# Past and contemporary introgression between two strongly differentiated *Ciona* species as revealed by a post-genomic SNPs panel

**Sarah Bouchemousse[1,2]*, Cathy Haag-Liautard[3,4], Nicolas Bierne[3,4] and Frédérique Viard[1,2,]***

**(1) Sorbonne Universités, UPMC Univ Paris 6**, **CNRS,** UMR 7144, Adaptation & Diversity in Marine Environment, team Div&Co Station Biologique de Roscoff, Place Georges Teissier, **29680 Roscoff, France**

**(2) Université Montpellier 2**, Station Méditerranéenne de l'Environnement Littoral, 2 Rue des Chantiers, **34200 Sète, France**

**(3) CNRS-UM2-IRD**, UMR 5554, Institut des Sciences de l'Evolution, Place Eugène Bataillon, **34095 Montpellier, France**

*Correspondence : UMR7144, Equipe Diversité et Connectivité dans le paysage marin Côtier (Div&Co), CNRS-UPMC, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

Email: viard@sb-roscoff.fr; sbouchemousse@sb-roscoff.fr

1

# Abstract

One important outcome of biological introductions is to bring into contact species that diverged in allopatry. For interfertile taxa, the evolutionary outcomes of such secondary contacts may be diverse (e.g. adaptive introgression from or into the introduced species) but are not yet well examined in the wild. In this context, the recent secondary contact between the non-native species *Ciona robusta* and the native species *C.intestinalis*, in the English Channel, provides an excellent case study to examine. By means of a population genomic approach, using 310 SNPs developed from full transcriptomes, the genetic diversity at population and species level was examined by studying 449 individuals ($N_{C.robusta}$ = 213, $N_{C.\ intestinalis}$ = 236) sampled in 12 sites from the English Channel, North Sea, NW Atlantic and SE Pacific where they are found either alone or living in the same locality and habitat (syntopy). As expected from previous analyses, *C. robusta* showed less polymorphism than *C. intestinalis*, a pattern that may partly be explained by its non-native status in most of the study localities. The results clearly showed an almost complete absence of contemporary gene flow between the two species in syntopic localities, with only one first generation hybrid and none other genotype compatible with recent backcrosses. Interestingly, introgression was also observed in allopatric populations of both species (i.e. where no contemporary hybridization can occur). Furthermore, one allopatric population sampled in SE Pacific exhibited a much higher introgression rate compared to all others *C. robusta* populations. Altogether, these results indicate that the observed inter-specific gene flow is the outcome of historical introgression, spread afterward at a worldwide scale. They also point out that efficient barriers are preventing hybridization in the wild between the introduced and native species in the English Channel, thus making adaptive introgression of the introduced species unlikely to favor the sustainable establishment of the study non-native species.

**Keywords:** Biological invasions, Tunicates, Hybridization, Species range, Population genomics

# Introduction

Speciation is a gradual spatio-temporal process during which geographical or ecological isolation decrease gene flow between groups of individuals thereby contributing to the emergence of new species (Abbott *et al.* 2013). Species range shifts in response to long-term environmental changes can deeply modifies the evolution of these emerging novel species by promoting the formation of contact zones (Hewitt 2004; Swenson & Howard 2005; Maggs *et al.* 2008). In cases of species that are not fully reproductively isolated, interspecific gene flow may create hybrid zones (Barton 1979). Hybridization and introgression processes in contact zones between species are particularly interesting to examine for studying intrinsic and extrinsic barriers responsible of the maintenance of species boundaries (Hewitt 1988; Orr & Smith 1998; Turelli *et al.* 2001; Abbott *et al.* 2013; Harrison & Larson 2014).

In last few years, next generation sequencing techniques has revolutionized the study of hybridization and speciation processes (for a review, Seehausen *et al.* 2014). For instance, recent population genomic studies provided evidences that adaptive introgression can occur between divergent species and may be probably more common than previously expected (Abbott *et al.* 2013; Hedrick 2013). The evolutionary history of the modern human (for a review, Racimo *et al.* 2015), the malaria vector mosquito *Anopheles gambiae* (Fontaine *et al.* 2015), *Heliconius* butterflies (Pardo-Diaz *et al.* 2012) or *Mytilus* mussels (Fraisse *et al.* 2015) are particularly well-documented cases illustrating such processes.

Most of these studies are concerned with historical interspecific gene flow which occurred during long-term range expansion (see Currat *et al.* (2008) for theoretical supports and review of empirical evidences). And yet adaptive introgression may theoretically occur on shorter time, particularly in light of biological introductions processes (i.e. transport of species away from their native range via human activities, Elton 1958) which modify species distribution at a global scale and an unprecedented rate (e.g. for a review in marine ecosystem, Molnar *et al.* 2008). Biological introductions processes are particularly interesting cases studies to examine regarding the fate of secondary contacts between previously allopatric and non-reproductively isolated species. A diverse set of consequences of hybridization between native and non-native taxa are expected (Allendorf *et al.* 2001) for instance, the extinction of the native species (Rhymer & Simberloff 1996) or the introgression of advantageous alleles from the native into the non-native species facilitating local adaptation of the non-native species to its new colonized environment. Opposite situations were reported also in the literature, i.e. the rapid fixation of non-native alleles, called "*super invasive alleles*", in the genome of native species (Hohenlohe et al. 2013), for example between the non-native Barred Tiger salamanders and the native California one (Fitzpatrick *et al.* 2010).

In this context, we consider two newly described although strongly differentiated species in the genus *Ciona*. These two species were first considered as cryptic species of the *Ciona intestinalis* species complex and named *C. intestinalis* type A and *C. intestinalis* type B (Nydam & Harrison 2007; Zhan et al. 2010). Following recent taxonomic revision, they are now formally accepted as two distinct species (WoRMS database) and named *C. robusta* and *C. intestinalis*, respectively (Brunetti *et al.* 2015). They display a divergence estimated at ca. 4 Mya that has been followed by a secondary contact estimated to have occurred 15,500 years ago (95% CI: 4,300 - 56,800) and since this event ca. 20% of loci presumably crossed the species barriers in both direction (estimations based on full transcriptomes data, Roux *et al.* 2013). Currently, the two species, and particularly *C. robusta*, display a large distribution over several distinct biogeographic regions because both have been introduced due to human-activities (Procaccini

*et al.* 2011). For instance, *C. robusta*, assumed to be native to NW Pacific (Japan), is described in SE Pacific (Chile and Peru), NE Pacific (California), SE and SW Atlantic (S. Africa and Brazil), NE Atlantic (Europe). The two taxa live in sympatry in one region only, namely in the Northeast Atlantic (i.e. Western English Channel and South of Brittany) due to the recent introduction *C. robusta* (ca. in recent decades; Nydam & Harrison 2011; J.D.D. Bishop, personal observation) into the European native range of *C. intestinalis*. The two taxa are thus excellent models to examine as having being repeatedly into secondary contacts, in past- and present time.

Despite their ancient divergence, the two species are apparently not yet reproductively isolated and F1 hybrids are easily obtained under laboratory conditions (Suzuki *et al.* 2005; Bouchemousse *et al.* 2015). Based on a few nuclear markers, recent field and molecular studies carried out in the only sympatric range described so far (i.e. NE Atlantic) however cast doubts about the likelihood of contemporary hybridization: despite a close syntopy and reproductive synchrony, by using a few set of nuclear markers, only very rare F1-hybrids were observed in the wild (Nydam & Harrison 2011; Bouchemousse *et al.* 2015). In addition, only a limited amount of interspecific admixture was observed leading to the hypothesis that these admixture patterns may be the outcome of historical past-introgression events rather than contemporary backcross events. To test for this hypothesis and better examine the relative extent of past- and contemporary inter-specific gene flow between the two species, we used a population genomic approach based on ca. 300 SNPs derived from full transcriptomic sequences (Roux *et al.* 2013) by examining the genetic diversity of the two species sampled in eight localities of the sympatric range and two localities for each species outside contact zones (i.e. allopatric populations). This study is the first population SNPs-based study in an ascidian species despite the importance of this taxonomic group in evolutionary biology, development biology and phylogeny (Shenkar & Swalla 2011; Satoh *et al.* 2014), and particularly regarding the *C. intestinalis* species complex (Procaccini *et al.* 2011).


## Materials and Methods

### *Sampling*

We aimed at examining *Ciona robusta* and *C. intestinalis* within their contemporary sympatric range (i.e. Western English Channel and South of Brittany). Sampling was done in seven localities where the two species are living in syntopy (i.e. living in the exact same habitat) and one locality where surveys carried out over several years never reported the presence of *C. robusta* (i.e. only *C. intestinalis* is present; no.9 in Fig. 1 and Table 1; Bouchemousse *et al.* 2015). For comparison, populations from localities outside of the contemporary contact zone (i.e. where a unique species has been recorded so far) were sampled: for *C. robusta*, two localities of the SE Pacific and Mediterranean Sea, and for *C. intestinalis*, two localities in the North Sea (one in shallow water and one at 20-meters depth) and one in the NW Atlantic (Fig. 1, Table 1). For each individual, DNA extraction was performed with Nucleospin® 96 Tissue Kit according to the manufacturer's protocol (Macherey-Nagel, Germany). A minimum of 24 individuals per population was selected based on the DNA quality following extraction. Altogether a total of 449 individuals, 213 for *C. robusta* and 236 for *C. intestinalis* were further analyzed. Note that the assignment to *C. robusta* and *C. intestinalis* was based both on morphological features (Sato *et al.* 2012; Brunetti *et al.* 2015) and a maternal species-diagnostic mitochondrial locus (mtCOI; Nydam & Harrison 2007). In addition to specimens sampled in natural populations, two F1-hybrids

4

produced from experimental crosses (Bouchemousse *et al.* 2015) were included as control for F1-hybrid genotype.

## *Loci selection and genotyping*

An Illumina BeadXpress® with Veracode™ technology (GoldenGate® Genotyping Assay) was used to genotype 384 single nucleotide polymorphisms (SNPs) developed from full transcriptomes of 10 individuals of *C. robusta* and 10 individuals of *C. intestinalis* (details in Roux *et al.* (2013)). The selection of these 384 loci was optimized to maximize the success of genotyping (i.e. not closed to exon borders, few degenerate bases in the primer sequences, minor allele frequency $\leq 0.1$ at each SNP). Based on the results by Roux *et al.* (2013), these loci could be sorted according to four categories of polymorphism (Table 2): 1) SNPs differentially fixed between the two species (sf), 2) SNPs polymorphic in *C. robusta* (sxA) but not in *C. intestinalis*, 3) SNPs polymorphic in *C. intestinalis* (sxB) but not in *C. robusta* and 4) SNPs displaying polymorphism in the two species (ss). The full SNP panel is intentionally not random: the 384 SNPs were selected to be spread over most of the chromosomes of the published genome of *C. robusta* (Dehal *et al.* 2002; note that the genome is still indicated as *C. intestinalis* type A or even *C. intestinalis* genome according to the genome databases) and 25 of them were localized in introgression hotspots identified by Roux *et al.* (2013). We enriched the panel with sf (101) and ss (47) SNPs and equalized the number of sxA (109) and sxB (127) SNPs as *C. intestinalis* is more polymorphic than *C. robusta*. However, we have a subset of 70 SNPs that strictly reflect the genome wide site frequency spectrum. Genotyping was performed using Genome Studio software (Illumina Inc.). Out of the 384 SNPs, 324 SNPs amplified successfully and 310 SNPs were retained for further statistical analyses as they displayed a high rate of genotyping success (a minimum of 97% of the individuals without missing data) and an unambiguous genotype assignment. Details regarding the physical mapping of these loci are provided in Figure S1 and Table S1.

## *Intraspecific analysis*

In order to compare populations of the two species between sympatric and allopatric areas or syntopic and non-syntopic localities, genetic studies at the species level were carried out for both *C. robusta* and *C. intestinalis*. To picture the overall distribution of the genetic diversity in the dataset (i.e. 310 loci), a Principal Component Analysis (PCA) was carried out using the R package ADEGENET v.1.4 (Jombart & Ahmed 2011; R Development Core Team 2010). Then, only those loci that are relevant for intra-specific analyses (i.e. polymorphic at the species level, neutral, in linkage equilibrium) were selected. Firstly, loci that were differentially fixed between the two species (i.e. 65 loci sf), and thus monomorphic in both species, were removed. Secondly, linkage disequilibrium between loci was investigated as some of them being on the same contig and chromosome. Linkage disequilibrium between loci was tested, for each species separately, over the whole dataset and in each population sample using GENEPOP v.4.0.6 (Rousset 2008) with default parameters. To adjust *P*-values from multiple tests, Q-values were computed using QVALUE package implemented in R. Thirdly, being interested by analyzing gene flow and genetic drift, outlier loci, i.e. loci displaying atypical genetic differentiation as compared to those expected under neutral expectations, were removed from the analyses. The outliers were identified using a Bayesian method with a logistic regression model implemented in BAYESCAN v.2.1 software (Foll & Gaggiotti 2008) with default parameters

5

*Genetic diversity.* At the intra-specific level, for each population, the number of polymorphic loci and the expected heterozygosity ($H_e$) were estimated using GENETIX v.4.05 software (Belkhir *et al.* 2004). Fixation index ($F_{IS}$) was estimated and departures from Hardy-Weinberg equilibrium were tested in each population using GENEPOP with default parameters for tests. *P*-values resulting of multiple tests were adjusted using QVALUE package. In addition, to compare the genetic diversity between species, $H_e$ was estimated using all the loci available across the two species (i.e. 239 loci, including monomorphic loci in one or the other of the two species). Genetic diversity indices were compared by permutation tests using the R package *coin* (Hothorn *et al.* 2008).

*Genetic structure.* Genetic structure between populations was analyzed by estimating the fixation index $F_{ST}$ (Wright 1951) using GENEPOP. Exact G test for population differentiation were carried out using 10,000 random permutations. To visualize the genetic structure between populations, a Principal Component Analysis (PCA) was computed for each species separately. The genetic structure was also investigated using a Bayesian clustering method implemented in STRUCTURE v.2.3 (Pritchard *et al.* 2000). This method used MCMC to generate posterior probabilities of assignment of each individual genotype to a number of clusters (K). For each K (from 1 to the number of sampling sites, 9 for *C. robusta* and 11 for *C. intestinalis*), ten replicates of 150,000 MCMC after a period of 50,000 burn-in cycles were ran. We employed the admixture model with correlated allele frequencies with an *a priori* on the population origin of individuals to assist the clustering (LOCPRIOR model) following recommendations of Hubisz *et al.* (2009) for species with low level of population structure. The most likely number of clusters was determined using ΔK method established by Evanno *et al.* (2005) implemented in the online software STRUCTURE HARVESTER v.06.94 (Earl & vonHoldrt 2011). Results were summarized across all replicate runs using CLUMPP v.1.1.2 (Jakobsson & Rosenberg 2007) and visualized with DISTRUCT v1.1 (Rosenberg 2004).

## *Inter-specific and genome admixture analysis*

To examine inter-specific gene flow at the genome level, we selected 115 loci that were the most differentiated loci according to $F_{ST}$ values computed between the two species using allopatric populations (as reference for the absence of contemporary interspecific gene flow; Figure S2). For each individual, the maximum likelihood value of hybrid index was estimated using the R package INTROGRESS (Gompert & Buerkle 2010). The hybrid index ($h$) is defined as the proportion of *C. intestinalis* alleles over all loci ($h = 0$ for individuals with *C. robusta* alleles only and $h = 1$ for individuals with *C. intestinalis* alleles only). To visualize and compare the genomic architecture of interspecific admixture at the individual level, the *mk.image* function implemented in INTROGRESS was used.

To identify putative F1- or recently introgressed individuals (product of several generations of backcrosses within the sympatric range), a Bayesian clustering method implemented in NEWHYBRID software v.1.1 was used (Anderson & Thompson 2002). Briefly, this method computes posterior probability for assigning a given individual to different hybrid categories (i.e. F1, F2-hybrids and backcrosses with parental *C. robusta* or *C. intestinalis* individuals) or parental species, using Markov chain Monte Carlo algorithms (MCMC). Here, 150,000 MCMC after a period of 50,000 burn-in cycles was ran.

6

# Results

## *Diversity of the SNP panel*

Overall, 451 individuals (including the two F1-hybrids from experimental crosses) were genotyped successfully at 310 SNPs defined from a transcriptome dataset of *Ciona robusta* and *C. intestinalis* (Roux *et al.* 2013). Following this genotyping, the distribution of SNPs across the categories, defined from a small sample of 10 specimens for each species, was modified, as shown in Table 2. The most substantial change was a decrease of the sf and sxA categories (i.e. a decrease of 31% and 22%, respectively) and a concomitant increase of the sxB and ss categories (17% and 22%, respectively). We considered these new categories in the analyses below.

## *Differences in intra-specific genetic diversity and structure in the two study species*

None loci pairs showed linkage disequilibrium in *C. intestinalis* and only one in *C. robusta* (snp373 and snp185). Locus no.373 showing the lowest value of minor allele frequency was removed from further analyses. Five outliers (one in *C. robusta* and four in *C. intestinalis*) were identified with BAYESCAN analyses: one of them showed signature of positive selection in the allopatric SE Pacific population (i.e. snp43 Figure S3; BAYESCAN analyses) and four loci showed strong evidence of positive selection in *C. intestinalis*, mostly differentiating the two North Sea populations (Figure S4). These five loci were removed from further analyses. Altogether, 239 loci were retained: 115 and 166 polymorphic, independent and neutral loci, in *C. robusta* and *C. intestinalis*, respectively, including 42 polymorphic loci found in the two dataset.

Table 2 provides genetic diversity indexes for each population studied of *C. robusta* and *C. intestinalis*. Comparisons of values of expected heterozygosity (*He*), estimated from the 239 loci showing polymorphism in at least one of the two species, revealed that expected heterozygosity is significantly higher in populations of *C. intestinalis* than those of *C. robusta*, over all localities (test per permutation, *test.value* = -3.275, *P* < 0.001) as well as within the sympatric range of the two species (*test.value* = -3.694, *P* < 0.001).

### *Diversity and genetic structure in populations of* C. robusta

Values of *He* were similar across populations of *C. robusta*, ranging from 0.233 (no.2) to 0.294 (no.1). No departure from Hardy-Weinberg equilibrium (HWE) was found in any of the study populations. Exact test of differentiation revealed significant differences in allele frequencies among all populations sampled and among populations of the sympatric range (Table 3). A major part of the genetic structure was explained by a significant genetic differentiation of the SE Pacific population (no.1) with all of the other populations (pairwise comparisons provided in Table S2a). PCA (Figure 2a) well-illustrated this finding, with a differentiation along the first axis explained by the genetic difference between population no.1 and the others populations. Individual Bayesian assignments from STRUCTURE analysis (Fig. 3a) with two clusters (the most likely number of clusters according to ΔK value) also pointed out that individuals from population no.1 all belong to one given genetic cluster whereas all the other study individuals are assigned to a second cluster. When examining the genetic configuration using three clusters, the individuals not sampled in SE Pacific (no.1; assigned to cluster 1) were assigned to one of the two other genetic clusters (cluster 2; dark red in Fig 3a) but with more admixture in the two populations from UK (no.5 and 6; individuals membership to cluster 3 ranged

7

between 24% and 48%). Altogether, SE Pacific and to a lesser extent UK populations were the most different genetically.

*Diversity and genetic structure in populations of* C. intestinalis

Values of *He* were similar among the study populations, ranging from 0.213 (no.6) to 0.225 (no.10), except for the populations from the North Sea which exhibited lower values of *He* (i.e. 0.185 and 0.165 for no.4a and 4b respectively). As for *C. robusta*, no departure from HWE was observed. Exact test of differentiation between *C. intestinalis* populations indicated significant differences among all populations but not between populations of the sympatric range (Table 3). The overall significant genetic structure was mainly due to 1) a strong and significant genetic differentiation of the populations sampled in the two allopatric regions (no.3, 4a and 4b) and 2) a significant genetic difference of one population sampled in the sympatric range (no.6) with all other populations (pairwise comparisons are provided in Table S2b). These patterns are summarized by PCAs (Fig.2 b, c) showing the separation of the Swedish populations (no.4a and 4b) along the first axis, of the population no.4a and no. 6 along the second axis, and of the population no.3 along the third axis. As for *C. robusta*, the best number of clusters explaining the overall genetic variance was two (Fig. 3b) with one cluster characterizing the individuals from populations no.4a and no.4b (83% of the individuals with a membership probability higher than 50%), and one cluster characterizing the individuals sampled in the populations sampled in the sympatric range (no. 5 to 12), the North American population (no.3) and one individual from population no.4a. With additional clusters retained (for improving the resolution of the assignment), 100% of the individuals of the NW Atlantic (no.3) were assigned to a third cluster (with K = 3, Fig. 3b) and 100% of the individuals of the populations no.4a were assigned with a probability higher than 50% to a fourth cluster (with K = 4, Fig. 3b).

## *Low hybrid index disregarding the regional category and population status*

A total of 115 loci showing a $F_{ST}$ strictly superior to 0.8, were used to examine the patterns of shared polymorphism between the two species, and the likelihood of contemporary inter-specific gene flow vs. past introgression events. At the species level (i.e. across all individuals for each species), values of *h* were very low, with a value of 0.0066 for *C. robusta* and 0.0119 for *C. intestinalis*. *h* was however significantly higher for individuals of *C. intestinalis* than *C. robusta* (test per permutation, *test.value* = -7.522, $P < 0.001$). Table 1 is providing the estimates of the hybrid index (*h*) averaged across individuals, for each population of *C. robusta* and *C. intestinalis*.

At the population level, for *C. robusta*, *h*-values averaged across individuals ranged from 0.0023 for the allopatric population of Mediterranean Sea (Etang de Thau, no.2) to 0.0181 for the allopatric population of SE Pacific (Guanaqueros, no.1). Populations sampled in syntopic localities showed intermediate *h*-values, from 0.0047 (no.7) to 0.0067 (no.5). The difference of *h* was significant between populations from syntopic and non-syntopic localities (test per permutation on *h* value of individuals, *test.value* = 4.582, $P < 0.001$) but this was mainly due to the large value of *h* in the non-syntopic population no.1, which was significantly higher than the others populations (pairwise comparison in Table S3a). For *C. intestinalis*, the average of *h* across individuals per population ranged from 0.0047 for the allopatric population of North Sea (Fiskebackskil – 20m depth, no.4b) to 0.0348 for the population of the syntopic locality no.11 (Camaret). A noticeable result was the presence of one individual in this latter population with an *h* value of 0.493. This individual was assigned with a

8

probability of 1 to the 'F1 hybrid' category with NEWHYBRIDS. When removing this individual from the $h$ estimation, the value in Camaret dropped to 0.0122, a value closed from the average values for *C. intestinalis* individuals (i.e. 0.0119). It is noteworthy that all the other study individuals were assigned to their respective parental 'species' categories with NEWHYBRIDS (Table 1). Significant lower values of $h$ was found in populations of non-syntopic localities compared to populations of syntopic localities (test per permutation, *test.value* = -2.931, *P* = 0.006), a result only explained by the value of $h$ in population no.4b which was significantly lower than the others populations (test per permutation without no.4b, *test.value* = -1.837, *P* = 0.061, pairwise comparison in Table S3b). When analyzing syntopic localities, no significant correlation was observed for $h$ between the populations of the two species sampled in these localities (Spearman's rank correlation, $r^2$ = -0.595, *P* = 0.563). To further examine the existence of backcrossed individuals, the relationship between $h$-value and the heterozygosity rate across the 115 study loci was plotted using a triangle plot displayed in Figure 4: all except one individual displayed extreme $h$-values (closed to 0 or 1) and an extremely low proportion of heterozygote loci for *C. robusta* and *C. intestinalis* alleles. The only exception is the individual sampled from Camaret (no.11) that was assigned by NEWHYBRIDS as a F1-hybrid: he showed both a high $h$-value and a high heterozygosity rate (i.e. 96.5%); these values were similar to the values observed for the two F1-hybrids experimentally produced in the laboratory (Fig. 4). Note that using all the markers available (i.e. 310 loci), the overall genetic variance was clearly distributed between the two study species (as defined based on morphological and mtDNA characteristics) with the putative wild F1-individual and the two experimentally produced F1 at an intermediate position as well-exemplified by the results of a PCA shown in figure 5.

## *Heterogeneous admixture at the genome level for every population*

Over the 115 loci used for admixture (inter-specific gene flow) analyses, 65 loci (56.5%) were differentially fixed between the two species (sf), 11 loci (9.6%) exhibited both specific alleles in *C. robusta* specimens, 34 (29.6%) showed admixture in *C. intestinalis* specimens and the remaining 5 loci (4.3%) showed admixture in specimens of the two species.

The extent of admixture varied along the genome across loci and individuals (Fig. 6a). For example, over the 8 admixed loci studied in the chromosome number 1, admixture patterns was different between species with two loci showing admixture in *C. robusta* only and 5 in *C. intestinalis* only whereas the eighth one showed admixed genotypes in the two species. Interestingly this latter locus was located in one introgression hotspot identified by Roux *et al.* (2013). The same pattern was also observed on the chromosome number 2: among the 9 study loci on this chromosome, only one showed admixture in *C. robusta* specimens whereas the eight other showed admixture in *C. intestinalis* specimens only (six being located in an introgression hotspot).

Admixture patterns were also informative regarding the syntopic and non-syntopic status of the study populations. Details of allele frequencies at each locus in the allopatric and sympatric ranges of the two species are provided in Table S4. The loci showing similar patterns among populations categories are summarized with a Venn diagram in Figure 7. When comparing populations for *C. robusta*, the admixture profile appeared to be remarkably stable across populations although with a noticeable exception for the SE Pacific population (no.1 in Fig. 6a). The highest number of admixed loci (13 of the 16 loci showing admixture in *C. robusta*) was observed in the SE Pacific allopatric population. Out of the 4 loci showing shared polymorphism in the two species, three of them were also specific of

9

the SE Pacific population. The two allopatric populations (no.1 and 2) shared a limited number of admixed loci but interestingly populations from allopatric and sympatric ranges share a substantial number of admixed loci (7 over the two allopatric sites and all with the allopatric populations from the Mediterranean Sea). The lack of differences in the overall admixture pattern between the allopatric and sympatric regions also holds for *C. intestinalis*. Over all *C. intestinalis* specimens, 39 loci showed admixture with 16 found in both sympatric and allopatric regions. Two admixed loci were also shared by every population and two between the three allopatric populations (i.e. no.3, no.4a and 4b). Although the highest number of admixed loci (38 of the 39 loci) was observed in the sympatric region, a substantial number of admixed loci was found in allopatric populations (17): this lowest number is likely explained by a lower sampling coverage of the allopatric range. It is also noteworthy that *C. intestinalis* specimens sampled in Aber Wrac'h (no. 9), located in the sympatric range but not in syntopy with *C. robusta*, displayed admixture patterns similar to syntopic populations found in the same region, i.e. the English Channel (Fig. 6a). Altogether, the results observed thus pointed out heterogeneous levels of introgression along the genome with admixture observed in both syntopic and non-syntopic sites.

## Discussion

In this study, which is the first population genomics study performed in an ascidian species, we used 310 SNPs 1) to compare the genetic diversity and structure of the two inter-fertile *Ciona* species living in sympatry in the Northeastern Atlantic, 2) to clarify the previous hypothesis of contemporary hybridization *versus* past introgression between the two species (Nydam & Harrison 2011; Bouchemousse *et al.* 2015) and 3) to analyze the introgression patterns within allopatric and sympatric ranges of the two species. These three points are discussed in turn below.

*Contrasting nuclear genetic diversity between the native and the non-native species in their distribution range*

At a global scale and in the range of sympatry (i.e. English Channel and South of Brittany), based on the 239 shared loci (Table 3), *C. intestinalis* displayed a significant higher genetic diversity than *C. robusta* confirming the diversity pattern observed by Roux *et al.* (2013) using full-transcriptomic data and by Bouchemousse *et al.* (submitted) using mtDNA at a global scale.

In the sympatric range, a significant genetic variance among populations of *C. robusta* was observed (Table 4). This pattern, not explained by an isolation by distance between populations (Mantel test with GENEPOP on the web: $r^2 = -0.501$, $P = 0.537$), is likely due to repeated introductions of *C. robusta* specimens from multiples origins. This hypothesis is well supported by individual Bayesian assignments (Fig. 4a) with individuals from the N. Atlantic introduction (sympatric) range assigned to two different clusters (especially for individuals sampled in UK). Such intra-specific mixture or admixture patterns are commonly observed in non-indigenous marine species (e.g. in ascidians; Bock *et al.* 2012; Ordonez *et al.* 2013) due to repeated introductions from multiples and diverse origins because of poorly controlled vectors of introductions (Molnar *et al.* 2008). At a global scale, both the strong genetic differentiation (Table 4; Fig. 3a) and the higher genetic diversity observed in the population of SE Pacific (Table 3) suggest an independent and older introduction history as compared to the European populations, confirming previous insights from mitochondrial data (Bouchemousse *et al.* submitted). In the population of SE Pacific, we also observed one locus showing an atypical behavior in terms of genetic differentiation, i.e. an outlier (Fig. S3). Selection from standing genetic variation following

introductions has sometimes, although rarely, been documented in introduced species (Facon *et al.* 2006; Rius & Darling 2014). Such a process has also been suggested as a plausible explanation for outliers observed in the marine introduced oyster *Crassostrea gigas* in Europe, here an outcome of positive selection in fjord-like environment (Rohfritsch *et al.* 2013). Similar processes may have contributed to the emergence of local adaptation in the site sampled in Chile. However, with only one population sampled in this region and in the absence of population samples from the native range (i.e. NW Pacific) in our study, we cannot further test for a signature of positive selection following introduction.

For *C. intestinalis*, populations of the sympatric range exhibit a higher genetic diversity compared to *C. robusta* but they were not genetically different between them (Table 4) which suggests a high connectivity associated to high effective size of populations in this region. Recreational sailing which is an important activity in the study area may have promoted human-mediated transport of *C. intestinalis* (a fouling organism) for a long time, and thus gene flow over a regional scale. At a larger scale, in the native range (i.e. NE Atlantic), populations of North Sea were however strongly differentiated from populations of the English Channel (Table S2b); a genetic structure not be explained by isolation by distance between populations (Mantel test without no.3: $r^2 = 0.683\ 1$, $P = 0.002$). The population from the NW Atlantic (Nahant no. 3) exhibited a lower genetic differentiation with populations of the English Channel than with populations of North Sea, a pattern already observed with mtDNA (data not showed). North Sea populations also displayed a signature of positive selection at several loci (snp204 and snp70, Fig. S4). Our data also provided evidence of reduced gene flow between the two sub-localities sampled in North Sea, one in shallow water (no.4a) and one at 20-meters depth (no.4b). To our knowledge, no published data report contrasted genetic features between local shallow and deep populations of *C. intestinalis*. However, such pattern is reported in the doctoral thesis summary by Elin Renborg (https://gupea.ub.gu.se/ handle/ 2077/ 35128? locale=sv). It is also noteworthy that similar patterns have already been documented in other coastal marine species showing extended distribution along depth gradient (Jennings *et al.* 2013; Pivotto *et al.* 2015). Signature of positive selection found in population no.4b at snp70 may suggest local adaptation to deep environment (i.e. as a response to low light conditions or decreased temperature and thermal variation). Adaptations to local conditions have been already suggested in native range of emblematic marine invaders at global scale (e.g. thermal tolerance in *Carcinus maenas*; Tepolt & Palumbi 2015). However, we cannot exclude that the outlier locus is associated with genetic incompatibilities which restrict gene flow at the fraction of the genome wherein the locus is located (Bierne *et al.* 2011). This is for instance the most likely explanation of the outliers found in the mollusk *Crepidula fornicata* in its native range (Riquet *et al.* 2013). More populations from deep and shallow environment in different regions need to be studied, in addition to experimental approach, to further examine the hypothesis of local adaptation to depth in *C. intestinalis*.

Another noticeable result is the absence of departure from Hardy Weinberg Equilibrium (HWE) in all populations studied whatever the species. This finding differed from several published studies of *C. robusta* and *C. intestinalis* based on microsatellite loci. Heterozygote deficiencies have indeed been regularly reported (Caputi *et al.* 2007; Sordino *et al.* 2008; Zhan *et al.* 2010; Zhan *et al.* 2012; Affinito *et al.* 2014). This discrepancy can be explained by the choice of marker used (microsatellite *versus* SNPs); null alleles are a known pitfall of microsatellite loci (Dakin & Avise 2004; Chapuis & Estoup 2007). With the high divergence between the two *Ciona* species (i.e. 14% of substitutions per synonymous site based on transcriptomic data of the two species; Roux *et al.* 2013) and a substantial amount of past-introgression between the two species (Roux *et al.* 2013, this study), the presence of null alleles is likely to be important because of a low conservation of the microsatellite regions between the

11

two species. By using bi-allelic markers chosen to avoid genotyping failure, SNPs appear as more reliable markers to avoid bias due to null alleles (i.e. departure of HWE, reduction of genetic diversity within populations, estimation of population differentiation, see for a review Chapuis & Estoup (2007)) and thus to study genetic diversity and structure at once in the two *Ciona* species, as demonstrated for several others taxa (Liu *et al.* 2005; Haasl & Payseur 2011; Dufresne *et al.* 2014).

## *Absence of interspecific gene flow in the contemporary sympatric range*

In previous studies that examined interspecific gene flow in the sympatric range, very low admixture rates have been observed (i.e. 4.2% and one putative F1-hybrid over 730 individuals (Nydam & Harrison 2011); 6.3% over 288 individual sampled in the same locality (Sato *et al.* 2014); 4.3% and one putative F1 hybrid over ca. 3,000 individuals (Bouchemousse *et al.* 2015)), suggesting, according to these authors, a lack of effective reproduction between the two species in syntopic localities. However, these authors used only few (supposed to be) species-diagnostic nuclear markers (between 3 and 6 loci according to the study). Consequently, discriminating the footprint left by historical introgression or contemporary hybridization was particularly difficult (Nydam & Harrison 2011; Bouchemousse *et al.* 2015). Using 115 loci and over the 449 individuals studied, only one individual, found in the syntopic population of Camaret (no.11), had a multi-locus genotype expected for a F1-hybrid. This result was supported by both NEWHYBRIDS and INTROGRESS analyses (Table 1, Fig. 4 and 6a). The mtDNA type of this individual is typical of *C. intestinalis* and in many studies (Suzuki *et al.* 2005; Bouchemousse *et al.* 2015), F1 hybrids produced in laboratory experiments are obtained in one direction only, which is the one corresponding to crosses involving oocytes of *C. intestinalis* and sperm of *C. robusta* (ca. 80% of fertilization rate against < 6% in the opposite direction, Bouchemousse *et al.* 2015). The presence of one F1 hybrid only in our study confirms the hypothesis by Sato *et al.* (2014) and Bouchemousse *et al.* (2015) of the existence of strong pre zygotic isolation mechanisms preventing contemporary hybridization in the wild. It is noteworthy that the *h*-value is not equal to 0.5 and that the proportion of heterozygote loci is not 100%, as should have been theoretically expected in F1-hybrids. But as shown in Fig. 4, this was also the case for the two experimental F1 hybrids. This pattern is likely explained by the fact that some of the loci have been historically introgressed (see next section) and that the parents were either heterozygotes for *C. intestinalis* and *C. robusta* alleles or homozygotes for the alternative genomic background. We also observed non-null *h* values in non-syntopic populations (Table 1 and pairwise comparison in Table S3). In addition, NEWHYBRIDS and INTROGRESS analyses did not point out any individuals derived from recent interspecific gene flow (i.e. F2 hybrids and backcrosses with parental lineage). Altogether these results confirm that no contemporary gene flow is occurring between the two species. Presence of F1 hybrid only is compatible with recessive Dobzhansky-Muller incompatibilities (Turelli & Orr 2000) which are expressed by recessive mutations in subsequent generations of hybridization (e.g. Fishman & Willis 2001; Bierne *et al.* 2006; and for a review, Maheshwari & Barbash 2011). This hypothesis proposed by Roux *et al.* (2013) fits well with our data pointing out the absence of backcrossed individuals.

## *Admixture observed result of past introgression between the two species*

Hybrid index values were very low but never null whatever the region (sympatric or allopatric) and locality status (syntopic vs. non-syntopic, Table 1). Indeed, admixture is observed in all populations including in localities of allopatric regions (for *C. robusta*: SE Pacific and Mediterranean Sea, and for

*C. intestinalis*: North Sea and NW Atlantic, Fig. 4a). With one exception in SE Pacific (see below), the admixture profile is also remarkably stable across populations sampled.

Shared polymorphisms observed between two species may result from 1) incomplete lineage sorting of ancestral polymorphism, 2) past inter specific gene flow (past introgressive hybridization) or 3) contemporary hybridization events. In the case of the two *Ciona* species studies here, the third explanation can be reasonably excluded as discussed above. Concerning incomplete lineage sorting of ancestral polymorphism, it would have mean that the polymorphism observed nowadays would have been maintained randomly across loci after the allopatric divergence estimated to have occurred during the Pliocene (between 2.59 and 5.33 My; Roux *et al.* 2013). Considering the long time elapsed since the divergence, the probability of occurrence of the two ancestral alleles in both daughter species is likely to be extremely low under a neutral model (Pamilo & Nei 1988). However, high effective population sizes moderates the effect of genetic drift and so the probability of fixation of alleles over the time (Pamilo & Nei 1988; Maddison 1997). *Ciona* species and their common ancestor were characterized by high effective population sizes, estimated in Roux *et al.* (2013), as between 115,000 and 395,000 for *C. robusta,* 748,000 - 1,022,000 for *C. intestinalis* and 1,606,000 - 2,220,000 for the common ancestor. We thus cannot exclude that for the few loci showing moderates values of the minor alleles frequency and shared polymorphism in most populations (as snp150, Table S4), the admixture observed is a feature of ancestral polymorphism. However for the vast majority of the loci used as species-diagnostic (or sub-diagnostic) loci (i.e. loci with *Fst* value higher than 0.80), it is more likely that the observed admixture is the footprint of past introgressive hybridization that occurred during the secondary contact described by Roux *et al.* (2013) and estimated to have occurred 15,500 years ago (95% CI: 4,300 - 56,800). Following this secondary contact, Roux *et al.* (2013) estimated that ca. 20% crossed the species barrier in both directions. Besides similarities in admixture patterns across localities, the hypothesis of an ancient admixture event is also supported by the presence of admixed loci in introgression hotspots (i.e. loci pointed by an asterisk in Fig. 7a) in which are ca. 11% of the loci showing evidence of interspecific migration in the study by Roux *et al.* (2013). At a genome scale, variations of the extent of admixture across loci over 449 individuals confirm the genome-wide heterogeneity of introgression rate observed by Roux *et al.* (2013) but on a much smaller set of individuals.

Our finding is also interesting to consider in light of previous studies (e.g. Nydam & Harrison 2011; Sato *et al.* 2014; Bouchemousse *et al.* 2015) which used a small set of putative diagnostic markers to examine contemporary interspecific gene flow. Our results indeed cast doubts about the status of these markers and their ability to reliably distinguish the two species. To better investigate the properties of these markers, we genotyped most of the individuals of the present study (all except individuals of populations no. 5 for *C. robusta* and no. 3 for *C. intestinalis*) using three of these putative diagnostic loci; namely Hox5 (Caputi *et al.* 2007), vAChTP and CesA (Nydam & Harrison 2007, 2010). We used these three loci in a previous study (Bouchemousse *et al.* 2015). As expected from the results of this study, the three markers were not heterozygotes in the experimentally produced F1 hybrids, which should have been the case if they are species-diagnostic markers. Concerning the natural populations, patterns of admixture are displayed in Fig. 6b and showed inconsistent patterns as compared to expectations. For instance, CesA showed a homozygote genotype with two *C. robusta* alleles in the single F1 individual otherwise identified with the complete set of SNPs. They also revealed admixture in one allopatric locality (i.e. in SE Pacific on vAChTP). Altogether the three nuclear markers are not strictly speaking "species diagnostic markers" and may falsely identify backcrossed individuals. Note however that their introgression rate, using all individuals successful genotyped, is low: minor allele

13

frequency observed for Hox5, vAChTP and CesA is 0.2, 0.2 and 1.4% for *C. intestinalis*, and 7.2, 1.3 and 0.5% for *C. robusta*, respectively. This explains why in experimental crosses they often appear as diagnostic markers (S. Bouchemousse, G. Dubois and F. Viard, unpublished data).

The admixture pattern is also asymmetrical across loci: for Hox5, most of the admixture is observed in individuals that are morphologically and maternally (mtDNA) assigned to *C. robusta*. And yet, laboratory experiments showed that inter-specific crosses are mostly successful in the opposite direction only. It is thus likely to be a locus in which alleles were ancestrally transferred from *C. intestinalis* to *C. robusta*. The opposite pattern and introgression direction characterized the two other loci. These results highlight the risks of using putative species-diagnostic markers without preliminary knowledge about the likelihood of past introgression between two study taxa. The species complex of *Mytilus* species is another well-known case study: *Glu* and *mac-1* loci were mistakenly considered as diagnostic makers for *M. galloprovincialis* and *M. edulis* at a global scale (Borsa *et al.* 2007; Borsa *et al.* 2012), but were later shown to have been historically introgressed during secondary contact(s) caused by glacial oscillations (Roux *et al.* 2014).

*Interaction between range shift and human-mediated introductions*

Admixture profiles were remarkably stable across populations of allopatric and sympatric ranges (except for the population of SE Pacific). This widespread interspecific admixture suggest that range expansion of the two species, through both natural range shifts (with long-term environmental changes) and/or human-mediated introductions, occurred after a primary episode of secondary contact between the two taxa, during which interspecific gene flow occurred. One population however displayed a distinctive pattern: the single population sampled in the SE Pacific. This *C. robusta* population showed the highest number of loci with polymorphism shared with *C. intestinalis* and the highest *h*-values over all *C. robusta* populations. Shared polymorphism between the SE Pacific *C. robusta* population and populations of *C. intestinalis* could result from 1) incomplete lineage sorting of ancestral polymorphism, 2) molecular convergence or 3) adaptive introgression (Hedrick 2013). Considering the time elapse since the divergence (as explained above), the first hypothesis on incomplete lineage sorting of ancestral polymorphism is unlikely. Molecular convergence have been documented between closed or distant related taxa (e.g. O'HUigin *et al.* 2002; and for a review, Wood *et al.* 2005) but for *C. intestinalis* and *C. robusta* for which mutation rate has been estimated at $10^{-8}$ per base pairs per year (Tsagkogeorga *et al.* 2012), this hypothesis is highly unlikely considering the short time elapsed since their historical secondary contact (i.e. between 4,300 and 56,800 yrs.). Adaptive introgression following secondary contact, a process documented in several recent studies (Mendez *et al.* 2012; Pardo-Diaz *et al.* 2012; Fontaine *et al.* 2015), is thus the most likely hypothesis to consider. In the SE Pacific, several loci exhibited atypical higher frequency of *C. intestinalis* alleles compared to others populations (SNPs in bold in Table S4). None of these loci matches with genes coding for a known phenotypic or physiological trait (Table S4). The date for the first report of *C. robusta* along the Chilean coasts is 1885 (Traustedt (1885) cited in Castilla *et al.* (2005)). We thus cannot exclude that adaptive introgression have occurred in the source population(s) of the populations introduced in Chile rather than after the introduction (as an outcome of selection in the Chilean introduction range). Further investigations are needed, using for instance modelling methods such as those performed by Fraisse *et al.* (2014) to examine the likelihood of introgressive adaptation in a *Mytilus* sp. hybrid zone. A much larger number of populations representative of the global distribution of *C. robusta* is also needed to investigate the processes that occurred in the SE Pacific as compared to the other regions where *C. robusta* is nowadays distributed.

In conclusion, our study confirmed the almost complete absence of contemporary gene flow in the human-mediated contact zone wherein *C. robusta* and *C. intestinalis* co-exist in syntopy. Because efficient barriers prevent hybridization in the wild, the evolution of the non-native species through adaptive introgression, following the transfer of *C. intestinalis* alleles in its genome, is unlikely. Our study also provide evidence that genetic admixture observed is the outcome of historical interspecific gene flow redistributed at global scale by natural range shifts and human-mediated introductions. Atypical introgression patterns were observed in the population sampled in the SE Pacific. This result paves the way for further work to investigate introgressive adaptation processes in other regions, in light of the range shift history of *C. robusta*.

# Acknowledgment

# References

Abbott R, Albach D, Ansell S*, et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology* **26**, 229-246.

Affinito O, Andreakis N, Caputi L*, et al.* (2014) High connectivity and directional gene flow in European Atlantic and Mediterranean populations of *Ciona intestinalis* sp. A. *Marine Ecology*, 1-14, doi: 10.1111/maec.12226 **in press**

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution* **16**, 613-622.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**, 1217-1229.

Barton N (1979) Dynamics of hybrid zones. *Heredity* **43**, 341-359.

Belkhir K, Borsa P, Goudet J*, et al.* (2004) *Genetix* 4.05, logiciel sous WindowsTM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Bierne N, Bonhomme F, Boudry P, Szulkin M, David P (2006) Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis. Proceedings of the Royal Society B-Biological Sciences* **273**, 1253-1260.

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

Bock DG, MacIsaac HJ, Cristescu ME (2012) Multilocus genetic analyses differentiate between widespread and spatially restricted cryptic species in a model ascidian. *Proceedings of the Royal Society B-Biological Sciences* **279**, 2377-2385.

Borsa P, Daguin C, Bierne N (2007) Genomic reticulation indicates mixed ancestry in Southern-Hemisphere *Mytilus* spp. mussels. *Biological Journal of the Linnean Society* **92**, 747-754.

Borsa P, Rolland V, Daguin-Thiebaut C (2012) Genetics and taxonomy of Chilean smooth-shelled mussels, *Mytilus* spp. (Bivalvia: Mytilidae). *Comptes Rendus Biologies* **335**, 51-61.

Bouchemousse S, Lévêque L, Dubois G, Viard F (2015) Co-occurrence and reproductive synchrony do not ensure hybridization between an alien tunicate and its interfertile native congener. *Evolutionary Ecology*, 1-19, doi: 10.1007/s10682-015-9788-1 **in press**

Brunetti R, Gissi C, Pennati R*, et al.* (2015) Morphological evidence that the molecularly determined *Ciona intestinalis* type A and type B are different species: *Ciona robusta* and *Ciona intestinalis*. *Journal of Zoological Systematics and Evolutionary Research* **53**, 186-193.

Caputi L, Andreakis N, Mastrototaro F*, et al.* (2007) Cryptic speciation in a model invertebrate chordate. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9364-9369.

Castilla JC, Uribe M, Bahamonde N*, et al.* (2005) Down under the southeastern Pacific: marine non-indigenous species in Chile. *Biological Invasions* **7**, 213-232.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24**, 621-631.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution* **62**, 1908-1920.

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity* **93**, 504-509.

Dehal P, Satou Y, Campbell RK*, et al.* (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**, 2157-2167.

Dufresne F, Stift M, Vergilino R, Mable BK (2014) Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology* **23**, 40-69.

Earl D, vonHoldrt B (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.

Elton CS (1958) The ecology of invasions by animals and plants. Methuen publ. Reprinted 2000 by University of Chicago Press

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Facon B, Genton BJ, Shykoff J*, et al.* (2006) A general eco-evolutionary framework for understanding bioinvasions. *Trends in Ecology & Evolution* **21**, 130-135.

Fishman L, Willis JH (2001) Evidence for Dobzhansky-Muller incompatibilites contributing to the sterility of hybrids between *Mimulus guttatus* and *M.nasutus*. *Evolution* **55**, 1932-1942.

Fitzpatrick BM, Johnson JR, Kump DK*, et al.* (2010) Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 3606-3610.

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

Fontaine MC, Pease JB, Steele A*, et al.* (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **347**, 1258524.

Fraisse C, Belkhir K, Welch J, Bierne N (2015) Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Molecular Ecology,* 1-18, doi: 10.1111/mec.13299 **in press**

Fraisse C, Roux C, Welch JJ, Bierne N (2014) Gene-flow in a mosaic hybrid zone: is local introgression adaptive? *Genetics* **197**, 939-951.

Gompert Z, Buerkle CA (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* **10**, 378-384.

Haasl RJ, Payseur BA (2011) Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity* **106**, 158-171.

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity* **105**, 795-809.

Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology* **22**, 4606-4618.

Hewitt GM (1988) Hybrid zones: natural laboratories for evolutionary studies. *Trends in Ecology & Evolution* **3**, 158-167.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **359**, 183-195.

Hothorn T, Hornik K, van de Wiel MA, Zeileis A (2008) Implementing a class of permutation tests: The *coin* package. *Journal of Statistical Software* **28**, 1-23.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.

Jennings RM, Etter RJ, Ficarra L (2013) Population differentiation and species formation in the deep sea: the potential role of environmental gradients and depth. *Plos One* **8**, e77594.

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

Liu NJ, Chen L, Wang S, Oh CG, Zhao HY (2005) Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *Bmc Genetics* **6**, S26.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology* **46**, 523-536.

Maggs CA, Castilho R, Foltz D, *et al.* (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology* **89**, S108-S122.

Maheshwari S, Barbash DA (2011) The genetics of hybrid incompatibilities. In: *Annual Review Genetics,* Vol 45 (eds. Bassler BL, Lichten M, Schupbach G), pp. 331-355.

Mendez FL, Watkins JC, Hammer MF (2012) A haplotype at STAT2 introgressed from Neanderthals and serves as a candidate of positive selection in Papua New Guinea. *American Journal of Human Genetics* **91**, 265-274.

Molnar JL, Gamboa RL, Revenga C, Spalding MD (2008) Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment* **6**, 485-492.

Nydam ML, Harrison RG (2007) Genealogical relationships within and among shallow-water *Ciona* species (Ascidiacea). *Marine Biology* **151**, 1839-1847.

Nydam ML, Harrison RG (2010) Polymorphism and divergence within the ascidian genus *Ciona*. *Molecular Phylogenetics and Evolution* **56**, 718-726.

Nydam ML, Harrison RG (2011) Introgression despite substantial divergence in a broadcast spawing marine invertebrate. *Evolution* **65**, 429-442.

O'HUigin C, Satta Y, Takahata N, Klein J (2002) Contribution of homoplasy and of ancestral polymorphism to the evolution of genes in anthropoid primates. *Molecular Biology and Evolution* **19**, 1501-1513.

Ordonez V, Pascual M, Rius M, Turon X (2013) Mixed but not admixed: a spatial analysis of genetic variation of an invasive ascidian on natural and artificial substrates. *Marine Biology* **160**, 1645-1660.

Orr MR, Smith TB (1998) Ecology and speciation. *Trends in Ecology & Evolution* **13**, 502-506.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5**, 568-583.

Pardo-Diaz C, Salazar C, Baxter SW, *et al.* (2012) Adaptive introgression across species boundaries in *Heliconius* butterflies. *Plos Genetics* **8**, e1002752.

Pivotto ID, Nerini D, Masmoudi M, *et al.* (2015) Highly contrasted responses of Mediterranean octocorals to climate change along a depth gradient. *Royal Society open science* **2**, 140493-140493.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Procaccini G, Affinito O, Toscano F, Sordino P (2011) A new animal model for merging ecology and evolution. In: *Evolutionary biology: concepts, biodiversity, macroevolution and genome evolution* (ed. Pontarotti G), pp. 91-106, Springer Verlag, Berlin.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sanchez E (2015) Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics* **16**, 359-371.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* **27**, 83-109.

Riquet F, Daguin-Thiebaut C, Ballenghien M, Bierne N, Viard F (2013) Contrasting patterns of genome-wide polymorphism in the native and invasive range of the marine mollusc *Crepidula fornicata*. *Molecular Ecology* **22**, 1003-1018.

Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution* **29**, 233-242.

Rohfritsch A, Bierne N, Boudry P, *et al.* (2013) Population genomics shed light on the demographic and adaptive histories of European invasion in the Pacific oyster, *Crassostrea gigas*. *Evolutionary Applications* **6**, 1064-1078.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.

Rousset F (2008) Genepop ' 007: a complete re-implementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Roux C, Fraisse C, Castric V, *et al.* (2014) Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *Journal of Evolutionary Biology* **27**, 1662-1675.

Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution* **30**, 1574-1587.

Sato A, Satoh N, Bishop JDD (2012) Field identification of 'types' A and B of the ascidian *Ciona intestinalis* in a region of sympatry. *Marine Biology* **159**, 1611-1619.

Sato A, Shimeld SM, Bishop JDD (2014) Symmetrical reproductive compatibility of the two species in the *Ciona intestinalis* (Acidiacea) species complex, a model for marine genomics and developmental biology. *Zoological Science* **31**, 369-374.

Satoh N, Rokhsar D, Nishikawa T (2014) Chordate evolution and the three-phylum system. *Proceedings of the Royal Society B-Biological Sciences* **281**, 20141729

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

Shenkar N, Swalla BJ (2011) Global diversity of Ascidiacea. *Plos One* **6**, e20657

Sordino P, Andreakis N, Brown ER, *et al.* (2008) Natural variation of model mutant phenotypes in *Ciona intestinalis*. *Plos One* **3**, e2344

Suzuki MM, Nishikawa T, Bird A (2005) Genomic approaches reveal unexpected genetic divergence within *Ciona intestinalis*. *Journal of Molecular Evolution* **61**, 627-635.

Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *American Naturalist* **166**, 581-591.

Tepolt CK, Palumbi SR (2015) Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader. *Molecular Ecology* **24**, 4145-4158.

Traustedt MPA (1885) Ascidiae simplices fra det Stille Ocean Vidensk. In: *Meddr dansk naturh.*, pp. 1-160, Foren, Kjobenhavn.

Tsagkogeorga G, Cahais V, Galtier N (2012) The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. *Genome Biology and Evolution* **4**, 852-861.

Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology & Evolution* **16**, 330-343.

Turelli M, Orr HA (2000) Dominance, epistasis and the genetics of postzygotic isolation. *Genetics* **154**, 1663-1679.

Wood TE, Burke JM, Rieseberg LH (2005) Parallel genotypic adaptation: when evolution repeats itself. *Genetica* **123**, 157-170.

Wright S (1951) The genetical structure of populations. *Annuals of Augenics* **15**, 323-354.

Zhan A, Darling JA, Bock DG, *et al.* (2012) Complex genetic patterns in closely related colonizing invasive species. *Ecology and Evolution* **2**, 1331-1346.

Zhan A, Macisaac HJ, Cristescu ME (2010) Invasion genetics of the *Ciona intestinalis* species complex: from regional endemism to global homogeneity. *Molecular Ecology* **19**, 4678-4694.

Table 1. **Study localities, hybrid index (*h*) and number of hybrids *sensu lato*** (i.e. F1, F2 hybrids and backcrosses with parental species) **in each population of *Ciona robusta* and *C. intestinalis*.**

*Regional status* and *locality status* indicate if the two species have been reported to co-exist at a regional scale (allopatric vs. sympatric) or at the locality level (syntopic vs. non-syntopic). *h* values were averaged across individuals for each sampled localities. Analyses were done with 115 SNPs selected for interspecific gene flow analyses ($F_{ST} > 0.8$).

| N° | Locality | Region | Introduced vs. native status | Regional status | Locality status | Sampling year | Nind | Hybrid index (mean ± SD) | Number of hybrids |
|---|---|---|---|---|---|---|---|---|---|
| *C. robusta* | | | | | | | | | |
| 1- | Guanaqueros, Chile | South Eastern Pacific | Introduced | Allopatric | Non-syntopic | 2012 | 24 | 0.0181 ± 0.0081 | 0 |
| 2- | Etang de Thau | Mediterranean Sea | Introduced | Allopatric | Non-syntopic | 2013 | 23 | 0.0023 ± 0.0037 | 0 |
| 5- | Falmouth | English Channel | Introduced | Sympatric | Syntopic | 2013 | 24 | 0.0067 ± 0.0044 | 0 |
| 6- | Plymouth | English Channel | Introduced | Sympatric | Syntopic | 2011 | 24 | 0.0054 ± 0.0047 | 0 |
| 7- | St Vaast | English Channel | Introduced | Sympatric | Syntopic | 2012 | 23 | 0.0047 ± 0.0037 | 0 |
| 8- | Perros Guirec | English Channel | Introduced | Sympatric | Syntopic | 2011 | 24 | 0.0062 ± 0.0046 | 0 |
| 10- | Moulin blanc | Bay of Brest | Introduced | Sympatric | Syntopic | 2012 | 24 | 0.0051 ± 0.0044 | 0 |
| 11- | Camaret | Bay of Brest | Introduced | Sympatric | Syntopic | 2011, 2012 | 24 | 0.0053 ± 0.0041 | 0 |
| 12- | Quiberon | Bay of Biscay | Introduced | Sympatric | Syntopic | 2012, 2013 | 23 | 0.0057 ± 0.0036 | 0 |
| | **Total** | | | | | | **213** | **0.0066 ± 0.0063** | |
| | | | | | | | | | |
| *C. intestinalis* | | | | | | | | | |
| 3- | Nahant | North Western Atlantic | Cryptogenic | Allopatric | Non-syntopic | 2013 | 24 | 0.0114 ± 0.0078 | 0 |
| 4a- | Fiskebackskil – surface | North Sea | Native | Allopatric | Non-syntopic | 2010 | 12 | 0.0069 ± 0.0051 | 0 |
| 4b- | Fiskebackskil - 20m depth | | | | | 2010 | 12 | 0.0047 ± 0.0034 | 0 |
| 5- | Falmouth | English Channel | Native | Sympatric | Syntopic | 2011 | 24 | 0.0107 ± 0.0056 | 0 |
| 6- | Plymouth | English Channel | Native | Sympatric | Syntopic | 2011 | 24 | 0.0103 ± 0.0067 | 0 |
| 7- | St Vaast | English Channel | Native | Sympatric | Syntopic | 2012 | 23 | 0.0127 ± 0.0079 | 0 |
| 8- | Perros Guirec | English Channel | Native | Sympatric | Syntopic | 2011, 2012 | 24 | 0.0140 ± 0.0068 | 0 |
| 9- | Aber Wrac'h | English Channel | Native | Sympatric | Non-syntopic | 2011, 2012 | 23 | 0.0121 ± 0.0064 | 0 |
| 10- | Moulin blanc | Bay of Brest | Native | Sympatric | Syntopic | 2011 | 24 | 0.0158 ± 0.0082 | 0 |
| 11- | Camaret | Bay of Brest | Native | Sympatric | Syntopic | 2011, 2012 | 22 | 0.0348 ± 0.1060 | **1 (F1-hybrid)** |
| | *without F1 -hybrid* | | | | | | 21 | 0.0122 ± 0.0064 | *0* |
| 12- | Quiberon | Bay of Biscay | Native | Sympatric | Syntopic | 2011, 2012 | 24 | 0.0140 ± 0.0090 | 0 |
| | **Total** | | | | | | **236** | **0.0140 ± 0.0331** | **1 (F1-hybrid)** |
| | *Total (without F1-hybrid)* | | | | | | *235* | *0.0119 ± 0.0074* | *0* |

Table 2. **Number of SNPs per categories defined according to their polymorphism patterns:** 1) sf: loci differentially fixed between the two species; 2) sxA: loci polymorphic in *Ciona robusta* only; 3) sxB: loci polymorphic in *C. intestinalis* only; 4) ss: polymorphism shared by the two species. The number of SNPs in each category is given according to the sorting made using the dataset from which the loci were designed (Roux *et al.* 2013) and according to the results of the present study.

| | | Distribution based on the results of the present study | | | | | Not genotyped or polymorphic | Total |
|---|---|---|---|---|---|---|---|---|
| | | sf | sxA | sxB | ss | **Total** | | |
| Distribution based on data from Roux *et al.* (2013) | sf | 64 | 3 | 26 | 1 | **94** | 7 | **101** |
| | sxA | 0 | 71 | 1 | 17 | **89** | 20 | **109** |
| | sxB | 0 | 0 | 99 | 3 | **102** | 25 | **127** |
| | ss | 1 | 1 | 2 | 21 | **25** | 22 | **47** |
| | **Total** | **65** | **75** | **128** | **42** | **310** | **74** | **384** |

Table 3. **Genetic diversity indices and fixation index** of each study populations for *Ciona robusta* and *C. intestinalis*.

| N° | Locality | Introduced vs. native status | $P_{loc}$ | $H_e$ (mean ± SD) | $H_e$ (mean ± SD) −239 shared loci | $F_{IS}$ |
|---|---|---|---|---|---|---|
| **C. robusta** | | | | | | |
| 1- | Guanaqueros | Introduced | 107 | 0.294 ± 0.177 | 0.142 ± 0.192 | -0.014 |
| 2- | Etang de Thau | Introduced | 77 | 0.233 ± 0.207 | 0.112 ± 0.185 | -0.013 |
| 5- | Falmouth | Introduced | 80 | 0.247 ± 0.212 | 0.119 ± 0.192 | -0.052 |
| 6- | Plymouth | Introduced | 80 | 0.239 ± 0.205 | 0.115 ± 0.186 | -0.048 |
| 7- | St Vaast | Introduced | 80 | 0.239 ± 0.202 | 0.115 ± 0.185 | -0.078 |
| 8- | Perros Guirec | Introduced | 81 | 0.245 ± 0.205 | 0.118 ± 0.189 | -0.024 |
| 10- | Moulin blanc | Introduced | 80 | 0.236 ± 0.202 | 0.114 ± 0.183 | -0.042 |
| 11- | Camaret | Introduced | 82 | 0.245 ± 0.205 | 0.118 ± 0.188 | -0.022 |
| 12- | Quiberon | Introduced | 83 | 0.253 ± 0.206 | 0.122 ± 0.191 | -0.026 |
| | *Total (Sympatric pop.)* | | *87* | *0.248 ± 0.200* | *0.119 ± 0.186* | |
| | **Total** | | **115** | **0.259 ± 0.189** | **0.125 ± 0.185** | |
| | | | | | | |
| **C. intestinalis** | | | | | | |
| 3- | Nahant | Cryptogenic | 116 | 0.215 ± 0.191 | 0.147 ± 0.189 | -0.016 |
| 4a- | Fiskebackskil - surface | Native | 99 | 0.185 ± 0.194 | 0.126 ± 0.182 | 0.023 |
| 4b- | Fiskebackskil - 20m depth | | 83 | 0.165 ± 0.196 | 0.112 ± 0.178 | -0.007 |
| 5- | Falmouth | Native | 122 | 0.216 ± 0.191 | 0.148 ± 0.187 | -0.029 |
| 6- | Plymouth | Native | 131 | 0.213 ± 0.187 | 0.146 ± 0.184 | -0.009 |
| 7- | St Vaast | Native | 131 | 0.217 ± 0.186 | 0.149 ± 0.185 | -0.023 |
| 8- | Perros Guirec | Native | 128 | 0.220 ± 0.193 | 0.151 ± 0.190 | -0.047 |
| 9- | Aber Wrac'h | Native | 126 | 0.223 ± 0.191 | 0.153 ± 0.189 | -0.021 |
| 10- | Moulin blanc | Native | 130 | 0.225 ± 0.188 | 0.155 ± 0.187 | 0.009 |
| 11- | Camaret | Native | 127 | 0.220 ± 0.189 | 0.151 ± 0.187 | -0.020 |
| 12- | Quiberon | Native | 125 | 0.224 ± 0.192 | 0.153 ± 0.190 | -0.019 |
| | *Total (Sympatric pop.)* | | *166* | *0.223 ± 0.182* | *0.152 ± 0.183* | |
| | **Total** | | **166** | **0.224 ± 0.181** | **0.152 ± 0.182** | |

$P_{loc}$: number of polymorphic loci and $H_e$: expected heterozygosity over 115 and 166 loci retained for intra-specific analyses in *C. robusta* and *C. intestinalis* respectively; $H_e$-239 loci: expected heterozygosity over the 239 loci shared by both species (115 and 166 polymorphic, independent and neutral loci in *C. robusta* and *C. intestinalis*, respectively including 42 polymorphic loci found in the two dataset), $F_{IS}$: fixation index calculated (no deviation from Hardy-Weinberg equilibrium; exact test, $P < 0.05$)

Table 4. **Genetic structure among populations for *Ciona robusta* and *C. intestinalis*.**

|  | $F_{ST}$ | *P*-value |
|---|---|---|
| ***C. robusta*** | | |
| All sampled populations (9 populations) | 0.049 | *P* < 0.001 |
| All populations without Guanaqueros (all except no.1) | 0.023 | *P* < 0.001 |
| Sympatric populations (all except no.1 and 2) | 0.021 | *P* < 0.001 |
| | | |
| ***C. intestinalis*** | | |
| All sampled populations (11 populations) | 0.039 | *P* < 0.001 |
| All populations without Fiskebackskil (all except no.4a and 4b) | 0.017 | *P* < 0.001 |
| Sympatric populations (all except no.3, 4a and 4b) | 0.011 | *P* = 0.273 |
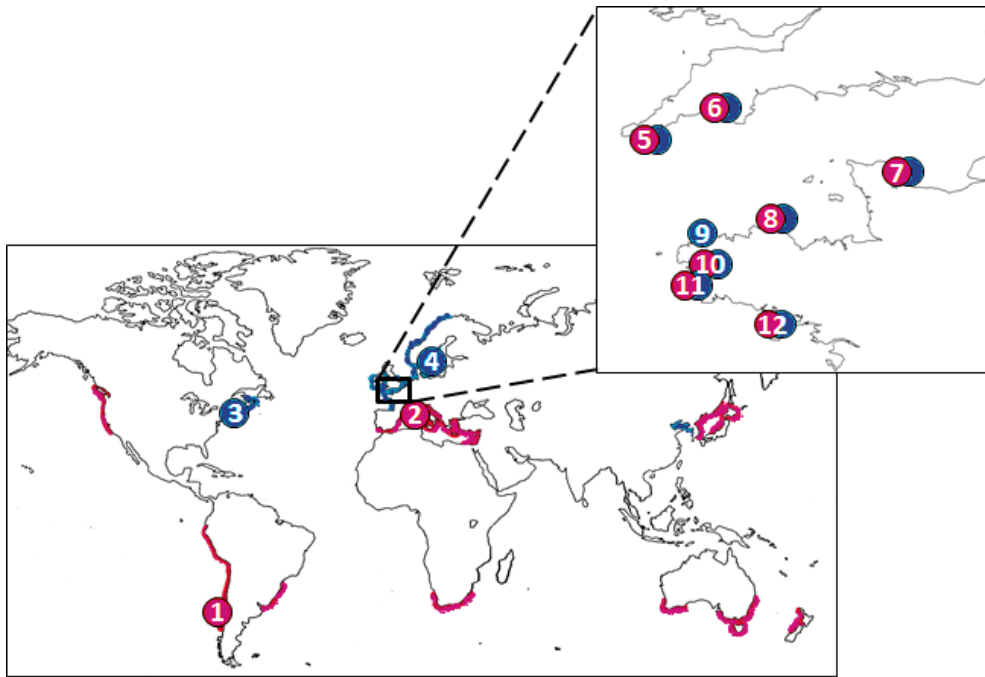
Figure 1. **Distribution** (pink line for *C. robusta* and blue line for *C. intestinalis*) **and study locations** (red points for *C. robusta* and blue points for *C. intestinalis*). See Table 1 for population labels.
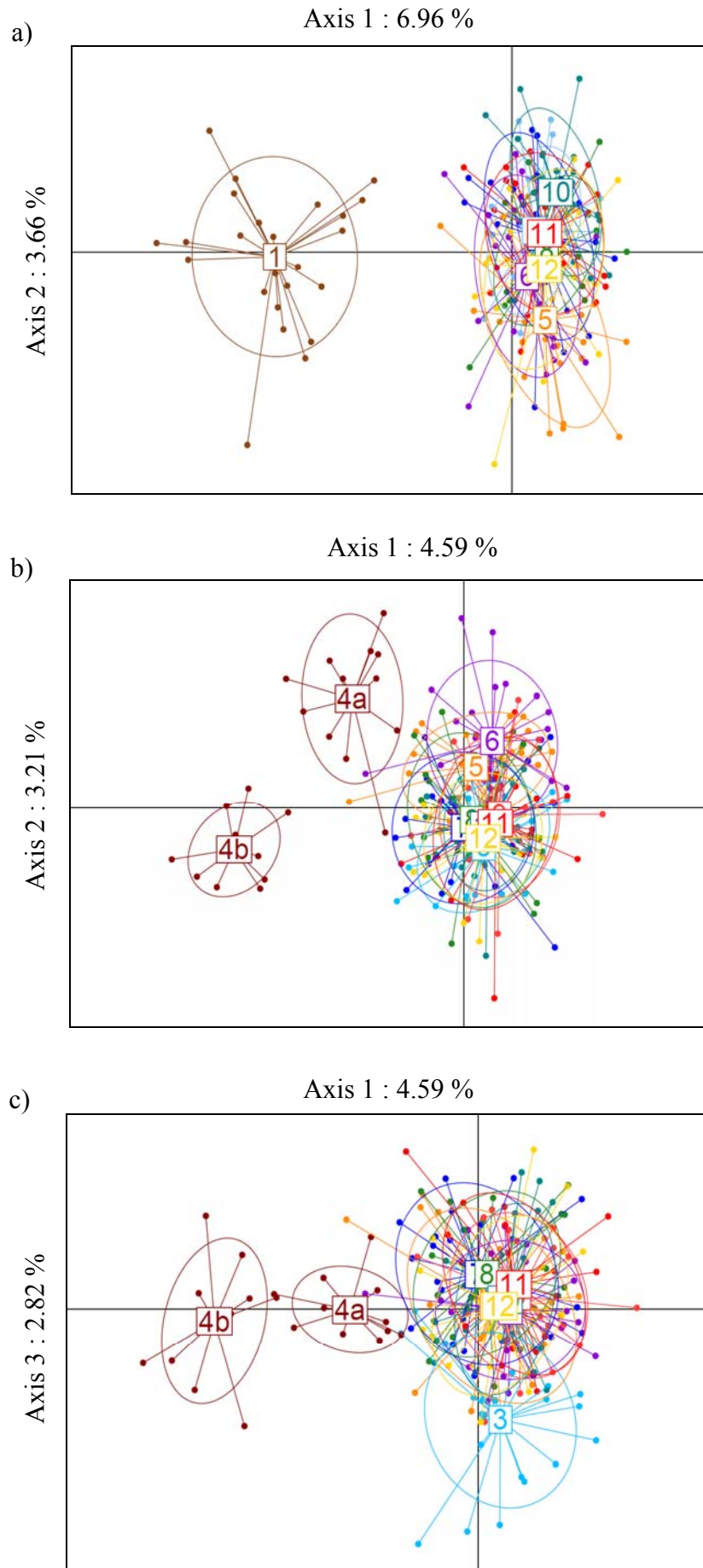
Figure 2. **Principal Components Analysis** among populations of *C. robusta* (a) and *C. intestinalis* (b, c).
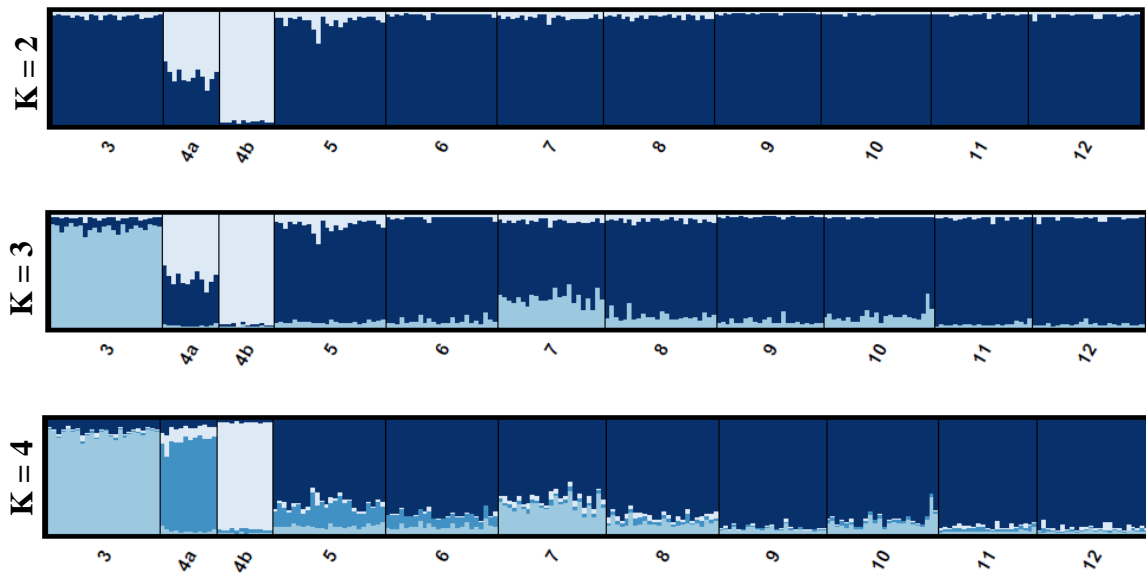
a) *C. robusta*
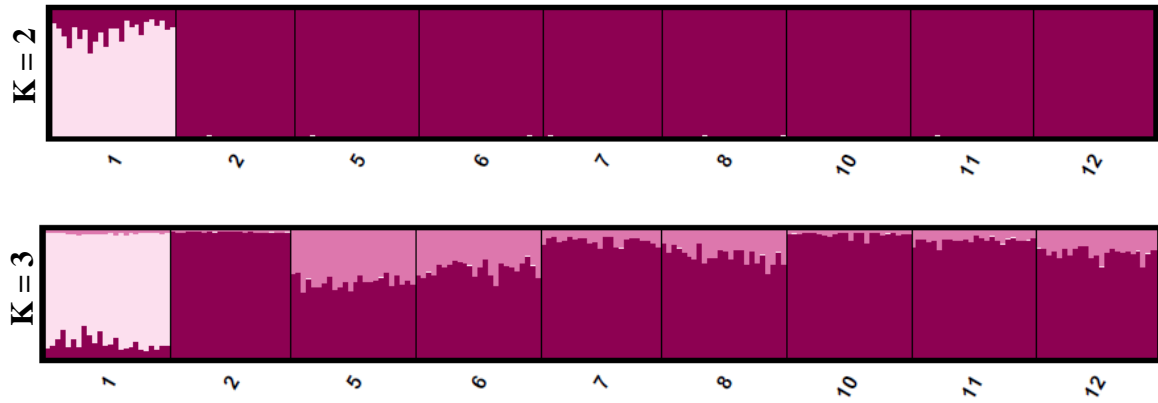


b) *C. intestinalis*



Figure 3. **Individual Bayesian assignment proportion** for *C. robusta* (a) and *C. intestinalis* (b) individuals using 115 and 166 polymorphic loci, respectively. Note that for both species, the most likely number of clusters K according to ΔK method of Evanno *et al.* (2005) was K = 2. Additional K highlighted the intraspecific clustering; to K = 3 for *C. robusta* and to K = 4 for *C. intestinalis*.
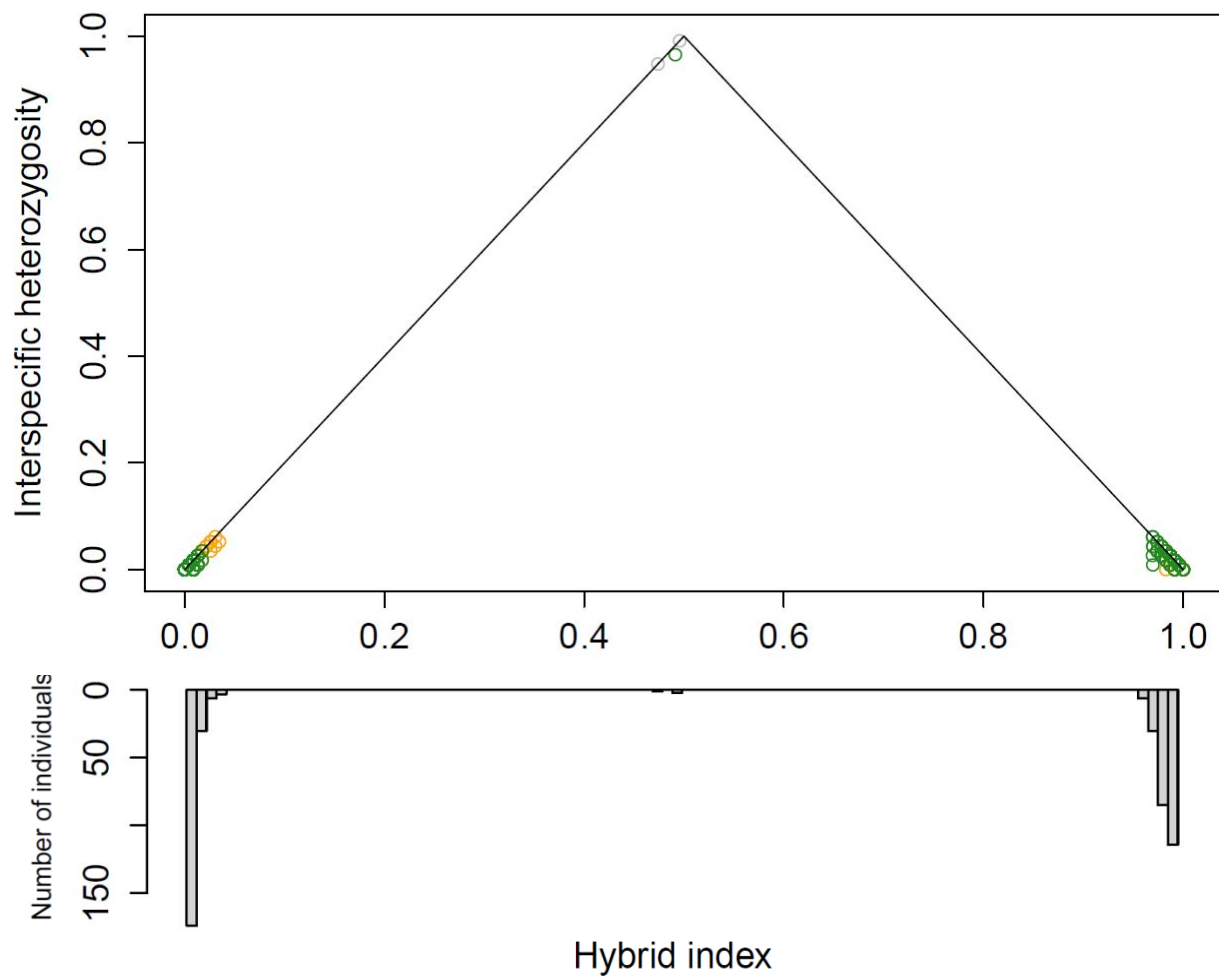
Figure 4. **Triangle plot showing the relationship between heterozygosity rate across loci and hybrid index for each individual**. Green and orange circles are individuals from syntopic and non-syntopic localities, respectively. At the top of the triangle, one green circle is picturing one individuals from the locality no. 11 and the two gray circles are F1-hybrids from experimental crosses.
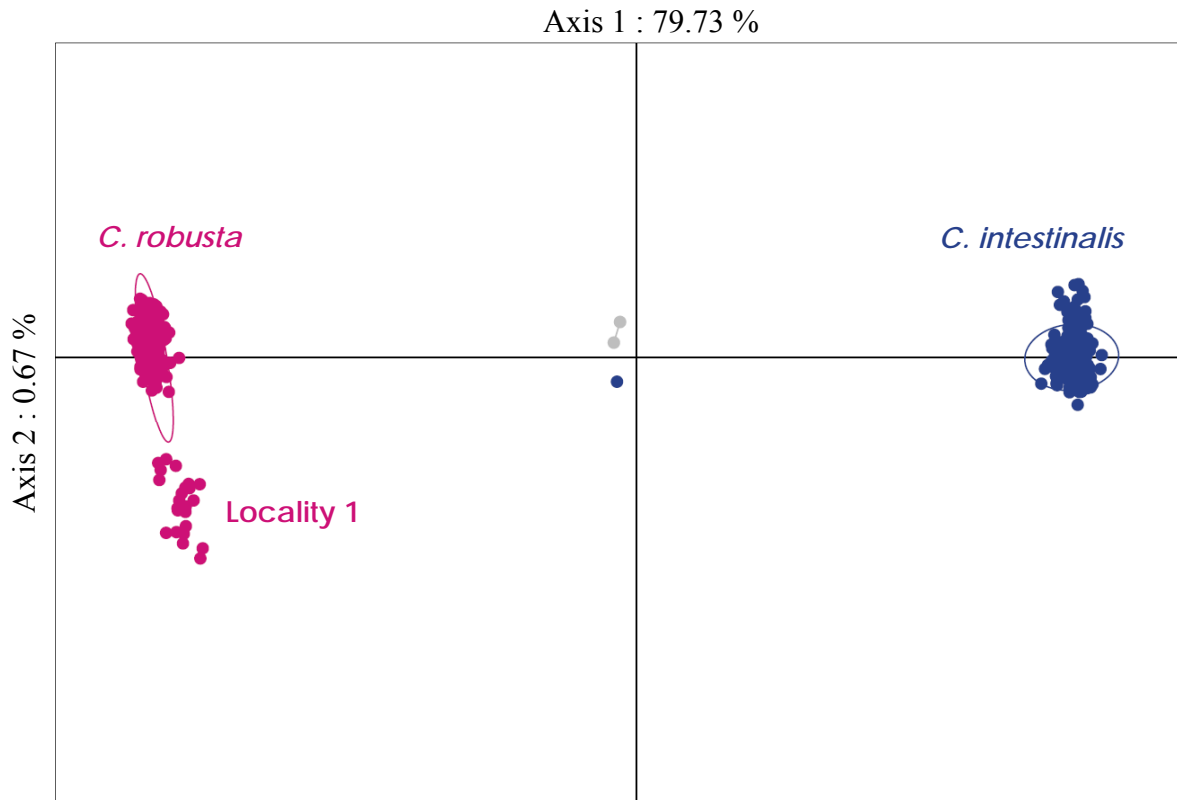
Axis 1 : 79.73 %



Figure 5. **Principal Components Analysis** using the all dataset (i.e. 310 SNPs, 449 individuals from natural population and in grey two F1 hybrids from experimental crosses).
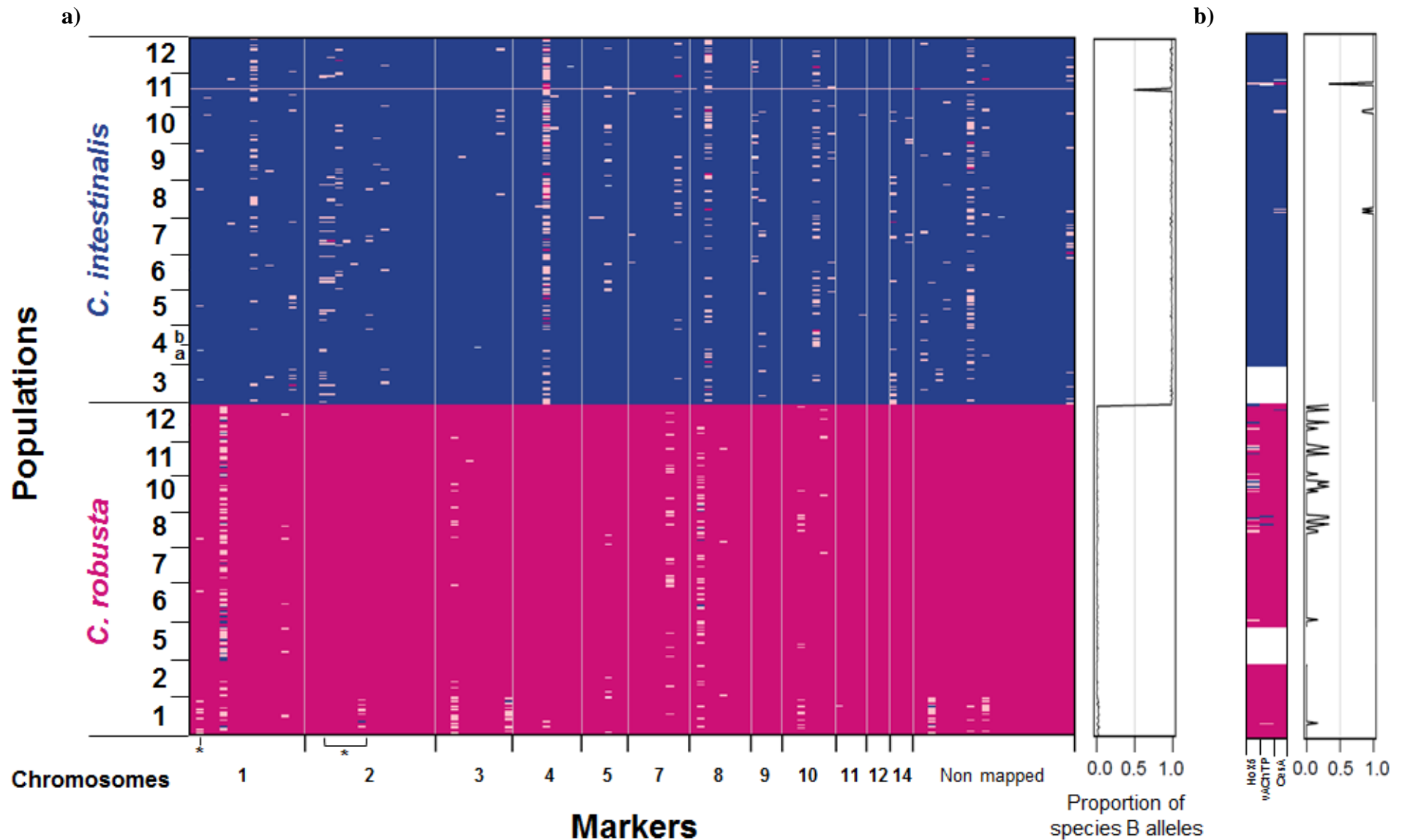
Figure 6. **a) Genomic architecture using 115 loci selected for inter-specific analyses**. Markers (x-axis) are ordered following physical position on chromosomes. Individuals (y-axis) are ordered per population. Dark pink cases indicate homozygote genotype on *C. robusta* alleles; dark blue, homozygote genotype on *C. intestinalis* alleles; light pink, heterozygotes for *C. robusta* and *C. intestinalis* alleles. Asterisks indicate loci located in introgression hotspots defined by Roux *et al.* (2013). **b) Pattern of admixture for 3 nuclear loci** (Hox5, vAChTP, CesA) **analyzed by PCR and PCR-RFLP**, already used in previous studies (Nydam & Harrison 2011; Sato *et al.* 2014; Bouchemousse *et al.* 2015). Note that for two populations (no. 5 for *C. robusta* and no.3 for *C. intestinalis*), genotypic data for these three loci are missing.
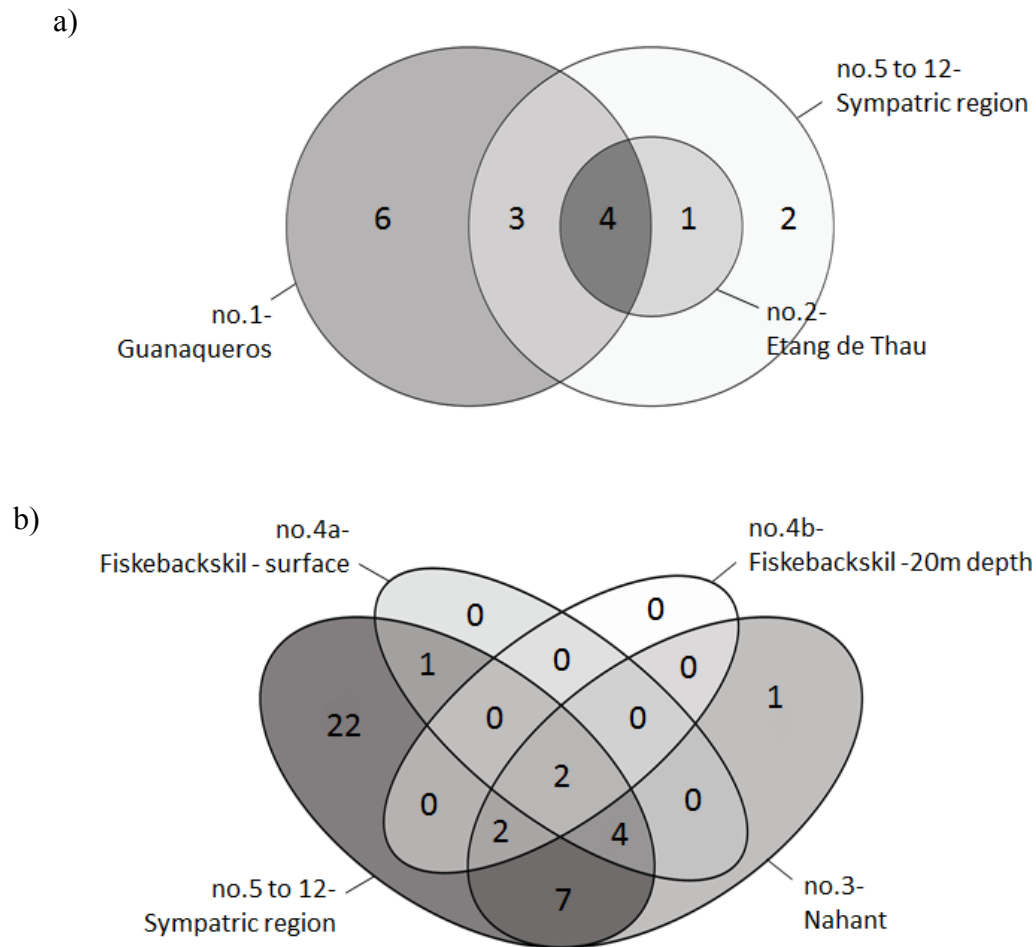
a)



b)



Figure 7. **Venn diagram showing the number and distribution of loci with shared polymorphism in allopatric vs. sympatric populations** over the 115 SNPs selected for interspecific gene flow analyses, for *Ciona robusta* (a) and *C. intestinalis* (b) individuals.