latter included mostly (90%) individuals from the same location used in the array design.

The NimbleGen/Roche Dual label kit was used to label the reference sample (Cy5 dye) and the DNA of each specimen (Cy3 dye) following manufacturer’s instructions. Hybridization procedures, array scanning and quality control of the images obtained were performed as in the expression arrays, but increasing the amount of labeled DNA (20 µg) and incubation time (72 h). The whole experiment included 72 Galician snails (12 per ecotype and locality) for which genomic DNA extracts were individually hybridized to the array. The data from scanning pictures generated by NimbleScan were parsed using `ringo`⁶³, an R/Bioconductor package. The signal intensity data for each channel was corrected for the local background signal using the `normexp + offset` method⁶⁴, `log2`-transformed, and quantile normalized using the method proposed for two channels⁶⁵, as implemented in the package `limma` for R/Bioconductor⁶⁶. We performed a probe-level data analysis to test DNA sequence differences between the distinct gene fragments included in a probe set and the hybridized DNA.

Low hybridization signals (below 10.35) in the *L. saxatilis* microarray may correspond in some instances to probes spanning exon boundaries and/or untranslated regions⁵⁸. To account for this possible source of noise in our data, and also to exclude probes that were not accurately detected in the array, we have filtered these sequences by removing probes with an average hybridization signal lower than the “background signal” (i.e. 10.7) of each channel. Untranslated regions would similarly generate low hybridization signals in the expression study, and these were also removed from the data (see above). Only genes containing probes that simultaneously passed genome and expression profiling filters were used in the subsequent analyses, to ensure that all the probes/genes only span coding sequences.

Statistical analysis. Differential expression (genes) and genomic divergence (probes) were determined using the linear modeling analysis for microarrays implemented in the `limma` package⁶⁶ with empirical Bayes adjustment to the variance. For both expression and sequence divergence records, three different linear models were fitted to the data and contrast matrices were created to identify (i) differences between ecotypes, localities and their interaction, (ii) differences between ecotypes within localities, and (iii) differences between localities for each ecotype. The resulting *p*-values were corrected for multiple tests using the binomial sequential goodness of fit procedure (SGoF)⁶⁷ at $\alpha = 0.05$.

For genes/probes showing significant differences between ecotype pairs in the three localities examined, we computed the *p*-value that the observed parallelism could be due to chance alone using both a randomization test⁶⁸ and the algorithm developed by Derome *et al.*⁶⁹ modified to include three localities (P. Duchesne, personal communication). Randomization tests were also used to estimate the random expectation of parallel and nonparallel changes, and of directional and nondirectional changes. For each randomization test, data were sorted 200,000 times and the corresponding outcome was obtained after multitest correction. An unpaired *t* test⁷⁰ was used to test that intrapopulation variance differs between genes/probes showing directional versus nondirectional parallel changes. The program `Blast2GO`⁷¹ was used to identify which GO terms were significantly over-represented in those genes or probes showing significant differences for each analysis. General patterns of gene expression and sequence divergence were visualized with heat maps using R/Bioconductor.

Results

We examined transcriptomes from pools including snails from the “crab” or “wave” ecotypes, and variation in the coding sequences of single snails. Snails were collected from three isolated, independently evolved population pairs of sympatric “crab” and “wave” ecotypes (Fig. 1) that previously showed a repeatable morphological divergence by parallel evolution^{33,35,40}. Variation in expression and genomic sequence was determined for the same genes using a microarray specifically developed for *L. saxatilis*. After quality control of the hybridized arrays, we retained 22 out of 24 pools for gene expression, 69 out of 72 individuals for coding sequence divergence, and 17,431 genes.

Locality	Transcriptomics	Genomics	Transcriptomics and Genomics
Burela	17.2% (2998)	13% (2266)	1.9% (331)
Roncudo	6.4% (1115)	25.5% (4445)	2% (349)
Silleiro	5.9% (1028)	9.7% (1691)	1.1% (192)
N° of genes with parallel changes	0.8% (146)	1.2% (216)	0.08% (15)

Table 1. Number of genes (percentage and absolute values) showing transcriptomic or genomic sequence differences between pairs of ecotypes in each single locality, and across the three localities simultaneously (parallel changes), after using SGoF multitest correction ($\alpha = 0.05$). The last column shows the genes that were divergent for gene expression and genomic sequence simultaneously. The total number of genes investigated is 17,431.

Table 1 shows the proportion of genes displaying expression and genomic sequence differences between pairs of ecotypes for the three localities examined after using SGoF multitest correction ($\alpha = 0.05$). Pairs of ecotypes living in the same site displayed significant differences in expression and genomic sequence, respectively, for up to 17.2 and 25.5% of all assayed genes. Transcriptomic differences were more prevalent than genomic differences in only one of the three localities assayed. The majority of divergent genes (93–94%) were divergent either for gene expression or genomic sequence, but not for both simultaneously. The number of differences between ecotype pairs varied among localities ($P < 0.05$; G test), and many of these differences (40.6–79% for gene expression; 68–71% for genomic divergence) occurred only in a single locality. Therefore, we tested whether differences between ecotype pairs frequently involved the same genes in the three localities (i.e. parallel changes). The observed numbers of genes with parallel changes in expression and genomic sequence were, respectively, 146 (0.8% of all assayed genes) and 216 (1.2%). Such levels of parallelism are highly unlikely just by chance ($p < 10^{-5}$ for both expression and genomic data using a permutation test, or the algorithm by Derome *et al.*⁶⁹), and therefore consistent with repeatable genetic differentiation by natural selection. Remarkably, as few as 15 genes displayed simultaneous parallel changes in expression and genomic sequence, representing 4% of all genes with parallel changes.

We examined the directionality of observed parallel differences. To increase precision, genomic divergence is referred to in the subsequent analyses in terms of the 354 probes rather than the 216 genes that showed parallel changes (see methods). The majority of parallel differences between ecotype pairs were due to changes in the same direction (directional changes), whereas only a few were due to differences in opposite directions (non-directional changes) (Fig. 2). Specifically, up to 132 (90%) of all genes displaying parallel differences in expression showed directional changes (54% of which were up-regulated in the “crab ecotype”). Similarly, 294 (83%) of all probes with parallel variation in genomic sequence also showed directional changes (75% of which displayed a higher hybridization signal in “crab” than “wave” snails). The large over-representation of directional parallel differences for both expression and sequence divergence data is highly unlikely just by chance (each $p < 0.0001$), as determined by randomization tests resorting expression or genomic data sets (Fig. 2).

We also determined whether the mean intrapopulation variance differs between genes/probes showing directional versus nondirectional parallel changes. We found that variance in expression and sequence divergence for directional changes was twice less than that observed for nondirectional changes (Fig. 3). These differences were statistically significant for both variation in expression ($p = 0.0007$) and genomic sequence ($p = 0.0185$) using a randomization test, and also using 2-tailed *t* tests (all $p < 0.027$).

To further assess the nature of evolutionary forces underlying parallel variation, we determined which proportion of genes/probes showing parallel and nonparallel differences among ecotype pairs also showed a significant geographic differentiation among the three localities for the “crab” or “wave” ecotypes. We found that, independently of the ecotype considered, genes/probes with parallel changes showed more frequently geographic differentiation than genes/probes with nonparallel changes after SGoF multitest correction ($\alpha = 0.05$), and that this pattern was more pronounced for genomic than transcriptomic differences (Fig. 4). The proportion of genes/probes with parallel changes that displayed geographic differentiation deviated more strongly ($p < 0.001$) from the random expectation than the proportion observed for nonparallel changes.

We used an enrichment analysis with BLAST2GO to test whether parallel differentiation is linked with specific functional groups. This analysis did not identify enriched gene/probe sets after correction for multiple testing when the whole data set or only intra-site GO enrichment tests were considered. Nevertheless, some genes/probes with the most extreme parallel directional changes in hybridization signal included annotations expected to be involved in adaptation. For example, the Dermatoptin 2 (for gene expression profiling) and the Keratin-associated protein 4–3 (for sequence divergence profiling) are involved, respectively, in the formation of the shell⁷² and the operculum⁷³, key features defining differences between ecotype pairs (Supplementary Tables S1 and S2). We also tested whether the differences between ecotype pairs that are unique to each locality are linked with specific functional groups. For this analysis, significant enrichment GO terms were observed only for gene expression profiling after correction for multiple testing. We observed an important enrichment in energetic metabolism GO terms for Burela, but almost no GO terms were shared among pairs of localities, and none between the three localities simultaneously, either for the categories of molecular function, biological process, or cellular component (Supplementary Figs S1 and S2).

Discussion

Parallelism in gene expression and coding sequences. We report evidence that parallel differences in expression and sequence divergence of a limited set of genes underlay the repeated phenotypic divergence of replicate pairs of *L. saxatilis* ecotypes. The observed numbers of parallel differences in gene expression and sequence divergence largely exceeded the random expectation. Several results suggest that adaptive selection played a role, direct or indirect, in the process of molecular divergence among ecotypes.

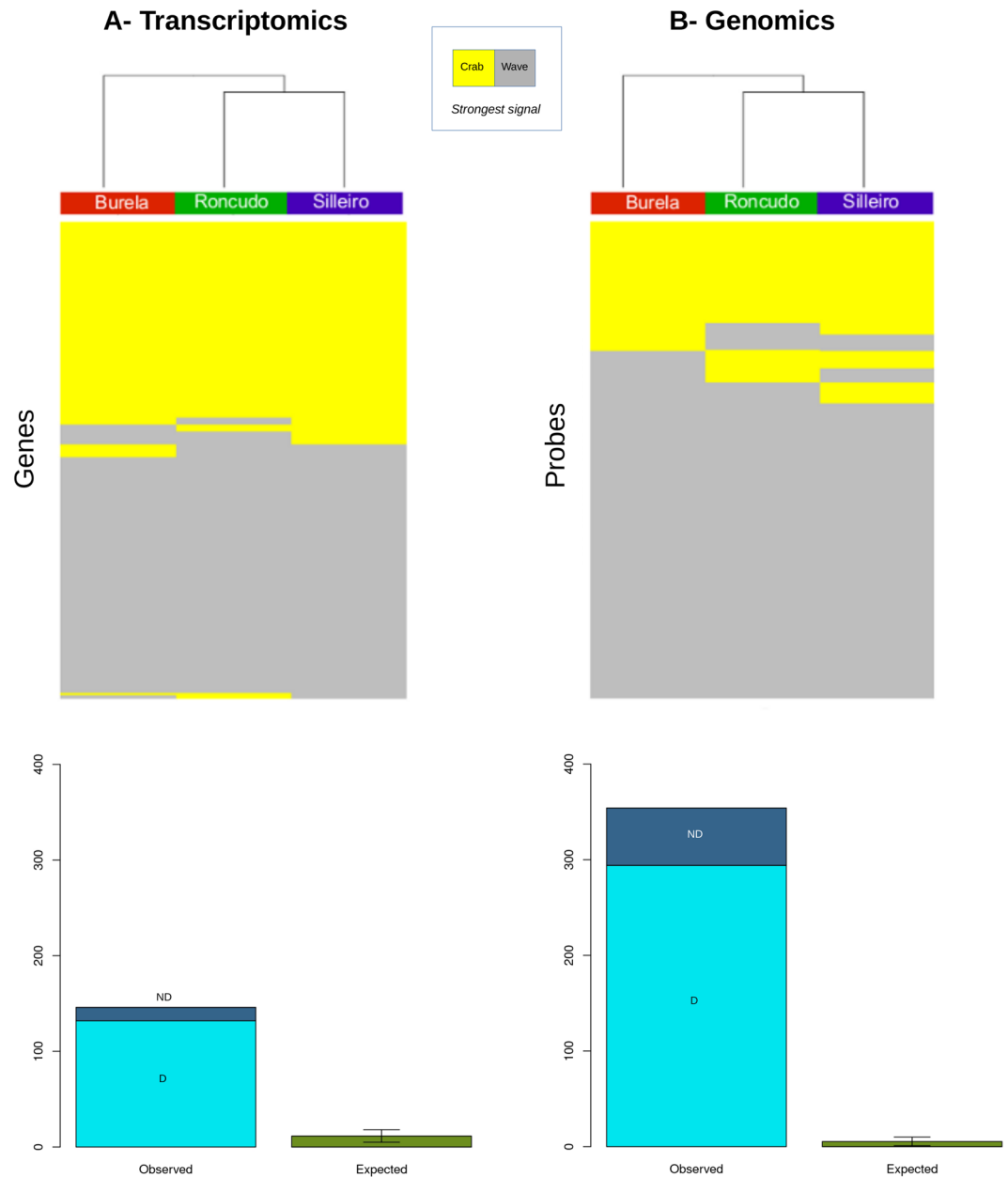


Figure 2. Directionality of observed parallel differences for (A) the transcriptomic and (B) genomic sequence scans. Top panels: heat maps of hybridization signals across localities for the 146 genes (left) and 354 probes (right) that displayed significant parallel changes after multitest correction. Yellow indicates a greater hybridization signal for the “wave ecotype”, and grey a higher hybridization signal for the “crab ecotype”. Directional parallel changes appear as straight horizontal lines of one single color (yellow or grey). Nondirectional parallel changes display horizontal lines scattered of yellow and grey stretches. Bottom panels: barplots indicating the observed number of directional (D) and nondirectional (ND) changes versus the random expectation, obtained using a randomization test (see methods). Vertical error lines indicate the 95% confidence intervals across all replicates of the randomization test.

We expect that genes repeatedly recruited by strong natural selection would show striking habitat-associated differences⁷⁴, would display less variation than those under weaker selection⁶⁹, and would show a higher geographical differentiation⁷⁵. Overall, our results fit these expectations and are consistent with a scenario in which the same subset of genes, or regulatory regions, were repeatedly recruited by natural selection in populations adapted to similar habitats. Parallel changes in hybridization signal were nearly restricted to directional changes, denoting a repeated and significant habitat-association among independently evolving populations of similar phenotype that cannot be explained by chance. Therefore, directional parallel changes showed a lower intrapopulation variance than nondirectional parallel changes, as expected from a stronger impact of selection in the former^{69,76}. Also, the distinctive higher geographic differentiation in expression and coding sequence for genes

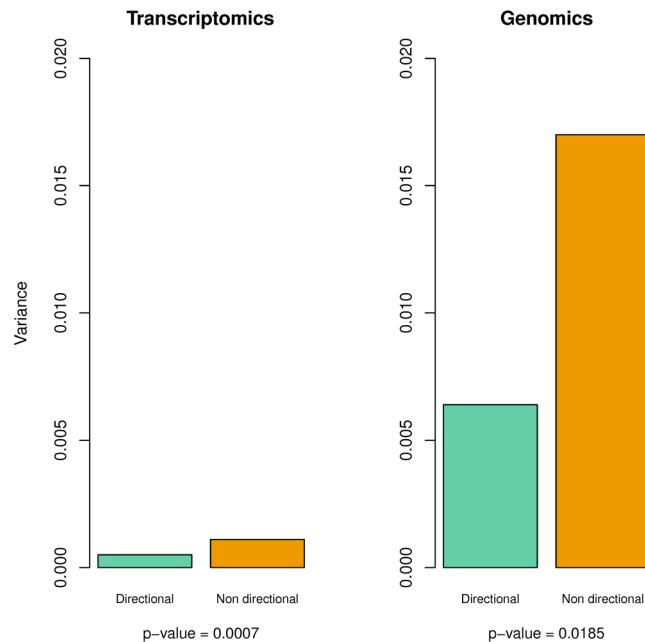


Figure 3. Mean intrapopulation variance in hybridization signal for directional parallel changes and nondirectional parallel changes. Variance for directional changes was significantly lower than that observed for nondirectional changes for both transcriptomic and genomic sequence scans.

displaying parallelism did not match the random expectation. This in turn suggest that geographic differentiation for genes showing parallelism is determined by the joint action of divergent selection and stochastic forces, whereas geographic differentiation at nonparallel genes is mostly driven by stochastic forces. Last, we examined the function of genes with parallel divergence. Although annotation was very incomplete due to the poor representation of mollusk sequences in public databases⁷⁷, some of the genes that could be annotated exhibited functions related with well know adaptive phenotypic characters, such as the formation of the snail shell and the operculum.

Despite the observed parallelism, the majority of differences in gene expression and coding sequence were not shared among localities. Low sharing of divergent genes contrasts with the expectation of high gene reuse for the parallel evolution of individual phenotypic traits among closely related taxa and populations⁹. Several non-mutually exclusive factors may account for this discrepancy. First, we might have underestimated the parallelism existing in natural populations. For example, parallelism owing to low diverged alleles, or to alleles equally diverged from the reference but carrying mutations at different sequence positions, could remain somewhat undetected using microarrays. Sequence mismatches due to sequence polymorphisms could also affect the ability to detect parallelism in gene expression. Therefore, the number of genes showing parallelism in our study should be viewed as conservative. Yet, the impact of these challenges on our patterns of parallelism seems to be modest since we detected many differences between ecotype pairs of a very recent origin within each locality, and still only a minor fraction of these differences were repeatable among localities. Moreover, expression measurements in different species did not reveal a consistent variation in signal intensity due to sequence mismatches^{24,78}, since the expression of each gene is calculated as the average intensity for each probe set. Second, if divergent traits in *Littorina* (e.g. shell size and shell shape) are highly polygenic, then they may show greater genetic redundancy than traits determined by a single gene or molecular pathway. As such, changes in different pathways of a complex polygenic trait could lead to similar phenotypes and show less repeatable genetic signatures of adaptation^{3,22}. Third, patterns of parallel evolution could be more common at higher levels of biological organization⁷⁹. Thus, sharing of physiological processes, biochemical pathways, or organismal functions may therefore be more prevalent than observed at the gene or regulatory level^{7,80–83}. Last, a number of biases could have inflated the very high expectation of gene reuse, such as publication bias against non-sharing genetic patterns, or an emphasis on genes of large effect that may not be illustrative of the true spectrum of phenotypes^{3,9,84}.

Other studies in a number of different organisms have similarly demonstrated little sharing of sequence divergence^{10,13,85,86} and gene expression patterns^{69,87} linked to recent parallel evolution. Our findings are consistent with recent genome scan studies in *Littorina* indicating a low sharing of genomic divergence among ecotypes that arose in parallel in different parts of Europe but also, as shown in Sweden, in geographically close localities^{42,44}. Moreover, parallelism between ecotype pairs mostly involved genomic regions under strong selection^{42,43}, thus supporting our hypothesis that genes showing shared genomic and expression divergence are likely targeted by natural selection.

About 10% of sequence differences in the *Littorina* array are expected to be copy number variants⁵⁸. Thus, processes such as duplication and subsequent neofunctionalization might also play a role in the divergence among ecotypes^{4,22,88}. Our results show that the *Littorina* microarray is able to detect more sequence differences among

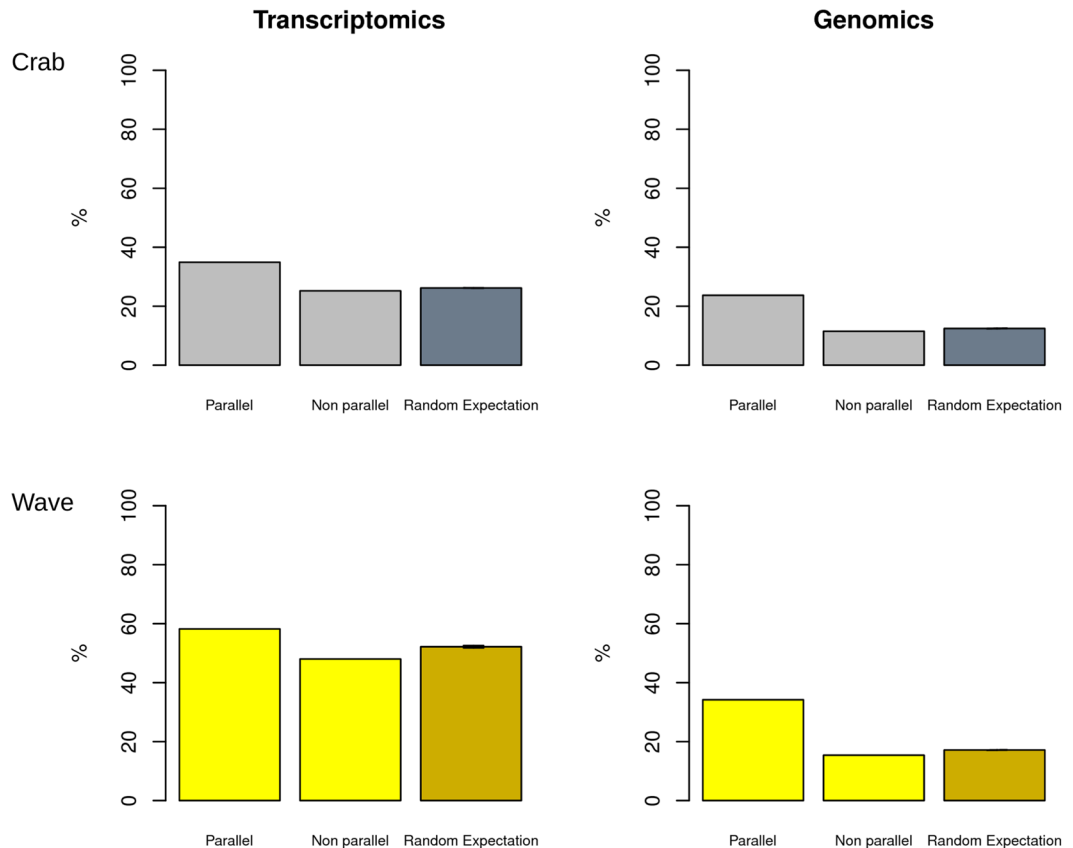


Figure 4. Geographical differentiation for genes (left) and probes (right) that displayed parallelism in gene expression or genomic sequence, and for those not showing parallelism. The y axis indicates the percentage of cases for which these genes/probes showed statistically significant differences among the three localities. Vertical error lines indicate the 95% confidence intervals for the random expectation (almost not appreciable in the figure), obtained resampling 1000 times at random the same number of genes/probes showing either parallel or nonparallel changes from the total set of genes/probes.

ecotype pairs than reported in a previous study using this same microarray⁵⁸. Several reasons explain this gain in power. In the former study, a reference sample was not used and data was not filtered, thus increasing the inter-array variance due to technical noise effects⁵⁹. Also, a probe-based analysis was not used to assess sequence differences. Instead, the relative hybridization signal for each gene represented on the array was calculated as the *average* intensity for *each probe set*. Thus, any mismatch signal resulting from a target DNA polymorphism affecting one single probe would be averaged with the remaining gene probes and therefore would be difficult to detect. In addition, sequence comparisons between ecotype pairs within a single locality were based on only four individuals from each ecotype. To obtain more power, in the present study the sample size was increased to 12 “crab” and 12 “wave” individuals per locality (72 individuals in total versus 8 in the former study for Galician snails). Finally, we used the limma package⁶⁶ with empirical Bayes adjustment to the variance, that allows an improved estimation of variance respect to the conventional ANOVA tests previously used.

The chances of successfully capturing adaptive loci are greater when targeting functionally important regions. In this study, we simultaneously screened patterns of expression and sequence variation for the coding fraction of the genome. Thus, although some of the genes displaying parallel evolution could be false positives, the likelihood of identifying genes directly or indirectly (through linked selection) targeted by selection may be substantial. Remarkably, a large number of divergence events occurred in a single ecotype pair. An unknown proportion of this non-shared divergence could have resulted from stochastic processes, adaptive changes, or a combination of these factors. Overall, our results suggest that the genomic architecture underlying parallel phenotypic divergence probably followed a complex evolutionary path, affecting multiple loci in a mosaic pattern of both repeatable and idiosyncratic divergence, and where the repeated element involved many regions affected by natural selection.

Divergence in gene expression is decoupled from divergence in coding sequence. Our results showed that patterns of differentiation in gene expression and coding sequence were markedly dissimilar. The majority of divergent genes were divergent either for gene expression or genomic sequence, but not for both simultaneously. Remarkably, as few as 15 genes displayed simultaneous parallel changes in expression and genomic divergence, representing 4% of all genes with parallel changes. Such result is consistent with the hypothesis that expression and gene sequence differences underlying phenotypic divergence could, at least to a certain extent, be considered decoupled processes³¹.

One concern is that the comparison between expression and sequence variation could have been partly affected by misleading expression measurements resulting from sequence mismatches between the samples used for expression analysis and the reference upon which the array was designed. However, sequence mismatches cannot account for the dissimilarity in patterns of differentiation, since such mismatches should also be present in the samples used for sequence differentiation and would generate a correlated signal between gene expression and sequence divergence⁹⁰. It might be also possible that our genome scan was not sensitive enough to pick up all the genes carrying a single nucleotide variant difference. However, this lack of sensitivity should equally affect the coding regions of genes displaying either expression or no expression differences, and thus cannot explain the dissimilarity. Also, for gene duplications where both genes are retained, similar patterns of differentiation are expected for gene expression and gene sequence if both diverge at clock-like rates⁹¹. It is also unlikely that power differences between expression and sequence divergence studies can account for the dissimilarity in patterns of differentiation, as they should lead to consistently larger differences between ecotype pairs for one such level (expression or sequence divergence) in the three localities examined and, therefore, genes with significant differences in the less powerful study should also display concordant significant differences for the most powerful one. These patterns are not observed in our data (Table 1). These considerations further support that, independently of the source of variation or error considered, gene expression and coding sequences appear to evolve differently as ecotypes repeatedly adapt to complex ecological gradients.

The decoupling between gene expression and coding sequence differentiation is consistent with the existence of trans-regulation factors driving gene expression evolution, but also with the evolutionary decoupling of cis-regulatory regions and coding sequences. For example, in *Drosophila melanogaster*, gene expression differences among distinct strains are correlated with 5 prime sequences but not with coding sequences, thus supporting that differentiation at cis-regulatory regions is decoupled from differentiation at coding regions⁹². Our results are in line with what was observed among closely related ecotypes of lake whitefish²⁰, rainwater killifish⁹³, and woody sunflower²⁹, where differentiation of gene expression and coding sequences were also decoupled. However, what distinguishes our study from these previous ones is that we focus on genes displaying parallel evolution across similar environmental gradients. As such, the genes we identify are more likely to underlie variability related with traits implied in a relevant adaptive response.

Functional interpretations of the decoupling between gene expression and sequence divergence should be taken cautiously, as array data do not allow to tell apart effects due to nonsynonymous mutations that alter the amino acid sequence from those due to synonymous mutations that do not affect the amino acid composition. Yet, even if most of changes occurred at synonymous sites, it would be needed to explain why in our data differentially expressed genes do not show such changes. This would point to the existence, even for synonymous sites, of selective constraints slowing down the evolution of coding sequences for genes displaying parallel changes in expression. Indeed, evidence exists indicating that synonymous sites appear to evolve slower than expected under neutrality in a way apparently consistent with weak selection in organisms as diverse as insects, yeast, worms, chicken or mammals^{94–98}. For example, in *D. melanogaster*, 22% of four-fold synonymous sites are evolving under strong constraints, and genes with such constrained sites tend to be especially relevant, highly expressed, and often involved in developmental networks⁹⁹. Therefore, our results may point to the possibility of some division of tasks underlying coding and regulatory regions, as previously hypothesized¹⁰⁰.

Genome-wide data on expression variation versus sequence divergence are uncommon. Thus, this study provides a rare opportunity to determine the relative contribution of expression and coding changes underlying parallel phenotypic evolution. Our results suggest that both coding and expression changes contribute to parallel divergence among pairs of ecotypes. The relative contribution of expression and sequence changes varied among localities, but there was not an overall preeminent role of expression over coding sequence differences across all localities. Our results differ from other studies in three-spined sticklebacks providing a major role to gene expression variation (up to 83% of all differences) over coding sequence variation in the evolution of parallel phenotypic divergence¹⁶. This suggests that differences in life history features and the number, location and interactions among genes and regulatory regions, may generate very diverse outcomes in the molecular fingerprint underlying phenotypic adaptation²³.

Data Availability

The expression and genomic divergence dataset is available in the NCBI gene expression Omnibus under the accessions GSE120697 and GSE120698 respectively.

References

- Brandvain, Y. & Wright, S. I. The limits of natural selection in a nonequilibrium world. *Trends Genet.* **32**, 201–210 (2016).
- Schluter, D. & Nagel, L. M. Parallel speciation by natural selection. *Am. Nat.* **146**, 292–301 (1995).
- Elmer, K. R. & Meyer, A. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol. Evol.* **26**, 298–306 (2011).
- Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013).
- Faria, R. *et al.* Advances in ecological speciation: an integrative approach. *Mol. Ecol.* **23**, 513–521 (2014).
- Westram, A. & Johannesson, K. *Parallel speciation In: Encyclopedia of Evolutionary Biology* (ed. Kliman, R.) 212–219 (Elsevier, 2016).
- Losos, J. B. Convergence, adaptation, and constraint. *Evolution* **65**, 1827–1840 (2011).
- Wagner, M. R. & Mitchell-Olds, T. Repeated phenotypic changes highlight molecular targets of convergent evolution. *Genome Biol.* **12**, 124 (2011).
- Conte, G. L., Arnegard, M. E., Peichel, C. L. & Schluter, D. The probability of genetic parallelism and convergence in natural populations. *Proc. R. Soc. Lond. B Biol. Sci.* **279**, 5039–5047 (2012).
- Kautt, A. F., Elmer, K. R. & Meyer, A. Genomic signatures of divergent selection in a “natural experiment”, the young parallel radiations of Nicaraguan crater lake cichlid fishes. *Mol. Ecol.* **21**, 4770–4786 (2012).

11. Roda, F. *et al.* Convergence and divergence during the adaptation to similar environments by an Australian groundsel. *Evolution* **67**, 2515–29 (2013).
12. Perreault-Payette, A. *et al.* Investigating the extent of parallelism in morphological and genomic divergence among lake trout ecotypes in Lake Superior. *Mol. Ecol.* **26**, 1477–1497 (2017).
13. Soria-Carrasco, V. *et al.* Stick insect genomes reveal natural selection's role in parallel speciation. *Science* **344**, 738–742 (2014).
14. Bradic, M., Teotónio, H. & Borowsky, R. L. The population genomics of repeated evolution in the blind cavefish *Astyanax mexicanus*. *Mol. Biol. Evol.* **30**, 2383–2400 (2013).
15. Manceau, M., Domingues, V. S., Linnen, C. R., Rosenblum, E. B. & Hoekstra, H. E. Convergence in pigmentation at multiple levels: mutations, genes and function. *Philos. Trans. R. Soc Lond B. Biol. Sci.* **365**, 2439–2450 (2010).
16. Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
17. Renaut, S., Grassa, C., Moyers, B., Kane, N. & Rieseberg, L. The population genomics of sunflowers and genomic determinants of protein evolution revealed by RNA-seq. *Biology* **1**, 575–596 (2012).
18. Holliday, J. A., Zhou, L., Bawa, R., Zhang, M. & Oubida, R. W. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. *New Phytol.* **209**, 1240–1251 (2016).
19. Khaitovich, P. *et al.* Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309**, 1850–1854 (2005).
20. Jeukens, J., Renaut, S., St-Cyr, J., Nolte, A. W. & Bernatchez, L. The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Mol. Ecol.* **19**, 5389–5403 (2010).
21. Zhao, L., Wit, J., Svetec, N. & Begun, D. J. Parallel gene expression differences between low and high latitude populations of *Drosophila melanogaster* and *D. simulans*. *PLoS Genet.* **11**, e1005184, <https://doi.org/10.1371/journal.pgen.1005184> (2015).
22. Yeaman, S. *et al.* Convergent local adaptation to climate change in distantly related conifers. *Science* **353**, 1431–1433 (2016).
23. Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo-devo and the genetics of adaptation. *Evolution* **61**, 995–1016 (2007).
24. Nuzhdin, S. V., Wayne, M. L., Harmon, K. L. & McIntyre, L. M. Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol. Biol. Evol.* **21**, 1308–1317 (2004).
25. Jordan, I. K., Mariño-Ramírez, L. & Koonin, E. V. Evolutionary significance of gene expression divergence. *Gene* **345**, 119–126 (2005).
26. Hunt, B. G., Ometto, L., Keller, L. & Goodisman, A. D. Evolution at two levels in fire ants: the relationships between patterns of gene expression and protein sequence evolution. *Mol. Biol. Evol.* **30**, 263–271 (2012).
27. Hodgins, K. A., Yeaman, S., Nurkowski, K. A., Riesenber, L. H. & Aitken, S. N. Expression divergence is correlated with sequence evolution but not positive selection in conifers. *Mol. Biol. Evol.* **33**, 1502–1516 (2016).
28. Tirosh, I. & Barkai, N. Evolution of gene sequence and gene expression are not correlated in yeast. *Trends Genet.* **24**, 109–113 (2008).
29. Moyers, B. T. & Riesenber, L. H. Divergence in gene expression is uncoupled from divergence in coding sequence in a secondarily woody sunflower. *Int. J. Plant. Sci.* **174**, 1079–1089 (2013).
30. Warnefors, M. & Kaessmann, H. Evolution of the correlation between expression divergence and protein divergence in mammals. *Genome Biol. Evol.* **5**, 1324–1335 (2013).
31. Tirosh, I., Reikhav, S., Levy, A. & Barkai, N. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* **324**, 659–62 (2009).
32. Haygood, R., Babbitt, C. C., Fedrigo, O. & Wray, G. A. Contrasts between adaptive coding and noncoding changes during human evolution. *Proc. Natl. Acad. Sci USA* **107**, 7853–7857 (2010).
33. Rolán-Alvarez, E. *et al.* Nonallopatric and parallel origin of local reproductive barriers between two snail ecotypes. *Mol. Ecol.* **13**, 3415–24 (2004).
34. Panova, M., Hollander, J. & Johannesson, K. Site-specific genetic divergence in parallel hybrid zones suggests non-allopatric evolution of reproductive barriers. *Mol. Ecol.* **15**, 4021–4031 (2006).
35. Butlin, R. K. *et al.* Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution* **68**, 935–49 (2014).
36. Rolán-Alvarez, E., Austin, C. & Boulding, E. G. The contribution of *Littorina* to the field of Evolutionary Ecology. *Oceanography and Marine Biology, an Annual Review* **53**, 157–214 (2015).
37. Johannesson, K., Rolán-Alvarez, E. & Ekendahl, A. Incipient reproductive isolation between two sympatric morphs of the intertidal snail *Littorina saxatilis*. *Evolution* **49**, 1180–1190 (1995).
38. Johannesson, K. *et al.* Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. *Philos. Trans. R. Soc Lond. B Biol. Sci.* **365**, 1735–1747 (2010).
39. Rolán-Alvarez, E., Johannesson, K. & Erlandsson, J. The maintenance of a cline in the marine snail *Littorina saxatilis*: the role of home site advantage and hybrid fitness. *Evolution* **51**, 1838–1847 (1997).
40. Quesada, H., Posada, D., Caballero, A., Morán, P. & Rolán-Alvarez, E. Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution* **61**, 1600–1612 (2007).
41. Pérez-Pereira, N., Quesada, H. & Caballero, A. Can parallel ecological speciation be detected with phylogenetic analyses? *Mol. Phylogenet. Evol.* **166**, 149–156 (2017).
42. Westram, A. M. *et al.* Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Mol. Ecol.* **23**, 4603–4616 (2014).
43. Westram, A. M., Panova, M., Galindo, J. & Butlin, R. K. Targeted re-sequencing reveals geographic patterns of differentiation for loci implicated in parallel evolution. *Mol. Ecol.* **25**, 3169–3186 (2016).
44. Ravinet, M. *et al.* Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol. Ecol.* **25**, 287–305 (2016).
45. Goodwin, S., McPherson, J. D. & McCombie, R. C. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
46. Gresham, D., Dunham, M. J. & Botstein, D. Comparing whole genomes using DNA microarrays. *Nat. Rev. Genet.* **9**, 291–302 (2008).
47. Renn, S. C. P. *et al.* Using comparative genomic hybridization to survey genomic sequence divergence across species: a proof-of-concept from *Drosophila*. *BMC Genomics* **11**, 271 (2010).
48. Bottomly, D. *et al.* Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-seq and microarrays. *PLoS One* **6**, e17820 (2011).
49. Sirbu, A., Kerr, G., Crane, M. & Ruskin, H. J. RNA-seq vs dual- and single-channel microarray data: sensitivity analysis for differential expression and clustering. *PLoS One* **7**, e50986 (2012).
50. Robinson, D. G., Wang, J. Y. & Storey, J. D. A nested parallel experiment demonstrates differences in intensity-dependence between RNA-seq and microarrays. *Nucleic. Acids. Res.* **43**, e131 (2015).
51. Hoen, P. A. *et al.* Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic. Acids. Res.* **36**, e141 (2008).
52. Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and microarray in transcriptomic profiling of activated T cells. *PLoS One* **9**, e78644 (2014).

53. Willenbrock, H. *et al.* Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing. *RNA* **15**, 2028–2034 (2009).
54. McIntyre, L. M. *et al.* RNA-seq: technical variability and sampling. *BMC Genomics* **12**, 293 (2011).
55. Tirado, T., Saura, M., Rolán-Alvarez, E. & Quesada, H. Historical biogeography of the marine snail *Littorina saxatilis* inferred from haplotype and Shell morphology evolution in NW Spain. *Plos One*, **11**, e0161287, <https://doi.org/10.1371/journal.pone.0161287> (2016).
56. Diz, A. P., Páez de la Cadena, M. & Rolán-Alvarez, E. Proteomic evidence of a paedomorphic evolutionary process within a marine snail species: a strategy for adapting to extreme ecological conditions? *J. Evol. Biol.* **25**, 2569–2581 (2012).
57. Wilding, C. S., Butlin, R. K. & Grahame, J. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J. Evol. Biol.* **14**, 611–619 (2001).
58. Panova, M. *et al.* Species and gene divergence in *Littorina* snails detected by array comparative genomic hybridization. *BMC Genomics* **15**, 687 (2014).
59. Canbäck, B. *et al.* The *Littorina* sequence database (LSD) – an online resource for genomic data. *Mol. Ecol. Res.* **12**, 142–148 (2012).
60. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363–2367 (2010).
61. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
62. Stolc, V., Samanta, M. P., Tongprasit, W. & Marshall, W. F. Genome-wide transcriptional analysis of flagellar regeneration in *Chlamydomonas reinhardtii* identifies orthologs of ciliary disease genes. *Proc. Natl. Acad. Sci. USA* **102**, 3703–3707 (2005).
63. Toedling, J. *et al.* Ringo - an R/Bioconductor package for analyzing ChIP-chip readouts. *BMC Bioinformatics* **8**, p. 221 (2007).
64. Ritchie, M. E. *et al.* A comparison of background correction methods for two-color microarrays. *Bioinformatics* **23**, 2700–2707 (2007).
65. Yang, Y. H. & Thome, N. P. Normalization for two-color cDNA microarray data. *Lect. Notes Monogr. Ser.* **40**, 403–418 (2003).
66. Smyth, G. K. Limma: Linear Models for Microarray Data in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds Gentleman, R., Carey, V. J., Dudoit, S., Irizarry, R. & Huber, W.) 397–420 (Springer, New York, 2005).
67. Carvajal-Rodríguez, A., De Uña-Alvarez, J. & Rolán-Alvarez, E. A new multitest correction (SGoF) that increases its statistical power when increasing the number of tests. *BMC Bioinformatics* **10**, 209 (2009).
68. Manly, B. F. J. *Randomization, bootstrap and Monte Carlo methods in biology*, 3rd edn. (Chapman & Hall, London, 2006).
69. Derome, N., Duchesne, P. & Bernatchez, L. Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchell) ecotypes. *Mol. Ecol.* **15**, 1239–1249 (2006).
70. Draghici, S. *Data Analysis Tools for DNA Microarrays* (Chapman & Hall, London, 2003).
71. Conesa, A. *et al.* Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
72. Sarashina, I. *et al.* Molecular evolution and functionally important structures of molluscan Dermatopontin: Implications for the origins of molluscan shell matrix proteins. *J. Mol. Evol.* **62**, 307–318 (2006).
73. Grégoire, C. *Structure of the molluscan shell in Chemical Zoology IV* (eds Florkin, M. & Sheer, B. T.) 45–102 (Academic Press, NY, 1972).
74. Nosil, P. *Ecological speciation*. Oxford University Press, Oxford. (2012).
75. Nielsen, R. Molecular signatures of natural selection. *Ann. Rev. Genet.* **39**, 197–218 (2005).
76. Falconer, D. S. Introduction to quantitative genetics. Longman, Harlow, 438 pp. (1989).
77. Ekblom, R. & Galindo, J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**, 1–15 (2011).
78. García, C., Avila, V., Quesada, H. & Caballero, A. Candidate transcriptome sources of inbreeding depression in *Drosophila melanogaster*. *Plos One* **8**, e70067 (2013).
79. Edelman, G. M. & Gally, J. A. Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. USA* **98**, 13763–13768 (2001).
80. Renaut, S., Owens, G. L. & Riesenberger, L. H. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Mol. Ecol.* **23**, 311–324 (2014).
81. Moczek, A. P. & Nagy, L. M. Diverse developmental mechanisms contribute to different levels of diversity in horned beetles. *Evol. Dev.* **7**, 175–185 (2005).
82. Steiner, C. C., Römpler, H., Boettger, L. M., Schöneberg, T. & Hoekstra, H. E. The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Mol. Biol. Evol.* **26**, 35–45 (2009).
83. Ralph, P. L. & Coop, G. Convergent evolution during local adaptation to patchy landscapes. *Plos Genet.* **11**, e1005630 (2015).
84. Hendry, A. P. *Eco-evolutionary dynamics*. Princeton University Press (2016).
85. Renaut, S., Nolte, A. W., Rogers, S. M., Derome, N. & Bernatchez, L. SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus* spp.). *Mol. Ecol.* **20**, 545–559 (2010).
86. Deagle, B. E. *et al.* Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proc. R. Soc Lond. B Biol. Sci.* **279**, 1277–1286 (2012).
87. St-Cyr, J., Derome, N. & Bernatchez, L. The transcriptomics of life-history trade-offs in whitefish pairs (*Coregonus* sp.). *Mol. Ecol.* **17**, 1850–1870 (2008).
88. Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M. & Andolfatto, P. Parallel molecular evolution in an herbivore community. *Science* **28**, 1634–1637 (2012).
89. Eisen, M. B. & Brown, P. O. DNA arrays for analysis of gene expression. *Methods Enzymol.* **303**, 179–205 (1999).
90. Lemos, B., Meiklejohn, C. D., Caceres, M. & Hartl, D. L. Rates of divergence in gene expression profiles of primates, mice, and flies: Stabilizing selection and variability among functional categories. *Evolution* **59**, 126–137 (2005).
91. Wagner, A. Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate. *Proc. Natl. Acad. Sci. USA* **97**, 6579–6584 (2000).
92. Kohn, M. H., Shapiro, J. & Wu, C. I. Decoupled differentiation of gene expression and coding sequence among *Drosophila* populations. *Genes Genet. Syst.* **83**, 265–273 (2008).
93. Kozak, G. M., Brennan, R. S., Berdan, E. L., Fuller, R. C. & Whitehead, A. Functional and population genomic divergence within and between two species of killifish adapted to different osmotic niches. *Evolution* **68**, 63–80 (2013).
94. Akashi, H. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* **139**, 1067–1076 (1995).
95. Chamary, J., Parmley, J. L. & Hurst, L. D. Hearing silence: Non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* **7**, 98–108 (2006).
96. Eöry, L., Halligan, D. L. & Keightley, P. D. Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. *Mol. Biol. Evol.* **27**, 177–192 (2010).
97. Zhou, T., Gu, W. & Wilke, C. O. Detecting positive and purifying selection at synonymous sites in yeast and worm. *Mol. Biol. Evol.* **27**, 1912–1922 (2010).
98. Künstner, A., Nabholz, B. & Ellegren, H. Significant selective constraint at 4-fold degenerate sites in the avian genome and its consequence for detection of positive selection. *Genome Biol. Evol.* **3**, 1381 (2011).

99. Lawrie, D. S., Messer, P. W., Hershberg, R. & Petrov, D. A. Strong purifying selection at synonymous sites in *D. melanogaster*. *Plos Genet.* **9**, e1003527 (2013).
100. Stern, D. L. & Orgogozo, V. The loci of evolution: how predictable is genetic evolution. *Evolution* **62**, 2155–2177 (2008).

Acknowledgements

We would like to thank the ECIMAT Marine Research Center (University of Vigo) for providing marine laboratory facilities. We are grateful to Pierre Duchesne for extending from two to three localities the algorithm for calculating the probability that the observed parallelism could be due to chance alone and help in calculating the corresponding *p*-values. This work was supported by Ministerio de Economía y Competitividad (codes BFU2013-44635-P, CGL2016-75482-P and CGL2016-75904-C2-1), Axudas do programa de consolidación e estruturación de unidades de investigacións competitivas do SUG, Xunta de Galicia (ED431C 2016-037), Fondos Feder: “Unha maneira de facer Europa”, Xunta de Galicia (INCITE09 310 006 PR) and the Swedish Research Councils VR and Formas (Linnaeus grant Formas 217-2008-1719).

Author Contributions

H.Q., K.J. and E.R.-A. conceived the study. H.Q. wrote the manuscript, with input from all authors. M.J.R. and M.S. conducted the experiments. M.J.R. and A.P.-F. performed the bioinformatic analyses with advises and support from H.Q., E.R.-A. and A.C. M.P., C.A. and T.J. designed the microarray and assisted in data analyses.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-33897-8>.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018