Proteomic analysis of F\textsubscript{1} hybrids and intermediate variants in a *Littorina saxatilis* hybrid zone

Angel P. Diz\textsuperscript{a,*},*, Mónica R. Romero\textsuperscript{a, #}, Juan Galindo\textsuperscript{a}, María Saura\textsuperscript{b}, David O.F. Skibinski\textsuperscript{c, #}, & Emilio Rolán-Alvarez\textsuperscript{a, #}

\textsuperscript{a}Centro de Investigación Mariña (CIM-UVIGO), Universidade de Vigo, 36310 Vigo, España; \textsuperscript{b}Departamento de Mejora Genética Animal, INIA, 28040 Madrid, España; \textsuperscript{c}Institute of Life Science, Swansea University Medical School, Swansea SA2 8PP, UK

*Address Correspondence to Angel P. Diz, E-mail: angel.p.diz@uvigo.es, #Equal contribution

Handling editor: Nicolas Bierne

Received on 8 January 2021; accepted on 3 July 2021

Abstract

Proteomic analysis was carried out on the Crab (upper-shore) and Wave (lower-shore) ecotypes of *Littorina saxatilis* from a hybrid zone at Silleiro Cape, Spain. Proteome profiles of individual snails were obtained. Protein expression in F\textsubscript{1} hybrid snails bred in the laboratory and snails with intermediate shell phenotypes collected from the mid-shore were compared with Crab and Wave ecotypes using analytical approaches used to study dominance. Multivariate analysis over many protein spots showed that the F\textsubscript{1} snails are distinct from both ecotypes but closer to the Wave ecotype. The intermediate snails are highly variable, some closer to the Crab and others to the Wave ecotype. Considered on a protein by protein basis, some proteins are significantly closer in expression to the Crab and others to the Wave ecotype for both F\textsubscript{1} and intermediate snails. Furthermore, a significant majority of proteins were closer in expression to the Wave ecotype for the F\textsubscript{1}, consistent with the multivariate analysis. No such significant majority towards either the Crab or Wave ecotype was observed for the intermediate snails. The closer similarity of F\textsubscript{1} and Wave ecotype expression patterns could be the result of similar selective pressures in the similar mid-shore and low-shore environments. For a significantly larger number of proteins intermediate snails were closer in expression to the ecotype having the lower expression, for both Crab and Wave ecotypes. This is somewhat unexpected as lower expression might be expected to be an indication of impairment of function and lower fitness. Proteomic analysis could be important for the identification of candidate proteins useful for gaining improved understanding of adaptation and barriers to gene flow in hybrid zones.

**Keywords:** marine invertebrates, reproductive isolation, F\textsubscript{1} hybrids, speciation, gene expression, molecular phenotype
The study of hybridisation and hybrid zones is important for understanding the causes of speciation (Barton and Hewitt 1985; Harrison 1990, 1993; Gompert et al. 2017). Different hybrid zones between ecologically divergent ecotypes of the marine intertidal gastropod *Littorina saxatilis* have been well studied across three regions from NE Atlantic coasts (Sweden, Britain, Spain) (reviewed in Galindo and Grahame 2014; Rolán-Alvarez et al. 2015).

Here we study the *L. saxatilis* hybrid zone from Galicia (NW Spain), where a ridged and banded ecotype is thought to have evolved to resist crab predation in the upper-shore (Crab ecotype) due to its large and robust shell with a small aperture, and a smooth and unbanded ecotype has evolved to resist wave action in the lower-shore (Wave ecotype) due to its small and thin shell with a large aperture (Johannesson et al. 1993; Rolán-Alvarez et al. 1997; Butlin et al. 2014). They meet and hybridize at the mid-shore (reviewed in Rolán-Alvarez 2007) and have been described as an example of *in situ* divergence due to ecologically-based divergent selection existing at these different microhabitats (crab predation and wave action) in the face of gene flow (Rolán-Alvarez et al. 2004; Quesada et al. 2007; Galindo et al. 2009). They show partial reproductive barriers due to divergent natural selection and positive size-assortative mate choice (Johannesson et al. 2010; Rolán-Alvarez et al. 2015; Boulding et al. 2017) and can be crossed successfully in the laboratory to produce F<sub>1</sub> hybrids (Saura et al. 2011). In the field, at the mid-shore, it is possible to observe a small proportion of mating pairs between ecotypes and also low frequencies of intermediate snails with mixed shell characters (ridged-unbanded, smooth-banded; referred to as intermediates hereafter) (reviewed in Rolán-Alvarez 2007). These intermediates show viabilities, fertilities and sexual selection on phenotypic traits which are the average of those of the pure ecotypes (Rolán-Alvarez et al. 1997; Johannesson et al. 2000; Cruz et al. 2001). Use of genome-wide AFLP markers in this hybrid zone in Galicia showed that these phenotypic intermediates are not necessarily the offspring (F<sub>1</sub>) of a heterospecific cross between the pure ecotypes, but certain snails showed introgression between the parental ecotypes. This pattern was highly dependent on environmental characteristics of the locality studied (Galindo et al. 2013). Moreover, selection on shell morphology of these intermediates was variable even at the scale of micro-habitat within locality (Galindo et al. 2014). A more recent study using SNPs derived from ddRAD markers detected a very low proportion of introgressed individuals (1.8%) within the studied intermediates (Kess et al. 2018). The genetic constitution and ancestry of these intermediate snails between the Crab and Wave ecotypes is thus not fully understood in the Galician hybrid zone of *L. saxatilis*. In the current study we use proteomics to study expression in these intermediate snails compared with the two ecotypes as a complementary approach to previous genetic analyses (Galindo et al. 2013; Kess et al. 2018).

Proteomic analysis has been performed so far only on the parental ecotypes (Crab and Wave) (Martínez-Fernández et al. 2008, 2010; Diz et al. 2012a; Diz and Rolán-Alvarez 2014). Those studies provide evidence that up to 30% of proteins show significant differences in expression between ecotypes which persist from embryonic through to adult stages. The current study is novel in analysing the proteome of F<sub>1</sub> snails bred in the laboratory for comparison with the proteome of the intermediate snails collected on the shore and in determining the proteome of individual snails, the earlier studies having used protein samples pooled from more than one individual. We also test the hypothesis of sex difference in protein expression and its interaction with ecotype.

The study of hybrid trait values in relation to parental values in divergently selected traits is key to understanding speciation mechanisms (Thompson et al. 2020). An analytical approach to this is that used to quantify dominance and is applied here to the intermediate snails as well as the laboratory bred F<sub>1</sub> snails. In a review of nearly 200 studies of F<sub>1</sub> hybrids and their parent populations specific trends in relation to dominance in divergently-selected traits have been noted in diverse species (Thompson et al. 2020). Hybrids usually resemble one parent more closely than the other, the
direction of resemblance differs between traits, and dominance of higher and lower trait values is equally likely. Here we consider these three trends as hypotheses to test in the *L. saxatilis* Galician hybrid zone treating each of the many proteins in a proteomics dataset as a separate trait. We discuss the proteomic nature of the intermediate snails and the relevance of these results to our understanding of speciation and adaptation in *Littorina*.

**Material and Methods**

**Sampling and laboratory acclimation**

Samples were collected in July 2010 across the *Littorina saxatilis*’ hybrid zone at Silleiro Cape (42°06’16.7”N, 8°53’56.3”W; NW Spain). Snails were randomly collected from each of the typical parental microhabitats, the barnacle belt in the upper-shore (Crab ecotype; large, banded and ridged shell; referred also as RB ecotype in previous studies using populations from NW Spain) and from the mussel belt in the lower-shore (Wave ecotype; small, unbanded and smooth shell; referred also as SU ecotype in previous studies using populations from NW Spain). Phenotypically intermediate snails (called intermediates hereafter) were collected in the mid-shore, a patchy environment (barnacles and mussels). Their phenotypes include smooth shells with dark bands, ridged and unicolored shells (unbanded), and ridged with two or more incomplete bands (following Johannesson et al. 1993 and Cruz et al. 2001).

After the sampling, all the snails were brought alive to the ECIMAT marine laboratory (Toralla Marine Science Station, University of Vigo) in order to keep them in controlled laboratory conditions. The snails were randomly assigned to different aquaria where individual phenotypes remained clearly identifiable. Further details on the culture system are given in Saura et al. (2011). After one week, the individuals were flash frozen in liquid nitrogen and kept at -80°C until total protein extraction.

**Rearing F₁ hybrids in the laboratory**

We also studied three F₁ hybrids that were reared in the ECIMAT marine laboratory (Toralla Marine Science Station, University of Vigo) for 6 months. *L. saxatilis* is a dioic species with internal fertilisation, the female carries the developing embryos inside the shell (ovoviparity) until they are born as crawlaways (Reid 1996). These F₁ individuals are from three different broods resulting from several attempted laboratory crosses between a Crab ecotype female and a Wave ecotype male following Saura et al. (2011), both ecotypes also being collected at Silleiro Cape as were the intermediate individuals. The crosses were carried out using virgin juvenile females in order to avoid the possibility that the females had previously mated in the field, as *L. saxatilis* presents multiple paternity (Panova et al. 2010). The three F₁ hybrids were flash frozen in liquid nitrogen and kept at -80°C until total protein extraction.

It is possible that any protein expression differences between the individuals collected from the shore and the laboratory reared F₁ individuals could be attributed in part to environmental differences. The former were maintained under laboratory conditions for one week after collection, whereas the latter were reared for several months under laboratory conditions. However this is an unlikely explanation for the results we report because we have previously shown that patterns of protein expression in these *L. saxatilis* ecotypes are only marginally affected by the environment where the snails live (Martínez-Fernández et al. 2010).
Proteomic analysis

The proteomic analysis of individual snails was carried out by two-dimensional electrophoresis (2-DE) following the procedure described in García et al. (2013) with slight modifications. Prior to total protein extraction, the shells were removed from the snails and the sex was determined. Whole snails without shells were submerged individually into 1 mL of lysis buffer (7M urea, 2M thiourea, 4% CHAPS, 1% DTT and 1% IPG), the samples were sonicated (10 x 5 sec pulses with 10 sec breaks) using a Branson Digital Sonifier 250 (Branson) and centrifuged for 20 min at 21,000 g (4°C), then the supernatant containing the proteins was stored at -80°C.

A total of 24 intermediate, 6 Crab ecotype, 6 Wave ecotype and three laboratory reared F1 snails were analysed individually by 2-DE. Equal numbers of males and females were analysed within each category of the snails collected from Silleiro Cape. The same 452 protein spots were quantified in all of the individual snails. Representative images showing the distinct differences between the Crab, Wave and intermediate snail shells are shown in Figure 1. Protein concentration was measured using a modification of the Bradford method (Ramagli and Rodríguez 1985). Then, approximately 100 μg of total protein were separated through isoelectric focusing (first-dimension separation) on IPG (immobilized pH gradient) strips (ReadyStrip™, pH 5-8/17 cm; Bio-Rad) using a Protean IEF System (Bio-Rad) following manufacturer’s instructions. The second-dimension was carried out by electrophoresis (constant current of 12 W/gel for 5 h) on 12.5% polyacrylamide gels using an Ettan DALTsix electrophoresis system (GE Healthcare) connected to a refrigerated circulator system (20°C). Protein spots were visualized using a modified version of the silver staining method, then the gels were scanned using a GS-800™ Calibrated Imaging Densitometer (Bio-Rad). The analysis of the gel images was performed with the software Progenesis SameSpots v.3.3 (Nonlinear Dynamics Ltd.) which was used for semi-automatic alignment of the gels, protein spot detection and spot volume measurements. A final spot-filtering step was carried out in order to exclude technical artefacts or spots that were not well-defined. Absolute spot volumes for each gel were corrected by background subtraction and normalized using the default method that takes into account the differences in the total spot volume across 2-DE gels in Progenesis SameSpots v.3.3 for further statistical analyses. Figure S1 gives an example of a 2-DE gel on which is marked the position of 38 spots which were a focus of study as described in the Results section. Normalized spot volumes were transformed to a logarithmic scale to fit normality and homoscedasticity assumptions of parametric tests (see Diz et al. 2017 and references therein). Hierarchical clustering (using the UPGMA method) and heat map analyses were carried out using correlation values for distance calculations and the average linkage criterion for clustering on log-normalized protein spot volume data in ClustVis (Metsalu and Vilo 2015). Principal component analysis (PCA) was also carried out on the logarithm of normalised spot volume using ClustVis with unit variance scaling and singular value decomposition (SVD) with imputation to calculate the principal components.

Dominance analyses

In the present study, the logarithm of normalized protein spot volume was used to study expression differences between the ecotypes and F1 and intermediate snails. Hereafter the abbreviation Hy (for hybrid) is used as a convenience to stand for both the F1 individuals which are true hybrids and the intermediates recognising that there is uncertainty about the genetic status of the latter. Similarly we use the word dominance to describe the position of Hy for the intermediates as a convenience to avoid introducing new terminology. Thus the word dominance is used as a generic term when applied to both F1 and intermediates. In addition, the words dominance, overdominance and underdominance are used to
Diz et al.: Proteomic variation in a Littorina hybrid zone

describe the phenotypic value of $\text{Hy}$ relative to that of the two parental ecotypes and the mid-parent value (the average of the Crab and Wave ecotypes) (Figure 2). In the context of our descriptive use of the word (following Thompson et al. 2020), dominance occurs when the value of $\text{Hy}$ is different from the mid-parent and closer to that of one ecotype than the other, but has a value between that of the two parental ecotypes. Overdominance occurs when $\text{Hy}$ has a spot volume greater than both ecotypes, and underdominance occurs when $\text{Hy}$ has spot volume less than both ecotypes. These descriptions of phenotypic dominance make no assumptions about the genetic basis of protein expression in terms of the number of underlying loci and the dominance relationships of the alleles at these loci. For a given spot, eight possible general patterns illustrating dominance, overdominance and underdominance are possible if we take into account that the Wave value could be greater than Crab (patterns 1-4) or the other way round (patterns 5-8) (Figure 2).

If spot volumes of Crab, Wave and $\text{Hy}$ are regarded as phenotypes for the purpose of estimating dominance by analogy with a quantitative genetics model (Falconer and Mackay 1996; Caballero 2020), then values of $+a$ or $-a$ can be assigned to the two parental ecotypes and the value $d$ to $\text{Hy}$. The values of $a$ and $d$ for the eight patterns of dominance are shown in Figure 2. If the value of $\text{Hy}$ coincides with the mid-parent, then $d=0$, representing the lack of dominance.

Transformations of the scale of measurement of the trait, such as the conversion to log values of the protein spot volume in this study will affect the value of $d$. Then if $\text{Hy}$ lies between both parental ecotypes this could change the sign of $d$. However this will rarely happen if $\text{Hy}$ is situated far from the mid-parent and will not happen at all with overdominance or underdominance which is prevalent for significant spots here, as previously mentioned. Logarithmic transformation will also compress higher values of spot volume and this could lead to an overestimation of the number of spots showing dominance of higher expression. However, dominance of lower expression predominates in our results (see Results section), which thus could not be explained by the transformation.

Statistical analysis of dominance

Two methods were used to assess dominance statistically. In the first method, referred to here as the mid-parent method, one-way analysis of variance (ANOVA) was used to compare log-normalized spot volume between $\text{Hy}$ ($F_1$ or intermediates) and the mean of Crab and Wave ecotypes on a spot by spot basis (mid-parent value; see Figure 2). The ANOVA was carried out with $F_1$ hybrids and intermediates separately. A significant result provides evidence that $d$ is not equal to zero, and therefore evidence of dominance as defined in the previous section, that is the $F_1$ or intermediates have spot volume which differs from the mid-parent. In each analysis a $P$-value was obtained for every protein spot analysed. Then two different multiple testing adjustment methods were used, the SGoF method (Sequential Goodness of Fit) using the SGoF software (Carvajal-Rodriguez et al. 2009) and the Stouffer method (Z scores; referred to as SStouffer hereafter) (see Whitlock 2005) applied sequentially and implemented in Excel. Using more than one multiple testing methods provides some flexibility in assessing the number of significant results depending on research priorities (Diz et al. 2011). A significance level of $P < 0.05$ was used with both multiple testing methods.

In the second method for assessing dominance statistically, referred to here as the “SNK method” because it incorporates the Student-Newman-Keuls (SNK) test, a two-way ANOVA, again applied on a spot by spot basis, was first used to compare log-normalized spot volume between the parental ecotypes (Crab vs Wave; fixed factor Ecotype) and sexes (Female vs Male; fixed factor Sex). $\text{Hy}$ individuals were excluded from this first step. This ANOVA was used to determine candidate proteins accounting for ecotype differentiation and which would also thus be good candidates in
the search for dominance. Then, multiple test adjustments (SGoF and SStouffer) were carried out and the significant spots (P < 0.05) for the factor Ecotype, were considered as candidate proteins to assess dominance. In the SNK method these candidate spots were used as the starting point to assess dominance. One-way ANOVAs for each candidate spot, including both parental ecotypes and Hy (F1 or intermediates in separate analyses) were carried out. Then, the significance of the three comparisons between Crab, Wave and Hy (F1 or intermediates) was determined by the post-hoc studentized range distribution Student-Newman-Keuls (SNK) test. Two different results were considered as being compatible with a dominance effect. First, when the spot volume of Hy (F1 or intermediates) is not significantly different to one of the two parental ecotypes, but both of these are significantly different from the other ecotype. For example, if Hy (F1 or intermediates) and Wave are not significantly different from each other, but both are significantly greater than Crab, then Wave dominance is concluded (in line with patterns 1 and 2 in Figure 2A and patterns 7 and 8 in Figure 2B). The second criterion is when the spot volume of Hy is either significantly greater than both ecotypes (patterns 1 and 5; overdominance), or significantly less than both ecotypes (patterns 4 and 8; underdominance). For protein spots that do not meet any of the above criteria, it is concluded that there is no evidence for dominance. In presenting the results of both the mid-parent and SNK dominance assessment methods the word dominance is used as a generic term to describe statistically significant dominance, overdominance and underdominance as defined above.

The frequencies of protein spots showing significant Crab or Wave dominance after multiple testing were compared against a 1:1 expectation using a G-test (G-test calculator spreadsheets; see McDonald 2014). A test against a 1:1 expectation was also performed to compare the frequency of spots showing dominance for higher expression (higher spot volume; patterns 1-2, 5-6, Figure 2) and those showing dominance for lower expression (lower spot volume; patterns 3-4, 7-8, Figure 2).

Statistical analyses for the mid-parent and SNK methods were accomplished in SPSS ver. 24 (IBM Corp.) software.

Results

In the mid-parent method, the number of protein spots showing significant deviation from the mid-parent after multiple testing adjustment (SGoF and SStouffer) are shown in Table 1. SStouffer generates a greater number of significant spots than SGoF. For both datasets (F1 and intermediates) there is no evidence of a difference in frequency of Hy (ie F1 or intermediate) being closer to Crab or Wave than the mid-parent (non-significant P-values, Table 1). However, for the intermediate snails the ecotype with lower expression is closer to the mid-parent for a significant majority of spots (P < 0.001). An overall picture of the possible patterns of phenotype expression comparing Crab, Wave and Hy for the mid-parent method is given in Figure 3A for the analysis of F1 hybrids and in Figure 3B for intermediates. In each case, the genotypic values of a and d are plotted for each of the 452 protein spots. The spots which are significant for dominance, that is for d in the test of Hy against mid-parent, are highlighted in the plot and their frequencies correspond to those shown in Table 1. The majority of spots showing significant dominance are those which fall above (top quadrants) or below (bottom quadrants) the curves for a or -a, and thus show overdominance and underdominance respectively. For the intermediates (Figure 3B) it is clear that the number of significant spots showing underdominance is greater than those showing overdominance, and this is reflected in the significant result of the G-test between high and low expression (Table 1).
In the SNK method for the factor Ecotype, 38 spots (8.4%) are significant after SGoF and 75 (16.6%) are significant after SStouffer (Table 1). For the factor Sex, 9 spots (2.0%) are significant after SGoF and 28 (6.2%) are significant after SStouffer. The corresponding values for the interaction (Ecotype x Sex) are 11 (2.4%) and 43 (9.5%) spots after SGoF and SStouffer, respectively. The significance of the factor Ecotype in the ANOVA analysis derives from the comparison Crab vs Wave, thus separation of these ecotypes in clustering or PCA component plots is to be expected. The F₁ and intermediates are excluded from the ANOVA analysis, thus their positioning in cluster or PCA analysis in relation to the ecotypes is not predictable a priori.

The results of the dominance analysis using the SNK method are displayed in Table 1. The number of spots showing significant dominance is lower than for the mid-parent method. This might be due in part to the initial screening for candidate loci using two-way ANOVA in the SNK method. The spots that are significant with SNK tend to have high absolute values of the genotypic value $a$, the wide separation of the two ecotypes favouring a statistically significant difference between them. In the dataset using F₁ hybrids, the results of the G-test for Crab:Wave against the 1:1 expectation gives significant p-values indicating a preponderance of Wave dominance over Crab for both SGoF and SStouffer (Table 1). This significant result was not observed with the mid-parent method. In line with the results for intermediates with the mid-parent method, the G-tests provide evidence for a greater number of spots showing dominance of lower expression, with one quite low p-value for SGoF ($P = 0.085$) and a significant p-value for SStouffer.

In summary, Table 1 presents evidence for significant dominance with different protein spots showing dominance of Crab or Wave and lower or higher expression. Both the mid-parent and SNK methods suggest a significant preponderance of spots showing dominance of lower expression in intermediates, and the SNK method in addition provides evidence of a preponderance of spots showing dominance of Wave over Crab for the F₁ hybrids.

The heat map constructed from the 38 spots declared significant after SGoF shows that the Crab and Wave ecotypes are clearly differentiated into separate clusters as expected, and these in turn are differentiated from the cluster of the three F₁ individuals (Figure 4). The intermediates fall into three distinct clusters. The largest cluster links closely to the three F₁ individuals. Of the other two clusters one links to Crab the other to Wave. A similar heat map was obtained for the 75 candidate spots significant after SStouffer (Figure S2). The corresponding PCA plots of scores of component 2 against component 1 are shown in Figure 5, separately for F₁ hybrids and intermediates. Crab and Wave are separated mainly on component 1. The F₁ hybrids appear to lie between Crab and Wave on component 1 and are shifted down somewhat on component 2. They are clearly distinct from Crab but much closer to Wave. Although the Wave and F₁ ellipses overlap, the three F₁ individuals nevertheless lie outside the cloud of Wave individuals. The cloud of points for the intermediates lies between Crab and Wave on component 1. Although the ellipses for Wave and the intermediates overlap the two clouds of points are separated.

Discussion

This study focused on comparing the proteome profiles of F₁ hybrid Littorina saxatilis bred in the laboratory with their Crab (upper-shore) and Wave (lower-shore) ecotype parents. These profiles are also compared with those of phenotypically intermediate individuals, whose ancestral status is uncertain, collected from the mid-shore where the
Crab and Wave ecotypes overlap and hybridise. The proteome is analysed over all proteins using multivariate techniques, and the proteins are also considered individually. Analytical techniques applied to the analysis of dominance are used to quantify the similarity of the F$_1$ and intermediates to the Crab and Wave ecotypes.

In the multivariate analysis the distinct clustering apart of the true F$_1$ hybrids bred in the laboratory from the parental Crab and Wave ecotypes (Figure 4, Figure S2) is generally consistent with their hybrid genetic status, lying between the ecotypes for component 1 in the PCA analysis (Figure 5, Figure S3), which is consistent with what is expected of F$_1$ hybrids. For example, additivity of expression in proteomic data in F$_1$ hybrids has been observed previously in polychaetes (Blank et al. 2012). By contrast in a study of proteomic patterns of naturally occurring hybrids between the mussels *Mytilus edulis* and *M. galloprovincialis* (Diz and Skibinski 2007) the two parental species were more closely similar to each other than either were to the hybrids. Thus the tendency for hybrids to be intermediate proteomically does not always occur.

The greater similarity of the F$_1$ hybrids to Wave on the PCA plots (Figure 5, Figure S3) supports our hypothesis, from Thompson et al. (2020), that hybrids resemble one parent more closely than the other. Dominance in F$_1$ hybrids towards the Wave ecotype might be favoured by natural selection, as environmental conditions (e.g. temperature, humidity) at the mid-shore are more similar to the lower-shore (Wave ecotype) than to the upper-shore (Crab ecotype), and therefore affecting the physiological response of the organism through different protein pathways. Moreover, at the mid-shore, wave action still represents an important selective pressure and crab predation is low, and this is also reflected by a cline in shell size for the Crab ecotype, being smaller and more similar in size to the Wave ecotype at the mid-shore (Johannesson et al. 1993; Boulding et al. 2017).

The distinct clustering of the naturally occurring intermediate snails (Figure 4, Figure S2) and high variation on component 1 located between the ecotypes (Figure 5, Figure S3), though without such marked closer similarity to Wave, possibly suggests a high heterogeneity in protein expression, which is also in line with the heterogeneity previously detected for AFLP markers and at the shell phenotypic level (Galindo et al. 2013, 2014), with some intermediate individuals more similar to Crab, others more similar to Wave. Although, we cannot make inferences about the genetic variation of the intermediate snails studied here, the proteomic profiles should also give important information as a proxy for their genetic and gene expression variation.

Another possible explanation for why the expression of the intermediate snails lie between those of the Crab and Wave ecotype on component 1 in the PCA plots is that these snails are adapted to the microenvironment of the mid-shore, which is different from the habitats where the ecotypes were sampled (upper and lower-shore). In this circumstance we can suppose that their phenotypes, have been selected in this environment, but this could be the case also for Crab and Wave ecotypes from the mid-shore that have not been included in this study. The phenotypes of these intermediates are not intermediate between Crab and Wave on any quantitative scale. Similarly, the main selective factors, crab predation in the upper-shore and wave action in the lower-shore, are not opposites on a quantitative scale. Why then should the multivariate proteomic phenotype of the intermediate snails lie clearly between Crab and Wave on the first component in the PCA plots (Figure 5, Figure S3)? Further studies should include individuals from the whole environmental cline in order to shed light into these questions.

In our results it is important to highlight that there is one correspondence between the multivariate analysis and the spot by spot analysis of the proteomic data. The closer similarity of the F$_1$ hybrid snails to Wave than Crab in the PCA plots (Figure 5, Figure S3) is consistent with the significantly higher number of spots showing Wave than Crab
Diz et al.: Proteomic variation in a *Littorina* hybrid zone

dominance for the SNK method (Table 1). These results may reflect a particular general property of the Wave ecotype gene pool, but it should be noted that this generalisation cannot apply to all spots studied as by contrast other spots show dominance of Crab. What would be relevant theoretically is whether the majority effect is the leading one or whether the results should be considered on a spot by spot basis. Still, because hybridisation occurs outside of the parental micro-habitats, it is possible that natural selection is maintaining certain levels of genetic variation (Crab and Wave) within the mid-shore. On the other hand, the corresponding results for the intermediate snails are consistent with the heterogeneity previously described, showing spots where intermediates are closer to Crab and other spots closer to Wave (Figure 5, Figure S3, Table 1), suggesting the heterogeneity in selection found in previous studies (Galindo et al. 2013, 2014).

Two other observations made in this study are worthy of attention. The first is that some spots showed significance for the factors Sex and Sex x Ecotype. This is expected given the difference in the nature of the reproductive tissues between sexes, and despite the greater representation of muscular foot tissue in homogenates. The significant interaction implies that the difference in protein expression between sexes itself differs between the two ecotypes. This may be related to the observation that sexual size dimorphism differs between the ecotypes, the Wave ecotype showed the greatest size differences at maturity, females being larger than males (see fig. S4 in Boulding et al. 2017; Perini et al. 2020). The second observation is the difference in results obtained with the two distinctly different methods to detect expression differences between ecotypes and the F$_1$ and intermediate snails (mid-parent and SNK methods). These methods both identify significant spots that are not significant using the other method (Table 1, Figure 3), demonstrating the utility of this approach. Furthermore two different multiple testing methods were used (SGoF and SSStouffer) with similar results but also some differences (Table 1). The additional information that can be got in this approach has been noted previously (Diz et al. 2011) and might be particularly useful in identifying candidate proteins for further study. We set out in our Introduction three other hypotheses that we can test and are derived from Thompson et al. (2020). The first is that the expression of F$_1$ hybrids and intermediate snails are significantly closer in phenotype to one of the ecotypes is supported by the current results when considered on a spot by spot basis. Many spots are significantly closer to either Crab or Wave, refuting the hypothesis that they are intermediate (Figure 2, Table 1, Figure 3). For some spots there is closer resemblance to Crab, for others closer resemblance to Wave. This is in accord with the second hypothesis derived from Thompson et al. (2020) that the direction of closer resemblance differs between traits. This is observed for both the laboratory reared F$_1$ hybrids (Figure 3A) and the intermediate snails (Figure 3B).

The third hypothesis derived from Thompson et al. (2020) that the expression of F$_1$ and intermediate snails is just as likely to be closer to the ecotype with higher expression as to the ecotype with lower expression. This is supported for the F$_1$ hybrids where spots with dominance for both higher and lower expression are observed without any significant bias (Table 1, Figure 3A). For the intermediate snails there is however a bias with a significant excess of spots showing closer expression to the ecotype with lower expression for both Crab and Wave (Table 1, Figure 3B). Traditional theories of the evolution of dominance might associate lower expression and less gene product with some impairment of function and reduction in fitness. Patterns of lower expression in hybrids (reviewed in Landry et al. 2007; Ortiz-Barrientos et al. 2007), by means of gene expression (misexpression), have been previously described, for example between divergent ecotypes of whitefish *C. clupeaformis* (Renaut et al. 2009) and also in a recent adaptive radiation of *Cyprinodon* pupfishes (McGirr and Martin 2020). The authors associated these phenotypes to transgressive segregation effects and linked this lower expression and misexpression to potential postzygotic barriers to gene flow. Assuming that the intermediate snails are involved in barriers to gene flow in this *L. saxatilis* hybrid zone these might be in addition to those barriers already described in previous work (e.g. divergent selection, assortative mating; reviewed in
Rolán-Alvarez et al. 2015). Lower expression in intermediate snails and its relation to barriers to gene flow needs to be studied further by measuring the fitness of $F_1$ hybrids and intermediate snails with specific protein expression patterns, for example by using transplant experiments in the field (Cruz et al. 2004) or measuring hybrid female fecundity (e.g. clutch size, embryo abortion) (Cruz et al. 1998; Johannesson et al. 2000, 2020). The results for both $F_1$ hybrids and intermediate snails show a large proportion of protein spots lying outside the range of the parental ecotypes, (Figure 3), possibly representing transgressive and thus potential maladaptive phenotypes. The genetic basis of these transgressive phenotypes in hybrids and the proteins involved, also represent interesting future areas of investigation.

In comparison to genome and transcriptome analysis, less attention has been generally provided to the proteome in evolutionary ecology studies, despite proteins representing the molecular phenotype, a more direct target for natural selection (Diz et al. 2012b; Baer and Millar 2016). Further studies of the proteome of ecotypes, hybrids and intermediate snails may thus be important in gaining knowledge of isolating barriers and speciation in *Littorina* despite the complexity of investigating the molecular genetics of a trait such as spot volume which might be a function of the joint effect of variation at many loci differing between the ecotypes, particularly modifier loci affecting the level of gene expression.

Acknowledgements

We thank Mary Riádigos for administrative contribution. This work was supported by *Xunta de Galicia* (ED431C 2020/05), FONDOS FEDER (“*unha maneira de facer Europa*”) and *Ministerio de Economía y Competitividad* (CGL2016-75482-P). Marine Research Center (CIM-UVIGO) is funded by the Galician Regional Government through the "Excellence in Research (INUGA)" Program and ERDF Operational European Union Program Galicia 2014-2020. J Galindo was funded by a JIN project (*Ministerio de Ciencia, Innovación y Universidades*, code RTI2018-101274-J-100).

References

Baer B, Millar AH, 2016. Proteomics in evolutionary ecology. J Proteomics 135:4-11.

Barton NH, Hewitt GM, 1985. Analysis of hybrid zones. Ann Rev Ecol Syst. 16:113-148.

Blank M, Mikkat S, Verleih M, Bastrop R, 2012. Proteomic comparison of two invasive Polychaeta species and their naturally occurring F1-hybrids. J Proteome Res, 11: 897-905.

Boulding EG, Rivas MJ, González-Lavín N, Rolán-Alvarez E, Galindo J, 2017. Size selection by a gape-limited predator of a marine snail: Insights into magic traits for speciation. Ecol Evol. 7:674-688.

Butlin RK, Saura M, Charrier G, Jackson B, André C et al., 2014. Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. Evolution 68:935-949.

Caballero A, 2020. *Quantitative Genetics*. Cambridge: Cambridge University Press.
Carvajal-Rodríguez A, de Uña-Alvarez J, Rolán-Alvarez E, 2009. A new multitest correction (SGoF) that increases its statistical power when increasing the number of tests. BMC Bioinformatics 10:209.

Cruz R, Rolán-Alvarez E, Garcia C, 1998. Natural selection on a vertical environmental gradient in Littorina saxatilis: analysis of fecundity. Hydrobiologia 378:89-94.

Cruz R, Rolán-Alvarez E, Garcia C, 2001. Sexual selection on phenotypic traits in a hybrid zone of Littorina saxatilis (Olivi). J Evol Biol. 14:773-785.

Cruz R, Vilas C, Mosquera J, Garcia C, 2004. Relative contribution of dispersal and natural selection to the maintenance of a hybrid zone in Littorina. Evolution 58:2734-2746.

Diz AP, Truebano M, Skibinski DOF, 2009. The consequences of sample pooling in proteomics: an empirical study. Electrophoresis 30:2967-2975.

Diz AP, Carvajal-Rodríguez A, Skibinski DOF, 2011. Multiple hypothesis testing in proteomics: A strategy for experimental work. Mol Cell Proteomics 10(3). doi:10.1074/mcp.M110.004374

Diz AP, Páez de la Cadena M, Rolán-Alvarez E, 2012a. 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.

Diz AP, Martínez-Fernández M, Rolán-Alvarez E, 2012b. Proteomics in evolutionary ecology: Linking the genotype with the phenotype. Mol Ecol 21:1060-1080.

Diz AP, Rolán-Alvarez E, 2014. High proteome variation between ecotypes of Littorina saxatilis cannot be explained by tissue heterogeneity or a common-garden×ecotype effect. J Exp Mar Bio Ecol. 451:69-73.

Diz AP, Álvarez-Rodríguez M, Romero MR, Rolán-Alvarez E, Galindo J, 2017. Limited proteomic response in the marine snail Melarhaphe neritoides after long-term emersion, Curr Zool. 63:487-493.

Falconer DS, Mackay TFC, 1996. Introduction to Quantitative Genetics. 4th edn. Harlow, Essex: Longmans Green.

Galindo J, Morán P, Rolán-Alvarez E, 2009. Comparing geographical genetic differentiation between candidate and noncandidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail Littorina saxatilis. Mol Ecol. 18:919-930.

Galindo J, Martínez-Fernández M, Rodríguez-Ramilo ST, Rolán-Alvarez E, 2013. The role of local ecology during hybridization at the initial stages of ecological speciation in a marine snail. J Evol Biol. 26:1472-1487.

Galindo J, Grahame JW, 2014. Ecological speciation and the intertidal snail Littorina saxatilis. Adv Ecol. (Article ID 239251).

Galindo J, Rivas MJ, Saura M, Rolán-Alvarez E, 2014. Selection on hybrids of ecologically divergent ecotypes of a marine snail: The relative importance of exogenous and endogenous barriers. Biol J Linn Soc. 111:391-400.

García SD, Diz AP, Sá-Pinto A, Rolán-Alvarez E, 2013. Proteomic and morphological divergence in micro-allopatric morphotypes of Melarhaphe neritoides in the absence of genetic differentiation. Mar Ecol Prog Ser. 475:145-153.
Diz et al.: Proteomic variation in a *Littorina* hybrid zone

Gompert Z, Mandeville EG, Buerkle A, 2017. Analysis of population genomic data from hybrid zones. Annu Rev Ecol Evol Syst. 48:207-229.

Harrison RG, 1990. Hybrid zones: Windows on evolutionary process. Oxford Surveys in Evol Biol 7:69-128.

Harrison RG, 1993. *Hybrid Zones and the Evolutionary Process*. New York: Oxford Univ. Press.

Johannesson K, Johannesson B, Rolán-Alvarez E, 1993. Morphological differentiation and genetic cohesiveness over a microenvironmental gradient in the marine snail *Littorina saxatilis*. Evolution 47:1770-1787.

Johannesson K, Larsson A, Cruz R, Garcia C, Rolán-Alvarez E, 2000. Hybrid fitness seems not to be an explanation for the partial reproductive isolation between ecotypes of Galician *Littorina saxatilis*. J Molluscan Stud. 66:149-156.

Johannesson K, Panova M, Kemppainen P, André C, Rolán-Alvarez E et al., 2010. Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. Philos Trans R Soc Lond B Biol Sci. 365:1735-1747.

Johannesson K, Zagrodzka Z, Faria R, Westram AM, Butlin RK, 2020. Is embryo abortion a post-zygotic barrier to gene flow between *Littorina* ecotypes?. J Evol Biol. 33:342-351.

Kess T, Galindo J, Boulding EG, 2018. Genomic divergence between Spanish *Littorina saxatilis* ecotypes unravels limited admixture and extensive parallelism associated with population history. Ecol Evol. 8:8311-8327.

Landry C, Hartl D, Ranz J, 2007. Genome clashes in hybrids: insights from gene expression. Heredity 99:483-493.

McDonald JH, 2014. *Handbook of Biological Statistics*. 3rd edn. Baltimore: Sparky House Publishing.

McGirr JA, Martin CH, 2020. Ecological divergence in sympatry causes gene misexpression in hybrids. Mol Ecol 29:2707-2721.

Martínez-Fernández M, Rodríguez-Piñero AM, Oliveira E, Páez de la Cadena M, Rolán-Alvarez E, 2008. Proteomic comparison between two marine snail ecotypes reveals details about the biochemistry of adaptation. J Proteome Res. 7:4926-4934.

Martínez-Fernández M, Páez de la Cadena M, Rolán-Alvarez E, 2010. The role of phenotypic plasticity on the proteome differences between two sympatric marine snail ecotypes adapted to distinct micro-habitats. BMC Evol Biol. 10:65.

Metsalu T, Vilo J, 2015. ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. Nucleic Acids Res. 43(W1):W566-W570.

Ortiz-Barrientos D, Counterman BA, Noor MAF, 2007. Gene expression divergence and the origin of hybrid dysfunctions. Genetica 129:71-81.

Panova M, Boström J, Hofving T, Areskoug T, Eriksson A et al., 2010. Extreme female promiscuity in a non-social invertebrate species. Plos One 5:e9640.
Diž et al.: Proteomic variation in a *Littorina* hybrid zone

Perini S, Rafajlović M, Westram AM, Johannesson K, Butlin RK, 2020. Assortative mating, sexual selection, and their consequences for gene flow in *Littorina*. Evolution 74:1482-1497.

Quesada H, Posada D, Caballero A, Morán P, Rolán-Alvarez E, 2007. Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. Evolution 61:1600-1612.

Ramagli LS, Rodriguez LV, 1985. Quantitation of microgram amounts of protein in two-dimensional polyacrylamide gel electrophoresis sample buffer. Electrophoresis 6:559-563.

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

Renaut S, Nolte AW, Bernatchez L, 2009. Gene expression divergence and hybrid misexpression between lake whitefish species pairs (*Coregonus* spp. Salmonidae). Mol Biol Evol 26:925–936.

Rolán-Alvarez E, Johannesson K, Erlandsson J, 1997. The maintenance of a cline in the marine snail *Littorina saxatilis*: the role of home site advantage and hybrid fitness. Evolution 51:1838-1847.

Rolán-Alvarez E, Carballo M, Galindo J, Morán P, Fernández B, et al., 2004. Nonallopatric and parallel origin of local reproductive barriers between two snail ecotypes. Mol Ecol. 13:3415-3424.

Rolán-Alvarez E, 2007. Sympatric speciation as a by-product of ecological adaptation in the Galician *Littorina saxatilis* hybrid zone. J Mollusc Stud. 73:1-10.

Rolán-Alvarez E, Austin CJ, Boulding EG, 2015. The contribution of the genus *Littorina* to the field of evolutionary ecology. Oceanogr Mar Biol Annu Rev. 53:157-214.

Saura M, Martinez-Fernández M, Rivas MJ, Caballero A, Rolán-Alvarez E, 2011. Lack of early laboratory postzygotic reproductive isolation between two ecotypes of *Littorina saxatilis* (Mollusca, Gastropoda) showing strong premating sexual isolation. Hydrobiologia 675:13.

Thompson KA, Urquhart-Cronish M, Whitney KD, Rieseberg LH, Schulter D, 2020. Patterns, predictors, and consequences of dominance in hybrids. Am Nat. doi:10.1086/712603

Whitlock MC, 2005. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. J Evol Biol. 18:1368-1373.
Diz et al.: Proteomic variation in a *Littorina* hybrid zone

Table 1. Summary of results of proteomic analysis with the number of protein spots showing dominance of Crab or Wave ecotype, and High or Low expression for F$_1$ and Intermediates for the mid-parent and SNK methods of dominance assessment, with SGoF and SSTouffer multiple testing adjustment methods. Note that the word dominance is used generically to describe the similarity to Crab or Wave ecotype for both F$_1$ and intermediate snails (see M&M). The final two columns show $p$-values for $G$-test against 1:1 expectations (see text).

| Analyses     | Multiple test correction | N   | No Dominance | Crab | Wave | Crab | Wave | Crab:Wave | Higher:Lower |
|--------------|---------------------------|-----|---------------|------|------|------|------|-----------|--------------|
|              |                           |     |               |      |      |      |      |           |              |
|              | SGoF                      | F$_1$ | 452       | 413  | 8    | 12   | 7    | 12        | 0.148        | 0.873        |
|              |                           | Intermediates | 452 | 348  | 12   | 13   | 42   | 37        | 0.695        | 0.000        |
|              | SSTouffer                 | F$_1$ | 452       | 352  | 20   | 35   | 26   | 19        | 0.423        | 0.317        |
|              |                           | Intermediates | 452 | 285  | 21   | 26   | 62   | 58        | 0.938        | 0.000        |
|              | SGoF                      | F$_1$ | 38        | 19   | 3    | 10   | 2    | 4         | 0.035        | 0.104        |
|              |                           | Intermediates | 38  | 21   | 1    | 4    | 6    | 6         | 0.466        | 0.085        |
|              | SNK                       | F$_1$ | 75        | 50   | 3    | 11   | 3    | 8         | 0.008        | 0.548        |
|              |                           | Intermediates | 75  | 44   | 5    | 5    | 9    | 12        | 0.590        | 0.046        |
Figure 1. Shell images of representative individuals of *Littorina saxatilis* for the different categories used in the proteomic (2-DE) analysis. Crab ecotype (large, ridged and banded), laboratory reared F₁ hybrid, intermediate snails collected in the field and Wave ecotype (small, smooth, unbanded).
**Figure 2.** Scheme of the patterns (1-8) of dominance, overdominance and underdominance that could be obtained from the proteomic analysis performed in this study. The mean of log-normalized protein spot volume of the Hy individuals (F1 or intermediates) is shown in relation to that of the parental ecotypes (Crab and Wave) and the mid-parent (the average of Crab and Wave). The scale (4.5 – 7.5) is an approximate representation of the values of log-normalized spot volume observed in the data. When Hy has a value between that of the two parental ecotypes there is dominance, if Hy exceeds both Crab and Wave there is overdominance and when Hy is less than both ecotypes there is underdominance. The values of a (additive genetic variance) and d (dominance variation) are indicated for each pattern. If \( a = -1 \), the spot volume of Wave is greater than the Crab (A, patterns 1-4) and if \( a = +1 \), the Crab is greater than the Wave (B, patterns 5-8). Patterns of dominance for higher expression (spot volume) are shown in dark orange (e.g. 6.5, 7.5) and for lower expression in light orange (e.g. 4.5, 5.5).
**Figure 3.** Values of $a$ and $d$ plotted for $F_1$ hybrids (A) and intermediates (B) for 452 protein spots. The spots are aligned on the x-axis in order of increasing $a$ from left to right, according to rank value of $a$. The right hand y axis gives the value $d$. The four quadrants on the graph correspond to the patterns of Figure 2. Thus top left corresponds to Wave and high expression being dominant (patterns 1 and 2), bottom left to Crab and low expression dominant (patterns 3 and 4), top right to Crab and high expression dominant (patterns 5 and 6), and bottom right to Wave and low expression dominant (patterns 7 and 8). Descriptive statements using $<$ and $>$ symbols are given in the four quadrants corresponding to the equations at the foot of Figure 2. The value of both $a$ (thick curved line) and $-a$ (thin curved line) are also plotted with the value of $a$ specified by the left hand y axis. Points lying between these two lines show dominance; points lying above or below both lines show overdominance or underdominance respectively. Points significant according to the Stouffer method are indicated with specific symbols: red circles; significant with mid-parent and SNK; yellow triangles; significant with mid-parent but not SNK; blue squares; significant with SNK but not mid-parent; small green circles; not significant with either mid-parent or SNK.

(A) $F_1$ hybrids

(B) Intermediates
Figure 4. Hierarchical clustering and heat map of samples analysed based on log normalized protein spot volumes of 38 candidate spots showing statistically significant differences between Crab (red) and Wave (purple) ecotypes after applying SGoF multiple testing correction. The original sample codes for the individual snails analysed are shown at the bottom of the heat map, and the protein spot codes are shown at the right. The individual codes also indicate the sex (F or M). Clustering patterns of F1 (blue) and intermediate snails (green) are also shown.
**Figure 5.** Principal component analysis of log normalised protein spot volumes of 38 candidate spots showing statistically significant differences between Crab and Wave ecotypes after applying SGoF multiple testing correction. Scores on principal component 2 (PC2) are plotted against scores on principal component 1 (PC1). The proportion of the total variance explained by a component is shown in brackets on the axes. The ellipses are such that a new observation from the same group would have a probability of 0.95 of falling within the ellipse. F$_1$ hybrids (A), intermediates (B).
Figure 1. Shell images of representative individuals of Littorina saxatilis for the different categories used in the proteomic (2-DE) analysis.

191x99mm (400 x 400 DPI)
Figure 2. Scheme of the patterns (1-8) of dominance, overdominance and underdominance that could be obtained from the proteomic analysis performed in this study.

203x147mm (300 x 300 DPI)
Figure 3. Values of $a$ and $d$ plotted for F1 hybrids (A) and intermediates (B) for 452 protein spots.

118x159mm (300 x 300 DPI)
Figure 4. Hierarchical clustering and heat map of samples analysed based on log normalized protein spot volumes of 38 candidate spots showing statistically significant differences between Crab (red) and Wave (purple) ecotypes after applying SGoF multiple testing correction.
Figure 5. Principal component analysis of log normalised protein spot volumes of 38 candidate spots showing statistically significant differences between Crab and Wave ecotypes after applying SGoF multiple testing correction.

239x92mm (300 x 300 DPI)